Econometrica, Vol. 55, No. 3 (May, 1987), 703-708

A SIMPLE, POSITIVE SEMI-DEFINITE, HETEROSKEDASTICITY AND AUTOCORRELATION CONSISTENT COVARIANCE MATRIX

BY WHITNEY K. NEWEY AND KENNETH D. WEST[1]

MANY RECENT RATIONAL EXPECTATIONS MODELS have been estimated by the techniques developed by Hansen (1982), Hansen and Singleton (1982), Cumby, Huizinga, and Obstfeld (1983), and White and Domowitz (1984). These estimation techniques make use of an orthogonality condition $Eh_t(\theta^*)=0$, where θ^* is a $(k\times 1)$ vector of unknown parameters and $h_t(\theta)$ is a $(r\times 1)$ vector of functions of the data and parameters, where $r\geq k$. This orthogonality condition can be employed to form a generalized method of moments (GMM, Hansen (1982)) estimator of θ^* by choosing $\hat{\theta}$ as the solution to

$$\text{(1)} \qquad \min_\theta h_T(\theta)'\hat{W}_T h_T(\theta),$$

where $h_T(\theta)=\sum_{t=1}^T h_t(\theta)/T$ is the vector of sample moments of $h_t(\theta)$ and $\hat{W}_T$ is a (possibly) random, symmetric weighting matrix.

As shown in Cumby, Huizinga, and Obstfeld (1983), Hansen (1982), and White and Domowitz (1984), the asymptotic covariance matrix of $\hat{\theta}$ is given by

$$\text{(2)} \qquad V_T=(H_T'W_TH_T)^{-1}H_T'W_TS_TW_TH_T(H_T'W_TH_T)^{-1}$$

where $H_T=\sum_{t=1}^T E[h_{t\theta}(\theta^*)]/T$ and $h_{t\theta}(\theta)$ is the $(r\times k)$ matrix of partial derivatives of $h_t(\theta)$, W_T is a nonrandom matrix such that plim $(\hat{W}_T-W_T)=0$, and $S_T=\sum_{s=1}^T\sum_{t=1}^T E[h_t(\theta^*)h_s(\theta^*)']/T$. Consistent estimation of this asymptotic covariance matrix is essential for the construction of asymptotic confidence intervals and hypothesis tests. Estimation of H_T and W_T is straightforward, since $\hat{W}_T$ forms a natural estimator of W_T and under the regularity conditions in Hansen (1982) or White and Domowitz (1984) it will be the case that

$$\text{(3)} \qquad H_T-\sum_{t=1}^T h_{t\theta}(\hat{\theta})/T \xrightarrow{P} 0.$$

Estimation of S_T is more difficult, and is also more important. As shown by Hansen (1982), an optimal GMM estimator (in the sense that V_T is as small as possible) is obtained when $\hat{W}_T$ is a consistent estimator of $(S_T)^{-1}$, so that estimation of S_T is also important for the formation of an optimal GMM estimator.

The simplest estimator of S_T that has been proposed takes the form

$$\text{(4)} \qquad \tilde{S}_T=\hat{\Omega}_0+\sum_{j=1}^m[\hat{\Omega}_j+\hat{\Omega}_j'], \quad \hat{\Omega}_j=\sum_{t=j+1}^T \hat{h}_t\hat{h}_{t-j}'/T,$$

where $\hat{h}_t=h_t(\hat{\theta})$. The bound m on the number of sample autocovariances $\hat{\Omega}_j$ used to form $\tilde{S}_T$, is in many studies equal to the number of nonzero autocorrelations of $h_t(\theta^*)$, which is known a priori (e.g., Cumby, Huizinga, and Obstfeld (1983), Hansen and Singleton (1982), and West (1986a)). In some studies the number of nonzero autocorrelations is not known a priori and may not even be finite (e.g., West (1985, 1987)). In such cases S_T may still be consistently estimated by $\tilde{S}_T$ (i.e., $\tilde{S}_T-S_T\xrightarrow{p}O$) if m is chosen to be a function $m(T)$ of sample size and is allowed to grow slowly enough with the sample size (see White and Domowitz (1984) and Theorem 2 below).

While $\tilde{S}_T$ is consistent, it need not be positive semi-definite in any finite sample when m is not zero. It follows that an estimator of V_T that uses $\tilde{S}_T$ as the middle matrix need

[1] We are grateful to Stephen Cecchetti, Lars Hansen, John Huizinga, and two referees for helpful comments. We are also grateful to the NSF for research support under Grants SES-8410249 and SES-8511070. This paper was revised while West was a National Fellow at the Hoover Institution.

not be positive semi-definite. This property of $\tilde{S}_T$ interferes with asymptotic confidence interval formation and hypothesis testing. Estimated variances and test statistics will be negative for some linear combinations of $\hat{\theta}$ when the estimated covariance matrix is not positive semi-definite. In addition, an estimator of S_T that is not positive semi-definite may be troublesome because, as pointed out to us by John Huizinga, iterative techniques for computing an optimal GMM estimator with $\hat{W}_T=(\tilde{S}_T)^{-1}$ may behave poorly if $\tilde{S}_T$ is not positive semi-definite.

Time domain techniques to calculate an estimator of S_T that is positive semi-definite have been suggested by Eichenbaum, Hansen, and Singleton (1985) and by Cumby, Huizinga, and Obstfeld (1983). These techniques appear to be difficult to apply in practice. Hansen (1982) suggested the use of spectral methods for the estimation of S_T, motivated by the fact that in the covariance stationary case the limit of S_T is 2π times the spectral density of $h_t(\theta^*)$ at frequency zero. Although frequency domain techniques for estimating S_T are cumbersome, a time domain approach turns out to be very useful. As in West (1985) we consider an estimator $\hat{S}_T$ of S_T that is as simple to compute as $\tilde{S}_T$:

$$(5) \qquad \hat{S}_T=\hat{\Omega}_0+\sum_{j=1}^{m} w(j,m)[\hat{\Omega}_j+\hat{\Omega}_j'], \quad w(j,m)=1-[j/(m+1)].$$

This estimator is numerically equal to 2π times an estimator of the spectral density of $h_t(\theta^*)$ at frequency zero, where the modified Bartlett weights are used to smooth the sample autocovariance function; see Anderson (1971, Section 9.2). Note that $\hat{S}_T$ is obtained in a similar fashion to $\tilde{S}_T$, except that the sample autocovariances are weighted by $w(j,m)=1-[j/(m+1)]$, which declines as j increases. Such a covariance smoothing approach to estimation of S_T has been suggested by Doan and Litterman (1983).[2] That $\hat{S}_T$ is positive semi-definite follows from the positive semi-definiteness of the sample autocovariance function.

THEOREM 1: *$\hat{S}_T$ is positive semi-definite.*

PROOF: For any $(r\times 1)$ vector c, $c'\hat{S}_Tc=\omega_0+2\sum_{j=1}^{m} w(j,m)\omega(j)$, where $\omega(j)=\sum_{t=j+1}^{T}(c'\hat{h}_t)(c'\hat{h}_{t-j})/T\,(j=0,1,\ldots,T-1)$. Let $P=[p_{ij}]$ be the $(m+1)$-dimensional symmetric matrix with $p_{ij}=\omega(|i-j|)$. Positive semi-definiteness of P is proven, for example, in McLeod and Jimenez (1984). Letting e be a $([m+1]\times 1)$ vector of ones, we then have

$$(6) \qquad c'\hat{S}_Tc=e'Pe/(m+1)\geq 0.$$

Other choices of the weight function $w(j,m)$ will also yield positive semi-definite estimators of S_T. If the vector of ones in the proof of Theorem 1 is replaced by $(v(0,m),\ldots,v(m,m))$, where each $v(j,m)$ is an arbitrary number, then we find that the following choice of weights will also yield a positive semi-definite estimator of S_T:

$$(7) \qquad w(j,m)=\left[\sum_{\ell=0}^{m-j} v(\ell,m)v(\ell+j,m)\right]\Big/\left[\sum_{\ell=0}^{m} v(\ell,m)^2\right].$$

Also, if $w(j,m)$ is chosen to be a weight function that would generate a nonnegative spectral density estimate for a univariate time series then the resulting estimator of S_T will be positive semi-definite. Anderson (1971, Section 9.2) discusses the relative merits of different weighting schemes under a different set of regularity conditions than those we consider below. Gallant (1985) also discusses the choice of weights and presents results similar to ours.[3]

Note that for fixed j the weight $w(j,m)=1-[j/(m+1)]$ approaches one as m grows. It is reasonable to expect that estimators of S_T that are formed by smoothing sample

[2] Doan and Litterman (1983) do not assert or show that S_T is positive semi-definite, nor do they establish consistency.

[3] Gallant's (1985) manuscript came to our attention after this note was submitted.

autocovariances with weights that approach one as m grows should be consistent if m is allowed to grow with the sample size. The consistency of such estimators of S_T can be shown to hold under regularity conditions like those of White and Domowitz (1984), where the interested reader is referred for the notation and definitions that relate to mixing conditions. For a matrix $A=[a_{ij}]$, let $|A|$ denote the norm $\max_{i,j}|a_{ij}|$.

THEOREM 2: *Suppose that*:

(i) $h_t(\theta)=h(z_t,\theta)$, *where* $h(z,\theta)$ *is measurable in* z *for all* θ, *and continuously differentiable in* θ *for all* θ *in a neighborhood* N *of* θ^*, *with probability one*;

(ii) (*a*) *there is a measurable function* $m(z)$ *such that* $\sup_N|h_t(\theta)|<m(z)$ *and* $\sup_N|h_{t\theta}(\theta)|<m(z)$, *where for some finite constant* D, $E[m(z_t)^2]<D$ *for all* t; (b) *there are finite constants* D, $\delta>0$ *and* $r\geqslant 1$, *such that for all* t, $E[|h_t(\theta^*)|^{4(r+\delta)}]<D$;

(iii) z_t *is a mixing sequence with either* $\phi(\ell)$ *of size* $2r/(2r-1)$ *or* $\alpha(\ell)$ *of size* $2r/(r-1)$, $r>1$;

(iv) *for all* t, $E[h_t(\theta^*)]=0$, *and* $\sqrt{T}(\hat{\theta}-\theta^*)$ *is bounded in probability*;

(v) *the weights* $w(j,m)$, $(m=1,2,\ldots,j=1,\ldots,m)$ *satisfy* $|w(j,m)|\leqslant C$ *for finite constant* C *and for each* j, $\lim_{m\to\infty}w(j,m)=1$.

Then if m *is chosen to be a function* $m(T)$ *of sample size such that* $\lim_{T\to\infty}m(T)=+\infty$ *and* $\lim_{T\to\infty}[m(T)/T^{1/4}]=0$, *it follows that*

$$\left\{\hat{\Omega}_0+\sum_{j=1}^{m(T)}w(j,m(T))[\hat{\Omega}_j+\hat{\Omega}_j']\right\}-S_T\xrightarrow{P}0. \tag{8}$$

The proof of Theorem 2 is given below.

The assumptions of Theorem 2 require that $h_t(\theta)$ and $h_{t\theta}(\theta)$ be dominated by a function of z_t that has uniformly bounded second moment, that $h_t(\theta^*)$ have uniformly bounded moments of up to slightly more than the fourth order, and that the dependence between observations go to zero at certain rates as the distance between observations increases. Consistency follows if $m(T)$ goes to infinity with T more slowly than $T^{1/4}$.

Note that choosing $w(j,m)$ equal to one for each j and m yields the estimator $\tilde{S}_T$ of equation (4), a special case of which was considered by White and Domowitz (1984). The consistency result of Theorem 2 differs from that of Theorem 3.5 of White and Domowitz (1984) in two respects. First, the slower rate of growth of $m(T)$ required in Theorem 2, with $m(T)$ required to grow slower than $T^{1/4}$ rather than slower than $T^{1/3}$, results from a slight correction to the arguments in White and Domowitz (1984), and not from allowing for a general class of weights. Second, the above consistency result allows for general forms of nonlinearity in the parameters.

It should be noted that the derivation of the slower than $T^{1/4}$ growth rate for $m(T)$ depends heavily on the use of mixing conditions. If $h_t(\theta^*)$ is an infinite order moving average with absolutely summable coefficients and i.i.d. innovations, where the innovations have finite fourth moments, then the proof of Theorem 2 and Theorem 7.2.3 in Fuller (1976) can be combined to show that a growth rate of slower than $T^{1/2}$ for $m(T)$ will suffice for consistency of $\hat{S}_T$. On the other hand, as pointed out to us by Lars Hansen, it may be difficult to obtain an appropriate growth rate for $m(T)$ under weaker dependence restrictions than mixing, such as the stationary, ergodic situation considered in Hansen (1982).

The specification of an appropriate growth rate for $m(T)$ gives little guidance concerning the choice of m in practice. Cross-validation methods (e.g., Wahba and Wold (1975)) and the testing approach suggested by White and Domowitz (1984) may prove useful. The assessment of such suggestions using Monte Carlo work or more refined asymptotics is an important topic of future research. It would also be useful to know if the estimators suggested by Cumby, Huizinga, and Obstfeld (1983) and Eichenbaum, Hansen, and Singleton (1984) provide better estimators of S_T than $\hat{S}_T$ when the number of nonzero autocorrelations is known a priori.

PROOF OF THEOREM 2: A sequence of symmetric matrices $\{A_T\}$ converges to a symmetric matrix A_0 if and only if $c'A_Tc \to c'A_0c$ for all comfortable vectors c. Then taking a linear combination $c'\hat{h}_t$ where, e.g., $|c'\hat{h}_t| \leq r|c||\hat{h}_t|$, we can restrict attention to the scalar case with $r=1$.

Let $\bar{S}_T = \sum_{t=1}^T \hat{h}_t^2/T + 2\sum_{j=1}^m w(j,m)\sum_{t=j+1}^T \hat{h}_t\hat{h}_{t-j}/T$ and $h_t = h_t(\theta^*)$. For notational convenience we will suppress the T argument in $m(T)$. It follows by the triangle inequality and the form of $\bar{S}_T$ that

$$
\begin{aligned}
(9)\qquad |\bar{S}_T - S_T| \leq {} & \left| \bar{S}_T - \left[\sum_{t=1}^T h_t^2/T + 2\sum_{j=1}^m w(j,m) \sum_{t=j+1}^T h_t h_{t-j}/T \right] \right| \\
& + \left| \sum_{t=1}^T (h_t^2 - E[h_t^2])/T \right. \\
& \left. + 2\sum_{j=1}^m w(j,m) \sum_{t=j+1}^T (h_t h_{t-j} - E[h_t h_{t-j}])/T \right| \\
& + \left| \sum_{t=1}^T E[h_t^2]/T + 2\sum_{j=1}^m w(j,m) \sum_{t=j+1}^T E[h_t h_{t-j}]/T - S_T \right| \\
\leq {} & \left| \bar{S}_T - \left[\sum_{t=1}^T h_t^2/T + 2\sum_{j=1}^m w(j,m) \sum_{t=j+1}^T h_t h_{t-j}/T \right] \right| \\
& + \left| \sum_{t=1}^T (h_t^2 - E[h_t^2])/T \right. \\
& \left. + 2\sum_{j=1}^m w(j,m) \sum_{t=j+1}^T (h_t h_{t-j} - E[h_t h_{t-j}])/T \right| \\
& + 2\sum_{j=1}^m |w(j,m) - 1| \sum_{t=j+1}^T |E[h_t h_{t-j}]|/T \\
& + 2\sum_{j=m+1}^{T-1} \sum_{t=j+1}^T |E[h_t h_{t-j}]|/T.
\end{aligned}
$$

The fourth term goes to zero as T goes to infinity by Lemma 6.17 of White (1984) and $\lim_{T\to\infty} m = +\infty$.

By Corollary 6.16 in White (1984), there is a sequence $\gamma(\ell)$ ($\ell = 1, \ldots, \infty$), and a finite constant D' such that $|E[h_t h_{t-j}]| < D'\gamma(j)$ for all T and for all j, with $\sum_{\ell=1}^\infty \gamma(\ell) < +\infty$. Then $\sum_{t=j+1}^T |E[h_t h_{t-j}]|/T < D'\gamma(j)$ for all T and j. Since $\lim_{T\to\infty} w(j,m) = 1$ for each j follows from assumption (v), the dominated convergence theorem, applied to the counting measure on the positive integers, implies that the third term in equation (9) goes to zero as T goes to infinity.

Let $Z_{tj} = h_t h_{t-j} - E[h_t h_{t-j}]$. Assumption (ii) (b) implies that there is a finite constant D' such that $E(|Z_{tj}|^{2(r+\delta)}) < D'$ for all t and j. The proof of Lemma 6.19 in White (1984) is incorrect as stated, and so cannot be used to show that the second term in equation (9) converges in probability to zero. Nevertheless, if one replaces (in our notation) the double sum $\sum_{\ell=j+1}^{T-1}\sum_{t=\ell+1}^T$ on page 153 of White (1984) with the correctly indexed sum $2\sum_{\ell=1}^{T-j-1}\sum_{t=j+1+\ell}^T$ and applies the same argument as in the proof of Lemma 6.19 in White (1984), one finds that there is a finite constant D^* such that for all j between zero and T, and for all T,

$$
(10)\qquad E\left\{\left[\sum_{t=j+1}^T Z_{tj}\right]^2\right\} \leq (T-j)(j+1)D^* \leq T(j+1)D^*, \quad j \geq 0.
$$

It follows from $w(j, m)$ uniformly bounded by C that $\sum_{j=1}^{m}|w(j,m)| \leqslant mC$. Then for any $\varepsilon > 0$, the triangle inequality, the implication rule (i.e., if the occurrence of event A implies that event B has occurred, then Prob $(A) \leqslant$ Prob (B)), the fact that the probability of the union of several events is less than or equal to the sum of the probabilities, Chebyshev's inequality, and equation (10) imply

$$
\begin{aligned}
(11)\qquad P\left(\left|\sum_{j=1}^{m} w(j,m) \sum_{t=j+1}^{T} Z_{tj}\right| \Big/ T > \varepsilon\right) &\leqslant P\left(\sum_{j=1}^{m} |w(j,m)| \left|\sum_{t=j+1}^{T} Z_{tj}/T\right| > \varepsilon\right) \\
&\leqslant \sum_{j=1}^{m} P\left(\left|\sum_{t=j+1}^{T} Z_{tj}/T\right| > \varepsilon/Cm\right) \\
&\leqslant \sum_{j=1}^{m} (Cm/\varepsilon)^2 D^*(j+1)/T = D^* C^2 m^3 (m+3)/(2\varepsilon^2 T).
\end{aligned}
$$

Then the second term in equation (9) converges in probability to zero by the fact that m grows more slowly than $T^{1/4}$, equation (10) (with $j = 0$) applied to $\sum_{t=1}^{T}(h_t^2 - E[h_t^2])/T$, and the triangle inequality.

By (iv), $\hat{\theta}$ lies in N with probability approaching one as T grows, so that with probability approaching one it is possible to obtain a mean value expansion of $\tilde{S}_T$ around θ^*. Let $\tilde{h}_t = h_t(\tilde{\theta})$ and $\tilde{h}_{t\theta} = h_{t\theta}(\tilde{\theta})$, where $\tilde{\theta}$ is the mean value from this expansion. Then with probability approaching one, the first term after the inequality in equation (9) can be written as

$$
\begin{aligned}
&2\left|\left[\sum_{t=1}^{T} \tilde{h}_t \tilde{h}_{t\theta} + \sum_{j=1}^{m} w(j,m) \sum_{t=j+1}^{T} (\tilde{h}_t \tilde{h}_{t-j\theta} + \tilde{h}_{t-j} \tilde{h}_{t\theta})\right](\hat{\theta} - \theta^*)\right| \Big/ T \\
&\leqslant 2\left[\sum_{t=1}^{T} m(z_t)^2 + \sum_{j=1}^{m} |w(j,m)| \sum_{t=j+1}^{T} 2m(z_t)m(z_{t-j})\right] \cdot |\hat{\theta} - \theta^*|/T \\
&\leqslant 2\left[\sum_{t=1}^{T} m(z_t)^2 + \sum_{j=1}^{m} |w(j,m)| \sum_{t=j+1}^{T} \{m(z_t)^2 + m(z_{t-j})^2\}\right] \cdot |\hat{\theta} - \theta^*|/T \\
&\leqq 2[(2Cm+1)/\sqrt{T}] \cdot \left[\sum_{t=1}^{T} m(z_t)^2/T\right] \cdot \sqrt{T}|\hat{\theta} - \theta^*|.
\end{aligned}
$$

Note that $\sqrt{T}|\hat{\theta} - \theta^*|$ is bounded in probability by assumption (iv) and that $\sum_{t=1}^{T} m(z_t)^2/T$ is bounded in probability by Markov's inequality and assumption (ii)(a). Then the first term in equation (9) converges in probability to zero, since the fact that m grows more slowly than $T^{1/4}$ implies that $(2Cm+1)/\sqrt{T}$ converges to zero.

The conclusion now follows from equation (9), since we have shown that each of the terms on the right-hand side of the second inequality converges in probability to zero.

Department of Economics, Princeton University, Princeton, NJ 08544, U.S.A.
and
Woodrow Wilson School, Princeton University, Princeton, NJ 08544, U.S.A.

Manuscript received May, 1985; final revision received March, 1986.

REFERENCES

ANDERSON, T. W. (1971): *The Statistical Analysis of Time Series.* New York: John Wiley and Sons.

CUMBY, R. E., J. HUIZINGA, AND M. OBSTFELD (1983): "Two-Step Two-Stage Least Squares Estimation in Models with Rational Expectations," *Journal of Econometrics*, 21, 333–355.

DOAN, T. A., AND R. B. LITTERMAN (1983): *RATS User's Manual.* Minneapolis: VAR Econometrics.
EICHENBAUM, M. S., L. P. HANSEN, AND K. J. SINGLETON (1985): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," manuscript.
FULLER, W. A. (1976): *Introduction to Statistical Time Series.* New York: John Wiley and Sons, 1976.
GALLANT, A. R. (1985): "Dynamic Nonlinear Models," Ch. 9 of *Nonlinear Statistical Models,* North Carolina State University, Institute of Statistics Mimeograph Series No. 1667.
HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica,* 50, 1029-1054.
HANSEN, L. P., AND K. J. SINGLETON (1982): "Generalized Instrumental Variable Estimation of Nonlinear Rational Expectations Models," *Econometrica,* 50, 1269-1286.
MCLEOD, I. A., AND C. JIMENEZ (1984): "Nonnegative Definiteness of the Sample Autocovariance Function," *American Statistician,* 38, 297-298.
WEST, K. D. (1985): "A Specification Test for Speculative Bubbles," Princeton University Financial Research Memorandum No. 58.
——— (1986): "A Variance Bounds Test of the Linear Quadratic Inventory Model," *Journal of Political Economy,* 94, 374-401.
———(1987): "Dividend Innovations and Stock Price Volatility," *Econometrica,* forthcoming.
WHABA, G. AND S. WOLD (1975): "A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation," *Communications in Statistics,* 4, 1-17.
WHITE, H., AND I. DOMOWITZ (1984): "Nonlinear Regression with Dependent Observations," *Econometrica,* 52, 143-161.
WHITE, H. (1984): *Asymptotic Theory for Econometricians.* New York: Academic Press.