The Hausman Test and Weak Instruments*

Jinyong Hahn — UCLA
John Ham — Maryland
Hyungsik Roger Moon — USC

August 3, 2009

Abstract

We consider the following problem. There is a structural equation of interest that contains an explanatory variable that theory predicts is endogenous. There are one or more instrumental variables that credibly are exogenous with regard to this structural equation, but which have limited explanatory power for the endogenous variable. Further, there is one or more potentially 'strong' instrument, which has much more explanatory power but which may not be exogenous. Hausman (1978) provided a test for the exogeneity of the second instrument when none of the instruments are weak. Here we focus on how the standard Hausman test does in the presence of weak instruments using the Staiger-Stock asymptotics. It is natural to conjecture that the standard version of the Hausman test would be invalid in the weak instruments case, which we confirm. However, we provide a version of the Hausman test that is valid even in the presence of weak IV. We show that the situation we analyze occurs in several important economic examples. Our Monte Carlo experiments show that this procedure works relatively well in finite samples. We should note that our test is not consistent, although we believe that it is impossible to construct a consistent test with weak instruments.

1 Introduction

The weak instruments problem has led to development to two strands of research, each of which is characterized by a different alternative asymptotic approximation. The first of these, which we will call the many-instrument asymptotics, emphasizes the finite sample distortion which can be explained by the approximation where the number of instruments grows to infinity as a function of the sample size. This literature often concludes that the IV estimators are still approximately normal, but that the asymptotic variance estimators need to address the finite sample issue.[1] Because the many-instrument

*We thank Takeshi Amemiya, two referees, and Joris Pinkse for helpful comments and suggestions, and Martin Weidner for proofreading. Hahn and Moon acknowledge the support from National Science Foundation. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[1] See Bekker (1994), Donald and Newey (2001), Hahn and Hausman (2002), among others.

asymptotics still produces a normal approximation for the estimators, the implication for practitioners is more or less a simple message that the standard error calculations need to be refined. On the other hand, there is a concern that the many-instrument asymptotics may not be relevant for situations where the degree of overidentification is mild. When the model is just or only mildly overidentified, and the explanatory power of the instruments is small, the alternative approximation due to Staiger and Stock (1997) is intuitively appealing.[2] This approximation is characterized by alternative asymptotics where the first stage coefficient shrinks to zero as a function of the square root of the sample size. We will call this approximation the weak-instrument asymptotics.

Under Staiger and Stock's asymptotic approximation, many usual statistics have nonstandard asymptotic distribution. For example, it is well-known that IV estimators, and t-statistics have nonstandard distributions. Staiger and Stock (1997) also considered test of overidentification under their asymptotics, and established that standard tests of overidentification do not have chi-square (χ^2 hereafter) distributions either. On the other hand, they showed that, in the context of comparing the weak IV against OLS, a version of Hausman test statistic as usually practiced has a correct asymptotic size, although they did observe that the test is not consistent. This finding is important because there is no standard test in the literature to determine whether conventional asymptotics or Staiger and Stock's alternative asymptotics is more appropriate for a given finite sample. The version of Hausman test has the identical asymptotic distribution under both asymptotics, and is thus a exception to the rule of thumb that test statistics tend to have nonstandard distribution under weak instrument asymptotics. It is a useful exception in that practitioners do not need to worry about the weakness of IV and its potentially complicated consequence.

In this paper, we extend Staiger and Stock's (1997) analysis, and document further exceptional cases. We consider the standard Hausman test that examines the difference of two IV estimators based on two different sets of instruments, and show that it possesses a certain robustness property in that its asymptotic distribution is invariant to whether conventional or weak instrument asymptotics is adopted. We consider Hausman test that compares weak IV against strong IV. It is well-known that the test statistic has a χ^2 distribution under conventional asymptotics. We establish that a version of Hausman test continues to have the χ^2 distribution even under the weak instrument asymptotics. We go further and show that a version of overidentification test, which we interpret to be a natural generalization of Hausman test, has such robustness. Finally, we also provide empirical researchers with a version of the Hausman test that can be used with heteroscedasticity under both conventional and Staiger-Stock asymptotics; although quite straightforward theoretically, neither case is currently available in the literature.

Besides being of theoretical interest, our result has substantial practical implications because empirical researchers often face the following problem.[3] They have a structural equation of interest that

[2] See also Kleibergen (2000), Moreira (2003), and Andrews, Moreira, and Stock (2004).

[3] See Section 3 below.

contains an explanatory variable that theory predicts is endogenous. They want to obtain a confidence interval for the estimated coefficient on the structural parameter, or for a set of coefficients from the structural equation. On the one hand they have one or more instrumental variables that credibly are exogenous with regard to this structural equation, but which have limited explanatory power for the endogenous variable. On the other hand they have one or more 'strong instrument', which has much more explanatory power but which may not be exogenous. Researchers currently can take one of two tacks. First, if the researcher only uses the weak instruments, the standard errors on the structural equation calculated by standard methods may be very large. Moreover, it may be the case that the standard asymptotic distribution for IV estimators is invalid because of the weak instrument problem. Second, in the vast majority of cases, empirical researchers use the strong instrument since it is simple to use and likely to produce statistically significant results. Thus it has obvious appeal to the researcher, but also has the obvious disadvantage that the researcher may obtain inconsistent results if the strong instrument is not a valid instrument. We would propose that researchers take a third approach in their work: use the strong instrument but provide a diagnostic via a Hausman test comparing the results using the strong and weak instrument. However, this approach raises the concern of whether the Hausman test is valid when one of the instruments is weak, which our result naturally addresses.

The outline of the paper is as follows. We outline our model and assumptions in Section 2. In Section 3 we motivate the paper by showing that the situation we analyze arises in several important economic examples: i) estimating models of life cycle labor supply behavior; ii) estimating dynamic models such as a health production function for individuals in a developing country and iii) estimating the return to schooling. In Section 4 we consider the conventional Hausman test under weak IV asymptotics and show that, in general, it will not have the standard χ^2 distribution, but if the model is exactly identified given the weak instruments, one of the standard tests can be used without modification. In section 5 we provide a modification of the Hausman test when the model is overidentified given the weak instruments; this modification has the standard χ^2 distribution. The results of our Monte Carlo experiments are presented in Section 7. They show that there is indeed a problem with the standard tests when the model is overidentified, and that our general procedure works relatively well in this case in finite samples.

2 Model and Assumptions

We consider a simultaneous equation linear regression model

$$y_1 = Y_2\beta + \varepsilon \text{ and}$$
$$Y_2 = Z\Pi + V,$$

where ε and V are mean zero unobserved error matrices, y_1 is an $n-$vector of dependent variables, Y_2 is an $n \times K$ matrix of regressors that are correlated with ε, Z is an $n \times L$ matrix of IV's with $L \geq K$ that are independent of V.[4] The sample size is denoted by n and all the asymptotic results of the paper are based on $n \to \infty$.

We assume that the IV's consist of two components, $Z = [W, S]$, where W is an $n \times L_w$ matrix that contains 'weak' IV's and S is an $n \times L_s$ matrix that contains strong, but potentially invalid, IV's. Further, $\widetilde{S}$ is the "residual" when S is projected on W in the population,

$$S = W\Gamma_w + \widetilde{S}. \tag{1}$$

We also denote y_{1i}, Y_{2i}', w_i', s_i', $\widetilde{s}_i'$, ε_i, and v_i' to be the i^{th} row of $y_1, Y_2, W, S, \widetilde{S}, \varepsilon$, and V, respectively. We assume that $L_w \geq K$.

Throughout this section, we will assume that W is orthogonal to the regression error ε, that is, $E[w_i\varepsilon_i] = 0$. The main object of interest in this paper is test for the validity of the IV's in S. In this case, the hypotheses that we are testing are

$$H_0 : E[s_i\varepsilon_i] = 0 \text{ (or } E[\widetilde{s}_i\varepsilon_i] = 0) \tag{2}$$

$$H_1 : E[s_i\varepsilon_i] \neq 0 \text{ (or } E[\widetilde{s}_i\varepsilon_i] \neq 0). \tag{3}$$

If the exclusion restriction is violated, then it is only through the possible correlation between $\widetilde{s}_i$ and ε_i.

Let $\rho_s = [E(\widetilde{s}_i\widetilde{s}_i')]^{-1} E(\widetilde{s}_i\varepsilon_i)$ denote the coefficient of projection of ε on $\widetilde{S}$. We write that

$$\varepsilon = \widetilde{S}\rho_s + V\rho_v + e, \tag{4}$$

where $\rho_v = [E(v_iv_i')]^{-1} E(v_i\varepsilon_i)$ denotes the coefficient of projection of ε_i on v_i. We will assume that e_i is uncorrelated with $\widetilde{s}_i$ and v_i and has mean zero and variance σ_e^2. Our null and alternative hypotheses can then be rewritten as

$$H_0 : \rho_s = 0$$

$$H_1 : \rho_s \neq 0.$$

The basic idea of the Hausman test statistics for the null hypothesis (2) is based on the difference of the following two estimators[5]:

$$\widehat{\beta}_w = \left(Y_2'P_WY_2\right)^{-1} Y_2'P_Wy_1$$

$$\widehat{\beta}_z = \left(Y_2'P_ZY_2\right)^{-1} Y_2'P_Zy_1.$$

[4] We follow the standard approach, and assume that included exogenous variables are 'partialled out' - see the online appendix for more details.

[5] Given a matrix A, we use notation $P_A = A(A'A)^{-1}A$ and $M_A = I - P_A$ throughout the note.

When conventional asymptotic approximation is valid, then $\widehat{\beta}_z$ is an efficient but non-robust estimator, while $\widehat{\beta}_w$ is a less efficient, but robust, estimator. Then, the conventional Hausman test statistics measures the difference $\widehat{\beta}_w - \widehat{\beta}_z$ using various weight matrices. We first consider three versions of the Hausman test that are used widely in the literature:

$$\mathcal{H}_1 = \left(\widehat{\beta}_w - \widehat{\beta}_z\right)' \left[\widehat{\sigma}^2_{\varepsilon,w}\left(Y_2' P_W Y_2\right)^{-1} - \widehat{\sigma}^2_{\varepsilon,z}\left(Y_2' P_Z Y_2\right)^{-1}\right]^{-1} \left(\widehat{\beta}_w - \widehat{\beta}_z\right),$$
$$\mathcal{H}_2 = \widehat{\sigma}^{-2}_{\varepsilon,w}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)' \left[\left(Y_2' P_W Y_2\right)^{-1} - \left(Y_2' P_Z Y_2\right)^{-1}\right]^{-1} \left(\widehat{\beta}_w - \widehat{\beta}_z\right),$$
$$\mathcal{H}_3 = \widehat{\sigma}^{-2}_{\varepsilon,z}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)' \left[\left(Y_2' P_W Y_2\right)^{-1} - \left(Y_2' P_Z Y_2\right)^{-1}\right]^{-1} \left(\widehat{\beta}_w - \widehat{\beta}_z\right),$$

where

$$\widehat{\sigma}^2_{\varepsilon,w} = \frac{1}{n}\left(y_1 - Y_2\widehat{\beta}_w\right)'\left(y_1 - Y_2\widehat{\beta}_w\right) \tag{5}$$

and

$$\widehat{\sigma}^2_{\varepsilon,z} = \frac{1}{n}\left(y_1 - Y_2\widehat{\beta}_z\right)'\left(y_1 - Y_2\widehat{\beta}_z\right). \tag{6}$$

Under conventional asymptotics, these test statistics all converge to χ^2_K, a (central) chi-square distribution with d.f. K, under the null. Therefore, the conventional asymptotics suggests that we compare these test statistics with the critical value from χ^2_K.

Our contribution is to consider the properties of the test statistics under the assumption that W is 'weak' and S is 'strong' but potentially invalid under the asymptotics developed by Staiger and Stock (1997).

3 Economic Examples

The problem we analyze arises in many empirical studies; here we show this for three important cases.

3.1 Life Cycle Labor Supply Models

Researchers often consider the following model to describe the (annual) intertemporal labor supply function for prime-aged males[6]

$$\Delta \ln(h_{it}) = \delta \Delta \ln(w_{it}) + \alpha + \beta \Delta X_{it} + \Delta e_{it} + \delta \eta_{it}, \tag{7}$$

where Δ denotes the first difference. In (7) h_{it} are hours of work in year t for individual i, w_{it} is his real hourly wage rate in that year, X_{it} are time changing demographic variables, e_{it} is an idiosyncratic error term, and η_{it} is a 'rational expectations' error term which is orthogonal to all variables known in period $t-1$; thus $\Delta \ln(w_{it})$ is correlated with η_{it}.We also expect that $\Delta \ln(w_{it})$ is correlated with

[6] See MaCurdy 1981, Altoni 1986, Ham 1986, Ham and Reilly 2002. Corner solutions of zero hours are not important for this group and thus a regression framework is appropriate.

Δe_{it} since variable w_{it} is formed by dividing annual earnings by h_{it}, and the latter is thought to contain substantial measurement error. MaCurdy (1981) used polynomials in age as IV for $\Delta \ln(w_{it})$, but Altonji (1986) argued that MaCurdy's instruments were weak in the sense of not being jointly significant in the first stage equation. Instead Altonji considered a direct measure wm_{it} of the wage which is obtained from a question put to individuals in the sample 'what is your hourly wage rate?' He assumes that the measurement error in w_{it} and wm_{it} are independent, and thus only considering the error term Δe_{it} , the variable $\Delta \ln(wm_{it})$ is a valid IV for $\Delta \ln(w_{it})$. However, as Altonji noted, this potential instrument will not be independent of η_{it} unless wm_{it} is known in period $t-1$. He next considered $\Delta \ln(wm_{it-1})$ as an IV for $\Delta \ln(w_{it})$, since it will be orthogonal to η_{it}, but finds that the correlation between $\Delta \ln(w_{it})$ and $\Delta \ln(wm_{it-1})$ is too weak to be empirically useful. Instead he assumes that the wage is known one period in advance so that $\Delta \ln(wm_{it})$ is indeed an appropriate IV for $\Delta \ln(w_{it})$. Thus our procedure could be used to offer readers a diagnostic test whether $\Delta \ln(wm_{it})$ is indeed a valid IV, using either (or both) MaCurdy's polynomial in age or $\Delta \ln(wm_{it-1})$ as the weak instrument.

3.2 Dynamic Models

Researchers often consider dynamic panel data regressions of the form

$$y_{it} = \gamma y_{it-1} + \beta X_{it} + u_{it}. \tag{8}$$

In (8) y_{it} is a scaler dependent variable for individual i in year t, X_{it} is a vector of exogenous explanatory variables, and u_{it} is an error term. Since it is unreasonable to assume that the error term u_{it} is independent over time for the same person, y_{it-1} must be treated as endogenous. Natural instruments are lagged values of X_{it}, but researchers often find that these lagged values of X_{it} do a poor job of explaining y_{it-1}. Instead they often assume a $MA(k)$ structure for u_{it}, which implies that y_{it-k-1} is a a valid IV for y_{it-1}. However, the choice of k is usually arbitrary, since economic theory does not provide any guidance on this issue. Again our test can be used here, where y_{it-k-1} is the strong instrument and the lags of X_{it} are the weak instruments. An example of such an equation is given in Strauss and Thomas (1995), where (8) is a human capital production function for individuals in a developing country, and the X_{it} represents variables such as distance to the village health clinic. They used the strong instruments (lagged y_{it}), but could use our procedure below to obtain a diagnostic for their approach.

3.3 Estimating the Return to Schooling

Consider the wage equation

$$\ln(w_i) = \alpha S_i + \gamma A_i + \beta X_i + e_i. \tag{9}$$

In (9) the variables w_i, S_i and A_i represent the hourly wage, years of schooling, and ability (as measured by a test such as the AFQT in the case of the NLS data), and X_i represents variables such as race, experience, and experienced squared. The problem here is that even conditional on ability A_i, S_i and e_i may be correlated. For example, an increase in ambition may increase both S_i and e_i, leading to a positive correlation between these variables. One possible instrument for S_i is the father's education FE_i, which Willis and Rosen (1979) use to identify a more complicated version of (9). They argue that children from wealthier families have a lower discount rate than poorer children, since wealthy parents are more likely to help finance their children's education. In practice FE_i will be an important determinant of S_i conditional on A_i and X_i. However, it may be an invalid IV since it can also reflect the father's ambition, which he may pass on to his children; if so, FE_i will be correlated with e_i and thus will be an invalid IV. An alternative IV is the father's age FA_i in the year that the individual turned eighteen, since this will also affect the family's ability to help its children pay for college, but is unlikely to be correlated with e_i. Unfortunately FA_i may have little predictive power for S_i conditional on A_i and X_i.

4 Hausman Tests under Weak IV Asymptotics

We now investigate the theoretical properties of the conventional Hausman test for the hypothesis of (2) under the assumption that W is weak. More specifically, we assume that the coefficient of the population projection of Y_2 on W shrinks to zero at the rate $\frac{1}{\sqrt{n}}$, while the coefficient of population projection of Y_2 on S does not. For this purpose, we adopt the following parameterization:

$$Y_2 = W\frac{C}{\sqrt{n}} + \widetilde{S}\Pi_s + V. \tag{10}$$

We assume that $L_s \geq K$. Suppose that we use the convention of using $\mathcal{H}_1$, $\mathcal{H}_2$, or $\mathcal{H}_3$ but adopt Staiger and Stock's (1997) alternative asymptotic approximation. Because their asymptotics implies that the asymptotic distribution of $\widehat{\beta}_w$ is not normal, it is natural to conjecture that the asymptotic size of the conventional procedure would be distorted.[7] Not surprisingly, the $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$ are usually not distributed as χ^2_K under the weak-instrument asymptotic approximation. See Theorem 4 in the Appendix C.1.

Our first main contribution is to recognize that there is an important exception. We show that the $\mathcal{H}_3$ is asymptotically χ^2_K under the null despite the presence of weak instruments if the model is exactly identified with only the weak IV:

Theorem 1 *Assume Conditions 1 and 2 in the Appendix A.*[8] *Suppose that $L_w = K$ and $Y_2'W$ has full rank K. Then, (a) under the null hypothesis (2),*

$$\mathcal{H}_3 \Rightarrow \mathcal{Z}'\mathcal{Z} \equiv \chi^2_K,$$

[7] Proposition 4 in the Appendix confirms this conjecture.

[8] We impose standard regularity conditions, which are discussed in the Appendix.

(b) under the alternative hypothesis (3),

$$\mathcal{H}_3 \Rightarrow (\mathcal{Z}+\kappa)'(\mathcal{Z}+\kappa) \equiv \chi^2_K(\kappa),$$

where $\mathcal{Z} \sim N(0, I_K)$ *and the noncentrality parameter* κ *is defined in (15) in the Appendix.*

Proof. In Appendix. ■

Theorem 1 indicates that, as long as the weak instrument W exactly identifies β ($L_w = K$), the standard practice of using $\mathcal{H}_3$ along with a critical value from χ^2_K is asymptotically valid even under the weak-instrument asymptotics. The weak-instrument asymptotic distribution under the null is identical to the standard asymptotic distribution. The weak-instrument asymptotic distribution under the alternative $\chi^2_K(\kappa)$, a noncentral chi-square distribution with d.f. K, dominates the asymptotic distribution under the null χ^2_K, and therefore, the test is unbiased under the weak-instrument asymptotics.

5 Generalized Hausman Test

The results in Section 4 implies that a version of Hausman test, i.e., $\mathcal{H}_3$ (but not $\mathcal{H}_1$ and $\mathcal{H}_2$), combined with a critical value from χ^2_K, is asymptotically valid even under the weak-instrument asymptotics as long as β is exactly identified by weak instruments ($L_w = K$). On the other hand, Theorem 4 in the Appendix C.1 shows that neither $\mathcal{H}_1$, $\mathcal{H}_2$, nor $\mathcal{H}_3$, along with a critical value from χ^2_K, is valid under the weak-instrument asymptotics if the weak IV's overidentify β, that is, when $L_w > K$. One might argue that the *overidentified* case is not of practical concern because a practitioner can always choose a subset of weak instruments from W that exactly identify β and thus $\mathcal{H}_3$. Although one can resolve the situation in this fashion, it is not clear which K weak instruments should be chosen out of L_w. We show in this section that there is a version of the specification test which can be used with a critical value from χ^2 even under the weak-instrument asymptotics and when the model is overidentified with the weak IV's.

Suppose that the model is in fact overidentified with the weak instruments ($L_w \geq K$). Under the null that the strong IV's are valid, we have the moment condition

$$E\left[w_i\left(y_{1i} - Y_{2i}' \operatorname{plim} \widehat{\beta}_z\right)\right] = 0.$$

However, under the alternative that the strong IV's are not valid, we have

$$E\left[w_i\left(y_{1i} - Y_{2i}' \operatorname{plim} \widehat{\beta}_z\right)\right] \neq 0,$$

since the probability limit of $\widehat{\beta}_z$ will be different from β under the alternative. From these observations, we might consider a test statistic based on

$$\frac{1}{\sqrt{n}} W'\left(y_1 - Y_2 \widehat{\beta}_z\right).$$

With some algebra, it can be shown that

Lemma 1 *Assume Conditions 1 and 2 in the Appendix. Under the null (2) and conventional asymptotics,*

$$\frac{1}{\sqrt{n}}W'\left(y_1 - Y_2\widehat{\beta}_z\right) \Rightarrow N\left(0, \sigma_\varepsilon^2 \Psi\right),$$

where Ψ *is the probability limit of* $\frac{1}{n}\widehat{\Psi}$ *and*

$$\widehat{\Psi} = W'W - \left(W'Y_2\right)\left(Y_2'P_ZY_2\right)^{-1}\left(Y_2'W\right).$$

Proof. In the Online Appendix. ■

Therefore, the test statistic is equal to

$$\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right) = \frac{1}{\widehat{\sigma}_\varepsilon^2}\left(y_1 - Y_2\widehat{\beta}_z\right)' W\widehat{\Psi}^{-1}W'\left(y_1 - Y_2\widehat{\beta}_z\right),$$

where $\widehat{\Psi} = W'W - (W'Y_2)(Y_2'P_ZY_2)^{-1}(Y_2'W)$ and $\widehat{\sigma}_\varepsilon^2$ denotes *some* consistent estimator for σ_ε^2. In light of Lemma 1, it is straightforward to conclude that the (conventional) asymptotic distribution of $\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right)$ is $\chi^2_{L_w}$. In other words, researchers can use $\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right)$ to obtain a χ^2-test with standard critical values even when the model is overidentified with the weak IV's.

Proposition 1 gives an interpretation of the new statistic $\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right)$.

Proposition 1 *When* $L_w = K$,

$$\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right) = \frac{1}{\widehat{\sigma}_\varepsilon^2}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)'\left[\left(Y_2'P_WY_2\right)^{-1} - \left(Y_2'P_ZY_2\right)^{-1}\right]^{-1}\left(\widehat{\beta}_w - \widehat{\beta}_z\right).$$

Proof. In the Online Appendix. ■

From Proposition 1, we can conclude that $\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right)$ can be understood as *a version* of the Hausman test in a special case where $L_w = K$. Depending on the estimator $\widehat{\sigma}_\varepsilon^2$ used, the statistic $\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right)$ can be understood to be an extension of $\mathcal{H}_2$ or $\mathcal{H}_3$.[9]

Recall $\widehat{\sigma}_{\varepsilon,z}^2$ in (6). It turns out that $\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right)$, which is comparable to $\mathcal{H}_3$, has desirable asymptotic properties:

Theorem 2 *Assume Conditions 1 and 2 in the Appendix. Assume that* $L_w \geq K$. *(a) Under the null hypothesis,*

$$\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right) \Rightarrow \chi^2_{L_w}.$$

(b) Under the alternative hypothesis,

$$\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right) \Rightarrow (\kappa + \mathcal{Z})'(\kappa + \mathcal{Z}),$$

where $\mathcal{Z} \sim N\left(0, I_{L_w}\right)$ *and* κ *is the same noncentrality parameter in Theorem 1.*

[9] When $L_w = K$, we have $\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,w}^2\right) = \mathcal{H}_2$ and $\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right) = \mathcal{H}_3$.

Proof. In Appendix. ■

Theorem 2 indicates that using $\mathcal{H}\left(\widehat{\sigma}^2_{\varepsilon,z}\right)$ along with a critical value from χ^2 is asymptotically valid even under the weak-instrument asymptotics. The weak-instrument asymptotic distribution under the null is identical to the standard asymptotic distribution. The weak-instrument asymptotic distribution under the alternative dominates the asymptotic distribution under the null, and therefore, the test is unbiased under the weak-instrument asymptotics. On the other hand, the test statistic does not diverge to infinity under the alternative, as is the case with standard asymptotics, and therefore the test is not consistent under weak-instrument asymptotics regardless of whether the model is exactly identified or over-identified by the weak IV's.

6 Discussion

We first consider two deviations from our assumptions. First, we consider the case where the strong IV is valid under both null and alternative hypotheses, while the weak IV is valid only under the null.[10] Second, we examine the consequences of heteroscedasticity. After this, we consider the issue of improving power.

6.1 When the Weak IV are Valid Only Under the Alternative Hypothesis

We may want to consider an alternative scenario, where the strong IV are valid both under the null and the alternative, and the weak IV are valid only under the null. Although this scenario is unlikely to be common in practice, we address this situation for its theoretical interest. In this case, the model could be modified as $y_1 = Y_2\beta + \varepsilon$ and $Y_2 = \tilde{W}\frac{C}{\sqrt{n}} + S\Pi_s + V$, where the alternative hypothesis is now written $\varepsilon = \widetilde{W}\rho_w + V\rho_v + e$. Here $\widetilde{W}$ is the population projection "residual" of W on S: $W = S\Gamma_s + \widetilde{W}$.

Here we consider the properties of (generalized) Hausman test statistic $\mathcal{H}\left(\widehat{\sigma}^2_{\varepsilon,z}\right)$.[11] It can be shown that the $\mathcal{H}\left(\widehat{\sigma}^2_{\varepsilon,z}\right)$ is distributed as $\chi^2_{L_s}$ under the null, but diverges to ∞ under the alternative.[12] In other words, the $\mathcal{H}\left(\widehat{\sigma}^2_{\varepsilon,z}\right)$ has the identical properties as under the conventional asymptotics!

6.2 Heteroscedasticity

It is well-known that 2SLS is not efficient under heteroscedasticity, and the usual form of the Hausman test would no longer be valid even under the null. This implies that, even with conventional asymptotics, the Hausman test has to be modified. We note that there does not exist a standard modification of Hausman test to accommodate heteroscedasticity. We consider one possible modification here.

[10] We thank an anonymous referee who suggested this agenda.

[11] Given that the role of w and s is switched, we note that our test would be based on $\frac{1}{\sqrt{n}}S'\left(y_1 - Y_2\widehat{\beta}_z\right)$.

[12] Proof available in Online Appendix C.1.

To simplify our notation, assume that β is a scalar. Since the size of the Hausman test is valid only when $L_w = K$ even under homoscedasticity, we assume that the weak IV is a scalar also (that is, $L_w = 1$). Given that the Hausman test has an interpretation of comparison between $\widehat{\beta}_w$ and $\widehat{\beta}_z$, a natural modification of the Hausman test statistic would take the form

$$\frac{n\left(\widehat{\beta}_w - \widehat{\beta}_z\right)^2}{\widehat{\text{Var}}\left(\sqrt{n}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)\right)} \tag{11}$$

where $\widehat{\text{Var}}\left(\sqrt{n}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)\right)$ denotes a consistent estimator of the asymptotic variance of $\sqrt{n}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)$ under *conventional* asymptotics. To see this in more detail, by definition, under the null we have

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)\\
&= \sqrt{n}\left(\widehat{\beta}_w - \beta\right) - \sqrt{n}\left(\widehat{\beta}_z - \beta\right)\\
&= \left(\frac{W'Y_2}{n}\right)^{-1}\left(\frac{W'\varepsilon}{\sqrt{n}}\right) - \left(\frac{Y_2'P_Z Y_2}{n}\right)^{-1}\left(\frac{Y_2'P_Z\varepsilon}{\sqrt{n}}\right),
\end{aligned}$$

and its asymptotic variance under *conventional* asymptotics is

$$\frac{E\left[w_i^2\varepsilon_i^2\right]}{\left(E\left[w_iY_{2i}\right]\right)^2} - 2\frac{E\left(Y_{2i}z_i'\right)E\left(z_iw_i\varepsilon_i^2\right)}{E\left[w_iY_{2i}\right]\left[E\left(Y_{2i}z_i'\right)\left(E\left(z_iz_i'\right)\right)^{-1}E\left(z_iY_{2i}\right)\right]} + \frac{E\left(Y_{2i}z_i'\right)E\left(z_iz_i'\varepsilon_i^2\right)E\left(z_iY_{2i}\right)}{\left[E\left(Y_{2i}z_i'\right)\left(E\left(z_iz_i'\right)\right)^{-1}E\left(z_iY_{2i}\right)\right]^2}.$$

One can use the idea behind White's heteroscedasticity corrected standard error, using the standard IV estimator's residuals $\hat{\varepsilon}_z = y_1 - Y_2\widehat{\beta}_z$. A natural choice for a consistent estimator of $\text{Var}\left(\sqrt{n}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)\right)$ is

$$\begin{aligned}
&\widehat{\text{Var}}\left(\sqrt{n}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)\right)\\
&= \frac{\left(\frac{1}{n}\sum_{i=1}^n w_i^2\widehat{\varepsilon}_i^2\right)}{\left(\frac{1}{n}\sum_{i=1}^n w_iY_{2i}\right)^2} - 2\frac{\left(\frac{1}{n}\sum_{i=1}^n Y_{2i}z_i'\right)\left(\frac{1}{n}\sum_{i=1}^n z_iw_i\widehat{\varepsilon}_i^2\right)}{\left(\frac{1}{n}\sum_{i=1}^n w_iY_{2i}\right)\left[\left(\frac{1}{n}\sum_{i=1}^n Y_{2i}z_i'\right)\left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_iY_{2i}\right)\right]}\\
&+ \frac{\left(\frac{1}{n}\sum_{i=1}^n Y_{2i}z_i'\right)\left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\hat{\varepsilon}_i^2\right)\left(\frac{1}{n}\sum_{i=1}^n z_iY_{2i}\right)}{\left[\left(\frac{1}{n}\sum_{i=1}^n Y_{2i}z_i'\right)\left(\frac{1}{n}\sum_{i=1}^n z_iz_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_iY_{2i}\right)\right]^2}.
\end{aligned}$$

In the Online appendix we show that under the Staiger-Stock asymptotics

$$\mathcal{H}_{hetero}\left(\hat{\varepsilon}_z\right) = \frac{n\left(\widehat{\beta}_w - \widehat{\beta}_z\right)^2}{\widehat{\text{Var}}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)} \Rightarrow \left(\mathcal{Z} + \kappa_{hetero}\right)^2, \tag{12}$$

where $\mathcal{Z} \sim N(0,1)$ and

$$\kappa_{hetero} = \frac{-\Sigma_{ww}C\tau}{\lim_n \frac{1}{n}\sum_{i=1}^n E\left[w_i^2\left(\tilde{s}_i\left(\rho_s - \Pi_s\tau\right) + v_i\left(\rho_v - \tau\right) + e_i\right)^2\right]}.$$

Since $\kappa_{hetero} = 0$ $(\tau = 0)$ under the null,

$$\mathcal{H}_{hetero}\left(\hat{\varepsilon}_z\right) \Rightarrow \chi_1^2.$$

Therefore, the test is valid under the null. On the other hand, under the alternative,

$$\mathcal{H}_{hetero}\left(\hat{\varepsilon}_z\right) \Rightarrow \chi_1^2\left(\kappa_{hetero}\right),$$

and thus the test $\mathcal{H}_{hetero}\left(\hat{\varepsilon}_z\right)$ becomes asymptotically unbiased.

6.3 Improving Power of $\mathcal{H}_3$

In Section 4, we noted that the $\mathcal{H}_3$ is asymptotically unbiased. On the other hand, the test statistic does not diverge to infinity under the alternative, as is the case with standard asymptotics, and therefore the test is not consistent under weak-instrument asymptotics. Given that the consistency of a test is usually understood to be a necessity, a researcher may conclude that the test using $\mathcal{H}_3$ is deficient. We should note, though, that a consistent test is probably impossible to construct given the nature of weak instrument. Many other tests are inconsistent with weak IV's, and a lack of consistency is not limited to the weak IV literature. (For example, a recent test by Andrews (2003) exhibits similar properties.)

The lack of consistency suggests that it would be a useful endeavor to try to improve the power of the test while maintaining its good size properties. For this purpose, we propose the following version of the Hausman test:

$$\mathcal{H}_4 = \widetilde{\sigma}_{\varepsilon,z}^{-2}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)'\left[\left(Y_2'P_WY_2\right)^{-1} - \left(Y_2'P_ZY_2\right)^{-1}\right]^{-1}\left(\widehat{\beta}_w - \widehat{\beta}_z\right),$$

where $\widetilde{\sigma}_{\varepsilon,z}^2$ is obtained by the following algorithm:

1. Using the IV estimator $\widehat{\beta}_z$, we get the IV residual $\widehat{\varepsilon}_z = y_1 - Y_2\widehat{\beta}_z$.

2. Regress the IV residual $\widehat{\varepsilon}_z$ on $Z = [S, W]$ to get residual $M_Z\widehat{\varepsilon}_z$ and define

$$\widetilde{\sigma}_{\varepsilon,z}^2 = \frac{1}{n}\widehat{\varepsilon}_z'M_Z\widehat{\varepsilon}_z = \widehat{\sigma}_{\varepsilon,z}^2 - \frac{1}{n}\widehat{\varepsilon}_z'P_Z\widehat{\varepsilon}_z. \tag{13}$$

From (13), one can see that the proposed estimator $\widetilde{\sigma}_{\varepsilon,z}^2$ modifies $\widehat{\sigma}_{\varepsilon,z}^2$ by subtracting $\frac{1}{n}\widehat{\varepsilon}_z'P_Z\widehat{\varepsilon}_z$. Although it is generally preferable to use the modified estimator $\widetilde{\sigma}_{\varepsilon,z}^2$, there are two special cases where such modification is unnecessary and $\widehat{\sigma}_{\varepsilon,z}^2$ is enough. The first case is when $\Sigma_{\widetilde{s}\widetilde{s}}^{1/2}\rho_s$ belongs to the space spanned by the columns of $\Sigma_{\widetilde{s}\widetilde{s}}^{1/2}\Pi_s$ (or Π_s). We then have $\Sigma_{\widetilde{s}\widetilde{s}}^{1/2}\rho_s = \Sigma_{\widetilde{s}\widetilde{s}}^{1/2}\Pi_s\left(\Pi_s'\Sigma_{\widetilde{s}\widetilde{s}}\Pi_s\right)^{-1}\Pi_s'\Sigma_{\widetilde{s}\widetilde{s}}\rho_s$, so that $\rho_s = \Pi_s\tau$. This coincidence depends on the alternative, which is not known to the practitioner, so it probably has little practical importance. The second case is of more practical significance. Suppose that the model is exactly identified by the strong instrument s_i, that is, $L_s = K$, and Π_s is invertible.

We then have $\rho_s = \Pi_s \tau$ and $\sigma_*^2 = \sigma_{**}^2$, and there is no need for the second step modification above and we can use $\widehat{\sigma}_{\varepsilon,z}^2$ in place of $\widehat{\sigma}_{\varepsilon}^2$. We believe that the second case is empirically more relevant than the first case, because in many applications the endogenous regressor Y_{2i} and the strong IV s_i are scalars.

Theorem 3 below shows that $\mathcal{H}_4$ thus defined has the usual χ_K^2 under the null, and its asymptotic distribution under the null stochastically dominates χ_K^2.

Theorem 3 *Assume Conditions 1 and 2 in the Appendix. Suppose that* $L_w = K$. *Then* $\mathcal{H}_4 \Rightarrow \chi_K^2$ *under the null. Under the alternative hypothesis (3),* $\mathcal{H}_4 \Rightarrow \frac{\sigma_{**}^2}{\sigma_*^2} (\mathcal{Z} + \kappa)' (\mathcal{Z} + \kappa)$, *where* $\mathcal{Z} \sim N(0, I_K)$, κ *is the same noncentrality parameter in Theorem 1, (b)* $\sigma_{**}^2 = p \lim \widehat{\sigma}_{\varepsilon,z}^2$ *under the alternative and* $\sigma_*^2 = p \lim \widetilde{\sigma}_{\varepsilon,z}^2$ *under the alternative.*

Proof. Omitted because Theorem 3 is an immediate consequence of Theorem 2 in Section 5. ■

In the Appendix we show that under the alternative, $\widetilde{\sigma}_{\varepsilon,z}^2 \to_p \sigma_*^2 = (\rho_v - \tau)' \Sigma_{vv} (\rho_v - \tau) + \sigma_e^2$ and $\widehat{\sigma}_{\varepsilon,z}^2 \to_p \sigma_{**}^2 = (\rho_s - \Pi_s \tau)' \Sigma_{\widetilde{s}\widetilde{s}} (\rho_s - \Pi_s \tau) + \sigma_*^2$, where $\tau = \text{plim} \left(\widehat{\beta}_z - \beta \right) = \left(\Pi_s' \Sigma_{\widetilde{s}\widetilde{s}} \Pi_s \right)^{-1} \Pi_s' \Sigma_{\widetilde{s}\widetilde{s}} \rho_s$. Here it is obvious to see $\sigma_*^2 / \sigma_{**}^2 \leq 1$, which implies that the asymptotic power of the modified test $\mathcal{H}_4$ is larger than $\mathcal{H}_3$.

In Section 5, we proposed a generalized Hausman test statistic $\mathcal{H}(\cdot)$ for the case where $L_w > K$.. A natural question is whether the power of $\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right)$ is dominated by $\mathcal{H}\left(\widetilde{\sigma}_{\varepsilon,z}^2\right)$. Using Theorem 2, it is easy to see that $\mathcal{H}\left(\widetilde{\sigma}_{\varepsilon,z}^2\right)$ is asymptotically unbiased and its power dominates that of $\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right)$. As such, $\mathcal{H}\left(\widetilde{\sigma}_{\varepsilon,z}^2\right)$ is a desirable test.

We note that, if $L_w = K$, $\mathcal{H}\left(\widetilde{\sigma}_{\varepsilon,z}^2\right)$ simplifies to

$$\mathcal{H}\left(\widetilde{\sigma}_{\varepsilon,z}^2\right) = \widetilde{\sigma}_{\varepsilon,z}^{-2} \left(\widehat{\beta}_w - \widehat{\beta}_z \right)' \left[\left(Y_2' P_W Y_2 \right)^{-1} - \left(Y_2' P_Z Y_2 \right)^{-1} \right]^{-1} \left(\widehat{\beta}_w - \widehat{\beta}_z \right) = \mathcal{H}_4 \tag{14}$$

by Proposition 1. Based on this equality, we will, without too much loss of generality, define $\mathcal{H}_4 = \mathcal{H}\left(\widetilde{\sigma}_{\varepsilon,z}^2\right)$ even for the overidentified case. In the Online Appendix, we show how to construct $\mathcal{H}_4$ in STATA.[13]

7 Monte Carlo Simulations

The data generating process used in the Monte Carlo simulations is

$$\begin{aligned} y_{1i} &= Y_{2i} \beta + s_i' \rho_s + \varepsilon_i \\ Y_{2i} &= w_i' \Pi_w + s_i' \Pi_s + v_i, \qquad i = 1, \ldots, n \end{aligned}$$

where $(w_i', s_i') \overset{\text{iid}}{\sim} N(0, I)$, $(\varepsilon_i, v_i) \overset{\text{iid}}{\sim} N\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$, and y_{1i} and Y_{2i} are scalars. Further, $\beta = 1$ and Π_w and Π_s are proportional to vectors consisting of ones. They are related to the (partial) first stage

[13] From this appendix one sees that a test based on $\mathcal{H}_4$ is quite easy to construct in STATA or similar programs.

R^2 by

$$R_w^2 = \frac{\Pi_w' \Pi_w}{\Pi_w' \Pi_w + 1}, \qquad R_s^2 = \frac{\Pi_s' \Pi_s}{\Pi_s' \Pi_s + 1}.$$

We fixed $R_s^2 = 0.2$ throughout the simulation, and we considered $R_w^2 = 0.01, 0.02$. We set $\rho_s = 0$ under the null, and $\rho_s = (\gamma_s, 0, \ldots, 0)$ under the alternative. We consider $n = 100, 200, 500$, $\rho = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, $R_w^2 = 0.01$, 0.02 0.03, 0.05, 0.1, 0.2, and $\gamma_s = 1$. The dimensions of w_i and s_i are $L_w = 1, 5$ and $L_s = 1, 2, 5$ respectively. The nominal size of the test is 5%. (Additional cases are considered in the Online Appendix.) All the simulation results are based on 5000 runs.

Table A looks at the size of the test for $\mathcal{H}_1-$ $\mathcal{H}_4$ when there is one weak IV and $\rho = \frac{1}{4}$ for different sample sizes and numbers of strong instruments. Thus we first consider the case where the model is exactly identified using the weak instruments and the degree of endogeneity is relatively small. The first section of the table considers this specification for the three different sample sizes and the six values of R_w^2. Ideally each entry should be 0.05, so that we see that in each the size with low R_w^2 is much too small for $\mathcal{H}_1$ and $\mathcal{H}_2$ but is dead on for $\mathcal{H}_3$ and $\mathcal{H}_4$. As R_w^2 and the sample size increase, the size distortion of $\mathcal{H}_1$ and $\mathcal{H}_2$ deceases in each case. The bottom two sections of Table A consider the case of two and five strong instruments respectively. Now the size of $\mathcal{H}_3$ and $\mathcal{H}_4$ are still equal to 0.05 or 0.06 in the rest of the cases, while $\mathcal{H}_1$ and $\mathcal{H}_2$ continue to be under-sized.

Table B looks at the size of the test for $\mathcal{H}_1-$ $\mathcal{H}_4$ when there are five weak IV and $\rho = \frac{3}{4}$; i.e. a case where the model is overidentified under the weak instruments and the degree of endogeneity is considerably higher. Now the size of each test is biased upwards - this is especially true for $\mathcal{H}_1$ and $\mathcal{H}_2$ with low R_w^2. However, it is interesting to note that $\mathcal{H}_4$ substantially outperforms $\mathcal{H}_3$ for most of the cases, which is intuitively plausible since $\mathcal{H}_4$ was developed for the case where the model is potentially overidentified using the weak instruments. Comparing the results in Tables A and B does raise an interesting dilemma. On the one hand, researchers can improve the size of the test by using only one of the weak IV's when the model is overidentified under the weak IV. On the other hand, since different researchers are likely to make different choices in terms of which weak IV to use, they will obtain different test results for identical models. Further, there is also the potential problem of researchers running all five regressions when there are five weak instruments and one endogenous variable, and choosing the results they like the best.

In Table C we consider the power properties of $\mathcal{H}_3$ and $\mathcal{H}_4$ when there is one weak IV, five strong IV and $\rho_s = (1, 0, \ldots, 0)$; i.e. only one of the strong instruments is invalid. Note that this is a conservative example in that it will be harder to reject the null when it is false than if all the strong IV were invalid. Given that we have weak instruments, it is unrealistic to expect the entries in Table C to be close to one. Not surprisingly, the power of each test rises with the sample size and the explanatory power of the weak instruments. It is also interesting to note that when R_w^2 is low, the power of $\mathcal{H}_4$ is often more than double that of $\mathcal{H}_3$ when $n = 100$, a little less than double that of $\mathcal{H}_3$ when $n = 200$, and about 50% greater than that of $\mathcal{H}_3$ when $n = 500$. Thus in terms of power with low R_w^2, $\mathcal{H}_4$ substantially out performs $\mathcal{H}_3$ for all sample sizes in our example. This is to be expected as the model is overidentified under the *strong* IV, and $\mathcal{H}_4$ was developed with power considerations in mind.[14] (Recall that there is no need to use the modified version of $\widetilde{\sigma}^2_{\varepsilon,z}$ when the model is exactly identified under the strong instruments.) When R^2_w is high, the power gain decreases. However, in this case, the power itself is much higher than that of the case with low R^2_w.

8 Conclusion

Hausman (1978) provides a test for whether an instrument(s) is valid given that the model is identified by other instruments which can be treated as exogenous. However, as we show in a series of examples, researchers often face the problem that the most acceptable instruments are also quite weak. Using Staiger-Stock asymptotics, we show that the standard Hausman test for this case may have size distortion under the null in the presence of weak instruments unless the model is exactly identified using the weak instruments. We then provide a form of the Hausman test that eliminates the problem for the overidentified case. Finally, we show in an Online Appendix that this test is easy for empirical researchers to implement using a program like STATA. Our Monte Carlo results suggest that there is indeed a problem with the standard tests, and that our general procedure works relatively well in finite samples.

[14] The corresponding power statistics for $\mathcal{H}_4$ when the model is overidentified under the weak instruments are somewhat higher than those in Table C.

Appendix

A Regularity Conditions

Condition 1 *We assume the following. (i)* $\frac{1}{n}Z'Z \to_p \Sigma_{zz} > 0$; $\frac{1}{n}\widetilde{S}'\widetilde{S} \to_p \Sigma_{\widetilde{s}\widetilde{s}} > 0$; $\frac{1}{n}V'V \to_p \Sigma_{vv} > 0$; $\frac{1}{n}Y_2'Y_2 \to_p \Sigma_{22}$, *(ii)* $\frac{1}{n}Z'V \to_p 0$; $\frac{1}{n}Z'e \to_p 0$; $\frac{1}{n}\widetilde{S}'V \to_p 0$; $\frac{1}{n}\widetilde{S}'e \to_p 0$, *and (iii)* $\frac{1}{n}\varepsilon'\varepsilon \to_p \sigma_\varepsilon^2 > 0$; $\frac{1}{n}e'e \to_p \sigma_e^2 > 0$, *where* $\Sigma_{zz} = \begin{bmatrix} \Sigma_{ww} & \Sigma_{ws} \\ \Sigma_{sw} & \Sigma_{ss} \end{bmatrix}$ *and notation "* > 0 *" in (i) signifies positive definiteness of the matrices.*

Remark 1 *Condition 1 assumes the weak law of large numbers of the variables in* Z, $\widetilde{S}$, V, *and* Y_2. *The asymptotic orthogonalities in Condition 1(ii) reflect the definitions of the parameterizations in (10), (1), and (4). In Condition 1(iii) we assume homoscedasticity of "errors", as is common in the literature. We discuss heteroscedasticity in Section 6.*

Condition 2 *We also assume that*

$$\begin{bmatrix} \frac{1}{\sqrt{n}}\operatorname{vec}\left(W'\widetilde{S}\right) \\ \frac{1}{\sqrt{n}}\operatorname{vec}\left(W'V\right) \\ \frac{1}{\sqrt{n}}W'e \end{bmatrix} \Rightarrow \begin{bmatrix} \operatorname{vec}\left(\mathcal{Z}_{w\widetilde{s}}\right) \\ \operatorname{vec}\left(\mathcal{Z}_{wv}\right) \\ \mathcal{Z}_{we} \end{bmatrix} \equiv N\left(0, \operatorname{diag}\left(\Sigma_{\widetilde{s}\widetilde{s}} \otimes \Sigma_{ww},\ \Sigma_{vv} \otimes \Sigma_{ww},\ \sigma_e^2\Sigma_{ww}\right)\right),$$

where $\operatorname{diag}\left(\Sigma_{\widetilde{s}\widetilde{s}} \otimes \Sigma_{ww},\ \Sigma_{vv} \otimes \Sigma_{ww},\ \sigma_e^2\Sigma_{ww}\right)$ *is a block diagonal matrix consisting of* $\Sigma_{\widetilde{s}\widetilde{s}} \otimes \Sigma_{ww}$, $\Sigma_{vv} \otimes \Sigma_{ww}$, *and* $\sigma_e^2\Sigma_{ww}$ *as blocks.*

B Preliminaries

Proofs of the lemmas below are available in an Online Appendix.

Lemma 2 *Assume Condition 1. The following hold both under the null and under the alternative.*
(a) $\frac{1}{n}Z'Z \to_p \begin{bmatrix} \Sigma_{ww} & \Sigma_{ww}\Gamma_w \\ \Gamma_w'\Sigma_{ww} & \Gamma_w'\Sigma_{ww}\Gamma_w + \Sigma_{\widetilde{s}\widetilde{s}} \end{bmatrix}$.
(b) $\frac{1}{n}Z'Y_2 \to_p \begin{bmatrix} 0 \\ \Sigma_{\widetilde{s}\widetilde{s}}\Pi_s \end{bmatrix}$.
(c) $\frac{1}{n}Y_2Z'\left(\frac{1}{n}Z'Z\right)^{-1}\frac{1}{n}Z'Y_2 \to_p \Pi_s'\Sigma_{\widetilde{s}\widetilde{s}}\Pi_s$.

Lemma 3 *Assume Conditions 1 and 2. The following hold under the null hypothesis in Section 4.*
(a) $\frac{1}{\sqrt{n}}W'\varepsilon \Rightarrow \mathcal{Z}_{wv}\rho_v + \mathcal{Z}_{we}$.
(b) $\begin{bmatrix} Y_2'P_WY_2 \\ Y_2'P_W\varepsilon \end{bmatrix} \Rightarrow \begin{bmatrix} \left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)'\Sigma_{ww}^{-1}\left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right) \\ \left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)'\Sigma_{ww}^{-1}\left(\mathcal{Z}_{wv}\rho_v + \mathcal{Z}_{we}\right) \end{bmatrix}$.

(c) $\frac{1}{n}Z'\varepsilon \to_p 0$.

(d) $\frac{1}{n}Y_2'\varepsilon \to_p \Sigma_{vv}\rho_v$.

(e) $\widehat{\beta}_z \to_p \beta$.

(f) $\widehat{\sigma}^2_{\varepsilon,z} \to_p \sigma^2_\varepsilon$.

(g) $\widetilde{\sigma}^2_{\varepsilon,z} \to_p \sigma^2_\varepsilon$.

Lemma 4 *Assume Conditions 1 and 2. The following hold under the alternative hypothesis in Section 4.*

(a) $\frac{1}{n}Z'\varepsilon \to_p \left[\begin{array}{c} 0 \\ \Sigma_{\widetilde{s}\widetilde{s}}\rho_s \end{array}\right]$.

(b) $\widehat{\beta}_z \to_p \beta + \tau$*, where* $\tau = \left(\Pi_s'\Sigma_{\widetilde{s}\widetilde{s}}\Pi_s\right)^{-1}\Pi_s'\Sigma_{\widetilde{s}\widetilde{s}}\rho_s$.

(c) $\frac{1}{\sqrt{n}}W'\left(\varepsilon - Y_2\tau\right) \Rightarrow \mathcal{Z}_{w\widetilde{s}}\left(\rho_s - \Pi_s\tau\right) - \Sigma_{ww}C\tau + \mathcal{Z}_{wv}\left(\rho_v - \tau\right) + \mathcal{Z}_{we}$.

(d) $\widetilde{\sigma}^2_{\varepsilon,z} \to_p \left(\rho_v - \tau\right)'\Sigma_{vv}\left(\rho_v - \tau\right) + \sigma^2_e$.

(e) $\widehat{\sigma}^2_{\varepsilon,z} \to_p \left(\rho_s - \Pi_s\tau\right)'\Sigma_{\widetilde{s}\widetilde{s}}\left(\rho_s - \Pi_s\tau\right) + \left(\rho_v - \tau\right)'\Sigma_{vv}\left(\rho_v - \tau\right) + \sigma^2_e$.

C Proofs of the Results in Section 4

We begin by presenting a rather natural result on the properties of the Hausman test with weak IV. In Theorem 4 below, we show that the Hausman test does not have the usual χ^2 distribution under the null.

C.1 Asymptotic Distribution of Hausman Test

Before presenting Theorem 4, we introduce the following definitions: define

$$\left[\begin{array}{c} \mathcal{D}_w \\ \mathcal{N}_w \end{array}\right] = \left[\begin{array}{c} \left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)'\Sigma_{ww}^{-1}\left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right) \\ \left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)'\Sigma_{ww}^{-1}\left(\mathcal{Z}_{wv}\rho_v + \mathcal{Z}_{we}\right) \end{array}\right],$$

$$\begin{aligned} &\xi\left(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv}, \mathcal{Z}_{we}\right) \\ &= \left(1 + \frac{1}{\sigma^2_\varepsilon}\left(\Sigma_{22}^{1/2}\mathcal{D}_w^{-1}\mathcal{N}_w - \Sigma_{22}^{-1/2}\Sigma_{vv}\rho_v\right)'\left(\Sigma_{22}^{1/2}\mathcal{D}_w^{-1}\mathcal{N}_w - \Sigma_{22}^{-1/2}\Sigma_{vv}\rho_v\right) - \frac{1}{\sigma^2_\varepsilon}\rho_v'\Sigma_{vv}\Sigma_{22}^{-1}\Sigma_{vv}\rho_v\right), \end{aligned}$$

and

$$\begin{aligned} \zeta\left(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv}\right) = &\frac{1}{\sigma_e}\left[\left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)'\Sigma_{ww}^{-1}\left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)\right]^{-1/2} \\ &\times\left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)'\Sigma_{ww}^{-1}\mathcal{Z}_{wv}\rho_v. \end{aligned}$$

$$\begin{aligned} \mathcal{H}_1 &= \left(\widehat{\beta}_w - \widehat{\beta}_z\right)'\left[\widehat{\sigma}^2_{\varepsilon,w}\left(Y_2'P_WY_2\right)^{-1} - \widehat{\sigma}^2_{\varepsilon,z}\left(Y_2'P_ZY_2\right)^{-1}\right]^{-1}\left(\widehat{\beta}_w - \widehat{\beta}_z\right), \\ \mathcal{H}_2 &= \widehat{\sigma}^{-2}_{\varepsilon,w}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)'\left[\left(Y_2'P_WY_2\right)^{-1} - \left(Y_2'P_ZY_2\right)^{-1}\right]^{-1}\left(\widehat{\beta}_w - \widehat{\beta}_z\right), \\ \mathcal{H}_3 &= \widehat{\sigma}^{-2}_{\varepsilon,z}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)'\left[\left(Y_2'P_WY_2\right)^{-1} - \left(Y_2'P_ZY_2\right)^{-1}\right]^{-1}\left(\widehat{\beta}_w - \widehat{\beta}_z\right). \end{aligned}$$

Theorem 4 *Assume Conditions 1 and 2. Suppose that $\mathcal{Z}$ denotes a random vector of $N(0, I_K)$ that is independent of $\zeta(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv})$. Under the null hypothesis (2),*
(a) $\mathcal{H}_1, \mathcal{H}_2 \Rightarrow \frac{\sigma_e^2}{\sigma_\varepsilon^2}\xi(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv}, \mathcal{Z}_{we})^{-1}(\zeta(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv}) + \mathcal{Z})'(\zeta(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv}) + \mathcal{Z})$.
(b) $\mathcal{H}_3 \Rightarrow \frac{\sigma_e^2}{\sigma_\varepsilon^2}(\zeta(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv}) + \mathcal{Z})'(\zeta(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv}) + \mathcal{Z})$.
Suppose that $L_w = K$, that is, W exactly identifies β. Then, under the null hypothesis (2),
(c) $\mathcal{H}_1, \mathcal{H}_2 \Rightarrow \xi(\mathcal{Z}_{w\widetilde{s}}, \mathcal{Z}_{wv}, \mathcal{Z}_{we})^{-1}\mathcal{Z}'\mathcal{Z}$.
(d) $\mathcal{H}_3 \Rightarrow \mathcal{Z}'\mathcal{Z} \equiv \chi_K^2$.

Proof. Part (a): Here we show only the limit of $\mathcal{H}_1$. The limit of $\mathcal{H}_2$ can be derived by similar fashion and we omit the proof. By Lemma 3(b), we have

$$\begin{bmatrix} Y_2'P_W Y_2 \\ Y_2'P_W\varepsilon \end{bmatrix} \Rightarrow \begin{bmatrix} \mathcal{D}_w \\ \mathcal{N}_w \end{bmatrix}.$$

From this, we can deduce that $\widehat{\beta}_w - \beta \Rightarrow \mathcal{D}_w^{-1}\mathcal{N}_w$. Also, by Lemma 3(d), (e), and (f), and Lemma 2(c), we have $\widehat{\beta}_z = \beta + o_p(1)$, $\widehat{\sigma}_{\varepsilon,z}^2 = \sigma_\varepsilon^2 + o_p(1)$, $\frac{1}{n}Y_2'\varepsilon = \Sigma_{vv}\rho_v + o_p(1)$, and $\frac{Y_2'P_Z Y_2}{n} = O_p(1)$. Therefore, we obtain

$$\begin{aligned}
\widehat{\sigma}_{\varepsilon,w}^2 &= \frac{1}{n}\varepsilon'\varepsilon - 2\left(\widehat{\beta}_w - \beta\right)'\frac{1}{n}Y_2'\varepsilon + \left(\widehat{\beta}_w - \beta\right)'\left(\frac{1}{n}Y_2'Y_2\right)\left(\widehat{\beta}_w - \beta\right) \\
&\Rightarrow \sigma_\varepsilon^2 - 2\mathcal{N}_w'\mathcal{D}_w^{-1}\Sigma_{vv}\rho_v + \mathcal{N}_w'\mathcal{D}_w^{-1}\Sigma_{22}\mathcal{D}_w^{-1}\mathcal{N}_w \\
&= \sigma_\varepsilon^2 + \left(\Sigma_{22}^{1/2}\mathcal{D}_w^{-1}\mathcal{N}_w - \Sigma_{22}^{-1/2}\Sigma_{vv}\rho_v\right)'\left(\Sigma_{22}^{1/2}\mathcal{D}_w^{-1}\mathcal{N}_w - \Sigma_{22}^{-1/2}\Sigma_{vv}\rho_v\right)' - \rho_v'\Sigma_{vv}\Sigma_{22}^{-1}\Sigma_{vv}\rho_v,
\end{aligned}$$

Now note that

$$\begin{aligned}
\mathcal{H}_1 &= \left(\widehat{\beta}_w - \widehat{\beta}_z\right)'\widehat{\sigma}_{\varepsilon,z}^{-2}\left(Y_2'P_Z Y_2\right)\left[\widehat{\sigma}_{\varepsilon,z}^{-2}\left(Y_2'P_Z Y_2\right) - \widehat{\sigma}_{\varepsilon,w}^{-2}\left(Y_2'P_W Y_2\right)\right]^{-1}\widehat{\sigma}_{\varepsilon,w}^{-2}\left(Y_2'P_W Y_2\right)\left(\widehat{\beta}_w - \widehat{\beta}_z\right) \\
&= \left(\widehat{\beta}_w - \beta + o_p(1)\right)'\left\{\widehat{\sigma}_{\varepsilon,z}^{-2}\left(Y_2'P_Z Y_2\right)\left[\widehat{\sigma}_{\varepsilon,z}^{-2}\left(Y_2'P_Z Y_2\right) - \widehat{\sigma}_{\varepsilon,w}^{-2}\left(Y_2'P_W Y_2\right)\right]^{-1}\right\} \\
&\quad\times \widehat{\sigma}_{\varepsilon,w}^{-2}\left(Y_2'P_W Y_2\right)\left(\widehat{\beta}_w - \beta + o_p(1)\right) \\
&= \left(\widehat{\beta}_w - \beta + o_p(1)\right)'\left\{\widehat{\sigma}_{\varepsilon,z}^{-2}\left(\frac{Y_2'P_Z Y_2}{n}\right)\left[\widehat{\sigma}_{\varepsilon,z}^{-2}\left(\frac{Y_2'P_Z Y_2}{n}\right) - \widehat{\sigma}_{\varepsilon,w}^{-2}\left(\frac{Y_2'P_W Y_2}{n}\right)\right]^{-1}\right\} \\
&\quad\times \widehat{\sigma}_{\varepsilon,w}^{-2}\left(Y_2'P_W Y_2\right)\left(\widehat{\beta}_w - \beta + o_p(1)\right) \\
&= \widehat{\sigma}_{\varepsilon,w}^{-2}\left(\varepsilon'P_W Y_2\right)\left(Y_2'P_W Y_2\right)^{-1}\left(Y_2'P_W\varepsilon\right) + o_p(1),
\end{aligned}$$

where the last equality follows from $\frac{Y_2'P_W Y_2}{n} = o_p(1)$. We therefore obtain

$$\mathcal{H}_1 \Rightarrow \left(\sigma_\varepsilon^2 + \left(\Sigma_{22}^{1/2}\mathcal{D}_w^{-1}\mathcal{N}_w - \Sigma_{22}^{-1/2}\Sigma_{vv}\rho_v\right)'\left(\Sigma_{22}^{1/2}\mathcal{D}_w^{-1}\mathcal{N}_w - \Sigma_{22}^{-1/2}\Sigma_{vv}\rho_v\right) - \rho_v'\Sigma_{vv}\Sigma_{22}^{-1}\Sigma_{vv}\rho_v\right)^{-1}\mathcal{N}_w'\mathcal{D}_w^{-1}\mathcal{N}_w.$$

Let

$$\mathcal{Z} = \frac{1}{\sigma_e}\left[\left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)'\Sigma_{ww}^{-1}\left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)\right]^{-1/2}\left(\Sigma_{ww}C + \mathcal{Z}_{w\widetilde{s}}\Pi_s + \mathcal{Z}_{wv}\right)'\Sigma_{ww}^{-1}\mathcal{Z}_{we}.$$

Then, $\mathcal{Z} \equiv N(0, I_K)$ and $\mathcal{Z}$ is independent of $\zeta(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv})$. Recalling the definitions of $\xi(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv}, \mathcal{Z}_{we})$ and $\zeta(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv})$, the limit of $\mathcal{H}_1$ is presented as

$$\frac{\sigma_e^2}{\sigma_\varepsilon^2}\xi(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv}, \mathcal{Z}_{we})^{-1}(\zeta(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv}) + \mathcal{Z})'(\zeta(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv}) + \mathcal{Z}).$$

Part (b): Notice that under the null hypothesis, by Lemma 3(f), $\widehat{\sigma}_{\varepsilon,z}^2 \to_p \sigma_\varepsilon^2$. Using similar arguments in the proof of Part (a), we can show that

$$\begin{aligned}
\mathcal{H}_3 &= \sigma_\varepsilon^{-2}\left(\widehat{\beta}_w - \beta\right)' Y_2' P_W Y_2 \left(\widehat{\beta}_w - \beta\right) + o_p(1) \\
&= \sigma_\varepsilon^{-2}\left(\varepsilon' P_W Y_2\right)\left(Y_2' P_W Y_2\right)^{-1}\left(Y_2' P_W \varepsilon\right) + o_p(1) \\
&\Rightarrow \sigma_\varepsilon^{-2}\mathcal{N}_w' \mathcal{D}_w^{-1}\mathcal{N}_w \\
&= \frac{\sigma_e^2}{\sigma_\varepsilon^2}(\zeta(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv}) + \mathcal{Z})'(\zeta(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv}) + \mathcal{Z}).
\end{aligned}$$

Part (c): When W exactly identifies β, $\mathcal{N}_w' \mathcal{D}_w^{-1}\mathcal{N}_w = (\mathcal{Z}_{wv}\rho_v + \mathcal{Z}_{we})\Sigma_{ww}^{-1}(\mathcal{Z}_{wv}\rho_v + \mathcal{Z}_{we})$. In this case, define

$$\mathcal{Z} = \frac{1}{\sigma_\varepsilon}\Sigma_{ww}^{-1/2}(\mathcal{Z}_{wv}\rho_v + \mathcal{Z}_{we}) \sim N(0, I_K).$$

Then, the limit distribution of $\mathcal{H}_1$ (and $\mathcal{H}_2$) is now $\xi(\mathcal{Z}_{w\tilde{s}}, \mathcal{Z}_{wv}, \mathcal{Z}_{we})^{-1}\mathcal{Z}'\mathcal{Z}$ as required. ■

C.2 Proof of Theorem 1

Part (a): We proceed as in Part (c) in the proof of Theorem 4. Define

$$\mathcal{Z} = \frac{1}{\sigma_\varepsilon}\Sigma_{ww}^{-1/2}(\mathcal{Z}_{wv}\rho_v + \mathcal{Z}_{we}) \sim N(0, I_K).$$

Then, the limit distribution of $\mathcal{H}_3$ is then $\mathcal{Z}'\mathcal{Z} \equiv \chi_K^2$. ■

Part (b): Using similar argument in the proof of Theorem 4 Part (a), we have

$$\begin{aligned}
\mathcal{H}_3 &= \widehat{\sigma}_{\varepsilon,z}^{-2}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)'\left(Y_2' P_Z Y_2\right)\left[\left(Y_2' P_Z Y_2\right) - \left(Y_2' P_W Y_2\right)\right]^{-1}\left(Y_2' P_W Y_2\right)\left(\widehat{\beta}_w - \widehat{\beta}_z\right) \\
&= \widehat{\sigma}_{\varepsilon,z}^{-2}\left(\widehat{\beta}_w - \beta - \tau + o_p(1)\right)'\left(\frac{Y_2' P_Z Y_2}{n}\right)\left[\left(\frac{Y_2' P_Z Y_2}{n}\right) - \left(\frac{Y_2' P_W Y_2}{n}\right)\right]^{-1}\left(Y_2' P_W Y_2\right)\left(\widehat{\beta}_w - \beta - \tau + o_p(1)\right) \\
&= \widehat{\sigma}_{\varepsilon,z}^{-2}\left(\widehat{\beta}_w - \beta - \tau + o_p(1)\right)'\left(Y_2' P_W Y_2 + o_p(1)\right)\left(\widehat{\beta}_w - \beta - \tau + o_p(1)\right) \\
&= \widehat{\sigma}_{\varepsilon,z}^{-2}\left((\varepsilon - Y_2\tau)' P_W Y_2\right)\left(Y_2' P_W Y_2\right)^{-1}\left(Y_2' P_W(\varepsilon - Y_2\tau)\right) + o_p(1) \\
&= \widehat{\sigma}_{\varepsilon,z}^{-2}(\varepsilon - Y_2\tau)' P_W(\varepsilon - Y_2\tau) + o_p(1),
\end{aligned}$$

where the second line holds since $\widehat{\beta}_z = \beta + \tau + o_p(1)$ by Lemma 4(b), the third line holds since $\frac{Y_2' P_W Y_2}{n} = o_p(1)$, and the last line follows since the dimension of W and dimension of Y_2 are the same and $Y_2' W$ is full rank.

By Lemmas 4 (c) and (e) and Condition 1, we can write

$$\frac{1}{\sqrt{n}}W'(\varepsilon - Y_2\tau) \Rightarrow \mathcal{Z}_{w\widetilde{s}}(\rho_s - \Pi_s\tau) + \mathcal{Z}_{wv}(\rho_v - \tau) + \mathcal{Z}_{we} - \Sigma_{ww}C\tau$$

and

$$\begin{aligned}\widehat{\sigma}^2_{\varepsilon,z} &\to_p (\rho_s - \Pi_s\tau)'\Sigma_{\widetilde{s}\widetilde{s}}(\rho_s - \Pi_s\tau) + (\rho_v - \tau)'\Sigma_{vv}(\rho_v - \tau) + \sigma^2_e \\ &= \sigma^2_{**}, \text{ say.}\end{aligned}$$

Define

$$\mathcal{Z} = \sigma^{-1}_{**}\Sigma^{-1/2}_{ww}\left(\mathcal{Z}_{w\widetilde{s}}(\rho_s - \Pi_s\tau) + \mathcal{Z}_{wv}(\rho_v - \tau) + \mathcal{Z}_{we}\right)$$

and

$$\kappa = -\frac{1}{\sigma_{**}}\Sigma^{1/2}_{ww}C\tau. \tag{15}$$

Then,

$$\mathcal{H}_3 \Rightarrow (\mathcal{Z}+\kappa)'(\mathcal{Z}+\kappa),$$

where $\mathcal{Z} \sim N(0, I_K)$. ■

D Proofs of Results in Section 5

We introduce a few lemmas that are helpful in proving Theorem 2. Lemmas 5 and 6 assume that the estimator $\widehat{\sigma}^2_\varepsilon$ is consistent under the null even when we adopt the weak IV asymptotics , and find the limit of the test statistic $\mathcal{H}\left(\widehat{\sigma}^2_\varepsilon\right)$ under the null and under the alternative, respectively. In Lemma 7 we provide an estimator $\widehat{\sigma}^2_\varepsilon$ that is consistent under the null even when we adopt the weak IV asymptotics.

Lemma 5 *Assume Conditions 1 and 2. Suppose that $\widehat{\sigma}^2_\varepsilon$ is consistent for σ^2_ε under the null using weak instrument asymptotics. Then $\mathcal{H}\left(\widehat{\sigma}^2_\varepsilon\right) \Rightarrow \chi^2_{L_w}$.*

Proof. The result easily follows from the proof of Lemma 6 by noting that $\rho_s = 0$, $\tau = 0$, and $\varepsilon = e$ under the null. ■

In Lemma 5, we make the additional assumption that $\widehat{\sigma}^2_\varepsilon$ is consistent for σ^2_ε under the weak instrument asymptotics in order to isolate the properties of the "numerator". This is inspired by the discussion in the previous section, where we have seen that the $\mathcal{H}_2$ failed to converge to a central chi-square distribution (see Proposition 4) because the estimator $\widehat{\sigma}^2_{\varepsilon,w}$ in (5) is inconsistent for σ^2_ε.

It turns out that the properties of $\widehat{\sigma}^2_\varepsilon$ have implications for the the power property of $\mathcal{H}\left(\widehat{\sigma}^2_\varepsilon\right)$ under the weak instrument asymptotics. Define

$$\sigma^2_* = (\rho_v - \tau)'\Sigma_{vv}(\rho_v - \tau) + \sigma^2_e$$

and

$$\kappa\left(\mathcal{Z}_{w\widetilde{s}}\right)=\sigma_*^{-1}\Sigma_{ww}^{-1/2}\left(\mathcal{Z}_{w\widetilde{s}}\left(\rho_s-\Pi_s\tau\right)-\Sigma_{ww}C\tau\right),$$

where

$$\tau=\operatorname{plim}\left(\widehat{\beta}_z-\beta\right)=\left(\Pi_s'\Sigma_{\widetilde{s}\widetilde{s}}\Pi_s\right)^{-1}\Pi_s'\Sigma_{\widetilde{s}\widetilde{s}}\rho_s$$

denotes the asymptotic bias of $\widehat{\beta}_z$ under the alternative hypothesis.

Lemma 6 *Assume Conditions 1 and 2. Under the alternative hypothesis (3),*

$$\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right)\Rightarrow\left(\kappa+\mathcal{Z}\right)'\left(\kappa+\mathcal{Z}\right),$$

where $\mathcal{Z}\sim N\left(0,I_{L_w}\right)$ and κ is the same noncentrality parameter in Theorem 1 (b).

Proof. Recall the definition

$$\mathcal{H}\left(\widehat{\sigma}_\varepsilon^2\right)=\frac{1}{\widehat{\sigma}_\varepsilon^2}\left(y_1-Y_2\widehat{\beta}_z\right)'W\widehat{\Psi}^{-1}W'\left(y_1-Y_2\widehat{\beta}_z\right)=\frac{\mathcal{N}}{\widehat{\sigma}_\varepsilon^2},\text{ say.}$$

We start with the analysis of

$$\mathcal{N}=\left(y_1-Y_2\widehat{\beta}_z\right)'W\widehat{\Psi}^{-1}W'\left(y_1-Y_2\widehat{\beta}_z\right).$$

Note that

$$\frac{1}{\sqrt{n}}W'\left(y_1-Y_2\widehat{\beta}_z\right)=\frac{1}{\sqrt{n}}W'\varepsilon-\frac{1}{\sqrt{n}}W'Y_2\cdot\left(\frac{1}{n}Y_2'Z\cdot\left(\frac{1}{n}ZZ\right)^{-1}\cdot\frac{1}{n}Z'Y_2\right)^{-1}\frac{1}{n}Y_2'Z\cdot\left(\frac{1}{n}Z'Z\right)^{-1}\cdot\frac{1}{n}Z'\varepsilon.$$

Using Lemmas 2, 3, 4, we can write

$$\begin{aligned}\frac{1}{\sqrt{n}}W'\left(y_1-Y_2\widehat{\beta}_z\right)&=\frac{1}{\sqrt{n}}W'\left(\varepsilon-Y_2\tau\right)+o_p\left(1\right)\\&\Rightarrow\mathcal{Z}_{w\widetilde{s}}\left(\rho_s-\Pi_s\tau\right)+\mathcal{Z}_{wv}\left(\rho_v-\tau\right)+\mathcal{Z}_{we}-\Sigma_{ww}C\tau\end{aligned}$$

Because

$$\frac{1}{n}Y_2'P_ZY_2=O_p\left(1\right),\quad\frac{1}{n}W'Y_2=O_p\left(\frac{1}{\sqrt{n}}\right),$$

we have

$$\frac{1}{n}\widehat{\Psi}=\frac{1}{n}W'W-\left(\frac{1}{n}W'Y_2\right)\left(\frac{1}{n}Y_2'P_ZY_2\right)^{-1}\left(\frac{1}{n}Y_2'W\right)=\Sigma_{ww}+o_p\left(1\right)$$

under both the null and the alternative. We may therefore write that

$$\begin{aligned}\mathcal{N}=&\left(\mathcal{Z}_{w\widetilde{s}}\left(\rho_s-\Pi_s\tau\right)+\mathcal{Z}_{wv}\left(\rho_v-\tau\right)+\mathcal{Z}_{we}-\Sigma_{ww}C\tau\right)'\Sigma_{ww}^{-1}\\&\cdot\left(\mathcal{Z}_{w\widetilde{s}}\left(\rho_s-\Pi_s\tau\right)+\mathcal{Z}_{wv}\left(\rho_v-\tau\right)+\mathcal{Z}_{we}-\Sigma_{ww}C\tau\right)+o_p\left(1\right).\end{aligned}$$

Also, under the alternative, by Lemma 4(e),

$$\widehat{\sigma}_\varepsilon^2\rightarrow_p\sigma_{**}^2,$$

where

$$\sigma_{**}^2 = (\rho_s - \Pi_s\tau)' \Sigma_{\tilde{s}\tilde{s}} (\rho_s - \Pi_s\tau) + (\rho_v - \tau)' \Sigma_{vv} (\rho_v - \tau) + \sigma_e^2.$$

Now let

$$\mathcal{Z} = \sigma_{**}^{-1} \Sigma_{ww}^{-1/2} \left(\mathcal{Z}_{w\widetilde{s}} (\rho_s - \Pi_s\tau) + \mathcal{Z}_{wv} (\rho_v - \tau) + \mathcal{Z}_{we} \right),$$

and

$$\kappa = -\sigma_{**}^{-1} \Sigma_{ww}^{1/2} C\tau.$$

Then, it is easy to see that $\mathcal{Z} \sim N(0, I_K)$. By writing

$$\mathcal{H}\left(\widehat{\sigma}_{\varepsilon}^2\right) = \frac{\mathcal{N}}{\widehat{\sigma}_{\varepsilon}^2} = (\kappa + \mathcal{Z})' (\kappa + \mathcal{Z}) + o_p(1),$$

we obtain the desired conclusion. ■

Lemmas 5 and 6 imply that it is important to choose $\widehat{\sigma}_{\varepsilon}^2$ such that it is consistent for σ_{ε}^2 under the null, and consistent for σ_*^2 under the alternative. To see this, suppose that we use $\widehat{\sigma}_{\varepsilon,z}^2$ in (6) for $\widehat{\sigma}_{\varepsilon}^2$. It can be shown[15] that $\widehat{\sigma}_{\varepsilon,z}^2 \to_p \sigma_{\varepsilon}^2$ under the null, but $\widehat{\sigma}_{\varepsilon,z}^2 \to_p (\rho_s - \Pi_s\tau)' \Sigma_{\widetilde{s}\widetilde{s}} (\rho_s - \Pi_s\tau) + \sigma_*^2$ under the alternative. In other words, we have $\sigma_*^2/\sigma_{**}^2 \leq 1$ if we use $\widehat{\sigma}_{\varepsilon,z}^2$. This implies that the asymptotic distribution of $\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right)$ under the alternative is the mixture of χ^2 distributions $\left(\kappa\left(\mathcal{Z}_{w\widetilde{s}}\right) + \mathcal{Z}\right)' \left(\kappa\left(\mathcal{Z}_{w\widetilde{s}}\right) + \mathcal{Z}\right)$ multiplied by a constant less than or equal to 1. Therefore, the test $\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right)$ may be asymptotically biased.[16]

Below, we present asymptotic properties of $\widetilde{\sigma}_{\varepsilon,z}^2$ developed in (13). It turns out that if we use $\widetilde{\sigma}_{\varepsilon,z}^2$ as an estimate of σ_{ε}^2, then the ratio $\sigma_*^2 / \sigma_{**}^2$ converges to 1 under the alternative:

Lemma 7 *Assume Conditions 1 and 2. Under the null,* $\widetilde{\sigma}_{\varepsilon,z}^2 \to_p \sigma_{\varepsilon}^2$*, and under the alternative,* $\widetilde{\sigma}_{\varepsilon,z}^2 \to_p \sigma_*^2 = (\rho_v - \tau)' \Sigma_{vv} (\rho_v - \tau) + \sigma_e^2$*.*

Proof. The required results follow by Lemma 3(g) and Lemma 4(d). ■

D.1 Proof of Theorem 2

Proof. Part(a) follows by Lemmas 5 and 7. Part (b) follows by Lemmas 6 and 7. ■

E Computational Issue

For convenience to practitioners, we provide below an alternative algorithm to compute $\mathcal{H}_4$ in the special case when the weak instruments W exactly identify the coefficient. The algorithm is based on the characterization (14) in Section 5: For a general over-identified case, please see the Online Appendix.

[15] See Lemma 3(f) and Lemma 4(e) in the appendix.

[16] Recall that $\mathcal{H}\left(\widehat{\sigma}_{\varepsilon,z}^2\right) = \mathcal{H}_3$ when $L_w = K$. The upshot is that unless $L_w = L_s = K$, $\mathcal{H}_4$ will be more powerful than $\mathcal{H}_3$; hence our focus on calculating $\mathcal{H}_4$ in the online appendix.

- Compute the 2SLS $\widehat{\beta}_z$ by using the instrument $Z = [W, S]$. Let $\widehat{V}_z = \widehat{\sigma}^2_{\varepsilon,z}\left(Y_2' P_Z Y_2\right)^{-1}$ denote the standard variance estimator of $\widehat{\beta}_z$, where $\widehat{\sigma}^2_{\varepsilon,z} = \frac{1}{n}\widehat{\varepsilon}_z'\widehat{\varepsilon}_z$ denote the standard estimator of σ^2_{ε}.

- Computation of $\widetilde{\sigma}^2_{\varepsilon,z}$:

  1. Using the IV estimator $\widehat{\beta}_z$, get the IV residual $\widehat{\varepsilon}_z = y_1 - Y_2\widehat{\beta}_z$.
  2. Regress the IV residual $\widehat{\varepsilon}_z$ on $Z = [S, W]$, and get the residual $\widetilde{\varepsilon}_z = M_Z\widehat{\varepsilon}_z$.
  3. Calculate $\widetilde{\sigma}^2_{\varepsilon,z} = \frac{1}{n}\widetilde{\varepsilon}_z'\widetilde{\varepsilon}_z$.

- Let

$$\widetilde{V}_z = \frac{\widetilde{\sigma}^2_{\varepsilon,z}}{\widehat{\sigma}^2_{\varepsilon,z}}\widehat{V}_z$$

- Compute the 2SLS $\widehat{\beta}_w$ by using the instrument W. Let $\widehat{V}_w = \widehat{\sigma}^2_{\varepsilon,w}\left(Y_2' P_W Y_2\right)^{-1}$ denote the standard variance estimator of $\widehat{\beta}_w$, where $\widehat{\sigma}^2_{\varepsilon,w} = \frac{1}{n}\widehat{\varepsilon}_w'\widehat{\varepsilon}_w$ denote the standard estimator of σ^2_{ε}.

- Let

$$\widetilde{V}_w = \frac{\widetilde{\sigma}^2_{\varepsilon,z}}{\widehat{\sigma}^2_{\varepsilon,w}}\widehat{V}_w$$

- $\mathcal{H}_4$ can now be calculated as

$$\mathcal{H}_4 = \left(\widehat{\beta}_w - \widehat{\beta}_z\right)'\left[\widetilde{V}_w - \widetilde{V}_z\right]^{-1}\left(\widehat{\beta}_w - \widehat{\beta}_z\right)$$

References

[1] Altonji, Joseph G. (1986): "Intertemporal Substitution in Labor Supply: Evidence from Micro Data", *Journal of Political Economy,* 94, S176-215.

[2] Andrews, Donald W.K. (2003): "End-of-Sample Instability Tests," *Econometrica*, 71, 1661-1694

[3] Andrews, Donald W.K., Marcelo Moreira, and James H. Stock (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression", *Econometrica*, 74, 715-752.

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

[5] Bekker, Paul A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators", *Econometrica*, 62, 657-681.

[6] Donald, Stephen D. and Whitney K. Newey (2001): "Choosing the Number of Instruments", *Econometrica*, 69, 1161-1191.

[7] Hahn, Jinyong and Jerry A. Hausman (2002): "A New Specification Test for the Validity of Instrumental Variables", *Econometric*a, 70, 163-189.

[8] Ham, John C. (1986): "Testing Whether Unemployment Represents Life-Cycle Labor Supply." *Review of Economic Studies*, 53, 559-78.

[9] Ham, John C. and Kevin T. Reilly (2002): "Testing Intertemporal Substitution, Implicit Contracts, and Hours Restriction Models of the Labor Market Using Micro Data." *American Economic Review*, 92, 905-927.

[10] Hausman, Jerry A. (1978): "Specification Tests in Econometrics", *Econometrica*, 46, 1251-1271.

[11] Kleibergen, Frank (2000): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression", *Econometrica*, 70, 1781-1803.

[12] MaCurdy, Thomas E. (1981): "An Empirical Model of Labour Supply in a Life-Cycle Setting." *Journal of Political Economy*, 89, 1059-85.

[13] Moreira, Marcelo (2003): "A Conditional Likelihood Ratio Test for Structural Models", *Econometrica*, 71, 1027-1048.

[14] Staiger, Douglas., and James H. Stock (1997): "IV Regression with Weak Instruments," *Econometrica*, 65, 557-586.

[15] Strauss, John and Duncan Thomas. (1995): “Human Resources: Empirical Modeling of Household and Family Decisions” , in T.N. Srinivasan and Jere Behrman eds., *Handbook of Development Economics*, Volume 3A, North Holland Press, 1883-2024.

[16] Willis, Robert and Sherwin Rosen. (1979): “Education and Self-Selection,” *Journal of Political Economy*, 87, s7-s36.

Table A: Size of Test, # weak IV = 1, ρ=.25

n	*# strong IV*	R^2	*H1*	*H2*	*H3*	*H4*
100	1	0.01	0.00	0.00	0.05	0.05
100	1	0.02	0.01	0.01	0.05	0.06
100	1	0.03	0.01	0.01	0.05	0.06
100	1	0.05	0.01	0.01	0.05	0.06
100	1	0.1	0.02	0.02	0.05	0.06
100	1	0.2	0.03	0.04	0.05	0.05
200	1	0.01	0.01	0.01	0.05	0.05
200	1	0.02	0.01	0.01	0.05	0.05
200	1	0.03	0.01	0.01	0.05	0.05
200	1	0.05	0.02	0.02	0.05	0.05
200	1	0.1	0.03	0.03	0.05	0.05
200	1	0.2	0.04	0.04	0.05	0.05
500	1	0.01	0.01	0.01	0.05	0.05
500	1	0.02	0.02	0.02	0.05	0.05
500	1	0.03	0.03	0.03	0.05	0.05
500	1	0.05	0.03	0.03	0.05	0.06
500	1	0.1	0.04	0.04	0.05	0.05
500	1	0.2	0.05	0.05	0.06	0.06
100	2	0.01	0.00	0.00	0.05	0.06
100	2	0.02	0.01	0.01	0.05	0.06
100	2	0.03	0.01	0.01	0.05	0.05
100	2	0.05	0.01	0.01	0.05	0.06
100	2	0.1	0.02	0.02	0.05	0.06
100	2	0.2	0.03	0.03	0.06	0.06
200	2	0.01	0.01	0.01	0.05	0.06
200	2	0.02	0.01	0.01	0.05	0.06
200	2	0.03	0.01	0.01	0.05	0.05
200	2	0.05	0.02	0.02	0.05	0.05
200	2	0.1	0.03	0.03	0.05	0.05
200	2	0.2	0.04	0.04	0.05	0.05
500	2	0.01	0.01	0.01	0.06	0.06
500	2	0.02	0.02	0.02	0.06	0.06
500	2	0.03	0.03	0.03	0.06	0.06
500	2	0.05	0.04	0.04	0.06	0.06
500	2	0.1	0.04	0.05	0.05	0.06
500	2	0.2	0.05	0.05	0.06	0.06
100	5	0.01	0.00	0.00	0.05	0.05
100	5	0.02	0.00	0.00	0.05	0.06
100	5	0.03	0.01	0.01	0.05	0.06
100	5	0.05	0.01	0.01	0.05	0.06
100	5	0.1	0.02	0.02	0.05	0.06
100	5	0.2	0.03	0.03	0.05	0.06
200	5	0.01	0.01	0.01	0.05	0.06
200	5	0.02	0.01	0.01	0.05	0.06
200	5	0.03	0.01	0.01	0.05	0.06
200	5	0.05	0.02	0.02	0.05	0.06
200	5	0.1	0.03	0.03	0.06	0.06
200	5	0.2	0.04	0.05	0.06	0.06
500	5	0.01	0.01	0.01	0.05	0.05
500	5	0.02	0.02	0.02	0.05	0.05
500	5	0.03	0.02	0.02	0.05	0.05
500	5	0.05	0.03	0.03	0.05	0.05
500	5	0.1	0.04	0.04	0.05	0.05
500	5	0.2	0.04	0.04	0.05	0.05

R^2 is the partial R^2 of the weak IV.

Table B: size of Test, # weak IV = 5, ρ=.75

n	*# strong IV*	R^2	*H1*	*H2*	*H3*	*H4*
100	1	0.01	0.37	0.32	0.23	0.11
100	1	0.02	0.34	0.29	0.20	0.11
100	1	0.03	0.31	0.27	0.18	0.11
100	1	0.05	0.27	0.22	0.14	0.11
100	1	0.1	0.20	0.15	0.10	0.11
100	1	0.2	0.14	0.09	0.07	0.10
200	1	0.01	0.38	0.35	0.23	0.09
200	1	0.02	0.33	0.30	0.19	0.09
200	1	0.03	0.29	0.25	0.15	0.09
200	1	0.05	0.24	0.19	0.12	0.09
200	1	0.1	0.16	0.12	0.08	0.09
200	1	0.2	0.11	0.07	0.06	0.09
500	1	0.01	0.30	0.28	0.16	0.07
500	1	0.02	0.22	0.21	0.12	0.07
500	1	0.03	0.18	0.16	0.10	0.07
500	1	0.05	0.14	0.12	0.07	0.07
500	1	0.1	0.10	0.08	0.06	0.07
500	1	0.2	0.07	0.06	0.05	0.07
100	2	0.01	0.35	0.31	0.23	0.12
100	2	0.02	0.33	0.28	0.20	0.12
100	2	0.03	0.30	0.25	0.17	0.11
100	2	0.05	0.25	0.20	0.14	0.11
100	2	0.1	0.18	0.13	0.09	0.11
100	2	0.2	0.12	0.07	0.06	0.10
200	2	0.01	0.35	0.32	0.20	0.09
200	2	0.02	0.30	0.27	0.16	0.08
200	2	0.03	0.26	0.22	0.13	0.08
200	2	0.05	0.21	0.17	0.10	0.08
200	2	0.1	0.14	0.10	0.07	0.08
200	2	0.2	0.09	0.06	0.06	0.08
500	2	0.01	0.30	0.28	0.16	0.06
500	2	0.02	0.22	0.20	0.11	0.06
500	2	0.03	0.18	0.16	0.09	0.06
500	2	0.05	0.13	0.11	0.07	0.06
500	2	0.1	0.09	0.07	0.06	0.06
500	2	0.2	0.07	0.06	0.05	0.06
100	5	0.01	0.26	0.23	0.18	0.12
100	5	0.02	0.23	0.20	0.15	0.12
100	5	0.03	0.21	0.17	0.13	0.12
100	5	0.05	0.17	0.13	0.09	0.12
100	5	0.1	0.11	0.07	0.06	0.12
100	5	0.2	0.06	0.04	0.04	0.12
200	5	0.01	0.30	0.27	0.18	0.09
200	5	0.02	0.25	0.22	0.13	0.09
200	5	0.03	0.20	0.18	0.11	0.09
200	5	0.05	0.15	0.13	0.08	0.08
200	5	0.1	0.09	0.07	0.06	0.08
200	5	0.2	0.06	0.04	0.05	0.08
500	5	0.01	0.27	0.26	0.16	0.07
500	5	0.02	0.20	0.19	0.11	0.07
500	5	0.03	0.16	0.14	0.09	0.07
500	5	0.05	0.11	0.10	0.07	0.07
500	5	0.1	0.08	0.06	0.06	0.07
500	5	0.2	0.06	0.05	0.05	0.07

R^2 is the partial R^2 of the weak IV.

Table C: Power of Test, # weak IV = 1, # strong IV = 5, γ_s=1

n	R^2	ρ	*H3*	*H4*
100	0.01	0.25	0.10	0.22
100	0.02	0.25	0.14	0.27
100	0.03	0.25	0.17	0.32
100	0.05	0.25	0.24	0.40
100	0.1	0.25	0.37	0.56
100	0.2	0.25	0.55	0.71
200	0.01	0.25	0.13	0.25
200	0.02	0.25	0.21	0.36
200	0.03	0.25	0.29	0.45
200	0.05	0.25	0.43	0.59
200	0.1	0.25	0.64	0.78
200	0.2	0.25	0.81	0.90
500	0.01	0.25	0.27	0.42
500	0.02	0.25	0.46	0.62
500	0.03	0.25	0.61	0.76
500	0.05	0.25	0.79	0.89
500	0.1	0.25	0.95	0.98
500	0.2	0.25	0.99	1.00
100	0.01	0.5	0.11	0.26
100	0.02	0.5	0.16	0.33
100	0.03	0.5	0.21	0.39
100	0.05	0.5	0.29	0.50
100	0.1	0.5	0.45	0.65
100	0.2	0.5	0.62	0.78
200	0.01	0.5	0.16	0.32
200	0.02	0.5	0.26	0.46
200	0.03	0.5	0.35	0.56
200	0.05	0.5	0.50	0.69
200	0.1	0.5	0.71	0.86
200	0.2	0.5	0.86	0.93
500	0.01	0.5	0.32	0.53
500	0.02	0.5	0.54	0.74
500	0.03	0.5	0.70	0.85
500	0.05	0.5	0.86	0.95
500	0.1	0.5	0.97	0.99
500	0.2	0.5	0.99	1.00
100	0.01	0.75	0.13	0.36
100	0.02	0.75	0.20	0.45
100	0.03	0.75	0.27	0.54
100	0.05	0.75	0.37	0.63
100	0.1	0.75	0.55	0.77
100	0.2	0.75	0.70	0.84
200	0.01	0.75	0.20	0.46
200	0.02	0.75	0.33	0.61
200	0.03	0.75	0.45	0.71
200	0.05	0.75	0.61	0.82
200	0.1	0.75	0.80	0.91
200	0.2	0.75	0.89	0.95
500	0.01	0.75	0.41	0.70
500	0.02	0.75	0.67	0.88
500	0.03	0.75	0.81	0.94
500	0.05	0.75	0.93	0.98
500	0.1	0.75	0.98	0.99
500	0.2	0.75	1.00	1.00

R^2 is the partial R^2 of the weak IV.