Manuscript submitted to *The Econometrics Journal*, pp. 1–28.

Testing overidentifying restrictions with many instruments and heteroskedasticity using regularized Jackknife IV

Marine Carrasco[†] and Mohamed Doukali[‡]

[†] *Department of Economics, University of Montreal, CP 6128, succ Centre Ville, Montreal, QC H3C3J7, Canada.*
E-mail: marine.carrasco@umontreal.ca

[‡] *Department of Economics, McGill University, and Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Montreal, Canada.*
E-mail: mohamed.doukali@mail.mcgill.ca

Summary This paper proposes a new overidentifying restrictions test in a linear model when the number of instruments (possibly weak) may be smaller or larger than the sample size n or even infinite in a heteroskedastic framework. The proposed J test combines two techniques: the Jackknife method and the regularization technique which consists in stabilizing the projection matrix. We theoretically show that our new test achieves the asymptotically correct size in the presence of many instruments. The simulation results demonstrate that our modified J statistic test has better empirical properties in small samples than existing J tests. We also propose a regularized F-test to assess the strength of the instruments, which is robust to heteroskedasticity and many instruments.

Keywords: *Overidentification tests, Many instruments, Weak instruments, Heteroskedasticity, Regularization method.*

1. INTRODUCTION

When the number of the instruments grows, it is well known that the conventional J test for overidentifying restrictions performs poorly. It was shown that the asymptotic behavior of the conventional J test of Hansen (1982) gives a limit distribution which is not standard when the number of instruments or moment conditions is very large (see Kunitomo et al. (1983) and Burnside and Eichenbaum (1996)). Here, we focus on linear models with many instruments.

We propose a modified version of the J test which remains valid in presence of many (semi-)weak instruments and when the error is heteroskedastic. We construct our proposed test by using regularization to compute the inverse involved in the projection matrix P, instead of using the usual projection matrix (see Carrasco et al. (2007) for a review on inverse problems). For that purpose, we apply the Tikhonov regularization method, which is also known as the ridge regression. It depends on a tuning or regularization parameter α. To compute the residual of the regression, we replace the unknown regression coefficient by the regularized Jackknife IV estimator (RJIVE) proposed by Carrasco and Doukali (2017). We show that our test has correct asymptotic size provided that the regularization parameter α goes to zero at a certain rate which depends on the strength of the instruments. Interestingly, no restrictions are imposed on the number of instruments which can be larger or smaller than the sample size. In practice, the tuning parameter, α, is chosen so that it minimizes the cross-validation approximation of the mean squared error (MSE) derived in Carrasco and Doukali (2017). Our Monte Carlo study shows that our proposed J test performs favorably compared to other existing J tests. Indeed, its empirical size remains close to the theoretical one even when the number of instruments is large and its power is large.

We also develop a new test to assess the strength of the instruments. This test based on Jackknife and regularization is robust to many instruments and heteroskedasticity of the error. Following Stock and Yogo (2005), the critical value is selected so that the bias of the Jackknife estimator does not exceed 10%.

Other regularization techniques could have been used in this framework such as the Landweber-Fridman technique which is an iterative method or the principal component which consists in selecting the eigenvectors associated with the largest eigenvalues. Carrasco (2012) used those regularization techniques to estimate a linear model in the presence of many instruments in a consistent and efficient way. Carrasco and Doukali (2017) proposed a new estimator which they called the regularized Jackknife instrumental variable estimator (RJIVE) when the number of available instruments is very large in linear models.

There are many studies related to this paper. Lee and Okui (2012) proposed a modification of the Sargan (1958)'s test of overidentifying restrictions in a homoskedastic framework when the number of instruments L grows with the sample size n. They established the asymptotic null distribution of their proposed test statistic and studied its local power under some regularity conditions. Anatolyev and Gospodinov (2011) proposed a modification of the Anderson-Rubin (AR) test and of the conventional J test for overidentifying restrictions in linear models with homoskedasticity assumption under many instruments asymptotics. They consider an alternative way to compute the critical values of the chi-squared distribution. In a recent paper, Carrasco and Tchuente (2018) propose to use regularization techniques to construct a robust Anderson Rubin (AR) test in linear models when the number of instruments is large. Their inference relies on a new restricted efficient boostrap method and simulated Monte Carlo test. The closest paper to our approach is Chao et al. (2014), where they propose a new version of the J test that is robust to many instruments and heteroskedasticity. Their test is based on subtracting out the diagonal terms in the numerator of the test statistic. They consider the heteroskedasticity robust version of the Fuller (1977) estimator of Hausman et al. (2012). Here, we consider instead the regularized Jackknife instrumental variable estimator (RJIVE). We choose this estimator because of its good properties (see Carrasco and Doukali (2017) for more details) and we implement the Tikhonov technique to stabilize the projection matrix P that appears in the numerator of the test statistic in order to improve the accuracy of the overidentifying restrictions test. The advantage of the regularization is that it permits to handle the case where the number of instruments exceeds the sample size.

Our F-test for weak instruments is closely related to a recent paper by Mikusheva and Sun (2020) who propose a pre-test for weak identification which also uses Jackknife and is robust to many instruments and heteroskedasticity. However, it does not rely on regularization and hence needs to restrict the number of instruments to be smaller than the sample size.

The remainder of this paper is organized as follows. Section 2 describes the model and the test statistic. Section 3 establishes asymptotic results. Section 4 reports Monte Carlo simulation results. In Section 5, we propose a regularized F-test for weak instruments.

Empirical applications are illustrated in section 6. Section 7 concludes. All of the proofs are provided in the appendix.

2. MODEL, ESTIMATOR, AND TEST STATISTIC

This section presents the model, the estimator, and the regularized J test.
Consider the linear IV regression model:

$$y_i = X_i'\delta_0 + \epsilon_i \tag{2.1}$$

$$X_i = \Upsilon_i + U_i \tag{2.2}$$

$i = 1, \dots, n$. The vector of interest is δ_0 which is a $p \times 1$ vector for some fixed p. y_i is the scalar outcome variable. The vector Υ_i is the optimal instrument, which is typically unknown. We assume that y_i and X_i are observed but the Υ_i is not and $E(X_i\epsilon_i) \neq 0$. The estimation will be based on a sequence of instruments $Z_i = Z(\tau; \nu_i)$ where ν_i is a vector of exogenous variables and τ is an index taking countable values.
For the estimation of δ_0, we consider the Tikhonov Jackknife estimator proposed in Carrasco and Doukali (2017) because of its good properties relative to other existing IV estimators in the presence of many instruments. First we recall the expression of the Jackknife estimator (JIVE) proposed by Angrist et al. (1999) when the number of instruments is finite.

$$\hat{\delta} = (\hat{\Upsilon}'X)^{-1}(\hat{\Upsilon}'Y) \tag{2.3}$$

$$= (\sum_{i=1}^{n} \hat{\Upsilon}_i X_i')^{-1} \sum_{i=1}^{n} \hat{\Upsilon}_i y_i \tag{2.4}$$

The leave-one-out estimator $\hat{\Upsilon}_i$ is defined as $\hat{\Upsilon}_i = Z_i'\hat{\pi}_{-i}$, where $\hat{\pi}_{-i} = (Z'Z - Z_iZ_i')^{-1}(Z'X - Z_iX_i')$ is the OLS coefficient from running a regression of X on Z using all but the i^{th} observation.
The JIVE estimator can alternatively be written as:

$$\hat{\delta} = (\sum_{i=1}^{n} \hat{\pi}'_{-i} Z_i X_i')^{-1} \sum_{i=1}^{n} \hat{\pi}'_{-i} Z_i y_i \tag{2.5}$$

with

$$\hat{\pi}'_{-i} Z_i = (X'Z(Z'Z)^{-1}Z_i - P_{ii}X_i)/(1 - P_{ii}) = \textstyle\sum_{j \neq i}^{n} P_{ij}X_j/(1 - P_{ii})$$

where P is a $n \times n$ matrix defined as $P = Z(Z'Z)^{-1}Z'$ and P_{ij} denotes the *(i,j)*th element of P.
Then, the JIVE estimator is given by:

$$\hat{\delta} = \hat{H}^{-1} \sum_{i \neq j}^{n} X_i P_{ij} (1 - P_{jj})^{-1} y_j,$$

where $\hat{H} = \sum_{i \neq j}^{n} X_i P_{ij} (1 - P_{jj})^{-1} X_j'$, and $\sum_{i \neq j}$ denotes the double sum $\sum_i \sum_{j \neq i}$. When the number of the instruments is large, the inverse of $Z'Z$ needs to be regularized because it is singular or nearly singular.
Now let us suppose that the number of moment conditions is finite or countable infinite. Here are some examples of Z_i.

- If $Z_i = \nu_i$ where ν_i is a L-vector of exogenous variables with a fixed L, then $Z(\tau;\nu_i)$ denotes the τth element of ν_i.
- $Z(\tau;\nu_i) = (\nu_i)^{\tau-1}$ with $\tau \in N$, thus we have an infinite countable sequence of instruments.

We note that, unlike the other existing test statistics, the number of moment conditions is not restricted and may be smaller or larger than the sample size.
The expression of the Tikhonov Jackknife IV estimator $\hat{\delta}^\alpha$ is

$$\hat{\delta}^\alpha = \hat{H}^{-1} \sum_{i \neq j}^{n} X_i P_{ij}^\alpha (1 - P_{jj}^\alpha)^{-1} y_j, \tag{2.6}$$

$$\hat{H} = \sum_{i \neq j}^{n} X_i P_{ij}^\alpha (1 - P_{jj}^\alpha)^{-1} X_j' \tag{2.7}$$

where P^α is a $n \times n$ matrix defined as

$$P^\alpha = Z(Z'Z + \alpha I)^{-1} Z', \tag{2.8}$$

and P_{ij}^α denotes the (i,j)th element of P^α. The Tikhonov Jackknife estimator depends on a regularization term α. In practice, we choose α that minimizes the mean square error (MSE) as in Carrasco and Doukali (2017).

Remark 2.1. *It is useful to write the RJIVE as*

$$\hat{\delta}^\alpha = \hat{H}^{-1} \sum_{i,j=1}^{n} X_i C_{ji}^\alpha y_j, \tag{2.9}$$

where $\hat{H} = \sum_{i,j=1}^{n} X_i C_{ji}^\alpha X_j'$, *and* $C^\alpha = (C_{ij}^\alpha) = \begin{cases} \frac{P_{ij}^\alpha}{1-P_{ii}^\alpha} & if \quad i \neq j \\ C_{ii}^\alpha = 0 & if \quad i = j \end{cases}$. *Then, we obtain:*

$$\sqrt{n}(\hat{\delta}^\alpha - \delta_0) = \left(\frac{X'C^{\alpha\prime}X}{n}\right)^{-1} \left(\frac{X'C^{\alpha\prime}\epsilon}{\sqrt{n}}\right). \tag{2.10}$$

The test statistic.
Chao et al. (2014) proposed a modified J statistic with many instruments based on the heteroskedasticity-robust version of the Fuller (1977) estimator, which is known as HFUL estimator. Their test statistic takes the form:

$$J_{CHNSW} = \frac{\hat{\epsilon}' P \hat{\epsilon} - \sum_{i=1}^{n} P_{ii} \hat{\epsilon}_i^2}{\sqrt{\hat{V}}} + L \tag{2.11}$$

with

$$\hat{V} = \frac{\hat{\epsilon}(2)' P(2) \hat{\epsilon}(2) - \sum_{i=1}^{n} P_{ii}^2 \hat{\epsilon}_i^4}{tr(P)} = \frac{\sum_{i \neq j}^{n} \hat{\epsilon}_i^2 P_{ij}^2 \hat{\epsilon}_j^2}{L}$$

where L is the number of instruments, P is the projection matrix, $\hat{\epsilon}_i = y_i - X_i'\hat{\delta}$, $\hat{\epsilon}(2) = (\hat{\epsilon}_1^2, ..., \hat{\epsilon}_n^2)$, $P(2)$ is the n-dimensional square matrix with ijth component equal to P_{ij}^2. Note that the numerator of the test statistic, $\sum_{i \neq j}^{n} \hat{\epsilon}_i P_{ij} \hat{\epsilon}_j$, is the numerator of the traditional Sargan test without the observation i. The denominator is a heteroskedastic consistent estimator of the variance of $\sum_{i \neq j}^{n} \hat{\epsilon}_i P_{ij} \hat{\epsilon}_j$. The test rejects the null hypothesis when J_{CHNSW} is greater than the critical value of a chi-squared distribution with $L-p$ degrees of freedom. Chao et al. (2014), Anatolyev and Gospodinov (2011) and Lee and Okui (2012) have proposed tests that allow for many instruments but they impose that the number of moment conditions L cannot be larger than n, which is not the case in our present work.

In this paper, we assume that the number of moment conditions L is large relatively to n. The inverse of $Z'Z$ needs to be stabilized because it is nearly singular or even not invertible whenever $L \geq n$. The main contribution is the use of the Tikhonov regularization method to stabilize the inverse of $(Z'Z)$ in presence of many instruments. Let P^α be defined as (2.8) when the number of instruments is finite and as (7.16) in Appendix A when the number of instruments is infinite. We note here that the Tikhonov technique involves a tuning parameter α. The case $\alpha = 0$ corresponds to the case without regularization. We obtain $P^0 = P = Z(Z'Z)^\dagger Z$, where $\dagger$ denotes the Moore-Penrose generalized inverse. The regularization parameter needs to go to zero at a certain rate characterized in Section 3.

To describe our proposed test statistic, let $P^\alpha(2)$ be the n-dimensional square matrix with (i,j) element equal to $(P^\alpha_{ij})^2$.

The test statistic we propose is

$$J_{Tikh} = \frac{\hat{\epsilon}' P^\alpha \hat{\epsilon} - \sum_{i=1}^n P^\alpha_{ii} \hat{\epsilon}_i^2}{\sqrt{\hat{V}}} + tr(P^\alpha) \tag{2.12}$$

with

$$\hat{V} = \frac{\hat{\epsilon}(2)' P^\alpha(2) \hat{\epsilon}(2) - \sum_{i=1}^n (P^\alpha_{ii})^2 \hat{\epsilon}_i^4}{tr(P^\alpha)} = \frac{\sum_{i \neq j}^n \hat{\epsilon}_i^2 (P^\alpha_{ij})^2 \hat{\epsilon}_j^2}{tr(P^\alpha)}, \tag{2.13}$$

where $\hat{\epsilon}_i = y_i - X_i'\hat{\delta}^\alpha$ where $\hat{\delta}^\alpha$ is the regularized Jackknife estimator of Carrasco and Doukali (2017). It will be shown in the next section that J_{Tikh} follows asymptotically a chi-squared with $tr(P^\alpha) - p$ degrees of freedom. Let $q_r(\tau)$ be the τth quantile of chi-squared distribution with r degrees of freedom. We reject the null hypothesis of our test with the asymptotic rejection frequency β if $J_{Tikh} \geq q_{tr(P^\alpha)-p}(1-\beta)$.

Our test has the same form as Chao et al. (2014) test with the projection matrix P replaced by the regularized projection matrix P^α and the number of instruments L replaced by the trace of P^α, i.e $tr(P^\alpha)$.

3. ASYMPTOTIC DISTRIBUTION

This section presents the asymptotic theory under which we establish the limiting behaviour of our proposed test statistic in the presence of many moment conditions. We consider many weak instruments asymptotic as in Chao et al. (2014).

Let K be the covariance operator defined in Appendix A. For a finite number of instruments, $K = Z'Z/n$.

ASSUMPTION 3.1. *(i) The operator K is nuclear. (ii) There exists a constant $\bar{C}$ such that $P^\alpha_{ii} \leq \bar{C} < 1$, $i = 1, ..., n$.*

Assumption 3.1 (i) is the same as in Carrasco (2012). Condition (i) means that the eigenvalues of the covariance operator K are summable. Condition (ii) is reminiscent of Assumption 1 in Chao et al. (2014): "for some $\bar{C} < 1$, $P_{ii} < \bar{C}$, $i = 1, ..., n$". However it is much less restrictive. Indeed, $P_{ii} < \bar{C} < 1$ implies that $\sum_i \frac{P_{ii}}{n} = \frac{L}{n} < 1$, $L =$rank(Z), which restricts the number of instruments. Our condition $P_{ii}^{\alpha} \leq \bar{C} < 1$ implies that trace$(P^{\alpha}) = \sum_i q_i < n$, which implies a condition on α, where $q_j = \frac{\lambda_j^2}{\lambda_j^2+\alpha}$, and λ_j are the eigenvalues of K. Recall that from Carrasco (2012) $\sum_i q_i = O(\frac{1}{\alpha})$. So Assumption (ii) implies $\frac{1}{\alpha n} < 1$.

The next assumption allows for the presence of many weak instruments. A measure of the strength of the instruments is the concentration parameter, which can be seen as a measure of the information contained in the instruments. If one could approximate the reduced form Υ by a sequence of instruments Z, so that $X = Z'\pi + u$ where $E[u^2|Z] = \sigma_u^2$, the concentration parameter would be given by

$$\mu_n^2 = \frac{\pi' Z' Z \pi}{\sigma_u^2}.$$

The following assumption generalizes this notion.

Assumption 3.2. *$\Upsilon_i = S_n f_i/\sqrt{n}$ where $S_n = \hat{S}_n diag(\mu_{1n}, \dots, \mu_{pn})$ such that $\hat{S}_n$ is a $p \times p$ bounded matrix, the smallest eigenvalue of $\hat{S}_n \hat{S}_n'$ is bounded away from zero, for each j, either $\mu_{jn} = \sqrt{n}$ (strong identification) or $\frac{\mu_{jn}}{\sqrt{n}} \to 0$ (weak identification). Moreover $\mu_n = \underset{1<j<p}{min}\, \mu_{jn} \to \infty$ and $1/(\sqrt{\alpha}\mu_n^2) \to 0$, $\alpha \to 0$. Also there is a constant $\bar{C}$ such that $|| \sum_{i=1}^n f_i f_i'/n|| \leq \bar{C}$ and $\lambda_{min}(\sum_{i=1}^n f_i f_i'/n) \geq 1/\bar{C}$, a.s.n.*

Assumption 3.2 allows for both strong and weak instruments. If $\mu_{jn} = \sqrt{n}$, the instrument j is strong. If μ_{jn}^2 is growing slower than n, this leads to a weak identification as that of Chao and Swanson (2005). f_i defined in Assumption 3.2 is unobserved and has the same dimension as the infeasible optimal instrument, Υ_i. Then f_i can be seen as a rescaled version of this optimal instrument.

An illustration of Assumption 3.2 is as follows. Let us consider the simple linear model $y_i = z_{i1}\delta_1 + \delta_{0p} x_{i2} + \epsilon_i$, where z_{i1} is an included instruments and x_{i2} is an endogenous variable. Suppose that x_{i2} is a linear combination of the included instrumental z_{i1} and an unknown excluded instruments z_{ip}, i.e $x_{i2} = \pi_1 z_{i1} + (\frac{\mu_n}{\sqrt{n}}) z_{ip}$. The reduced form is:

$$\Upsilon_i = \begin{pmatrix} z_{i1} \\ x_{i2} \end{pmatrix} = \begin{pmatrix} z_{i1} \\ \pi_1 z_{i1} + (\frac{\mu_n}{\sqrt{n}}) z_{ip} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \pi_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \frac{\mu_n}{\sqrt{n}} \end{pmatrix} \begin{pmatrix} z_{i1} \\ z_{ip} \end{pmatrix}$$

with

$$\hat{S}_n = \begin{pmatrix} 1 & 0 \\ \pi_1 & 1 \end{pmatrix}, \ \mu_{jn} = \begin{cases} \sqrt{n} & , \quad j = 1 \\ \mu_n & , \quad j = 2 \end{cases}, \text{ with } \frac{\mu_n}{\sqrt{n}} \to 0, \text{ and } f_i = \begin{pmatrix} z_{i1} \\ z_{ip} \end{pmatrix}.$$

Assumption 3.3. *There is a constant $C > 0$ such that $(\epsilon_1, U_1), ..., (\epsilon_n, U_n)$ are independent, with $E[\epsilon_i] = 0$, $E[U_i] = 0$, $E[\epsilon_i \Upsilon_i] = 0$, $E[\epsilon_i^2] < C$, $E[||U_i||^2] \leq C$, $Var((\epsilon_i, U_i')') = diag(\Omega_i, 0)$, and $\lambda_{min}(\sum_{i=1}^n \Omega_i/n) \geq 1/C$.*

Note that (ϵ_i, U_i) are independent but not necessarily identically distributed. This assumption allows for heteroskedasticity but requires the second moment of the disturbances to be bounded. It also imposes uniform nonsingularity of the variance of the reduced form disturbances.

Assumption 3.4. *There exists a π_L such that $\sum_{i=1}^n ||f_i - \pi_L Z_i||^2/n \to 0$.*

Assumptions 3.1 and 3.4 imply that the structural parameters are identified asymptotically. Although Assumption 3.4 implies that f_i belongs to the closure of the linear span of instruments, it does not imply that f_i is a finite linear combination of the instruments.

ASSUMPTION 3.5. *There is a constant $C > 0$ such that, with probability one, $\sum_{i=1}^{n} ||f_i||^4/n^2 \to 0$, $E[\epsilon_i^4] \leq C$ and $E[||U_i||^4] \leq C$.*

Assumption 3.5 can be found in Chao et al. (2014). It simplifies the asymptotic theory in the sense that certain terms vanish asymptotically.

ASSUMPTION 3.6. *α goes to zero and $1/(\alpha\mu_n^2) \to C$ for a finite C.*

Note that Assumptions 3.1, 3.2, and 3.6 imply some restrictions on α, namely α needs to go to zero but not too fast.
Define $\sigma_i^2 = E[\epsilon_i^2]$, $H_n = \sum_i f_i f_i'/n$, $\Omega_n = \sum_i f_i f_i' \sigma_i^2/n$,
$\Psi_n = S_n^{-1} \sum_{i \neq j}^{n} (P_{ij}^{\alpha})^2 (E[U_i U_i'] \sigma_j^2 (1-P_{jj})^{-2} + E[U_i \epsilon_i](1-P_{ii})^{-1} E[U_j \epsilon_j](1-P_{jj})^{-1}) S_n'^{-1}$.

THEOREM 3.1. *Suppose that Assumptions 3.1-3.6 are satisfied. Then,*
$V_n^{-1/2}(\hat{\delta}^{\alpha} - \delta_0) \xrightarrow{d} \mathcal{N}(0, I_p)$, *where* $V_n = H_n^{-1}(\Omega_n + \Psi_n) H_n^{-1}$.

Proof: See the proof in Carrasco and Doukali (2017) (Theorem 2).

REMARK 3.1. *As in Chao et al. (2012), the term Ψ_n in the asymptotic variance of $\hat{\delta}^{\alpha}$ accounts for the presence of many instruments. The order of this term is $\frac{1}{\alpha\mu_n^2}$. So if $\frac{1}{\alpha\mu_n^2} \to 0$, the term Ψ_n vanishes asymptotically and the asymptotic variance becomes $V_n = H_n^{-1}\Omega_n H_n^{-1}$.*

THEOREM 3.2. *Let $q_{tr(P^{\alpha})-p}(1-\beta)$ be the $(1-\beta)$ quantile of a chi-square distribution with $tr(P^{\alpha}) - p$ degrees of freedom. If assumptions 3.1-3.6 are satisfied then $Pr(\hat{T} \geq q_{tr(P^{\alpha})-p}(1-\beta)) \to \beta$.*

Proof: See Appendix.
Theorem 3.2 shows that, under the many instruments asymptotic condition, our modified J test achieves the correct asymptotic critical value β. We can see this test as a specification test for the linear instrumental variables regression (see Hansen (1982)). If the model is correctly specified, all the moment conditions (including the overidentifying restrictions) should be close to zero. The novelty of our proposed test is that it is robust to many instruments in the sense that we do not make any assumption on the number of instruments.

Related Literature.
In the literature on testing overidentifying restrictions in linear models with many instruments, the J test performs poorly when one increases the number of the instruments. To deal with this problem, Anatolyev and Gospodinov (2011) proposed a new J test that guarantees the asymptotical sizes, but their test is valid only under the homoskedasticity assumption and when the number of instruments is a fraction of the sample size $0 < \frac{L}{n} < 1$. Lee and Okui (2012) proposed a modification of the Sargan (1958) test in

the presence of a large number of instruments. They gave the limiting behavior of their proposed test statistic when the number of instruments and the sample size go to infinity, but they still maintained the assumption $0 < \frac{L}{n} < 1$. Donald et al. (2003) established the asymptotic distribution of some parameter and specification tests in models when the number of instruments L increases asymptotically, but again slowly relative to the sample size n. They called this assumption a moderately many instruments, but the validity of their test fails in the case of the many instruments theory of Bekker (1994). Hahn and Hausman (2002) developed a new specification test for the validity of instrumental variables in linear models. They compared the difference of the forward (conventional) 2SLS estimator with the reverse 2SLS estimator under the assumption $0 < \frac{L}{n} < 1$.
In this paper, we consider the case when the number of instruments is potentially very large. The matrix $Z'Z$ may be nearly singular or possibly not invertible, so the projection matrix $P = Z(Z'Z)^{-1}Z'$ that appears in the numerator of the J test may affect the precision of the test statistic. Inverting $Z'Z$ can be seen as solving an ill-posed problem. We implement the Tikhonov technique to stabilize the projection matrix. The advantage of the regularization is that we can use all the available information and we do not need to discard some instruments a priori. This yields an improved performance of the J test as illustrated in the simulation study.

4. SIMULATION STUDY ON REGULARIZED J TEST

The goal of our simulation study is to demonstrate the finite-sample performance of the proposed J test and compare it to other existing J tests. We consider a linear model with one regressor and L instruments. The J statistic is interpreted as a test of the validity of the $L-1$ overidentifying restrictions. We investigate two cases: the homoskedastic and heteroskedastic case.
Homoskedastic case. The data generating process (DGP) is generated as follows:

$$\begin{aligned} y_i &= \delta X_i + \epsilon_i \\ X_i &= z_i'\pi + u_i, \end{aligned}$$

where $(\epsilon_i, u_i) \overset{iid}{\sim} N(0, \sum)$ and $\sum = \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{pmatrix}$, $z_i \overset{iid}{\sim} N(0, I_L)$, $\delta = 1$, and $\pi = \frac{1}{\sqrt{L}}\iota_L$, where ι_L is an L-vector of ones.
Heteroskedastic case. Now the error is allowed to be heteroskedastic. We keep the same DGP except that the errors are now generated as follows:
$u_i \overset{iid}{\sim} N(0,1)$, $\epsilon_i = \rho u_i + \sqrt{\frac{1-\rho^2}{\phi^2+0.86^4}}(\phi v_{1i} + 0.86 v_{2i})$, where $v_{1i} \overset{iid}{\sim} N(0, z_{1i}^2)$ and $v_{2i} \overset{iid}{\sim} N(0, (0.86)^2)$. We choose $\rho = 0.3$, $\phi = 0.2$.
Tables 1 and 2 present the empirical size at 5% nominal level of J, J_{Corr}, J_{CHNSW} and J_{Tikh} tests which denote respectively the conventional J test, the modified J test proposed in Anatolyev and Gospodinov (2011), the modified J test proposed in Chao et al. (2014), and the Tikhonov J test proposed in this paper. These results are based on 5000 Monte Carlo replications. We consider values of $\lambda = \frac{L}{n}$ equal to 0.2, 0.5, 0.8, 0.95, and 1.1. The values of λ are used in combination with sample sizes of 100, 200 and 500. For the Tikhonov J test, the regularization parameter α is chosen by minimizing[1] the cross-validation approximation of the mean squared error (MSE) as in Carrasco and

[1]The regularization parameter α is searched over the interval [0.01,0.5] with 0.01 increment.

Doukali (2017) (Equation 7):

$$\hat{S}(\alpha) = \hat{\sigma}_{\varepsilon}^2 \frac{1}{n} \|X - C^{\alpha} X\|^2 + \hat{\sigma}_{u\varepsilon}^2 \frac{tr(C^{\alpha 2})}{n}$$

where $\hat{\sigma}_{\varepsilon}^2$ and $\hat{\sigma}_{u\varepsilon}^2$ are consistent estimators of σ_{ε}^2 and $\sigma_{u\varepsilon}^2$.

Description of the other tests:

Hansen-Sargan J test.

Let $\hat{\delta}_{2SLS} = (X'PX)^{-1}X'Py$ be the two stage least-squared estimator and $\hat{\epsilon} = y - X\hat{\delta}_{2SLS}$. The Hansen-Sargan J test takes the following form:

$$J = \frac{\hat{\epsilon}' P \hat{\epsilon}}{\hat{\sigma}^2}, \tag{4.14}$$

with $\hat{\sigma}^2 = \hat{\epsilon}'\hat{\epsilon}/(n-p)$. The decision rule of Hansen-Sargan J test consists in rejecting the null hypothesis if J exceeds the critical value given by the chi-square distribution with $L-p$ degrees of freedom.

Anatolyev and Gospodinov (2011)'s J test.

They suggest to use the same J statistic as in (4.14) with $\hat{\epsilon} = y - X\hat{\delta}_{LIML}$ where $\hat{\delta}_{LIML}$ is the limited information maximum likelihood estimator of δ but the critical value is modified. The decision rule consists in rejecting H_0 at the level β if J exceeds the quantile of a chi-square distribution with $L-p$ degrees of freedom and probability $\Phi(\sqrt{1 - \frac{L}{n}}\Phi^{-1}(\beta))$, where Φ is the distribution function of the standard normal.

Chao et al. (2014)'s J test.

J_{CHNSW} uses the test described in Equation (2.11) with $\hat{\epsilon} = y - X\hat{\delta}_{HFULL}$, where $\hat{\delta}_{HFULL}$ is the heteroskedasticity-robust version of the Fuller (1977) estimator of Hausman et al. (2012).

Table 1. Empirical rejection rates at 0.05 nominal level of the J test - homoskedastic case

λ	0.2	0.5	0.8	0.95	1.1
$n = 100$					
J	0.059	0.017	0	0	NA
J_{Corr}	0.069	0.056	0.049	0.049	NA
J_{CHNSW}	0.072	0.057	0.049	0.057	NA
J_{Tikh}	0.073	0.066	0.070	0.062	0.063
$n = 200$					
J	0.054	0.027	0	0	NA
J_{Corr}	0.058	0.059	0.045	0.040	NA
J_{CHNSW}	0.061	0.057	0.043	0.029	NA
J_{Tikh}	0.063	0.063	0.059	0.061	0.053
$n = 500$					
J	0.058	0.044	0	0	NA
J_{Corr}	0.055	0.057	0.047	0.039	NA
J_{CHNSW}	0.056	0.056	0.047	0.031	NA
J_{Tikh}	0.056	0.062	0.053	0.050	0.048

Tables 1 and 2 report the empirical sizes of the four tests in the homoskedastic and the heteroskedastic cases respectively. We remark that the performance of the conventional J test is sensitive to the number of instruments, i.e. the rejection frequencies for the J test is not close to the nominal value 5% throughout these tables. We also remark that

Table 2. Empirical rejection rates at 5% nominal level of the J test - heteroskedastic case

λ	0.2	0.5	0.8	0.95	1.1
$n = 100$					
J	0.049	0.007	0	0	NA
J_{Corr}	0.062	0.047	0.070	0.045	NA
J_{CHNSW}	0.067	0.050	0.073	0.098	NA
J_{Tikh}	0.069	0.058	0.046	0.046	0.050
$n = 200$					
J	0.044	0.012	0	0	NA
J_{Corr}	0.056	0.046	0.053	0.067	NA
J_{CHNSW}	0.058	0.046	0.051	0.082	NA
J_{Tikh}	0.061	0.055	0.046	0.050	0.046
$n = 500$					
J	0.045	0.016	0	0	NA
J_{Corr}	0.054	0.052	0.046	0.067	NA
J_{CHNSW}	0.054	0.051	0.045	0.064	NA
J_{Tikh}	0.057	0.052	0.049	0.047	0.046

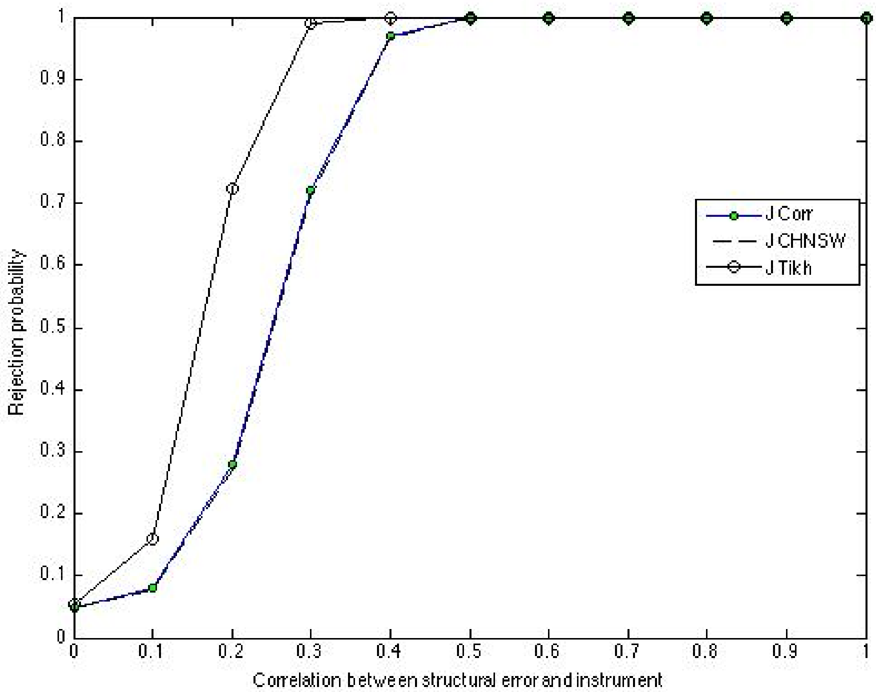

Figure 1. Power curves of J tests, n=500, $\lambda = 0.8$, homoskedastic case.

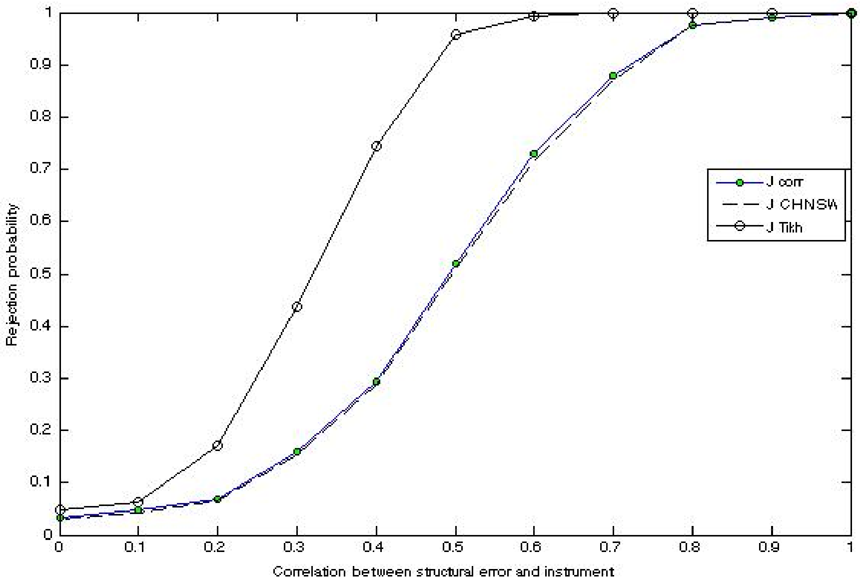

Figure 2. Power curves of J tests, n=500, $\lambda = 0.8$, heteroskedastic case.

Anatolyev and Gospodinov (2011)'s J test, the J_{CHNSW} and the J_{Tikh} perform very well when the number of instruments increase as long as L is not too large. However, J, J_{Corr}, and J_{CHNSW} tests exhibit a large size distortion when λ is close to 1 (i.e. λ= 0.95), which is worse in the heteroskedastic case. Our regularized J_{Tikh} has almost correct size even with very large number of instruments. When the number of instruments is larger than the sample size, the J, J_{Corr}, and J_{CHNSW} cannot be computed. Tables 1 and 2 show also that our proposed regularized J test performs well when $L > n$, in the sense that the empirical rejection rates are close to the nominal value 5%.

To compare the powers of the different J tests, we consider the same design as before, but the structural error is giving by: $\xi_i = \epsilon_i + \rho_z z_{1i}$. We allow the correlation ρ_z between structural error and instrument to vary between 0 and 1. We choose $n = 500$ and $\lambda = 0.8$. The rejection frequencies under the null hypothesis (ρ_z=0) are 0.047, 0.047, 0.053 respectively for J_{Corr} J_{CHNSW} and the J_{Tikh} for homoskedastic case. For the heteroskedastic case they are 0.046, 0.045, 0.049. The power curves (rejection frequencies) are plotted in Figures 1 and 2. We see that J_{Tikh} statistic has clearly better power properties than the J_{Corr} and J_{CHNSW}.

In conclusion, simulations suggest that the implementation of the Tikhonov regularization can increase the power, while controlling for the size. Thus, the regularization provides a correction to size distortions for the J test arising from the use of many instruments.

5. DETECTION OF WEAK INSTRUMENTS

In this section, we propose a regularized F-test to assess the strength of the instruments in the first stage equation. We will consider the case where there is a single endogenous regressor (case where δ is scalar) and we will use the notations x_i and u_i to emphasize the fact that X_i and U_i are scalar. The first stage equation is then

$$x_i = \Upsilon_i + u_i = \pi' z_i + u_i$$

where $\Upsilon_i = \pi' z_i$ and π is a $L \times 1$ vector. When the number of instruments is countable infinite, then

$$x_i = \langle \pi (.) , z_i (.) \rangle + u_i$$

where $\langle , \rangle$ denotes the inner product in $L^2 (\omega)$ for some pdf ω and π and z_i are elements of $L^2 (\omega)$ (see Appendix A for more details). The remaining of the section will present the test using vector notations.

First, we develop a test for $H_0 : \pi = 0$. We propose a F-test robust to heteroskedasticity and many instruments.

$$F_{Tikh} = \frac{\sum_{i=1}^n \sum_{j \neq i} P_{ij}^\alpha x_i x_j}{\sqrt{2 \sum_{i=1}^n \sum_{j \neq i} \left(P_{ij}^\alpha\right)^2 \widehat{u}_i^2 \widehat{u}_j^2}}$$

where $\hat{u} = (I - P^\alpha) X = X - Z\widehat{\pi}^\alpha$, $\widehat{\pi}^\alpha = (Z'Z + \alpha I)^{-1} Z'X$ is the ridge estimator of π. Let

$$\gamma^2 = \frac{\frac{1}{n} \sum_{i=1}^n \sum_{j \neq i} \widetilde{\pi}' z_i P_{ij}^\alpha z_j' \widetilde{\pi}}{\sqrt{2 \sum_{i=1}^n \sum_{j \neq i} \left(P_{ij}^\alpha\right)^2 E\left(u_i^2\right) E\left(u_j^2\right)}}.$$

Assumption 5.1. *(a) Υ_i satisfies the condition*

$$\frac{\sum_{i=1}^n \left| \sum_{j \neq i} P_{ij}^\alpha \Upsilon_j \right|^3}{\left(\sum_{i=1}^n \left(\sum_{j \neq i} P_{ij}^\alpha \Upsilon_j \right)^2 E\left(u_i^2\right) \right)^{3/2}} \underset{n \to \infty}{\to} 0.$$

(b) Let $\Upsilon (z_i) \equiv \Upsilon_i$, $\hat{\Upsilon}^\alpha (z_i) \equiv \widehat{\pi}^{\alpha\prime} z_i$. Let D be the domain of the distribution of z_i. Then,

$$\sup_{z \in D} \left| \Upsilon (z) - \hat{\Upsilon}^\alpha (z) \right| \xrightarrow{P} 0.$$

Assumption 5.1(a) is a Lyapunov's condition needed in the proof of the asymptotic normality of F_{Tikh}. Assumption 5.1(b) is used to show that $\sum_{i=1}^n \sum_{j \neq i} \left(P_{ij}^\alpha\right)^2 \widehat{u}_i^2 \widehat{u}_j^2$ is a consistent estimator of $\sum_{i=1}^n \sum_{j \neq i} \left(P_{ij}^\alpha\right)^2 E\left(u_i^2\right) E\left(u_j^2\right)$, once rescaled. It is satisfied under some regularity conditions on $\Upsilon (.)$, see Carrasco, Florens, and Renault (2007) and Hall and Horowitz (2007). Both conditions imply restrictions on the rate of convergence of α depending on how regular (or smooth) the function Υ is.

Theorem 5.1. *Let $q_\gamma (1 - \beta)$ be the $1 - \beta$ quantile of a normal distribution with mean γ^2 and variance 1. Assume Assumption 3.1 and 3.7 hold, that u_i is independent with mean 0 and there exists a constant $C > 0$ such that $E\left(u_i^4\right) < C$, and that $\alpha \to 0$ as n goes to infinity. Under the weak instrument assumption $\pi = \widetilde{\pi}/\sqrt{n}$, we have*

$$P_r \left(F_{Tikh} \geq q_\gamma (1 - \beta)\right) \to \beta$$

as n goes to infinity.

Remark 5.1. *1. The expression of γ^2 may seem complicated. However, it can be bounded* *by a simple expression. Using*

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\widetilde{\pi}' z_i P_{ij}^{\alpha} z_j' \widetilde{\pi} &= \frac{1}{n}\sum_{i=1}^{n}\widetilde{\pi}' z_i z_i' \left(Z'Z+\alpha I\right)^{-1}\sum_{j\neq i} z_j z_j' \widetilde{\pi} \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\widetilde{\pi}' z_i z_i' \widetilde{\pi} \\
&= \widetilde{\pi}' \left(\frac{Z'Z}{n}\right)\widetilde{\pi}.
\end{aligned}$$

We obtain

$$\gamma^2 \leq \frac{\widetilde{\pi}'\left(\frac{Z'Z}{n}\right)\widetilde{\pi}}{\sqrt{2\sum_{i=1}^{n}\sum_{j\neq i}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}}.$$

This upper bound is equal to

$$\frac{\widetilde{\pi}'\left(\frac{Z'Z}{n}\right)\widetilde{\pi}}{\sqrt{2V(u_i)^2\left(tr\left(P^{\alpha 2}\right)-\sum_{i=1}^{n}P_{ii}^{\alpha 2}\right)}}$$

in the homoskedastic case. We recognize the usual concentration parameter normalized by a term which is of the same order as $\sqrt{trP^{\alpha}}$, i.e. $1/\sqrt{\alpha}$.
2. The expression of the test statistic is similar to that of Mikusheva and Sun (2020) (Equation 5). The main difference is in the numerator where they use a different estimator of the variance based on cross-fit. They derive the joint distribution of the Wald test on δ and the F-test in order to control the size of the two step procedure using the F-test as pre-test. Here, we will not investigate the Wald test. Another difference with Mikusheva and Sun (2020) is that we use regularization which permits to handle an arbitrary number of instruments, while, in their paper, the number of instruments has to be smaller than the sample size.
3. The term γ^2 is nonnegative for L large enough so that the test can be treated as a one-sided test.

An important question is which critical value to use. The critical value based on $\pi = 0$ (similarly on $\gamma^2 = 0$) would be too small as it is well-known that the estimators of δ have bad properties when π is close to zero. We follow Stock and Yogo (2005) and motivate our choice of the critical value based on the bias. We wish that the absolute bias of the Jackknife estimator does not exceed 10%. Here, we focus on JIVE2 estimator proposed by Angrist, Imbens, and Krueger (1999) because it has a simpler expression than the JIVE. The regularized version of the JIVE2 estimator is given by

$$\hat{\delta}_{JIV2} = \left(\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^{\alpha}x_i x_j\right)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^{\alpha}x_i y_j.$$

To characterize the value of γ^2 yielding a 10% bias, we need to restrict ourselves to the case with normal errors and constant correlation.

Assumption 5.2.

$$\left(\begin{array}{c}\epsilon_i\\ u_i\end{array}\right) \sim iidN\left(\left(\begin{array}{c}0\\0\end{array}\right), \left(\begin{array}{cc}\sigma_{\epsilon i} & \sigma_{\epsilon ui}\\ \sigma_{\epsilon ui} & \sigma_{ui}\end{array}\right)\right)$$

and $\sigma_{\epsilon ui}/\left(\sigma_{\epsilon i}\sigma_{ui}\right) = \rho$ *does not depend on* i.

Ideally, we would like to compute the absolute bias:

$$B = \lim_{n\to\infty}\left|E\left(\hat{\delta}_{JIV2}\right) - \delta\right|.$$

But caution is in order here because the JIVE estimator does not have any moments, see Davidson and MacKinnon (2007). The regularization may help in that matter, for instance Carrasco and Tchuente (2015) show that the regularized LIML estimator has moments under certain conditions. However, it is not clear whether the regularized JIVE estimator has moments. So instead of computing B, we compute the bias of the leading terms of the distribution of $\hat{\delta}_{JIV2} - \delta$ using an Edgeworth expansion similar to that of Rothenberg (1984, p.920). Montiel Olea and Pflueger (2013) use a similar approach based on Nagar approximation in the context of a finite number of weak instruments.

Theorem 5.2. *Under the assumptions of Theorem 5.1 and assuming Assumption 5.2 holds, the asymptotic absolute bias based on the leading terms is given by*

$$B_{LT} = \left|\frac{\rho}{\gamma^4}\right|$$

where ρ *is the correlation between* u_i *and* ϵ_i.

Remark 5.2. *1. Interestingly, the asymptotic bias depends on* α *and the number of instruments, only through* γ^4.
2. The instruments will be deemed strong if they lead to a bias smaller than 10%. Given $|\rho| \leq 1$, *we obtain a bias* $B_{LT} \leq 0.1$ *for* $\gamma^2 = \sqrt{10}$. *This value of* γ^2 *is an upper bound and could be quite a bit smaller if* ρ *is small. We can deduce the critical value of the* F_{Tikh} *with level 5% by adding* γ^2 *to 1.64. If* F_{Tikh} *exceeds this critical value, 4.8, we can conclude that the instruments are strong enough to lead to a reliable estimation of* δ.
3. In the weak instrument literature, it is customary to consider the relative bias with respect to the ordinary least-squares estimator (OLS), namely $\lim_{n\to\infty}\left|\frac{E\left(\hat{\delta}_{JIV2}\right)-\delta}{E\left(\hat{\delta}_{OLS}\right)-\delta}\right|$ *to determine the critical value for the F test. However, this ratio would depend on* σ_u/σ_ϵ *which is not estimable. Therefore, we use the absolute bias instead of relative bias. Stock and Yogo (2005) mention that both measures can be used interchangeably.*

As an illustration, we performed a small simulation. The model is as in (2.1) and (2.2) with $\delta = 1$, $(\epsilon_i, u_i) \overset{iid}{\sim} N(0,\Sigma)$, $\Sigma = \begin{pmatrix}\sigma_\epsilon^2 & \sigma_{\epsilon u}\\ \sigma_{\epsilon u} & \sigma_u^2\end{pmatrix}$, with $\sigma_{\epsilon u} = 0.5$, $\sigma_\epsilon^2 = 1$, and $\sigma_u^2 = 1$ $z_i \overset{iid}{\sim} N(0, I_L)$ and $\pi = \frac{c}{\sqrt{n}}\iota_L$, where c is chosen such that the absolute bias is near to 0.1, and ι_L is an L-vector of ones. We set the sample size $n = 500, 800$, and 1000 and show the results in Table 3 for 1,000 Monte Carlo replications. We report the mean and standard deviation of the proposed F-test, the rejection frequency of the proposed F-test, the absolute mean bias of the JIVE estimator and of the OLS estimator, the parameter

γ^2, and the concentration parameter CP. The regularization parameter α is selected by minimizing the MSE for the first simulation, then this value of α is kept fixed for the other simulations. This chosen α may be too small in same cases explaining why the standard deviation of F_{Tikh} is larger than 1. We find that the rejection frequency of the F-test using our critical value is near to 5% at the 5% nominal level. Table 4 reports the same statistics for two cases where γ^2 is larger. We observe that our F test displays good power in these cases.

Table 3. Simulations results when $\pi = \frac{c}{\sqrt{n}}\iota_L$.

n	L	F_{Tikh} mean	F_{Tikh} st.	Rej freq	OLS bias	$JIVE$ bias	γ^2	CP
500	250	1.54	1.67	3.1%	0.76	0.108	1.29	30
800	450	1.42	1.73	3.2%	0.77	0.098	1.12	35
1000	800	1.51	1.95	5.4%	0.77	0.103	0.94	40

Table 4. Simulations results under the alternative

n	L	F_{Tikh} mean	F_{Tikh} st.	Rej freq	OLS bias	$JIVE$ bias	γ^2	CP
500	250	11.04	2.87	99.3%	0.58	0.02	9.91	230
800	450	9.56	2.52	98.3%	0.63	0.04	7.99	250

6. EMPIRICAL APPLICATIONS.

6.1. Institutions and Growth

We consider the empirical work of Hall and Jones (1999). In their paper, they argue that the difference between output per worker across countries is mainly due to the differences in institution and government policies - the so-called social infrastructure. They write "Countries with corrupt government officials, severe impediments to trade, poor contract enforcement, and government interference in production will be unable to achieve levels of output per worker anywhere near the norms of western Europe, northern America, and eastern Asia." Their linear IV model is given as follows.

$$y = c + \delta S + \epsilon$$
$$S = b + \beta' Z + u$$

where y is an $n \times 1$ vector of log income per capita, S is $n \times 1$ vector which is the proxy for social infrastructure, c, b and δ are scalars. Z is an $n \times L$ matrix of instruments. Hall and Jones (1999) use four instruments $Z = (EnL, EuL, Lt, FR)$, where EnL is the fraction of population speaking English at birth, EuL is the fraction of population speaking one of the five major European languages at birth, Lt is the distance from the equator, and the geography-predicted trade intensity (FR). These instruments are intended to capture the influences of colonial origin on current institutional quality. To address the issue of weak identification, we increased the number of instruments from 4 to 38 by including interactions and power functions[2]. The use of many instruments increased the concentration parameter (a measure of the weakness of instruments) from $\hat{\mu}_n^2 = 28.6$ to $\hat{\mu}_n^2 = 80.05$. We apply our proposed $F-$test to assess whether instruments are weak. We find that the regularized $F-$test (58.76) is larger than the critical value 4.8, which means that the instruments are strong enough. As the regularized JIVE estimator corrects the bias due to the use of many instruments, the JIVE should provide better point estimates. We use a sample of 79 countries for which no data were imputed[3].

Table 5 reports the test statistics corresponding to different J tests. We find that the conventional J test, the J_{corr}, and the J_{CHNSW} are larger than chi-square critical value, which means that the null hypothesis is rejected. However, our proposed Tikhonov J test is smaller than the chi-square critical value, then we can conclude that the model is correctly specified.

It may seem surprising that the J_{Tikh} is so much smaller than other J tests. One possible explanation is the presence of heteroskedasticity. The errors are found to be heteroskedastic according to the F test (p-value= 0). The J and J_{Corr} are not robust to heteroskedasticity which may explain the difference of conclusions. However, J_{CHNSW} is robust to heteroskedasticity. An explanation for the difference between J_{CHNSW} and J_{Tikh} may be that the matrix $Z'Z$ is very ill-conditioned. The condition number[4], which is the ratio of the largest eigenvalue on the smallest eigenvalue of $\underline{Z}'\underline{Z}/n$, is an indicator on how ill-posed the matrix $\underline{Z}'\underline{Z}/n$. The higher the condition number, the more imprecise the inverse of $\underline{Z}'\underline{Z}/n$ will be. The smallest possible condition number is 1 (which corresponds to the identity matrix). In this application, the condition number is equal to $3.42\ 10^{16}$.

Table 5. Estimated J statistics for the Institutions' Model.

	J	J_{Corr}	J_{CHNSW}	J_{Tikh}
J **statistic**	361.56	361.56	144.11	22.93

Note: The chi-square critical value= 52.19 (level=5% and the degree of freedom=37). Critical value of the J_{corr}= 47.22 (level=5% and the degree of freedom=37). $tr(P^{\alpha}) = 15.51$, the critical value for the $J_{Tikh} = 23.04$.

[2]The 38 instruments used in our regression are derived from Z and are given by $\underline{Z} = [Z, Z.^2, Z.^3, Z.^4, Z.^5, Z.^6, Z.^7, Z.^8, Z(:,1)\star Z(:,2), Z(:,1)\star Z(:,3), Z(:,1)\star Z(:,4), Z(:,2)\star Z(:,3), Z(:,2)\star Z(:,4), Z(:,3)\star Z(:,4)]$. All the instruments are standardized, which means that the instruments are divided with their standard deviation. Such standardizations are customary whenever regularizations are used, see for instance De Mol et al. (2008), and Stock and Watson (2012).

[3]The data were downloaded from Charles Jones' webpage: https://web.stanford.edu/~chadj/HallJones400.asc

[4]The condition number is scale invariant.

6.2. Elasticity of intertemporal substitution

The elasticity of intertemporal substitution (EIS) in consumption is crucial in macroeconomics and finance. We follow the specification in Yogo (2004)[5] who analyzes the problem of the estimation of the EIS using the linearized Euler equation.
The estimated model is as follows:

$$\Delta c_{t+1} = \tau + \psi r_{f,t+1} + \xi_{t+1} \tag{6.15}$$

$$r_{f,t+1} = \mu + \frac{1}{\psi}\Delta c_{t+1} + \eta_{t+1}, \tag{6.16}$$

where ψ is the EIS, Δc_{t+1} is the consumption growth at time $t+1$, $r_{f,t+1}$ is the real return on a risk free asset, τ and μ are constants, and ξ_{t+1} and η_{t+1} are the innovations to consumption growth and asset return respectively.
Yogo (2004) explains how weak instruments have been the cause of the EIS empirical puzzle. He shows that, using conventional IV methods, the estimated EIS, ψ, is significantly less than 1 but its reciprocal is not different from 1. Carrasco and Tchuente (2015) estimate EIS using regularized LIML estimator. They increase the number of instruments[6] from 4 to 18 by including interactions and power functions. As a result, the concentration parameters is increased in the following way: from $\hat{\mu}_n^2 = 11.06$ to $\hat{\mu}_n^2 = 68.77$ for model (6.14) and from $\hat{\mu}_n^2 = 9.66$ to $\hat{\mu}_n^2 = 33.54$ for model (6.15). We apply our regularized F$-$test, and find that its value[7] is larger than the critical value 4.8 for models (6.14) and (6.15). We conclude that the instruments are strong enough. Moreover, the point estimates are similar to those used for macro calibrations.
According to Table 6, the J statistic of the conventional J test, the J_{corr}, and J_{CHNSW} are larger than chi-square critical value, which means that they reject the null hypothesis. However, the Tikhonov J test is smaller than the chi-square critical value. We can conclude that the model is correctly specified according to our proposed test, so the instruments used in the model seem to be exogenous. The difference in the conclusion may be due to the fact that the matrix $\underline{Z}'\underline{Z}$ is very ill-conditioned. In this application, the condition number is equal to $5.06\ 10^5$. Moreover, the F test for heteroskedasticity reveals that the errors are heteroskedastic (p-value=0).

Table 6. Estimated J statistics for the EIS Model.

	J	J_{Corr}	J_{CHNSW}	J_{Tikh}
ψ	34.46	34.84	32.68	1.09
1/ ψ	48.79	56.30	41.48	0.49

Note: The chi-square critical value= 26.29 (level=5% and the degree of freedom=16). Critical value of the J_{corr}= 25.72 (level=5% and the degree of freedom=16). $tr(P^{\alpha}) = 1.88$, the critical value for $J_{Tikh} = 3.53$.

[5] Yogo (2004) used quarterly data from 1947.3 to 1998.4 for the United States.

[6] The instruments used by Yogo (2004) are: the twice lagged, nominal interest rate (r), inflation (i), consumption growth (c) and log dividend rate (p). We denote this bloc of instruments by Z=[r, i, c, p]. The 18 instruments used in our regression are derived from Z and are given by $\underline{Z} = [Z, Z.^2, Z.^3, Z(:,1) \star Z(:,2), Z(:,1) \star Z(:,3), Z(:,1) \star Z(:,4), Z(:,2) \star Z(:,3), Z(:,2) \star Z(:,4), Z(:,3) \star Z(:,4)]$.

[7] The value of our proposed F$-$test for weak instruments is 7.14 for model (6.15) and 99.02 for model (6.16).

7. CONCLUSION

The J test for overidentifying restrictions is a popular test to assess the correct specification of a model. However, it exhibits important size distortions when the number of instruments is large. This paper proposes a new J test, based on Tikhonov regularization and studies its properties under many possibly weak instruments and heteroskedasticity. Simulations results show that the proposed test performs very well. Its empirical size is close to the theoretical size and its power is greater than that of competing tests. We recommend the use of this modified J test in applied studies because of its ease of implementation and its robustness. We also propose a regularized F-test robust to heteroskedasticity and many instruments to assess the strength of instruments.

ACKNOWLEDGEMENTS

The authors are grateful for comments from Victor Chernozhukov, Prosper Dovonon, Benoit Perron, and a referee. We benefited from discussions with participants at the CIREQ Econometrics Conference in honor of Jean-Marie Dufour, the economics department at Pompeu Fabra University, and the SCSE2018. Carrasco thanks SSHRC for partial financial support.

REFERENCES

Anatolyev, S. and N. Gospodinov (2011). Specification testing in models with many instruments. *Econometric Theory 27*(02), 427–441.

Angrist, J., G. Imbens, and A. Krueger (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics* (14), 57–67.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica: Journal of the Econometric Society*, 657–681.

Burnside, C. and M. Eichenbaum (1996). Small-sample properties of gmm-based wald tests. *Journal of Business & Economic Statistics 14*(3), 294–308.

Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics 170*(2), 383–398.

Carrasco, M. and M. Doukali (2017). Efficient estimation using regularization jackknife estimator. *Annals of economics and statistics* (128), 109–149.

Carrasco, M., J.-P. Florens, and E. Renault (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics 6*, 5633–5751.

Carrasco, M. and G. Tchuente (2015). Regularized LIML for many instruments. *Journal of Econometrics 186*(2), 427–442.

Carrasco, M. and G. Tchuente (2018). Regularization based Anderson Rubin tests for many instruments. *Mimeo*.

Chao, J. C., J. A. Hausman, W. K. Newey, N. R. Swanson, and T. Woutersen (2014). Testing overidentifying restrictions with many instruments and heteroskedasticity. *Journal of Econometrics 178*, 15–21.

Chao, J. C. and N. R. Swanson (2005). Consistent estimation with a large number of weak instruments. *Econometrica 73*(5), 1673–1692.

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012). Asymptotic distribution of jive in a heteroskedastic iv regression with many instruments. *Econometric Theory*, 42–86.

De Mol, C., D. Giannone, and L. Reichlin (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics 146*(2), 318–328.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics 117*(1), 55–93.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica: Journal of the Econometric Society*, 939–953.

Hahn, J. and J. A. Hausman (2002). A new specification test for the validity of instrumental variables. *Econometrica 70*(1), 163–189.

Hall, P. and J. Horowitz (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Hall, R. E. and C. I. Jones (1999). Why do some countries produce so much more output per worker than others? *The quarterly journal of economics 114*(1), 83–116.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1029–1054.

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics 3*(2), 211–255.

Kress, R. (1999). Linear integral equations, volume 82 of applied mathematical sciences.

Kunitomo, N., K. Morimune, and Y. Tsukuda (1983). Asymptotic expansions of the distributions of the test statistics for overidentifying restrictions in a system of simultaneous equations. *International Economic Review*, 199–215.

Lee, Y. and R. Okui (2012). Hahn–hausman test as a specification test. *Journal of Econometrics 167*(1), 133–139.

Mikusheva, A. and L. Sun (2020). Inference with many weak instruments. *Working paper*.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, 393–415.

Stock, J. H. and M. W. Watson (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics 30*(4), 481–493.

Yogo, M. (2004). Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics 86*(3), 797–810.

APPENDIX

A. Presentation of the Tikhonov Regularization.

Here we consider the general case where the estimation is based on a sequence of instruments $Z_i = Z(\tau; \nu_i)$ with $\tau \in N$. Assume τ lies in a space Ξ ($\Xi = \{1, .., L\}$ or $\Xi = \mathbb{N}$) and let ω be a positive measure on Ξ. Let K be the covariance operator of the instruments from $L^2(\omega)$ to $L^2(\omega)$ such that:

$$(Kg)(\tau) = \sum_{l=1}^{L} E(Z(\tau, \nu_i)Z(\tau_l, \nu_i))g(\tau_l)\omega(\tau_l).$$

where $L^2(\omega)$ denotes the Hilbert space of square integrable functions with respect to ω. The inner product in $L^2(\omega)$ denoted $\langle v, w\rangle$ is $\sum_l v_l w_l \omega(l)$. K is supposed to be a nuclear operator which means that its trace is finite. The operator can be estimated by K_n defined as:

$$K_n : L^2(\omega) \to L^2(\omega)$$
$$(K_n g)(\tau) = \sum_{l=1}^{L} \frac{1}{n} \sum_{i=1}^{n} (Z(\tau, \nu_i) Z(\tau_l, \nu_i)) g(\tau_l) \omega(\tau_l).$$

If the number of instruments L is large relatively to n, inverting the operator K is considered as an ill-posed problem, which means that the inverse is not continuous. To solve this problem, we need to stabilize the inverse of K_n using regularization. A regularized inverse of an operator K is defined as: $R_\alpha : L^2(\omega) \to L^2(\omega)$ such that $lim_{\alpha \to 0} R_\alpha K \rho = \rho, \forall \rho \in L^2(\omega)$, where α is the regularization parameter (see Kress (1999) and Carrasco et al. (2007)). Let λ_j and ϕ_j, $j = 1...$ be respectively the eigenvalues (ordered in decreasing order) and the orthogonal eigenfunctions of K_n.

Tikhonov regularization

We consider the Tikhonov regularization scheme.

$$(K_n^\alpha)^{-1} = (K_n^2 + \alpha I)^{-1} K_n.$$
$$(K_n^\alpha)^{-1} r = \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \alpha} \langle r, \phi_j \rangle \phi_j.$$

where $\alpha > 0$ and I is the identity operator. The Tikhonov regularization is related to ridge regularization. Ridge method was first proposed to improve the properties of the OLS estimator in regressions with many regressors. The aim was to stabilize the inverse of XX' by replacing XX' by $XX' + \alpha I$. However, the reduction of variance was obtained at the expense of a bias relative to OLS estimator. In the IV regression, the 2SLS estimator has already a bias and the use of many instruments usually increases its bias. So, the Tikhonov regularization tends to reduce the bias of the IV estimator (at the expense of a larger variance).

Let $(K_n^\alpha)^{-1}$ be the regularized inverse of K_n and P^α a $n \times n$ matrix as defined in Carrasco (2012) by

$$P^\alpha = T(K_n^\alpha)^{-1} T^* \tag{7.16}$$

where $T : L^2(\omega) \to R^n$ with
$Tg = (< Z_1, g >, < Z_2, g >', ..., < Z_n, g >')'$
and $T^* : R^n \to L^2(\omega)$ with
$T^* v = \frac{1}{n} \sum_j^n Z_j v_j$
such that $K_n = T^* T$ and TT^* is a $n \times n$ matrix with typical element $\frac{<Z_i, Z_j>}{n}$. Let ϕ_j, $\lambda_1 \geq \lambda_2 \geq ..... \geq 0$, $j = 1, 2, ...$ be the orthonormalized eigenfunctions and eigenvalues of K_n and ψ_j the eigenfunctions of TT^*. We then have $T\phi_j = \sqrt{\lambda}_j \psi_j$ and $T^* \psi_j = \sqrt{\lambda}_j \phi_j$.
For $v \in R^n$, $P^\alpha v = \sum_j^\infty q(\alpha, \lambda_j^2) < v, \psi_j > \psi_j$ where $q(\alpha, \lambda_j^2) = \frac{\lambda_j^2}{\lambda_j^2 + \alpha}$.
Remark that the case when $\alpha = 0$ corresponds to no regularization Thus we have $q(0, \lambda_j^2) = 1$ and $P^0 = Z(Z'Z)^+ Z'$, where $(.)^+$ represents the Moore-Penrose generalized inverse.

B. Proofs

Our proof of Theorem 3.2 follows the same steps as the proofs of Theorem 1 in Chao et al. (2014). However, our results are not a straightforward application of Chao et al. (2014). In their paper, there is no regularization. Instead, the number of instruments plays the role of the regularization parameter and the matrix $P = Z\left(Z'Z\right)^{-1} Z'$ is a projection matrix. Their results rely often on the properties of projection matrices. In our paper, the regularization parameter is α and the regularized matrix $P^{\alpha} = \sum_j q(\alpha, \lambda_j^2) < v, \psi_j > \psi_j$ is not a projection matrix any longer. So we need to derive some properties on the elements of P^{α} in Lemma 7.1 below. This lemma corresponds to Lemma A0 of Carrasco and Doukali (2017).

Lemma 7.1. *If Assumptions 3.1-3.3 are satisfied, then :*
i) $P_{ii}^{\alpha} < 1$ *for* $\alpha > 0$,
ii) $\sum_{i\neq j} \left(P_{ij}^{\alpha}\right)^2 = O(1/\alpha)$,
iii) $\sum_{i\neq j} P_{ij}^{\alpha} = O(1/\alpha)$.
iv) $\sum_{i,l,k,r} P_{ik}^{\alpha} P_{kl}^{\alpha} P_{lr}^{\alpha} P_{ri}^{\alpha} = O(1/\alpha)$.
v) $\sum_{i,j} \left(P_{ij}^{\alpha}\right)^4 = O(1/\alpha)$.

Proof of Lemma 7.1. The proof can be found in Carrasco and Doukali (2017).
Let us define some notations that will be used in the following Lemmas. For random variables[8] W_i, Y_i, η_i, let $\bar{w}_i = E[W_i]$, $\bar{y}_i = E[Y_i]$, $\bar{\eta}_i = E[\eta_i]$, $\tilde{W}_i = W_i - \bar{w}_i$ and $\tilde{Y}_i = Y_i - \bar{y}_i$, $\tilde{\eta}_i = \eta_i - \bar{\eta}_i$, $\bar{w}_n = E[(W_1, ..., W_n)']$, $\bar{y}_n = E[(Y_1, ..., Y_n)']$, $\bar{\mu}_W = max_{i\leq n}|\bar{w}_i|$, $\bar{\mu}_Y = max_{i\leq n}|\bar{y}_i|$, $\bar{\mu}_\eta = max_{i\leq n}|\bar{\eta}_i|$, $\bar{\sigma}_{W_n}^2 = max_{i\leq n} var(W_i)$, $\bar{\sigma}_{Y_n}^2 = max_{i\leq n} var(Y_i)$, $\bar{\sigma}_{\eta}^2 = max_{i\leq n} var(\eta_i)$.
Define the norm: $||W||_{L_2}^2 = \sqrt{E[W^2]}$, and let M, CS, T denote the Markov inequality, the Cauchy-Schwarz inequality, and the triangle inequality, respectively. In the sequel, C denotes a constant, which may be different from place to place, $\hat{\delta}$ denotes the regularized Jackknife IV estimator previously denoted $\hat{\delta}^{\alpha}$ (the dependence in α is hidden for simplicity).

Lemma 7.2. *Suppose the following conditions hold:*
(i) $P^{\alpha}v = Z(Z'Z + \alpha I)^{-1}Z'v$ *or* $\sum_j^{\infty} q(\alpha, \lambda_j^2) < v, \psi_j > \psi_j$ *as defined in Appendix A.*
(ii)$(W_{1n}, U_1, \epsilon_1), ..., (W_{nn}, U_n, \epsilon_n)$ *are independent, and* $D_{1,n} := \sum_{i=1}^n E[W_{in}W_{in}']$ *satisfies* $||D_{1,n}|| < C$,
(iii) $E[W_{in}'] = 0$, $E[U_i] = 0$, $E[\epsilon_i] = 0$, *and there is a constant C such that* $E[||U_i||^4] \leq C$ *and* $E[\epsilon_i^4] \leq C$,
(iv) $\sum_{i=1}^n E[||W_{in}||^4] \to 0$ *a.s.*
(v) $\alpha \to 0$ *as* $n \to \propto$.
Then for:
$D_{2,n} := \alpha \sum_{i\neq j}^n (P_{ij}^{\alpha})^2 (E[U_iU_i']E[\epsilon_j^2] + E[U_i\epsilon_i]E[U_j'\epsilon_j])$
and any sequences c_{1n} *and* c_{2n} *with* $||c_{1n}|| \leq C$, $||c_{2n}|| \leq C$, *and* $\sum_n = c_{1n}' D_{1,n} c_{1n} + c_{2n}' D_{2,n} c_{2n} > 1/C$, *it follows that:*
$\bar{Y}_n = \sum_n{}^{-1/2} (c_{1n}' \sum_{i=1}^n W_{i,n} + \sqrt{\alpha} c_{2n}' \sum_{i\neq j}^n U_i (P_{ij}^{\alpha})^2 \epsilon_j) \xrightarrow{d} N(0,1)$

[8] Note that here W_i and η_i are arbitrary scalar variables that will take various forms in the sequel.

Proof of Lemma 7.2. This is Lemma A2 of Carrasco and Doukali (2017) when Z and Υ are not random.

LEMMA 7.3. *If assumptions 3.1-3.3 are satisfied then:*
(i) $S_n^{-1}\sum_{i\neq j}^n X_i P_{ij}^\alpha X_j' S_n^{-1'} = O_p(1)$.
(ii) $S_n^{-1}\sum_{i\neq j}^n X_i P_{ij}^\alpha \epsilon_j = O_p(1+\frac{1}{\sqrt{\alpha}\mu_n})$.

Proof of Lemma 7.3. Consider first (i): We have $S_n^{-1}\sum_{i\neq j}^n X_i P_{ij}^\alpha X_j' S_n^{-1'} = \sum_{i\neq j}^n f_i P_{ij}^\alpha f_j'/n + o_p(1)$.
We also have $\sum_{i\neq j}^n f_i P_{ij}^\alpha f_j'/n = f'P^\alpha f/n - \sum_i^n f_i f_i' P_{ii}^\alpha/n$, and both $f'P^\alpha f/n \leq f'f/n$ and
$\sum_i^n f_i f_i' P_{ii}^\alpha/n \leq f'f/n$ are bounded, giving the first conclusion.
(ii) holds by Lemma A5 of Carrasco and Doukali (2017) and (i) of Lemma 7.1.

LEMMA 7.4. *If* $\hat{\delta} \xrightarrow{p} \delta$, $E[||X_i||^2] \leq C$, $E[\epsilon_i^4] \leq C$, $\epsilon_1, ..., \epsilon_n$ *are mutually independent, and either* $\alpha \to 0$ *or* $max_{i\leq n}\ P_{ii}^\alpha \to 0$ *then:*
$\alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2 \hat{\epsilon}_i^2\hat{\epsilon}_j^2 - \alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2 \sigma_i^2\sigma_j^2 \xrightarrow{P} 0$.

Proof of Lemma 7.4. By $\hat{\delta} \xrightarrow{p} \delta$ we have $||\hat{\delta}-\delta||^2 \leq ||\hat{\delta}-\delta||$ with probability one. Denote $d_i = 2|\epsilon_i|||X_i|| + |||X_i||^2$, we have:

$$\begin{aligned}\hat{\epsilon}_i &= y_i - X_i'\hat{\delta}\\ &= X_i'\delta + \epsilon_i - X_i'\hat{\delta}\\ &= \epsilon_i - X_i'(\hat{\delta}-\delta).\end{aligned}$$

It follows that:
$\hat{\epsilon}_i^2 = \epsilon_i^2 - 2\epsilon_i X_i'(\hat{\delta}-\delta) + (\hat{\delta}-\delta)'X_iX_i'(\hat{\delta}-\delta)$.
Then:
$\hat{\epsilon}_i^2 - \epsilon_i^2 = -2\epsilon_i X_i'(\hat{\delta}-\delta) + (\hat{\delta}-\delta)'X_iX_i'(\hat{\delta}-\delta)$.
$|\hat{\epsilon}_i^2 - \epsilon_i^2| \leq 2|\epsilon X_i'(\hat{\delta}-\delta)| + |(\hat{\delta}-\delta)'X_iX_i'(\hat{\delta}-\delta)|$.
$|\hat{\epsilon}_i^2 - \epsilon_i^2| \leq 2|\epsilon_i|\ ||X_i||\ ||\hat{\delta}-\delta|| + |||X_i||^2||\hat{\delta}-\delta||^2 \leq d_i||\hat{\delta}-\delta||$.
Also by (ii) of Lemma 7.1, $\sum_{i\neq j}^n (P_{ij}^\alpha)^2 = O(1/\alpha)$,
$\alpha E[\sum_{i\neq j}^n (P_{ij}^\alpha)^2 d_i d_j] \leq \alpha C\sum_{i\neq j}^n P_{ij}^2 \leq C$,
$\alpha E[\sum_{i\neq j}^n (P_{ij}^\alpha)^2 \epsilon_i^2 d_j] \leq C$.
Then by M,
$\alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2 d_i d_j = O_p(1)$, $\alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2 \epsilon_i^2 d_j = O_p(1)$,
Therefore, for $\hat{V}_n = \alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2\hat{\epsilon}_i^2\hat{\epsilon}_j^2$, $\tilde{V}_n = \alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2\epsilon_i^2\epsilon_j^2$, we have
$|\hat{V}_n - \tilde{V}_n| \leq \alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2|\hat{\epsilon}_i^2\hat{\epsilon}_j^2 - \epsilon_i^2\epsilon_j^2|$
$|\hat{V}_n - \tilde{V}_n| \leq \alpha||\hat{\delta}-\delta||^2\sum_{i\neq j}^n (P_{ij}^\alpha)^2 d_i d_j + 2\alpha||\hat{\delta}-\delta||\sum_{i\neq j}^n (P_{ij}^\alpha)^2\epsilon_i^2 d_j \to 0$.
Let $V_n = \alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2\sigma_i^2\sigma_j^2$ and $v_i = \epsilon_i^2 - \sigma_i^2$. We have:
$\sum_{i\neq j}^n (P_{ij}^\alpha)^2\epsilon_i^2\epsilon_j^2 - \sum_{i\neq j}^n (P_{ij}^\alpha)^2\sigma_i^2\sigma_j^2 = 2\sum_{i\neq j}^n (P_{ij}^\alpha)^2 v_i\sigma_j^2 + \sum_{i\neq j}^n (P_{ij}^\alpha)^2 v_i v_j$.
We note that $E[v_i^2] \leq E[\epsilon_i^4] \leq C$, so we have:
$E[(\alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2 v_i\sigma_j^2)^2] = \alpha^2\sum_i\sum_{i\neq j}\sum_{k\neq i}(P_{ij}^\alpha)^2(P_{ik}^\alpha)^2 E[v_i^2]\sigma_i^2\sigma_k^2$
$E[(\alpha\sum_{i\neq j}^n (P_{ij}^\alpha)^2 v_i\sigma_j^2)^2] \leq C\alpha^2\sum_i\sum_j (P_{ij}^\alpha)^2\sum_k (P_{ik}^\alpha)^2$

We note that $P^{\alpha}_{ij} = P^{\alpha}_{ji}$, and $\sum_i \sum_j (P^{\alpha}_{ij})^2 \sum_k (P^{\alpha}_{ik})^2 = O(1/\alpha)$ by Lemma 7.1 (vi). So:
$E[(\alpha \sum_{i\neq j}^n (P^{\alpha}_{ij})^2 v_i \sigma_j^2)^2] = C\alpha \to 0$.
Also by CS, $max_{ij}(P^{\alpha}_{ij})^2 \leq max_{ii}(P^{\alpha}_{ii})^2$, so that:
$E[(\alpha \sum_{i\neq j}^n (P^{\alpha}_{ij})^2 v_i v_j)^2] = 2\alpha^2 \sum_{i\neq j}^n (P^{\alpha}_{ij})^4 E[v_i^2]E[v_j^2] \leq C\alpha^2 \sum_{i\neq j}^n (P^{\alpha}_{ij})^4 \leq C\alpha^2 O(1/\alpha) \to 0$. Because of (v) of Lemma 7.1.
Then by T and M we have $\tilde{V}_n - V_n \xrightarrow{\text{p}} 0$. The conclusion then follows by T.

Proof of Theorem 3.2.

$$\begin{aligned}\sqrt{\alpha}\sum_{i\neq j}^n \hat{\epsilon}_i P^{\alpha}_{ij}\hat{\epsilon}_j &= \sqrt{\alpha}\sum_{i\neq j}^n [\epsilon_i - X_i'(\hat{\delta}-\delta)]P^{\alpha}_{ij}[\epsilon_j - X_j'(\hat{\delta}-\delta)] \\ &= \sqrt{\alpha}\sum_{i\neq j}^n \epsilon_i P^{\alpha}_{ij}\epsilon_j + (\hat{\delta}-\delta)' S_n \times \sqrt{\alpha}[S_n^{-1}\sum_{i\neq j}^n X_i P^{\alpha}_{ij} X_j' S_n'^{-1}] S_n'(\hat{\delta}-\delta) \\ &\quad + 2\sqrt{\alpha}(\hat{\delta}-\delta)' S_n [S_n^{-1}\sum_{i\neq j}^n X_i P^{\alpha}_{ij}\epsilon_j].\end{aligned}$$

If $1/(\alpha\mu_n^2) \to C < \infty$, then by Theorem 2 of Carrasco and Doukali (2017) we have $S_n'(\hat{\delta}-\delta) = O_p(1)$. Then by Lemma 7.3 we have:

$$\sqrt{\alpha}\sum_{i\neq j}^n \hat{\epsilon}_i P^{\alpha}_{ij}\hat{\epsilon}_j = \sqrt{\alpha}\sum_{i\neq j}^n \epsilon_i P^{\alpha}_{ij}\epsilon_j + o_p(1)$$

Next, note that $\sigma_i^2 \geq C$ by Assumption 3.3 and $P^{\alpha}_{ii} \leq C < 1$ by Assumption 3.1 so that:

$$\begin{aligned}V_n = \alpha\sum_{i\neq j}^n \sigma_i^2 (P^{\alpha}_{ij})^2 \sigma_j^2 &> C(\alpha\sum_{i,j}^n (P^{\alpha}_{ij})^2 - \sum_i^n (P^{\alpha}_{ii})^2) \\ &= C\alpha\sum_i^n P^{\alpha}_{ii}(1-P^{\alpha}_{ii}) > C(1-C) > 0.\end{aligned}$$

Moreover, $E[\epsilon_i^4] \leq C$ and,

$$\begin{aligned}E[\sum_{i\neq j}^n (\epsilon_i P^{\alpha}_{ij}\epsilon_j)^2] &= E[\sum_{i\neq j}\sum_{k\in\{i,j\}} P^{\alpha}_{ik}P^{\alpha}_{jk}\epsilon_i\epsilon_j'\epsilon_k^2 + \sum_{i\neq j}^n P^{\alpha}_{ij}\epsilon_i^2\epsilon_j^2 \\ &= E[2\sum_{i\neq j}^n (P^{\alpha}_{ij}\epsilon_i^2\epsilon_j^2)] = 2\sum_{i\neq j}^n P^{\alpha}_{ij}\sigma_i^2\sigma_j^2 = 2tr(P^{\alpha})V_n\end{aligned}$$

It follows from Lemma 7.2 with $W_{in} = 0$, $c_{1n} = 0$, $c_{2n} = 1$, $U_i = \epsilon_i$ that :

$$\frac{\sum_{i\neq j}^n \epsilon_i P^{\alpha}_{ij}\epsilon_j}{\sqrt{2tr(P^{\alpha})V_n}} \xrightarrow{\text{d}} N(0,1).$$

Next by Theorem 1 of Carrasco and Doukali (2017), we have $\hat{\delta} \xrightarrow{\text{p}} \delta$. Moreover by Lemma 7.1 (iii), $tr(P^{\alpha}) = O(\frac{1}{\alpha})$. Hence, by Lemma 7.4, $\hat{V}_n - V_n \xrightarrow{\text{p}} 0$. Then by V_n bounded and bounded away from zero, $\sqrt{\frac{V_n}{\hat{V}_n}} \to 1$. Therefore by Slutsky theorem,

$$\frac{\sum_{i\neq j}^{n}\hat{\epsilon}_i P_{ij}^{\alpha}\hat{\epsilon}_j}{\sqrt{2tr(P^{\alpha})\hat{V}_n}} = \frac{\sum_{i\neq j}^{n}\epsilon_i P_{ij}^{\alpha}\epsilon_j}{\sqrt{2tr(P^{\alpha})\hat{V}_n}} + \frac{o_p(1)}{2\hat{V}_n} = \sqrt{\frac{V_n}{\hat{V}_n}}\frac{\sum_{i\neq j}^{n}\epsilon_i P_{ij}^{\alpha}\epsilon_j}{\sqrt{2tr(P^{\alpha})V_n}} + o_p(1) \xrightarrow{\mathrm{d}} N(0,1)$$

Next note that $\hat{T} \geq q_{(tr(P^{\alpha})-p)}(1-\beta)$ if and only if

$$\frac{\sum_{i\neq j}^{n}\hat{\epsilon}_i P_{ij}^{\alpha}\hat{\epsilon}_j}{\sqrt{2tr(P^{\alpha})\hat{V}_n}} \geq \frac{q_{(tr(P^{\alpha})-p}(1-\beta) - tr(P^{\alpha})}{\sqrt{2tr(P^{\alpha})}}$$

Using the fact that $tr(P^{\alpha}) = O(\frac{1}{\alpha})$, we have, as $\alpha \to 0$, $q_{(tr(P^{\alpha})-p)}(1-\beta) - (tr(P^{\alpha}) - p)/\sqrt{2(tr(P^{\alpha})-p)} \to q(1-\beta)$ where $q(1-\beta)$ is the $1-\beta$ quantile of the standard normal distribution, also, we have:

$$\sqrt{\frac{(tr(P^{\alpha}))-p}{tr(P^{\alpha})}}(\frac{q_{(tr(P^{\alpha}))-p}(1-\beta)-(tr(P^{\alpha})-p)}{\sqrt{2tr(P^{\alpha})-p}}) - \frac{p}{\sqrt{2tr(P^{\alpha})}} \to q(1-\beta).$$

The conclusion now follows.

Proof of Theorem 5.1.
We will use the following two limiting distributions.
First, using Lemma A2 of Chao et al (2012) and results on P^{α} from Carrasco and Doukali (2017), we have

$$\frac{\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^{\alpha}u_i u_j}{\sqrt{2\sum_{i=1}^{n}\sum_{j\neq i}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}} \xrightarrow{d} \mathcal{N}\left(0,1\right).$$

Next, using Lindeberg theorem and Lyapunov's condition which is satisfied by Assumption 5.1(a), we have

$$\frac{\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^{\alpha}u_i\Upsilon_j}{\sqrt{\sum_{i=1}^{n}\left(\sum_{j\neq i}P_{ij}^{\alpha}\Upsilon_j\right)^2 E\left(u_i^2\right)}} \xrightarrow{d} \mathcal{N}\left(0,1\right).$$

Moreover,

$$\sum_{i\neq j}P_{ij}^{\alpha}x_i x_j = \sum_{i\neq j}P_{ij}^{\alpha}\Upsilon_i\Upsilon_j + \sum_{i\neq j}P_{ij}^{\alpha}u_i u_j + 2\sum_{i\neq j}P_{ij}^{\alpha}u_i\Upsilon_j.$$

We have

$$\frac{\sum_{i\neq j}P_{ij}^{\alpha}u_i\Upsilon_j}{\sqrt{2\sum_{i=1}^{n}\sum_{j\neq i}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}}$$
$$= \frac{\sum_{i\neq j}P_{ij}^{\alpha}u_i\Upsilon_j}{\sqrt{\sum_{i=1}^{n}\left(\sum_{j\neq i}P_{ij}^{\alpha}\Upsilon_j\right)^2 E\left(u_i^2\right)}}\frac{\sqrt{\sum_{i=1}^{n}\left(\sum_{j\neq i}P_{ij}^{\alpha}\Upsilon_j\right)^2 E\left(u_i^2\right)}}{\sqrt{2\sum_{i=1}^{n}\sum_{j\neq i}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}}.$$

Given $C > E\left(u_i^2\right) > 0$, it suffices to study

$$\frac{\sum_{i=1}^{n}\left(\sum_{j\neq i} P_{ij}^{\alpha}\Upsilon_j\right)^2}{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2} = \frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2\Upsilon_j^2}{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2} + \frac{\sum_{i=1}^{n}\left(\sum_{j\neq i} P_{ij}^{\alpha}\Upsilon_j\right)\left(\sum_{l\neq i} P_{il}^{\alpha}\Upsilon_l\right)}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2} \tag{7.17}$$

$$= O\left(\frac{1}{n}\right) + O\left(\frac{\gamma^2}{\sqrt{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}}\right) \tag{7.18}$$

$$= o\left(1\right) \tag{7.19}$$

because $\Upsilon_j = z_j'\widetilde{\pi}/\sqrt{n}$ and the fact that $\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 = O\left(1/\alpha\right)$ by Lemma 7.1.

So we get

$$F_{Tikh} = \frac{\sum_{i\neq j} P_{ij}^{\alpha}x_i x_j}{\sqrt{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}} \frac{\sqrt{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}}{\sqrt{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 \widehat{u}_i^2\widehat{u}_j^2}}$$

where

$$\frac{\sum_{i\neq j} P_{ij}^{\alpha}x_i x_j}{\sqrt{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}} = \gamma^2 + \frac{\sum_{i\neq j} P_{ij}^{\alpha}u_i u_j}{\sqrt{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}} + o\left(1\right). \tag{7.20}$$

Hence, the term on the l.h.s. of (7.20) minus γ^2 converges to a normal with mean 0 and variance 1.

Finally, we need to prove that

$$\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 \widehat{u}_i^2\widehat{u}_j^2}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)} \xrightarrow{P} 1. \tag{7.21}$$

The proof of (7.21) is done in two steps. First, we establish

$$\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 \widehat{u}_i^2\widehat{u}_j^2}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2} - \frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 u_i^2 u_j^2}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2} \xrightarrow{P} 0. \tag{7.22}$$

Second, we show that

$$\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 u_i^2 u_j^2}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2} - \frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2 E\left(u_i^2\right)E\left(u_j^2\right)}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2} \xrightarrow{P} 0. \tag{7.23}$$

Using $P_{ij}^{\alpha}=P_{ji}^{\alpha}$, we have

$$\begin{aligned}\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2\widehat{u}_i^2\widehat{u}_j^2-\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2u_i^2u_j^2 &= 4\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2u_i\left(\Upsilon_i-\hat{\Upsilon}_i^{\alpha}\right)u_j\left(\Upsilon_j-\hat{\Upsilon}_j^{\alpha}\right)\\ &+4\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2u_i\left(\Upsilon_i-\hat{\Upsilon}_i^{\alpha}\right)\left(\Upsilon_j-\hat{\Upsilon}_j^{\alpha}\right)^2\\ &+\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2\left(\Upsilon_i-\hat{\Upsilon}_i^{\alpha}\right)^2\left(\Upsilon_j-\hat{\Upsilon}_j^{\alpha}\right)^2\\ &+4\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2u_i\left(\Upsilon_i-\hat{\Upsilon}_i^{\alpha}\right)u_j^2\\ &+2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2\left(\Upsilon_i-\hat{\Upsilon}_i^{\alpha}\right)^2u_j^2.\end{aligned}$$

Consider the first term on the r.h.s.:

$$\begin{aligned}\frac{\left|\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2u_i\left(\Upsilon_i-\hat{\Upsilon}_i^{\alpha}\right)u_j\left(\Upsilon_j-\hat{\Upsilon}_j^{\alpha}\right)\right|}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2} &\leq \frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2|u_i|\,|u_j|\left|\Upsilon_i-\hat{\Upsilon}_i^{\alpha}\right|\left|\Upsilon_j-\hat{\Upsilon}_j^{\alpha}\right|}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2}\\ &\leq\left(\sup_{z\in D}\left|\Upsilon(z)-\hat{\Upsilon}^{\alpha}(z)\right|\right)^2\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2|u_i|\,|u_j|}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2}\\ &=o_p(1)\,O_p(1)\end{aligned}$$

because of Assumption 5.1(b) and the fact that $E|u_i|<C$, so that

$$E\left[\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2|u_i|\,|u_j|}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2}\right]<C$$

and hence $\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2|u_i||u_j|}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2}=O_p(1)$ by Markov inequality. Handling the other terms in the same fashion yields the result (7.22).

Now, we turn our attention towards (7.23). Let $v_i=u_i^2-E\left(u_i^2\right)$. We have

$$\begin{aligned}&\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2u_i^2u_j^2-\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2E\left(u_i^2\right)E\left(u_j^2\right)\\ &=\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2v_iv_j+2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2v_iE\left(u_j^2\right).\end{aligned}$$

Then, using $E\left(v_i^2\right)\leq E\left(u_i^4\right)<C$,

$$\begin{aligned}E\left[\left[\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2v_iv_j}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2}\right]^2\right] &= \frac{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^4E\left(v_i^2\right)E\left(v_j^2\right)}{\left(\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2\right)^2}\\ &<2C^2\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^4}{\left(\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2\right)^2}\to 0\end{aligned}$$

as $\alpha\to 0$ because $\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^4=O(1/\alpha)$ and $\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^2=O(1/\alpha)$ from Lemma 7.1.

Moreover,

$$E\left[\left[\frac{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^{2}v_{i}E\left(u_{j}^{2}\right)}{\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^{2}}\right]^{2}\right]=\frac{\sum_{i}\sum_{j\neq i}\sum_{k\neq i}\left(P_{ij}^{\alpha}\right)^{2}\left(P_{ik}^{\alpha}\right)^{2}E\left(v_{i}^{2}\right)E\left(u_{j}^{2}\right)E\left(u_{k}^{2}\right)}{\left(\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^{2}\right)^{2}}$$
$$\leq C^{3}\frac{\sum_{i}\sum_{j\neq i}\sum_{k\neq i}\left(P_{ij}^{\alpha}\right)^{2}\left(P_{ik}^{\alpha}\right)^{2}}{\left(\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^{2}\right)^{2}}$$

which goes to 0 as α goes to zero. Then, by the Triangle inequality and Markov inequality, the result (7.23) follows. This completes the proof of Theorem 5.1.

Proof of Theorem 5.2.
Let $\Xi_n=\sqrt{2\sum_{i\neq j}\left(P_{ij}^{\alpha}\right)^{2}E\left(u_{i}^{2}\right)E\left(u_{j}^{2}\right)}$.

$$\hat{\delta}_{JIV2}-\delta=\frac{\sum_{i\neq j}P_{ij}^{\alpha}\left(\Upsilon_{i}+u_{i}\right)\epsilon_{j}}{\sum_{i\neq j}P_{ij}^{\alpha}\left(\Upsilon_{i}+u_{i}\right)\left(\Upsilon_{j}+u_{j}\right)}$$
$$=\frac{\Xi_{n}^{-1}\sum_{i\neq j}P_{ij}^{\alpha}\Upsilon_{i}\epsilon_{j}+\Xi_{n}^{-1}\sum_{i\neq j}P_{ij}^{\alpha}u_{i}\epsilon_{j}}{\gamma^{2}+2\Xi_{n}^{-1}\sum_{i\neq j}P_{ij}^{\alpha}u_{i}\Upsilon_{j}+\Xi_{n}^{-1}\sum_{i\neq j}P_{ij}^{\alpha}u_{i}u_{j}}.$$

It follows that

$$\gamma^{2}\left(\hat{\delta}_{JIV2}-\delta\right)=\frac{A+B}{1+\frac{D}{\gamma^{2}}+\frac{E}{\gamma^{2}}}$$

where $A=\Xi_{n}^{-1}\sum_{i\neq j}P_{ij}^{\alpha}\Upsilon_{i}\epsilon_{j}$, $B=\Xi_{n}^{-1}\sum_{i\neq j}P_{ij}^{\alpha}u_{i}\epsilon_{j}$, $D=2\Xi_{n}^{-1}\sum_{i\neq j}P_{ij}^{\alpha}u_{i}\Upsilon_{j}$, and $E=\Xi_{n}^{-1}\sum_{i\neq j}P_{ij}^{\alpha}u_{i}u_{j}$. Instead of doing an expansion for n large, we do the expansion for γ^2 large. When γ^2 is large enough, we can use the following expansion:

$$\gamma^{2}\left(\hat{\delta}_{JIV2}-\delta\right)=(A+B)\left(1-\frac{D}{\gamma^{2}}-\frac{E}{\gamma^{2}}+\frac{1}{\gamma^{4}}(D+E)^{2}\right)+\frac{R}{\gamma^{6}}$$

where R is a polynomial of normal distributions and hence satisfies condition (3.8) of Rothenberg (1984) with γ^2 replacing $1/n$ and can be neglected.

Moreover, we observe that, because of the independence assumption, $E(A)=E(B)=E(BD)=E(AE)=E((A+B)D^{2})=E((A+B)E^{2})=E(BDE)=0$. Therefore, $\gamma^{2}E\left(\hat{\delta}_{JIV2}-\delta\right)$ can be approximated by

$$-\frac{E(AD)}{\gamma^{2}}-\frac{E(BE)}{\gamma^{2}}+\frac{2E(ADE)}{\gamma^{4}}.$$

$$
\begin{aligned}
\frac{E\left(BE\right)}{\gamma^2} &= \frac{1}{\gamma^2\Xi_n^2}E\left[\left(\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^{\alpha}u_i\epsilon_j\right)\left(\sum_{l=1}^{n}\sum_{k\neq i}P_{lk}^{\alpha}u_l u_k\right)\right] \\
&= \frac{2}{\gamma^2\Xi_n^2}\sum_{i=1}^{n}E\left(u_i^2\right)\sum_{j\neq i}P_{ij}^{\alpha 2}E\left(\epsilon_j u_j\right) \\
&= \frac{2\rho}{\gamma^2\Xi_n^2}\sum_{i=1}^{n}E\left(u_i^2\right)\sum_{j\neq i}P_{ij}^{\alpha 2}E\left(u_j^2\right) \\
&= \frac{\rho}{\gamma^2}
\end{aligned}
$$

using $E\left(\epsilon_j u_j\right) = \rho E\left(u_j^2\right)$ which follows from the joint normality assumption. This term will be the dominant term as we will show below.

We have

$$
\begin{aligned}
\frac{E\left(AD\right)}{\gamma^2} &= \frac{2}{\gamma^2\Xi_n^2}E\left[\left(\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^{\alpha}\epsilon_i\Upsilon_j\right)\left(\sum_{l=1}^{n}\sum_{k\neq i}P_{lk}^{\alpha}u_l\Upsilon_k\right)\right] \\
&= \frac{2}{\gamma^2\Xi_n^2}\sum_{i=1}^{n}E\left(\epsilon_i u_i\right)\left(\sum_{j\neq i}P_{ij}^{\alpha}\Upsilon_j\right)^2 \\
&= \frac{2\rho}{\gamma^2\Xi_n^2}\sum_{i=1}^{n}E\left(u_i^2\right)\left(\sum_{j\neq i}P_{ij}^{\alpha}\Upsilon_j\right)^2 \\
&= o\left(\frac{\rho}{\gamma^2}\right)
\end{aligned}
$$

by Equations (7.17), (7.18), and (7.19).

We have

$$
\begin{aligned}
\frac{E\left(ADE\right)}{\gamma^4} &= \frac{1}{\gamma^4\Xi_n^3}E\left[\left(\sum_{i=1}^{n}\sum_{j\neq i}P_{ij}^{\alpha}\epsilon_i\Upsilon_j\right)\left(\sum_{l=1}^{n}\sum_{k\neq l}P_{lk}^{\alpha}u_l\Upsilon_k\right)\left(\sum_{i'=1}^{n}\sum_{j'\neq i'}P_{i'j'}^{\alpha}u_{i'}u_{j'}\right)\right] \\
&= \frac{2\rho}{\gamma^4\Xi_n^3}\sum_{i=1}^{n}E\left(u_i^2\right)\left(\sum_{j\neq i}P_{ij}^{\alpha}\Upsilon_j\right)\left(\sum_{k\neq i}P_{ik}^{\alpha}E\left(u_k^2\right)\right)\left(\sum_{j'\neq k}P_{kj'}^{\alpha}\Upsilon_{j'}\right) \\
&= \frac{2\rho C^2}{\gamma^4\Xi_n^3}\sum_{i=1}^{n}\left(\sum_{j\neq i}P_{ij}^{\alpha}\Upsilon_j\right)\left(\sum_{k\neq i}P_{ik}^{\alpha}\right)\left(\sum_{j'\neq k}P_{kj'}^{\alpha}\Upsilon_{j'}\right)
\end{aligned}
$$

using the fact that $E\left(u_i^2\right) < C$. For α small, the matrix P^{α} is almost idempotent and the term $\frac{E(ADE)}{\gamma^4}$ can be approximated by $\frac{2\rho C^2}{\gamma^2\Xi_n^2}$ which is negligeable compared to $\frac{\rho}{\gamma^2}$. So the bias of the dominant term is simply $-\frac{\rho}{\gamma^4}$. This completes the proof of Theorem 5.2.