Department of Economics
University of Southampton
Southampton SO17 1BJ
UK

Discussion Papers in Economics and Econometrics

TESTING THE EXOGENEITY ASSUMPTION IN PANEL DATA MODELS WITH "NON CLASSICAL" DISTURBANCES

Raymond O'Brien
Eleonora Patacchini

No. 0302

Testing the Exogeneity Assumption in Panel Data Models with "Non Classical" Disturbances

Raymond O'Brien   Eleonora Patacchini
University of Southampton

February 20, 2003

Abstract

This paper is concerned with the use of the Durbin-Wu-Hausman test for correlated effects with panel data. The assumptions underlying the construction of the statistic are too strong in many empirical cases. The consequences of deviations from the basic assumptions are investigated. The size distortion is assessed. In the case of measurement error, the Hausman test is found to be a test of the difference in asymptotic biases of between and within group estimators. However, its 'size' is sensitive to the relative magnitude of the intra-group and inter-group variations of the covariates, and can be so large as to preclude the use of the statistic in this case. We show to what extent some assumptions can be relaxed in a panel data context and we discuss an alternative robust formulation of the test. Power considerations are presented.

Keywords: models with panel data, Hausman test, minimum variance estimators, quadratic forms in normal variables, Monte Carlo simulations

JEL Classification: C23, C12, C16, C15

1 Introduction

The Hausman test is the standard procedure used in empirical work in order to discriminate between the fixed effects and random effects model. It can be described as follows.[1]

Suppose that we have two estimators for a certain parameter θ of dimension $K \times 1$. One of them , $\widehat{\vartheta}_r$, is robust, i.e. consistent under both the null hypothesis H_0 and the alternative H_1, the other, $\widehat{\vartheta}_e$, is efficient and consistent under H_0 but inconsistent under H_1. The difference between the two is then used as the basis for testing. It can be shown (Hausman, 1978) that, under appropriate assumptions, under H_0 the statistic h based on $\left(\widehat{\vartheta}_R - \widehat{\vartheta}_E\right)$ has a limiting chi-squared distribution:

$$h = \left(\widehat{\vartheta}_r - \widehat{\vartheta}_e\right)' \left[\widehat{Var}\left(\widehat{\vartheta}_r - \widehat{\vartheta}_e\right)\right]^{-1} \left(\widehat{\vartheta}_r - \widehat{\vartheta}_e\right) \overset{a}{\sim} \chi^2_k.$$

If this statistic lies in the upper tail of the chi-square distribution we reject H_0. If the variance matrix is consistently estimated, the test will have power against any alternative under which $\widehat{\vartheta}_r$ is robust and $\widehat{\vartheta}_e$ is not. Holly (1982) discusses the power in the context of maximum likelihood.

In a panel data context the test can be used as a test for correlated effects. The null hypothesis assumes lack of correlation between the individual effect η_i and explanatory variable x_{it} :

$$H_0 : Cov(x_{it}, \eta_i) = 0.$$

The *Within Groups* estimator, $\widehat{\beta}_{wg}$, is robust regardless of the correlation between η_i and x_i. The *Balestra-Nerlove* estimator, $\widehat{\beta}_{BN}$, is efficient under H_0 but inconsistent under H_1 :

$$H_1 : Cov(x_{it}, \eta_i) \neq 0.$$

The Hausman statistic in this case takes the form

$$h_1 = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right)' \left[\widehat{Var}\left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right)\right]^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right) \overset{a}{\sim} \chi^2_k. \quad (1)$$

If we cannot reject the null hypothesis then the most reasonable model for the data at hand is the random effects model, otherwise the fixed effects model is more justified. However, using the results in Hausman (1978), the statistic used in practice

$$h_2 = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right)' \left(\widehat{V}_{wg} - \widehat{V}_{BN}\right)^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right), \quad (2)$$

where $V_{wg} = Var\left(\widehat{\beta}_{wg}\right)$ and $V_{BN} = Var\left(\widehat{\beta}_{BN}\right)$. It is based on the result that the variance of the difference between an estimator and an efficient estimator is equal to

[1] This approach is also used by Durbin (1954) and Wu (1973). For this reason tests based on the comparison of two sets of parameter estimates are also called Durbin-Wu-Hausman tests, or DWH. For simplicity of exposition we will refer to the Hausman (1978) set up.

the differences of the variances:

$$Var\left(\widehat{\beta}_{wg}-\widehat{\beta}_{BN}\right)=V_{wg}-V_{BN}. \tag{3}$$

In the time series-cross section model considered in Hausman (1978) this equality holds because $\widehat{\beta}_{BN}$ is an efficient estimator in the sense that it attains the Cramér-Rao Lower Bound for fixed λ (defined below), and $Cov\left(\widehat{\beta}_{wg},\widehat{\beta}_{BN}\right)=Var\left(\widehat{\beta}_{BN}\right)$. This implies

$$\begin{aligned}Var\left(\widehat{\beta}_{wg}-\widehat{\beta}_{BN}\right) &= Var\left(\widehat{\beta}_{wg}\right)+Var\left(\widehat{\beta}_{BN}\right)-2Cov\left(\widehat{\beta}_{wg},\widehat{\beta}_{BN}\right)\\ &= Var\left(\widehat{\beta}_{wg}\right)+Var\left(\widehat{\beta}_{BN}\right)-2Var\left(\widehat{\beta}_{BN}\right)\\ &= Var\left(\widehat{\beta}_{wg}\right)-Var\left(\widehat{\beta}_{BN}\right)=V_{wg}-V_{BN}.\end{aligned}$$

However, in applied studies, this may not always be the case and one should be careful in using h_2 automatically. If equality (3) does not hold, h_2 does not follow an asymptotic chi-squared distribution, even under H_0.

This paper considers the effects on the Hausman statistic used in applied panel data studies, h_2, of deviations from the conditions required in Lemma 2.1 in Hausman (1978), which guarantees that equality (3) holds. The lemma is stated as follows.

Lemma 1 Consider two estimators $\widehat{\beta}_0$, $\widehat{\beta}_1$ which are both consistent and asymptotically normally distributed with $\widehat{\beta}_0$ attaining the asymptotic Cramér-Rao bound so that $\sqrt{T}\left(\widehat{\beta}_0-\beta\right)\overset{a}{\sim}N\left(0,V_0\right)$ and $\sqrt{T}\left(\widehat{\beta}_1-\beta\right)\overset{a}{\sim}N\left(0,V_1\right)$ where V_0 is the inverse of Fisher's information matrix. Consider $\widehat{q}=\widehat{\beta}_1-\widehat{\beta}_0$. Then the limiting distributions of $\sqrt{T}\left(\widehat{\beta}_0-\beta\right)$ and $\sqrt{T}\widehat{q}$ have zero covariance, $Cov\left(\widehat{\beta}_0,\widehat{q}\right)=0$, a null matrix.

The plan of the paper is as follows.

Regarding the attainment of the Cramér-Rao Lower Bound, in Section 2 we prove that if we want to compare different estimators within a specific set, the assumption of full efficiency is not necessary. A relative lower bound for the variance can play the role. The variance of the difference between two estimators belonging to such a set is still equal to the difference of the variances if one of the two is the minimum variance estimator in the specific set considered. The algebraic derivation of this result is provided in the panel data framework. The Lemmas contained in Appendix 1 prove that this holds both in the exact and in the limiting case. Given that the *Balestra-Nerlove* estimator can be obtained as a matrix weighted average of the *Between Groups,* $\widehat{\beta}_{bg}$, and the *Within Groups* estimators (Maddala, 1971), we consider the set of estimators which is defined by a matrix weighted average of two unbiased (or consistent in the limiting case) estimators.

However, even the attainment of a minimum variance bound may be a strong assumption in empirical studies. This circumstance is related to assumptions about the error term. A failure of the assumption of spherical disturbances is quite common circumstance in practice. Section 3 presents a robust formulation of the Hausman test for correlated effects, which is based on the construction of an auxiliary regression. We explain and discuss to what extent the use of artificial regressions may allow us to construct tests based on the difference between two estimators in a panel data model without making strong assumptions about the disturbances. The motivation underlying the implementation of the robust test is that the size distortion of the standard Hausman test, h_2, in cases of misspecification of the variance-covariance matrix of the disturbances may be serious. This is investigated in Section 5.

The failure of the consistency of the two estimators under the null is discussed in Section 4. Such discussion is extremely relevant because a possible failure of the consistency of the *Within Groups* and the *Balestra-Nerlove* estimators, not related to the source of endogeneity being tested, is almost never raised in empirical studies. We explain to what extent the econometrics of panel data, offering a variety of different estimators for the same parameter, can help us to deal with this issue.

Section 6 compares the power of the standard Hausman test and the robust formulation presented in Section 3 using a Monte Carlo experiment. Section 7 concludes.

2 The Failure of the Assumption of Full Efficiency

Consider the following model

$$y_{it} = x_{it}^{'}\beta + \eta_i + v_{it}, \qquad i = 1, ..., N, \qquad t = 1, ..., T \tag{4}$$

where x_{it} is a $K \times 1$ vector of stochastic regressors, $\eta_i \sim iid\left(0, \sigma_\eta^2\right)$, $v_{it} \sim iid\left(0, \sigma^2\right)$ are uncorrelated with x_{it} and $Cov\left(\eta_i, v_{it}\right) = 0$.

Defining the disturbance term

$$\varepsilon_{it} = \eta_i + v_{it},$$

the variance-covariance matrix of the errors is

$$\underset{(NT\times NT)}{\Sigma} = I_N \otimes \Omega$$

where

$$\Omega = \begin{pmatrix} \sigma_\eta^2 + \sigma^2 & \dots & \sigma_\eta^2 \\ \vdots & \ddots & \vdots \\ \sigma_\eta^2 & \dots & \sigma_\eta^2 + \sigma^2 \end{pmatrix} = \sigma^2 I_T + \sigma_\eta^2\, \iota\iota^{'} \tag{5}$$

and ι is a column vector of T ones.

The unobserved heterogeneity implies correlation over time for single units, but there is no correlation across units.

Hausman and Taylor (1981) propose three different specification tests for the hypothesis of uncorrelated effects: one based on the difference between the *Within Groups* and the *Balestra-Nerlove* estimator, another on the difference between the *Balestra-Nerlove* and the *Between Groups* and a third on the difference between the *Within Groups* and the *Between Groups*. They show that the *chi-square* statistics for the three tests are numerically identical. We now analyze the Hausman statistic constructed on the difference between the *Within Groups* and the *Balestra-Nerlove* estimator, commonly used in empirical work.

Hereafter, we define as fully efficient an estimator that reaches the Cramér-Rao Lower Bond and as minimum variance the one that has the minimum variance within a specific class. Let

$$\lambda = \frac{\sigma^2}{\sigma^2 + T\sigma_\eta^2}.$$

If we assume normality in model (4), it is well-known that the *Balestra-Nerlove* estimator, i.e. the generalized least square estimator, is fully efficient if the variance-ratio parameter λ is known, and asymptotically fully efficient if λ is consistently estimated. (A distributional assumption is required in order to obtain the Cramér-Rao Bound.) Therefore the hypothesis underlying the construction of the Hausman statistic are satisfied and the results of the test are reliable. However, we will demonstrate that even without assuming normality of the ε_{it} the results of the standard Hausman test are reliable, the key assumption being (5). We will use the panel data framework as an example. In what follows we take λ as known. The same result holds asymptotically if a consistent estimator $\widehat{\lambda}$ is available. It is implied by the Hausman-Taylor result that we can construct the same test using $\widehat{\beta}_{wg}$ and $\widehat{\beta}_{bg}$, as will be clarified below.

We write the *Balestra-Nerlove* estimator (Balestra and Nerlove, 1966) as a function of the variables in levels

$$\widehat{\beta}_{BN} = \left(X'QX + \lambda X'MX\right)^{-1}\left(X'Q + \lambda X'M\right)Y \tag{6}$$

where

$$\begin{aligned} Q &= I_N \otimes Q^+, \\ Q^+ &= I_T - \frac{1}{T}ii', \\ M &= I_N \otimes M^+, \\ M^+ &= \frac{1}{T}ii' = I_T - Q^+, \end{aligned}$$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix},\ Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix},\ X_i = \begin{bmatrix} x'_{i1} \\ x'_{i2} \\ \vdots \\ x'_{iT} \end{bmatrix},\ y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}.$$

Q^{+} is the matrix that transforms the data to deviations from the individual time mean, M^{+} is the matrix that transforms the data to averages. Rearranging

$$\widehat{\beta}_{BN} = \left[X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1} X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]Y. \tag{7}$$

The variance is

$$\begin{aligned} Var(\widehat{\beta}_{BN}) &= \left\{\left[X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1} X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]\right\} Var(Y) \\ &\times\left\{\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\left[X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1}\right\}. \end{aligned} \tag{8}$$

Using a simplified version of the Sherman-Morrison-Woodbury formula (Golub and Loan, 1983, p.50) one can show, that, under assumption (5), the variance of y_i can be written as[2]

$$\begin{aligned} Var(y_i) &= \sigma^2\left[I_T - \frac{\sigma_\eta^2}{\sigma^2 + T\sigma_\eta^2}\iota\iota^{'}\right]^{-1} = \sigma^2\left[I_T - \frac{1}{T}(1-\lambda)\iota\iota^{'}\right]^{-1} \\ &= \sigma^2\left[\left(I_T - \frac{1}{T}\iota\iota^{'}\right) + \lambda\frac{1}{T}\iota\iota^{'}\right]^{-1}. \end{aligned}$$

This can also be obtained by ignoring time effects, and thus setting $\omega = 0$, in Nerlove (1971). Using the matrices involved in formula (6), we can rewrite this expression as

$$\begin{aligned} Var(y_i) &= \sigma^2\left[\left(I_T - M^{+}\right) + \lambda M^{+}\right]^{-1} && (9) \\ &= \sigma^2\left[Q^{+} + \lambda I_T - \lambda Q^{+}\right]^{-1} && (10) \\ &= \sigma^2\left[\lambda I_T + (1-\lambda)Q^{+}\right]^{-1}. && (11) \end{aligned}$$

Thus

$$Var(Y) = I_N \otimes Var(y_i) = \sigma^2\left[\lambda I_{NT} + (1-\lambda)Q\right]^{-1}.$$

Substituting (11) in (8), we obtain

$$\begin{aligned} & Var(\widehat{\beta}_{BN}) \\ &= \sigma^2\left[X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1} X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]\left[\lambda I_{NT} + (1-\lambda)Q\right]^{-1} && (12) \\ &\quad\times\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\left[X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1} \\ &= \sigma^2\left[X^{'}\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1}. && (13) \end{aligned}$$

Similarly, using the Q matrix defined in formula (6), we write also the *Within Groups* estimator as a function of the initial variables in levels

$$\widehat{\beta}_{wg} = \left[X^{'}QX\right]^{-1} X^{'}QY. \tag{14}$$

[2]See Appendix 2 for further details.

The variance is

$$Var(\widehat{\beta}_{wg}) = \left[X'QX\right]^{-1} X'Q(VarY)Q'X \left[X'QX\right]^{-1}. \tag{15}$$

If we transform the data into deviations, the variance of y_i can be written as

$$Var(Q^{+}y_i) = Q^{+}Var(y_i)Q^{+'} = \sigma^2 Q^{+}\left[I_T + \theta\iota\iota'\right]Q^{+} = \sigma^2 Q^{+}Q^{+} = \sigma^2 Q^{+} \tag{16}$$

where $\theta = \sigma_\eta^2/\sigma^2$ and $Q^{+}\iota = 0$, a vector of zeros. Thus

$$Var(QY) = \sigma^2 I_N \otimes Q^{+} = \sigma^2 Q$$

Plugging (16) in (15), we obtain[3]

$$\begin{aligned} Var(\widehat{\beta}_{wg}) &= \sigma^2 \left[X'QX\right]^{-1} X'QQQ'X \left[X'QX\right]^{-1} \\ &= \sigma^2 \left[X'QX\right]^{-1}. \end{aligned} \tag{17}$$

Hence, from (13) and (17)

$$Var(\widehat{\beta}_{wg}) - Var(\widehat{\beta}_{BN}) = \sigma^2 \left\{ \left[X'QX\right]^{-1} - \left[X'\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1} \right\}. \tag{18}$$

Next we show that such expression is exactly equal to the variance of the difference between the two estimators.

$$Var(\widehat{\beta}_{BN} - \widehat{\beta}_{wg}) = Var(\widehat{\beta}_{BN}) - Cov(\widehat{\beta}_{BN}, \widehat{\beta}_{wg}) - Cov(\widehat{\beta}_{wg}, \widehat{\beta}_{BN}) + Var(\widehat{\beta}_{wg}).$$

>From (7) and (14)

$$\begin{aligned} & Cov(\widehat{\beta}_{BN}, \widehat{\beta}_{wg}) \\ = \; & \sigma^2 \left[X'\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1} X'\left[\lambda I_{NT} + (1-\lambda)\,Q\right] \\ & \times \left[\lambda I_{NT} + (1-\lambda)Q\right]^{-1} QX \left[X'QX\right]^{-1} \\ = \; & \sigma^2 \left[X'\left[\lambda I_{NT} + (1-\lambda)\,Q\right]X\right]^{-1} = Var(\widehat{\beta}_{BN}). \end{aligned}$$

This is symmetric, and thus equal to $Cov(\widehat{\beta}_{wg}, \widehat{\beta}_{BN})$. Thus using (11) and (16), we obtain

$$\begin{aligned} Var(\widehat{\beta}_{BN} - \widehat{\beta}_{wg}) &= Var(\widehat{\beta}_{BN}) - Var(\widehat{\beta}_{BN}) - Var(\widehat{\beta}_{BN}) + Var(\widehat{\beta}_{wg}) \\ &= Var(\widehat{\beta}_{wg}) - Var(\widehat{\beta}_{BN}) \end{aligned} \tag{19}$$

[3] Recall that Q is an idempotent matrix.

as required. We have proved that equality (3) holds for λ known or otherwise fixed.

As we said, the case of estimated λ can be treated by using the Hausman-Taylor result that an algebraically identical test statistic can be constructed using the difference between $\widehat{\beta}_{wg}$ and the *Between Groups* estimator $\widehat{\beta}_{bg}$. We obtain

$$(\widehat{\beta}_{wg} - \widehat{\beta}_{bg})' \left[Var(\widehat{\beta}_{wg}) + Var(\widehat{\beta}_{bg})\right]^{-1} (\widehat{\beta}_{wg} - \widehat{\beta}_{bg})$$

as the estimators have zero covariance. In this form, we can see that estimating σ^2 and λ (or σ_η^2) affects only the variance matrix of the test statistic. We thus obtain the same test statistic whatever λ is, and (2) remains correct. It does not follow from these arguments that the equality (3) can be made exact for estimated λ. If $\widehat{\beta}_{BN}$ and $\widehat{\beta}_{wg}$ were independent of $\widehat{\lambda}$, the result would follow, but this requires normality of the disturbances. Viewing $\widehat{\beta}_{BN}$ as a feasible *GLS* estimator, Kakwani (1967) implies it is unbiased. However, conditional on $\widehat{\lambda}$ it may or may not be unbiased. Further, the variances obtained are for λ fixed, not conditional on $\widehat{\lambda}$. So attempts to obtain unconditional variances from conditional variances and variances of conditional expectations do not seem fruitful. So it would appear that the exact result (3) may require normality of the ε_{it} or λ fixed. Equality (3) implies that for fixed and known λ, and known σ^2, under normality h would have an exact χ^2 distribution. If λ is estimated, and/or the ε_{it} are not normal, h is asymptotically χ^2 as long as x_{it} are sufficiently well-behaved to ensure that $\widehat{\beta}_{BN}$ and $\widehat{\beta}_{wg}$ are asymptotically normal, and σ^2 and σ_η^2 (or equivalently λ) are appropriately estimated. This is less restrictive than the assumptions required for the identification of the Cramér-Rao bound. We obtain the result (3) without assuming normality because we compare two linear unbiased estimators, one of them achieving the minimum variance for a linear estimator. Lemma 4 in Appendix 1 shows that the variance result depends only on minimum variance properties, not on normality or achievement of a particular (Cramér-Rao) bound. However, in order to get a panel data generalized version of Lemma 1, it is necessary to prove a similar result in the limiting case. This aim is achieved in Lemma 10 in Appendix 1. The minimum variance property required is within a set of the form

$$\mathcal{T} = \{t : t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2\}$$

where t_1 and t_2 are estimators of the parameter vector θ. For completeness, Lemma 9 establishes that sets of this form will contain minimum variance members.

We can summarize as follows. If we want to use the Hausman statistic to compare two different estimators, e.g. one linear and one non linear, the assumption of normality may be crucial because it allow us to find an absolute lower bound for the variance of the estimators. However, if we want to compare different estimators within a set of the form of $\mathcal{T}$ neither the assumption of normality nor the attainment of the Cramér-Rao Lower Bound, even in the limiting case, is crucial. A lower bound for the variance can play the required role. The variance of the difference between two estimators belonging to the same set is still equal to the difference of the variances if one of the two is the minimum variance estimator in the specific set. Lemma 10 in Appendix 1 allows us to rely on the results provided by a traditional Hausman test in a more general set-up.

It is worth noting that we are not removing the assumption of asymptotic normality of the estimators in Lemma 1, which is needed to obtain the χ^2 distribution of the Hausman statistic. Our generalization applies for estimators that are asymptotically normally distributed but that do not reach the Cramér-Rao Bound.

We prove the result for a specific set of estimators but this does not rule out the possibility of extending the result to wider contexts. For instance, the GMM estimator is asymptotically normally distributed and attains the asymptotic Cramér-Rao Lower Bound only in some cases. Nevertheless, if we compare an arbitrary GMM estimator, e.g. using the identity matrix, and the one which uses the optimal weighting matrix (Hansen, 1982), Lemma 10 implies that Hansen's GMM can be used as basis for a Hausman test.

3 The Failure of the Assumption of Spherical Disturbances

In the previous section, we relaxed the assumption of full efficiency in Lemma 1. However, even the assumption that one of the two estimators has the minimum variance or that both are consistent under the null hypothesis can be still too strong in many empirical cases. In the panel data framework above considered (model (4)), the crucial assumption for (3) to hold is (5). In other words, the form of the covariance matrix has to be assumed. In cases of misspecification, i.e. if $Var(y) = \Omega^* \neq \Omega$, equality (3) does not hold any longer.

As Hausman clearly states at the very beginning of his article (Hausman, 1978), the specification test he presents takes the hypothesis that the disturbances have a spherical covariance matrix. He considers the standard regression framework

$$y = X\beta + \varepsilon, \tag{20}$$

where

$$E(\varepsilon/X) = 0, \tag{21}$$

and

$$Var(\varepsilon/X) = \sigma^2 I. \tag{22}$$

In most of the articles that followed, assumption (22) is never relaxed. The emphasis of this part of literature is placed in testing the orthogonality assumption, i.e. $E(\varepsilon/X) = 0$. In the panel data framework a test of the assumption (21) is a test for random versus fixed effects. Also in this context the assumption (22) is maintained.

The reason is straightforward if we consider the comparison between the *Within Groups* estimator and the *Balestra-Nerlove* estimator as a comparison between an *OLS* and a *GLS* estimator. One basic assumption in the construction of the Hausman statistic (Lemma 2.1 in Hausman, 1978) is that one of the two estimators has to reach the asymptotic Cramér-Rao Lower Bound or, using the generalization provided in Lemma 4 in Appendix 1, that at least has to be the minimum variance estimator in a specific class. In the panel data framework the *Balestra-Nerlove*, that is the generalized least square estimator, is the BLUE estimator if the *GLS* transformation produces spherical disturbances. This is the case if the correlation in the covariance matrix of the initial errors is due only to the omission of the individual effects, i.e. if the initial disturbances are spherical.

To make it clear, we analyze in detail the construction of the *Balestra-Nerlove* estimator. In practice the *Balestra-Nerlove* estimator can be calculated running an *OLS* regression on a transformed model. Assuming model (4), which implies the disturbances variance covariance matrix (5), the transformation of the y_i and the x_i is the following

$$\Omega^{-\frac{1}{2}}y_i = \begin{bmatrix} y_{i1} - \theta\overline{y}_{i.} \\ y_{i2} - \theta\overline{y}_{i.} \\ \vdots \\ y_{iT} - \theta\overline{y}_{i.} \end{bmatrix}$$

where

$$\Omega^{-\frac{1}{2}} = I - \frac{\theta}{T}ii', \quad \theta = 1 - \frac{\sigma}{\left(\sigma^2 + T\sigma_\eta^2\right)^{\frac{1}{2}}}$$

and likewise for the rows of x_i.

Under assumption (5), which implies initial spherical disturbances, this is a *GLS* transformation that produces a model with spherical disturbances. Hence running *OLS* on such a model we obtain the BLUE estimator. However, if assumption (5) does not hold, the *GLS* transformation does not guarantee that the new disturbances are spherical. In this case the *GLS* estimator, namely the *Balestra-Nerlove*, is still consistent but it may not be the minimum variance estimator. The consequence is that we can no longer be sure that the equality (3) still holds. In these circumstance the results of the test may not be reliable. However, if the two estimators remain consistent the comparison can still be conducted, but the methodology needs to be adjusted in an appropriate way.

In what follows, we present a robust version of the Hausman test for panel data. It is based on the use of an artificial regression. Keeping the assumption of consistency of the two estimators, it allows us to compare different estimators without assuming normality or ranking them in terms of efficiency. Specifically, such methodology does not use the hypothesis that the variance of the difference of the two estimators is equal to the difference of the variances. It estimates directly the variance of the difference of the two estimators. It simply uses the statistic (1) instead of (2). Moreover, it provides estimators for the variances that are consistent and robust to heteroskedasticity and/or serial correlation of arbitrary form in the covariance matrix of the random disturbances. These estimators are obtained using White's formulae (White, 1984). It will be made clear to what extent the application of White's heteroskedasticity consistent estimators of covariance matrices in a panel data framework may also allow for the presence of dynamic effects.

Different artificial regressions have been proposed in the panel data literature to test for the presence of random individual effects, such as a Gauss-Newton regression by Baltagi (1996) or that proposed by Ahn and Lo (1996). However, the assumption of initial spherical disturbances has not been relaxed. As shown by Baltagi (1997, 1998), under the assumption of spherical disturbances, the three approaches, i.e. the Hausman specification test, the Gauss-Newton regression and the regression proposed by Ahn and Lo, yield exactly the same test statistic. Arellano (1993) first noted in the same panel data framework that an auxiliary regression can also be used to obtain a generalized test for correlated effects which is robust to heteroskedasticity and correlation of arbitrary forms in the disturbances. Davidson and MacKinnon (1993) list at least five different uses for artificial regressions including the calculation of estimated covariances matrices. We will use this device to estimate directly the variance between the two estimators without using equality (3). Furthermore, the application of White's formulae (White, 1984) in the panel data case will lead to heteroskedasticity and autocorrelation consistent estimators of such variance. Therefore, we can use an artificial regression to construct a test for the comparison of different estimators which is robust to deviations from the assumption of spherical disturbances. From now on we will call this technique the *HR-test*, for Hausman-Robust test.

Next we present the auxiliary regression that was proposed by Arellano (1993) to test for random versus fixed effects in a static panel data model.

Consider the general panel data model for individual i

$$\underset{(T\times1)}{y_i} = \underset{(T\times K)}{X_i}\beta + \underset{(T\times1)}{v_i}, \qquad i = 1, ..., N.$$

This system of T equations in levels can be transformed into $(T-1)$ equations in deviations and one in averages. We obtain

$$\begin{cases} y_i^* = x_i^*\beta + \mu_i^* \longrightarrow (T-1) \text{ equations} \\ \overline{y_i} = \overline{x_i}\beta + \overline{\mu_i} \longrightarrow \quad 1 \;\; \text{equation.} \end{cases}$$

Estimating by *OLS* the $N(T-1)$ equations in orthogonal deviations from individual time-means we obtain the *Within Groups* estimator, i.e. $\widehat{\beta}_{wg}$. Estimating by *OLS* the N average equations we obtain the *Between Groups* estimator, i.e. $\widehat{\beta}_{bg}$.

Let

$$\beta_{wg} = E\left(\widehat{\beta}_{wg}\right)$$

and

$$\beta_{bg} = E\left(\widehat{\beta}_{bg}\right).$$

Rewrite the system as

$$\begin{cases} y_i^* = x_i^* \beta_{wg} + \mu_i^* - x_i^* \beta_{bg} + x_i^* \beta_{bg} \\ \overline{y_i} = \overline{x_i} \beta_{bg} + \overline{\mu_i}. \end{cases}$$

Rearranging, we obtain

$$\begin{cases} \overline{y_i} = x_i^* \left(\beta_{wg} - \beta_{bg}\right) + x_i^* \beta_{bg} + \mu_i^* \\ \overline{y_i} = \overline{x_i} \beta_{bg} + \overline{\mu_i}. \end{cases}$$

Call

$$Y_i^+ = \begin{pmatrix} y_i^* \\ \overline{y_i} \end{pmatrix}, W_i^+ = \begin{pmatrix} x_i^* & x_i^* \\ 0 & \overline{x_i} \end{pmatrix},$$

$$\beta^+ = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_{wg} - \beta_{bg} \\ \beta_{bg} \end{pmatrix}, \mu_i^+ = \begin{pmatrix} \mu_i^* \\ \overline{\mu_i} \end{pmatrix}.$$

The augmented auxiliary model is

$$Y_i^+ = W_i^+ \beta^+ + \mu_i^+, \qquad i = 1, ..., N. \tag{23}$$

If we estimate β^+ by *OLS*, we obtain directly the variance of the difference of the two estimators in the upper left part of the variance-covariance matrix of β^+. If we then estimate this covariance matrix using the White's formulae and we perform a Wald test on appropriate coefficients, we obtain a reliable *HR-test* comparing the two estimators we are interested in, namely $\widehat{\beta}_{wg}$ and $\widehat{\beta}_{bg}$. As first noted by Arellano (1993), under the assumption of spherical disturbances a Wald test on appropriate coefficients in the auxiliary regressions is equivalent to the standard Hausman test. Appendix 4 provides an analytical derivation of this result. The following Lemma is proved.

Lemma 2 Given model (23),

$$\widehat{\beta}_1 = \widehat{\beta}_{wg} - \widehat{\beta}_{bg}, \tag{24}$$

$$Var(\widehat{\beta}_1) = Var\left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right), \tag{25}$$

An appropriate estimator $\widehat{Var}(\widehat{\beta}_1)$ *consistently estimates* $Var(\widehat{\beta}_1)$. (26)

It is shown that, in order to get a consistent estimate of the variance, the first set of equations has to be scaled.

In what follows, we will clarify to what extent an application of White's formulae for estimators of covariances matrices (White, 1984) in a panel data context provides a consistent estimator which is robust to heteroskedasticity and arbitrary correlation in the covariance matrix of the random disturbances. It may also control for the presence of fixed effects. This latter possibility may be accommodated if we make further assumptions, i.e. cross-sectional heteroskedasticity which takes on a finite number of different values.

Consider a simple panel data framework without fixed effects

$$\begin{aligned} y_{i1} &= \beta x_{i1} + \varepsilon_{i1} \\ y_{i2} &= \beta x_{i2} + \varepsilon_{i2}, \\ &\vdots \\ y_{iT} &= \beta x_{iT} + \varepsilon_{iT}, \qquad i = 1, ..., N, \end{aligned}$$

where

$$E(\epsilon_i \epsilon_i^{'}) = \begin{pmatrix} \sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2 \end{pmatrix} = \sigma^2 I_T = \Sigma.$$

Assume that in the complete model

$$\underset{(NT \times NT)}{\Omega} = I \otimes \Sigma = \begin{pmatrix} \Sigma & 0 & \dots & 0 \\ 0 & \Sigma & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma \end{pmatrix}. \tag{27}$$

Define

$$\underset{(T\times 1)}{X_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{pmatrix}} \qquad \underset{(T\times 1)}{y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}} \qquad \underset{(T\times 1)}{\epsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}}$$

and rewrite the model as

$$\underset{(T\times 1)}{y_i} = \underset{(T\times 1)}{X_i}\beta + \underset{(T\times 1)}{\epsilon_i}, \qquad i = 1, ..., N. \tag{28}$$

This formulation allows us to consider panel data in the framework defined in White (1984). If we assume no cross-sectional correlation and $N \to \infty$, all the hypotheses underlying the derivation of White's results are satisfied. Hence, Proposition 7.2 in White (1984, p. 165) applies.

$$\widehat{\Sigma} = N^{-1} \sum_{i=1}^{N} \widehat{\epsilon_i} \widehat{\epsilon_i}^{'} \xrightarrow{p} \Sigma \tag{29}$$

and

$$\widehat{\Omega} = I \otimes \widehat{\Sigma} \xrightarrow{p} \Omega.$$

However, while with uni-dimensional data sets we obtain heteroskedasticity consistent estimators because ϵ_i is a scalar, in the two dimensional case ϵ_i is a vector and we obtain a consistent estimator of the whole matrix Σ. Hence, by applying the result (29) in the panel data case we obtain a consistent estimator of the variance covariance matrix of the disturbances that also allows for the presence of dynamic effects within groups.

Therefore, the estimators of the variance of the *OLS* estimators of β in the panel data model (28) can be obtained by

$$\widehat{Var(\beta)} = \left[\sum_{i=1}^{N}\left(X_i'X_i\right)\right]^{-1}\sum_{i=1}^{N}X_i'\widehat{\Omega}X_i\left[\sum_{i=1}^{N}\left(X_i'X_i\right)\right]^{-1}. \tag{30}$$

As stated by Arellano (1993), they are heteroskedasticity and autocorrelation consistent. Such estimators are the ones used in the implementation of the *HR-test*. This case is referred in White (1984) as *contemporaneous covariance estimation*.

However, White (1984) also implements consistent estimators in another case that explicitly takes into consideration a grouping structure of the data. Consider again the panel data model (28). Replace assumption (27) by

$$\underset{(NT\times NT)}{\Omega} = \begin{pmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \Sigma_N \end{pmatrix}.$$

In this context, in a slightly different notation from that used by White (1984, p.172-173), suitable for the panel data framework, we can obtain consistent estimators of the covariance matrix Ω using

$$\widehat{\Omega} = diag(\widehat{\Sigma}_1, \widehat{\Sigma}_{2,} \dots \widehat{\Sigma}_N)$$

where

$$\widehat{\Sigma}_i = T^{-1}\widehat{\epsilon_i}\widehat{\epsilon_i}'.$$

In other words, a consistent estimator for the covariance matrix of group i is constructed by averaging the group residuals over only the observations in group i. In the balanced panel data case, their number is constant between groups and equal to T. This estimator is not only robust to autocorrelation of arbitrary form within groups but it also allows for the possibility that individual error covariance matrices may differ according to observable characteristics (such as region, union, race, etc....).

4 The Failure of the Orthogonality Assumption between Regressors and Random Errors

The previous section discusses the use of the Hausman test when there are reasons to think that one of the assumptions, namely that one estimator is the minimum variance one, is too strong, as it is often the case in empirical work.

This section refers to the use of the test in circumstances where even the consistency of the estimators under the null hypotheses cannot be assured. A possible failure of the consistency of the two estimators, not related to the source of endogeneity being test, is almost never considered in empirical studies. It is worthwhile noting that the question addressed by the Hausman test is whether the parameters of interest have been estimated consistently. Thus, the test detects the presence of any possible endogeneity problem (Davidson and MacKinnon, 1989), not necessarily induced by a correlation between the regressors and the individual effects. Rejection may be also caused, for instance, by the presence of measurement errors-in-variables. Almost always in the widespread use of the Hausman test for correlated effects in static panel data modelling, the consistency of the *Within Groups* and the *Balestra-Nerlove* estimators under the null is not questioned. However if for instance we are in presence of measurement errors-in-variables, least square estimators do not lose only their efficiency but also their consistency. Our claim is that in such contexts the use of the standard Hausman test is not correct. In the presence of arbitrary measurement errors-in-variables, if we compare the *Within Groups* estimator and the *Balestra-Nerlove* estimators to test for uncorrelated individual effects, we may be comparing two inconsistent estimators. Moreover, the *Within Groups* estimator and the *Balestra-Nerlove* estimator are *OLS* estimators constructed on different transformations of the data. Measurement errors can have different impact using different transformations of the data. For instance, if we use first differences the bias can be magnified (Griliches and Hausman, 1986). As a consequence, the probability limits of two estimators calculated on different transformations of the data may be different. In this case the null distribution of the Hausman test will depend on this difference, and thus on the (unknown) parameters. In other words, in presence of measurement errors-in-variables the widespread practice of using the standard Hausman statistics based on the comparison between the *Within Groups* and the *Balestra-Nerlove* estimator is not methodologically correct and it can lead to unreliable results.

An analysis of the causes that lead to a failure of the assumption of consistency is quite delicate because they are often related to unobservable factors often difficult to detect and to treat properly. The econometrics of panel data, offering a variety of different estimators for the same parameter, can help us to deal with this issue. The structure of a panel data set can be useful to distinguish among different sources of bias and can allow us to control for the effects of different kinds of unobservable factors. Using the "repeated measurement property" of a panel data set, i.e. each cross sectional observation is followed over time, we can construct different kinds of instrumental variables from the data set. Assuming a specific structure of the measurement errors, we can find instrumental variables estimators that remain consistent. Hence it is still possible to use the Hausman test framework using appropriate estimators and gain some knowledge on the most reliable model specification. Patacchini (2002) presents a sequential test for panel data aiming to distinguish between an endogeneity problem caused by measurement errors-in-variables and an endogeneity problem caused by correlation between regressors and individual effects. It is based on the use of appropriate *HR-tests* in a particular sequence.

5 The Size of the Test

In this section we investigate the size distortion which occurs in the use of the standard Hausman test when the basic assumptions (Lemma 2.1 in Hausman 1978) are not satisfied.

Consider the panel data model (4) presented in Section 3. The Hausman test investigates the presence of specification errors of the form $E(\eta_i|x_{it}) \neq 0$. The robust version proposed in Section 3 tests such orthogonality assumption between explanatory variables and disturbances in presence of other forms of misspecification. In particular we are interested in a possible misspecification in the variance-covariance matrix of the disturbances arising, for instance, from the presence of measurement errors in variables. This case may be the rule rather than the exception in applied studies.

We want to test the hypothesis

$$H_o : E(\varepsilon_{it}|x_{it}) = 0 \tag{31}$$

against the alternative

$$H_1 : E(\varepsilon_{it}|x_{it}) \neq 0,$$

when

$$Var(\varepsilon_i|x_{it}) \neq \Omega_i. \tag{32}$$

Hausman (1978) shows that under H_o the test statistic

$$h = \widehat{q}'\widehat{V}(\widehat{q})^{-1}\widehat{q} \sim \chi^2_k \tag{33}$$

where, $V(\widehat{q})$ is the asymptotic variance of q, and k is the length of q. The same test statistic is obtained if we consider the vector $\widehat{q}$ equal to

$$\begin{aligned} \widehat{q}_1 &= (\widehat{\beta}_{wg} - \widehat{\beta}_{BN}), \\ \text{or } \widehat{q}_2 &= (\widehat{\beta}_{bg} - \widehat{\beta}_{BN}), \\ \text{or } \widehat{q}_3 &= (\widehat{\beta}_{wg} - \widehat{\beta}_{bg}). \end{aligned}$$

As Hausman and Taylor (1981) pointed out they are all nonsingular transformations of one another. The estimate of the variance covariance matrix used in the three cases is

$$\begin{aligned} \widehat{V}(\widehat{q}_1) &= \widehat{V}(\widehat{\beta}_{wg}) - \widehat{V}(\widehat{\beta}_{BN}), \\ \text{or } \widehat{V}(\widehat{q}_2) &= \widehat{V}\widehat{\beta}_{bg}) - \widehat{V}(\widehat{\beta}_{BN}), \\ \text{or } \widehat{V}(\widehat{q}_3) &= \widehat{V}(\widehat{\beta}_{wg}) + \widehat{V}(\widehat{\beta}_{bg}). \end{aligned}$$

If we are in presence of misspecification of the form (32), none of the above expressions gives a consistent estimate of the variance-covariance matrix, even under H_o. The distribution of the test statistic under H_o need to be investigated. The nominal size may be quite different from the observed one.

To investigate the size distortion under normality, we use the distributions of quadratic forms in normal random variables.[4] In particular, we use the following Lemma.[5]

Lemma 3 *(in Lemma 3.2 in Vuong, 1989). Let $x \sim N_K(0, V)$, with rank $(V) \leq K$, and let A be an $K \times K$ symmetric matrix. Then the random variable $x'Ax$ is distributed as a weighted sum of chi-squares with parameters (K, γ), where γ is the vector of eigenvalues of AV.*

This implies that $x'Ax$ is χ^2_r, where $r = rank(A)$, if and only if AV is idempotent (Muirhead, 1982, Theorem 1.4.5).

If $A = V^{-1}$, i.e. in cases of no misspecification, AV is idempotent. The theorem is satisfied and result (33) holds. The test statistic gives correct significance levels.

If $A \neq V^{-1}$ but AV is idempotent then $rank\,(A) < K$ and/or $rank\,(V) < K$ but still (33) holds. We omit this case for simplicity of exposition.

If $A \neq V^{-1}$ and AV is not idempotent, implying that the eigenvalues of AV are not 0 or 1, the asymptotic distribution of the Hausman test under H_o is a weighted sum of central chi-squares

$$h \sim \sum_{i=1}^{K} d_i z_i^2$$

where $z_i^2 \sim \chi_1^2$ and d_i are the eigenvalues of AV. This implies that the significance levels of the standard Hausman test are not correct.

Consider first the limiting case where $d_1 \to K$, $d_i \to 0$, $i = 2, .., K$. Figure 1 illustrates numerically that

$$\Pr\left[K\chi_1^2 > \chi_{K,\alpha}^2\right],$$

where $\chi_{K,\alpha}^2$ is the critical value for a test of size α under the χ_r^2 distribution. In this illustration α is set equal to 0.05.

In general we distinguish two effects: a scale effect if $\sum\limits_{i=1}^{K} d_i \neq K$, which is predictable (e.g. if $d_i = 2\ \forall\ i,\ h \sim 2\chi_K^2$) and a dispersion effect if $d_i \neq d_j$, even if $\sum\limits_{i=1}^{K} d_i = K$. We normalize the weights and we conjecture that the dispersion effect is maximized in the limit if we put all the weight on the largest eigenvalue, say the first one.

[4] See, among others, Murihead (1982, Ch. 1), Johnson and Kotz (1970, Ch.29).

[5] Both this Lemma and the following one hold also in the asymptotic case (using the Continuous Mapping Theorem, e.g. White, 1984, Lemma 4.27).

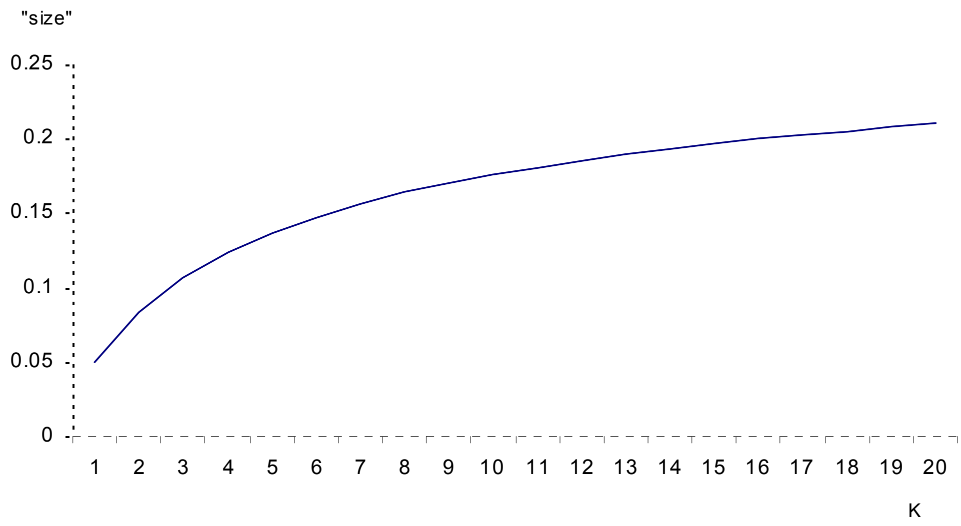

Figure 1: $\Pr\left[K\chi_1^2 > \chi_{K,\alpha}^2\right]$

Figure 1 illustrates this case, i.e. the tail area of a χ_K^2 is compared with the maximum tail area of $K\chi_1^2$. The graph shows that the size distortion is an increasing function of K. For instance, if K is equal to 14, an inappropriate use of the Hausman test will give a probability of rejecting a true hypothesis of exogeneity which is almost 4 time larger than the nominal size.

In certain simple contexts an expression for the eigenvalues of AV can be analytically derived. For instance, a common source of misspecification in the variance covariance matrix occurs when elements of the regressor matrix contain measurement errors.

Suppose the true model is

$$y_{it} = z_{it}^{'}\beta + \eta_i + v_{it}, \qquad i = 1, ..., N, \qquad t = 1, ..., T \tag{34}$$

where $z_{it}^{'}$ is a $1 \times K$ vector of theoretical variables, $\eta_i \sim iid\left(0, \sigma_\eta^2\right)$, $v_{it} \sim iid\left(0, \sigma^2\right)$ uncorrelated with the columns of z_{it} and $Cov\left(\eta_i, v_{it}\right) = 0$. The observed variables are

$$x_{it} = z_{it} + m_{it},$$

where m_{it} is a vector of measurement errors uncorrelated with η_i and v_{it}. The estimated model is

$$y_{it} = x_{it}^{'}\beta + \eta_i + v_{it} - \beta' m_{it}, \qquad i = 1, ..., N, \qquad t = 1, ..., T. \tag{35}$$

In the case of exact measurement, i.e. $m_{it} = 0$,

$$\begin{aligned} Var(y_{it}) &= E(\eta_i + v_{it})^2 = \sigma_\eta^2 + \sigma^2, \\ Cov(y_{it}, y_{it-s}) &= Cov(x_{it}^{'}\beta + \eta_i + v_{it}, x_{it-s}^{'}\beta + \eta_i + v_{it-s}) \\ &= \sigma_\eta^2 \qquad \forall s. \end{aligned}$$

The variance-covariance matrix is matrix (5). It can be written as

$$\underset{(NT\times NT)}{\Sigma} = I_N \otimes \Omega_i,$$

where

$$\Omega_i = \sigma^2 I_T + \sigma_\eta^2\ \iota\iota^{'} = \sigma^2[I_T + \vartheta_1 \iota\iota^{'}], \tag{36}$$

and

$$\vartheta_1 = \frac{\sigma_\eta^2}{\sigma^2}.$$

If we assume that $m_{it} \sim iid\,(0, \Sigma_M)$, we obtain

$$\begin{aligned} Var(y_{it}) &= E(\eta_i + v_{it} - \beta m_{it})^2 = \sigma_\eta^2 + \sigma^2 + \beta'\Sigma_M\beta, \\ Cov(y_{it}, y_{it-s}) &= Cov(x_{it}^{'}\beta + \eta_i + v_{it} - \beta' m_{it}, x_{it-s}^{'}\beta + \eta_i + v_{it-s} - \beta' m_{it-s}) \\ &= \sigma_\eta^2 \qquad \forall s \neq 0. \end{aligned}$$

So

$$\Omega_i = \left(\sigma^2 + \beta'\Sigma_M\beta\right) I_T + \sigma_\eta^2\ \iota\iota^{'} = \left(\sigma^2 + \beta'\Sigma_M\beta\right)(I_T + \vartheta_2\iota\iota^{'}), \tag{37}$$

and

$$\vartheta_2 = \frac{\sigma_\eta^2}{\sigma^2 + \beta'\Sigma_M\beta}.$$

Consider now the exogeneity test based, for instance, on the comparison between $\widehat{\beta}_{BG}$ and $\widehat{\beta}_{WG}$. In this case, the measurement errors render $\widehat{\beta}_{BG}$ and $\widehat{\beta}_{WG}$ inconsistent. If we assume that

$$p\lim(\widehat{\beta}_{BG} - \beta) = p\lim(\widehat{\beta}_{WG} - \beta) = \left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1}\Sigma_M\beta = \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}\Sigma_M\beta$$

we show in Appendix 5 that if the rows $M_i \backsim NID(0, \Sigma_M)$

$$\begin{aligned} &\sqrt{N}(\widehat{\beta}_{WG} - \widehat{\beta}_{BG}) \xrightarrow{D} N(0, [1/(T-1)\}\left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \times \\ &\left[(\sigma^2 + \beta'\Sigma_M\beta)\Sigma_{ZQZ}/(T-1) + \sigma^2\Sigma_M + \{\Sigma_M\beta\beta'\Sigma_M + (\beta'\Sigma_M\beta)\Sigma_M\}\right] \times \\ &\left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} + \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \times \\ &[T\sigma_\eta^2\Sigma_{ZMZ} + (\sigma^2 + \beta'\Sigma_M\beta)\Sigma_{ZMZ} + \sigma_\eta^2 T\Sigma_M + \sigma^2\Sigma_M + \{\Sigma_M\beta\beta'\Sigma_M + (\beta'\Sigma_M\beta)\Sigma_M\}] \\ &\left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}). \end{aligned}$$

The Hausman test

$$\begin{aligned} h &= (\widehat{\beta}_{WG} - \widehat{\beta}_{BG})' \left[\widehat{Var}(\widehat{\beta}_{WG}) + \widehat{Var}(\widehat{\beta}_{BG})\right]^{-1} (\widehat{\beta}_{WG} - \widehat{\beta}_{BG}) \\ &= \sqrt{N}(\widehat{\beta}_{WG} - \widehat{\beta}_{BG})' \left[N\widehat{Var}(\widehat{\beta}_{WG}) + N\widehat{Var}(\widehat{\beta}_{BG})\right]^{-1} \sqrt{N}(\widehat{\beta}_{WG} - \widehat{\beta}_{BG}) \end{aligned}$$

will have the same asymptotic distribution as

$$h_a = \sqrt{N}(\widehat{\beta}_{WG} - \widehat{\beta}_{BG})' p\lim \left[N\widehat{Var}(\widehat{\beta}_{WG}) + N\widehat{Var}(\widehat{\beta}_{BG})\right]^{-1} \sqrt{N}(\widehat{\beta}_{WG} - \widehat{\beta}_{BG})$$

and we also show in Appendix 5 that

$$\begin{aligned} &N\widehat{Var}(\widehat{\beta}_{WG}) \\ &\xrightarrow{p} \{\sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M \left[\frac{1}{(T-1)}\Sigma_{ZQZ} + \Sigma_M\right]^{-1} \Sigma_M\beta\} \times \\ &[\Sigma_{ZQZ} + (T-1)\Sigma_M]^{-1} \end{aligned}$$

and

$$\begin{aligned} &N\widehat{Var}(\widehat{\beta}_{BG}) \\ &\xrightarrow{p} \{T\sigma_\eta^2 + \sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \Sigma_M\beta\} \times \\ &[\Sigma_{ZMZ} + \Sigma_M]^{-1} \end{aligned}$$

Thus in terms of the notation of Lemma 3, for the asymptotic distribution

$$\begin{aligned} V \;=\; & [1/(T-1)\} \left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \times \\ & \left[(\sigma^2 + \beta'\Sigma_M\beta)\Sigma_{ZQZ}/(T-1) + \sigma^2\Sigma_M + \{\Sigma_M\beta\beta'\Sigma_M + (\beta'\Sigma_M\beta)\Sigma_M\}\right] \times \\ & [\Sigma_{ZQZ}/(T-1) + \Sigma_M]^{-1} + [\Sigma_{ZMZ} + \Sigma_M]^{-1} \times \\ & [T\sigma_\eta^2\Sigma_{ZMZ} + (\sigma^2 + \beta'\Sigma_M\beta)\Sigma_{ZMZ} + \sigma_\eta^2 T\Sigma_M + \sigma^2\Sigma_M + \{\Sigma_M\beta\beta'\Sigma_M + (\beta'\Sigma_M\beta)\Sigma_M\}] \\ & [\Sigma_{ZMZ} + \Sigma_M]^{-1}). \end{aligned}$$

and

$$A = \left[\begin{array}{c} \{\sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M \left[\frac{1}{(T-1)}\Sigma_{ZQZ} + \Sigma_M\right]^{-1} \Sigma_M\beta\} \times [\Sigma_{ZQZ} + (T-1)\Sigma_M]^{-1} \\ +\{T\sigma_\eta^2 + \sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M [\Sigma_{ZMZ} + \Sigma_M]^{-1} \Sigma_M\beta\} \times [\Sigma_{ZMZ} + \Sigma_M]^{-1} \end{array}\right]^{-1}$$

Consider first the case when $\beta = 0$.

$$\begin{aligned} V \;=\; & [1/(T-1)] \left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \times \\ & \left[\sigma^2\Sigma_{ZQZ}/(T-1) + \sigma^2\Sigma_M\right] \times [\Sigma_{ZQZ}/(T-1) + \Sigma_M]^{-1} \\ & + [\Sigma_{ZMZ} + \Sigma_M]^{-1} \times [T\sigma_\eta^2\Sigma_{ZMZ} + \sigma^2\Sigma_{ZMZ} + \sigma_\eta^2 T\Sigma_M + \sigma^2\Sigma_M] \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \\ =\; & [1/(T-1)]\sigma^2 \left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \\ & + [\Sigma_{ZMZ} + \Sigma_M]^{-1} (T\sigma_\eta^2 + \sigma^2)[\Sigma_{ZMZ} + \Sigma_M] \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \\ =\; & [1/(T-1)]\sigma^2 \left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \\ & +(T\sigma_\eta^2 + \sigma^2) \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \end{aligned}$$

$$A = \left[\sigma^2 \left[\Sigma_{ZQZ} + (T-1)\Sigma_M\right]^{-1} + \{T\sigma_\eta^2 + \sigma^2\} \times \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}\right]^{-1}$$

so $AV = I$. As a check, when $\Sigma_M = 0$,

$$V = \sigma^2[1/(T-1)\} \left[\Sigma_{ZQZ}/(T-1)\right]^{-1} + \left[T\sigma_\eta^2 + \sigma^2\right]\left[\Sigma_{ZMZ}\right]^{-1}$$

$$A = \left[\sigma^2\Sigma_{ZQZ}^{-1} + \{T\sigma_\eta^2 + \sigma^2\}\Sigma_{ZMZ}^{-1}\right]^{-1}$$

which can be compared with Appendix 3.

Now let $\Sigma_Q = \Sigma_{ZQZ}/(T-1)$, $\sigma^{*2} = \sigma^2 + \beta'\Sigma_M\beta$, $\mathbf{c} = \Sigma_M\beta$, $\sigma^{**2} = \sigma^{*2} + T\sigma_\eta^2$, so

$$\begin{aligned} V &= [1/(T-1)]\left[\Sigma_Q + \Sigma_M\right]^{-1}\left[\sigma^{*2}[\Sigma_Q + \Sigma_M] + \mathbf{cc}'\right]\left[\Sigma_Q + \Sigma_M\right]^{-1} + \\ &\quad \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}\left[\sigma^{**2}[\Sigma_{ZMZ} + \Sigma_M] + \mathbf{cc}'\right]\left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \\ &= [1/(T-1)]\left[\sigma^{*2}[\Sigma_Q + \Sigma_M]^{-1} + \mathbf{dd}'\right] + \left[\sigma^{**2}[\Sigma_{ZMZ} + \Sigma_M]^{-1} + \mathbf{ee}'\right] \end{aligned}$$

where $\mathbf{d} = \left[\Sigma_Q + \Sigma_M\right]^{-1}\mathbf{c}$, and $\mathbf{e} = [\mathbf{\Sigma_{ZMZ}} + \mathbf{\Sigma_M}]^{-\mathbf{1}}\mathbf{c}$. These are just the inconsistencies, which we are assuming equal.

$$A = \begin{bmatrix} 1/(T-1)\{\sigma^{*2} - \mathbf{c}'\left[\Sigma_Q + \Sigma_M\right]^{-1}\mathbf{c}\} \times \left[\Sigma_Q + \Sigma_M\right]^{-1} \\ +\{\sigma^{**2} - \mathbf{c}'\left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}\mathbf{c}\} \times \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \end{bmatrix}^{-1}$$

The simplest case to examine is when $\Sigma_Q = \Sigma_{ZMZ} \Leftrightarrow p\lim\widehat{\beta}_{WG} = p\lim\widehat{\beta}_{BG}$ for all β; let $\Sigma_{QM} = \Sigma_Q + \Sigma_M = \Sigma_{ZMZ} + \Sigma_M$. Noting $\mathbf{d} = \mathbf{e}$, we have

$$V = \sigma^{+2}{\Sigma_{QM}}^{-1} + 2\mathbf{dd}'$$

where

$$\begin{aligned} \sigma^{+2} &= [1/(T-1)]\sigma^{*2} + \sigma^{**2} \\ &= [T/(T-1)]\sigma^{*2} + T\sigma_\eta^2 \end{aligned}$$

$$A = [\sigma^{++2}\Sigma_{QM}^{-1}]^{-1}$$

where

$$\begin{aligned} \sigma^{++2} &= [1/(T-1)]\{\sigma^{*2} - \mathbf{c}'\Sigma_{QM}^{-1}\mathbf{c}\} + \sigma^{**2} - \mathbf{c}'\Sigma_{QM}^{-1}\mathbf{c}, \\ &= [T/(T-1)][\sigma^{*2} - \mathbf{c}'\Sigma_{QM}^{-1}\mathbf{c}] + T\sigma_\eta^2 \end{aligned}$$

and AV has the same eigenvalues as

$$A^{1/2}VA^{1/2} = \frac{\sigma^{+2}}{\sigma^{++2}}I + \frac{2}{\sigma^{++2}}{\Sigma_{QM}}^{1/2}\mathbf{dd}'{\Sigma_{QM}}^{1/2}$$

and has $K-1$ eigenvalues of

$$k = \sigma^{+2}/\sigma^{++2}$$

and one of

$$
\begin{aligned}
k + (2/\sigma^{++2})\mathbf{d}'\mathbf{\Sigma_{QM}d} &= k+(2/\sigma^{++2})\mathbf{c}'\mathbf{\Sigma_{QM}^{-1}c} \\
&= k+(2/\sigma^{++2})\boldsymbol{\beta}'\Sigma_M\mathbf{\Sigma_{QM}^{-1}\Sigma}_M\beta.
\end{aligned}
$$

Thus the size distortion depends on scalar quantities,

$$
\begin{aligned}
k &= \frac{\sigma^{+2}}{\sigma^{++2}} = \frac{1}{1-k^*}, \\
k^* &= \frac{\sigma^{+2}-\sigma^{++2}}{\sigma^{+2}} = \frac{\boldsymbol{\beta}'\Sigma_M\mathbf{\Sigma_{QM}^{-1}\Sigma}_M\beta}{[T/(T-1)]\{\sigma^2+\beta'\Sigma_M\beta\}+T\sigma_\eta^2}
\end{aligned}
$$

and the larger root is

$$
\frac{\sigma^{+2}}{\sigma^{++2}} + \frac{2}{\sigma^{++2}}k^*\sigma^{+2} = \frac{1}{1-k^*}\left[1+2k^*\right].
$$

$$
\begin{aligned}
\beta'\Sigma_M\Sigma_{QM}^{-1}\Sigma_M\beta &= \beta'\Sigma_M^{1/2}[\Sigma_M^{1/2}(\Sigma_Q+\Sigma_M)^{-1}\Sigma_M^{1/2}]\Sigma_M^{1/2}\beta \\
&= \beta'\Sigma_M^{1/2}[\Sigma_M^{-1/2}\Sigma_Q\Sigma_M^{-1/2}+\mathbf{I}]^{-1}\Sigma_M^{1/2}\beta
\end{aligned}
$$

If we now write

$$
\gamma = \Sigma_M^{1/2}\beta
$$

γ is the vector of parameters in the model

$$
\begin{aligned}
y_i &= [Z_i+M_i]\,\Sigma_M^{-1/2}\Sigma_M^{1/2}\beta + \eta_i\mathbf{i} + \varepsilon_i \\
&= Z_i^*\gamma + M_i^*\gamma + \eta_i\mathbf{i} + \varepsilon_i
\end{aligned}
$$

where the rows of M_i are $NID(\mathbf{0},\mathbf{I})$ and $Z_i^* = Z_i\Sigma_M^{-1/2} \Rightarrow Z_i = Z_i^*\Sigma_M^{1/2} \Rightarrow Z_i'M^+Z_i = \Sigma_M^{1/2}Z_i^{*\prime}M^+Z_i^*\Sigma_M^{1/2}$

$$
\begin{aligned}
k^* &= \gamma'[\Sigma_M^{-1/2}\Sigma_{ZMZ}\Sigma_M^{-1/2}+\mathbf{I}]^{-1}\boldsymbol{\gamma}/\sigma^{+2} \\
&= \gamma'\left[\Sigma_{Z^*MZ^*}+I\right]^{-1}\gamma/\left[[T/(T-1)]\{\sigma^2+\gamma'\gamma\}+T\sigma_\eta^2\right] \qquad (38)
\end{aligned}
$$

The components of the variance of $y_{i,t}$ are

$$
Var(y_{i,t}) = \gamma'\gamma + \sigma_\eta^2 + \sigma^2
$$

so an interpretation of our result is that if one takes one component of the variance, $\gamma'\gamma$, downweights it by the between sums of squares of the unobserved 'true' variables (in the model with standardised measurement errors), to produce $\gamma'\left[\Sigma_{Z^*MZ^*}+I\right]^{-1}\gamma$, then the 'size' distortion depends on k^*, as in (38), and the asymptotic distribution of the Hausman test is not $\chi^2(K)$, but a weighted sum of K $\chi^2(1)$, $K-1$ weights

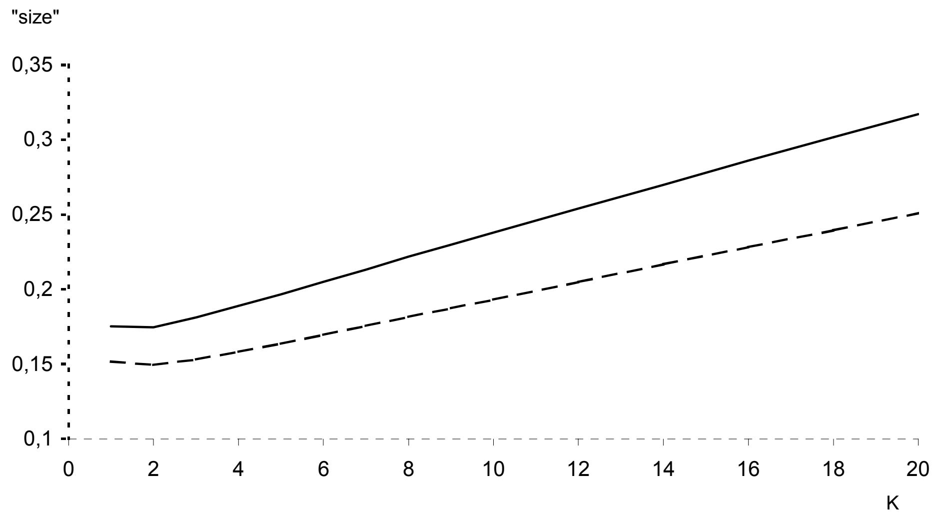

Figure 2: 'size' *vs* K

being $1/(1-k^*)$, with one of $[1+3k^*/(1-k^*)]$. It also follows that a lower bound to the distortion is provided by multiplying a $\chi^2(k)$ by $1/(1-k^*)$.

A number of qualifications are in order. This only occurs if the inconsistency of within and between estimators is equal, and, further, the within group sum of squares matrix, and between group sum of squares matrix, are equal:

$$\Sigma_{ZMZ}, = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} Z_i M^+ Z_i, = \Sigma_Q, = \frac{1}{T-1}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} Z_i Q^+ Z_i.$$

The equality of $p\lim(\widehat{\beta}_{BG}-\beta)$ and $p\lim(\widehat{\beta}_{WG}-\beta)$ is required to ensure that the asymptotic 'size' is not 1. (Thus the Hausman test can be regarded as a (consistent) test of equality of these 'inconsistencies').The equality of Σ_{ZMZ} and Σ_Q simplifies the result and is an aid to interpretability. We also assume that the rows of M_i, the measurement errors, are $NID(0,\Sigma_M)$. Some assumption about fourth moments is required, and this appears the simplest.

We can plot the size distortion for assumed values of $T, K, \gamma'\gamma, \gamma'\left[\Sigma_{Z^*MZ^*}+I\right]^{-1}\gamma, \sigma^2_\eta$ and σ^2. If $T=5$ or $10, 1\leq K\leq 10, \gamma'\gamma=1, \sigma^2_\eta=\sigma^2=0.1$, and $\gamma'\left[\Sigma_{Z^*MZ^*}+I\right]^{-1}\gamma=0.5$, we have Figure 2, evaluated by Monte Carlo (1 million replications).

We can relax the assumtion that $\Sigma_Q=\Sigma_{ZMZ}$ by observing that V is of the form

$$V = k_1 B + k_2 C + \mathbf{d}^*\mathbf{d}^{*\prime}$$

and A is of the form

$$A = (k_3 B + k_4 C)^{-1}$$

where

$$\begin{aligned}
B &= [\Sigma_Q + \Sigma_M]^{-1}, \quad C = [\Sigma_{ZMZ} + \Sigma_M]^{-1} \\
k_1 &= [1/(T-1)]\sigma^{*2}, \quad k_2 = \sigma^{**2} \\
\mathbf{d}^* &= \{1 + 1/(T-1)\}^{1/2}\mathbf{d} = \{T/(T-1)\}^{1/2}\mathbf{d}, \\
k_3 &= 1/(T-1)\{\sigma^{*2} - \mathbf{c}'B^{-1}\mathbf{c}\}, < k_1 \\
k_4 &= \{\sigma^{**2} - \mathbf{c}'C^{-1}\mathbf{c}\}, < k_2
\end{aligned}$$

and B and C are positive definite. We see that A is "too small", and the test will be oversized.

$$\begin{aligned}
V &= B^{1/2}[k_1 I + k_2 B^{-1/2}CB^{-1/2} + B^{-1/2}\mathbf{d}^*\mathbf{d}^{*\prime}B^{-1/2}]B^{1/2} \\
A^{-1} &= B^{1/2}[k_3 I + k_4 B^{-1/2}CB^{-1/2}]B^{1/2}
\end{aligned}$$

Let

$$D = B^{-1/2}CB^{-1/2} = P\Lambda P'$$

where P is orthogonal, Λ diagonal, with as diagonal elements λ_i the eigenvalues of D. Then

$$\begin{aligned}
V &= B^{1/2}P[k_1 I + k_2\Lambda + P'B^{-1/2}\mathbf{d}^*\mathbf{d}^{*\prime}B^{-1/2}P]P'B^{1/2} \\
A &= \left[B^{1/2}P[k_3 I + k_4\Lambda]P'B^{1/2}\right]^{-1} = B^{-1/2}P[k_3 I + k_4\Lambda]^{-1}P'B^{-1/2}
\end{aligned}$$

and thus

$$\begin{aligned}
AV &= B^{-1/2}P[k_3 I + k_4\Lambda]^{-1}[k_1 I + k_2\Lambda + P'B^{-1/2}\mathbf{d}^*\mathbf{d}^{*\prime}B^{-1/2}P]P'B^{1/2} \\
&= B^{-1/2}P[diag([k_3 + k_4\lambda_i]^{-1})\{diag(k_1 + k_2\lambda_i) \\
&\quad + P'B^{-1/2}\mathbf{d}^*\mathbf{d}^{*\prime}B^{-1/2}P\}]P'B^{1/2}
\end{aligned}$$

which has the same eigenvalues as

$$\begin{aligned}
&\{diag(\frac{k_1 + k_2\lambda_i}{k_3 + k_4\lambda_i}) \\
&+ diag([k_3 + k_4\lambda_i]^{-1}\{P'B^{-1/2}\mathbf{d}^*\mathbf{d}^{*\prime}B^{-1/2}P\}
\end{aligned}$$

The second matrix has rank 1, and the eigenvalues of the whole matrix are bounded between the smallest of $k_{0,i} = (k_1 + k_2\lambda_i)/(k_3 + k_4\lambda_i)$ and the largest of $k_{0,i} + \mathbf{d}^{*\prime}B^{-1}\mathbf{d}^*/(k_3 + k_4\lambda_i)$. λ_i are the eigenvalues of $D = B^{-1/2}CB^{-1/2}$, or of $B^{-1}C = [\Sigma_Q + \Sigma_M][\Sigma_{ZMZ} + \Sigma_M]^{-1}$. $\mathbf{d} = [\Sigma_Q + \Sigma_M]^{-1}\,\mathbf{c} = [\boldsymbol{\Sigma}_{\mathbf{ZMZ}} + \boldsymbol{\Sigma}_{\mathbf{M}}]^{-\mathbf{1}}\mathbf{c} = B\mathbf{c} = C\mathbf{c}$

$$\begin{aligned}
\mathbf{d}^{*\prime}B^{-1}\mathbf{d}^* &= \{T/(T-1)\}\mathbf{c}'B\mathbf{c} = \{T/(T-1)\}\beta'\Sigma_M[\Sigma_{ZMZ} + \Sigma_M]^{-1}\Sigma_M\beta \\
&= \{T/(T-1)\}\gamma'\left[\Sigma_{Z^*MZ^*} + I\right]^{-1}\gamma
\end{aligned}$$

$$
\begin{aligned}
k_1 &= [1/(T-1)]\sigma^{*2}, \sigma^{*2} = \sigma^2 + \beta'\Sigma_M\beta = \sigma^2 + \gamma'\gamma \\
k_2 &= \sigma^{**2} = \sigma^{*2} + T\sigma_\eta^2 \\
k_3 &= 1/(T-1)\{\sigma^{*2} - \mathbf{c}'B^{-1}\mathbf{c}\} = 1/(T-1)\left[\sigma^2 + \gamma'\gamma - \gamma'\left[\Sigma_{Z^*MZ^*} + I\right]^{-1}\gamma\right] < k_1 \\
k_4 &= \{\sigma^{**2} - \mathbf{c}'C^{-1}\mathbf{c}\} = \sigma^2 + \gamma'\gamma + T\sigma_\eta^2 - \gamma'\left[\Sigma_{Z^*MZ^*} + I\right]^{-1}\gamma < k_2
\end{aligned}
$$

Thus

$$
\begin{aligned}
\sigma^{+2} &= [1/(T-1)]\sigma^{*2} + \sigma^{**2} = k_1 + k_2 \\
\sigma^{++2} &= [1/(T-1)]\{\sigma^{*2} - \mathbf{c}'\Sigma_{QM}^{-1}\mathbf{c}\} + \sigma^{**2} - \mathbf{c}'\Sigma_{QM}^{-1}\mathbf{c} = k_3 + k_4
\end{aligned}
$$

$$
\begin{aligned}
k_{0,i} &= \frac{k_1 + k_2\lambda_i}{k_3 + k_4\lambda_i} = \frac{k_1 + k_2 + k_2(\lambda_i - 1)}{k_3 + k_4 + k_4(\lambda_i - 1)} \\
&= \frac{\sigma^{+2}[1 + k_2(\lambda_i - 1)/\sigma^{+2}]}{\sigma^{++2}[1 + k_4(\lambda_i - 1)/\sigma^{++2}]} = k\frac{[1 + k_2(\lambda_i - 1)/\sigma^{+2}]}{[1 + k_4(\lambda_i - 1)/\sigma^{++2}]}
\end{aligned}
$$

Thus comparing this case with the $B = C$ case, we are introducing more variability into the eigenvalues, which as we have seen , may well increase the 'size' of the test. (Thus the 'size' is sensitive to the relative magnitude of the intra-group and inter-group variations of the covariates, Σ_{ZQZ} and Σ_{ZMZ}). Our conclusion is somewhat dispiriting: a significant Hausman statistic may arise from measurement error, as it is implicitly comparing the inconsistencies: but cannot be used to test if the inconsistencies are equal, as the 'size' may considerably exceed its nominal value, even when the inconsistencies *are* equal.

6 A Power Comparison

The possible serious size distortion of the standard Hausman test motivates the formulation of the *HR-test*. Using the White estimators for the variance-covariance matrix, the test is robust to the presence of common sources of misspecification of the variance-covariance matrix, i.e. to arbitrary patterns of within groups dependence. In other words, using the notation in Lemma 3, AV is idempotent and the nominal size is equal to the observed one. We now use a simulation experiment to investigate the relative power of the Hausman test and the *HR-test*. We are interested in a quantitative assessment of the possible power loss that may incur in using a robust version of the test, in absence of misspecification.

The postulated data generation process is the following.

We consider the model

$$y = \alpha x + \beta z + u,$$

where y, x, u and z are $(NT \times 1)$. The null hypothesis of the Hausman test is

$$E(u|x,z) = 0.$$

We assume z exogenous variable and we generate x correlated with u, so that the null hypothesis above is not satisfied. We consider

$$x = \gamma w + \varepsilon, \tag{39}$$

where x, w, ε are $(NT \times 1)$, w is an exogenous variable and (u, ε) are drawn from a bivariate normal distribution with a specified correlation structure.

The values from the exogenous regressors and the range of values for the parameters comes from the empirical case of study analyzed in Patacchini (2002). Using UK data, the following model is estimated.

$$lfillv_{it} = c + \gamma lunfv_{it} + \pi lutot_{it} + e_{it}, \qquad i = 1, ..., 275;\ t = 1, ..., 63$$

where $lfillv$ is the logarithm of filled vacancies, $lunfv$ is the number of unfilled vacancies (stock variable) and *lutot* is the number of unemployed in the area i at time t, both expressed in logs, c is a constant term, e indicates a disturbance term. The estimates of γ and π, 0.5 and 0.4, have been used in the simulation experiment for α and β respectively. Also, the best prediction for unfilled vacancies ($lunfv$) is found to be

$$lunfv_{it} = 1.2 \ln otv_{it}, \qquad i = 1, ..., 275;\ t = 1, ..., 63,$$

where *lnotv* is the log of the number of monthly notified vacancies (flow variable). In our experiment design, the real values for *lutot* and *lnotv* have been used as exogenous variables, i.e. respectively z and w. The endogenous variable $lunfv$, i.e. x, has been constructed according to the structure (39)

$$x = 1.2w + \varepsilon.$$

The equation estimated is

$$y = 0.5x + 0.4z + u,$$

where (u, ε) are constructed as draws from a multivariate normal distribution with the specified correlation coefficient rho of $(0, 0.05, 0.10, \ldots, 0.95)$.

Six sample sizes, typically encountered in applied panel data studies are used. The experiment is repeated 5000 times for each sample size and level of correlation. Figures 3 to 5 contain the results of the simulation experiment. The power is expressed in percentage.

The tables displayed compare *H_pow*, the power of the Hausman statistic (*H-test*):

$$h = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right)' \left(\widehat{V}_{wg} + \widehat{V}_{bg}\right)^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right)$$

Table 1: N=25, T=4

rho	rho^	H_pow	HR_pow
0.00	0.00	4.90	4.80
0.05	0.03	5.10	4.90
0.10	0.06	7.90	7.40
0.15	0.09	9.20	9.30
0.20	0.12	14.40	13.80
0.25	0.15	19.90	20.90
0.30	0.17	25.50	26.80
0.35	0.20	32.20	32.50
0.40	0.23	34.50	38.50
0.45	0.26	43.60	45.80
0.50	0.29	50.10	57.40
0.55	0.32	70.10	70.80
0.60	0.35	78.20	79.90
0.65	0.37	87.90	89.70
0.70	0.40	94.10	92.70
0.75	0.43	98.50	98.90
0.80	0.46	99.90	100.00
0.85	0.49	100.00	100.00
0.90	0.52	100.00	100.00
0.95	0.55	100.00	100.00

Table 2: N=25, T=10

rho	rho^	H_pow	HR_pow
0.00	0.00	4.60	4.50
0.05	0.04	6.50	5.40
0.10	0.08	8.10	6.10
0.15	0.11	12.50	9.20
0.20	0.15	16.40	13.90
0.25	0.17	20.60	20.10
0.30	0.21	25.40	27.50
0.35	0.25	31.50	32.50
0.40	0.28	40.10	43.30
0.45	0.32	50.20	55.50
0.50	0.35	57.20	61.90
0.55	0.39	70.20	72.70
0.60	0.42	82.40	85.40
0.65	0.46	88.60	90.00
0.70	0.49	99.80	96.70
0.75	0.53	99.90	99.40
0.80	0.56	99.90	99.90
0.85	0.60	100.00	99.90
0.90	0.64	100.00	100.00
0.95	0.67	100.00	100.00

Figure 3: Simulation Results

Table 3: N=25, T=20

rho	rho^	H_pow	HR_pow
0.00	0.00	4.80	4.70
0.05	0.04	6.80	5.90
0.10	0.07	9.00	8.10
0.15	0.10	17.80	16.50
0.20	0.14	27.80	27.00
0.25	0.18	36.10	36.40
0.30	0.21	46.20	48.10
0.35	0.25	66.20	66.50
0.40	0.28	79.00	79.60
0.45	0.32	87.20	87.90
0.50	0.35	95.00	93.90
0.55	0.39	97.80	97.70
0.60	0.42	99.10	98.70
0.65	0.46	99.90	99.80
0.70	0.50	99.90	100.00
0.75	0.53	100.00	100.00
0.80	0.57	100.00	100.00
0.85	0.60	100.00	100.00
0.90	0.64	100.00	100.00
0.95	0.67	100.00	100.00

Table 4: N=275, T=4

rho	rho^	H_pow	HR_pow
0.00	0.00	4.90	5.00
0.05	0.03	6.30	6.40
0.10	0.06	9.60	8.80
0.15	0.09	18.20	17.60
0.20	0.11	29.10	28.90
0.25	0.15	45.10	48.10
0.30	0.17	57.20	62.50
0.35	0.20	72.40	78.20
0.40	0.23	86.00	89.10
0.45	0.26	93.60	96.20
0.50	0.29	97.90	98.00
0.55	0.32	99.80	99.80
0.60	0.34	99.80	100.00
0.65	0.37	100.00	100.00
0.70	0.40	100.00	100.00
0.75	0.43	100.00	100.00
0.80	0.46	100.00	100.00
0.85	0.49	100.00	100.00
0.90	0.52	100.00	100.00
0.95	0.55	100.00	100.00

Figure 4: Simulation Results

Table 5: N=275, T=10

rho	rho^	H_pow	HR_pow
0.00	0.00	5.00	4.90
0.05	0.03	9.80	6.40
0.10	0.06	26.10	15.10
0.15	0.09	61.00	34.00
0.20	0.12	87.80	55.10
0.25	0.15	97.80	74.10
0.30	0.18	98.90	86.50
0.35	0.20	99.80	93.40
0.40	0.23	99.90	97.90
0.45	0.26	100.00	98.90
0.50	0.29	100.00	99.90
0.55	0.32	100.00	100.00
0.60	0.35	100.00	100.00
0.65	0.38	100.00	100.00
0.70	0.41	100.00	100.00
0.75	0.44	100.00	100.00
0.80	0.47	100.00	100.00
0.85	0.50	100.00	100.00
0.90	0.53	100.00	100.00
0.95	0.55	100.00	100.00

Table 6: N=275, T=20

rho	rho^	H_pow	HR_pow
0.00	0.00	5.10	4.70
0.05	0.03	18.40	6.40
0.10	0.06	59.70	18.90
0.15	0.09	91.10	40.10
0.20	0.12	99.80	62.40
0.25	0.15	99.90	75.50
0.30	0.18	99.90	87.40
0.35	0.20	100.00	94.10
0.40	0.23	100.00	98.90
0.45	0.26	100.00	100.00
0.50	0.29	100.00	100.00
0.55	0.32	100.00	100.00
0.60	0.35	100.00	100.00
0.65	0.38	100.00	100.00
0.70	0.41	100.00	100.00
0.75	0.44	100.00	100.00
0.80	0.47	100.00	100.00
0.85	0.50	100.00	100.00
0.90	0.53	100.00	100.00
0.95	0.56	100.00	100.00

Figure 5: Simulation Results

with *HR_pow,* the power of the robust Hausman statistic (*HR-test*) obtained using the auxiliary regression detailed in Section (3):

$$hr = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right)' \left[\widehat{Var\left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right)}\right]^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}\right),$$

with different sample sizes. Figures 6 to 11 contained in Appendix 6 illustrate the relative power functions. The significance level has been fixed at 5%. $rho\hat{}$ is the estimated level of correlation between x and u conditioned upon w. For each level of rho, H_pow and HR_pow indicate the percentage of times we reject a false hypothesis if we use respectively the *H-test* or the *HR-test*.

In Table 1, 2 and 3 the number of cross-sectional units is held fixed at 25 and the number of time periods is varied respectively between 4, 10 and 20. In Table 4, 5 and 6 the number of cross-sectional units is held fixed at 275 and the number of time periods is varied respectively between 4, 10 and 20. Table 1 to 4 show that the performance of the *HR-test* is comparable with the one of the *H-test*, even better for values of *rho* greater than 0.3. In larger samples (Table 5 and 6) the performance of the *H-test* is superior but the power loss of the *HR-test* is not serious. The *HR-test* gives a very high rejection frequency for the false hypothesis of absence of correlation between x and u, starting from levels of correlation around 0.3 (86.5% and 87.4% respectively in Table 5 and 6) and it detects the endogeneity problem almost surely as soon as *rho* is higher than 0,4 (97.9% and 98.9% respectively in Table 5 and 6). Taking the results as a whole, the simulation experiment provides evidence that the performance of the *HR-test* in terms of power is satisfying in large samples and even better than the one given by the *H-test* in small samples.

In addition, it is worthwhile noting that a version of the Hausman test implemented in most econometric software, which is generally used in empirical studies, is the one based on the comparison between $\widehat{\beta}_{wg}$ and $\widehat{\beta}_{BN}$, i.e.

$$h = \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right)' \left(\widehat{V}_{wg} - \widehat{V}_{bg}\right)^{-1} \left(\widehat{\beta}_{wg} - \widehat{\beta}_{BN}\right).$$

The problem related with this approach is that, in finite samples, the difference between the two estimated variance-covariance matrices of the parameters estimates (i.e. $\widehat{V}_{wg} - \widehat{V}_{bg}$) may not be positive definite. In this cases, the use of a code implementing a different Hausman statistic or the formulation of the Hausman test using an auxiliary regressions (e.g. the one proposed by Davidson and McKinnon (1993, p. 236), which is now already implemented in some statistical packages, or the (robust) one presented in this paper) are the only possibilities to get a test outcome.

7 Conclusions

This paper has presented a methodological revision on the use of the Hausman test for correlated effects with panel data. The relevance of the discussion is both theoretical and empirical.

From a theoretical point of view, it is shown that the assumptions in Lemma 2.1. in Hausman (1978) are sufficient but not necessary. The main result is that the attainment of the absolute Fisher lower bound can be replaced by the attainment of a relative minimum variance bound.

From an empirical point of view, the main implication of this paper is a *caveat* on the use of the standard Hausman test framework for correlated effects in applied panel data studies. Our claim is that the application of this test is often not correct from a methodological point of view. The assumptions underlying the construction of the Hausman statistic (Hausman, 1978) may be rarely satisfied in empirical work. An analytical investigation of the size of the test shows that, at least in some cases, the distortion is substantial. The econometrics of panel data offers a variety of estimators for the same parameters. Our recommendation is to use the Hausman test framework for the comparison of appropriate panel data estimators, but to construct a version of the test robust to deviations from the classical errors assumption. This test, the *HR-test*, gives correct significance levels in common cases of misspecification of the variance-covariance matrix and has a power comparable to the Hausman test when no evidence of misspecification is present. The power of the *HR-test* is even higher in small samples. It can be easily implemented using a standard econometric package.

References

[1] Arellano, M. (1993). On the Testing of Correlated Effects with Panel Data, *Journal of Econometrics*, 59, 87-97.

[2] Ahn, S.C. and Low, S. (1996). A Reformulation of the Hausman Test for Regression Models with Pooled Cross Sectional Time Series Data, *Journal of Econometrics*, 71, 309-319.

[3] Balestra, P. and Nerlove, M. (1966). Pooling Cross Section and Time Series Data, *Econometrica*, 34, 585-612.

[4] Baltagi, B.H. (1996). Testing for Random Individual and Time Effects using a Gauss-Newton Regression, *Economics Letters*, 50, 189-192.

[5] Baltagi, B.H. (1997). Problems and Solutions, *Econometric Theory*, 13, 757.

[6] Baltagi, B.H. (1998). Problems and Solutions, *Econometric Theory*, 14, 527.

[7] Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*, Oxford University Press.

[8] Durbin, J. (1954). Errors in Variables. *Review of the International Statistical Institute*, 22, 23-32.

[9] Golub, G.H. and Van Loan, C.F. (1983). *Matrix Computations*, third edition, North Oxford Academic Press.

[10] Griliches, Z. and Hausman, J.A. (1986). Error in Variables in Panel Data, *Journal of Econometrics*, 31, 93-118.

[11] Hansen, L. (1982). Large Sample Properties of Generalized Method of Moment Estimators, *Econometrica*, 50, 517-533.

[12] Hausman, J.A. (1978). Specification Tests in Econometrics, *Econometrica*, 46, 1251-1271.

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

[14] Holly, A. (1982). A Remark on Hausman's Specification Test, *Econometrica*, 50, 749-760.

[15] Kakwani, N.C. (1967). The Unbiasedness of Zellner's Seemingly Unrelated Regression Equation Estimators, *Journal of the American Statistical Association*, 82, 141-2

[16] Maddala, G.S. (1971). The Use of Variance Components Models in Pooling Cross Section and Time Series Data, E*conometrica*, 39, 341-358.

[17] Magnus, J.R. and H. Neudecker, (1988) *Matrix Differential Calculus* John Wiley and Sons.

[18] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*, New York: John Wiley and Sons.

[19] Nerlove, M. (1971). A Note on Error Components Models, *Econometrica*, 39, 383-396.

[20] Patacchini, E. (2002). Unobservable Factors and Panel Data Sets: the Case of Matched Employee-Employer Data, Discussion Papers in Economics and Econometrics n.0202, University of Southampton.

[21] Johnson, N.L. and Kotz, S. (1970). *Continuous Univariate Distributions*, New York: John Wiley and Sons.

[22] Vuong, Q. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, 57, 307-333.

[23] White, H. (1984). *Asymptotic Theory for Econometricians*, Academic Press, New York.

[24] Wu, D. (1973). Alternatives Tests of Independence between Stochastic Regressors and Disturbances. *Econometrica*, 41, 733-750.

8 Appendix 1

Lemma 4 *If t_1 and t_2 are unbiased estimators of $\theta \in R^p$, with t_1 minimum variance (MV) at least in the set*

$$\mathcal{T} = \{t : t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2\}$$

then

$$Cov(t_1, t - t_1) = \mathbf{0}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{0}$ a null matrix, and $\mathbf{A} \in R^{p\times p}$ is fixed.

Proof.

$$\begin{aligned} t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2 &= t_1 + (\mathbf{I} - \mathbf{A})(t_2 - t_1) \\ &= t_1 + \mathbf{B}d, \text{ say, } \mathbf{B} \in R^{p\times p} \end{aligned}$$

$$\begin{aligned} Var(t) &= E\{[t_1 - \theta + \mathbf{B}d]\,[t_1 - \theta + \mathbf{B}d]'\} \\ &= Var(t_1) + Cov(t_1, d)\mathbf{B}' + \mathbf{B}Cov(d, t_1) + \mathbf{B}Var(d)\mathbf{B}'. \end{aligned}$$

Thus we can write

$$Var(t) - Var(t_1) = \mathbf{C}\mathbf{B}' + \mathbf{B}\mathbf{C}' + \mathbf{B}\mathbf{D}\mathbf{B}'.$$

The minimum variance property of t_1 implies this difference is positive semi-definite, and thus for every $\lambda \in R^p$, and $\mathbf{B} \in R^{p\times p}$,

$$Q = \lambda' \left(\mathbf{C}\mathbf{B}' + \mathbf{B}\mathbf{C}' + \mathbf{B}\mathbf{D}\mathbf{B}'\right) \lambda \geq 0.$$

However, for the particular case of

$$\mathbf{B} = -\mathbf{C}\mathbf{D}^{-1}$$

$$\begin{aligned} Q &= \lambda'(-\mathbf{C}\mathbf{D}^{-1}\mathbf{C}' - \mathbf{C}\mathbf{D}^{-1}\mathbf{C}' + \mathbf{C}\mathbf{D}^{-1}\mathbf{D}\mathbf{D}^{-1}\mathbf{C}')\lambda \\ &= \lambda'(-\mathbf{C}\mathbf{D}^{-1}\mathbf{C}')\lambda \end{aligned}$$

which satisfies the required inequality if and only if

$$\mathbf{C} = \mathbf{0}.$$

Further, for any $\mathbf{B} \in R^{p\times p}$

$$\begin{aligned} t - t_1 &= \mathbf{B}d, \\ Cov(t_1, t - t_1) &= \mathbf{C}\mathbf{B}' = \mathbf{0}. \end{aligned}$$

■

Remark 5 *We exclude the case where* $\mathbf{D}$ *is singular, as in that case replacing* $\mathbf{D}^{-1}$ *with a pseudo-inverse* $\mathbf{D}^{+}$ *such that* $\mathbf{D}^{+}\mathbf{D}\mathbf{D}^{+} = \mathbf{D}^{+}$ *reveals that all that is required is* $\mathbf{C}\mathbf{D}^{+}\mathbf{C}' = \mathbf{0}$*, or that* $\mathbf{C}$ *has rows orthogonal to the eigenvectors of* $\mathbf{D}$ *corresponding to the non-zero roots. As an example, consider the case where some elements of* t_1 *and* t_2 *coincide. It is simplest to exclude the coincident elements, and apply the argument above to the reduced vectors so formed.*

Remark 6 *This lemma implies that the MV unbiased estimator is uncorrelated with its difference from any other unbiased estimator, and the MV linear unbiased estimator is uncorrelated similarly.*

We next show that a set of the form $\mathcal{T}$ in Lemma 1 contains a minimum variance estimator. First, it is convenient to re-write the basis of the set in terms of t_1 and t_3, where $Cov(t_3, t_1) = \mathbf{0}$.

Lemma 7 *If* t_1 *and* t_2 *are unbiased estimators of* $\theta \in R^p$ *with covariance matrix* $\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$*, the set*

$$\mathcal{T} = \{t : t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2\}$$

can also be defined in terms of t_1 *and*

$$t_3 = \mathbf{B}t_1 + (\mathbf{I} - \mathbf{B})t_2$$

where

$$Cov(t_3, t_1) = \mathbf{0}$$

as

$$\mathcal{T} = \{t : t = \mathbf{C}t_1 + (\mathbf{I} - \mathbf{C})t_3\}$$

with

$$\mathbf{B} = -\mathbf{V}_{21}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}, \mathbf{I} - \mathbf{B} = \mathbf{V}_{11}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}$$
$$Var(t_3) = -\mathbf{D}\mathbf{V}_{11}^{-1}\mathbf{D}' + \mathbf{D}\mathbf{V}_{21}^{-1}\mathbf{V}_{22}\mathbf{V}_{12}^{-1}\mathbf{D}', \mathbf{D} = \left[\mathbf{V}_{21}^{-1} - \mathbf{V}_{11}^{-1}\right]^{-1}$$
$$\mathbf{C} = \mathbf{A}(\mathbf{V}_{11} - \mathbf{V}_{21}) + \mathbf{V}_{21})\mathbf{V}_{11}^{-1}, \mathbf{I} - \mathbf{C} = (\mathbf{I} - \mathbf{A})(\mathbf{V}_{11} - \mathbf{V}_{12})\mathbf{V}_{11}^{-1}$$
$$Var(t) = \mathbf{C}\mathbf{V}_{11}\mathbf{C}' + (\mathbf{I} - \mathbf{C})Var(t_3)(\mathbf{I} - \mathbf{C})'$$

Proof.

$$\begin{aligned} Cov(t_3, t_1) &= E\{[\mathbf{B}t_1 + (\mathbf{I} - \mathbf{B})t_2 - \theta][t_1 - \theta]'\} \\ &= \mathbf{B}\mathbf{V}_{11} + (\mathbf{I} - \mathbf{B})\mathbf{V}_{21} \\ &= -\mathbf{V}_{21}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}\mathbf{V}_{11} + \mathbf{V}_{11}(\mathbf{V}_{11} - \mathbf{V}_{21})^{-1}\mathbf{V}_{21} \end{aligned}$$

Now

$$\left[\mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{21}\right]^{-1} = \mathbf{V}_{21}^{-1}(\mathbf{V}_{11}-\mathbf{V}_{21})\mathbf{V}_{11}^{-1} = \mathbf{V}_{21}^{-1} - \mathbf{V}_{11}^{-1}$$

and

$$\left[\mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{11}\right]^{-1} = \mathbf{V}_{11}^{-1}(\mathbf{V}_{11}-\mathbf{V}_{21})\mathbf{V}_{21}^{-1} = \mathbf{V}_{21}^{-1} - \mathbf{V}_{11}^{-1}.$$

It follows that

$$\mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{21} = \mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{11} \quad (1.1)$$

and thus

$$Cov(t_3, t_1) = \mathbf{0}$$

To find $Var(t_3)$, as

$$t_3 = \mathbf{B}t_1 + (\mathbf{I}-\mathbf{B})t_2$$

$$Var(t_3) = \mathbf{B}\mathbf{V}_{11}\mathbf{B}' + (\mathbf{I}-\mathbf{B})\mathbf{V}_{21}\mathbf{B}' + \mathbf{B}\mathbf{V}_{12}(\mathbf{I}-\mathbf{B})' + (\mathbf{I}-\mathbf{B})\mathbf{V}_{22}(\mathbf{I}-\mathbf{B})'$$

$$\mathbf{B}\mathbf{V}_{11}\mathbf{B}' = \mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1\prime}\mathbf{V}_{21}'$$

$$(\mathbf{I}-\mathbf{B})\mathbf{V}_{21}\mathbf{B}' = -\mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1\prime}\mathbf{V}_{21}'$$

Identity (40) implies equality between these expressions.

$$\mathbf{B}\mathbf{V}_{12}(\mathbf{I}-\mathbf{B})' = -\mathbf{V}_{21}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{12}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1\prime}\mathbf{V}_{11}'$$

Transposing (40), this becomes the same as the expression for $\mathbf{B}\mathbf{V}_{11}\mathbf{B}'$.

$$(\mathbf{I}-\mathbf{B})\mathbf{V}_{22}(\mathbf{I}-\mathbf{B})' = \mathbf{V}_{11}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1}\mathbf{V}_{22}(\mathbf{V}_{11}-\mathbf{V}_{21})^{-1\prime}\mathbf{V}_{11}'$$

This suggests writing the matrix in (40) as

$$\mathbf{D} = \left[\mathbf{V}_{21}^{-1} - \mathbf{V}_{11}^{-1}\right]^{-1}$$

to give

$$Var(t_3) = -\mathbf{D}\mathbf{V}_{11}^{-1}\mathbf{D}' + \mathbf{D}\mathbf{V}_{21}^{-1}\mathbf{V}_{22}\left(\mathbf{V}_{21}'\right)^{-1}\mathbf{D}'$$

■

Remark 8 *Again, we are assuming non-singularity, in particular of* $\mathbf{V}_{21}$*. One could apply the steps above to zero a single non-zero element of* $\mathbf{V}_{21}$*, by shrinking* t_1 *and* t_2 *to the corresponding elements. Repeated application would then replace* $\mathbf{V}_{21}$ *with a null matrix.*

We can now show that $\mathcal{T}$ always contains a minimum variance unbiased estimator.

Lemma 9 *If t_1 and t_2 and $\mathcal{T}$ are as in Lemma (2) but with $\mathbf{V}_{12} = \mathbf{0}$ then t has the minimum variance in $\mathcal{T}$ if*

$$\mathbf{A} = [\mathbf{V}_{11}^{-1} + \mathbf{V}_{22}^{-1}]^{-1}\mathbf{V}_{11}^{-1}$$

Proof. Let this value of t be t_M, the corresponding $\mathbf{A}$ be $\mathbf{A}_M$, and $\mathbf{V}_M = Var(t_M)$. Let

$$\mathbf{A}_M = \mathbf{E}\mathbf{V}_{11}^{-1}, \Rightarrow \mathbf{I} - \mathbf{A}_M = \mathbf{E}\mathbf{V}_{22}^{-1}$$

We have

$$\begin{aligned} Var(t_M) &= \mathbf{E}\mathbf{V}_{11}^{-1}\mathbf{V}_{11}\mathbf{V}_{11}^{-1}\mathbf{E} + \mathbf{E}\mathbf{V}_{22}^{-1}\mathbf{V}_{22}\mathbf{V}_{22}^{-1}\mathbf{E} \\ &= \mathbf{E}\left[\mathbf{V}_{11}^{-1} + \mathbf{V}_{22}^{-1}\right]\mathbf{E} = \mathbf{E}. \end{aligned}$$

Moreover,

$$\begin{aligned} Cov(t_M, t_1 - t_2) &= Cov(\mathbf{A}_M t_1 + (\mathbf{I} - \mathbf{A}_M)t_2, t_1 - t_2) \\ &= E[\{\mathbf{E}\mathbf{V}_{11}^{-1}(t_1 - \theta) + \mathbf{E}\mathbf{V}_{22}^{-1}(t_2 - \theta)\}\{t_1' - t_2'\}] \\ &= E[\mathbf{E}(\mathbf{V}_{11}^{-1}\mathbf{V}_{11} - \mathbf{V}_{22}^{-1}\mathbf{V}_{22})] = \mathbf{0}. \end{aligned}$$

If $t \in \mathcal{T}$,

$$\begin{aligned} t &= \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2 \\ &= (\mathbf{A}_M + \mathbf{A} - \mathbf{A}_M)t_1 + (\mathbf{I} - \mathbf{A}_M - \mathbf{A} + \mathbf{A}_M)t_2 \\ &= t_M + (\mathbf{A} - \mathbf{A}_M)(t_1 - t_2) \end{aligned}$$

Thus

$$Var(t) = Var(t_M) + (\mathbf{A} - \mathbf{A}_M)Var(t_1 - t_2)(\mathbf{A} - \mathbf{A}_M)'$$

and thus $Var(t)$ exceeds $Var(t_M)$ by a positive semi-definite difference, and thus t_M is the minimum variance estimator in $\mathcal{T}$. ■

Finally, we establish the large sample equivalent of Lemma 1.

Lemma 10 *Consider $t_*' = [t_1', t_2'], \theta_*' = [\theta', \theta']$*

$$\sqrt{n}(t_* - \theta_*) \xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix})$$

where $\mathbf{V}_{11}$ is the 'asymptotic variance', Avar, of t_1 and $\mathbf{V}_{12}$ is the 'asymptotic covariance' of t_1 and t_2, $Acov(t_1, t_2)$. If t_1 is asymptotically minimum variance at least in the class

$$\mathcal{T} = \{t : t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2\}, \mathbf{A} \in R^{p \times p}, \text{ fixed,}$$

then if $t_d' = [t_1', [t - t_1]'], \theta_d' = [\theta', \mathbf{0}']$

$$\sqrt{n}(t_d - \theta_d) \xrightarrow{D} (0, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{0} \\ \mathbf{0} & Var(t) - \mathbf{V}_{11} \end{bmatrix})$$

Proof. Let $t_d = \begin{bmatrix} t_1 \\ t_2 - t_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$, so, as $\theta_d = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \theta_*$,

$$\begin{aligned}\sqrt{n}(t_d - \theta_d) &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \sqrt{n}(t_* - \theta_*) \\ &\xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} - \mathbf{V}_{11} \\ \mathbf{V}_{21} - \mathbf{V}_{11} & \mathbf{V}_{11} - \mathbf{V}_{12} - \mathbf{V}_{21} + \mathbf{V}_{22} \end{bmatrix}) \\ &\xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{C} \\ -\mathbf{C}' & \mathbf{D} \end{bmatrix}), \text{ say}\end{aligned}$$

:

$$\begin{aligned} t = \mathbf{A}t_1 + (\mathbf{I} - \mathbf{A})t_2 &= t_1 + (\mathbf{I} - \mathbf{A})(t_2 - t_1) \\ &= t_1 + \mathbf{B}d, \text{ say, } \mathbf{B} \in R^{p \times p} \end{aligned}$$

$$\begin{bmatrix} t_1 \\ t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{B} \end{bmatrix} t_d$$

$$\theta_* = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{B} \end{bmatrix} \theta_d$$

$$\begin{aligned}\sqrt{n}(\begin{bmatrix} t_1 \\ t \end{bmatrix} - \theta_*) &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{B} \end{bmatrix} \sqrt{n}(t_d - \theta_d) \\ &\xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{11} + \mathbf{C}\mathbf{B}' \\ \mathbf{V}_{11} + \mathbf{B}\mathbf{C}' & \mathbf{V}_{11} + \mathbf{B}\mathbf{D}\mathbf{B}' + \mathbf{B}\mathbf{C}' + \mathbf{C}\mathbf{B}' \end{bmatrix})\end{aligned} \tag{40}$$

so we can write

$$Avar(t) - Avar(t_1) = \mathbf{C}\mathbf{B}' + \mathbf{B}\mathbf{C}' + \mathbf{B}\mathbf{D}\mathbf{B}'.$$

The minimum variance property of t_1 implies this difference is positive semi-definite, and thus for every $\lambda \in R^p$, and $\mathbf{B} \in R^{p \times p}$,

$$Q = \lambda' \left(\mathbf{C}\mathbf{B}' + \mathbf{B}\mathbf{C}' + \mathbf{B}\mathbf{D}\mathbf{B}'\right) \lambda \geq 0.$$

However, for the particular case of

$$\mathbf{B} = -\mathbf{C}\mathbf{D}^{-1}$$

$$\begin{aligned} Q &= \lambda'(-\mathbf{C}\mathbf{D}^{-1}\mathbf{C}' - \mathbf{C}\mathbf{D}^{-1}\mathbf{C}' + \mathbf{C}\mathbf{D}^{-1}\mathbf{D}\mathbf{D}^{-1}\mathbf{C}')\lambda \\ &= \lambda'(-\mathbf{C}\mathbf{D}^{-1}\mathbf{C}')\lambda \end{aligned}$$

which satisfies the required inequality if and only if

$$\mathbf{C} = \mathbf{0}.$$

Further, for any $\mathbf{B} \in R^{p\times p}$

$$t - t_1 = \mathbf{B}d,$$

so as

$$\begin{bmatrix} t_1 \\ t - t_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \begin{bmatrix} t_1 \\ d \end{bmatrix}$$

$$\begin{aligned} \sqrt{n}(\begin{bmatrix} t_1 \\ t - t_1 \end{bmatrix} - \theta_d) &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \sqrt{n}(t_d - \theta_d) \\ &\xrightarrow{D} (\mathbf{0}, \begin{bmatrix} \mathbf{V}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{BDB}' \end{bmatrix}) \end{aligned}$$

where, as $\mathbf{C} = \mathbf{0}, \mathbf{V}_{11} = \mathbf{V}_{12} = \mathbf{V}_{21}, \mathbf{D} = \mathbf{V}_{22} - \mathbf{V}_{11}$. Moreover, from (40)

$$Var(t) = \mathbf{BDB}' + \mathbf{V}_{11} \Rightarrow Var(t - t_1) = \mathbf{BDB}' = Var(t) - Var(t_1)$$

as required. ■

Remark 11 *The assumption that* $\mathbf{A}$ *is fixed can be replaced by a stochastic matrix* $\mathbf{A}_n$ *with* $plim(\mathbf{A}_n) = \mathbf{A}$

Remark 12 *This lemma implies that an asymptotically MV consistent estimator is uncorrelated in large samples with its difference from any other consistent estimator.*

9 Appendix 2

In this Appendix we give further details about the expression for $Var(y_i)$ used in Section 2.

As

$$Var(y_i) = \Omega_i = \sigma^2 I_T + \sigma_\eta^2\ \iota\iota',$$

we can use the formula (see, e.g., Golub and van Loan (1983, p.50))

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1}U)^{-1}V^T A^{-1}$$

which simplifies for vector u, v to

$$(A + uv^T)^{-1} = A^{-1} - \frac{1}{1 + v^T A^{-1} u} A^{-1} u v^T A^{-1}.$$

It follows that, if $\theta = \sigma_\eta^2/\sigma^2$

$$\begin{aligned} \Omega_i &= \sigma^2[I_T + \theta\iota\iota'] = \sigma^2 \left[I_T - \frac{\theta}{1 + T\theta}\iota\iota' \right]^{-1} \\ &= \sigma^2 \left[I_T - \frac{\sigma_\eta^2}{\sigma^2 + T\sigma_\eta^2}\iota\iota' \right]^{-1}. \end{aligned}$$

10 Appendix 3

In Section 2 we focused our attention on Hausman test constructed using the contrast between the *Within Groups* and the *Balestra-Nerlove* estimator. In this Appendix we show the derivation of the Hausman statistic for the comparison between the *Within Groups* and the *Between Groups* estimator. Using the notation in Section 2, the *Between Groups* estimator can be written as

$$\widehat{\beta}_{bg} = (X'MX)X'MY.$$

The variance is

$$Var(\widehat{\beta}_{bg}) = \left[X'MX\right]^{-1} X'M(VarY)M'X \left[X'MX\right]^{-1}. \tag{41}$$

Further

$$\begin{aligned} Var(M^{+}y_i) &= M^{+}Var(y_i)M^{+'} = \sigma^2 M^{+}\left[I_T + \theta\iota\iota'\right]M^{+} && (42)\\ &= \sigma^2 M^{+}\left[I_T + \theta T M^{+}\right]M^{+} = \sigma^2(1+\theta T)M^{+}, && (43)\end{aligned}$$

where $\theta = \sigma_\eta^2/\sigma^2$. Thus

$$Var(MY) = \sigma^2(1+\theta T)I_N \otimes M^{+} = \sigma^2(1+\theta T)M.$$

Plugging (16) in (15), we obtain

$$\begin{aligned} Var(\widehat{\beta}_{bg}) &= \sigma^2(1+\theta T)\left[X'MX\right]^{-1} X'MX \left[X'MX\right]^{-1}\\ &= \sigma^2(1+\theta T)\left[X'MX\right]^{-1}. && (44)\end{aligned}$$

In addition

$$\begin{aligned} Cov(\widehat{\beta}_{bg}, \widehat{\beta}_{wg}) &= \left[X'MX\right]^{-1} X'M(VarY)Q'X\left[X'QX\right]^{-1} && (45)\\ &= \sigma^2\left[X'MX\right]^{-1} X'M\left[I_{NT} + \theta TM\right]QX\left[X'QX\right]^{-1} = 0 && (46)\end{aligned}$$

So

$$\begin{aligned} Var(\widehat{\beta}_{bg} - \widehat{\beta}_{wg}) &= Var(\widehat{\beta}_{bg}) + Var(\widehat{\beta}_{wg})\\ &= \sigma^2(1+\theta T)\left[X'MX\right]^{-1} + \sigma^2\left[X'QX\right]^{-1}.\end{aligned}$$

Thus we have as a test

$$(\widehat{\beta}_{wg} - \widehat{\beta}_{bg})\left[\sigma^2(1+\theta T)\left[X'MX\right]^{-1} + \sigma^2\left[X'QX\right]^{-1}\right]^{-1}(\widehat{\beta}_{wg} - \widehat{\beta}_{bg}).$$

11 Appendix 4

Lemma 13 *If*

$$\widehat{\beta} = (X'X)^{-1}X'y, \widehat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y,$$
$$X^* = XA, |A| \neq 0, \widehat{\varepsilon} = y - X\widehat{\beta}, \widehat{\varepsilon}^* = y - X^*\widehat{\beta}^*,$$

then

$$(X^{*'}X^*)^{-1} = A^{-1}(X'X)^{-1}A'^{-1}$$
$$\widehat{\beta}^* = A^{-1}\widehat{\beta}$$
$$\widehat{\varepsilon}^* = \widehat{\varepsilon}$$

Proof.

$$(X^{*'}X^*)^{-1} = (A'X'XA)^{-1} = A^{-1}(X'X)^{-1}A'^{-1}.$$
$$\widehat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y = A^{-1}(X'X)^{-1}A'^{-1}A'X'y = A^{-1}\widehat{\beta}.$$
$$\widehat{\varepsilon}^* = y - X^*\widehat{\beta}^* = y - XAA^{-1}\widehat{\beta} = y - X\widehat{\beta} = \widehat{\varepsilon}$$

■

Lemma 14 *If*

$$\widehat{\beta}_A = (X_A'X_A)^{-1}X_A'y_A, \widehat{\beta}_B = (X_B'X_B)^{-1}X_B'y_B,$$
$$\widehat{\varepsilon}_A = y_A - X_A\widehat{\beta}_A, \widehat{\varepsilon}_B = y_B - X_B\widehat{\beta}_B$$
$$X^* = \begin{bmatrix} X_A & X_A \\ 0 & X_B \end{bmatrix}, y^* = \begin{bmatrix} y_A \\ y_B \end{bmatrix}, \widehat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y$$
$$\widehat{\varepsilon}^* = y^* - X^*\widehat{\beta}^*$$

then

$$\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_A - \widehat{\beta}_B \\ \widehat{\beta}_B \end{bmatrix}, \widehat{\varepsilon}^* = \begin{bmatrix} \widehat{\varepsilon}_A \\ \widehat{\varepsilon}_B \end{bmatrix}$$

Proof. Let

$$X = \begin{bmatrix} X_A & 0 \\ 0 & X_B \end{bmatrix}, \Rightarrow X^* = X\begin{bmatrix} I & I \\ 0 & I \end{bmatrix} = XA \text{ say}$$

$$A^{-1} = \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix}$$

Further, it is an exercise in elementary matrix algebra to show that

$$\widehat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} \widehat{\beta}_A \\ \widehat{\beta}_B \end{bmatrix}, \widehat{\varepsilon} = y - X\widehat{\beta} = \begin{bmatrix} \widehat{\varepsilon}_A \\ \widehat{\varepsilon}_B \end{bmatrix}.$$

So applying Lemma 13,

$$\widehat{\beta}^* = A^{-1}\widehat{\beta} = \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} \widehat{\beta}_A \\ \widehat{\beta}_B \end{bmatrix} = \begin{bmatrix} \widehat{\beta}_A - \widehat{\beta}_B \\ \widehat{\beta}_B \end{bmatrix}$$

and

$$\widehat{\varepsilon}^* = \widehat{\varepsilon} = \begin{bmatrix} \widehat{\varepsilon}_A \\ \widehat{\varepsilon}_B \end{bmatrix}$$

■

Return now to model (23). Results (24) and (25) in Lemma 2 directly follow from the application of Lemma 13 and 14. Next, we will prove the remaining result in Lemma 2, i.e. equality (26).

Let

$$\begin{aligned} H^+ &= \frac{1}{T}i', H = I_N \otimes H^+, H'H = \frac{1}{T}M \\ \widehat{\beta}_{bg} &= [(HX)'(HX)]^{-1}(HX)'(HY) = (X'MX)^{-1}X'MY \\ \widehat{\beta}_{wg} &= [(QX)'(QX)]^{-1}(QX)'(QY) = (X'QX)^{-1}X'QY \end{aligned}$$

Further, let G^+ be Arellano and Bover's (1990) forward orthogonal deviations matrix, $(T-1) \times T$, such that

$$\begin{aligned} G^+i &= 0, G^+G^{+\prime} = I_{(T-1)}, G^{+\prime}G^+ = Q^+ = I_T - \frac{1}{T}ii' \\ G &= I_N \otimes G^+, G'G = Q, GG' = I_N \otimes I_{(T-1)} = I_{N(T-1)} \\ \widehat{\beta}_{wg} &= [(GX)'(GX)]^{-1}(GX)'(GY) = (X'QX)^{-1}X'QY \end{aligned}$$

and identifying HX and HY with X_A and Y_A, GX and GY with X_B and Y_B, we see that the artificial regression of $Y^* = \begin{bmatrix} HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} HX & HX \\ 0 & GX \end{bmatrix}$ gives coefficients $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{bg} - \widehat{\beta}_{wg} \\ \widehat{\beta}_{wg} \end{bmatrix}$. In this case,

$$Var(Y^*) = \begin{bmatrix} HVar(Y)H' & 0 \\ 0 & GVar(Y)G' \end{bmatrix}.$$

If $\theta = \sigma_\eta^2/\sigma^2$we have

$$\begin{aligned} GVar(Y)G' &= \sigma^2 G(I_{NT} + \theta I_N \otimes ii')G' \\ &= \sigma^2 GG' \text{ as } G^+i = 0 \\ &= \sigma^2 I_{N(T-1)} \end{aligned}$$

and

$$\begin{aligned} HVar(Y)H' &= \sigma^2 H(I_{NT} + \theta I_N \otimes ii')H' \\ &= \sigma^2 [I_N \otimes H^+](I_{NT} + \theta I_N \otimes ii')[I_N \otimes H^{+\prime}] \\ &= \sigma^2 [I_N \otimes (H^+ H^{+\prime}) + \theta I_N \otimes (H^+ ii' H^{+\prime}). \end{aligned}$$

As

$$H^+ = \frac{1}{T} i', H^+ i = 1, H^+ H^{+\prime} = \frac{1}{T},$$

$$HVar(Y)H' = \sigma^2 [\frac{1}{T} I_N + \theta I_N] = \frac{\sigma^2}{T}(1 + T\theta) I_N.$$

Assembling our results,

$$Var(Y^*) = \left[\begin{array}{cc} \frac{\sigma^2}{T}(1+T\theta)I_N & 0 \\ 0 & \sigma^2 I_{N(T-1)} \end{array} \right].$$

If now $\widetilde{X} = \left[\begin{array}{cc} HX & 0 \\ 0 & GX \end{array} \right]$,

$$\begin{aligned} Var(\widehat{\beta}^*) &= (X^{*\prime} X^*)^{-1} X^{*\prime} Var(Y^*) X^* (X^{*\prime} X^*)^{-1} \\ &= A^{-1} (\widetilde{X}' \widetilde{X})^{-1} \widetilde{X}' Var(Y^*) \widetilde{X} (\widetilde{X}' \widetilde{X})^{-1} A^{-1\prime}. \end{aligned}$$

Next, we calculate this variance by separating the different components.

$$\widetilde{X}' Var(Y^*) \widetilde{X} = \left[\begin{array}{cc} X'H' & 0 \\ 0 & X'G' \end{array} \right] \left[\begin{array}{cc} \frac{\sigma^2}{T}(1+T\theta)I_N & 0 \\ 0 & \sigma^2 I_{N(T-1)} \end{array} \right] \left[\begin{array}{cc} HX & 0 \\ 0 & GX \end{array} \right]$$

$$= \sigma^2 \left[\begin{array}{cc} X'H' & 0 \\ 0 & XG' \end{array} \right] \left[\begin{array}{cc} (\theta + 1/T)HX & 0 \\ 0 & GX \end{array} \right]$$

$$= \sigma^2 \left[\begin{array}{cc} (\theta/T + 1/T^2)X'MX & 0 \\ 0 & X'QX \end{array} \right].$$

$$(\widetilde{X}' \widetilde{X})^{-1} = \left[\begin{array}{cc} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{array} \right].$$

Thus

$$(\widetilde{X}' \widetilde{X})^{-1} \widetilde{X}' Var(Y^*) \widetilde{X}' (\widetilde{X}' \widetilde{X})^{-1}$$

$$= \sigma^2 \left[\begin{array}{cc} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{array} \right] \times$$

$$\left[\begin{array}{cc} (\theta/T + 1/T^2)X'MX & 0 \\ 0 & X'QX \end{array} \right] \left[\begin{array}{cc} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{array} \right]$$

$$= \sigma^2 \left[\begin{array}{cc} (T\theta + 1)(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{array} \right]$$

and

$$A^{-1} (\widetilde{X}' \widetilde{X})^{-1} \widetilde{X}' Var(Y^*) \widetilde{X}' (\widetilde{X}' \widetilde{X})^{-1} A^{-1\prime}$$

$$
\begin{aligned}
&= \sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} (T\theta+1)(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} (T\theta+1)(X'MX)^{-1} & -(X'QX)^{-1} \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} (T\theta+1)(X'MX)^{-1} + (X'QX)^{-1} & -(X'QX)^{-1} \\ -(X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix}. \qquad (47)
\end{aligned}
$$

We now need to find the variance-covariance matrix the artificial regression will assume. This will be proportional to

$$
\begin{aligned}
(X^{*\prime}X^*)^{-1} &= (A'\widetilde{X}'\widetilde{X}A)^{-1} = A^{-1}(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
&= \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} T(X'MX)^{-1} & -(X'QX)^{-1} \\ 0 & (X'QX)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \begin{bmatrix} T(X'MX)^{-1} + (X'QX)^{-1} & -(X'QX)^{-1} \\ -(X'QX)^{-1} & (X'QX)^{-1} \end{bmatrix}. \qquad (48)
\end{aligned}
$$

By comparing (47) with (48) it appears that an artificial regression is a valuable device to estimate a suitable variance-covariance matrix. This variance is estimated using a (White) robust *OLS* estimator which uses a consistent estimator of $X^{*\prime}Var(Y^*)X^*$ under the assumption that $Var(Y^*)$ is diagonal. Next, we derive this consistent estimator. Following the steps used in the derivation of $Var(\widehat{\beta}^*)$ above, we separate the different components.

$$
\begin{aligned}
&\widetilde{X}'Var(Y^*)\widetilde{X} \\
&= \begin{bmatrix} X'H' & 0 \\ 0 & X'G' \end{bmatrix} \begin{bmatrix} \sigma^2\Omega & 0 \\ 0 & \sigma^2\Omega \end{bmatrix} \begin{bmatrix} HX & 0 \\ 0 & GX \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} X'H' & 0 \\ 0 & X'G' \end{bmatrix} \begin{bmatrix} \Omega & 0 \\ 0 & \Omega \end{bmatrix} \begin{bmatrix} HX & 0 \\ 0 & GX \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} X'H'\Omega & 0 \\ 0 & X'G'\Omega \end{bmatrix} \begin{bmatrix} HX & 0 \\ 0 & GX \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} XH'\Omega HX & 0 \\ 0 & X'G'\Omega GX \end{bmatrix}.
\end{aligned}
$$

$$
(\widetilde{X}'\widetilde{X})^{-1} = \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix}.
$$

Thus

$$
(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}
$$

$$
\begin{aligned}
&= \sigma^2 \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \times \\
&\begin{bmatrix} XH'\Omega HX & 0 \\ 0 & X'G'\Omega GX \end{bmatrix} \begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} T(X'MX)^{-1}(XH'\Omega HX) & 0 \\ 0 & (X'QX)^{-1}(X'G'\Omega GX) \end{bmatrix} \times \\
&\begin{bmatrix} T(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1} \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} T^2(X'MX)^{-1}(XH'\Omega HX)(X'MX)^{-1} & 0 \\ 0 & (X'QX)^{-1}(X'G'\Omega GX)(X'QX)^{-1} \end{bmatrix}.
\end{aligned}
$$

Let

$$B = T^2(X'MX)^{-1}(XH'\Omega HX)(X'MX)^{-1}$$

and

$$D = (X'QX)^{-1}(X'G'\Omega GX)(X'QX)^{-1}.$$

So

$$
\begin{aligned}
& A^{-1}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Var(Y^*)\widetilde{X}'(\widetilde{X}'\widetilde{X})^{-1}A^{-1\prime} \\
&= \sigma^2 \begin{bmatrix} I & -I \\ 0 & I \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} B & -D \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} B+D & -D \\ -D & D \end{bmatrix}.
\end{aligned}
$$

The residuals from this regression of $Y^* = \begin{bmatrix} HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} HX & HX \\ 0 & GX \end{bmatrix}$ to give coefficients $\widehat{\beta}^* = \begin{bmatrix} \widehat{\beta}_{bg} - \widehat{\beta}_{wg} \\ \widehat{\beta}_{wg} \end{bmatrix}$ can be obtained by stacking those from HY on HX above those from GY on GX. The first set will yield sum of squares

$$
\begin{aligned}
RSS_A &= (HY)'[I_N - (HX)T(X'MX)^{-1}(X'H')]HY \\
&= \frac{1}{T}Y'(M - MX(X'MX)^{-1}X'M)Y.
\end{aligned}
$$

Note that $(M - MX(X'MX)^{-1}X'M) = M_P$ is idempotent, and $M_P MX = 0$.

Note than if the write the model as

$$Y = X\beta + E$$

we get

$$MY = MX\beta + ME,$$

$$M_P M Y = M_P E$$

and

$$RSS_A = \frac{1}{T} E' M_P E.$$

The expectation is given by

$$\begin{aligned} ERSS_A &= \frac{1}{T} trace\,[M_P Var(E)] = \frac{1}{T} trace\,[M_P Var(Y)] \\ &= \frac{\sigma^2}{T} trace\,[M_P \{I_{NT} + \theta I_N \otimes ii'\}]\,. \end{aligned}$$

As

$$M(I_N \otimes ii') = (I_N \otimes \frac{1}{T} ii')(I_N \otimes ii') = I_N \otimes ii' = TM,$$

$$ERSS_A = \frac{\sigma^2}{T}(1+\theta T) trace(M_P) = \frac{\sigma^2}{T}(1+\theta T)(N-K).$$

Similarly, if

$$\begin{aligned} RSS_B &= (GY)'[I_{NT} - GX(X'QX)^{-1}X'G']GY \\ &= Y'[Q - QX(X'QX)^{-1}X'Q)]Y, \end{aligned}$$

$$\begin{aligned} ERSS_B &= \sigma^2 trace\,[Q_P\{I_{NT} + \theta I_N \otimes ii'\}] \\ &= \sigma^2 trace\,[Q_P] = \sigma^2[N(T-1) - K]. \end{aligned}$$

Accordingly, there is no multiple of $RSS_A + RSS_B$ with expectation σ^2. However, if in the first regression Y_A and X_A are scaled by

$$k = \sqrt{T/(1+\theta T)}$$

the coefficients will be unchanged, their variance will be unchanged, $(X_A' X_A)^{-1}$ will be scaled by $1/k^2 = (1+\theta T)/T$. So instead of

$$[(HX)'HX]^{-1} = T(X'MX)^{-1}$$

we will now have

$$(X_A' X_A)^{-1} = (1+\theta T)(X'MX)^{-1}.$$

Further,

$$k^2 ERSS_A = \frac{T}{(1+\theta T)} \frac{\sigma^2}{T}(1+\theta T)(N-K) = \sigma^2(N-K)$$

and $(k^2 RSS_A + RSS_B)/(NT-2K)$ is an unbiased estimator of σ^2.

Thus given a consistent estimator $\widehat{\theta}$ of θ, and thus $\widehat{k}$ of k, we can construct the Hausman test by carrying out the artificial regression of $Y^* = \begin{bmatrix} \widehat{k}HY \\ GY \end{bmatrix}$ on $X^* = \begin{bmatrix} \widehat{k}HX & \widehat{k}HX \\ 0 & GX \end{bmatrix}$, and constructing a Wald test on the first K coefficients.

12 Appendix 5

We define between groups and within groups estimators as usual:

$$\begin{aligned}\widehat{\beta}_{BG} &= \left(X^{'}MX\right)^{-1}X^{'}MY \\ \widehat{\beta}_{WG} &= \left(X^{'}QX\right)^{-1}X^{'}QY\end{aligned}$$

where

$$\begin{aligned}Q &= I_N \otimes Q^{+}, \\ Q^{+} &= I_T - \frac{1}{T}ii^{'}, \\ M &= I_N \otimes M^{+}, \\ M^{+} &= \frac{1}{T}ii^{'} = I_T - Q^{+},\end{aligned}$$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}, \ Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \ X_i = \begin{bmatrix} x_{i1}^{\prime} \\ x_{i2}^{\prime} \\ \vdots \\ x_{iT}^{\prime} \end{bmatrix}, \ y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}.$$

Q^{+} is the matrix that transforms the data to deviations from the individual time mean, M^{+} is the matrix that transforms the data to averages.

Suppose the true model is

$$y_{it} = z_{it}^{'}\beta + \eta_i + v_{it}, \qquad i = 1, ..., N, \qquad t = 1, ..., T$$

where $z_{it}^{'}$ is a $1 \times K$ vector of theoretical variables, $\eta_i \sim iid\left(0, \sigma_\eta^2\right)$, $v_{it} \sim iid\left(0, \sigma^2\right)$ uncorrelated with the columns of z_{it} and $Cov\left(\eta_i, v_{it}\right) = 0$. The observed variables are

$$x_{it} = z_{it} + m_{it},$$

where m_{it} is a $K \times 1$ measurement error uncorrelated with η_i and v_{it}. The estimated model is

$$y_{it} = x_{it}^{'}\beta + \eta_i + v_{it} - m_{it}^{\prime}\beta, \qquad i = 1, ..., N, \qquad t = 1, ..., T.$$

In the case of exact measurement, $m_{it} = 0$. So

$$y_i = X_i\beta + \eta_i\mathbf{i} + \boldsymbol{\nu}_i - M_i\beta = X_i\beta + \zeta_i,$$

say, where $\mathbf{i}$ is a column of T 1s,

$$\boldsymbol{\nu}_i = \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iT} \end{bmatrix}, M_i = \begin{bmatrix} m_{i1}^{\prime} \\ m_{i2}^{\prime} \\ \vdots \\ m_{iT}^{\prime} \end{bmatrix}.$$

To consider the 'generic' estimator, let

$$\begin{aligned}\widehat{\beta}_{AG} &= \left(X'AX\right)^{-1} X'AY \\ &= \left[\sum_{i=1}^{N} X_i'A^+X_i\right]^{-1} \sum_{i=1}^{N} X_i'A^+y_i \\ &= \beta + \left[\sum_{i=1}^{N} X_i'A^+X_i\right]^{-1} \sum_{i=1}^{N} X_i'A^+\zeta_i,\end{aligned}$$

where $A = Q$ or M as appropriate, and

$$\begin{aligned}\sum_{i=1}^{N} X_i'A^+\zeta_i &= \sum_{i=1}^{N} X_i'A^+\left[\eta_i\mathbf{i} + \boldsymbol{\nu}_i - M_i\beta\right] \\ &= \sum_{i=1}^{N} \left[Z_i + M_i\right]' A^+ \left[\eta_i\mathbf{i} + \boldsymbol{\nu}_i - M_i\beta\right],\end{aligned}$$

where

$$Z_i = \begin{bmatrix} z_{i1}' \\ z_{i2}' \\ \vdots \\ z_{iT}' \end{bmatrix}.$$

Given our assumptions,

$$E\left[\sum_{i=1}^{N} X_i'A^+\zeta_i\right] = -E\left[\sum_{i=1}^{N} M_i'A^+M_i\beta\right].$$

Let

$$M_i = [M_{i1}, .., M_{iK}]$$

so the $r-s-$th element of $M_i'A^+M_i$ has expectation

$$tr(E[M_{ir}'A^+M_{is}]) = E[tr(M_{is}M_{ir}'A)] = tr(\sigma_{Mrs}I_TA) = \sigma_{Mrs}tr(A)$$

if we assume m_{ij} are possibly correlated only contemporaneously within groups. Thus

$$E\left[\sum_{i=1}^{N} X_i'A^+\zeta_i\right] = -\left[tr(A)\sum_{i=1}^{N}\Sigma_M\beta\right] = -tr(A)N\Sigma_M\beta.$$

If we write

$$\begin{aligned}X_i &= Z_i + M_i \\ X_i'A^+X_i &= (Z_i+M_i)'A^+(Z_i+M_i) = Z_i'A^+Z_i + Z_i'A^+M_i + M_i'A^+Z_i + M_i'A^+M_i\end{aligned}$$

and taking the Z as non-stochastic,

$$E(X_i' A^+ X_i) = Z_i' A^+ Z_i + E(M_i' A^+ M_i) = Z_i' A^+ Z_i + tr(A)\Sigma_M.$$

If we make appropriate assumptions about Z_i to ensure that $(1/N)\sum_{i=1}^N Z_i' A^+ Z_i$ converges to an appropriate limit, say Σ_{ZAZ}, then

$$\begin{aligned}
\widehat{\beta}_{AG} &= \beta + \left[\sum_{i=1}^N X_i' A^+ X_i\right]^{-1} \sum_{i=1}^N X_i' A^+ \zeta_i \\
&\xrightarrow{p} \beta + p\lim\left[\frac{1}{N}\sum_{i=1}^N X_i' A^+ X_i\right]^{-1} p\lim \frac{1}{N}\sum_{i=1}^N X_i' A^+ \zeta_i \\
&= \beta - \left[\Sigma_{ZAZ} + tr(A^+)\Sigma_M\right]^{-1} tr(A^+)\Sigma_M \beta.
\end{aligned}$$

For $\widehat{\beta}_{BG}$,

$$A^+ = M^+ = \frac{1}{T} ii' \Rightarrow tr(A) = 1$$

so

$$\widehat{\beta}_{BG} \xrightarrow{p} \beta - \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \Sigma_M \beta.$$

For $\widehat{\beta}_{WG}$,

$$A^+ = Q^+ = I_T - \frac{1}{T} ii' \Rightarrow tr(A) = T - 1$$

so

$$\begin{aligned}
&\widehat{\beta}_{WG} \xrightarrow{p} \beta - (T-1)\left[\Sigma_{ZQZ} + (T-1)\Sigma_M\right]^{-1} \Sigma_M \beta \\
&= \beta - \left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \Sigma_M \beta
\end{aligned}$$

These formulae are comparable, as Σ_{ZQZ} grows with T. Indeed, if $\Sigma_{ZQZ}/(T-1) \approx \Sigma_{ZMZ}$, that is, the between sum of squares and the within sum of squares are roughly proportional to the number of terms contributing to each, then

$$\delta = p\lim(\widehat{\beta}_{BG} - \widehat{\beta}_{WG}) \approx \mathbf{0}.$$

We turn next to the estimation of the variance of the disturbances. For the generic estimation

$$\begin{aligned}
\widehat{\varepsilon}_{AG} &= AY - AX\widehat{\beta}_{AG} = AY - AX\left(X'AX\right)^{-1} X'AY, \\
Y &= X\beta + \zeta
\end{aligned}$$

where

$$\zeta = \left[\begin{array}{ccc} \zeta_1', & \cdots & \zeta_N' \end{array}\right]'.$$

Substituting

$$\begin{aligned}\widehat{\varepsilon}_{AG} &= AX\beta + A\zeta - AX\left(X'AX\right)^{-1}X'A(X\beta+\zeta) \\ &= A\zeta - AX\left(X'AX\right)^{-1}X'A\zeta.\end{aligned}$$

Consider

$$\begin{aligned}\widehat{\varepsilon}'_{AG}\widehat{\varepsilon}_{AG} &= \zeta'A\zeta - \zeta'AX\left(X'AX\right)^{-1}X'A\zeta \\ &= \sum_{i=1}^{N}\zeta_i'A^+\zeta_i - \zeta'AX\left(X'AX\right)^{-1}X'A\zeta.\end{aligned}$$

As

$$A^+\zeta_i = A^+\left[\eta_i\mathbf{i} + \boldsymbol{\nu}_i - M_i\beta\right],$$

$$\frac{1}{N(T-1)}\sum_{i=1}^{N}\zeta_i'A^+\zeta_i \xrightarrow{p} \frac{1}{T-1}\left[\sigma_\eta^2\mathbf{i}'A^+\mathbf{i}+tr(A^+)\{\sigma^2+\beta'\Sigma_M\beta\}\right].$$

If $A^+ = Q^+ = I_T - \frac{1}{T}\mathbf{i}\mathbf{i}'$, $tr(A^+) = T-1$,

$$\frac{1}{N(T-1)}\sum_{i=1}^{N}\zeta_i'Q^+\zeta_i \xrightarrow{p} \{\sigma^2+\beta'\Sigma_M\beta\}.$$

If $A^+ = M^+ = \frac{1}{T}ii' = I_T - Q^+$, $tr(A^+) = 1$,

$$\frac{1}{N}\sum_{i=1}^{N}\zeta_i'M^+\zeta_i \xrightarrow{p} \left[\sigma_\eta^2\mathbf{i}'A^+\mathbf{i}+tr(A^+)\{\sigma^2+\beta'\Sigma_M\beta\}\right] = T\sigma_\eta^2+\sigma^2+\beta'\Sigma_M\beta.$$

The other component in the 'natural' variance estimate $\widehat{\varepsilon}'_{AG}\widehat{\varepsilon}_{AG}/(N(T-1))$ is

$$\zeta'AX\left(X'AX\right)^{-1}X'A\zeta$$

We are assuming that $(1/N)\sum_{i=1}^{N}Z_i'A^+Z_i$ converges to an appropriate limit, say Σ_{ZAZ}, and thus

$$\frac{1}{N}\sum_{i=1}^{N}X_i'A^+X_i \xrightarrow{p} \Sigma_{ZAZ} + tr(A^+)\Sigma_M$$

Further, that

$$p\lim\frac{1}{N}\sum_{i=1}^{N}X_i'A^+\zeta_i = tr(A^+)\Sigma_M\beta$$

Thus

$$\begin{aligned}
\frac{1}{N}\zeta' AX\left(X'AX\right)^{-1}X'A\zeta &= \left[\frac{1}{N}\sum_{i=1}^{N}X_i'A^{+}\zeta_i\right]'\left[\frac{1}{N}\sum_{i=1}^{N}X_i'A^{+}X_i\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}X_i'A^{+}\zeta_i \\
&\xrightarrow{p}\left[tr(A^{+})\Sigma_M\beta\right]'\left[\Sigma_{ZAZ}+tr(A^{+})\Sigma_M\right]^{-1}tr(A^{+})\Sigma_M\beta \\
&= tr(A^{+})\beta'\Sigma_M\left[\frac{1}{tr(A^{+})}\Sigma_{ZAZ}+\Sigma_M\right]^{-1}\Sigma_M\beta
\end{aligned}$$

If $A^{+}=Q^{+}=I_T-\frac{1}{T}\mathbf{ii}', tr(A^{+})=T-1,$

$$\frac{1}{N(T-1)}\zeta' QX\left(X'QX\right)^{-1}X'Q\zeta\xrightarrow{p}\beta'\Sigma_M\left[\frac{1}{(T-1)}\Sigma_{ZQZ}+\Sigma_M\right]^{-1}\Sigma_M\beta.$$

If $A^{+}=M^{+}=\frac{1}{T}ii'=I_T-Q^{+}, tr(A^{+})=1,$

$$\frac{1}{N}\zeta' MX\left(X'MX\right)^{-1}X'M\zeta\xrightarrow{p}\beta'\Sigma_M\left[\Sigma_{ZMZ}+\Sigma_M\right]^{-1}\Sigma_M\beta.$$

Thus

$$\frac{1}{N(T-1)}\widehat{\varepsilon}_{WG}'\widehat{\varepsilon}_{WG}\xrightarrow{p}\{\sigma^2+\beta'\Sigma_M\beta-\beta'\Sigma_M\left[\frac{1}{(T-1)}\Sigma_{ZQZ}+\Sigma_M\right]^{-1}\Sigma_M\beta\} \quad (49)$$

and

$$\frac{1}{N}\widehat{\varepsilon}_{BG}'\widehat{\varepsilon}_{BG}\xrightarrow{p}T\sigma_\eta^2+\sigma^2+\beta'\Sigma_M\beta-\beta'\Sigma_M\left[\Sigma_{ZMZ}+\Sigma_M\right]^{-1}\Sigma_M\beta. \quad (50)$$

Finally, we require $Var(\widehat{\beta}_{AG})$. We have

$$\widehat{\beta}_{AG}=\beta-\left[\Sigma_{ZAZ}+tr(A^{+})\Sigma_M\right]^{-1}tr(A^{+})\Sigma_M\beta.$$

and thus under appropriate assumptions

$$\begin{aligned}
&\sqrt{N}\left[\widehat{\beta}_{AG}-\beta+\left[\Sigma_{ZAZ}+tr(A^{+})\Sigma_M\right]^{-1}tr(A^{+})\Sigma_M\beta\right] \\
&\xrightarrow{D}N(0,\left[\Sigma_{ZAZ}+tr(A^{+})\Sigma_M\right]^{-1}\times \\
&Var\left[\frac{1}{\sqrt{N}}\left\{\sum_{i=1}^{N}X_i'A^{+}\zeta_i-tr(A^{+})\Sigma_M\beta\right\}\right]\left[\Sigma_{ZAZ}+tr(A^{+})\Sigma_M\right]^{-1}).
\end{aligned}$$

As

$$\sum_{i=1}^{N}X_i'A^{+}\zeta_i=\sum_{i=1}^{N}\left[Z_i+M_i\right]'A^{+}\left[\eta_i\mathbf{i}+\boldsymbol{\nu}_i-M_i\beta\right],$$

$$E\left[\sum_{i=1}^{N}X_i'A^{+}\zeta_i\right]=-\left[tr(A^{+})\sum_{i=1}^{N}\Sigma_M\beta\right]=-tr(A^{+})N\Sigma_M\beta.$$

$$
\begin{aligned}
Var\left[\sum_{i=1}^{N} X_i^{'} A^{+}\zeta_i\right] &= Var\left[\sum_{i=1}^{N}[Z_i' A^{+} + M_i' A^{+}]\left[\eta_i \mathbf{i} + \boldsymbol{\nu}_i - M_i\beta\right]\right] \\
&= Var\left[\sum_{i=1}^{N}\left[\begin{array}{c} Z_i' A^{+}\eta_i \mathbf{i} + Z_i' A^{+}\nu_i - Z_i' A^{+} M_i\beta + \\ M_i' A^{+}\eta_i \mathbf{i} + M_i' A^{+}\boldsymbol{\nu}_i - M_i' A^{+} M_i\beta \end{array}\right]\right]
\end{aligned}
$$

where $A^{+} = Q^{+} = I_T - \frac{1}{T}ii'$ or $= M^{+} = \frac{1}{T}ii' = I_T - Q^{+}$. We are assuming no correlation between groups. We thus need to evaluate

$$
\begin{aligned}
Var(X_i^{'} A^{+}\zeta_i) &= Var\left[\begin{array}{c} Z_i' A^{+}\eta_i \mathbf{i} + Z_i' A^{+}\nu_i - Z_i' A^{+} M_i\beta \\ + M_i' A^{+}\eta_i \mathbf{i} + M_i' A^{+}\boldsymbol{\nu}_i - M_i' A^{+} M_i\beta \end{array}\right] \\
&= Var[\mathbf{a} + \mathbf{b} + \mathbf{c} + \mathbf{d} + \mathbf{e} + \mathbf{f}]
\end{aligned}
$$

say, where

$$
E(X_i^{'} A^{+}\zeta_i) = -tr(A^{+})\Sigma_M\beta = E(\mathbf{f})
$$

and if $\mathbf{u}$ is a random vector,

$$
Var(\mathbf{u}) = \mathbf{E}\left\{[\mathbf{u} - E(\mathbf{u})][\mathbf{u} - E(\mathbf{u})]'\right\}.
$$

Thus

$$
\begin{aligned}
Var(X_i^{'} A^{+}\zeta_i) &= E(\mathbf{aa}') + E(\mathbf{bb}') + E(\mathbf{cc}') \\
&\quad + Cov(\mathbf{cf}') + Cov(\mathbf{fc}') + E(\mathbf{dd}') \\
&\quad + E(\mathbf{ee}') + Var(\mathbf{f}).
\end{aligned}
$$

as

$$
\begin{gathered}
E(\mathbf{ad}') = \sigma_\eta^2 Z_i' A^{+}\mathbf{ii}' A^{+} E(M_i) = \mathbf{0}, \\
E(\mathbf{be}') = \sigma^2 Z_i' A^{+} E(M_i) = \mathbf{0}, \\
E(\mathbf{cd}') = E(-Z_i' A^{+} M_i\beta v_i' A^{+} M_i) = \mathbf{0}, \\
E(\mathbf{ce}') = E(-Z_i' A^{+} M_i\beta \mathbf{i}\eta_i A^{+} M_i) = \mathbf{0}, \\
E(\mathbf{de}') = E(M_i' A^{+}\eta_i \mathbf{i}\boldsymbol{\nu}_i' A^{+} M_i) = \mathbf{0}, \\
Cov(\mathbf{df}') = Cov(M_i' A^{+}\eta_i \mathbf{i}, M_i' A^{+} M_i\beta') = \mathbf{0},
\end{gathered}
$$

and

$$
Cov(\mathbf{ef}') = E(M_i' A^{+}\boldsymbol{\nu}_i, M_i' A^{+} M_i\beta') = \mathbf{0},
$$

assuming that the appropriate fourth order cross moments are zero, or, more strongly, that ν_i, η_i, and M_i are independent. Of the 36 possible terms, 8 are non-zero, and 2 of these are obtained by transposition. Further,

$$
\begin{aligned}
& Var(X_i^{'} A^{+}\zeta_i) \\
= \; & \sigma_\eta^2 Z_i' A^{+}\mathbf{ii}' A^{+} Z_i + \sigma^2 Z_i' A^{+} Z_i + E(Z_i' A^{+} M_i\beta\beta' M_i' A^{+} Z_i) \\
& + Cov(\mathbf{cf}') + Cov(\mathbf{fc}') + \sigma_\eta^2 E(M_i' A^{+}\mathbf{ii}' A^{+} M_i) + \\
& \sigma^2 E(M_i' A^{+} M_i) + Var(\mathbf{f})
\end{aligned}
$$

We show below that

$$E(Z_i'A^+M_i\beta\beta'M_i'A^+Z_i) = \boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta}Z_i'A^+Z_i$$
$$E(M_i'A^+M_i) = tr(\mathbf{A}^+)\Sigma_M$$
$$E(M_i'A^+\mathbf{ii}'A^+M_i) = \mathbf{i}'A^+\mathbf{i}\Sigma_M$$

and that under assumptions of normality, that is if $\mathbf{W}$ is a matrix with i.i.d. rows $\backsim N(0, \Sigma_W)$

$$Var(\mathbf{W}'\mathbf{A}\mathbf{W}\boldsymbol{\beta}) = tr(\mathbf{A}^2)\left\{\boldsymbol{\Sigma}_W\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_W + (\boldsymbol{\beta}'\boldsymbol{\Sigma}_W\boldsymbol{\beta})\boldsymbol{\Sigma}_W\right\}.$$

Thus

$$\begin{aligned} Var(\mathbf{f}) &= Var(M_i'A^+M_i\beta) \\ &= tr(\mathbf{A}^{+2})\left\{\boldsymbol{\Sigma}_M\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M + (\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\boldsymbol{\Sigma}_M\right\} \end{aligned}$$

Under these assumptions,

$$Cov(\mathbf{cf}') = Cov(Z_i'A^+M_i\beta, M_i'A^+M_i\beta) = \mathbf{0}.$$

Thus

$$\begin{aligned} & Var(X_i'A^+\zeta_i) \\ = \; & \sigma_\eta^2 Z_i'A^+\mathbf{ii}'A^+Z_i + \sigma^2 Z_i'A^+Z_i + \\ & \boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta}Z_i'A^+Z_i + \sigma_\eta^2\mathbf{i}'A^+\mathbf{i}\Sigma_M + \\ & \sigma^2 tr(\mathbf{A}^+)\Sigma_M + tr(\mathbf{A}^{+2})\left\{\boldsymbol{\Sigma}_M\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M + (\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\boldsymbol{\Sigma}_M\right\}. \end{aligned}$$

Finally

$$\begin{aligned} & \sqrt{N}\left[\widehat{\beta}_{AG} - \beta + \left[\Sigma_{ZAZ} + tr(A^+)\Sigma_M\right]^{-1} tr(A^+)\Sigma_M\beta\right] \\ & \xrightarrow{D} N(0, \left[\Sigma_{ZAZ} + tr(A^+)\Sigma_M\right]^{-1} \times \\ & \left[\begin{array}{c} T\sigma_\eta^2\Sigma_{ZAMAZ} + (\sigma^2 + \boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\Sigma_{ZAZ} + \sigma_\eta^2\mathbf{i}'A^+\mathbf{i}\Sigma_M + \\ \sigma^2 tr(\mathbf{A}^+)\Sigma_M + tr(\mathbf{A}^{+2})\left\{\boldsymbol{\Sigma}_M\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M + (\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\boldsymbol{\Sigma}_M\right\} \end{array}\right] \times \\ & \left[\Sigma_{ZAZ} + tr(A^+)\Sigma_M\right]^{-1}). \end{aligned}$$

where

$$\lim_{N\to\infty}\frac{1}{N}Z_i'A^+\mathbf{ii}'A^+Z = T\lim_{N\to\infty}\frac{1}{N}Z_i'A^+\mathbf{M}A^+Z$$

and $A^+ = Q^+ = I_T - \frac{1}{T}ii'$ or $= M^+ = \frac{1}{T}ii' = I_T - Q^+$. So

$$\begin{aligned} & \sqrt{N}\left[\widehat{\beta}_{WG} - \beta + \left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1}\Sigma_M\beta\right] \\ & \xrightarrow{D} N(0, \left[\Sigma_{ZQZ} + (T-1)\Sigma_M\right]^{-1} \times \\ & \left[\begin{array}{c} (\sigma^2 + \boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\Sigma_{ZQZ} + \\ \sigma^2(T-1)\Sigma_M + (T-1)\left\{\boldsymbol{\Sigma}_M\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M + (\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\boldsymbol{\Sigma}_M\right\} \end{array}\right] \times \\ & \left[\Sigma_{ZQZ} + (T-1)\Sigma_M\right]^{-1}). \end{aligned}$$

and the variance matrix can be written

$$\begin{array}{l}[1/(T-1)\}\,[\Sigma_{ZQZ}/(T-1)+\Sigma_M]^{-1}\times\\ \left[(\sigma^2+\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\Sigma_{ZQZ}/(T-1)+\sigma^2\Sigma_M+\{\boldsymbol{\Sigma}_M\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M+(\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\boldsymbol{\Sigma}_M\}\right]\times\\ [\Sigma_{ZQZ}/(T-1)+\Sigma_M]^{-1}.\end{array}$$

Thus

$$\begin{array}{l}\sqrt{N}\left[\widehat{\beta}_{BG}-\beta+[\Sigma_{ZMZ}+\Sigma_M]^{-1}\Sigma_M\beta\right]\\ \xrightarrow{D} N(0,[\Sigma_{ZMZ}+\Sigma_M]^{-1}\times\\ \left[\begin{array}{c}T\sigma_\eta^2\Sigma_{ZMZ}+(\sigma^2+\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\Sigma_{ZMZ}+\sigma_\eta^2T\Sigma_M+\\ \sigma^2\Sigma_M+\{\boldsymbol{\Sigma}_M\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M+(\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\boldsymbol{\Sigma}_M\}\end{array}\right]\times\\ [\Sigma_{ZMZ}+\Sigma_M]^{-1}).\end{array}$$

To complete our analysis, and obtain the limiting variance of $\widehat{\beta}_{WG}-\widehat{\beta}_{BG}$, we need $Cov(\widehat{\beta}_{WG},\widehat{\beta}_{BG})$. This would be zero except for the measurement error. Accordingly, we require only some terms of

$$Cov(X_i'M^+\zeta_i,X_i'Q^+\zeta_i).$$

Remembering

$$\begin{array}{rl} & X_i'A^+\zeta_i\\ = & [Z_i'A^++M_i'A^+]\,[\eta_i\mathbf{i}+\boldsymbol{\nu}_i-M_i\beta]\\ = & Z_i'A^+\eta_i\mathbf{i}+Z_i'A^+\boldsymbol{\nu}_i-Z_i'A^+M_i\beta+M_i'A^+\eta_i\mathbf{i}+M_i'A^+\boldsymbol{\nu}_i-M_i'A^+M_i\beta\end{array}$$

and as $Q^+=I_T-\frac{1}{T}ii'$, $M^+=\frac{1}{T}ii'=I_T-Q^+$, $M^+\mathbf{i}=\mathbf{i}$, $Q^+\mathbf{i}=\mathbf{0}$, $M^+Q^+=\mathbf{0}$, consider

$$\begin{array}{rl} & X_i'M^+\zeta_i\\ = & Z_i'\eta_i\mathbf{i}+Z_i'M^+\boldsymbol{\nu}_i-Z_i'M^+M_i\beta+M_i'\eta_i\mathbf{i}+M_i'M^+\boldsymbol{\nu}_i-M_i'M^+M_i\beta\\ = & \mathbf{a}_M+\mathbf{b}_M+\mathbf{c}_M+\mathbf{d}_M+\mathbf{e}_M+\mathbf{f}_M,\text{ say}\end{array}$$

and

$$\begin{array}{rl} & X_i'Q^+\zeta_i\\ = & Z_i'Q^+\boldsymbol{\nu}_i-Z_i'Q^+M_i\beta+M_i'Q^+\boldsymbol{\nu}_i-M_i'Q^+M_i\beta\\ = & \mathbf{b}_Q+\mathbf{c}_Q+\mathbf{e}_Q+\mathbf{f}_Q,\text{ say.}\end{array}$$

Of the 24 possible covariances in $Cov(X_i'M^+\zeta_i,X_i'Q^+\zeta_i)$, $\mathbf{a}_M$ and $\mathbf{d}_M$ have zero covariance with $X_i'Q^+\zeta_i$ under our assumption that η_i is uncorrelated with $\boldsymbol{\nu}_i$ and M_i,

$$Cov(\mathbf{b}_M,\mathbf{b}_Q)=E(Z_i'M^+\boldsymbol{\nu}_i(Z_i'Q^+\boldsymbol{\nu}_i)')=\mathbf{0}$$

as $E(\boldsymbol{\nu}_i\boldsymbol{\nu}_i') = \sigma^2\mathbf{I}_T$ and $M^+Q^+ = \mathbf{0}$. Similarly $Cov(\mathbf{b}_M, \mathbf{e}_Q) = Cov(\mathbf{e}_M, \mathbf{b}_Q) = Cov(\mathbf{e}_M, \mathbf{e}_Q) = \mathbf{0}$. $Cov(\mathbf{b}_M, \mathbf{c}_Q) = Cov(\mathbf{b}_M, \mathbf{f}_Q) = Cov(\mathbf{e}_M, \mathbf{c}_Q) = Cov(\mathbf{e}_M, \mathbf{f}_Q)$ under our assumption that $\boldsymbol{\nu}_i$ is uncorrelated with M_i. Again, $Cov(\mathbf{c}_M, \mathbf{b}_Q) = Cov(\mathbf{c}_M, \mathbf{e}_Q) = Cov(\mathbf{f}_M, \mathbf{b}_Q) = Cov(\mathbf{f}_M, \mathbf{e}_Q) = \mathbf{0}$. This just leaves the 4 terms involving the measurement error,

$$\begin{aligned}
Cov(\mathbf{c}_M, \mathbf{c}_Q) &= E(Z_i'M^+M_i\beta(Z_i'Q^+M_i\beta)') \\
Cov(\mathbf{c}_M, \mathbf{f}_Q) &= E(Z_i'M^+M_i\beta(M_i'Q^+M_i\beta)') \\
Cov(\mathbf{f}_M, \mathbf{c}_Q) &= E(M_i'M^+M_i\beta(Z_i'Q^+M_i\beta)') \\
Cov(\mathbf{f}_M, \mathbf{f}_Q) &= Cov(M_i'M^+M_i\beta, M_i'Q^+M_i\beta)
\end{aligned}$$

Taking them in order, $E(M_i\beta\beta'M_i)$ has j,k-th element

$$E(m_{i(j)}'\beta m_{i(k)}\beta) = \beta' E(m_{i(k)}m_{i(j)}')\beta = \delta_{j,k}\beta'\Sigma_M\beta$$

where $m_{i(k)}'$ is the k-th row of M_i, and thus

$$\begin{aligned}
Cov(\mathbf{c}_M, \mathbf{c}_Q) &= Z_i'M^+E(M_i\beta\beta'M_i)Q^+Z_i)') \\
&= \beta'\Sigma_M\beta Z_i'M^+I_TQ^+Z_i)') = \mathbf{0}.
\end{aligned}$$

$$\begin{aligned}
Cov(\mathbf{c}_M, \mathbf{f}_Q) &= E(Z_i'M^+M_i\beta\beta'M_i'Q^+M_i) \\
&= Z_i'M^+E(M_i\beta\beta'M_i'Q^+M_i).
\end{aligned}$$

The expectation has j,k-th element

$$\begin{aligned}
& E(m_{i(j)}'\beta\beta'M_i'Q^+m_{i(k)}) \\
=\; & E(m_{i(j)}'\beta m_{i(k)}'Q^+M_i)\beta \\
=\; & E[(m_{i(j)}\otimes m_{i(k)}')'(\beta\otimes Q^+M_i\beta) \\
=\; & E\sum_{r=1}^{K} m_{i,j,r}\sum_{s=1}^{K} m_{i,k,s}\left[\beta_r q_{(k)}^{+\prime}M_i\beta\right] \\
=\; & E\sum_{r=1}^{K}\beta_r m_{i,j,r}\sum_{s=1}^{K} m_{i,k,s}\sum_{t=1}^{T} q_{kt}^{+}\sum_{u=1}^{K} m_{i,t,u}\beta_u \\
=\; & E\sum_{r=1}^{K}\beta_r\sum_{t=1}^{T} q_{kt}^{+}\sum_{u=1}^{K}\beta_u\sum_{s=1}^{K} m_{i,j,r}m_{i,k,s}m_{i,t,u}
\end{aligned}$$

using

$$\mathbf{a}'\mathbf{b}\mathbf{c}'\mathbf{d} = (\mathbf{a}\otimes\mathbf{c})'(\mathbf{b}\otimes\mathbf{d}).$$

We are assuming all third order moments are zero, and thus this expectation will be zero, as will $Cov(\mathbf{f}_M, \mathbf{c}_Q)$. This leaves

$$Cov(\mathbf{f}_M, \mathbf{f}_Q) = Cov(M_i'M^+M_i\beta, M_i'Q^+M_i\beta)$$

However, under assumptions of Normality, $M^{+}Q^{+}=\mathbf{0}$ ensures that $M_i'M^{+}M_i$ and $M_i'Q^{+}M_i$ are independently distributed, and hence the covariance will be zero.

We can now assemble the Hausman test statistic for the measurement error case. One would calculate

$$
\begin{aligned}
& h \\
= & (\widehat{\beta}_{WG}-\widehat{\beta}_{BG})' \left[\widehat{Var}(\widehat{\beta}_{WG})+\widehat{Var}(\widehat{\beta}_{BG})\right]^{-1}(\widehat{\beta}_{WG}-\widehat{\beta}_{BG}) \\
= & \sqrt{N}(\widehat{\beta}_{WG}-\widehat{\beta}_{BG})' \left[N\widehat{Var}(\widehat{\beta}_{WG})+N\widehat{Var}(\widehat{\beta}_{BG})\right]^{-1}\sqrt{N}(\widehat{\beta}_{WG}-\widehat{\beta}_{BG})
\end{aligned}
$$

$$
\begin{aligned}
& N\widehat{Var}(\widehat{\beta}_{WG}) \\
= & N\frac{1}{N(T-1)}\widehat{\varepsilon}'_{WG}\widehat{\varepsilon}_{WG}\left[\sum_{i=1}^{N}X_i'Q^{+}X_i\right]^{-1} \\
= & \frac{1}{N(T-1)}\widehat{\varepsilon}'_{WG}\widehat{\varepsilon}_{WG}\left[\frac{1}{N}\sum_{i=1}^{N}X_i'Q^{+}X_i\right]^{-1} \\
& \xrightarrow{p}\{\sigma^2+\beta'\Sigma_M\beta-\beta'\Sigma_M\left[\frac{1}{(T-1)}\Sigma_{ZQZ}+\Sigma_M\right]^{-1}\Sigma_M\beta\}\times \\
& [\Sigma_{ZQZ}+(T-1)\Sigma_M]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
& N\widehat{Var}(\widehat{\beta}_{BG}) \\
= & N\frac{1}{N}\widehat{\varepsilon}'_{BG}\widehat{\varepsilon}_{BG}\left[\sum_{i=1}^{N}X_i'M^{+}X_i\right]^{-1} \\
= & \frac{1}{N}\widehat{\varepsilon}'_{BG}\widehat{\varepsilon}_{BG}\left[\frac{1}{N}\sum_{i=1}^{N}X_i'M^{+}X_i\right]^{-1} \\
& \xrightarrow{p}\{T\sigma_\eta^2+\sigma^2+\beta'\Sigma_M\beta-\beta'\Sigma_M\left[\Sigma_{ZMZ}+\Sigma_M\right]^{-1}\Sigma_M\beta\}\times \\
& [\Sigma_{ZMZ}+\Sigma_M]^{-1}
\end{aligned}
$$

using $A^{+}=Q^{+}=I_T-\frac{1}{T}ii'$ or $=M^{+}=\frac{1}{T}ii'=I_T-Q^{+}$,

$$\frac{1}{N}\sum_{i=1}^{N}X_i'A^{+}X_i\xrightarrow{p}\Sigma_{ZAZ}+tr(A^{+})\Sigma_M$$

$$\frac{1}{N(T-1)}\widehat{\varepsilon}'_{WG}\widehat{\varepsilon}_{WG} \tag{51}$$

$$\xrightarrow{p}\{\sigma^2+\beta'\Sigma_M\beta-\beta'\Sigma_M\left[\frac{1}{(T-1)}\Sigma_{ZQZ}+\Sigma_M\right]^{-1}\Sigma_M\beta\} \tag{52}$$

$$\frac{1}{N}\widehat{\varepsilon}'_{BG}\widehat{\varepsilon}_{BG} \tag{53}$$
$$\xrightarrow{p} T\sigma_\eta^2 + \sigma^2 + \beta'\Sigma_M\beta - \beta'\Sigma_M\left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}\Sigma_M\beta. \tag{54}$$

However, under the assumption that

$$\left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1}\Sigma_M\beta = \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}\Sigma_M\beta$$

$$\begin{aligned}
&\sqrt{N}(\widehat{\beta}_{WG} - \widehat{\beta}_{BG}) \\
&\xrightarrow{D} N(0, [1/(T-1)\}\left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} \times \\
&\left[(\sigma^2 + \boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\Sigma_{ZQZ}/(T-1) + \sigma^2\Sigma_M + \{\boldsymbol{\Sigma}_M\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M + (\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\boldsymbol{\Sigma}_M\}\right] \times \\
&\left[\Sigma_{ZQZ}/(T-1) + \Sigma_M\right]^{-1} + \left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1} \times \\
&[\sigma_\eta^2\Sigma_{ZMZ} + (\sigma^2 + \boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\Sigma_{ZMZ} + \sigma_\eta^2 T\Sigma_M + \sigma^2\Sigma_M + \{\boldsymbol{\Sigma}_M\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M + (\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta})\boldsymbol{\Sigma}_M\}] \\
&\left[\Sigma_{ZMZ} + \Sigma_M\right]^{-1}).
\end{aligned}$$

12.1 A matrix result

If $\mathbf{U}$ is a random, $K \times K$ matrix, and $\boldsymbol{\beta}$ is a fixed $K \times 1$ vector, $Var(\mathbf{U}\boldsymbol{\beta})$ has i,i-th element

$$Var(\mathbf{u}'_{(i)}\boldsymbol{\beta}) = \boldsymbol{\beta}'Var(\mathbf{u}_{(i)})\boldsymbol{\beta}$$

if $\mathbf{u}'_{(i)}$ is the i-th row of $\mathbf{U}$. Similarly,$Var(\mathbf{U}\boldsymbol{\beta})$ has i,j-th element

$$Cov(\mathbf{u}'_{(i)}\boldsymbol{\beta}, \mathbf{u}'_{(\mathbf{j})}\boldsymbol{\beta}) = \boldsymbol{\beta}'Cov(\mathbf{u}_{(i)}, \mathbf{u}_{(j)})\boldsymbol{\beta}.$$

Considering $K = 2$,

$$\begin{aligned}
Var(\mathbf{U}\boldsymbol{\beta}) &= \begin{bmatrix} \boldsymbol{\beta}' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}' \end{bmatrix} \begin{bmatrix} Var(\mathbf{u}_{(1)}) & Cov(\mathbf{u}_{(1)}, \mathbf{u}_{(2)}) \\ Cov(\mathbf{u}_{(2)}, \mathbf{u}_{(1)}) & Var(\mathbf{u}_{(2)}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta} \end{bmatrix} \\
&= (\mathbf{I}_2 \otimes \boldsymbol{\beta}')Var(vec(\mathbf{U}'))(\mathbf{I}_2 \otimes \boldsymbol{\beta})
\end{aligned}$$

Thus we can see that in general

$$Var(\mathbf{U}\boldsymbol{\beta}) = (\mathbf{I}_K \otimes \boldsymbol{\beta}')Var(vec(\mathbf{U}'))(\mathbf{I}_K \otimes \boldsymbol{\beta})$$

So $Var((\mathbf{W}'\mathbf{A}\mathbf{W}\boldsymbol{\beta})$ can be written in terms of $Var(vec(\mathbf{W}'\mathbf{A}\mathbf{W}))$, as $\mathbf{A}$ is symmetric.

12.2 $Cov(\mathbf{Z}'\mathbf{A}\mathbf{W}\boldsymbol{\beta}, \mathbf{W}'\mathbf{A}\mathbf{W}\beta)$

This covariance matrix is given by

$$E(\mathbf{Z}'\mathbf{A}\mathbf{W}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{W}'\mathbf{A}\mathbf{W}) = \mathbf{Z}'\mathbf{A}E(\mathbf{W}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{W}'\mathbf{A}\mathbf{W}).$$

$\mathbf{W}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{W}'\mathbf{A}\mathbf{W}$ has i-th row $\mathbf{w}'_{(i)}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{W}'\mathbf{A}\mathbf{W}$ and thus i,j-th element

$$\begin{aligned}\mathbf{w}'_{(i)}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{W}'\mathbf{A}\mathbf{w}_j &= \left(\sum_{l=1}^{K} w_{il}\beta_l\right)\left(\sum_{m=1}^{K}\beta_m\mathbf{w}'_m\mathbf{A}\mathbf{w}_j\right)\\ &= \left(\sum_{l=1}^{K} w_{il}\beta_l\right)\left(\sum_{m=1}^{K}\beta_m\sum_{t=1}^{T}\sum_{s=1}^{T} w_{tm}a_{ts}w_{sj}\right)\end{aligned}$$

The product $w_{il}w_{tm}w_{sj}$ always has zero expectation, under the assumption that odd moments of order 3 and 4 are zero.

12.3 $E(Z'A^+M\beta\beta'M'A^+Z)$

$$E(Z'A^+M\beta\beta'M'A^+Z) = Z_i'A^+E(M\beta\beta'M')A^+Z$$

$M\beta\beta'M'$ has i,j-th element

$$\mathbf{m}'_{(i)}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{m}_{(j)} = \left[\sum_{r=1}^{K} m_{ir}\beta_r\right]\left[\sum_{s=1}^{K} m_{js}\beta_s\right].$$

$$E(\mathbf{m}'_{(i)}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{m}_{(j)}) = \delta_{ij}\sum_{r=1}^{K}\sum_{s=1}^{K}\sigma_{Mrs}\beta_r\beta_s = \delta_{ij}\boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta}.$$

Thus

$$E(Z'A^+M\beta\beta'M'A^+Z_i) = \boldsymbol{\beta}'\boldsymbol{\Sigma}_M\boldsymbol{\beta}Z'A^+Z$$

12.4 $E(M'A^+M)$

$M'A^+M$ has i,j-th element

$$\begin{aligned}\mathbf{m}_i'\mathbf{A}^+\mathbf{m}_j &= \sum_{t=1}^{T}\sum_{s=1}^{T} m_{ti}a_{ts}m_{sj},\\ E(\mathbf{m}_i'\mathbf{A}^+\mathbf{m}_j) &= \sum_{t=1}^{T}\sum_{s=1}^{T}\delta_{ts}a_{ts}\sigma_{Mij} = \sigma_{Mij}tr(\mathbf{A}^+)\\ E(M'A^+M) &= tr(\mathbf{A}^+)\Sigma_M\end{aligned}$$

12.5 $E(M_i'A^+\mathbf{i}\mathbf{i}'A^+M_i)$

$M'A^+\mathbf{i}\mathbf{i}'A^+M$ is of the form $M'\mathbf{a}\mathbf{a}'M$. Following the analysis for $E(M'A^+M)$, we obtain

$$\begin{aligned}E(M'\mathbf{a}\mathbf{a}'M) &= tr(\mathbf{a}\mathbf{a}')\Sigma_M = \mathbf{a}'\mathbf{a}\Sigma_M\\ E(M_i'A^+\mathbf{i}\mathbf{i}'A^+M_i) &= \mathbf{i}'A^+\mathbf{i}\Sigma_M\end{aligned}$$

12.6 $Var(vec(\mathbf{W}'\mathbf{AW}))$ and $Var(\mathbf{W}'\mathbf{AW}\boldsymbol{\beta})$ under Normality

Magnus and Neudecker (1988), p. 251, Theorem 12, provide, if $x \backsim N(0, \Omega)$, and A is $n \times n$, symmetric,

$$\begin{aligned} E(x'Ax) &= tr(A\Omega) \\ Var(x'Ax) &= 2tr(A\Omega A\Omega) + 4\mu' A\Omega A\mu \end{aligned}$$

We need to generalise this to a matrix normal W, of order $T \times K$, but assume that the rows of W are $NID(0, \Sigma)$. The typical covariance required is $Cov(w_i'Aw_j, w_r'Aw_s)$ where w_i is the i-th column of W. Consider

$$Q_{i,j} = \begin{bmatrix} w_i' & w_j' \end{bmatrix} \begin{bmatrix} 0 & A \\ A & 0 \end{bmatrix} \begin{bmatrix} w_i \\ w_j \end{bmatrix} = 2w_i'Aw_j.$$

As

$$\begin{aligned} \begin{bmatrix} w_i \\ w_j \end{bmatrix} &\backsim N(0, \begin{bmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{jj} \end{bmatrix} \otimes I_T) \\ Var(Q_{i,j}) &= 2tr\left[\left\{\left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes A\right)\left(\begin{bmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{jj} \end{bmatrix} \otimes I_T\right)\right\}^2\right] \\ &= 2tr\left[\left\{\left(\begin{bmatrix} \sigma_{ij} & \sigma_{jj} \\ \sigma_{ii} & \sigma_{ij} \end{bmatrix} \otimes A\right)\right\}^2\right] \\ Var(w_i'Aw_j) &= \frac{1}{2}tr\left[\begin{bmatrix} \sigma_{ij} & \sigma_{jj} \\ \sigma_{ii} & \sigma_{ij} \end{bmatrix}^2\right] tr(A^2). \end{aligned}$$

using $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$, and $tr(A \otimes B) = tr(A)tr(B)$. Consider next

$$Q_{i,j,r,s} = \begin{bmatrix} w_i' & w_j' & w_r' & w_s' \end{bmatrix} \begin{bmatrix} 0 & A & 0 & 0 \\ A & 0 & 0 & 0 \\ 0 & 0 & 0 & A \\ 0 & 0 & A & 0 \end{bmatrix} \begin{bmatrix} w_i \\ w_j \\ w_r \\ w_s \end{bmatrix} = 2w_i'Aw_j + 2w_r'Aw_s.$$

Now

$$Var(Q_{i,j,r,s}) = 4Var(w_i'Aw_j) + 4Var(w_r'Aw_s) + 8Cov(w_i'Aw_j, w_r'Aw_s).$$

Applying our previous result

$$\begin{aligned} Var(Q_{i,j,r,s}) &= 2tr\left[\left\{\left(\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \otimes A\right)\left(\begin{bmatrix} \sigma_{ii} & \sigma_{ij} & \sigma_{ir} & \sigma_{is} \\ \sigma_{ij} & \sigma_{jj} & \sigma_{jr} & \sigma_{js} \\ \sigma_{ir} & \sigma_{jr} & \sigma_{rr} & \sigma_{rs} \\ \sigma_{is} & \sigma_{js} & \sigma_{rs} & \sigma_{ss} \end{bmatrix} \otimes I_T\right)\right\}^2\right] \\ &= 2tr\left[\begin{bmatrix} \sigma_{ij} & \sigma_{jj} & \sigma_{jr} & \sigma_{js} \\ \sigma_{ii} & \sigma_{ij} & \sigma_{ir} & \sigma_{is} \\ \sigma_{is} & \sigma_{js} & \sigma_{rs} & \sigma_{ss} \\ \sigma_{ir} & \sigma_{jr} & \sigma_{rr} & \sigma_{rs} \end{bmatrix}^2\right] tr(A^2) \end{aligned}$$

Thus

$$Cov(w_i'Aw_j, w_r'Aw_s) = \frac{1}{4}tr(A^2)\left\{ tr\begin{bmatrix} \sigma_{ij} & \sigma_{jj} & \sigma_{jr} & \sigma_{js} \\ \sigma_{ii} & \sigma_{ij} & \sigma_{ir} & \sigma_{is} \\ \sigma_{is} & \sigma_{js} & \sigma_{rs} & \sigma_{ss} \\ \sigma_{ir} & \sigma_{jr} & \sigma_{rr} & \sigma_{rs} \end{bmatrix}^2 \right.$$
$$\left. -tr\left(\begin{bmatrix} \sigma_{ij} & \sigma_{jj} \\ \sigma_{ii} & \sigma_{ij} \end{bmatrix}^2\right) - tr\left(\begin{bmatrix} \sigma_{rs} & \sigma_{ss} \\ \sigma_{rr} & \sigma_{rs} \end{bmatrix}^2\right)\right\}.$$

In general,

$$\begin{aligned} & tr\left(\begin{bmatrix} B & C \\ C^* & D \end{bmatrix}^2\right) - tr(B^2) - tr(C^2) \\ = & \; tr\left(\begin{bmatrix} B & C \\ C^* & D \end{bmatrix}\begin{bmatrix} B & C \\ C^* & D \end{bmatrix}\right) - tr(B^2) - tr(C^2) \\ = & \; tr\left(\begin{bmatrix} B^2 + CC^* & ? \\ ? & D^2 + C^*C \end{bmatrix}\right) - tr(B^2) - tr(C^2) \\ = & \; tr(B^2 + CC^*) + tr(D^2 + C^*C) - tr(B^2) - tr(C^2) \\ = & \; tr(CC^*) + tr(C^*C) = 2tr(CC^*). \end{aligned}$$

In our case

$$\begin{aligned} CC^* &= \begin{bmatrix} \sigma_{jr} & \sigma_{js} \\ \sigma_{ir} & \sigma_{is} \end{bmatrix}\begin{bmatrix} \sigma_{is} & \sigma_{js} \\ \sigma_{ir} & \sigma_{jr} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{jr}\sigma_{is} + \sigma_{js}\sigma_{ir} & 2\sigma_{jr}\sigma_{js} \\ 2\sigma_{ir}\sigma_{is} & \sigma_{jr}\sigma_{is} + \sigma_{js}\sigma_{ir} \end{bmatrix} \\ Cov(w_i'Aw_j, w_r'Aw_s) &= tr(A^2)\left[\sigma_{jr}\sigma_{is} + \sigma_{js}\sigma_{ir}\right] \end{aligned}$$

This can be verified with some algebra, using $\mathbf{x} = [X_1, .., X_n]' \backsim N(0, \Omega) \Rightarrow$

$$E(X_iX_jX_kX_l) = \omega_{ij}\omega_{kl} + \omega_{ik}\omega_{kl} + \omega_{il}\omega_{jk}$$

(Anderson, 1958, p. 39). Our result also exhibits the necessary invariance to the ordering of i, j, r, s. As $W'AW$ has $w_i'Aw_j$ as i, j-th element, $vec(W'AW)$ has i varying more rapidly than j, so if

$$k = n(j-1) + i, l = n(s-1) + r$$

then $Var(vec(W'AW)$ has k, l-th element $Cov(w_i'Aw_j, w_r'Aw_s) = tr(A^2)\left[\sigma_{jr}\sigma_{is} + \sigma_{js}\sigma_{ir}\right]$
If $K = 2$, the pattern of subscripts is

$$\begin{bmatrix} i, j\backslash l, m & 1,1 & 2,1 & 1,2 & 2,2 \\ 1,1 & 1,1,1,1+1,1,1,1 & 1,2,1,1+1,1,1,2 & 1,1,1,2+1,2,1,1 & 1,2,1,2+1,2,1,2 \\ 2,1 & 1,1,2,1+1,1,2,1 & 1,2,2,1+1,1,2,2 & 1,1,2,2+1,2,2,1 & 1,2,2,2+1,2,2,2 \\ 1,2 & 2,1,1,1+2,1,1,1 & 2,2,1,1+2,1,1,2 & 2,1,1,2+2,2,1,1 & 2,2,1,2+2,2,1,2 \\ 2,2 & 2,1,2,1+2,1,2,1 & 2,2,2,1+2,1,2,2 & 2,1,2,2+2,2,2,1 & 2,2,2,2+2,2,2,2 \end{bmatrix},$$

and

$$
\begin{aligned}
& Var(vec(\mathbf{W}'\mathbf{A}\mathbf{W})) \\
= \;& tr(\mathbf{A}^2)\begin{bmatrix} 2\sigma_{1,1}^2 & 2\sigma_{1,1}\sigma_{1,2} & 2\sigma_{1,2}\sigma_{1,1} & 2\sigma_{1,2}^2 \\ 2\sigma_{1,1}\sigma_{2,1} & \sigma_{1,1}\sigma_{2,2}+\sigma_{12}^2 & \sigma_{1,1}\sigma_{2,2}+\sigma_{12}^2 & 2\sigma_{1,2}\sigma_{2,2} \\ 2\sigma_{2,1}\sigma_{1,1} & \sigma_{1,1}\sigma_{2,2}+\sigma_{12}^2 & \sigma_{1,1}\sigma_{2,2}+\sigma_{12}^2 & 2\sigma_{2,2}\sigma_{1,2} \\ 2\sigma_{1,2}^2 & 2\sigma_{2,1}\sigma_{2,2} & 2\sigma_{2,2}\sigma_{2,1} & 2\sigma_{2,2}^2 \end{bmatrix}. \qquad (55)
\end{aligned}
$$

We notice that the symmetry of $W'AW$ implies an implicit duplication in the *vec* operator, and ensures that in the last array, column 2 = column 3 and row 2 = row 3. Now stacking

$$Cov(w_i'Aw_j, w_r'Aw_s) = tr(A^2)\left[\sigma_{jr}\sigma_{is} + \sigma_{js}\sigma_{ir}\right]$$

vertically, first on i, we have

$$Cov(W'Aw_j, w_r'Aw_s) = tr(A^2)\left[\sigma_{jr}\boldsymbol{\sigma}_s + \sigma_{js}\boldsymbol{\sigma}_r\right]$$

then with respect to j,

$$Cov(vec(W'AW), w_r'Aw_s) = tr(A^2)\left[\boldsymbol{\sigma}_r \otimes \boldsymbol{\sigma}_s + \boldsymbol{\sigma}_s \otimes \boldsymbol{\sigma}_r\right].$$

where $\boldsymbol{\sigma}_r$ is the r-th column of Σ. Then we stack horizontally, first with respect to r,

$$Cov(vec(W'AW), W'Aw_s) = tr(A^2)\left[\Sigma \otimes \boldsymbol{\sigma}_s + \boldsymbol{\sigma}_s \otimes \Sigma\right]$$

then with respect to s,

$$
\begin{aligned}
& Cov(vec(W'AW, vec(W'AW) \\
= \;& tr(A^2)\{\begin{bmatrix} \Sigma \otimes \boldsymbol{\sigma}_1 & \cdots & \Sigma \otimes \boldsymbol{\sigma}_K \end{bmatrix} + \Sigma \otimes \Sigma\} \\
= \;& Var(vec(W'AW)).
\end{aligned}
$$

Finally, we need

$$
\begin{aligned}
& Var(W'AW\beta) \\
= \;& (I_K \otimes \beta')tr(A^2)\{\begin{bmatrix} \Sigma \otimes \boldsymbol{\sigma}_1 & \cdots & \boldsymbol{\Sigma} \otimes \boldsymbol{\sigma}_K \end{bmatrix} + \Sigma \otimes \Sigma\}(I_K \otimes \beta) \\
= \;& tr(A^2)\left\{\begin{bmatrix} \Sigma \otimes \beta'\boldsymbol{\sigma}_1 & \cdots & \Sigma \otimes \beta'\boldsymbol{\sigma}_K \end{bmatrix} + \Sigma \otimes \beta'\Sigma\right\}(I_K \otimes \beta) \\
= \;& tr(A^2)\left\{\begin{bmatrix} \Sigma \otimes \beta'\boldsymbol{\sigma}_1 & \cdots & \Sigma \otimes \beta'\boldsymbol{\sigma}_K \end{bmatrix}(I_K \otimes \beta) + \Sigma \otimes (\beta'\Sigma\beta)\right\} \\
= \;& tr(A^2)\left\{\begin{bmatrix} \beta'\boldsymbol{\sigma}_1 \otimes \Sigma & \cdots & \boldsymbol{\beta'\sigma}_K \otimes \Sigma \end{bmatrix}(I_K \otimes \beta) + \Sigma \otimes (\beta'\Sigma\beta)\right\} \\
= \;& tr(A^2)\left\{([\beta'\Sigma] \otimes \Sigma)(I_K \otimes \beta) + \Sigma \otimes (\beta'\Sigma\beta)\right\} \\
= \;& tr(A^2)\left\{(\beta'\Sigma) \otimes \Sigma\beta) + (\beta'\Sigma\beta)\Sigma\right\} = tr(A^2)\left\{\Sigma\beta\beta'\Sigma + (\beta'\Sigma\beta)\Sigma\right\}.
\end{aligned}
$$

Checking this for $K = 2$,

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix},$$

and after some algebra

$$
\begin{aligned}
& Var(W'AW\beta) \\
= \;& tr(\mathbf{A}^2)\begin{bmatrix} 2\beta_1^2\sigma_{11}^2 + 4\beta_1\beta_2\sigma_{11}\sigma_{12} + \beta_2^2(\sigma_{11}\sigma_{22} + \sigma_{12}^2) & 2\beta_1^2\sigma_{12}\sigma_{11} + \beta_2\beta_1(3\sigma_{12}^2 + \sigma_{11}\sigma_{22}) + 2\beta_2^2\sigma_{12}\sigma_{22} \\ 2\beta_1^2\sigma_{12}\sigma_{11} + \beta_2\beta_1(3\sigma_{12}^2 + \sigma_{11}\sigma_{22}) + 2\beta_2^2\sigma_{12}\sigma_{22} & 2\beta_2^2\sigma_{22}^2 + 4\beta_1\beta_2\sigma_{22}\sigma_{12} + \beta_1^2(\sigma_{22}\sigma_{11} + \sigma_{12}^2) \end{bmatrix}
\end{aligned}
$$

which again can be obtained directly from (55).

13 Appendix 6

This appendix contains the graphs of the power curve of the standard Hausman test (*H-test*) versus the one displayed by the robust formulation presented in Section 3 (*HR-test*) with different sample sizes.

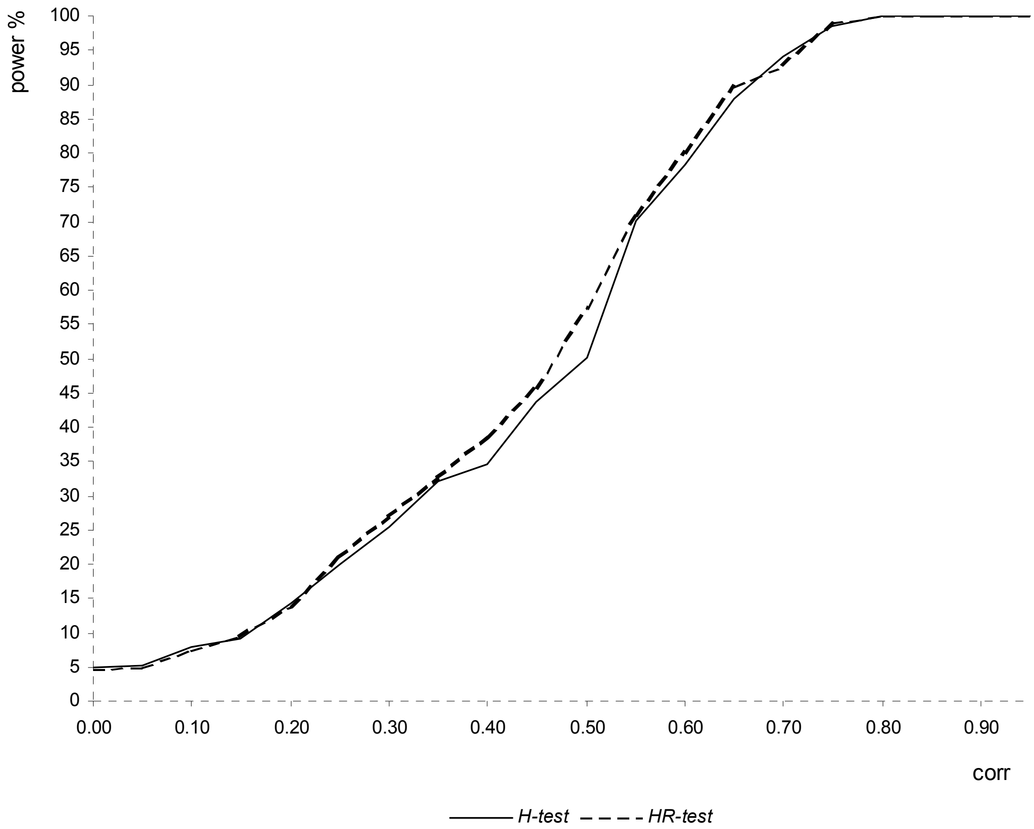

Figure 6: Power function comparison when N=25, T=4

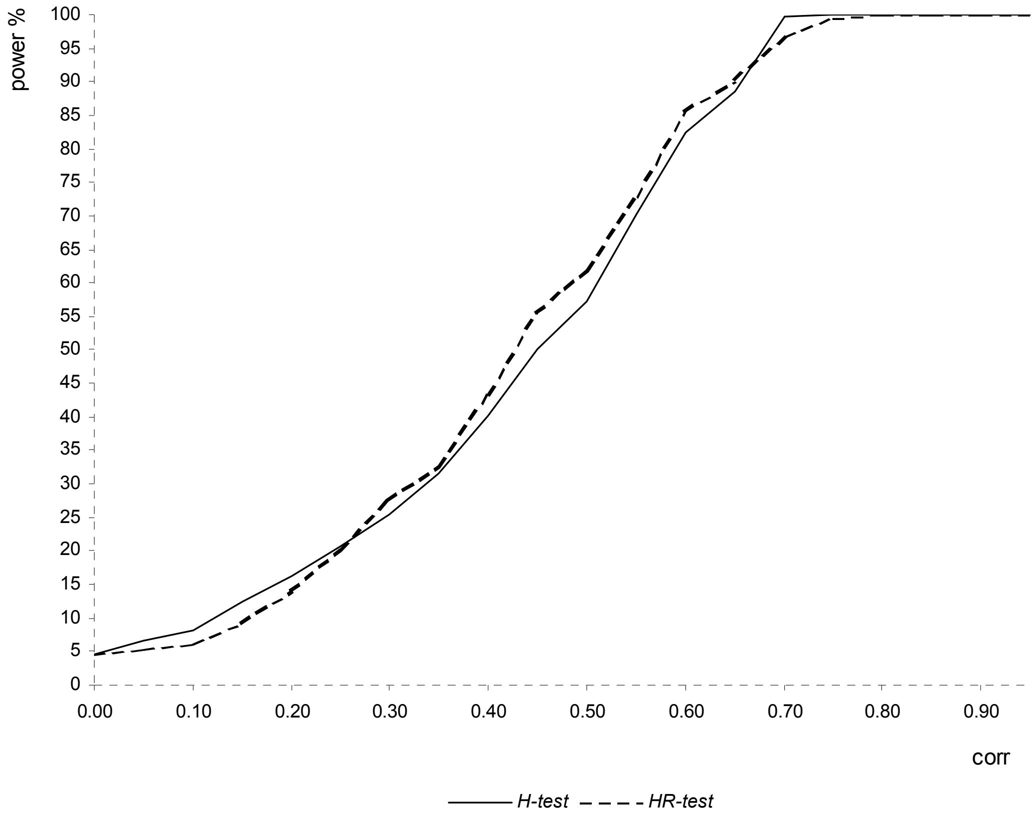

Figure 7: Power function comparison when N=25, T=10

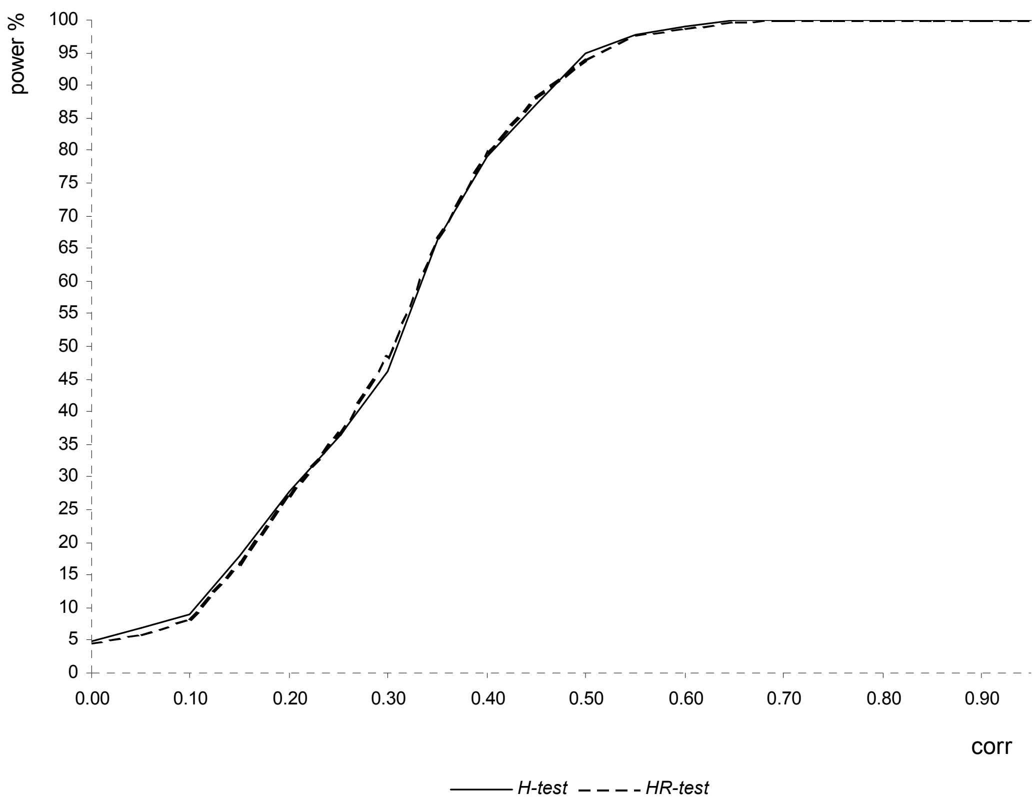

Figure 8: Power function comparison when N=25, T=20

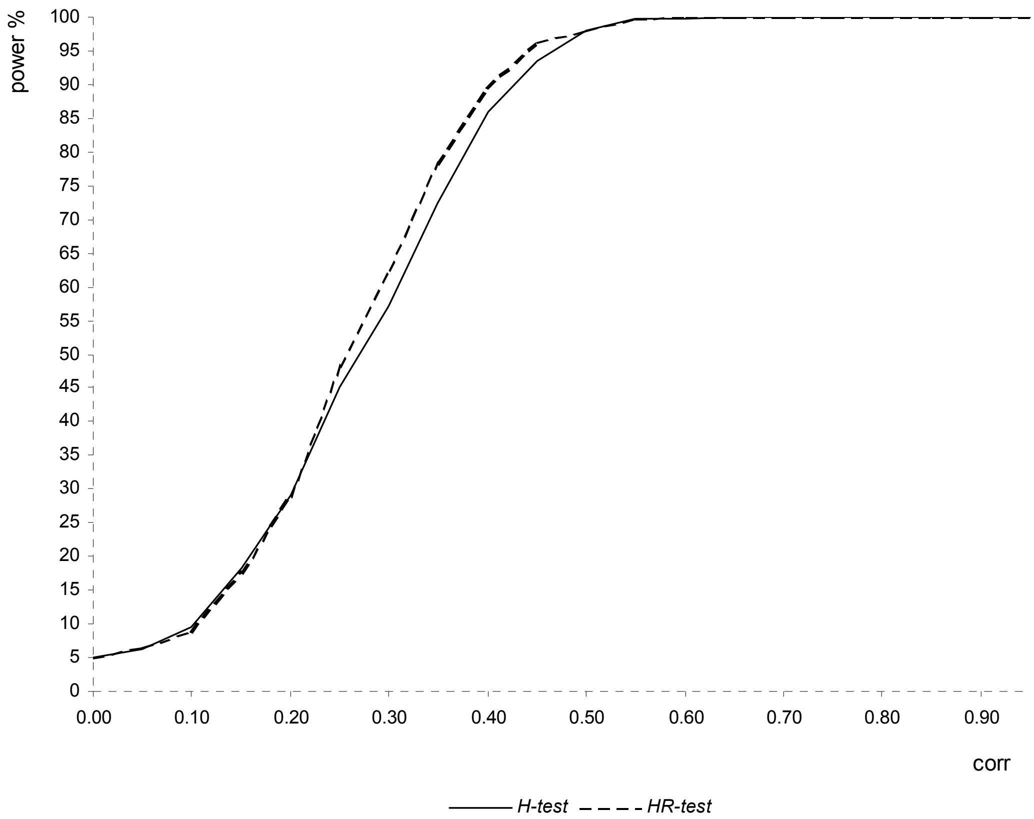

Figure 9: Power function comparison when N=275, T=4

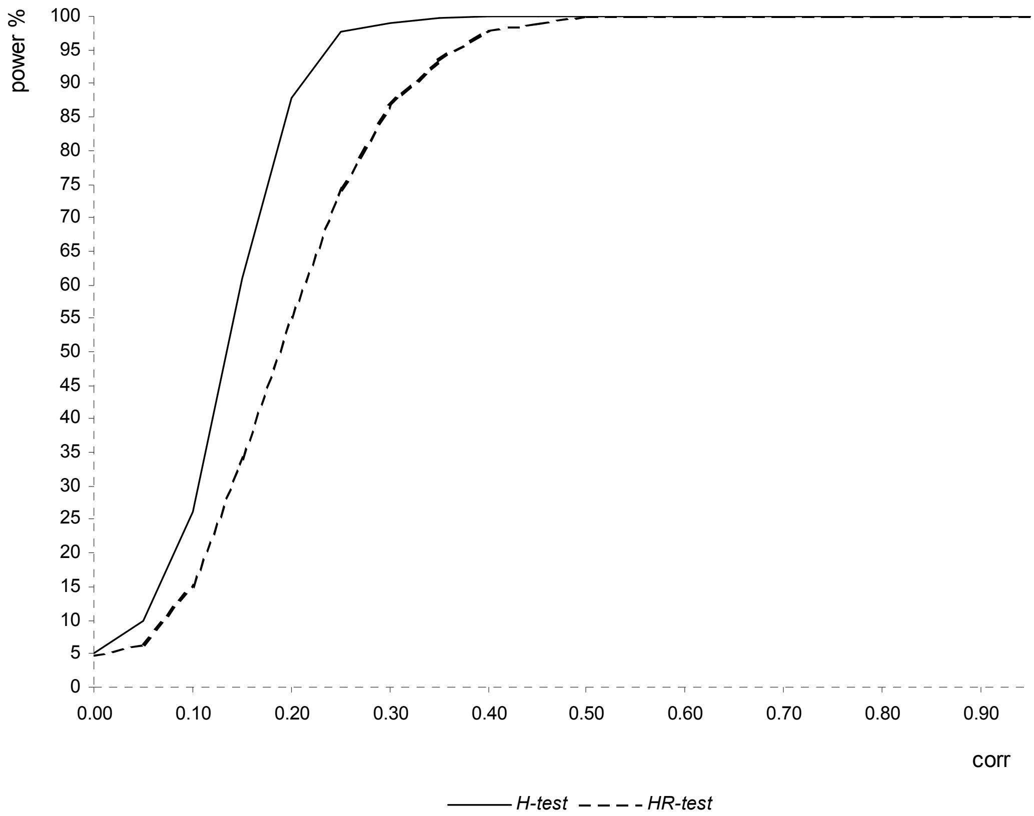

Figure 10: Power function comparison when N=275, T=10

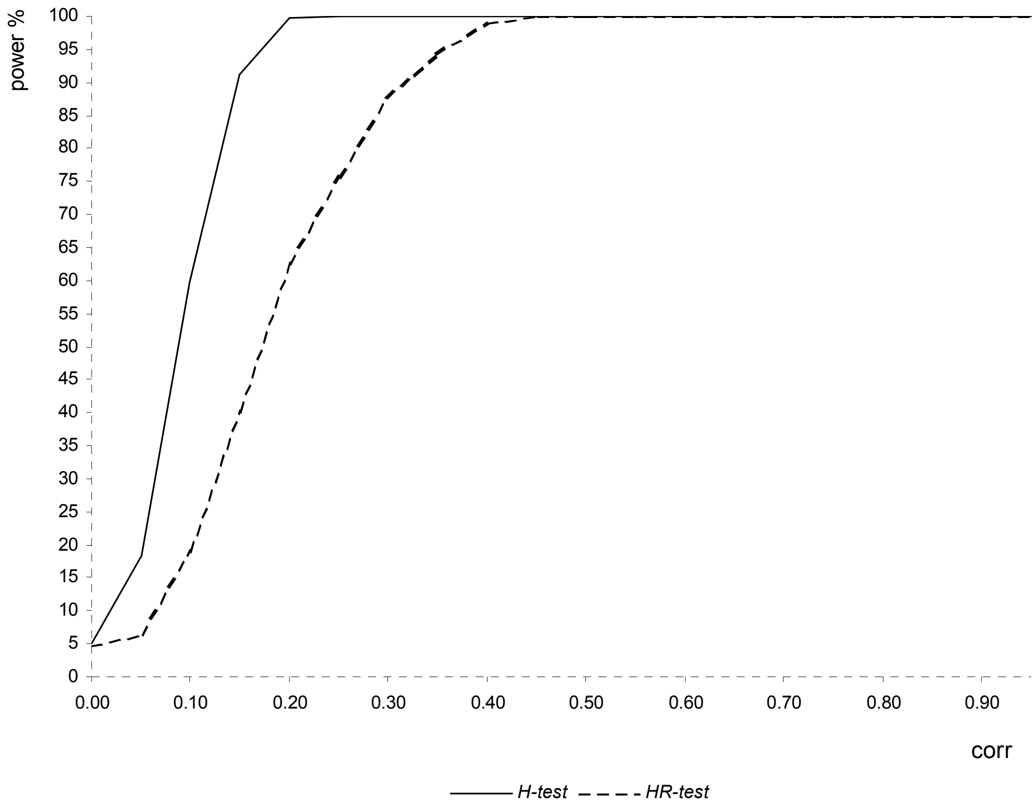

Figure 11: Power function comparison when N=275, T=20