A New Paradigm: A Joint Test of Structural and Correlation Parameters in Instrumental Variables Regression When Perfect Exogeneity is Violated

Mehmet Caner and Melinda Morrill

North Carolina State University

14. August 2009

A Modified T-Ratio for Inference in Instrumental Variables Regression when Perfect Exogeneity is Violated

Mehmet Caner
North Carolina State University *

Melinda Morrill
North Carolina State University †

July 7, 2009

Abstract

In empirical research, applied economists benefit from instrumental variable methods and use a simple t-ratio test statistic to infer whether there is a causal relationship among the variables analyzed. However, the t-test gives unreliable results even when there is only a slight violation of exogeneity. This paper demonstrates that it is possible to modify the t-ratio in a simple way so that causal inference can still be drawn under a violation of perfect exogeneity, thus providing applied researchers with the necessary robustness property for inference.

*Mehmet Caner: Department of Economics, 4168 Nelson Hall, Raleigh, NC 27518. email: mcaner@ncsu.edu.

†Melinda Morrill: Department of Economics, 4144 Nelson Hall, Raleigh, NC 27695-8110. Email: melinda_morrill@ncsu.edu

1 Introduction

Economists frequently apply instrumental variable methods to draw inferences about whether or not some variable influences an economic outcome. For example, labor economists employ varied instruments, including quarter and year of birth (Angrist and Krueger, 1991), tuition and distance to nearest college (Kane and Rouse, 1995, Card, 1993), attending reform school (Meghir and Palene, 2005) and birth year interacted with school buildings in region of birth (Dufflo, 2001) to measure the extent to which a person's education influences her salary and wages. In a distinct but related literature that combines macro-economics, political economy and comparative institutions, economists employ instruments including early settler mortality (Acemoglu, Johnson and Robinson, 2001), ethnic capital (Hall and Jones, 1999), ethno-linguistic fractionalization (Mauro, 1995) and legal families (Djankov et al., 2003, and Acemoglu and Johnson, 2006) to determine whether or not the quality of institutions influences long term growth and investment. Instrumental variable methods are used to identify causal relationships by isolating changes in an endogenous variable (or variables) that are unrelated to potential unobserved factors. To identify a causal relationship, instruments must be exogenous; that is, they are not related to the outcome variable after controlling for relevant explanatory variables. For example, early settler mortality is exogenous if it is only related to long term growth through its impact on institutions, after controlling for relevant variables such as latitude. This requirement is strong because it means that settler mortality can only influence long term growth indirectly through the quality of contemporary institutions. The exogeneity of early settler mortality, however, is controversial: for example, as noted by Glaeser et al. (2004), early settler mortality could also influence long term growth through its impact on the unobservable human capital of the early settlers. Whether or not the exclusion restriction is perfectly satisfied is debatable for many (and perhaps most) applications of instrumental variables.

In empirical research, applied economists benefit from instrumental variable methods and use a simple t-ratio test statistic to infer whether there is causality among the variables analyzed. However, the t-test gives unreliable results even there is slight violation of exogeneity, as established recently in a paper by Berkowitz, Caner and Fang (Economics Letters, 2008). That paper only shows there is a problem. Here we extend this research to show that it is possible to modify the t-ratio in a simple way so that causal inference can still be drawn under a violation of perfect exogeneity. In Table 1, we show that using standard t-ratio is not desirable when the instrument is endogenous. The table shows the actual size of the test when the single instrument is endogenous.

Table 1: Size (5% level), Standard t test

$corr_0 =$	-0.5	-0.3	-0.1	0	0.1	0.3	0.5
$n = 1000$	100.0	100.0	87.9	5.3	88.8	100.0	100.0
$n = 200$	100.0	99.0	26.9	5.3	32.9	99.2	100.0
$n = 100$	100.0	85.2	15.4	5.3	19.9	89.6	99.9

Note: $corr_0$ represents the true correlation between the single instrument and the error (second stage equation). The first column header is true correlation, all the other column headers are specific true correlation values.

When we increase the sample size the problem gets worse. For example with true correlation of 0.1, at $n = 100$ (sample size is n) the size of the test is 19.9%, then if the sample size increases to $n = 1000$, the size is 88.8%. This is a massive size distortion. The deatils of the setup is described in (13)(14).

We propose a new test that modifies the t-test in a very simple way, yet is robust to instrument validity concerns. The test we propose depends on the idea that we can subtract the drift from the a version of t test and the drift depends on the value of the true correlation between the structural error and the instrument. Since the true correlation is between -1 and 1 we can do a grid search. Also a good property of the test is, it is monotonic and continuous in the value of true correlation. This gives rise to good inference when the null is true. We can understand the neighborhood of true correlation in that case and concur that the test fails to reject unlike the regular t. If the alternative is true our test rejects the false null regardless of correlation values as does the regular t. In finite samples to have power the modified t test works best when we have strong instruments and the true correlation is in between -0.5 to 0.5. Since the other correlation values are large, these are not plausible in a given study as long as the researcher is careful about instrument choice.

Berkowitz, Caner and Fang (2009) analyze the Anderson-Rubin test with a new resampling scheme when there is violation of exogeneity. The main assumption in that paper is violation of exogeneity, but this is local to zero. In large samples exogeneity is kept intact. In finite samples, the block size choice is important in their resampling scheme. Here, in this paper we consider a modified version of t statistics. The violation of exogeneity is not mild, this is allowed even in large samples. Our test depends on the correlation between the instrument and the error in the second stage regression. But the modified t is monotonic and continuous in the correlation, so a grid search is very helpful in inference.

The intuition behind this test is that, depending on the relative signs of the correlations and bias, there are "good" and "bad" directions of non-exogeneity. In the Acemoglu and Johnson example described above, early settler mortality might fail the strict exogeneity test because mortality might be related to morbidity, which could influence long term growth by stunting the human capital accumulation of early settlers (for a discussion see Glaeser et al., 2004). So, outside of its influence on institutions, log settler mortality decreases GDP through deficient human capital accumulation.

In linear models the instrumental variables estimate, β_{IV}, is simply the ratio of the reduced form to the first stage, $\beta_{IV} = \frac{\beta_{RF}}{\beta_{FS}}$. Here β_{RF} is the coefficient of early settler mortality on GDP and β_{FS} is the coefficient of early settler mortality on the institutional parameter (here the constraint on the executive). Since both coefficients are negative, the IV estimate is positive. If the exclusion restriction is not met because of the unobserved human capital mechanism, then the coefficient β_{RF} is "too big" in the sense that the negative effect operates both through deficient institutions and through deficient human capital accumulation. If log settler mortality were to benefit an unobserved variable, say through culling or a selection effect (e.g., only the most robust and talented individuals move where log settler mortality is high), then this reduced form coefficient would be biased toward zero, and hence the IV result would be biased toward zero. So, in effect, our test suggests that if the correlation between the instrument and the error term in the structural equation is positive (or at least not too negative) then we can still draw inference. According to the empirical exercise described in Section 5, as long as the correlation is not in the range [-1, -0.5] the modified t-ratio still indicates a rejection of the null hypothesis.

This is a very powerful result because it demonstrates that as long as the violations of exogeneity are within a set boundary, we are still able to make inferences. Since a perfectly exogenous instrument is very difficult (or perhaps impossible) to find, this test allows researchers to estimate how robust their findings are to modest violations of exogeneity.

In addition to the Acemoglu and Johnson example, we also demonstrate this method using a well-known example from labor economics. In Card (1995), the author argues that proximity to college can be used as an instrument for college attendance when calculating the returns to schooling on wages. One potential violation of exogeneity is that proximity to college is correlated with other unobserved factors that are positively associated with high earnings, such as having well-educated parents or having a higher quality public primary and secondary education. Thus the violation of perfect exogeneity is in the "bad" direction, so that proximity to college leads to higher wages both through college attendance *and* potentially through these other channels. This leads to an overstatement of results. If instead proximity were negatively correlated with the structural error, the IV estimate would be biased toward zero. We show that the modified t-ratio still rejects the null of no effect if the correlation is between [-1,0], but fails to reject with *any* positive correlation. This is true even with correlations as small as 0.01. Hence, in this case, the researcher must be cautious of any positive correlations between the instrument and the error term, even if they are only slight. The results in Card (1995) are surprising in that they indicate a larger return to college than standard OLS estimates. Most economists believe that returns to schooling are overstated, since schooling is correlated with unobserved ability, which in turn leads to higher wages. Our discussion here indicates that potential violations of the exclusion restriction, even at very small levels, could lead to incorrect inference without adjusting the t-ratios.

Section **??** provides a theoretical basis for the empirical technique. In Section **??**, we present an algorithm that suggests a simple and tractable modification of the standard t-ratio statistic. Section **??** presents simulation results which justify the practicality and efficiency of our estimates. Section 5 presents the two examples of the application of this technique in empirical research. Section 6 provides a discussion and concludes. Note that the Stata code used to implement the test statistic is included in Appendix 7.

2 The Model and Assumptions

We consider the following linear simultaneous equations model

$$y = X\beta_0 + u, \tag{1}$$

$$X = Z\pi_0 + V, \tag{2}$$

where (1) is the structural equation, and (2) is the reduced form one. X represents $n \times k$ matrix, where X_i is a $k \times 1$ vector of endogenous variables. Z is $n \times l$ matrix, and Z_i is $l \times 1$ vector of instruments, $l \geq k$. Also π_0 is of full column rank k. The errors $u_i, V_{ij}, i = 1, \cdots, n, j = 1 \cdots k$ are correlated. Control variables may be added to the system. If this is the case, simply projecting them out works in the analysis below. So in order not to complicate the analysis we abstract away from them. The variance matrix $EV_iV_i' = \Sigma_{VV} < \infty$, and nonsingular. $Eu_i^2 = \sigma_u^2 < \infty$, and $Eu_i = EV_i = 0$. Let $\hat{\beta}$ represent the two-stage least squares estimate of β_0, and $\hat{\pi}$ is the LS estimate of π_0.

In this part we discuss and present our assumptions.

Assumption 1. (i). (Violation of Exogeneity)

$$EZ_i u_i = C,$$

where C is $l \times 1$ vector with $C = (C_1, \cdots, C_m, \cdots, C_l)'$, and each $C_m \neq 0$ for $m = 1, 2, \cdots, l$ and finite.

(ii). We also have

$$EZ_i V_i' = 0.$$

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

(i).

$$\left(n^{-1} \sum_{i=1}^{n} u_i^2, n^{-1} \sum_{i=1}^{n} V_i u_i, n^{-1} \sum_{i=1}^{n} V_i V_i' \right) \xrightarrow{p} (\sigma_u^2, \Sigma_{Vu}, \Sigma_{VV}),$$

where $\sigma_u^2, \Sigma_{Vu}, \Sigma_{VV}$ are scalar, $k \times 1$, and $k \times k$ matrix, respectively. The scalar is positive , the vector is nonzero and finite, and the matrix is positive definite and finite.

(ii). We have the following law of large numbers

$$\hat{Q}_{zz} = n^{-1} \sum_{i=1}^{n} Z_i Z_i' \xrightarrow{p} Q_{zz},$$

where Q_{zz} is a positive definite and finite $k \times k$ matrix.

(iii). We have the following central limit theorem

$$\left(n^{-1/2} \sum_{i=1}^{n} (Z_i u_i - EZ_i u_i), n^{-1/2} \sum_{i=1}^{n} Z_i V_i' \right) \xrightarrow{d} (\Psi_{zu}, \Psi_{ZV}),$$

$$\begin{pmatrix} \Psi_{Zu} \\ \Psi_{ZV} \end{pmatrix} \equiv N[0, \Sigma \otimes Q_{zz}],$$

and

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

Note that Assumption 1i is the main issue of this paper. The perfect exogeneity that is used instrumental variable analysis is a knife-edge, and unrealistic assumption for applied work. Even though the researcher is careful in selecting the "perfectly exogenous" instrument there can still be unavoidable violations of exogeneity. There will be more discussion about that assumption in the next section.

Another possibility is the case of near exogeneity (a local to zero violation). This is analyzed in Berkowitz, Caner and Fang (2008) and it is shown that t-test is affected and inference becomes unreliable. But there is no solution that is proposed in that paper. In this paper, with a more realistic assumption we propose a solution to inference with Assumption 1.

Assumption 2 is basically law of large numbers and a central limit theorem. These hold under primitive conditions, such as moment conditions on the instruments and the errors, for these see Davidson (1994). One important fact to remember is the central limit theorem that we have is for $Z_i u_i - EZ_i u_i$, so that we can get a zero mean.

3 Test Statistics

In this section we discuss and analyze Assumption 1i and based on that introduce three cases of interest in applied work. First we cover one of the most widely used case of just identified system with one endogenous regressor and one instrument ($k = l = 1$). Then in our second case we consider one endogenous variable with more than one instrument ($k = 1, l \geq k$). The last case involves the general case where we may have more than one endogenous variable, and more than one instrument ($k > 1, l \geq k$).

3.1 The Just Identified Case with One Endogenous Regressor

We start with two issues that are related to standard t statistic. In this section, we propose solutions to both problems. First, the covariance between the instruments and the structural error is the issue. We can convert that to a measure involving correlation. The correlation is standardized so we can try a grid search. In that respect, assume $cov(Z_i, u_i) = C$ where C is scalar now since $k = l = 1$, for all $i = 1, \cdots n$.

See that

$$corr(Z_i, u_i) = \frac{cov(Z_i, u_i)}{\sigma_u \sqrt{var(Z_i)}}.$$

Then

$$C = \sigma_u \sigma_{zz} corr_0, \tag{3}$$

where $corr_0$ denotes the true unknown correlation, and $varZ_i = \sigma_{zz}^2$ for all $i = 1 \cdots n$.

In other words, we can write the covariance in terms of correlation, and since the correlation is between -1 to 1, we can evaluate t-test at each correlation as long as we can estimate σ_u, and σ_{zz}

consistently. At true correlation, we show that a modified t-test converges to a standard normal law, and if we make a mistake and choose the wrong correlation modified t-test diverges to infinity. So we can differentiate between the true and wrong correlation under the null distribution. This is the result in Theorem 1. The details about grid, the finite sample issues will be discussed after Theorem 1. We also consider the power issues in this section as well. There will be extensive discussions about the this modified test and the limit after Theorem 1.

The second issue is that in regular t-test, the estimator $\hat{\sigma}_u^2$ is inconsistent which is shown in (22) below. Then we impose the null of $H_0 : \beta = \beta_0$, and have the following consistent estimate under the null

$$\tilde{\sigma}_u^2 = n^{-1} \sum_{i=1}^{n} (y_i - X_i \beta_0)^2 \xrightarrow{p} E u_i^2 = \sigma_u^2.$$

We propose the following modified t-test evaluated at true correlation ($corr_0$), the issue of the using wrong correlation is shown in Theorem below.

$$modt_0 = \frac{\sqrt{n}(\hat{\beta} - \beta_0)}{\tilde{\sigma}_u [|\hat{\pi}|^{-1} \hat{Q}_{zz}^{-1/2}]} - \sqrt{n} sgn(\hat{\pi}) f(z) corr_0, \tag{4}$$

where $sgn(\hat{\pi})$ is the sign of the least squares estimate $\hat{\pi}$ in (2) for the scalar case, and

$$f(z) = \sqrt{\frac{\hat{\sigma}_{zz}^2}{\hat{\sigma}_{zz}^2 + \bar{Z}^2}},$$

where $\hat{\sigma}_{zz}^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 / n, \bar{Z} = \sum_{i=1}^n Z_i / n$.

Note that we replace the unknown C by two components. First we use consistent estimators such as $\tilde{\sigma}_u$, $\hat{\sigma}_{zz}$. Then for the true correlation since we cannot estimate it we use a grid search so that we can observe different values and this may help us in building the test statistic. So if we know the true correlation ($i.e. corr_0$)

$$\hat{C} = \tilde{\sigma}_u \hat{\sigma}_{zz} corr_0 \xrightarrow{p} C = \sigma_u \sigma_{zz} corr_0. \tag{5}$$

Of course if we do not know $corr_0$ this will not be consistent, but still see that if we choose $corr_1 = corr_0 + d/n^{1/2}$, with d a nonzero constant, still we have consistent estimation of C. In the test statistics, theorems and the discussions below, (5) will be a good guide. The main idea is if the null is true, when the test statistics is evaluated at true correlation or neighborhood of that in the grid search, the test will not be able to reject the true null, and if we make a mistake in choice of correlation values, it will reject the true null. So we can look at all the results from the grid and say that the null is not rejected. The power issues related to this approach, and the size discussion above will be substantiated with Theorems and the implications of them.

The key question is how do we obtain (4)? Why it is built in the way it is? This will be answered rigorously in the proof of Theorem 1, but here we provide a brief sketch. Lemma A.1i shows that bias in two-stage least squares estimate by using Assumption 1 is

$$(\pi_0^2 Q_{zz})^{-1}\pi_0 C,$$

when we have $k = l = 1$. If we know C, we can then subtract the least squares estimate of the bias from $\hat{\beta} - \beta_0$ and setup the test. So

$$modt = \frac{\sqrt{n}[\hat{\beta} - \beta_0 - (\hat{\pi}^2\hat{Q}_{zz})^{-1}\hat{\pi}C]}{\tilde{\sigma}_u[\hat{\pi}^2\hat{Q}_{zz}]^{-1/2}}.$$

Since we do not know C using (3) (5)

$$modt_0 = \frac{\sqrt{n}(\hat{\beta} - \beta_0)}{\tilde{\sigma}_u|\hat{\pi}|^{-1}\hat{Q}_{zz}^{-1/2}} - \sqrt{n}sgn(\hat{\pi})f(z)corr_0. \tag{6}$$

We also define another test statistic $modt_1$. In that test statistic, we make a mistake in the correlation choice (i.e. in (4)) instead of choosing the true correlation ($corr_0$), we use $corr_1 \neq corr_0$, and call the test statistic: $modt_1$. In the following Theorem, we consider the case of $k = 1$ and $l = k$ which is an empirically relevant case in most of the applied research.

Theorem 1. *Under Assumptions 1-2, with (3), and the null of $H_0 : \beta = \beta_0$, when $k = 1, l = k$*

(i).

$$modt_0 \xrightarrow{d} N(0,1).$$

(ii).

$$modt_1 \to \infty.$$

Remarks.

1. Theorem 1 shows that the modified t-ratio converges to a standard normal limit if the true correlation is used (i.e. $corr_0$). Otherwise we diverge to infinity. This theorem can help the applied researchers in their efforts for inference. It is clear that in large samples by looking at t- test value, we will be able to differentiate true correlation under the null of $H_0 : \beta = \beta_0$. We conduct some simulations to show these also. Basically, in large samples if the null is true, then at the true correlation level we do not reject the null and all the other values of the correlation we reject the

null hypotheses. We can have a very fine grid, and this helps us, as it can be seen from Figures 1-2 in simulations.

An important issue in practice is the size distortion of the regular t-test due to the violation of exogeneity. Our modified t-test remedies this problem. If the null hypothesis is true, in our grid search, either at true correlation value ($corr_0$) or in a range of correlation values around true correlation value the modified t test does not reject the true null, and prevent the size distortions. The size and power issues will be discussed at length in the Remarks below.

2. An important issue is that what if we miss the true correlation in the grid search when the null hypotheses is true? We know that $modt_1 \to \infty$, and only if we know the true correlation $corr_0$ and hence we use $modt_0$ we do not reject the null if it is true. Capturing the true correlation may be considered a remote possibility. But we argue that given a very fine grid, and strong instruments this is possible, if not we can at least pinpoint the neighborhood of the true correlation.

To show these, one interesting fact is that the modified t test is monotonic in the value of the correlation which is clear from (4). In other words, when we start the grid search from -1, and go toward 1, modt will either decrease or increase depending on $-sgn(\hat{\pi})$. This is good news if we miss the true correlation in our grid search. This prompts us to do a finer grid search. To illustrate this point assume that modt test is -2 at correlation 0, and 2 at correlation 0.1. So even though the null is true, we will be inclined to reject based on coarse grid. But since the test is monotonic and continuous we have to check and see the test values between correlations 0 and 0.1. Via Intermediate Value Theorem, the modified t statistics evaluated between 0 and 0.1 correlations are less than 5% critical value, hence the null will not be rejected. The power issue is analyzed in Remarks 7 and 8 below.

3. We should remember that the regular t-test uses $corr_0 = 0$ (accepts that that is the true correlation) and then tests the null of $H_0 : \beta = \beta_0$. Also the regular t-test assumes the two stage least squares estimator is consistent and builds $\hat{\sigma}_u^2$ on that information. Here we do not assume that two stage least squares estimate is consistent, and incorporate various values of correlation. For an extensive comparison between the modified t and the regular t-test we refer the readers to next subsection.

4. Another important point is the local analysis. What will happen to the test statistic if we are in $n^{-1/2}$ neighborhood of $corr_0$: $corr_1 = corr_0 + d/n^{1/2}$, where $d \neq 0$. To state the case rigorously, we build the modified local t-test

$$
\begin{aligned}
modt_l &= \frac{\sqrt{n}(\hat{\beta}-\beta_0)}{\tilde{\sigma}_u[|\hat{\pi}|^{-1}\hat{Q}_{zz}^{-1/2}]} - \sqrt{n} sgn(\hat{\pi}) f(z)(corr_0 + d/n^{1/2}) \\
&= modt_0 - sgn(\hat{\pi}) f(z) d \\
&\xrightarrow{d} N(D,1),
\end{aligned}
$$

where we use (10), and $D = -sgn(\hat{\pi}) f(z) d$.

So instead of $N(0,1)$ distribution as in $modt_0$ (where $modt_0$ assumes that we know the true correlation level) the true distribution is again normal with variance 1, but shifted to left or right. Then the question is: can we conduct inference? We try to answer this question in this remark and Remark 6 below.

If the null is true, then two things can happen. First since we use wrong critical values (i.e. N (0,1)) and the truth is $N(D,1)$, then $modt_l$ value may be large (compared with $modt_0$) but still we do not reject the null hypothesis of $H_0 : \beta = \beta_0$. So this is recorded as non rejection in our grid search of correlation values. The second possibility is $modt_l$ is much larger than the $N(0,1)$ critical values, and leads us to reject the null falsely at that specific correlation level (i.e. at $corr_1$, where $corr_1 = corr_0 + d/n^{1/2}$). Assuming that in our grid search we do not miss the true $corr_0$,(see Remark 2 above) by using $modt_0$ we do not reject the null. Note that choosing a very fine grid in an application reduces the probability of the second possibility.

So if the null is true, either we have a point in the grid that tells no rejection, or a neighborhood of the true correlation (a range) that shows no rejection. In either scenario we can achieve the right conclusion. This is good news from an applied perspective. The power issue relating to this choice will be discussed in Remark 7 below.

5. An important issue here is if the difference between the two-stage least squares estimate and β_0 is positive and if the sign of the true correlation is reverse of the sign of the reduced form estimate, there is no need to be concerned about inference if we reject the null with regular t test, since the modified t will be much larger (see that standard errors will be close since the bias is small). The same issue is true if the difference between the two-stage least squares estimate and β_0 is negative and the sign of the correlation has the same sign of the reduced form estimate then if we reject the null with regular t, we will reject with modified t as well. So violation of exogeneity will not change the results of the inference.

6. If the alternative is true (i.e. β_1 is true value and $\beta_0 \neq \beta_1$), then clearly $modt_0 \to \infty$. There is power against the fixed alternatives at true value of the correlation (i.e. $corr_0$). The $modt_0$ is

consistent.

As an additional fact, when H_0 is false, and the true value of β is β_1 and if we impose β_0, through Assumptions 1 and 2,

$$\tilde{\sigma}_u^2 - \sigma_u^2 \xrightarrow{p} a < \infty,$$

where $a \neq 0$,

$$a = (\beta_1 - \beta_0)^2(\pi_0^2 Q_{zz} + \Sigma_{VV}) - 2(\beta_1 - \beta_0)(\pi_0 C + \Sigma_{uV}).$$

Under the alternative the $\tilde{\sigma}_u^2$ is not consistent, however this does not affect the consistency of the $modt_0$, when we have fixed alternatives.

We now conduct a local power analysis. Set the true β as $\beta_1 = \beta_0 + c/n^{1/2}$, $c \neq 0$, note that then $a \to 0$, so $\tilde{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$.

at the true correlation

$$\begin{aligned}
\frac{\sqrt{n}(\hat{\beta} - \beta_1)}{\tilde{\sigma}_u |\hat{\pi}|^{-1} \hat{Q}_{zz}^{-1}} &- \sqrt{n} sgn(\hat{\pi}) f(z) corr_0 - \frac{\sqrt{n}(\beta_1 - \beta_0)}{\tilde{\sigma}_u |\hat{\pi}|^{-1} \hat{Q}_{zz}^{-1}} \\
&\xrightarrow{d} N(0,1) - \frac{c}{\sigma_u |\pi_0|^{-1} Q_{zz}^{-1}} \\
&\equiv N(\tilde{c}, 1),
\end{aligned}$$

where $\tilde{c} = -c|\pi_0|Q_{zz}/\sigma_u$.

So we have local power in $modt_0$ test. This also shows through $\tilde{c}$ that with strong instruments, the power will be larger. Note that with strong instruments $\tilde{c}$ (a shift in the mean compared with standard normal) will be large and it will be easy to differentiate the alternative from the null.

7. If the alternative is true, and if we make a mistake in selecting true correlation, use $corr_1 \neq corr_0$, is it plausible to have this test fail to reject the false null? So we consider $modt_1$, and analyze whether it is plausible to fail to show that alternative is true. Below we show that this is probable at implausible large correlation values only when we select strong instruments. The simulations also confirm this.

To show this issue, analyze the just identified case for simplicity. There if the true value of the structural parameter is $\beta_1 \neq \beta_0$ (alternative hypotheses is true), and we use $corr_1 \neq corr_0$

$$\begin{aligned}
modt_1 &= \frac{\sqrt{n}(\hat{\beta} - \beta_0)}{\tilde{\sigma}_u [|\hat{\pi}|^{-1} \hat{Q}_{zz}^{-1/2}]} - \sqrt{n} sgn(\hat{\pi}) f(z) corr_1 \\
&= \frac{\sqrt{n}(\hat{\beta} - \beta_1) - \sqrt{n}(\beta_1 - \beta_0)}{\tilde{\sigma}_u [|\hat{\pi}|^{-1} \hat{Q}_{zz}^{-1/2}]} - \sqrt{n} sgn(\hat{\pi}) f(z) corr_1 - \sqrt{n} sgn(\hat{\pi}) f(z)(corr_0 - corr_1).
\end{aligned}$$

It is possible then that

$$\frac{\sqrt{n}(\beta_1 - \beta_0)}{\tilde{\sigma}_u[|\hat{\pi}|^{-1}\hat{Q}_{zz}^{-1/2}]} \cong \sqrt{n} sgn(\hat{\pi}) f(z)(corr_1 - corr_0),$$

where these two terms may be equal to each other and cancel each other in the test statistic. Then $modt_1$ will not diverge to infinity but converge to a normal distribution. This may result in fail to reject the false null. Now we show that with strong instruments, this issue may occur at large correlation values. By analyzing the left hand side of the above, and using Estimator of Concentration Parameter= $CP : n\hat{\pi}^2\hat{Q}_{zz}/\tilde{\sigma}_u^2$ we have that

$$\sqrt{CP}(\beta_1 - \beta_0) \cong \sqrt{n} sgn(\hat{\pi}) f(z)(corr_1 - corr_0).$$

So if the concentration parameter is large, then the possible non rejection of the false null occurs at correlation values near -1 or +1. These are nearly implausible values in any given application (given that instruments are selected carefully, not randomly). So the problem can be avoided with large n, or using strong instruments.

8. Related to Remark 7 above and Remark 2 above, we may have non-rejection (of the null H_0) region at certain correlation values if $corr_1 = corr_0 + d/\sqrt{n}$, and if the alternative hypotheses is true. This will not be a practical issue as we show. This is related to formula above

$$\sqrt{CP}(\beta_1 - \beta_0) \cong sgn(\hat{\pi}) f(z) d. \tag{7}$$

But this may be avoided with large n or strong instruments, where the left hand and right hand sides will be far apart in (7). 9. In practice we work with finite samples and plausible correlation values are in the range of [-0.3, 0.3]. So we may do a grid search and choose the modified t test with the smallest absolute value and then compare that to standard normal distribution and conduct inference.

From an applied perspective, if we combine Remark 8 with Remark 2 above, if the null is true, the possibilities are a point of non rejection or a range of correlation values that do not reject the true null. If the alternative is right, then we may avoid non-rejection with large samples or strong instruments or both.

3.2 The Comparison Between The Regular t and the Modified t

In both test statistics, we consider $H_0 : \beta = \beta_0$, and also the analysis applies to overidentified case, but to make the comparison better we prefer to use the just identified case. The standard t test is given $(k = l = 1)$

$$t = \frac{\sqrt{n}(\hat{\beta} - \beta_0)}{\hat{\sigma}_u |\hat{\pi}|^{-1} \hat{Q}_{zz}^{-1/2}},$$

where $\hat{\sigma}_u^2 = n^{-1} \sum_{i=1}^n (y_i - x_i \hat{\beta})^2$. The modified t is

$$modt = \frac{\sqrt{n}(\hat{\beta} - \beta_0)}{\tilde{\sigma}_u |\hat{\pi}|^{-1} \hat{Q}_{zz}^{-1/2}} = \sqrt{n} sgn(\hat{\pi}) corr,$$

where $corr$ can be the true correlation ($corr_0$), then the test is $modt_0$, if it is the wrong correlation ($corr_1 \neq corr_0$), then the test is called $modt_1$ to differentiate the results. The differences between the two are clear from the equations above. First, as discussed also before, $\hat{\sigma}_u \neq \tilde{\sigma}_u$, and they are asymptotically equivalent only in the case of $\hat{\beta} \xrightarrow{p} \beta_0$. The second difference is the subtraction of the drift in modt. The regular t specifically assumes that $corr_0 = 0$, the modified t does not assume that. It tries to find the neighborhood of the true correlation if the null is true, and if the alternative is true, the finding of the true correlation is not important, since modt is consistent at all correlation levels.

Clearly, since the regular t test assumes $corr_0 = 0$, if this is not true, and the truth is some other correlation, then under the null $t \to \infty$. This is also well illustrated in the simulation in Table 1. So the size distortions with regular t test is huge, and we can almost always reject the true null. The situation gets worse with increasing the sample size. In the modified t test if we know the true correlation and this is not equal to zero, then modt converges to standard normal distribution, and has excellent size (see Tables 2-4).

Then the next question is under the true null, what if there is a mistake in the true correlation choice in the modified t test? In large samples, there are two possibilities, with a large mistake $corr_1 \neq corr_0$, the modified t test ($modt_1$ in that case) diverges to infinity as shown in Theorem 1ii. If we have a fine grid search we can catch the treu correlation since mod t values with wrong correlation will be very high, and with true correlation modt will be between the critical values in standard normal. This point is also discussed in Remark 2 after Theorem 1. If the correlation is local to $corr_0$, then the distribution is a normal distribution with drift, so we may reject the true null or not depending on the magnitude of the drift. In the regular t ratio, if the true correlation is not 0, but local to some other number, then again the regular t diverges to infinity in this case, only if the true correlation is 0 and we put $corr_1 = 0 + d/n^{1/2}$, then we have the same distribution as modt. This is the distribution in Remark 4 after Theorem 1, and Theorem 1 in Berkowitz, Caner, and Fang (2008).

Next, if the null is true, what can we say about the performance of regular t and the modified t in finite samples? Here we compare them in a simulation. In Table 1, for $n = 100$, at 5% nominal size regular t rejects the true null 20% at $corr_0 = 0.1$, and 90% at $corr_0 = 0.3$. In Table 4, at true $corr_0 = 0.1$, $modt_0$ rejects at 5%, and at $corr_0 = 0.3$ $modt_0$ rejects at 3%. There is still a very large difference between two. Even if we make a mistake in the choice of true correlation, still modified t does better. For example, if we choose a correlation of 0 or 0.2, when the truth is 0.1 the modified t rejects the true null at 16-17% compared with 20% rejection of the regular t. At true correlation of 0.3, if we make a mistake and use correlation of 0.2 or 0.4 in our modified t test, the size is 14-15%, where as the regulat t has 90% size.

Another issue is that if the regular t fails to reject the true null, is that true for the modified t as well? In large samples, regular t test chooses the correct null only when $corr_0 = 0$, this is true for $modt_0$ test as well as it is clear from Theorem 1i. In small samples, with $n = 100$ and $corr_0 = 0$, the size of $modt_0$ is 4.9% at the nominal 5% level (not shown in Tables). For standard t test, this is 5.3% as seen in Table 1. What if we make a mistake in $corr_0 = 0$ in the modified t? It is shown in Remark 2 after Theorem 1, that through Intermediate Value Theorem, we can have a fine grid and get a very close neighborhood of $corr_0$, where the modified t does not reject the true null.

When the alternative is right, both standard t and the modified t is consistent. Modified t can have some power losses in finite samples but the discussion in Remarks 7-8 after Theorem 1 shows that this can be prevented through a choice of strong instruments.

3.3 The Overidentified Case of One Endogenous Regressor

In this case, since $k = 1, l \geq k$, we assume that

$$corr(Z_{im}, u_i) = \frac{cov(Z_{im}, u_i)}{\sigma_u \sqrt{var(Z_{im})}}, \tag{8}$$

for all $m = 1, \cdots, l$. Using Assumption 1 we can rewrite (8) as, for all $i = 1, \cdots n$,

$$C_m = \sigma_u \sqrt{var(Z_{im})} corr_0. \tag{9}$$

So with (9) we assume two things in addition to the first case analyzed in section 3.1. First, the instruments are such that $cov(Z_{im}, Z_{ip}) = 0$ for all $m \neq p$, $m = 1, \cdots, l$, $p = 1, \cdots l$, $i = 1, 2, \cdots n$. The instruments are not correlated with each other. In finite samples we can handle this through simple projections as discussed in Remark 2, after Theorem 2. The second assumption in (9) is for all instruments true correlation between the structural error and the instrument is the same

($corr_0$). This is similar to regular instrumental variable estimation, there the claim is this correlation between u_i and Z_{im} is the same and 0 for all instruments. So we extend to nonzero correlations. Our results will go through with different correlations but we need multiple grid searches. For many instruments case, this is not practical.

Now we setup the test statistic. Note that this also covers the former case as well ($k = l = 1$). The reason that we have a separate section for the simple just identified case is the simplicity of the test in that case as a subcase of the following test. The following test can be built using Lemma A.1i. If we know true C in Assumption 1

$$modt = \frac{\sqrt{n}(\hat{\beta} - \beta_0 - (\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\hat{\pi}'C)}{\tilde{\sigma}_u(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1/2}}. \tag{10}$$

We can replace the infeasible test in (10) with the following by Assumption 1, (9), and extending (5) to a vector

$$modt_0 = \frac{\sqrt{n}(\hat{\beta} - \beta_0 - (\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\hat{\pi}'[\tilde{\sigma}_u\sqrt{\widehat{var(Z_1)}}, \cdots, \tilde{\sigma}_u\sqrt{\widehat{var(Z_l)}}]'corr_0)}{\tilde{\sigma}_u(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1/2}},$$

where $\tilde{\sigma}_u$ is the square root of the estimator $\tilde{\sigma}_u^2$, and $\widehat{var(Z_m)} = \frac{1}{n}\sum_{i=1}^n (Z_{im} - \bar{Z}_m)^2$ where $\bar{Z}_m = n^{-1}\sum_{i=1}^n Z_{im}$ for $m = 1, \cdots l$. We can further simplify the test above as

$$modt_0 = \frac{\sqrt{n}(\hat{\beta} - \beta_0 - (\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\bar{\pi}corr_0)}{\tilde{\sigma}_u(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1/2}}, \tag{11}$$

where $\bar{\pi} = \hat{\pi}'[\tilde{\sigma}_u\sqrt{\widehat{var(Z_1)}}, \cdots, \tilde{\sigma}_u\sqrt{\widehat{var(Z_l)}}]'$ which is scalar.

Note that (10) simplifies to (4) as can be seen from (6).

In the same way as in the just identified case, define another test statistic $modt_1$. In that test statistic, we make a mistake in the correlation choice (i.e. in (11)) instead of choosing the true correlation ($corr_0$), we use $corr_1 \neq corr_0$, and call the test statistic: $modt_1$.

Theorem 2. *Under Assumptions 1-2, with (9), and the null of $H_0 : \beta = \beta_0$, when $k = 1, l \geq k$*

(i).

$$modt_0 \xrightarrow{d} N(0, 1).$$

(ii).

$$modt_1 \to \infty.$$

Remarks. 1. Theorem 2 also shows that $modt_0$ still works when $k = 1, l \geq k$. In large samples at true correlation level test statistic does not reject the null if H_0 is true. At other values of correlation the test rejects the null. In the finite samples, this case is exactly the same as the just identified case. Choosing a fine grid with strong instruments ensures good size and power. If we choose the wrong correlation, and still do not reject H_0 (when H_0 is false) then as in the just identified case, choosing strong instruments solves the problem. If we choose the wrong correlation and this time if the null is true, and our test rejects the true null this is easily fixed. Since the test is monotonic in the correlation, choosing a fine grid and conducting the tests in these new correlation values, we should be able to not reject the true null.

2. Note that in finite samples, instruments may be correlated. So we can use the following. Assume that we have two instruments: Z_{i1}, Z_{i2}, for $i = 1, \cdots n$. We regress (least squares) Z_{i1} on Z_{i2} and define the residual as $Z_{i1\perp}$. Then we use $Z_{i1\perp}$ and Z_{i2} in the test statistic.

3. The local analysis in Remark 4 for the simple just identified case carries over if the null is true. So if $corr_1 = corr_0 + d/n^{1/2}$, we can get no rejection of the null since the modified t will converge in distribution to a normal law with a constant non zero mean (variance 1). So if the null is true, still test may not reject H_0, if it rejects the true null, then we know that at true $corr_0$ we do not reject due to Theorem 2i. So in any case, we do not reject the true null at a point or in a range of possible correlation values. Missing the true correlation in the grid issue is handled in the same way through a monotonicity argument as in Remark 2 after Theorem 1.

4. The overall advice to applied researcher is to try plausible correlation values [-0.3, 0.3] in a very fine grid and record no rejection of the null at certain correlation value/s. If there are such values we do not reject the null, if all the values of a fine grid rejects the null then the alternative is true, and we reject the false null.

5. Note that at $corr_0 = 0$, $modt_0$ and standard t is not the same due to the usage of $\tilde{\sigma}$ versus $\hat{\sigma}$ respectively in their denominators which is explained in section 3.1. Asymptotically they converge to the same limit under the null if $corr_0 = 0$.

3.4 The General Case

For testing individual coefficients when $k > 1$ (multiple endogenous variables), the modified t-ratio test does not work, since to get a consistent estimate for σ_u (i.e.$\hat{\sigma}_u$) we need to impose for all $\beta = \beta_0$. This is only plausible when $k = 1$, and we are using modified t-ratio test for the only structural coefficient, or if we have a joint modified Wald test for $H_0 : R\beta = R\beta_0$, where R is a $j \times k$ matrix. This test will include all parameters corresponding to endogenous regressors in the structural equation.

So we now introduce the modified Wald test evaluated at $corr_0$ (true correlation). This is constructed in the same way as modified t-test, and the proof is the same, and hence is skipped. Still we impose (9) with Assumption 1

$$C_m = \sigma_u \sqrt{var Z_{im}} corr_0, \tag{12}$$

for all $m = 1, \cdots l, i = 1, \cdots n$.

We want to test $H_0 : R\beta = R\beta_0$. Define $modW_0$ as follows, by extending (5) to a vector

$$\begin{aligned} modW_0 &= n[R\hat{\beta} - R\beta_0 - R(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\hat{\pi}'[\tilde{\sigma}_u\sqrt{\widehat{var(Z_1)}}, \cdots, \tilde{\sigma}_u\sqrt{\widehat{var(Z_l)}}]' corr_0]' \\ &\times (R\tilde{\sigma}_u^2(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}R')^{-1} \\ &\times [R\hat{\beta} - R\beta_0 - R(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\hat{\pi}'[\tilde{\sigma}_u\sqrt{\widehat{var(Z_1)}}, \cdots, \tilde{\sigma}_u\sqrt{\widehat{var(Z_l)}}]' corr_0]. \end{aligned}$$

Let $modW_1$ be the modified Wald, when we make a mistake in correlation choice, and use $corr_1$ instead of true correlation $corr_0$.

Theorem 3. *Under Assumptions 1-2, with (12), and under the joint null of $R\beta = R\beta_0$, we have*

(i).

$$modW_0 \xrightarrow{d} \chi^2_j.$$

(ii).

$$modW_1 \to \infty.$$

Remarks.

1.When we use $corr_1$ (make a mistake in selection of true correlation) then $modW_1 \to \infty$. as in Theorem 1 for modified t-test, where $modW_1$ represents the modified Wald where $corr_1$ is used instead of $corr_0$ in the modified Wald test above.

2. If we make a minor mistake in our correlation choice can we still have "no rejection" of the true null? If we set $corr_1 = corr_0 + d/n^{1/2}$, and denote the modified Wald as $modW_l$, then following the analysis in Remark 4, we get a non central χ^2 distribution. So it is possible to reject the true null,(at least in certain correlation values in the neighborhood of $corr_0$) and hence make a mistake. But if we have a very fine grid, we can definitely evaluate the test statistic at certain values of correlation near the true correlation, and do not reject the true null in some of them, and make a correct inference.

3. The test is consistent with a fixed true alternative. This can be shown easily. With a local alternative, also there is still power if we are at true correlation. When we use correlation values different than the true correlation $corr_1 \neq corr_0$, then it is possible to fail to reject the false null. However, as shown in Remarks 7-8 in the just identified case, using strong instruments solves the problem.

4 Simulation

In this part of the paper we conduct simulations. We try to answer the following questions. First, can we verify the results of Theorem 1? Namely, can we see that $modt_0 \stackrel{d}{\to} N(0,1)$, and $modt_1 \to \infty$ in large samples?. The second issue is the finite sample behavior of the test statistic $modt_0$. The issue is whether in the finite samples given a grid search (it may be a very fine grid search, with very small steps in a given empirical application) is the smallest rejection level still corresponds to $modt_0$? (Since $modt_1 \to \infty$, wrong choice of correlation can result in large rejection rates). The third question is related to power of the test. Is there a power loss at certain grid points as discussed after Theorem 1? If there is can we also see that they are near extreme correlation values for a given application. If this power loss occurs away from [-0.3, 0.3] range of correlations that power loss may not be important. We generate the data with one instrument ($l = 1$), one endogenous regressor ($k = 1$) and no control variables.

$$y_i = X_i\beta_0 + u_i, \tag{13}$$

$$X_i = Z_i\pi + V_i, \tag{14}$$

where $\beta_0 = 0$ (for the size exercise), and $\pi = 2$. The structural error u_i, the reduced form error V_i, and the instrument are iid. These are generated from the same joint normal distribution with $N(0, \Lambda)$, where

$$\Lambda = \begin{bmatrix} 1 & cov(Z_i, u_i) & 0 \\ cov(Z_i, u_i) & 1 & cov(V_i, u_i) \\ 0 & cov(u_i, V_i) & 1 \end{bmatrix},$$

since $varZ_i = 1$, $varu_i = 1$, $cov(Z_i, u_i)$ is also the correlation between Z_i, u_i. This is denoted as $corr_0$ in the other sections. The covariance between V_i, u_i is set at 0.5. Since the variances are set at 1, the true correlation between the structural error and the instrument varies among -0.5, -0.3, -0.1, 0.1, 0.3, 0.5. The grid step is 0.1 for the Tables. For the graphs the true correlation is set at

-0.1, 0.25, 0.52 and the grid step is 0.01. The sample sizes are $n = 100, 200, 1000$. The iteration number is 10000. For the size exercise, we report the percentage of rejections at 5% critical values from the standard normal distribution (-1.96, +1.96).

Table 1 provides the size of $modt_0, modt_1$ tests at $n = 1000$. In Table 1, $corr_0$ represents the true correlation between the structural error (u_i) and the instrument (Z_i). The first column is the grid values of the correlation "Grid". When the grid value is equal to $corr_0$ then the size of the test($modt_0$) should be 5% at that level ideally. Otherwise if the grid value of the correlation is not equivalent to $corr_0$ then the size of the test (now the test becomes $modt_1$) should be near 100% according to Theorem 1. We see that the results in Table 1 confirm Theorem 1. Namely, the size of the $modt_0$ test is at 1-5% level. (i.e. at $corr_0 = -0.5, modt_0$ is the one that corresponds to $Grid = -0.5$. Otherwise when $Grid = corr_1 \neq corr_0$ the test is named $modt_1$. When we look at $modt_1$ test the rejection rate is 88-100% at 5% nominal level. So if we have a grid search of the correlation, then only at true value we get the 5% rejection at nominal level, otherwise we almost always reject the null. In that sense, we can differentiate the true correlation by looking at the absolute value of the modified t statistic. We can choose the one with the smallest absolute value and compare it with standard normal distribution. To see how reliable is this in finite samples, we conduct the same exercise with $n = 100, n = 200$ observations. For $n = 200, n = 100$ in Tables 2-3 we see that $modt_0$ test achieve 1-5% size at 5% nominal level. This is very good, and confirms that even in the finite samples the asymptotic approximation is very good. For $modt_1$ tests (i.e. when $corr_1 \neq corr_0$) the situation is different than the one in Table 1. Table 2 shows the size of the tests at $n = 200$. For example, at true correlation of $corr_0 = -0.1$, the $modt_0$ has the size of 4.5%, and $modt_1$ ($corr_1 = -0.2$) the size is 29.6% rather than near 100%. But still there is substantial difference between $modt_1$ and $modt_0$ test sizes. So picking up and using the smallest modified t test in inference (in absolute terms) in a correlation search makes sense. At $n = 100$ in Table 3, still $modt_0$ has the smallest level. Tables 2-3 support our claim in Remarks 2 and 4 (in just identified case) of the possibility of region of no rejection of the true null when we select correlation values near the true correlation but miss the true correlation. This region is around the true correlation value. We also report size results with a much finer grid of 0.01, these are shown in Figures ?.

Tables 4-7, report the percentage of the rejections of the false null hypothesis for $modt_0$ and $modt_1$ tests. We have the same number of iterations as the size exercise, and the same critical values are used. The true values of $\beta = -2, -1, 1, 2$, and we test $H_0 : \beta = 0$, and $n = 100, 1000$. The results confirm the remarks after Theorem 1. Namely, the power of $modt_0$ is very good almost all the relevant correlation levels for applications (-0.5, +0.5). Even with a mistake (using $corr_1 = Grid \neq corr_0$, and the test is $modt_1$), the power is still very good at the range of the correlation values of [-0.5, 0.5]. There are certain power losses around high implausible correlation levels, but as can be seen with large sample size this problem is less important. We also experiment with increasing the concentration parameter estimate by putting $\pi = 5$, and this gives much better power results. We also experiment with $\beta = -0.5, -0.3, 0.3, 0.5$, the results are very similar even in this close neighborhood of 0. These are not reported.

Overall, we think that the applied researcher may use this method for a very fine grid between [-0.3, 0.3] for the modified t-ratio test that is described. If s/he gets a region of no rejection of the null this is a good check that the null hypothesis is not rejected. If that region shows only rejection for all values of the correlation grid for the null, then the alternative is true.

5 Empirical Examples

Since the modified t test can choose the wrong correlation, this may cause problems. However, if the null is true, we can learn the neighborhood of the true correlation as described in Remarks 2 and 4 after Theorem 1. In that neighborhood, we do not reject the true null. If the alternative is true, whether we use the true correlation or not the modified t will reject H_0.

We apply this technique to two empirical examples. First, we replicate the results from Acemoglu and Johnson (2005), hereafter AJ. As discussed in the introduction, the main results in AJ utilize log settler mortality to instrument for institutions when measuring the effect of institutions on economic growth as measured by GDP per capita. For our study, we have obtained the data used by AJ on 64 countries. In this discussion we focus on Table 2 of AJ which provides estimates for the just identified case of one instrument and one endogenous variable. In Table 2, Panel C, Column (3) of AJ, the two-stage least squares estimate of the effect of the constraint on executive power on GDP per capita is 0.76 with a standard error of 0.15. This coefficient is interpreted as highly statistically significant under standard inference. However, as shown in Berkowitz et al. (2008), this estimate is inconsistent and the standard t-test is biased.

To resolve this, we implemented our modified t-test procedure, as shown in Table 4 Column (1). As long as the correlation is not in the range [-1, -0.5] then the modified t-ratio still indicates a rejection of the null hypothesis. In other words, our test indicates that as long as the correlation (or non-exogeneity) is not too extreme, the estimate is robust and we can be assured that the true

Table 1: Size (5% level), $modt_0, modt_1$ $n = 1000$

Grid	$corr_0 = -0.5$	$corr_0 = -0.3$	$corr_0 = -0.1$	$corr_0 = 0.1$	$corr_0 = 0.3$	$corr_0 = 0.5$
-1	100.0	100.0	100.0	100.0	100.0	100.0
-0.9	100.0	100.0	100.0	100.0	100.0	100.0
-0.8	100.0	100.0	100.0	100.0	100.0	100.0
-0.7	100.0	100.0	100.0	100.0	100.0	100.0
-0.6	94.7	100.0	100.0	100.0	100.0	100.0
-0.5	0.8	100.0	100.0	100.0	100.0	100.0
-0.4	94.1	90.8	100.0	100.0	100.0	100.0
-0.3	100.0	3.7	100.0	100.0	100.0	100.0
-0.2	100.0	90.0	88.7	100.0	100.0	100.0
-0.1	100.0	100.0	4.8	100.0	100.0	100.0
0.0	100.0	100.0	88.7	88.1	100.0	100.0
0.1	100.0	100.0	100.0	4.9	100.0	100.0
0.2	100.0	100.0	100.0	88.2	90.4	100.0
0.3	100.0	100.0	100.0	100.0	3.1	100.0
0.4	100.0	100.0	100.0	100.0	91.4	94.4
0.5	100.0	100.0	100.0	100.0	100.0	0.9
0.6	100.0	100.0	100.0	100.0	100.0	95.1
0.7	100.0	100.0	100.0	100.0	100.0	100.0
0.8	100.0	100.0	100.0	100.0	100.0	100.0
0.9	100.0	100.0	100.0	100.0	100.0	100.0
1.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: Grid represents the grid correlation values that we put into the modified t-tests. When $Grid = corr_0$, the we have $modt_0$ test, otherwise the tests are $modt_1$. The critical values are -1.96, +1.96. We set $\pi = 2$. For example, in column 2, $corr_0 = -0.5$, when $Grid = -0.5$, the test is called $modt_0$, otherwise the tests are called $modt_1$.

Table 2: Size (5% level), $modt_0, modt_1$ $n = 200$

Grid	$corr_0 = -0.5$	$corr_0 = -0.3$	$corr_0 = -0.1$	$corr_0 = 0.1$	$corr_0 = 0.3$	$corr_0 = 0.5$
-1	100.0	100.0	100.0	100.0	100.0	100.0
-0.9	100.0	100.0	100.0	100.0	100.0	100.0
-0.8	100.0	100.0	100.0	100.0	100.0	100.0
-0.7	88.8	100.0	100.0	100.0	100.0	100.0
-0.6	24.0	100.0	100.0	100.0	100.0	100.0
-0.5	1.1	83.4	100.0	100.0	100.0	100.0
-0.4	22.7	27.0	99.0	100.0	100.0	100.0
-0.3	86.5	3.1	80.9	100.0	100.0	100.0
-0.2	100.0	26.4	28.9	98.9	100.0	100.0
-0.1	100.0	82.8	4.9	80.7	100.0	100.0
0.0	100.0	100.0	29.5	28.0	100.0	100.0
0.1	100.0	100.0	80.8	4.5	82.1	100.0
0.2	100.0	100.0	98.8	29.1	27.5	100.0
0.3	100.0	100.0	100.0	81.4	3.0	86.2
0.4	100.0	100.0	100.0	100.0	27.8	22.4
0.5	100.0	100.0	100.0	100.0	83.4	1.0
0.6	100.0	100.0	100.0	100.0	100.0	24.3
0.7	100.0	100.0	100.0	100.0	100.0	89.2
0.8	100.0	100.0	100.0	100.0	100.0	100.0
0.9	100.0	100.0	100.0	100.0	100.0	100.0
1.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: Grid represents the grid correlation values that we put into the modified t-tests. When $Grid = corr_0$, the we have $modt_0$ test, otherwise the tests are $modt_1$. The critical values are -1.96, +1.96. We set $\pi = 2$. For example, in column 2, $corr_0 = -0.5$, when $Grid = -0.5$, the test is called $modt_0$, otherwise the tests are called $modt_1$.

Table 3: Size (5% level), $modt_0, modt_1$ $n = 100$

Grid	$corr_0 = -0.5$	$corr_0 = -0.3$	$corr_0 = -0.1$	$corr_0 = 0.1$	$corr_0 = 0.3$	$corr_0 = 0.5$
-1	100.0	100.0	100.0	100.0	100.0	100.0
-0.9	100.0	100.0	100.0	100.0	100.0	100.0
-0.8	93.5	100.0	100.0	100.0	100.0	100.0
-0.7	52.8	100.0	100.0	100.0	100.0	100.0
-0.6	11.2	88.5	100.0	100.0	100.0	100.0
-0.5	1.0	51.5	98.3	100.0	100.0	100.0
-0.4	8.1	15.8	85.7	100.0	100.0	100.0
-0.3	51.7	2.7	51.9	97.9	100.0	100.0
-0.2	90.3	13.6	16.6	85.1	100.0	100.0
-0.1	100.0	52.1	4.5	51.6	98.5	100.0
0.0	100.0	87.0	16.5	16.2	86.6	100.0
0.1	100.0	98.5	51.3	4.6	52.4	100.0
0.2	100.0	100.0	85.4	16.8	13.8	90.1
0.3	100.0	100.0	97.9	51.2	2.9	51.7
0.4	100.0	100.0	100.0	85.7	14.9	8.1
0.5	100.0	100.0	100.0	98.4	52.2	1.2
0.6	100.0	100.0	100.0	100.0	88.4	11.7
0.7	100.0	100.0	100.0	100.0	100.0	52.5
0.8	100.0	100.0	100.0	100.0	100.0	93.7
0.9	100.0	100.0	100.0	100.0	100.0	100.0
1.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: Grid represents the grid correlation values that we put into the modified t-tests. When $Grid = corr_0$, the we have $modt_0$ test, otherwise the tests are $modt_1$. The critical values are -1.96, +1.96. We set $\pi = 2$. For example, in column 2, $corr_0 = -0.5$, when $Grid = -0.5$, the test is called $modt_0$, otherwise the tests are called $modt_1$.

Table 4: Rejection percentage of $H_0 : \beta = 0$, $modt_0, modt_1$, $corr_0 = 0.1$

	$n = 1000$				$n = 100$			
Grid	$\beta = -2$	$\beta = -1$	$\beta = 1$	$\beta = 2$	$\beta = -2$	$\beta = -1$	$\beta = 1$	$\beta = 2$
-1	0.0	100.0	100.0	100.0	0.0	0.0	100.0	100.0
-0.9	0.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0
-0.8	0.0	100.0	100.0	100.0	0.0	0.0	100.0	100.0
-0.7	79.2	100.0	100.0	100.0	79.2	24.9	100.0	100.0
-0.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.5	100.0	100.0	100.0	100.0	100.0	100.0	94.7	100.0
0.6	100.0	100.0	100.0	100.0	100.0	100.0	25.1	90.1
0.7	100.0	100.0	76.6	100.0	100.0	100.0	0.0	0.7
0.8	100.0	100.0	0.8	0.5	100.0	100.0	0.1	0.0
0.9	100.0	100.0	100.0	41.1	100.0	100.0	7.2	0.1
1.0	100.0	100.0	100.0	100.0	100.0	100.0	81.3	14.4

Note: Grid represents the grid correlation values that we put into the modified t-tests. When $corr_0 = Grid$, then we have $modt_0$ test, otherwise the tests are $modt_1$. The critical values are -1.96, +1.96. We set $\pi = 2$. Here $modt_0$ is when $corr_0 = Grid = 0.1$.

Table 5: Rejection percentage of $H_0 : \beta = 0$, $modt_0, modt_1$ $corr_0 = -0.1$

	$n = 1000$				$n = 100$			
Grid	$\beta = -2$	$\beta = -1$	$\beta = 1$	$\beta = 2$	$\beta = -2$	$\beta = -1$	$\beta = 1$	$\beta = 2$
-1	0.0	100.0	100.0	100.0	0.0	100.0	100.0	100.0
-0.9	0.0	0.0	100.0	100.0	0.0	74.3	100.0	100.0
-0.8	0.0	100.0	100.0	100.0	0.0	13.3	100.0	100.0
-0.7	89.5	100.0	100.0	100.0	90.3	0.5	100.0	100.0
-0.6	100.0	100.0	100.0	100.0	100.0	0.3	100.0	100.0
-0.5	100.0	100.0	100.0	100.0	100.0	31.2	100.0	100.0
-0.4	100.0	100.0	100.0	100.0	100.0	89.5	100.0	100.0
-0.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.5	100.0	100.0	100.0	100.0	100.0	100.0	80.7	100.0
0.6	100.0	100.0	100.0	100.0	100.0	100.0	7.7	80.2
0.7	100.0	100.0	5.4	100.0	100.0	100.0	0.0	0.2
0.8	100.0	100.0	45.9	0.0	100.0	100.0	0.6	0.0
0.9	100.0	100.0	100.0	86.0	100.0	100.0	24.3	0.1
1.0	100.0	100.0	100.0	100.0	100.0	100.0	95.3	26.8

Note: Grid represents the grid correlation values that we put into the modified t-tests. When $corr_0 = Grid$, then we have $modt_0$ test, otherwise the tests are $modt_1$. The critical values are -1.96, +1.96. We set $\pi = 2$. Here $modt_0$ is when $corr_0 = Grid = -0.1$.

Table 6: Rejection percentage of $H_0 : \beta = 0$, $modt_0, modt_1$ $corr_0 = 0.3$

	$n = 1000$				$n = 100$			
Grid	$\beta = -2$	$\beta = -1$	$\beta = 1$	$\beta = 2$	$\beta = -2$	$\beta = -1$	$\beta = 1$	$\beta = 2$
-1	100.0	100.0	100.0	100.0	0.0	1.3	100.0	100.0
-0.9	0.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0
-0.8	100.0	90.1	100.0	100.0	0.0	0.0	100.0	100.0
-0.7	100.0	100.0	100.0	100.0	6.7	9.9	100.0	100.0
-0.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.6	100.0	100.0	100.0	100.0	100.0	100.0	55.1	95.7
0.7	100.0	100.0	100.0	100.0	100.0	100.0	0.0	2.5
0.8	100.0	100.0	0.0	3.1	100.0	100.0	0.0	0.0
0.9	100.0	100.0	100.0	41.8	100.0	100.0	1.0	0.0
1.0	100.0	100.0	100.0	100.0	100.0	100.0	53.4	6.8

Note: Grid represents the grid correlation values that we put into the modified t-tests. When $corr_0 = Grid$, then we have $modt_0$ test, otherwise the tests are $modt_1$. The critical values are -1.96, +1.96. We set $\pi = 2$. Here $modt_0$ is when $corr_0 = Grid = 0.3$.

Table 7: Rejection percentage of $H_0 : \beta = 0$, $modt_0, modt_1$ $corr_0 = -0.3$

	$n = 1000$				$n = 100$			
Grid	$\beta = -2$	$\beta = -1$	$\beta = 1$	$\beta = 2$	$\beta = -2$	$\beta = -1$	$\beta = 1$	$\beta = 2$
-1	98.6	100.0	100.0	100.0	0.0	0.0	100.0	100.0
-0.9	0.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0
-0.8	100.0	100.0	100.0	100.0	0.0	0.0	100.0	100.0
-0.7	100.0	100.0	100.0	100.0	96.5	91.8	100.0	100.0
-0.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
-0.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.4	100.0	100.0	100.0	100.0	100.0	100.0	97.2	100.0
0.5	100.0	100.0	100.0	100.0	100.0	100.0	56.4	100.0
0.6	100.0	100.0	100.0	100.0	100.0	100.0	1.5	67.3
0.7	100.0	100.0	0.0	100.0	100.0	100.0	0.1	0.0
0.8	100.0	100.0	98.3	0.0	100.0	100.0	3.7	0.0
0.9	100.0	100.0	100.0	98.9	100.0	100.0	50.1	0.4
1.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	40.7

Note: $corr_1$ represents the grid correlation values that we put into the modified t-tests. When $corr_1 = corr_0$, then we have $modt_0$ test, otherwise the tests are $modt_1$. The critical values are -1.96, +1.96. $\pi = 2$. Here $modt_0$ is when $corr_1 = corr_0 = -0.3$.

Table 8: Modified T-Scores for Empirical Examples

Correlation	Acemoglu and Johnson	Card No Covars	Card Full
-1	-2.50	70.19	62.42
-0.9	-1.65	64.18	56.41
-0.8	-0.81	58.17	50.40
-0.7	0.04	52.16	44.39
-0.6	0.89	46.15	38.38
-0.5	1.74	40.13	32.37
-0.4	2.59	34.12	26.36
-0.3	3.44	28.11	20.35
-0.2	4.29	22.10	14.34
-0.1	5.13	16.09	8.33
0.0	5.98	10.08	2.32
0.1	6.83	4.07	-3.69
0.2	7.68	-1.94	-9.71
0.3	8.53	-7.95	-15.72
0.4	9.38	-13.96	-21.73
0.5	10.23	-19.97	-27.74
0.6	11.07	-25.98	-33.75
0.7	11.92	-32.00	-39.76
0.8	12.77	-38.01	-45.77
0.9	13.62	-44.02	-51.78
1	14.47	-50.03	-57.79

effect is statistically different from zero. This is a very powerful result because it demonstrates that as long as the violations of exogeneity are within a set boundary, we are still able to make inferences.

Next we consider David Card's 1995 paper using proximity to a college as an instrument for educational attainment. This paper finds much larger returns to education relative to previous work. As in AJ, the instrumental variable in Card (1995) may not be completely exogenous, leading to somewhat biased results. We have obtained the original data set used in this analysis and have replicated the main results in Table 3, Column (5). We present two sets of modified t-ratios for the Card (1995) results. The first, in Table 4, Column (2) is a specification with no covariates, while Column (3) includes covariates as in the original Card paper. With the no covariates case, we reject the null if the correlation is in the range [-1,0.13], while with covariates we reject only for non-positive correlations. This is a more extreme result than for AJ, where small correlations did not affect inference. In the Card example, even small amounts of correlation between the instrument and the structural error lead to non-rejection.

6 Discussion and Conclusion

REFERENCES

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001), "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369-1401.

Acemoglu, Daron and Simon Johnson (2006), "Unbundling Institutions," *Journal of Political Economy*, 113, 949-995.

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

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

Card, David (1995), "Using Geographic Variation in College Proximity to Estimate the Returns to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honor of John Vanderkamp*, eds. L. N. Christofides et al. Toronto: University of Toronto Press. 201-221.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer (2003), "Courts," *Quarterly Journal of Economics*, 118, 453-517.

Duflo, Esther (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment,," *American Economic Review*, Vol. 91(4), pages 795-813, September.

Glaeser, Edward, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer (2004), "Do Institutions Cause Growth?" *Journal of Economic Growth*, 9, 271-303.

Guiso, Luigi, Paola Sapienza and Luigi Zingales (2006), "Does Culture Affect Economic Outcomes," *Journal of Economic Perspectives*, vol. 20(2), pages 23-48, Spring.

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

Kane, Thomas J. and Cecelia E. Rouse (1995), "Labor-Market Returns to Two- and Four-Year Colleges," *American Economic Review*, 85, 600-614.

Mauro, Paolo (1995), "Corruption and Growth," *Quarterly Journal of Economics*, 110, 681-712.

Meghir, Costas and Marten Palme (2005), "Assessing the Effect of Schooling on Earnings Using a Social Experiment," *American Economic Review*, March 2005, pp 414-424.

Appendix

Before the proof of main Theorem, we need the following result that will help us in understanding the main result. This holds for both the simple case just identified case, $k = l = 1$, and the overidentified case $k = 1, l > k$.

Lemma A.1. *Under Assumptions 1-2, and under the null hypotheses of $H_0 : \beta = \beta_0$,*

(i).

$$\hat{\beta} - \beta_0 \xrightarrow{p} [\pi_0' Q_{zz} \pi_0]^{-1} [\pi_0' C] \neq 0.$$

(ii).

$$\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2 - 2C_\pi(\pi_0' C + \Sigma_{Vu}) + C_\pi^2 (\pi_0' Q_{zz} \pi_0 + \Sigma_{VV}).$$

(iii).

$$t_{2sls} \to \infty,$$

where t_{2sls} represents the regular two stage least squares based t-test.

Proof of Lemma A.1. We analyze a system with $k = 1$, and multiple instruments ($l \geq k$). First we show that $\hat{\beta}$ is inconsistent given Assumption 1.

$$\hat{\beta} - \beta_0 = \left[\left(\frac{X'Z}{n} \right) \left(\frac{Z'Z}{n} \right)^{-1} \left(\frac{Z'X}{n} \right) \right]^{-1} \left[\left(\frac{X'Z}{n} \right) \left(\frac{Z'Z}{n} \right)^{-1} \left(\frac{Z'u}{n} \right) \right]. \tag{15}$$

See that by reduced form equation and Assumption 2

$$\begin{aligned} \frac{Z'X}{n} &= \frac{Z'Z}{n} \pi_0 + \frac{Z'V}{n} \\ &\xrightarrow{p} Q_{zz} \pi_0. \end{aligned} \tag{16}$$

$$n^{-1} \sum_{i=1}^n Z_i u_i \xrightarrow{p} E Z_i u_i = C < \infty. \tag{17}$$

Use (16)(17) in (15) to have

$$\hat{\beta} - \beta_0 \xrightarrow{p} [\pi_0' Q_{zz} \pi_0]^{-1} [\pi_0' C] \neq 0, \tag{18}$$

as long as $C \neq 0$. Next we show that $\hat{\sigma}_u^2$ is not a consistent estimator for σ_u^2. First

$$\begin{aligned} n^{-1} \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2 &= n^{-1} \sum_{i=1}^n (u_i - (x_i(\hat{\beta} - \beta_0))^2 \\ &= n^{-1} \sum_{i=1}^n u_i^2 - 2(\hat{\beta} - \beta_0) n^{-1} \sum_{i=1}^n x_i u_i \\ &+ (\hat{\beta} - \beta_0)^2 n^{-1} \sum_{i=1}^n x_i^2. \end{aligned} \tag{19}$$

See that from (18), set $C_\pi = [\pi_0' Q_{zz} \pi_0]^{-1} [\pi_0' C]$, by Assumption 2i, Assumption 1, and using reduced form equation

$$\begin{aligned} \frac{X'u}{n} &= \frac{\pi_0' Z'u}{n} + \frac{V'u}{n} \\ &\xrightarrow{p} \pi_0' C + \Sigma_{Vu}, \end{aligned} \tag{20}$$

$$\begin{aligned} \frac{X'X}{n} &= \frac{\pi_0' Z' Z \pi_0}{n} + \frac{2\pi_0' Z' V}{n} + \frac{V'V}{n} \\ &\xrightarrow{p} \pi_0' Q_{zz} \pi_0 + \Sigma_{VV}. \end{aligned} \tag{21}$$

Use (18)(20)(21) in (19)

$$\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2 - 2C_\pi(\pi_0' C + \Sigma_{Vu}) + C_\pi^2(\pi_0' Q_{zz}\pi_0 + \Sigma_{VV}) < \infty. \tag{22}$$

So the last two terms are nonzero (unless they cancel each other in special empirical cases). We cannot use $\hat{\sigma}_u^2$ as a consistent estimator. Next we show that under Assumption 1, the t-test for $H_0 : \beta = \beta_0$

$$t_{2sls} = \frac{\sqrt{n}(\hat{\beta} - \beta_0)}{\hat{\sigma}_u (X' P_Z X)^{-1/2}} \to \infty,$$

by (15)(18)(22). **Q.E.D.**

Proof of Theorem 1. The proof is a subcase of the proof of Theorem 2, since that proof is for $k = 1, l \geq k$.

Proof of Theorem 2. This proof is for $k = 1, l \geq k$, and hence covers the cases of $k = l = 1$, and $k = 1, l > k$. Now we show that a modified t-test converges in distribution to standard normal distribution. In that respect, we first try to understand the numerator of the new test statistic. See that

$$\begin{aligned} n^{1/2}(\hat{\beta} - \beta_0) &= \left[\left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{Z'X}{n}\right) \right]^{-1} \\ &\times \left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{1}{n^{1/2}} \sum_{i=1}^n (Z_i u_i - EZ_i u_i) + n^{1/2} EZ_i u_i \right) \\ &= \{ \left[\left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{Z'X}{n}\right) \right]^{-1} \left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{1}{n^{1/2}} \sum_{i=1}^n (Z_i u_i - EZ_i u_i) \right) \} \\ &+ \{ \left[\left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \left(\frac{Z'X}{n}\right) \right]^{-1} \left(\frac{X'Z}{n}\right) \left(\frac{Z'Z}{n}\right)^{-1} \sqrt{n} C \} \\ &= A_1 + A_2, \end{aligned} \tag{23}$$

where A_1, A_2 represent the first and the second terms with curly bracket expressions. By using Assumption 2, (16)

$$A_1 \xrightarrow{d} N(0, \sigma_u^2 (\pi_0' Q_{zz} \pi_0)^{-1}). \tag{24}$$

Then by (16) and Assumptions 1, 2ii

$$A_2 \to \infty.$$

So we definitely have to subtract A_2 from $\sqrt{n}(\hat{\beta} - \beta_0)$ term. But the real issue is the handling of C. So we handle that by the arguments in the main text. Given (16)(23)(24) we have (if we had known true C)

$$modt_0 = \frac{\sqrt{n}(\hat{\beta} - \beta_0 - (\hat{\pi}' \hat{Q}_{zz} \hat{\pi})^{-1} \hat{\pi}' C)}{\tilde{\sigma}_u (\hat{\pi}' \hat{Q}_{zz} \hat{\pi})^{-1/2}} \xrightarrow{d} N(0, 1). \tag{25}$$

Equivalently via (10)(11), by writing the modified t-test in (25) as

$$modt_0 = \frac{\sqrt{n}(\hat{\beta} - \beta_0 - (\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\bar{\pi}corr_0)}{\tilde{\sigma}_u(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1/2}}.$$

So if the true correlation is $corr_0$, then

$$modt_0 \xrightarrow{d} N(0,1),$$

as shown above.

If we had used $corr_1 \neq corr_0$ in our grid search

$$\begin{aligned} modt_1 &= \frac{\sqrt{n}(\hat{\beta} - \beta_0 - (\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\bar{\pi}corr_1)}{\tilde{\sigma}_u(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1/2}} \\ &= \frac{\sqrt{n}(\hat{\beta} - \beta_0 - (\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\bar{\pi}corr_0)}{\tilde{\sigma}_u(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1/2}} \\ &+ \frac{\sqrt{n}(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\bar{\pi}(corr_1 - corr_0)}{\tilde{\sigma}_u(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1/2}} \\ &= modt_0 + \frac{(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1}\bar{\pi}n^{1/2}(corr_1 - corr_0)}{\tilde{\sigma}_u(\hat{\pi}'\hat{Q}_{zz}\hat{\pi})^{-1/2}} \\ &\rightarrow \infty. \end{aligned} \tag{26}$$

Note that the second term on the right hand side of the above equation diverges to infinity since $\sqrt{n}(corr_1 - corr_0) \rightarrow \infty$. **Q.E.D.**

7 Stata Code

In this code, x is the endogenous variable of interest, y is the dependent variable, and z is the instrument. For simplicity, let ‘covars’ be a local macro for all exogenous covariates and ‘corr’ be the correlation value that you are testing.

```
% ivreg y (x = z) ‘covars’
% scalar b2sls = _b[x]
% scalar N = e(N)

% reg x ‘covars
% predict xresid, resid

% reg y ‘covars
% predict yresid, resid

% reg z ‘covars
% predict zresid, resid

% egen ssyresid = sum(yresid^2)
% scalar sigmatilda = sqrt(1/(N-1) * ssyresid)

% reg xresid yresd
```

```
% scalar phihat = _b[zresid]

% egen ssz = sum(zresid^2)
% scalar modt = N^(1/2)*b2sls/(sigmatilda*(abs(phihat)^(-1)*(ssz/N)&(-1/2)))

% scalar modttest = modt - sqrt(N)*`corr'*sign(phihat)
```

Alternatively, you can do a grid search:

```
% forvalues i = -1(.01)1{
% scalar modttest`i' = modt - sqrt(N)*`i'*sign(phihat)
% }
```