Inference for Alternating Time Series

Ursula U. Müller[1], Anton Schick[2], and Wolfgang Wefelmeyer[3]

[1] Department of Statistics
Texas A&M University
College Station, TX 77843-3143, USA
(e-mail: `uschi@stat.tamu.edu`)
[2] Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902-6000, USA
(e-mail: `anton@math.binghamton.edu`)
[3] Mathematisches Institut
Universität zu Köln
Weyertal 86-90
50931 Köln, Germany
(e-mail: `wefelm@math.uni-koeln.de`)

Abstract. Suppose we observe a time series that alternates between different autoregressive processes. We give conditions under which it has a stationary version, derive a characterization of efficient estimators for differentiable functionals of the model, and use it to construct efficient estimators for the autoregression parameters and the innovation distributions. We also study the cases of equal autoregression parameters and of equal innovation densities.
Keywords: Autoregression, Local asymptotic normality, Semiparametric model, Efficiency, Adaptivity.

1 Introduction

By an *alternating AR(1) process* of *period* m we mean a time series X_t, $t = 0, 1, \ldots$, that alternates periodically between m possibly different AR(1) processes,

$$X_{jm+k} = \vartheta_k X_{jm+k-1} + \varepsilon_{jm+k}, \quad j = 0, 1, \ldots, \quad k = 1, \ldots, m, \tag{1}$$

where the innovations ε_t, $t \in \mathbb{N}$, are independent with mean zero and finite variances, and ε_{jm+k} has a positive density f_k. Then the m-dimensional process $\mathbf{X}_j = (X_{(j-1)m+1}, \ldots, X_{jm})^\top$, $j \in \mathbb{N}$, is a homogeneous Markov chain. Its transition density from $\mathbf{X}_{j-1}$ to $\mathbf{X}_j = \mathbf{x} = (x_1, \ldots, x_m)^\top$ depends only on the last component of $\mathbf{X}_{j-1}$, say x_0, and is given by

$$(x_0, \mathbf{x}) \mapsto \prod_{k=1}^{m} f_k(x_k - \vartheta_k x_{k-1}).$$

Note that an alternating AR(1) process is not a multivariate autoregressive process, which would require a representation $\mathbf{X}_j = \Theta \mathbf{X}_{j-1} + \boldsymbol{\varepsilon}_j$ for a matrix Θ and i.i.d. vectors $\boldsymbol{\varepsilon}_j$.

If we replace X_{jm+k-1} in (1) by its autoregressive representation and iterate this $m-1$ times, we arrive at the representation

$$X_{jm+k} = \tau(\boldsymbol{\vartheta}) X_{(j-1)m+k} + \eta_{jm+k}, \quad j \in \mathbb{N}, \tag{2}$$

where

$$\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_m)^\top, \quad \tau(\boldsymbol{\vartheta}) = \prod_{k=1}^{m} \vartheta_k, \quad \eta_{jm+k} = \sum_{t=0}^{m-1} \tau_{kt}(\boldsymbol{\vartheta}) \varepsilon_{jm+k-t},$$

and, setting $\vartheta_s = \vartheta_t$ if $s = t \mod m$,

$$\tau_{kt}(\boldsymbol{\vartheta}) = \prod_{s=0}^{t-1} \vartheta_{k-s}.$$

In particular, $\tau_{k0}(\boldsymbol{\vartheta}) = 1$ and $\tau_{k,jm}(\boldsymbol{\vartheta}) = \tau^j(\boldsymbol{\vartheta})$. The innovations η_{jm+k}, $j \in \mathbb{N}$, in (2) are independent with positive density. Hence for each $k = 1, \ldots, m$ the subseries X_{jm+k}, $j = 0, 1, \ldots,$ is AR(1) and an irreducible and aperiodic Markov chain, and positive Harris recurrent if and only if $|\tau(\boldsymbol{\vartheta})| < 1$. In particular, we do not need $|\vartheta_k| < 1$ for all k. We obtain that the m-dimensional Markov chain $\mathbf{X}_j$, $j \in \mathbb{N}$, is irreducible and aperiodic, and positive Harris recurrent if and only if $|\tau(\boldsymbol{\vartheta})| < 1$. In this case we also have infinite-order moving average representations

$$X_{jm+k} = \sum_{t=0}^{\infty} \tau_{kt}(\boldsymbol{\vartheta}) \varepsilon_{jm+k-t}, \quad j = 0, 1, \ldots, \quad k = 1, \ldots, m. \tag{3}$$

In the following sections we derive efficient estimators for submodels of alternating AR(1) processes. We treat dependencies between the autoregression parameters and also consider the cases of equal autoregression parameters and of equal innovation densities. In Section 2 we give conditions under which the alternating AR(1) model is locally asymptotically normal, and characterize efficient estimators of vector-valued functionals. In Section 3 we construct efficient estimators for the autoregression parameters and the innovation distributions. Section 4 considers submodels with equal innovation densities.

2 Characterization of efficient estimators

In order to describe possible dependencies between the autoregression parameters $\vartheta_1, \ldots, \vartheta_m$, we reparametrize them as follows. Let $p \leq m$ and $A \subset \mathbb{R}^p$ open, let $\boldsymbol{\vartheta} : A \to \mathbb{R}^m$, and set $\vartheta_k = \vartheta_k(\boldsymbol{\varrho})$ for $\boldsymbol{\varrho} \in A$. Set $\mathbf{f} = (f_1, \ldots, f_m)^\top$. Our model is semiparametric; its distribution is determined by $(\boldsymbol{\varrho}, \mathbf{f})$.

Fix $\boldsymbol{\varrho} \in A$ with $|\tau(\boldsymbol{\vartheta}(\boldsymbol{\varrho}))| < 1$. Assume that $\boldsymbol{\vartheta} : A \to \mathbb{R}^m$ has continuous partial derivatives at $\boldsymbol{\varrho}$, and write $\dot{\boldsymbol{\vartheta}}$ for the $m \times p$ matrix of partial derivatives and $\dot{\boldsymbol{\vartheta}}_k$ for its k-th row. Assume that $\dot{\boldsymbol{\vartheta}}$ is of full rank. Fix innovation densities $f_1, \dots, f_m$. Assume that the f_k are absolutely continuous with a.e. derivative f_k' and finite Fisher information $J_k = E[\ell_k^2(\varepsilon_k)]$, where $\ell_k = -f_k'/f_k$. Introduce perturbations $\boldsymbol{\varrho}_{n\mathbf{t}} = \boldsymbol{\varrho} + n^{-1/2}\mathbf{t}$ with $\mathbf{t} \in \mathbb{R}^p$, and $f_{knu_k}(x) = f(x)(1 + n^{-1/2}u_k(x))$ with u_k in the space U_k of bounded measurable functions such that $E[u_k(\varepsilon_k)] = 0$ and $E[\varepsilon_k u_k(\varepsilon_k)] = 0$. These two conditions guarantee that f_{knu_k} is a mean zero probability density for n sufficiently large. The transition density from $X_{jm+k-1} = x_{k-1}$ to $X_{jm+k} = x_k$ is $f_k(x_k - \vartheta_k x_{k-1})$. The perturbed transition density

$$(x_{k-1}, x_k) \mapsto f_{knu_k}(x_k - \vartheta_k(\boldsymbol{\varrho}_{n\mathbf{t}})x_{k-1})$$

is Hellinger differentiable with derivative

$$(x_{k-1}, x_k) \mapsto \dot{\boldsymbol{\vartheta}}_k \mathbf{t} x_{k-1} \ell_k(x_k - \vartheta_k x_{k-1}) + u_k(x_k - \vartheta_k x_{k-1}).$$

Here and in the following we write $\boldsymbol{\vartheta}$ for $\boldsymbol{\vartheta}(\boldsymbol{\varrho})$. Set $\mathbf{U} = U_1 \times \cdots \times U_m$, $\mathbf{u} = (u_1, \dots, u_m)^\top$ and $\mathbf{f}_{n\mathbf{u}} = (f_{1nu_1}, \dots, f_{mnu_m})^\top$. Suppose we observe $X_0, \mathbf{X}_1, \dots, \mathbf{X}_n$. Let P_n and $P_{n\mathbf{tu}}$ denote their joint laws under $(\boldsymbol{\varrho}, \mathbf{f})$ and $(\boldsymbol{\varrho}_{n\mathbf{t}}, \mathbf{f}_{n\mathbf{u}})$, respectively. Following [Koul and Schick, 1997], who treat non-alternating autoregression, we obtain *local asymptotic normality*

$$\begin{aligned}\log \frac{dP_{n\mathbf{tu}}}{dP_n} = n^{-1/2} \sum_{j=1}^{n} \sum_{k=1}^{m} \big(\dot{\boldsymbol{\vartheta}}_k \mathbf{t} X_{jm+k-1} \ell_k(\varepsilon_{jm+k}) + u_k(\varepsilon_{jm+k})\big) \\ -\frac{1}{2}\mathbf{t}^\top \dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}} \mathbf{t} - \frac{1}{2}\sum_{k=1}^{m} E[u_k^2(\varepsilon_k)] + o_p(1),\end{aligned} \tag{4}$$

where D is the diagonal matrix with entries $E[X_1^2]J_1, \dots, E[X_m^2]J_m$. Here we have used that $X_0\ell_1(\varepsilon_1), \dots, X_{m-1}\ell_m(\varepsilon_m), u_1(\varepsilon_1), \dots, u_m(\varepsilon_m)$ are uncorrelated.

We can now characterize efficient estimators as follows, using results originally due to Hájek and LeCam, for which we refer to Section 3.3 of the monograph [Bickel *et al.*, 1998]. Let $\bar{U}_k$ denote the closure of U_k in $L_2(f_k)$ and set $\bar{\mathbf{U}} = \bar{U}_1 \times \cdots \times \bar{U}_m$. The squared norm of $(\mathbf{t}, \mathbf{u})$ on the right-hand side of (4) determines how difficult it is, asymptotically, to distinguish between $(\boldsymbol{\varrho}, \mathbf{f})$ and $(\boldsymbol{\varrho}_{n\mathbf{t}}, \mathbf{f}_{n\mathbf{u}})$. It defines an inner product on $\mathbb{R}^p \times \bar{\mathbf{U}}$. A real-valued functional φ of $(\boldsymbol{\varrho}, \mathbf{f})$ is called *differentiable* at $(\boldsymbol{\varrho}, \mathbf{f})$ with *gradient* $(\mathbf{t}_\varphi, \mathbf{u}_\varphi) \in \mathbb{R}^p \times \bar{\mathbf{U}}$ if

$$n^{1/2}(\varphi(\boldsymbol{\varrho}_{n\mathbf{t}}, \mathbf{f}_{n\mathbf{u}}) - \varphi(\boldsymbol{\varrho}, \mathbf{f})) \to \mathbf{t}_\varphi^\top \dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}} \mathbf{t} + \sum_{k=1}^{m} E[u_{\varphi k}(\varepsilon_k) u_k(\varepsilon_k)] \tag{5}$$

for all $(\mathbf{t}, \mathbf{u}) \in \mathbb{R}^p \times \mathbf{U}$. An estimator $\hat{\varphi}$ of φ is called *regular* at $(\boldsymbol{\varrho}, \mathbf{f})$ with *limit* L if

$$n^{1/2}(\hat{\varphi} - \varphi(\boldsymbol{\varrho}_{n\mathbf{t}}, \mathbf{f}_{n\mathbf{u}})) \Rightarrow L \text{ under } P_{n\mathbf{tu}}, \quad (\mathbf{t}, \mathbf{u}) \in \mathbb{R}^p \times \mathbf{U}. \tag{6}$$

The convolution theorem of Hájek and LeCam says that L is distributed as the convolution of some random variable with a normal random variable N that has mean 0 and variance

$$\mathbf{t}_\varphi^\top \dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}} \mathbf{t}_\varphi + \sum_{k=1}^m E[u_{\varphi k}^2(\varepsilon_k)].$$

This justifies calling $\hat{\varphi}$ *efficient* if L is distributed as N.

An estimator $\hat{\varphi}$ of φ is called *asymptotically linear* at $(\boldsymbol{\varrho}, \mathbf{f})$ with *influence function* g if $g \in L_2(P_1)$ with $E(g(X_0, \mathbf{X}_1)|X_0) = 0$ and

$$n^{1/2}(\hat{\varphi} - \varphi(\boldsymbol{\varrho}, \mathbf{f})) = n^{-1/2} \sum_{j=1}^n g(X_{(j-1)m}, \mathbf{X}_j) + o_p(1).$$

It follows from the convolution theorem that an estimator $\hat{\varphi}$ is regular and efficient if and only if it is asymptotically linear with *efficient influence function*

$$g(x_0, \mathbf{x}) = \sum_{k=1}^m \big(\dot{\boldsymbol{\vartheta}}_k \mathbf{t}_\varphi x_{k-1} \ell_k(x_k - \vartheta_k x_{k-1}) + u_{\varphi k}(x_k - \vartheta_k x_{k-1})\big).$$

The inner product in (5) decomposes into $m+1$ inner products on $\mathbb{R}^p$ and $\bar{U}_1, \dots, \bar{U}_m$. This implies that the gradient of a functional φ of $\boldsymbol{\varrho}$ only is the same for each submodel in which some or all of the f_k are known. Hence asymptotically we cannot estimate φ better in these submodels. In this sense, functionals $\varphi(\boldsymbol{\varrho})$ are *adaptive* with respect to $\mathbf{f}$. Similarly, functionals of f_k are adaptive with respect to the other parameters.

For a q-dimensional functional $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_q)^\top$ of $(\boldsymbol{\varrho}, \mathbf{f})$, differentiability is understood componentwise. For an estimator $\hat{\boldsymbol{\varphi}}$ of $\boldsymbol{\varphi}$, asymptotic linearity is also understood componentwise, and regularity is defined as in (6), now with L a q-dimensional random vector. It is then convenient to write the gradient of $\boldsymbol{\varphi}$ as a matrix $(T_{\boldsymbol{\varphi}}, U_{\boldsymbol{\varphi}})$ whose s-th *row* is the gradient of φ_s; so differentiability (5) reads

$$n^{1/2}(\boldsymbol{\varphi}(\boldsymbol{\varrho}_{n\mathbf{t}}, \mathbf{f}_{n\mathbf{u}}) - \boldsymbol{\varphi}(\boldsymbol{\varrho}, \mathbf{f})) \to T_{\boldsymbol{\varphi}} \dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}} \mathbf{t} + \sum_{k=1}^m E[U_{\boldsymbol{\varphi}, \cdot k}(\varepsilon_k) u_k(\varepsilon_k)] \tag{7}$$

for all $(\mathbf{t}, \mathbf{u}) \in \mathbb{R}^p \times \mathbf{U}$. The convolution theorem then says that L is distributed as the convolution of some random vector with a normal random vector N that has mean vector 0 and covariance matrix

$$T_{\boldsymbol{\varphi}} \dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}} T_{\boldsymbol{\varphi}}^\top + \sum_{k=1}^m E[U_{\boldsymbol{\varphi}, \cdot k}(\varepsilon_k) U_{\boldsymbol{\varphi}, \cdot k}(\varepsilon_k)].$$

Finally, $\hat{\boldsymbol{\varphi}}$ is regular and efficient if and only if it is asymptotically linear with (q-dimensional) *efficient influence function*

$$g(x_0, \mathbf{x}) = \sum_{k=1}^m \big(T_{\boldsymbol{\varphi}} \dot{\boldsymbol{\vartheta}}_k^\top x_{k-1} \ell_k(x_k - \vartheta_k x_{k-1}) + U_{\boldsymbol{\varphi}, \cdot k}(x_k - \vartheta_k x_{k-1})\big).$$

3 Construction of efficient estimators

Autoregression parameters. Suppose we want to estimate $\boldsymbol{\varrho}$. By adaptivity, the gradient of the functional $\boldsymbol{\varphi}(\boldsymbol{\varrho}) = \boldsymbol{\varrho}$ is obtained from (7) as $(T_{\boldsymbol{\varphi}}, 0)$ with $T_{\boldsymbol{\varphi}}$ solving

$$n^{1/2}(\boldsymbol{\varrho}_{n\mathbf{t}} - \boldsymbol{\varrho}) = \mathbf{t} = T_{\boldsymbol{\varphi}} \dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}} \mathbf{t} \quad \mathbf{t} \in \mathbb{R};$$

so $T_{\boldsymbol{\varphi}} = (\dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}})^{-1}$, and the efficient influence function is

$$g(x_0, \mathbf{x}) = (\dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}})^{-1} \sum_{k=1}^{m} \dot{\boldsymbol{\vartheta}}_k^\top x_{k-1} \ell_k(x_k - \vartheta_k x_{k-1}). \tag{8}$$

Hence the asymptotic variance of an efficient estimator is $(\dot{\boldsymbol{\vartheta}}^\top D \dot{\boldsymbol{\vartheta}})^{-1}$.

Following [Koul and Schick, 1997], we can construct an efficient estimator of $\boldsymbol{\varrho}$, with this influence function, by the *Newton–Raphson* procedure. This is a one-step improvement of a root-n consistent initial estimator. As initial estimator of $\boldsymbol{\varrho}$ we can take e.g. the *least squares estimator* $\tilde{\boldsymbol{\varrho}}$, the minimum in $\boldsymbol{\varrho}$ of

$$\sum_{j=1}^{n} \sum_{k=1}^{m} \big(X_{jm+k} - \vartheta_k(\boldsymbol{\varrho}) X_{jm+k-1}\big)^2,$$

i.e. a solution of the martingale estimating equation

$$\sum_{j=1}^{n} \sum_{k=1}^{m} \dot{\boldsymbol{\vartheta}}_k^\top(\boldsymbol{\varrho}) X_{jm+k-1} \big(X_{jm+k} - \vartheta_k(\boldsymbol{\varrho}) X_{jm+k-1}\big) = 0.$$

An efficient estimator is then

$$\hat{\boldsymbol{\varrho}} = \tilde{\boldsymbol{\varrho}} + (\dot{\boldsymbol{\vartheta}}(\tilde{\boldsymbol{\varrho}})^\top \tilde{D} \dot{\boldsymbol{\vartheta}}(\tilde{\boldsymbol{\varrho}}))^{-1} \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{m} \dot{\boldsymbol{\vartheta}}_k^\top(\tilde{\boldsymbol{\varrho}}) X_{jm+k-1} \tilde{\ell}_k(\tilde{\varepsilon}_{jm+k}).$$

Here we have estimated ℓ_k by $\tilde{\ell}_k = -\tilde{f}_k'/\tilde{f}_k$ with $\tilde{f}_k$ an appropriate kernel estimator of f_k based on residuals $\tilde{\varepsilon}_{jm+k} = X_{jm+k} - \vartheta_k(\tilde{\boldsymbol{\varrho}}) X_{jm+k-1}$ for $j = 1, \dots, n$, and we have estimated D by plugging in empirical estimators for $\gamma_k = E[X_k^2]$ and J_k,

$$\tilde{\gamma}_k = \frac{1}{n} \sum_{j=1}^{n} X_{jm+k}^2, \quad \tilde{J}_k = \frac{1}{n} \sum_{j=1}^{n} \tilde{\ell}_k^2(\tilde{\varepsilon}_{jm+k}).$$

A special case is the alternating AR(1) model (1) with equal autoregression parameters $\vartheta_1 = \cdots = \vartheta_m = \vartheta$. This is described by the reparametrization $\boldsymbol{\vartheta}(\vartheta) = (\vartheta, \dots, \vartheta)^\top$, with ϑ playing the role of $\boldsymbol{\varrho}$. Then $\dot{\boldsymbol{\vartheta}} = (1, \dots, 1)^\top$ and

$\dot{\boldsymbol{\vartheta}}^\top D\dot{\boldsymbol{\vartheta}} = D_0 = E[X_1^2]J_1 + \cdots + E[X_m^2]J_m$, and the efficient influence function (8) reduces to

$$g(x_0, \mathbf{x}) = D_0^{-1} \sum_{k=1}^{m} x_{k-1}\ell_k(x_k - \vartheta x_{k-1}).$$

Hence the asymptotic variance of an efficient estimator is D_0^{-1}.

An initial estimator for ϑ is the least squares estimator

$$\tilde{\vartheta} = \sum_{t=1}^{nm} X_{t-1}X_t \Big/ \sum_{t=1}^{nm} X_{t-1}^2,$$

and an efficient estimator is the one-step improvement

$$\hat{\vartheta} = \tilde{\vartheta} + \tilde{D}_0^{-1} \frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{m} X_{jm+k-1}\tilde{\ell}_k(\tilde{\varepsilon}_{jm+k})$$

with $\tilde{\varepsilon}_t = X_t - \tilde{\vartheta}X_{t-1}$ and $\tilde{D}_0 = \tilde{\gamma}_1\tilde{J}_1 + \cdots + \tilde{\gamma}_m\tilde{J}_m$.

Innovation distributions. Suppose we want to estimate a linear functional $\varphi(f_k) = E[h(\varepsilon_k)] = \int h(x)f_k(x)\,dx$ of the innovation distribution, where $h \in L_2(f_k)$. By adaptivity, the gradient of φ is obtained from (5) as $(0, \mathbf{u}_\varphi)$ with $u_{\varphi i} = 0$ for $i \neq k$ and $u_{\varphi k}$ solving

$$n^{1/2}(\varphi(f_{knu_k}) - \varphi(f_k)) = E[h(\varepsilon_k)u_k(\varepsilon_k)] = E[u_{\varphi k}(\varepsilon_k)u_k(\varepsilon_k)], \quad u_k \in U_k;$$

so $u_{\varphi k}$ is the projection of h onto $\bar{U}_k$,

$$u_{\varphi k}(x) = h(x) - E[h(\varepsilon_k)] - \frac{E[\varepsilon_k h(\varepsilon_k)]}{E[\varepsilon_k^2]}x.$$

Hence the efficient influence function is

$$g(x_0, \mathbf{x}) = h(x_k - \vartheta_k x_{k-1}) - E[h(\varepsilon_k)] - \frac{E[\varepsilon_k h(\varepsilon_k)]}{E[\varepsilon_k^2]}(x_k - \vartheta_k x_{k-1}),$$

and the asymptotic variance of an efficient estimator is

$$E[u_{\varphi k}^2(\varepsilon_k)] = \operatorname{Var} h(\varepsilon_k) - \frac{(E[\varepsilon_k h(\varepsilon_k)])^2}{E[\varepsilon_k^2]}.$$

An efficient estimator, with influence function g, is

$$\hat{\varphi} = \frac{1}{n}\sum_{j=1}^{n} h(\tilde{\varepsilon}_{jm+k}) - \frac{\sum_{j=1}^{n}\tilde{\varepsilon}_{jm+k}h(\tilde{\varepsilon}_{jm+k})}{\sum_{j=1}^{n}\tilde{\varepsilon}_{jm+k}^2}\frac{1}{n}\sum_{j=1}^{n}\tilde{\varepsilon}_{jm+k}.$$

This requires that h or f_k is sufficiently smooth. For appropriate assumptions we refer to [Schick and Wefelmeyer, 2002]. An alternative to the above additive correction of an empirical estimator are *weighted empirical estimators* $\frac{1}{n}\sum_{j=1}^{n} w_j h(\tilde{\varepsilon}_{jm+k})$, where the random weights w_j are chosen such that $\sum_{j=1}^{n} w_j\tilde{\varepsilon}_{jm+k} = 0$; see [Owen, 2001] and [Müller *et al.*, 2005].

4 Equal innovation densities

Suppose that the innovation distributions are known to be equal, $f_1 = \cdots = f_m = f$, say. As in Section 2, assume that f is absolutely continous with finite Fisher information $J = E[\ell^2(\varepsilon)]$, where $\ell = -f'/f$. Introduce perturbations $f_{nu}(x) = f(x)(1 + n^{-1/2}u(x))$ with u in the space U of bounded measurable functions such that $E[u(\varepsilon)] = 0$ and $E[\varepsilon u(\varepsilon)] = 0$. Then local asymptotic normality (4) reduces to

$$\log \frac{dP_{n\mathbf{t}u}}{dP_n} = n^{-1/2} \sum_{j=1}^{n} \sum_{k=1}^{m} \dot{\boldsymbol{\vartheta}}_k \mathbf{t} X_{jm+k-1} \ell(\varepsilon_{jm+k}) + n^{-1/2} \sum_{t=1}^{nm} u(\varepsilon_t)$$
$$- \frac{1}{2} \mathbf{t}^\top \dot{\boldsymbol{\vartheta}}^\top D_1 \dot{\boldsymbol{\vartheta}} \mathbf{t} - \frac{m}{2} E[u^2(\varepsilon)] + o_p(1),$$

where D_1 is the diagonal matrix with entries $E[X_1^2]J, \dots, E[X_m^2]J$.

Let $\bar{U}$ denote the closure of U in $L_2(f)$. A real-valued functional φ of $(\boldsymbol{\varrho}, f)$ is *differentiable* at $(\boldsymbol{\varrho}, f)$ with *gradient* $(\mathbf{t}_\varphi, u_\varphi) \in \mathbb{R}^p \times \bar{U}$ if

$$n^{1/2}(\varphi(\boldsymbol{\varrho}_{n\mathbf{t}}, f_{nu}) - \varphi(\boldsymbol{\varrho}, f)) \to \mathbf{t}_\varphi^\top \dot{\boldsymbol{\vartheta}}^\top D_1 \dot{\boldsymbol{\vartheta}} \mathbf{t} + mE[u_\varphi(\varepsilon) u(\varepsilon)]$$

for all $(\mathbf{t}, u) \in \mathbb{R}^p \times U$. The factor m is there because we count the observations in blocks of length m. The inner product on the right-hand side decomposes into two inner products on $\mathbb{R}^p$ and on $\bar{U}$. Hence functionals of $\boldsymbol{\varrho}$ or f are adaptive with respect to the other parameter.

Autoregression parameters. The efficient influence function (8) of $\boldsymbol{\varrho}$ reduces to

$$g(x_0, \mathbf{x}) = (\dot{\boldsymbol{\vartheta}}^\top D_1 \dot{\boldsymbol{\vartheta}})^{-1} \sum_{k=1}^{m} \dot{\boldsymbol{\vartheta}}^\top x_{k-1} \ell(x_k - \vartheta_k x_{k-1}).$$

An efficient estimator $\hat{\boldsymbol{\varrho}}$ of $\boldsymbol{\varrho}$ is again obtained as one-step improvement of a root-n consistent initial estimator $\tilde{\boldsymbol{\varrho}}$,

$$\hat{\boldsymbol{\varrho}} = \tilde{\boldsymbol{\varrho}} + (\dot{\boldsymbol{\vartheta}}(\tilde{\boldsymbol{\varrho}})^\top \tilde{D}_1 \dot{\boldsymbol{\vartheta}}(\tilde{\boldsymbol{\varrho}}))^{-1} \sum_{k=1}^{m} \dot{\boldsymbol{\vartheta}}_k^\top(\tilde{\boldsymbol{\varrho}}) X_{jm+k-1} \tilde{\ell}(\tilde{\varepsilon}_{jm+k}).$$

Here we can estimate ℓ by $\tilde{\ell} = -\tilde{f}'/\tilde{f}$ with $\tilde{f}$ a kernel estimator based on all residuals $\tilde{\varepsilon}_{jm+k} = X_{jm+k} - \vartheta_k(\tilde{\boldsymbol{\varrho}}) X_{jm+k-1}$ for $j = 1, \dots, n$ and $k = 1, \dots, m$, and we can estimate J by

$$\tilde{J} = \frac{1}{nm} \sum_{t=1}^{nm} \tilde{\ell}^2(\tilde{\varepsilon}_t).$$

Innovation distribution. Suppose we want to estimate a linear functional $\varphi(f) = E[h(\varepsilon)]$ for $h \in L_2(f)$. By adaptivity, the gradient of φ is $(0, u_\varphi)$ with $u_\varphi \in \bar{U}$ solving

$$n^{1/2}(\varphi(f_{nu}) - \varphi(f)) = E[h(\varepsilon)u(\varepsilon)] = mE[u_\varphi(\varepsilon)u(\varepsilon)], \quad u \in U;$$

so mu_φ is the projection of h onto $\bar{U}$,

$$mu_\varphi(x) = h(x) - E[h(\varepsilon)] - \frac{E[\varepsilon h(\varepsilon)]}{E[\varepsilon^2]}x.$$

Hence the efficient influence function is

$$g(x_0, \mathbf{x}) = \frac{1}{m}\sum_{k=1}^{m}\Big(h(x_k - \vartheta_k x_{k-1}) - E[h(\varepsilon)] - \frac{E[\varepsilon h(\varepsilon)]}{E[\varepsilon^2]}(x_k - \vartheta_k x_{k-1})\Big),$$

and the asymptotic variance of an efficient estimator is

$$E[g^2(X_0, \mathbf{X}_1)] = \frac{1}{m}\Big(\operatorname{Var} h(\varepsilon) - \frac{(E[\varepsilon h(\varepsilon)])^2}{E[\varepsilon^2]}\Big).$$

An efficient estimator is obtained, similarly as in Section 3, as

$$\hat{\varphi} = \frac{1}{nm}\sum_{t=1}^{nm} h(\tilde{\varepsilon}_t) - \frac{\sum_{t=1}^{nm}\tilde{\varepsilon}_t h(\tilde{\varepsilon}_t)}{\sum_{t=1}^{nm}\tilde{\varepsilon}_t^2}\frac{1}{nm}\sum_{t=1}^{nm}\tilde{\varepsilon}_t.$$

Of course, if both the autoregression parameters and the innovation densities are equal, $\vartheta_1 = \cdots = \vartheta_m = \vartheta$ and $f_1 = \cdots = f_m = f$, then the alternating AR(1) model reduces to the usual AR(1) model $X_t = \vartheta X_{t-1} + \varepsilon_t$, where the ε_t are independent with density f, and the sample size is nm.

Acknowledgment. Anton Schick was supported in part by NSF Grant DMS 0405791.

References

[Bickel *et al.*, 1998]P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models.* Springer, New York, 1998.

[Koul and Schick, 1997]H.L. Koul and A. Schick. Efficient estimation in nonlinear autoregressive time series models. *Bernoulli*, 3:247–277, 1997.

[Müller *et al.*, 2005]U.U Müller, A. Schick, and W. Wefelmeyer. Weighted residual-based density estimators for nonlinear autoregressive models. *Statistica Sinica*, 15:177–195, 2005.

[Owen, 2001]A.B. Owen. *Empirical Likelihood.* Chapman & Hall, London, 2001.

[Schick and Wefelmeyer, 2002]A. Schick and W. Wefelmeyer. Estimating the innovation distribution in nonlinear autoregressive models. *Annals of the Institute of Statistical Mathematics*, 54:245–260, 2002.