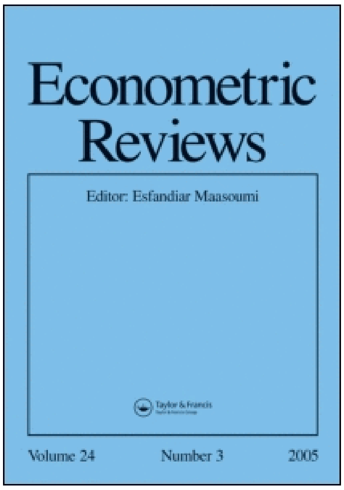

Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597248

Uniformly adaptive estimation for models with arma errors

Dougas G. Steigerwald[a]
[a] Department of Economics, University of California, Santa Barbara, CA

ECONOMETRIC REVIEWS, 16(4), 393-409 (1997)

Uniformly Adaptive Estimation for Models with ARMA Errors

Douglas G. Steigerwald
Department of Economics, University of California
Santa Barbara, CA 93106

Keywords and Phrases: adaptive; ARMA; semiparametric; uniform convergence.
JEL Classificiation: C14; C22.

Abstract

A semiparametric estimator based on an unknown density is *uniformly adaptive* if the expected loss of the estimator converges to the asymptotic expected loss of the maximum likelihood estimator based on the true density (MLE), and if convergence does not depend on either the parameter values or the form of the unknown density. Without uniform adaptivity, the asymptotic expected loss of the MLE need not approximate the expected loss of a semiparametric estimator for *any* finite sample. I show that a two-step semiparametric estimator is uniformly adaptive for the parameters of nonlinear regression models with autoregressive moving average errors.

1. Introduction

I study a two-step semiparametric estimator for nonlinear regression models for data that consist of observations $z_t = (y_t, x_t')$ in which the *conditional mean* of a period-t one-dimensional variable y_t is a known function h of period-t k-dimensional variables x_t and parameter vector $\beta = (\beta_0, \beta_1')' \in \mathbb{R}^{k+1}$. It is assumed that y_t is related to x_t and a period-t error e_t as

$$y_t = h(x_t, \beta) + e_t, \tag{1.1}$$

for all integers t. The error e_t is assumed to follow a stationary stochastic autoregressive moving- average (ARMA) process of known order (p, q) with parameter vectors $\rho \in \mathbb{R}^p$ and $\theta \in \mathbb{R}^q$, such that both ρ_p and θ_q are not equal to zero. The random variable that drives the ARMA process, u_t, is assumed independent and identically distributed (i.i.d.) with density g *that is known only to be a member of a family* $\mathcal{G}$. The value e_t is related to past values of e_t and to current and past values of u_t by the difference equation

$$e_t = \sum_{i=1}^{p} \rho_i \, e_{t-i} + u_t + \sum_{j=1}^{q} \theta_j u_{t-j}, \tag{1.2}$$

for all integers t.

The researcher uses a two-step procedure to estimate $\psi = (\beta', \rho', \theta')'$ from a T-period sample $\{y_t, x_t'\}_{t=1}^{T}$ with observed initial values $(u_{1-q}, ..., u_0, e_{1-p}, ..., e_0)$.[1] The first step is to assume that g is normal and to construct a quasi-maximum-likelihood estimator of ψ from (1.1)-(1.2). The second step is to: first, construct a nonparametric estimator of g from the residuals from the quasi-maximum likelihood estimator; second, use the nonparametric estimator of g to construct estimators of the score for ψ (the partial derivative of the true log-likelihood for the T-period sample with respect to ψ) and the information matrix for ψ (the expected value of the outer product of the score), both evaluated at the quasi-maximum likelihood estimator; and third, construct the two-step semiparametric estimator as the sum of the quasi-maximum likelihood estimator with the product of the estimators of the score and the inverse information matrix.

To judge the performance of such semiparametric estimators, relative to any other estimators of ψ based on the assumption that g is unknown, it is common to compute the asymptotic expected loss of the estimators over neighborhoods of the true parameter values and true density (henceforth, the truth). The minimum asymptotic expected loss that can be attained by these estimators is the asymptotic efficiency bound. The asymptotic efficiency bound can be no less than the asymptotic expected loss of the maximum likelihood estimator based on the *true* density (hereafter termed the maximum likelihood estimator or MLE). If the asymptotic efficiency bound is given by the asymptotic expected loss of the MLE and if a two-step semiparametric estimator achieves the asymptotic efficiency bound, then a two-step semiparametric estimator is asymptotically as efficient as the most efficient estimator of ψ based on the assumption that g is known. In this case, a two-step semiparametric estimator is *uniformly adaptive.*

To express the definition of uniform adaptivity symbolically, let $f = f(\cdot, \psi, g)$ be the density for y_t, where dependence of f on t is suppressed for notational simplicity. Let Ψ and $\mathcal{C}$ denote neighborhoods of the true parameter values and the true density, respectively, and let $f^T = f(\cdot, \psi^T, g^T)$ be a density in this neigh-

[1] Two-step procedures are commonly used to estimate nonlinear ARMAX models, see Harvey [1981].

borhood. (Throughout, superscript T denotes values in a neighborhood of the truth (ψ, g) defined for a T-period sample, and subscript T denotes an estimator based on a T-period sample.) Let l be a continuous and bounded loss function that takes $\hat{\psi}_T$ as an argument and returns a nonnegative real number and let $\mathcal{J}_\psi(\psi, g)$ be the information matrix for ψ. The T-period sample expected loss of a two-step semiparametric estimator $\hat{\psi}_T$ if the unknown parameters and density assume values in a neighborhood of the truth is $E_{f^T}\{l[\sqrt{T}(\hat{\psi}_T - \psi^T)]\}$. The asymptotic distribution of the MLE is $Z^* \sim N(0, \mathcal{J}_\psi(\psi, g)^{-1})$, so the asymptotic expected loss of the MLE is $E[l(Z^*)]$.

Because the asymptotic efficiency bound can be no less than the asymptotic expected loss of the MLE, convergence of $E_{f^T}\{l[\sqrt{T}(\hat{\psi}_T - \psi^T)]\}$ to $E[l(Z^*)]$ implies both that the asymptotic efficiency bound is given by the asymptotic expected loss of the MLE and that a two-step semiparametric estimator achieves the asymptotic efficiency bound. Thus a two-step semiparametric estimator is uniformly adaptive if

$$E_{f^T}\{l[\sqrt{T}(\hat{\psi}_T - \psi^T)]\} \to E[l(Z^*)]. \tag{1.3}$$

Because l is continuous and bounded, (1.3) follows from the weak convergence of the law of $\sqrt{T}(\hat{\psi}_T - \psi^T)$ to Z^*, denoted $\sqrt{T}(\hat{\psi}_T - \psi^T) \Rightarrow Z^*$ as $T \to \infty$ for all $\{(\psi^T, g^T)\} \in \Psi \times \mathcal{C}$ where $\Rightarrow$ denotes weak convergence.

Uniform adaptivity ensures that an adaptive estimator is robust to local perturbations in g. As Begun et al. (1983, p. 483) state uniform adaptivity "captures the local uniformity that should reasonably be required of adaptive estimates when $g \in \mathcal{G}$ is unknown." Local uniformity has an implication that is important for applied work. If a two-step semiparametric estimator is uniformly adaptive and is in a neighborhood of the true parameters and density, the finite sample expected loss of the estimator is in a neighborhood of the asymptotic expected loss of the MLE. If a two-step semiparametric estimator is not uniformly adaptive, however, the asymptotic expected loss of the MLE need not approximate the expected loss of a two-step semiparametric estimator for *any* finite sample.

Uniform adaptivity implies but is not implied by adaptivity. Proof that an estimator is uniformly adaptive requires that one establish (1.3). Proof that an estimator is adaptive (e.g. Bickel (1982), Kreiss (1987), and Steigerwald (1992)) requires that one establish a condition that is similar to (1.3) but is weaker, in that the expectation is taken only over neighborhoods of the true parameters and not over neighborhoods of the true density.

I ask whether a two-step semiparametric estimator is uniformly adaptive. In Section 2, I describe in detail a two-step estimator. Neighborhoods of the truth, details of the loss function, and asymptotic efficiency bounds over the neighborhoods are in Section 3. In Section 4, I prove that a two-step semiparametric estimator for all parameters in a nonlinear ARMA model is uniformly adaptive if the unknown density is symmetric about zero. Proofs of Lemmas 1-4 and Theorem 1 are contained in the appendix.

2. A Two-Step Semiparametric Estimator

The first-step estimator, denoted $\bar{\psi}_T$, is the quasi-maximum likelihood estimator under the assumption that g is a normal density. The residuals from the first-step estimator, $\left\{u_t(\bar{\psi}_T)\right\}_{t=1}^{T}$, are needed to construct a nonparametric density estimator. The equation for a period-t residual is obtained by first inverting the moving-average polynomial $\theta(z) = \sum_{j=0}^{q}\theta_j z^j$, where $\theta_0 = 1$, to obtain an autoregression of infinite order that has coefficients γ. Construction of a period-t residual from an autoregression of infinite order is difficult because it requires an infinite number of past values of e_t. To overcome this difficulty, a recursion relating γ to the coefficient vector θ is derived (a complete derivation is in the appendix) that reduces the order of the autoregression so that a period-t residual is a function only of data observable at period t:

$$u_t(\psi) = \sum_{i=0}^{t-1}\gamma_i(e_{t-i} - \rho_1 e_{t-1-i} - \cdots - \rho_p e_{t-p-i}) + \sum_{s=0}^{q-1} u_{-s}\sum_{k=0}^{s}\theta_k\gamma_{t+s-k}. \qquad (2.1)$$

The period-t residual from $\bar{\psi}_T$ is obtained by replacing the values of $\{\gamma', \beta', \rho', \theta'\}$ with $\{\bar{\gamma}_T', \bar{\beta}_T', \bar{\rho}_T', \bar{\theta}_T'\}$ in (2.1), where $\bar{\beta}_T$, $\bar{\rho}_T$, and $\bar{\theta}_T$ are the elements of $\bar{\psi}_T$ that correspond to β, ρ, and θ, respectively, and $\bar{\gamma}_T$ is the vector of coefficients from a power-series expansion of $\bar{\theta}_T(z)^{-1}$.

The second step begins with the construction of a nonparametric estimator of g, denoted $\hat{g}$. The estimator $\hat{g}$ is a kernel density estimator defined for all u in a neighborhood of each $u_t(\bar{\psi}_T)$ as:

$$\hat{g}_t(u, \bar{\psi}_T) = (T-1)^{-1}\sum_{s\neq t}^{T} J(u - u_s(\bar{\psi}_T); s_T),$$

where $J(u)$ is the kernel and s_T controls the degree of smoothing.[2] Let $\hat{g}_t^{(1)} = \partial\hat{g}_t/\partial u$, $\hat{g}_t^{(1)}(u, \bar{\psi}_T)/\hat{g}_t(u, \bar{\psi}_T) = \hat{\tau}_t(u, \bar{\psi}_T)$, and let $\hat{\tau}_t(u_t(\bar{\psi}_T), \bar{\psi}_T) \equiv \hat{\tau}_t(u_t(\bar{\psi}_T))$ and $\hat{g}_t(u_t(\bar{\psi}_T), \bar{\psi}_T) \equiv \hat{g}_t(u_t(\bar{\psi}_T))$.[3]

The estimator is constructed from the sample score and information matrix for ψ, denoted $S_\psi(z_T, \psi, g)$ and $\mathcal{J}_\psi(z_T, \psi, g)$, respectively, which depend on T-period sample, parameter value, and density function g. The sample score and information matrix for ψ (evaluated at $\bar{\psi}_T$) are estimated by

$$S_\psi(z_T, \bar{\psi}_T, \hat{g}) = T^{-\frac{1}{2}}\sum_{t=1}^{T}\hat{\tau}_t(u_t(\bar{\psi}_T))d_t(\bar{\psi}_T),$$
$$\mathcal{J}_\psi(z_T, \bar{\psi}_T, \hat{g}) = \mathcal{I}(z_T, \bar{\psi}_T, \hat{g})\Gamma_\psi(z_T, \bar{\psi}_T),$$

respectively, where $\mathcal{I}(z_T, \bar{\psi}_T, \hat{g}) = T^{-1}\sum_{t=1}^{T}\hat{\tau}_t(u_t(\bar{\psi}_T))^2$, and $\Gamma_\psi(z_T, \bar{\psi}_T)$ equals $T^{-1}\sum_{t=1}^{T} d_t(\bar{\psi}_T)d_t(\bar{\psi}_T)'$ with $d_t(\bar{\psi}_T) = \frac{\partial u_t(\psi)}{\partial\psi}\,|_{\bar{\psi}_T}$ and $u_t(\psi)$ given by (2.1). To ensure that these estimators are well behaved, extreme values of $\hat{\tau}_t(u, \bar{\psi}_T)$ are

[2]The asymptotic distribution of the semiparametric estimator is independent of kernel choice.

[3]Throughout, functions superscripted by (i) are ith derivatives of the function.

trimmed. That is, $\hat{\tau}_t(u,\bar{\psi}_T)$ equals $\hat{g}_t^{(1)}(u,\bar{\psi}_T)/\hat{g}_t(u,\bar{\psi}_T)$ if: (i) $\hat{g}_t(u,\bar{\psi}_T) \geq \lambda_{1T}$; (ii) $|u| \leq \lambda_{2T}$; and (iii) $|\hat{g}_t^{(1)}(u,\bar{\psi}_T)| \leq \lambda_{3T}\hat{g}_t(u,\bar{\psi}_T)$, and $\hat{\tau}_t(u,\bar{\psi}_T)$ equals zero otherwise. The parameters $(\lambda_{1T},\lambda_{2T},\lambda_{3T})$ are trimming parameters.

For technical reasons I restrict attention to discretized estimators developed by LeCam (1970). A discretized estimator, denoted $\bar{\psi}_T^d$, is defined as the nearest vertex to $\bar{\psi}_T$ in the c-dimensional lattice of integers scaled by $T^{-\frac{1}{2}}$. The advantage of using discretized estimators is that to derive the limiting properties of a two-step semiparametric estimator I need fewer differentiability and boundedness assumptions. Kreiss (1987) Lemma 4.4 proves that for any sequence of random variables $a_T(\cdot)$, if $a_T(\psi^T) = o_P(1)$ with $\left|T^{\frac{1}{2}}\left(\psi^T - \psi\right)\right| \leq c$ for some constant $c > 0$, then $a_T(\bar{\psi}_T^d) = o_P(1)$ for any discrete $T^{\frac{1}{2}}$-consistent estimator $\bar{\psi}_T^d$.

The two-step semiparametric estimator is then

$$\hat{\psi}_T = \bar{\psi}_T^d + T^{-\frac{1}{2}}\left[\mathcal{J}_\psi(z_T,\bar{\psi}_T^d,\hat{g})\right]^{-1} S_\psi(z_T,\bar{\psi}_T^d,\hat{g}). \tag{2.2}$$

To understand intuitively why the two-step semiparametric estimator is sensible, consider a two-step estimator for which the T-period sample score and information matrix are constructed under the assumption that g is normal. The resulting estimator is consistent even if g is not normal. The two-step semiparametric estimator retains this consistency and, because $\hat{g}$ converges pointwise to g, is asymptotically more efficient than a two-step estimator that is constructed under a fixed incorrect assumption for g.[4]

3. Asymptotic Efficiency Bound

The neighborhoods of the truth shrink as the sample size grows because it is known that semiparametric estimators do not converge uniformly over neighborhoods that are fixed independent of sample size. The neighborhood, Ψ, of ψ is the union $\Psi = \cup_m\{\Psi(m)\}$ of sets

$$\Psi(m) = \{\{\psi^T\} :| \sqrt{T}(\psi^T - \psi) - m | \rightarrow 0 \text{ as } T \rightarrow \infty\},$$

where $|\cdot|$ is the Euclidean norm and $m \in \mathbb{R}^c$. The elements of m conform in dimension to ψ so that $m = (m_\beta', m_\rho', m_\theta')'$.

As is standard, I use the square-root of the density in defining the neighborhood of the true density to ensure that the derivative of the log-likelihood with respect to the square-root of the density is square integrable.[5] Let $\tau(u) = g^{(1)}(u)/g(u)$ and let $\tau^T(u) = g^{T(1)}(u)/g^T(u)$. The neighborhood, $\mathcal{C}$, of g is the

[4]The two-step estimators that I consider take one step from the initial estimator. Of course, the procedure could be iterated, leading to estimators that take more than one step from the initial estimator. The first-order asymptotic theory is identical for these estimators, but the finite sample performance may differ.

[5]Because the density integrates to one, the square-root of the density, and hence the derivative of the log-likelihood with respect to the square-root of the density, is square integrable.

union $\mathcal{C} = \cup_\varsigma\{\mathcal{C}(\varsigma)\}$ of sets

$$\mathcal{C}(\varsigma) = \{\{g^T\} : \| \sqrt{T}(\sqrt{g^T} - \sqrt{g}) - \varsigma \|_v \to 0 \text{ and } E\left[\tau(g^T) - \tau(g)\right]^2 \to 0 \text{ as } T \to \infty\},$$

where $g^T \in \mathcal{G}$ and where ς is square integrable and is orthogonal to $\sqrt{g}$.[6]

The continuous and bounded loss function l is defined to be subconvex, that is, that $\{z : l(z) \leq b\}$ is closed, convex, and symmetric for all $b \geq 0$.

To derive the asymptotic efficiency bound over the neighborhoods of the truth, I make the following three assumptions. The first two assumptions restrict the parameter space and function space over which Ψ and $\mathcal{C}$ are defined. Let $\psi \in B \subset \mathbf{R}^c$, with $c = \dim(\psi)$, and partition B as $B_1 \times B_2$ where $\beta \in B_1$ and $(\rho', \theta')' \in B_2$. Let the lag polynomial $\rho(z)$ be $\rho(z) = \sum_{i=0}^{p} \rho_i z^i$, where $\rho_0 = 1$. Let $w \in \mathbf{R}^c$ and let $A(c) = \{w : w'w \leq c\}$. Finally, let $S_\psi(\psi, g) = \lim_{T\to\infty} ES_\psi(z_T, \psi, g)$ and let $\mathcal{J}_\psi(\psi, g) = \lim_{T\to\infty} E\mathcal{J}_\psi(z_T, \psi, g)$.

Assumption 1:

a) B_1 is a compact subspace of $\mathbf{R}^{k+1}$.

b) B_2 is a subspace of $\mathbf{R}^{p+q}$ and the elements of B_2 satisfy:

$\rho(z) \neq 0$ and $\theta(z) \neq 0$ for all $|z| \leq 1$;

$\rho(z)$ and $\theta(z)$ have 1 as their greatest common left divisor.

Assumption 2:

a) for all $g \in \mathcal{G}$, $g^T \in \mathcal{G}$, g and g^T are mutually absolutely continuous.

for all $g \in \mathcal{G}$:

b) g is absolutely continuous with respect to Lebesgue measure v.

c) $g(u) > 0$ for all $u \in \mathbf{R}^1$.

d) $\mathcal{I}(g) = \int [g^{(1)}(u)]^2 / g(u) du < \infty$.

e) $\lim_{w\to 0} \int [\tau(u - w) - \tau(u)]^2 g(u) du = 0$.

f) $\lim_{w\to 0} w^{-1} \int [\tau(u - w) - \tau(u)] g(u) du = \mathcal{I}(g)$.

g) $\lim_{c\to 0} \sup_{A(c)} w^{-1} \int \tau(u + w) g(u) du = -\mathcal{I}(g)$.

h) g is symmetric about zero.

Remark: Assumption 1b ensures that the ARMA process for e_t is stationary and invertible with no common roots in lag polynomials. Assumptions 2d-2g are smoothness restrictions on $g(\cdot)$ of the "quadratic-mean differentiability type". Lind and Roussas (1977) relate quadratic-mean differentiability condi-

[6] I use $\| \cdot \|_v$ and $\| \cdot \|_\mu$ to denote the norms in $\mathcal{L}^2(\nu)$ and $\mathcal{L}^2(\mu)$, respectively, where $\mathcal{L}^2(\nu)$ and $\mathcal{L}^2(\mu)$ are spaces of square-integrable functions with respect to the measures v and μ.

tions to Cramér-type conditions; the latter are pointwise derivative conditions for $\ln g(\cdot)$. Although Kreiss (1987) is able to relax the assumption that $g(\cdot)$ be symmetric about zero for autoregressive models, Manski (1984) shows that for models of the type I consider Assumption 2h cannot be relaxed. Specifically, unless a free intercept is included in the conditional mean, adaptivity requires Assumption 2h.

The third assumption restricts the density f.[7]

Assumption 3:

a) f is absolutely continuous with respect to Lebesgue measure μ.

b) $f(\cdot,\psi,\cdot) > 0$ for all $T \in \{1,2,...\}$ and $\psi \in B$.

c) $f(\cdot,\psi_0,\cdot) \neq f(\cdot,\psi_1,\cdot)$ whenever at least one element of ψ_0 does not equal the corresponding element of ψ_1.

d) $\sqrt{f}$ is Hellinger-differentiable at (ψ, g).

Remark: Assumption 3d ensures that $f^{\frac{1}{2}}$ is asymptotically quadratic over local neighborhoods of the truth. To understand Hellinger-differentiability, note that Assumption 3d is implied by the following two conditions. The first condition, which restricts the temporal dependence introduced by the ARMA error, is that $\| \sqrt{T}(\sqrt{f^T} - \sqrt{\tilde{f}^T}) \|_\mu \to 0$ as $T \to \infty$, where $\tilde{f}^T = \tilde{f}(\cdot, \alpha^T, g^T)$ is the density for $\tilde{y}$ defined as $\tilde{y}_t = \alpha^T + u_t$. The second condition, which ensures that the density $\tilde{f}^T$ is differentiable with respect to both the parameter vector and the density for u_t, is that there exists a function A, the score for α, that is square integrable, and there exists a bounded linear operator, $\mathcal{S}$, on the set of square-integrable functions such that, with $\tilde{f}^T = \tilde{f}(\cdot, \alpha^T, g^T)$ and $\tilde{f} = \tilde{f}(\cdot, \alpha, g)$:

$$\frac{\| \sqrt{\tilde{f}^T} - \sqrt{\tilde{f}} - (A \cdot (\alpha^T - \alpha) + \mathcal{S} \cdot (\sqrt{g^T} - \sqrt{g})) \|_\mu}{| \alpha^T - \alpha | + \| \sqrt{g^T} - \sqrt{g} \|_\nu} \to 0,$$

as $T \to \infty$ for all sequences $\alpha^T \to \alpha$ and $\sqrt{g^T} \to \sqrt{g}$ in $\mathcal{L}^2(\nu)$ where $g^T \in \mathcal{G}$.

To combine the local neighborhoods of the true parameter values and the true density into a neighborhood of f, note that if $f^{\frac{1}{2}}$ is Hellinger-differentiable at (ψ, g) and $\{(\psi^T, g^T)\}_{T\geq 1} \in \Psi(m) \times \mathcal{C}(\varsigma)$ for some $m \in \mathbb{R}^c$ and $\varsigma \in \mathcal{L}^2(\upsilon)$, then

$$\| \sqrt{T}(\sqrt{f^T} - \sqrt{f}) - a \|_\mu \to 0, \tag{3.1}$$

as $T \to \infty$ with $a \in \mathcal{L}^2(\mu)$ and $a = Sm + \mathcal{S}\varsigma$, where the square-integrable function S is the score for ψ. Let $H = \{a \in \mathcal{L}^2(\mu) : a = Sm + \mathcal{S}\varsigma$ for some $m \in \mathbb{R}^c$ and

[7] I separate the assumptions on $f(\cdot,\cdot)$ and $g(\cdot)$ because the regressors are assumed to be weakly exogenous rather than fixed (Assumption 4a), so the distribution for y_t is not a simple transformation of the distribution for u_t.

$\varsigma \in \mathcal{L}^2(\upsilon)\}$. For $a \in H$, let $\mathcal{F}(a)$ be the collection of all sequences $\{f^T\}$ such that (3.1) holds. Let $\mathcal{F} = \cup_a\{\mathcal{F}(a)\}$. The local neighborhood of f is given by $\mathcal{F}^T(c) = \{f^T \in \mathcal{F} : \sqrt{T} \parallel \sqrt{f^T} - \sqrt{f} \parallel_\mu \leq c\}$.

An immediate consequence of Hellinger differentiability of $\sqrt{f}$ is that the log-likelihood ratio over local neighborhoods of the truth, defined for a sample of T observations as

$$L^T = \ln \frac{f_0(u_{1-q}, \ldots, u_0, e_{1-p}, \ldots, e_0; \psi^T, g^T)}{f_0(u_{1-q}, \ldots, u_0, e_{1-p}, \ldots, e_0; \psi, g)} \frac{f^T(y_1, \ldots, y_T)}{f(y_1, \ldots, y_T)}, \tag{3.2}$$

is asymptotically normal. That is, the log-likelihood ratio is locally asymptotically normal.

Lemma 1. *If Assumptions 1-3 are satisfied, then the log-likelihood ratio defined in (3.2) is locally asymptotically normal.*

Computation of the asymptotic efficiency bound also depends on Hellinger differentiability of $f^{\frac{1}{2}}$. Because $f^{\frac{1}{2}}$ is Hellinger-differentiable at (ψ, g) and each ς that satisfies the definition of $\mathcal{C}(\varsigma)$ is square integrable, Theorem 3.2 from Begun et al. (1983) implies that the asymptotic efficiency bound for estimators of ψ based on the assumption that g is unknown (the estimators are generically denoted as $\tilde{\psi}_T$) is

$$\lim_{c\to\infty} \lim_{T\to\infty} \inf_{\tilde{\psi}_T} \sup_{f^T \in \mathcal{F}^T(c)} E_{f^T}\{l[\sqrt{T}(\tilde{\psi}_T - \psi^T)]\} \geq E\{l(Z)\}, \tag{3.3}$$

where $Z \sim N(0, V(\psi, g)^{-1})$.

From (1.3), a two-step semiparametric estimator $\hat{\psi}_T$ is uniformly adaptive if it attains the asymptotic efficiency bound (3.3) and if $V(\psi, g) = \mathcal{J}_\psi(\psi, g)$, where $\mathcal{J}_\psi(\psi, g)^{-1}$ is the covariance matrix of the asymptotic distribution of the MLE. To understand the role of the condition $V(\psi, g) = \mathcal{J}_\psi(\psi, g)$, note that parametric estimators do not make use of information about ψ contained in the unknown density; this information equals the population projection of $S_\psi(\psi, g)$ onto the tangent set, $\mathcal{T}$. The tangent set consists of the scores for all possible unknown densities, so it consists of functions of u_t because each $\mathcal{S}$ is a function of u_t.[8] Let $R_\psi(\psi, g)$ be the vector of residuals from the population projection of $S_\psi(\psi, g)$ onto $\mathcal{T}$. Then $R_\psi(\psi, g)$ represents the information about ψ contained only in the parametric part of the specification, so $V(\psi, g) = R_\psi(\psi, g)R_\psi(\psi, g)'$. If the population projection of $S_\psi(\psi, g)$ onto $\mathcal{T}$ is zero, the unknown density contains no information about ψ, so $R_\psi(\psi, g) = S_\psi(\psi, g)$ and $V(\psi, g) = \mathcal{J}_\psi(\psi, g)$. Thus if $S_\psi(\psi, g)$ is orthogonal to $\mathcal{T}$, then $V(\psi, g) = \mathcal{J}_\psi(\psi, g)$ and the asymptotic efficiency bound is given by the asymptotic expected loss of the MLE.

[8]Formally, the tangent set is the mean-square closure, over all possible underlying densities, of the set of linear combinations of S that conform in dimension to S.

4. Uniformly Adaptive Estimation

To provide sufficient conditions for $\hat{\psi}_T$ to be uniformly adaptive I begin by proving that $\hat{\psi}_T$ is regular. Specifically, $\hat{\psi}_T$ is regular if

$$T^{\frac{1}{2}}(\hat{\psi}_T - \psi) - \mathcal{J}_\psi(\psi, g)^{-1} S_\psi(z_T, \psi, g) = o_P(1). \tag{4.1}$$

The proof that $\hat{\psi}_T$ is regular uses the following assumptions on the conditional mean, the moments of u_t, and the nonparametric density estimator. Let $h^{(1)}_{\beta_i}$ be the derivative of h with respect to β_i, let $h^{(2)}_{\beta_i\beta_j}$ be the derivative of h with respect to β_i and β_j, and let M^s_{ij} be a $(k+1) \times (k+1)$ matrix.

Assumption 4:

a) $\{x_t\}_{t=1}^T$ *is weakly exogenous for* ψ.

b) $h(\cdot, \beta)$ *is twice continuously differentiable with respect to* β *for all* $\beta \in B_1$.

c) $T^{-1}\sum_{t=1}^T h^{(1)}_{\beta_i}(x_{t+s}, \beta) h^{(1)}_{\beta_j}(x_t, \beta) \to M^s_{ij}$ *for each* $i, j \in (0, ..., k)$ *and every integer* s *as* $T \to \infty$ *uniformly in* $\beta \in B_1$, *where* $M^s_{ij} \to 0$ *as* $s \to \infty$.

d) $\max_t T^{-\frac{1}{2}} h^{(2)}_{\beta_l\beta_l}(x_t, \cdot) \to 0$ *for each* $l \in (0, ..., k)$ *as* $T \to \infty$ *uniformly in* $\beta \in B_1$.

e) $\left|h^{(1)}(x_t, \beta)\right| \leq c_1 + c_2 |x_t|$.

f) $\lim_{t\to\infty} \sup_T \sum_{i=0}^{t-1} \gamma_i^T x_{t-i} < \infty$.

g) $\left|\frac{\partial}{\partial\beta} h^{(1)}_{\beta_l}(x_t, \beta^T)\right| \leq d_1 + d_2 |x_t|$ *for each* $l \in (0, ..., k)$ *uniformly in* T.

h) $E[u_t] = 0$, $E[u_t^2] > 0$, $E[u_t^4] < \infty$.

Remark: Assumptions 4c-4f ensure that the nonlinear conditional mean function satisfies conditions for consistent estimation of the covariance matrix (note that $\sum_{i=0}^\infty \mid \gamma_i^T \mid < \infty$ from Lemma 6.1 in Kreiss (1987)). Note that the regressors do not have to be fixed according to Assumption 4; they need only be weakly exogenous. Assumption 4e allows specifications of the form $y_t = \beta_0 + \beta_1 z_t^3 + e_t$. For this specification $x_t = z_t^3$ and $h(\cdot, \cdot)$ is a linear function.

The fifth assumption restricts the nonparametric density estimator.

Assumption 5:

a) $J(\cdot)$ *is twice continuously differentiable.*

b) $(s_T, \lambda_{1T}, s_T\lambda_{3T}) \to 0$, $(\lambda_{2T}, \lambda_{3T}) \to \infty$, $s_T^{-4}\lambda_{2T} = o(T)$, *and* $s_T^{\frac{9}{2}} = O(\frac{1}{T})$ *as* $T \to \infty$.

Remark: The last two conditions in Assumption 5b are the result of the dependence that arises because both $d_t(\cdot)$ and $\hat{g}_t(\cdot,\cdot)$ are functions of $\{u_t(\cdot)\}_{t=1}^T$.

To prove that a two-step semiparametric estimator is regular, I require the following three lemmas.

Lemma 2. *If Assumptions 1-5 are satisfied and $\bar{\psi}_T^d$ is a discrete $T^{\frac{1}{2}}$-consistent estimator of ψ, then*

$$\mathcal{I}(z_T, \bar{\psi}_T^d, \hat{g}) - \mathcal{I}(g) = o_P(1) \text{and } \Gamma_\psi(z_T, \bar{\psi}_T^d) - \Gamma_\psi(\psi) = o_P(1),$$

so that $\mathcal{J}_\psi(z_T, \bar{\psi}_T^d, \hat{g})$ is a consistent estimator of $\mathcal{J}_\psi(\psi, g)$.
The third lemma establishes consistency of the semiparametric score.

Lemma 3. *If Assumptions 1-5 are satisfied and $\bar{\psi}_T^d$ is a discrete $T^{\frac{1}{2}}$-consistent estimator of ψ, then*

$$S_\psi(z_T, \bar{\psi}_T^d, \hat{g}) - S_\psi(z_T, \bar{\psi}_T^d, g) = o_P(1).$$

The next lemma establishes asymptotic linearity.

Lemma 4. *If Assumptions 1-5 are satisfied, then*

$$\mathcal{J}_\psi(\psi, g)T^{\frac{1}{2}}(\bar{\psi}_T^d - \psi) + S_\psi(z_T, \bar{\psi}_T^d, g) - S_\psi(z_T, \psi, g) = o_P(1),$$

with $\psi^T = \psi + T^{-\frac{1}{2}}m$ for any $m \in \mathbf{R}^c$.

With these preliminary lemmas, I prove that a two-step semiparametric estimator is regular.

Lemma 5. *If Assumptions 1-5 are satisfied, then the two-step semiparametric estimator $\hat{\psi}_T$ defined in (2.2) is regular.*

I now prove that local asymptotic normality of the log-likelihood ratio and regularity of $\hat{\psi}_T$ are sufficient for $\hat{\psi}_T$ to be uniformly adaptive.

Theorem 1. *If Assumptions 1-5 are satisfied, then the two-step semiparametric estimator $\hat{\psi}_T$ is uniformly adaptive for nonlinear ARMA models of (1.1)-(1.2).*

Remark: To understand why a two-step semiparametric estimator is adaptive for the model of (1.1)-(1.2), note that because $\mathcal{G}$ is a family of densities that are symmetric about zero under Assumption 2, $\mathcal{S}$ is an even function of u_t so $\mathcal{T}$ consists of even functions of u_t. Also because $\mathcal{G}$ is a family of densities that are symmetric about zero, $\tau(u_t)$ is an odd function of u_t. Therefore $\tau(u_t)$ is orthogonal to $\mathcal{T}$, so the projection of $\tau(u_t)$ onto $\mathcal{T}$ is zero. Because the remaining component of $S_\psi(\psi, g)$ is independent of u_t, $S_\psi(\psi, g)$ is orthogonal to $\mathcal{T}$ and the asymptotic efficiency bound is given by the asymptotic expected loss of the MLE.

5. Conclusion

For the parameters of a nonlinear regression model with ARMA errors, I establish the minimum asymptotic expected loss for a semiparametric estimator. The asymptotic expected loss is constructed over neighborhoods of the true parameter values and the true density. Because the minimum asymptotic expected loss of a semiparametric estimator is equivalent to the asymptotic expected loss of the MLE, the semiparametric estimator is uniformly adaptive.

ACKNOWLEDGMENTS

I thank Zari Rachev and an anonymous referee for helpful comments.

References

[1] Begun, J., W. Hall, W. Huang, and J. Wellner. (1983) Information and asymptotic efficiency in parametric-nonparametric models. *Annals of Statistics* 11: 432-452.

[2] Bickel, P. (1982) On adaptive estimation. *Annals of Statistics* 10: 647-671.

[3] Billingsley, P. (1968) *Convergence of Probability Measures.* New York: John Wiley and Sons.

[4] Fabian, V. and J. Hannan. (1982) On estimation and adaptive estimation for locally asymptotically normal families. *Probability Theory and Related Fields* 59: 459-478.

[5] Hájek, J. and Z. Sidák. (1967) *Theory of Rank Tests.* New York: Academic Press.

[6] Harvey, A. (1981) *The Econometric Analysis of Time Series.* Oxford: Philip Allan Publishers.

[7] Kreiss, J. (1987) On adaptive estimation in stationary ARMA processes. *Annals of Statistics* 15: 112-133.

[8] Kreiss, J. (1987) On adaptive estimation in autoregressive models when there are nuisance functions. *Statistics and Decisions* 5: 59-76.

[9] LeCam, L. (1970) On the assumption used to prove asymptotic normality of maximum likelihood estimators. *Annals of Mathematical Statistics* 41: 802-828.

[10] Lind, B. and G. Roussas. (1977) Cramér-type conditions and quadratic mean differentiability. *Annals of the Institute of Statistical Mathematics* 29: 189-201.

[11] Manski, C. (1984) Adaptive estimation of non-linear regression models. *Econometric Reviews* 3: 145-194.

[12] Steigerwald, D. (1992) Adaptive estimation in time series regression models. *Journal of Econometrics* 54: 251-275.

[13] Steigerwald, D. (1996) Uniformly adaptive estimation for models with ARMA errors. Manuscript. Department of Economics. University of California, Santa Barbara.

A. Appendix

To derive the formula for the residual given in (2.1), note that from the definition of the ARMA error

$$u_t = (1 + \theta_1 L + \cdots + \theta_q L^q)^{-1}(e_t - \rho_1 e_{t-1} - \cdots \rho_p e_{t-p}), \tag{A.1}$$

where L is the lag operator. Under Assumption 1, for each $(\cdot, \theta) \in B_2$ there exists an $\eta > 1$ such that $\theta(z)^{-1} = \sum_{i=0}^{\infty} \gamma_i z^i$, with $\gamma_0 = 1$, for all $| z |< \eta$. Replace $\theta(L)^{-1}$ with $\sum_{i=0}^{\infty} \gamma_i L^i$ in (A.1) and note that $\sum_{j=0}^{p} \rho_j e_{t-j} = \sum_{k=0}^{q} \theta_k u_{t-k}$:

$$u_t(\psi) = \sum_{i=0}^{t-1} \gamma_i L^i \sum_{j=0}^{p} \rho_j e_{t-j} + \sum_{i=t}^{\infty} \gamma_i L^i \sum_{k=0}^{q} \theta_k u_{t-k}. \tag{A.2}$$

The power series coefficients γ satisfy the recursion

$$\gamma_i + \theta_1 \gamma_{i-1} + \cdots + \theta_q \gamma_{i-q} = 0, \tag{A.3}$$

for all $i > 0$ where $\gamma_s = 0$ if $s < 0$, which in turn implies that $\sum_{s=0}^{q-1} u_{-s} \sum_{k=0}^{s} \theta_k \gamma_{t+s-k}$ $\sum_{i=t}^{\infty} \gamma_i L^i \sum_{k=0}^{q} \theta_k u_{t-k}$. Replacing $\sum_{i=t}^{\infty} \gamma_i L^i \sum_{k=0}^{q} \theta_k u_{t-k}$ with $\sum_{s=0}^{q-1} u_{-s} \sum_{k=0}^{s} \theta_k \gamma_{t+s-k}$ on the right-hand side of (A.2) yields (2.1).

In Steigerwald (1996) it is shown that

$$u_t(\psi) - u_t(\psi^T) =$$

$$(\psi^T - \psi)' \sum_{i=0}^{t-1} \gamma_i^T v_t(\psi^T) + O_P(T) = -\{d_t(\psi^T)'[\mathcal{J}_\psi(\psi, g)T]^{-\frac{1}{2}} m_T + O_P(T)\}, \tag{A.4}$$

where $v_t(\psi^T)' = [h_{\beta_0}^{(1)}(x_{t-i}, \beta^T) - \sum_{j=1}^{p} \rho_j^T h_{\beta_0}^{(1)}(x_{t-i-j}, \beta^T), \ldots, h_{\beta_k}^{(1)}(x_{t-i}, \beta^T) - \sum_{j=1}^{p} \rho_j^T h_{\beta_k}^{(1)}(x_{t-i-j}, \beta^T), e_{t-i-1}(\psi^T), \ldots, e_{t-i-q}(\psi^T), u_{t-i-1}(\psi^T), \ldots, u_{t-i-q}(\psi^T)]$.

In the proofs that follow: I_c is the $c \times c$ identity matrix; $\sum_t$ denotes $\sum_{t=1}^{T}$; and $\rightarrow^P$ denotes convergence in probability.

A.1. Proof of Lemma 1

Because $f^{\frac{1}{2}}$ is Hellinger-differentiable at (ψ, g), the densities f and f^T satisfy (3.1). The result then follows from Lemma 2.1 in Begun et al. (1983). □

A.2. Proof of Lemma 2

Let $\mathcal{J}_\psi(\psi, g) = \mathcal{I}(g)\Gamma_\psi(\psi)$, where $\Gamma_\psi(\psi) = E\left[\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\gamma_i\gamma_j v_1(\psi)v_1(\psi)'\right]$. I first show that $\mathcal{I}(z_T, \bar{\psi}_T^d, \hat{g}) - \mathcal{I}(g) = T^{-1}\sum_t \hat{\tau}_t(u_t(\bar{\psi}_T^d))^2 - \mathcal{I}(g) = o_P(1)$. If ψ^T is the true value, then $\{\tau(u_t(\psi^T))\}_{t\geq 1}$ are independent and identically distributed random variables, so $T^{-1}\sum_t \tau(u_t(\psi^T))^2 - \mathcal{I}(g) = o_P(1)$. By Lemma 4.1 in Bickel (1982) $T^{-1}\sum_t \hat{\tau}_t(u_t(\psi^T))^2 - \mathcal{I}(g) = o_P(1)$. Because L^T is locally asymptotically normal (Lemma 1), convergence with (ψ^T, g) as true values implies convergence with (ψ, g) as true values. Finally, by Lemma 4.4 in Kreiss (1987) I replace ψ^T with $\bar{\psi}_T^d$ in $T^{-1}\sum_t \hat{\tau}_t(u_t(\psi^T))^2 - \mathcal{I}(g) = o_P(1)$, yielding $T^{-1}\sum_t \hat{\tau}_t(u_t(\bar{\psi}_T^d))^2 - \mathcal{I}(g) = o_P(1)$.

I next establish that $T^{-1}\sum_t d_t(\bar{\psi}_T^d)d_t(\bar{\psi}_T^d)' - \Gamma_\psi(\psi) = o_P(1)$. The vector $d_t(\bar{\psi}_T^d)$ is partioned such that the first $k+1$ components (for β) are of the form

$$d_{\beta_l,t}(\bar{\psi}_T^d) = \sum_{i=0}^{t-1}\bar{\gamma}_{iT}^d\left[h_{\beta_l}^{(1)}(x_{t-i}, \bar{\beta}_T^d) - \sum_{j=1}^{p}\bar{\rho}_{j,T}^d h_{\beta_l}^{(1)}(x_{t-i-j}, \bar{\beta}_T^d)\right],$$

the next p components (for ρ) are of the form

$$d_{\rho_j,t}(\bar{\psi}_T^d) = \sum_{i=0}^{t-1}\bar{\gamma}_{iT}^d e_{t-i-j}(\bar{\psi}_T^d),$$

and the last q components (for θ) are of the form

$$d_{\theta_k,t}(\bar{\psi}_T^d) = \sum_{i=0}^{t-1}\bar{\gamma}_{iT}^d u_{t-i-k}(\bar{\psi}_T^d).$$

Kreiss expression (6.11) establishes convergence for the last $p+q$ components of $d_t(\psi^T)$ (Lemma 6.1 in Kreiss is necessary for the result). Under Assumptions 4e and 4f the behavior of $h_{\beta_l}^{(1)}(x_{t-i}, \beta^T) - \sum_{j=1}^{p}\rho_j^T h_{\beta_l}^{(1)}(x_{t-i-j}, \beta^T)$ is restricted so that by Lemma 6.1 from Kreiss $\{d_{\beta_l,t}(\psi^T)\}_{t\geq 1}$ forms a convergent sequence. Hence the result for the last $p+q$ components of $d_t(\psi^T)$ is preserved for the first $k+1$ components of $d_t(\psi^T)$, which are the components corresponding to β, so by Lemma 4.4 from Kreiss $T^{-1}\sum_t d_t(\bar{\psi}_T^d)d_t(\bar{\psi}_T^d)' - \Gamma_\psi(\psi) = o_P(1)$. ◫

A.3. Proof of Lemma 3

By definition $E\left\|S_\psi(z_T, \psi^T, \hat{g}) - S_\psi(z_T, \psi^T, g)\right\|^2$ equals

$$\left(T^{-1}\left\{\sum_t d_t(\psi^T)\left[\hat{\tau}_t\left(u_t(\psi^T)\right) - \tau\left(u_t(\psi^T)\right)\right]\right\}'\left\{\sum_t d_t(\psi^T)\left[\hat{\tau}_t\left(u_t(\psi^T)\right) - \tau\left(u_t(\psi^T)\right)\right]\right\}\right)$$
$$= \sum_{j=1}^{c} T^{-1}E\{\sum_t d_{tj}(\psi^T)\left[\hat{\tau}_t\left(u_t(\psi^T)\right) - \tau\left(u_t(\psi^T)\right)\right]\}^2.$$

Theorem 5.1 in Kreiss establishes that $T^{-1}E\{\sum_t d_{tj}(\psi^T)\left[\hat{\tau}_t\left(u_t(\psi^T)\right) - \tau\left(u_t(\psi^T)\right)\right]\}^2 = o_P(1)$ for $j = k+2, \ldots, c$. To ensure that this result holds for $j = 1, \ldots, k+1$, I need only show that $\sum_{l=0}^{k}\left\{\sum_{i=0}^{t-1}\gamma_i^T\left[h_{\beta_l}^{(1)}(x_{t-i},\beta^T)-\sum_{j=1}^{p}\rho_j^T h_{\beta_l}^{(1)}(x_{t-i-j},\beta^T)\right]\right\}^2$ is bounded uniformly in t. As shown in the proof of Lemma 2, Assumptions 4e and 4f ensure that $\sum_{i=0}^{t-1}\gamma_i^T\left[h_{\beta_l}^{(1)}(x_{t-i},\beta^T)-\sum_{j=1}^{p}\rho_j^T h_{\beta_l}^{(1)}(x_{t-i-j},\beta^T)\right]$ is bounded uniformly in t because k is fixed independently of t. Thus

$$E\left\|S_\psi(z_T,\psi^T,\hat{g})-S_\psi(z_T,\psi^T,g)\right\|^2=o_P(1),$$

so by Lemma 4.4 in Kreiss

$$S_\psi(z_T,\bar{\psi}_T^d,\hat{g})-S_\psi(z_T,\bar{\psi}_T^d,g)=o_P(1).$$

□

A.4. Proof of Lemma 4

Observe that $\mathcal{J}_\psi(\psi,g)T^{\frac{1}{2}}(\bar{\psi}_T^d-\psi)+S_\psi(z_T,\bar{\psi}_T^d,g)-S_\psi(z_T,\psi,g)$ equals

$$Q_\psi(z_T,\bar{\psi}_T^d)-Q_\psi(z_T,\psi)+Y_\psi(z_T,\bar{\psi}_T^d),$$

where I use the notation

$$\begin{gathered}Q_\psi(z_T,\bar{\psi}_T^d)=T^{-\frac{1}{2}}\textstyle\sum_t\{\tau(u_t(\bar{\psi}_T^d))d_t(\bar{\psi}_T^d)-E[\tau(u_t(\bar{\psi}_T^d))d_t(\bar{\psi}_T^d)\mid\mathcal{F}_{t-1}]\},\\ Y_\psi(z_T,\bar{\psi}_T^d)=T^{-\frac{1}{2}}\textstyle\sum_t\{E[\tau(u_t(\bar{\psi}_T^d))d_t(\bar{\psi}_T^d)\mid\mathcal{F}_{t-1}]-E[\tau(u_t(\psi))d_t(\psi)\mid\mathcal{F}_{t-1}]\}\\ +\mathcal{J}_\psi(\psi,g)T^{\frac{1}{2}}(\bar{\psi}_T^d-\psi),\end{gathered}$$

with $\mathcal{F}_{t-1}=\sigma\{x_t,z_{t-1},z_{t-2},\ldots\}$. By construction Lemma 4 follows from

$$Q_\psi(z_T,\bar{\psi}_T^d)-Q_\psi(z_T,\psi)=o_P(1), \tag{A.5}$$

and

$$Y_\psi(z_T,\bar{\psi}_T^d)=o_P(1). \tag{A.6}$$

Proof of (A.5). I first show that

$$Q_\psi(z_T,\psi^T)-Q_\psi(z_T,\psi)=o_P(1). \tag{A.7}$$

Let $q_t(\psi^T)=\tau(u_t(\psi^T))d_t(\psi^T)-E[\tau(u_t(\psi^T))d_t(\psi^T)\mid\mathcal{F}_{t-1}]$, so $Q_\psi(z_T,\psi^T)=T^{-\frac{1}{2}}\sum_t q(\psi^T)$. Because $\{q_t(\psi^T)-q_t(\psi)\}$ is a martingale difference sequence conditional on $\mathcal{F}_{t-1}$,

$$E\left\|Q_\psi(z_T,\psi^T)-Q_\psi(z_T,\psi)\right\|^2=T^{-1}\sum_{j=1}^{c}\sum_t E\left[q_{tj}(\psi^T)-q_{tj}(\psi)\right]^2.$$

Kreiss Lemma 6.4 establishes that $T^{-1}\sum_t E\left[q_{tj}(\psi^T)-q_{tj}(\psi)\right]^2=o_P(1)$ for $j=k+2,\ldots,c$, which are the components of $q_t(\psi^T)-q_t(\psi)$ that correspond to ρ and θ (Lemma 6.1 in Kreiss is necessary for the result). Under Assumption 4g it is the case that $| h^{(1)}_{\beta_l}(x_{t-i}, \beta^T) - h^{(1)}_{\beta_l}(x_{t-i}, \beta) | \leq | \beta^T - \beta | (d_1 + d_2 | x_{t-i} |)$, so by Lemma 6.1 from Kreiss the result for the last $p+q$ components of $q_t(\psi^T) - q_t(\psi)$ is preserved for the first $k+1$ components of $q_t(\psi^T) - q_t(\psi)$, which are the components corresponding to β. Thus $Q_\psi(z_T, \psi^T) - Q_\psi(z_T, \psi) = o_P(1)$, so by Lemma 4.4 from Kreiss $Q_\psi(z_T, \bar{\psi}^d_T) - Q_\psi(z_T, \psi) = o_P(1)$.

Proof of (A.6). I first show that

$$Y_\psi(z_T, \psi^T) = o_P(1).$$

Because g is symmetric about zero

$$E\left[\tau(u_t(\psi))d_t(\psi) \mid \mathcal{F}_{t-1}\right] = 0.$$

For notational convenience, let $w_{T,t} = d_t(\psi^T)'[\mathcal{J}_\psi(\psi, g)T]^{-\frac{1}{2}} m_T$ (recall (A.4)), so

$$E\left[\tau(u_t(\psi^T))d_t(\psi^T) \mid \mathcal{F}_{t-1}\right] = T^{-\frac{1}{2}} \sum_t d_t(\psi^T) \int \tau(u + w_{T,t})g(u)du + o_P(1), \quad \text{(A.8)}$$

$$= T^{-1} \sum_t d_t(\psi^T) d_t(\psi^T)' \mathcal{J}_\psi(\psi, g)^{-\frac{1}{2}} m_T [w_{T,t}]^{-1} \int \tau(u + w_{T,t}) g(u) du + o_P(1),$$

which is bounded by (recall that $A(c) = \{w : w'w \leq c\}$, so as $T \to \infty$, $w_{T,t} \in A(c)$):

$$T^{-1} \sum_t d_t(\psi^T) d_t(\psi^T)' \mathcal{J}_\psi(\psi, g) m_T \sup_{A(c)} w^{-1} \int \tau(u + w) g(u) du + o_P(1).$$

From the definition of ψ^T it follows that $\mathcal{J}_\psi(\psi, g)T^{\frac{1}{2}}(\psi^T - \psi) = \mathcal{J}_\psi(\psi, g)^{\frac{1}{2}} m_T$, so

$$\begin{aligned} Y_\psi(z_T, \psi^T) \leq{} & T^{-1} \textstyle\sum_t d_t(\psi^T) d_t(\psi^T)' \mathcal{J}_\psi(\psi, g) m_T \sup_{A(c)} w^{-1} \int \tau(u + w) g(u) du \\ & + \mathcal{J}_\psi(\psi, g)^{\frac{1}{2}} m_T + o_P(1). \end{aligned}$$

For large values of T ($T \to \infty$) and small values of c ($c \to 0$):

$$Y_\psi(z_T, \psi^T) \leq T^{-1} \sum_t d_t(\psi^T) d_t(\psi^T)' \mathcal{J}_\psi(\psi, g)^{-\frac{1}{2}} m_T (-\mathcal{I}(g)) + \mathcal{J}_\psi(\psi, g)^{\frac{1}{2}} m_T + o_P(1),$$

by Assumption 2g. By Lemma 2:

$$\begin{aligned} Y_\psi(z_T, \psi^T) &\leq -\mathcal{J}_\psi(\psi, g) \mathcal{J}_\psi(\psi, g)^{-\frac{1}{2}} m_T + \mathcal{J}_\psi(\psi, g)^{\frac{1}{2}} m_T + o_P(1) \\ &= o_P(1). \end{aligned}$$

By Lemma 4.4 from Kreiss $Y_\psi(z_T, \bar{\psi}^d_T) = o_P(1)$.

A.5. Proof of Lemma 5

To show that $\hat{\psi}_T$ is regular, I use (2.2) to rewrite the left-hand side of the regularity condition (4.1) as

$$\mathcal{J}_\psi(\psi,g)T^{\frac{1}{2}}(\bar{\psi}_T^d-\psi)+\mathcal{J}_\psi(\psi,g)\mathcal{J}_\psi(z_T,\bar{\psi}_T^d,\hat{g})^{-1}S_\psi(z_T,\bar{\psi}_T^d,\hat{g})-S_\psi(z_T,\psi,g). \quad (A.9)$$

Because the functional $\mathcal{J}_\psi()$ is strictly positive, Lemma 2 implies that (A.9) equals

$$\mathcal{J}_\psi(\psi,g)T^{\frac{1}{2}}(\bar{\psi}_T^d-\psi)+S_\psi(z_T,\bar{\psi}_T^d,\hat{g})(1+o_P(1))-S_\psi(z_T,\psi,g). \quad (A.10)$$

By Lemma 3, (A.10) equals

$$\begin{aligned}\mathcal{J}_\psi(\psi,g)T^{\frac{1}{2}}(\bar{\psi}_T^d-\psi)+S_\psi(z_T,\bar{\psi}_T^d,g)(1+o_P(1))-S_\psi(z_T,\psi,g),\\ =o_P(1),\end{aligned} \quad (A.11)$$

where the last equality follows by Lemma 4. ◫

A.6. Proof of Theorem 1

To prove Theorem 1, I prove that regularity of $\hat{\psi}_T$ is sufficient for $\hat{\psi}_T$ to be uniformly adaptive. By Lemma 5, $\hat{\psi}_T$ is regular, that is

$$T^{\frac{1}{2}}(\hat{\psi}_T-\psi)-\mathcal{J}_\psi(\psi,g)^{-1}S_\psi(z_T,\psi,g)=o_P(1). \quad (A.12)$$

From the definition of ψ^T:

$$[\mathcal{J}_\psi(\psi,g)T]^{\frac{1}{2}}(\hat{\psi}_T-\psi^T)=[\mathcal{J}_\psi(\psi,g)T]^{\frac{1}{2}}(\hat{\psi}_T-\psi)-m_T,$$

so regularity of $\hat{\psi}_T$ implies that

$$[\mathcal{J}_\psi(\psi,g)T]^{\frac{1}{2}}(\hat{\psi}_T-\psi^T)-\mathcal{J}_\psi(\psi,g)^{-\frac{1}{2}}S_\psi(z_T,\psi,g)-m_T=o_P(1). \quad (A.13)$$

From Lemma 1, L^T is locally asymptotically normal, so that convergence in (A.13) with (ψ,g) as true values implies convergence under the assumption that (ψ^T,g^T) are true values. As a result, (A.13) implies that

$$[\mathcal{J}_\psi(\psi^T,g^T)T]^{\frac{1}{2}}(\hat{\psi}_T-\psi^T)-\mathcal{J}_\psi(\psi^T,g^T)^{-\frac{1}{2}}S_\psi(z_T,\psi^T,g^T)-m_T=o_P(1). \quad (A.14)$$

From Lemmas 4.2 and 4.3 in Fabian and Hannan (1982)

$$\mathcal{J}_\psi(\psi^T,g)^{-\frac{1}{2}}S_\psi(z_t,\psi^T,g)-m_T\Rightarrow N(0,I_c). \quad (A.15)$$

Because $\tau^T(u)\rightarrow^P\tau(u)$ (from the definition of $\mathcal{C}(\varsigma)$ in Section 3) and because both $S_\psi(\cdot,\cdot,\cdot)$ and $\mathcal{J}_\psi(\cdot,\cdot)$ are continuous with respect to $\tau(u)$ (note that both $S_\psi(\cdot,\cdot,\cdot)$ and $\mathcal{J}_\psi(\cdot,\cdot)$ are linear in $\tau(u)$ and $\tau(u)^2$, respectively) it follows that

$$\mathcal{J}_\psi(\psi^T,g^T)^{-\frac{1}{2}}S_\psi(z_T,\psi^T,g^T)-\mathcal{J}_\psi(\psi^T,g)^{-\frac{1}{2}}S_\psi(z_T,\psi^T,g)=o_P(1). \quad (A.16)$$

Together (A.15) and (A.16) imply

$$\mathcal{J}_\psi(\psi^T,g^T)^{-\frac{1}{2}}S_\psi(z_T,\psi^T,g^T)-m_T\Rightarrow N(0,I_c). \quad (A.17)$$

Together (A.14) and (A.17) imply

$$[\mathcal{J}_\psi(\psi^T,g^T)T]^{\frac{1}{2}}(\hat{\psi}_T-\psi^T)\Rightarrow N(0,I_c),$$

and consequently that

$$T^{\frac{1}{2}}(\hat{\psi}_T - \psi^T) \Rightarrow N(0, \mathcal{J}_{\psi}(\psi, g)^{-1}) \tag{A.18}$$

for all $(\psi^T, g^T) \in \Psi \times \mathcal{C}$.

Because l is a continuous and bounded loss function, (A.18) together with Theorem 5.2 from Billingsley (1968) imply

$$E_{f^T}\left[l\left(\sqrt{T}(\hat{\psi}_T - \psi^T)\right)\right] \to E\left[l\left(Z^*\right)\right], \tag{A.19}$$

for all $(\psi^T, g^T) \in \Psi \times \mathcal{C}$. □