Asymptotic Sizes of Subset Anderson-Rubin Tests with Weakly Identified Nuisance Parameters and General Covariance Structure

Joonhwan Lee

MIT

1 Introduction

Making inference on structural parameters in the linear instrumental variables (IVs) regression models has been one of the classic problems of econometrics. One of the biggest problems that many applications of the linear IV models encounter is that instruments are often weak, i.e. they are poorly correlated with the corresponding endogenous variables. Classical asymptotics has very bad finite sample behavior with weak instruments and thus the classical inference is practically unreliable. (See Stock, Wright and Yogo (2002)) Naturally, the problem of developing inference procedures that are robust to weak instruments has been one of the central questions of econometrics for the last 15 years.

There has been rich progress in constructing robust test statistics, most notably, the AR statistic by Anderson and Rubin (1949), the Lagrange multiplier (LM) statistic by Kleibergen (2002), and the conditional likelihood ratio (CLR) statistic by Moreira

(2003). An important shortcoming of the above methods is that they are designed to test only the simple full vector hypothesis in the form of $H_0 : \beta = \beta_0$ where β contains the coefficients for all the endogenous variables. Testing for a subset of parameters is not straightforward because the unrestricted structural parameters enter as additional nuisance parameters. Projection based tests are general solution for such problems but they are often very conservative especially when the number of dimensions projected out is large. When unrestricted structural parameters are strongly identified, the above test statistics can be adapted to have correct asymptotic size and improved power compared to projection based tests. See Stock and Wright (2000), Kleibergen (2004) and Guggenberger and Smith (2005) among others.

The problem of testing without any assumption on the identification of unrestricted structural parameters was a long standing question. Guggenberger et al (2012) provided an partial answer to this question. They showed that with a Kronecker product structure on a certain covariance matrix, the subset AR statistic with LIML (limited information maximum likelihood) estimator plugged in has the correct asymptotic size and power improvement over projection based tests. The Kronecker product structure, however, essentially implies the conditional homoskedasticity among reduced form disturbances. Thus, the result of Guggenberger et al (2012) is not practically useful because most economic data involve high degree of heteroskedasticity or serial correlation. Also, an important question arises: Will the result of Guggenberger et al (2012) hold with general covariance structure?

This paper provides an answer to this question by documenting a counter-example. I consider a reduced model of the linear IV model with normal disturbances and perform thorough simulations. It is shown that the result of Guggenberger et al (2012) breaks down in wide range of covariance structure if the Kronecker product assumption is removed. Moreover, it is demonstrated via simulation that the projection based tests have sharp asymptotic size. The range of covariance structure where the break-down is observed in this paper, however, necessarily imply serial correlation among reduced form disturbances. Thus, the implication of this paper may shed light on the weak identification robust inference procedures with times series data.

The paper is organized as follows. Section 2 briefly discusses the model and the problem of interest. Section 3 considers a simplification of the model to make it tractable for analysis and simulations. Section 4 reiterates the result of Guggenberger et al (2012) in the simplified model and show that the Kronecker structure is actually isomorphic to the identity matrix in the context of the test statistic. Section 5 and 6 discuss the counter example to their result with general covariance structure and thorough simulation results.

2 Linear Instrumental Regression Model and Weak IV

Hausman(1983) wrote that an IV regression model can be represented as limited information simultaneous equations model, in which we only specify single structural equation of interest. The full structured model is

$$\begin{aligned} y &= Y\beta + W\gamma + \epsilon \\ Y &= Z\Pi_Y + V_Y \\ W &= Z\Pi_W + V_W \end{aligned}$$

where y, Y, W are $T \times 1$, $T \times m_Y$, $T \times m_W$ matrices that contain endogenous variables. W is consisted of solely endogenous variables, while Y may contain some exogenous variables that is of interest. Z is a $T \times k$ matrix of instruments. We assume away any other included exogenous variable in the structural equation by regarding all the variables to be pre-multiplied by $M_X = I_T - X(X'X)^{-1}X'$, where X is a $T \times m_X$ matrix of exogenous variables that is not contained in Y. As usual, we assume that Z is a full rank with $k \geq m_Y + m_W$ to satisfy the rank condition. The hypothesis that we are interested in is

$$H_0 : \beta = \beta_0 \quad \text{v.s.} \quad H_1 : \beta \neq \beta_0.$$

With appropriate re-parametrization, we can also test a general linear restriction in this framework as well. If we have a test that have correct asymptotic size, we can construct a corresponding confidence interval of β by inverting the test. Under classic asymptotics when we have fixed full rank matrix of $[\Pi_Y \, \Pi_W]$ and sample size T increases to infinity, we can easily establish asymptotic normality of the estimator for β and γ. We can test these parameters or any function of them with conventional Wald (or t) statistics.

Testing the parameters under potential weak identification, that is when $[\Pi_Y \, \Pi_W]$ is close to degenerate along some direction, is problematic because usual asymptotic approximation does not work well even with very large T . Staiger and Stock (1997), among others, examine this problem by considering an alternative asymptotics where $[\Pi_Y \, \Pi_W]$ are changing with sample size T with order of $\frac{1}{\sqrt{T}}$. More recent works have focused on finding a set of robust tests that have asymptotically correct size under arbitrarily weak identification. These robust tests are tests based on Anderson-Rubin statistic (Anderson and Rubin, 1949), conditional likelihood ratio statistic (Moreira, 2003) and a Lagrange multiplier statistic (Kleibergen, 2002). The aforementioned statistics are known to have limiting distributions that do not depend on nuisance parameters when testing a hypothesis that contains the whole set of endogenous variables, in our case that would be $H_0 : \beta = \beta_0$, $\gamma = \gamma_0$.

Contrary to classic asymptotics, it is not straightforward to perform a test on a subset of parameters based on weak-instrument robust statistics. This is due to the fact that unrestricted structural parameters constitute additional nuisance parameters in the testing problem. In our model, the hypothesis of interest is

$$H_0 : \beta = \beta_0,$$

while allowing γ to be unrestricted. If γ is strongly identified, the robust tests above can be adapted by replacing γ with $\hat{\gamma}$, which is a consistent estimator of γ. Stock and Wright (2000) show that such modification of AR statistic in GMM setting provides a valid test. Kleibergen (2004) extends the result to CLR and LM statistic for a linear regression model. Guggenberger and Smith (2005) and Otsu (2006) address the similar issue in a more general GEL(Generalized Empirical Likelihood) framework.

Without the assumption of strong identification of γ, a natural approach is to apply projection type tests. See Dufour (1997) and Dufour and Taamouti (2005), among others. Projection test based on AR statistic can be described as follows. Consider AR statistic for both β and γ, $AR(\beta, \gamma)$. For testing the hypothesis $H_0 : \beta = \beta_0$, the projection test rejects the null when $AR(\beta_0, \gamma) > \chi^2(k)_{1-\alpha}$ for all values of γ. Thus, the corresponding test statistic is

$$AR(\beta_0) = \min_{\gamma \in \mathbb{R}^{m_W}} AR(\beta_0, \gamma)$$

The problem of the projection approach is that it does not provide efficient test if γ happens to be strongly identified, i.e. it has lower power than potentially optimal tests in some sense. One can note that level-α projected AR test uses $\chi^2(k)_{1-\alpha}$ as a critical value while the subset AR test with the assumption that γ_0 is strongly identified uses $\chi^2(k - m_W)_{1-\alpha}$ while having the same test statistic.

Guggenberger et al(2012) show that we can actually improve upon projection tests even with weakly identified γ. They show that under a Kronecker product covariance of (ϵ, V_W), that is

$$E\left[vec(Z_i U_i')\left((vec(Z_i U_i'))'\right] = E[U_i U_i'] \otimes E[Z_i Z_i']$$

where $U_i = (\epsilon_i, V_W')'$, the same subset AR test statistic $AR(\beta_0) = \min_{\gamma \in \mathbb{R}^{m_W}} AR(\beta_0, \gamma)$ has the limit distribution that is stochastically dominated by $\chi^2(k - m_W)$. Along with the fact that the limit distribution is exactly $\chi^2(k - m_W)$ with strong identification of γ, one can conclude that the test based on the subset AR statistic with critical value of $\chi^2(k - m_W)_{1-\alpha}$ has correct asymptotic size of α, and provides power improvements over the projected AR test.

The crucial Kronecker product assumption essentially corresponds to conditional homoskedasticity of $U_i = (\epsilon_i, V_W')'$. This can be very restrictive in many empirical applications, especially where the weak instrument robust procedures are widely used. For example, Kleibergen and Mavroeidis (2009) applied the subset AR test based on $\chi^2(k - m_W)$ critical value for times series data to make inference on New Keynesian Phillips curve. Presumably, the data has significant auto-correlation and conditional heteroskedasticity which are common in any time series data. They applied subset tests with AR statistic based on a conjecture that the power improvement over projection tests would hold in the case of general covariance structure. However, as pointed out later by Guggenberger et al (2012), subset LM test and CLR test does not give correct asymptotic size even under the conditional homoskedasticity and the positive result holds only for subset AR test with the Kronecker product covariance assumption. A question whether the result of Guggenberger et al (2012) holds in case of general covariance structure which allows heteroskedasticity and auto-correlation remains open. If this question has a negative answer, in other words, if the stochastic domination by

$\chi^2(k - m_W)$ does not hold with general covariance structure, then we are back to the lower power of projection tests unless we are willing to accept the very restrictive assumption of conditional homoskedasticitiy.

This paper tries to address the question by providing an counter-example and show that the stochastic domination by $\chi^2(k - m_W)$ breaks down in wide range of non-Kronecker covariance structure. Also, a thorough Monte-Carlo simulation experiment is done to examine the region of parameters that causes break-down of Guggenberger et al (2012)'s result.

3 Simplification of the Model

Here, I analyze the reduced form in case of fixed instruments, normal errors and a known covariance matrix. The model is canonical in the literature for several reasons. First, it provides a benchmark that allows simple exposition and finite sample analysis of the statistic of interest. Second, it is a ground for asymptotics of more general models. See Moreira(2003, 2009), Andrews, Moreira and Stock(2006) and Guggenberger et al(2012) among others. Since the purpose of the paper is to provide an counter-example along with a thorough Monte-Carlo study, the notations and definitions of the following benchmark model will be used in the remaining parts.

We can rewrite the model in reduced form as

$$\begin{pmatrix} y \\ Y \\ W \end{pmatrix} = \begin{pmatrix} Z\Pi_1 + U \\ Z\Pi_Y + V_Y \\ Z\Pi_W + V_W \end{pmatrix},$$

where $U = \epsilon + V_Y\beta_0 + V_W\gamma_0$ and $\Pi_1 = \Pi_Y\beta_0 + \Pi_W\gamma_0$. Since error terms are normal and instruments are fixed, the model can be reduced to

$$\begin{pmatrix} \hat{\Pi}_1 \\ vec(\hat{\Pi}_Y) \\ vec(\hat{\Pi}_W) \end{pmatrix} \sim N\left(\begin{pmatrix} \Pi_Y\beta_0 + \Pi_W\gamma_0 \\ vec(\Pi_Y) \\ vec(\Pi_W) \end{pmatrix}, \Sigma\right),$$

where

$$\Sigma = \left(I_{(1+m_Y+m_W)} \otimes (Z'Z)^{-1}\right) Var\left(vec\left(Z'(U\, V_Y\, V_W)\right)\right) \left(I_{(1+m_Y+m_W)} \otimes (Z'Z)^{-1}\right)$$

and $(\hat{\Pi}_1, \hat{\Pi}_Y, \hat{\Pi}_W)$ are OLS estimator of (Π_1, Π_Y, Π_W). Under the null hypothesis of $H_0 : \beta = \beta_0$, we can further concentrate the model by incorporating the information of true β. We have

$$\begin{pmatrix} \hat{\Pi}_1 - \hat{\Pi}_Y\beta_0 \\ vec(\hat{\Pi}_W) \end{pmatrix} \sim N\left(\begin{pmatrix} \Pi_W\gamma_0 \\ \Pi_W \end{pmatrix}, \tilde{\Sigma}\right).$$

Note that

$$\tilde{\Sigma} = \left(I_{(1+m_W)} \otimes (Z'Z)^{-1}\right) Var\left(vec\left(Z'(\tilde{U}\, V_W)\right)\right) \left(I_{(1+m_W)} \otimes (Z'Z)^{-1}\right),$$

where $\tilde{U} = \epsilon + V_W\gamma$. Thus, there is no need to specify the covariance structure between V_Y and other stochastic terms to analyze this model under the null. That is exactly why Guggenberger et al (2012) did not need Kronecker product assumption for terms involving V_Y. Note also that the reduced form covariance matrix $\tilde{\Sigma}$ can be consistently estimable and thus we can treat $\tilde{\Sigma}$ as known in asymptotic analysis of the model. Here, unknown parameters are Π_W and γ_0.

The statistic of interest is the subset AR test statistic defined as

$$AR(\beta_0) = \min_{\gamma} \min_{\Pi_W} \begin{pmatrix} \xi - \Pi_W \gamma \\ vec(\hat{\Pi}_W) - vec(\Pi_W) \end{pmatrix}' \tilde{\Sigma}^{-1} \begin{pmatrix} \xi - \Pi_W \gamma \\ vec(\hat{\Pi}_W) - vec(\Pi_W) \end{pmatrix},$$

where $\xi = \hat{\Pi}_1 - \hat{\Pi}_Y \beta_0$.

We can decompose $\tilde{\Sigma}^{-1} = Q'Q$ where Q can be represented as

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ 0 & Q_{22} \end{pmatrix}.$$

Such decomposition includes Cholesky decomposition, which makes Q an upper-diagonal matrix with positive diagonal entries. Then we can rewrite the statistic as

$$AR(\beta_0) = \min_{\gamma} \min_{\Phi} \begin{pmatrix} \tilde{\xi} - H(\gamma) vec(\Phi) \\ vec(\hat{\Phi}) - vec(\Phi) \end{pmatrix}' \begin{pmatrix} \tilde{\xi} - H(\gamma) vec(\Phi) \\ vec(\hat{\Phi}) - vec(\Phi) \end{pmatrix},$$

where

$$\begin{aligned} \tilde{\xi} &= Q_{11}\xi + Q_{12} vec(\hat{\Pi}_W); \\ \hat{\Phi} &= Q_{22} vec(\hat{\Pi}_W); \\ \Phi &= Q_{22} vec(\Pi_W); \\ H(\gamma) &= (\gamma' \otimes Q_{11}) Q_{22}^{-1} + Q_{12} Q_{22}^{-1}, \end{aligned}$$

and

$$\begin{pmatrix} \tilde{\xi} \\ vec(\hat{\Phi}) \end{pmatrix} \sim N\left(\begin{pmatrix} H(\gamma_0) vec(\Phi) \\ vec(\Phi) \end{pmatrix}, I_{k(1+m_W)} \right).$$

The parameter Φ can be straightforwardly concentrated out because the statistic is a quadratic function of $vec(\Phi)$ given value of γ. The first order condition with respect to $vec(\Phi)$ is,

$$2H(\gamma)'H(\gamma)vec(\Phi) - 2H(\gamma)'\tilde{\xi} - 2(vec(\hat{\Phi}) - vec(\Phi)) = 0,$$

which gives,

$$vec(\Phi^*) = (H(\gamma)'H(\gamma) + I_{km_W})^{-1}\left(H(\gamma)'\tilde{\xi} + vec(\hat{\Phi})\right).$$

Plugging in the optimal value of Φ, we have

$$\begin{aligned}
AR(\beta_0) &= \min_{\gamma} -\left(H(\gamma)'\tilde{\xi} + vec(\hat{\Phi})\right)'(H(\gamma)'H(\gamma) + I_{km_W})^{-1}\left(H(\gamma)'\tilde{\xi} + vec(\hat{\Phi})\right) + \left\|vec(\hat{\Phi})\right\|^2 + \left\|\tilde{\xi}\right\|^2 \\
&= \min_{\gamma}\left(\tilde{\xi} - H(\gamma)vec(\hat{\Phi})\right)'(H(\gamma)H(\gamma)' + I_k)^{-1}\left(\tilde{\xi} - H(\gamma)vec(\hat{\Phi})\right),
\end{aligned}$$

where the second equality uses the following decomposition of $I_{k(m_W+1)}$,

$$I_{k(m_W+1)} = \begin{pmatrix} H(\gamma)' & I_{km_W} \\ I_k & -H(\gamma) \end{pmatrix}'\begin{pmatrix} (H(\gamma)'H(\gamma) + I_{km_W})^{-1} & 0_{km_W} \\ 0_k & (H(\gamma)H(\gamma)' + I_k)^{-1} \end{pmatrix}\begin{pmatrix} H(\gamma)' & I_{km_W} \\ I_k & -H(\gamma) \end{pmatrix}$$

(footnote) Note that for any $m \times n$ matrix H, the following holds.

$$(I_n + H'H)^{-1} = I_n - H'(I_m + HH')^{-1}H.$$

4 Dominance of $AR(\beta_0)$ by $\chi^2(k - m_W)$ with Conditional Homoskedasticity

Here I briefly discuss the result and proof of Guggenberger et al (2012) using the concentration and simplification I developed above for the linear IV model with conditional homoskedasticity. Given the Kronecker product assumption (i.e. conditional homoskedasticity), we have

$$\tilde{\Sigma} = \Omega \otimes (Z'Z)^{-1},$$

where $\Omega = Var\left((U_i,\, V_{W,i})'\right)$. Let $P'P = \Omega^{-1}$ be the Cholesky decomposition of Ω^{-1}. Consider the following decomposition

$$\tilde{\Sigma}^{-1} = (P \otimes (Z'Z)^{\frac{1}{2}})'(P \otimes (Z'Z)^{\frac{1}{2}}).$$

Also, let

$$P = \begin{pmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{pmatrix},$$

where P_{11} is a scalar, P_{22} is a $m_W \times m_W$ matrix. One can note that in this case,

$$\begin{aligned} H(\gamma) &= (\gamma' P_{11} P_{22}^{-1} + P_{12} P_{22}^{-1}) \otimes I_k \equiv \tilde{\gamma}' \otimes I_k, \\ H(\gamma) vec(\Phi) &= \Phi \tilde{\gamma}, \end{aligned}$$

with a re-parametrization. The AR statistic can be written as

$$AR(\beta_0) = \min_{\tilde{\gamma}} \frac{\left\| \tilde{\xi} - \hat{\Phi}\tilde{\gamma} \right\|^2}{1 + \tilde{\gamma}'\tilde{\gamma}}.$$

This shows the Kronecker product assumption basically makes the model equivalent to a model from $\tilde{\Sigma} = I_{k(1+m_W)}$ with different set of parameters. That is, the class of model $\{(\gamma_0, \Pi_W, \tilde{\Sigma})|\tilde{\Sigma} = \Omega \otimes (Z'Z)^{-1}\}$ is equivalent to the class of model $\{(\tilde{\gamma}_0, \Phi, I_{k(m_W+1)})\}$ and we can can achieve significant reduction in dimensionality by assuming the Kronecker product structure. Now let us define

$$\begin{aligned}
\eta_1 &= \tilde{\xi} - \Phi\tilde{\gamma}_0, \\
\eta_2 &= \hat{\Phi} - \Phi.
\end{aligned}$$

The proof of the statement that the statistic $AR(\beta_0)$ is dominated by $\chi^2(k - m_W)$ in Guggenberger et al (2012), hinges on the following equivalence,

$$\min_{\tilde{\gamma}} \frac{\left\|\tilde{\xi} - \hat{\Phi}\tilde{\gamma}\right\|^2}{1 + \tilde{\gamma}'\tilde{\gamma}} = \min_{d_1, d_2} \frac{\left\|\epsilon_1 d_1 + \tilde{\Phi} d_2\right\|^2}{d_1^2 + d_2' d_2} \quad \text{s.t. } d_1^2 + d_2' d_2 = C,$$

where

$$\begin{aligned}
\epsilon_1 &= \frac{1}{\sqrt{1 + \tilde{\gamma}_0'\tilde{\gamma}_0}}(\eta_1 - \eta_2\tilde{\gamma}_0), \\
\epsilon_2 &= \frac{1}{\sqrt{1 + \tilde{\gamma}_0'\tilde{\gamma}_0}}(\eta_1\tilde{\gamma}_0' + \eta_2), \\
\tilde{\Phi} &= \sqrt{1 + \tilde{\gamma}_0'\tilde{\gamma}_0}\Phi + \epsilon_2,
\end{aligned}$$

and C is any positive number. By plugging in

$$d_1^* = 1, \quad d_2^* = -(\tilde{\Phi}'\tilde{\Phi})^{-1}\tilde{\Phi}'\epsilon_1,$$

we have

$$\min_{\tilde{\gamma}} \frac{\left\|\tilde{\xi} - \hat{\Phi}\tilde{\gamma}\right\|^2}{1+\tilde{\gamma}'\tilde{\gamma}} \leq \frac{\epsilon_1' \left(I_k - \tilde{\Phi}(\tilde{\Phi}'\tilde{\Phi})^{-1}\tilde{\Phi}'\right)\epsilon_1}{1+d_2^{*'}d_2^*} \leq \epsilon_1' \left(I_k - \tilde{\Phi}(\tilde{\Phi}'\tilde{\Phi})^{-1}\tilde{\Phi}'\right)\epsilon_1 \sim \chi^2(k-m_W),$$

since ϵ_1 and ϵ_2 are independent. Obviously, all the elements above is not feasible but that is not of concern here because we basically use them to show that there exists $\tilde{\gamma}*$ for every realization of η_1 and η_2 such that the criterion function evaluated at $\tilde{\gamma}^*$ be dominated by $\chi^2(k-m_W)$. That is sufficient to show that AR statistic is indeed dominated by $\chi^2(k-m_W)$.

Negative Result in General Covariance Case

The above result, however, does not hold generally with potential heteroskedasticity and auto-correlation. Here, I document a case where the above result breaks down in a case of general covariance structure. Since one counter-example is enough to prove the claim, I demonstrate such an example with simplest possible setting, i.e. with $m_W = 1$ and $k = 2$. I set parameters to be

$$\gamma_0 = 1 \quad \Phi = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \quad Q_1 \equiv Q_{11}Q_{22}^{-1} = I_2 \quad Q_2 \equiv Q_{12}Q_{22}^{-1} = 4\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}.$$

Note that the parameter setup seems to be not much a departure from the homoskedastic case except that $Q_{12}Q_{22}^{-1}$ is not a scalar multiple of I_2. Also, the value of Φ indicates that the strength of identification is very weak. As shown later in more thorough simulation experiments, the upward departure from $\chi^2(k-m_W)$ is more pronounced as Φ and γ_0 get closer to zero. I ran a Monte-Carlo simulation based on the concentrated and reduced model

$$\begin{pmatrix} \tilde{\xi} \\ \hat{\Phi} \end{pmatrix} \sim N\left(\begin{pmatrix} H(\gamma_0)\Phi \\ \Phi \end{pmatrix}, I_{2k} \right),$$

and examined the distribution of AR statistic

$$AR(\beta_0) = \min_{\gamma} \left(\tilde{\xi} - H(\gamma)\hat{\Phi} \right)' \left(H(\gamma)H(\gamma)' + I_k \right)^{-1} \left(\tilde{\xi} - H(\gamma)\hat{\Phi} \right).$$

Projection method guarantee that $AR(\beta_0)$ is dominated by $\chi^2(2)$, and the question we wish to address is whether $AR(\beta_0)$ is dominated by $\chi^2(1)$. Although the minimization involved is over just a single dimensional space, the criterion function has potentially many humps and thus guaranteeing global minimization for every simulation draw is not an easy task. I employed a Newton type algorithm with multiple starting values to do the task. Figure 1 shows the quantile function of $AR(\beta_0)$ along with that of $\chi^2(1)$ and $\chi^2(2)$. As we can see from Figure 1, there is a clear stochastic ordering in this case. The statistic $AR(\beta_0)$ is stochastically larger than $\chi^2(1)$ and thus using $\chi^2(1)$ critical values would clearly produce over rejection, i.e. size distortion.

5 Simulation Results

In this section, I report some Monte-Carlo simulation results with varying model parameters γ_0, Φ, Q_1 and Q_2 as well as the number of instruments k. For most part, I look into the case of $k = 2$ for simplicity of specification. As in the previous example,

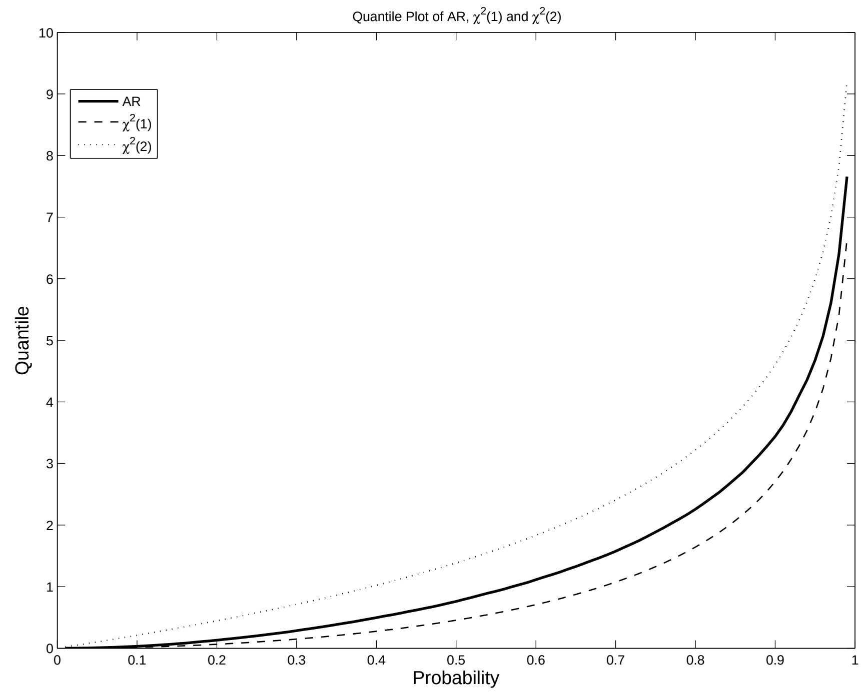

Figure 1: The case of AR dominating $\chi^2(k - m_W)$
100,000 simulation draws. Parameters are set to
$\gamma_0 = 1$, $\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$, $Q_1 = I_2$, $Q_2 = 4\begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}$. This case is with $k = 2$ and $m_W = 1$.

I set $Q_1 = I_2$ and consider the combinations of the following set of parameters.

$$\begin{aligned}
\gamma_0 &\in \{0,\ 0.1,\ 0.5,\ 1,\ 5,\ 20\},\\
\Phi &\in \{\lambda\mu | \mu = \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix},\ \lambda = 0,\ 0.1,\ 0.5,\ 1,\ 5,\ 20\},\\
Q_2 &\in \{\eta R | R = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix},\ \eta = 0,\ ,1,\ 4,\ 10\}.
\end{aligned}$$

Due to the nature of the $AR(\beta_0)$, the distribution of it is expected to be continuous in the change of the above parameters. The criterion function is a continuous function in those parameters. Since minimum is a continuous functional, one can argue that the distribution of $AR(\beta_0)$ and its functionals are also continuous in all the parameters. [1]

Before examining the simulation result, it may be helpful to see what the above parameter represent in terms of the original model of interest. The reduced form is written

$$\begin{pmatrix} y \\ Y \\ W \end{pmatrix} = \begin{pmatrix} Z\Pi_1 + U \\ Z\Pi_Y + V_Y \\ Z\Pi_W + V_W \end{pmatrix},$$

and in case of the model we consider for the simulation, where $k = 2$ and $m_W = 1$, under the null hypothesis of $H_0 : \beta = \beta_0$, the model is reduced to

$$\begin{pmatrix} \hat{\Pi}_1 - \hat{\Pi}_Y \beta_0 \\ \hat{\Pi}_W \end{pmatrix} \sim N\left(\begin{pmatrix} \Pi_W \gamma_0 \\ \Pi_W \end{pmatrix}, \tilde{\Sigma}\right).$$

If we assume that Z is normalized so that $Z'Z = I_k$, we have

$$\tilde{\Sigma} = Var\left(vec\left(Z'(\tilde{U}\, V_W)\right)\right),$$

where $\tilde{U} = U - \beta_0 V_Y$.

We can see how Q_1 and Q_2 are related to $\tilde{\Sigma}$, the reduced form covariance as follows. Since the model is a reduced version, there are infinitely many $\tilde{\Sigma}$ that corresponds to the same value of Q_1 and Q_2. If we normalize so that $\tilde{\Sigma}_{22} = I_2$, then we have

$$\tilde{\Sigma} = \begin{pmatrix} (Q_1'(I + Q_2 Q_2')^{-1} Q_1)^{-1} & -(Q_1' Q_1)^{-1} Q_1' Q_2 \\ -Q_2' Q_1 (Q_1' Q_1)^{-1} & I_2 \end{pmatrix} = \begin{pmatrix} (1+\eta^2) I_2 & -\eta R \\ -\eta R' & I_2 \end{pmatrix}.$$

Note that when Q_2 is zero matrix, that is when $\eta = 0$, then we have conditional homoskedasticity. Larger values of η means that the reduced form error of $Z'\tilde{U}$ has

[1] Note that the minima γ^* is not necessarily continuous in parameters.

much larger variance (at the order of η) than that of $Z'V_W$. Also, larger values of η translate into higher correlation, either negative or positive, between $Z_1'\tilde{U}$ and $Z_2'\tilde{V}_W$, and $Z_2'\tilde{U}$ and $Z_1'\tilde{V}_W$ because the correlation coefficients are

$$\pm\frac{\eta}{\sqrt{1+\eta^2}}.$$

Given the value of R in the simulation, we have opposite signs in those two correlations. The matrix R is chosen in this way because it produced the most amount of upward deviation from $\chi^2(k-m_W)$. However, a slight change in the value of R also generated such deviation and therefore the deviation is not a singularity, but occurs over a wide range of set of parameters. One can ask whether the structure of $\tilde{\Sigma}$ that generates the deviation is plausible or realistic in empirical application. This question will be addressed more thoroughly in the later sections.

Vector parameterΦ indicates the strength of identification because under conditional homoskedasticity with $\tilde{\Sigma} = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \otimes (Z'Z)^{-1}$, we have

$$\begin{aligned} \Phi &= Q_{22}\Pi_W = \frac{1}{\sigma_v}(Z'Z)^{\frac{1}{2}}\Pi_W, \\ ||\Phi||^2 &= \Pi_W' Z'Z\Pi_W/\sigma_v^2, \end{aligned}$$

i.e., the norm of Φ is the familiar concentration parameter for W under the assumption that $H_0 : \beta = \beta_0$ holds. See Stock, Wright and Yogo (2002) for the definition. In general case, we can note that

$$||\Phi||^2 = \Pi_W \tilde{\Sigma}_{22}^{-1}\Pi_W,$$

where $\tilde{\Sigma}_{22}$ is the variance-covariance matrix of $vec(\hat{\Phi})$. In case of $\Phi = 20\mu$ or $\Phi = 5\mu$ in this experiment setting, they correspond to the concentration parameter of 400 and 25, which are regarded as strong identification in most empirical literature in case of $k = 2$. The value of γ_0 also presumably affects the distribution of $AR(\beta_0)$. For conditional homoskedastic case, I showed that a model with $(\gamma, \Pi, \tilde{\Sigma})$ is equivalent to $(\tilde{\gamma}, \Phi, I)$ for some values of $\tilde{\gamma}$ and Φ in terms of the $AR(\beta_0)$ distribution. As shown above, the value of $\tilde{\gamma}$ is affected by the elements of $\tilde{\Sigma}$. The simulation results clearly indicates that the value ofγ_0 indeed affects the distribution of $AR(\beta_0)$ in a subtle manner.

The number of simulation draw is $50,000$ for each combination of parameters and I tabulate the 95% and 90% quantile values of empirical distribution of the simulated $AR(\beta_0)$, denoted by $AR95$ and $AR90$ respectively, along with the standard errors. Also, I report $P(AR(\beta_0) > \chi^2(1)_{1-\alpha})$ for $\alpha = 0.05$ and 0.1, which is the true size of the test based on $\chi^2(1)_{1-\alpha}$ critical values. Those values will show the size distortion when we apply the subset test in Guggenberger et al(2012) to a described model without the assumption of Kronecker product. Through the simulation experiment, I found that the behavior of $AR95$ or $AR90$ is a good representation of the behavior of the whole distribution of $AR(\beta_0)$. That is, at least in this experiment, when $AR95$ is above $\chi^2(1)_{0.95}$, then the distribution of $AR(\beta_0)$ stochastically dominates $\chi^2(1)$. Also, when $AR95$ converges to $\chi^2(1)_{0.95}$ in some change of parameter, then the distribution of $AR(\beta_0)$ is also found to converge to $\chi^2(1)$. Therefore, the simulation results can be interpreted accordingly.

The simulation results in Table 1 and Table 2 show some notable tendencies. First, as Φ is further apart from zero, the distribution of $AR(\beta_0)$ converge to $\chi^2(1)$. This has a natural explanation that $||\Phi||^2$ is the concentration parameter and larger value of it indicates strong identification. This is consistent with the result of Stock and Wright (2000) that we have the subset test statistic following $\chi^2(k - m_W)$ with strongly identified γ_0. However, the speed of convergence varies for different values of γ and

	$\eta = 0$						$\eta = 1$					
	$\gamma_0 = 0$	$\gamma_0 = 0.1$	$\gamma_0 = 0.5$	$\gamma_0 = 1$	$\gamma_0 = 5$	$\gamma_0 = 20$	$\gamma_0 = 0$	$\gamma_0 = 0.1$	$\gamma_0 = 0.5$	$\gamma_0 = 1$	$\gamma_0 = 5$	$\gamma_0 = 20$
$\lambda = 0$	1.67 (0.014)	1.68 (0.014)	1.68 (0.014)	1.66 (0.014)	1.70 (0.015)	1.68 (0.015)	3.42 (0.029)	3.37 (0.028)	3.41 (0.029)	3.37 (0.028)	3.40 (0.029)	3.37 (0.029)
$\lambda = 0.1$	1.69 (0.015)	1.67 (0.014)	1.69 (0.015)	1.66 (0.014)	1.76 (0.015)	2.70 (0.023)	3.37 (0.028)	3.40 (0.029)	3.37 (0.028)	3.42 (0.029)	3.39 (0.029)	3.42 (0.032)
$\lambda = 0.5$	1.80 (0.016)	1.80 (0.016)	1.81 (0.016)	1.89 (0.016)	3.08 (0.026)	3.80 (0.032)	3.44 (0.030)	3.41 (0.029)	3.42 (0.030)	3.42 (0.030)	3.37 (0.031)	3.79 (0.032)
$\lambda = 1$	2.06 (0.018)	2.07 (0.018)	2.12 (0.018)	2.33 (0.020)	3.70 (0.032)	3.81 (0.033)	3.42 (0.031)	3.43 (0.031)	3.39 (0.031)	3.36 (0.030)	3.40 (0.028)	3.79 (0.032)
$\lambda = 5$	3.69 (0.031)	3.67 (0.031)	3.78 (0.033)	3.82 (0.033)	3.85 (0.033)	3.88 (0.033)	3.61 (0.030)	3.67 (0.031)	3.63 (0.031)	3.72 (0.032)	3.79 (0.032)	3.78 (0.032)
$\lambda = 20$	3.82 (0.032)	3.84 (0.033)	3.84 (0.033)	3.81 (0.033)	3.81 (0.032)	3.86 (0.033)	3.80 (0.032)	3.78 (0.032)	3.79 (0.032)	3.80 (0.032)	3.85 (0.033)	3.84 (0.033)
	$\eta = 4$						$\eta = 10$					
	$\gamma_0 = 0$	$\gamma_0 = 0.1$	$\gamma_0 = 0.5$	$\gamma_0 = 1$	$\gamma_0 = 5$	$\gamma_0 = 20$	$\gamma_0 = 0$	$\gamma_0 = 0.1$	$\gamma_0 = 0.5$	$\gamma_0 = 1$	$\gamma_0 = 5$	$\gamma_0 = 20$
$\lambda = 0$	5.21 (0.035)	5.21 (0.035)	5.26 (0.036)	5.19 (0.036)	5.22 (0.036)	5.18 (0.036)	5.70 (0.038)	5.67 (0.037)	5.66 (0.037)	5.67 (0.037)	5.69 (0.038)	5.66 (0.037)
$\lambda = 0.1$	5.24 (0.036)	5.20 (0.036)	5.17 (0.035)	5.21 (0.036)	5.15 (0.036)	4.82 (0.035)	5.59 (0.037)	5.62 (0.037)	5.63 (0.037)	5.66 (0.038)	5.59 (0.037)	5.44 (0.036)
$\lambda = 0.5$	4.74 (0.034)	4.79 (0.035)	4.78 (0.035)	4.87 (0.036)	4.54 (0.036)	3.56 (0.033)	5.06 (0.037)	4.92 (0.035)	5.03 (0.037)	5.03 (0.037)	4.88 (0.036)	4.29 (0.037)
$\lambda = 1$	4.31 (0.036)	4.35 (0.036)	4.30 (0.036)	4.29 (0.036)	3.90 (0.036)	3.66 (0.031)	4.35 (0.036)	4.40 (0.036)	4.41 (0.036)	4.32 (0.035)	4.31 (0.037)	3.66 (0.035)
$\lambda = 5$	3.61 (0.030)	3.61 (0.030)	3.61 (0.031)	3.60 (0.030)	3.82 (0.033)	3.83 (0.032)	3.60 (0.030)	3.59 (0.030)	3.62 (0.030)	3.58 (0.030)	3.69 (0.031)	3.83 (0.033)
$\lambda = 20$	3.83 (0.033)	3.81 (0.032)	3.85 (0.033)	3.80 (0.033)	3.87 (0.033)	3.84 (0.033)	3.89 (0.033)	3.83 (0.033)	3.86 (0.033)	3.87 (0.033)	3.81 (0.033)	3.78 (0.032)

Table 1: Simulation Results of $AR95$ for $k = 2$, $m_W = 1$
95% quantile of $\chi^2(1)$ is 3.84 and that of $\chi^2(2)$ is 5.99. The result is from 50,000 simulation draws for each configuration.

Q_2. Most notably, higher values of γ_0 seems to be highly correlated with the speed of convergence. Figure 2 shows how $AR95$ changes in the identification strength of γ_0. In homoskedastic case where $Q_2 = 0$, we can observe monotonic increase of $AR95$ until it converges to $\chi^2(k-1)_{0.95}$. In general covariance case where $Q_2 = 4R$, we can observe $AR95$ is well above $\chi^2(1)_{0.95}$ with weaker identification and then decreases way below $\chi^2(1)_{0.95}$, and finally increase monotonically while converging. In both cases, the speed of convergence is faster when the value of γ_0 is larger. We can also note that the speed of convergence is much slower in case of $Q_2 = 4R$.

	$\eta = 0$						$\eta = 1$					
	$\gamma_0 = 0$	$\gamma_0 = 0.1$	$\gamma_0 = 0.5$	$\gamma_0 = 1$	$\gamma_0 = 5$	$\gamma_0 = 20$	$\gamma_0 = 0$	$\gamma_0 = 0.1$	$\gamma_0 = 0.5$	$\gamma_0 = 1$	$\gamma_0 = 5$	$\gamma_0 = 20$
$\lambda = 0$	1.17 (0.009)	1.18 (0.009)	1.17 (0.009)	1.15 (0.009)	1.18 (0.009)	1.17 (0.009)	2.44 (0.020)	2.42 (0.019)	2.43 (0.020)	2.41 (0.019)	2.43 (0.019)	2.38 (0.019)
$\lambda = 0.1$	1.19 (0.010)	1.17 (0.009)	1.18 (0.010)	1.17 (0.009)	1.23 (0.010)	1.92 (0.015)	2.41 (0.019)	2.43 (0.019)	2.41 (0.019)	2.42 (0.019)	2.41 (0.019)	2.33 (0.019)
$\lambda = 0.5$	1.24 (0.010)	1.26 (0.010)	1.25 (0.010)	1.32 (0.011)	2.19 (0.017)	2.69 (0.021)	2.44 (0.020)	2.40 (0.019)	2.43 (0.020)	2.41 (0.020)	2.29 (0.019)	2.65 (0.021)
$\lambda = 1$	1.43 (0.011)	1.43 (0.011)	1.48 (0.012)	1.62 (0.013)	2.58 (0.020)	2.69 (0.021)	2.35 (0.019)	2.37 (0.020)	2.33 (0.019)	2.32 (0.019)	2.42 (0.019)	2.70 (0.022)
$\lambda = 5$	2.59 (0.020)	2.58 (0.020)	2.64 (0.021)	2.66 (0.021)	2.68 (0.021)	2.71 (0.021)	2.57 (0.020)	2.59 (0.021)	2.53 (0.020)	2.64 (0.021)	2.70 (0.021)	2.68 (0.021)
$\lambda = 20$	2.71 (0.021)	2.70 (0.021)	2.68 (0.021)	2.69 (0.021)	2.67 (0.021)	2.71 (0.021)	2.66 (0.021)	2.67 (0.021)	2.69 (0.021)	2.69 (0.021)	2.72 (0.022)	2.72 (0.022)
	$\eta = 4$						$\eta = 10$					
	$\gamma_0 = 0$	$\gamma_0 = 0.1$	$\gamma_0 = 0.5$	$\gamma_0 = 1$	$\gamma_0 = 5$	$\gamma_0 = 20$	$\gamma_0 = 0$	$\gamma_0 = 0.1$	$\gamma_0 = 0.5$	$\gamma_0 = 1$	$\gamma_0 = 5$	$\gamma_0 = 20$
$\lambda = 0$	3.95 (0.026)	3.95 (0.026)	3.95 (0.026)	3.92 (0.025)	3.93 (0.026)	3.90 (0.025)	4.34 (0.026)	4.31 (0.026)	4.34 (0.026)	4.35 (0.027)	4.33 (0.026)	4.35 (0.027)
$\lambda = 0.1$	3.93 (0.026)	3.92 (0.026)	3.94 (0.026)	3.94 (0.026)	3.86 (0.025)	3.59 (0.025)	4.29 (0.027)	4.28 (0.026)	4.30 (0.026)	4.28 (0.026)	4.27 (0.026)	4.13 (0.026)
$\lambda = 0.5$	3.52 (0.025)	3.57 (0.025)	3.56 (0.025)	3.54 (0.025)	3.27 (0.025)	2.40 (0.020)	3.74 (0.026)	3.71 (0.026)	3.74 (0.026)	3.71 (0.026)	3.58 (0.025)	3.05 (0.025)
$\lambda = 1$	3.09 (0.025)	3.12 (0.025)	3.11 (0.025)	3.07 (0.025)	2.71 (0.023)	2.58 (0.020)	3.13 (0.025)	3.14 (0.025)	3.17 (0.025)	3.12 (0.025)	3.04 (0.024)	2.48 (0.021)
$\lambda = 5$	2.56 (0.020)	2.57 (0.020)	2.57 (0.020)	2.54 (0.020)	2.66 (0.021)	2.70 (0.021)	2.57 (0.020)	2.55 (0.020)	2.56 (0.020)	2.57 (0.020)	2.60 (0.021)	2.66 (0.021)
$\lambda = 20$	2.70 (0.021)	2.68 (0.021)	2.68 (0.021)	2.65 (0.021)	2.70 (0.021)	2.71 (0.022)	2.75 (0.022)	2.70 (0.021)	2.71 (0.022)	2.73 (0.022)	2.69 (0.021)	2.68 (0.021)

Table 2: Simulation Results of $AR90$ for $k = 2$, $m_W = 1$

90% quantile of $\chi^2(1)$ is 2.71 and that of $\chi^2(2)$ is 4.61. The result is from 50,000 simulation draws for each configuration. Standard error from simulation in parenthesis.

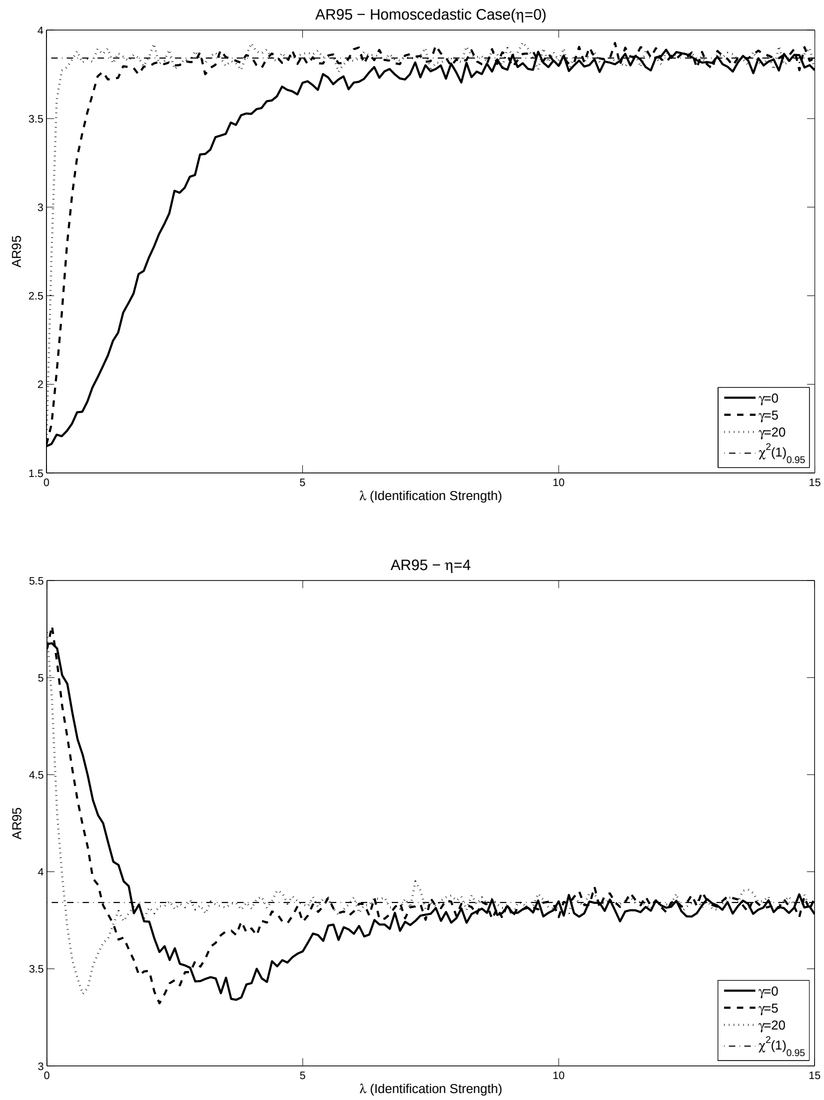

Figure 2: Change of $AR95$ in Identification Strength
For $\eta = 0$ (Conditional homoskedastic case) and $\eta = 4$ (Non-Kronecker case). 50,000 simulation draws for each configuration.

Break-down of the dominance by $\chi^2(k - m_W)$ is observed in the weak identification region of Φ and it is more pronounced when η is larger. Figure 3 shows that $AR(\beta_0)$ can indeed get very close to $\chi^2(2)$ when η is sufficiently large and there is no identification. This example demonstrate that at least with $k = 2$ and $m_W = 1$, the asymptotic size of the subset AR test based on critical value of the projection AR test sharply equals

α if we consider every possible $\tilde{\Sigma}$, not just a class of $\tilde{\Sigma}$'s that have Kronecker product structure. In fact, in the next section, it is demonstrated that this is generally true for $k > 2$ and $m_W > 1$. These results show that applying the critical value of $\chi^2(k - m_W)_{1-\alpha}$ in empirical applications may generate significant size distortion when we have weakly identified γ_0 and high degree of heteroskedasticity and auto-correlation. Even in the most severe case of break-down in this setup, the upward deviation from $\chi^2(1)$ seems to disappear when $\lambda \geq 2$ which can be translated to concentration parameter of 4. This seems pretty small number but obviously one cannot use such threshold to decide whether we are safe to use the $\chi^2(k - m_W)$ critical values in empirical applications.

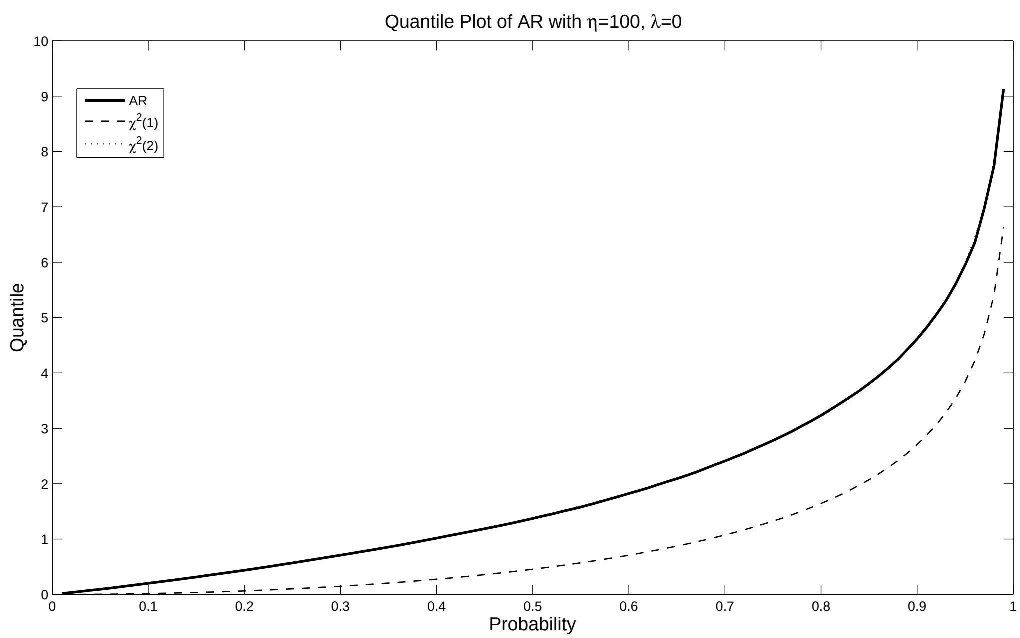

Figure 3: Case of AR being close to $\chi^2(k)$

Note that the value of γ_0 does not matter here because we set $\lambda = 0$ so that $\Phi = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

6 Simulation of More General Setting

The previous section discussed the simulation results for $k = 2$ and Q_2 of scalar multiple of $R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. One may ask if we can observe the break-down of dominance by $\chi^2(k - m_W)$ in wider range of parameters. The answer is positive both for $Q \neq \eta R$ and $k > 2$. This section is devoted to exploration of the region of parameters that generates $AR(\beta_0)$ distribution dominating $\chi^2(k - m_W)$ and discussion of whether the region of parameter is plausible in empirical applications.

First, I consider Q_2 that is not a scalar multiple of R in case of $k = 2$. Instead, I consider a class of Q_2,

$$Q_2 \in \left\{ \eta R(\theta) | \eta \in \mathbb{R},\, \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}],\, R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \right\},$$

that is, Q_2 is a scalar multiple of a rotation matrix in $\mathbb{R}^2$. The class contains the values of Q_2 used in the previous section as a special case of $\theta = \frac{\pi}{2}$. Figure 4 shows how $AR95$ changes over different values of θ, where other parameters are set to $\{Q_1 = I_2,\, \Phi = 0.1\mu,\, \gamma_0 = 1,\, \eta \in \{2, 8, 20\}\}$. Interestingly, $AR95$ is increasing to the rotation angle and it takes maximum at $\theta = \pm\frac{\pi}{2}$. We can note that when η is sufficiently large, we can observe $AR95 > \chi^2(1)_{0.95}$ for wide range of θ. For the case of $\eta = 2$, $|\theta| > \frac{\pi}{3}$. It should be noted that there exists Q_2 that produces AR statistic dominating $\chi^2(1)$ outside the class considered here. The class of scalar multiple of rotation matrices is considered just for convenience of characterization. In fact, any Q_2 with complex eigenvalues whose imaginary parts are sufficiently large could generate AR statistic that dominates $\chi^2(1)$. The exact mechanic of this, however, could not be clearly described analytically.

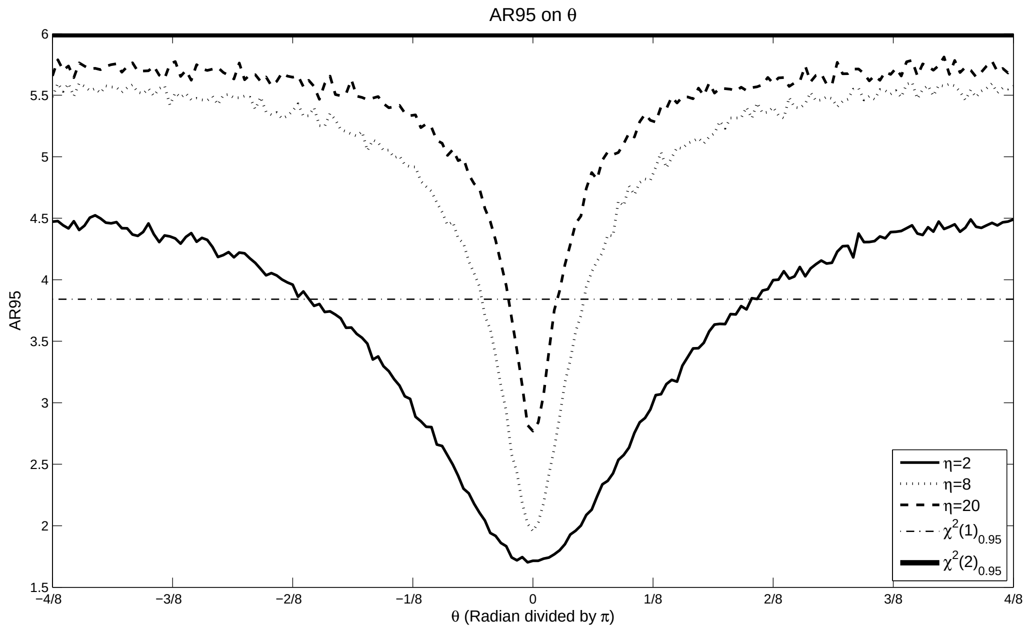

Figure 4: Change of $AR95$ on θ
50,000 simulation draws with increment of $\pi/180$ of θ.

Second, I consider the case of $k > 2$ and $k = 2l$, i.e. k is even with l being a positive integer. Then an obvious extension of Q_2 with $\eta R(\theta)$ is

$$Q_2 = I_l \otimes \eta R(\theta),$$

where $\eta = 20$, $\theta = \frac{\pi}{2}$ and and 0_k is a $k \times 1$ vector of zeros. The other parameters are set as

$$Q_1 = I_k, \quad \gamma_0 = 0, \quad \Phi = 0_k.$$

The Table 3 shows the simulation results for $k \in \{4, 6, 10, 30\}$. We can see that both $AR95$ and $AR90$ exceeds the value of $\chi^2(k-1)_{0.95}$ and $\chi^2(k-1)_{0.9}$ by a significant margin in all cases.. Note that Q_2's in these cases have quite a sparse structure when k is larger, and they still generate the $AR(\beta_0)$ statistic that stochastically dominates $\chi^2(k-1)$.

	$\chi^2(k-1)_{0.90}$	$AR90$	$\chi^2(k)_{0.90}$	$\chi^2(k-1)_{0.95}$	$AR95$	$\chi^2(k)_{0.95}$
$k=4$	6.25	7.56 (0.023)	7.78	7.81	9.28 (0.033)	9.49
$k=6$	9.24	10.42 (0.027)	10.64	11.07	12.36 (0.037)	12.59
$k=10$	14.68	15.75 (0.033)	15.99	16.92	18.05 (0.044)	18.31
$k=30$	39.09	39.93 (0.051)	40.26	42.56	43.45 (0.067)	43.77

Table 3: $AR95$ and $AR90$ in case of $k=2l$

For $k=2l+1$, it is also possible to observe a breakdown of $\chi^2(k-1)$ dominance. By setting the parameters as

$$Q_1 = \begin{pmatrix} \epsilon & 0'_{k-1} \\ 0_{k-1} & I_{k-1} \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 1 & 0'_{k-1} \\ 0_{k-1} & I_l \otimes \eta R(\theta) \end{pmatrix}, \quad \gamma_0 = 0, \quad \Phi = 0_k,$$

where $\eta = 20$ and $\theta = \frac{\pi}{2}$. The table 4 shows the simulation results for $k \in \{3, 7, 11, 31\}$ with $\epsilon = 0.0001$. I could not observe a case with Q_1 not being close to singular matrix for $k = 2l+1$. When k is a odd number, I was not able to find any Q_2 that generates $AR > \chi^2(k-1)$ with $Q_1 = I_k$ among 5000 randomly generated Q_2's. Such cases was only observed with Q_1 being near singular. Q_1 being above with very small ϵ implies that one of the instruments has nearly no information.

The whole region of parameters where I found the break-down of $\chi^2(k-m_W)$ dominance, however, implies some degree of auto-correlation in errors. If we only have conditional heteroskedasticity without any auto-correlation, we have

$$\begin{aligned} \tilde{\Sigma} &= \left(I_{(1+m_W)} \otimes (Z'Z)^{-1}\right) Var\left(vec\left(Z'\omega\right)\right) \left(I_{(1+m_W)} \otimes (Z'Z)^{-1}\right) \\ &= \left(I_{(1+m_W)} \otimes (Z'Z)^{-1}\right) E[E[\omega_i \omega_i' | z_i] \otimes (z_i z_i')] \left(I_{(1+m_W)} \otimes (Z'Z)^{-1}\right), \end{aligned}$$

	$\chi^2(k-1)_{0.90}$	$AR90$	$\chi^2(k)_{0.90}$	$\chi^2(k-1)_{0.95}$	$AR95$	$\chi^2(k)_{0.95}$
$k=3$	4.61	5.83 (0.022)	6.25	5.99	7.37 (0.030)	7.81
$k=7$	10.64	11.65 (0.029)	12.02	12.59	13.69 (0.039)	14.07
$k=11$	15.99	16.85 (0.034)	17.28	18.31	19.21 (0.045)	19.68
$k=31$	40.26	41.02 (0.052)	41.42	43.77	44.53 (0.067)	44.99

Table 4: $AR95$ and $AR90$ in case of $k = 2l + 1$

which implies all $k \times k$ blocks that consist $\tilde{\Sigma}$ should be symmetric. It can be shown that the parameters values where I documented the stochastic dominance of AR statistic over $\chi^2(k - m_W)$ are not feasible under the block-wise symmetry of $\tilde{\Sigma}$. This does not necessarily imply that AR statistic is dominated by $\chi^2(k - m_W)$ when we have only conditional heteroskedasticity. As Guggenberger et al(2012) noted, proving the result analytically is not an easy feat. Thus, the findings in this paper have something to say for applying subset AR test for data prone to auto-correlation, e.g. time series data. A notable application is inference on New-Keynesian Phillips curve as in Kleibergen and Mavroeidis (2009).

7 Conclusion

Reducing the degree of freedom for testing a subset of parameters with weak identification robust test statistic and weakly identified nuisance parameters has been a challenging problem in the literature. Such dimension reduction is important in practical perspective because the efficiency of a test can be substantially improved. Recent work of Guggenberger et al (2012) showed that subset Anderson-Rubin statistic can be applied with reduced degree of freedom with conditional homoskedasticity, or Kronecker product structure. This paper is first to document that the result of Guggenberger et al (2012) does not hold in general covariance structure. With a thorough simulation study, I show that the projection based tests have sharp asymptotic size and this cannot be improved without further assumptions on the covariance structure. Also, it is shown that the break down of the result is most pronounced when the identification of nuisance parameters is weak.

The region of parameters that I found the break down, however, necessarily imply there is some degree of auto-correlation in errors. This leaves an important question: can we have the dimension reduction with only conditional heteroskedasticity? The simulation results suggest that the answer might be positive. Theoretical proof may be a daunting task, but it will definitely allow subset AR test to be applied to much wider range of problems.

References

[1] Anderson, Theodore W., and Herman Rubin. "Estimation of the parameters of a single equation in a complete system of stochastic equations." The Annals of Mathematical Statistics 20.1 (1949): 46-63.

[2] Andrews, Donald WK, Marcelo J. Moreira, and James H. Stock. "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression." Econometrica 74.3 (2006): 715-752.

[3] Guggenberger, Patrik, and Richard J. Smith. "Generalized empirical likelihood estimators and tests under partial, weak, and strong identification." Econometric Theory 21.04 (2005): 667-709.

[4] Guggenberger, Patrik, et al. "On the asymptotic sizes of subset Anderson–Rubin and Lagrange multiplier tests in linear instrumental variables regression." Econometrica 80.6 (2012): 2649-2666.

[5] Dufour, Jean Marie, "Some impossibility theorems in econometrics with applications to structural and dynamic models." Econometrica 65.6 (1997): 1365-1387.

[6] Dufour, Jean-Marie, and Mohamed Taamouti. "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments." Econometrica 73.4 (2005): 1351-1365.

[7] Hausman, Jerry A. "Specification and estimation of simultaneous equation models." Handbook of econometrics 1.1 (1983): 391-448.

[8] Kleibergen, Frank. "Pivotal statistics for testing structural parameters in instrumental variables regression." Econometrica 70.5 (2002): 1781-1803.

[9] Kleibergen, Frank. "Testing parameters in GMM without assuming that they are identified." Econometrica 73.4 (2005): 1103-1123.

[10] Kleibergen, Frank, and Sophocles Mavroeidis. "Weak instrument robust tests in GMM and the new Keynesian Phillips curve." Journal of Business & Economic Statistics 27.3 (2009): 293-311.

[11] Moreira, Marcelo J. "A conditional likelihood ratio test for structural models." Econometrica 71.4 (2003): 1027-1048.

[12] Moreira, Marcelo J. Tests with correct size when instruments can be arbitrarily weak. Center for Labor Economics, University of California, Berkeley, 2001.

[13] Otsu, Taisuke. "Generalized empirical likelihood inference for nonlinear and time series models under weak identification." Econometric Theory 22.03 (2006): 513-527.

[14] Staiger, Douglas, and James H. Stock. "Instrumental Variables Regression with Weak Instruments." Econometrica 65.3 (1997): 557-586.

[15] Stock, James H., and Jonathan H. Wright. "GMM with weak identification." Econometrica 68.5 (2000): 1055-1096.

[16] Stock, James H., Jonathan H. Wright, and Motohiro Yogo. "A survey of weak instruments and weak identification in generalized method of moments." Journal of Business & Economic Statistics 20.4 (2002).