The Hausman Test and Weak Instruments‡

Jinyong Hahn† John C. Ham§ Hyungsik Roger Moon¶

September 7, 2010

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’ instruments, 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 presence of weak instruments, which we confirm. However, we provide a version of the Hausman test that is valid even in the presence of weak IV, and illustrate how to implement the test in the presence of heteroskedasticity. We show that the situation we analyze occurs in several important economic examples. Our Monte Carlo experiments show that our 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.

JEL Classification: C12

Key Words: Hausman Test, Weak Instruments.

1 Introduction

The weak instruments problem has led to the development to two strands of research, each of which is characterized by a different asymptotic approximation. The first of these, which we will call the...

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

†All the omitted proofs and derivations are available in our Online Appenex (Hahn, Ham, and Moon (2010)) which is available at www-rcf.usc.edu/~moonr or upon request by email to the authors.

§Department of Economics, University of California, Los Angeles.

∥Department of Economics, University of Maryland, IZA, IRP (UW-Madison).

¶Department of Economics, University of Southern California
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.\footnote{See Bekker (1994), Donald and Newey (2001), and Hahn and Hausman (2002), among others.} Because the many-instrument 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 the 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.\footnote{See also Kleibergen (2000), Moreira (2003), and Andrews, Moreira, and Stock (2004).} 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 refer to this approximation as the weak-instruments asymptotics or Staiger and Stock asymptotics.

Under Staiger and Stock’s asymptotic approximation, many usual statistics have nonstandard asymptotic distributions. For example, it is well-known that IV estimators and t-statistics have nonstandard distributions. Staiger and Stock (1997) also considered tests of overidentification under their asymptotics, and established that standard tests of overidentification do not have chi-square ($\chi^2$ 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 employed 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 the IV and its potentially complicated consequences.

In this paper, we extend Staiger and Stock’s (1997) analysis and document further exceptional cases. We next 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 are adopted. We consider a Hausman test that compares weak IV against strong IV. It is well-known that the test statistic has a $\chi^2$ distribution under conventional asymptotics. We establish that a version of Hausman test continues to have the $\chi^2$ distribution even under the weak instrument asymptotics. We then show that a version of the overidentification test, which we interpret to be a natural generalization of the 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. They have a structural equation of interest that 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 instruments’, which have 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 some of the instruments are weak, which our paper 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 $\chi^2$ 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 that has a standard $\chi^2$ distribution. In Section 6 we first consider the case where it is the weak instrument, and not the strong instrument, that may not be valid. We then extend our test to the case where the errors are heteroskedastic. Finally we extend the test

---

3See Section 3 below.

4To carry out this extension, we must first derive the standard Hausman test to for the case where both the exogenous and potentially endogenous instruments are strong and the errors are heteroskedastic. Although this derivation is quite straightforward from the perspective of econometric theory, it should be helpful to applied researchers since this test is not currently available in the literature.
statistic from Section 5 to improve its power while maintaining its good size properties. 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 there are weak instruments which overidentify the model, 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

\[ y_1 = Y_2\beta + \varepsilon \quad \text{and} \quad Y_2 = Z\Pi + V, \]

where \( \varepsilon \) 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 \( \varepsilon \), \( Z \) is an \( n \times L \) matrix of IV’s with \( L \geq K \) that are independent of \( V \). 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, \( \tilde{S} \) is the “residual” when \( S \) is projected on \( W \) in the population,

\[ S = WT_{w} + \tilde{S}. \]

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

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

\[ H_0 : E[s_i\varepsilon_i] = 0 \quad \text{or} \quad E[\tilde{s}_i\varepsilon_i] = 0, \]

\[ H_1 : E[s_i\varepsilon_i] \neq 0 \quad \text{or} \quad E[\tilde{s}_i\varepsilon_i] \neq 0. \]

Thus, if the exclusion restriction is violated, then it is only through the possible correlation between \( \tilde{s}_i \) and \( \varepsilon_i \).

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

\[ \varepsilon = \tilde{S}\rho_s + V\rho_v + \varepsilon, \]

\footnote{We follow the standard approach, and assume that included exogenous variables are ‘partialled out’ - see the online appendix for more details.}

\footnote{The main focus of the paper is not to obtain a post-specification test inference, but rather to investigate the validity of various versions of the Hausman tests themselves with weak IV’s. See Guggenberger(2008), e.g., for issues on post-specification test inference.}
where \( \rho_v = [E(v_i'v_i)]^{-1}E(v_i\varepsilon_i) \) denotes the coefficient of projection of \( \varepsilon_i \) on \( v_i \). We will assume that \( \varepsilon_i \) is uncorrelated with \( \tilde{s}_i \) and \( v_i \) and has mean zero and variance \( \sigma^2_e \). 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:\footnote{Given a matrix \( A \), we use notation \( P_A = A(A' A)^{-1} A \) and \( M_A = I - P_A \) throughout.}

\[
\hat{\beta}_w = (Y_2' P_W Y_2)^{-1} Y_2' P_W y_1, \\
\hat{\beta}_z = (Y_2' P_Z Y_2)^{-1} Y_2' P_Z y_1.
\]

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

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

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

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

where

\[
\hat{\sigma}_{\varepsilon,w}^2 = \frac{1}{n} (y_1 - Y_2 \hat{\beta}_w)' (y_1 - Y_2 \hat{\beta}_w) \tag{5}
\]

and

\[
\hat{\sigma}_{\varepsilon,z}^2 = \frac{1}{n} (y_1 - Y_2 \hat{\beta}_z)' (y_1 - Y_2 \hat{\beta}_z). \tag{6}
\]

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

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:\(^8\)

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

where \(\Delta\) 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, and \(e_{it}\) is an idiosyncratic error term. Further, \(\eta_{it}\) 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 \(\eta_{it}\). We also expect that \(\Delta \ln(w_{it})\) is correlated with \(\Delta e_{it}\) 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 \(\Delta e_{it}\), 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 \(\eta_{it}\) 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 \(\eta_{it}\), 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 instruments.

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 valid IV for \(y_{it-1}\). However, the choice of \(k\) is usually arbitrary, since economic theory does not

\(^8\)See MaCurdy 1981, Altoni 1986, Ham 1986, Ham and Reilly 2002. Corner solutions at zero hours are not important for this group and thus a regression framework is appropriate.
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 health 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 have used 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), while \( X_i \) 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 \( X_i \), and again out test provides a diagnostic for Willis and Rosen’s assumption that \( FE_i \) is a valid IV.

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:

\[
Y_2 = W \frac{C}{\sqrt{n}} + \bar{S} \Pi_s + V. \tag{10}
\]

We assume that \( L_s \geq K \). Suppose that we consider \( \mathcal{H}_1, \mathcal{H}_2, \) or \( \mathcal{H}_3 \), but adopt Staiger and Stock’s (1997) alternative asymptotic approximation. Because their asymptotics implies that the asymptotic
distribution of $\hat{\beta}_w$ is not normal, it is natural to conjecture that the asymptotic size of the conventional procedure would be distorted. Not surprisingly, $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{H}_3$ are usually not distributed as $\chi^2_K$ under the weak-instrument asymptotic approximation – see Theorem 4 in 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 $\chi^2_K$ 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 Appendix A. Suppose that $L_w = K$ and $Y_2'W$ has full rank $K$. Then, (a) under the null hypothesis (2),

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

(b) under the alternative hypothesis (3),

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

where $Z \sim N(0, I_K)$ and the noncentrality parameter $\kappa$ is defined in (16) in the Appendix.

**Proof.** In Appendix C.2. ■

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

The invalidity of $\mathcal{H}_1$ and $\mathcal{H}_2$ (even under exact identification) can be attributed to the failure to estimate $\sigma^2_\varepsilon$ consistently under the null. On the other hand, even $\mathcal{H}_3$, which is based on a consistent estimator of $\sigma^2_\varepsilon$ under the null, is invalid without exact identification. This is one of the main differences between our results and the results in Staiger and Stock (1997), who test for exogeneity of the regressors (there the strong IV’s are the regressors). In the next section, we develop a modification of the Hausman test that does not require exact identification. This modification requires $\sigma^2_\varepsilon$ to be estimated consistently under the null.

5 **Generalized Hausman Test**

The results in Section 4 imply 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 $\chi^2_K$, is asymptotically valid even under the weak-instrument asymptotics as

---

9We impose standard regularity conditions, which are discussed in the Appendix.
long as \( \beta \) is exactly identified by weak instruments \((L_w = K)\). On the other hand, Theorem 4 in the Appendix C.1 shows that neither \( H_1 \), \( H_2 \), nor \( H_3 \), along with a critical value from \( \chi^2_K \), is valid under the weak-instrument asymptotics if the weak IV’s overidentify \( \beta \), 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 identifies \( \beta \) and thus \( H_3 \) can be used. Although one can resolve the situation in this fashion, it is not clear which of the \( K \) 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 a chi-square distribution even under the weak-instrument asymptotics 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

\[
E \left[ w_i \left( y_{1i} - Y_2 \text{plim} \hat{\beta}_z \right) \right] = 0.
\]

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

\[
E \left[ w_i \left( y_{1i} - Y_2 \text{plim} \hat{\beta}_z \right) \right] \neq 0
\]

because the probability limit of \( \hat{\beta}_z \) will be different from \( \beta \) under the alternative. From these observations, we might consider a test statistic based on

\[
\frac{1}{\sqrt{n}} W' \left( y_1 - Y_2 \hat{\beta}_z \right).
\]

With some algebra, it can be shown that

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

\[
\frac{1}{\sqrt{n}} W' \left( y_1 - Y_2 \hat{\beta}_z \right) \Rightarrow N \left( 0, \sigma_z^2 \Psi \right),
\]

where \( \Psi \) is the probability limit of \( \frac{1}{n} \hat{\Psi} \) and

\[
\hat{\Psi} = W'W - (W'Y_2) (Y_2' P_Z Y_2)^{-1} (Y_2' W).
\]

**Proof.** In the Online Appendix. □

Therefore, the test statistic is equal to

\[
\mathcal{H} \left( \hat{\sigma}_z^2 \right) = \frac{1}{\hat{\sigma}_z^2} \left( y_1 - Y_2 \hat{\beta}_z \right)' W \hat{\Psi}^{-1} W' \left( y_1 - Y_2 \hat{\beta}_z \right),
\]

where \( \hat{\Psi} = W'W - (W'Y_2) (Y_2' P_Z Y_2)^{-1} (Y_2' W) \) and \( \hat{\sigma}_z^2 \) denotes some consistent estimator for \( \sigma_z^2 \). In light of Lemma 1, it is straightforward to conclude that the (conventional) asymptotic distribution of \( \mathcal{H} \left( \hat{\sigma}_z^2 \right) \) is \( \chi^2_{L_w} \). In other words, researchers can use \( \mathcal{H} \left( \hat{\sigma}_z^2 \right) \) to obtain a \( \chi^2 \)-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 \( \mathcal{H} \left( \hat{\sigma}_z^2 \right) \).
Proposition 1  When \( L_w = K \),
\[
\mathcal{H} \left( \hat{\sigma}_e^2 \right) = \frac{1}{\hat{\sigma}_e^2} \left( \hat{\beta}_w - \hat{\beta}_z \right)' \left[ (Y'_2 P_W Y_2)^{-1} - (Y'_2 P_Z Y_2)^{-1} \right]^{-1} \left( \hat{\beta}_w - \hat{\beta}_z \right),
\]

**Proof.** In the Online Appendix. ■

From Proposition 1, we can conclude that \( \mathcal{H} \left( \hat{\sigma}_e^2 \right) \) can be understood as a version of the Hausman test in a special case where \( L_w = K \). Depending on the estimator \( \hat{\sigma}_e^2 \) used, the statistic \( \mathcal{H} \left( \hat{\sigma}_e^2 \right) \) can be understood to be an extension of \( \mathcal{H}_2 \) or \( \mathcal{H}_3 \).\(^{10}\)

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

**Theorem 2**  Assume Conditions 1 and 2 in Appendix A. Assume that \( L_w \geq K \). (a) Under the null hypothesis,
\[
\mathcal{H} \left( \hat{\sigma}_{\varepsilon,z}^2 \right) \Rightarrow \chi^2_{L_w},
\]
(b) under the alternative hypothesis,
\[
\mathcal{H} \left( \hat{\sigma}_{\varepsilon,z}^2 \right) \Rightarrow (\kappa + Z)' (\kappa + Z),
\]
where \( Z \sim N(0, I_{L_w}) \) and \( \kappa \) is the same noncentrality parameter as in Theorem 1.

**Proof.** In Appendix D.1. ■

Theorem 2 indicates that using \( \mathcal{H} \left( \hat{\sigma}_{\varepsilon,z}^2 \right) \) along with a critical value from \( \chi^2 \) is asymptotically valid even under the weak-instrument asymptotics, and 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.\(^{11}\) Second, we examine the consequences of heteroscedasticity. After this, we consider the issue of improving power.

\(^{10}\)When \( L_w = K \), we have \( \mathcal{H} \left( \hat{\sigma}_{\varepsilon,w}^2 \right) = \mathcal{H}_2 \) and \( \mathcal{H} \left( \hat{\sigma}_{\varepsilon,z}^2 \right) = \mathcal{H}_3 \).

\(^{11}\)We thank an anonymous referee for suggesting this question.
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 $Y_2 = \tilde{W}\frac{C}{\sqrt{n}} + \Pi s + V$, where the alternative hypothesis is now written $\varepsilon = \tilde{W}\rho w + V\rho v + e$. Here $\tilde{W}$ is the population projection “residual” of $W$ on $S$: $W = S\Gamma s + \tilde{W}$.

Here we consider the properties of (generalized) Hausman test statistic $H(\hat{\sigma}^2_{\varepsilon,z})$. It can be shown that $H(\hat{\sigma}^2_{\varepsilon,z})$ is distributed as $\chi^2_{L_w}$ under the null, but diverges to $\infty$ under the alternative. In other words, the $H(\hat{\sigma}^2_{\varepsilon,z})$ 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.

Since the size of the Hausman test is valid only when $L_w = K$ even under homoscedasticity, we assume that the dimension of the weak IV’s is the same as the dimension of $\beta$, that is, $L_w = K$. Here for expositional simplicity we assume that $\beta$ is a scalar (that is, $K = 1$). The extension to a general $K$ is straightforward and has been placed in the Online appendix.

Given that the Hausman test has an interpretation of a comparison between $\hat{\beta}_w$ and $\hat{\beta}_z$, a natural modification of the Hausman test statistic would take the form

$$n \left( \hat{\beta}_w - \hat{\beta}_z \right)^2 \over \text{Var} \left( \sqrt{n} \left( \hat{\beta}_w - \hat{\beta}_z \right) \right)$$

(12)

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

$$\sqrt{n} \left( \hat{\beta}_w - \hat{\beta}_z \right) = \sqrt{n} \left( \hat{\beta}_w - \beta \right) - \sqrt{n} \left( \hat{\beta}_z - \beta \right)$$

$$= \left( \frac{W'Y_2}{n} \right)^{-1} \left( \frac{W'\varepsilon}{\sqrt{n}} \right) - \left( \frac{Y_2'PZY_2}{n} \right)^{-1} \left( \frac{Y_2'PZ\varepsilon}{\sqrt{n}} \right),$$

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

13 The proof is available in our Online Appendix.
and its asymptotic variance under conventional asymptotics is

\[
\frac{E \left[ w_i^2 \varepsilon_i^2 \right]}{(E[w_i Y_{2i}])^2} - 2 \frac{E (Y_{2i} \varepsilon_i) (E (z_i \varepsilon_i))^{-1} E (z_i w_i \varepsilon_i^2)}{E[w_i Y_{2i}]} \left( E(Y_{2i} \varepsilon_i) (E (z_i \varepsilon_i))^{-1} E (z_i Y_{2i}) \right) \frac{E(Y_{2i} \varepsilon_i) (E (z_i \varepsilon_i))^{-1} E (z_i Y_{2i})}{\left[ E(Y_{2i} \varepsilon_i) (E (z_i \varepsilon_i))^{-1} E (z_i Y_{2i}) \right]^2}.
\]

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

\[
\text{Var} \left( \sqrt{n} \left( \hat{\beta}_w - \hat{\beta}_z \right) \right) = \frac{\left\{ \frac{1}{n} \sum_{i=1}^{n} w_i^2 \varepsilon_i^2 \right\} - 2 \frac{\left\{ \frac{1}{n} \sum_{i=1}^{n} Y_{2i} \varepsilon_i \right\} (\frac{1}{n} \sum_{i=1}^{n} z_i \varepsilon_i)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} z_i w_i \varepsilon_i^2 \right)}{\left\{ \frac{1}{n} \sum_{i=1}^{n} w_i Y_{2i} \right\} \left( \frac{1}{n} \sum_{i=1}^{n} Y_{2i} \varepsilon_i \right) \left( \frac{1}{n} \sum_{i=1}^{n} z_i \varepsilon_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} z_i Y_{2i} \right)} + \frac{\left\{ \frac{1}{n} \sum_{i=1}^{n} Y_{2i} \varepsilon_i \right\} (\frac{1}{n} \sum_{i=1}^{n} z_i \varepsilon_i)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} z_i \varepsilon_i \right)^2 \left( \frac{1}{n} \sum_{i=1}^{n} z_i \varepsilon_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} z_i Y_{2i} \right)}{\left\{ \frac{1}{n} \sum_{i=1}^{n} z_i \varepsilon_i \right\} \left( \frac{1}{n} \sum_{i=1}^{n} z_i \varepsilon_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} z_i Y_{2i} \right)} \right]}
\]

Although the above derivation is quite straightforward from the perspective of econometric theory, it should be helpful to applied researchers since the standard Hausman term where both the exogenous and potentially endogenous instruments are strong is not currently available in the literature when there is heteroskedasticity. In the Online appendix we show that under the Staiger-Stock asymptotics

\[
H_{\text{hetero}} (\hat{\varepsilon}_z) = \frac{n \left( \hat{\beta}_w - \hat{\beta}_z \right)^2}{\text{Var} (\hat{\beta}_w - \hat{\beta}_z)} \Rightarrow (Z + \kappa_{\text{hetero}})^2,
\]

where \( Z \sim N (0, 1) \) and

\[
\kappa_{\text{hetero}} = \frac{-\Sigma_{ww} C_T}{\left( \lim_{n} \frac{1}{n} \sum_{i=1}^{n} E \left[ w_i^2 (s_i (\rho_s - \Pi_s \tau) + v_i (\rho_v - \tau) + \epsilon_i)^2 \right] \right)^{1/2}}.
\]

where \( \tau = (\Pi_s \Sigma_{ss} \Pi_s)^{-1} \Pi_s \Sigma_{ss} \rho_s \). Since \( \kappa_{\text{hetero}} = 0 (\tau = 0) \) under the null,

\[
H_{\text{hetero}} (\hat{\varepsilon}_z) \Rightarrow \chi_1^2.
\]

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

\[
H_{\text{hetero}} (\hat{\varepsilon}_z) \Rightarrow \chi_1^2 (\kappa_{\text{hetero}}),
\]

and thus the test \( H_{\text{hetero}} (\hat{\varepsilon}_z) \) becomes asymptotically unbiased.
6.3 Improving the Power of \( \mathcal{H}_3 \)

In Section 4, we noted that \( \mathcal{H}_3 \) is asymptotically unbiased for the case where \( L = K \). 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 a 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 instruments. 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:

\[
\mathcal{H}_4 = \tilde{\sigma}^2_{\varepsilon,z} \left( \tilde{\beta}_w - \tilde{\beta}_z \right)' \left[ (Y'_2 P_W Y_2)^{-1} - (Y'_2 P_Z Y_2)^{-1} \right]^{-1} \left( \tilde{\beta}_w - \tilde{\beta}_z \right),
\]

where \( \tilde{\sigma}^2_{\varepsilon,z} \) is obtained by the following algorithm:

1. Using the IV estimator \( \tilde{\beta}_z \), we get the IV residual \( \varepsilon_1 = y_1 - Y_2 \tilde{\beta}_z \).

2. Regress the IV residual \( \varepsilon_1 \) on \( Z = [S, W] \) to get the residual \( M_Z \varepsilon_1 \) and define

\[
\tilde{\sigma}^2_{\varepsilon,z} = \frac{1}{n} \varepsilon_1' M_Z \varepsilon_1 = \sigma^2_{\varepsilon,z} = \frac{1}{n} \varepsilon_1' P_Z \varepsilon_1.
\]

(14)

From (14), one can see that the proposed estimator \( \tilde{\sigma}^2_{\varepsilon,z} \) modifies \( \sigma^2_{\varepsilon,z} \) by subtracting \( \frac{1}{n} \varepsilon_1' P_Z \varepsilon_1 \).

Although it is generally preferable to use the modified estimator \( \tilde{\sigma}^2_{\varepsilon,z} \), there are two special cases where such modification is unnecessary and \( \sigma^2_{\varepsilon,z} \) does not need to be modified. The first case is when \( \Sigma^{1/2}_{ss} \rho_s \) belongs to the space spanned by the columns of \( \Sigma^{1/2}_{ss} \Pi_s \) (or \( \Pi_s \)). Recall the definition \( \tau = (\Pi'_s \Sigma_{ss}^{-1} \Pi_s) \). We then have \( \Sigma^{1/2}_{ss} \rho_s = \Sigma^{1/2}_{ss} \Pi_s (\Pi'_s \Sigma_{ss}^{-1} \Pi_s)^{-1} \Pi'_s \Sigma_{ss} \rho_s \), 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 \( \Pi_s \) is invertible. We then have \( \rho_s = \Pi_s \tau \) and \( \sigma^2_s = \sigma^2_{s,s} \), and there is no need for the second step modification above – we can use \( \sigma^2_{\varepsilon,z} \) in place of \( \tilde{\sigma}^2_{\varepsilon,z} \).

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 \( \chi^2_K \) under the null, and its asymptotic distribution under the null stochastically dominates \( \chi^2_K \).
Theorem 3 Assume Conditions 1 and 2 in Appendix A. Suppose that \( L_w = K \). Then \( \mathcal{H}_4 \Rightarrow \chi^2_K \) under the null. Under the alternative hypothesis \( (3) \), \( \mathcal{H}_4 \Rightarrow \frac{\sigma^2_{z}}{\sigma^2_{z}} (Z + \kappa)' (Z + \kappa) \), where \( Z \sim N (0, I_K) \), \( \kappa \) is the same noncentrality parameter as in Theorem 1. Also, \( \sigma^2_{s*} = \text{plim} \sigma^2_{z,z} \) under the alternative and \( \sigma^2_{s} = \text{plim} \sigma^2_{z,z} \) 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, \( \sigma^2_{z,z} \rightarrow_p \sigma^2_s = (\rho_v - \tau)' \Sigma_{vv} (\rho_v - \tau) + \sigma^2_e \) and \( \sigma^2_{s,s} \rightarrow_p \sigma^2_{s,s} = (\rho_s - \Pi_s \tau)' \Sigma_{SS} (\rho_s - \Pi_s \tau) + \sigma^2_s \), where \( \tau = \text{plim} (\beta - \beta) = (\Pi_s' \Sigma_{SS} \Pi_s)^{-1} \Pi_s' \Sigma_{SS} \rho_s \).

Here it is obvious to see \( \sigma^2_s/\sigma^2_{s,s} \leq 1 \), 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 \( \mathcal{H} (\sigma^2_{z,z}) \) is dominated by \( \mathcal{H} (\sigma^2_{s,s}) \). Using Theorem 2, it is easy to see that \( \mathcal{H} (\sigma^2_{z,z}) \) is asymptotically unbiased and its power dominates that of \( \mathcal{H} (\sigma^2_{s,s}) \). As such, \( \mathcal{H} (\sigma^2_{z,z}) \) is a desirable test.

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

\[
\mathcal{H} (\sigma^2_{z,z}) = \sigma^{-2}_{z,z} (\beta - \beta)' [ (Y_2' P_W Y_2)^{-1} - (Y_2' P_Z Y_2)^{-1} ]^{-1} (\beta - \beta) = \mathcal{H}_4
\]

by Proposition 1. Based on this equality, we will, without too much loss of generality, define \( \mathcal{H}_4 = \mathcal{H} (\sigma^2_{z,z}) \) even for the overidentified case.\(^{14}\)

7 Monte Carlo Simulations

The data generating process used in the Monte Carlo simulations is

\[
y_{1i} = Y_{2i} \beta + s_i \rho_s + \varepsilon_i
\]

\[
Y_{2i} = w_i' \Pi_w + s_i' \Pi_s + v_i, \quad i = 1 , \ldots , n
\]

where \((w_i', s_i') \sim N (0, I), (\varepsilon_i , v_i) \sim N \left( \begin{array}{c} 1 \\ \rho \\ 1 \end{array} \right) \), and \( y_{1i} \) and \( Y_{2i} \) are scalars. Further, \( \beta = 1 \) and \( \Pi_w \) and \( \Pi_s \) are proportional to vectors consisting of ones. They are related to the (partial) first stage \( R^2 \) by

\[
R^2_w = \frac{\Pi_w' \Pi_w}{\Pi_w' \Pi_w + 1}, \quad R^2_s = \frac{\Pi_s' \Pi_s}{\Pi_s' \Pi_s + 1}.
\]

We fixed \( R^2_s = 0.2 \) throughout the simulation, and we considered \( R^2_w = 0.01, 0.02 \). We set \( \rho_s = 0 \) under the null, and \( \rho_s = (\gamma_s , 0 , \ldots , 0) \) under the alternative. We consider \( n = 100, 500, \rho = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \)

\[
R^2_w = 0.01, 0.02, 0.03, 0.05, 0.1, 0.2, \text{ and } \gamma_s = 1. \] The dimensions of \( w_i \) and \( s_i \) are \( L_w = 1, 5 \) and

\(^{14}\)From our Online Appendix one sees that a test based on \( \mathcal{H}_4 \) is quite easy to construct in Stata or similar programs.
\( L_s = 1, 2, 5 \) respectively. The nominal size of the test is 5\%. (Additional cases are considered in our Online Appendix.) All of 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 \( R^2_w \). Ideally each entry should be 0.05, so that we see that in each case the size with low \( R^2_w \) 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 \( R^2_w \) and the sample size increase, the size distortion of both \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) deceases. 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. In this case, for tests \( \mathcal{H}_3 \) and \( \mathcal{H}_4 \) we consider \( \mathcal{H}_3 = \mathcal{H}(\hat{\sigma}^2_{e,z}) \) and \( \mathcal{H}_4 = \mathcal{H}(\hat{\sigma}^2_{e,z}) \), where the test statistic \( \mathcal{H}(\cdot) \) is defined in (11). Now the size of each test is biased upwards - this is especially true for \( \mathcal{H}_1 \) and \( \mathcal{H}_2 \) with low \( R^2_w \). 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 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 \( R^2_w \) is low, the power of \( \mathcal{H}_4 \) is often more than double that of \( \mathcal{H}_3 \) when \( n = 100 \), and about 50\% greater than that of \( \mathcal{H}_3 \) when \( n = 500 \). Thus in terms of power with low \( R^2_w \), \( \mathcal{H}_4 \) 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.\(^{15}\) (Recall that there is no need to use the modified

\(^{15}\) 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.
version of $\overline{\sigma}^2_{\epsilon,z}$ 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 for 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, while most strong instruments are potentially endogenous. Using Staiger-Stock asymptotics, we show that the standard Hausman test for this case may have a 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 our 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.
Appendix

A Regularity Conditions

Condition 1 We assume the following. (i) \( \frac{1}{n} Z' Z \rightarrow_p \Sigma_{zz} > 0; \) \( \frac{1}{n} \tilde{S}' \tilde{S} \rightarrow_p \Sigma_{\tilde{S}\tilde{S}} > 0; \) \( \frac{1}{n} V' V \rightarrow_p \Sigma_{vv} > 0; \)
\( \frac{1}{n} Y_2' Y_2 \rightarrow_p \Sigma_{22}; \) (ii) \( \frac{1}{n} Z' e \rightarrow_p 0; \) \( \frac{1}{n} \tilde{S}' e \rightarrow_p 0; \) \( \frac{1}{n} V' e \rightarrow_p 0, \) and (iii) \( \frac{1}{n} e' e \rightarrow_p \sigma_e^2 > 0; \)
\( \frac{1}{n} e' e \rightarrow_p \sigma_e^2 > 0, \) where \( \Sigma_{zz} = \left[ \begin{array}{cc} \Sigma_{ww} & \Sigma_{ws} \\ \Sigma_{sw} & \Sigma_{ss} \end{array} \right] \) and the 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 \( Z, \tilde{S}, V, \) 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 the “errors”, as is common in the literature. We discuss heteroscedasticity in Section 6.

Condition 2 We also assume that
\[
\begin{bmatrix}
\frac{1}{\sqrt{n}} \text{vec} \left( W' \tilde{S} \right) \\
\frac{1}{\sqrt{n}} \text{vec} \left( W' V \right) \\
\frac{1}{\sqrt{n}} W' e
\end{bmatrix} \Rightarrow \begin{bmatrix}
\text{vec} \left( \Sigma_{ws} \right) \\
\text{vec} \left( \Sigma_{wv} \right) \\
\Sigma_{we}
\end{bmatrix} = N \left( 0, \text{diag} \left( \Sigma_{\tilde{S}\tilde{S}} \otimes \Sigma_{ww}, \Sigma_{wv} \otimes \Sigma_{ww}, \sigma_e^2 \Sigma_{ww} \right) \right),
\]
where \( \text{diag} \left( \Sigma_{\tilde{S}\tilde{S}} \otimes \Sigma_{ww}, \Sigma_{wv} \otimes \Sigma_{ww}, \sigma_e^2 \Sigma_{ww} \right) \) is a block diagonal matrix consisting of \( \Sigma_{\tilde{S}\tilde{S}} \otimes \Sigma_{ww}, \Sigma_{wv} \otimes \Sigma_{ww}, \) and \( \sigma_e^2 \Sigma_{ww} \) as blocks.

B Preliminaries

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

Lemma 2 Assume Condition 1 holds. The following hold both under the null and under the alternative.
\( \frac{1}{n} Z' Z \rightarrow_p \left[ \begin{array}{ccc}
\Sigma_{ww} & \Sigma_{ww} \Gamma_w \\
\Gamma_w' \Sigma_{ww} & \Gamma_w' \Sigma_{ww} \Gamma_w + \Sigma_{\tilde{S}\tilde{S}} \end{array} \right]. \)

(b) \( \frac{1}{n} Z' Y_2 \rightarrow_p \left[ \begin{array}{c}
0 \\
\Sigma_{\tilde{S}\tilde{S}} \Pi_s
\end{array} \right]. \)

(c) \( \frac{1}{n} Y_2' Z' \left( \frac{1}{n} Z' Z \right)^{-1} \frac{1}{n} Z' Y_2 \rightarrow_p \Pi_s' \Sigma_{\tilde{S}\tilde{S}} \Pi_s. \)

Lemma 3 Assume Conditions 1 and 2 hold. The following hold under the null hypothesis in Section 4.
\( \frac{1}{\sqrt{n}} W' e \Rightarrow Z_{ww} \rho_v + Z_{we}. \)
Before presenting Theorem 4, we introduce the following definitions: define

\[
\begin{align*}
(b) & \quad Y_2^t P W Y_2 
\Rightarrow \begin{bmatrix} (\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv})' \Sigma_{wv}^{-1} (\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv}) \\
\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv})' \Sigma_{wv}^{-1} (\Sigma_{wv} \rho_v + Z_{wve})
\end{bmatrix}. \\
(c) & \quad \frac{1}{n} Z' \varepsilon \xrightarrow{p} 0. \\
(d) & \quad \frac{1}{n} Y_2^t \varepsilon \xrightarrow{p} \Sigma_{wv} \rho_v. \\
(f) & \quad \tilde{\sigma}_{\varepsilon, \varepsilon}^2 \xrightarrow{p} \sigma_{\varepsilon}^2. \\
(g) & \quad \tilde{\sigma}_{\varepsilon, \varepsilon}^2 \xrightarrow{p} \sigma_{\varepsilon}^2. 
\end{align*}
\]

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

(a) \( \frac{1}{n} Z' \varepsilon \xrightarrow{p} \begin{bmatrix} 0 \\
\Sigma_{wv} \rho_v 
\end{bmatrix} \).

(b) \( \tilde{\beta}_z \xrightarrow{p} \beta + \tau \), where \( \tau = (\Pi_s' \Sigma_{wv} \Pi_s)^{-1} \Pi_s' \Sigma_{wv} \rho_v \).

(c) \( \frac{1}{\sqrt{n}} W' (\varepsilon - Y_2 \tau) \Rightarrow Z_{wv} (\rho_v - \Pi_s \tau) - \Sigma_{wv} C \tau + Z_{wv} (\rho_v - \tau) + Z_{wve}. \)

(d) \( \tilde{\sigma}_{\varepsilon, \varepsilon}^2 \xrightarrow{p} (\rho_v - \tau)' \Sigma_{wv} (\rho_v - \tau) + \sigma_{\varepsilon}^2. \)

(e) \( \tilde{\sigma}_{\varepsilon, \varepsilon}^2 \xrightarrow{p} (\rho_v - \Pi_s \tau)' \Sigma_{wv} (\rho_v - \Pi_s \tau) + (\rho_v - \tau)' \Sigma_{wv} (\rho_v - \tau) + \sigma_{\varepsilon}^2. \)

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 \( \chi^2 \) distribution under the null.

C.1 The Asymptotic Distribution of Hausman Test

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

\[
\begin{align*}
D_w &= \begin{bmatrix} 0 \\
\Sigma_{wv} \rho_v 
\end{bmatrix} \\
N_w &= \begin{bmatrix} (\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv})' \Sigma_{wv}^{-1} (\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv}) \\
(\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv})' \Sigma_{wv}^{-1} (\Sigma_{wv} \rho_v + Z_{wve})
\end{bmatrix}.
\end{align*}
\]

\[
\xi (Z_{w}, Z_{wv}, Z_{we}) = \left( 1 + \frac{1}{\sigma_{\varepsilon}^2} \left( \Sigma_{w2} \Sigma_{wv}^{-1} N_w - \Sigma_{w2} \Sigma_{wv}^{-1} \Sigma_{wv} \rho_v \right)' \left( \Sigma_{w2} \Sigma_{wv}^{-1} N_w - \Sigma_{w2} \Sigma_{wv}^{-1} \Sigma_{wv} \rho_v \right) \right) - \frac{1}{\sigma_{\varepsilon}^2} \rho_v' \Sigma_{wv} \Sigma_{w2} \Sigma_{wv}^{-1} \rho_v,
\]

\[
\zeta (Z_{w}, Z_{wv}) = \frac{1}{\sigma_{\varepsilon}^2} \left[ (\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv})' \Sigma_{wv}^{-1} (\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv}) \right]^{-1/2} \times (\Sigma_{wv} C + Z_{wv} \Pi_s + Z_{wv})' \Sigma_{wv}^{-1} \Sigma_{wv} \rho_v,
\]

18
Now note that
\[ \mathcal{H}_1 = \left( \mathbf{w}_e - \mathbf{z}_e \right)^T \left[ \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 (Y_2'P \mathbf{Y}_2)^{-1} - \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 (Y_2'P \mathbf{Z}_2)^{-1} \right]^{-1} \left( \mathbf{w}_e - \mathbf{z}_e \right), \]
\[ \mathcal{H}_2 = \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( \mathbf{w}_e - \mathbf{z}_e \right)^T \left[ (Y_2'P \mathbf{Y}_2)^{-1} - (Y_2'P \mathbf{Z}_2)^{-1} \right]^{-1} \left( \mathbf{w}_e - \mathbf{z}_e \right), \]
and \[ \mathcal{H}_3 = \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( \mathbf{w}_e - \mathbf{z}_e \right)^T \left[ (Y_2'P \mathbf{Y}_2)^{-1} - (Y_2'P \mathbf{Z}_2)^{-1} \right]^{-1} \left( \mathbf{w}_e - \mathbf{z}_e \right). \]

**Theorem 4** Assume Conditions 1 and 2 hold, and suppose that \( Z \) denotes a random vector of \( N(0, I_K) \) that is independent of \( \zeta (Z_{wz}, Z_{wv}) \). Under the null hypothesis (2),
(a) \( \mathcal{H}_1, \mathcal{H}_2 \Rightarrow \frac{\sigma^2}{\sigma^2} \xi (Z_{wz}, Z_{wv}, Z_{wv})^{-1} (\zeta (Z_{wz}, Z_{wv}) + \mathbf{Z}) \) \( \zeta (Z_{wz}, Z_{wv}) + \mathbf{Z} \).
(b) \( \mathcal{H}_3 \Rightarrow \frac{\sigma^2}{\sigma^2} (G_{wz}, G_{wv})^{-1} (\zeta (Z_{wz}, Z_{wv}) + \mathbf{Z}) \).
Suppose that \( L_w = K \), that is, \( W \) exactly identifies \( \beta \). Then, under the null hypothesis (2),
(c) \( \mathcal{H}_1, \mathcal{H}_2 \Rightarrow \xi (Z_{wz}, Z_{wv}, Z_{wv})^{-1} Z'Z \).
(d) \( \mathcal{H}_3 \Rightarrow Z'Z = \chi^2_K \).

**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
\[ \begin{bmatrix} Y_2'P \mathbf{Y}_2 \\ Y_2'P \mathbf{w} \end{bmatrix} \Rightarrow \begin{bmatrix} \mathcal{D}_{w} \\ \mathcal{N}_{w} \end{bmatrix}. \]
From this, we can deduce that \( \mathbf{w}_e - \beta \Rightarrow \mathcal{D}_{w}^{-1} \mathcal{N}_{w} \). Also, by Lemma 3(d), (e), and (f), and Lemma 2(c), we have \( \mathbf{w}_e = \beta + o_p(1), \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 = \sigma_{\mathbf{z},\mathbf{z}}^2 + o_p(1), \frac{1}{n} Y_2' \mathbf{w} = \Sigma_{vw} \rho_v + o_p(1), \) and \( \frac{Y_2'P \mathbf{Z}_2}{n} = O_p(1) \). Therefore, we obtain
\[ \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 = \frac{1}{n} \mathbf{w}_e - 2 \left( \beta - \beta \right) \left( \frac{1}{n} Y_2' \mathbf{w} \right) \left( \frac{1}{n} Y_2' \mathbf{w} \right)^T \left( \beta - \beta \right) \]
\[ \Rightarrow \sigma_{\mathbf{z},\mathbf{z}}^2 - 2 N_{w} \mathcal{D}_{w}^{-1} \Sigma_{vw} \rho_v + N_{w} \mathcal{D}_{w}^{-1} \Sigma_{22} \mathcal{D}_{w}^{-1} \mathcal{N}_{w} \]
\[ = \sigma_{\mathbf{z},\mathbf{z}}^2 + \left( \Sigma_{22} \mathcal{D}_{w}^{-1} \mathcal{N}_{w} - \Sigma_{22} \mathcal{D}_{w}^{-1} \mathcal{N}_{w} \right) \left( \Sigma_{22} \mathcal{D}_{w}^{-1} \mathcal{N}_{w} - \Sigma_{22} \mathcal{D}_{w}^{-1} \mathcal{N}_{w} \right) \]
\[ - \rho_v \Sigma_{vw} \Sigma_{22} \mathcal{D}_{w}^{-1} \mathcal{N}_{w}. \]
Now note that
\[ \mathcal{H}_1 = \left( \beta - \mathbf{z}_e \right)^T \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( Y_2'P \mathbf{Z}_2 \right) \left[ \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( Y_2'P \mathbf{Z}_2 \right) - \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( Y_2'P \mathbf{Y}_2 \right) \right]^{-1} \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( Y_2'P \mathbf{Y}_2 \right) \left( \beta - \mathbf{z}_e \right) \]
\[ = \left( \beta - \beta + o_p(1) \right) \left( \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( Y_2'P \mathbf{Z}_2 \right) - \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( Y_2'P \mathbf{Y}_2 \right) \right) \left( \beta - \beta + o_p(1) \right) \]
\[ = \left( \beta - \beta + o_p(1) \right) \left( \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( Y_2'P \mathbf{Z}_2 \right) - \hat{\sigma}_{\mathbf{z},\mathbf{z}}^2 \left( Y_2'P \mathbf{Y}_2 \right) \right) \left( \beta - \beta + o_p(1) \right) \]
\[ = \sigma_{\mathbf{z},\mathbf{z}}^2 \left( \epsilon'P \mathbf{Y}_2 \right) \left( Y_2'P \mathbf{Y}_2 \right)^{-1} \left( Y_2'P \mathbf{w} \right) + o_p(1). \]
where the last equality follows from $\frac{Y_2'PWY_2}{n} = o_p(1)$. We therefore obtain

$$H_1 \Rightarrow \left( \sigma_2^2 + \left( \Sigma_{21}^{1/2} \mathcal{D}_w^{-1} N_w - \Sigma_{21}^{1/2} \Sigma_{vV} \rho_v \right) ' \left( \Sigma_{22}^{1/2} \mathcal{D}_w^{-1} N_w - \Sigma_{21}^{1/2} \Sigma_{vV} \rho_v \right) - \rho_v ' \Sigma_{vV} \Sigma_{22}^{1/2} \Sigma_{vV} \rho_v \right)^{-1} \mathcal{N}_w \mathcal{D}_w^{-1} N_w.$$  

Let

$$Z = \frac{1}{\sigma_e} \left[ \left( \Sigma_{ww} C + \Sigma_{ww} \Pi_s + \Sigma_{ww} ' \Sigma_{ww} \Pi_s + \Sigma_{ww} \right)^{-1/2} \left( \Sigma_{ww} C + \Sigma_{ww} \Pi_s + \Sigma_{ww} \right) \Sigma_{ww} ' \Sigma_{ww} \right].$$

Then, $Z \equiv N(0, I_K)$ and $Z$ is independent of $\zeta(z_{w,\tilde{w}}, z_{w,v}, z_{w,v})$. Recalling the definitions of $\xi(z_{w,\tilde{w}}, z_{w,v}, z_{w,v})$ and $\zeta(z_{w,\tilde{w}}, z_{w,v}, z_{w,v})$, the limit of $H_1$ is

$$\frac{\sigma_2^2}{\sigma_e^2} \xi(z_{w,\tilde{w}}, z_{w,v}, z_{w,v})^{-1} \left( \zeta(z_{w,\tilde{w}}, z_{w,v}) + Z \right) ' \left( \zeta(z_{w,\tilde{w}}, z_{w,v}) + Z \right).$$

**Part (b):** Notice that under the null hypothesis, by Lemma 3(f), $\frac{\sigma_2^2}{\sigma_e^2} \rightarrow_p 1$. Using similar arguments to those used in the proof of Part (a), we can show that

$$H_3 = \sigma_e^{-2} \left( \beta_w - \beta \right) ' Y_2' PWY_2 \left( \beta_w - \beta \right) + o_p(1)$$

$$= \sigma_e^{-2} \left( \varepsilon ' PWY_2 \right) \left( Y_2 ' PWY_2 \right)^{-1} \left( Y_2 ' PW \varepsilon \right) + o_p(1)$$

$$\Rightarrow \sigma_e^{-2} \mathcal{N}_w \mathcal{D}_w^{-1} N_w$$

$$= \frac{\sigma_e^2}{\sigma_e^2} \left( \xi(z_{w,\tilde{w}}, z_{w,v}) + Z \right) ' \left( \xi(z_{w,\tilde{w}}, z_{w,v}) + Z \right).$$

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

$$Z = \frac{1}{\sigma_e} \Sigma_{ww}^{-1/2} \left( \Sigma_{ww} \rho_v + \Sigma_{we} \right) \sim N(0, I_K).$$

Then, the limit distribution of $H_1$ (and $H_2$) is now $\xi(z_{w,\tilde{w}}, z_{w,v}, z_{w,v})^{-1} Z ' Z$ as required. $\blacksquare$

### C.2 Proof of Theorem 1

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

$$Z = \frac{1}{\sigma_e} \Sigma_{ww}^{-1/2} \left( \Sigma_{ww} \rho_v + \Sigma_{we} \right) \sim N(0, I_K).$$

Then, the limit distribution of $H_3$ is then $Z ' Z \equiv \chi_2^2$. $\blacksquare$

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

$$H_3 = \frac{\sigma_e^{-2}}{\sigma_e^2} \left( \beta_w - \beta \right) ' \left( Y_2 ' PWY_2 \right) \left( (Y_2 ' PWY_2) - (Y_2 ' PWY_2)^{-1} (Y_2 ' PWY_2) \right) \left( \beta_w - \beta \right)$$

$$= \frac{\sigma_e^{-2}}{\sigma_e^2} \left( \beta_w - \beta - \tau + o_p(1) \right) ' \left( Y_2 ' PWY_2 \right) \left( \left( Y_2 ' PWY_2 \right) - \left( Y_2 ' PWY_2 \right) \right)^{-1} \left( Y_2 ' PWY_2 \right) \left( \beta_w - \beta - \tau + o_p(1) \right)$$

$$= \frac{\sigma_e^{-2}}{\sigma_e^2} \left( \varepsilon - Y_2 \tau \right) ' \left( Y_2 ' PWY_2 + o_p(1) \right) \left( \beta_w - \beta - \tau + o_p(1) \right)$$

$$= \frac{\sigma_e^{-2}}{\sigma_e^2} \left( \varepsilon - Y_2 \tau \right) ' \left( Y_2 ' PWY_2 + o_p(1) \right) \left( \beta_w - \beta - \tau + o_p(1) \right)$$

$$= \frac{\sigma_e^{-2}}{\sigma_e^2} \left( \varepsilon - Y_2 \tau \right) ' \left( Y_2 ' PWY_2 + o_p(1) \right) \left( \beta_w - \beta - \tau + o_p(1) \right).$$
where the second line holds since \( \hat{\beta}_z = \beta + \tau + o_p(1) \) by Lemma 4(b), the third line holds since \( \frac{Y_2'P_0W_0Y_2}{n} = o_p(1) \), 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

\[
\frac{1}{\sqrt{n}} W' (\varepsilon - Y_2 \tau) \Rightarrow Z_w \tilde{\sigma}_w \tau + Z_{wv} \rho_v - Z_{we} - \Sigma_{wv} C \tau
\]

and

\[
\hat{\sigma}_{\varepsilon \tau}^2 \rightarrow_p \left( \rho_s - \Pi_s \tau \right)' \Sigma_{ss} \left( \rho_s - \Pi_s \tau \right) + \left( \rho_v - \tau \right)' \Sigma_{vv} \left( \rho_v - \tau \right) + \sigma_e^2 = \sigma_{**}^2, \text{ say.}
\]

Define

\[
Z = \sigma_{**}^{-1} \Sigma_{vw}^{-1/2} \left( Z_w \tilde{\sigma}_w \tau + Z_{wv} \rho_v - Z_{we} \right)
\]

and

\[
\kappa = -\frac{1}{\sigma_{**}} \Sigma_{vw}^{1/2} C \tau.
\]

Then,

\[
H_3 \Rightarrow (Z + \kappa)' (Z + \kappa),
\]

where \( Z \sim N(0, I_K). \)

\[\text{D Proofs of the Results in Section 5}\]

\[\text{D.1 Some Useful Lemmas}\]

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

**Lemma 5** Assume Conditions 1 and 2 hold. Suppose that \( \hat{\sigma}_e^2 \) is consistent for \( \sigma_e^2 \) under the null using weak instrument asymptotics. Then \( H \left( \hat{\sigma}_e^2 \right) \Rightarrow \chi^2_{\Sigma_{ww}}.\)

**Proof** The result easily follows from the proof of Lemma 6 below by noting that \( \rho_s = 0, \tau = 0, \) and \( \varepsilon = \epsilon \) under the null. ■

In Lemma 5, we make the additional assumption that \( \hat{\sigma}_e^2 \) is consistent for \( \sigma_e^2 \) 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 saw that $\mathcal{H}_2$ failed to converge to a central chi-square distribution (see Proposition 4) because the estimator $\hat{\sigma}_{e^2}^2$ in (5) is inconsistent for $\sigma_e^2$.

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

$$
\sigma_{e^2}^2 = (\rho_v - \tau)' \Sigma_{vv} (\rho_v - \tau) + \sigma_e^2
$$

and

$$
\kappa(Z_{\text{w}\hat{\beta}}) = \sigma_s^{-1} \Sigma_{wv}^{-1/2} (Z_{\text{w}\hat{\beta}} (\rho_s - \Pi_s \tau) - \Sigma_{wv} \Sigma \tau),
$$

where

$$
\tau = \lim(\hat{\beta}_{e^2} - \beta) = (\Pi'_s \Sigma_{\text{ss}} \Sigma_s)^{-1} \Pi'_s \Sigma_{\text{ss}} \rho_s
$$
denotes the asymptotic bias of $\hat{\beta}_{e^2}$ under the alternative hypothesis.

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

$$
\mathcal{H}(\hat{\sigma}_{e^2}^2) \Rightarrow (\kappa + \mathcal{Z})' (\kappa + \mathcal{Z}),
$$

where $\mathcal{Z} \sim N(0, I_{Lw})$ and $\kappa$ is the same noncentrality parameter as in Theorem 1 (b).

**Proof** Recall the definition

$$
\mathcal{H}(\hat{\sigma}_{e^2}^2) = \frac{1}{\hat{\sigma}_{e^2}^2} (y_1 - Y_2 \hat{\beta}_{e^2})' W \hat{\Psi}^{-1} W' (y_1 - Y_2 \hat{\beta}_{e^2}) = \frac{N}{\hat{\sigma}_{e^2}^2}, \text{ say.}
$$

We start with the analysis of

$$
N = (y_1 - Y_2 \hat{\beta}_{e^2})' W \hat{\Psi}^{-1} W' (y_1 - Y_2 \hat{\beta}_{e^2}).
$$

Note that

$$
\frac{1}{\sqrt{n}} W' (y_1 - Y_2 \hat{\beta}_{e^2}) = \frac{1}{\sqrt{n}} W' \varepsilon - \frac{1}{\sqrt{n}} W' Y_2 - \frac{1}{\sqrt{n}} W' Z \left( \frac{1}{n} Z Z' \right)^{-1} - \frac{1}{n} Z' Y_2
$$

Using Lemmas 2, 3, and 4, we can write

$$
\frac{1}{\sqrt{n}} W' (y_1 - Y_2 \hat{\beta}_{e^2}) = \frac{1}{\sqrt{n}} W' (\varepsilon - Y_2 \tau) + o_p(1)
$$

$$
\Rightarrow Z_{w\hat{\beta}} (\rho_s - \Pi_s \tau) + Z_{wv} (\rho_v - \tau) + Z_{we} - \Sigma_{wv} \Sigma \tau.
$$

Because

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

we have

$$
\frac{1}{n} \hat{\Psi} = \frac{1}{n} W' W - \left( \frac{1}{n} W' Y_2 \right) \left( \frac{1}{n} Y_2' P Z Y_2 \right)^{-1} \left( \frac{1}{n} Y_2' W \right) = \Sigma_{wv} + o_p(1)
$$
under both the null and the alternative. We may therefore write that
\[ N = (Z_{\omega} (\rho_s - \Pi_s \tau) + Z_{uv} (\rho_v - \tau) + Z_{we} - \Sigma_{ww} C \tau)' \Sigma_{ww}^{-1} \times (Z_{\omega} (\rho_s - \Pi_s \tau) + Z_{uv} (\rho_v - \tau) + Z_{we} - \Sigma_{ww} C \tau) + o_p(1). \]

Also, under the alternative, by Lemma 4(e),
\[ \hat{\sigma}_\varepsilon^2 \rightarrow_p \sigma^2_{**}, \]
where
\[ \sigma^2_{**} = (\rho_s - \Pi_s \tau)' \Sigma_{\omega} (\rho_s - \Pi_s \tau) + (\rho_v - \tau)' \Sigma_{uv} (\rho_v - \tau) + \sigma^2_v. \]

Now let
\[ Z = \sigma^{-1}_{**} \Sigma_{ww}^{-1/2} (Z_{\omega} (\rho_s - \Pi_s \tau) + Z_{uv} (\rho_v - \tau) + Z_{we}), \]
and
\[ \kappa = -\sigma^{-1}_{**} \Sigma_{ww}^{1/2} C \tau. \]
Then, it is easy to see that \( Z \sim N(0, I_K) \). By writing
\[ \mathcal{H} (\hat{\sigma}_\varepsilon^2) = \frac{N}{\hat{\sigma}_\varepsilon^2} = (\kappa + Z)' (\kappa + Z) + o_p(1), \]
we obtain the desired conclusion. ■

Lemmas 5 and 6 imply that it is important to choose \( \hat{\sigma}_\varepsilon^2 \) such that it is consistent for \( \sigma^2_\varepsilon \) under the null, and consistent for \( \sigma^2_\varepsilon \) under the alternative. To see this, suppose that we use \( \hat{\sigma}_\varepsilon^2 \) in (6) for \( \hat{\sigma}_\varepsilon^2 \). It can be shown\(^{16}\) that \( \hat{\sigma}_\varepsilon^2 \rightarrow_p \sigma^2_\varepsilon \) under the null, but \( \hat{\sigma}_\varepsilon^2 \rightarrow_p (\rho_s - \Pi_s \tau)' \Sigma_{\omega} (\rho_s - \Pi_s \tau) + \sigma^2_v \) under the alternative. In other words, we have \( \sigma^2_\varepsilon / \sigma^2_v \leq 1 \) if we use \( \hat{\sigma}_\varepsilon^2 \). This implies that the asymptotic distribution of \( \mathcal{H} (\hat{\sigma}_\varepsilon^2) \) under the alternative is the mixture of \( \chi^2 \) distributions \( (\kappa (Z_{\omega}) + Z)' (\kappa (Z_{\omega}) + Z) \) multiplied by a constant less than or equal to 1. Therefore, the test \( \mathcal{H} (\hat{\sigma}_\varepsilon^2) \) may be asymptotically biased.\(^{17}\)

Below, we present asymptotic properties of \( \hat{\sigma}_{\varepsilon,z}^2 \) developed in (14). It turns out that if we use \( \hat{\sigma}_{\varepsilon,z}^2 \) as an estimate of \( \sigma^2_\varepsilon \), then the ratio \( \sigma^2_*/\sigma^2_{**} \) converges to 1 under the alternative:

**Lemma 7** Assume Conditions 1 and 2. Under the null, \( \hat{\sigma}_{\varepsilon,z}^2 \rightarrow_p \sigma^2_\varepsilon \), and under the alternative, \( \hat{\sigma}_{\varepsilon,z}^2 \rightarrow_p (\rho_v - \tau)' \Sigma_{uv} (\rho_v - \tau) + \sigma^2_v. \)

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

\(^{16}\)See Lemma 3(f) and Lemma 4(e) in the appendix.

\(^{17}\)Recall that \( \mathcal{H} (\hat{\sigma}_{\varepsilon,z}^2) = \mathcal{H}_3 \) 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.
D.2 Proof of Theorem 2

Proof Part (a) follows by Lemmas 5 and 7 above in Section D.1. Part (b) follows by Lemmas 6 and 7 from Section D.1.

E Computational Issue

For convenience to practitioners, we provide below an alternative algorithm to compute $H_4$ in the special case when the weak instruments $W$ exactly identify the coefficient. The algorithm is based on the characterization (15) in Section 5: For a general over-identified case, see our Online Appendix.

- Compute the 2SLS $\hat{\beta}_z$ by using the instruments $Z = [W, S]$. Let $\hat{V}_z = \hat{\sigma}_{\hat{\varepsilon}, z}^2 (Y_2' P_2 Y_2)^{-1}$ denote the standard variance estimator of $\hat{\beta}_z$, where $\hat{\sigma}_{\hat{\varepsilon}, z}^2 = \frac{1}{n} \hat{\varepsilon}' \hat{\varepsilon}$ denotes the standard estimator of $\sigma_{\varepsilon}^2$.

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

- Let
  \[ \tilde{V}_z = \frac{\hat{\sigma}_{\hat{\varepsilon}, z}^2}{\bar{V}_z}. \]

- Compute the 2SLS estimate $\hat{\beta}_w$ by using the instruments $W$. Let $\hat{V}_w = \hat{\sigma}_{\hat{\varepsilon}, w}^2 (Y_2' P_2 Y_2)^{-1}$ denote the standard variance estimator of $\hat{\beta}_w$, where $\hat{\sigma}_{\hat{\varepsilon}, w}^2 = \frac{1}{n} \hat{\varepsilon}' \hat{\varepsilon}$ denotes the standard estimator of $\sigma_{\varepsilon}^2$.

- Let
  \[ \tilde{V}_w = \frac{\hat{\sigma}_{\hat{\varepsilon}, w}^2}{\bar{V}_w}. \]

- $H_4$ can now be calculated as
  \[ H_4 = (\hat{\beta}_w - \hat{\beta}_z)' [\tilde{V}_w - \tilde{V}_z]^{-1} (\hat{\beta}_w - \hat{\beta}_z). \]
References


