PART I

Dynamic structural modeling

CHAPTER 1

Efficient instrumental variables estimation of systems of implicit heterogeneous nonlinear dynamic equations with nonspherical errors

Charles Bates and Halbert White

1 Introduction

First used by the zoologist Sewall Wright (1925), the method of instrumental variables (IVs) was initially given a formal development by Reiersøl (1941, 1945) and Geary (1949). It is now one of the most useful and widely applied estimation methods of modern econometrics. Major contributions to the early development of this method in econometrics are those of Theil (1953), Basmann (1957), and Sargan (1958) for systems of linear simultaneous equations with independent identically distributed (i.i.d.) errors. A major concern of this development was to make efficient use of the available instrumental variables by finding instrumental variables estimators that have minimal asymptotic variance.

This concern is also evident in the subsequent work of Zellner and Theil (1962), Brundy and Jorgenson (1974), and Sargan (1964), who consider contemporaneous correlation in linear systems; Sargan (1959), Amemiya (1966), Fair (1970), Hansen and Sargent (1982), and Hayashi and Sims (1983), who consider specific forms of serial correlation for errors of linear equations or systems of equations; and Amemiya (1983), Bowden and Turkington (1985), and White (1984, Chapter 7), who consider specific forms of heteroscedasticity. For systems of linear equations White (1984, Chapter 4; 1986) considers general forms of nonsphericality in an instrumental variables context. Investigation of the properties of instrumental variables estimators in nonlinear contexts was undertaken in seminal work

The authors are grateful to William Barnett, Lars Hansen, Alberto Holly, Whitney Newey, Kenneth Singleton, and the participants of the Austin Symposium in Economic Theory and Econometrics for helpful comments. Errors remain the authors' responsibility. Research support for White was provided by NSF grant SES-8510637.

by Amemiya (1974, 1977) for single equations and for systems of nonlinear equations with contemporaneous correlation. Further important contributions are those of Barnett (1976), Gallant (1977), Gallant and Jorgenson (1979), and Burguete, Gallant, and Souza (1982). Research initiated by Hansen (1982) has recently culminated in a general and elegant theory of efficient instrumental variables estimation for systems of nonlinear dynamic equations with errors exhibiting general forms of conditional nonsphericality (Hansen 1985).

Hansen's results exploit techniques of Gordin (1969) for approximating stationary ergodic sequences with sums of martingale difference sequences and thus apply directly to models relating stationary ergodic processes. Recent work of Eichenbaum and Hansen (1986) extends Hansen's results to allow for the inclusion of deterministic trends. However, Hansen's approach does not readily lend itself to situations involving heterogeneous time series. Because nonexplosive economic time series cannot be guaranteed to be stationary, it is useful to investigate whether results similar to Hansen's hold for heterogeneous time series. We obtain such results here using the method of Bates and White (1987) to obtain the efficient estimator in an appropriate class of instrumental variables estimators. In the general case of heterogeneous time series processes, one can follow the approach of Gallant and White (in press), who obtain results for IV estimators in the context of near-epoch-dependent functions of mixing processes. The near epoch dependence approach provides an approximation procedure for heterogeneous processes analogous to the martingale difference approximation procedure of Gordin (1969) and Hansen (1985) for stationary ergodic sequences. Our results are formulated so as to allow sample-based transformation procedures. Proofs of results are given in the appendix. We illustrate the content of our results with several simple examples.

2 Nonlinear instrumental variables estimators

In this section we consider a class of nonlinear instrumental variables (NLIV) estimators for the parameters of a system of implicit nonlinear dynamic equations with possibly nonspherical errors. We provide conditions ensuring that the NLIV estimators form a regular consistent asymptotically normal indexed (RCANI) class (Bates and White 1987). We then find the efficient NLIV estimator using Corollary 2.11 of Bates and White.

We first provide definitions that make it possible to specify precisely the models of interest to us. We specify the properties of the exogenous variables and errors determining the dependent variables of interest in the following way.

Definition 2.1. Let $(\Omega, \mathfrak{F})$ be a measurable space, and let $e_{nt}: \Omega \to \mathbb{R}^p$ and $X_{nt}: \Omega \to \mathbb{R}^q$ be functions measurable with respect to $\mathfrak{F}$, $t = 1, \dots, n$, $n = 1, 2, \dots$.

We say that $\{(X_{nt}, e_{nt})\}$ *admits a family* $\mathcal{P}$ *of probability measures generating instrumental variable candidates and removable nonsphericalities* if and only if there exists a nonempty collection $\mathcal{P}$ of probability measures P on $(\Omega, \mathfrak{F})$ satisfying the following conditions:

(i) For $t = 1, \dots, n$, $n = 1, 2, \dots$, there exists $W_{nt}: \Omega \to \mathbb{R}^{w_{nt}}$, $w_{nt} \in \mathbb{N}$, measurable-$\mathfrak{F}$ such that with $\mathcal{G}_{nt} \equiv \sigma(W_{nt}) \subset \mathfrak{F}$

$$E(e_{nt} \mid \mathcal{G}_{nt}) = 0 \text{ a.s.-}P$$

and $e_{nth} e_{n\tau g}$ is not measurable-$\mathcal{G}_{nt\tau} \equiv \mathcal{G}_{nt} \vee \mathcal{G}_{n\tau}$ for all $h, g = 1, \dots, p$; and for any $\tilde{\mathcal{G}}_{nt} \supset \mathcal{G}_{nt}$, $\tilde{\mathcal{G}}_{nt} \subset \mathfrak{F}$, either

$$P[E(e_{nt} \mid \tilde{\mathcal{G}}_{nt}) \neq 0] > 0$$

or $e_{nth} e_{n\tau g}$ is measurable-$\tilde{\mathcal{G}}_{nt\tau} \equiv \tilde{\mathcal{G}}_{nt} \vee \tilde{\mathcal{G}}_{n\tau}$ for some n, t, τ, h, g.

(ii) There exists a matrix function $B_n: \Omega \to \mathbb{R}^{pn \times pn}$ measurable-$\mathfrak{F}$ and nonsingular a.s.-P, $n = 1, 2, \dots$, such that with $B_n \equiv [b_{nt\tau}]$, $B_n^{-1} \equiv [b_n^{t\tau}]$, $b_{nt\tau}: \Omega \to \mathbb{R}^{p \times p}$ and $b_n^{t\tau}: \Omega \to \mathbb{R}^{p \times p}$ are measurable-$\mathcal{G}_{nt}$, $\tau = 1, \dots, n$, $n = 1, 2, \dots$, and with $e_n^* \equiv B_n e_n$, where e_n is the $np \times 1$ vector

$$e_n' = (e_{n11}, \dots, e_{n1p}, \dots, e_{nth}, \dots, e_{nn1}, \dots, e_{nnp}),$$

there exists $W_{nt}^*: \Omega \to \mathbb{R}^{w_{nt}^*}$, $w_{nt}^* \in \mathbb{N}$, measurable-$\mathfrak{F}$ such that with $\mathcal{G}_{nt}^* \equiv \sigma(W_{nt}^*)$

$$E(e_{nt}^* \mid \mathcal{G}_{nt}^*) = 0 \text{ a.s.-}P$$

and $e_{nth}^* e_{n\tau g}^*$ is not measurable-$\mathcal{G}_{nt\tau}^* \equiv \mathcal{G}_{nt}^* \vee \mathcal{G}_{n\tau}^*$ for all $h, g = 1, \dots, p$; and for any $\tilde{\mathcal{G}}_{nt}^* \supset \mathcal{G}_{nt}^*$, $\tilde{\mathcal{G}}_{nt}^* \subset \mathfrak{F}$, either

$$P[E(e_{nt}^* \mid \tilde{\mathcal{G}}_{nt}^*) \neq 0] > 0$$

or $e_{nth}^* e_{n\tau g}^*$ is measurable-$\tilde{\mathcal{G}}_{nt\tau}^* \equiv \tilde{\mathcal{G}}_{nt}^* \vee \tilde{\mathcal{G}}_{n\tau}^*$ for some n, t, τ, h, g. Further,

$$\mathcal{G}_{nt} = \bigwedge_{\{1 \le \tau \le n, 1 \le g \le p : P[b_n^{\tau tgh} \neq 0] > 0\}} \mathcal{G}_{n\tau}^*, \quad h = 1, \dots, p,$$

and

$$E(e_{nt}^* e_{n\tau}^{*\prime} \mid \mathcal{G}_{nt\tau}^*) = \Sigma_{nt}^* 1[t = \tau] \text{ a.s.-}P,$$

where Σ_{nt}^* is nonsingular a.s.-P and $P[E(e_{nt}^* e_{n\tau}^{*\prime} \mid \tilde{\mathcal{G}}_{nt\tau}^*) \neq \Sigma_{nt}^*] > 0$ for $\tilde{\mathcal{G}}_{nt}^* \subset \mathcal{G}_{nt}^*$ for some n, t. □

Heuristically, the errors e_{nt} admit instrumental variables candidates W_{nt}, that is, $E(e_{nt} \mid W_{nt}) = 0$, and also possess possible nonsphericalities in that $E(e_{nt} e_{n\tau}' \mid W_{nt}, W_{n\tau}) \neq I_p 1[t \neq \tau]$ with positive probability. These nonsphericalities may include autocorrelation, heteroscedasticity, and/or contemporaneous correlation. At this point, the available number of instrumental variables candidates is not critical. Although there may well be available strictly exogenous random variables $\{X_{nt}\}$ that can serve as instrumental variables in the presence of serially correlated errors, the constant unity [which generates the trivial information set $(\Omega, \varnothing)$] is always available.

The transformation B_n "removes" correlation between the errors in such a way that the elements e^*_{nt} of the transformed errors $e^*_n = B_n e_n$ share a common set of instrumental variables candidates W^*_{nt} and exhibit only possible heteroscedasticity and/or contemporaneous correlation conditional on this set of instrumental variables candidates. The instrumental variables candidates W^*_{nt} may contain more information than W_{nt}. Thus, one may be able to use predetermined rather than only strictly exogenous variables as instrumental variables candidates after a suitable transformation.

In particular contexts (recursive systems of structural equations or panel observations in which dynamics are important), a different set of instrumental variables candidates may be available for each observation. In such cases, the analogous results obtain after an additional transformation, yielding the complete set of instrumental variables candidates W^{**}_{nth}. Alternatively, analogous results obtain by first removing serial correlation (across t) and then applying the theory given here with observations reindexed so that $p=1$. We do not treat this case explicitly because of the unwieldy notation required.

In contrast to the case for the untransformed errors, the information content of the instrumental variables candidates with respect to the transformed errors turns out to be important, as indicated in what follows.

We now specify how the dependent variables Y_{nt} are generated.

Definition 2.2. Let Θ be a compact subset of $\mathbb{R}^k$, $k \in \mathbb{N}$, and let F_{nt}: $\mathbb{R}^{tp} \times \mathbb{R}^{nq} \times \Theta \to \mathbb{R}^p$ define the sequence $Y_{nt}: \Theta \to \mathbb{R}^p$, $t = 1, \ldots, n$, uniquely by the implicit relation

$$e_{nt} = F_{nt}(Y^t_n(\theta), X^n_n, \theta), \quad t = 1, \ldots, n, \;\; n = 1, 2, \ldots,$$

with $X^n_n \equiv (X_{n1}, \ldots, X_{nn})$ and $Y^t_n \equiv (Y_{n1}, \ldots, Y_{nt})$, where F_{nt} is such that there exists G_{nt}: $\mathbb{R}^{tp} \times \mathbb{R}^{nq} \times \Theta \to \mathbb{R}^p$ for which

$$Y_{nt}(\theta) = G_{nt}(e^t_n, X^n_n, \theta), \quad t = 1, \ldots, n, \;\; n = 1, 2, \ldots,$$

where for each θ in Θ, $G_{nt}(\cdot, \cdot, \theta)$ is measurable-$\mathbb{B}$ $(\mathbb{R}^{tp} \times \mathbb{R}^{nq})$.

Then we say that $\{F_{nt}, X_{nt}, e_{nt}\}$ *implicitly determines* $\{Y_{nt}\}$. □

This is a fairly standard system of nonlinear implicit simultaneous equations, except for the dependence of all the functions on n. Note that dynamics are allowed by letting all lags of Y_{nt} appear in F_{nt}. Further dynamics are allowed by letting e_{nt} be autocorrelated. The "strictly exogenous" variables X_{nt} are allowed to enter at all lags and leads, although it is not necessary to have all of these. The index t appearing on G_{nt} and F_{nt} allows given initial values to enter the determination of Y_{nt}, and also allows for the heterogeneity introduced by regime changes.

The model of interest is obtained by letting the underlying probability measure and parameters of interest range over the relevant values.

Definition 2.3. Let $(\Omega, \mathfrak{F})$ be a measurable space and let $\Theta \subseteq \mathbb{R}^k$, $k \in \mathbb{N}$, as in the preceding. Let $\mathcal{P}$ be a family of probability measures on $(\Omega, \mathfrak{F})$.

We say that $(\mathcal{P} \times \Theta)$ *induces a regular probability model* $\mathcal{Q} = \{\mathcal{Q}_n\}$ *for* $\{Y_{nt}\}$ *with instrumental variable candidates and removable nonsphericalities* if and only if $\{Y_{nt}\}$ is determined implicitly by $\{F_{nt}, X_{nt}, e_{nt}\}$ and $\{X_{nt}, e_{nt}\}$ admits the family $\mathcal{P}$ generating instrumental variable candidates and removable nonsphericalities, where for each element (P^o, θ^o) of $\mathcal{P} \times \operatorname{int} \Theta$:

(i) $\epsilon^o_{nt}(\cdot) \equiv F_{nt}(Y^t_n(\theta^o), X^n_n, \cdot)$ is continuously differentiable on Θ a.s.-P^o, $t = 1, \ldots, n$, $n = 1, 2, \ldots$;

(ii) there exist measurable functions $Z_{nth}: \Omega \to \mathbb{R}^k$, $h = 1, \ldots, p$, $t = 1, \ldots, n$, $n = 1, 2, \ldots$, such that the rows $\ddot{Z}_{nth}$ of $\ddot{Z}_n \equiv B_n'^{-1} Z_n$ (where Z_n has rows Z_{nth}) are measurable-$\mathcal{G}^*_{nt}$ and for which

(a) $Z_n' \epsilon^o_n(\theta) - \bar{\psi}^o_n(\theta) \to^{P^o} 0$ uniformly on Θ, where $\bar{\psi}^o_n(\theta) \equiv E^o(Z_n' \epsilon^o_n(\theta)/n)$ is $O(1)$,

(b) $V_n^{o\,-1/2} n^{-1/2} Z_n' e_n \sim^{A^o} N(0, I_k)$, where $V^o_n \equiv \operatorname{var}^o[n^{-1/2} Z_n' e_n]$ is $O(1)$ and uniformly positive definite, and

(c) $Z_n' \nabla \epsilon^o_n(\theta) - \nabla \bar{\psi}^o_n(\theta) \to^{P^o} 0$ uniformly on Θ, where $\nabla \bar{\psi}^o_n(\theta) \equiv E^o(Z_n' \nabla \epsilon^o_n(\theta)/n)$ is $O(1)$;

(iii) for Z_n as in (ii)

(a) $\bar{\psi}^o_n(\theta)$ satisfies the generalized rank condition at θ^o;

(b) $\nabla \bar{\psi}^o_n(\theta^o)$ is nonsingular uniformly in n.

The probability model $\mathcal{Q}_n$ is the set of all probability measures for $\{Y^n_n(\theta), X^n_n\}$ on $(\mathbb{R}^{pn} \times \mathbb{R}^{qn}, \mathbb{B}(\mathbb{R}^{pn} \times \mathbb{R}^{qn}))$ induced by letting P range over $\mathcal{P}$ and θ over $\operatorname{int} \Theta$. The parameters of interest are defined by the mapping $\mathfrak{I}: \mathfrak{M} \to \operatorname{int} \Theta$, where $\mathfrak{M}$ is the space of sequences of probability measures $\mathcal{Q}_n$ on $(\mathbb{R}^{pn} \times \mathbb{R}^{qn}, \mathbb{B}(\mathbb{R}^{pn} \times \mathbb{R}^{qn}))$ for which $\mathfrak{I}(\{\mathcal{Q}_n\}) = \theta$ for any sequence $\{\mathcal{Q}_n\}$ such that $\mathcal{Q}_n$ is induced by a given $\theta \in \Theta$ for any element P in $\mathcal{P}$. □

In this definition, the notations E^o, var^o, and $\sim^{A^o}$ refer to expectation, variance, and convergence in distribution, respectively, under P^o. The "regularity" imposed in this definition is a set of conditions ensuring that there exists a consistent, asymptotically normal NLIV estimator

$$\hat{\theta}_{n,\text{NLIV}} \equiv \text{argmin } \epsilon_n(\theta)' Z_n V_n^- Z_n' \epsilon_n(\theta),$$

where $\epsilon_n(\theta) \equiv [F_{nt}(Y_n^t, X_n^n, \theta)]$ and V_n^- is a generalized inverse for V_n (e.g., the Moore–Penrose inverse) provided that V_n has the property that $V_n - V_n^o \to^{P^o} 0$ for all (P^o, θ^o) in $\mathcal{P} \times \text{int } \Theta$.

This NLIV estimator is easily recognized as a particular form of the generalized method of moments (GMM) estimator (Hansen 1982), arrived at by focusing on moment conditions $E^o(Z_n' e_n) = 0$ and by explicitly using V_n^- in forming the objective function defining the estimator. Use of V_n^- yields a minimum χ^2 estimator, which is known to give the efficient estimator for a given choice of instruments (Hansen 1982; White 1982; Bates and White 1987). These restrictions allow us to focus attention directly on choosing the instruments.

The instruments of interest are specified in condition 2.3(ii) as satisfying particular measurability requirements. To indicate the possibilities allowed by this condition, we give the following result.

Proposition 2.4. *Let $Z_n = A_n S_n$, where A_n is an $np \times np$ matrix with $A_n \equiv [a_{nt\tau}]$, $a_{nt\tau} \equiv [a_{nt\tau hg}]$, such that $a_{nt\tau hg}$ is measurable-$\mathcal{G}_{n\tau}$, $h, g = 1, \ldots, p$, $t = 1, \ldots, n$, and S_n is an $np \times k$ matrix with typical row S_{nth}.*

Suppose either that $A_n = B_n'$ and S_{nth} is measurable-$\mathcal{G}_{nt}^*$ or that S_{nth} is measurable-$\mathcal{G}_{nt}$ and $C_n \equiv B_n'^{-1} A_n$ is such that $P[c_{nt\tau hg} \neq 0] > 0$ if and only if $P[b_n^{\tau tgh} \neq 0] > 0$, where

$$C_n \equiv [c_{nt\tau}], \qquad c_{nt\tau} \equiv [c_{nt\tau hg}], \qquad c_{nt\tau hg} = \sum_{\lambda=1}^{n} \sum_{\gamma=1}^{p} b_n^{\lambda t \gamma h} a_{n\lambda\tau\gamma g}.$$

Then $\ddot{Z}_n = B_n'^{-1} Z_n$ has rows $\ddot{Z}_{nth}$ that are measurable-$\mathcal{G}_{nt}^*$, $h = 1, \ldots, p$, $t = 1, \ldots, n$, $n = 1, 2, \ldots$. □

This proposition establishes that instruments satisfying 2.3(ii) may be formed either as specific linear combinations of functions of the instrumental variables candidates compatible with the transformed errors $e_n^* = B_n e_n$ ($A_n = B_n'$, S_{nth} measurable-$\mathcal{G}_{nt}^*$) or as linear combinations of functions of the instrumental variables candidates compatible with the untransformed errors. A key restriction in this latter case is that C_n must have zeros in the same locations as B_n'. In particular, if B_n is lower triangular, C_n must be upper triangular. This is guaranteed if A_n is upper triangular.

Primitive conditions ensuring the convergences in condition 2.3(ii.a–c) will differ from case to case depending on the behavior of the specific stochastic processes involved. The following result uses definitions and results of Gallant and White (in press) to give primitive conditions ensuring 2.3(ii.a–c) in the general context of dependent heterogeneous double arrays.

Proposition 2.5. *Let $\mathcal{P}$ and Θ be as previously given, let $\{F_{nt}, X_{nt}, e_{nt}\}$ implicitly determine $\{Y_{nt}\}$, and let $V_t: \Omega \to \mathbb{R}^v$, $v \in \mathbb{N}$, be measurable-$\mathfrak{F}$, $t = 1, \ldots$. Suppose condition* 2.4(i) *holds and that for all P^o in $\mathcal{P}$ $\{V_t\}$ is a mixing process with either ϕ_m of size $-r/(r-1)$ or α_m of size $-2r/(r-2)$, $r > 2$. If for each (P^o, θ^o) in $\mathcal{P} \times \text{int}\,\Theta$ the elements of $Z_{nth}'\epsilon_{nth}^o(\theta)$ and of $Z_{nth}'\nabla\epsilon_{nth}^o(\theta)$ are*

- (a) *r-dominated on Θ with respect to P^o uniformly in $t = 1, \ldots, n$, $n = 1, 2, \ldots$, $r > 2$ for $h = 1, \ldots, p$;*
- (b) *Lipschitz-L_1* a.s.-*P^o;*
- (c) *near-epoch-dependent-P^o on $\{V_t\}$ of size -1 uniformly on $(\Theta, |\cdot|)$; and*
- (d) *$V_n^o \equiv \text{var}^o(n^{-1/2} Z_n' e_n)$ is $O(1)$ and uniformly positive definite,*

then conditions 2.3(ii.a–c) *hold.* □

In less general settings, one may be able to state weaker moment and/or smoothness conditions ensuring 2.3(ii.a–c).

Conditions 2.3(iii.a, b) are fundamental identification conditions for θ^o [see Bates and White (1985) for the generalized rank conditions] ensuring that there are "enough" instrumental variables.

In general, there will exist more than one consistent asymptotically normal NLIV estimator. We consider the following class of all such NLIV estimators.

Definition 2.6. Let $(\mathcal{P} \times \Theta)$ induce a regular probability model $\mathcal{Q}$ for $\{Y_{nt}\}$ with instrumental variables candidates and removable nonsphericalities.

Let Γ_n be the set of all functions $\hat{Z}_n: \Omega \to \mathbb{R}^{pn \times k}$ measurable-$\mathfrak{F}$, $n = 1, 2, \ldots$, and denote a typical element of Γ_n as $\gamma_n = \hat{Z}_n$.

The class of regular NLIV estimators $\mathcal{E}_{\text{NLIV}}(\Gamma, \mathcal{Q}, \mathfrak{J})$ is the set of all sequences $\hat{\theta}_{\text{NLIV}} = \{\hat{\theta}_n(\gamma_n, \cdot): \Omega \to \mathbb{R}^k\}$ such that

$$\hat{\theta}_n(\gamma_n, \cdot) \equiv \text{argmin}_\Theta\, \epsilon_n(\theta)' \hat{Z}_n \hat{V}_n(\gamma_n)^- \hat{Z}_n' \epsilon_n(\theta)$$

for $\gamma \equiv \{\gamma_n\} \in \Gamma \subseteq \Gamma^\infty \equiv \times_{n=1}^{\infty} \Gamma_n$, where Γ is the set of all sequences γ in Γ^∞ such that for all (P^o, θ^o) in $\mathcal{P} \times \text{int}\,\Theta$

(i) $n^{-1}(\hat{Z}_n' \epsilon_n^o(\theta) - Z_n' \epsilon_n^o(\theta)) \to^{P^o} 0$ uniformly on Θ;
(ii) $n^{-1}(\hat{Z}_n' \nabla \epsilon_n^o(\theta) - Z_n' \nabla \epsilon_n^o(\theta)) \to^{P^o} 0$ uniformly on Θ;
(iii) $n^{-1/2}(\hat{Z}_n' e_n - Z_n' e_n) \to^{P^o} 0$; and
(iv) $\hat{V}_n(\gamma_n)^- - V_n^{o-1} \to^{P^o} 0$,

where $\{Z_n : \Omega \to \mathbb{R}^{pn \times k}\}$ is a sequence satisfying conditions 2.3(ii) and 2.3(iii) and $\hat{V}_n(\gamma_n) : \Omega \to \mathbb{R}^{k \times k}$ is measurable-$\mathfrak{F}$ for each $\gamma_n \in \Gamma_n$, $n = 1, 2, \dots$. □

The convergence conditions 2.6(i)–(iv) permit sample-based procedures in which feasible NLIV estimators are constructed in several stages. A typical situation occurs when parametric models for the conditional covariance structure of e_n are posited and estimated, as in the method of three-stage least squares or as in the heteroscedastic error model of White (1980), in which variances differ across cross-sectional groups. Unknown parameters may also be embedded in the expression for the desired instruments, as in the binary endogenous variable example of White (1986). The present conditions allow for their estimation in a prior stage as well.

Although parametric models are often useful in prior stages, it may be possible to use nonparametric techniques in arriving at $\hat{Z}_n$, as in Newey (1986) and Robinson (1987). Given the usual stage of ignorance in economics, such techniques might be very useful indeed. Conditions 2.6(i)–(iv) permit the use of appropriately regular nonparametric methods in constructing instruments.

Essentially, these conditions ensure that $\hat{Z}_n$ functions in a manner asymptotically equivalent to non-sample-based instruments Z_n, constructed in a manner compatible with the conditions of Definition 2.3. Note that a given sequence $\{Z_n\}$ is associated with a particular element γ of Γ, although we usually leave this dependence implicit for notational convenience.

The conditions of Definition 2.6 ensure that the class of NLIV estimators is a regular consistent asymptotically normal indexed (RCANI) class, as defined by Bates and White (1987, Definition 2.3). This is formally stated as follows.

Theorem 2.7. *The class of regular NLIV estimators* $\mathcal{E}_{\text{NLIV}}(\Gamma, \mathcal{Q}, \mathfrak{J})$ *is a RCANI class, with*

$$C_n^o(\gamma)^{-1/2}(\hat{\theta}_n(\gamma_n, \cdot) - \theta^o) \overset{A^o}{\sim} N(0, I_k),$$

where

$$C_n^o(\gamma) \equiv \nabla\tilde{\psi}_n^o(\gamma)'^{-1} V_n^o(\gamma)\, \nabla\tilde{\psi}_n^o(\gamma)^{-1},$$

$$\nabla\tilde{\psi}_n^o(\gamma) \equiv E^o(Z_n(\gamma)'\, \nabla\epsilon_n^o(\theta^o)/n). \qquad \square$$

Bates and White (1987) define the asymptotically efficient estimator in a RCANI class as one having a maximal concentration property (their Definition 2.7) and show that this is equivalent to having the smallest asymptotic covariance matrix (their Theorem 2.8) in the usual sense. Theorem 2.9 and Corollary 2.11 of Bates and White (1987) make it possible to find the efficient estimator in a RCANI class. In the present case we have the following result.

Theorem 2.8. *Suppose that* $\hat{\theta}^*_{\mathrm{NLIV}} \in \mathcal{E}_{\mathrm{NLIV}}(\Gamma, \mathfrak{Q}, \mathfrak{I})$, *where* $\hat{\theta}^*_{\mathrm{NLIV}} = \{\hat{\theta}^*_n\}$,

$$\hat{\theta}^*_n \equiv \mathrm{argmin}_{\Theta}\, \epsilon_n(\theta)'\hat{Z}^*_n \hat{V}^{*-}_n \hat{Z}^{*\prime}_n \epsilon_n(\theta)$$

with $\hat{Z}^*_n$ *such that conditions* 2.6(i)–(iv) *hold with* $Z^*_n \equiv [Z^*_{nt}]$ *given by*

$$Z^*_{nt} = \sum_{\tau=1}^{n} b_{n\tau t}\, \Sigma^{*-1}_{n\tau} E^o(\nabla\epsilon^*_{n\tau}(\theta^o) \mid \mathcal{G}^*_{n\tau}),$$

where $\nabla\epsilon^*_n(\theta^o) \equiv [\nabla\epsilon^*_{nt}(\theta^o)] = B_n \nabla\epsilon^o_n(\theta^o)$.

Then $\hat{\theta}^*_{\mathrm{NLIV}}$ *is the efficient NLIV estimator. In particular,*

$$\mathrm{avar}^o\, \hat{\theta}^*_{\mathrm{NLIV}} = E^o(\nabla\epsilon^o_n(\theta^o)'\Omega_n^{-1}\nabla\epsilon^o_n(\theta^o)/n)^{-1},$$

where $\Omega_n \equiv B_n^{-1}\Sigma^*_n B_n'^{-1}$. $\square$

The formula for $\mathrm{avar}^o\, \hat{\theta}^*_{\mathrm{NLIV}}$ gives an asymptotic variance bound identical to that found by Hansen (1985). Whereas Hansen's result applies to stationary ergodic processes, the present results apply to heterogeneous and/or dependent sequences. Note the clear similarity to the formula for the variance of the GLS estimator [replace $\nabla\epsilon^o_n(\theta^o)$ with X].

An efficient estimator that is not necessarily feasible obtains by setting $\hat{Z}^*_n = Z^*_n$, $\hat{V}^*_n = V^o_n$. Leaving feasibility aside for the moment, we see that in the context of Proposition 2.4, the best choice for A_n is B_n', which removes correlations in e_n, and the best choice for $S_{n\tau}$ is $\Sigma^{*-1}_{n\tau}E(\nabla\epsilon^*_{n\tau}(\theta^o) \mid \mathcal{G}^*_{n\tau})$, which is measurable-$\mathcal{G}^*_{n\tau}$, as required. Note that $\Sigma^{*-1}_{n\tau}$ is acting to remove heteroscedasticity and/or contemporaneous correlation conditional on the instrumental variables candidates for the transformed series e^*_n.

A useful feature of the present result is that it specifies (an equivalence class of) feasible estimators that may achieve the efficiency bound. We see how this is possible in the next section.