Efficient Semiparametric Estimation of Expectations in Dynamic Nonlinear Systems*

Bryan W. Brown
Department of Economics
Rice University

and

Byung M. Jeon
Department of Economics
Rice University

Abstract

Semiparametric estimation of the expectations of a general class of dynamic functions is considered. Such expectation functionals that are of interest for dynamic models are one- and multi-period ahead forecasting functions, distribution functions, and covariance matrices. The semiparametric efficiency bound for this problem is established and an estimator which attains the bound is developed. The explicit form of the semiparmetric efficient expectation estimator is worked out for several explicit assumptions regarding the degree of dependence between the predetermined variables and the disturbances of the model. Under the assumption of independence, the one- and multi-period ahead residual-based predictors proposed by Brown and Mariano (1989) are shown to be semiparametric efficient. Under unconditional mean zero assumption, we propose an improved heteroskedastic autocorrelation consistent estimator.

Key words: Dynamic nonlinear systems, estimation of expectations, semiparametric efficiency bound, residual-based estimation.
* Earlier drafts of this paper were given at the 2001 Texas Econometrics Camp and the 2001 North American Winter Meetings of the Econometric Society.

1 Introduction

Models of expectations have an important role in econometric theory and applications because of a need to estimate functions which can be represented as conditional or unconditional expectations. These include point predictions and any target function which results from the integration of random variables, for example, distribution functions, covariance matrices.

In a fully parametric models, where both the structural equations and distribution function are specified, the estimation of an expectation function is simple. When the target expectation function is available in closed form, the problem can be solved by substituting the estimated parameters into the expectation function. If the expectation functions is not available in closed form, then the target function can be estimated by averaging a sample of Monte Carlo stochastic simulations of the random variables in the model.

However, semiparametric models, models which have both parametric and nonparametric components, have received increasing attention because of the problem of misspecification and measurement error. The semiparametric approach to those problems is to allow the functional form of some components of the model to be unrestricted. Specifically, we want to look at a semiparametric model with specified structure and unrestricted error distribution except for general restrictions such as unconditional and conditional mean zero. In reality, we do seldom have a complete specification of the distribution of these models except for general restrictions such as unconditional or conditional mean zero. So it is interesting to take a semiparametric approach which allows us to take advantage of the specified structural equations.

For the expectations of static systems, Brown and Mariano (1984) propose a residual-based simulation procedure as an alternative to Monte Carlo simulation for point prediction problem. This procedure avoids the need to specify an explicit distribution for the disturbances. Brown and Newey (1998) establish the semiparametric efficiency bounds of target functions and suggest a feasible estimator which attains the bounds for unconditional expectations. For the expectations of dynamic nonlinear systems, Brown and Mariano (1989) shows that the residual-based predictors are quite promising alternative to Monte Carlo simulation when applied to conditional prediction problems under independence assumption between predetermined variables and error distribution.

In this paper, we propose general procedures for optimal estimation of expectation functions in nonlinear dynamic models under relaxed distributional assumptions. This also extends Brown and Newey (1998) to the class of dynamic nonlinear systems. An important application for the systems is in the construction of *ex ante* predictions. Typically, the system is dynamic with either lagged endogenous variables or serially correlated disturbances (or both). We develop semiparametric efficiency bounds and an estimator which achieves the bound for the expectation of dynamic nonlinear systems. The explicit form of the semiparametric efficient expectation estimator is worked out for several explicit assumptions regarding the degree of dependence between the predetermined variables and the disturbances of the model.

2 Model

Suppose that we observe the stationary and ergodic $((g+k) \times 1)$ vector $z_t = (y_t', x_t')'$ for $t = 1, \ldots, n.$ and the following dynamic nonlinear equations holds:

$$\rho(z_t, z_{t-1}, \ldots, z_{t-l}; \beta) = \varepsilon_t \tag{1}$$

where ρ is a possibly non-linear relationship of known functional form, ε_t is a $g \times 1$ disturbances vector which is possibly serially correlated, and β is an unknown p-dimensional vector of parameters of interest with true value β_0. We assume that we can invert the function ρ and solve for y_t in terms of the following reduced form model:

$$y_t = \pi(\varepsilon_t, z_{t-1}, \ldots, z_{t-l}, x_t; \beta) \tag{2}$$

We can simplify above models as

$$\begin{aligned} \rho(y_t, w_t, \beta) &= \varepsilon_t & (3) \\ y_t &= \pi(\varepsilon_t, w_t, \beta) & (4) \end{aligned}$$

where $w_t = (y_{t-1}', \ldots, y_{t-l}', x_t', x_{t-1}', \ldots, x_{t-l}')'$ and $\beta \in int\ \Theta \subset R^p$.

The problem studied in this paper is efficient estimation of expectation of a known dynamic function of the observable variables when the density and parameters are unknown. Formally, the functional which is the object of interest has the following representation

$$\begin{aligned} \mu(\beta, h) &= E_{\beta,h}\left[m(z_t, z_{t-1}, \ldots, z_{t-v}, \beta)\right] & (5) \\ &= \int m\left(z_t^{-v}, \beta\right) \Pi_{r=0}^{\infty} f(z_{t-r}|z_{t-r}^{-l}; \beta, h) dz_{t-r} \end{aligned}$$

where $m(\cdot)$ is a known $(q \times 1)$ function and $f(\cdot)$ is a density function with respect to some measure. Let $\nu = \max\{l, v\}$ and redefine w_t be $(y_{t-1}', \ldots, y_{t-\nu}', x_t', x_{t-1}', \ldots, x_{t-\nu}')'$. Then the (5) will be $E_{\beta,h}\left[m(y_t, w_t, \beta)\right]$. The presence of h allow the form of $f(\cdot)$ to be unrestricted except for general restrictions on densities. Lots of expectation functionals can be included in general framework of (5). Under some regularity conditions and independence between ε and w, the conditional expectation of y given w_τ is the unconditional expectation of $m(y, w_\tau, \beta) = \pi(\rho(y, w, \beta), w_\tau, \beta)$ where w_τ is treated as fixed. This also applies to conditional covariance matrix, $m(y, w_\tau, \beta) = \pi(\rho(y, w, \beta), w_\tau, \beta) \cdot \pi(\rho(y, w, \beta), w_\tau, \beta)'$, and conditional distribution function, $m(y, w, \beta) = 1(\pi(\rho(y, w, \beta), w_\tau, \beta) \leq c)$. Under nonindependence assumption, conditional expectation becomes difficult but a number of interesting unconditional expectations remain. covariance matrix of the disturbances, distribution function of the disturbances, and distribution function of the observable variables are unconditional expectations and plays an important inferential role in the model. These are the expectations of $m(y, w, \beta) = \rho(y, w, \beta) \cdot \rho(y, w, \beta)'$, $m(y, w, \beta) = 1(\rho(y, w, \beta) \leq c)$, and $m(y, w, \beta) = 1(\pi(\rho(y, w, \beta), w, \beta) \leq c)$, respectively.

3 The semiparametric efficiency bound

Semiparametric efficiency bound is developed by Stein (1956), Kposhevnik and Levi (1976), Pfanzagl and Wefelmeyer (1982), Begun et al. (1983), and Bickel et al. (1992). We can define a parametric submodel, $f(z|z^{-l};\beta_i, h(\eta_i))$, where $z^{-l} = (z_{-1}, \ldots, z_{-l})$ and η_i is a finite-length vector of shape parameters for parametric submodel i, that satisfies the semiparametric assumptions and contains the truth. $f(z|z^{-l};\beta_0, h_0)$ is the true density of z where a zero subscript indicates the true parameter value, $h_0 = h(\eta_0)$. The 'sub' prefix refers that it is a subset of the model consisting of all distributions satisfying the assumptions. For each parametric submodel, we can obtain classical Cramer-Rao bound. Any consistent and asymptotically normal semiparametric estimator has an asymptotic variance that is comparable to the Cramer-Rao bound of a parametric submodel. It is no smaller than the bound for a parametric submodel. Since a semiparmetric model can be represented by infinite number of parametric submodels, the asymptotic variance of any semiparametric estimator is no smaller than the supremum of the Cramer-Rao bounds for all parametric submodels. The supremum of the Cramer-rao bounds is a lower bound on the asymptotic variance of any semiparametric estimator. Some regularity conditions are necessary to guarantee that the Cramer-Rao bound is well-defined and gives an asymptotic efficiency bound. The regularity conditions for parametric submodels are mean-square differentiability with respect to $\theta = (\beta', \eta')'$, nonsingular information matrix, and some additional smoothness conditions, for example existence of variance. A regular estimator is one that the limiting distribution of $\sqrt{n}(\widehat{\beta} - \beta_n)$ does not depend on a sequence of true parameter values $\theta_n = (\beta_0', \eta_0')' + \xi/\sqrt{n}$, where $\xi > 0$. The precise definition of the efficiency bound is that it is the supremum of the Cramer-Rao bounds of all regular parametric submodels. The class of regular estimators also excludes superefficient estimators that has an asymptotic variance less than that of the maximum-likelihood estimator for some true parameter values.

Assuming that we have some initial observations $\{z_0, z_{-1}, \ldots, z_{-l}\}$, and that the density of these initial conditions is asymptotically negligible in the analysis of the likelihood function. The parametric submodel also satisfies mean-square continuous differentiability of the square-root of the likelihood function and has a nonsingular information matrix. We can base our derivation of semiparametric efficiency bound on the analysis of the following likelihood for $\{z_t\}_{t=1}^n$, conditioning on the initial conditions:

$$\mathcal{L}\left(\{z_t\}_{t=1}^n, \beta, h\right) = \Pi_{t=1}^n f_{z|z^{-l}}(z_t|z_t^{-l};\beta, h) \tag{6}$$

We are assuming that the density $f_{z|z^{-l}}$ is unknown, so we have written the likelihood for a parametric submodel in which h represents some parameterization that contains the true densities. Let s_θ denote the score for θ, then we have $\sum_{t=1}^n s_\theta(y_t, w_t) = \sum_{t=1}^n \partial \ln f_{z|z^{-l}}(z_t|z_t^{-l};\theta_0)/\partial\theta'$ where $w_t = (x_t', z_{t-1}', \ldots, z_{t-l}')$. Note that we have to use the sum of scores in the followings because of possible correlation among scores. In parametric estimation theory, local asymptotic normality (LAN) condition played a very important rule in establishing general lower bounds on the accuracy of estimates (Le Cam (1972) and Hajek (1972)). Levit (1975) applies this concept for nonparametric estimation theory. We use Ibragimov and Khas'minskii (1991) LAN condition[1] to establish the semiparametric efficiency bound. Given notations, we can state the following result.

Lemma 3.1 *Assume that* $f_{z|z^{-l}}(z_t|z_t^{-l};\beta,h)$ *is three times differentiable with respect to* β *and has bounded third derivative, that* $E\left[\frac{1}{n}\sum_{t=1}^{n}s_\theta(y_t,w_t)\sum_{s=1}^{n}s_\theta(y_s,w_s)'\right]$ *is finite. Then the data* $\{z_t\}_{t=1}^{n}$ *falls into locally asymptotically normal (LAN) family.*

Lemma 3.2 *Assume that* $\{z_t\}_{t=1}^{n}$ *is* $\rho-$*mixing*[2] *process. Then* $E\left[p(z_t)\sum_{s=-\infty}^{n}q(z_s)'\right]$ $= E\left[\frac{1}{n}\sum_{t=1}^{n}p(z_t)\sum_{s=1}^{n}q(z_s)'\right]+o(1)$ *where* $p(\cdot)$*.and* $q(\cdot)$ *are functions of their argument.*

Let $E_n[\cdot]$ denote the expectation taken at θ_n, *i.e.*, $\mu_n = E_n[m(y,w,\beta)]$. Define an estimator $\widehat{\mu}$ to be asymptotically linear if it is asymptotically equivalent to a sample average, i.e., there is a function $\psi_\mu(y,w)$ at the truth

$$\sqrt{n}(\widehat{\mu}-\mu_0)=\sqrt{n}\overline{\psi}_\mu+o_p(1),\ \overline{\psi}_\mu=\frac{1}{n}\sum_{t=1}^{n}\psi_\mu(y_t,w_t) \tag{7}$$

where $E\left[\sqrt{n}\overline{\psi}_\mu\right]=0$, and $E\left[n\overline{\psi}_\mu\overline{\psi}_\mu'\right]$ is finite and nonsingular. For a matrix A let $\|A\|\equiv[trace(A'A)]^{1/2}$.

Lemma 3.3 *Assume that the data* $\{z_t\}_{t=1}^{n}$ *is LAN family and* $\rho-$*mixing process, that* $\widehat{\mu}$ *is asymptotically linear and for all regular parametric submodels* $\mu(\theta)$ *is differentiable, and that* $E_\theta\left[\left\|\sqrt{n}\overline{\psi}_\mu\right\|^2\right]$ *exist and is continuous on a neighborhood of* θ_0*. Then* $\widehat{\mu}$ *is regular if and only if, for all regular parametric submodels,*

$$\frac{\partial\mu(\theta)}{\partial\theta'}=E\left[n\overline{\psi}_\mu\overline{s}_\theta'\right]+o_p(1) \tag{8}$$

Equation (8) is fundamental in the sense that it gives an important formula for the Cramer-Rao bound of a parametric submodel j. The Cramer-Rao bound for θ is $\left(E\left[n\overline{s}_\theta^j\overline{s}_\theta^{j\prime}\right]\right)^{-1}$. Furthermore, if $\mu(\theta)$ is differentiable, then the Cramer-Rao bound for μ is $V_\mu^j=\left(\partial\mu(\theta_0)/\partial\theta'\right)\left(E\left[n\overline{s}_\theta^j\overline{s}_\theta^{j\prime}\right]\right)^{-1}(\partial\mu(\theta_0)/\partial\theta)$ by the invariance of maximum-likelihood and delta method. It can be rewritten as follows

$$\begin{aligned} V_\mu^j &= E\left[n\overline{\psi}_\mu\overline{s}_\theta^{j\prime}\right]\left(E\left[n\overline{s}_\theta^j\overline{s}_\theta^{j\prime}\right]\right)^{-1}E\left[n\overline{s}_\theta^j\overline{\psi}_\mu{}'\right] \\ &= E\left[\sqrt{n}\overline{\psi}_{\mu,\theta}\sqrt{n}\overline{\psi}_{\mu,\theta}'\right] \text{ where } \sqrt{n}\overline{\psi}_{\mu,\theta}=E\left[n\overline{\psi}_\mu\overline{s}_\theta{}^{j\prime}\right]\left(E\left[n\overline{s}_\theta^j\overline{s}_\theta^{j\prime}\right]\right)^{-1}\sqrt{n}\overline{s}_\theta^j \end{aligned}$$

[1] For formal definition, see Ibragimov and Khas'minskii (1991 p.1682)

[2] If $lim_{n\to\infty}\ \rho_n=0$, then the sequence $\{\xi_n\}$ is defined to be ρ-mixing where $\rho_n=\sup\{|E[\xi\eta]|\,\xi\in\mathcal{F}_k,E[\xi]=0,\|\xi\|\leq 1,\eta\in\mathcal{G}_{k+n},E[\eta]=0,\|\eta\|\leq 1\}$ $\mathcal{F}_n=\sigma\{\xi_k:k\leq n\},\mathcal{G}_n=\sigma\{\xi_k:k\geq n\}$.

This is nothing but the variance matrix of a projected value from the population regression of $\sqrt{n}\overline{\psi}_{\mu}$ on the sum of scores. For the efficiency bound of semiparametric estimation, i.e., supremum of Cramer-Rao bound of all parametric submodels, we define the complete tangent set $\mathcal{S}_q$ to be the mean square closure of all $q-$dimensional linear combinations of scores s_θ for smooth parametric submodels.

$$\mathcal{S}_q = \left\{ \omega_\theta \in \mathsf{R}^q : E\left[\left\|\omega_\theta\right\|^2\right] < \infty, \exists\ A_j, \sqrt{n}\overline{s}_\theta^j \text{ with } \lim_{j\to\infty} E\left[\left\|\omega_\theta - A_j\sqrt{n}\overline{s}_\theta^j\right\|^2\right] = 0 \right\} \tag{9}$$

where A_j is a constant matrix with q rows. Since the tangent set is an infinite-dimensional set that includes all parametric submodels, the projected value from $n\overline{\psi}_{\mu}$ on the tangent set should have larger variance than the projected value for any parametric submodel. Let $\sqrt{n}\overline{\psi}_{\mu}^* = \text{Proj}\left(\sqrt{n}\overline{\psi}_{\mu}\,|\mathcal{S}_q\right)$, then we can write

$$\sqrt{n}\overline{\psi}_{\mu} = \sqrt{n}\overline{\psi}_{\mu}^* + \sqrt{n}\overline{\xi} \tag{10}$$

where $\sqrt{n}\overline{\xi}$ is orthogonal to the tangent set $\mathcal{S}_q$. Equation (8) will be

$$\frac{\partial\mu(\theta)}{\partial\theta'} = E\left[n\overline{\psi}_{\mu}^*\overline{s}_{\theta j}{}'\right] + o_p(1) \text{ for all } j. \tag{11}$$

Furthermore, the linearity of the tangent set gives us a unique projection onto the tangent set, $\sqrt{n}\overline{\psi}_{\mu}^*$, for all asymptotically linear and regular semiparametric estimators. Let μ^+ be other aymptotically linear regular estimator, so $\sqrt{n}(\mu^+ - \mu_0) = \sqrt{n}\overline{\psi}_{\mu}^+(y_t, w_t) + o_p(1) \to_d N(0, V_{\mu}^+)$. By linearity of the tangent set $\left(\sqrt{n}\overline{\psi}_{\mu}^+ - \sqrt{n}\overline{\psi}_{\mu}^*\right) = \sqrt{n}\overline{\psi}_{\mu}^+ - \text{Proj}\left(\sqrt{n}\overline{\psi}_{\mu}^+\,|\mathcal{S}_q\right) + \text{Proj}\left(\sqrt{n}\overline{\psi}_{\mu}^+ - \sqrt{n}\overline{\psi}_{\mu}\,|\mathcal{S}_q\right)$. Also, by the regularity two estimators satisfy $\partial\mu(\theta)/\partial\theta' = E\left[n\overline{\psi}_{\mu}\overline{s}_{\theta}{}^{j\prime}\right] + o_p(1) = E\left[n\overline{\psi}_{\mu}^+\overline{s}_{\theta}{}^{j\prime}\right] + o_p(1)$. So that $E\left[\left(\sqrt{n}\overline{\psi}_{\mu}^+ - \sqrt{n}\overline{\psi}_{\mu}\right)\sqrt{n}\overline{s}_{\theta}{}^{j\prime}\right] + o_p(1) = 0$ for all j. We have a unique projection $\sqrt{n}\overline{\psi}_{\mu}^*$ for all asymptotically linear regular estimators. Combining the previous results, we obtain the efficiency bound for estimation of $E_{\beta,h}\left[m(y,w,\beta)\right]$ as follows.

Theorem 3.1 *Assume that the data $\{z_t\}_{t=1}^n$ is LAN family and $\rho-$mixing process, that $\widehat{\mu}$ is regular and asymptotically linear, that for all regular parametric submodels $\mu(\theta)$ is differentiable, $E_\theta\left[\left\|\sqrt{n}\overline{\psi}_{\mu}\right\|^2\right]$ exists, and is continuous on a neighborhood of θ_0, and that $\mathcal{S}_q$ is linear and $E\left[n\overline{\psi}_{\mu}^*\overline{\psi}_{\mu}^{*\prime}\right]$ is nonsingular for the projection $\sqrt{n}\overline{\psi}_{\mu}^*$ of $\sqrt{n}\overline{\psi}_{\mu}$ on $\mathcal{S}_q$. Then $V_{\mu}^* = E\left[n\overline{\psi}_{\mu}^*\overline{\psi}_{\mu}^{*\prime}\right]$.*

4 Calculating the efficiency bound

Using the structure of our model, the likelihood function, equation (6), can be rewritten in terms of the density of ε, the change of variables formula gives us:

$$\mathcal{L}\left(\{z_t\}_{t=1}^n, \beta, h\right) = \Pi_{t=1}^n \widetilde{J}(y_t, w_t, \beta) \cdot f_{\varepsilon|w}(\varepsilon_t|w_t; h) \cdot f_{x|z^{-l}}(x_t|z_t^{-l}; h) \tag{12}$$

where $\widetilde{J}(y,w,\beta) = \left|\det \frac{\partial \rho(y,w,\beta)}{\partial y'}\right|$.

We can then write the scores of the likelihood for all observations (y,x) with respect to β and η as

$$\begin{aligned}
\sum_{t=1}^{n} s_\beta(y_t,w_t) &= \sum_{t=1}^{n}\left\{J_\beta(y_t,w_t,\beta_0) + \frac{\partial \ln f_{\varepsilon|w}(\varepsilon_t|w_t;h_0)}{\partial \varepsilon}\cdot \rho_\beta(y_t,w_t,\beta_0)\right\} \\
\sum_{t=1}^{n} s_\eta(y_t,w_t) &= \sum_{t=1}^{n}\left\{\frac{\partial \ln f_{\varepsilon|w}(\varepsilon_t|w_t;h_0)}{\partial \eta} + \frac{\partial \ln f_{x|z^{-l}}(x_t|z_t^{-l};h_0)}{\partial \eta}\right\}
\end{aligned} \tag{13}$$

where $J(y,w,\beta) = \ln \widetilde{J}(y,w,\beta)$ and the β subscripts on J and ρ denote partial derivatives. Note that the terms in sum of nuisance scores $\sum_{t=1}^{n} s_\eta(y_t,w_t)$ are unrestricted (except for the zero mean property of sum of scores) functions of their arguments, ε_t and (x_t, z_t^{-l}).

To get efficient scores, we need to orthogonalize $\sqrt{n}\bar{s}_\beta$ to the nuisance scores $\sqrt{n}\bar{s}_\eta$, where normalization term, $n^{-1/2}$, is required to satisfy finiteness of the second moments. Since the nuisance scores could come from any parametric submodel that includes the truth, this requires orthogonalization with respect to the space spanned by suitable linear transformations of all, in a sense that includes all parametric submodels, nuisance parameter scores. This space is known as the tangent set and is the linear Hilbert space given by

$$\mathcal{T}_q = \left\{ t = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{t_1(\varepsilon_t,w_t) + t_2(x_t,z_t^{-l})\right\} : E\left[t\right] = 0 \right\} \tag{14}$$

Since $t_1(\varepsilon_t,w_t)$ is an unrestricted function of its arguments, any functions orthogonal to the set of $t_1(\varepsilon_t,w_t)$ must also be orthogonal to the set of $t_2(x_t,z_t^{-l})$. The efficient score is given by the residual of $\sqrt{n}\bar{s}_\beta$ less its projection on this space, which, for arbitrary function $R(z)$, is given by

$$\operatorname{Proj}\left(R(z)|\mathcal{T}_q\right) = E\left[R(z)\,\middle|\sqrt{n}\bar{t}_1\right] - E\left[R(z)\right] \tag{15}$$

By the mean zero property of scores we can show that $\operatorname{Proj}\left(\sqrt{n}\bar{s}_\beta|\mathcal{T}_q\right) = E\left[\sqrt{n}\bar{s}_\beta|\sqrt{n}\bar{t}_1\right]$, whereupon we have

$$\begin{aligned}
\sqrt{n}\bar{s} &= \sqrt{n}\bar{s}_\beta - \operatorname{Proj}\left(\sqrt{n}\bar{s}_\beta\middle|\,\mathcal{T}_q\right) \\
&= \sqrt{n}\bar{s}_\beta - E\left[\sqrt{n}\bar{s}_\beta\,\middle|\sqrt{n}\bar{t}_1\right]
\end{aligned} \tag{16}$$

as the efficient score. This is different from that of static problems, so any estimators derived from one efficient score do not necessarily consistent in dynamic problems.

The efficiency bound for semiparametric estimation of β is then

$$V_\beta^* = \left(E\left[\sqrt{n}\bar{s}\bar{s}'\right]\right)^{-1} \tag{17}$$

The semiparametric efficiency bound is a lower bound on the asymptotic covariance matrix among regular, consistent, and asymptotic normal estimators. There is another way to establish the bound without regularity. This method is not to restrict allowable estimators but to use the local asymptotic minimax criteria to derive optimality. An estimator whose asymptotic covariance matrix is V_β^* will be optimal under both of approaches.

The previous discussion shows how the semiparametric variance bound can be computed for the finite dimensional parameter β of a semiparametric model $f(z|z^{-l};\beta,h)$. Now, we want to compute the bound for an dynamic nonlinear function like $\mu(\beta,h) = E_{\beta,\eta}[m(y,w,\beta)]$ using the general result of previous section. Combining this result with Theorem 3.1, we obtain the efficiency bound for estimation of $E[m(y,w,\beta)]$ as follows

Theorem 4.1 *Suppose that the assumptions of Theorem 3.1 are satisfied, and that* $M = \partial E[m(y,w,\beta)]/\partial\beta|_{\beta_0}$ *exists. Then the semiparametric efficiency bound of the model* $\mu(\theta) = E[m(y,w,\beta_0)]$ *is given by* $V_\mu^* = V_p + \widetilde{M}\cdot V_\beta^* \cdot \widetilde{M}'$ *where*

$$\begin{aligned} V_p &= E\left[\text{Proj}\left(n^{-1/2}\sum_{t=1}^{n} m(y_t,w_t,\beta_0)|\mathcal{T}_q\right)\cdot \text{Proj}\left(n^{-1/2}\sum_{t=1}^{n} m(y_t,w_t,\beta_0)|\mathcal{T}_q\right)'\right] \quad \textit{and} \\ \widetilde{M} &= M + E\left[\frac{1}{n}\sum_{t=1}^{n} m(y_t,w_t,\beta_0)\cdot\sum_{s=1}^{n} s(y_s,w_s)'\right]. \end{aligned}$$

5 Efficient Estimation

Given the efficiency bounds in the previous section, our objective is to develop feasible estimators that attain them. Since the bound is based on the projection on tangent set $\mathcal{T}_q$. the tangent set and the projection of $\sqrt{n}\overline{m}$ on $\mathcal{T}_q$ have to be calculated. The calculation of tangent set is usually straightforward. It is often to conjecture a form of tangent set from the restrictions on the scores implied by the semiparametric model. This conjecture can be verified by showing that this set contains the scores and can approximate the scores for parameters of interest arbitrarily well in mean square. But the calculation of the projection is difficult though it is easy in several interesting examples. The simple model is that the estimation problems satisfy $(m(y,w,\beta_0)-\mu_0)\in\mathcal{T}_q$. The efficiency bound will be a more simplified form.

Corollary 5.1 *If* $(m(y,w,\beta_0)-\mu_0)\in\mathcal{T}_q$*, then the semiparametric efficiency bound of the model will be* $V_m + M\cdot V_\beta^* M'$ *where*
$V_m = E\left[\left\{(n^{-1/2}\sum_{t=1}^n(m(y_t,w_t,\beta_0)-\mu_0)\right\}\left\{(n^{-1/2}\sum_{t=1}^n(m(y_t,w_t,\beta_0)-\mu_0)\right\}'\right].$

Satisfaction of this condition depends on the form of $m(\cdot)$ and the nature of function that defines $\mathcal{T}_q$. Other interesting examples are under several semiparametric assumptions regarding the stochastic dependence between ε and w : ε independent of w and $E[\varepsilon]=0$. Since we are dealing with unconditional expectation of dynamic functions, we can modify Brown and Newey (1998) result. As a general feasible estimator that attains the efficiency bound for $\mu(\beta,h)$, we propose

$$\widehat{\widehat{\mu}}\left(\widehat{\beta},\widehat{h}\right) = \frac{1}{\sqrt{n}}\text{Proj}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} m(y_t,w_t,\widehat{\beta})|\mathcal{T}_q\right) + \mu\left(\widehat{\beta},\widehat{h}\right) \quad (18)$$

Under restriction on the dependence of the limiting distribution of $\widehat{\widetilde{\mu}}$ on $\widehat{h}$, and other more standard assumptions, it turns out that this estimator is efficient if $\widehat{\beta}$ is semiparametric efficient. For simplicity, we define $\widetilde{m}(\beta, h) = n^{-1/2}\,\text{Proj}\left(n^{-1/2}\sum_{t=1}^{n} m(y_t, w_t, \beta)|\mathcal{I}_q\right) + \mu(\beta, h)$, $\widetilde{\mu}(\beta) = \int \widetilde{m}(y, w, \beta, h_0)\,\Pi_{t=-\infty}^{n} f(z_t|z_t^{-l}; \beta_0, h_0)dz$, and $\mu_0 = \int m(y, w, \beta_0)\Pi_{t=-\infty}^{n} f(z_t|z_t^{-l}; \beta_0, h_0)dz$.

Theorem 5.1 *Suppose that (i) $\{z_t\}_{t=1}^{n}$ is $\rho-$mixing process and LAN family, (ii) $\widehat{\beta}$ is regular and asymptotically linear where $E\left[\sqrt{n}\overline{\psi}_\beta\right] = 0$ and $V_\beta = E\left[n\overline{\psi}_\beta\overline{\psi}_\beta'\right]$ is finite, (iii) $\widetilde{M}(\beta) = \partial\widetilde{\mu}(\beta)/\partial\beta'|_{\beta_0}$ is bounded and continuous on a neighborhood of β_0, (iv) $n^{-1/2}\sum_{t=1}^{n}[\{\widetilde{m}(\beta, h_0) - \widetilde{\mu}(\beta)\} - \{\widetilde{m}(\beta_0, h_0) - \widetilde{\mu}(\beta_0)\}]$ is stochastically equicontinuous at $\beta = \beta_0$, (v) $n^{-1/2}\sum_{t=1}^{n}[\widetilde{m}(\widehat{\beta}, \widehat{h}) - \widetilde{m}(\widehat{\beta}, h_0)] = o_p(1)$, and (vi) $V_p = E[\text{Proj}\left(n^{-1/2}\sum_{t=1}^{n}(m(y_t, w_t, \beta_0)|\mathcal{I}_q\right)$ $\text{Proj}\left(n^{-1/2}\sum_{t=1}^{n}(m(y_t, w_t, \beta_0)|\mathcal{I}_q\right)']$ is finite and continuous on a neighborhood of β_0. Then $n^{1/2}\left(\widehat{\widetilde{\mu}} - \mu_0\right) \to^d N\left(0, V_{\widetilde{\mu}}\right)$ where $V_{\widetilde{\mu}} = V_p + \widetilde{M}\cdot V_\beta \cdot \widetilde{M}$.*

Assumption (i) is a primitive condition for asymptotic normality of the function, $n^{1/2}\left[\widetilde{m}(\beta_0, h_0) - \widetilde{\mu}(\beta_0)\right]$. Asymptotic independence of $m(\cdot)$ comes from mixing $\{z_t\}$ process and measurable mapping $m(\cdot)$ that depends on finite sequences of $\{z_t\}$, i.e., $m(\cdot)$ is a function of (y_t, w_t). This holds for $m(\cdot)$ function which depends on infinite sequences of $\{z_t\}$ if it can be approximated by finite sequences with small errors (Billingsley (1999)).

A sufficient condition of (iv) is the following Lipschitz condition in probability. This condition provides easily verifiable conditions for Theorem 5.1. Define

$$\widehat{Q}_n(\widehat{\beta}) = n^{1/2}\{\widetilde{m}\left(\widehat{\beta}, h_0\right) - \widetilde{\mu}(\widehat{\beta})\}$$

and

$$\widehat{Q}_n(\beta_0) = n^{1/2}\{\widetilde{m}(\beta_0, h_0) - \widetilde{\mu}(\beta_0)\},$$

then the Lipschitz condition is

$$\left|\widehat{Q}_n(\widehat{\beta}) - \widehat{Q}_n(\beta_0)\right| \leq B_n\left\|\widehat{\beta} - \beta_0\right\|^\alpha$$

where $\alpha > 0$ and $B_n = O_p(1)$. If the derivative of $\widehat{Q}_n(\widehat{\beta})$ can be replaced by the derivative of the limit of $\widehat{Q}_n(\widehat{\beta})$, a sufficient condition of stochastic equicontinuity is boundedness of the derivative.

Assumption (v) indicates that estimation of h should not have effect on the asymptotic variance. From the Newey and McFadden (1994), this condition requires sufficiently fast convergence rate of $\widehat{h}$ to its target. If $\widehat{h}$ depends on a finite number of estimated parameters, η, then the derivatives of $E\left[\widetilde{m}(\beta_0, h(\eta))\right]$ with respect to η is zero, a standard condition for estimation of h to have no effect on asymptotic variance.

The asymptotic variance matrix of Theorem 5.1 is that of $\widehat{\widetilde{\mu}}(\widehat{\beta}, h_0)$. Structure of the covariance matrix shows that the efficiency of $\widehat{\widetilde{\mu}}(\widehat{\beta}, h_0)$ depends on that of parameter estimator. If the parameter estimator is semiparametric efficient, the estimator attains a lower bound for asymptotic covariance, V_μ^* given in Theorem 4.1. With the addition of assumptions to guarantee regularity, we have following corollary.

Corollary 5.2 *Suppose the conditions of Theorem 5.1 are satisfied and* $E_{\beta,\eta}\left[\left\|\psi^*_\mu(y,w)\right\|^2\right]$ *is finite and continuous on a neighborhood of* (β_0,η_0)*. Then* $\widehat{\widetilde{\mu}}$ *based on a semiparametric efficient estimator of* β *is regular and attains the lower bound* V^*_μ*.*

Since the result of Corollary 5.2 needs $\widehat{\widetilde{\mu}}(w_{n+1})$ based on a semiparametric efficient estimator of β, we need those estimators for following examples. We assume for each case that a semiparametric efficient estimator of β is available[3].

5.1 ε is independent of w

Here we deal with a dynamic forecasting function given w where the function depends on (y,w) only through $\varepsilon=\rho(y,w,\beta_0)$. Assume that ε is i.i.d. and that $\{x_t\}_{-\infty}^{\infty}$ is exogenous and given. It is a well-known result that conditional prediction is a efficient estimator for forecasting in mean squared error sense. So we work on semiparametric efficient estimation of conditional prediction. First, a conditional prediction function with a unknown variable is considered, $m(\varepsilon,w_{n+1},\beta)$ where w_{n+1} is given. This, one-period ahead prediction function, can be represented by the function of β and $f(\varepsilon|w_{n+1};h)$.

$$\begin{aligned} \mu_0(w_{n+1}) &= \int m(\varepsilon,w_{n+1},\beta_0)f(\varepsilon|w_{n+1};h_0)d\varepsilon \\ &= \int m(\varepsilon,w_{n+1},\beta_0)f(\varepsilon;h_0)d\varepsilon \end{aligned} \tag{19}$$

where h denotes nonparametric components and the second equality comes from independency. So, we can use $\widehat{\widetilde{\mu}}\left(\widehat{\beta},\widehat{h}\right)=\widetilde{m}\left(\widehat{\beta},\widehat{h}\right)$ as a efficient estimator of target function. Although realized ε_t correlated with the final value w_{n+1}, an analogy with initial value problem of processes indicates that the final value w_{n+1} should be negligible in the analysis of ε_t. This requires asymptotic independence and symmetry of $m(\cdot)$. Symmetry condition tells us that past and future of processes can be interchanged. Time reversibility of $m(\cdot)$ process is immediate from the definition of $\rho-$mixing process. The assumption (i) of theorem 5.1 is also important to asymptotic distribution of $\widehat{\beta}$. To have the same form of asymptotic variance as unconditional expectation, the asymptotic distribution of $\sqrt{n}(\widehat{\beta}-\beta_0)\Big|_{w_{n+1}}$ should be same as that of $\sqrt{n}(\widehat{\beta}-\beta_0)$. Asymptotic independency is a sufficient condition of having same asymptotic distribution. Assumption (v) of theorem 5.1 can be verified easily under the independence assumption between ε and w.

$$\begin{aligned} \left.\frac{\partial E\left[\widetilde{m}\left(\varepsilon,w_{n+1},\beta_0,\eta\right)\right]}{\partial\eta'}\right|_{\eta_0} &= \frac{\partial}{\partial\eta'}\left\{\int \widetilde{m}\left(\varepsilon,w_{n+1},\beta_0,\eta\right)f_{\varepsilon|w}(\varepsilon|w_{n+1},\beta_0,\eta)d\varepsilon\right\} \\ &= \frac{\partial}{\partial\eta'}\left\{\int m\left(\varepsilon,w_{n+1},\beta_0\right)f_{\varepsilon}(\varepsilon,\beta_0)d\varepsilon\right\} \\ &= 0 \end{aligned} \tag{20}$$

[3]Brown and Hodgson (2000) proposed a locally efficient estimator under elliptical symmetry assumption.

where the second equality comes from independency and zero mean property of tangent set $\mathcal{T}$.

Specific form of efficient estimator under independence assumption can be worked out by imposing this additional restriction on the general tangent set. Assume that the parameter matrix β does not include an intercept. Instead the distribution of the disturbances ε are allowed to have a nonzero location parameter. Then equation (14) is given by

$$\mathcal{T}_q = \left\{ t = \frac{1}{\sqrt{n}} \sum_{s=1}^{n} \left\{ t_1(\varepsilon_s) + t_2(x_s, x_s^{-l}) \right\} : E\left[\sqrt{n} t_1(\varepsilon_s)\right] = E\left[\sqrt{n} t_2(x_s, x_s^{-l})\right] = 0 \right\} \quad (21)$$

where $\sqrt{n}t_1(\cdot)$ and $\sqrt{n}t_2(\cdot)$ are unrestricted except for the mean zero property. The projection onto this set of the sum of an arbitrary function of $R(\varepsilon_t)$ consists of following two parts

$$\text{Proj}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} R(\varepsilon_t) \middle| \frac{1}{\sqrt{n}} \sum_{s=1}^{n} t_1(\varepsilon_s) \right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E[R(\varepsilon_t)|\varepsilon_t] - \sqrt{n} E[R(\varepsilon)]$$

$$\text{Proj}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} R(\varepsilon_t) \middle| \frac{1}{\sqrt{n}} \sum_{s=1}^{n} t_2(x_s, x_s^{-l}) \right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E\left[R(\varepsilon_t)|x_t, x_t^{-l}\right] - \sqrt{n} E[R(\varepsilon)]$$

These are indeed projections to the tangent set that are closest to the original function with minimum mean squared error. Note that the projection to the space of $n^{-1/2} \sum_{s=1}^{n} t_2(x_s, x_s^{-l})$ is zero because $R(\varepsilon_t)$ is a function of only ε. Using this result, we obtain

$$\begin{aligned} \widetilde{m}(\varepsilon, w_{n+1}, \beta, h) &= n^{-1/2} \text{Proj}\left(n^{-1/2} \sum_{t=1}^{n} m(\varepsilon_t, w_{n+1}, \beta) | \mathcal{T}_q \right) + \mu(\beta_0, h_0) \quad (22) \\ &= n^{-1} \sum_{t=1}^{n} E[m(\varepsilon_t, w_{n+1}, \beta)|\varepsilon] \\ &= n^{-1} \sum_{t=1}^{n} m(\varepsilon_t, w_{n+1}, \beta) \end{aligned}$$

Given a semiparametric efficient $\widehat{\beta}$, we obtain

$$\widehat{\widehat{\mu}}(w_{n+1}) = \frac{1}{n} \sum_{t=1}^{n} m(\widehat{\varepsilon}_t, w_{n+1}, \widehat{\beta}) \quad (23)$$

as the optimal semiparametric estimator of $\mu(w_{n+1}, \beta_0, \eta_0)$. This is a method of moment estimator that integrates out unknown distribution of ε using empirical distribution. Naturally, the next question will be about efficient estimators for multi-period ahead prediction. For two-period ahead prediction, we have 2 unknown disturbance vectors, i.e., $\widetilde{m}(\varepsilon_{s+1}, \varepsilon_s, w_{n+1}; \beta, h)$. The nonparametric tangent set is the same as before. The projection onto this set of an arbitrary function of $n^{-1/2} \sum_{s=1}^{n} R(\varepsilon_{s+1}, \varepsilon_s)$ results in

$$\begin{aligned} \text{Proj}\left(n^{-1/2} \sum_{s=1}^{n} R(\varepsilon_{s+1}, \varepsilon_s) | \mathcal{T}_q \right) &= n^{-1/2} \sum_{s=1}^{n} \{E[R(\varepsilon_{s+1}, \varepsilon_s)|\varepsilon_{s+1}] + E[R(\varepsilon_{s+1}, \varepsilon_s)|\varepsilon_s] \quad (24) \\ &\quad -2E[R(\varepsilon_{s+1}, \varepsilon_s)]\} \end{aligned}$$

Thus we have

$$\widetilde{m}(\varepsilon_{s+1},\varepsilon_s,w_{n+1},\beta,h) = \frac{1}{\sqrt{n}}\operatorname{Proj}\left(n^{-1/2}\sum_{s=1}^{n}R(\varepsilon_{s+1},\varepsilon_s)|\mathcal{T}_q\right)+\mu(\beta,h) \tag{25}$$

$$= n^{-1}\sum_{s=1}^{n}\{E[R(\varepsilon_{s+1},\varepsilon_s)|\varepsilon_{s+1}]+E[R(\varepsilon_{s+1},\varepsilon_s)|\varepsilon_s]\}-E[R(\varepsilon_{s+1},\varepsilon_s)]$$

$$= \int m(\varepsilon_{s+1},\varepsilon_s,w_{n+1},\beta)f_\varepsilon(\varepsilon_s)d\varepsilon_s$$

$$+\int m(\varepsilon_{s+1},\varepsilon_s,w_{n+1},\beta)f_\varepsilon(\varepsilon_{s+1})d\varepsilon_{s+1}$$

$$-\int\left[\int m(\varepsilon_{s+1},\varepsilon_s,w_{n+1},\beta)f_\varepsilon(\varepsilon_s)d\varepsilon_s\right]f_\varepsilon(\varepsilon_{s+1})d\varepsilon_{s+1}$$

Since we do not know the nature of $f_\varepsilon(\varepsilon)$ and ε is independent of w_{n+1} $\sqrt{n}-$consistent estimates of these integrals given $\widehat{\beta}$ and w_{n+1} are obtained by utilizing the empirical distribution functions of ε. Therefore, we have

$$\widetilde{m}(\varepsilon_{s+1},\varepsilon_s,w_{n+1},\widehat{\beta},\widehat{h}) = n^{-1}\sum_{j=1}^{n}m(\widehat{\varepsilon}_i,\widehat{\varepsilon}_j,w_{n+1},\widehat{\beta})+n^{-1}\sum_{j=1}^{n}m(\widehat{\varepsilon}_j,\widehat{\varepsilon}_i,w_{n+1},\widehat{\beta})$$

$$-n^{-1}\sum_{i=1}^{n}n^{-1}\sum_{j=1}^{n}m(\widehat{\varepsilon}_i,\widehat{\varepsilon}_j,w_{n+1},\widehat{\beta})$$

where ε_i is $\varepsilon(y_i,w_i)$. When we average this function, the second and the third terms cancel and we obtain

$$\widehat{\widehat{\mu}} = n^{-1}\sum_{i=1}^{n}n^{-1}\sum_{j=1}^{n}m(\widehat{\varepsilon}_i,\widehat{\varepsilon}_j,w_{n+1},\widehat{\beta}) \tag{26}$$

as the optimal semiparametric estimator of two-period ahead prediction function $\mu(w_{n+1},\beta_0,h_0)$.

By the same analogy, multi-period, s, ahead case, we obtain

$$\widehat{\widehat{\mu}} = n^{-1}\sum_{t_1=1}^{n}\cdots n^{-1}\sum_{t_s=1}^{n}m(\widehat{\varepsilon}_{t_1},\ldots,\widehat{\varepsilon}_{t_s},w_n,\widehat{\beta}) \tag{27}$$

as an efficient estimator of $\mu(w_{n+1},\beta_0,h_0)$. This is the residual-based estimator of Brown and Mariano (1989). Here, we have established the semiparametric optimality of the estimator when the underlying parameter β is estimated semiparametric efficiently.

Similarly, the estimators

$$\widehat{\Omega} = \frac{1}{n}\left[\sum_{t=1}^{n}\pi(\widehat{\varepsilon}_t,w_{n+1},\widehat{\beta})\pi(\widehat{\varepsilon}_t,w_{n+1},\widehat{\beta})'\right] \tag{28}$$

and

$$\widehat{F}(c|w_{n+1}) = \frac{1}{n}\sum_{t=1}^{n}1\left(\pi(\widehat{\varepsilon}_t,w_{n+1},\widehat{\beta})\le c\right) \tag{29}$$

are the optimal semiparametric estimators of the conditional covariance matrix and the conditional distribution function of endogenous variables.

5.2 Expectation of ε is zero

This is the type of assumption that would be utilized to obtain general heteroskedasticity autocovariance consistent (HAC) estimators. A important problem is generalized method of moments (GMM) estimators of β under $E\left[\rho(y,w,\beta_0)\right]=0$. The optimal weight matrix for GMM is given by the inverse of $\Omega=E\left[n\overline{\rho}(y,w,\beta_0)\overline{\rho}(y,w,\beta_0)'\right]$ and the efficiency of GMM estimates depends on that of estimates for Ω. One of the most popular estimates for Ω was proposed by Newey and West (1987). Their positive semi-definite, heteroskedasticity and autocorrelation consistent estimator is given by

$$\widehat{S}=\widehat{\Omega}_0+\sum_{j=1}^{m}p(j,m)\left[\widehat{\Omega}_j+\widehat{\Omega}_j'\right],\quad p(j,m)=1-\frac{j}{m+1} \tag{30}$$

where $\widehat{\Omega}_0=\frac{1}{n}\sum_{i=1}^{n}\rho(y_i,w_i,\widehat{\beta})\rho(y_i,w_i,\widehat{\beta})'$, $\widehat{\Omega}_j=\frac{1}{n}\sum_{i=j+1}^{n}\rho(y_i,w_i,\widehat{\beta})\rho(y_{i-j},w_{i-j},\widehat{\beta})$, and m, the bound on the number of sample autocovariances, is equal to the number of nonzero autocorrelations of $\rho(y_i,w_i,\beta_0)$.

However, this is not a efficient estimator because it's based on the sample moments. Using the method proposed in the previous section, we can obtain the efficient estimate of the covariance matrix. Let $\rho_t(\beta)=\rho(y_t,w_t,\beta)$ and $\rho_t=\rho_t(\beta_0)$. To derive the nonparametric tangent set, we, first, differentiate $E\left[\rho(y_t,w_t,\beta_0)\right]=0$ with respect to η.

$$\begin{aligned}
0 &= \int\rho_s\Pi_{t=-\infty}^{n}\frac{\partial f_{z|z^{-l}}(z_t|z_t^{-l};\beta_0,h_0)}{\partial\eta}dz \\
&= E\left[\rho_s\sum_{t=-\infty}^{n}s_\eta(y_t,w_t)'\right] \\
&= E\left[\frac{1}{n}\sum_{s=1}^{n}\rho_s\sum_{t=1}^{n}s_\eta(y_t,w_t)'\right]
\end{aligned} \tag{31}$$

Then, combining the result of section 3, the nonparametric tangent set is give by

$$\mathcal{T}_q=\left\{\frac{1}{\sqrt{n}}\sum_{s=1}^{n}t(y_s,w_s):E\left[\frac{1}{\sqrt{n}}\sum_{s=1}^{n}t(y_s,w_s)\right]=0,E\left[\frac{1}{n}\sum_{s=1}^{n}t(y_s,w_s)\sum_{t=1}^{n}\rho_t'\right]=0\right\} \tag{32}$$

where $t(y_t,w_t)$ is any function of arguments which satisfies mean zero and imposed semiparametric assumption.

The corresponding projection of an arbitrary function onto this set has the form

$$\begin{aligned}
\text{Proj}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}R(y_t,w_t)|\mathcal{T}_q\right) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}R(y_t,w_t)-E\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}R(y_t,w_t)\right] \\
&\quad -E\left[\frac{1}{n}\sum_{t=1}^{n}R(y_t,w_t)\sum_{\tau=1}^{n}\rho_\tau'\right]\Omega^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\rho_t
\end{aligned} \tag{33}$$

where $\Omega=E\left[\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\rho_t\rho_s'\right]$.

To improve the efficiency of Newey and West (1987) estimator we estimate $\Omega_j = E\left[\rho_t \rho'_{t-j}\right] = \mu(\beta, h)$ efficiently. Let $m_j(y_t, w_t, \beta) = vec(\rho_t \rho'_{t-j})$. Using the previous result, we have

$$\begin{aligned} \widetilde{m}_j(\beta, h) &= \frac{1}{\sqrt{n}} \text{Proj}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} m_j(y_t, w_t) | \mathcal{T}_q\right) + \mu(\beta, h) \\ &= \frac{1}{n} \sum_{t=1}^{n} m_j(y_t, w_t) - E\left[\frac{1}{n} \sum_{\tau=1}^{n} m_j(y_\tau, w_\tau) \sum_{s=1}^{n} \rho'_s\right] \Omega^{-1} \frac{1}{n} \sum_{t=1}^{n} \rho_t \\ &= \frac{1}{n} \sum_{t=1}^{n} \left\{m_j(y_t, w_t) - C\Omega^{-1} \rho_t\right\} \end{aligned} \tag{34}$$

where $C = E\left[n^{-1} \sum_{\tau=1}^{n} m_j(y_\tau, w_\tau) \sum_{s=1}^{n} \rho'_s\right]$. An optimal semiparametric estimator of Ω_j is given by

$$\widehat{\Omega}_j = \frac{1}{n} \sum_{t=1}^{n} \left[m_j(y_t, w_t) - \widehat{C} \widehat{\Omega}^{-1} \rho_t\right] \tag{35}$$

where consistent estimators for $\widehat{C}$ and $\widehat{\Omega}$ can be obtained using Newey and West's (1987) method. Note that this estimator differs from the standard approach by the second term in the summation. This estimator will be asymptotically efficient if semiparametric efficient estimates of β are used. By substituting all components, $\widehat{\Omega}_j$, of Newey and West (1987) estimator, we will get a more efficient estimator of Ω.

6 Conclusion

This paper explores scores, semiparametric efficiency bound, and efficient estimation of expectation in dynamic nonlinear systems. The semiparametric efficiency bound is developed for dynamic problems. The explicit form of the semiparametric efficient expectation estimator is worked out for two important semiparametric assumptions. Under the assumption of independence between disturbances and predetermined variables, the residual-based predictors proposed by Brown and Mariano (1989) are shown to be semiparametric efficient. Under unconditional mean zero assumption, we developed improved heteroskedasticity, autocorrelation consistent (HAC) estimators.

Appendix: Mathematical Proofs

Proof of Lemma 3.1. Define the vector of unknown parameters

$$\delta = \left(\beta, f_{z|z^{-l}}(z|z^{-l})\right) \in \Delta = \Theta \times \Xi$$

Define the sequence of absolute continuous probability measures $\{P_{\delta,n}\}$, which represent the distribution of the sample of size n when δ is the parameter vectors. By Radon-Nikodym theorem, there exist densities relative to a measure μ. Define the Hilbert space $\mathsf{H} = \overline{\mathsf{H}}_1 + \overline{\mathsf{H}}_2$ where $\overline{\mathsf{H}}_1$ and $\overline{\mathsf{H}}_2$ are the Hilbert space containing functions of the form

$$h_1(y, w, \beta) = \kappa' \frac{1}{\sqrt{n}} \sum_{t=1}^{n} s_\beta(y_t, w_t, \beta) \sqrt{\Pi_{t=1}^{n} f_{z|z^{-l}}(z_t|z_t^{-l})}$$

where κ is a vector of constants with dimensionality equal to that of β. We further define H_2 as the set of all bounded, square-integrable functions $h_2(y, w, \beta)$ having the form

$$h_2(y, w, \beta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} t(y_t, w_t, \beta) \sqrt{\Pi_{t=1}^{n} f_{z|z^{-l}}(z_t|z_t^{-l})}$$

such that $\int h_2(y, w, \beta) \sqrt{\Pi_{t=1}^{n} f_{z|z^{-l}}(z_t|z_t^{-l})} d\overline{z} = 0$ where $\overline{z} = (z_1, \cdots, z_n)$.

Define the norm of an element $h \in \mathsf{H}$ by

$$\|h\|_H = \left\{ \int (h_1(y, w, \beta) + h_2(y, w, \beta))^2 \, d\overline{y} d\overline{w} \right\}^{1/2}$$

The following sequence of linear operators $\{A_n\}$ maps H into $R^p \times L_2$:

$$A_n(h) = n^{-1/2} \left[\begin{array}{c} V_\beta^* \int \frac{1}{\sqrt{n}} \sum_{t=1}^{n} s(y_t, w_t) \frac{1}{\sqrt{n}} \sum_{s=1}^{n} h_1' \sqrt{\Pi_{t=1}^{n} f_{z|z^{-l}}(z_t|z_t^{-l})} d\overline{z} \\ h_2(y, w, \beta) \sqrt{\Pi_{t=1}^{n} f_{z|z^{-l}}(z_t|z_t^{-l})} \end{array} \right]$$

For every $h \in \mathsf{H}$, we have

$$\delta + A_n(h) = \delta + n^{-1/2} \left[\begin{array}{c} \kappa \\ h_2(y, w, \beta) \sqrt{\Pi_{t=1}^{n} f_{z|z^{-l}}(z_t|z_t^{-l})} \end{array} \right]$$

To obtain our LAN theory, we have to verify that conditions 1-3 of Ibragimov and Khas'minskii (1991, p.1682) hold for our model. Condition 1 states that $\lim_{n\to\infty} \|A_n(h)\| = 0$, for all $h \in \mathsf{H}$, which holds by the boundedness and integrability of $h_2(y, w, \beta)$. Condition 2 will follow if we can show that for every $h \in \mathsf{H}$, there exists n sufficiently large that $\delta + A_n(h) \in \Delta$. We note that $\beta + n^{-1/2}\kappa \in \Theta$, since $\beta \in int\ \Theta$ for n sufficiently large. It's not hard to see that for n sufficiently large, $f_{z|z^{-l}}(z_t|z_t^{-l}) + n^{-1/2} h_2(y, w, \beta) \sqrt{\Pi_{t=1}^{n} f_{z|z^{-l}}(z_t|z_t^{-l})}$ is a density function. So, Condition 2 holds.

To check Condition 3, we analyze the asymptotic behavior of the likelihood ratio $\Lambda_n(\delta + A_n(h), \delta) = \frac{dP_{\delta+A_n(h)}}{dP_\delta}$. With appropriate assumptions on the distribution of the initial conditions, and defining $\beta_n = \beta + n^{-1/2}k$, we can approximate the likelihood ratio by

$$\Lambda_n(\delta + A_n(h), \delta) \cong \Pi_{t=1}^n \left\{ \frac{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta_n)}{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)} + n^{-1/2} \frac{h_2(y_t, w_t, \beta_n)\sqrt{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta_n)}}{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)} \right\}$$

Note that the following Taylor expansion;

$$\begin{aligned} f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta_n) &= f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta) + n^{-1/2}\kappa' \frac{\partial f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)}{\partial \beta} \\ &\quad + \frac{1}{2n}\kappa' \frac{\partial^2 f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)}{\partial \beta \partial \beta'} \kappa + O_p(n^{-3/2}) \end{aligned}$$

Now define

$$r_t(\beta) = h_2(y_t, w_t, \beta)\sqrt{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)}$$

We have

$$n^{-1/2} r_t(\beta_n) = n^{-1/2} r_t(\beta) + n^{-1}\kappa' \frac{\partial r_t(\beta)}{\partial \beta} + O_p(n^{-3/2})$$

Substituting the above expansion into the likelihood ratio equation, we get

$$\begin{aligned} \Lambda_n(\delta + A_n(h), \delta) \cong \Pi_{t=1}^n \Bigg\{ & 1 + n^{-1/2}\kappa' \frac{\partial f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)/\partial \beta}{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)} \\ & + n^{-1/2} \frac{h_2(y_t, w_t, \beta)}{\sqrt{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)}} \\ & + \frac{1}{2n}\kappa' \frac{\partial^2 f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)/\partial \beta \partial \beta'}{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)} \kappa \\ & + \frac{1}{n} \frac{\kappa' \partial r_t(\beta)/\partial \beta}{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)} + o_p(1) \Bigg\} \end{aligned}$$

We note that

$$\exp(x) = 1 + x + \frac{1}{2}x^2 + \cdots,$$

that

$$n^{-1} \sum_{t=1}^n \frac{\partial^2 f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)/\partial \beta \partial \beta'}{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)} = o_p(1),$$

and that

$$n^{-1} \sum_{t=1}^n \frac{\partial r_t(\beta)/\partial \beta}{f_{z|z^{-\iota}}(z_t|z_t^{-l}; \beta)} = o_p(1).$$

Using above results, we obtain

$$\Lambda_n(\delta + A_n(h), \delta) = \exp\left\{ n^{-1/2} \sum_{t=1}^{n} \kappa' \frac{\partial f_{z|z^{-l}}(z_t|z_t^{-l};\beta)/\partial\beta}{f_{z|z^{-l}}(z_t|z_t^{-l};\beta)} + n^{-1/2} \sum_{t=1}^{n} \frac{h_2(y_t, w_t, \beta)}{\sqrt{f_{z|z^{-l}}(z_t|z_t^{-l};\beta)}} - \frac{1}{2n} \sum_{t=1}^{n} \left\{ \kappa' \frac{\partial f_{z|z^{-l}}(z_t|z_t^{-l};\beta)/\partial\beta}{f_{z|z^{-l}}(z_t|z_t^{-l};\beta)} + \frac{h_2(y_t, w_t, \beta)}{\sqrt{f_{z|z^{-l}}(z_t|z_t^{-l};\beta)}} \right\}^2 + o_p(1) \right\}$$

Defining the quantity

$$\Delta_n(h) = n^{-1/2} \sum_{t=1}^{n} \left\{ \kappa' \frac{\partial f_{z|z^{-l}}(z_t|z_t^{-l};\beta)/\partial\beta}{f_{z|z^{-l}}(z_t|z_t^{-l};\beta)} + \frac{h_2(y_t, w_t, \beta)}{\sqrt{f_{z|z^{-l}}(z_t|z_t^{-l};\beta)}} \right\}$$

We can show that

$$\Delta_n(h) \to_d N\left(0, \|h\|_H^2\right)$$

by the stationarity and boundedness of sum of covariances, where

$$\frac{1}{n} \sum_{t=1}^{n} \left\{ \kappa' \frac{\partial f_{z|z^{-l}}(z_t|z_t^{-l};\beta)/\partial\beta}{f_{z|z^{-l}}(z_t|z_t^{-l};\beta)} + \frac{h_2(y_t, w_t, \beta)}{\sqrt{f_{z|z^{-l}}(z_t|z_t^{-l};\beta)}} \right\}^2 \to_P \|h\|_H^2$$

It follows that

$$\Lambda_n(\delta + A_n(h), \delta) = \exp\left\{ \Delta_n(h) - \frac{1}{2} \|h\|_H^2 + o_p(1) \right\}$$

so that the LAN conditions of Ibragimov and Khas'minskii (1991) are satisfied. ■

Proof of Lemma 3.2. By adding and subtracting,

$$\begin{aligned} E\left[\frac{1}{n}\sum_{t=1}^{n} p(z_t) \sum_{s=1}^{n} q(z_s)'\right] &= \frac{1}{n}\sum_{t=1}^{n} E\left[p(z_t) \sum_{s=1}^{n} q(z_s)'\right] \\ &= \frac{1}{n}\sum_{t=1}^{n} E\left[p(z_t) \sum_{s=-\infty}^{n} q(z_s)'\right] - \frac{1}{n}\sum_{t=1}^{n} E\left[p(z_t) \cdot \sum_{s=-\infty}^{0} q(z_s)'\right] \\ &= \frac{1}{n}\sum_{t=1}^{n} E\left[p(z_t) \cdot \sum_{s=-\infty}^{n} q(z_s)'\right] + o(1) \end{aligned}$$

To show the last line, we decompose each term as follows

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n} E\left[p(z_t) \sum_{s=-\infty}^{0} q(z_s)'\right] &= \frac{1}{n}\sum_{t=1}^{n} E\left[p(z_t) \sum_{s=-\infty}^{-r} q(z_s)'\right] + \frac{1}{n}\sum_{t=1}^{n} E\left[p(z_t) \sum_{s=-r+1}^{0} q(z_s)'\right] \\ &= \frac{1}{n}\sum_{t=1}^{n} E\left[p(z_t) \sum_{s=-\infty}^{-r} q(z_s)'\right] + \frac{1}{n}\sum_{t=r+1}^{n} E\left[p(z_t) \sum_{s=-r+1}^{0} q(z_s)'\right] \\ &\quad + \frac{1}{n}\sum_{t=1}^{r} E\left[p(z_t) \sum_{s=-r+1}^{0} q(z_s)'\right] \\ &< O(\xi(r)) + O\left(\frac{r^2}{n}\right) \end{aligned}$$

where $\xi(r) > 0$. ■

Proof of Lemma 3.3. Consider an local data generating process with parameter θ_n. Since local data generating process for regular parametric submodels are contiguous to the process with $\theta_n = \theta_0$, $\sqrt{n}(\widehat{\mu} - \mu_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\mu(y_t, w_t) + o_p(1)$ also holds under the process. Then by the addition and subtraction,

$$\begin{aligned}
\sqrt{n}(\widehat{\mu} - \mu_n) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\psi_\mu(y_t, w_t) - E_n\left[\psi_\mu(y, w)\right]\right) + \sqrt{n}(\mu_0 - \mu_n) \\
&\quad + \sqrt{n} E_n\left[\psi_\mu(y, w)\right] + o_p(1)
\end{aligned}$$

where $\mu_n = \mu(\theta_n)$. By regularity $f(\theta_n) \to^{a.s} f(\theta_0)$, so that $\lim_{n\to\infty}\left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\psi_\mu(y_t, w_t) - E_n\left[\psi_\mu(y, w)\right]\right)\right] = 0$. Also, by LAN condition, $\sqrt{n}\left(\widehat{\theta} - \theta_n\right) \to_d N(0, V_\theta)$. Therefore, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\psi_\mu(y_t, w_t) - E_n\left[\psi_\mu(y, w)\right]\right) \to_d N(0, V_\mu)$$

where $V_\mu = E\left[\frac{1}{n} \sum_{t=1}^{n} \psi_\mu(y_t, w_t) \sum_{s=1}^{n} \psi_\mu(y_s, w_s)'\right]$. We can expand $\sqrt{n} E_n\left[\psi_\mu(y, w)\right]$ around θ_0.

$$\begin{aligned}
\sqrt{n} E_n\left[\psi_\mu(y, w)\right] &= \sqrt{n}\left\{E\left[\psi_\mu(y, w)\right] + E\left[\psi_\mu(y, w) \sum_{t=-\infty}^{n} s_\theta(y_t, w_t)'\right](\theta_n - \theta_0)\right\} \\
&\quad + o\left(\sqrt{n}\left\|\theta_n - \theta_0\right\|\right) \\
&= E\left[\psi_\mu(y, w) \sum_{t=-\infty}^{n} s_\theta(y_t, w_t)'\right] \sqrt{n}\left(\theta_n - \theta_0\right) + o_p(1) \\
&= E\left[\frac{1}{n} \sum_{t=1}^{n} \psi_\mu(y_t, w_t) \sum_{t=1}^{n} s_\theta(y_t, w_t)'\right] \sqrt{n}\left(\theta_n - \theta_0\right) + o_p(1)
\end{aligned}$$

The last line comes from Lemma 3.2. Also, by differentiability of $\mu(\theta)$,

$$\begin{aligned}
\sqrt{n}(\mu_0 - \mu_n) &= \sqrt{n}\left[-\frac{\partial \mu(\theta_0)}{\partial \theta'}(\theta_n - \theta_0) + o\left(\theta_n - \theta_0\right)\right] \\
&= -\frac{\partial \mu(\theta_0)}{\partial \theta'} \sqrt{n}(\theta_n - \theta_0) + o(1)
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\sqrt{n}(\widehat{\mu} - \mu_n) &\to {}_d N(0, V_\mu) \\
&\quad + \left\{E\left[\frac{1}{n} \sum_{t=1}^{n} \psi_\mu(y_t, w_t) \sum_{t=1}^{n} s_\theta(y_t, w_t)'\right] - \frac{\partial \mu(\theta_0)}{\partial \theta'}\right\} \sqrt{n}(\theta_n - \theta_0) + o(1)
\end{aligned}$$

Since $\sqrt{n}(\widehat{\mu} - \mu_0) \to_d N(0, V_\mu)$ for $\theta_n = \theta_0$, it follows from the above that the limiting distribution of $\sqrt{n}(\widehat{\mu} - \mu(\theta_n))$ exists and does not depend on the sequence $\{\theta_n\}$ if and only if

$$\left\{E\left[\frac{1}{n} \sum_{t=1}^{n} \psi_\mu(y_t, w_t) \sum_{t=1}^{n} s_\theta(y_t, w_t)'\right] - \frac{\partial \mu(\theta_0)}{\partial \theta'}\right\} \sqrt{n}(\theta_n - \theta_0) = o(1)$$

This equation holds for all sequences such that $\sqrt{n}(\theta_n - \theta_0)$ is bounded if and only if $\frac{\partial \mu(\theta_0)}{\partial \theta'} = E\left[\frac{1}{n}\sum_{t=1}^{n}\psi_\mu(y_t, w_t)\sum_{t=1}^{n}s_\theta(y_t, w_t)'\right] + o_p(1)$. ■

Proof of Theorem 3.1. For any asymptotically linear and regular estimator $\widehat{\mu}$ the variance will be

$$\begin{aligned} V_\mu &= \lim_{n\to\infty} Var\left(\sqrt{n}\overline{\psi}_\mu\right) \\ &= \lim_{n\to\infty}\left\{Var\left(\sqrt{n}\overline{\psi}_\mu^*\right) + Var\left(\sqrt{n}\overline{\xi}\right)\right\} \\ &= V_\mu^* + \lim_{n\to\infty} Var\left(\sqrt{n}\overline{\xi}\right) \end{aligned}$$

Therefore, the minimum variance will be V_μ^*. ■

Proof of Theorem 4.1. Let $\mathcal{S}_q = \left\{B\sqrt{n}\overline{s}_\beta + \sqrt{n}\overline{t} : \sqrt{n}\overline{t} \in \mathcal{T}_q \text{ and } B \text{ is a constant } q \times p \text{ matrix}\right\}$, which is linear. By definition, $B\sqrt{n}\overline{s}_\beta = B\sqrt{n}\overline{s} + B\,\text{Proj}\left(\sqrt{n}\overline{s}_\beta|\mathcal{T}_q\right) = B\sqrt{n}\overline{s} + \sqrt{n}\overline{t}$. So we can rewrite $\mathcal{S}_q = \left\{B\sqrt{n}\overline{s} + \sqrt{n}\overline{t} : \sqrt{n}\overline{t} \in \mathcal{T}_q\right\}$. From the differentiability of any regular parametric submodels, $\partial\mu(\theta)/\partial\theta'$ can be written as follows

$$\begin{aligned} \frac{\partial \mu(\theta)}{\partial \theta'} &= \left.\frac{\partial\left\{\int m(y,w,\beta)\Pi_{t=-\infty}^{n} f(z_t|z_{t-1},\beta_0,h_0)dz\right\}}{\partial\beta'}\right|_{\beta_0} \cdot \frac{\partial\beta}{\partial\theta'} \\ &\quad + \int m(y,w,\beta_0)\cdot\frac{\partial\Pi_{t=-\infty}^{n} f(z_t|z_{t-1},\beta_0,h_0)/\partial\theta'}{\Pi_{t=-\infty}^{n} f(z_t|z_{t-1},\beta_0,h_0)}\Pi_{t=-\infty}^{n} f(z_t|z_{t-1},\beta_0,h_0)dz \\ &= M\left[I_p, 0\right] + E\left[m(y,w,\beta_0)\sum_{t=-\infty}^{n} s_\theta(y_t,w_t,\beta_0)'\right] \\ &= M\left(E\left[n\overline{s}\overline{s}'\right]\right)^{-1}E\left[n\overline{s}\overline{s}_\theta'\right] + E\left[\frac{1}{n}\sum_{t=1}^{n} m(y_t,w_t,\beta_0)\sum_{t=1}^{n} s_\theta(y_t,w_t,\beta_0)'\right] + o_p(1) \\ &= E\left[\left(M\left(E\left[n\overline{s}\overline{s}'\right]\right)^{-1}\sqrt{n}\overline{s} + \sqrt{n}\overline{m}\right)\sqrt{n}\overline{s}_\theta'\right] + o_p(1) \end{aligned}$$

The third line follows by $E\left[n\overline{s}\overline{s}_\theta'\right] = \left(E\left[n\overline{s}\overline{s}'\right], 0\right)$ and Lemma 3.2. Therefore, $\sqrt{n}\overline{\psi}_\mu = M\left(E\left[n\overline{s}\overline{s}'\right]\right)^{-1}\sqrt{n}\overline{s} + \sqrt{n}\overline{m}$. Since $B\sqrt{n}\overline{s}$ is orthogonal to $\mathcal{T}_q$ the projection on $\mathcal{S}_q$ is direct summation of the projection on the linear space of $B\sqrt{n}\overline{s}$ and the projection on $\mathcal{T}_q$. The projection $\sqrt{n}\overline{\psi}_\mu$ on $\mathcal{S}_q$ is

$$\begin{aligned} \text{Proj}\left(\sqrt{n}\overline{\psi}_\mu|\mathcal{S}_q\right) &= M\left(E\left[n\overline{s}\overline{s}'\right]\right)^{-1}\sqrt{n}\overline{s} + \text{Proj}\left(\sqrt{n}\overline{m}|\mathcal{S}_q\right) \\ &= M\left(E\left[n\overline{s}\overline{s}'\right]\right)^{-1}\sqrt{n}\overline{s} + E\left[n\overline{m}\overline{s}'\right]\left(E\left[n\overline{s}\overline{s}'\right]\right)^{-1}\sqrt{n}\overline{s} + \text{Proj}\left(\sqrt{n}\overline{m}|\mathcal{T}_q\right) \\ &= \text{Proj}\left(\sqrt{n}\overline{m}|\mathcal{T}_q\right) + \left(M + E\left[n\overline{m}\overline{s}'\right]\right)\left(E\left[n\overline{s}\overline{s}'\right]\right)^{-1}\sqrt{n}\overline{s} \\ &= \sqrt{n}\overline{\psi}_\mu^* \end{aligned}$$

Therefore, the semiparametric efficiency bound is the covariance of $\sqrt{n}\overline{\psi}_\mu^*$. Since $\sqrt{n}\overline{s}$ is orthogonal to the tangent set $\mathcal{T}_q$, then it has zero covariance with $\text{Proj}\left(\sqrt{n}\overline{m}|\mathcal{T}_q\right)$. We have the form given in the theorem for the efficiency bound. ■

Proof of Corollary 5.1. By the mean zero property of score functions and the orthogonality of $\mathcal{T}_q$ and $\sqrt{n}\overline{s}$, we have $E\left[n\overline{m}\overline{s}'\right]=0$. By the assumption, we have $\text{Proj}(\sqrt{n}\overline{m}|\mathcal{T}_q)=n^{-1/2}\sum_{t=1}^{n}(m(y_t,w_t,\beta_0)-\mu_0)$. Therefore, the efficiency bound is

$$\begin{aligned}V_p+\widetilde{M}\cdot V_\beta^*\cdot\widetilde{M}' &= E\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(m(y_t,w_t,\beta_0)-\mu_0)\right)\cdot\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(m(y_t,w_t,\beta_0)-\mu_0)\right)'\right]\\ &\quad +M\cdot V_\beta^*\cdot M\end{aligned}$$

■

Proof of Theorem 5.1. Adding and subtracting, we get the expression

$$\begin{aligned}n^{1/2}\left(\widehat{\widetilde{\mu}}-\widetilde{\mu}(\beta_0)\right) &= n^{1/2}\left[\widetilde{m}\left(\beta_0,h_0\right)-\widetilde{\mu}(\beta_0)\right]+n^{1/2}\left[\widetilde{\mu}(\widehat{\beta})-\widetilde{\mu}(\beta_0)\right]\\ &\quad +n^{1/2}\left\{\left[\widetilde{m}\left(\widehat{\beta},h_0\right)-\widetilde{\mu}(\widehat{\beta})\right]-\left[\widetilde{m}\left(\beta_0,h_0\right)-\widetilde{\mu}(\beta_0)\right]\right\}\\ &\quad +n^{1/2}\left[\widetilde{m}\left(\widehat{\beta},\widehat{h}\right)-\widetilde{m}\left(\widehat{\beta},h_0\right)\right]\end{aligned}$$

By the mean value theorem, the second term yields

$$\widetilde{\mu}(\widehat{\beta})-\widetilde{\mu}(\beta_0)=\widetilde{M}(\beta^*)n^{1/2}(\widehat{\beta}-\beta_0),$$

where β^* lies between $\widehat{\beta}$ and β_0. The continuity of $\widetilde{M}(\beta^*)$ together with (ii) implies

$$\widetilde{M}(\beta^*)n^{1/2}(\widehat{\beta}-\beta_0)=\widetilde{M}(\beta_0)n^{-1/2}\Sigma_{t=1}^{n}\psi_\beta(y_t,w_t)+o_p(1)$$

Assumptions (iv) and (v) imply that the last two lines are $o_p(1)$, whereupon we have

$$n^{1/2}\left(\widehat{\widetilde{\mu}}-\widetilde{\mu}(\beta_0)\right)=n^{1/2}\left\{\left[\widetilde{m}\left(\beta_0,h_0\right)-\widetilde{\mu}(\beta_0)\right]+\widetilde{M}(\beta_0)n^{-1/2}\Sigma_{t=1}^{n}\psi_\beta(y_t,w_t)\right\}+o_p(1)$$

Now, asymptotic normality of the first term is given by (i) and $\sqrt{n}\overline{\psi}_\beta$ must be orthogonal to elements of $\mathcal{T}_q$ for $E\left[\sqrt{n}\overline{\psi}_\beta\right]=0$. Thus $\sqrt{n}\overline{\psi}_\beta$ and $\text{Proj}\left(n^{-1/2}\sum_{t=1}^{n}m(y_t,w_t,\beta_0)|\mathcal{T}_q\right)$ are orthogonal and the covariance matrix exists by (ii) and (vi) with the form given in the theorem.

To show $\widetilde{\mu}(\beta_0)=\mu_0$, we know

$$\begin{aligned}\widetilde{\mu}(\beta) &= \int\widetilde{m}\left(\beta,h_0\right)\Pi_{t=-\infty}^{n}f(z_t|z_{t-1};\beta_0,h_0)dz\\ &= \mu(\beta,h_0)\end{aligned}$$

identically in β since a projection on the tangent set has mean zero. So,

$$\widetilde{\mu}(\beta_0)=\mu(\beta_0,h_0)=\mu_0.$$

By differentiating this equation with respect to β, we have

$$\widetilde{M}(\beta_0)+E\left[\widetilde{m}\left(\beta_0,h_0\right)\sum_{t=-\infty}^{n}s_\beta(y_t,w_t)'\right]=M+E\left[m(y,w,\beta_0)\sum_{t=-\infty}^{n}s_\beta(y_t,w_t)'\right]$$

Since

$$\begin{aligned} E\left[\widetilde{m}\left(\beta_0, h_0\right) \sum_{t=-\infty}^{n} s_\beta(y_t, w_t)\right] &= \frac{1}{n}\sum_{t=1}^{n} E\left[\text{Proj}\left(m(y_t, w_t, \beta_0)|\mathcal{T}_q\right) \sum_{t=-\infty}^{n} s_\beta(y_t, w_t)\right] \\ &= E\left[m(y, w, \beta_0)\,\text{Proj}\left(\sum_{t=-\infty}^{n} s_\beta(y_t, w_t)|\mathcal{T}_p\right)'\right], \end{aligned}$$

we have

$$\begin{aligned} \widetilde{M}(\beta_0) &= M + E\left[m(y, w, \beta_0)\left\{\sum_{t=-\infty}^{n} s_\beta(y_t, w_t) - \text{Proj}\left(\sum_{t=-\infty}^{n} s_\beta(y_t, w_t)|\mathcal{T}_p\right)\right\}'\right] \\ &= M + E\left[m(y, w, \beta_0) \sum_{t=-\infty}^{n} s(y_t, w_t)\right] \\ &\cong M + E\left[\frac{1}{n}\sum_{t=1}^{n} m(y_t, w_t, \beta_0) \sum_{s=1}^{n} s(y_s, w_s)\right] \end{aligned}$$

■

Proof of Corollary 5.2. Under the assumptions, Corollary 5.2 follows directly from Corollary 1 in Brown and Newey (1998). ■

References

[1] Andrews, D (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465-1471.

[2] ____ (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden. New York: North Holland, 2247-2294.

[3] Back, K., and D. Brown (1993): "Implied Probabilities in GMM Estimators," *Econometrica*, 61,971-975.

[4] Bickel, P., C. Klassen, Y. Ritov, and J. Wellner (1998): *Efficient and Adaptive Estimation for Semiparametric Models*, Springer.

[5] Billingsley, P (1999): *Convergence of Probability Measures*, Second edition, Wiley.

[6] Brown, B., and D. Hodgson (2000): "Efficient Semiparametric Estimation of Dynamic Nonlinear Systems Under Elliptical Symmetry," *Mimeo*, Economics Department, Rice University.

[7] Brown, B., and S. Maital (1981): "What Do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica*, 49, 491-504.

[8] Brown, B., and R. Mariano (1984): "Residual-Based Stochastic Prediction and Estimation in a Nonlinear Simultaneous System," *Econometrica*, 52, 321-343.

[9] ____ (1989): "Predictors in Dynamic Nonlinear Models: Large-Sample Behavior," *Econometric Theory*, 5, 430-452.

[10] Brown, B., and W. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66, 453-464.

[11] Cramér, H. (1970): *Random Variables and Probability Distributions*, Cambridge University Press.

[12] Ibragimov, I., and R. Khas'minskii (1991): "Asymptotically Normal Families of Distributions and Efficient Estimation," *Annals of Statistics*, 19, 1681-1724.

[13] Ibragimov, I., and R. Has'minskii (1981): *Statistical Estimation: Asymptotic Theory*, New York, Springer-Verlag.

[14] Newey, W. (1989): "Locally Efficient, Residual-Based Estimation of Nonlinear Simultaneous Equations," *Mimeo*, Department of Economics, Princeton University.

[15] ____ (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.

[16] ____ (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161-1167.

[17] Newey, W., and D. Macfadden (1994): "Estimation and Inference in Large Samples," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. Macfadden. New York: North Holland, 2111-2445.

[18] Newey, W., and K. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55,703-708.

[19] Reed, M., and B. Simon (1972): *Methods of Modern Mathematical Physics I: functional Analysis*, New York, Academic Press.

[20] White, H., and I. Domowitz (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, 52, 143-162.