ELSEVIER

Journal of Econometrics 76 (1997) 171–191

JOURNAL OF Econometrics

Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator

Kenneth D. West*

Department of Economics, University of Wisconsin, Madison, WI 53706, USA

(Received August 1994; final version received August 1995)

Abstract

A $\sqrt{T}$-consistent estimator of a heteroskedasticity and autocorrelation consistent covariance matrix estimator is proposed and evaluated. The relevant applications are ones in which the regression disturbance follows a moving average process of known order. In a system of l equations, this 'MA-l' estimator entails estimation of the moving average coefficients of an l-dimensional vector. Simulations indicate that the MA-l estimator's finite sample performance is better than that of the estimators of Andrews and Monahan (1992) and Newey and West (1994) when cross-products of instruments and disturbances are sharply negatively autocorrelated, comparable or slightly worse otherwise.

Key words: Moving average; Time series; Serial correlation; Spectral density; Inference; Hypothesis test
JEL classification: C12; C22; C32

1. Introduction

This paper proposes and evaluates an estimator of a heteroskedasticity- and autocorrelation-consistent covariance matrix that is positive semidefinite by construction. The estimator is applicable when the regression disturbance follows a moving average (MA) process of known order, and the innovations in this moving average process have zero mean conditional on past disturbances and

I thank Ka-fu Wong and Chia-Yang Hueng for excellent research assistance and Ka-fu Wong, two referees, and an Associate Editor for helpful comments and discussion. I also thank the National Science Foundation and the University of Wisconsin Graduate School for financial support.

PII S0304-4076(95)01788-X

current and past instruments. I prove that the estimator, which is parametric, is $\sqrt{T}$-consistent under mild conditions. This means that it is asymptotically more efficient than the nonparametric estimators emphasized in recent work such as Andrews (1991), Andrews and Monahan (1992), and Newey and West (1994).

Simulations are used to evaluate the finite sample performance of hypothesis tests about a parameter in a linear model. Consistent with some asymptotic calculations worked out for a simple example, these simulations indicate that the estimator works relatively well – has relatively accurately sized tests – when cross-products of instruments and disturbances are sharply negatively correlated. The simulations also indicate that these are precisely the circumstances under which an estimator known as the 'truncated' one is likely to fail to be positive semidefinite. Since repeated occurrence of this failure in empirical work was one of the major spurs to development of alternative covariance matrix estimators, I take the implication to be that such circumstances are empirically relevant ones.[1] The simulations also indicate, however, that when the estimator's asymptotic advantages relative to nonparametric estimators are relatively small (but still nonzero), the estimator works comparably or slightly worse than the nonparametric ones.

A second contribution of the simulations is to evaluate some existing estimators when cross-products of instruments and disturbances are negatively autocorrelated. When the negative autocorrelation is sufficiently strong, some earlier estimators have a tendency to reject too infrequently, rejecting at the 5 percent level, for example, in distinctly less than 5 percent of the simulations. This complements the Andrews and Monahan (1992) and Newey and West (1994) result that strong positive autocorrelation tends to cause the nonparametric estimators to reject too often.

The proposed estimator, which generalizes one suggested by Hodrick (1991), is more restrictive than the nonparametric ones now in common use. It is not applicable when the order of the moving average of the disturbance is not known or is infinite, as sometimes happens in empirical work. But in many studies the null specification implies a moving average of known order. Examples in which this is the case include: evaluation of multi-period forecasts, using either financial market (e.g., Hansen and Hodrick, 1980), or survey data (e.g., Brown and Maital, 1981), Euler equations (first-order conditions) from rational

[1] Unfortunately, empirical papers typically do not provide enough information to allow one to deduce the autocorrelations of cross-products of instruments and disturbances. But perhaps the unpublished calculations underlying some of my own work on monthly, aggregate inventories is representative. West and Wilcox (1996) considered a model with an MA(2) disturbance. Underlying the estimates in Table 7 of West and Wilcox (1996) are cross-products of instruments and disturbances whose estimated first-order autocorrelations are around -0.6. In the simulations, the proposed estimator performs relatively well with an MA(2) process calibrated to the West and Wilcox estimates. See Table 2, panel C, column (5) below.

expectations models when there are costs of adjustment (e.g., West, 1986), nonseparable utility (e.g., Eichenbaum et al., 1988), and/or unobservable moving average shocks (e.g., Kollintzas, 1993), time-aggregated models (e.g., Hansen and Singleton, 1990).

As was noted above, the estimator also requires that the innovation in the regression disturbance have a zero mean conditional on past disturbances and current and past instruments. This means that the best predictor of the disturbance is the same as the best linear predictor, and so is not implied by a conventional stationarity assumption. This condition thus is not invariably maintained in empirical work. But it is consistent with popular parametric models for regression disturbances, including for example GARCH models. It is to be emphasized that the estimator allows for heteroskedasticity of the disturbance conditional on the instruments.

In a system with l equations, the estimator requires obtaining the moving average coefficients of the l-dimensional vector of disturbances. In a single-equation system, then, one fits a univariate MA model, regardless of the size of the parameter or instrument vector. Software to fit univariate MA models of course is widely available. Software to fit multivariate MA models is less widely available, so computational considerations may well call for use of other techniques such as nonparametric ones in systems with many equations.

Section 2 describes the estimator, Section 3 presents simulation results, and Section 4 concludes. For clarity of exposition, the formal econometric theory – not only proofs but precise statement of technical conditions as well – is in an appendix.

2. The new estimator

2.1. Mechanics

I first illustrate this estimator with a simple scalar example, and then define it in the general case. Precise statement of technical conditions may be found in Appendix A. Let $y_t = x_t\beta + u_t$ be a scalar regression model, where u_t is the unobservable disturbance and β is an unknown parameter. For a sample of size T, let β be estimated by instrumental variables using as an instrument a scalar z_t, $\hat{\beta} = (\sum_{t=1}^{T} z_t x_t)^{-1} \sum_{t=1}^{T} z_t y_t$; $z_t = x_t$ if OLS is run. Thus, $Ez_t u_t = 0$ is an orthogonality condition used to estimate β. Let $\{z_t x_t\}$ and $\{z_t u_t\}$ be covariance-stationary. For inference about β, one needs an estimate of the asymptotic variance–covariance matrix $(Ez_t x_t)^{-2} S$, where $S = \sum_{j=-\infty}^{\infty} Ez_t u_t u_{t-j} z_{t-j}$ $= Ez_t^2 u_t^2 + 2\sum_{j=1}^{\infty} Ez_t u_t u_{t-j} z_{t-j}$. (The last equality follows since $z_t u_t$ is a stationary scalar). Estimation of $Ez_t x_t$ is straightforward, since under very mild conditions $T^{-1} \sum_{t=1}^{T} z_t x_t \xrightarrow{p} Ez_t x_t$.

Estimation of S is more problematical, and is the subject of this paper. To illustrate the approach, let u_t follow an MA(1) process, $u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$, and suppose that $\{z_t \varepsilon_t\}$ and $\{z_t \varepsilon_{t-1}\}$ are mean zero and stationary. (To prevent confusion, it may be worth noting the dating convention: if ε_{t-1} is a shock realized in period $t-1$, and is orthogonal to z's that are realized in period $t-2$ and earlier, then z_t must be realized in period $t-2$ or earlier.) Suppose further that ε_t has zero mean conditional on past ε_t's and z_{t+1}'s: $E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, z_{t+1}, z_t, \ldots) = 0$.[2] (In other words, suppose that ε_t is a martingale difference sequence with respect to past ε_t's and z_{t+1}'s.) Since $u_t \sim$ MA(1) and $E(\varepsilon_t | \varepsilon_{t-1}, \ldots, z_{t+1}, \ldots) = 0$, the autocorrelations of $z_t u_t$ are zero for lags greater than 1. Hence, the Wold representation of $z_t u_t$ is an MA(1), and $S = Ez_t^2 u_t^2 + 2Ez_t u_t u_{t-1} z_{t-1}$. We have

$$\begin{aligned} Ez_t^2 u_t^2 &= E[z_t^2 (\varepsilon_t + \theta_1 \varepsilon_{t-1})^2] \\ &= E[z_t^2 (\varepsilon_t^2 + 2\theta_1 \varepsilon_{t-1} + \theta_1^2 \varepsilon_{t-1}^2)] \\ &= E[z_t^2 (\varepsilon_t^2 + \theta_1^2 \varepsilon_{t-1}^2)] \\ &= Ez_t^2 \varepsilon_t^2 + \theta_1^2 Ez_{t+1}^2 \varepsilon_t^2. \end{aligned} \tag{2.1}$$

The equality at the beginning of the third line follows from $E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, z_{t+1}, z_t, \ldots) = 0$, the last equality from stationarity. Similarly, $Ez_t u_t u_{t-1} z_{t-1} = \theta_1 Ez_t z_{t+1} \varepsilon_t^2$. Thus,

$$\begin{aligned} S &= Ez_t^2 u_t^2 + 2Ez_t u_t u_{t-1} z_{t-1} \\ &= Ez_t^2 \varepsilon_t^2 + \theta_1^2 Ez_{t+1}^2 \varepsilon_t^2 + 2(\theta_1 Ez_t z_{t+1} \varepsilon_t^2) \\ &= E[(z_t \varepsilon_t + \theta_1 z_{t+1} \varepsilon_t)^2] \\ &\equiv Ed_{t+1}^2, \qquad d_{t+1} \equiv (z_t + \theta_1 z_{t+1}) \varepsilon_t. \end{aligned} \tag{2.2}$$

One then estimates S by a sample average of a measure of d_t, as illustrated below.

The general case proceeds as follows. There is a regression model

$$f_t(\beta) = u_t, \tag{2.3}$$

where the $l \times 1$ vector f_t depends on data observable at time t and the $(k \times 1)$ unknown parameter vector β, and u_t is an $l \times 1$ unobservable disturbance vector. In a linear model, for example, $f_t(\beta) = y_t - X_t'\beta$, where the $l \times 1$ vector y_t and the $k \times l$ matrix X_t are observable. Let Z_t be a $(q \times l)$ vector of instruments used to estimate β. In the common case in which an $(r \times 1)$ vector of instruments R_t is

[2] It should be noted that with ε_t the univariate innovation in u_t, the zero mean condition $Ez_t \varepsilon_t = 0$ will be violated in some applications (see Hayashi and Sims, 1983).

orthogonal to each of the elements of u_t (e.g., Hansen and Singleton, 1982), $Z_t = R_t \otimes I_l$ and $q = rl$.

To motivate the present study, suppose that a technique such as that in Hansen (1982) is used to estimate β, under Hansen's conditions (although the present technique is not necessarily tied to Hansen's estimation technique and technical conditions). Then $\hat{\beta}$ solves $\min_\beta\{[\sum_{t=1}^T Z_t f_t(\beta)]' W_T [\sum_{t=1}^T Z_t f_t(\beta)]\}$, where W_T is a $(q \times q)$ symmetric positive semidefinite matrix. Let W_T converge in probability to a $(q \times q)$ symmetric positive definite matrix W, let F_t denote the $(k \times l)$ matrix of derivatives of f_t evaluated at the true parameter vector, and let $H = \mathrm{E}Z_t F_t'$. Then $\sqrt{T}(\hat{\beta} - \beta) \overset{A}{\sim} \mathrm{N}(0, V)$, $V \equiv (H'WH)^{-1}H'WSWH(H'WH)^{-1\prime}$, $S \equiv \sum_{j=-\infty}^{\infty} \mathrm{E}Z_t u_t u_{t-j}' Z_{t-j}'$. Thus, here and in other contexts, one needs to estimate S.

Let the disturbance follow an MA process of known order n,

$$u_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_n \varepsilon_{t-n}, \tag{2.4}$$

where ε_t is $l \times 1$, the θ_i's are $l \times l$, and $I + \theta_1 L + \cdots + \theta_n L^n$ is invertible. I assume $\mathrm{E}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, Z_{t+n}, Z_{t+n-1}, \ldots) = 0$, which implies that $Z_t u_t \sim \mathrm{MA}(n)$. Then

$$S = \Gamma_0 + \sum_{j=1}^{n} (\Gamma_j + \Gamma_j'), \quad \Gamma_j \equiv \mathrm{E}(Z_t u_t u_{t-j}' Z_{t-j}').$$

Define the $(q \times 1)$ vector $d_{t+n} = (Z_t + Z_{t+1}\theta_1 + \cdots + Z_{t+n}\theta_n)\varepsilon_t$. It is easily established that $\mathrm{E}d_t d_t' = S$. It is to be emphasized that $\mathrm{E}d_t d_t' = S$ even if u_t is heteroskedastic conditional on Z_t, so that $\mathrm{E}Z_t u_t u_t' Z_t' \neq \mathrm{E}Z_t(\mathrm{E}u_t u_t')Z_t'$.
Let

$$\hat{u}_t = f_t(\hat{\beta}),$$

where $\hat{\beta}$ is a consistent estimate of β. Let $\hat{\theta}_1, \ldots, \hat{\theta}_n$ be consistent estimates of $\theta_1, \ldots, \theta_n$, and let $\hat{\varepsilon}_t$ satisfy $\hat{u}_t = \hat{\varepsilon}_t + \hat{\theta}_1 \hat{\varepsilon}_{t-1} + \cdots + \hat{\theta}_n \hat{\varepsilon}_{t-n}$. In the case where $l = 1$ and u_t is a scalar, the $\hat{\theta}$'s and $\hat{\varepsilon}$'s may be obtained, for example, by nonlinear least squares applied to $\hat{u}_t$, with $\hat{\varepsilon}_t \equiv 0$ for $t \leq 0$; Hannan and Deistler (1988) discuss algorithms applicable for vector MA models. For $t = 1, \ldots, T - n$, define the $(q \times 1)$ vector $\hat{d}_{t+n}$ as

$$\hat{d}_{t+n} = (Z_t + Z_{t+1}\hat{\theta}_1 + \cdots + Z_{t+n}\hat{\theta}_n)\hat{\varepsilon}_t. \tag{2.5}$$

Estimate S as

$$\hat{S} \equiv (T - n)^{-1} \sum_{t=1}^{T-n} \hat{d}_{t+n}\hat{d}_{t+n}'. \tag{2.6}$$

Evidently, $\hat{S}$ is positive semidefinite.

Hodrick (1991) suggests a similar estimator, in the case of a certain linear model in which it is known that $\theta_1 = \cdots = \theta_n = 1$.

2.2. Discussion

If $\hat{\beta}$ and $\hat{\theta}_1, \ldots, \hat{\theta}_n$ are obtained by $T^{1/2}$-consistent estimators (and some other mild conditions hold), this estimator is $T^{1/2}$-consistent for S. (See the Appendix.[3]) By conventional asymptotic efficiency criteria, then, this method dominates the positive semidefinite nonparametric estimators proposed by Andrews (1991), Andrews and Monahan (1992), and Newey and West (1994), which are T^{α}-consistent for some $\alpha < \frac{1}{2}$.

Under the present assumption that $Z_t u_t \sim \text{MA}(n)$, the $T^{1/2}$ rate of convergence is, however, shared by the truncated kernel. This kernel works as follows. Let $\hat{u}_t \equiv f_t(\hat{\beta})$ be the regression residual. For $j = 0, \ldots, n$ let $\hat{\Gamma}_j = T^{-1} \sum_{t=j+1}^{T} Z_t \hat{u}_t \hat{u}'_{t-j} Z'_{t-j}$, with $\Gamma_j \equiv \text{E} Z_t u_t u'_{t-j} Z'_{t-j}$ the corresponding population moment. The truncated kernel estimates S as

$$\tilde{S} = \hat{\Gamma}_0 + (\hat{\Gamma}_1 + \hat{\Gamma}'_1) + \cdots + (\hat{\Gamma}_n + \hat{\Gamma}'_n). \tag{2.7}$$

As is well-known, $\tilde{S}$ need not be positive semidefinite, a point I return to below.

To get a feel for how $\hat{S}$ compares to $\tilde{S}$, I computed the asymptotic variances of $\hat{S}$ and $\tilde{S}$ in a scalar linear model in which the only regressor is the constant term. In this model, which is described in detail in the notes to Table 1, $f_t(\beta) = y_t - \beta$, $l = n = k = q = 1$, and $Z_t \equiv 1$. Appendix B outlines the algebra used to derive the asymptotic variances.[4]

It may be seen that the new estimator – which I call the MA-l estimator – is dramatically more efficient when θ_1 is near -1. This is essentially the following well-known result from Box–Jenkins analysis: Suppose that one wants to estimate θ in the MA(1) model $x_t = v_t + \theta v_{t-1}$, where $v_t \sim$ i.i.d. Then, if θ is near -1, nonlinear least squares (NLLS) is dramatically more efficient than is the simple estimator that relies on the one-to-one mapping between the MA coefficient and the first autocorrelation (e.g., Brockwell and Davis, 1991, p. 254). The textbook intuition for this result is that NLLS exploits information in the sample autocorrelations beyond the first (Fuller, 1976, p. 343), intuition that seems to carry over here as well.

Note that the proposed estimator involves estimation of the moving average coefficients of $\hat{u}_t$ and not $Z_t \hat{u}_t$. In the general, and empirically plausible, case in

---

[3] The appendix also shows that $\hat{S}$ is consistent as long as $\hat{\beta}$ and $\hat{\theta}_1, \ldots, \hat{\theta}_n$ are consistent. The implication is that one will be able to obtain the $\hat{\theta}_i$'s by inverting the estimates of the autoregressive representation of $\hat{u}_t$, provided one lets the order of the autoregression increase at an appropriate rate. Such a procedure might be computationally convenient when the number of equations l is large. In this paper I do not, however, attempt to establish what this rate might be.

[4] The table assumes that ε_t is i.i.d. normal. Suppose more generally that ε_t is i.i.d. with $\text{E}\varepsilon_t^3 = 0$, $\text{E}\varepsilon_t^4 = \kappa$. Then the ratios reported in Table 1 will continue to be greater than one, but will shrink if $\kappa > 3$, grow if $\kappa < 3$.

Table 1
Inefficiency of the truncated estimator relative to the MA-l estimator

θ_1					
−0.9	−0.6	−0.3	0.3	0.6	0.9
840.08	8.25	1.47	1.16	1.52	2.04

This table presents the ratio of the asymptotic variance of the truncated estimator (Eq. (2.7)) to that of the MA-l estimator (Eq. (2.6)). That the entries are greater than one indicates that the truncated estimator is less efficient asymptotically. The calculations assume OLS estimation of a scalar model with an MA(1) disturbance, whose only regressor is the constant term: $y_t = \beta + u_t$, $u_t = \varepsilon_t + \theta_1\varepsilon_{t-1}$, where ε_t is an i.i.d. normal variable. The reported ratios are invariant to the scale of ε_t. In the MA-l model, it is assumed that $\hat{\theta}_1$ is obtained by nonlinear least squares or an asymptotically equivalent procedure. The object of interest is $S \equiv Eu_t^2 + 2Eu_tu_{t-1}$.

which Z_t is stochastic, a positive semidefinite estimator at least as efficient as the one I propose results from fitting an MA(n) to ($q \times 1$) vector $Z_t\hat{u}_t$ and estimating S as the usual quadratic form in the variance–covariance matrix of the innovation to $Z_t\hat{u}_t$ (see, e.g., Fuller, 1976, p. 166, for the population formula). Why then do I not propose applying a multivariate analogue of NLLS to $Z_t\hat{u}_t$? The reason is computational. Since $q \geqslant l$ fitting an MA(n) to the l-vector $\hat{u}_t$ obviously is computationally simpler than fitting an MA(n) to the q-vector $Z_t\hat{u}_t$, and in practice it is often the case that $q \gg l$. In Eichenbaum et al. (1988), for example, $q = 14$ and $l = 2$.

It should be noted that the circumstances under which the new estimator is relatively efficient are precisely those under which the truncated estimator tends to yield an estimate that is not p.s.d. This is indicated by the Monte Carlo simulations reported in the next section, and is suggested by some algebra given in a footnote.[5]

In any case, one should expect the asymptotic comparison in Table 1 to provide at best a rough guide to actual performance. One obvious reason is that the example is so simple and stylized. When there are multiple, stochastic regressors, efficiency of the MA-l estimator will of course be affected not only by serial correlation properties of the disturbance but by those of the instruments as well. In general, then, there will not be a simple scalar that indexes relative efficiency of the MA-l estimator for any and all hypothesis tests. A second reason

[5] Suppose that $u_t = \varepsilon_t + \theta_1\varepsilon_{t-1}$ is a scalar ($l = 1$) MA(1). For $j = 0, 1$ let $\gamma_j = Eu_tu_{t-j}$ and $\hat{\gamma}_j \equiv T^{-1}\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}$ be the population and sample autocovariance of u_t. Assume that the first element of Z_t is the constant term $\Rightarrow \tilde{S}(1,1) = \hat{\gamma}_0 + 2\hat{\gamma}_1$. Then $\tilde{S}(1,1) < 0 \Leftrightarrow \hat{\gamma}_0 + 2\hat{\gamma}_1 < 0$; given that $\gamma_0 + 2\gamma_1 \to 0$ as $\theta_1 \to -1$, it is not unreasonable that when θ_1 is nearer -1 sampling error will more likely cause $\tilde{S}(1,1) < 0$. The same logic applies to the other diagonal elements of $\tilde{S}$, at least when u_t is conditionally homoskedastic and the relevant element of Z_t is highly positively autocorrelated.

is that the precision of estimation of the asymptotic variance–covariance matrix is affected by the precision of estimation of the expectation of cross-products of Z_t and the gradient of f_t (that is, of EZ_tX_t' in the linear model $y_t = X_t'\beta + u_t$ – see the discussion of (2.3)); given that this estimate typically will also converge at rate $\sqrt{T}$, there is no a priori reason to expect performance to be dominated by the precision of estimation of S.

As we shall see, the simulations nonetheless indicate a broad connection between serial correlation properties of the disturbance and of cross-products of instruments and disturbances on the one hand, and performance of the MA-l estimator on the other.

3. Monte Carlo results

3.1. Description of data generating processes and estimators

The data generating processes (DGPs) and hypothesis tests are very similar to some reported in Andrews and Monahan (1992). The experimental design was chosen in large part because the simplicity of the Andrews and Monahan (1992) DGPs allowed me to cleanly extend their analysis of DGPs with positive autocorrelation of cross-products of instruments and disturbances to ones with negative autocorrelation.

As in Andrews and Monahan (1992): all experiments involve the linear regression $y_t = \beta_1 + \beta_2 z_{2t} + \beta_3 z_{3t} + \beta_4 z_{4t} + \beta_5 z_{5t} + u_t \equiv Z_t'\beta + u_t$, $t = 1, \ldots, T$, $T = 128$; $E(u_t|Z_t) = 0$ and least squares is the estimator $\Rightarrow \hat{\beta} = (\sum_{t=1}^{T} Z_t Z_t')^{-1} \sum_{t=1}^{T} Z_t y_t$; without loss of generality, β is set equal to zero; the hypothesis of interest is H_0: $\beta_2 = 0$. Let $\Gamma_j \equiv EZ_t u_t u_{t-j} Z_{t-j}'$. In all experiments, $Z_t u_t \sim MA(n)$ for $n = 1$ or $n = 2 \Rightarrow$

$$S = \Gamma_0 + \Gamma_1 + \Gamma_1' \qquad \text{(MA(1) specifications)},$$

$$S = \Gamma_0 + \Gamma_1 + \Gamma_1' + \Gamma_2 + \Gamma_2' \qquad \text{(MA(2) specifications)}. \tag{3.1}$$

Let $\hat{V}$ be an estimate of the asymptotic variance–covariance matrix of $\hat{\beta}$,

$$\hat{V} = \left(T^{-1}\sum_{t=1}^{T} Z_t Z_t'\right)^{-1} (\text{estimate of } S) \left(T^{-1}\sum_{t=1}^{T} Z_t Z_t'\right)^{-1}. \tag{3.2}$$

The relevant test statistic is $T\hat{\beta}_2^2/\hat{V}(2,2) \overset{A}{\sim} \chi^2(1)$. In all experiments, the number of replications was 1000.

The regressors (= the instruments) follow independent AR(1) processes with common parameter ϕ: for $i = 2, \ldots, 5$, $z_{it} = \phi z_{it-1} + e_{it}$. Two values of ϕ were used: $\phi = 0.5$ and $\phi = 0.9$. An autocorrelation of 0.5 is approximately that of growth rates of some macroeconomic variables, such as GDP; that of 0.9 is characteristic of many undifferenced macroeconomic variables. For each value of ϕ, the variance of the i.i.d. normal variable e_{it} was chosen so that $Ez_t^2 = 1$.

In the *homoskedastic* models, $u_t = \varepsilon_t + \theta_1\varepsilon_{t-1}$ or $u_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$ where ε_t is i.i.d. normal and $E\varepsilon_t e_{is} = 0$ for all t, i, s. For various values of θ_1 and θ_2, the variance of ε_t was chosen so that $Eu_t^2 = 1$. In the MA(1) model, θ_1 ranged over the seven values -0.9, -0.6, $-0.3, 0, 0.3, 0.6$, and 0.9. The values towards the lower end perhaps capture some important characteristics of applications in which the truncated estimator of S fails to be p.s.d., because, as we shall see, these tend to cause such a failure. The smaller positive values might arise from time aggregation. The larger positive values are for comparison. In the MA(2) model, three sets of parameters were used: $\theta_1 = -1.3$, $\theta_2 = 0.5$; $\theta_1 = -1.0$, $\theta_2 = 0.2$; $\theta_1 = 0.67$, $\theta_2 = 0.33$. The first two sets come from estimates from inventory data in West and Wilcox (1996); the last is suggested by Andrews and Monahan (1992). Thus the total number of homoskedastic MA(1) models is 14 (= 2 values of the regressors's autoregressive parameter ϕ times 7 values of the disturbance's moving average coefficient θ_1), the total number of MA(2) models was 6 (= 2×3). When $\theta_1 = \theta_2 = 0$, $u_t \sim$ i.i.d.; to prevent possible misunderstanding, I note that if this fact were known, one would not use an autocorrelation consistent estimator.

The *heteroskedastic* models, which were suggested by a similar model in Andrews and Monahan (1992), are identical to the homoskedastic models except that

$$u_t = (1/\sqrt{3})[z_{2t}^2\varepsilon_t + \theta_1(z_{2t-1}^2\varepsilon_{t-1})], \tag{3.3a}$$

$$u_t = (1/\sqrt{3})[z_{2t}^2\varepsilon_t + \theta_1(z_{2t-1}^2\varepsilon_{t-1}) + \theta_2(z_{2t-2}^2\varepsilon_{t-2})]. \tag{3.3b}$$

(The factor of $1/\sqrt{3}$ keeps the variance of u_t at unity.)

For future reference, I note the following about the serial correlation properties of u_t and $Z_t u_t$, all of which may be established with a little bit of algebra. First, for given θ_i's the autocorrelations of u_t are identical for the homoskedastic and heteroskedastic models. Second, the signs of the first-order autocorrelations of u_t and $z_{it}u_t$ ($i = 2,3,4,5$) are the same as those of θ_1, for both MA(1) and MA(2) models; all the MA(2) models happen to have (small) positive second-order autocorrelations. Third, the autocorrelations of $z_{it}u_t$ are smaller in absolute value for $\phi = 0.5$ than for $\phi = 0.9$.

Four estimators are considered. The new estimator is implemented by applying nonlinear least squares to the least squares residuals, with presample values of ε_t set to zero.[6] A second estimator was the truncated. I checked whether

[6] I rarely encountered numerical problems using this estimator. Of 40,000 sets of estimates, only 21 did not converge using a canned nonlinear search algorithm (the OPTMUM procedure of GAUSS). Of the 21 cases of nonconvergence, eight occurred for the heteroskedastic MA(1) model with $\theta_1 = 0.9$; for no other parameter configuration did nonconvergence occur more than three times. Rather than attempt to tune the search for these 21 data sets, I omitted them altogether from the size calculations reported in Tables 2 and 3: with so few cases of nonconvergence per parameter configuration, the character of the results would not change in the slightest, no matter what test statistics would result from playing with the search algorithm until it converged for these data sets.

the estimate (2.7) was positive definite. If so, I used (2.7) in computing the variance–covariance matrix (3.2); if not, I computed (3.2) setting the estimate of S to $\hat{\Gamma}_0$ (i.e., I ignored the autocorrelation in u_t and Z_t). A similar procedure was used in the simulations reported in the working paper version of Cumby and Huizanga (1992).

The third and fourth estimators are the prewhitened QS estimator suggested in Andrews and Monahan (1992, Sec. 3) and the prewhitened Bartlett estimator suggested in Newey and West (1994, Sec. 2). For details on these estimators, see the original papers; here, I limit myself to a brief outline. These two estimators: (1) Prewhiten $Z_t\hat{u}_t$ by fitting a vector autoregression of order 1 to $Z_t\hat{u}_t$. Let $h_t^{\dagger}$ denote the (5×1) vector of residuals to this vector autoregression. (2) Estimate the spectrum of $h_t^{\dagger}$ by taking weighted sums of the sample autocovariances of this residual. After defining (in the notation of Andrews and Monahan, 1992; Newey and West, 1994) $w_1 = 0$, $w_2 = w_3 = w_4 = w_5 = 1$, the weights are determined by a procedure that is asymptotically optimal in a certain precise sense. The two estimators differ in the weighting scheme used. Call the resulting estimate $S^{\dagger}$. (3) Use $S^{\dagger}$ and the matrix of autoregressive coefficients estimated in step 1 to estimate S.

3.3. Simulation results

Table 2 presents sizes of nominal 1, 5, and 10 percent tests for the homoskedastic models. First consider panel A, in which $\phi = 0.5$ so the regressors are mildly positively autocorrelated. When $\theta_1 \geq -0.3$, so that the autocorrelations of the disturbance, and of cross-products of instruments and disturbance, are positive or mildly negative, all the estimators display a tendency to overreject. For the QS and Bartlett estimators, such a tendency also characterized the simulations in Andrews and Monahan (1992) and Newey and West (1994). With the possible exception of the truncated, the estimators seem to perform better for $\theta_1 = -0.9$ and $\theta_1 = -0.6$. Overall, the estimators seem to perform comparably, with the possible exception of the truncated estimator when $\theta_1 = -0.9$.

Panel B considers the MA(1) model when the regressors are strongly positively autocorrelated. When $\theta_1 \geq 0$, the estimators show a mild tendency to overreject; for such values of θ_1, the Bartlett performs worse than the other three, which seem about comparable. When $\theta_1 = -0.9$ or $\theta_1 = -0.6$, so that the disturbance, and cross-products of instruments and disturbances, are strongly negatively autocorrelated, the Bartlett overrejects and the MA-l estimator is relatively accurately sized; the QS and truncated estimators substantially underreject. With $\theta_1 = -0.9$, for example, the test statistic generated by QS was greater than 3.84 (the 5 percent value for a $\chi^2(1)$) in only 7 of the 1000 replications (the ideal is 50).

A comparison of panels A and B suggests that for given θ_1, the MA-l estimator is not sensitive to ϕ, the autocorrelation coefficient of the instruments;

for given ϕ, the estimator seems to perform a little better when $\theta_1 = -0.9$. The other estimators seem sensitive to both ϕ and θ_1, with the QS and truncated estimators tending to underreject when the product $\phi\theta_1$ is near -1 – that is, when cross-products of instruments and disturbances are sharply negatively autocorrelated.

Panel C tells a similar story for the MA(2) specifications. The performance of the MA-l estimator seems insensitive to ϕ, but for given ϕ is better when $\theta_1 < 0$. When $\theta_1 < 0$, the QS and truncated estimators underreject mildly for $\phi = 0.5$ (columns (1) and (2) of panel C), substantially for $\phi = 0.9$ (columns (4) and (5)). All four estimators tend to overreject when both ϕ and θ_1 are positive.

I also experimented with an MA(1) DGP in which the instruments were strongly negatively autocorrelated ($\phi = -0.9$) and the disturbance was strongly positively autocorrelated ($\theta_1 = 0.9$). Such strong negative autocorrelation of the instrument is not common in the economic data that I am familiar with. I used this DGP nonetheless to see whether the key characteristic that leads to relatively good performance of the MA-l estimator is strong negative autocorrelation of cross-products of instruments and disturbance. And, indeed, the MA-l estimator seemed insensitive to this change in parameters, while the QS and truncated estimators tended to underreject. Rejection rates for nominal 5 percent tests, for example, were: Bartlett, 5.2; QS, 0.4; truncated, 0.5; MA-l, 3.6.

A broadly similar story is told in the heteroskedastic simulations reported in Table 3. While the performance of the MA-l estimator is somewhat worse here than in the homoskedastic simulations, so, too, is the performance of the other estimators. And the MA-l estimator continues to perform relatively well when the autocorrelation of the disturbance is rather negative (θ_1 near -1 for MA(1) models, $\theta_1 < 0$ for MA(2) models): in all three panels, the QS and truncated estimators underreject when cross-products of instruments and disturbances have sharp negative autocorrelation. See columns (1) and (2) in panels A and B, and columns (1), (2), (4), and (5) in panel C. All the estimators show a tendency to overreject when both ϕ and θ_1 are positive.

I summarize the simulations in Tables 2 and 3 and the asymptotic calculations in Table 1 as indicating that the MA-l estimator tends to perform relatively well when cross-products of instruments and disturbances are strongly negatively autocorrelated, although the magnitude of autocorrelation is by no means a sufficient statistic for performance.

These points are illustrated in Fig. 1, which plots the actual size of tests of nominal size 0 to 25, for selected experiments. A comparison of graphs (1) and (2) in each panel illustrates the insensitivity of the MA-l estimator to autocorrelation of the instrument, and the tendency of QS to underreject when cross-products of instruments and disturbances display sharp negative autocorrelation. Graphs (1) and (3) of both panels illustrate that the MA-l and QS estimators behave quite similarly in many of the simulations. Especially in

Table 2
Size of nominal 1, 5, and 10 percent tests, homoskedastic models

A. MA(1), $\phi = 0.5$

	(1) $\theta_1 = -0.9$			(2) $\theta_1 = -0.6$			(3) $\theta_1 = -0.3$			(4) $\theta_1 = 0.0$			(5) $\theta_1 = 0.3$			(6) $\theta_1 = 0.6$			(7) $\theta_1 = 0.9$		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Bartlett	1.6	6.1	10.2	2.0	6.8	11.9	1.4	6.8	12.9	3.2	9.1	16.7	2.5	8.3	13.7	3.3	8.0	15.5	2.3	8.0	12.9
QS	1.0	4.5	7.6	1.1	5.0	9.5	1.1	5.8	10.7	2.0	8.1	14.3	1.7	6.8	12.1	2.4	7.4	13.2	1.8	6.7	11.4
Truncated	0.8	2.6	6.2	0.6	4.6	7.8	1.4	6.5	11.7	2.4	8.8	14.3	2.4	7.3	12.7	2.6	7.8	14.3	2.0	7.4	13.0
MA-*l*	1.6	4.8	9.3	1.4	6.2	11.5	1.4	6.2	10.9	2.0	7.1	13.6	2.0	6.8	12.0	2.0	7.8	11.9	1.8	6.3	11.4

B. MA(1), $\phi = 0.9$

	(1) $\theta_1 = -0.9$			(2) $\theta_1 = -0.6$			(3) $\theta_1 = -0.3$			(4) $\theta_1 = 0.0$			(5) $\theta_1 = 0.3$			(6) $\theta_1 = 0.6$			(7) $\theta_1 = 0.9$		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Bartlett	2.6	7.0	11.8	3.4	8.6	15.2	4.3	10.0	16.0	4.9	12.0	18.9	3.7	9.6	15.5	4.6	11.3	18.0	5.0	11.0	18.2
QS	0.0	0.7	2.1	0.1	1.0	3.9	1.1	4.1	9.7	1.5	7.9	14.7	1.7	5.5	11.4	1.4	4.7	10.4	1.4	4.4	9.6
Truncated	0.0	0.2	0.4	0.1	0.8	1.6	1.6	5.8	10.3	2.4	9.3	16.5	1.8	7.0	12.8	1.9	8.0	13.7	2.3	8.5	14.5
MA-*l*	0.6	4.2	9.2	1.1	6.5	13.2	2.4	7.8	12.7	1.9	8.7	15.2	1.2	5.9	11.1	1.3	5.6	11.5	1.5	6.9	11.8

C. MA(2), $\phi = 0.5$ and $\phi = 0.9$

	(1) ϕ, θ_1, θ_2			(2) ϕ, θ_1, θ_2			(3) ϕ, θ_1, θ_2			(4) ϕ, θ_1, θ_2			(5) ϕ, θ_1, θ_2			(6) ϕ, θ_1, θ_2		
	0.5, −1.3, 0.5			0.5, −1.0, 0.2			0.5, 0.67, 0.33			0.9, −1.3, 0.5			0.9, −1.0, 0.2			0.9, 0.67, 0.33		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Bartlett	1.8	6.6	12.1	1.2	4.6	8.4	3.1	9.4	15.5	2.2	6.1	11.4	1.6	5.4	10.3	5.4	11.9	17.2
QS	0.7	3.8	7.4	0.6	3.8	6.3	2.4	7.3	13.9	0.2	1.2	2.2	0.0	0.9	2.1	1.9	7.5	11.8
Truncated	0.6	2.5	4.9	0.5	2.5	5.1	2.9	9.0	15.4	0.0	0.1	0.2	0.0	0.1	0.5	3.7	10.9	16.6
MA-l	0.9	4.8	9.5	0.7	3.4	8.3	1.7	7.1	12.9	0.8	4.3	9.1	0.6	3.4	8.2	2.4	7.8	14.1

For each of 1000 data sets, the following was done: (1) A sample of size 128 was generated for a linear model with five regressors and an i.i.d. normal disturbance. As discussed in Section 3.1, the parameter ϕ indexes the degree of autocorrelation in the regressors; θ_1 and (in panel C) θ_2 are the parameters of the MA(1) (panels A and B) or MA(2) (panel C) disturbance. (2) OLS was used to estimate the regression vector. (3) A test was performed of the hypothesis that a specific regression coefficient is equal to its population value. The relevant variance–covariance matrix was computed according to (3.2), using either the Bartlett, QS, truncated, or MA-l estimator of S (defined in (3.1)). See the text for additional details.

Asymptotically, each test statistic is $\chi^2(1)$. For the indicated estimators of S, the columns labelled 1, 5, and 10 report the percentage of the 1000 test statistics greater than 6.64, 3.84, and 2.71.

Table 3
Size of nominal 1, 5, and 10 percent tests, heteroskedastic models

A. MA(1), $\phi = 0.5$

	(1) $\theta_1 = -0.9$			(2) $\theta_1 = -0.6$			(3) $\theta_1 = -0.3$			(4) $\theta_1 = 0.0$			(5) $\theta_1 = 0.3$			(6) $\theta_1 = 0.6$			(7) $\theta_1 = 0.9$		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Bartlett	1.3	6.3	12.1	2.5	7.1	13.7	3.7	11.4	17.5	3.1	9.8	15.7	3.4	11.0	16.7	4.6	11.6	19.5	3.4	10.4	17.2
QS	0.3	3.2	6.2	0.9	4.6	9.4	2.1	8.0	14.8	2.1	8.1	13.2	2.9	9.2	15.1	3.0	9.9	17.6	1.9	7.8	14.7
Truncated	0.7	1.8	3.5	1.0	3.5	7.3	1.9	8.4	14.8	2.3	8.7	14.8	3.1	9.8	15.2	3.4	10.9	18.5	2.5	9.9	17.3
MA-l	0.4	4.1	8.3	1.8	5.8	12.3	2.2	8.7	17.5	2.7	8.4	13.6	2.5	8.7	14.5	3.2	9.5	16.7	2.5	10.3	16.6

B. MA(1), $\phi = 0.9$

	(1) $\theta_1 = -0.9$			(2) $\theta_1 = -0.6$			(3) $\theta_1 = -0.3$			(4) $\theta_1 = 0.0$			(5) $\theta_1 = 0.3$			(6) $\theta_1 = 0.6$			(7) $\theta_1 = 0.9$		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Bartlett	3.7	8.8	13.2	3.9	10.8	17.9	6.1	13.4	19.0	8.7	16.8	24.3	7.3	16.9	24.9	7.9	16.2	24.0	10.2	17.3	24.0
QS	0.2	0.4	1.1	0.4	1.5	3.7	1.1	5.1	9.3	3.0	9.3	15.6	2.6	10.0	16.0	2.3	6.8	12.4	3.4	7.9	14.3
Truncated	0.0	0.1	0.3	0.2	0.5	1.2	0.9	2.9	6.5	3.4	10.2	17.1	3.7	11.6	17.7	2.9	8.5	16.8	3.5	10.3	17.7
MA-l	1.0	3.1	7.2	2.3	8.7	15.7	2.9	10.1	14.9	3.6	10.6	17.3	2.7	9.8	16.5	2.4	7.4	14.1	2.5	9.3	16.2

C. MA(2), $\phi = 0.5$ and $\phi = 0.9$

	(1) ϕ, θ_1, θ_2			(2) ϕ, θ_1, θ_2			(3) ϕ, θ_1, θ_2			(4) ϕ, θ_1, θ_2			(5) ϕ, θ_1, θ_2			(6) ϕ, θ_1, θ_2		
	0.5, −1.3, 0.5			0.5, −1.0, 0.2			0.5, 0.67, 0.33			0.9, −1.3, 0.5			0.9, −1.0, 0.2			0.9, 0.67, 0.33		
	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
Bartlett	2.4	8.0	13.1	2.0	7.9	13.8	4.4	12.5	17.7	3.7	7.7	13.2	3.2	8.4	13.4	9.6	18.9	26.8
QS	0.4	3.4	7.5	0.6	4.1	8.2	3.3	10.1	15.8	0.1	1.3	2.6	0.3	0.8	1.8	4.2	9.9	16.8
Truncated	0.6	2.3	4.8	0.7	3.1	5.4	4.3	12.2	18.2	0.0	0.2	0.2	0.0	0.0	0.2	6.5	14.9	22.8
MA-*l*	0.7	6.4	12.6	5.0	13.9	21.4	5.7	14.4	20.4	3.3	10.2	16.8	5.3	14.0	22.2	3.6	11.1	18.6

See notes to Table 2.

The data generating processes underlying these results differ from those in Table 1 only in that the disturbance is heteroskedastic conditional on the instruments. See (3.3).

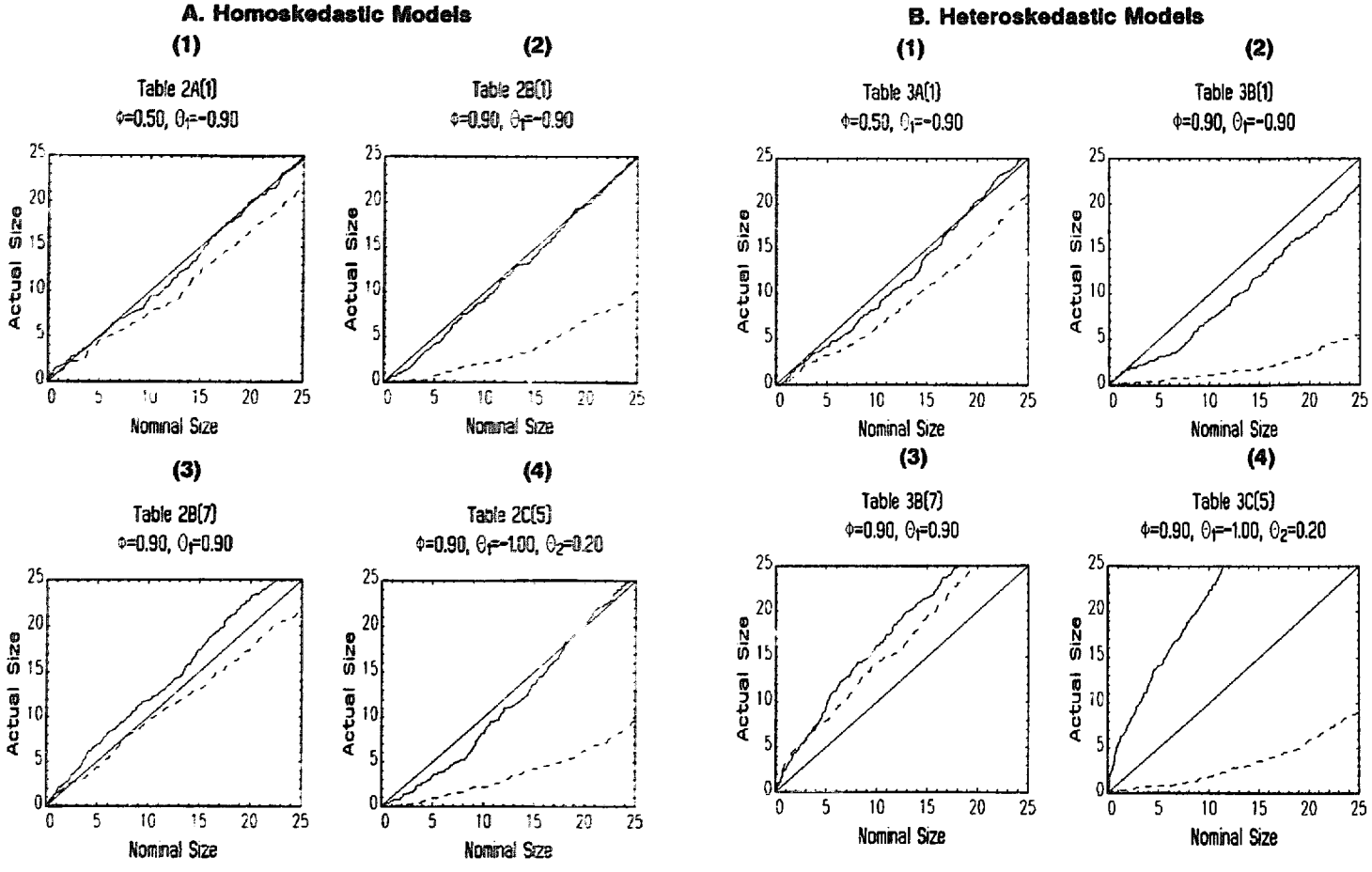

Fig. 1. Plots of actual versus nominal size, MA-*l* and QS, selected experiments.

The dashed line is the QS, one solid line the MA-*l* estimator, the other solid line the 45 degree line. Panel A, graph (1), for example, plots the actual size of tests whose 1, 5, and 10 percent actual sizes are also reported in Table 2, panel A, column 1.

The plot for an ideal estimator lies on the 45 degree line. When the plot is below this line, as is the case for all nominal sizes and for both estimators in panel B, graph (2), the estimator does not reject often enough. When the plot is above this line, as is the case for all nominal sizes and both estimators in panel B, graph (3), it rejects too often.

graph (4), a comparison of panel A to panel B illustrates that the estimators perform worse in the heteroskedastic simulations.

In those simulations in which the QS and truncated estimators performed poorly, use of formula (2.7) tended to generate truncated estimates that were not p.s.d. When $\theta_1 = -0.9$ and $\phi = 0.9$ (column (1) of panel B in Tables 2 and 3), for example, (2.7) was not p.s.d. in an astonishing 93.2 (Table 2) and 88.7 (Table 3) percent of the simulations. (See Appendix C.) Given that in such cases I set the estimate of S to $\hat{\Gamma}_0$, the tendency to underreject is unsurprising: in such DGPs $V(2, 2) < \Gamma_0(2, 2)$, so the estimator will underreject if $\Gamma_0(2, 2)$ is near $\Gamma_0(2, 2)$.

To get a feel for why the QS estimator also underrejected in these DGPs, I calculated the bias across the 1000 repetitions in the estimate of $S(2, 2)$. (Recall that in population, $V(2, 2) = S(2, 2)$.) My thought was that underrejection might be associated with estimates of $S(2, 2)$ that were too large, i.e., that QS was biased upwards in these DGPs. And this was indeed the case. For example, in the homoskedastic MA(1) process with $\phi = 0.9$, $\theta = -0.9$, the average differences between the estimated and population values of $S(2, 2)$, expressed as a fraction of the population value of $S(2, 2)$, were 3.86 for truncated, -0.20 for Bartlett, 0.96 for QS, and -0.35 for MA-l

On the other hand, in all but the DGPs with sharp negative correlation, QS tended to be biased downwards, as it was in Andrews and Monahan (1992). So, too were the other estimators, as is consistent with the general tendency to overreject that is evident in Tables 2 and 3.[7]

4. Conclusions

This paper has proposed and evaluated a positive semidefinite estimator of a heteroskedasticity- and autocorrelation-consistent covariance matrix. A requirement is that the regression disturbance follow a moving average (MA) process of known order. In a system of l equations, this 'MA-l' estimator entails estimation of the moving average coefficients of an l-dimensional vector; in a single-equation system, for example, one fits a univariate MA model, regardless of the size of the parameter or instrument vector. Simulations indicate that the estimator performs better than the nonparametric ones now in common use when cross-products of instruments and disturbances are sharply negatively autocorrelated, comparably or slightly worse otherwise.

---

[7] To prevent misunderstanding, let me note that many factors determine the small sample performance of the estimators. An downward (upward) bias in the estimate of $S(2, 2)$ may not and indeed did not always translate into overrejection (underrejection) at a given nominal significance level, let alone at all significance levels. For example, in the homoskedastic MA(1) process with $\phi = 0.9$, $\theta = 0.9$, the biases as a fraction of $S(2, 2)$ were -0.51 for truncated, -0.53 for Bartlett, -0.23 for QS, and -0.47 for MA-l. Thus, QS was biased downwards, but, as indicated in Table 2, still underrejected (slightly) at the 0.05 and 0.10 significance levels.

One priority for future work is to allow for disturbances whose moving average representation is of unknown, and possibly infinite, order. Such an estimator might be implemented in, say, a single-equation model as follows. First estimate a univariate autoregression. Then obtain the first n moving average coefficients from the autoregressive estimates in the usual way (e.g., Fuller, 1976, p. 74). Use these coefficients as described in Eqs. (2.5) and (2.6) above, with the autoregression's residuals used for $\hat{\varepsilon}_t$.[8] If the order of the ARMA process of the disturbance is unknown, one needs to let the number of coefficients in the autoregression and the number of moving average coefficients increase as a suitable function of sample size. The theoretical challenge is to determine this suitable function.

A second priority is to develop refined asymptotics that better characterize the finite sample distribution of the present estimator.

Appendix A

This appendix formally proves the consistency results stated in Section 2. To do so, it is helpful to denote the true value of the regression vector as β^* rather than β, the true value of the matrices of moving average parameters as θ_i^* rather than θ_i. So: Let Z_t be $q \times l$, let u_t and ε_t be $l \times 1$, with $u_t = \varepsilon_t + \theta_1^* \varepsilon_{t-1} + \cdots + \theta_n^* \varepsilon_{t-n}$; for $\varsigma \in \mathbb{C}$, $|I + \theta_1^* \varsigma + \cdots + \theta_n^* \varsigma^n| = 0 \Rightarrow |\varsigma| > 1$. Let $S = \sum_{j=-n}^{n} \mathrm{E} Z_t u_t u_{t-j}' Z_{t-j}'$. Let $f_t(\beta^*) = u_t$, where β^* is $(k \times 1)$.

Assumption 1. $\mathrm{E}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, Z_{t+n}, Z_{t+n-1}, \ldots) = 0$, and $\{(Z_t \varepsilon_t)', \ldots, (Z_{t+n} \varepsilon_t)'\}'$ is covariance-stationary and ergodic.

Assumption 2. In some open neighborhood around β^*, and with probability 1, $f_t(\beta)$ is measurable and continuously differentiable in β.

For notational simplicity, assume that none of the elements of $\theta_1^*, \ldots, \theta_n^*$ are known. (In some applications it may be known that some of the elements of the θ_i^*'s are, say, zero; in such cases, the argument presented here is easily adapted.) Let $\alpha^* = (\beta^{*\prime}, \mathrm{vec}(\theta_1^*)', \ldots, \mathrm{vec}(\theta_n^*)')'$; let

$$r \equiv (k + nl^2)$$

be the dimension of α^*. Methods for estimation of MA models vary in treatment of presample values of the unobservable disturbance. For concreteness, I assume that these are zero, both in the data and in the estimation method: $\varepsilon_0 = \varepsilon_{-1} = \cdots = \varepsilon_{-n+1} = \hat{\varepsilon}_0 = \hat{\varepsilon}_{-1} = \cdots = \hat{\varepsilon}_{-n+1} = 0$. Accordingly, for an estimate $\hat{\alpha}$ of α^* obtained from a sample of size T, define $\hat{\varepsilon}_t = \varepsilon_t(\hat{\alpha})$ by solving for

[8] Cumby et al. (1983) and Eichenbaum et al. (1988) suggest similar procedures, but require that an autoregression be estimated whose dimension is the number of orthogonality conditions rather than the number of equations.

the first $t-1$ autoregressive weights obtained by inverting the MA(n) lag polynomial $I + \hat{\theta}_1 L + \cdots + \hat{\theta}_n L^n$, $\varepsilon_t(\hat{\alpha}) = \sum_{j=0}^{t-1} \hat{\psi}_j f_{t-j}(\hat{\beta})$, where the autoregressive $\hat{\psi}_j$'s are defined by the usual recursion (e.g., Fuller, 1976, p. 74): $\hat{\psi}_0 = I$, $\hat{\psi}_1 = -\hat{\theta}_1$, $\hat{\psi}_2 = -\hat{\theta}_1\hat{\psi}_1 - \hat{\theta}_2$, ... For given $\alpha \in R^r$, define $\varepsilon_t(\alpha)$ analogously, and define $d_{t+n}(\alpha)$: $R^r \to R^q$ as $(Z_t + Z_{t+1}\theta_1 + \cdots + Z_{t+n}\theta_n)\varepsilon_t(\alpha)$. It is understood that 'd_t' means '$d_t(\alpha^*)$', '$\hat{d}_t$' means '$d_t(\hat{\alpha})$'. Let $\hat{S} = (T-n)^{-1}\sum_{t=1}^{T-n} \hat{d}_{t+n}\hat{d}'_{t+n}$.

By Assumption 2, there is a neighborhood N around α^* in which $d_t(\alpha)$ is continuously differentiable; for $\alpha \in N$, let $D_t(\alpha) \equiv \partial d_t(\alpha)/\partial\alpha$ denote the $(q \times r)$ matrix of partial derivatives of $d_t(\alpha)$. For any matrix $A = [a_{ij}]$, let $|A| = \max_{i,j}|a_{ij}|$.

Assumption 3. There exists a constant c and a measurable random variable m_t such that for all t, $\sup_{\alpha\in N}|d_t(\alpha)| < m_t$, $\sup_{\alpha\in N}|D_t(\alpha)| < m_t$, $Em_t^2 < c < \infty$.

Assumption 4. $T^{1/2}[(T-n)^{-1}\sum_{t=1}^{T-n} d_{t+n}d'_{t+n} - S] = O_p(1)$.

Proposition 1. *Under Assumptions 1–3, if* $\hat{\alpha} \xrightarrow{p} \alpha^*$, $\hat{S} \xrightarrow{p} S$.

Proposition 2. *Under Assumptions 1–4, if* $T^{1/2}|\hat{\alpha} - \alpha^*| = O_p(1)$, *then* $T^{1/2}(\hat{S} - S) = O_p(1)$.

Proof of Propositions 1 and *2.* Set $q = 1$ for notational simplicity. A mean value expansion of $(T-n)^{-1}\sum_{t=1}^{T-n}\hat{d}_{t+n}^2 \equiv (T-n)^{-1}\sum\hat{d}_{t+n}^2$ around $(T-n)^{-1}\sum d_{t+n}^2$ yields

$$\hat{S} - S = (T-n)^{-1}\sum d_{t+n}^2 - S + 2B_T,$$

$$B_T = (T-n)^{-1}\{\sum d_{t+n}(\tilde{\alpha})D_{t+n}(\tilde{\alpha})(\hat{\alpha} - \alpha^*)\},$$

where $\tilde{\alpha}$ is on the line between $\hat{\alpha}$ and α^*. By the ergodic theorem, $(T-n)^{-1}\sum d_{t+n}^2 \xrightarrow{p} Ed_t^2$; it is easily verified that $Ed_t^2 = S$. For $\hat{\alpha}$ sufficiently close to α^*, we have

$$|d_{t+n}(\tilde{\alpha})D_{t+n}(\tilde{\alpha})(\hat{\alpha} - \alpha^*)| \leq r|d_{t+n}(\tilde{\alpha})||D_{t+n}(\tilde{\alpha})||\hat{\alpha} - \alpha^*| \leq rm_t^2|\hat{\alpha} - \alpha^*|$$

$$\Rightarrow$$

$$|B_T| \leq r|\hat{\alpha} - \alpha^*|[(T-n)^{-1}\sum m_t^2].$$

Since $(T-n)^{-1}\sum m_t^2$ is $O_p(1)$ by Markov's inequality, $B_T \xrightarrow{p} 0$ under the conditions of Proposition 1, and $T^{1/2}B_T = O_p(1)$ under the conditions of Proposition 2.

Assumptions similar to Assumptions 2 and 3 are also made in Andrews and Monahan (1992) and Newey and West (1994). Assumption 4 follows from the assumption about summability of fourth cumulants made in Assumption A in Andrews and Monahan and in Assumption 2 in Newey and West.

For the reader unfamiliar with those papers, the following illustration may help in interpretation of my assumptions. Consider a scalar linear model,

$f_t(\beta^*) = y_t - X_t'\beta^*$ for some observable data y_t (a scalar) and X_t. Then Assumption 2 holds. Assumption 3 holds if $\sup_t E|Z_t|^4 < \infty$, $\sup_t E|X_t|^4 < \infty$, $\sup_t E|\varepsilon_t|^4 < \infty$. Assumption 4 holds if Z_t is stationary with moving average representation (say) $\sum_{j=0}^{\infty} g_j e_{t-j}$ and $\sum_{j=0}^{\infty} |g_j| < \infty$; for some m, the $(q+1)\times 1$ vector $\{e_t', \varepsilon_{t-m}\}'$ is i.i.d. with finite eighth moments, with $Ee_t\varepsilon_{t-m}$ possibly not zero. (See Section 2 on the dating convention, which accounts for a nonzero cross-correlation between e_t and ε_{t-m} occurring when $m \neq 0$.)

Appendix B

This appendix outlines the asymptotic theory used to compute the figures in Table 1. Let $\sigma^2 \equiv E\varepsilon_t^2$, $\Gamma_0 = Eu_t^2 = (1+\theta_1^2)\sigma^2$, $\Gamma_1 = Eu_t u_{t-1} = \theta_1\sigma^2$. From Fuller (1976, p. 239) the asymptotic variance of the truncated estimator $\hat{\Gamma}_0 + 2\hat{\Gamma}_1$ is $V_{11} + 4V_{12} + 4V_{22}$, $V_{11} = 2\Gamma_0^2 + 4\Gamma_1^2$, $V_{12} = 4\Gamma_0\Gamma_1$, $V_{22} = \Gamma_0^2 + 3\Gamma_1^2$.

In this example, the MA-l estimator is $(1+\hat{\theta}_1)^2(T-1)^{-1}\sum_{t=1}^{T-1}\hat{\varepsilon}_{t+1}^2 = (1+\hat{\theta}_1)^2\hat{\sigma}^2$, where $\hat{\varepsilon}_t$ is the NLLS residual. From Fuller (1976, pp. 346–349), one can conclude the following. After some rearrangement, a second-order mean value expansion of $\hat{S}$ around S gives $\sqrt{T}(\hat{S} - S) = g'\delta_T + o_p(1)$, $g \equiv [2(1+\theta_1)\sigma^2, (1+\theta_1)^2]'$, $\delta_T = \sqrt{T}(\hat{\theta}_1 - \theta_1, \hat{\sigma}^2 - \sigma^2)'$. For a certain (2×1) random vector C_T, $\delta_T = C_T + o_p(1)$, with $\lim_{T\to\infty} EC_TC_T' = C$, $C(1,1) = 1 - \theta_1^2$, $C(1,2) = C(2,1) = 0$, $C(2,2) = 2\sigma^4 \Rightarrow$ the asymptotic variance of $\hat{S}$ is $g'Cg = 4(1+\theta_1)^2(1-\theta_1^2)\sigma^4 + 2(1+\theta_1)^4\sigma^4$.

Appendix C

Percentage of truncated estimates that were not positive definite:

	Value of θ_1						
	−0.9	−0.6	−0.3	0.0	0.3	0.6	0.9
Table 2A	65.3	44.7	5.3	0.1	0.0	0.0	0.0
Table 2B	93.2	78.5	27.3	1.7	0.0	0.0	0.0
Table 3A	69.7	48.9	17.6	2.6	0.2	0.5	0.1
Table 3B	88.7	78.9	45.3	12.6	1.7	0.2	0.0

	Values of ϕ, θ_1, θ_2					
	0.5, −1.3, 0.5	0.5, −1.0, 0.2	0.5, 0.67, 0.33	0.9, −1.3, 0.5	0.5, −1.0, 0.2	0.9, 0.67, 0.33
Table 2C	68.6	63.3	0.0	95.6	93.6	0.0
Table 3C	72.8	68.6	0.6	94.6	93.4	0.7

References

Andrews, Donald W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817–858.

Andrews, Donald W.K. and J. Christopher Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 60, 953–966.

Brockwell, Peter J. and Richard A. Davis, 1991, Time series analysis: Theory and methods (Springer-Verlag, New York, NY).

Brown, Bryan W. and Shlomo Maital, 1981, What do economists know? An empirical study of experts' expectations, Econometrica 49, 491–504.

Cumby, Robert E. and John Huizanga, 1992, Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions, Econometrica 60, 185–196.

Cumby, Robert E., John Huizanga, and Maurice Obstfeld, 1983, Two-step, two-stage least squares estimation in models with rational expectations, Journal of Econometrics 21, 333–355.

Eichenbaum, Martin S., Hansen, Lars Peter, and Kenneth J. Singleton, 1988, A time series analysis of representative agent models of consumption and leisure choice under uncertainty, Quarterly Journal of Economics CIII, 51–78.

Fuller, Wayne A., 1976, Introduction to statistical time series (Academic Press, New York, NY).

Hannan, E.J. and Manfred Deistler, 1988, The statistical theory of linear systems (Wiley, New York, NY).

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029–1054.

Hansen, Lars Peter and Robert J. Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates: An econometric analysis, Journal of Political Economy 96, 829–853.

Hansen, Lars Peter and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica 50, 1269–1286.

Hansen, Lars Peter and Kenneth J. Singleton, 1990, Efficient estimation of linear asset pricing models with moving-average errors, NBER technical working paper no. 86.

Hayashi, Fumio and Christopher A. Sims, 1983, Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments, Econometrica 51, 783–798.

Hodrick, Robert J., 1991, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, NBER technical working paper no. 108.

Kollintzas, Tryphon, 1993, A generalized variance bounds test, with an application to the Holt et al. inventory model, Manuscript (Athens University of Economics and Business, Athens).

Newey, Whitney K. and Kenneth D. West, 1994, Automatic lag selection in covariance matrix estimation, Review of Economic Studies 61, 631–654.

West, Kenneth D., 1986, A variance bounds test of the linear quadratic inventory model, Journal of Political Economy 94, 374–401.

West, Kenneth D. and David W. Wilcox, 1996, A comparison of alternative instrumental variables estimators of a dynamic linear model, Journal of Business and Economic Statistics 14, 281–293.