A Robust Test for Weak Instruments

José Luis Montiel Olea
Harvard University
Cambridge, MA 02138
E-mail: montiel@fas.harvard.edu

Carolin Pflueger
University of British Columbia
Vancouver, BC V6T 1Z2, Canada
carolin.pflueger@sauder.ubc.ca.

April 2013

Abstract

We develop a test for weak instruments in linear instrumental variables regression that is robust to heteroskedasticity, autocorrelation, and clustering. Our test statistic is a scaled non-robust first stage F statistic. Instruments are considered weak when the Two-Stage Least Squares (TSLS) or the Limited Information Maximum Likelihood (LIML) Nagar bias is large relative to a benchmark. We apply our procedures to the estimation of the Elasticity of Intertemporal Substitution, where our test cannot reject the null of weak instruments in a larger number of countries than the test proposed in Stock and Yogo (2005).

Keywords: F Statistic; Heteroskedasticity; Autocorrelation; Clustered; Elasticity of Intertemporal Substitution.

1. INTRODUCTION

This paper proposes a simple test for weak instruments that is robust to heteroskedasticity, serial correlation, and clustering. Staiger and Stock (1997) and Stock and Yogo (2005) developed widely used tests for weak instruments under the assumption of conditionally homoskedastic serially uncorrelated model errors. However, applications with heteroskedasticity, time series autocorrelation, and clustered panel data are common. Our proposed test provides empirical researchers with a new tool to assess instrument strength for those applications.

The practical relevance of heteroskedasticity in linear instrumental variable (IV) regression has been by highlighted before by Antoine and Lavergne (2012), Chao and Newey (2012) and Hausman (2012). We show, more generally, that departures from the conditionally homoskedastic serially uncorrelated framework affect the weak instrument asymptotic distribution of both the Two-Stage Least Squares (TSLS) and the Limited Information Maximum Likelihood (LIML) estimators. Consequently, heteroskedasticity, autocorrelation, and/or clustering can further bias estimators and distort test sizes when instruments are potentially weak. At the same time, the first stage may falsely indicate that instruments are strong.

Under strong instruments, both TSLS and LIML are asymptotically unbiased, while such is generally not the case when instruments are weak. We follow the standard Nagar (1959) methodology to derive a tractable proxy for the asymptotic estimator bias that is defined for both TSLS and LIML. Our procedure tests the null hypothesis that the Nagar bias is large relative to a "worst-case" benchmark. Our benchmark coincides with the Ordinary Least Squares (OLS) bias benchmark when the model errors are conditionally homoskedastic and serially uncorrelated, but differs otherwise.

Our proposed test statistic, which we call the effective F statistic, is a scaled version of the non-robust first stage F statistic. The null hypothesis for weak instruments is rejected for large values of the effective F. The critical values depend on an estimate of the covariance matrix of the OLS reduced form regression coefficients, and on the covariance matrix of the reduced form errors, which can be estimated using standard procedures.

We consider two different testing procedures, generalized and simplified; both are asymptotically valid. Critical values for both procedures can be calculated either by Monte-Carlo methods, or by a curve-fitting methodology by Patnaik (1949). The generalized testing procedure applies to both TSLS and LIML, and has increased power, but is computationally more demanding. In contrast, the simplified procedure applies only to TSLS. The simplified procedure is conservative, because it protects against the worst type of heteroskedasticity, serial correlation, and/or clustering in the second stage.

Empirical researchers frequently report the robust F statistic as a simple way of adjusting the Staiger and Stock (1997) and Stock and Yogo (2005) pre-tests for heteroskedasticity, serial correlation, and clustering, and compare them to the homoskedastic critical values. To the best of our knowledge, there is no theoretical or analytical support for this practice, as cautioned in Baum et al. (2007). Our proposed procedures adjust the critical values. While our proposed test statistic corresponds to the robust F statistic in the just identified case, it differs in the over-identified case.

Our baseline implementation tests the null hypothesis that the Nagar bias exceeds 10% of a"worst-case" bias with a size of 5%. The simplified procedure for TSLS has critical values between 11 and 23.1 that depend only on the covariance matrix of the first stage reduced form coefficients. Thus a simple, asymptotically valid rule of thumb is available for TSLS that rejects when the effective F is greater than 23.1.

We apply weak instrument pre-tests to a well-known empirical example, the IV estimation of the Elasticity of Intertemporal Substitution (EIS) (Yogo, 2004; Campbell, 2003). Our empirical results are consistent with Yogo (2004)'s finding that the EIS is small and close to zero. However, for several countries in our sample, conditionally homoskedastic serially uncorrelated pre-tests indicate strong instruments, while our proposed test cannot reject the null hypothesis of weak instruments.

There is a large literature on inference when IVs are weak; see Stock et al. (2002) and Andrews and Stock (2006) for overviews. Our paper is closest to Staiger and Stock (1997) and Stock and Yogo (2005). Zhan (2010) provides another interesting approach, which, unlike ours, proposes to test the null hypothesis of strong instruments. Bun and de Haan (2010) point out the invalidity of pre-tests based on the first stage F statistic in two particular examples of non-homoskedastic and serially correlated errors, but do not provide a valid pre-test.

Robust methods for inference about the coefficients of a single endogenous regressor when IVs are weak and errors are heteroskedastic and/or serially correlated are also available (Andrews and Stock, 2006; Kleibergen, 2007). A pre-test for weak instruments followed by standard inference procedures can be less computationally demanding, and the use of this two-stage decision rule is widespread because of its simplicity. We therefore view this paper as complementary to robust inference methods.

It is well-known that pre-tests can induce uniformity problems (Leeb and Poetscher, 2005; Guggenberger, 2010*a*,*b*). However, Stock and Yogo (2005) have shown that in the conditionally homoskedastic and serially uncorrelated case the first stage F statistic can be used to control the Wald test size distortion. In this case, uniformity problems are therefore not a first order concern.

The rest of the paper is organized as follows. Section 2 introduces the model, and presents the generalized and simplified testing procedures. Section 3 derives asymptotic distributions, and shows that conditional heteroskedasticity and serial correlation can effectively weaken instruments in an illustrative example. Section 4 derives the expressions for the TSLS and LIML Nagar biases, and describes the test statistic and critical values. Section 5 discusses the implementation of the critical values by Monte Carlo simulation and Patnaik (1949)'s methodology. Section 6 applies the pre-testing procedure to the IV estimation of the EIS. Section 7 concludes. All proofs are collected in Appendix A.

2. MODEL AND SUMMARY OF TESTING PROCEDURE

2.1 Model and Assumptions

We consider a linear IV model in reduced form with one endogenous regressor and K instruments

$$\mathbf{y} = \mathbf{Z}\mathbf{\Pi}\beta + \mathbf{v}_1 \quad (1)$$

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{v}_2 \quad (2)$$

The structural parameter of interest is $\beta \in \mathbb{R}$, while $\mathbf{\Pi} \in \mathbb{R}^K$ denotes the unknown first stage parameter vector. The sample size is S and the econometrician observes the data set $\{y_s, Y_s, \mathbf{Z}_s\}_{s=1}^S$. We denote observations of the outcome variable, the endogenous regressor, and the vector of instruments by y_s, Y_s and $\mathbf{Z}_s$, respectively. The unobserved reduced form errors have realizations v_{js}, $j \in \{1, 2\}$. We stack the realized variables in matrices $\mathbf{y} \in \mathbb{R}^S$, $\mathbf{Z} \in \mathbb{R}^{S \times K}$, and $\mathbf{v}_j \in \mathbb{R}^S$, $j \in \{1, 2\}$.

Our analysis extends straightforwardly to a model with additional exogenous regressors. In the presence of additional exogenous regressors, TSLS and LIML estimators are unchanged if we replace all variables by their projection errors onto those exogenous regressors. TSLS and LIML are also invariant to normalizing the instruments to be orthonormal. We can therefore assume without loss of generality that there are no additional exogenous regressors, and that $\mathbf{Z}'\mathbf{Z}/S = \mathbb{I}_K$. When implementing the pre-test, an applied researcher needs to normalize the data.

We model weak instruments by assuming that the IV first stage relation is local to zero, following the modeling strategy in Staiger and Stock (1997).

Assumption $\mathbf{L}_\Pi$. (**Local to Zero**) $\mathbf{\Pi} = \mathbf{\Pi}_S = \mathbf{C}/\sqrt{S}$, where $\mathbf{C}$ is a fixed vector $\mathbf{C} \in \mathbb{R}^K$.

Additional high-level assumptions allow us to derive asymptotic distributions for IV estimators and F statistics. TSLS and LIML estimators and first stage F statistics depend on the statistics $\mathbf{Z}'\mathbf{v}_j/\sqrt{S}$, and estimates of the covariance matrices $\mathbf{W}$ and $\mathbf{\Omega}$ as defined below.

Assumption HL. (**High Level**) The following limits hold as $S \to \infty$.

1. $\begin{pmatrix} \mathbf{Z}'\mathbf{v}_1/\sqrt{S} \\ \mathbf{Z}'\mathbf{v}_2/\sqrt{S} \end{pmatrix} \xrightarrow{d} \mathcal{N}_{2K}(\mathbf{0}, \mathbf{W})$ for some positive definite $\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_{12} \\ \mathbf{W}_{12}' & \mathbf{W}_2 \end{pmatrix}$
2. $[\mathbf{v}_1, \mathbf{v}_2]'[\mathbf{v}_1, \mathbf{v}_2]/S \xrightarrow{p} \mathbf{\Omega}$ for some positive definite $\mathbf{\Omega} \equiv \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix}$
3. There exists a sequence of positive definite estimates $\{\widehat{\mathbf{W}}(S)\}$, measurable with respect to $\{y_s, Y_s, \mathbf{Z}_s\}_{s=1}^S$, such that $\widehat{\mathbf{W}}(S) \xrightarrow{p} \mathbf{W}$ as $S \to \infty$

Assumption HL is satisfied under various primitive conditions on the joint distribution of $(\mathbf{Z}, \mathbf{v}_1, \mathbf{v}_2)$; see Supplementary Materials C.2 for examples. Assumption HL.1 is satisfied as long as a Central Limit Theorem holds for $\mathbf{Z}'\mathbf{v}_j/\sqrt{S}$. Assumption HL.2 holds under a Weak Law of Large Numbers for $[\mathbf{v}_1, \mathbf{v}_2]'[\mathbf{v}_1, \mathbf{v}_2]/S$. Assumption HL.3 assumes that we can consistently estimate the covariance matrix $\mathbf{W}$ from the observable variables.

Assumption HL allows for a general form of $\mathbf{W}$, similarly to the models in Müller (2011) and Mikusheva (2010). This is our key generalization from the model in Staiger and Stock (1997), who require $\mathbf{W}$ to have the form $\mathbf{\Omega} \otimes I_K$. The Kronecker form arises naturally only in the context of a conditionally homoskedastic serially uncorrelated model. Our generalization is therefore relevant for practitioners working with heteroskedastic, time series, or panel data, and it is consequential for econometric practice.

2.2 Implementing the Testing Procedure

2.2.1 Generalized Test

The generalized testing procedure can be implemented in four simple steps. When rejecting the null, the empirical researcher can conclude that the estimator Nagar bias is small relative to the benchmark. Under the null hypothesis, the Nagar bias of TSLS or LIML is greater than a fraction τ of the benchmark. Critical values for the effective F statistic depend on the desired threshold τ, the desired level of significance α, and estimates for the matrices $\widehat{\mathbf{W}}$, $\widehat{\mathbf{\Omega}}$. Critical values also vary between TSLS and LIML. In our numerical results, we focus on $\tau = 10\%$ and $\alpha = 5\%$.

1. If there are additional exogenous regressors, replace all variables by their projection residuals onto those exogenous regressors. Normalize instruments to be orthonormal.

2. Obtain $\widehat{\mathbf{W}}$ as the estimate for the asymptotic covariance matrix of the reduced form OLS coefficients. Standard statistical packages estimate this matrix (divided by the sample size S) under different distributional assumptions. For cross-sectionally heteroskedastic applications, use a heteroskedasticity robust estimate; for time series applications, use a heteroskedasticity and autocorrelation consistent (HAC) estimate; and for panel data applications, use a "clustered" estimate.

3. Compute the test statistic, the *Effective F Statistic*

$$\widehat{F}_{eff} \equiv \frac{1}{S} \frac{\mathbf{Y}'\mathbf{Z}\mathbf{Z}'\mathbf{Y}}{tr\left(\widehat{\mathbf{W}}_2\right)} \tag{3}$$

   where $tr(\cdot)$ denotes the trace operator.

4. Estimate the *effective degrees of freedom*

$$\widehat{K}_{eff} \equiv \frac{\left[tr\left(\widehat{\mathbf{W}}_2\right)\right]^2(1+2x)}{tr\left(\widehat{\mathbf{W}}_2'\widehat{\mathbf{W}}_2\right)+2xtr\left(\widehat{\mathbf{W}}_2\right)\max eval(\widehat{\mathbf{W}}_2)} \tag{4}$$

$$\text{where } x = B_e(\widehat{\mathbf{W}},\widehat{\mathbf{\Omega}})/\tau \text{ for } e \in \{\text{TSLS}, \text{LIML}\} \tag{5}$$

Here, $\max eval(\widehat{\mathbf{W}}_2)$ denotes the maximum eigenvalue of the lower diagonal $K \times K$ block of the matrix $\widehat{\mathbf{W}}$. The function $B_e(\widehat{\mathbf{W}},\widehat{\mathbf{\Omega}})$ is closely related to the supremum of the Nagar bias relative to the benchmark; see Theorem 1.2. The numerical implementation of $B_e(\widehat{\mathbf{W}},\widehat{\mathbf{\Omega}})$ is discussed in Remark 5, Theorem 1. A fast numerical MATLAB routine is available for the function $B_e(\widehat{\mathbf{W}},\widehat{\mathbf{\Omega}})$.

The generalized test rejects the null hypothesis of weak instruments when $\widehat{F}_{eff}$ exceeds a critical value that can be obtained by either of the following procedures:

a) Monte Carlo methods, as described in Section 5;

b) Patnaik (1949)'s curve-fitting methodology; Patnaik critical values obtain as the upper α quantile of $\chi^2_{\widehat{K}_{eff}}\left(x\widehat{K}_{eff}\right)/\widehat{K}_{eff}$, where $\chi^2_{\widehat{K}_{eff}}\left(x\widehat{K}_{eff}\right)$ denotes a noncentral χ^2 distribution with $\widehat{K}_{eff}$ degrees of freedom and noncentrality parameter $x\widehat{K}_{eff}$. Table 1 tabulates 5% Patnaik critical values.

2.2.2 Simplified Test

A simplified conservative version of the test is available for TSLS. The simplified procedure follows the same steps, but sets $x = 1/\tau$ in Step 4. For a given effective degrees of freedom $\widehat{K}_{eff}$, the simplified 5% critical value can be conveniently read off Table 1. For instance, the critical value for a threshold $\tau = 10\%$ can be found in the column with $x = 10$. The simplified test does not require numerical evaluation of $B_e(\widehat{\mathbf{W}},\widehat{\mathbf{\Omega}})$, for it uses the bound

Table 1
Critical Values:
Upper 5% Quantile of $\chi^2_{K_{eff}}(xK_{eff})/K_{eff}$

K_{eff}	$x = 3.33$	$x = 5$	$x = 10$	$x = 20$
1	12.05	15.06	23.11	37.42
2	9.57	12.17	19.29	32.32
3	8.53	10.95	17.67	30.13
4	7.92	10.23	16.72	28.85
5	7.51	9.75	16.08	27.98
6	7.21	9.40	15.62	27.35
7	6.98	9.14	15.26	26.86
8	6.80	8.92	14.97	26.47
9	6.65	8.74	14.73	26.15
10	6.52	8.59	14.53	25.87
11	6.41	8.47	14.36	25.64
12	6.32	8.36	14.21	25.44
13	6.24	8.26	14.08	25.26
14	6.16	8.17	13.96	25.10
15	6.10	8.10	13.86	24.96
16	6.04	8.03	13.77	24.83
17	5.99	7.96	13.68	24.71
18	5.94	7.91	13.60	24.60
19	5.89	7.85	13.53	24.50
20	5.85	7.80	13.46	24.41
21	5.81	7.76	13.40	24.33
22	5.78	7.72	13.35	24.25
23	5.74	7.68	13.29	24.18
24	5.71	7.64	13.24	24.11
25	5.68	7.61	13.20	24.05
26	5.66	7.57	13.15	23.98
27	5.63	7.54	13.11	23.93
28	5.61	7.51	13.07	23.87
29	5.58	7.49	13.04	23.82
30	5.56	7.46	13.00	23.77

NOTE: Critical values computed by Patnaik (1949) method. For generalized and simplified testing procedures, estimate K_{eff} as in (4). For a Nagar bias threshold τ (e.g. $\tau = 10\%$) set $x = 1/\tau$ for the simplified procedure. For the generalized procedure, set $x = B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau$; see Step 4 in Section 2.2.1.

$B_{TSLS}(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}}) \leq 1$, proved in Theorem 1.3. The matrix $\widehat{\mathbf{W}}$ enters only through the lower $K \times K$ block $\widehat{\mathbf{W}}_2$.

2.2.3 Comparison with Stock and Yogo (05) Critical Values

We compare the generalized and simplified TSLS critical values to those proposed by Stock and Yogo (2005) for the case when the data is conditionally homoskedastic and serially uncorrelated. For this comparison, we assume $\mathbf{W} = \mathbf{\Omega} \otimes \mathbb{I}_K$ and $\mathbf{W}$ and $\mathbf{\Omega}$ known, as in Stock and Yogo (2005). It then follows from (3) and (4) that the effective and non-robust F statistics are equal, and that the effective number of degrees of freedom K_{eff} equals the number of instruments K.

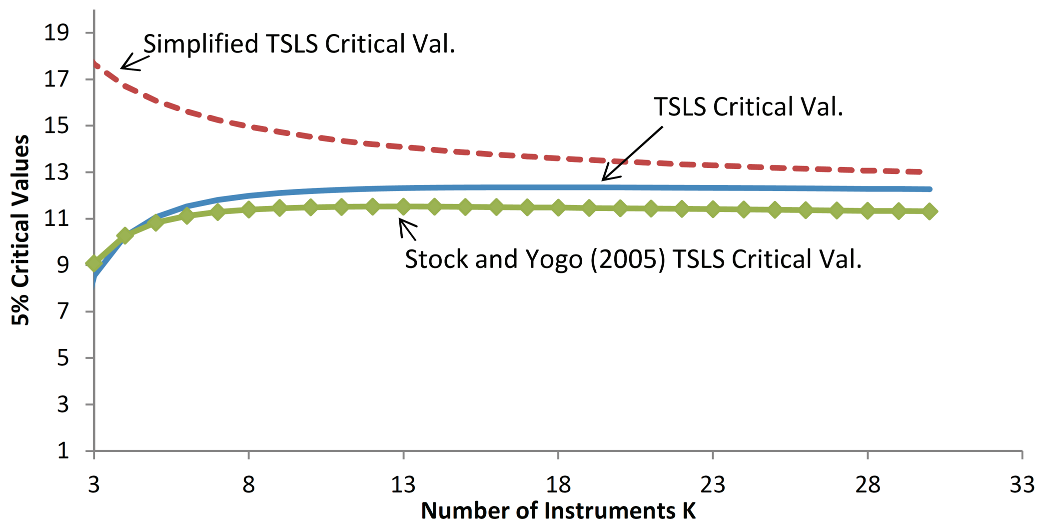

Figure 1 shows the 5% TSLS critical value for testing the null hypothesis that the asymptotic estimator bias exceeds 10% of the benchmark, the 5% critical value for the corresponding simplified test, and the Stock and Yogo (2005) 5% critical value for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias. The Stock and Yogo (2005) critical value is defined when the degree of over identification is at least two and we therefore show critical values for $3 \leq K \leq 30$. The TSLS critical value increases from 8.53 for $K = 3$ to 12.27 for $K = 30$. By comparison, the Stock and Yogo (2005) critical value increases from 9.08 for

$K = 3$ to 11.32 for $K = 30$. The simplified TSLS critical value is strictly larger than the TSLS critical value for all K shown, illustrating that the simplified test can be strictly less powerful than the generalized procedure. The difference between the simplified critical value and the TSLS and Stock and Yogo (2005) critical values decreases as K becomes large.

3. ASYMPTOTIC DISTRIBUTIONS AND AN EXAMPLE

3.1 Illustrative Example

A simple example illustrates that heteroskedasticity and serial correlation impact the entire asymptotic distribution of both TSLS and LIML estimators, and can weaken the performance of the estimators. In this example, the first stage F statistic rejects the null hypothesis of weak instruments too often, while the effective F statistic allows for testing for weak instruments with asymptotically correct size.

For the sake of exposition, assume $\beta = 0$. Also assume that the departure from the conditionally homoskedastic serially uncorrelated framework takes the particularly simple form

$$\mathbf{W} = a^2(\mathbf{\Omega} \otimes I_K) \tag{6}$$

a is a scalar parameter and for $a = 1$ the expression (6) reduces to the conditionally homoskedastic case.

Remark 1. We can generate example (6) with a purely conditionally heteroskedastic data-generating process. Let $\{\mathbf{Z}_s, \tilde{v}_{1s}, \tilde{v}_{2s}\}$ identically and independently distributed (i.i.d.). Let instruments independent with $\mathbb{E}[Z_{ks}] = 0$, $\mathbb{E}[Z_{ks}^2] = 1$, $\mathbb{E}[Z_{ks}^3] = 0$, $\mathbb{E}[Z_{ks}^4] = a^2$. Let $(\tilde{v}_{1s}, \tilde{v}_{2s}) \sim N_2\left((0,0)', \mathbf{\Omega}\right)$ independently of $\mathbf{Z}_s$. Let the reduced form errors $v_{1s} = \tilde{v}_{1s}\Pi_{k=1}^{K} Z_{ks}$, $v_{2s} = \tilde{v}_{2s}\Pi_{k=1}^{K} Z_{ks}$. Then $\mathbb{E}\left([v_{1s}, v_{2s}]\,[v_{1s}, v_{2s}]'\right) = \mathbf{\Omega}$ and $\mathbb{E}\left([v_{1s}, v_{2s}]\,[v_{1s}, v_{2s}]' \otimes \mathbf{Z}_s\mathbf{Z}_s'\right) = a^2\mathbf{\Omega} \otimes \mathbb{I}_K$. HL.1, HL.2, and (6) follow from the Central Limit Theorem and the Weak Law of Large Numbers.

Remark 2. We can alternatively generate (6) with a simple serially correlated data-generating process. Assume that instruments and reduced form errors follow independent AR(1) processes $Z_{ks+1} = \rho_Z Z_{ks} + \epsilon_{ks+1}, k = 1, ...K$ and $v_{js+1} = \rho_v v_{js} + \eta_{js+1}, j = 1, 2$. Let ϵ_{ks} and η_{js} serially uncorrelated with mean zero, $\mathbb{E}(\epsilon_s \epsilon_s') = (1-\rho_Z^2) \times I_K$ and $\mathbb{E}[\eta_{1s}, \eta_{2s}]'[\eta_{1s}, \eta_{2s}] = (1-\rho_V^2) \times \mathbf{\Omega}$. Then $\mathbb{E}\left[v_{1s}, v_{2s}\right]\left[v_{1s}, v_{2s}\right]' = \mathbf{\Omega}$ and $\mathbb{E}\left(\mathbf{Z}_s \mathbf{Z}_s'\right) = \mathbb{I}_K$. HL.1, HL.2 follow from the Central Limit Theorem and the Weak Law of Large Numbers. Expression (6) holds with $a = (1+\rho_v\rho_Z)/(1-\rho_Z\rho_v)$. Serial correlation in both the instruments and the errors is required for $a \neq 1$. As a numerical example, moderate serial correlation of $\rho_v = \rho_Z = 0.5$ gives rise to $a = 1.67$.

With Assumptions L_Π and HL the asymptotic distribution of the TSLS estimator

$$\widehat{\beta}_{TSLS} \equiv \left[\mathbf{Y}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Y}\right]^{-1}\mathbf{Y}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{v}_1 \tag{7}$$

$$= \frac{\omega_1}{\omega_2}\left[\left(\frac{\mathbf{C}}{a\omega_2} + \frac{\mathbf{Z}'\mathbf{v}_2/\sqrt{S}}{a\omega_2}\right)'\left(\frac{\mathbf{C}}{a\omega_2} + \frac{\mathbf{Z}'\mathbf{v}_2/\sqrt{S}}{a\omega_2}\right)\right]^{-1} \tag{8}$$

$$\times \left(\frac{\mathbf{C}}{a\omega_2} + \frac{\mathbf{Z}'\mathbf{v}_2/\sqrt{S}}{a\omega_2}\right)'\frac{\mathbf{Z}'\mathbf{v}_1/\sqrt{S}}{a\omega_1} \tag{9}$$

$$\xrightarrow{d} \frac{\omega_1}{\omega_2}\left[\boldsymbol{\psi}_2'\boldsymbol{\psi}_2\right]^{-1}\boldsymbol{\psi}_2'\boldsymbol{\psi}_1 \tag{10}$$

where $\begin{pmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \end{pmatrix} \sim \mathcal{N}_{2K}\left(\begin{pmatrix} \mathbf{0}_K \\ \mathbf{C}/(a\,\omega_2) \end{pmatrix}, \begin{pmatrix} 1 & \omega_{12}/(\omega_1\omega_2) \\ \omega_{12}/(\omega_1\omega_2) & 1 \end{pmatrix} \otimes \mathbb{I}_K\right)$.

The asymptotic TSLS distribution depends only on the elements of the non-central Wishart matrix $[\boldsymbol{\psi}_1, \boldsymbol{\psi}_2]'[\boldsymbol{\psi}_1, \boldsymbol{\psi}_2]$ Hence, the vector of first stage coefficients $\mathbf{C}$ and the parameter a enter into the asymptotic distribution in (10) only through the noncentrality parameter $\mathbf{C}'\mathbf{C}/a^2\omega_2^2$, so $\mathbf{C}'\mathbf{C}/a^2\omega_2^2$ summarizes instrument strength.

In this example, heteroskedasticity and serial correlation affect the biases and test size distortion of TSLS and LIML estimators in the same way as a weaker first stage relationship.

The conditionally homoskedastic serially uncorrelated case obtains for $a = 1$, so the TSLS estimator is asymptotically distributed as if the errors were conditionally homoskedastic serially uncorrelated, and the first stage coefficients were reduced by a factor of a. We prove an analogous result for LIML in Appendix A.

Consider a null hypothesis for weak instruments of the form $(\mathbf{C}'\mathbf{C}/\omega_2^2 a^2 K) < x$. In the presence of conditional heteroskedasticity or serial correlation of the form (6), the first stage F statistic is asymptotically distributed as $a^2\chi^2_K\left(\mathbf{C}'\mathbf{C}/\omega_2^2 a^2\right)/K$. As a increases without bound, the noncentrality parameter goes to zero and instruments become arbitrarily weak, but the first stage F statistic diverges to infinity almost surely. On the other hand, the effective F statistic is asymptotically distributed as a $\chi^2_K\left(\mathbf{C}'\mathbf{C}/\omega_v^2 a^2\right)/K$, so we can reject the null hypothesis of weak instruments with confidence level α whenever $\widehat{F}_{eff}$ exceeds the upper α quantile of $\chi^2_K\left(x \times K\right)/K$.

3.2. Asymptotic Distributions

Definition 1. Denote the projection matrix onto $\mathbf{Z}$ by $\mathbf{P_Z} = \mathbf{ZZ}'/S$ and the complementary matrix by $\mathbf{M_Z} = \mathbb{I}_S - \mathbf{P_Z}$.

1. The *Two-Stage Least Squares* (TSLS) estimator

$$\widehat{\beta}_{TSLS} \equiv (\mathbf{Y}'\mathbf{P_Z}\mathbf{Y})^{-1}(\mathbf{Y}'\mathbf{P_Z}\mathbf{y}) \tag{11}$$

2. The *Limited Information Likelihood* (LIML) estimator

$$\widehat{\beta}_{LIML} = (\mathbf{Y}'(\mathbb{I}_S - k_{LIML}\mathbf{M_Z})\mathbf{Y})^{-1}(\mathbf{Y}'(\mathbb{I}_S - k_{LIML}\mathbf{M_Z})\mathbf{y}) \tag{12}$$

where k_{LIML} is the smallest root of the determinantal equation

$$\left| [\mathbf{y},\mathbf{Y}]'[\mathbf{y},\mathbf{Y}] - k[\mathbf{y},\mathbf{Y}]'\mathbf{M_Z}[\mathbf{y},\mathbf{Y}] \right| = 0 \tag{13}$$

3. The *non-robust first stage F statistic*

$$\widehat{F} \equiv \frac{\mathbf{Y}'\mathbf{P_Z}\mathbf{Y}}{K\,\widehat{\omega}_2^2} \tag{14}$$

   where $\widehat{\omega}_2^2 \equiv \frac{(\mathbf{Y}-\mathbf{P_Z}\mathbf{Y})'(\mathbf{Y}-\mathbf{P_Z}\mathbf{Y})}{S-K-1}$

4. The *robust first stage F statistic*

$$\widehat{F}_r \equiv \frac{\mathbf{Y}'\mathbf{Z}\widehat{\mathbf{W}}_2^{-1}\mathbf{Z}'\mathbf{Y}}{K \times S} \tag{15}$$

   where $\widehat{\mathbf{W}}_2$ is the lower diagonal $K \times K$ block of the matrix $\widehat{\mathbf{W}}$.

5. The *effective first stage F statistic*

$$\widehat{F}_{eff} \equiv \frac{\mathbf{Y}'\mathbf{P_Z}\mathbf{Y}}{tr\left(\widehat{\mathbf{W}}_2\right)} \tag{16}$$

Lemma 1 derives asymptotic distributions for these statistics, generalizing Theorem 1 in Staiger and Stock (1997).

Lemma 1. *Write* $\sigma_1^2 = \omega_1^2 - 2\beta\omega_{12} + \beta^2\omega_2^2$, $\sigma_{12} = \omega_{12} - \beta\omega_2^2$, $\sigma_2^2 = \omega_2^2$ *and* $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$. *Under Assumptions* L_Π *and* HL *the following limits hold jointly as* $S \to \infty$.

1. $\widehat{\beta}_{TSLS} - \beta \xrightarrow{d} \beta^*_{TSLS} = (\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2)^{-1}\boldsymbol{\gamma}_2'(\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2)$

2. $\widehat{\beta}_{LIML} - \beta \xrightarrow{d} \beta^*_{LIML} = (\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2 - \kappa_{LIML}\omega_2^2)^{-1}(\boldsymbol{\gamma}_2'(\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2) - \kappa_{LIML}(\omega_{12} - \beta\omega_2^2))$

   where κ_{LIML} is the smallest root of $|[\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_2]'[\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2, \boldsymbol{\gamma}_2] - \kappa\mathbf{\Sigma}| = 0$

3. $\widehat{F} \xrightarrow{d} F^* \equiv \boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2/K\omega_2^2$

4. $\widehat{F}_r \xrightarrow{d} F_r^* \equiv \boldsymbol{\gamma}_2{}'\mathbf{W}_2^{-1}\boldsymbol{\gamma}_2/K$

5. $\widehat{F}_{eff} \xrightarrow{d} F_{eff}^* \equiv \boldsymbol{\gamma}_2{}'\boldsymbol{\gamma}_2/tr(\mathbf{W}_2)$

Where

$$\begin{pmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{pmatrix} \sim \mathcal{N}_{2K}\left(\begin{pmatrix} \beta\mathbf{C} \\ \mathbf{C} \end{pmatrix}, \mathbf{W}\right) \tag{17}$$

Proof. See Appendix A. □

The limiting distributions are functions of a multivariate normal vector whose distribution depends on the parameters $(\beta, \mathbf{C})$, and on the matrix W. We treat the asymptotic distributions in Lemma 1 as a limiting experiment in the sense of Müller (2011), and use it to analyze inference problems regarding $(\beta, \mathbf{C})$.

4. TESTING THE NULL HYPOTHESIS OF WEAK INSTRUMENTS

We base our null hypothesis of weak instruments on a bias criterion. We follow the standard methodology in Nagar (1959), and approximate the asymptotic TSLS and LIML distributions to obtain the Nagar bias. Under standard asymptotics, the Nagar bias for both estimators is zero everywhere in the parameter space, but under weak instrument asymptotics, the bias may be large in some regions of the parameter space. We consider instruments to be weak when the estimator Nagar bias is large relative to a benchmark, extending the OLS bias criterion in Stock and Yogo (2005).

4.1 Nagar Approximation

Theorem 1. ***(Nagar Approximation)*** *Let* $\mathbf{W} \in \mathbb{R}^{2K\times 2K}$ *positive definite. Write* $\mathbf{C} \in \mathbb{R}^K$ *as* $\mathbf{C} = \|\mathbf{C}\|\,\mathbf{C}_0$ *and let* $\mu^2 \equiv ||\mathbf{C}||^2/tr(\mathbf{W}_2)$. *Define* $\mathbf{S}_1 = \mathbf{W}_1 - 2\beta\mathbf{W}_{12} + \beta^2\mathbf{W}_2$, $\mathbf{S}_{12} =$

$\mathbf{W}_{12} - \beta \mathbf{W}_2$, $\mathbf{S}_2 = \mathbf{W}_2$ *and the benchmark* $BM(\beta, \mathbf{W}) \equiv \sqrt{tr(\mathbf{S}_1)/tr(\mathbf{S}_2)}$. *We write* $\mathcal{S}^{K-1}$ *for the* $K-1$ *dimensional unit sphere.*

1. *For* $e \in \{TSLS, LIML\}$ *the Taylor expansion of* β_e^* *around* $\mu^{-1} = 0$ *gives the Nagar (1959) bias*

$$N_e(\beta, \mathbf{C}, \mathbf{W}, \mathbf{\Omega}) = \mu^{-2} n_e(\beta, \mathbf{C}_0, \mathbf{W}, \mathbf{\Omega}) \tag{18}$$

with

$$n_{TSLS}(\beta, \mathbf{C}_0, \mathbf{W}, \mathbf{\Omega}) = \frac{tr(\mathbf{S}_{12})}{tr(\mathbf{S}_2)} \left[1 - 2\frac{\mathbf{C}_0' \mathbf{S}_{12} \mathbf{C}_0}{tr(\mathbf{S}_{12})}\right] \tag{19}$$

$$n_{LIML}(\beta, \mathbf{C}_0, \mathbf{W}, \mathbf{\Omega}) = \frac{tr(\mathbf{S}_{12}) - \frac{\sigma_{12}}{\sigma_1^2} tr(\mathbf{S}_1) - \mathbf{C}_0' \left(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}\mathbf{S}_1\right) \mathbf{C}_0}{tr(\mathbf{S}_2)} \tag{20}$$

2. *For* $e \in \{TSLS, LIML\}$*:*

$$B_e(\mathbf{W}, \mathbf{\Omega}) \equiv \sup_{\beta \in \mathbb{R}, \mathbf{C}_0 \in \mathcal{S}^{K-1}} \left(\frac{\left| n_e(\beta, \mathbf{C}_0, \mathbf{W}, \mathbf{\Omega}) \right|}{BM(\beta, \mathbf{W})} \right) < \infty \tag{21}$$

3. $B_{TSLS}(\mathbf{W}, \mathbf{\Omega}) \leq 1$

Proof. See Appendix A. □

Remark 3. The Nagar bias is the bias of an approximating distribution. It equals the expectation of the first three terms in the Taylor series expansion of the asymptotic estimator distribution under weak instrument asymptotics. It is therefore always defined and bounded for both TSLS and LIML. While the asymptotic estimator bias may not always exist, our test is still performing well. Under the null hypothesis, the Nagar bias can be large, but under the alternative hypothesis, the Nagar bias is small; see Section 4.2. Under certain conditions, we can also prove that the Nagar bias approximates the asymptotic estimator bias as the concentration parameter μ^2 goes to infinity; see Supplementary Materials C.1.

Remark 4. We interpret the benchmark $BM(\beta, \mathbf{W}) = \sqrt{tr(\mathbf{S}_1)/tr(\mathbf{S}_2)}$ as a "worst-case" bias. An ad-hoc approximation of $\mathbb{E}[\beta^*_{TSLS}]$ as a ratio of expectations as in Staiger and Stock (1997) helps convey the intuition:

$$\mathbb{E}[\beta^*_{TSLS}] \approx \frac{tr(\mathbf{S}_{12})}{tr(\mathbf{S}_2)[1+\mu^2]} \tag{22}$$

$$\approx \frac{1}{[1+\mu^2]} \frac{tr(\mathbf{S}_{12})}{\sqrt{tr(\mathbf{S}_2)}\sqrt{tr(\mathbf{S}_1)}} \sqrt{\frac{tr(\mathbf{S}_1)}{tr(\mathbf{S}_2)}} \tag{23}$$

The first factor is maximized when instruments are completely uninformative and μ^2=0, while the second factor is maximized when first and second stage errors are perfectly correlated (Liu and Neudecker (1995)).

Remark 5. In the implementation of our generalized testing procedure, we use the function $B_e(\mathbf{W}, \mathbf{\Omega})$ to bound the Nagar bias relative to the benchmark. We provide a fast and accurate numerical MATLAB routine for $B_e(\mathbf{W}, \mathbf{\Omega})$. For any given value of the structural parameter β, we compute the supremum over $\mathbf{C}_0 \in \mathcal{S}^{K-1}$ analytically using matrix diagonalization. We then compute the limits of $\sup_{\mathbf{C}_0 \in \mathcal{S}^{K-1}} |n_e(\beta, \mathbf{C}_0, \mathbf{W}, \mathbf{\Omega})| / BM(\beta, \mathbf{W})$ as $\beta \to \pm\infty$. Finally, we numerically maximize the function $\sup_{\mathbf{C}_0 \in \mathcal{S}^{K-1}} |n_e(\beta, \mathbf{C}_0, \widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})| / BM(\beta, \mathbf{\Omega})$ over $\beta \in [-X, X]$, where $X \in \mathbb{R}^+$ is chosen sufficiently large.

4.2 Null hypothesis

For a given threshold $\tau \in [0, 1]$ and matrix $\mathbf{W} \in \mathbb{R}^{2K \times 2K}$ we define the null and alternative hypotheses for $e \in \{TSLS, LIML\}$

$$H_e^0 : \mu^2 \in \mathcal{H}_e(\mathbf{W}, \mathbf{\Omega}) \quad v.s. \quad H_e^1 : \mu^2 \notin \mathcal{H}_e(\mathbf{W}, \mathbf{\Omega}) \tag{24}$$

where

$$\mathcal{H}_e(\mathbf{W}, \mathbf{\Omega}) \equiv \left\{ \mu^2 \in \mathbb{R}_+ : \sup_{\beta \in \mathbb{R}, C_0 \in \mathcal{S}^{K-1}} \frac{\left| N_e(\beta, \mu\sqrt{tr\mathbf{W}_2}\mathbf{C}_0, \mathbf{W}, \mathbf{\Omega}) \right|}{BM(\beta, \mathbf{W})} > \tau \right\} \tag{25}$$

Under the null hypothesis, the Nagar bias exceeds a fraction τ of the benchmark for at least some value of the structural parameter β and some direction of the first stage coefficients $\mathbf{C}_0$. On the other hand, under the alternative, the Nagar bias is at most a fraction τ of the benchmark for any values $(\beta, \mathbf{C}_0)$.

4.3 Testing Procedures

We base our test on the statistic $\widehat{F}_{eff}$, which is asymptotically distributed as a quadratic form in normal random variables with mean $1 + \mu^2$; see Lemma 1. For a survey of this class of distributions, see Johnson et al. (1995, chap. 29). Denote by $F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)$ the upper α quantile of the distribution $\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2/tr(\mathbf{W}_2)$, where $\boldsymbol{\gamma}_2 \sim \mathcal{N}_K(\mathbf{C}, \mathbf{W}_2)$ and let

$$c(\alpha, \mathbf{W}_2, x) \equiv \sup_{\mathbf{C} \in \mathbb{R}^K} \left\{ F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha) \mathbb{1}_{\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2) < x} \right\} \tag{26}$$

$\mathbb{1}_A(\cdot)$ denotes the indicator function over a set A. We base the generalized test on the observation that $\mathcal{H}_e(\mathbf{W}, \mathbf{\Omega}) = [0, B_e(\mathbf{W}, \mathbf{\Omega})/\tau)$. The generalized procedure is applicable to both TSLS and LIML, and it rejects the null hypothesis H_e^0 whenever

$$\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau) \tag{27}$$

Lemma 2. *Under Assumptions L_Π and HL the generalized procedure is pointwise asymptotically valid, i.e.*

$$\sup_{\mathcal{H}_e(\mathbf{W},\mathbf{\Omega})} \lim_{S \to \infty} \mathbb{P}\Big(\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau) \Big) \leq \alpha$$

Furthermore, provided that $B(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})$ is bounded in probability

$$lim_{\mu^2 \to \infty} lim_{S \to \infty} \mathbb{P}\left(\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau\right) = 1 \tag{28}$$

Proof. See Appendix A. □

The inequality in Theorem 1.3 implies a simplified asymptotically valid test for TSLS, which rejects the null hypothesis $\mathcal{H}_e(\mathbf{W}, \mathbf{\Omega})$ whenever

$$\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, 1/\tau) \tag{29}$$

With $c(\alpha, \widehat{\mathbf{W}}_2, 1/\tau) \geq c(\alpha, \widehat{\mathbf{W}}_2, B_{TSLS}(\mathbf{W}, \mathbf{\Omega})/\tau)$ the simplified procedure is asymptotically valid and weakly less powerful than the generalized procedure. The simplified test is conservative, in the sense that under the alternative hypothesis, the TSLS Nagar bias is lower than the threshold for any degree of dependence in the second stage.

5. COMPUTATION OF CRITICAL VALUES

We provide two simple methods to compute the critical value $c(\alpha, \mathbf{W}_2, x)$. Our first method generates Monte Carlo critical values $c_m(\alpha, \mathbf{W}_2, x)$. We obtain estimates of $F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)$ as the sample upper α point from a large number of draws from the distribution of $\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2/tr(\widehat{\mathbf{W}}_2)$, and then maximize over a large set of $\mathbf{C}$, such that $\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2) \leq x$.

The second procedure is based on a curve-fitting methodology first suggested by Patnaik (1949). Patnaik (1949) and Imhof (1961) approximate the critical values of a weighted sum of independent non-central chi-squared distributions by a central χ^2 with the same first and second moments. We analogously approximate the distribution $F_{\mathbf{C},\mathbf{W}_2}$ by a non-central χ^2 with the same first and second moments. Our approximation errors are therefore bounded by the original Patnaik errors through a triangle inequality. We use

$$F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha) \approx \frac{1}{K_{eff}} F^{-1}_{\chi^2_{K_{eff}}(K_{eff}\mu^2)}(\alpha) \tag{30}$$

where K_{eff} is possibly fractional with

$$K_{eff} = [tr(\mathbf{W}_2)]^2 \frac{1+2\mu^2}{tr(\mathbf{W}_2^2) + 2\mathbf{C}'\mathbf{W}_2\mathbf{C}} \tag{31}$$

There is a large literature that approximates distributions by choosing a family of distributions and selecting the member that fits best, often by matching lower order moments of the original distribution (Satterthwaite, 1946; Theil and Nagar, 1961; Henshaw, 1966; Pearson, 1959; Grubbs, 1964; Conerly and Mansfield, 1988; Liu et al., 2009). The non-central chi-squared distribution is a natural choice, because it is exact in the homoskedastic case.

While it is hard to assess the accuracy of these curve-fitting approximations analytically, they are often simple and numerically highly accurate (Rothenberg, 1984). Authors demonstrate the degree of accuracy of their approximations using numerical examples. In the Supplementary Materials B.1, we verify that the approximation (30) is numerically as accurate as the original central Patnaik distribution for the quadratic forms considered in Imhof (1961); approximation errors are at most 0.7 % points in the important upper 15% tail of the distributions.

Numerical results, such as in Table 1, clearly indicate that upper α quantiles of (30) are decreasing in K_{eff}. Moreover, the upper α quantile in (30) is nondecreasing in the noncentrality parameter μ^2 (Ghosh, 1973). Taking the supremum over $\mathbf{C}$ with $\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2) < x$, suggests the Patnaik critical value.

Definition 2. (**Patnaik Critical Value**) Define the Patnaik critical value as

$$c_P(\alpha, \mathbf{W}_2, x) \equiv F^{-1}_{(1/K_{eff})\chi^2_{K_{eff}}(xK_{eff})}(\alpha) \tag{32}$$

with the effective number of degrees of freedom

$$K_{eff} \equiv \frac{tr(\mathbf{W}_2)^2(1+2x)}{tr(\mathbf{W}_2^2) + 2tr(\mathbf{W}_2)\max eval(\mathbf{W}_2)x} \tag{33}$$

We numerically analyze the sizes of Monte Carlo and Patnaik critical values for benchmark parameter values $\alpha = 5\%$ and $x = 10$, and find that size distortions are small for both methodologies. Monte Carlo critical values are computed with 40000 draws from $\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2/tr(\mathbf{W}_2)$, and we replace the infinite set of vectors $\mathbf{C}$ s.t. $\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2) < x$ by a finite set of size 500. We use code for $F_{\mathbf{C},\mathbf{W}_2}(x)$ available at http://elsa.berkeley.edu/∼ruud/cet/pgms.htm (Imhof, 1961; Koerts and Abrahamse, 1969; Farebrother, 1990; Ruud, 2000). For 400 matrices $\mathbf{W}_2$ from a diffuse prior with $K \in \{1, 2, 3, 4, 5\}$ our numerical values for $\max_{\mathbf{C}'\mathbf{C}/tr\mathbf{W}_2<x} F_{\mathbf{C},\mathbf{W}_2}(c_m)$ range between 4.77% and 5.26%, and our numerical values for $\max_{\mathbf{C}'\mathbf{C}/tr\mathbf{W}_2<x} F_{\mathbf{C},\mathbf{W}_2}(c_P)$ range between 5.00% and 5.02%. For further details and MATLAB routines, see Supplementary Materials B.2-B.5.

Our generalized and simplified critical values differ from those proposed by Stock and Yogo (2005) for the TSLS bias, even when first and second stage errors are perfectly conditionally homoskedastic and serially uncorrelated. In this case, the effective F statistic coincides with the Stock and Yogo (2005) test statistic. We obtain different critical values because, unlike them, we use an approximation to evaluate the weak instrument TSLS bias. Moreover, estimating $\widehat{\mathbf{W}}$ and $\widehat{\boldsymbol{\Omega}}$ also generates differences in critical values. The difference between our generalized TSLS critical values and analogous Stock and Yogo (2005) critical values becomes small as the number of instruments becomes large.

In the Supplementary Materials B.6, we tabulate Stock and Yogo (2005) 5% critical values for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias and our generalized and simplified critical values with a threshold of 10% and size 5%, assuming conditional homoskedasticity and no serial correlation. TSLS critical values are smaller than Stock and Yogo (2005) critical values for $K = 3, 4$, but larger than Stock and Yogo (2005) critical values for $K \geq 5$. The difference between the TSLS and Stock and Yogo (2005) critical values is always less than 1. The LIML critical values decline more rapidly with the number of instruments than either the TSLS or simplified critical values. The simplified critical values exceed the generalized TSLS critical values, because they use a bound that applies for any form of the matrix $\mathbf{W}$.

6. EMPIRICAL APPLICATION: ESTIMATING THE ELASTICITY OF INTERTEMPORAL SUBSTITUTION

We now apply our pre-testing procedure to an empirical example, and show that allowing for heteroskedasticity and time series correlation can affect pre-testing conclusions.

The literature has focused on estimating the linearized Euler equation in two standard IV frameworks (Hansen and Singleton, 1983; Campbell and Mankiw, 1989; Hall, 1988; Campbell, 2003).

$$\Delta c_{t+1} = \nu + \psi r_{t+1} + u_{t+1} \text{ and } \mathbb{E}[\mathbf{Z}_{t-1} u_{t+1}] = 0 \tag{34}$$

$$r_{t+1} = \xi + (1/\psi)\Delta c_{t+1} + \eta_{t+1} \text{ and } \mathbb{E}[\mathbf{Z}_{t-1} \eta_{t+t}] = 0 \tag{35}$$

ψ is the Elasticity of Intertemporal Substitution (EIS), Δc_{t+1} is consumption growth at time $t+1$, r_{t+1} is a real asset return, and ν is a constant. The vector of instruments is denoted by $\mathbf{Z}_{t-1}$. We follow the preferred choice of variables in Yogo (2004), using as r_t the real return on the short-term interest rate, and as instruments the nominal interest rate, inflation, consumption growth and the log dividend-price ratio, all lagged twice. We use quarterly data from Yogo (2004).

The EIS determines an agent's willingness to substitute consumption over time. Its magnitude is important for understanding the dynamics of consumption and asset returns (Epstein and Zin, 1989, 1991; Campbell, 2003). While time-varying volatility can introduce additional bias into the estimation of the EIS (Bansal and Yaron, 2004), Yogo (2004) argues that under certain types of conditional heteroskedasticity the EIS can still be identified.

Table 2 compares pre-tests for weak instruments for 11 countries. Panel A shows weak instrument pre-tests with the ex-post real interest rate as the endogenous variable, while Panel B shows weak instrument pre-tests with consumption growth as the endogenous variable. The non-robust first stage F statistic in column 1 is shown in bold whenever it exceeds the Stock and Yogo (2005) critical value 10.27. This is the 5% critical value for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias under the assumption of conditional homoskedasticity and no serial correlation. As in Yogo (2004), this homoskedastic pre-test indicates strong instruments in Panel A, but cannot reject weak instruments in Panel B for almost all countries in the sample.

The second and third columns report the HAC robust first stage F statistic and the effective F statistic computed with a Newey-West kernel and six lags. We show 5% critical values for TSLS, LIML, and simplified pre-tests for the null hypothesis that the respective Nagar bias exceeds 10% of the "worst-case" benchmark.

In Panel A, we see that allowing for heteroskedasticity and serial correlation changes the pre-testing results for some countries, while for other countries all pre-tests yield the same conclusion. The effective F statistic can be smaller or larger than the regular or robust F statistics. Simplified critical values always exceed TSLS critical values. LIML critical values tend to be smallest.

The results in Table 2A for the U.S. are particularly striking. While the U.S. regular F statistic clearly exceeds the homoskedastic threshold of 10.27, the robust and effective F statistics are significantly smaller. The effective F does not exceed the simplified, TSLS,

Table 2
Estimating the Elasticity of Intertemporal Substitution:
Weak Instrument Pre-Tests

Panel A: $\Delta c_{t+1} = \nu + \psi r_{t+1} + u_{t+1}$ and $\mathbb{E}[\mathbf{Z}_{t-1} u_{t+1}] = 0$

Country	Sample Period	$\widehat{F}$	$\widehat{F}_r$	$\widehat{F}_{eff}$	c_{Simp}	c_{TSLS}	c_{LIML}	$\widehat{\psi}_{\text{TSLS}}$	$\widehat{\psi}_{\text{LIML}}$
USA	1947.3-1998.4	**15.53**	8.60	7.94	18.20	15.49	9.68	0.06	0.03
AUL	1970.3-1998.4	**21.81**	27.56	17.52	18.36	**16.64**	**10.25**	0.05	0.03
CAN	1970.3-1999.1	**15.37**	11.58	12.95	18.95	17.38	**11.44**	-0.30	-0.34
FR	1970.3-1998.3	**38.43**	41.67	40.29	**19.51**	**17.01**	**12.89**	-0.08	-0.08
GER	1979.1-1998.3	**17.66**	12.47	11.66	18.24	16.30	10.01	-0.42	-0.44
ITA	1971.4-1998.1	**19.01**	25.09	19.44	**19.26**	**17.37**	**12.98**	-0.07	-0.07
JAP	1970.3-1998.4	8.64	8.32	5.09	21.66	20.24	18.71	-0.04	-0.05
NTH	1977.3-1998.4	**12.05**	9.31	10.53	18.89	17.18	11.28	-0.15	-0.14
SWD	1970.3-1999.2	**17.08**	28.86	19.82	**19.04**	**15.59**	**11.65**	0.00	0.00
SWT	1976.2-1998.4	8.55	6.68	7.19	18.49	15.80	10.38	-0.49	-0.50
UK	1970.3-1999.1	**17.04**	11.78	7.65	20.18	18.72	14.57	0.17	0.16

Panel B: $r_{t+1} = \xi + (1/\psi)\Delta c_{t+1} + \eta_{t+1}$ and $\mathbb{E}[\mathbf{Z}_{t-1}\eta_{t+1}] = 0$

Country	Sample Period	$\widehat{F}$	$\widehat{F}_r$	$\widehat{F}_{eff}$	c_{Simp}	c_{TSLS}	c_{LIML}	$\widehat{\psi}^{-1}_{\text{TSLS}}$	$\widehat{\psi}^{-1}_{\text{LIML}}$
USA	1947.3-1998.4	2.93	3.37	2.58	17.61	13.99	10.23	0.68	34.11
AUL	1970.3-1998.4	1.79	2.87	2.31	19.89	17.25	15.70	0.50	30.03
CAN	1970.3-1999.1	3.03	5.99	2.70	18.19	15.89	9.77	-1.04	-2.98
FR	1970.3-1998.3	0.17	0.39	0.22	19.83	18.08	14.09	-3.12	-12.38
GER	1979.1-1998.3	0.83	2.48	1.13	18.58	16.98	14.19	-1.05	-2.29
ITA	1971.4-1998.1	0.73	0.39	0.47	19.05	16.96	11.63	-3.34	-14.81
JAP	1970.3-1998.4	1.18	2.17	2.00	17.94	13.93	15.58	-0.18	-21.56
NTH	1977.3-1998.4	0.89	3.62	1.84	19.00	16.13	15.30	-0.53	-6.94
SWD	1970.3-1999.2	0.48	0.81	0.83	17.24	12.51	9.73	-0.10	-399.86
SWT	1976.2-1998.4	0.97	2.28	1.56	20.21	18.76	16.47	-1.56	-2.00
UK	1970.3-1999.1	2.52	3.95	2.55	17.94	15.64	14.50	1.06	6.21

NOTE: Δc is consumption growth and r is the ex-post real short-term interest rate. We instrument using twice lagged nominal interest rate, inflation, dividend-price ratio, and consumption growth. HAC variance-covariance matrix $\widehat{\mathbf{W}}$ estimated with OLS and Newey-West kernel with six lags. F statistic in bold when it exceeds the critical value of 10.27. This is the 5% critical value for testing the null hypothesis that the TSLS bias exceeds 10% of the OLS bias under the assumption of conditional homoskedasticity and no serial correlation (Stock and Yogo, 2005). We show simplified, TSLS, and LIML critical values $c_{Simp} = c_P(5\%, \widehat{\mathbf{W}}_2, 10)$, $c_{\text{TSLS}} = c_P\left(5\%, \widehat{\mathbf{W}}_2, 10 \times B_{TSLS}(\widehat{\mathbf{W}}, \widehat{\boldsymbol{\Omega}})\right)$, and $c_{\text{LIML}} = c_P\left(5\%, \widehat{\mathbf{W}}_2, 10 \times B_{\text{LIML}}(\widehat{\mathbf{W}}, \widehat{\boldsymbol{\Omega}})\right)$. Critical values are in bold when exceeded by $\widehat{F}_{eff}$. $\hat{\psi}_{\text{TSLS}}$, $\hat{\psi}_{\text{LIML}}$, $\widehat{(1/\psi)}_{\text{TSLS}}$ and $\widehat{(1/\psi)}_{\text{LIML}}$ are TSLS and LIML estimates of the EIS and its inverse.

or LIML critical values, so we cannot reject the null hypothesis of weak instruments under heteroskedasticity and serial correlation.

Panel B shows weak instrument pre-tests for the instrumental variable estimation of the inverse of the EIS. For this estimation, the results are consistent between homoskedastic and HAC weak instrument pre-tests. We cannot reject that instruments are weak for any of the countries in the sample.

The last two columns in Table 2 show the point estimates for $\boldsymbol{\psi}$ and $1/\boldsymbol{\psi}$. For those cases where we can reject weak instruments under heteroskedasticity and serial correlation, the corresponding EIS point estimates are close to zero and often negative. Additional caution is, however, warranted in this interpretation, because as the number of countries increases, we are more and more likely to reject weak instruments at least once.

Our results confirm Yogo (2004)'s finding that the EIS is small and close to zero. However, we also note that conditional heteroskedasticity and serial correlation may further weaken instruments and may affect TSLS and LIML bias in several of the country-specific regressions.

7. CONCLUSION

Heteroskedasticity, serial correlation, and panel data clustering can affect instrument strength. This paper develops a robust test for weak instruments that allows empirical researchers to test the null hypothesis that the TSLS or LIML Nagar bias is large relative to a benchmark.

The test is based on a scaled version of the regular F statistic. Critical values depend on the covariance matrix of the reduced form coefficients and errors. Our general test requires computational work to evaluate the Nagar bias of TSLS or LIML. A simplified conservative version does not require this step, but is only available for TSLS. Critical values can then be implemented as quantiles of a non-central chi-squared distribution with non-integer degrees of freedom.

Pre-tests based on the robust (or non-robust) first stage F statistic with Stock and Yogo (2005) critical values are commonly applied outside the conditionally homoskedastic serially uncorrelated framework. However, to the best of our knowledge, there is no analysis supporting this practice. This paper offers an alternative: a simple, asymptotically valid test that should be used for conditionally heteroskedastic, time series, and clustered panel data.

ACKNOWLEDGMENTS

We are grateful to John Campbell, Matias Cattaneo, Gary Chamberlain, Eduardo Davila, Joshua Gottlieb, Stephanie Hurder, Rustam Ibragimov, Guido Imbens, Michal Kolesar, Ignacio Lobato, Joshua Mendel, Marcelo Moreira, Daniel Pollmann, Doug Staiger, Luis Viceira, the editors Keisuke Hirano and Jonathan Wright, an anonymous associate editor, two anonymous referees, and seminar participants at the Harvard Econometrics Lunch and at the North American Summer Meeting of the Econometric Society 2011 for helpful comments and suggestions. We are especially grateful to Jim Stock for his invaluable advice and guidance. The usual disclaimer applies.

APPENDIX A

A.1 Proof of Lemma 1

First note the preliminary result that under Assumptions L_{Π} and HL

$$\frac{1}{\sqrt{S}}\begin{pmatrix}\mathbf{Z}'\mathbf{y}\\ \mathbf{Z}'\mathbf{Y}\end{pmatrix} = \begin{pmatrix}\beta\mathbf{C}+\mathbf{Z}'\mathbf{v}_1/\sqrt{S}\\ \mathbf{C}+\mathbf{Z}'\mathbf{v}_2/\sqrt{S}\end{pmatrix} \tag{36}$$

$$\xrightarrow{d} \begin{pmatrix}\boldsymbol{\gamma}_1\\ \boldsymbol{\gamma}_2\end{pmatrix} \tag{37}$$

1. $\widehat{\beta}_{TSLS} \equiv (\mathbf{Y}'\mathbf{P_Z}\mathbf{Y})^{-1}(\mathbf{Y}'\mathbf{P_Z}\mathbf{y}) = (\mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y})^{-1}(\mathbf{Y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y})$. Since we have assumed that $\mathbf{Z}'\mathbf{Z}/S = \mathbb{I}_K$, the result follows from (37) and the continuous mapping theorem.

2. Write $\mathbf{J} = \begin{bmatrix}1 & 0\\ -\beta & 1\end{bmatrix}$ and $\kappa = S(k-1)$. Note that $\mathbf{J}$ is nonsingular and so the roots of $|[\mathbf{y},\mathbf{Y}]'[\mathbf{y},\mathbf{Y}] - k[\mathbf{y},\mathbf{Y}]'\mathbf{M_Z}[\mathbf{y},\mathbf{Y}]| = 0$ are the same as of

$|\mathbf{J}'[\mathbf{y},\mathbf{Y}]'[\mathbf{y},\mathbf{Y}]\mathbf{J} - k\mathbf{J}'[\mathbf{y},\mathbf{Y}]'\mathbf{M}_\mathbf{Z}[\mathbf{y},\mathbf{Y}]\mathbf{J}| = 0$. Moreover

$[\mathbf{y},\mathbf{Y}]'[\mathbf{y},\mathbf{Y}] - (1+\kappa/S)[\mathbf{y},\mathbf{Y}]'\mathbf{M}_\mathbf{Z}[\mathbf{y},\mathbf{Y}] = [\mathbf{y},\mathbf{Y}]'\mathbf{P}_\mathbf{Z}[\mathbf{y},\mathbf{Y}] - \kappa[\mathbf{y},\mathbf{Y}]'\mathbf{M}_\mathbf{Z}[\mathbf{y},\mathbf{Y}]/S$

$\xrightarrow{d}[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2]'[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2] - \kappa\boldsymbol{\Omega}$ uniformly in κ over compact sets. The solutions of

$|[\mathbf{y},\mathbf{Y}]'[\mathbf{y},\mathbf{Y}] - (1+\kappa/S)[\mathbf{y},\mathbf{Y}]'\mathbf{M}_\mathbf{Z}[\mathbf{y},\mathbf{Y}]| = 0$ therefore converge to those of

$|\mathbf{J}'[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2]'[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2]\mathbf{J} - \kappa\mathbf{J}'\boldsymbol{\Omega}\mathbf{J}| = 0$. With $\mathbf{J}'\boldsymbol{\Omega}\mathbf{J} = \boldsymbol{\Sigma}$ thus $S(\hat{k}_{LIML}-1) \xrightarrow{d} \kappa_{LIML}$ where κ_{LIML} is as given in Lemma 1.2.

Then $\widehat{\beta}_{LIML} - \beta =$

$\left[\mathbf{Y}'(\mathbb{I}_S - \hat{k}_{LIML}\mathbf{M}_\mathbf{Z})\mathbf{Y}\right]^{-1}[\mathbf{Y}'(\mathbb{I}_S - \hat{k}_{LIML}\mathbf{M}_\mathbf{Z})(\mathbf{y}-\beta\mathbf{Y})]$

$=\left[\mathbf{Y}'\mathbf{P}_\mathbf{Z}\mathbf{Y} - S(\hat{k}_{LIML}-1)\frac{\mathbf{Y}'\mathbf{M}_\mathbf{Z}\mathbf{Y}}{S}\right]^{-1}[\mathbf{Y}'\mathbf{P}_\mathbf{Z}(\mathbf{y}-\beta\mathbf{Y}) - S(\hat{k}_{LIML}-1)\frac{\mathbf{Y}'\mathbf{M}_\mathbf{Z}(\mathbf{y}-\beta\mathbf{Y})}{S}]$

$\xrightarrow{d}[\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2 - \kappa_{LIML}\sigma_2^2]^{-1}[\boldsymbol{\gamma}_2(\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2) - \kappa_{LIML}\sigma_{12}]$

3. Note that $\widehat{\omega}_2^2 \equiv (\mathbf{Y}-\mathbf{P}_\mathbf{Z}\mathbf{Y})'(\mathbf{Y}-\mathbf{P}_\mathbf{Z}\mathbf{Y})/(S-K-1) = (\mathbf{v}_2-\mathbf{P}_\mathbf{Z}\mathbf{v}_2)'(\mathbf{v}_2-\mathbf{P}_\mathbf{Z}\mathbf{v}_2)/(S-K-1)$

   $\xrightarrow{d}\omega_2^2$ by Assumptions L_Π and HL. The result follows from (37) and the continuous mapping theorem.

4. and 5. follow from (37), the continuous mapping theorem, and Assumptions L_Π and HL.

A.2 LIML Distribution in Illustrative Example

We show that in the illustrative example heteroskedascity and serial correlation can effectively make instruments weaker for LIML. Assume $\mathbf{W} = a^2\boldsymbol{\Omega}\otimes\mathbb{I}_K$. Remember that $\hat{\beta}_{LIML} = arg\min_{\tilde{\beta}}(\mathbf{y}-\tilde{\beta}Y)'\mathbf{P}_\mathbf{Z}(\mathbf{y}-\tilde{\beta}\mathbf{Y})/(\mathbf{y}-\tilde{\beta}\mathbf{Y})'(\mathbf{y}-\tilde{\beta}\mathbf{Y})$. We will analyze the weak instrument limit of the LIML objective function. Note that, using assumptions L_Π and HL $\mathbf{Z}'(\mathbf{y}-\tilde{\beta}\mathbf{Y})/\sqrt{S}$ $\xrightarrow{d}\boldsymbol{\gamma}_1 - \tilde{\beta}\boldsymbol{\gamma}_2$.

Moreover, $(\mathbf{y}-\tilde{\beta}\mathbf{Y})'(\mathbf{y}-\tilde{\beta}\mathbf{Y})/S \xrightarrow{p} \omega_1^2 - 2\tilde{\beta}\omega_{12} + \tilde{\beta}^2\omega_2^2$ uniformly in $\tilde{\beta}$ over compact sets.

Hence β^*_{LIML} is distributed according to

$$arg\min_{\tilde{\beta}} a^2 \frac{\left(\omega_1 \boldsymbol{\psi}_1 - \tilde{\beta}\omega_2\boldsymbol{\psi}_2\right)' \left(\omega_1 \boldsymbol{\psi}_1 - \tilde{\beta}\omega_2\boldsymbol{\psi}_2\right)}{\omega_1^2 - 2\tilde{\beta}\omega_{12} + \tilde{\beta}^2\omega_2^2}$$

Just as for the β_{TSLS}, the vector of first stage coefficients $\mathbf{C}$ and the parameter a enter into the asymptotic distribution β^*_{LIML} only through the noncentrality parameter $\mathbf{C}'\mathbf{C}/(a^2\omega_2^2)$.

A.3 Proof of Theorem 1

A.3.1 Proof of Theorem 1.1

We follow Rothenberg (1984) in developing the Nagar (1959) moments for the TSLS and LIML estimators. We need to expand β^*_{TSLS} and β^*_{LIML} as second order Taylor expansions in μ^{-1} around $\mu^{-1} = 0$.

We start by developing the Taylor expansion for κ_{LIML}. Write $\mathbf{z}_u = \mathbf{S}_1^{-1/2}(\boldsymbol{\gamma}_1 - \beta\boldsymbol{\gamma}_2)$ and $\mathbf{z}_v = \mathbf{S}_2^{-1/2}(\boldsymbol{\gamma}_2 - C)$ so $\mathbf{z}_u$ and $\mathbf{z}_v$ are standard multivariate normal. Also write $\boldsymbol{\lambda} = \mu tr\left(\mathbf{S}_2\right)^{1/2}\mathbf{S}_2^{-1/2}\mathbf{C}_0$ where $\mathbf{C}_0 = \mathbf{C}/\left\|\mathbf{C}\right\|$.

κ_{LIML} is defined as the smallest root of the determinantal equation

$$\det\left(\mathbf{A} - \kappa_{LIML}\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right) = 0 \tag{38}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{z}_u'\mathbf{S}_1\mathbf{z}_u & \mathbf{z}_u'\mathbf{S}_1^{1/2}\mathbf{S}_2^{1/2}\left(\boldsymbol{\lambda} + \mathbf{z}_v\right) \\ \mathbf{z}_u'\mathbf{S}_1^{1/2}\mathbf{S}_2^{1/2}\left(\boldsymbol{\lambda} + \mathbf{z}_v\right) & \left(\mathbf{z}_v + \boldsymbol{\lambda}\right)'\mathbf{S}_2\left(\mathbf{z}_v + \boldsymbol{\lambda}\right) \end{bmatrix}$$

We can rewrite this as a quadratic equation

$$\left(\frac{\kappa_{LIML}}{\mu^2}\right)^2 - \frac{\sigma_v^2 a_{11} + \sigma_1^2 a_{22} - 2a_{12}\sigma_{12}}{\mu^2 \det \mathbf{\Sigma}} \frac{\kappa_{LIML}}{\mu^2} + \frac{\det \mathbf{A}}{\mu^4 \det \mathbf{\Sigma}} = 0 \tag{39}$$

We use the method of undetermined coefficients. Write

$$\kappa_{LIML}\mu^{-2} = c_0 + c_1\mu^{-1} + c_2\mu^{-2} + O\left(\mu^{-3}\right) \tag{40}$$

for unknown constants c_0, c_1, c_2. Similarly write

$$d(\mu) \equiv \frac{\sigma_2^2 a_{11} + \sigma_u^2 a_{22} - 2a_{12}\sigma_{12}}{\mu^2 \det \mathbf{\Sigma}} = d_0 + d_1\mu^{-1} + d_2\mu^{-2} + O\left(\mu^{-2}\right) \tag{41}$$

$$e(\mu) \equiv \frac{\det \mathbf{A}}{\mu^4 \det \mathbf{\Sigma}} = \frac{\det \mathbf{A}}{\mu^4 \det \mathbf{\Sigma}} = e_0 + e_1\mu^{-1} + e_2\mu^{-2} + O\left(\mu^{-3}\right) \tag{42}$$

where the Taylor series expansions for d and e give $d_0 = \sigma_1^2 tr\left(\mathbf{S}_2\right)/\det \mathbf{\Sigma}$, $e_0 = 0$, $e_1 = 0$, and $e_2 = tr\left(\mathbf{S}_2\right)\left[\mathbf{z}_u'\mathbf{S}_1\mathbf{z}_u - \left(\mathbf{z}_u'\mathbf{S}_1^{1/2}\mathbf{C}_0\right)^2\right]/\det \mathbf{\Sigma}$.

Substituting (40), (41) and (42) into the quadratic equation (39) and equating coefficients gives $c_0\left(c_0 - d_0\right) = 0$ Since we are interested in the smaller solution, we have $c_0 = 0$. Then $c_0 = 0, c_1 = 0, c_2 = \varepsilon_2/d_0$ and so $\kappa_{LIML}\mu^{-2} = \frac{1}{\sigma_1^2}\left[\mathbf{z}_u'\mathbf{S}_1\mathbf{z}_u - \left(\mathbf{z}_u'\mathbf{S}_1^{1/2}\mathbf{C}_0\right)^2\right]\mu^{-2} + O(\mu^{-3})$

We then expand β_{LIML}^*

$$\begin{aligned}\beta_{LIML}^* \quad = \quad & \mu^{-1}\frac{\mathbf{C}_0'\mathbf{S}_1^{1/2}\mathbf{z}_u tr\left(\mathbf{S}_2\right)^{1/2}}{tr\left(\mathbf{S}_2\right)} + \\ & +\mu^{-2}\frac{1}{tr\left(\mathbf{S}_2\right)}\left(\mathbf{z}_v'\mathbf{S}_2^{1/2}\mathbf{S}_1^{1/2}\mathbf{z}_u - 2\left(\mathbf{C}_0'\mathbf{S}_1^{1/2}\mathbf{z}_u\right)\left(\mathbf{C}_0'\mathbf{S}_2^{1/2}\mathbf{z}_v\right) - c_2\sigma_{12}\right) \\ & +O\left(\mu^{-3}\right)\end{aligned}$$

Taking the expectation of the first two terms gives the LIML Nagar bias as in the Theorem.

We can similarly derive the Taylor expansion for β^*_{TSLS} according to

$$\begin{aligned}
\beta^*_{TSLS} &= \mu^{-1}\frac{\mathbf{C}_0'\mathbf{S}_1^{1/2}\mathbf{z}_u tr\left(\mathbf{S}_2\right)^{1/2}}{tr\left(\mathbf{S}_2\right)}+ \\
&+\mu^{-2}\frac{1}{tr\left(\mathbf{S}_2\right)}\left(\mathbf{z}_v'\mathbf{S}_2^{1/2}\mathbf{S}_1^{1/2}\mathbf{z}_u-2\left(\mathbf{C}_0'\mathbf{S}_1^{1/2}\mathbf{z}_u\right)\left(\mathbf{C}_0'\mathbf{S}_2^{1/2}\mathbf{z}_v\right)\right) \\
&+O\left(\mu^{-3}\right)
\end{aligned}$$

The Nagar bias is defined as the expected value of the first two terms, and hence equals

$$N_{TSLS}(\beta,\mathbf{C},\mathbf{W},\boldsymbol{\Omega})=\frac{1}{tr\mathbf{S}_2}\left(tr\mathbf{S}_{12}-2\mathbf{C}_0'\mathbf{S}_{12}\mathbf{C}_0\right)\mu^{-2}$$

A.3.2 Proof of Theorems 1.2 and 1.3

We prove Theorem 1.3 first. We assume that $\mathbf{W}$ and $\boldsymbol{\Omega}$ are positive definite, so $\mathbf{S}$ and $\boldsymbol{\Sigma}$ are also positive definite. $\mathbf{S}_{12}$ is real valued but not necessarily symmetric. Note that

$$tr\mathbf{S}_{12}-2\mathbf{C}_0'\mathbf{S}_{12}\mathbf{C}_0=tr\mathbf{S}_{12}^{sym}-2\mathbf{C}_0'\mathbf{S}_{12}^{sym}\mathbf{C}_0$$

where $\mathbf{S}_{12}^{sym}=\frac{1}{2}(\mathbf{S}_{12}+\mathbf{S}_{12}')$ is the symmetric part of $\mathbf{S}_{12}$. Write $\boldsymbol{\Lambda}=\begin{bmatrix}\lambda_1 & 0 & ... & 0\\ 0 & \lambda_2 & ... & 0\\ 0 & 0 & ... & \lambda_K\end{bmatrix}$ for the diagonal matrix of eigenvalues of $\mathbf{S}_{12}^{sym}$. Assume the eigenvalues are ordered so $\lambda_1\geq\lambda_2\geq..\geq\lambda_K$. For any real matrix $\mathbf{M}$ we write $|\mathbf{M}|=\sqrt{\mathbf{M}'\mathbf{M}}$ so the Schatten 1-norm for matrices is defined as $\|\mathbf{M}\|_1 = tr|\mathbf{M}|$.

$$
\begin{aligned}
tr\mathbf{S}_{12}^{sym} - 2\mathbf{C}_0'\mathbf{S}_{12}^{sym}\mathbf{C}_0 &\leq \sum_{k=1}^{K} \lambda_k - 2\lambda_K \\
&= \sum_{k=1}^{K-1} \lambda_k - \lambda_K \\
&\leq \sum_{k=1}^{K} |\lambda_k| \\
&= \|\mathbf{S}_{12}^{sym}\|_1
\end{aligned}
$$

Similarly $tr\mathbf{S}_{12}^{sym} - 2\mathbf{C}_0'\mathbf{S}_{12}^{sym}\mathbf{C}_0 \geq -\|\mathbf{S}_{12}^{sym}\|_1$. Hence $|tr\mathbf{S}_{12}^{sym} - 2\mathbf{C}_0'\mathbf{S}_{12}^{sym}\mathbf{C}_0| \leq \|0.5\mathbf{S}_{12} + 0.5\mathbf{S}_{12}'\|_1$ $\leq \|\mathbf{S}_{12}\|_1$. The last step follows from the triangle inequality and from the fact that the eigenvalues of $\mathbf{S}_{12}'\mathbf{S}_{12}$ and $\mathbf{S}_{12}\mathbf{S}_{12}'$ are the same.

Now $tr(\mathbf{S}_{12}'\mathbf{S}_2^{-1}\mathbf{S}_{12}) \leq tr(\mathbf{S}_1)$, see e.g. Theorem 7.14 in Zhang (2010). By the matrix trace Cauchy inequality (Liu and Neudecker (1995), Theorem 1) then

$$
\begin{aligned}
\|\mathbf{S}_{12}\|_1^2 &= (tr|\mathbf{S}_{12}|)^2 \\
&\leq tr\mathbf{S}_2 tr\left(|\mathbf{S}_{12}|'\mathbf{S}_2^{-1}|\mathbf{S}_{12}|\right) \\
&= tr\mathbf{S}_2 tr(\mathbf{S}_{12}'\mathbf{S}_2^{-1}\mathbf{S}_{12})
\end{aligned}
$$

Putting this together, we get $\|\mathbf{S}_{12}\|_1 \leq \sqrt{tr\mathbf{S}_1 tr\mathbf{S}_2}$, proving Theorem 1.3.

The TSLS part of Theorem 1.2 follows from Theorem 1.3. For the LIML part note that $B_{LIML}(\mathbf{W}, \mathbf{\Omega}) = \sup_{\beta\in\mathbb{R}} g_{LIML}(\beta)$ where

$$
g_{LIML}(\beta) = max\left(\left|\frac{tr\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}tr\mathbf{S}_1 - maxeval\mathbf{M_B}}{\sqrt{tr\mathbf{S}_1}\sqrt{tr\mathbf{S}_2}}\right|, \left|\frac{tr\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}tr\mathbf{S}_1 - mineval\mathbf{M_B}}{\sqrt{tr\mathbf{S}_1}\sqrt{tr\mathbf{S}_2}}\right|\right) \tag{43}
$$

where $\mathbf{M_B} = \frac{1}{2}(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}\mathbf{S}_1) + \frac{1}{2}(2\mathbf{S}_{12} - \frac{\sigma_{12}}{\sigma_1^2}\mathbf{S}_1)'$ and

$$g_{LIML}(\beta) \to \frac{maxeval\mathbf{W}_2}{tr\mathbf{W}_2} \text{ as } \beta \to \pm\infty \tag{44}$$

For $\mathbf{W}$ and $\mathbf{\Omega}$ nonsingular g_{LIML} is continuous in β everywhere, and hence bounded.

A.4 Proof of Lemma 2

Assume that $\mathbf{W}$ and $\mathbf{\Omega}$ are nonsingular. We prove that the test that rejects if:

$$\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau) \tag{45}$$

is asymptotically valid, i.e. its asymptotic size is at most α.

Claim 1: The function $F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)$ is continuous in $\{\mathbf{C}, \mathbf{W}_2\}$.

Proof: : $\boldsymbol{\gamma}_2'\boldsymbol{\gamma}_2/tr(\mathbf{W}_2)$ is a continuous random variable with nonzero density on $\mathbb{R}_+$, and therefore $F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)$ is strictly decreasing and continuous in α everywhere. By Van der Vaart (2000, Lemma 21.2) the quantile function $F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)$ is continuous in $\{\mathbf{C}, \mathbf{W}_2\}$ for any fixed α.

Claim 2: The function $B_e(\mathbf{W}, \mathbf{\Omega})$ is lower semicontinuous.

Proof: The function $\|n_e(\beta, \mathbf{C}_0, \mathbf{W}, \mathbf{\Omega})\| / BM(\beta, \mathbf{W})$ is continuous in $\mathbf{W}$ and $\mathbf{\Omega}$. $B_e(\mathbf{W}, \mathbf{\Omega})$ is the supremum of continuous functions, and therefore is lower semicontinuous (Yeh, 2000, p. 274).

Claim 3: The function $c(\alpha, \mathbf{W}_2, x)$ is lower semicontinuous in $\{\mathbf{W}_2, x\}$.

Proof: The function $\mathbb{1}_{\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2)<x}$ is an indicator function of an open set, and therefore lower semicontinuous in $\{\mathbf{W}_2, x\}$. The function $F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)$ is continuous in $\mathbf{W}_2$ and greater than 0. Hence the product $F^{-1}_{\mathbf{C},\mathbf{W}_2}(\alpha)\mathbb{1}_{\mathbf{C}'\mathbf{C}/tr(\mathbf{W}_2)<x}$ is lower semicontinuous in $(\mathbf{W}_2, x)$ for any

fixed α. $c(\alpha, \mathbf{W}_2, x)$ is a supremum of lower semicontinuous functions, and therefore lower semicontinuous in $(\mathbf{W}_2, x)$ (Yeh, 2000, p. 274). $c(\alpha, \mathbf{W}_2, x)$ is also clearly nondecreasing in x.

Proof of Result: From the lower semicontinuity of $B(\mathbf{W}, \mathbf{\Omega})$ and the continuous mapping theorem, it follows that $min\left(B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}}), B_e(\mathbf{W}, \mathbf{\Omega})\right) \xrightarrow{p} B_e(\mathbf{W}, \mathbf{\Omega})$. Similarly, for any $\left(\widehat{\mathbf{W}}_2, \widehat{x}\right) \xrightarrow{p} (\mathbf{W}_2, x)$, the continuous mapping theorem implies that $min(c(\alpha, \widehat{\mathbf{W}}_2, \widehat{x}), c(\alpha, \mathbf{W}_2, x)) \xrightarrow{p} c(\alpha, \mathbf{W}_2, x)$. Then

$$
\begin{aligned}
& \mathbb{P}\left(\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau\right) && (46)\\
\leq\ & \mathbb{P}\left(\widehat{F}_{eff} > min\left(c\left(\alpha, \mathbf{W}_2, \frac{B_e(\mathbf{W}, \mathbf{\Omega})}{\tau}\right),\right.\right. && (47)\\
& \left.\left. c\left(\alpha, \widehat{\mathbf{W}}_2, \frac{min(B_e(\mathbf{W}, \mathbf{\Omega}), B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}}))}{\tau}\right)\right)\right) && (48)\\
\rightarrow\ & \mathbb{P}\left(\widehat{F}^*_{eff} > c\left(\alpha, \mathbf{W}_2, \frac{B_e(\mathbf{W}, \mathbf{\Omega})}{\tau}\right)\right) && (49)\\
=\ & \alpha && (50)
\end{aligned}
$$

Now we prove the second part of the Lemma. We first prove a bound for the critical values. Let $F^{-1}_{\chi^2_d(x)}(\alpha)$ the upper α point of a non-central χ^2 with d degrees of freedom and noncentrality parameter x. For any $\alpha \in [0, 1]$
$c(\alpha, \mathbf{W}_2, x) \leq x^* \equiv \left(\sqrt{max\left(F^{-1}_{\chi^2_1(0)}(\alpha), F^{-1}_{\chi^2_2(0)/2}(\alpha), ..., F^{-1}_{\chi^2_K(0)/K}(\alpha)\right)} + \sqrt{x}\right)^2$.

Let $X_i \sim N(0, 1)$ i.i.d., $i = 1, 2..., K$, and let $c \in A$ where $A = \{c \in \mathbb{R}^K \quad | \quad \sum_{i=1}^K c_i = 1, \quad c_i \geq 0, \quad \forall i\}$. From Szekely and Bakirov (2003) $\tilde{x} \in \mathbb{R}_+$ that $\inf_{c \in A} P(\sum_{i=1}^K c_i X_i^2 \leq \tilde{x}) = P(\chi^2_n / n(\tilde{x}) \leq \tilde{x})$, where the function $n(\tilde{x})$ is integer, nondecreasing, bounded by K and equal to 1 whenever $\tilde{x} > 1.536$. Let $Q = \sum_{i=1}^K c_i(X_i + b_i)^2$ a quadratic form in normal random variables and write $\sum_{i=1}^K c_i b_i^2 = \mu^2$. From the triangle inequality

$$\mathbb{P}[Q > x] = \mathbb{P}\left[\sum_{i=1}^{K} c_i(X_i + b_i)^2 > x\right] \leq \mathbb{P}\left[\left(\sqrt{\sum_{i=1}^{K} c_i X_i^2} + \mu\right)^2 > x\right]$$

Whenever $x > \mu^2$ then $\mathbb{P}[Q > x] \leq \mathbb{P}\left[\chi^2_{n(x_1)}/n(x_1) > x_1(\mu^2, x)\right]$, where $x_1(\mu^2, x) = (x^{1/2} - \mu)^2$. Moreover, this bound is increasing in μ^2 whenever $x > \mu^2$. Let x^* as above. Then $x_1(x, x^*) = \max\left(F^{-1}_{\chi^2_1(0)}(\alpha), F^{-1}_{\chi^2_2(0)/2}(\alpha), ..., F^{-1}_{\chi^2_K(0)/K}(\alpha)\right)$. Therefore, for $\mu^2 \leq x$

$$P[Q > x^*] \leq \mathbb{P}(\left[\chi^2_{n(x_1)}/n(x_1) > x_1(x, x^*)\right] \leq \alpha$$

Now assume that $B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})$ is bounded in probability. Then $c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}}))$ is bounded above in probability by some c^*. Then

$$min\left[\mathbb{P}\left(\widehat{F}_{eff} > c(\alpha, \widehat{\mathbf{W}}_2, B_e(\widehat{\mathbf{W}}, \widehat{\mathbf{\Omega}})/\tau))\right), \mathbb{P}\left(\widehat{F}_{eff} > c^*\right)\right] \xrightarrow{p} \mathbb{P}\left(F^*_{eff} > c^*\right) \tag{51}$$

But then by the triangle inequality

$$\mathbb{P}\left(F^*_{eff} > c^*\right) \geq \mathbb{P}\left(\mu > \sqrt{c^*} + \sqrt{\sum_{i=1}^{K} c_i X_i^2}\right) \tag{52}$$

where c_i are the eigenvalues of $\mathbf{W}_2$ and X_i are iid standard normal. The right-hand side in (52) clearly converges to 1 as $\mu^2 \to \infty$, proving the second part of the Lemma.

FIGURE CAPTIONS

Figure 1: TSLS and simplified 5% critical values assuming conditional homoskedasticity, no serial autocorrelation, and known $\mathbf{\Omega}$ and $\mathbf{W} = \mathbf{\Omega} \otimes I_K$. Under these assumptions, the effective number of degrees of freedom K_{eff} equals the number of instruments K, and the

effective F statistic equals the non-robust first stage F statistic. The null hypothesis is that the estimator Nagar bias exceeds 10% of the benchmark. Critical values are computed using the Patnaik (1949) methodology. For comparison, we show Stock and Yogo (2005) 5% critical values of the null hypothesis that the asymptotic TSLS bias exceeds 10% of the OLS bias.

SUPPLEMENTARY MATERIALS

- [A Robust Test for Weak Instruments: Supplementary Materials.]. Computational details and additional results. (PDF)
- [Files201200717.zip] MATLAB and STATA code to compute figures, tables and critical values. (Zip file)

References

Andrews, D. W. K., and Stock, J. H. (2006), "Inference with Weak Instruments," in *Advances in Economics and Econometrics, Theory and Applications (9th Congress of the Econometric Society)*, eds. R. Blundell, W. K. Newey, and T. Persson, Vol. 3 Cambridge University Press, chapter 6.

Antoine, B., and Lavergne, P. (2012), "Conditional moment models under semi-strong identification," *Working Paper, Simon Fraser University*, .

Bansal, R., and Yaron, A. (2004), "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

Baum, C. F., Schaffer, M. E., and Stillman, S. (2007), "Enhanced routines for instrumental variables/generalized method of moments estimation and testing," *The Stata Journal*, 7(4), 465–506.

Bun, M., and de Haan, M. (2010), "Weak Instruments and the First Stage F-statistic in IV Models with a Nonscalar Error Covariance Structure," *University of Amsterdam*, .

Campbell, J. (2003), "Consumption-Based Asset Pricing," *Handbook of the Economics of Finance*, 1, 803–887.

Campbell, J., and Mankiw, N. (1989), "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomics Annual 1989*, eds. O. J. Blanchard, and S. Fischer, Vol. 4 MIT Press.

Chao, John C., N. R. S. J. A. H., and Newey, W. K. (2012), "Asymptotic distribution of JIVE in the heteroskedastic IV regression with many instruments," *Econometric Theory*, 28, 42–86.

Conerly, M. D., and Mansfield, E. R. (1988), "An Approximate Test for Comparing Heteroscedastic Regression Models," *Journal of the American Statistical Association*, 83(403), 811–817.

Epstein, L., and Zin, S. (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–969.

Epstein, L., and Zin, S. (1991), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99(2), 263–286.

Farebrother, R. (1990), "Algorithm AS 256: The Distribution of a Auadratic Form in Normal Variables," *Applied Statistics*, 39(2), 294–309.

Ghosh, B. (1973), "Some monotonicity theorems for χ^2, F and t distributions with applications," *Journal of the Royal Statistical Society. Series B (Methodological)*, 35(3), 480–492.

Grubbs, F. E. (1964), "Approximate Circular and Noncircular Offset Probabilities of Hitting," *Operations Research*, 12(1), 51–62.

Guggenberger, P. (2010*a*), "The impact of a Hausman pretest on the asymptotic size of a hypothesis test," *Econometric Theory*, 26, 369–382.

Guggenberger, P. (2010*b*), "The impact of a Hausman pretest on the size of a hypothesis test: The panel data case," *Journal of Econometrics*, 156, 337–343.

Hall, R. E. (1988), "Intertemporal Substitution in Consumption," *The Journal of Political Economy*, 96(2), 339–357.

Hansen, L., and Singleton, K. (1983), "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *The Journal of Political Economy*, 91(2), 249–265.

Hausman, Jerry A., W. K. N. T. W. J. C. C. N. R. S. (2012), "Instrumental variable estimation with heteroskedasticity and many instruments," *Quantitative Economics*, 3, 211–255.

Henshaw, R. C. (1966), "Testing Single-Equation Least Squares Regression Models for Autocorrelated Disturbances," *Econometrica*, 34(3), 646–660.

Imhof, J. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48(3-4), 419.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995), *Continuous Univariate Distributions*, Vol. 2 John Wiley & Sons, New York.

Kleibergen, F. (2007), "Generalizing Weak Instrument Robust IV Statistics Rowards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," *Journal of Econometrics*, 139(1), 181–216.

Koerts, J., and Abrahamse, A. (1969), *On the Theory and Application of the General Linear Model* Rotterdam University Press.

Leeb, H., and Poetscher, B. M. (2005), "Model selection and inference: Facts and fiction," *Econometric Theory*, 21, 21–59.

Liu, H., Tang, Y., and Zhang, H. (2009), "A New Chi-Square Approximation to the Distribution of Non-Negative Definite Quadratic Forms in Non-Central Normal Variables," *Computational Statistics & Data Analysis*, 53(4), 853–856.

Liu, S., and Neudecker, H. (1995), "Matrix-Trace Cauchy-Schwarz Inequalities and Applications in Canonical Correlation Analysis," *Statistical Papers*, 36(1), 287–298.

Mikusheva, A. (2010), "Robust Confidence Sets in the Presence of Weak Instruments," *Journal of Econometrics*, 157(2), 236 – 247.

Müller, U. K. (2011), "Efficient Tests Under a Weak Convergence Assumption," *Econometrica*, 79(2), 395–435.

Nagar, A. (1959), "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27(4), 575–595.

Patnaik, P. (1949), "The Non-Central Chi-Square and F-Distributions and Their Applications," *Biometrika*, 36(1/2), 202.

Pearson, E. (1959), "Note on an Approximation to the Distribution of Non-Central Chi-Squared," *Biometrika*, 46(3/4), 364.

Rothenberg, T. J. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, eds. J. J. Heckman, and E. E. Leamer, Vol. 6A North-Holland, chapter 15, pp. 881–935.

Ruud, P. (2000), *An Introduction to Classical Econometric Theory* Oxford University Press New York.

Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2(6), 110–114.

Staiger, D., and Stock, J. (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

Stock, J., Wright, J., and Yogo, M. (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518–529.

Stock, J., and Yogo, M. (2005), "Testing for weak instruments in linear IV regression, Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg," *ed. D.W. Andrews and J.H. Stock*, pp. 80–108.

Szekely, G., and Bakirov, N. (2003), "Extremal Probabilities for Gaussian Quadratic Forms," *Probability Theory and Related Fields*, 126(2), 184–202.

Theil, H., and Nagar, A. (1961), "Testing the Independence of Regression Disturbances," *Journal of the American Statistical Association*, 56(296), 793–806.

Van der Vaart, A. (2000), *Asymptotic Statistics* Cambridge University Press.

Yeh, J. (2000), *Lectures on Real Analysis* World Scientific, Singapore.

Yogo, M. (2004), "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economics and Statistics*, 86(3), 797–810.

Zhan, Z. (2010), "Detecting Weak Identification by Bootstrap," *Brown University*, .

Zhang, F. (2010), *Matrix Theory: Basic Results and Techniques, Second Edition* Springer.