The Validity of Instruments Revisited

Daniel Berkowitz * Mehmet Caner**
University of Pittsburgh North Carolina State University
Ying Fang***
Xiamen University

October 18, 2010

Abstract

Valid instrumental variables must be relevant and exogenous. However, in practice it is difficult to find instruments that perfectly satisfy the orthogonality condition and at the same time are strongly correlated with the endogenous regressors. In this paper we show how a mild violation of the exogeneity assumption affects the limit of the Anderson-Rubin (1949) test. The Anderson-Rubin (AR) test statistic is frequently used because it is robust to identification problems. However, when there is a mild violation of exogeneity the AR test is oversized and with larger samples the problem gets worse. In order to correct this problem, we introduce the fractionally resampled Anderson-Rubin (FAR) test that is derived by modifying the resampling technique of Wu (1990). We select half of the sample for resampling and obtain valid but slightly conservative critical values. The simulations show that our technique does perform well even with moderate to large violations of exogeneity with a finite sample correction for the block size choice.

Keywords: Berry-Esseen Bound, Finite Sample of Random Variables, Near-Exogeneity
JEL Code: C3, C13, C30.

*Daniel Berkowitz: Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260, email: dmberk@pitt.edu. ** Mehmet Caner, Corresponding Author: Department of Economics, 4168 Nelson Hall, Raleigh, NC 27695. email: mcaner@ncsu.edu. *** Ying Fang: WISE, Xiamen University, China email: yifst1@gmail.com. We thank the co-editor Cheng Hsiao, an associate editor, and an anonymous referee for their comments. We are grateful to Patrik Guggenberger for pointing out an error in an earlier version of the paper. We also thank C.R. Rao and C.F.J. Wu for comments.
1 Introduction

Instrumental variable estimation is one of the most widely used methods in economics. Valid instruments must be relevant and exogenous. Regarding the relevance of instruments, there has recently been a growing interest in the asymptotics of weak instruments. One of the most widely used test statistics in that research line is the Anderson-Rubin (1949) test (for herein denoted the AR test). The AR test statistic can be used when instruments are weak as shown by Stock and Wright (2000). Regarding exogeneity, however, there is a growing consensus that even when researchers carefully pick instruments that are plausibly exogenous, it is still unlikely that an instrument perfectly satisfies the orthogonality condition. Nevertheless, applied researchers lack formal methods for drawing inferences when their instruments are "nearly exogenous".

Existing methods enable researchers to make adjustments for the consistency of their estimates, but do not enable researchers to adjust their inferences. For example, in an influential paper, Acemoglu et al (2008) use instruments to identify the impact of democracy on national income. Acemoglu et al (2008) note that their instrumental variable estimates identify an upper bound because there is a small positive correlation between their instrument and the structural error term. However, Acemoglu et al cannot say how this small violation of orthogonality affects inference. This is because standard instrumental variable methods that use t-statistics do not adjust the standard error of the endogenous regressor to account for correlations between the instrument and structural error term.

We introduce a method that enables applied researchers to make reliable inferences in a world where instruments are "nearly exogenous" and not necessarily "perfectly exogenous." Our test accounts for the strength of the violation of the orthogonality condition. In the weak instruments literature, it has been shown that inferring causality becomes more difficult as the instrument becomes weaker. We find that identifying causal effects becomes more difficult as the endogeneity of the instrument increases. Our method enables researchers to make inferences that account for both endogeneity and weakness.

We employ a violation of the exogeneity assumption that allows for a local to zero correlation between the instruments and the structural error. Clearly, our method assumes that instruments exhibit asymptotic orthogonality. However, we also show that our method holds in finite samples even when the correlation between the error and the instrument is constant and nonzero.

Our test is a variant of the Anderson-Rubin (1949) test. We derive the limit of the AR test and show that the limit depends on the correlation between the instruments and structural error term. We also derive the limit in pointwise as well as the uniform case over correlations. These are different from each other. Furthermore, in larger samples, using
critical values for the AR test based on the perfect exogeneity assumption creates massive size distortions when there is, in reality, a mild correlation between the instruments and structural error term.

To deal with these size distortions we propose a novel resampling technique for the AR test. This technique is based on the jackknife histogram estimator in section 2 of Wu (1990). Because the AR test can be written in terms of the sample mean (see equation (4)), we can modify the results in section 2 of Wu (1990). Section 2 in Wu (1990) uses sampling without replacement from the original sample by drawing a fraction of the sample size that is proportional to the full sample. In this setup, we propose a fractionally resampled version of the AR test. We show that choosing half the sample as the block size provides valid but conservative critical values. As a competing technique we show that subsampling is oversized. We also conjecture that the Kleibergen (2002) test may not be amenable to the resampling technique that we use. Simulations are conducted to check for the size properties and power of this FAR test. We find that a simple finite sample correction for the block size provides very good power.

Guggenberger (2009) analyzes several tests under local violation of exogeneity, and shows all of them have size problems. However, he finds that the Anderson-Rubin type of test has the best size properties. In related work, Kraay (2009) and Conley, Hansen and Rossi (2007) both use a Bayesian approach for solving the problem of working with instruments that do not perfectly satisfy the orthogonality condition. They clearly show that even a small violation of the orthogonality condition can lead to entirely different outcomes. When they allow there to be mild violation of exogeneity, they find that the confidence intervals for structural parameters are larger. Conley, Hansen and Rossi (2007) analyze the support of the correlation parameter (i.e., the correlation between the instrument and structural error) and, for each plausible parameter value, they find the confidence interval for the structural parameters and take the union of these intervals. This method provides a conservative solution. Conley, Hansen and Rossi (2007) also use a local to zero approach: here, they assume the correlation parameter comes from a normal distribution and they characterize its asymptotics. In a third approach, Conley et al, attach Bayesian priors to this parameter and derive the posterior distribution. Kraay (2009) takes a similar approach to this problem; however, his prior for the correlation parameter is not drawn from a normal distribution. In contrast to these methods that place priors on the correlation between the instrument and structural error term, our method is completely data dependent.

Section 2 describes the problem of making inferences with instruments that violate the exogeneity assumption and we develop a novel way of resampling the AR test. Section 3 considers subsampling and then shows that it will be oversized and will not solve the problem of drawing reliable inferences with instruments violating the exogeneity assumption. Section
3 also contains an analysis of some of the variants of subsampling. Section 4 contains Monte Carlo simulations. Section 5 concludes.

2 Inference and Violations of Exogeneity

We analyze a model that contains a specific violation of the exogeneity assumption. Similar assumptions about the violation of exogeneity have been used by Newey (1985) and Hall and Inoue (2003). The assumption in our model allows for a local to zero covariance between the instruments and the structural error term, and is more flexible than the knife-edged exogeneity assumption used in the instrumental variables estimation literature. The model that we use is:

\[ y = Y\theta_0 + u, \]  
\[ Y = Z\Pi + V, \]  

where \( \text{cov}(u, V) \neq 0 \), \( Y : n \times m, Z : n \times k, k \geq m \), we also assume \( EZ_iV'_i = 0 \), for \( i = 1, \cdots, n \). Then we assume, for \( i = 1, \cdots, n \),

\[ EZ_iu_i = \frac{C}{\sqrt{n}}, \]

where \( C \) is a \( k \times 1 \) vector, each component of that vector \( C_{nj} (j = 1, \cdots, k) \) is a constant and is in a compact subset (\( \Gamma \)) of \( R^k \), \( C \in \Gamma \subset R^k \). This assumption allows for a mild correlation between the instruments and the structural error. The correlation can be negative or positive for different instruments. Note that the number of instruments "k" is fixed (i.e. it does not grow with the sample size). So in a two instrument case, we may have \( C = (1, -2)' \).

We assume without losing any generality that the lower and upper bounds for the compact subset are the same: (for each \( j = 1, 2 \cdots, k \)) \( \gamma_l, \gamma_u \), respectively. We also set \( \gamma_l = -\gamma_u \), where \( \gamma_u > 0 \). The symmetry assumption about the compact subset simplifies the proofs.

This is the same near exogeneity assumption used by Berkowitz, Caner and Fang (2008) in which the covariance between the structural error and the instruments is "\( C/\sqrt{n} \)" and where \( C \) is a constant vector. However we also consider the case when the bounds \( \gamma_l, \gamma_u \) get large and we consider a uniform result for \( C \in \Gamma \) in that context for the first time in the literature.

Note that there are no exogenous control variables in the system. In order to simplify the notation, control variables are not included in (1)(2). Control variables can be projected out to get the same results as here. Our goal is to test \( H_0 : \theta = \theta_0 \). The Anderson-Rubin (1949) test is used to test \( H_0 \) and is described as follows:

\[ AR(\theta_0) = [(y - Y\theta_0)'Z/n^{1/2}]\hat{\Omega}^{-1}[Z'(y - Y\theta_0)/n^{1/2}], \]  

where \( \hat{\Omega} = \frac{1}{n} \sum_{i=1}^n Z_iZ'_i u_i^2 \).
The Anderson-Rubin test (AR ($\theta_0$)) can also be rewritten as

$$AR(\theta_0) = \bar{S}_n'(\text{var} \bar{S}_n)^{-1}\bar{S}_n = n\bar{S}_n'\hat{\Omega}^{-1}\bar{S}_n,$$

where $\bar{S}_n = \sum_{i=1}^n \frac{Z_i u_i}{n} = \frac{Z'(y-Y\theta_0)}{n}$, and $\text{var} \bar{S}_n = \hat{\Omega}/n$. We can also demean $Z_i$ in the variance formula and this does not change the asymptotics.

2.1 Assumptions

In this section we introduce our assumptions and discuss them.

**Assumption 1.** For $i = 1 \cdots n$,

(i).

$$EZ_i u_i = \frac{C}{\sqrt{n}},$$

where $C$ is $k \times 1$ vector and $C_j/\sqrt{n} \to 0$, as $n \to \infty$, for each $j = 1, \cdots k$. Note that $k$ is fixed. We allow for both positive and negative covariance at the same in the $C$ vector. $C \in \Gamma$ which is a compact subset of $R^k$. Assume for each $j = 1, \cdots k$, the bounds of $\Gamma$ are the same: $\gamma_l$ is the lower bound (scalar) and the upper bound is $\gamma_u$ (scalar). Also set $\gamma_l = -\gamma_u$, and $\gamma_u > 0$, and denote the upper bound vector as $\Gamma_u = (\gamma_u, \cdots, \gamma_u)'$ is of $k \times 1$ dimension, and the lower bound vector as $\Gamma_l = (\gamma_l, \cdots, \gamma_l)'$ which is of $k \times 1$ dimension.

(ii).

$$\Omega = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n EZ_i Z_i' u_i^2,$$

where $\Omega$ is positive definite and finite and we assume also

$$\hat{\Omega} \to \Omega \quad a.s.$$

(iii).

$$EZ_i V_i' = 0.$$

**Assumption 2.**

$$\lim_{n \to \infty} n^{-1} \sum_{i=1}^n E||Z_i u_i||^3 < \infty.$$

Assumption 1 allows for a small covariance between the instruments and the structural error: it is how we operationalize mild violations of exogeneity. Assumption 1 is discussed above in this section. Note that it will become clear in Lemma 1b-c, Theorem 1b-c below that the symmetry of the compact subset will simplify the notation. The limits will be slightly different without this assumption, but more complicated in terms of notation.

Assumption 2 is needed for the strong law of large numbers approximation for obtaining the Berry-Esseen bounds. This assumption is discussed in Zhao, Wang and Wu (2004), as
Remark 3 after Corollary 1 in their paper. This is a sufficient condition for the Berry-Esseen bound for the independent case. The triangular array case can also be obtained by Theorem 1 in Zhao, Wang and Wu (2004), and Assumption 2 is again sufficient. The Berry-Esseen bound is used to prove resampling Central Limit Theorem type of result as (28)(29).

Assumption 2 also provides the sufficient condition in Theorem 23.12 of Davidson (1994) which is sufficient condition for Lindeberg Central Limit Theorem.

2.2 Full Sample Result

In this subsection we derive the limiting distribution of the full sample Anderson-Rubin (1949) test under our violation of the exogeneity condition in Assumption 1. In this paper both a pointwise result as well as a uniform result are analyzed over $C \in \Gamma$. To the best our knowledge, this is a new result in the literature. Before describing the limits we describe the relevant notation. The uniform limit changes with the number of instruments. To understand this, note that the limit of $n^{-1/2} \sum_{i=1}^{n} Z_i u_i - EZ_i u_i$ is the zero mean normal random variable/vector $L$ with variance $\Omega$ and the limit for the uniform case will change with the position of the random variable/vector $L$. When we have $2$ instruments, $L = (L_1, L_2)'$ which is a bivariate normal, zero mean distribution with $\Omega$ as the variance covariance matrix. Then denote $\Gamma_{l u} = (\gamma_l, \gamma_u)'$, $\Gamma_{u l} = (\gamma_u, \gamma_l)'$, and $k$ represents the number of instruments.

**Lemma 1.** Under Assumptions 1 and 2, we have

a). $$AR(\theta_0) \xrightarrow{d} \chi^2_{k, C'\Omega^{-1}C}.$$ (5)

b). For $k = 1$,

$$\sup_{C \in \Gamma} AR(\theta_0) \xrightarrow{d} \frac{\chi^2}{1, \frac{1}{2\pi}} 1\{L \leq 0\} + \frac{\chi^2}{1, \frac{1}{2\pi}} 1\{L > 0\}.$$ (6)

c). For $k = 2$,

$$\sup_{C \in \Gamma} AR(\theta_0) \xrightarrow{d} \frac{\chi^2}{2, \frac{1}{2\pi}} 1\{L_1 \leq 0, L_2 \leq 0\} + \frac{\chi^2}{2, \frac{1}{2\pi}} 1\{L_1 > 0, L_2 > 0\}$$

$$+ \frac{\chi^2}{2, \frac{1}{2\pi}} 1\{L_1 > 0, L_2 \leq 0\} + \frac{\chi^2}{2, \frac{1}{2\pi}} 1\{L_1 \leq 0, L_2 > 0\}.$$ (7)

where $C$ is the vector indicated in Assumption 1i and the limit in Lemma 1a is a noncentral $\chi^2$ distribution with $k$ degrees of freedom and $C'\Omega^{-1}C$ as the noncentrality parameter. Note that the noncentral $\chi^2$ limit in Lemma 1a is explicitly written as

$$(L + C)'\Omega^{-1}(L + C) \equiv L'\Omega^{-1}L + 2C'\Omega^{-1}L + C'\Omega^{-1}C$$

5
\[ \equiv \chi_k^2 + 2C'\Omega^{-1}L + C'\Omega^{-1}C, \quad (8) \]

by (22)(23).

For Lemma 1b, \( L \) is a normal random variable with \( \Omega \) as the variance. The non central \( \chi^2 \) distribution is with one degree of freedom both when \( L \leq 0 \), and \( L > 0 \). The non centrality parameter is \( \gamma_l^2/2\Omega \) when \( L \leq 0 \), and when \( L > 0 \), the noncentrality parameter is \( \gamma_u^2/2\Omega \). Both \( \gamma_u, \gamma_l \) are described in Assumption 1i. The explicit limit in this case is by (25)

\[
[(L + \gamma_l)'\Omega^{-1}(L + \gamma_l)]1_{\{L \leq 0\}} + [(L + \gamma_u)'\Omega^{-1}(L + \gamma_u)]1_{\{L > 0\}} \\
\equiv (\chi_1^2 + 2\gamma_l\Omega^{-1}L + \gamma_l^2\Omega^{-1})1_{\{L \leq 0\}} \\
+ (\chi_1^2 + 2\gamma_u\Omega^{-1}L + \gamma_u^2\Omega^{-1})1_{\{L > 0\}}. \quad (9)
\]

The explicit limit in Lemma 1c is more complicated. The limit depends on \( L = (L_1, L_2)' \) which is bivariate normal with variance covariance matrix \( \Omega \). This follows from equation (27). The case for \( k > 2 \) can be proved easily, but the notation gets really burdensome. This can be deducted easily from the case and the proof of \( k = 2 \).

Lemma 1 shows that if we use the standard \( \chi^2_k \) critical values when there is a violation of the exogeneity assumption, then the AR test will be oversized. This is also what we observe in the simulations in Table 1 for the setups discussed in Section 4. This extends the limit result for contiguous sequences in Berkowitz, Caner and Fang (2008) from a pointwise to a uniform result.

Lemma 1a, 1b, 1c show that the pointwise, and uniform results are different. The limit case for \( k > 2 \) involves several joint intersections of normal random variables. Also observe that the new uniform limits’ support starts at \( \gamma_u^2/\Omega \), rather than 0 as in the case of pointwise limit in Lemma 1a. So we expect this to give very conservative results and to waste power.

These findings are related to a recent paper by Guggenberger (2009), which analyzes various identification robust tests when there is a local to zero violation of exogeneity. Guggenberger (2009) finds that the AR test has the best finite sample properties. Caner (2009) analyzes the AR test in a many weak moments setup that allows for violations of exogeneity. He finds that when the number of violations is limited, the AR-test does not have size problems.

\(^1\)We thank Patrik Guggenberger for pointing out an mistake in the earlier version, and making the point that the pointwise and uniform limits should be different.
2.3 Resampling Technique

This subsection contains a description of Wu’s (1990) jackknife histogram estimator. We show that this resampling method is useful for recovering the limit for tests of the population mean.

It is well known that in large samples, to test $H_0 : \mu = \mu_0$, where $\mu$ is the population mean, and $\mu_0$ is the true value of the mean, $\bar{X} - \mu_0 \sim N(0, \frac{\sigma}{n^{1/2}})$.

$$\frac{\bar{X} - \mu_0}{\sigma/n^{1/2}} \xrightarrow{d} N(0,1), \quad (10)$$

where $\bar{X}$ is the sample mean out of $n$ observations, and $\sigma$ is the standard deviation of individual $x_i$.

We first describe Wu’s (1990) resampling technique. We take a subset of size “$b$” (block size) from $n$ observations. We resample from data “$x$” where $x = (x_1, \cdots, x_n)$. The blocks in this resampling from “$x$” are ”$x_b$” with size “$b$”, and equal probability of $\binom{n}{b}^{-1}$. This is done via simple random sampling without replacement from the population. The size of the blocks plays a crucial role in our results. Denote this resampling technique by “*”. Notation such as $P_*$, $E_*$ refer to calculations under “*”.

The jackknife histogram estimator is basically the resampled counterpart of the simple $t$ test in (10) (see also p.1440, Wu, 1990). So instead of $\mu_0$ we use $\bar{X}$; and, instead of $\bar{X}$ we use $\bar{X}_b$ where this is the sample mean from $b$ observations drawn without replacement from the sample of $n$ observations. Next instead of $\frac{\sigma}{n^{1/2}}$ we use $(1 - f)\hat{\sigma}^2/b$, where $f$ is a fraction of the sample, $0 < f_l < f < f_u < 1$, $f_l$ is the lower bound, and $f_u$ is the upper bound for fractions. So the term $(1 - f)\hat{\sigma}^2/b$ is the variance of $\bar{X}_b$ under simple random sampling without replacement (Cochran, 1977), and $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{X})^2$. Next, Wu (1990) shows that under simple regularity conditions

$$\sup_t |P_* \left( \frac{\sqrt{b}(\bar{X}_b - \bar{X})}{((1 - f)\hat{\sigma}^2)^{-1/2}} \leq t \right) - \phi(t)| \to 0 \quad a.s.,$$

where $\phi(t)$ represents the cumulative distribution function of standard normal law. So the cumulative distribution of the jackknife histogram estimator converges to the cumulative distribution of standard normal law.

2.4 Fractionally Resampled Anderson-Rubin Test

For our setup, we are interested in resampling from the following quantity: $Z'u = \sum_{i=1}^n Z_iu_i$, where $u = y - Y\theta_0$. Denoting the sample average by $\bar{S}_n = n^{-1}\sum_{i=1}^n Z_iu_i$, $\bar{S}_b$ is the mean of the simple random sample of size ”$b$” drawn without replacement from ”$n$” observations.
(the mean of the b observations that are drawn out of the sample of n observations). Thus, for block size \( b = fn, 0 < f_l < f < f_u < 1 \). Note that we are not directly using the estimators in equations (1.2) and (1.3) of Wu (1990). Instead we benefit from section 2, equations (2.2)(2.3) of Wu (1990). We also extend his case to independent random variables and the extension from iid to triangular arrays is simple and is discussed after Theorem 1. We discuss our choice of block size after Corollary 1. Note that the researcher chooses the fraction of the sample "f". The fractionally resampled Anderson-Rubin test (FAR (\( \theta_0 \))) is described as:

\[
FAR(\theta_0) = \bar{S}_b'(\text{var}^*_\bar{S}_b)^{-1}\bar{S}_b = \frac{b\bar{S}_b'\hat{\Omega}^{-1}\bar{S}_b}{(1 - f)}, (11)
\]

where immediately after equation (2.2) on p.1440 of Wu (1990) or Theorem 2.2. of Cochran (1977) it is shown that \( \text{var}^*_\bar{S}_b = \frac{1-f}{b}\hat{\Omega} \). Observe that the right-hand side in (11) is slightly different than the right hand side in (4). This is due to the property of \( \text{var}^*_\bar{S}_b = \left[ \frac{\hat{\Omega}}{b} \right] (1 - f) \).

This will play an important role in the derivation of our main result. The limit for \( FAR(\theta_0) \) still does not recover the limit in Lemma 1. Theorem 1 provides a very strong direction and intuition about how to obtain this limit by using a variant of resampling technique. After Theorem 1 we discuss how we provide a valid but slightly conservative limit compared to the limits in Lemma 1. Before deriving the limit for \( FAR(\theta_0) \), we define a multivariate normal distribution \( L \equiv N(0, \Omega) \) as in Lemma 1.

**Theorem 1.** For \( 0 < f_l < f < f_u < 1 \), and the test in (11),

a). Define

\[
J_b(t) = P_*(FAR(\theta_0) \leq t).
\]

Then, under Assumptions 1 and 2, given (5),

\[
\sup_t |J_b(t) - \phi_{mf}(t)| \to 0 \quad a.s.,
\]

where \( \phi_{mf}(t) \) is the cumulative distribution function of following distribution:

\[
(1 + \frac{\sqrt{f}}{\sqrt{1-f}})^2 \chi^2_{k, \frac{f}{1+2\sqrt{f}\sqrt{1-f}}}^{c'\Omega^{-1}C},
\]

where \( \chi^2_{k, \frac{f}{1+2\sqrt{f}\sqrt{1-f}}}^{c'\Omega^{-1}C} \) is the non central \( \chi^2 \) limit with \( k \) degrees of freedom and the noncentrality parameter \( \frac{f}{1+2\sqrt{f}\sqrt{1-f}}^{c'\Omega^{-1}C} \).

---

\(^2\)We could set \( b = \lfloor fn \rfloor \), where \( b/n \to f \) and \( \lfloor \cdot \rfloor \) is the integer part of the number fn. We do not do this because it would make our notation unnecessarily complicated. The notation we employ follows the notation in section 2 of Wu (1990),.
b). Define $J_H(t) = P_*(\sup_{C \in \Gamma} \text{FAR}(\theta_0) \leq t)$. Then, under Assumptions 1 and 2, given (6), with $k = 1$, 

$$
\sup_{t} |J_H(t) - \phi_{flu}(t)| \to 0 \text{ a.s.,}
$$

where $\phi_{flu}(t)$ is the cumulative distribution of the limit

$$(1 + \frac{\sqrt{J}}{\sqrt{1 - J}})^2 \left( \frac{\chi^2_{\lfloor \frac{\gamma^2_{2,1+2\sqrt{J} \sqrt{1 - J}}}{f} \rfloor}}{\chi^2_{\lfloor \frac{\gamma^2_{2,1+2\sqrt{J} \sqrt{1 - J}}}{f} \rfloor}} \right) 1_{\{L + \sqrt{\frac{J}{1 - J}} L \leq 0\}} + (1 + \frac{\sqrt{J}}{\sqrt{1 - J}})^2 \left( \frac{\chi^2_{\lfloor \frac{\gamma^2_{2,1+2\sqrt{J} \sqrt{1 - J}}}{f} \rfloor}}{\chi^2_{\lfloor \frac{\gamma^2_{2,1+2\sqrt{J} \sqrt{1 - J}}}{f} \rfloor}} \right) 1_{\{L + \sqrt{\frac{J}{1 - J}} L > 0\}}.$$

where $\chi^2_{\lfloor \frac{\gamma^2_{2,1+2\sqrt{J} \sqrt{1 - J}}}{f} \rfloor}$ is the non central $\chi^2$ limit with 1 degree of freedom and the noncentrality parameter $\frac{\gamma^2_{2,1+2\sqrt{J} \sqrt{1 - J}}}{f}$ and where $\chi^2_{\lfloor \frac{\gamma^2_{2,1+2\sqrt{J} \sqrt{1 - J}}}{f} \rfloor}$ is the non central $\chi^2$ limit with 1 degree of freedom and the noncentrality parameter $\frac{\gamma^2_{2,1+2\sqrt{J} \sqrt{1 - J}}}{f}$.

c). Under Assumptions 1 and 2, given (7), with $k = 2$, 

$$
\sup_{t} |J_H(t) - \phi_{flu2}(t)| \to 0 \text{ a.s.,}
$$

where $\phi_{flu2}(t)$ is the cumulative distribution of the following limit, and $L = (L_1, L_2)'$,

$$(1 + \frac{\sqrt{J}}{\sqrt{1 - J}})^2 \left( \frac{\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}}{\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}} \right) 1_{\{L_1 + \sqrt{\frac{J}{1 - J}} L_1 \leq 0, L_2 + \sqrt{\frac{J}{1 - J}} L_2 \leq 0\}} + (1 + \frac{\sqrt{J}}{\sqrt{1 - J}})^2 \left( \frac{\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}}{\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}} \right) 1_{\{L_2 + \sqrt{\frac{J}{1 - J}} L_2 > 0, L_2 + \sqrt{\frac{J}{1 - J}} L_2 > 0\}} + (1 + \frac{\sqrt{J}}{\sqrt{1 - J}})^2 \left( \frac{\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}}{\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}} \right) 1_{\{L_1 + \sqrt{\frac{J}{1 - J}} L_1 \leq 0, L_2 + \sqrt{\frac{J}{1 - J}} L_2 > 0\}} + (1 + \frac{\sqrt{J}}{\sqrt{1 - J}})^2 \left( \frac{\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}}{\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}} \right) 1_{\{L_1 + \sqrt{\frac{J}{1 - J}} L_1 > 0, L_2 + \sqrt{\frac{J}{1 - J}} L_2 \leq 0\}}.$$

Note that $\Gamma_{flu} = (\gamma_l, \gamma_u)' : 2 \times 1$ vector, and $\Gamma_{flu} = (\gamma_l, \gamma_u)' : 2 \times 1$ vector. To explain the limit we explain one of the terms, the others are holds in the same way: $\chi^2_{\lfloor \frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2} \rfloor}$ is the non central $\chi^2$ limit with 2 degrees of freedom and the noncentrality parameter $\frac{\Gamma_{1}^T \Omega^{-1} \Gamma_t}{2}$.
Remarks.

1. Note that the limit in Theorem 1a is written explicitly as
\[
\chi_k^2(1 + \sqrt{\frac{f}{1-f}})^2 + 2 \left( 1 + \sqrt{\frac{f}{1-f}} \right) \sqrt{\frac{f}{1-f}} \left( C'\Omega^{-1}L \right) + \frac{f}{1-f} \left( C'\Omega^{-1}C \right),
\]
where \( \chi_k^2 \) is the central \( \chi^2 \) distribution with \( k \) degrees of freedom, and \( L \equiv N(0, \Omega) \). This is shown in detail in the proof of Theorem 1a through (33)-(34).

Also the limit in Theorem 1b is explicitly written by (37)-(25)
\[
\begin{align*}
\left[ \chi_1^2 \left( 1 + \sqrt{\frac{f}{1-f}} \right)^2 + 2 \left( 1 + \sqrt{\frac{f}{1-f}} \right) \left( \sqrt{\frac{f}{1-f}} \right) \gamma_1\Omega^{-1}L + \frac{f}{1-f} \gamma_1^2\Omega^{-1} \right] & 1_{\{L + \sqrt{f(1-f)}L \leq 0\}} \\
+ \left[ \chi_1^2 \left( 1 + \sqrt{\frac{f}{1-f}} \right)^2 + 2 \left( 1 + \sqrt{\frac{f}{1-f}} \right) \left( \sqrt{\frac{f}{1-f}} \right) \gamma_u\Omega^{-1}L + \frac{f}{1-f} \gamma_u^2\Omega^{-1} \right] & 1_{\{L + \sqrt{f(1-f)}L > 0\}}
\end{align*}
\]

2. An important point is to have inference by resampling. It seems that a fixed \( f \) will not match the limit in Theorem 1a with the one in Lemma 1a perfectly. But with \( f = 1/2 \) we get a valid solution pointwise. This is conservative but corrects the size distortion. When the fraction is \( 1/2 \), the limit in Theorem 1a, equation (12) becomes
\[
4\chi_k^2 + 4C'\Omega^{-1}L + C'\Omega^{-1}C.
\]
In Lemma 1a, the limit is by (8)
\[
\chi_k^2 + 2C'\Omega^{-1}L + C'\Omega^{-1}C.
\]

Comparing (14) with (15) and since both the \( AR(\theta_0), FAR(\theta_0) \) use right hand side critical values, then using the critical values from (14) prevent size distortions. However, this may become too conservative. So for the finite samples, we suggest the following: \( fn \to 1/2 \), where \( fn \) is a deterministic sequence. Specifically, we can choose \( fn = 1/2 - \kappa_n \), where \( \kappa_n > 0 \), for each \( n \) and \( \kappa_n \to 0 \) when \( n \to \infty \), \( \kappa_n \) is a deterministic sequence, not a random sequence. Since \( fn \to 1/2 \), the limit by using this fraction will be again (14). This can also be seen by following the proof of Theorem 1. But in small samples since \( fn < 1/2 \), it may provide some power. We will consider two choices for \( \kappa_n \) in the simulation section.

3. The proof uses Assumption 2 which is for independent data. Triangular arrays could also easily be used, and this is discussed after Assumption 2.

4. Note that as in Lemma 1, the uniform results are different than pointwise results, and hence depend on the behaviour of the random vector \( L \). The case for \( k > 2 \) can be built easily using the case for \( k = 2 \), but the limit expression will have lots of extra terms depending on the new intersections of the random variables in the \( k \) vector \( L \).
5. For the uniform result, again we see that for any \( f \) it is not difficult to see the limit in Theorem 1b will not provide valid critical values. This is also discussed in the pointwise case in Remark 1. But we can again set \( f = 1/2 \) in (13) and get a valid but conservative limit:

\[
\left[ 4\chi_1^2 + 4\gamma_l\Omega^{-1}L + \gamma_l^2\Omega^{-1} \right] 1_{\{2L \leq 0\}} + \left[ 4\chi_1^2 + 4\gamma_u\Omega^{-1}L + \gamma_u^2\Omega^{-1} \right] 1_{\{2L > 0\}}. \quad (16)
\]

The limit in Lemma 1b, (6), is the following:

\[
\left[ \chi_1^2 + 2\gamma_l\Omega^{-1}L + \gamma_l^2\Omega^{-1} \right] 1_{\{L \leq 0\}} + \left[ \chi_1^2 + 2\gamma_u\Omega^{-1}L + \gamma_u^2\Omega^{-1} \right] 1_{\{L > 0\}}.
\]

Since \( 1_{\{2L \leq 0\}} = 1_{\{L \leq 0\}} \) and \( 1_{\{2L > 0\}} = 1_{\{L > 0\}} \), the limit in Theorem 1b is a valid but conservative solution. So the case of \( f = 1/2 \) works even in the uniform case.

2.5 Power Issues of \( AR(\theta_0), FAR(\theta_0) \)

In order to simplify the notation and to show the effects of other fractions, the analysis of the power of the \( FAR(\theta_0) \) will be conducted with a fixed choice of \( f \). From that we can infer what may happen also at \( f = 1/2 \) specifically.

Here we briefly show that the \( AR(\theta_0) \) is consistent against fixed alternatives when there is strong identification and a mild violation of exogeneity. So in equation (2), assume \( \Pi \) has full column rank, but Assumption 1 still holds. If the true parameter value is \( \theta_1 \) (for the power exercise here), and \( \theta_1 \neq \theta_0, \theta_1 - \theta_0 = l \), where \( l \) is a nonzero constant, we test

\[ H_0 : \theta = \theta_0. \]

Then using equations (2) and (3) with \( \theta_1 - \theta_0 = l \neq 0 \) it follows that \( AR(\theta_0) \overset{P}{\rightarrow} \infty \). The same is true uniformly over \( C \in \Gamma \).

The key issue is whether the resampling technique that we suggest will have power, and whether this is consistent against fixed alternatives. The key term in the analysis is \( \tilde{S}_b \) in equation (11). Note that \( (\cdot)_* \) represents resampling \( b \) observations out of \( n \) without replacement. So we can express

\[
\tilde{S}_b = \frac{\sum_{i=1}^{b} (Z_iZ_i')}{\sqrt{b}} \Pi(\theta_1 - \theta_0) + \frac{\sum_{i=1}^{b} (Z_iV_i')}{\sqrt{b}} (\theta_1 - \theta_0) + \frac{\sum_{i=1}^{b} (Z_iu_i)}{\sqrt{b}}. \quad (17)
\]

Note that the first term on the right hand side can be rewritten as

\[
\sqrt{b} \left( \frac{\sum_{i=1}^{b} (Z_iZ_i')}{b} \right)_*.
\]
And, this term converges in probability to infinity at rate $\sqrt{b} = \sqrt{n}\sqrt{f}$. The test will not be consistent if we use the resampled critical values, even though the test diverges at rate $n$. However, in finite samples we may gain power with a smaller "f". So the simulation exercises are crucial for choosing the correct block size in small samples. And, since the main problem is size, a reasonable power loss is acceptable.

2.6 The Algorithm

Next we write the algorithm to test the null of $H_0 : \theta = \theta_0$ by using the critical values obtained from the empirical distribution function of the half-sampled $FAR(\theta_0)$.

Step 1: First calculate the terms $\hat{\Omega}$ from the full sample of $Z_i, u_i$ as described at the beginning of this section.

Step 2: Denote $y_{b_o}, Y_{b_o}, Z_{b_o}$ as draws of block size $b_o = n/2$ from full sample $y, Y, Z$ without replacement, respectively. Note that $y_{b_o} : n/2 \times 1, Y_{b_o} : n/2 \times m, Z_{b_o} : n/2 \times k$. Form

$$\bar{S}_{b_o} = [Z'_{b_o}(y_{b_o} - Y_{b_o}\theta_0)]/(n/2).$$

Step 3. Form $FAR_o(\theta_0)$ by using steps 1-2.

Step 4. Repeat steps 2-4, $J$ times. ($J$ may be 1000, or 5000) Then sort $J$ values of what is found in step 4 to form the empirical distribution function.

Step 5. For a 5% test find the 95 percentile of the empirical distribution function in step 4.

Step 6. Reject the null of $H_0 : \theta = \theta_0$, if the full sample $AR(\theta_0)$ as described in equation (3) is larger than the 95th percentile in step 5.

Note that in the above algorithm, $\hat{\Omega}$ is calculated from the full sample. Only the numerator of the test statistic, the score, has to be resampled. The main technical reason for that is shown in the proof of Theorem 1. Basically for any block size $b$, p.1440 of Wu (1990) or Theorem 2.2 of Cochran (1977) shows that $\text{var}_r \bar{S}_b = \frac{1}{n} f b \hat{\Omega}$ in our case.

For $f_n$, we use $f_n = 1/2 - \kappa_n$ instead of $f = 1/2$ in the above algorithm, if the fraction of the sample is not an integer we choose the next highest integer.

3 Comparison With Subsampling and Variants

In this section we compare the resampling technique employed in Theorem 1 with subsampling. We use the limit fraction $f$ in the analysis here just to simplify, this is also valid with $f_n^*$. Note that subsampling the AR($\theta_0$) test (or $supAR(\theta_0)$) will not work because it will be oversized. The main reason for this is subsampling described by Politis, Romano and Wolf (1999) allows $b \to \infty, n \to \infty, b/n \to 0$. So $f \to 0$ in our exercise.
A simple counterexample can be seen from our results since in that case \( b/n \to 0 \), as \( b \to \infty, n \to \infty \), meaning \( f \to 0 \). That means the noncentrality parameter in Theorem 1a

\[
\frac{f}{1 + 2\sqrt{f}\sqrt{1 - f}} \frac{C'\Omega^{-1}C}{2} \to 0,
\]

and

\[
\left(1 + \frac{\sqrt{f}}{\sqrt{1 - f}}\right)^2 \to 1.
\]

Hence, the subsampled \( AR(\theta_0) \) test statistics will converge to the standard \( \chi^2 \) limit. Clearly, this is oversized. The subsampled limit is stochastically less than the one in (5).

The work of Andrews and Guggenberger (2007) is important for this section because it analyzes cases in which the subsampling approach works and cases in which it fails. Now we illustrate the difference between subsampling and fractional resampling in a simple example used in section 2 of Andrews and Guggenberger (2007) which analyzes a simple boundary problem. The true parameter \( \theta_0 \) is nonnegative. Assume that \( X_i \) is iid with \( N(0,1) \), for \( i = 1, \cdots, n \). The Maximum Likelihood Estimator (MLE) of \( \theta_0 \) is \( \hat{\theta}_n = \max\{\bar{X}_n, 0\} \), and \( \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \). The distribution of \( \hat{\theta}_n \) is

\[
\hat{\theta}_n \sim \max\{Z_n, 0\}, \quad Z_n \sim N(\theta_0, \frac{1}{n}).
\]

Then we subsample \( (b_{n,s}/n \to 0, \text{ as } b_{n,s} \to \infty, n \to \infty) \), with \( b_{n,s} = o(n) \), the subsampled estimator is \( \hat{\theta}_{b_{n,s,j}} = \max\{\bar{X}_{b_{n,s,j}}, 0\} \) where \( \bar{X}_{b_{n,s,j}} = \frac{1}{b_{n,s}} \sum_{i=j}^{j+b_{n,s} - 1} X_i \) and the distribution is

\[
\hat{\theta}_{b_{n,s,j}} \sim \max\{Z_{b_{n,s}}, 0\}, \quad Z_{b_{n,s}} \sim N(\theta_0, \frac{1}{b_{n,s}}).
\]

It is clear that the distribution of \( \hat{\theta}_{b_{n,s,j}} \) does not replicate the distribution of \( \hat{\theta}_n \). This is thoroughly discussed in Andrews and Guggenberger (2007). The main reason is that the subsample estimator is closer to the boundary of parameter space than the full sample estimator. To see this \( \text{var} \hat{\theta}_{b_{n,s,j}} \) when \( Z_n > 0 \) is \( 1/b_{n,s} \), and \( \text{var} \hat{\theta}_n \) is \( 1/n \) when \( Z_n > 0 \). Since \( b_{n,s} = o(n) \), \( 1/b_{n,s} \) is larger than \( 1/n \), and hence more variable near the boundary.

Using fractional resampling for this specific example, we set the fractions in a more restrictive way. Set \( b_n = fn, f \in [1/2, f_u], f_u < 1 \), so \( b_n = O(n) \). When we use this technique in the case of the former example (with \( Z_{b_n} > 0 \)), the variance of the fractionally resampled estimator is:

\[
\text{var} \hat{\theta}_{b_n} = \frac{1 - f}{b_n}.
\]

To understand this, we use section 2 of Wu (1990). The variance of the resampled mean is \( \frac{(1-f)\text{var}X_i}{b_n} \), where \( \text{var}X_i = 1 \) in this example. Then note that compared to the original variance and the subsampled ones

\[
\text{var} \hat{\theta}_{b_{n,s,j}} > \text{var} \hat{\theta} \geq \text{var} \hat{\theta}_{b_n},
\]
since
\[
\frac{1}{b_{n,s}} > \frac{1}{n} \geq \frac{(1 - f)}{b_n} = \left(\frac{1 - f}{f}\right) \frac{1}{n},
\]
and \(f \in [1/2, f_u], f_u < 1\). At \(f = 1/2\) we have the optimal choice and capture the variance. This shows that variability in this technique is less than or equal to the subsampling technique. In addition to regular subsampling, Andrews and Guggenberger (2009) suggest hybrid subsampling, and size-corrected subsampling. We think that these methods may result in a loss of power in our problem since the critical values are adjusted to make them larger in magnitude in these methods.

4 Simulation

This section describes the small sample properties of the tests proposed in equation (11). We consider the algorithm in section 2.6, and we resample the critical values according to the deterministic fraction sequence \(f_n = 1/2 - \kappa_n\). We set \(\kappa_n = 1.5/\sqrt{n}\) or \(1/\sqrt{n}\). The other choices for \(\kappa_n\) are tried in terms of convergence rates but these two rates provide the best results. Note that with \(n = 100, 200\), the choice of \(\kappa_n = 1.5/\sqrt{n}\) leads to block sizes of 35, 79 respectively. For \(\kappa_n = 1/\sqrt{n}\) with \(n = 100, 200\), the block sizes are 40, 86 respectively.

In the uniform case we believe that resampling will waste power. This is true since the support of the uniform limit starts at \(\gamma^2_u/\Omega\) in the case of \(k = 1\), compared to the support of the pointwise case at 0.

We consider several block sizes and use the setup in section 2, namely
\[
y_i = Y_i\theta_0 + u_i,
\]
\[
Y_i = Z_i\Pi + V_i,
\]
for \(i = 1, \ldots, n\). The sample size is \(n\) and varies between 100 and 200. We consider the case of one instrument and one endogenous regressor, so \(k = 1, m = 1\) (exact identification). A case with overidentification is also considered, but not reported here because the results are very similar. \(\Pi\) can take the values of 1 (strong identification), and 0.1 (weak identification). The iid data \((Z_i, u_i, V_i)\) are generated from a joint normal distribution \(N(0, \Omega)\) where
\[
\Omega = \begin{bmatrix}
1 & cov(Z_i, u_i) & 0 \\
cov(Z_i, u_i) & 1 & 0.5 \\
0 & 0.5 & 1
\end{bmatrix}.
\]
So \(var Z_i = var u_i = var V_i = 1, cov(Z_i, V_i) = 0, cov(u_i, V_i) = 0.5\). For the size exercise \(\theta_0 = 0\), we test \(H_0 : \theta = 0\). For the power \(\theta_0 = -2, -1.5, -1, -0.5, 0.5, 1, 1.5, 2\).
We consider three setups for the $\text{cov}(Z_i, u_i)$ term. The first setup is consistent with Assumption 1

$$\text{cov}(Z_i, u_i) = \frac{C}{n^{1/2}},$$

(18)

and $C$ takes the values of 2, 3 and 5. As $C$ becomes larger, endogeneity becomes more problematic. And, the researcher picks a terrible instrument when $C = 5$.

In the second setup we have:

$$\text{cov}(Z_i, u_i) = D,$$

(19)

where $D$ is a constant and takes on the values 0.2, 0.3, 0.5. In this setup we expect large size distortions to emerge as the sample becomes larger, because the drift $D$ is multiplied with the square root of the block size in the score in the test statistic. We also used negative values for the covariance term, but the results do not change and hence are not reported.

In the third setup, we have something that is more consistent with $\Gamma_u$ approaching infinity.

$$\text{cov}(Z_i, u_i) = \frac{an^{1/3}}{n^{1/2}},$$

(20)

where $C_n = an^{1/3}$, and $a$ takes the values of 0.25, 0.5 and 1. At $n = 100$, these correspond to covariances (and correlations, since the variances are normalized at 1) of 0.12, 0.23 and 0.46 respectively for $a = 0.25, 0.5, 1$.

For the size exercise we set $\Pi = 2$ (strong identification), and also $\Pi = 0.2$ (weak identification). For the power exercise we only use strong identification case, since AR test is known to have low power when there is weak identification. In all exercises, we use 1000 iterations.

Table 1 reports the size of the full sample regular $AR(\theta_0)$ test in (3). This is compared with asymptotic critical values for $\chi^2_1$ distribution at 10% level. We report the rejection rates of the true null in Table 1. We see that both in setups 1 and 2 the actual size is very large. In setup 1, at $C = 2$, the size is 66% with $n = 200$. This shows there is a major size distortion problem if we use the $AR(\theta_0)$ test when there is a violation of exogeneity. This can also be seen for t-tests in Berkowitz, Caner and Fang (2008). The size calculations are done for $\Pi = 1$. Simulations for the case when $\Pi = 0.1$ are also done, but the results are not reported because they are very similar to the case where $\Pi = 1$. Another point is that size gets worse as the sample size increases in setup 2. This case sends an important warning to applied researchers who believe that increasing the sample size can correct for size distortions! In fact, as is very clear from setup 3, a larger sample size can also increase size distortions. With $a = 1$, and $n = 100$, the correlation is 0.46, the instrument is poorly selected and there is a huge size distortion.

Tables 2-4 show the size of the test under two possible $f_n$ sequences. With $f_n = 1/2 - 1.5/\sqrt{n}$, at small to moderate correlation levels size is always less than 10% when the
sample size is either 100 or 200. In setups 1 and 3, it is notable that when the sample size increases from 100 to 200, the size decreases from 33\% to 1.8\% at $C = 5$ in Table 2. Also in Table 4, we see that when $a = 1$ with $n = 100$, the size is 19.5\% but with $n = 200$ the size decreases to 7\%. We also observe with $f_n = 1/2 - 1/\sqrt{n}$, tests are heavily undersized, so this reflects a power problem with that specific choice.

Note that Tables 2-4 assume homoskedastic errors and sets up the test and the resampling in that way. In Tables 5-7, we introduce conditionally heteroskedastic case. For this we borrow a heteroskedastic setup used in Guggenberger and Smith (2005):

$$u^*_i = |Z_i|u_i,$$

where $u^*_i$ represents the new structural error. The conditional heteroskedasticity robust AR statistic and FAR statistic are defined in (3) and (7) respectively, where $\hat{\Omega} = \frac{1}{n} \sum_{i=1}^{n} Z_i Z'_i u^2_i$ is White’s conditional heteroskedasticity consistent estimator. Tables 5-8 report actual sizes of the conditional heteroskedasticity robust FAR test under three setups when instruments are strong and weak ($\Pi = 2, \Pi = 0.2$).

In Tables 5-7, the size of the test is much smaller at $n = 100$ compared with $n = 100$ in the homoskedastic case in Tables 2-4. We can see that at $n = 100$, with $D = 0.5$ in Table 6, the size is 4.1\% at 10\% nominal level with $\Pi = 2$. This is the case with $f_n = 1/2 - 1.5/\sqrt{n}$. When we use $f_n = 1/2 - 1/\sqrt{n}$, almost uniformly the size is 0. This will create power problems.

Tables 8-10 provide power results for the homoskedastic case. Clearly with $f_n = 1/2 - 1.5/\sqrt{n}$, we have very good power. For example at Table 8, with setup 1, the power is around 89-100\% under various alternatives. However, with the choice of $f_n = 1/2 - 1/\sqrt{n}$, the power declines substantially. At setup 1, the power is around 1-47\% under several alternatives.

Tables 11-13 show the power under heteroskedasticity. The results are similar to homoskedastic case. The main difference is that when $f_n = 1/2 - 1.5/\sqrt{n}$ in Table 11. The power is low at alternatives -0.5, 0.5. At other alternatives, the power is around 89-100\%.
Table 1: Size at 10%, $AR(\theta_0)$ test, $\Pi = 1$

| Sample Size | Setup 1 | | Setup 2 | | Setup 3 | |
|-------------|---------| |---------| |---------| |
|              | $C = 2$ | $C = 3$ | $C = 5$ | $D = 0.2$ | $D = 0.3$ | $D = 0.5$ | $a = 0.25$ | $a = 0.5$ | $a = 1$ |
| 100         | 65.0    | 93.0    | 100.0   | 64.0     | 93.0      | 100.0      | 32.8      | 77.9      | 93.0    |
| 200         | 66.0    | 93.0    | 100.0   | 88.0     | 99.0      | 100.0      | 42.2      | 91.7      | 100.0   |

Note: Setup 1 is explained in (18), Setup 2 is explained in (19). "D" represents the covariance between the instrument and the structural error. Setup 3 and constant "a" is explained in (20).

Table 2: Setup 1, Size at 10%, $FAR(\theta_0)$, Homoskedastic Case

| Correction Factor $\kappa_n$ = | $\Pi = 2$ | | $\Pi = 0.2$ | |
|------------------------------|-----------| |----------| |
| $\frac{1.5}{n^{1/2}}$      | $\frac{1}{n^{1/2}}$ | | $\frac{1.5}{n^{1/2}}$ | | $\frac{1}{n^{1/2}}$ | |
| $C = 2$                     | 0.2       | 0.0     | 0.1      | 0.0      | 0.0      |
| $C = 3$                     | 0.0       | 0.0     | 0.6      | 0.0      | 0.0      |
| $C = 5$                     | 33.1      | 1.8     | 29.5     | 1.8      | 0.0      |

Note: This is the test statistic in (11) and setup 1 is (18). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 3: Setup 2, Size at 10%, $FAR(\theta_0)$, Homoskedastic Case

| Correction Factor $\kappa_n$ = | $\Pi = 2$ | | $\Pi = 0.2$ | |
|------------------------------|-----------| |----------| |
| $\frac{1.5}{n^{1/2}}$      | $\frac{1}{n^{1/2}}$ | | $\frac{1.5}{n^{1/2}}$ | | $\frac{1}{n^{1/2}}$ | |
| $D = .2$                    | 0.0       | 0.0     | 0.0      | 0.0      | 0.0      |
| $D = .3$                    | 1.6       | 0.4     | 0.2      | 0.2      | 0.0      |
| $D = .5$                    | 29.6      | 34.1    | 34.5     | 35.9     | 0.0      |

Note: This is the test statistic in (11) and setup 1 is (19). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.
Table 4: Setup 3, Size at 10%, \( \text{FAR}(\theta_0) \), Homoskedastic Case

<table>
<thead>
<tr>
<th>Correction Factor ( \kappa_n = )</th>
<th>( \Pi = 2 )</th>
<th>( \Pi = 0.2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>( n = 200 )</td>
</tr>
<tr>
<td>( a = 0.25 )</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>( a = 0.50 )</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>( a = 1 )</td>
<td>19.5</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 1 is (20). Note that \( \kappa_n = 1.5/\sqrt{n} \), \( \kappa_n = 1/\sqrt{n} \) corresponds to \( f_n = 1/2 - 1.5/\sqrt{n} \) and \( f_n = 1/2 - 1/\sqrt{n} \) respectively.

Table 5: Setup 1, Size at 10%, \( \text{FAR}(\theta_0) \), Heteroskedastic Case

<table>
<thead>
<tr>
<th>Correction Factor ( \kappa_n = )</th>
<th>( \Pi = 2 )</th>
<th>( \Pi = 0.2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>( n = 200 )</td>
</tr>
<tr>
<td>( C = 2 )</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>( C = 3 )</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>( C = 5 )</td>
<td>3.7</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 1 is (18). Note that \( \kappa_n = 1.5/\sqrt{n} \), \( \kappa_n = 1/\sqrt{n} \) corresponds to \( f_n = 1/2 - 1.5/\sqrt{n} \) and \( f_n = 1/2 - 1/\sqrt{n} \) respectively.

5 Conclusion

Instruments that perfectly satisfy the knife-edge orthogonality assumption are few and far between. This paper shows that it is possible to conduct inference using instrumental variables when there is a mild violation of the exogeneity assumption. We use a novel resampling technique which draws random blocks from the all of the sample without replacement. The fractionally resampled Anderson-Rubin test (\( \text{FAR}(\theta_0) \)) does not overreject the null in large samples when the fraction is half of the sample. We also provide a finite sample correction for the block size that provides good power.
Table 6: Setup 2, Size at 10%, $FAR(\theta_0)$, Heteroskedastic Case

<table>
<thead>
<tr>
<th>Correction Factor $\kappa_n =$</th>
<th>$\Pi = 2$</th>
<th>$\Pi = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 100$</td>
<td>$n = 200$</td>
</tr>
<tr>
<td>$D = 0.2$</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$D = 0.3$</td>
<td>1.6</td>
<td>0.0</td>
</tr>
<tr>
<td>$D = 0.5$</td>
<td>4.1</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 1 is (19). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 7: Setup 3, Size at 10%, $FAR(\theta_0)$, Homoskedastic Case

<table>
<thead>
<tr>
<th>Correction Factor $\kappa_n =$</th>
<th>$\Pi = 2$</th>
<th>$\Pi = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 100$</td>
<td>$n = 200$</td>
</tr>
<tr>
<td>$a = 0.25$</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$a = 0.50$</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>$a = 1$</td>
<td>2.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 1 is (20). Note that $\kappa_n = 1.5/\sqrt{n}$, $\kappa_n = 1/\sqrt{n}$ corresponds to $f_n = 1/2 - 1.5/\sqrt{n}$ and $f_n = 1/2 - 1/\sqrt{n}$ respectively.

Table 8: Setup 1, Power, $FAR(\theta_0)$, Homoskedastic Case

<table>
<thead>
<tr>
<th>$\theta_0 =$</th>
<th>$f_n = 1/2 - 1.5/\sqrt{n}$, $n = 200$</th>
<th>$f_n = 1/2 - 1/\sqrt{n}$, $n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>100.0</td>
<td>47.3</td>
</tr>
<tr>
<td>-1.5</td>
<td>100.0</td>
<td>43.4</td>
</tr>
<tr>
<td>-1</td>
<td>99.6</td>
<td>37.4</td>
</tr>
<tr>
<td>-0.5</td>
<td>98.1</td>
<td>3.7</td>
</tr>
<tr>
<td>0.5</td>
<td>92.3</td>
<td>1.1</td>
</tr>
<tr>
<td>1</td>
<td>99.2</td>
<td>13.6</td>
</tr>
<tr>
<td>1.5</td>
<td>99.5</td>
<td>23.6</td>
</tr>
<tr>
<td>2</td>
<td>99.3</td>
<td>28.2</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 1 is (18).
Table 9: Setup 2, Power, $FAR(\theta_0)$, Homoskedastic Case

<table>
<thead>
<tr>
<th>$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$</th>
<th>$\theta_0 =$</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D = 0.2$</td>
<td></td>
<td>99.7</td>
<td>99.8</td>
<td>99.7</td>
<td>96.5</td>
<td>95.1</td>
<td>99.3</td>
<td>99.7</td>
<td>100.0</td>
</tr>
<tr>
<td>$D = 0.3$</td>
<td></td>
<td>100.0</td>
<td>100.0</td>
<td>99.7</td>
<td>92.0</td>
<td>97.7</td>
<td>99.8</td>
<td>97.7</td>
<td>100.0</td>
</tr>
<tr>
<td>$D = 0.5$</td>
<td></td>
<td>99.9</td>
<td>100.0</td>
<td>99.6</td>
<td>69.5</td>
<td>99.5</td>
<td>99.8</td>
<td>100.0</td>
<td>99.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$f_n = 1/2 - 1/\sqrt{n}, n = 200$</th>
<th>$\theta_0 =$</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D = 0.2$</td>
<td></td>
<td>46.3</td>
<td>40.7</td>
<td>35.4</td>
<td>1.6</td>
<td>1.7</td>
<td>13.4</td>
<td>24.1</td>
<td>27.5</td>
</tr>
<tr>
<td>$D = 0.3$</td>
<td></td>
<td>44.1</td>
<td>43.0</td>
<td>36.4</td>
<td>1.0</td>
<td>3.5</td>
<td>15.7</td>
<td>26.2</td>
<td>30.2</td>
</tr>
<tr>
<td>$D = 0.5$</td>
<td></td>
<td>43.2</td>
<td>41.0</td>
<td>34.0</td>
<td>0.0</td>
<td>12.1</td>
<td>22.5</td>
<td>29.0</td>
<td>31.8</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 2 is (19).

Table 10: Setup 3, Power, $FAR(\theta_0)$, Homoskedastic Case

<table>
<thead>
<tr>
<th>$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$</th>
<th>$\theta_0 =$</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a = 0.25$</td>
<td></td>
<td>100.0</td>
<td>99.9</td>
<td>99.9</td>
<td>97.6</td>
<td>90.2</td>
<td>99.5</td>
<td>99.7</td>
<td>100.0</td>
</tr>
<tr>
<td>$a = 0.50$</td>
<td></td>
<td>99.9</td>
<td>99.8</td>
<td>99.9</td>
<td>96.6</td>
<td>95.5</td>
<td>99.3</td>
<td>99.6</td>
<td>99.7</td>
</tr>
<tr>
<td>$a = 1.00$</td>
<td></td>
<td>99.8</td>
<td>99.9</td>
<td>99.9</td>
<td>83.2</td>
<td>98.6</td>
<td>99.7</td>
<td>99.7</td>
<td>99.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$f_n = 1/2 - 1/\sqrt{n}, n = 200$</th>
<th>$\theta_0 =$</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a = 0.25$</td>
<td></td>
<td>44.9</td>
<td>42.5</td>
<td>37.8</td>
<td>4.8</td>
<td>1.1</td>
<td>10.7</td>
<td>19.8</td>
<td>25.9</td>
</tr>
<tr>
<td>$a = 0.50$</td>
<td></td>
<td>42.1</td>
<td>46.2</td>
<td>34.7</td>
<td>1.0</td>
<td>2.5</td>
<td>16.2</td>
<td>24.6</td>
<td>31.8</td>
</tr>
<tr>
<td>$a = 1.00$</td>
<td></td>
<td>45.4</td>
<td>40.6</td>
<td>32.7</td>
<td>0.2</td>
<td>7.2</td>
<td>22.6</td>
<td>28.7</td>
<td>32.7</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 3 is (20).
Table 11: Setup 1, Power, FAR($\theta_0$), Heteroskedastic Case

<table>
<thead>
<tr>
<th>$\theta_0$ =</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C = 2$</td>
<td>99.5</td>
<td>99.9</td>
<td>99.0</td>
<td>30.1</td>
<td>35.7</td>
<td>88.6</td>
<td>97.4</td>
<td>98.1</td>
</tr>
<tr>
<td>$C = 3$</td>
<td>99.9</td>
<td>99.7</td>
<td>98.4</td>
<td>17.3</td>
<td>47.6</td>
<td>90.3</td>
<td>96.6</td>
<td>98.9</td>
</tr>
<tr>
<td>$C = 5$</td>
<td>100.0</td>
<td>99.6</td>
<td>99.6</td>
<td>2.0</td>
<td>61.8</td>
<td>90.8</td>
<td>96.3</td>
<td>97.3</td>
</tr>
</tbody>
</table>

$f_n = 1/2 - 1.5/\sqrt{n}, n = 200$

<table>
<thead>
<tr>
<th>$\theta_0$ =</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C = 2$</td>
<td>49.0</td>
<td>43.1</td>
<td>21.3</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
<td>7.2</td>
<td>12.3</td>
</tr>
<tr>
<td>$C = 3$</td>
<td>53.3</td>
<td>50.6</td>
<td>27.3</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
<td>8.2</td>
<td>10.7</td>
</tr>
<tr>
<td>$C = 5$</td>
<td>63.1</td>
<td>63.6</td>
<td>34.9</td>
<td>0.0</td>
<td>0.3</td>
<td>2.3</td>
<td>6.9</td>
<td>11.2</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 1 is (18).

Table 12: Setup 2, Power, FAR($\theta_0$), Heteroskedastic Case

<table>
<thead>
<tr>
<th>$\theta_0$ =</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D = 0.2$</td>
<td>100.0</td>
<td>99.9</td>
<td>98.1</td>
<td>16.4</td>
<td>45.4</td>
<td>90.5</td>
<td>95.8</td>
<td>98.0</td>
</tr>
<tr>
<td>$D = 0.3$</td>
<td>99.8</td>
<td>100.0</td>
<td>99.3</td>
<td>6.1</td>
<td>58.0</td>
<td>90.0</td>
<td>96.3</td>
<td>97.9</td>
</tr>
<tr>
<td>$D = 0.5$</td>
<td>100.0</td>
<td>100.0</td>
<td>99.4</td>
<td>0.0</td>
<td>69.4</td>
<td>90.6</td>
<td>94.8</td>
<td>96.9</td>
</tr>
</tbody>
</table>

$f_n = 1/2 - 1/\sqrt{n}, n = 200$

<table>
<thead>
<tr>
<th>$\theta_0$ =</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D = 0.2$</td>
<td>54.0</td>
<td>48.6</td>
<td>24.4</td>
<td>0.0</td>
<td>0.0</td>
<td>1.4</td>
<td>7.2</td>
<td>11.1</td>
</tr>
<tr>
<td>$D = 0.3$</td>
<td>59.3</td>
<td>57.3</td>
<td>32.7</td>
<td>0.0</td>
<td>0.0</td>
<td>1.8</td>
<td>8.1</td>
<td>12.2</td>
</tr>
<tr>
<td>$D = 0.5$</td>
<td>77.4</td>
<td>80.2</td>
<td>44.4</td>
<td>0.0</td>
<td>0.3</td>
<td>2.5</td>
<td>6.3</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 2 is (19).
Table 13: Setup 3, Power, FAR($\theta_0$), Heteroskedastic Case

<table>
<thead>
<tr>
<th>$\theta_0$ =</th>
<th>-2</th>
<th>-1.5</th>
<th>-1</th>
<th>-0.5</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a = 0.25$</td>
<td>99.8</td>
<td>99.7</td>
<td>98.8</td>
<td>37.4</td>
<td>30.0</td>
<td>90.4</td>
<td>96.9</td>
<td>98.4</td>
</tr>
<tr>
<td>$a = 0.50$</td>
<td>100.0</td>
<td>99.9</td>
<td>99.0</td>
<td>15.1</td>
<td>44.5</td>
<td>91.6</td>
<td>96.2</td>
<td>98.3</td>
</tr>
<tr>
<td>$a = 1.00$</td>
<td>100.0</td>
<td>99.9</td>
<td>99.5</td>
<td>0.3</td>
<td>63.5</td>
<td>92.4</td>
<td>95.0</td>
<td>98.4</td>
</tr>
</tbody>
</table>

Note: This is the test statistic in (11) and setup 3 is (20).

**APPENDIX**

**Proof of Lemma 1.**

a). By (1)(3) we have

$$AR(\theta_0) = \left( \frac{u'Z}{\sqrt{n}} \right) \hat{\Omega}^{-1} \left( \frac{Z'u}{\sqrt{n}} \right)$$

$$= [n^{-1/2} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) + n^{-1/2} \sum_{i=1}^{n} EZ_i u_i]' \hat{\Omega}^{-1}$$

$$\times [n^{-1/2} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) + n^{-1/2} \sum_{i=1}^{n} EZ_i u_i]. \quad (21)$$

First by Assumption 1

$$n^{-1/2} \sum_{i=1}^{n} EZ_i u_i = C.$$ 

Then (21) can be written as

$$AR(\theta_0) = [n^{-1/2} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) + C]' \hat{\Omega}^{-1} [n^{-1/2} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) + C]. \quad (22)$$

Then in (22) we use the Lindeberg Central Limit Theorem

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) \overset{d}{\to} N(0, \Omega) \equiv L. \quad (23)$$

Then by Assumption 1, (22)(23)

$$AR(\theta_0) \overset{d}{\to} \chi^2_{k, \Sigma_0^{-1} C}.$$ 

b). So $k = 1$ in the proof here. The test is
\[ \sup_{C \in \Gamma} AR(\theta_0) = \sup\{ [n^{-1/2} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) + C] \bar{\Omega}^{-1} [n^{-1/2} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) + C] \}. \] (24)

Next benefiting from (23) and (24)

\[ \sup_{C \in \Gamma} AR(\theta_0) \xrightarrow{d} \sup\{ [L + C] \bar{\Omega}^{-1} [L + C] \} \]
\[ = \{ (L + \gamma_l) \bar{\Omega}^{-1} (L + \gamma_l) \}[\{ L \leq 0 \}]
+ \{ (L + \gamma_u) \bar{\Omega}^{-1} (L + \gamma_u) \}[\{ L > 0 \}]
\]
\[ = \chi_{l, \gamma_l^2/2\Omega}^2[\{ L \leq 0 \}] + \chi_{u, \gamma_u^2/2\Omega}^2[\{ L > 0 \}]. \quad (25) \]

where \( \chi_{l, \gamma_l^2/2\Omega}^2 \) is the non-central \( \chi^2 \) distribution with 1 degree of freedom, and the non-centrality parameter \( \gamma_l^2/2\Omega \), \( \chi_{u, \gamma_u^2/2\Omega}^2 \) is the non-central \( \chi^2 \) distribution with 1 degree of freedom, and the non-centrality parameter \( \gamma_u^2/2\Omega \). Note that \( L \) is the limit of \( n^{-1/2} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) \).

c). Now we provide the proof for the case of \( k = 2 \). The case for \( k > 2 \) follows the same steps, but the terms in the limit become numerous, and hence skipped. In this case, denote \( Z_{i1}, Z_{i2} \) as the first and second instruments respectively. The drifts follow the same argument, \( C_1, C_2 \) are the first and second elements of the \( C \) vector. Rewrite the AR test as

\[ AR(\theta_0) = \begin{bmatrix} n^{-1/2} \sum_{i=1}^{n} (Z_{i1} u_i - EZ_{i1} u_i) + C_1 \\ n^{-1/2} \sum_{i=1}^{n} (Z_{i2} u_i - EZ_{i2} u_i) + C_2 \end{bmatrix}' \bar{\Omega}^{-1} \begin{bmatrix} n^{-1/2} \sum_{i=1}^{n} (Z_{i1} u_i - EZ_{i1} u_i) + C_1 \\ n^{-1/2} \sum_{i=1}^{n} (Z_{i2} u_i - EZ_{i2} u_i) + C_2 \end{bmatrix}. \]

Next denote the limit of

\[ \begin{bmatrix} n^{-1/2} \sum_{i=1}^{n} (Z_{i1} u_i - EZ_{i1} u_i) \\ n^{-1/2} \sum_{i=1}^{n} (Z_{i2} u_i - EZ_{i2} u_i) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}. \]

Then using the Central Limit Theorem result above

\[ \sup_{C \in \Gamma} AR(\theta_0) \xrightarrow{d} \sup\{ \begin{bmatrix} L_1 + C \\ L_2 + C \end{bmatrix}' \bar{\Omega}^{-1} \begin{bmatrix} L_1 + C \\ L_2 + C \end{bmatrix} \}. \quad (26) \]

Then the limit in (26) can be simplified as in the proof of Lemma 1b, using whether \( L_1, L_2 \) are positive or not.
To see this the limit on the right hand side of (26)

\[
\sup_{C \in \Gamma} \left\{ \begin{bmatrix} L_1 + C \\ L_2 + C \end{bmatrix}' \right\} \times \hat{\Omega}^{-1} \left[ \begin{bmatrix} L_1 + C \\ L_2 + C \end{bmatrix} \right] \\
\end{equation}

\[
\equiv \left[ \begin{bmatrix} L_1 + \gamma_l \\ L_2 + \gamma_l \end{bmatrix} \right]' \hat{\Omega}^{-1} \left[ \begin{bmatrix} L_1 + \gamma_l \\ L_2 + \gamma_l \end{bmatrix} \right] 1_{\{L_1 \leq 0, L_2 \leq 0\}} \\
+ \left[ \begin{bmatrix} L_1 + \gamma_u \\ L_2 + \gamma_u \end{bmatrix} \right]' \hat{\Omega}^{-1} \left[ \begin{bmatrix} L_1 + \gamma_u \\ L_2 + \gamma_u \end{bmatrix} \right] 1_{\{L_1 > 0, L_2 > 0\}} \\
+ \left[ \begin{bmatrix} L_1 + \gamma_l \\ L_2 + \gamma_u \end{bmatrix} \right]' \hat{\Omega}^{-1} \left[ \begin{bmatrix} L_1 + \gamma_l \\ L_2 + \gamma_u \end{bmatrix} \right] 1_{\{L_1 \leq 0, L_2 > 0\}} \\
+ \left[ \begin{bmatrix} L_1 + \gamma_u \\ L_2 + \gamma_l \end{bmatrix} \right]' \hat{\Omega}^{-1} \left[ \begin{bmatrix} L_1 + \gamma_u \\ L_2 + \gamma_l \end{bmatrix} \right] 1_{\{L_1 > 0, L_2 \leq 0\}} \\
\equiv \chi^2_{\frac{\gamma_u - \gamma_l}{\gamma_u + \gamma_l}} 1_{\{L_1 \leq 0, L_2 \leq 0\}} + \chi^2_{\frac{\gamma_l - \gamma_u}{\gamma_l + \gamma_u}} 1_{\{L_1 > 0, L_2 > 0\}} \\
+ \chi^2_{\frac{\gamma_u - \gamma_l}{\gamma_u + \gamma_l}} 1_{\{L_1 \leq 0, L_2 > 0\}} + \chi^2_{\frac{\gamma_u - \gamma_l}{\gamma_u + \gamma_l}} 1_{\{L_1 > 0, L_2 \leq 0\}}, \tag{27}
\end{equation}

where \( \Gamma_{tu} = (\gamma_l, \gamma_u)', \Gamma_{al} = (\gamma_u, \gamma_l)' \).

So the crucial issue is the position of random variables \( L_1, L_2 \).

Q.E.D.

**Proof of Theorem 1.**

a). Define as in Wu (1990)

\[
J(t) = P_s \left[ \frac{\sqrt{b}(S_b - \bar{S}_n)}{(1 - f) \text{var} S}^{1/2} \leq t \right]. \tag{28}
\]

Then since \( \Omega \) being finite and nonsingular, via Assumption 2, Corollary 1 of Zhao, Wang, Wu (2004)

\[
\sup_t |J(t) - \phi(t)| \to 0, \quad a.s. \tag{29}
\]

where \( \phi(t) \) is the standard normal distribution.

Rewrite \( \text{FAR}(\theta_0) \) as, where \( \text{var} S = \hat{\Omega} \),

\[
\text{FAR}(\theta_0) = \{\sqrt{b}(S_b - E_s \bar{S}_n + E_s S_b)^t(1 - f) \text{var} S)^{-1}[\sqrt{b}(S_b - E_s \bar{S}_n + E_s S_b)] \}. \tag{30}
\]

Then note that

\[
\sqrt{b} E_s S_b = \frac{\sqrt{b}}{\sqrt{n}} \left[ \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) \right] + \frac{\sqrt{b}}{\sqrt{n}} (\sqrt{n} EZ_i u_i) \\
= \frac{\sqrt{b}}{\sqrt{n}} \left[ \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i) \right] + \frac{\sqrt{b}}{\sqrt{n}} C. \tag{31}
\]
Next, using (28)(29)(31) in (30), $\sqrt{b} = \sqrt{f}\sqrt{n}$, by Assumption 1, Lindeberg Central Limit Theorem with $L \equiv N(0, \Omega)$, and since $J_b(t) = P_s(FAR(\theta_0) \leq t)$,

$$\sup_t |J_b(t) - \phi_{\chi^2}(t)| \to 0, \quad a.s.,$$

where $\phi_{mf}(t)$ is the cumulative distribution for the following limit:

$$[L + \frac{\sqrt{f}}{1-f}L + \frac{\sqrt{f}}{1-f}C]' \times \Omega^{-1}[L + \frac{\sqrt{f}}{1-f}L + \frac{\sqrt{f}}{1-f}C] = (1 + \frac{f}{1-f})^2 \chi_k^2 + 2 \left(1 + \frac{f}{1-f}\right) \sqrt{\frac{f}{1-f}} [C'\Omega^{-1}L] + \frac{f}{1-f}[C''\Omega^{-1}C].$$

Note that by using the left hand side term in (33) we can rewrite the limit in terms of a non-central $\chi^2$ distribution by combining the terms in $L$ and by dividing and multiplying with $(1 + \frac{\sqrt{f}}{1-f})^2$

$$[L + \frac{\sqrt{f}}{1-f}L + \frac{\sqrt{f}}{1-f}C]' \times \Omega^{-1}[L + \frac{\sqrt{f}}{1-f}L + \frac{\sqrt{f}}{1-f}C] = (1 + \frac{f}{1-f})^2 \chi_k^2 + \frac{f}{1+2\sqrt{f}\sqrt{1-f}} c^{\alpha-1}c.$$

In the last equivalence, note that we use

$$\left(\frac{\sqrt{f}}{1-f} \frac{1 + \frac{f}{1-f}}{1+2\sqrt{f}\sqrt{1-f}}\right)^2 = \frac{f}{1+2\sqrt{f}\sqrt{1-f}}.$$

In deriving the result above we use in (30)(E, $\bar{S}_b = \bar{S}_n$)

$$E_s[\bar{S}_b - E_s\bar{S}_b] = [E_s\bar{S}_b]^2 - [E_s\bar{S}_b]^2 = 0.$$

b). We prove the case of $k = 1$. The case for $k = 2$ is similar only takes into account the joint nature of the instruments and their behavior as in Lemma 1c. First note that (28)(29) is also true in the uniform case. The critical issue is to evaluate $\sup FAR(\theta_0)$. This depends on the behavior of the random variable $L$, as in Lemma 1b. Using (24),

$$\sup_{C \in \Gamma} FAR(\theta_0) = \sup_{C \in \Gamma} \left\{\sqrt{b}(\bar{S}_b - E_s\bar{S}_b) + \sqrt{f}\sqrt{n^{-1/2}} \sum_{i=1}^{n} (Z_iu_i - EZ_iu_i) + \sqrt{f}C\right\}$$

$$\times \left[(1-f)v\var S\right]^{-1} \times \left[\sqrt{b}(\bar{S}_b - E_s\bar{S}_b) + \sqrt{f}\sqrt{n^{-1/2}} \sum_{i=1}^{n} (Z_iu_i - EZ_iu_i) + \sqrt{f}C\right]\}. \quad \vdots \quad 36
Benefiting from the left hand side term in (33), and using Lindeberg Central Limit Theorem

$$\sup_{C \in \Gamma} \text{FAR}(\theta_0) \overset{d}{\rightarrow} \sup_{C \in \Gamma}[L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} C]\Omega^{-1}[L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} C].$$

The issue is then similar to the proof of Lemma 1b, whether $L + \frac{\sqrt{T}}{\sqrt{1-f}} L \leq 0$ or not. So when $L + \frac{\sqrt{T}}{\sqrt{1-f}} L \leq 0$, remembering $k = 1$

$$\sup_{C \in \Gamma}[L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} C] \times \Omega^{-1}[L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} C]
= [L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} \gamma L][L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} \gamma L]
= [1 + \frac{T}{f(1-f)} \chi_1^2 + 2[1 + \frac{T}{f(1-f)}] \sqrt{\gamma \Omega^{-1} L}] + \frac{T}{f(1-f)} \chi_1^2 \Omega^{-1}. \quad (37)$$

Furthermore we can express the limit in (37) as a certain non-central $\chi^2$ distribution. When $L + \frac{\sqrt{T}}{\sqrt{1-f}} L \leq 0$, using the left hand side of (37), and the analysis in (34) provides

$$[L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} \gamma L][L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} \gamma L]
= (1 + \frac{T}{f(1-f)} \chi_1^2 + 2[1 + \frac{T}{f(1-f)}] \sqrt{\gamma \Omega^{-1} L}) + \frac{T}{f(1-f)} \chi_1^2 \Omega^{-1}. \quad (38)$$

and for $L + \frac{\sqrt{T}}{\sqrt{1-f}} L > 0$

$$\sup_{C \in \Gamma}[L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} C] \times \Omega^{-1}[L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} C]
= [L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} \gamma L][L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} \gamma L]
= [1 + \frac{T}{f(1-f)} \chi_1^2 + 2[1 + \frac{T}{f(1-f)}] \sqrt{\gamma \Omega^{-1} L}] + \frac{T}{f(1-f)} \chi_1^2 \Omega^{-1}. \quad (39)$$

Next we can express the limit in (39) as a certain non-central $\chi^2$ distribution. When $L + \frac{\sqrt{T}}{\sqrt{1-f}} L > 0$, using the left hand side of (39), and the analysis in (34) provides

$$[L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} \gamma L][L + \frac{\sqrt{T}}{\sqrt{1-f}} L + \frac{\sqrt{T}}{\sqrt{1-f}} \gamma L]$$
\[ \begin{align*}
&= (1 + \frac{\sqrt{f}}{\sqrt{1 - f}})^2 [L + \frac{\sqrt{f}}{\sqrt{1 - f}} \gamma_u \Omega^{-1} [L + \frac{\sqrt{f}}{\sqrt{1 - f}} \gamma_u] \\
&\equiv (1 + \frac{\sqrt{f}}{\sqrt{1 - f}})^2 \chi^2_{1 + 2 \sqrt{f}/\sqrt{1 - f}, \gamma^2 \Omega^{-1}} \text{.} \quad (40)
\end{align*} \]

Defining \( J_H(t) = P_{\ast}(\sup_{C \in \Gamma} F A R(\theta_0) \leq t) \), and let \( \phi_{fiu}(t) \) represents the cumulative distribution function for addition of the limit in (37) multiplied by \( 1_{\{L + [\sqrt{f}/\sqrt{1 - f}] | L \leq 0\}} \) to the limit in (39) multiplied by \( 1_{\{L + [\sqrt{f}/\sqrt{1 - f}] | L > 0\}} \).

Next we see that (35) holds uniformly as well. So the result is proved for \( k = 1 \).

c). The proof for this case is the same as in proof of b), but taking into account joint nature of the normal random variables \((L_1, L_2)'\) as in the proof of Lemma 1c. The limit is the following:

\[ \begin{align*}
&\left[ \left( 1 + \frac{f}{1 - f} \right)^2 \chi^2_2 + 2 \left( 1 + \frac{f}{1 - f} \right) \left( \sqrt{\frac{f}{1 - f}} \right) \Gamma_{ul} \Omega^{-1} \Gamma_{ul} \right] L_1 \leq 0, L_2 + \sqrt{\frac{f}{1 - f}} L_2 \leq 0 \\
&+ \left[ \left( 1 + \frac{f}{1 - f} \right)^2 \chi^2_2 + 2 \left( 1 + \frac{f}{1 - f} \right) \left( \sqrt{\frac{f}{1 - f}} \right) \Gamma_{ul} \Omega^{-1} \Gamma_{ul} \right] L_1 > 0, L_2 + \sqrt{\frac{f}{1 - f}} L_2 > 0 \\
&+ \left[ \left( 1 + \frac{f}{1 - f} \right)^2 \chi^2_2 + 2 \left( 1 + \frac{f}{1 - f} \right) \left( \sqrt{\frac{f}{1 - f}} \right) \Gamma_{ul} \Omega^{-1} \Gamma_{ul} \right] L_1 \leq 0, L_2 + \sqrt{\frac{f}{1 - f}} L_2 > 0 \\
&+ \left[ \left( 1 + \frac{f}{1 - f} \right)^2 \chi^2_2 + 2 \left( 1 + \frac{f}{1 - f} \right) \left( \sqrt{\frac{f}{1 - f}} \right) \Gamma_{ul} \Omega^{-1} \Gamma_{ul} \right] L_1 > 0, L_2 + \sqrt{\frac{f}{1 - f}} L_2 \leq 0 \text{,}
\end{align*} \]

where \( \Gamma_{ul} = (\gamma_l, \gamma_u)', \Gamma_{ul} = (\gamma_u, \gamma_l)' \).

The notation in terms of non central \( \chi^2 \) distribution in the statement of Theorem 1c can be obtained as in (40) above. Q.E.D.

Remark. The only difference between the proof of iid case in Wu (1990) and the one here is the Berry-Esseen bounds. The iid case in Wu (1990) is satisfied under finite second moments as well. The extension to triangular arrays in (28)(29) can be done using Theorem 1 of Zhao, Wu and Wang (2004). The Berry-Esseen bounds are for a sample sum from a finite set of independent random variables as described in Zhao, Wu, Wang (2004).

References


Graduate School of Business, University of Chicago.


