Economics Letters 123 (2014) 291–294

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

# The maximum number of parameters for the Hausman test when the estimators are from different sets of equations

Kazumitsu Nawata [a,*], Michael McAleer [b,c,d,e]

[a] Graduate School of Engineering, University of Tokyo, Japan
[b] Department of Quantitative Finance, National Tsing Hua University, Taiwan
[c] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands
[d] Tinbergen Institute, The Netherlands
[e] Department of Quantitative Economics, Complutense University of Madrid, Spain

HIGHLIGHTS

- There exist a maximum number of parameters that can be used in the Hausman test.
- The asymptotic variance may converge to a singular matrix.
- The maximum number of parameters that can be used in the test is determined.

ARTICLE INFO

*Article history:*
Received 21 January 2014
Received in revised form
1 February 2014
Accepted 7 March 2014
Available online 13 March 2014

*JEL classification:*
C3
C12

*Keywords:*
Hausman test
Specification test
Number of parameters
Instrumental variable (IV) model
Box–Cox model
Sample selection bias

ABSTRACT

Hausman (1978) developed a widely-used model specification test that has passed the test of time. In this paper, we show that the asymptotic variance of the difference of the two estimators can be a singular matrix. Three illustrative examples are used, namely an exogeneity test for the linear regression model, a test for the Box–Cox transformation, and a test for sample selection bias.



## 1. Introduction

Hausman (1978) developed a widely-used model specification test that has passed the test of time. The test is based on two estimators, one being consistent under the null hypothesis but inconsistent under the alternative, and the other being consistent under both the null and alternative hypotheses.

The difference of two estimators and the corresponding variance are used to calculate the test statistic, which asymptotically follows the chi-squared distribution with degrees of freedom given by the number of parameters. Holly (1982, p. 749) wrote "Hausman's procedure seems to be more general than the classical procedures for it does not seem to require that the null hypothesis be given in a parametric form".

This paper considers the Hausman test for a case in which two estimators are obtained as roots of two different sets of equations. Some equations of the two sets may be the same, but at least one equation is different. The null hypothesis is that two estimators obtained from different sets of equations converge to the same values. It is shown that it may not be possible to use all the parameters in the model for the Hausman test. The asymptotic variance of the difference of the two estimators may converge to a singular matrix, and there exist a maximum number of parameters that can be used in the Hausman test. The maximum number of parameters

* Corresponding author.
*E-mail address:* kiichi815@msn.com (K. Nawata).

http://dx.doi.org/10.1016/j.econlet.2014.03.005
0165-1765/

that can be used in the test is determined by the number of different equations in the two sets. This result coincides with the case of a standard parametric test, where the degrees of freedom are given by the number of restrictions in the null hypothesis.

The remainder of the paper is given as follows. A Hausman test for a general model is discussed in Section 2, and three illustrative examples are given in Section 3.

## 2. A Hausman test for a general model

Let $\theta$ be a $k$-dimensional vector of unknown parameters. Consider two estimators, $\hat{\theta}$ and $\tilde{\theta}$, where $\hat{\theta}$ is consistent under the null hypothesis and inconsistent under the alternative, whereas $\tilde{\theta}$ is consistent under both the null and alternative hypotheses. Suppose that $\hat{\theta}$ is the root of $k$ equations given by:

$$f_T(\theta) = 0, \quad \text{and} \quad g_T(\theta) = 0 \tag{1}$$

where $f_T(\theta) = 0$ and $g_T(\theta) = 0$ express vectors of $k_1$ and $k_2$ different equations, respectively. On the other hand, $\tilde{\theta}$ is given by:

$$f_T(\theta) = 0, \quad \text{and} \quad h_T(\theta) = 0. \tag{2}$$

Standard conditions are assumed to hold. All models, whether linear or nonlinear, that are estimated using moment restrictions belong to this category.

Let $\theta_0$ be the true parameter value of $\theta$. Under the null hypothesis, it follows that:

$$f_T(\hat{\theta}) = f_T(\theta_0) + \left.\frac{\partial f_T}{\partial \theta'}\right|_{\theta_0} (\hat{\theta} - \theta_0) + o_P(1/\sqrt{T}), \tag{3}$$

$$f_T(\tilde{\theta}) = f_T(\theta_0) + \left.\frac{\partial f_T}{\partial \theta'}\right|_{\theta_0} (\tilde{\theta} - \theta_0) + o_P(1/\sqrt{T}).$$

Since $f_T(\hat{\theta}) = 0$ and $f_T(\tilde{\theta}) = 0$, it follows that:

$$\left.\frac{\partial f_T}{\partial \theta'}\right|_{\theta_0} \sqrt{T}(\hat{\theta} - \tilde{\theta}) = o_P(1). \tag{4}$$

This means that there are $k_1$ linear relations between $\sqrt{T}\hat{\theta}$ and $\sqrt{T}\tilde{\theta}$ asymptotically, and only $k_2$ elements are linearly independent asymptotically. Let $R$ be a $r \times k$ matrix and Rank$(R) = r$. If $r > k_2$, $T \cdot V\{R(\hat{\theta} - \tilde{\theta})\}$ will converge to a singular matrix under the null hypothesis, so that we cannot use the Hausman test in this case, and the maximum number of parameters that can be used in the test is $k_2$.

Although $\Omega_T = T \cdot V(\hat{\theta} - \tilde{\theta})$ converges to a singular matrix, it is generally a nonsingular matrix for finite values of $T$. Therefore, it will be necessary to set some eigenvalues to zero and define a new matrix $\Omega_T^*$ in order to use the generalized inverse matrix method. The results of this paper imply that the maximum number of nonzero eigenvalues of $\Omega_T^*$ cannot be greater than $k_2$.

## 3. Illustrative examples

In this section we give three illustrative examples of the Hausman test, namely an exogeneity test for the linear regression model, a test for the Box–Cox transformation, and a test for sample selection bias.

### 3.1. An exogeneity test for the linear regression model

As the first example, we consider a classical exogeneity test of the linear regression model,

$$y_t = x'_{1t}\beta_1 + x'_{2t}\beta_2 + u_t = x'_t\beta + u_t, \quad t = 1, 2, \ldots, T, \tag{5}$$

where $x'_t = (x'_{1t}, x'_{2t})$, $\beta' = (\beta'_1, \beta'_2)$, $x_{1t}$ is the $k_1$th dimensional vectors of the explanatory variables which is known to satisfy $\text{cov}(x_{1t}, u_t) = 0$, and $x_{2t}$ is the $k_2$th dimensional vectors of the explanatory variables which might be $\text{cov}(x_{2t}, u_t) \neq 0$.

For this model, we consider the test where the null and alternative hypotheses are given by:

$$H_0 : \text{cov}(x_{2t}, u_t) = 0, \qquad H_1 : \text{cov}(x_{2t}, u_t) \neq 0. \tag{6}$$

This test is a classical example of the Hausman test, and has been examined extensively (see, for example, Durbin, 1954, Wu, 1973, Smith, 1983, 1984, 1985, Holly, 1982, and Hausman and Taylor, 1981). However, the problem has not been examined in the context of this paper, where we can reach the conclusion much more simply than using existing methods.

Under the null hypothesis, the ordinary least squares (OLS) estimator is consistent and efficient if the error terms are independently and identically distributed (i.i.d.) normal random variables. However, the OLS estimator is inconsistent under the alternative hypothesis. On the other hand, the instrumental variables (IV) estimator is consistent under both the null and alternative hypotheses.

Let $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}$ be the OLS estimators of $\beta_1$, $\beta_2$ and $\beta$, $\tilde{\beta}_1$, $\tilde{\beta}_2$ and $\tilde{\beta}$ be the IV estimators and $k = k_1 + k_2$. The OLS and IV estimators are given by:

$$\begin{aligned} &\sum_t x_{1t}(y_t - x'_{1t}\beta_1 - x'_{2t}\beta_2) = 0 \quad \text{and} \\ &\sum_t x_{2t}(y_t - x'_{1t}\beta_1 - x'_{2t}\beta_2) = 0, \quad \text{and} \\ &\sum_t x_{1t}(y_t - x'_{1t}\beta_1 - x'_{2t}\beta_2) = 0 \quad \text{and} \\ &\sum_t z_t(y_t - x'_{1t}\beta_1 - x'_{2t}\beta_2) = 0, \end{aligned} \tag{7}$$

where $z_t$ is a vector of variables which satisfies $\sum_t z_t u_t/T \xrightarrow{P} 0$. The first $k_1$ equations are the same, and the differences arise in the latter $k_2$ equations. As the first $k_1$ equations yield the OLS estimators, we have:

$$\hat{\beta}_1 - \tilde{\beta}_1 = \left\{\sum_t (x_{1t}x'_{it})^{-1} \sum_t (x_{1t}x'_{2t})\right\} (\tilde{\beta}_2 - \hat{\beta}_2), \tag{8}$$

which is a linear function of $\hat{\beta}_2 - \tilde{\beta}_2$. Therefore, if we choose $q > k_2$, $V(\hat{\beta}^* - \tilde{\beta}^*)$ becomes a singular matrix.

### 3.2. A test for the Box–Cox transformation

The second example is the Box and Cox (1964) transformation model (BC model), which is given by:

$$Z_t = x'_t\beta + u_t, \qquad y_t > 0, \quad t = 1, 2, \ldots, T, \tag{9}$$

$$Z_t = \frac{y_t^\lambda - 1}{\lambda}, \quad \text{if } \lambda \neq 0,$$

$$Z_t = \log(y_t), \quad \text{if } \lambda = 0.$$

Generally, the likelihood function under the normality assumption (BC likelihood function) is misspecified, and the maximum likelihood estimator (BC MLE) is not consistent. However, the BC MLE can be a consistent estimator under a certain assumption. Nawata (2013) proposed an estimator which is consistent even if the assumption is not satisfied. Therefore, we can use the Hausman test for this model using these estimators. We will explain the asymptotic distributions of the BC MLE and Nawata's estimator. We then show that the Hausman test cannot be used for more than two parameters.

#### 3.2.1. BC MLE

The BC likelihood function is given by:

$$\log L(\theta) = \sum_t [\log \phi\{(z_t - x'_t)/\sigma\beta\} - \log \sigma] + (\lambda - 1)\sum_t \log y_t, \tag{10}$$