8 Summary and Conclusion

In the foregoing chapters we have presented a theory of estimation and inference which applies to a variety of econometric estimators of the parameters of possibly misspecified models of time-dependent heterogeneous processes. Although the results are fairly general with regard to the scope of estimators and the allowed behavior of the stochastic processes under study, they are nevertheless restrictive in a number of different ways. Relaxing these restrictions is beyond the scope of this work; such relaxations constitute a number of interesting directions for further research.

Specifically, in assumption DG we focus attention on discrete time series. Treatment of continuously recorded processes or processes observed at irregular intervals is certainly a useful area for subsequent investigation. In assumption OP we restrict the nature of the extremum estimators studied in a convenient way; however, we avoid an explicit general treatment of multistage estimators. The reason for this is primarily ease of notation. Nevertheless, certain important cases will require explicit results for multistage estimators which do not fall into the present context. Extension to these cases appears straightforward in most cases (see Andrews and Fair 1987). Results given here may prove helpful in this effort.

Assumption OP also imposes the requirement that the parameter space Θ be a compact subset of a finite dimensional Euclidean space. Compactness is a great convenience; however, analogous results are readily available for σ-compact sets (e.g. Perlman 1972; Hansen 1982). Finite dimensionality is also unnecessary, at least for consistency results. By allowing for infinite dimensional parameter spaces, certain nonparametric estimators (e.g. the method of sieves - see Geman and Hwang 1982) can be brought into the present context (e.g. Wooldridge and White 1985).

Assumption OP and later assumption SM impose continuity and further smoothness (differentiability and Lipschitz) conditions on the

functions q_t. Techniques for replacing continuity with semicontinuity or differentiability with differentiability in mean square are well known (e.g. Hoadley 1971; Roussas 1972). It should pose no more than a modest technical challenge to relax the smoothness conditions imposed here.

The mixing and near epoch dependence assumptions MX and NE are those which distinguish the present work from other work in this area. Together they allow for a degree of time dependence and heterogeneity of the stochastic processes studied not previously available in the econometric literature. Even so, it should be possible to establish laws of large numbers and central limit theorems under less restrictive mixing and near epoch dependence conditions than those used here. (Andrews 1987 accomplishes this nicely for a weak law of large numbers.) The techniques used here, and the underlying results for extremum estimators which we provide, should assist in obtaining similar results under conditions weaker than MX and NE.

One of the more restrictive aspects of the present results is the imposition of domination conditions in assumption DM. This rules out certain trending or explosive stochastic processes, thereby eliminating a very important class of linear and nonlinear data generating processes from consideration. Elegant results of Crowder (1976) and Weiss (1971; 1973) for maximum likelihood estimation of correctly specified models suggest techniques useful in extending our treatment of extremum estimators of misspecified models to allow for unbounded processes. Substantial progress in this direction has already been made by Wooldridge (1986).

The identification condition, assumption ID, is also stronger than necessary. A weaker condition is given by Wooldridge and White (1985), related to that of Perlman (1972). Perlman gives conditions under which such a condition is necessary and sufficient for identification.

The remaining assumptions are conditions relevant to the problems of covariance matrix estimation and/or the study of local power. Of these, only assumption TL deserves special comment. Recall that assumption TL requires that the truncation lag m_n for estimating the autocorrelation- and heteroskedasticity-consistent parameter covariance matrix tend to infinity but not too fast, as $m_n = o(n^{1/4})$. It is possible that this rate is slower than absolutely necessary, and that some more clever method of proof would yield a faster allowable rate. We have left this investigation to other work.

Whatever the limitations of our assumptions, our method of analysis is emphatically limited by its focus on asymptotic results. We make no apologies for this, as an asymptotic analysis is clearly an appropriate first step in the investigation of any general estimation problem. However, we must issue a clear warning: the asymptotic results may provide little or no guidance as to what to expect in samples of the size typically used by economists and econometricians. The asymptotic results provide especially little guidance in the area of covariance matrix estimation. Investigation of finite sample behavior of our statistics and techniques for improving the asymptotic approximations discussed here (by adjusting either the statistics or the approximating distribution) should be of highest priority.

Despite our desire to give a comprehensive treatment, there are a number of issues which we have neglected almost entirely, primarily because proper discussion of these issues within the current framework would require at least another book! The most important of these issues are the topics of asymptotic efficiency in estimation and of specification testing. General treatments of asymptotic efficiency can be found in the excellent books by Ibragimov and Has'minskii (1981) and Roussas (1972). Recent articles in the econometrics literature more or less closely related to the present context are Hansen (1985) and Bates and White (1987). The latter work makes use of results given here in obtaining optimal instrumental variables estimators for systems of nonlinear implicit simultaneous systems of equations with generally nonspherical errors, in a context allowing for general dependent heterogeneous stochastic processes.

The specification testing issue is elegantly treated by Bierens (1982; 1984; 1987) in a context closely related to that considered here. White (1987a; 1987b) provides an extensive discussion of this issue in a context to which the present results are immediately relevant. Specifically, our results here can be used to construct a variety of interesting new specification tests using the conditional moment approach of Newey (1985), having power against a variety of potentially serious misspecifications. To mention just one possibility, the present results support construction of analogues of information matrix and dynamic information matrix tests for models estimated by the generalized method of moments (Hansen 1982).

Of necessity, this discussion only barely scratches the surface of what we have neglected or left undone, and of the possibilities for future work. However, we offer our present results with the hope that they may prove helpful in further developing the theory of estimation and inference for nonlinear dynamic models.

REFERENCES

Andrews, D. W. K. 1987: Laws of large numbers for dependent non-identically distributed random variables, California Institute of Technology, Division of Humanities and Social Sciences, unpublished paper.

Andrews, D. W. K. and R. Fair 1987: Inference in econometric models with structural change, Yale University Cowles Foundation discussion paper 832.

Bates, C. and H. White 1987: Efficient instrumental variables estimation of systems of implicit heterogeneous nonlinear dynamic equations with nonspherical errors, in W. A. Barnett, E. R. Berndt, and H. White (eds), *International Symposia in Economic Theory and Econometrics*, vol. 3, New York: Cambridge University Press, forthcoming.

Bierens, H. 1982: Consistent model specification tests, *Journal of Econometrics* 20, 105–34.

Bierens, H. 1984: Model specification testing of time series regressions, *Journal of Econometrics* 26, 323–53.

Bierens, H. 1987: A consistent Hausman-type model specification test, Vrije Universiteit, Faculteit der Economische Wetenschappen research memorandum 1987-2.

Crowder, M. 1976: Maximum likelihood estimation for dependent observations, *Journal of the Royal Statistical Society*, series B 38, 45–53.

Geman, S. and C. H. Hwang 1982: Nonparametric maximum likelihood estimation by the method of sieves, *Annals of Statistics* 70, 401–14.

Hansen, L. P. 1982: Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–54.

Hansen L. P. 1985: A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators, *Journal of Econometrics* 30, 203–38.

Hoadley, B. 1971: Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case, *Annals of Mathematical Statistics* 42, 1977–91.

Ibragimov, I. A. and R. Z. Has'minskii 1981: *Statistical Estimation: asymptotic theory*. New York: Springer-Verlag.

Newey, W. 1985: Maximum likelihood specification testing and conditional moment tests, *Econometrica* 53, 1047–70.

Perlman, M. 1972: On the strong consistency of approximate maximum likelihood estimators, in Proceedings of the Sixth Berkeley Symposium in Mathematical Statistics and Probability, Berkeley: University of California Press, 263–81.

Roussas, G. 1972: *Contiguity of Probability Measures: some applications in statistics*. Cambridge: Cambridge University Press.

Weiss, L. 1971: Asymptotic properties of maximum likelihooc estimators in some nonstandard cases, *Journal of the American Statistical Associction* 66, 345–50.
Weiss, L. 1973: Asymptotic properties of maximum likelihooc estimators in some nonstandard cases, II, *Journal of the American Statistical Association* 68, 428–30.
White, H. 1987a: Specification testing in dynamic models, in Truman Bewley (ed.), *Advances in Econometrics*, New York: Cambridge University Press, forthcoming.
White, H. 1987b: *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press, forthcoming.
Wooldridge, J. 1986: Asymptotic properties of econometric estimators, University of California San Diego, Department of Economics PhD dissertation.
Wooldridge, J. and H. White 1985: Consistency of optimization estimators, University of California San Diego, Department of Economics discussion paper 85-29.

Appendix

For ease of reference, this appendix collects together all of the labeled assumptions given in the text. Assumptions are listed in the order in which they are introduced. The assumptions and their various subparts are not in force before the indicated page.

Assumption DG (data generation) (p. 7)

Let (Ω, F, P) be a complete probability space. The observed data are generated as a realization

$$x_t = X_t(\omega) = W_t(\ldots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \ldots)$$

of a stochastic process $X_t : \Omega \to \mathbb{R}^{w_t}$, $w_t \in \mathbb{N} \equiv \{1, 2, \ldots\}$, where $V_t : \Omega \to \mathbb{R}^v$, $v \in \mathbb{N}$, and $W_t : \times_{\tau=-\infty}^{\infty} \mathbb{R}^v \to \mathbb{R}^{w_t}$ are such that X_t is measurable-$F/B(\mathbb{R}^{w_t})$, $t = 0, \pm 1, \pm 2, \ldots$. □

Assumption OP (optimand) (p. 11)

Let Θ be a compact subset of $\mathbb{R}^k$, $k \in \mathbb{N}$. For $n = 1, 2, \ldots$ define the optimand $Q_n : \Omega \times \Theta \to \mathbb{R}$ as

$$Q_n(\omega, \theta) \equiv g_n(\psi_n(\omega, \theta)),$$

where $\psi_n(\omega, \theta) \equiv n^{-1} \sum_{t=1}^{n} q_t(\omega, \theta)$, and

(i) $g_n : \mathbb{R}^l \to \mathbb{R}$ is continuous on compact subsets of $\mathbb{R}^l$ uniformly in n;
(ii) $q_t : \Omega \times \Theta \to \mathbb{R}^l$ is such that $q_t(\cdot, \theta)$ is measurable-$F/B(\mathbb{R}^l)$ for each θ in Θ and $q_t(\omega, \cdot)$ is continuous on Θ almost surely, i.e. for all ω in $F_t \in F$, $P(F_t) = 1$, $t = 1, 2, \ldots$. □

Assumption MX (mixing) (p. 35)

$\{V_t\}$ is a mixing sequence such that either ϕ_m is of size $-r/(2r-2)$, $r \geqslant 2$ or α_m is of size $-r/(r-2)$ with $r > 2$. □

Assumption SM (smoothness) (p. 35)

(i) $\{q_t\}$ is almost surely Lipschitz-L_1 on Θ. □

Assumption DM (domination) (p. 35)

The elements of $q_t(\theta)$ are r-dominated on Θ uniformly in $t = 1, 2, \ldots$, $r \geqslant 2$. □

Assumption NE (near epoch dependence) (p. 36)

(i) The elements of $\{q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size $-1/2$ on (Θ, ρ), where ρ is any convenient norm on $\mathbb{R}^k$. □

Assumption ID (identification) (p. 36)

When the functions $\bar{Q}_n = g_n \circ \bar{\psi}_n$ exist, $n = 1, 2, \ldots$, the sequence $\{\bar{Q}_n(\theta)\}$ has identifiably unique minimizers $\{\theta_n^*\}$ on Θ and identifiably unique minimizers $\{\theta_n^o\}$ on $\{\Theta_n\}$. □

Assumption OP′ (p. 73)

Let Θ be a compact subset of $\mathbb{R}^k$. For $n = 1, 2, \ldots$ define the optimand $Q_n : \Omega \times \Theta \to \mathbb{R}$ as

$$Q_n(\omega, \theta) \equiv g_n(\psi_n(\omega, \theta)),$$

where $\psi_n(\omega, \theta) \equiv n^{-1} \sum_{t=1}^n q_t(\omega, \theta)$, and

(i) $\{g_n : \mathbb{R}^l \to \mathbb{R}\}$ is continuously differentiable of order 2 on compact subsets of $\mathbb{R}^l$ uniformly in n;
(ii) $q_t : \Omega \times \Theta \to \mathbb{R}^l$ is a random function continuously differentiable of order 2 on Θ a.s., $t = 1, 2, \ldots$. □

Assumption DM′ (p. 76)

(i) The elements of $\{q_t(\theta)\}$ are r-dominated on Θ uniformly in $t = 1, 2, \ldots, r > 2$.
(ii) The elements of $\{\nabla_\theta q_t(\theta)\}$ are r-dominated on Θ uniformly in $t = 1, 2, \ldots, r > 2$. □

Assumption MX′ (p. 77)

$\{V_t\}$ is a mixing sequence such that either ϕ_m is of size $-r/(r-1)$, $r \geqslant 2$ or α_m is of size $-2r/(r-2)$, $r > 2$. □

Assumption NE′ (p. 77)

(i) The elements of $\{q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size -1 uniformly on (Θ, ρ).
(ii) The elements of $\{\nabla_\theta q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size -1 uniformly on (Θ, ρ). □

Assumption PD (positive definiteness) (p. 78)

(i) For $\{\theta_n^o\}$ and $\{\theta_n^*\}$ as defined in assumption ID, the sequences $\{B_n^o\}$ and $\{B_n^*\}$ are uniformly positive definite. □

Assumption SM (p. 78)

(ii) $\{\nabla_\theta q_t(\theta)\}$ is a.s. Lipschitz-L_1. □

Assumption ID′ (p. 79)

(i) The sequence $\{\bar{Q}_n(\theta)\}$ has identifiably unique minimizers $\{\theta_n^*\}$ on Θ, interior to Θ uniformly in n. □

Assumption SM (p. 79)

(iii) $\{\nabla_\theta^2 q_t(\theta)\}$ is a.s. Lipschitz-L_1. □

Assumption DM′ (p. 80)

(iii) The elements of $\{\nabla_\theta^2 q_t(\theta)\}$ are r-dominated on Θ uniformly in $t = 1, 2, \ldots, r > 2$ □

Assumption NE (p. 80)

(ii) The elements of $\{\nabla_\theta q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size $-1/2$ on (Θ, ρ);

(iii) The elements of $\{\nabla_\theta^2 q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size $-1/2$ on (Θ, ρ). □

Assumption NE′ (p. 80)

(iii) Assumption NE(iii) holds. □

Assumption PD (p. 81)

(ii) $\{A_n^*\}$ and $\{A_n^o\}$ are uniformly positive definite. □

Assumption DM″ (p. 93)

(i) The elements of $\{q_t(\theta)\}$ are $2r$-dominated on Θ uniformly in $t = 1, 2, \ldots, r > 2$.
(ii) The elements of $\{\nabla_\theta q_t(\theta)\}$ are $2r$-dominated on Θ uniformly in $t = 1, 2, \ldots, r > 2$.
(iii) The elements of $\{\nabla_\theta^2 q_t(\theta)\}$ are $2r$-dominated on Θ uniformly in $t = 1, 2, \ldots, r > 2$. □

Assumption NE″ (p. 94)

(i) The elements of $\{q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size $-(r-1)/(r-2)$ uniformly on (Θ, ρ).
(ii) The elements of $\{\nabla_\theta q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size $-(r-1)/(r-2)$ uniformly on (Θ, ρ). □

Assumption ID′ (p. 94)

(ii) The sequence $\{\bar{Q}_n(\theta)\}$ has identifiably unique minimizers $\{\theta_n^o\}$ on $\{\Theta_n\}$, where $\{\theta_n^o\}$ is interior to Θ uniformly in n. □

Assumption NE‴ (p. 101)

(i) The elements of $\{q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size $-2(r-1)/(r-2)$ uniformly on (Θ, ρ).
(ii) The elements of $\{\nabla_\theta q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size $-2(r-1)/(r-2)$ uniformly on (Θ, ρ). □

Assumption TL (truncation lag) (p. 101)

$\{m_n\}$ is a sequence of integers such that $m_n \to \infty$ as $n \to \infty$ and $m_n = o(n^{1/4})$. □

Assumption WT (weights) (p. 101)

For a given sequence $\{m_n\}$ define

$$w_{n\tau} = \sum_{\lambda=\tau+1}^{m_n+1} a_{n\lambda} a_{n\lambda-\tau},$$

where $\{a_{n\lambda}\}$, $n = 1, 2, \ldots, \lambda = 1, \ldots, m_n + 1$ is any triangular array such that $|w_{n\tau}| \leqslant \Delta < \infty$, $n = 1, 2, \ldots, \tau = 1, \ldots, m_n$, and for each τ, $w_{n\tau} \to 1$ as $n \to \infty$. □

Assumption HT (hypothesis testing) (p. 121)

Let (Ω, F, P) be a complete probability space and let $\Theta \subset \mathbb{R}^k$, $k \in \mathbb{N}$, be a compact set. Assume:

(a) $Q_n : \Omega \times \Theta \to \mathbb{R}$ is a random function continuously differentiable of order 2 on Θ, a.s., $n = 1, 2, \ldots$.
(b) There exist sequences of functions $\{\bar{Q}_n : \Theta \to \mathbb{R}\}$ and $\{A_n : \Theta \to \mathbb{R}^{k \times k}\}$ such that $\bar{Q}_n$ is differentiable on Θ and

$$Q_n(\theta) - \bar{Q}_n(\theta) \to 0 \quad \text{a.s. uniformly on } \Theta,$$
$$\nabla_\theta Q_n(\theta) - \nabla_\theta \bar{Q}_n(\theta) \to 0 \quad \text{a.s. uniformly on } \Theta,$$
$$\nabla_\theta^2 Q_n(\theta) - A_n(\theta) \to 0 \quad \text{a.s. uniformly on } \Theta,$$

where $\{\bar{Q}_n\}$, $\{\nabla_\theta \bar{Q}_n\}$, and $\{A_n\}$ are continuous on Θ uniformly in n.
(c) $\{\bar{Q}_n\}$ has identifiably unique minimizers $\{\theta_n^*\}$ on Θ, interior to Θ uniformly in n. Define

$$\Theta_n \equiv \{\theta \in \Theta : h(\theta) = h_n^o\}$$

where $h : \Theta \to \mathbb{R}^q$, $q \in \mathbb{N}$, is continuously differentiable on Θ, $h_n^* \equiv h(\theta_n^*)$, and $\{h_n^o\}$ is chosen so that

$$\sqrt{(n)}(h_n^* - h_n^o) = O(1).$$

Assume that $\{\bar{Q}_n\}$ has identifiably unique minimizers $\{\theta_n^o\}$ on $\{\Theta_n\}$, interior to Θ uniformly in n.

(d) $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$.

(e) There exist sequences $\{B_n^o\}$ and $\{B_n^*\}$ of $O(1)$ uniformly positive definite symmetric $k \times k$ matrices such that

$$\sqrt{(n)}B_n^{o\,-1/2}(\nabla_\theta Q_n^o - \nabla_\theta \bar{Q}_n^o)' \overset{A}{\sim} N(0, I_k)$$
$$\sqrt{(n)}B_n^{*\,-1/2}\nabla_\theta Q_n^{*\prime} \overset{A}{\sim} N(0, I_k).$$

(f) There exist sequences $\{\tilde{B}_n : \Omega \to \mathbb{R}^{k\times k}\}$ and $\{\hat{B}_n : \Omega \to \mathbb{R}^{k\times k}\}$ measurable-$F/B(\mathbb{R}^{k\times k})$ and $O(1)$ nonstochastic sequences $\{U_n^o\}$ and $\{U_n^*\}$ such that

$$\tilde{B}_n - (B_n^o + U_n^o) \xrightarrow{P} 0$$
$$\hat{B}_n - (B_n^* + U_n^*) \xrightarrow{P} 0.$$

(g) There exists a closed sphere $S \subset \Theta$ of finite nonzero radius such that for some $\varepsilon > 0$

$$\cup_{n=1}^{\infty} \{\theta \in \Theta : |\theta - \theta_n^*| < \varepsilon\} \subset S$$

and $\{A_n(\theta)\}$ is $O(1)$ and uniformly positive definite uniformly on S.

(h) There exist sequences $\{\tilde{A}_n : \Omega \to \mathbb{R}^{k\times k}\}$ and $\{\hat{A}_n\ \Omega \to \mathbb{R}^{k\times k}\}$ measurable-$F/B(\mathbb{R}^{k\times k})$ such that

$$\tilde{A}_n - A_n^o \to 0 \quad \text{a.s.}$$
$$\hat{A}_n - A_n^* \to 0 \quad \text{a.s.}$$

where $A_n^o \equiv A_n(\theta_n^o)$, $A_n^* \equiv A_n(\theta_n^*)$. □

Assumption PD′ (p. 123)

Assumption PD(i) holds, and

(ii) There exists a closed sphere $S \subset \Theta$ of finite nonzero radius such that for some $\varepsilon > 0$

$$\cup_{n=1}^{\infty} \{\theta \in \Theta : |\theta - \theta_n^*| < \varepsilon\} \subset S$$

and $\{A_n(\theta)\}$ is $O(1)$ and uniformly positive definite uniformly on S. □

Assumption CN (constraint) (p. 123)

Suppose $h : \Theta \to \mathbb{R}^q$, $q \in \mathbb{N}$, is continuously differentiable of order 2 on Θ

with Jacobian $H(\cdot) = \nabla_\theta h(\cdot)$ such that the eigenvalues of $H(\theta)H(\theta)'$ are bounded below on S by $\delta > 0$ and above by $\Delta < \infty$.

For $q = k$, let h be one-to-one with a continuous inverse on S. For $q < k$, suppose there exists $r : \Theta \to \mathbb{R}^{k-q}$ continuous on Θ such that the mapping

$$(\rho', \tau') = (r(\theta)', h(\theta)')$$

has a continuous inverse

$$\theta = \Psi(\rho, \tau)$$

defined over $M = \{(\rho, \tau) : \rho = r(\theta), \tau = h(\theta), \theta \in S\}$. Moreover, $\Psi(\rho, \tau)$ has a continuous extension to the set

$$R \times T = \{\rho : \rho = r(\theta), \theta \in \Theta\} \times \{\tau : \tau = h(\theta), \theta \in \Theta\}. \qquad \square$$

Assumption DR (Pitman drift) (p. 124)

The sequence $\{h_n^o\}$ is chosen such that

$$\sqrt{(n)}(h_n^* - h_n^o) = O(1). \qquad \square$$

Author Index

Subject Index