On the sensitivity of the restricted least squares estimators to covariance misspecification

ALAN T.K. WAN†, GUOHUA ZOU‡ and HUAIZHEN QIN§

†Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong
E-mail: msawan@cityu.edu.hk

‡Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China, and Department of Biostatistics and Computational Biology, University of Rochester, NY 14642, USA
E-mail: Guohua_Zou@URMC.Rochester.edu

§Department of Mathematical Sciences, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931–1295, USA
E-mail: hqin@mtu.edu

First version received: July 2005; final version accepted: April 2007

Summary Traditional econometrics has long stressed the serious consequences of non-spherical disturbances for the estimation and testing procedures under the spherical disturbance setting, that is, the procedures become invalid and can give rise to misleading results. In practice, it is not unusual, however, to find that the parameter estimates do not change much after fitting the more general structure. This suggests that the usual procedures may well be robust to covariance misspecification. Banerjee and Magnus (1999) proposed sensitivity statistics to decide if the Ordinary Least Squares estimators of the coefficients and the disturbance variance are sensitive to deviations from the spherical error assumption. This paper extends their work by investigating the sensitivity of the restricted least squares estimator to covariance misspecification where the restrictions may or may not be correct. Large sample results giving analytical evidence to some of the numerical findings reported in Banerjee and Magnus (1999) are also obtained.

Key words: Autocorrelation, Non-spherical disturbances, Restrictions, Sensitivity.

1. INTRODUCTION

Historically, most of the strong results concerning estimators’ properties in econometrics have relied on the premise that the model’s disturbances are spherical, that is, they have uniform variance and are not correlated with one another. It is well known that Ordinary Least Squares (OLS) is inappropriate when the assumption of spherical disturbances is not met. There can hardly be a textbook in econometrics which does not emphasize that OLS is no longer best linear unbiased in the face of a non-scalar disturbance covariance matrix and misleading inference will result. Yet it is not uncommon to find that in practice, the coefficients’ estimates usually do not
change much after fitting the more general structure. The important question then is whether it is really necessary to discard OLS and other common procedures in the presence of non-spherical disturbances. With regard to this matter, Banerjee and Magnus (1999) gave a new perspective to the problem. Instead of concerning whether the white noise assumption of the disturbances has been violated, they examined whether it matters at all that the disturbances are not white noise. The authors derived sensitivity statistics to decide if the OLS estimators of the coefficients and variance are sensitive to deviations from the white noise assumption of the disturbances. When the disturbances are AR(1) or MA(1), their results suggested that the OLS estimator of the coefficients and the predictor are generally insensitive to covariance misspecification, but the OLS estimator of the variance can be very sensitive. Further, they showed that the sensitivity statistic for the variance estimator (which is essentially the Durbin–Watson statistic) is nearly orthogonal to the sensitivity statistic for the coefficient estimator. So it may well happen that even in the presence of significant autocorrelations, the OLS estimates of the coefficients are little affected. In a subsequent article, Banerjee and Magnus (2000) showed that the conventional $F$ and $t$ statistics are sensitive to covariance misspecification, but nonetheless accepting the null hypothesis using the $F$ test is a robust procedure. Again, the authors derived a sensitivity statistic and a median based decision rule for decisions on sensitivity.

Banerjee and Magnus’ contributions have opened up a new frontier for research on sensitivity analysis, underscoring one of the main messages in traditional econometrics that OLS and conventional testing procedures are invalid and should be avoided when confronted with a non-spherical covariance structure of the disturbances. The current paper extends Banerjee and Magnus’ (1999) results in two directions. First, we generalize Banerjee and Magnus’ (1999) sensitivity analysis to the case where the regression parameters are subject to linear restrictions. We introduce sensitivity statistics for the coefficient and variance estimators and discuss their properties for both the cases of correct and incorrect restrictions. When the restrictions are correctly imposed, our results can be obtained directly from the results of Banerjee and Magnus (1999) by reparametrization. But when the restrictions are incorrect, the derivation is a lot more complicated and cannot be deduced from existing results. Second, we provide an analytical justification for the simulation-based conclusion in Banerjee and Magnus (1999) regarding the correlation of the sensitivity statistics. Specifically, we prove the large sample near incorrelativity of these statistics in a more general context. In Section 2, we explain the model set-up, introduce the sensitivity statistics and present theorems relating to the theoretical properties of these statistics. Section 3 reports numerical results on the behaviour of the sensitivity statistics focusing on AR(1) and MA(1) disturbances for the cases of correct or incorrect restrictions. Section 4 concludes. Proofs of theorems are contained in an appendix.

2. MODEL FRAMEWORK AND SENSITIVITY STATISTICS

Consider the linear regression model,

\[ y = X\beta + u, \quad u \sim N(0, \sigma^2 \Omega(\theta)), \quad (1) \]

where $y$ and $u$ are $n \times 1$; $X$ is $n \times k$, non-stochastic and of rank $k(< n)$; $\beta$ is $k \times 1$; $\sigma^2 (>0)$ is a nuisance parameter; and $\Omega(\theta)$ is an $n \times n$ matrix function of a vector of parameters $\theta = (\theta_1, \ldots, \theta_p)'$, positive definite and differentiable at least in the neighbourhood of $\theta = 0$. Without loss of generality, we assume $\Omega(0) = I_n$. 

© Royal Economic Society 2007
If $\theta = 0$, the OLS estimator $\hat{\beta}(0) = (X'X)^{-1}X'y$ is known to possess optimal properties. On the other hand, if $\theta \neq 0$ but the elements of $\theta$ and the form of $\Omega$ are known, then the Generalized Least Squares (GLS) estimator $\hat{\beta}(\theta) = S^{-1}(\theta)X'\Omega^{-1}(\theta)y$, where $S(\theta) = X'\Omega^{-1}(\theta)X$, is a more efficient estimator than $\hat{\beta}(0)$, but if $\theta$ is unknown, then estimating $\beta$ by Estimated GLS (EGLS) is not necessarily a better alternative than OLS. The corresponding GLS estimator of $\sigma^2$ is given by $\hat{\sigma}^2(\theta) = (y - \hat{\gamma}(\theta))'\Omega^{-1}(\theta)(y - \hat{\gamma}(\theta))/(n - k)$, where $\hat{\gamma}(\theta) = X\hat{\beta}(\theta)$. Now, let the general linear restrictions on the regression coefficients be specified as $R\beta = r$, where $R$ is an $m \times k$ known matrix of rank $m$ and $r$ is an $m \times 1$ known vector. This yields

$$
\hat{\beta}(\theta) = \hat{\beta}(\theta) + S^{-1}(\theta)R'(RS^{-1}(\theta)R')^{-1}(r - R\hat{\beta}(\theta)),
$$

as the restricted GLS (RGLS) estimator of $\beta$. The estimator of $\sigma^2$ corresponding to $\hat{\beta}(\theta)$ is

$$
\hat{\sigma}^2(\theta) = (y - \bar{\gamma}(\theta))'\Omega^{-1}(\theta)(y - \bar{\gamma}(\theta))/(n - k + m),
$$

where $\bar{\gamma}(\theta) = X\bar{\beta}(\theta) = \hat{\gamma}(\theta) + XS^{-1}(\theta)R'(RS^{-1}(\theta)R')^{-1}(r - R\hat{\beta}(\theta))$.

The sensitivity of $\hat{\gamma}(\theta) = X\hat{\beta}(\theta)$ and $\hat{\sigma}^2(\theta)$ has been explored in Banerjee and Magnus (1999). The question to address is how sensitive these estimators are to the underlying assumption concerning the error process. Now, following Banerjee and Magnus’ (1999) work, we define, for $s = 1, \ldots, p$,

$$
\zeta_s = \frac{\partial \bar{\gamma}(\theta)}{\partial \theta_s} \bigg|_{\theta=0},
$$

and

$$
\bar{\zeta}_s = \frac{\partial \hat{\sigma}^2(\theta)}{\partial \theta_s} \bigg|_{\theta=0},
$$

as the sensitivity of $\bar{\gamma}(\theta)$ (or equivalently, $\bar{\beta}(\theta)$) and $\hat{\sigma}^2(\theta)$ with respect to small changes in $\theta_s$. The sensitivity statistic of $\bar{\gamma}(\theta)$ is given by

$$
\bar{B}_s = \frac{\zeta_s (\bar{C}_s \bar{C}_s')^+ \bar{z}_s}{(n - k + m)\hat{\sigma}^2(0)},
$$

where $\bar{C}_s = (I_n - \bar{M})A_s\bar{M}, (\bar{C}_s \bar{C}_s')^+$ is the generalized (Moore–Penrose) inverse of $(\bar{C}_s \bar{C}_s')$, $A_s = \partial \Omega(\theta)/\partial \theta_s |_{\theta=0}, \bar{M} = M + XS^{-1}R'(RS^{-1}R')^{-1}RS^{-1}X'$, and $M = I_n - XS^{-1}X'$. $S = X'X$ are both symmetric idempotent and of rank $n - k + m$ and $n - k$, respectively. It is straightforward to show that $\bar{M}M = M\bar{M} = M$, $\bar{C}_s \bar{M} = \bar{C}_s$ and $0 \leq \bar{r}_s = \text{rank}(\bar{C}_s) \leq \min\{k - m, n - k + m\}$.

Denote $y = y - XS^{-1}R'(RS^{-1}R')^{-1}r$. Then it can be shown, using theorem 3 of Magnus and Neudecker (1999, ch. 8), that

$$
\zeta_s = -\bar{C}_s y
$$

and

$$
\bar{B}_s = \frac{\bar{y} \bar{W}_s y}{\bar{y} \bar{M} y},
$$

© Royal Economic Society 2007
where \( W_s = C_s' (C_s C_s')^+ C_s \) is a symmetric idempotent matrix such that \( W_s M = MW_s = W_s \). Correspondingly, the sensitivity statistic of \( \sigma^2(\theta) \) is given by

\[
\bar{D}_s = \frac{\bar{\lambda}_s}{\sigma^2(0)} = \left. \frac{\partial \log(\sigma^2(\theta))}{\partial \theta_s} \right|_{\theta=0}.
\] (9)

For purposes of convenience, we can write

\[
\bar{\lambda}_s = -\bar{y} M_A y / n - k + m
\] (10)

and

\[
\bar{D}_s = -\frac{\bar{y} M_A y}{\bar{y} M y}.
\] (11)

**Theorem 1.**

(i) The sensitivity statistics \( \bar{B}_s \) and \( \bar{D}_s \) given in (8) and (11) have the equivalent expressions

\[
\bar{B}_s = \bar{v} Q W_s Q v / \bar{v} v \quad \text{and} \quad \bar{D}_s = -\bar{v} Q A_s Q v / \bar{v} v,
\]

where \( Q \) is some \( n \times (n - k + m) \) matrix such that \( Q Q = I_{n-k+m} \) and \( Q Q = \bar{M} \), and \( v = \sigma^{-1} Q (u + XS^{-1} R'(RS^{-1} R')^{-1} (R\beta - r)) \).

(ii) Evaluating the distribution of \( y \) at \( \theta = 0 \) reduces \( \bar{v} \) to a normal vector with mean \( Q XS^{-1} R'(RS^{-1} R')^{-1} (R\beta - r) / \sigma \) and covariance matrix \( I_{n-k+m} \). Moreover, if \( 0 < \bar{r}_s < n - k + m \), then

\[
\bar{B}_s \sim B''(\bar{r}_s/2, (n - k + m - \bar{r}_s)/2), \quad \delta_s^{(1)}(0), \quad \delta_s^{(2)}(0),
\] (12)

where \( B''(.) \) denotes the doubly non-central Beta distribution and 
\( \delta_s^{(1)}(0) = (R\beta - r)' (RS^{-1} R')^{-1} RS^{-1} X' W_s X S^{-1} R'(RS^{-1} R')^{-1} (R\beta - r) / \sigma^2 \) and 
\( \delta_s^{(2)}(0) = (R\beta - r)' (RS^{-1} R')^{-1} RS^{-1} X' (\bar{M} - W_s) X S^{-1} R'(RS^{-1} R')^{-1} (R\beta - r) / \sigma^2 \) are the associated non-centrality parameters.

(iii) If, in addition to ii), the linear restrictions \( R\beta = r \) are satisfied, then \( \bar{B}_s \sim B(\bar{r}_s/2, (n - k + m - \bar{r}_s)/2) \).

**Proof.** See the Appendix. \( \square \)

Clearly, if \( m = 0 \), then \( \bar{M} = M, C_s = C_s = (I_n - M) A_s M \) and \( \bar{W}_s = W_s = C_s' (C_s C_s')^+ C_s \). We then have \( \bar{B}_s = y' W_s y/y' M y = B_s \) and \( \bar{D}_s = -y' M_A y y' M y = D_s \), which are the statistics given in Banerjee and Magnus (1999) for measuring the sensitivity of the GLS estimators \( \hat{\beta}(\theta) \) and \( \sigma^2(\theta) \). Part (iii) of Theorem 1 then reduces to the corresponding theorem given in Banerjee and Magnus (1999). For an MA(1) or an AR(1) process, that is, \( A_s = T^{(1)} \), where \( T^{(1)} = (t^{(1)}_{ij}) \) is a symmetric Toeplitz matrix such that \( t^{(1)}_{ij} = 1 \) when \( |i - j| = 1 \) and equals 0 otherwise.

\(^1\)Some readers have commented that the analytical results on the sensitivity of the restricted estimators can be obtained from a direct reparametrization of (Banerjee and Magnus’ 1999) results. This is NOT true in general. Our results can be derived from those of Banerjee and Magnus (1999) only when the restrictions are correct. When the restrictions are incorrect, Banerjee and Magnus’ (1999) results cannot be applied directly.

© Royal Economic Society 2007
Banerjee and Magnus (1999) denoted the statistics $B_s$ and $D_s$ as $B1$ and $D1$, respectively. They showed via a simulation experiment that $B1$ and $D1$ are nearly independent. This is an important result, as $D1$ is essentially the Durbin–Watson statistic in an OLS regression, so even though the DW statistic shows significant autocorrelation, the estimates of $\beta$ may still be little affected. The following theorem provides corroborative analytical evidence to Banerjee and Magnus’ (1999) numerical findings by showing the near incorrelativity of $B_s$ and $D_s$ in large samples. The near incorrelativity of $B1$ and $D1$ is included as a special case.

**Theorem 2.**

(i) If the distribution of $y$ is evaluated at $\theta = 0$, then the covariance of $B_s$ and $D_s$ is given by

$$\text{cov}(B_s, D_s) = -\frac{1}{4} \exp\left(\frac{-\delta}{2}\right) \int_0^1 (1 - t)^{(n - k + m)/2 - 1} \{2\text{tr}(A_s W_s) + 2\bar{s}\text{tr}(A_s \bar{M})

+ \{\bar{s} + \mu' A_s \mu + \mu' W_s \mu\text{tr}(A_s \bar{M}) + 4\mu' W_s A_s \mu\text{tr}(A_s W_s)\} \exp\left(\frac{\delta}{2} t\right) dt

+ \frac{1}{4} \exp\left(\frac{-\delta}{2}\right) \left(\int_0^1 t^{(n - k + m)/2 - 1} [\bar{s} + \mu' W_s \mu] \exp\left(\frac{\delta}{2} t\right) dt\right)

\times \left(\int_0^1 t^{(n - k + m)/2 - 1} [\text{tr}(A_s \bar{M}) + \mu' A_s \mu] \exp\left(\frac{\delta}{2} t\right) dt\right),

(13)$$

where $\delta = (R\beta - r)'(RS^{-1}R)^{-1}(R\beta - r)/\sigma^2$, and $\mu = XS^{-1}R'(RS^{-1}R)^{-1}(R\beta - r)/\sigma$.

(ii) In the special case where $R \beta = r$ (i.e. $\delta = 0$), equation (13) reduces to

$$\text{cov}(B_s, D_s | \delta = 0) = -\frac{2\text{tr}(A_s W_s)}{(n - k + m)(n - k + m + 2)} + \frac{2\bar{s}\text{tr}(A_s \bar{M})}{(n - k + m)(n - k + m + 2)}. \tag{14}$$

(iii) If, in addition to the condition of (i), the eigenvalues of $A_s$ are bounded, and $\bar{A}_s = (n - k + m)\text{tr}(A_s \bar{M})^2 - (\text{tr}(A_s \bar{M}))^2 > c_{1s}^2$ for some positive constant $c_{1s}$, then $\rho(B_s, D_s)$, the correlation coefficient between $B_s$ and $D_s$ is of order $O(n^{-1/2})$, provided that $\delta$ is bounded.

**Proof.** See the Appendix.

The condition that the eigenvalues of $A_s$ are bounded is a very weak condition which will be satisfied in most cases in practice. This condition determines the divergence rate of $\bar{A}_s$, which in turn determines the convergence rate of $\rho(B_s, D_s)$ (see the proof of Theorem 2 in the Appendix for details). By Theorem 2, $\rho(B_s, D_s)$ tends to zero as $n$ tends to infinity, and the rate of convergence is of order $O(n^{-1/2})$. This implies that, provided that $\delta$ is bounded, $B_s$ and $D_s$ are approximately uncorrelated in large samples. One therefore must not draw any conclusion on the sensitivity of the $\beta$ estimates from the sensitivity of the $\sigma^2$ estimate, and vice versa. Since $B_s = B1_s$ and $D_s = D1_s$ when $m = 0$, the near incorrelativity of $B_s$ and $D_s$ is included in Theorem 2 as a special case. In particular, when $B_s = B1$ and $D_s =$

---

2Even if $\delta$ is unbounded, this conclusion still holds provided that $\delta = O(n^a)$, where $0 < a < 1$, but it does not hold when $\delta = O(n)$. In other words, if $\delta$ is of order $O(n)$, it is possible for $B_s$ and $D_s$ to be correlated in large samples.
D1 (i.e., $A_s = T^{(1)}$), the conditions in part (iii) of Theorem 2 are satisfied, and so $B1$ and $D1$ are nearly uncorrelated in large samples. The latter result is consistent with the numerical findings reported in Banerjee and Magnus (1999). A referee has suggested that one might be able to deduce the large sample independence of $\overline{B}_s$ and $\overline{D}_s$ from the near incorrelativity result by exploiting the fact that the disturbances are normal. But note that $\overline{B}_s$ and $\overline{D}_s$ do not asymptotically follow a joint normal distribution even when normality of disturbances is assumed. To see this, consider the special case of $\delta = 0$ and $F_s = 1$. Now, from Part (iii) of Theorem 1, $\overline{B}_s \sim \text{Beta}(1/2, (n - k + m - 1)/2)$, and $\overline{B}_s$ has the same distribution as $\xi_i^2 / \sum_{i=1}^{n-k+m} \xi_i^2$, where $\xi_i \sim \text{i.i.d. } N(0, 1)$. Thus, $n\overline{B}_s = \xi_1^2 / ((n - k + m) \sum_{i=1}^{n-k+m} \xi_i^2/(n(n - k + m))) \rightarrow_d \xi_1^2$, which shows that $\overline{B}_s$ and $\overline{D}_s$ do not have an approximate joint normal distribution. Therefore, the asymptotic uncorrelatedness of $\overline{B}_s$ and $\overline{D}_s$ does not directly imply their asymptotic independence.

Next, we consider the limiting behaviour of the sensitivity statistics. The following theorem is related to the case of AR(1) disturbances:

**Theorem 3.** Suppose that the disturbances in model (1) are generated from a stationary AR(1) process, that is, $u_t = \phi_1 u_{t-1} + \varepsilon_t$ and $|\phi_1| < 1$. The following results are observed:

(i) If $\overline{M}i \neq 0$, then

$$
\lim_{\phi_1 \to 1} \Pr(\overline{B}_1 > c_{\overline{B}_1}) = \begin{cases} 0 & \text{if } c_{\overline{B}_1} > \overline{b}_1 \\ 1 & \text{if } c_{\overline{B}_1} < \overline{b}_1 \end{cases}
$$

and

$$
\lim_{\phi_1 \to 1} \Pr(\overline{D}_1 \leq c_{\overline{D}_1}) = \begin{cases} 0 & \text{if } c_{\overline{D}_1} < \overline{d}_1 \\ 1 & \text{if } c_{\overline{D}_1} > \overline{d}_1 \end{cases}
$$

where $\overline{b}_1 = i^T\overline{W}^{(1)} i / i^T\overline{M}i$, $i$ is an $n \times 1$ vector of ones, $\overline{d}_1 = -i^T\overline{M}T^{(1)}\overline{M}i / i^T\overline{M}i$, $\overline{W}^{(1)} = \overline{C}^{(1)T}(C^{(1)}C^{(1)T})^{-1}C^{(1)}$ and $\overline{C}^{(1)} = (I_n - \overline{M})T^{(1)}\overline{M}$.

(ii) If $\overline{M}i = 0$, then

$$
\lim_{\phi_1 \to 1} \Pr(\overline{B}_1 > c_{\overline{B}_1}) = \Pr \left( \overline{B}^{(1)}(\eta) > c_{\overline{B}_1} \right),
$$

and

$$
\lim_{\phi_1 \to 1} \Pr(\overline{D}_1 \leq c_{\overline{D}_1}) = \Pr \left( \overline{D}^{(1)}(\eta) \leq c_{\overline{D}_1} \right),
$$

where $\overline{B}^{(1)}(\eta) = l^T\overline{W}^{(1)}l / l^T\overline{M}l$ and $\overline{D}^{(1)}(\eta) = -l^T\overline{M}T^{(1)}\overline{M}l / l^T\overline{M}l$, $l = \overline{P}\eta + \overline{m}$, $\overline{P} = JP$, $J$ is an $n \times (n - 1)$ matrix such that $J'r = [0 : I_{n-1}]$, $P$ is an $(n - 1) \times (n - 1)$ lower triangular matrix with ones on and below the diagonal and zeroes elsewhere, and $\eta \sim N(0, I_{n-1})$.

**Proof.** The proof is omitted here for brevity, but is available on request from the authors. Part of the proof is obtained along the lines of the proof to theorem B.1 of Banerjee and Magnus (1999).
only if $RS^{-1}X' \neq 0$, that is, at least one of the restrictions involves the intercept. Therefore, part (i) of Theorem 3 is relevant to the regression that either has restrictions which involve the intercept term, or has no intercept at all. In this case, the limiting probabilities of $\overline{B}1 > c_{\overline{B}1}$ and $\overline{D}1 \leq c_{\overline{D}1}$ take only the value one or zero depending on the conditions (15) and (16) which in turn depend on the underlying data matrix. This result is valid under both correct and incorrect restrictions. Second, $\overline{M}i = 0$ when $Mi = 0$ and $RS^{-1}X'i = 0$, which implies that the regression corresponding to part (ii) of Theorem 3 must contain an intercept but no restriction is placed on the intercept. In this case, the probabilities of $\overline{B}1 > c_{\overline{B}1}$ and $\overline{D}1 \leq c_{\overline{D}1}$ will tend to a constant lying between zero and one. Again, this finding holds irrespective of whether the restrictions are correct or incorrect, although the precise values of the probability limits can differ between the two cases.

3. NUMERICAL RESULTS

This section presents numerical results to illustrate the behaviour of $\overline{B}s$ and $\overline{D}s$ under situations of correct or incorrect restrictions. Two forms of disturbances are considered. In the first, $ut$ is generated by the stationary AR(1) process of the form $ut = \phi_1 u_{t-1} + \varepsilon_t ; |\phi_1| < 1$, and so

$$\Omega(\phi_1) = (\omega_{ij}(\phi_1)),$$

where $\omega_{ij}(\phi_1) = \begin{cases} 1/(1 - \phi_1^2), & \text{if } i = j, \\ \phi_1^{j-i}/(1 - \phi_1^2), & \text{if } i \neq j. \end{cases}$ (19)

In the second, $ut$ follows the MA(1) process $ut = \psi_1 \varepsilon_{t-1} + \varepsilon_t$, so that

$$\Omega(\psi_1) = (1 + \psi_1^2)I_n + \psi_1 T^{(1)},$$

where $T^{(1)}$ is the symmetric Toeplitz matrix defined earlier. For both AR(1) and MA(1) disturbances, $A_s = \partial \Omega(\phi_1)/\partial \phi_1|_{\phi_1=0} = \partial \Omega(\psi_1)/\partial \psi_1|_{\psi_1=0} = T^{(1)}$. We denote $\overline{B}s$ and $\overline{D}s$ as $\overline{B}1$ and $\overline{D}1$, respectively as counterparts to the statistics $B1$ and $D1$ in Banerjee and Magnus (1999). As in the unrestricted model, $\overline{D}1$ is essentially the DW test statistic in the restricted least squares regression. Thus, in addition to being a sensitivity measure, the probability $Pr(\overline{D}1 \leq c_{\overline{D}1})$ can also be interpreted as the power of the DW test in testing $\phi_1 = 0$ against $\phi_1 > 0$ in a restricted regression.

Recognizing the data dependence of each statistic’s distribution, the numerical evaluations we have undertaken involve a variety of data based on columns or linear combinations of columns from the following two data sets: the first comprises the eigenvectors $t_1, t_2, \ldots, t_n$ that correspond to eigenvalues of the $n \times n$ Toeplitz matrix $T^{(1)}$; in the second data set, the regressors are $s_1 = i_n/\sqrt{n}$, $s_\ell = (i_{\ell-1}, 1 - \ell, 0'_{n-\ell})/\sqrt{(\ell - 1)}$, where $i_{\ell}$ is an $\ell \times 1$ vector of ones, $0_{n-\ell}$ is an $(n - \ell) \times 1$ vector of zeroes, $2 \leq \ell \leq n$. Note that the constant dummy regressor $s_1$ may be used to represent an intercept term in the regression matrix $X$.

For each model, we set $n = 15, k = 5, m = 2$ and $R = [I_2 : 0_{2 \times 3}]$. Specifically, the design matrix $X = [X_1 : X_2]$ are column orthogonal, with the norm of each column vector being equal to unity. So we have $\overline{M} = I - X_2'X_2$ and $M = I - XX'$ for each data set comprising $X$. The restrictions

---

1In (15) and (16), $\overline{B}1$ and $\overline{D}1$ are, respectively the probability limits of $\overline{B}1$ and $\overline{D}1$ as $\phi_1 \to 1$.

2Our results hold also for the case of column non-orthogonal design matrices. In fact, when $R \neq r$, the restricted estimators’ sensitivity statistics are merely congruent transformations of those in the unrestricted regression. A proof is available from the authors upon request.

© Royal Economic Society 2007
are in the form of $\beta(2) = r$, where $\beta(2)$ is a $2 \times 1$ vector comprising the first two elements of $\beta$. The degree of restriction misspecification introduced is controlled through the elements in the vector $r$, the values of which have been varied to determine values of $\delta$, set to 0, 5, 34, 125 and 500. Non-zero values of $\delta$ measure departures from valid restrictions, as is used conventionally in the pre-test literature. To measure the sensitivity of the estimators, we first determine the values $c_{P_1}$ and $c_{T_1}$ such that $\Pr(B_1 > c_{P_1}) = \Pr(D_1 \leq c_{T_1}) = \alpha$ under white noise disturbances. Then we calculate the probabilities $\Pr(B_1 > c_{P_1})$ and $\Pr(D_1 \leq c_{T_1})$ for values of $\phi_1$ and $\psi_1$ lying between 0 and 1. These probabilities measure the sensitivity (or lack of robustness) of $\tilde{\gamma}(\theta)$ and $\sigma^2(\theta)$ with respect to $\phi_1$ or $\psi_1$. Any significant deviation of these probabilities from $\alpha$ indicates sensitivity of the estimators to covariance misspecification, and vice versa.

Figures 1–6 illustrate a selection of our results which are considered to be representative of the various patterns that have emerged. The regression models (1–4) corresponding to the figures are characterized in Table 1. In models 1 and 2, the regressions contain no intercept term. Model 3 is based on a regression with an intercept which is part of a restriction. In model 4, the regression has an intercept but it is not part of any restrictions. In all cases we set $\alpha = 0.05$.

There are several results of interest emerging from the numerical analysis. First, our numerical findings are in accord with the theoretical results of Theorem 3 in the case of AR(1) errors as $\phi_1 \rightarrow 1$. With the exception of those of model 4, the reported numerical results show limiting probability values of 0 or 1 (Figures 1–3). Whether the limit is 0 or 1 is determined by the conditions given in (15) and (16). For example, in the case of model 1 with $\delta = 0$, we have $c_{T_1} = -0.7753 > d_1 = -1.5632$, and hence $\Pr(D_1 \leq c_{T_1})$ tends to one as $\phi_1$ approaches unity (Figure 1). In model 4, the dummy regressor $s_1$ is not part of the $X_1$ matrix, and accordingly the limiting probabilities lie between 0 and 1 (Figure 4). When the restrictions are correct (i.e. $\delta = 0$),

\[\begin{array}{c}
\text{Figure 1. } D_1 \text{ under AR(1) errors (Model 1).}
\end{array}\]

In the case of $\delta > 0$, our calculations of $c_{P_1}$ and $c_{T_1}$ do not allow, a priori, that the investigator is ignorant of the restrictions being incorrect. In any case, there is not the issue of size-distortion here because the $\alpha$ level merely serves as a benchmark for comparison and is not intended as a measure of the Type I error of a hypothesis test.

Complete findings including a larger range of graphs are available from the authors on request.

© Royal Economic Society 2007
the probability curves have essentially the same characteristics as when there are no restrictions. Specifically, in the case of MA(1) disturbances, \( \Pr(\beta_1 > c_{\beta_1}) \) is fairly constant; it does not deviate significantly from 0.05 over the entire range of \( \psi_1 \) (Figure 5). This observation may be taken to imply that \( \overline{\psi}(\psi_1) \) and \( \overline{\beta}(\psi_1) \) are insensitive to changes in \( \psi_1 \). On the other hand, \( \Pr(D_1 \leq c_{D_1}) \) has a clear tendency to deviate from 0.05 as \( \psi_1 \) increases; the degree of deviation, however, is less acute under MA(1) disturbances than under AR(1) disturbances (Compare Figures 1 and 6).

7This result is expected because \( \overline{\beta}_1 \) and \( \overline{D}_1 \) are distributional invariant with respect to the form of \( X_1 \) whenever the restrictions \( \beta_{(2)} = r \) hold. In fact, for the general case of column non-orthogonal matrices, provided that the restrictions are correct, with a transformation of data, the sensitivity statistics of the restricted model have essentially the same properties as those of the unrestricted model. Proof is available from the authors on request.

© Royal Economic Society 2007
In the latter case, there is also a decisive tendency for $\Pr(\bar{B}_1 > c_{\bar{B}_1})$ to move away from 0.05 as $\phi_1 \to 1$.

With regard to the consequences of incorrect restrictions on the behaviour of the sensitivity statistics, it is found that imposing wrong restrictions generally has the effect of pulling down the probability curves of $\overline{D}_1$ for both the AR(1) and MA(1) cases. As is obvious from the figures, this occurs most markedly for the cases where the restriction misspecification is relatively serious. Exceptions occur, for example, in Figure 2, where an increase in $\delta$ from 34 to 125 or 500 can result in an upward shift in the probability curve of $\overline{D}_1$ in regions of the parameter space near the unit-root. Figure 2 also reveals a sudden dip in $\Pr(\overline{D}_1 \leq c_{\overline{D}_1})$ as $\phi_1$ tends to 1 for $\delta = 125$ and 500. This behaviour is somewhat unusual, but is nevertheless consistent with the results.
Sensitivity of the restricted least squares estimators

Figure 6. $\overline{D}1$ under MA(1) errors (Model 1).

Table 1. Regression models for Figures 1–6.

<table>
<thead>
<tr>
<th>Model</th>
<th>$X_1$</th>
<th>$X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$(t_3 - t_6, t_5 - t_{12})/\sqrt{2}$</td>
<td>$(t_1, t_2, t_4)$</td>
</tr>
<tr>
<td>2</td>
<td>$(t_{11}, t_{14})$</td>
<td>$(t_1 + t_{13}, t_2 + t_{12}, t_3 + t_{10})/\sqrt{2}$</td>
</tr>
<tr>
<td>3</td>
<td>$(s_1, s_{15})$</td>
<td>$(s_4 + s_{11}, s_5 + s_{10}, s_7 + s_9)/\sqrt{2}$</td>
</tr>
<tr>
<td>4</td>
<td>$(s_{11}, (s_{13} + s_{15})/\sqrt{2})$</td>
<td>$(s_1, (s_3 + s_{12})/\sqrt{2}, (s_6 + s_{10})/\sqrt{2})$</td>
</tr>
</tbody>
</table>

reported elsewhere of the DW test having power that drops sharply to zero near the unit-root (see, King 1985, Krämer 1985, Krämer and Zeisel 1990, Bartels 1992). Turning to the behaviour of $\Pr(\overline{B}1 > c_{\overline{B}1})$, it is apparent that in the case of MA(1) errors, the probability values are generally not significantly altered by introducing constraint specification errors (Figure 5). For AR(1) errors, the shifts in the probability curves are more pronounced as $\delta$ increases. Commonly, an increase in restriction specification errors has the effect of lessening the coefficient estimators’ sensitivity to covariance misspecification. Again, exceptions have been found to occur near the unit-root; in Figure 3, for example, the probability curves of $\overline{B}1$ for large values of $\delta$ are pulled significantly upwards as $\phi_1$ approaches unity. Altogether, the clear conclusion arising is that the estimators’ sensitivity to covariance misspecification is usually weakened but can also be intensified as a result of specification errors in the restrictions. Generally speaking, the effect of incorrect restrictions on the sensitivity is greater for AR(1) than for MA(1) errors. Comparing $\overline{B}1$ and $\overline{D}1$, the latter is found to be less robust against specification errors in the restrictions. In all cases the data matrix used is an important factor and has a large impact on the results.

4. CONCLUSIONS

This paper has addressed the sensitivity of the restricted least squares estimator in the linear regression model to misspecification in the error covariance matrix. Our results offer some interesting insights into the practical question of whether it matters at all if one fails to take
account of the fact that the errors have a non-scalar covariance matrix. Some exact analytical results are derived and these are evaluated for various data sets and for the cases of AR(1) and MA(1) errors. The principal conclusions to be drawn from these results may be stated quite briefly, and in some cases they reinforce the conclusions of Banerjee and Magnus (1999). First of all, while the restricted variance estimator is generally very sensitive to covariance misspecification, whether or not the corresponding estimators of the regression coefficients are sensitive depend largely on the error process and the underlying data matrix. It is clear that with MA(1) errors, \( \hat{\beta}(\theta) \) is insensitive regardless of the data. If the disturbances are AR(1), then \( \hat{\beta}(\theta) \) can still be quite robust against covariance misspecification for small to moderate values of the autocorrelation parameter. For highly correlated AR(1) errors, however, the results are somewhat mixed and depend largely on the underlying data matrix. The extent to which the \( X \) matrix affects the results is a notable feature of this study. The latter issue prevails when one considers the effect of restriction misspecification on the sensitivity of the estimators. With AR(1) errors, depending on the underlying data, specification errors in the restrictions usually weaken but sometimes also intensify the estimators’ sensitivity. It remains for future research to consider other forms of covariance misspecification frequently encountered in practice (e.g., ARCH errors) and the sensitivity of other common estimators in linear regression. A separate paper by the authors has addressed the sensitivity of the pre-test estimator which is a weighted average of the unrestricted and restricted estimators (Judge and Bock 1978).

ACKNOWLEDGMENTS

The authors thank Anurag Banerjee, Jan Magnus, Kazuhiro Ohtani, the Coordinating Editor Pravin Trivedi and the referees at various stages for comments and suggestions that have greatly improved the article. The first and second authors acknowledge financial support from the Hong Kong Research Grant Council (Grant No. 9040762) and the National Natural Science Foundation of China (Grant No. 70625004), respectively.

REFERENCES


APPENDIX

Proof of Theorem 1

We only give the proof to part (ii) of the theorem. Part (i) can be obtained rather straightforwardly by algebraic manipulation, and part (iii) follows directly from the proof to part (ii). Now, to prove part (ii), note that from (i) of Theorem 1, we may write,

\[ \overline{B}_s = \frac{Q_s^{(1)}(\overline{v})}{Q_s^{(1)}(\overline{v}) + Q_s^{(2)}(\overline{v})}, \]

where \( Q_s^{(1)}(\overline{v}) = \overline{v} \overline{Q} \overline{W}_s \overline{Q} \) and \( Q_s^{(2)}(\overline{v}) = \overline{v} (I_{n-k+m} - \overline{Q} \overline{W}_s \overline{Q}) \). Note that \( \overline{Q} \overline{W}_s \overline{Q} \) is a symmetric idempotent matrix. Hence \( (I_{n-k+m} - \overline{Q} \overline{W}_s \overline{Q}) \overline{W}_s \overline{Q} = 0 \) which implies that \( Q_s^{(1)}(\overline{v}) \) and \( Q_s^{(2)}(\overline{v}) \) are independent if \( u \sim N(0, \sigma^2 I_n) \), that is,

\[ \overline{v} \sim N \left( \sigma^{-1} \overline{Q} \chi_{\rho}^{-1}(R \overline{S}^{-1} R')^{-1}(R\beta - r), I_{n-k+m} \right) \] (A.1)

Clearly, \( \text{rank}(\overline{Q} \overline{W}_s \overline{Q}) = \text{tr}(\overline{Q} \overline{W}_s \overline{Q}) = \rho \) and \( \text{rank}(I_{n-k+m} - \overline{Q} \overline{W}_s \overline{Q}) = n - k + m - \rho \). So, if (A.1) holds, \( Q_s^{(1)}(\overline{v}) \sim \chi_{n-k+m-\rho}^2(\delta_s^{(1)}(0)) \) and \( Q_s^{(2)}(\overline{v}) \sim \chi_n^2(\delta_s^{(1)}(0)) \), where \( \delta_s^{(1)}(0) \) and \( \delta_s^{(2)}(0) \) are the non-centrality parameters of the doubly non-central Beta distribution given in (12). The result of (12) then follows by noting the independence of \( Q_s^{(1)}(\overline{v}) \) and \( Q_s^{(2)}(\overline{v}) \).

Proof of Theorem 2

The following lemmas form the basis of the proof to Theorem 2:

Lemma A.1.

Let Beta\((a, b)\) be the Beta function. Denote \( g(h) = \int_0^1 t^h \exp(\rho t)dt \) and \( \vartheta(\rho) = 1 - \rho \sum_{i=0}^{\infty} \frac{(-1)^i \rho^i}{i!} \text{Beta}(s + 1, h + 5), \) where \( h \geq 0 \) and \( \rho \geq 0 \). Then \( |\vartheta(\rho)| \leq \exp(\rho) \) and

\[ g(h) = (\frac{1}{h+1} + \frac{\rho}{(h+1)(h+2)} + \frac{\rho^2}{(h+1)(h+2)(h+3)} - \frac{\rho^3 \vartheta(\rho)}{(h+1)(h+2)(h+3)(h+4)}) \exp(\rho). \]

Proof.

Write \( g(h) = \exp(\rho) \int_0^1 t^h \exp(\rho(t - 1))dt \). By letting \( 1 - t = \tilde{t} \) in the integral and rewriting \( \tilde{t} \) as \( t \), we can write \( g(h) = \exp(\rho) \int_0^1 (1 - t)^h \exp(-\rho t)dt \). Applying Taylor’s expansion on \( \exp(-\rho t) \), we have

\[ g(h) = \exp(\rho) \sum_{i=0}^{\infty} \frac{(-1)^i \rho^i}{s!} \text{Beta}(s + 1, h + 1). \] (A.2)

Clearly, the first three terms of the infinite series in (A.2) correspond to the last expression of \( g(h) \) stated in the lemma; the remainder is

\[ \gamma(\rho, h) = \exp(\rho) \sum_{i=3}^{\infty} \frac{(-1)^i \rho^i}{s!} \text{Beta}(s + 1, h + 1) \]

\[ = - \frac{\rho^3 \exp(\rho) \vartheta(\rho)}{(h+1)(h+2)(h+3)(h+4)}, \]

© Royal Economic Society 2007
where
\[ \vartheta^*(\rho) = \sum_{s=0}^{\infty} \frac{(-1)^s \rho^s}{s!} a(s, h) \tag{A.3} \]
and \( a(s, h) = \frac{\Gamma(s+1)\Gamma(h+5)}{\Gamma(s+h+5)} \). It can be shown that \( \sup_{s \geq 0} |a(s, h)| = 1 \). Clearly, \( a(0, h) = 1 \). For any \( s \geq 1 \), \( a(s, h) = s \text{Beta}(s, h + 5) = s \int_0^1 t^{s-1}(1 - t)^{h+4} dt = \int_0^1 (1 - t^{1/s})^{h+4} dt < 1 \). So it follows from (A.3) that \( |\vartheta^*(\rho)| \leq \exp(\rho) \) and
\[ \vartheta^*(\rho) = 1 + \sum_{s=1}^{\infty} \frac{(-1)^s \rho^s}{s!} a(s, h) = 1 - \rho \sum_{s=1}^{\infty} \frac{(-1)^s \rho^{s-1}}{(s-1)!} a(s, h) \]
\[ = 1 - \rho \sum_{s=0}^{\infty} \frac{(-1)^s \rho^s}{s!} \text{Beta}(s+1, h+5) = \vartheta(\rho). \]

This completes the proof of Lemma A.1.

The next lemma is related to the formulation of product moments of two ratios in random variables, which is obtained along the lines of lemma 4 in Magnus (1986). The details of the proof are available from the authors upon request.

**Lemma A.2.** Let \( q_1, q_2 \) and \( q_3 \) be random variables subject to the condition that \( \Pr(q_3 > 0) = 1 \). Assume the joint moment generating function of \( (q_1, q_2, q_3) \), defined as \( \varphi(t_1, t_2, t_3) = \mathbb{E}\{\exp(t_1 q_1 + t_2 q_2 + t_3 q_3)\} \), exists for all \( |t_1| < \varepsilon, |t_2| < \varepsilon \) and \(-\infty < t_3 < \varepsilon\), where \( \varepsilon > 0 \) is an arbitrary constant. Assume further that the expectation of \( q_1 q_2 q_3^{-2} \) also exists, then we have
\[ \mathbb{E}(q_1 q_2 q_3^{-2}) = \int_0^\infty t_3 \left( \frac{\partial^2 \varphi(t_1, t_2, -t_3)}{\partial t_1 \partial t_2} \right) \left. \right|_{t_1=t_2=0} dt_3. \]

Using Lemmas A.1 and A.2, we prove Theorem 2 as follows:

(i) Write \( q_1 = v' \overline{W}_v, \quad q_2 = v' \overline{MA}_v, \quad q_3 = v' \overline{M} v \), where \( v = u/\sigma + X'S^{-1}R'(RS^{-1}R)^{-1}(R\beta - 1)/\sigma \). It follows that \( \overline{W} = q_1 / q_3 \) and \( \overline{M} = q_2 / q_3 \). In Theorem 2, \( \psi \) is evaluated at \( \theta = 0 \). Then, by lemma 5 in Magnus (1986), we obtain
\[ \varphi(t_1, t_2, t_3) = \left| I_n - G \right|^{-1/2} \exp \left( -\frac{\delta}{2} \right) \exp \left( \frac{1}{2} \left[ I_n - G \right]^{-1} \psi \right) \tag{A.4} \]
as the joint moment generating function of \( (q_1, q_2, q_3) \), where \( G = t_1 \overline{W}_v + t_2 \overline{MA}_v, \overline{M} = t_3 \overline{M} \). Now that \( \overline{M} \) is an idempotent matrix, there exists some orthogonal matrix \( H \) such that \( H \overline{M} H' = \text{diag}(I_{n-k+m}, 0) \) and \( H(I_n - \overline{M})H' = \text{diag}(0, I_{k-m}) \). Thus, we can write
\[ \varphi(t_1, t_2, -t_3) = \left| \Delta \right| \exp \left( -\frac{\xi}{2} \right) \exp \left( \frac{1}{2} \left[ I_n - I_{n-k+m} \right]^{-1} \xi \right) \tag{A.5} \]
where \( \Delta = \left( I_n + 2t_3 \text{diag}(I_{n-k+m}, 0) \right)^{-1/2}, \quad \xi = \Delta H \overline{M}, \quad \psi(t_1, t_2) = \left| I_n - t_1 R_1 - t_2 R_2 \right|^{-1/2} \exp \left( -\frac{\xi}{2} \right) \exp \left( \frac{1}{2} \left[ I_n - t_1 R_1 - t_2 R_2 \right]^{-1} \xi \right) \tag{A.6} \]
where \( R_1 = \Delta H \overline{W}_v, \Delta \) and \( R_2 = \Delta H \overline{MA}_v, \Delta H \overline{M} \). Equation (A.6) gives the joint moment generating function of \( (\eta R_1 \eta, \eta R_2 \eta) \), where \( \eta \sim N(\xi, I_n) \). Clearly, \( \psi(t_1, t_2, -t_3) \) depends on \( t_1 \) and \( t_2 \) only through

© Royal Economic Society 2007
\( \psi (t_1, t_2) \). By virtue of the properties of moment generating functions,

\[
\frac{\partial^2 \psi(t_1, t_2)}{\partial t_1 \partial t_2} \bigg|_{t_1=t_2=0} = E(\eta'R_1 \eta'R_2 \eta).
\] (A.7)

It then follows from (A.5) to (A.7) and Lemma A.2 that

\[
E(\overline{B}, \overline{D}_s) = -E(q_1 q_2 s^3) = -\exp \left( -\frac{\delta}{2} \right) \int_0^\infty t_3 |\Delta| \exp \left( \frac{\delta \xi}{2} \right) E(\eta'R_1 \eta'R_2 \eta) dt_3.
\] (A.8)

Applying (3.2d.11) of theorem 3.2d.3 in Mathai and Provost (1992), we observe that

\[
E(\eta'R_1 \eta'R_2 \eta) = 2tr(R_1 R_2) + 4\xi'R_1 R_2 \xi + (\xi'R_1 \xi + tr(R_1))(\xi'R_2 \xi + tr(R_2)).
\] (A.9)

From this and by some manipulations, we obtain

\[
E(\overline{B}, \overline{D}_s) = -\exp \left( -\frac{\delta}{2} \right) \int_0^\infty (1 + 2t_3)^{a-k+m/2} \left\{ \begin{array}{l}
2tr(A_s W_s) + r_s tr(A_s M) \\
+ r_s \mu A_s \mu + \mu W_s \mu tr(A_s M) + 4\mu W_s A_s \mu + \mu W_s \mu A_s \mu \\
\end{array} \right\} \times \exp \left( \frac{\delta}{2t_3} \right) dt_3.
\] (A.10)

Making use of the transformation \( t = 1/(1 + 2t_3) \), (A.10) becomes

\[
E(\overline{B}, \overline{D}_s) = -\frac{1}{4} \exp \left( -\frac{\delta}{2} \right) \int_0^1 (1-t)^{(n-k+m)/2-1} \left\{ 2tr(A_s W_s) + r_s tr(A_s M) \\
+ [r_s \mu A_s \mu + \mu W_s \mu tr(A_s M) + 4\mu W_s A_s \mu + \mu W_s \mu A_s \mu ]t \\
\times \exp \left( \frac{\delta}{2t} \right) dt.
\] (A.11)

Using theorem 6 of Magnus (1986) and after some manipulations, we have

\[
E(\overline{B}_s) = \frac{1}{2} \exp \left( -\frac{\delta}{2} \right) \int_0^1 t^{(n-k+m)/2-1} (\overline{r}_s + \mu W_s \mu t) \exp \left( \frac{\delta}{2t} \right) dt,
\] (A.12)

\[
E(\overline{D}_s) = -\frac{1}{2} \exp \left( -\frac{\delta}{2} \right) \int_0^1 t^{(n-k+m)/2-1} (tr(A_s M) + \mu A_s \mu t) \exp \left( \frac{\delta}{2t} \right) dt,
\] (A.13)

\[
E(\overline{B}_s^2) = \frac{1}{4} \exp \left( -\frac{\delta}{2} \right) \int_0^1 (1-t)^{(n-k+m)/2-1} (\overline{r}_s + \mu W_s \mu t)^2 \exp \left( \frac{\delta}{2t} \right) dt \\
+ \frac{1}{2} \exp \left( -\frac{\delta}{2} \right) \int_0^1 (1-t)^{(n-k+m)/2-1} (\overline{r}_s + 2\mu W_s \mu t) \exp \left( \frac{\delta}{2t} \right) dt
\] (A.14)
and
\[
E(D^2_s) = \frac{1}{4} \exp \left( -\frac{\delta}{2} \right) \int_0^1 (1-t)^{(n-k+m)/2-1} \left( \text{tr}(A_s\overline{M})\overline{\tau}_s + \overline{\mu} A_s\mu \right)^2 \exp \left( \frac{\delta}{2} t \right) dt \\
+ \frac{1}{2} \exp \left( -\frac{\delta}{2} \right) \int_0^1 (1-t)^{(n-k+m)/2-1} \left( \text{tr}[(A_s\overline{M})^2] + \overline{\mu} A_s\overline{M}A_s\mu \right) \exp \left( \frac{\delta}{2} t \right) dt
\]  
(A.15)

Thus, (A.11)–(A.13) give the final expression of (13).

(ii) The result in (14) follows straightforwardly by setting \( \delta = 0 \) in (13).

(iii) To evaluate the convergence rates of \( \text{cov}(B_s, D_s^2) \), \( \text{var}(B_s^2) \) and \( \text{var}(D_s^2) \), recall that \( \overline{W}_s \) is an idempotent symmetric matrix with rank \( rs \leq k - m \) and \( \delta = \overline{\mu} \mu < \infty \). Hence,
\[
\overline{\mu} \mu \leq \delta.
\]  
(A.16)

Note that the eigenvalues of \( A_s \) are bounded. Therefore,
\[
|\text{tr}(A_s\overline{M})| \leq \mu(n-k+m),
\]  
(A.17)
\[
|\text{tr}(A_s\overline{W}_s)| \leq \mu\overline{\tau}_s
\]  
(A.18)

and
\[
|\text{tr}[(A_s\overline{M})^2]| \leq \mu^2(n-k+m),
\]  
(A.19)

where \( \mu \) is the upper bound (in absolute value) of the eigenvalues of \( A_s \). For \( m = 0 \), results analogous to (A.17)–(A.19) are given in equation (22) of Banerjee and Magnus (2000). Also, by the same conditions on \( A_s \) and \( \delta \), we have
\[
|\overline{\mu} A_s\mu| \leq \mu\delta,
\]  
(A.20)
\[
\overline{\mu} A_s\overline{M}A_s\mu \leq \mu^2\delta,
\]  
(A.21)

and
\[
|\overline{\mu} \mu \overline{W}_s A_s\mu| \leq \mu\delta.
\]  
(A.22)

In fact, (A.20) follows from the fact that \( -\mu I_n \leq A_s \leq \mu I_n \). To prove (A.21) and (A.22), define \( ||V|| = \sup_{x \in \mathbb{R}^n} \sqrt{x'Vx} \) for any \( n \times n \) matrix \( V \). Then \( ||V|| \) is some kind of matrix norm of \( V \). If \( V \) is symmetric, then \( ||V|| \) is the maximum of the eigenvalues of \( V \) in absolute values. Accordingly, \( ||\overline{W}_s|| = ||\overline{M}|| = 1 \). By the consistency of defined matrix norm,
\[
||\overline{M}A_s\overline{M}A_s\overline{M}|| = ||\overline{M}A_s\overline{M}A_s\overline{M}|| \\
\leq ||A_s|| \times ||A_s|| \leq \mu^2.
\]  
(A.23)

Observe that the last inequality in (A.23) implies \( \overline{M}A_s\overline{M}A_s\overline{M} \leq \mu^2 I_n \), which in turn implies (A.21) is true. Now that \( A_s \) is symmetric, we have,
\[
2\overline{\mu} \overline{W}_s A_s\mu = \overline{\mu} \overline{W}_s A_s\mu + (\overline{\mu} \overline{W}_s A_s\mu)' = \overline{\mu} (\overline{W}_s A_s + A_s\overline{W}_s) \overline{\mu}.
\]  
(A.24)

By the triangular inequality of matrix norm, it follows that
\[
||\overline{W}_s A_s + A_s\overline{W}_s|| \leq ||\overline{W}_s A_s|| + ||A_s\overline{W}_s|| \leq ||\overline{W}_s|| \times ||A_s|| + ||A_s|| \times ||\overline{W}_s|| = 2 ||A_s|| \leq 2\mu.
\]  
(A.25)
The last inequality in (A.25) implies that $-2\mu_n \leq \mathbf{W}_s A_s + A_s \mathbf{W}_s \leq 2\mu_n$, from which (A.22) follows.

The above results are useful for evaluating the convergence rates of the underlying covariances and variances. By Lemma A.1, it follows from (13), (14), (A.16)–(A.18), (A.20) and (A.22) that

$$\text{cov}(\mathbf{B}_s, \mathbf{D}_s) = CV_0 - \frac{4\mu \mathbf{W}_s A_s \mu}{(n - k + m + 2)(n - k + m + 4)} + \frac{4(\mathbf{r}_s A_s \mu + \mu \mathbf{W}_s \mu \text{tr}(A_s \mathbf{M}))}{(n - k + m)(n - k + m + 2)(n - k + m + 4)} + \frac{2[3(n - k + m) + 10](\mu A_s \mu \mathbf{W}_s \mu)}{(n - k + m + 2)^2(n - k + m + 4)(n - k + m + 6)} + \delta O(n^{-3})$$

$$= O(n^{-2}).$$

(A.26)

where $CV_0 = \text{cov}(\mathbf{B}_s, \mathbf{D}_s|\delta = 0)$ which is defined in the theorem. Recall that $0 < \mathbf{r}_s \leq \min(k - m, n - k + m)$. We observe from (A.12), (A.14), (A.16) and Lemma A.1 that

$$\text{var}(\mathbf{B}_s) = \frac{2r_s}{(n - k + m)(n - k + m + 2)} \left[ 1 - \frac{\mathbf{r}_s}{n - k + m} \right] + \frac{4\mu \mathbf{W}_s \mu}{(n - k + m)(n - k + m + 4)} - \frac{8\mathbf{r}_s \mu \mathbf{W}_s \mu}{(n - k + m + 2)(n - k + m + 4)}$$

$$- \frac{2[3(n - k + m) + 10](\mu \mathbf{W}_s \mu)}{(n - k + m + 2)^2(n - k + m + 4)(n - k + m + 6)} + \delta O(n^{-3})$$

$$= O(n^{-2}).$$

(A.27)

By (A.13), (A.15), (A.17), (A.19)–(A.21) and Lemma A.1, we obtain

$$\text{var}(\mathbf{D}_s) = \frac{2\overline{\mathbf{r}}_s}{(n - k + m)(n - k + m + 2)} + \frac{4\mu' A_s \mathbf{M} A_s \mu}{(n - k + m + 2)(n - k + m + 4)} - \frac{8\mu' A_s \mathbf{M} A_s \mu}{(n - k + m)(n - k + m + 2)(n - k + m + 4)}$$

$$- \frac{2[3(n - k + m) + 10](\mu A_s \mu)}{(n - k + m + 2)^2(n - k + m + 4)(n - k + m + 6)} + \delta O(n^{-3}).$$

(A.28)

Note that the absolute values of the eigenvalues of $A_s$ are bounded. So we have $\overline{\Lambda}_s \leq c_2 s n^2$ for some constant $c_2 > 0$. If, in addition, we assume that $\overline{\Lambda}_s \geq c_1 s n^2$ for some constant $c_1 > 0$, then $\overline{\Lambda}_s$ is of order $n^2$. It follows that

$$\text{var}(\mathbf{D}_s) = O(n^{-1}).$$

(A.29)

Part (iii) of Theorem 2 then follows from (A.26), (A.27), (A.29) and the definition of $\text{corr}(\mathbf{B}_s, \mathbf{D}_s)$.

---

8The derivation of (A.27) and (A.28) are very tedious and cumbersome. Details are available upon request from the authors.

© Royal Economic Society 2007