Increasing the Power of Specification Tests

Tiemen Woutersen and Jerry A. Hausman*
University of Arizona and MIT

November 7th, 2017

Abstract. This paper shows how to increase the power of Hausman's (1978) specification test as well as the difference test against fixed alternatives. If the null hypothesis is true then the proposed test has the same distribution as the existing ones in large samples. If the hypothesis is false then the proposed test statistic is larger with probability approaching one as the sample size increases in several important applications, including testing for endogeneity in the linear model.

Preliminary; Do Not Quote

Keywords: Specification test, Hausman test, Power of a test.

JEL Codes: C01, C14, C18, C41

Use v in Hausman 1978: Proposed in H but the only discussion is that it is equivalent. Local asymptotics. Good properties not emphasized.

1. Introduction

Specification tests are important in applied econometrics. Such tests help the researcher to evaluate whether the estimate of a quantity of interest would change if the model is changed. The purpose of this paper is to derive tests that are more powerful than the existing ones in important applications. The main idea is to impose the restrictions of the null *and* alternative hypotheses when estimating the variation of the test statistic. In particular we impose the null and alternative hypotheses when calculating the Hausman (1978) test and other tests that are based on the difference between two estimators. Durbin (1954) and Wu (1973) propose such tests for the linear model while Hausman (1978) considers a more general framework. We argue that a more powerful test against fixed

*Comments are welcome. We thank Gary Solon for helpful discussions. Email woutersen@email.arizona.edu.

alternatives is available under the assumptions made in these papers. These specification tests are very popular in applied economics and econometrics.[1] A reason for the popularity of these specification tests is that one can make a judgement about whether the estimates of the quantity of interest differ in a scientifically significant way in the two economic models as well as whether this difference is statistically significant. Some tests such as score tests only yield whether a difference is statistically significant and are less suitable for economic interpretation.

So many papers use the Hausman (1978) test that we cannot review them all but Zapata et al. (2012) and Adkins et al. (2012) review applications of the Hausman (1978) test. Guggenberger (2010) considers using the Hausman test for pretesting but this paper is concerned with (i) testing whether two estimands are different and (ii) testing whether an efficient estimator differs from a robust one. A related paper is Woutersen (2016), which shows how to improve the power of the Hansen (1982) and Sargan (1958) tests. We extend the techniques of that paper to the Hausman (1978) test and to semi-nonparametric models. This paper is organized as follows: section 2 presents three examples, section 3 gives the theorem and section 4 concludes.

2. Examples

This section presents several examples that show how imposing the restrictions of the null *and* alternative hypotheses can yield a more powerful test.

2.1. Testing for Endogeneity. We first consider the linear model with a binary regressor that is potentially endogenous and a binary instrument. We first assume common coefficients and test for endogeneity. After that we consider a closely related model by Imbens and Angrist (1994) and test whether the average treatment effect equals the local average treatment effect. Suppose we observe a random sample $\{X_i, Y_i, Z_i\}, i = 1, ..., N$,

[1]For example Hausman's (1978) article is cited 4184 times in the Web of Science Core Collection as of the time of this writing; see also Kim, Morse, and Zingales (2006) for an earlier review. Some example of textbooks that review the Hausman (1978) test are Ruud (2000), Cameron and Trivedi (2005), Wooldridge. (2010), and Greene (2012); the test is also reviewed by any other graduate econometric textbook that we checked; Romano and Shaikh, and Wolf (2010) review tests in econometrics.

and that these random variables satisfy the following conditions,

$$
\begin{aligned}
X_i, Z_i &\in \{0,1\}, \\
P(Z_i &= 1) = \frac{1}{3},\ P(X_i = 1|Z_i = 1) = P(X_i = 0|Z_i = 0) = \delta, \\
Y_i &= \alpha + \beta X_i + \varepsilon_i, \\
\varepsilon_i &= \gamma \cdot 1(X_i \neq Z_i) + u_i,\ \text{and } u_i|X_i, Z_i \sim N(0, \sigma_{uu}),
\end{aligned} \tag{1}
$$

where α, β, γ and δ are scalars. If the error term u_i in equation (1) is uncorrelated with the regressor, i.e. if $\gamma = 0$, then there is no endogeneity issue and the least squares estimator (regressing Y on a constant and X) is consistent and asymptotically efficient. The two stage least squares estimator is the 'robust' estimator in Hausman's (1978) terminology and that estimator is consistent no matter what the value of γ is. The Hausman (1978) test statistic is based on the difference between the least squares estimator, β_{LS}, which is the efficient estimator, and the two stage least squares estimator, β_{2SLS}. In particular,

$$
H = N \cdot \frac{(\beta_{2SLS} - \beta_{LS})^2}{\widehat{\sigma^2_{2SLS}} - \widehat{\sigma^2_{LS}}} \tag{2}
$$

where $\widehat{\sigma^2_{2SLS}}$ and $\widehat{\sigma^2_{LS}}$ are consistent estimators of the asymptotic variances of β_{2SLS} and β_{LS}. Writing the Hausman test in terms of observables yields

$$
H = \frac{\left(\frac{\sum_i (Z_i - \bar{Z})Y_i}{\sum_i (Z_i - \bar{Z})X_i} - \frac{\sum_i (X_i - \bar{X})Y_i}{\sum_i (X_i - \bar{X})^2}\right)^2}{\frac{\sum_i e_i^2}{N-2}\left(\frac{\sum_i (Z_i - \bar{Z})^2}{\{\sum_i (Z_i - \bar{Z})X_i\}^2} - \frac{1}{\sum_i (X_i - \bar{X})^2}\right)} \tag{3}
$$

where $e_i, i = 1, .., N$, are the least square residuals from the regression of Y on a constant and X. Note that if $\gamma = 0$ then H has an F-distribution with $\{1, N-2\}$ degrees of freedom. The Hausman test does not impose the condition that the residuals should be uncorrelated with both the regressor *and* the instrument.[2] We believe that this observation has not been made before. We propose to regress Y_i on a constant, the regressor X_i and $1(X_i \neq Z_i)$. Let $w_i, i = 1, ..., N$ denote these residuals of this regression and consider the test statistic

$$
T = \frac{\left(\frac{\sum_i (Z_i - \bar{Z})Y_i}{\sum_i (Z_i - \bar{Z})X_i} - \frac{\sum_i (X_i - \bar{X})Y_i}{\sum_i (X_i - \bar{X})^2}\right)^2}{\frac{\sum_i w_i^2}{N-3}\left(\frac{\sum_i (Z_i - \bar{Z})^2}{\{\sum_i (Z_i - \bar{Z})X_i\}^2} - \frac{1}{\sum_i (X_i - \bar{X})^2}\right)}, \tag{4}
$$

[2] Note that the least square residuals $e_i = \varepsilon_i - \bar{\varepsilon} - \frac{\widehat{\sigma_{X\varepsilon}}}{\widehat{\sigma_{XX}}}(X_i - \bar{X})$, $i = 1, ..., N$, are correlated with the instrument if $\sigma_{XZ} \neq 0$, i.e. if the null hypothesis fails; $\widehat{\sigma_{XX}}$, $\widehat{\sigma_{X\varepsilon}}$, $\bar{\varepsilon}$, and $\bar{X}$ denote the sample variance, sample covariance, average value of the error term and average value of the regressor; σ_{XZ} denotes the covariance between the regressor X and the instrument Z.

which has an F-distribution with $\{1, N-3\}$ degrees of freedom. These two F-distributions, the H and T statistics, converge to the χ^2 distribution with one degree of freedom as N increases. In particular, even for moderate sample sizes like $N = 100$, all percentiles are almost the same. Next, in order to simplify the notation we define $S_i \equiv 1(X_i \neq Z_i)$. We now allow for $\gamma \neq 0$ and consider the ratio of the test statistics and its probability limit,

$$\begin{aligned} \frac{T}{H} &= \frac{\frac{\sum_i e_i^2}{N-2}}{\frac{\sum_i w_i^2}{N-3}}, \qquad (5) \\ plim_{N\to\infty}\frac{T}{H} &= \frac{\sigma_{uu} + \gamma^2(\sigma_{SS} - \frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}}, \end{aligned}$$

where σ_{SS} denotes the variance of S and σ_{SX} denotes the covariance between S and X. Suppose that Z does not perfectly predict X and that $\sigma_{uu} > 0$. If, in addition, $\gamma = 0$, then the probability limit of $\frac{T}{H}$ equals one. However, under the alternative we have that $\gamma \neq 0$, in which case the probability limit of $\frac{T}{H}$ is strictly larger than one. This ratio, $\frac{\sigma_{uu}+\gamma^2(\sigma_{SS}-\frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}}$, is called the relative efficiency or Pitman efficiency since this ratio shows how much more data you need for the Hausman test compared to the proposed test in order for the Hausman test to be as powerful as the proposed test in large samples. Van der Vaart (2000) reviews relative or Pitman efficiency. In particular, the last example uses the fact that, under the alternative hypothesis, the least squares residuals are correlated with the instrument and the two stage least squares residuals are correlated with the regressor.

A closely related way to analyze the difference between the test statistics is by using weak instrument asymptotics. Staiger and Stock (1997) introduce the weak instrument asymptotics to analyze the two stage least squares estimator. In our notation this means that $\delta = \frac{1}{2} + \frac{c}{\sqrt{N}}$ for some c. In this weak instrument asymptotics the Hausman test statistic converges to a random variable that we denote by V. The proposed test statistic converges[3] to $\frac{\sigma_{uu}+\gamma^2(\sigma_{SS}-\frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}}V$ and the ratio of the test statistics, $\frac{T}{H}$, converges to $\frac{\sigma_{uu}+\gamma^2(\sigma_{SS}-\frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}}$. Since $\frac{\sigma_{uu}+\gamma^2(\sigma_{SS}-\frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}} > 1$ if $\gamma \neq 0$, we have that the proposed is more powerful. In particular, if $\gamma = 0$ then both tests have the same χ^2-distribution with one degree of freedom in this weak instrument asymptotics. Thus, the two test statistics

[3] See the appendix for details.

use the same critical value. Under the alternative hypothesis, i.e. $\gamma \neq 0$, the proposed test is more likely to reject since $\frac{\sigma_{uu}+\gamma^2(\sigma_{SS}-\frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}}V > c_{critical}$ is more likely than $V > c_{critical}$ for any $c_{critical}$.

An alternative way to calculate the Hausman test statistic is to use the two stage least squares residuals in equation (2) rather than the least squares residuals. Let $e_{IV,i}$, $i = 1, ..., N$, denote these residuals. We can then replace $\frac{\sum_i e_i^2}{N-2}$ in equation (2) by $\frac{\sum_i e_{IV,i}^2}{N-2}$. Let $H_{\text{IV Residuals}}$ denote the resulting Hausman test statistic. We can then calculate the Pitman efficiency of the new test statistic versus $H_{\text{IV Residuals}}$,

$$plim_{N\rightarrow\infty}\frac{T}{H_{\text{IV Residuals}}} = \frac{\sigma_{uu}+\gamma^2\sigma_{SS}}{\sigma_{uu}}.$$

Note that $\frac{\sigma_{uu}+\gamma^2\sigma_{SS}}{\sigma_{uu}} > \frac{\sigma_{uu}+\gamma^2(\sigma_{SS}-\frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}} > 1$ so that H performs better than $H_{\text{IV Residuals}}$ and the new test statistic has the best relative (or Pitman) efficiency. Hausman (1978) discusses that using the least squares residuals, i.e. using H in our notation, corresponds to the Lagrange Multiplier test while using the two stage least squares residuals, i.e. using $H_{\text{IV Residuals}}$ in our notation, corresponds to the Wald tests.[4] Our proposed test is outside the framework of the Lagrange Multiplier, Wald, or Likelihood ratio tests. Hausman (1978) uses local misspecification, so in his case using the least squares residuals or two stage residuals yields the same asymptotic variance-covariance matrix. However, the weak instrument asymptotics and relative (or Pitman) efficiency shows the difference between the different approaches.

Imbens and Angrist (1994) show that the two stage least squares estimator estimates the local average treatment effect when the endogenous regressor and instruments are dummies. Equation (1) can be interpreted as a random coefficients model with negative correlation between the coefficients on the intercept and the slope coefficients. We will later discuss tools to apply the Hausman test to more general random coefficients models. The null hypothesis in this case is that the average treatment effect in the restrictive random coefficients model equals the local average treatment effect. A typical assumption concerning random coefficients is that these random coefficients do not depend on the regressor or the instrument. The proposed method implements the restriction that the

[4] Engle (1984) and Lehmann and Romano (2005) give very clear overviews of the Lagrange Multiplier and the Wald tests.

random coefficients are random by using the residuals from regressing Y_i on a constant, the regressor X_i and $1(X_i \neq Z_i)$. This yields a more powerful test than the Hausman (1978) test. It is noteworthy that observing $1(X_i \neq Z_i)$ identifies a 'non-complier' while there will typically also be non-compliers for which $X_i \neq Z_i$ is not observed (e.g. for some 'always takers' or 'never takers' we may have $X_i = Z_i$). Thus, the relative efficiency gain can be realized without having to observe the 'type' of an individual in the local average treatment effect model. Also, instead of regressing Y_i on a constant, the regressor X_i and $1(X_i \neq Z_i)$, one could regress Y_i on a constant, the regressor X_i, $1(X_i > Z_i)$, and $1(X_i < Z_i)$. In that case the finite sample F-distribution discussed above has $\{1, N-4\}$ instead of $\{1,N-3\}$ degrees of freedom. Using $1(X_i > Z_i)$, and $1(X_i < Z_i)$ as regressors also implements the restriction that the residuals or random effects are uncorrelated with the regressor and the instrument. The simulations below show that the proposed test improves on the Hausman test. These simulations use the following values for the parameters in equation (1): $\alpha = 0$, $\beta = \gamma = \sigma_{uu} = 1$; the larger values of δ correspond to a stronger instrument.

Table 1: 0.05 Rejection Frequencies

N	δ	*New Test*	*LS*	*2SLS*	*Pitman New Test*	*Pitman LS*	*Pitman 2SLS*
1000	0.60	0.203	0.157	0.132	1.214	1	0.955
2000	0.60	0.311	0.252	0.237	1.214	1	0.955
4000	0.60	0.495	0.427	0.418	1.214	1	0.955
1000	0.70	0.458	0.399	0.391	1.190	1	0.977
2000	0.70	0.703	0.649	0.645	1.190	1	0.977
4000	0.70	0.926	0.903	0.902	1.190	1	0.977

Results based on 100,000 simulations; $\alpha = 0$, $\beta = 1$, and plim$(\beta_{LS}) \approx 1.285$

In table 1 'Pitman New Test' stands for the Pitman efficiency ratio of the new test and the Pitman efficiency ratio of the Hausman test with least squares residuals normalized to be one. The new test has a higher Pitman efficiency than the existing test. The size of the tests in table 1 is 5%, i.e. the critical value is such that the probability of falsely rejecting the null hypothesis is 5%. In table 2 the size of the tests is reduced to 1% and we see the same pattern as in table 1. That is, the endogeneity bias of the least squares estimator is detected more frequently.

Table 2: 0.01 Rejection Frequencies

N	δ	$New\ Test$	LS	$2SLS$	$Pitman\ New\ Test$	$Pitman\ LS$	$Pitman\ 2SLS$
1000	0.60	0.0849	0.0536	0.0333	1.214	1	0.955
2000	0.60	0.152	0.105	0.0867	1.214	1	0.955
4000	0.60	0.295	0.223	0.206	1.214	1	0.955
1000	0.70	0.267	0.205	0.193	1.190	1	0.977
2000	0.70	0.505	0.426	0.417	1.190	1	0.977
4000	0.70	0.825	0.769	0.765	1.190	1	0.977

Results based on 100,000 simulations; $\alpha = 0$, $\beta = 1$, and plim$(\beta_{LS}) \approx 1.285$

Table 3 uses the Pitman efficiency ratio to choose the sample sizes. In particular, we compare the new test to the existing tests but use fewer observations for the new test than for the existing ones (e.g. in the first line we use 824 for the new test and 1000 and 1047 for the existing tests). The average rejecting probabilities are now very close which shows that the Pitman efficiency ratio captures the power gains of the new test.

Table 3: 0.05 Rejection Frequencies using $N = N^*/Pitman\ Efficiency$

N^*	δ	$New\ Test : N$	$New\ Test$	$LS : N$	LS	$2SLS : N$	$2SLS$
1000	0.60	824	0.186	1000	0.157	1047	0.140
2000	0.60	1648	0.277	2000	0.252	2094	0.245
4000	0.60	3296	0.437	4000	0.427	4188	0.437
1000	0.70	840	0.409	1000	0.399	1024	0.398
2000	0.70	1680	0.639	2000	0.649	2048	0.655
4000	0.70	3360	0.883	4000	0.903	4096	0.908

Results based on 100,000 simulations; $\alpha = 0$, $\beta = 1$, and plim$(\beta_{LS}) \approx 1.285$

We summarize this subsection by observing that the Hausman test for endogeneity has the feature that, under the alternative hypothesis, the least squares residuals are correlated with the instrument and the two stage least squares residuals are correlated with the regressor. These simulations show that imposing the conditions that the residuals are uncorrelated with the regressor and the instrument yields a more powerful test.

2.2. Panel Data. The initial conditions problem is an important issue in panel data econometrics, see for example Wooldridge (2010). The following example illustrates this issue and shows how to apply our method. Here we consider a linear panel data model. The first estimator is the least squares estimator and this estimator is consistent under the null hypothesis that all the unobservables are randomly distributed. The second estimator is the fixed effect least squares estimator. This estimator allows for time invariant heterogeneity that is correlated with the regressors. Thus, the fixed effect least squares estimator is consistent under the alternative hypothesis that allows for such correlation.

In the last example, we used the fact that under the alternative hypothesis, the least squares residuals are correlated with the instrument and that the two stage least squares residuals are correlated with the regressor. Here we use that the residuals of the least squares (or random effects) estimator are correlated with the regressors of the initial period under the alternative. Thus, we can reduce the variation of the residuals by using this correlation under the alternative.

We now specify the data generating process for the panel data example. Suppose we observe the regressors X_{it} and Q_{it} as well as the outcome Y_{it} for N individuals, $i = 1, ..., N$, and two periods, $t = 1, 2$. Let

$$\begin{aligned} Y_{it} &= \alpha + \beta X_{it} + \gamma Q_{it} + \varepsilon_{it} \\ \varepsilon_{it} &= \delta Q_{i1} + \kappa u_{it}. \end{aligned} \tag{6}$$

where u_{it} is an unobserved error term. If the parameter $\gamma \neq 0$ then we have a so called initial conditions problem. Suppose we are interested in the parameter β. Also, let the regressors be independently distributed across individuals and time periods but allow for contemporaneous correlation. In particular, let

$$\begin{pmatrix} X_{it} \\ Q_{it} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{regressor} \\ \rho_{regressor} & 1 \end{pmatrix} \right),$$

where $i = 1, ..., N$, and $t = 1, 2$. Let the error terms $\{u_{i1}, u_{i2}\}$ be jointly normally distributed and independently distributed across individuals,

$$\begin{pmatrix} u_{i1} \\ u_{i1} \end{pmatrix} | X_{i1}, X_{i2}, Q_{i1}, Q_{i2} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \tag{7}$$

where $i = 1, ..., N$. Let β_{FE} and γ_{FE} denote the fixed effect estimators of the slope coefficients β and γ. We thus can construct the residual $\hat{u}_{it} = Y_{it} - \beta_{FE} X_{it} - \gamma_{FE} Q_{it}$. Under the null hypothesis of random effects, this residual is uncorrelated with the regressors as is made explicit in the last equation. We can thus regresses $\hat{u}_{it}$ on a constant, X_{i1} and Q_{i1}. Let $\tilde{u}_{it}$ denote the residual from this regression. We propose to use $\tilde{u}_{it}$ to construct the variance estimator for the difference between the least squares estimator and fixed effects least squares estimator of β. We use the following parameter values for our simulation, $\alpha = \gamma = 0$, $\beta = \delta = 1$, $\kappa = 0.1$, $\rho_{regressor} = 0.1$, $\rho = 0$ (in some simulations) $\rho = 0.2$ (in other simulations). The simulations below show that the proposed test is more powerful than the Hausman test. For both tables we use clustered and heteroscedasticity robust standard errors and the null hypothesis is violated in every case so that it is desirable to have high rejection frequencies.

Table 4: 0.05 Rejection Frequencies

N	ρ	*New Test*	*Difference Test*	*Pitman New Test*	*Pitman Diff. Test*
1000	0	0.237	0.0663	1.961	1
2000	0	0.430	0.166	1.961	1
4000	0	0.713	0.409	1.961	1
1000	0.20	0.240	0.0668	1.964	1
2000	0.20	0.430	0.166	1.964	1
4000	0.20	0.714	0.411	1.964	1

Results based on 100,000 simulations

In table 4 the Pitman efficiency ratio of the Hausman test with least squares residuals is normalized to be one. The new test has a higher Pitman efficiency than the existing test. The size of the tests in table 4 is 5%. In table 5 the size of the tests is reduced to 1% and we see the same pattern as in table 4. That is, the bias of the least squares estimator that is caused by the initial condition problem is detected more frequently.

Table 5: 0.01 Rejection Frequencies

N	ρ	*New Test*	*Difference Test*	*Pitman New Test*	*Pitman Diff. Test*
1000	0	0.0914	0.00881	1.961	1
2000	0	0.211	0.0334	1.961	1
4000	0	0.476	0.136	1.961	1
1000	0.20	0.0928	0.00906	1.964	1
2000	0.20	0.213	0.0341	1.964	1
4000	0.20	0.479	0.138	1.964	1

Results based on 100,000 simulations

These simulations show that imposing the conditions that the residuals are uncorrelated with the initial conditions yields a more powerful test. Next we consider a semi-nonparametric model and we then generalize our results in a theorem.

2.3. Semi-Nonparametric Model. Hausman and Woutersen (2014a) use the maximum rank correlation estimator to estimate a semi-nonparametric hazard model. In that paper we propose a test for detecting the effect of misspecification on the parameter estimates. However, we were at the time not aware of the techniques described in this paper.

We now revisit our test for the semi-nonparametric hazard model and find that we can make the test somewhat more powerful. Consider the following transition or hazard rate

$$\theta(t|v,X) = (\frac{1}{\kappa})^X \lambda(t)$$

where $\lambda(t)$ is a nonparametric function of time. Suppose that $X \in \{0,1\}$ is a treatment that may affect the transition rate such as a bonus for finding a job quickly in the context of an unemployment duration model. Suppose that the applied researcher is interested in the effect of X on the job finding rate of women and has data on the unemployment duration of a group of women and a group of men. The question then is whether to pool the data or not. If the model $(\frac{1}{\kappa})^X \lambda(t)$ holds for both groups then pooling gives an estimate of the parameter κ that is more precise. This parameter κ can then be used to construct counterfactuals as is discussed in Hausman and Woutersen (2014b).

Many semi-nonparametric or semiparametric estimators, including those in Hausman and Woutersen (2014a), have score functions that are similar to the residuals discussed in the last two examples. In the last two examples we showed how imposing restrictions on these residuals yields a more powerful test. One can do the same with the score functions of semi-nonparametric or semiparametric models. For example, one can regress the contributions of each individual to the score function on a constant and variables that should be uncorrelated with these score function contributions. If the regressors are discrete, as in the duration example above, then one can estimate the nonparametric function for different values of the regressors. We consider this approach here since we already discussed score functions in the earlier examples.

We now specify the data generating process of this example. Let the treatment be binary, i.e. $X \in \{0,1\}$, and suppose we observe two groups, i.e. $Q \in \{0,1\}$. Let X and Q be statistically independently distributed and let $P(X=1) = P(Q=1) = \frac{1}{2}$. Let T denote the duration and suppose that we observe a random sample of $\{T_i, Q_i, X_i\}$ where $i = 1,..,N$. Let $T_i|X_i, Q_i$ be exponentially distributed with mean $1/\exp(X_i + \gamma Q_i)$. If the parameter γ is zero, then pooling the groups for which $Q=0$ and $Q=1$ is a good idea. However, if $\gamma \neq 0$ then pooling causes the estimator for the effect of X to be inconsistent. This differs from the linear model where ignoring an uncorrelated regressor does *not* cause the estimators for the other regressors to be inconsistent. We consider $\gamma = 0$, $\gamma = 1$, and $\gamma = 2$. We estimate several conditional distribution functions:

(a) Let $G_{00}(t)$ denote the empirical distribution function (= edf) of the durations $T_1, ..., T_N$ conditional on $X = 0$ and $Q = 0$, i.e. $G_{00}(t) = \frac{\sum_{i|X_i=0,Q_i=0} 1(T_i \leq t)}{\sum_{i|X_i=0,Q_i=0} 1}$. Note that the denominator is just the number of individuals for which $X = 0$ and $Q = 0$.

(b) Let $G_{X=0}(t)$ denote the edf of the durations $T_1, ..., T_N$ conditional on $X = 0$, i.e. $G_{X=0}(t) = \frac{\sum_{i|X_i=0} 1(T_i \leq t)}{\sum_{i|X_i=0} 1}$.

We use the following estimators for κ. The first estimator uses all the data, $\kappa_1 \equiv -[\frac{\sum_{i|X_i=1} \ln\{1-G_{X=0}(T_i)\}}{\sum_{i|X_i=1} 1}]$. The second estimator only uses data of individuals for which $Q = 0$, i.e. $\kappa_2 \equiv -[\frac{\sum_{i|X_i=1,Q_i=0} \ln\{1-G_{00}(T_i)\}}{\sum_{i|X_i=1,Q_i=0} 1}]$.

We generated 1,000 datasets and calculated confidence intervals for $\kappa_1 - \kappa_2$ in the following ways. We sample with replacement to generate 1,000 bootstrap data sets for each data set. We calculate $\kappa_1 - \kappa_2$ for each bootstrap dataset, i.e. calculate $\kappa_{1j} - \kappa_{2j}$ where $j = 1, ..., 1,000$.

(i) Percentile bootstrap: Consider $\kappa_{1j} - \kappa_{2j}$ where $j = 1, ..., 1,000$ and order these differences. Calculate the 95% (or 99%) confidence interval by removing the smallest and largest 2.5% (or 0.5%). Accept the null hypothesis that $\text{plim}(\kappa_1 - \kappa_2) = 0$ if zero is inside the confidence interval.

(ii) Regular bootstrap: Calculate the average value of the bootstrap sample, i.e. $\bar{\kappa}_1 - \bar{\kappa}_2 = \frac{1}{M}\{\sum_j \kappa_{1j} - \kappa_{2j}\}$. Then calculate the variation of $\kappa_1 - \kappa_2$ using all these bootstrap data sets, i.e. $var\widehat{\{\kappa_1 - \kappa_2\}} = \frac{1}{1000}\sum_j \{\kappa_{1j} - \kappa_{2j} - (\bar{\kappa}_1 - \bar{\kappa}_2)\}^2$. Calculate the t-statistic $\frac{\kappa_1 - \kappa_2}{\sqrt{var\widehat{\{\kappa_1 - \kappa_2\}}}}$ and reject if $\frac{\kappa_1 - \kappa_2}{var\widehat{\{\kappa_1 - \kappa_2\}}}$ is larger than 1.96 (or 2.57) in absolute value.

(iii) Impose the restriction that the data generating process does not depend on Q. We do this by adjusting the estimate of the variation. Define

$$\triangle_1 = \frac{1}{\sum_{i|X_i=1} 1}\left(\frac{\sum_{i|X_i=1,Q_i=0} 1}{\sum_{i|X_i=1} 1 - \sum_{i|X_i=1,Q_i=0} 1}\right)(\kappa_1 - \kappa_2)^2,$$

and

$$var\widetilde{\{\kappa_1 - \kappa_2\}} = var\widehat{\{\kappa_1 - \kappa_2\}} - \triangle_1.$$

We calculate the t-statistic $\frac{\kappa_1 - \kappa_2}{\sqrt{var\widetilde{\{\kappa_1 - \kappa_2\}}}}$ and reject if $\frac{\kappa_1 - \kappa_2}{\sqrt{var\widetilde{\{\kappa_1 - \kappa_2\}}}}$ is larger than 1.96 (or

2.57) in absolute value.

(iv) Impose the restriction that the data generating process does not depend on Q. We now do this by adjusting the estimate of the standard error that is based on the length of the percentile bootstrap confidence interval.

$$\overline{var\{\kappa_1 - \kappa_2\}} = \{\frac{\text{Length 95\% Percentile Bootstrap Confidence Interval } \kappa_1 - \kappa_2}{2 * 1.96}\}^2 - \triangle_1$$

where the correction to the variation, $\triangle_1$, is the same as above. We calculate the t-statistic $\frac{\kappa_1-\kappa_2}{\sqrt{\overline{var\{\kappa_1-\kappa_2\}}}}$ and reject if $\frac{\kappa_1-\kappa_2}{\sqrt{\overline{var\{\kappa_1-\kappa_2\}}}}$ is larger than 1.96 (or 2.57) in absolute value. Our findings are summarized in the following tables. The abbreviation "Rej. Freq." stands for rejection frequency, i.e. the size of the test.

Table 6: Correct Specification: $E(T_i|X_i, Q_i) = \exp(-X_i)$

N	Rej. Freq.	(i) Perc. Bootstrap	(ii) Regular Bootstrap	(iii) New Test I	(iv) New Test II
500	0.01	0.010	0.005	0.005	0.005
1000	0.01	0.012	0.008	0.008	0.010
2000	0.01	0.011	0.010	0.010	0.010
500	0.05	0.055	0.048	0.048	0.049
1000	0.05	0.050	0.050	0.050	0.048
2000	0.05	0.049	0.046	0.046	0.047

Results based on 1,000 bootstrap simulations on 1,000 data sets

Table 7: Misspecification I: $E(T_i|X_i, Q_i) = \exp(-X_i - Q_i)$

N	Rej. Freq.	(i) Perc. Bootstrap	(ii) Regular Bootstrap	(iii) New Test I	(iv) New Test II
500	0.01	0.113	0.116	0.124	0.121
1000	0.01	0.256	0.269	0.276	0.278
2000	0.01	0.500	0.530	0.535	0.538
500	0.05	0.242	0.240	0.244	0.252
1000	0.05	0.478	0.479	0.483	0.496
2000	0.05	0.749	0.760	0.761	0.767

Results based on 1,000 bootstrap simulations on 1,000 data sets

Table 8: Misspecification II: $E(T_i|X_i, Q_i) = \exp(-X_i - 2Q_i)$

N	Rej. Freq.	(i) Perc. Bootstrap	(ii) Regular Bootstrap	(iii) New Test I	(iv) New Test II
500	0.01	0.680	0.664	0.676	0.717
1000	0.01	0.958	0.942	0.944	0.964
2000	0.01	0.996	0.996	0.996	0.999
500	0.05	0.841	0.784	0.787	0.847
1000	0.05	0.997	0.968	0.968	0.998
2000	0.05	1.000	0.990	0.990	1.000

Results based on 1,000 bootstrap simulations on 1,000 data sets

The simulations show that imposing restrictions yields a somewhat more powerful test. In the next section we generalize the findings of the three examples of this section.

3. Theorem

The last section gave several examples where the power of the Hausman (1978) test could be improved. We now generalize these examples and state our theorem. Consider two information sets (or sets of assumptions) that are denoted by $A1$ and $A2$. Let the information sets $A1$ and $A2$ imply consistency of the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ respectively. For example, the information set $A1$ implies that the least squares estimator is consistent while $A2$ implies that the two stage least squares estimator is consistent. Earlier we saw that $A2$ may be informative about $\hat{\theta}_1$ when the null hypothesis is violated. In particular, the least squares residuals are correlated with the instrument in example 1 if the null hypothesis is violated. Similarly, $A1$ may be informative about $\hat{\theta}_2$ when the null hypothesis is violated. Using these restrictions can yield a more powerful test. Formally, let the variation of the difference between the two estimators be denoted by $Var\{(\hat{\theta}_1 - \hat{\theta}_2)\}$. Imposing the restrictions[5] implied by $A1$ and $A2$ yields a variation that can be denoted by $Var\{(\hat{\theta}_1 - \hat{\theta}_2)|A1, A2\}$. Since these restrictions do not increase the variation we have that

$$Var\{(\hat{\theta}_1 - \hat{\theta}_2)|A1, A2\} \leq Var\{(\hat{\theta}_1 - \hat{\theta}_2)\} \tag{8}$$

in the sense that the difference between these two matrices is positive semi-definite. Note that the last equation holds even if we use some but not all of the restrictions. Define $\Omega_{\text{Restr}} = Var\{(\hat{\theta}_1 - \hat{\theta}_2)|A1, A2\}$ and $\Omega = Var\{(\hat{\theta}_1 - \hat{\theta}_2)\}$. We can now define the test statistics. In particular, we first consider an infeasible test statistic that uses Ω and then consider a feasible one that uses $\hat{\Omega}$, a consistent estimator of Ω. Define

$$T_{New,\text{infeasible}} \equiv (\hat{\theta}_1 - \hat{\theta}_2)'\Omega_{\text{Restr}}^{-1}(\hat{\theta}_1 - \hat{\theta}_2) \tag{9}$$

$$T_{H,\text{infeasible}} \equiv (\hat{\theta}_1 - \hat{\theta}_2)'\Omega^{-1}(\hat{\theta}_1 - \hat{\theta}_2) \tag{10}$$

$$T_{New} \equiv (\hat{\theta}_1 - \hat{\theta}_2)'\hat{\Omega}_{\text{Restr}}^{-1}(\hat{\theta}_1 - \hat{\theta}_2) \tag{11}$$

$$T_H \equiv (\hat{\theta}_1 - \hat{\theta}_2)'\hat{\Omega}^{-1}(\hat{\theta}_1 - \hat{\theta}_2). \tag{12}$$

The theorem that follows states that $T_{New,\text{infeasible}}$ is at least as powerful as $T_{H,\text{infeasible}}$ and that T_{New} is larger than T_H in large samples if the null hypothesis is violated. Finally,

[5] The following theorem is more general but what we have in mind for this paper is imposing restrictions that, asymptotically, do not change the distribution of test statistic under the null; examples that satisfy this were presented in the previous section.

the theorem states that $T_{H,\text{infeasible}}$ and T_{New} have the same asymptotic distribution under the null.

Theorem

Let the second moments of the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ exist. Let $\Omega = Var\{(\hat{\theta}_1 - \hat{\theta}_2)\}$ and $\Omega_{\text{Restr}} = Var\{(\hat{\theta}_1 - \hat{\theta}_2)|A1, A2\}$, where Ω and Ω_{Restr} are positive definite. Let $\hat{\Omega} = \Omega + o_p(1)$ and $\hat{\Omega}_{\text{Restr}} = \Omega_{\text{Restr}} + o_p(1)$. Then

(i) $T_{New,\text{infeasible}} = (\hat{\theta}_1 - \hat{\theta}_2)'\Omega^{-1}_{\text{Restr}}(\hat{\theta}_1 - \hat{\theta}_2) \geq (\hat{\theta}_1 - \hat{\theta}_2)'\Omega^{-1}(\hat{\theta}_1 - \hat{\theta}_2) = T_{H,\text{infeasible}}$.

(ii) Moreover, if the conditioning reduces the variation, i.e. $\Omega = \Omega_{\text{Restr}} + M_1$ for some matrix M_1 that is positive definite and $\text{plim}(\hat{\theta}_1) \neq \text{plim}(\hat{\theta}_2)$, then

$\frac{T_{New}}{T_H} > 1$ with probability approaching one where $T_{New} = (\hat{\theta}_1 - \hat{\theta}_2)'\hat{\Omega}^{-1}_{\text{Restr}}(\hat{\theta}_1 - \hat{\theta}_2)$ and $T_H = (\hat{\theta}_1 - \hat{\theta}_2)'\hat{\Omega}^{-1}(\hat{\theta}_1 - \hat{\theta}_2)$.

(iii) If $\hat{\theta}_1 - \hat{\theta}_2$ is normally distributed with mean zero and variance $\Omega = \Omega_{\text{Restr}}$, then T_{New} and T_H are distributed as a χ^2-distribution with $\dim(\theta)$ degrees of freedom.

Proof: See appendix.

Remark 1: The theorem can also be applied to estimators that are asymptotically normally distributed. In these cases, the leading stochastic term is normally distributed, i.e. $\hat{\theta}_1 = \tau_1 + \frac{1}{\sqrt{N}}\eta_1 + o_p(\frac{1}{\sqrt{N}})$ and $\hat{\theta}_2 = \tau_2 + \frac{1}{\sqrt{N}}\eta_2 + o_p(\frac{1}{\sqrt{N}})$ where N is the sample size and the variation of $(\eta_1 - \eta_2)$ is Ω. In this case imposing the restrictions $A1$ and $A2$ may decrease the variation Ω.

Remark 2: For many estimators that use i.i.d. data, the estimator for the asymptotic variance-covariance matrix has the term $\hat{\Sigma} = \frac{1}{N}\sum_i h(X_i, Y_i, Z_i; \hat{\theta})h(X_i, Y_i, Z_i; \hat{\theta})'$ for some function $h(X_i, Y_i, Z_i; \hat{\theta})$ where we observe $\{X_i, Y_i, Z_i\}$, $i = 1, ..., N$. See for example Newey and McFadden (1994, section 3). If $E\{h(X_i, Y_i, Z_i; \theta_0|X_i, Z_i) = 0$ under the null hypothesis then one can regress $h(X_i, Y_i, Z_i; \hat{\theta})$ on a constant, X_i and Z_i (or regress $h(X_i, Y_i, Z_i; \hat{\theta})$ on functions of these variables), $i = 1, ..., N$. Let $\tilde{h}(X_i, Y_i, Z_i; \hat{\theta})$ denote the residual from this regression and let

$\hat{\Sigma}_{\text{Restricted}} = \frac{1}{N}\sum_i \tilde{h}(X_i, Y_i, Z_i; \hat{\theta})\tilde{h}(X_i, Y_i, Z_i; \hat{\theta})'$. Then $\hat{\Sigma}_{\text{Restricted}} \leq \hat{\Sigma}$ in the sense that the difference is positive semi-definite.

Remark 3: The theorem above also applies to subset inference. The Hausman (1978) test is a convenient way to do subset inference (i.e. test the parameters that you care about) and the new test has this advantage as well.

Remark 4: Hahn et al. (2011) consider the linear model with endogeneity and assume that the applied researcher has a weak set of instruments that is valid and a strong set of instruments that is invalid. They then use the local asymptotics that was introduced by Staiger and Stock (1997) and also regress the residuals on the instruments. Thus, our theorem can also be viewed as a generalization of that approach.

Remark 5: Hausman (1978) considers two estimators. An efficient one, say $\hat{\theta}_1$, and a robust one, say $\hat{\theta}_2$. The assumptions associated with $\hat{\theta}_2$ do not reduce the asymptotic variation of the efficient estimator under the null. The examples and theorem in this paper show that it can be worthwhile to use the assumptions associated with $\hat{\theta}_1$ and $\hat{\theta}_2$ to reduce the asymptotic variation of $(\hat{\theta}_1 - \hat{\theta}_2)$ under the alternative.

4. Conclusion

This paper shows how to increase the power of Hausman's (1978) specification test, as well as the difference test, against fixed alternatives. If the null hypothesis is true then the proposed test has the same distribution as the existing one in large samples. If the hypothesis is false then the proposed test statistic is larger with probability approaching one as the sample size increases in several important applications, including testing for endogeneity in the linear model. As the Hausman (1978) test is very popular in empirical work, we expect the current results to be useful as well.

References

[1] Abadir, K. M. and J. R. Magnus (2005): *Matrix Algebra, Econometric Exercises 1*, Cambridge University Press, Cambridge.

[2] Adkins, L.C., R. C. Campbell, V. Chmelarova, R. C. Hill (2012): "The Hausman Test, and Some Alternatives, with Heteroskedastic Data," in *Advances in Econometrics: Essays in Honor of Jerry Hausman*, edited by B. H. Baltagi, R. C. Hill, W. K. Newey, H. L. White, Volume 29, Emerald Group Publishing, UK, 515-546.

[3] Cameron, A. C. and P. K. Trivedi (2005): *Microeconometrics, Methods and Applications*, Cambridge University Press, Cambridge.

[4] Durbin, James (1954). "Errors in variables," *Review of the International Statistical Institute,* 22: 23–32.

[5] Engle, R. F., (1984): "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics," in: Z. Griliches & M. D. Intriligator (ed.), Handbook of Econometrics, edition 1, volume 2, chapter 13: 775-826. North-Holland, Amsterdam.

[6] Greene, W. (2012): *Econometric Analyses*, 7th edition, Prentice Hall, New York.

[7] Guggenberger, P. (2010): "The Impact of a Hausman Pretest on the Asymptotic Size of a Hypothesis Test", *Econometric Theory* 26:369-382.

[8] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica,* 50: 1029-1054.

[9] Hahn, J., J. C. Ham, and H. R. Moon (2011): "The Hausman test and weak instruments", *Journal of Econometrics*, 160:289–299

[10] Hausman, J.A., (1978): "Specification Tests in Econometrics", *Econometrica*, 46, 1251-1272.

[11] Hausman J.A., and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, vol. 49(6), 1377-1398.

[12] Hausman, J. A. and T. Woutersen (2014a): "Estimating a Semi-Parametric Duration Model without Specifying Heterogeneity," *Journal of Econometrics*, 178: 114–131.

[13] Hausman, J. A. and T. Woutersen (2014b): "Estimating the Derivative Function and Counterfactuals in Duration Models with Heterogeneity," *Econometric Reviews*, 33: 472-496.

[14] Holly, A. (1982): "A Remark on Hausman's Specification Test," *Econometrica*, 50: 749-759.

[15] Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62: 467-475.

[16] Kim, E. H. , A. Morse, and L. Zingales (2006): "What Has Mattered to Economics since 1970," *Journal of Economic Perspectives*, 20(4): 189-202.

[17] Lehmann, E. L. and J. P. Romano (2005): *Testing Statistical Hypotheses*, Third Edition, Springer-Verlag, New York.

[18] Newey, W. K. (1985a): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53: 1047-1070.

[19] Newey, W. K. (1985b): "Generalized Method of Moments Specification Testing," *Journal of Econometrics,* 29: 229-256.

[20] Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in the *Handbook of Econometrics,* Vol. 4, ed. by R. F. Engle and D. MacFadden. North-Holland, Amsterdam.

[21] Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55: 703–708.

[22] Romano, J. P., A. M. Shaikh,and M. Wolf (2010): "Hypothesis Testing in Econometrics," *Annual Review of Economics,* 75-104.

[23] Ruud, Paul A. (2000): *An Introduction to Classical Econometric Theory.* New York: Oxford University Press. pp. 578–585.

[24] Sargan, J.D. (1958): "The estimation of economic relationships using instrumental variables," *Econometrica,* 26:393-415.

[25] Staiger, D., and Stock, J. (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65:557-586.

[26] Van der Vaart, A. W. (2000): *Asymptotic Statistics,* Cambridge University Press: Cambridge.Romano, J. P. and A. M. Shaikh, and M. Wolf (2010): "Hypothesis Testing in Econometrics," *Annual Review of Economics*, 75–104.

[27] Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70: 331-341.

[28] Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data,* 2nd edition, MIT Press, Cambridge MA.

[29] Woutersen, T. (2016): "Increasing the Power of Moment-based Tests," manuscript.

[30] Wu, De-Min (July 1973): "Alternative Tests of Independence between Stochastic Regressors and Disturbances". *Econometrica* 41:733–750.

[31] Zapata, H. O., and C. M. Caminita (2012): "The Diffusion of Hausman's Econometric Ideas," in *Advances in Econometrics: Essays in Honor of Jerry Hausman*, edited by B. H. Baltagi, R. C. Hill, W. K. Newey, H. L. White, Volume 29, Emerald Group Publishing, UK, 1-29.

Appendix 1: Weak Instrument Asymptotics in example 1

The Hausman test statistic, H, is stated in the text,

$$H = \frac{\left(\frac{\sum_i (Z_i-\bar{Z})Y_i}{\sum_i (Z_i-\bar{Z})X_i} - \frac{\sum_i (X_i-\bar{X})Y_i}{\sum_i (X_i-\bar{X})^2}\right)^2}{\frac{\sum_i e_i^2}{N-2}\left(\frac{\sum_i (Z_i-\bar{Z})^2}{\{\sum_i (Z_i-\bar{Z})X_i\}^2} - \frac{1}{\sum_i (X_i-\bar{X})^2}\right)}.$$

Using $\delta = \frac{1}{2} + \frac{c}{\sqrt{N}}$ implements the weak instrument asymptotics that was introduced by Staiger and Stock (1997). The first claim in the main text is that H converges to a random variable (that can be denoted by V) in this asymptotics. Note that the least squares estimator is not a function of the instrument and is not affected by the weak instruments asymptotics. Thus, under the assumptions in section 2.1, the least squares estimator $\frac{\sum_i (X_i-\bar{X})Y_i}{\sum_i (X_i-\bar{X})^2}$ converges in probability to $\beta + \frac{\sigma_{X\varepsilon}}{\sigma_{XX}}$ and $\frac{\sum_i e_i^2}{N-2}$ converges in probability to $\sigma_{\varepsilon\varepsilon} - \frac{\sigma_{X\varepsilon}^2}{\sigma_{XX}}$. Moreover, $\frac{N}{\sum_i (X_i-\bar{X})^2}$, $\frac{\sum_i (Z_i-\bar{Z})^2}{N}$, and $\bar{Z}$ converge to $\frac{1}{\sigma_{XX}}$, σ_{ZZ}, and $\mu_Z = E(Z)$ respectively. This gives

$$H = \frac{\{\beta + \frac{\sum_i (Z_i-\mu_Z)\varepsilon_i}{\sum_i (Z_i-\mu_Z)X_i} - (\beta + \frac{\sigma_{X\varepsilon}}{\sigma_{XX}})\}^2}{(\sigma_{\varepsilon\varepsilon} - \frac{\sigma_{X\varepsilon}^2}{\sigma_{XX}})(\frac{N\cdot\sigma_{ZZ}}{\{\sum_i (Z_i-\mu_Z)X_i\}^2} - \frac{1}{\sigma_{XX}})} + o_p(1).$$

Note that $\frac{\sum_i (Z_i-\mu_Z)\varepsilon_i}{\sum_i (Z_i-\mu_Z)X_i}$ and $\frac{\sigma_{ZZ}}{\{\sum_i (Z_i-\mu_Z)X_i\}^2}$ converge to random variables and so does H. In the text this random variable is denoted by V. Next, note that

$$\begin{aligned} T &= \frac{\left(\frac{\sum_i (Z_i-\bar{Z})Y_i}{\sum_i (Z_i-\bar{Z})X_i} - \frac{\sum_i (X_i-\bar{X})Y_i}{\sum_i (X_i-\bar{X})^2}\right)^2}{\frac{\sum_i w_i^2}{N-3}\left(\frac{\sum_i (Z_i-\bar{Z})^2}{\{\sum_i (Z_i-\bar{Z})X_i\}^2} - \frac{1}{\sum_i (X_i-\bar{X})^2}\right)} \qquad (13) \\ &= \frac{\frac{\sum_i e_i^2}{N-2}}{\frac{\sum_i w_i^2}{N-3}} H. \qquad (14) \end{aligned}$$

The ratio $\frac{\frac{1}{N-2}\sum_i e_i^2}{\frac{1}{N-3}\sum_i w_i^2}$ converges in probability to $\frac{\sigma_{\varepsilon\varepsilon} - \frac{\sigma_{X\varepsilon}^2}{\sigma_{XX}}}{\sigma_{uu}}$. Using $\varepsilon_i = \gamma \cdot 1(X_i \neq Z_i) + u_i$, and $u_i|X_i, Z_i \sim N(0, \sigma_{uu})$ yields that $\frac{\sigma_{\varepsilon\varepsilon} - \frac{\sigma_{X\varepsilon}^2}{\sigma_{XX}}}{\sigma_{uu}} = \frac{\sigma_{uu} + \gamma^2(\sigma_{SS} - \frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}}$ so that

$$\begin{aligned} T &= \frac{\sigma_{uu} + \gamma^2(\sigma_{SS} - \frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}} H + o_p(1) \\ &= \frac{\sigma_{uu} + \gamma^2(\sigma_{SS} - \frac{\sigma_{SX}^2}{\sigma_{XX}})}{\sigma_{uu}} V + o_p(1). \end{aligned}$$

Appendix 2: Proof of Theorem

By assumption, the second moments of the estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ exist. Conditioning on the information sets $A1$ and $A2$ does not increase the variation of the estimators so that $\Omega = \Omega_{\text{Restr}} + M_1$ for some matrix M_1 that is positive semi-definite. If all the elements of the matrix M_1 are zero, then $\Omega_{\text{Restr}}^{-1} = \Omega^{-1}$ and $T_{New,\text{infeasible}} = T_{H,\text{infeasible}}$. For a positive definite M_1, we use Abadir and Magnus (2005, exercise and solution 12.16), i.e. $\{\Omega - \Omega_{\text{Restr}}\}$ is positive definite implies that $\{\Omega_{\text{Restr}}^{-1} - \Omega^{-1}\}$ is positive definite. In other words we have that $\Omega_{\text{Restr}}^{-1} = \Omega^{-1} + M_2$ for some matrix M_2 that is positive definite. This gives

$$\begin{aligned} T_{New,\text{infeasible}} &\equiv (\hat{\theta}_1 - \hat{\theta}_2)'\Omega_{\text{Restr}}^{-1}(\hat{\theta}_1 - \hat{\theta}_2) \\ &= (\hat{\theta}_1 - \hat{\theta}_2)'\Omega^{-1}(\hat{\theta}_1 - \hat{\theta}_2) + (\hat{\theta}_1 - \hat{\theta}_2)'M_2(\hat{\theta}_1 - \hat{\theta}_2) \\ &= T_{H,\text{infeasible}} + (\hat{\theta}_1 - \hat{\theta}_2)'M_2(\hat{\theta}_1 - \hat{\theta}_2) \\ &\geq T_{H,\text{infeasible}}. \end{aligned}$$

This proves (i).

In order to prove (ii) note that now $\Omega = \Omega_{\text{Restr}} + M_1$ for some matrix M_1 that is positive definite so that, using Abadir and Magnus (2005, exercise and solution 12.16), $\Omega_{\text{Restr}}^{-1} = \Omega^{-1} + M_2$ for some matrix M_2 that is positive definite. Let δ be a row vector of length $\dim(\theta)$. We thus have that for any $\delta \neq 0$ (i.e. all elements of δ are unequal to zero) that $\delta'\Omega_{\text{Restr}}^{-1}\delta > \delta'\Omega^{-1}\delta$. Using $\text{plim}(\hat{\theta}_1) \neq \text{plim}(\hat{\theta}_2)$ and the Slutsky theorem yields that with probability approaching one $\frac{T_{New}}{T_H} > 1$.

We prove (iii) by noting that (iii) $\hat{\theta}_1 - \hat{\theta}_2 \sim N(0, \Omega)$ where $\Omega = \Omega_{\text{Restr}}$. Thus, $T_{New} = (\hat{\theta}_1 - \hat{\theta}_2)'\hat{\Omega}_{\text{Restr}}^{-1}(\hat{\theta}_1 - \hat{\theta}_2) = T_H + o_p(1)$ and the result follows from the properties of the χ^2-distribution (see, e.g., Lehmann and Romano (2005)).

4.1. Notes on Appendix 1.

$$
\begin{aligned}
X_i, Z_i &\in \{0,1\}, \\
P(Z_i &= 1) = \frac{1}{3},\ P(X_i = 1|Z_i = 1) = P(X_i = 0|Z_i = 0) = \delta, \\
Y_i &= \alpha + \beta X_i + \varepsilon_i, \qquad (15) \\
\varepsilon_i &= \gamma \cdot 1(X_i \neq Z_i) + u_i,\ \text{and}\ u_i|X_i, Z_i \sim N(0, \sigma_{uu}),
\end{aligned}
$$

$$
E\{1(X_i \neq Z_i)\} = E\{1(X_i \neq Z_i)|Z_i = 1\} \cdot \frac{1}{3} + E\{1(X_i \neq Z_i)|Z_i = 0\} \cdot \frac{2}{3} = 1 - \delta.
$$

$$
var\{1(X_i \neq Z_i)\} = \delta(1-\delta).
$$

1

$$
\begin{aligned}
\sigma_{\varepsilon\varepsilon} &= \gamma^2 var\{1(X_i \neq Z_i)\} + \sigma_{uu} \\
&= \gamma^2 \delta(1-\delta) + \sigma_{uu}
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{X\varepsilon} &= \gamma E[\{1(X_i \neq Z_i) - \delta\}X_i] = \\
&= \gamma^2 \delta(1-\delta) + \sigma_{uu}
\end{aligned}
$$

$\varepsilon_i = \gamma \cdot 1(X_i \neq Z_i) + u_i$, and $u_i|X_i, Z_i \sim N(0, \sigma_{uu})$ yields that $\frac{\sigma_{\varepsilon\varepsilon} - \frac{\sigma^2_{X\varepsilon}}{\sigma_{XX}}}{\sigma_{uu}} = \frac{\sigma_{uu} + \gamma^2(\sigma_{SS} - \frac{\sigma^2_{SX}}{\sigma_{XX}})}{\sigma_{uu}}$[checked last equality]

bias LS $\frac{\sigma_{X\varepsilon}}{\sigma_{XX}} = \frac{\gamma\sigma_{SX}}{\sigma_{XX}}$

noncentrality parameter: $N\frac{\gamma\sigma_{SX}}{\sigma_{XX}}$

In this local asymptotics the Hausman test statistic converges to a random variable that we denote by V.

n this local asymptotics the Hausman test statistic converges to a random variable that we denote by V. The proposed test statistic converges[6] to $\frac{\sigma_{uu} + \gamma^2(\sigma_{SS} - \frac{\sigma^2_{SX}}{\sigma_{XX}})}{\sigma_{uu}} V$ and the ratio of the test statistics, $\frac{T}{H}$, converges to $\frac{\sigma_{uu} + \gamma^2(\sigma_{SS} - \frac{\sigma^2_{SX}}{\sigma_{XX}})}{\sigma_{uu}}$. Since $\frac{\sigma_{uu} + \gamma^2(\sigma_{SS} - \frac{\sigma^2_{SX}}{\sigma_{XX}})}{\sigma_{uu}} > 1$ if $\gamma \neq 0$ we have that the proposed is more powerful.

[6] See the appendix for details.