The Annals of Statistics
1997, Vol. 25, No. 2, 786–817

ADAPTIVE ESTIMATION IN TIME-SERIES MODELS

BY FEIKE C. DROST,[1] CHRIS A. J. KLAASSEN AND BAS J. M. WERKER[2]

Tilburg University and University of Amsterdam

In a framework particularly suited for many time-series models we obtain a LAN result under quite natural and economical conditions. This enables us to construct adaptive estimators for (part of) the Euclidean parameter in these semiparametric models. Special attention is directed to group models in time series with the important subclass of models with time varying location and scale. Our set-up is confronted with the existing literature and, as examples, we reconsider linear regression and ARMA, TAR and ARCH models.

1. Introduction. Consider estimation of a Euclidean parameter θ in a semiparametric model parametrized by θ and an infinite-dimensional nuisance parameter g. To study what is best possible asymptotically, one needs a bound on the asymptotic performance of estimators of θ and an estimator attaining this bound. A vast majority of models is Locally Asymptotically Normal (LAN). Then the Hájek–Le Cam convolution theorem yields an appropriate bound. On an ad hoc basis, it is often possible to find estimators of θ that have the right rate of consistency. Typically, such estimators may be used to construct efficient estimators, which attain the bound of the convolution theorem. If this bound is the same as in the parametric model with g known, then such estimators are called adaptive.

For the i.i.d. case, a comprehensive account on the present theory along these lines is given in Bickel, Klaassen, Ritov and Wellner (1993) (henceforth BKRW). Survey papers in an econometric setting are, for example, Robinson (1988), Newey (1990) and Stoker (1991). There are only a few papers in a general time-series context. Kreiss (1987a, b) has developed the theory for ARMA models. Engle and González-Rivera (1991) and Linton (1993) have discussed ARCH models under rather stringent conditions and Steigerwald (1992) [see also Pötscher (1995) and Steigerwald (1995)] has obtained some results for more general time series. The most general set-up is by Jeganathan (1995). His Assumptions (A.1)–(A.5) in Section 4 are comparable to our Assumptions A–D in Section 2 and lead to generalizations of LAN. The LAN situations Jeganathan studies in detail are of the special location-scale type (4.1) in Section 4 below with the scale fixed at $\sigma_t(\theta) = 1$, and he derives adaptive estimators for these situations under symmetric densities. Adaptive estimators in these location type time-series models have also been con-

---

Received November 1994; revised June 1996.

[1] Research fellow of the Royal Netherlands Academy of Arts and Sciences (K.N.A.W.).

[2] Partially supported by the Netherlands Organization for Scientific Research (NWO).

AMS 1991 *subject classifications*. Primary 62G05; secondary 62F12, 62M10.

Key words and phrases. LAN in time series, adaptive estimation, semiparametrics.

structed by Koul and Schick (1995). Quite another approach is followed by Wefelmeyer (1994, 1996), who obtained efficiency results in a Markovian context when only some moments are given and the innovations are assumed to be martingale differences. Here we will focus on adaptive estimation within a semiparametric framework which is well suited for time-series models with i.i.d. errors, which yields the LAN property and which is more general than the particular LAN situations studied in Jeganathan (1995) and in Koul and Schick (1995). Typically, adaptive estimation of (part of) the Euclidean parameter θ is possible in these models thanks to the independence of the present innovation and the past. Criteria for parameters to be adaptively estimable in principle have been given for these models in Drost, Klaassen and Werker (1994) (henceforth DKW). Their necessary condition is derived from the specific structure of the score function in time-series models. Here we will show that this necessary condition is also sufficient and we will construct an adaptive estimator, indeed. Thus, we construct adaptive and hence efficient estimators of (part of) the Euclidean parameter in a very broad class of semiparametric models. In this way we include several well-known adaptation results both in i.i.d. models, for example, regression with i.i.d. regressors, and in models with time dependent data, for example, time-series models such as ARMA and ARCH. Despite the generality of our set-up, it results in conditions for the existence of adaptive estimators which are even (a little bit) weaker in some examples than the classical conditions and never stronger.

To carry out the program of proving a convolution theorem and constructing adaptive estimators, we need a uniform LAN property, which is uniform in θ but assumes g fixed. This LAN result, Theorem 2.1, is given in Section 2, together with a discussion of the consequences of the convolution theorem. A one-step improvement procedure for $\sqrt{n}$-consistent estimators in order to obtain adaptive estimators is given in Theorem 3.1 of Section 3. It appears that the $\sqrt{n}$-unbiasedness condition (7.8.19), page 395 of BKRW [see also Klaassen (1987)], which is necessary to construct semiparametric estimators in an i.i.d. context, may be generalized to our time-series framework; see Remarks 3.1 and 3.2. To enhance the interpretation of our general theory, we have included several illuminating results for the class of group models in time series; see the end of Section 2 and Example 3.1. Generally the parameters of interest can be split up into two components for such models. The first, say, component will describe the influence of past observations on the model while the second part is comparable to the parameter in the corresponding i.i.d. group model. In this paper we will focus on estimation of the parameter describing the dependence structure of the model, and we will show that this part of the parameter is adaptively estimable, that is, knowledge of the infinite-dimensional parameter g will not help. Given such an adaptive estimator, the estimation problem of the remaining i.i.d. type parameters is completely equivalent to the estimation problem in i.i.d. group models. Therefore, we do not repeat these respective results here, but we refer to Section

4.2, pages 88–103 of BKRW. To relate our results to the existing literature we reconsider, in Section 4, several examples fitting into the location-scale group time-series framework: namely linear regression and ARMA, TAR, and ARCH time-series models. The Appendix contains a few minor technical details.

2. LAN in time-series models. In this section we will derive a general uniform LAN theorem for a time-series model $\mathscr{P} = \{P_{\theta g}\colon\ \theta \in \Theta,\ g \in \mathscr{G}\}$, where θ denotes an unknown Euclidean parameter in some open subset Θ of $\mathbb{R}^p$ and where g denotes the unknown density of the innovations in the time-series model. As usual (in a time-series context), we suppose that the rather complex probability structure $P_{\theta g}$ can be obtained from the innovations in the following manner. Let $(\Omega, \mathscr{S})$ be a measurable space and let $\{P_g\colon g \in \mathscr{G}\}$ be a set of probability measures on $(\Omega, \mathscr{S})$. Let the innovations $\varepsilon_1, \dots, \varepsilon_n$ be i.i.d. from a distribution with density g (under P_g) and independent of the random vector X_n^*, all defined on $(\Omega, \mathscr{S})$. Assume that $\mathscr{F}_0^n = \mathscr{F}(X_n^*)$ defines a filtration and put $\mathscr{F}_t^n = \mathscr{F}_{t-1}^n \vee \mathscr{F}(\varepsilon_t)$, $t = 1, \dots, n$. The observed random variables $Y_1, \dots, Y_n$ and the observed starting conditions and/or exogenous variables X_n of the time series are supposed to be measurable functions of θ and the random variables on the underlying probability space. More precisely stated: $X_n = X(X_n^*, \theta) \in \mathscr{F}_0^n$ and $Y_t = Y(X_n^*, \varepsilon_1, \dots, \varepsilon_t, \theta) \in \mathscr{F}_t^n$, $t = 1, \dots, n$. These functions are assumed to be invertible, that is, under $P_{\theta g}$, $X_n^* = X^*(X_n, \theta)$ and $\varepsilon_t(\theta) = \varepsilon(X_n, Y_1, \dots, Y_t, \theta) = \varepsilon_t$, $t = 1, \dots, n$.

To prove a uniform LAN theorem, let $\theta_0 \in \Theta$ be fixed, let θ_n denote the true parameter point, suppose $\theta_n \to \theta_0$, let $\tilde{\theta}_n$ be such that $\sqrt{n}(\tilde{\theta}_n - \theta_n) \to \lambda$, and let Λ_n denote the log-likelihood ratio statistic of the observations $X_n, Y_1, \dots, Y_n$ for $\tilde{\theta}_n$ with respect to θ_n. The density g of the innovations is fixed; thus we actually study a parametric model. It will be convenient to embed g into some parametric family $\mathscr{Q} = \{Q_\zeta\colon \zeta \in Z \subset \mathbb{R}^q\}$ with dominating measure μ and corresponding densities $q(\zeta) = dQ_\zeta/d\mu$, such that $g = q(\zeta_0)$. Put $l(\zeta) = \log q(\zeta)$. Often, possibly after a suitable reparametrization, the family $\mathscr{Q}$ may be chosen such that the log-likelihood ratio statistic Λ_n can be written in the following form:

$$(2.1)\quad \Lambda_n = \sum_{t=1}^{n} \Big\{ l\big(\zeta_0 + W_{nt}'\big(\tilde{\theta}_n - \theta_n\big)\big)\big(\varepsilon_t(\theta_n)\big) - l(\zeta_0)\big(\varepsilon_t(\theta_n)\big) \Big\} + \Lambda_n^s,$$

where W_{nt} and Λ_n^s will be defined below. This expansion holds especially true for group models in time series; see the discussion at the end of this section. Observe that, under $P_{\theta_n g}$, the innovations ε_t are equal to the random variables $\varepsilon_t(\theta_n)$, $t = 1, \dots, n$. So, when studying Λ_n under $P_{\theta_n g}$, we may delete the argument θ_n of $\varepsilon_t(\theta_n)$. Unless indicated otherwise, expectations and limits will be taken under the probability measure $P_{\theta_n g}$. We introduce the following notation and assumptions.

NOTATION AND ASSUMPTIONS.

A. $\Lambda_n^s \in \mathscr{F}_0^n$ denotes some term with respect to X_n; it satisfies

$$\Lambda_n^s \to_P 0. \tag{2.2}$$

B. The parametric model $\mathscr{Q} = \{Q_\zeta\colon \zeta \in Z \subset \mathbb{R}^q\}$ is regular at ζ_0. Put $s(\zeta) = q(\zeta)^{1/2}$ and let $\dot{s}(\zeta_0) \in L_2(\mu)$ be the Fréchet derivative of $s(\zeta)$ at ζ_0. The score is denoted by $\psi = 2\dot{s}(\zeta_0)/s(\zeta_0)1_{\{s(\zeta_0)>0\}}$ and the information matrix $J = E\psi(\varepsilon_t)\psi(\varepsilon_t)'$ is nonsingular.

C. The square-integrable random $p \times q$-matrix $W_{nt} = W_t(\theta_n, \tilde{\theta}_n)$ is measurable with respect to the past, that is, $W_{nt} \in \mathscr{F}_{t-1}^n$ and depends both on θ_n and $\tilde{\theta}_n$.

D. There exists a continuous matrix-valued information function $I\colon \Theta \to \mathbb{R}^{p\times p}$ and square-integrable $p \times q$-matrices $W_t = W_t(\theta_n) \in \mathscr{F}_{t-1}^n$, $t = 1,\dots,n$, satisfying

$$n^{-1}\sum_{t=1}^{n} W_t J W_t' \to_P I(\theta_0) > 0, \forall\, \delta > 0\colon n^{-1}\sum_{t=1}^{n} |W_t|^2 1_{\{n^{-1/2}|W_t|>\delta\}} \to_P 0, \tag{2.3}$$

such that W_{nt} converges to W_t, as $n \to \infty$, in the following sense:

$$\sum_{t=1}^{n} |(W_{nt} - W_t)'(\tilde{\theta}_n - \theta_n)|^2 \to_P 0. \tag{2.4}$$

E. The score of the time-series model $\mathscr{P}$ will be denoted by

$$\dot{l}_t(\theta) = W_t(\theta)\psi(\varepsilon_t(\theta)), \qquad t = 1,\dots,n. \tag{2.5}$$

To obtain the desired uniform LAN theorem at $\theta_0 \in \Theta$ and $g \in \mathscr{G}$ we need the assumptions above for arbitrary sequences $\{\theta_n\}$ and $\{\tilde{\theta}_n\}$ satisfying the prescribed conditions.

Assumption A states that the influence of the starting values and/or exogenous variables has a negligible effect on the asymptotic behavior of Λ_n. In our examples Assumption B will be fulfilled with ζ a location and/or scale parameter. The random matrix W_{nt} in C will be obtained when the likelihood of the observations is written as the product of the conditional likelihoods. Typically, the matrix W_t in Condition D can be obtained from W_{nt} by taking limits for $\tilde{\theta}_n$ to θ_n in the definition of W_{nt}. Equation (2.3) allows us to apply suitable weak laws of large numbers and central limit theorems to the score (2.5). Assumption (2.4) will serve as the main key to prove LAN and the existence of efficient estimators in these time series. This assumption appears to be rather weak; if $W_t(\theta_n, \tilde{\theta}_n)$ is smooth in $\tilde{\theta}_n$, then the difference $W_{nt} - W_t$ will be $O_p(n^{-1/2})$ and typically the averaged sum of squares in (2.4) will be $O_p(n^{-1})$. The structure of the score, $\dot{l} = W\psi$, is a multivariate generalization of the one considered in DKW.

We have the following result for time series; compare the uniform LAN Proposition 2.1.2, page 16 of BKRW.

THEOREM 2.1. *Assume* A–E *and* (2.1) *and write*

$$\Lambda_n = \lambda' n^{-1/2} \sum_{t=1}^{n} \dot{l}_t(\theta_n) - \tfrac{1}{2} n^{-1} \sum_{t=1}^{n} |\lambda' \dot{l}_t(\theta_n)|^2 + R_n; \tag{2.6}$$

then, under $P_{\theta_n g}$,

$$R_n \to_P 0 \quad \text{and} \quad \Lambda_n \to_{\mathscr{L}} N\big(-\tfrac{1}{2}\lambda' I(\theta_0)\lambda, \lambda' I(\theta_0)\lambda\big). \tag{2.7}$$

The sequences of probability measures $\{P_{\theta_n g}\}$ *and* $\{P_{\tilde{\theta}_n g}\}$ *are contiguous. If* A–E *hold for all sequences* $\theta_n = \theta_0 + O(n^{-1/2})$ *and* $\tilde{\theta}_n = \theta_0 + O(n^{-1/2})$, *then the smoothness condition holds*:

$$\sqrt{n}\left\{\tilde{\theta}_n - \theta_n + I\big(\tilde{\theta}_n\big)^{-1} n^{-1} \sum_{t=1}^{n} \dot{l}_t\big(\tilde{\theta}_n\big) - I(\theta_n)^{-1} n^{-1} \sum_{t=1}^{n} \dot{l}_t(\theta_n)\right\} \to_P 0. \tag{2.8}$$

PROOF. For notational simplicity, define $U_{nt} = W_{nt}'\sqrt{n}(\tilde{\theta}_n - \theta_n)$, $U_t = W_t'\lambda$, and observe that by (2.3), (2.4) and the nonsingularity of J,

$$n^{-1} \sum_{t=1}^{n} |U_t|^2 = O_P(1), \qquad n^{-1} \sum_{t=1}^{n} |U_t|^2 1_{\{n^{-1/2}|U_t| > \delta\}} \to_P 0 \tag{2.9}$$

and

$$n^{-1} \sum_{t=1}^{n} |U_{nt} - U_t|^2 \to_P 0. \tag{2.10}$$

Write

$$T_{nt} = 2\big\{s\big(\zeta_0 + n^{-1/2} U_{nt}\big)(\varepsilon_t)/s(\zeta_0)(\varepsilon_t) - 1\big\} \quad \text{and} \quad \psi_t = \psi(\varepsilon_t). \tag{2.11}$$

Under $P_{\theta_n g}$, the log-likelihood ratio statistic can be written as [compare BKRW, proof of Proposition 2.1.2, Appendix 9, page 511]:

$$\begin{aligned}
\Lambda_n &= 2 \sum_{t=1}^{n} \log\big(1 + \tfrac{1}{2} T_{nt}\big) + \Lambda_n^s \\
&= \lambda' n^{-1/2} \sum_{t=1}^{n} \dot{l}_t(\theta_n) - \tfrac{1}{2} n^{-1} \sum_{t=1}^{n} |\lambda' \dot{l}_t(\theta_n)|^2 \\
&\quad + \Lambda_n^s + n^{-1/2} \sum_{t=1}^{n} (U_{nt} - U_t)' \psi_t \\
&\quad + \tfrac{1}{2} n^{-1} \sum_{t=1}^{n} \big\{|U_t'\psi_t|^2 - |U_{nt}'\psi_t|^2\big\} \\
&\quad + \sum_{t=1}^{n} \big\{T_{nt} - n^{-1/2} U_{nt}'\psi_t - E(T_{nt} \,|\, \mathscr{F}_{t-1}^n)\big\} \\
&\quad + \sum_{t=1}^{n} \big\{E(T_{nt} \,|\, \mathscr{F}_{t-1}^n) + \tfrac{1}{4} n^{-1} |U_{nt}'\psi_t|^2\big\} \\
&\quad - \tfrac{1}{4} \sum_{t=1}^{n} \big\{T_{nt}^2 - n^{-1} |U_{nt}'\psi_t|^2\big\} + \tfrac{1}{6} \sum_{t=1}^{n} \alpha_{nt} T_{nt}^3.
\end{aligned} \tag{2.12}$$

Observe that the assumptions imply that, under $P_{\theta_n g}$, the random character of Λ_n is completely determined by X_n^* and the i.i.d. innovations $\varepsilon_1, \dots, \varepsilon_n$. Hence we may consider (2.12) under a fixed probability measure on $(\Omega, \mathscr{S})$, say $P_{\theta_0 g}$. In the remainder of the proof we will restrict attention to this probability measure when studying the limiting behavior of Λ_n (as a function of the nonobservable random variables $X_n^*, \varepsilon_1, \dots, \varepsilon_n$).

To prove the required result we need the following four convergence results.

LEMMA 2.1. *Suppose* $\to$ *refers to either* $\to_P$, $\to$ $(a.s.)$ *or* $\to_1$ *and* $O(1)$ *to either tightness, boundedness* $(a.s.)$ *or bounded expectations, respectively, in equations* (2.13)–(2.15).

Let X_{nt} *and* Y_{nt}, $1 \le t \le n$, $n \in \mathbb{N}$, *be random* k*-vectors satisfying*

$$(2.13) \quad n^{-1} \sum_{t=1}^{n} |Y_{nt}|^2 = O(1), \quad \forall\, \delta > 0: n^{-1} \sum_{t=1}^{n} |Y_{nt}|^2 1_{\{n^{-1/2}|Y_{nt}| > \delta\}} \to 0$$

and

$$(2.14) \qquad n^{-1} \sum_{t=1}^{n} |X_{nt} - Y_{nt}|^2 \to 0.$$

If $\varphi: \mathbb{R}^k \to \mathbb{R}$ is continuous at 0 with $\varphi(0) = 0$ and if $\varphi(\cdot)$ is bounded, then

$$(2.15) \qquad n^{-1} \sum_{t=1}^{n} |X_{nt}|^2 \varphi\big(n^{-1/2} X_{nt}\big) \to 0.$$

If the process $\{Y_{nt} = Y_t: t \in \mathbb{N}\}$ is stationary with finite second moments then (2.13) is satisfied.

PROOF. Let $\varepsilon > 0$. Without loss of generality, we assume $0 \le \varphi(u) \le 1$. Since φ is continuous at 0, there exists a $\delta > 0$ such that $0 \le \varphi(u) \le \varepsilon + 1_{\{|u| > \delta\}}$. Since for all vectors a and b,

$$(2.16) \qquad |a + b|^2 1_{\{|a+b| > \delta\}} \le 4\Big(|a|^2 1_{\{|a| > \delta/2\}} + |b|^2 1_{\{|b| > \delta/2\}}\Big),$$

this yields

$$\begin{aligned} n^{-1} \sum_{t=1}^{n} |X_{nt}|^2 \varphi\big(n^{-1/2} X_{nt}\big) & \\ &\le \varepsilon n^{-1} \sum_{t=1}^{n} |X_{nt}|^2 + n^{-1} \sum_{t=1}^{n} |X_{nt}|^2 1_{\{n^{-1/2}|X_{nt}| > \delta\}} \\ &\le 2\varepsilon n^{-1} \sum_{t=1}^{n} |Y_{nt}|^2 + 2\varepsilon n^{-1} \sum_{t=1}^{n} |X_{nt} - Y_{nt}|^2 \\ &\quad + 4 n^{-1} \sum_{t=1}^{n} |Y_{nt}|^2 1_{\{n^{-1/2}|Y_{nt}| > \delta/2\}} \\ &\quad + 4 n^{-1} \sum_{t=1}^{n} |X_{nt} - Y_{nt}|^2 1_{\{n^{-1/2}|X_{nt} - Y_{nt}| > \delta/2\}}. \end{aligned}$$

For the three different modes of convergence, the conclusion (2.15) is immediate from assumptions (2.13) and (2.14).

To prove the final assertion concerning stationary sequences, note that the first condition in (2.13) is trivially met. The second one follows from (for n large and some fixed M):

$$n^{-1}\sum_{t=1}^{n}|Y_t|^2 1_{\{n^{-1/2}|Y_t|>\delta\}} \le n^{-1}\sum_{t=1}^{n}|Y_t|^2 1_{\{|Y_t|>M\}} \to E\big(|Y|^2 1_{\{|Y|>M\}}\,|\mathscr{F}\big),$$

where $\mathscr{F}$ denotes the invariant sigma field. The second condition of (2.13) is obtained by letting $M\to\infty$. □

LEMMA 2.2. *Let* $\{(X_{nt}, F_t^n)\colon 1\le t\le n\}$ *be an adapted, square-integrable process. If*

$$\sum_{t=1}^{n} E\big(X_{nt}^2\,|\mathscr{F}_{t-1}^n\big) \to_P 0, \tag{2.17}$$

then

$$\sum_{t=1}^{n} X_{nt}^2 \to_P 0, \tag{2.18}$$

$$\sum_{t=1}^{n}\{X_{nt} - E(X_{nt}\,|\mathscr{F}_{t-1}^n)\} \to_P 0. \tag{2.19}$$

PROOF. Convergence (2.18) is implied by

$$\sum_{t=1}^{n} X_{nt}^2 \le 2\sum_{t=1}^{n}\{X_{nt} - E(X_{nt}\,|\mathscr{F}_{t-1}^n)\}^2 + 2\sum_{t=1}^{n} E\big(X_{nt}^2\,|\mathscr{F}_{t-1}^n\big)$$

and Theorem 2.23, page 44, of Hall and Heyde (1980). Convergence (2.19) follows from Corollary 3.1, page 58, of Hall and Heyde (1980). Compare Theorem VIII.1, page 171, of Pollard (1984). □

LEMMA 2.3. *Under* (2.9) *and* (2.10),

$$\begin{aligned} n^{-1}\sum_{t=1}^{n} E\big(|U_t'\psi_t|^2\,|\mathscr{F}_{t-1}^n\big) &\to_P \lambda' I(\theta_0)\lambda,\\ n^{-1}\sum_{t=1}^{n}|U_t'\psi_t|^2 &\to_P \lambda' I(\theta_0)\lambda, \end{aligned}\tag{2.20}$$

$$\begin{aligned} n^{-1}\sum_{t=1}^{n} E\big(|U_{nt}'\psi_t|^2\,|\mathscr{F}_{t-1}^n\big) &\to_P \lambda' I(\theta_0)\lambda,\\ n^{-1}\sum_{t=1}^{n}|U_{nt}'\psi_t|^2 &\to_P \lambda' I(\theta_0)\lambda, \end{aligned}\tag{2.21}$$

$$\forall \delta > 0: n^{-1} \sum_{t=1}^{n} E\Big(|U_t'\psi_t|^2 1_{\{n^{-1/2}|U_t'\psi_t| > \delta\}} \,|\, \mathscr{F}_{t-1}^n\Big) \to_P 0, \tag{2.22}$$

$$\forall \delta > 0: n^{-1} \sum_{t=1}^{n} E\Big(|U_{nt}'\psi_t|^2 1_{\{n^{-1/2}|U_{nt}'\psi_t| > \delta\}} \,|\, \mathscr{F}_{t-1}^n\Big) \to_P 0, \tag{2.23}$$

with convergence under $P_{\theta_n g}$.

PROOF. First we prove the four conditional statements in (2.20)–(2.23). Relation (2.20) is already given in (2.3); (2.21) follows from (2.20), (2.10) and Cauchy–Schwarz. Equation (2.22) follows from Lemma 2.1 with $X_{nt} = Y_{nt} = U_t$ and

$$\varphi(x) = E|\psi_t|^2 1_{\{|x'\psi_t| > \delta\}}; \tag{2.24}$$

(2.23) follows similarly with $X_{nt} = U_{nt}$ and $Y_{nt} = U_t$.

The unconditional statements in (2.20) and (2.21) are immediate consequences of the conditional statements and Theorem 2.23, page 44 of Hall and Heyde (1980). □

LEMMA 2.4. *With the notation* (2.11), *under* $P_{\theta_n g}$,

$$\sum_{t=1}^{n} E\Big(|T_{nt} - n^{-1/2}U_{nt}'\psi_t|^2 \,|\, \mathscr{F}_{t-1}^n\Big) \to_P 0, \tag{2.25}$$

$$\sum_{t=1}^{n} |T_{nt} - n^{-1/2}U_{nt}'\psi_t|^2 \to_P 0, \tag{2.26}$$

$$\max_{1 \le t \le n} |T_{nt}| \to_P 0. \tag{2.27}$$

PROOF. Where possible we try to copy the proof of Lemma A.9.5, page 509 in BKRW and, therefore, we introduce a conditioning argument. We refer to BKRW for the details on bounding the conditional terms and we will concentrate on the remaining part. We obtain

$$\begin{aligned} &E\Big(|T_{nt} - n^{-1/2}U_{nt}'\psi_t|^2 \,|\, \mathscr{F}_{t-1}^n\Big) \\ &\quad \le 4\|s(\zeta_0 + n^{-1/2}U_{nt}) - s(\zeta_0) - n^{-1/2}U_{nt}'\dot{s}(\zeta_0)\|^2 \\ &\quad \le n^{-1}|U_{nt}|^2\varphi\big(n^{-1/2}U_{nt}\big), \end{aligned}$$

with

$$\begin{aligned} \varphi(x) = {} & 4|x|^{-2}\|s(\zeta_0 + x) - s(\zeta_0) - x'\dot{s}(\zeta_0)\|^2 1_{\{|x| \le 1\}} \\ & + 4|x|^{-2}\big(2 + |x| \cdot \|\, |\dot{s}(\zeta_0)| \,\|\big)^2 1_{\{|x| > 1\}}. \end{aligned}$$

Application of (2.9), (2.10), Assumption B and Lemma 2.1 yields convergence (2.25).

This implies (2.26) by the first part of Lemma 2.2.

To prove (2.27), note that by (2.16),

$$\sum_{t=1}^{n} P(|T_{nt}| > \delta \,|\, \mathscr{F}_{t-1}^{n}) \le 4\delta^{-2} \sum_{t=1}^{n} E\big(|T_{nt} - n^{-1/2}U_{nt}'\psi_t|^2 \,|\, \mathscr{F}_{t-1}^{n}\big) + 4\delta^{-2}n^{-1} \sum_{t=1}^{n} E\Big(|U_{nt}'\psi_t|^2 1_{\{n^{-1/2}|U_{nt}'\psi_t| > \delta/2\}} \,|\, \mathscr{F}_{t-1}^{n}\Big).$$

The first term at the right-hand side converges to zero by (2.25) and the second one by (2.23). The conclusion (2.27) is obtained from Dvoretzky's lemma [see Lemma 2.5, page 45, Hall and Heyde (1980)],

$$P\Big\{\max_{1\le t\le n} |T_{nt}| > \delta\Big\} \le \varepsilon + P\bigg\{\sum_{t=1}^{n} P\big(|T_{nt}| > \delta \,|\, \mathscr{F}_{t-1}^{n}\big) > \varepsilon\bigg\} \to \varepsilon.$$

This completes the proof of the Lemma. □

Now we continue the proof of our theorem. As in the previous lemma we rely heavily on Appendix A.9 of BKRW, especially pages 509–513.

The convergence of the first two terms in the expansion (2.12) to the required normal distributions is obvious from (2.20), (2.22) and application of Corollary 3.1, page 58, of Hall and Heyde (1980).

The remainder term R_n defined in (2.6) consists of the last seven terms at the right-hand side of (2.12). The term Λ_n^s concerning the initial variables converges to zero by assumption. The second term of R_n converges to zero because of (2.10) and application of the second part of Lemma 2.2. Convergence to zero of the third remainder term follows from Lemma 2.3. Convergence of the next term follows from (2.25) and the second part of Lemma 2.2. To kill the fifth remainder term, note that [cf. (c), page 459 of BKRW]:

$$\begin{aligned}
&\bigg|\sum_{t=1}^{n} \Big\{E(T_{nt} \,|\, \mathscr{F}_{t-1}^{n}) + \tfrac{1}{4}n^{-1}|U_{nt}'\psi_t|^2\Big\}\bigg| \\
&\quad\le \tfrac{1}{4}\sum_{t=1}^{n} \Big|E\big(T_{nt}^2 \,|\, \mathscr{F}_{t-1}^{n}\big) - n^{-1}E\big(|U_{nt}'\psi_t|^2 \,|\, \mathscr{F}_{t-1}^{n}\big)\Big| \\
&\qquad + \tfrac{1}{4}n^{-1}\bigg|\sum_{t=1}^{n} \Big\{E\big(|U_{nt}'\psi_t|^2 \,|\, \mathscr{F}_{t-1}^{n}\big) - |U_{nt}'\psi_t|^2\Big\}\bigg| \\
&\qquad + \sum_{t=1}^{n} \int_{\{s(\zeta_0 = 0\}} s^2\big(\zeta_0 + n^{-1/2}U_{nt}\big)\, d\mu \\
&\quad\le \tfrac{1}{4}\sum_{t=1}^{n} E\Big(\big|T_{nt}^2 - n^{-1}|U_{nt}'\psi_t|^2\big| \,|\, \mathscr{F}_{t-1}^{n}\Big) \\
&\qquad + \tfrac{1}{4}n^{-1}\bigg|\sum_{t=1}^{n} \Big\{E\big(|U_{nt}'\psi_t|^2 \,|\, \mathscr{F}_{t-1}^{n}\big) - |U_{nt}'\psi_t|^2\Big\}\bigg| \\
&\qquad + 2\sum_{i\in\{-1,1\}} \sum_{t=1}^{n} \Big\| s\big(\zeta_0 + in^{-1/2}U_{nt}\big) - s(\zeta_0) - in^{-1/2}U_{nt}'\dot{s}(\zeta_0)\Big\|^2.
\end{aligned}$$

The second term at the right-hand side of this inequality converges to zero by Lemma 2.3 and the last term at the right-hand side has been treated in the argument leading to (2.25). The first term converges to zero by (2.25), Cauchy–Schwarz and the tightness of $n^{-1}\sum_{t=1}^{n}|U_{nt}'\psi_t|^2$ as implied by Lemma 2.3. Similarly we obtain from (2.26) that the sixth remainder term of (2.12) is negligible.

To show that the final term of R_n also converges to zero observe that (2.27) implies that we can restrict attention to the set

$$A_n = \left\{ \max_{1\le t\le n} |T_{nt}| \le \tfrac{1}{2} \right\}.$$

Then the variables α_{nt} are all bounded by one. Moreover, note that (2.26) and (2.21) imply that $\sum_{t=1}^{n} T_{nt}^2$ is tight. Now the required relationship follows from Slutsky's lemma, (2.27) and

$$\left| \sum_{t=1}^{n} \alpha_{nt} T_{nt}^3 \right| \le \max_{1\le t\le n} |T_{nt}| \sum_{t=1}^{n} T_{nt}^2.$$

The contiguity and smoothness can be obtained along similar lines as in the i.i.d. case; compare, for example, BKRW, page 513. This completes the proof of the theorem. □

The uniform LAN property as shown above allows us to apply the convolution theorem. Hence, if g is known, the lower bound at θ_0 for regular estimators of θ is given by $I(\theta_0)^{-1}$. In the parametric model with g known, this is also the best attainable lower bound. In a semiparametric context, with g unknown, the actual lower bound might be higher. In this case one also needs the LAN property for submodels where g is allowed to vary with sample size n. However, as we will see in Section 3 below, this is not necessary in the time-series context discussed in this paper: (part of) the parameter will be adaptively estimable and, hence, the same lower bound applies as in the parametric case. We conclude this section with several remarks with regard to group models in time series, since they fit perfectly into the framework introduced in the first part of this section. Our main assumption in this endeavor will be that the time-series model can be obtained as a group model with predictable time dependent Euclidean parameter. This will be explained in detail below.

To show that group models in time series fit into the setting discussed above, we adopt the notation for i.i.d. group models as given in Section 4.2, pages 88–103 of BKRW. Let $\mathscr{A}$ be a group of measurable transformations on $\mathbb{R}^m$ under composition. We assume that the elements a of $\mathscr{A}$ are parametrized by a parameter $\zeta \in Z \subset \mathbb{R}^q$. The parameter value ζ_0 yields the identity transformation. Let $g \in \mathscr{G}$ determine a fixed probability distribution G on $\mathbb{R}^m$. Then the group model induced by $(\mathscr{A}, G)$ is

$$\mathscr{Q} = \left\{ Q_\zeta = G a_\zeta^{-1} \colon \zeta \in Z \right\}.$$

Let μ be a measure on the Borel sets of $\mathbb{R}^m$ such that μa^{-1} is equivalent to μ for all $a \in \mathscr{A}$ and such that μ dominates G. Then, we have

$$\frac{dGa_\zeta^{-1}}{d\mu}(y) = \frac{dG}{d\mu}\left(a_\zeta^{-1}y\right)\frac{d\mu a_\zeta^{-1}}{d\mu}(y).$$

This i.i.d. group model generates a time-series model in the following way. Let X denote some observed starting conditions and/or exogenous variables with distribution $P_{\theta g}^0$. Define $\mathscr{F}_0 = \mathscr{F}(X)$. Furthermore, we observe $Y_1, \dots, Y_n$. Let $\mathscr{F}_t$ denote the filtration generated by $X, Y_1, \dots, Y_t$, $t = 1, \dots, n$. Finally, we are given a Z-valued predictable process $Z_t(\theta)$ depending on the unknown parameter θ such that the conditional distribution of Y_t given $\mathscr{F}_{t-1}$ under $P_{\theta g}$ is given by $Q_{Z_t(\theta)}$, $t = 1, \dots, n$.

It will be convenient to introduce innovations in the following way. Define $\varepsilon_t(\zeta) = a_\zeta^{-1}(Y_t)$, $\zeta \in Z$ and $\varepsilon_t^n = \varepsilon_t(Z_t(\theta_n))$. Note that under θ_n the ε_t^n are i.i.d. with distribution G and independent of X. Moreover, we have $Y_t = a_{Z_t(\theta_n)}\varepsilon_t^n$. All this implies that the log-likelihood ratio statistic Λ_n of the observations $X, Y_1, \dots, Y_n$ for $\tilde{\theta}_n$ with respect to θ_n can be written as

$$\Lambda_n = \sum_{t=1}^n \left\{\log\left(\frac{dG}{d\mu}\left(a_{Z_t(\tilde{\theta}_n)}^{-1}a_{Z_t(\theta_n)}\varepsilon_t^n\right)\frac{d\mu a_{Z_t(\tilde{\theta}_n)}^{-1}/d\mu}{d\mu a_{Z_t(\theta_n)}^{-1}/d\mu}\left(a_{Z_t(\theta_n)}\varepsilon_t^n\right)\right) - \log\frac{dG}{d\mu}(\varepsilon_t^n)\right\} + \Lambda_n^s,$$

where Λ_n^s denotes the log-likelihood ratio contribution of X.

Since by definition $\mu a(B) = \mu(aB)$ for all $a \in \mathscr{A}$ and B Borel, we have

$$\frac{d\mu\tilde{a}^{-1}}{d\mu a^{-1}}(a\varepsilon) = \frac{d\mu\tilde{a}^{-1}a}{d\mu}(\varepsilon).$$

Moreover, the group structure allows us to write

$$a_{Z_t(\tilde{\theta}_n)}^{-1}a_{Z_t(\theta_n)} = a_{\zeta_0 + W_{nt}'(\tilde{\theta}_n - \theta_n)}^{-1}$$

for some W_{nt} and hence the log-likelihood ratio statistic can be written as

$$\begin{aligned}\Lambda_n &= \sum_{t=1}^n \left\{\log\left(\frac{dG}{d\mu}\left(a_{\zeta_0 + W_{nt}'(\tilde{\theta}_n - \theta_n)}^{-1}\varepsilon_t^n\right)\frac{d\mu a_{\zeta_0 + W_{nt}'(\tilde{\theta}_n - \theta_n)}^{-1}}{d\mu}(\varepsilon_t^n)\right) - \log\frac{dG}{d\mu}(\varepsilon_t^n)\right\} + \Lambda_n^s \\ &= \sum_{t=1}^n \left\{\log\frac{dGa_{\zeta_0 + W_{nt}'(\tilde{\theta}_n - \theta_n)}^{-1}}{d\mu}(\varepsilon_t^n) - \log\frac{dG}{d\mu}(\varepsilon_t^n)\right\} + \Lambda_n^s.\end{aligned}$$

In this way we have shown the desired expansion (2.1) for group models in time series.

In the group models $\mathscr{P}$ in time series as introduced above, the Euclidean parameter will typically consist of two parts, $\theta = (\nu', \eta')'$, where η and ζ

have the same interpretation and dimension. The latter part of the vector $Z_t(\theta)$ consists, for example, of the parameter $\zeta = \eta$. In this case estimation of η is equivalent to estimation of ζ in the i.i.d. model $\mathscr{Q}$ and the second block of $W = (W_1', W_2')'$ (corresponding to the parameter η) is a square, invertible $q \times q$-matrix not depending on t. The special location-scale group structure will be considered in more detail in Section 4. If the parametrization is not of the given form, necessary and sufficient conditions can be derived along the lines of DKW to obtain a reparametrization of the required form. For the estimation problem of η, we will rely on the existing literature on i.i.d. models. Here we will concentrate on the time-series parameter ν, describing the dependence structure of $\mathscr{P}$. The convolution theorem implies that the upper left $(p-q) \times (p-q)$-block $(I(\theta_0)^{-1})_{11}$ of $I(\theta_0)^{-1}$ is a lower bound for ν. By Assumption D and the fixed character of the invertible matrix W_2 this bound is equal to the probability limit of

$$\begin{aligned}(2.28)\qquad &\Big\{n^{-1}\sum_{t=1}^{n} W_{1t}JW_{1t}' \\ &-n^{-1}\sum_{t=1}^{n} W_{1t}JW_2'[W_2JW_2']^{-1}n^{-1}\sum_{s=1}^{n} W_2JW_{1s}'\Big\}^{-1} \\ &= \left\{n^{-1}\sum_{t=1}^{n}\left(W_{1t}-n^{-1}\sum_{s=1}^{n}W_{1s}\right)J\left(W_{1t}-n^{-1}\sum_{s=1}^{n}W_{1s}\right)'\right\}^{-1}.\end{aligned}$$

As a corollary of the results in the following section, we may construct an estimator of ν attaining the lower bound implied by (2.28) when g is not known. That is, adaptive estimation of ν with respect to g is possible. Estimation of ν in the model with both θ and g unknown is not harder than estimation of ν in the model with only θ unknown (and g known). See Example 3.1.

Adaptiveness of ν with respect to the nuisance g in the model with θ unknown does not imply that ν is also adaptive with respect to both η and g. If both η and g are known, the lower bound for ν decreases (due to the absence of the nuisance parameter η) to the probability limit of

$$\left\{n^{-1}\sum_{t=1}^{n} W_{1t}JW_{1t}'\right\}^{-1}.$$

However, both limits agree if the average of the W_{1t}'s tends to zero and we will have full adaptation of ν with respect to both η and g. See also Example 3.1.

An efficient estimator for θ in our semiparametric model can be obtained by combining the efficient estimator of η and the adaptive estimator of ν.

3. Adaptive estimation. In this section the density $g \in \mathscr{G}$ of the innovations is assumed to be a nuisance parameter. We will construct adaptive estimators for linear functions of θ under some additional assumptions. In Section 4 we show the validity of these assumptions in a wide class of examples. The structure of the score, $\dot{l} = W\psi$, as given in Assumption E, is a multivariate generalization of the one considered in DKW. Their derivations concerning the possibility of adaptation carry over to this case.

In the typical situation discussed at the end of Section 2, the adaptiveness condition (2.4) of DKW is satisfied since the covariance matrix

$$n^{-1} \sum_{t=1}^{n} \left(W_t - n^{-1} \sum_{s=1}^{n} W_s\right)\left(W_t - n^{-1} \sum_{s=1}^{n} W_s\right)'$$

has q zero eigenvalues (recall that W_2 does not depend on t); see (3.4) of their Theorem 3.1. This suggests that adaptive estimation of ν with respect to g (in the presence of the nuisance parameter η) is likely to be possible. In fact, we will later obtain an estimator of ν attaining the lower bound. Knowledge of the infinite-dimensional parameter is not helpful (at least asymptotically); see Example 3.1.

In the more general situation of estimating linear functions of θ, we will adapt the method proposed by Schick (1986) for i.i.d. models. As before, expectations and limits will be taken under the probability measure $P_{\theta_n g}$. To apply this method we need the following additional notation and assumptions.

NOTATION AND ASSUMPTIONS.

F. There exists an estimator $\bar{\psi} = \bar{\psi}(\varepsilon_1, \dots, \varepsilon_n)$ of ψ such that

$$E\left(|\bar{\psi}(\varepsilon) - \psi(\varepsilon)|^2 \mid \varepsilon_1, \dots, \varepsilon_n\right) \to_P 0, \tag{3.1}$$

where $\varepsilon, \varepsilon_1, \dots, \varepsilon_n$ are i.i.d. from a distribution with density g.

Let $C \in \mathbb{R}^{q \times r}$ be an orthogonal matrix of full rank, where $0 \le r \le q$ and $C = 0_q$ if $r = 0$, such that the $\sqrt{n}$-unbiasedness condition holds for the estimator $C'\bar{\psi}$ of $C'\psi$,

$$C'\sqrt{n}\, E\left(\bar{\psi}(\varepsilon) \mid \varepsilon_1, \dots, \varepsilon_n\right) \to_P 0. \tag{3.2}$$

G. $\tilde{\theta}_n$ is a $\sqrt{n}$-consistent discretized estimator (under $P_{\theta_n g}$).

H. The matrices $W_t = W_t(\theta_n)$, $t = 1, \dots, n$, do not depend on g and there exists a square-integrable $W_0 = W_0(\theta_0, g)$ such that

$$n^{-1} \sum_{t=1}^{n} W_t \to_P W_0. \tag{3.3}$$

The first part of Assumption F is often satisfied in applications; see, for example, Proposition 7.8.1, page 400 of BKRW. The second part of this assumption is merely notation. Observe that the matrix C may span only a part of the space where $\sqrt{n}$-unbiased estimation of ψ is possible. In principle, we would like to use a (nonparametric) estimate of ψ which is completely $\sqrt{n}$-unbiased. Often this will not be possible or we do not know the exact bias properties of $\bar{\psi}$. The choice $C = 0_q$ renders (3.2) void, but forces some additional (possibly not needed) orthogonality restrictions in (3.6) below. However, if (3.6) is fulfilled with $C = 0_q$, then it is not necessary to bother about an "optimal" estimator of ψ. The $\sqrt{n}$-unbiasedness property of $\bar{\psi}$ and the orthogonality (3.6) yield a $\sqrt{n}$-unbiasedness condition for time series similar to the one in i.i.d. models; see Remark 3.1. Assumption G describes the existence of a preliminary estimator needed to be able to apply a Newton–Raphson step. Generally, moment estimators or estimators based on quasi maximum likelihood methods yield the $\sqrt{n}$-consistency. Finally, Assumption H requires convergence of averages in addition to the tightness implied by Assumption D.

In our construction of efficient estimators, we will adopt the principle of sample splitting, not because of its elegancy but because it yields a relatively easy way to obtain such estimators under minimal conditions. See also Remark 3.4 for references to other constructions which might be preferable for small sample sizes. Using $\tilde{\theta}_n$, compute the residuals $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$ with $\tilde{\varepsilon}_t = \varepsilon(X_n, Y_1, \dots, Y_t, \tilde{\theta}_n)$, $t = 1, \dots, n$. Let α_n be an integer such that $\alpha_n/n \to \alpha \in (0,1)$ as $n \to \infty$. Split the time series of residuals in two parts: $(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_{\alpha_n})$ and $(\tilde{\varepsilon}_{\alpha_n+1}, \dots, \tilde{\varepsilon}_n)$. Use Assumption F to estimate ψ twice, using the first and second part of the residuals separately. Call these estimators $\hat{\psi}_{n1} = \bar{\psi}(\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_{\alpha_n})$ and $\hat{\psi}_{n2} = \bar{\psi}(\tilde{\varepsilon}_{\alpha_n+1}, \dots, \tilde{\varepsilon}_n)$, respectively. If the calculated residuals would have been the innovations themselves with density g then we would have (by Assumption F)

$$\int |\hat{\psi}_{nj}(x) - \psi(x)|^2 g(x)\, dx \to_P 0, \qquad j = 1, 2.$$

This relation implies also that, if $\tilde{\theta}_n$ is nonrandom, this convergence holds true under $P_{\tilde{\theta}_n g}$.

To stress the dependence of the information matrix $I(\theta)$ on g we will write $I_g(\theta)$. Under the present conditions it is possible to estimate the information matrix consistently; there exist estimators $\hat{I}_n$ satisfying

$$\hat{I}_n \to_P I_g(\theta_0). \tag{3.4}$$

An explicit construction of such an estimator is given in the following lemma. Abbreviate $\psi_t = \psi(\varepsilon_t)$, $\tilde{\psi}_t = \psi(\tilde{\varepsilon}_t)$, $\hat{\psi}_{jt} = \hat{\psi}_{nj}(\tilde{\varepsilon}_t)$ and denote $\tilde{W}_t = W_t(\tilde{\theta}_n)$.

LEMMA 3.1. *Assume* A–G *for all sequences* $\theta_n = \theta_0 + O(n^{-1/2})$ *and* $\tilde{\theta}_n = \theta_0 + O(n^{-1/2})$. *Then the estimator*

$$\hat{I}_n = n^{-1}\left\{ \sum_{t=1}^{\alpha_n} \tilde{W}_t \hat{\psi}_{2t} \hat{\psi}'_{2t} \tilde{W}'_t + \sum_{t=\alpha_n+1}^{n} \tilde{W}_t \hat{\psi}_{1t} \hat{\psi}'_{1t} \tilde{W}'_t \right\} \tag{3.5}$$

of the information matrix satisfies (3.4).

PROOF. By Assumption G it suffices to consider this equation for all nonrandom $\tilde{\theta}_n$ such that $\sqrt{n}(\tilde{\theta}_n - \theta_n) = O(1)$; see, for example, Theorem 2.5.2, page 44, of BKRW. We prove that the first sum in (3.5) converges to $I(\theta_0)$ under $P_{\tilde{\theta}_n g}$. Write, with $a \in \mathbb{R}^p$,

$$\begin{aligned}
&\left|\frac{1}{\alpha_n}\sum_{t=1}^{\alpha_n}|a'\tilde{W}_t\hat{\psi}_{2t}|^2 - a'I_g(\theta_0)a\right| \\
&\quad\le \left|\frac{1}{\alpha_n}\sum_{t=1}^{\alpha_n}|a'\tilde{W}_t\tilde{\psi}_t|^2 - a'I_g(\theta_0)a\right| + \left|\frac{1}{\alpha_n}\sum_{t=1}^{\alpha_n}\left\{|a'\tilde{W}_t\hat{\psi}_{2t}|^2 - |a'\tilde{W}_t\tilde{\psi}_t|^2\right\}\right| \\
&\quad\le \left|\frac{1}{\alpha_n}\sum_{t=1}^{\alpha_n}|a'\tilde{W}_t\tilde{\psi}_t|^2 - a'I_g(\theta_0)a\right| + \frac{1}{\alpha_n}\sum_{t=1}^{\alpha_n}\left|a'\tilde{W}_t\left(\hat{\psi}_{2t} - \tilde{\psi}_t\right)\right|^2 \\
&\qquad + \left|\frac{1}{\alpha_n}\sum_{t=1}^{\alpha_n}2a'\tilde{W}_t\left(\hat{\psi}_{2t} - \tilde{\psi}_t\right)\tilde{\psi}_t'\tilde{W}_t'a\right|.
\end{aligned}$$

The first term at the right-hand side is negligible by the unconditional part of (2.20). The other terms converge to zero by Cauchy–Schwarz, Assumptions D and F and the first part of Lemma 2.2 using the filtration $\mathscr{F}_t^n \vee \mathscr{F}(\varepsilon_{\alpha_n+1}, \dots, \varepsilon_n)$. Note that we use here that, under $P_{\tilde{\theta}_n g}$, $\hat{\psi}_{2t} - \tilde{\psi}_t = \bar{\psi}(\varepsilon_{\alpha_n+1}, \dots, \varepsilon_n)(\varepsilon_t) - \psi_t$ and that, with E_0 denoting expectation under $P_{\theta_0 g}$, $E_0(|\bar{\psi}(\varepsilon_{\alpha_n+1}, \dots, \varepsilon_n)(\varepsilon_t) - \psi_t|^2 \,|\, \mathscr{F}_t^n, \varepsilon_{\alpha_n+1}, \dots, \varepsilon_n)$ is the same for $t = 1, \dots, \alpha_n$. The second sum in (3.5) can be treated similarly. Since $a \in \mathbb{R}^p$ is arbitrary this implies convergence of $\hat{I}_n$ under $P_{\tilde{\theta}_n g}$. Contiguity yields the required convergence (3.4). □

Now we are ready to present our adaptive estimator.

THEOREM 3.1. *Assume* A–H *for all sequences* $\theta_n = \theta_0 + O(n^{-1/2})$ *and* $\tilde{\theta}_n = \theta_0 + O(n^{-1/2})$ *and let A be a matrix satisfying*

$$\forall\, \theta \in \Theta,\ g \in \mathscr{G}\colon A'I_g(\theta)^{-1}W_0(\theta, g)(I - CC') = 0, \qquad P_{\theta g} - a.s. \tag{3.6}$$

Then, with $\hat{I}_n$ some estimator of the information matrix satisfying (3.4),

$$\begin{aligned}
\hat{\nu}_n = A'\tilde{\theta}_n + A'\hat{I}_n^{-1}n^{-1}\Bigg\{&\sum_{t=1}^{\alpha_n}\tilde{W}_t\hat{\psi}_{2t} + \sum_{t=\alpha_n+1}^{n}\tilde{W}_t\hat{\psi}_{1t} \\
&- \frac{1}{\alpha_n}\sum_{s=1}^{\alpha_n}\tilde{W}_s(I - CC')\sum_{t=1}^{\alpha_n}\hat{\psi}_{2t} \\
&- \frac{1}{n - \alpha_n}\sum_{s=\alpha_n+1}^{n}\tilde{W}_s(I - CC')\sum_{t=\alpha_n+1}^{n}\hat{\psi}_{1t}\Bigg\}
\end{aligned} \tag{3.7}$$

is an adaptive estimator of $\nu = A'\theta$, *that is, under* $P_{\theta_n g}$,

$$\sqrt{n}\left\{\hat{\nu}_n - A'\theta_n - A'I_g(\theta_n)^{-1}n^{-1}\sum_{t=1}^{n} W_t\psi_t\right\} \to_P 0. \tag{3.8}$$

Consequently,

$$\sqrt{n}\,(\hat{\nu}_n - A'\theta_n) \to_{\mathscr{L}} N\Big(0, A'I_g(\theta_0)^{-1}A\Big). \tag{3.9}$$

Proof. In view of the LAN property Theorem 2.1 and the Hájek–Le Cam convolution theorem we need to prove (3.8) [cf. Hájek (1970) and Theorem 2.3.1, page 24 of BKRW]. Insert the expression (3.7) for $\hat{\nu}_n$ into the left-hand side of (3.8) and observe that, by Assumption G, it suffices to consider this equation for all nonrandom $\tilde{\theta}_n$ such that $\sqrt{n}\,(\tilde{\theta}_n - \theta_n) = O(1)$; see, for example, Theorem 2.5.2, page 44 of BKRW. Contiguity obtained in the LAN Theorem 2.1 implies that it suffices to show convergence to zero under $P_{\tilde{\theta}_n g}$. Let $\tilde{E}$ denote expectation under $P_{\tilde{\theta}_n g}$ and use (3.6) to rewrite (3.8) and obtain

$$\begin{aligned}
&\sqrt{n}\left\{\hat{\nu}_n - A'\theta_n - A'I_g(\theta_n)^{-1}n^{-1}\sum_{t=1}^{n} W_t\psi_t\right\}\\
&\quad= A'\sqrt{n}\left\{\tilde{\theta}_n - \theta_n + I_g\big(\tilde{\theta}_n\big)^{-1}n^{-1}\sum_{t=1}^{n}\tilde{W}_t\tilde{\psi}_t - I_g(\theta_n)^{-1}n^{-1}\sum_{t=1}^{n} W_t\psi_t\right\}\\
&\qquad + A'\Big\{\hat{I}_n^{-1} - I_g\big(\tilde{\theta}_n\big)^{-1}\Big\}n^{-1/2}\sum_{t=1}^{n}\tilde{W}_t\tilde{\psi}_t\\
&\qquad - A'\Big\{\hat{I}_n^{-1} - I_g(\theta_0)^{-1}\Big\}W_0(I - CC')n^{-1/2}\sum_{t=1}^{n}\tilde{\psi}_t\\
&\qquad + A'\hat{I}_n^{-1}\left\{\left(W_0 - \frac{1}{\alpha_n}\sum_{s=1}^{\alpha_n}\tilde{W}_s\right)(I - CC')n^{-1/2}\sum_{t=1}^{\alpha_n}\tilde{\psi}_t\right.\\
&\qquad\qquad \left. + \left(W_0 - \frac{1}{n-\alpha_n}\sum_{s=\alpha_n+1}^{n}\tilde{W}_s\right)(I - CC')n^{-1/2}\sum_{t=\alpha_n+1}^{n}\tilde{\psi}_t\right\}\\
&\qquad + A'\hat{I}_n^{-1}\left\{n^{-1/2}\sum_{s=1}^{\alpha_n}\tilde{W}_s CC'\tilde{E}\Big(\hat{\psi}_{21}\,|\,\tilde{\varepsilon}_{\alpha_n+1},\dots,\tilde{\varepsilon}_n\Big)\right.\\
&\qquad\qquad \left. + n^{-1/2}\sum_{s=\alpha_n+1}^{n}\tilde{W}_s CC'\tilde{E}\Big(\hat{\psi}_{1n}\,|\,\tilde{\varepsilon}_1,\dots,\tilde{\varepsilon}_{\alpha_n}\Big)\right\}
\end{aligned}$$

$$
\begin{aligned}
&- A'\hat{I}_n^{-1}\Bigg\{\frac{1}{\alpha_n}\sum_{s=1}^{\alpha_n}\tilde{W}_s(I - CC')n^{-1/2} \\
&\qquad\qquad \times \sum_{t=1}^{\alpha_n}\Big\{\hat{\psi}_{2t} - \tilde{\psi}_t - \tilde{E}\Big(\hat{\psi}_{2t} \mid \tilde{\varepsilon}_{\alpha_n+1}, \dots, \tilde{\varepsilon}_n\Big)\Big\} \\
&\qquad + \frac{1}{n-\alpha_n}\sum_{s=\alpha_n+1}^{n}\tilde{W}_s(I - CC')n^{-1/2} \\
&\qquad\qquad\qquad \times \sum_{t=\alpha_n+1}^{n}\Big\{\hat{\psi}_{1t} - \tilde{\psi}_t - \tilde{E}\Big(\hat{\psi}_{1t} \mid \tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_{\alpha_n}\Big)\Big\}\Bigg\} \\
&+ A'\hat{I}_n^{-1}\Bigg\{n^{-1/2}\sum_{t=1}^{\alpha_n}\tilde{W}_t\Big\{\hat{\psi}_{2t} - \tilde{\psi}_t - \tilde{E}\Big(\hat{\psi}_{2t} \mid \tilde{\varepsilon}_{\alpha_n+1}, \dots, \tilde{\varepsilon}_n\Big)\Big\} \\
&\qquad + n^{-1/2}\sum_{t=\alpha_n+1}^{n}\tilde{W}_t\Big\{\hat{\psi}_{1t} - \tilde{\psi}_t - \tilde{E}\Big(\hat{\psi}_{1t} \mid \tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_{\alpha_n}\Big)\Big\}\Bigg\}.
\end{aligned}
$$

Now, this expression converges to zero under $P_{\tilde{\theta}_n g}$ (and hence, by contiguity, under $P_{\theta_n g}$ as desired) in view of smoothness (2.8), convergence of $\hat{I}_n$ to $I_g(\theta_0)$, continuity and invertibility of $I_g(\cdot)$, the LAN property, tightness of $\hat{I}_n^{-1}$, convergence to zero of expressions like

$$W_0 - \frac{1}{\alpha_n}\sum_{s=1}^{\alpha_n}\tilde{W}_s,$$

the implied tightness of partial averages, the $\sqrt{n}$-unbiasedness (3.2) of $C'\bar{\psi}$ and by application of Assumptions D and F and (2.19) of Lemma 2.2 with different filtrations to the final four sums at the right-hand side. This completes the proof of the efficiency of $\hat{\nu}_n$. □

REMARK 3.1. Together with (3.2) of Assumption F, the additional assumption (3.6) in Theorem 3.1 may be viewed as the $\sqrt{n}$-unbiasedness condition in time series, since it implies

$$A'I_g(\theta)^{-1}W_0(\theta, g)\sqrt{n}\,E\big(\bar{\psi}(\varepsilon) \mid \varepsilon_1, \dots, \varepsilon_n\big) \to_P 0.$$

Just as with the $\sqrt{n}$-unbiasedness condition in i.i.d. models [see (7.8.19) and Theorem 7.8.2, pages 395–397 of BKRW], this latter expression arises from the efficient score $A'I_g(\theta)^{-1}n^{-1}\Sigma_{t=1}^{n}W_t\psi_t$. Observe that (3.2) and (3.6) establish a trade-off between the $\sqrt{n}$-unbiasedness condition for the (nonparametric) estimator $\bar{\psi}$ and the orthogonality relation (3.6). Together they will yield $\sqrt{n}$-unbiasedness for time series. To obtain this property, it is not necessary to have a $\sqrt{n}$-unbiased estimator of the complete nonparametric part ψ of the score $W\psi$. This observation will be particularly useful in time-series analysis when the error distributions are not symmetric or when the volatility of the process is time dependent; see Section 4.

Finally, note that the matrix C in (3.2) is not uniquely determined. Apart from trivial orthogonal transformations of C (which will have no effect on the estimator $\hat{\nu}_n$), it is possible that the estimator $\bar{\psi}$ is also $\sqrt{n}$-unbiased in directions for which the orthogonality property (3.6) holds true. Such directions may be either included or excluded from the linear span of C; the adaptiveness of $\hat{\nu}_n$ will not be affected. In practical situations one will choose C such that joint verification of (3.2) and (3.6) is rather easy; see also Example 4.2.

REMARK 3.2. Theorem 3.1 shows that the $\sqrt{n}$-unbiasedness condition (3.6) is a sufficient condition for adaptiveness of ν with respect to the infinite-dimensional parameter g. This condition is also necessary in the following sense. Adaptiveness of the parameter ν requires that the projection of the tangent space with respect to g onto the tangent space with respect to θ is contained in the tangent space with respect to η [this is a multivariate generalization of the necessary condition (2.4) for adaptation in DKW]; note that this has to hold at all θ and g. Let the fixed orthogonal matrix C from Assumption F have maximal rank (possibly $C = 0$) such that $C'\psi$ is orthogonal to the tangent space implied by g, again at all g, suggesting that $\sqrt{n}$-unbiased estimation of $C'\psi$ is possible. If moreover at all g, no linear combination of the components of $(I - CC')\psi$ is orthogonal to the tangent space generated by g, then one may verify (along the lines of Theorem 3.1 of DKW) that the multivariate version of (2.4) of DKW is equivalent to (3.6) above.

REMARK 3.3. To improve on the initial estimator $\tilde{\theta}_n$ by a one-step Newton–Raphson method, the efficient influence function for ν, $A'I_g(\theta_n)^{-1}n^{-1}\sum_{t=1}^{n}W_t\psi_t$, is estimated by

$$A'\hat{I}_n^{-1}n^{-1}\Bigg\{\sum_{t=1}^{\alpha_n}\tilde{W}_t\hat{\psi}_{2t} + \sum_{t=\alpha_n+1}^{n}\tilde{W}_t\hat{\psi}_{1t} - \frac{1}{\alpha_n}\sum_{s=1}^{\alpha_n}\tilde{W}_s(I - CC')\sum_{t=1}^{\alpha_n}\hat{\psi}_{2t} - \frac{1}{n-\alpha_n}\sum_{s=\alpha_n+1}^{n}\tilde{W}_s(I - CC')\sum_{t=\alpha_n+1}^{n}\hat{\psi}_{1t}\Bigg\}.$$

At first sight the averages $(1/\alpha_n)\sum_{s=1}^{\alpha_N}\tilde{W}_s$ and $1/(n-\alpha_n)\sum_{s=\alpha_n+1}^{n}\tilde{W}_s$ appearing in this estimated influence function are surprising. However, writing the influence function for ν as $A'I_g(\theta_n)^{-1}n^{-1}\sum_{t=1}^{n}\{W_t - W_0(\theta_n, g)(I - CC')\}\psi_t$ the structure of $\hat{\nu}_n$ becomes more transparent. Furthermore, the proof of Theorem 3.1 shows that the bias terms, such as $\tilde{E}(\hat{\psi}_{2t} \mid \tilde{\varepsilon}_{\alpha_n+1}, \dots, \tilde{\varepsilon}_n)$, cancel due to the inclusion of the averages. The presence of these averages is essential as may be seen by the following argument. If these averages are excluded from our estimator $\hat{\nu}_n$, we are confronted with two additional remainder terms:

$$A'\hat{I}_n^{-1} n^{-1/2} \frac{1}{\alpha_n} \sum_{s=1}^{\alpha_n} 1 \tilde{W}_s (I - CC') \sum_{t=1}^{\alpha_n} \hat{\psi}_{2t}$$
$$= A' \left\{ \left(\hat{I}_n^{-1} - I_g(\theta_0)^{-1} \right) n^{-1/2} \sum_{s=1}^{\alpha_n} \tilde{W}_s \right.$$
$$\left. + I_g(\theta_0)^{-1} n^{-1/2} \sum_{s=1}^{\alpha_n} \left(\tilde{W}_s - W_0 \right) \right\} (I - CC') \tilde{E}\left(\hat{\psi}_{21} \mid \tilde{\varepsilon}_{\alpha_n+1}, \dots, \tilde{\varepsilon}_n \right)$$
$$+ o_P(1),$$

where the equality follows from Lemma 2.2. These remainder terms should be of order $o_P(1)$ for $\hat{\nu}_n$ to be an efficient estimator. One readily verifies that the remainder term is in fact the product of a term of order $o_P(n^{1/2})$ and a bias term which is of order $o_P(1)$. Examples show that the $o_P(n^{1/2})$ rate cannot be improved, in general. Hence, if the averaged $\tilde{W}_s$ are excluded from our estimator, it would be necessary to require the additional condition $\sqrt{n}\, E(\overline{\psi}(\varepsilon_1, \dots, \varepsilon_n)(\varepsilon) \mid \varepsilon_1, \dots, \varepsilon_n) = O_P(1)$. This condition is close to requiring (3.2) with $C = I_q$. In time-series models this condition will often be violated while orthogonality relations like (3.6) are available. We conclude that deletion of the average terms distorts the adaptiveness of $\hat{\nu}_n$.

Even if the limit W_0 of averaged $\tilde{W}_s$ is known, for example, $W_0 = 0$, replacement of the averages by the limit W_0 itself will usually destroy the efficiency of $\hat{\nu}_n$. In that case one of the two remainder terms is given by

$$A'\hat{I}_n^{-1} n^{1/2} \frac{1}{\alpha_n} \sum_{s=1}^{\alpha_n} \left(\tilde{W}_s - W_0 \right) (I - CC') \sum_{t=1}^{\alpha_n} \hat{\psi}_{2t}$$
$$= A'\hat{I}_n^{-1} n^{-1/2} \sum_{s=1}^{\alpha_n} \left(\tilde{W}_s - W_0 \right) (I - CC') \tilde{E}\left(\hat{\psi}_{21} \mid \tilde{\varepsilon}_{\alpha_n+1}, \dots, \tilde{\varepsilon}_n \right) + o_P(1),$$

which will not converge to zero, in general. In special circumstances, however, this term will converge to zero. This happens, for example, if the matrices W_t, $t = 1, \dots, n$, satisfy a central limit theorem, $n^{-1/2} \sum_{t=1}^{n} (W_t - W_0) = O_p(1)$, or if the $\sqrt{n}$-unbiasedness condition (3.2) is fulfilled with $C = I_q$. In such cases replacement of the estimated averages by W_0 itself does not harm, but it will not improve the first order asymptotic properties of $\hat{\nu}_n$ either.

REMARK 3.4. As noted in Schick (1987) [see also BKRW, Theorem 7.8.3, page 403], it is possible in i.i.d. models to avoid sample splitting as in (3.7), however, at the cost of stronger conditions. See Koul and Schick (1995) for results along these lines in the linear time-series models of type (4.1) with $\sigma_t(\theta) = 1$.

Next we will verify condition (3.6) in some specific situations where the classical $\sqrt{n}$-unbiasedness of $\bar{\psi}$ is not needed ($C = 0_q$). The case of group models in time series, as introduced at the end of Section 2, will be a particular example; see Example 3.1 and also the discussion of various known models in Section 4.

Corollary 3.1. *Assume* A–H *for all sequences* $\theta_n = \theta_0 + O(n^{-1/2})$ *and* $\tilde{\theta}_n = \theta_0 + O(n^{-1/2})$. *Let* $A \in \mathbb{R}^{p\times r}$ *and* $B \in \mathbb{R}^{p\times q}$ *be matrices satisfying* $A'B = 0$, *with r arbitrary and* $q \le p$. *If* $B'W_t(\theta)$ *is an invertible matrix not depending on t for all* $\theta \in \Theta$, *then* $\nu = A'\theta$ *is adaptively estimable with respect to g.*

Proof. Without loss of generality we may assume that A and B are orthogonal matrices satisfying $(A, B)^{-1} = (A, B)'$. Note that Assumption H implies $B'W_t = B'W_0$, $t = 1, \dots, n$. Hence, $B'W_0$ is also invertible and Assumptions H and D yield

$$\begin{aligned} A'W_0 &= A'W_0 JW_0'B\{B'W_0 JW_0'B\}^{-1} B'W_0 \\ &\leftarrow_P A'n^{-1}\sum_{t=1}^{n} W_t JW_t'B\left\{B'n^{-1}\sum_{t=1}^{n} W_t JW_t'B\right\}^{-1} B'W_0 \\ &\rightarrow_P A'I_g(\theta_0)B\{B'I_g(\theta_0)B\}^{-1}B'W_0, \end{aligned}$$

under $P_{\theta_n g}$. Contiguity gives

$$A'W_0 = A'I_g(\theta_0)B\{B'I_g(\theta_0)B\}^{-1}B'W_0 \quad P_{\theta_0 g}\text{-a.s.} \tag{3.10}$$

Using the formula for the inverse of a partitioned matrix, we obtain condition (3.6) of Theorem 3.1:

$$\begin{aligned} &A'I_g(\theta_0)^{-1}W_0 \\ &\quad = (I_r, 0_{r\times(p-r)})\{(A, B)'I_g(\theta_0)(A, B)\}^{-1}(A, B)'W_0 \\ &\quad = \left\{A'I_g(\theta_0)A - A'I_g(\theta_0)B\{B'I_g(\theta_0)B\}^{-1}B'I_g(\theta_0)A\right\}^{-1} \qquad (3.11) \\ &\qquad \times\left\{A'W_0 - A'I_g(\theta_0)B\{B'I_g(\theta_0)B\}^{-1}B'W_0\right\} \\ &\quad = 0 \quad P_{\theta_0 g}\text{-a.s.} \end{aligned}$$

To complete the proof of the corollary, take $C = 0_q$ in Theorem 3.1. □

COROLLARY 3.2. *Assume* A–H *for all sequences* $\theta_n = \theta_0 + O(n^{-1/2})$ *and* $\tilde{\theta}_n = \theta_0 + O(n^{-1/2})$.

A. *Let* A *and* B *be matrices with* p *rows. Suppose that* (A, B) *has rank* p *and that* $A'B = 0$. *If*

$$\forall\, \theta \in \Theta,\ g \in \mathscr{G}\colon A' I_g(\theta) B = 0 \quad \text{and} \quad A' W_0(\theta, g) = 0 \quad P_{\theta g}\text{-}a.s., \tag{3.12}$$

then $\nu = A'\theta$ *is adaptively estimable with respect to both* $\eta = B'\theta$ *and* g.

B. *If* $W_0(\theta, g) = 0$, $P_{\theta g}$-*a.s., for all* $\theta \in \Theta$, $g \in \mathscr{G}$, *then* θ *is adaptively estimable with respect to* g.

PROOF. Formula (3.10) is trivially implied by (3.12) and, hence, (3.11) yields adaptiveness of ν with respect to g. The block-diagonality of $(A, B)' I_g(\theta)^{-1}(A, B)$ implies also the adaptiveness with respect to η. □

These corollaries are particularly useful in group models in time series. In that case the matrix (A, B) is the identity matrix and the adaptivity results apply to the first component of θ. This is summarized in the following example.

EXAMPLE 3.1. Consider the typical time-series model discussed at the end of Section 2, where $\theta = (\nu', \eta')' \in \mathbb{R}^{(p-q)+q}$ and $W = (W_1', W_2')' \in \mathbb{R}^{((p-q)+q)\times q}$ are partitioned and where W_2 is a square nonsingular matrix not depending on t. Then ν is adaptively estimable with respect to g (Corollary 3.1).

Assume in addition also $n^{-1}\sum_{t=1}^{n} W_{1t} \to_P 0$. Then ν is adaptively estimable with respect to both η and g (Corollary 3.2.A).

4. Examples. In this section we will apply the results of Sections 2 and 3 to several well-known classes of time-series models. These examples include many of the most important known results on semiparametric time-series analysis and also some new (to the best of our knowledge) results. Since location-scale models with completely unknown error distribution are often used in econometric time-series applications, we will discuss these models rather extensively. Subsequently we will discuss elliptic models shortly.

An important special case of our general theory is the class of location-scale models. This class contains regression models, ARMA models and GARCH-type models. We maintain the notation of Section 2 and consider applications fitting into the following framework:

$$Y_t = \mu_t(\theta) + \sigma_t(\theta)\varepsilon_t, \tag{4.1}$$

where the time dependent location-scale parameter $(\mu_t(\theta), \sigma_t(\theta))' \in \mathscr{F}_{t-1}^n$, $t = 1, \dots, n$, is supposed to depend on θ, the observed starting value $X_n = X(X_n^*, \theta) \in \mathscr{F}_0^n$ and the first $t-1$ observations, $Y_1, \dots, Y_{t-1}$. The ε_t are i.i.d. with completely unknown distribution (apart from some regularity, to be discussed later). The time dependent location-scale parameter is supposed to be differentiable with respect to θ. In order to show how W_{nt} as given in (2.1) is obtained in this location-scale context, write $\tilde{\theta}_n^k = (\tilde{\theta}_{n1}, \dots, \tilde{\theta}_{nk}, \theta_{n,k+1}, \dots, \theta_{np})'$, $k = 0, \dots, p$, and define the $p \times 2$-matrix W_{nt} by its rows, $k = 1, \dots, p$,

$$(4.2)\quad W_{ntk} = \begin{cases} \sigma_t(\theta_n)^{-1}\big(\tilde{\theta}_{nk} - \theta_{nk}\big)^{-1} \\ \quad \times\Big(\mu_t\big(\tilde{\theta}_n^k\big) - \mu_t\big(\tilde{\theta}_n^{k-1}\big), \sigma_t\big(\tilde{\theta}_n^k\big) - \sigma_t\big(\tilde{\theta}_n^{k-1}\big)\Big), & \text{if } \tilde{\theta}_{nk} \neq \theta_{nk}, \\ \sigma_t(\theta_n)^{-1} \dfrac{\partial}{\partial \theta_k}\big(\mu_t(\theta), \sigma_t(\theta)\big)\big|_{\theta = \tilde{\theta}_n^k}, & \text{if } \tilde{\theta}_{nk} = \theta_{nk}. \end{cases}$$

Note that we now indeed have

$$W_{nt}'\big(\tilde{\theta}_n - \theta_n\big) = \sigma_t(\theta_n)^{-1}\Big(\mu_t\big(\tilde{\theta}_n\big) - \mu_t(\theta_n), \sigma_t\big(\tilde{\theta}_n\big) - \sigma_t(\theta_n)\Big)'.$$

With $\zeta = (\mu, \sigma)'$ a location-scale parameter and $\zeta_0 = (0, 1)'$ the log-likelihood ratio statistic of the observations, $X_n, Y_1, \dots, Y_n$, for $\tilde{\theta}_n$ with respect to θ_n satisfies (2.1),

$$(4.3)\quad \Lambda_n = \sum_{t=1}^{n} \Big\{ l\Big(\zeta_0 + W_{nt}'\big(\tilde{\theta}_n - \theta_n\big)\Big)\big(\varepsilon_t(\theta_n)\big) - l(\zeta_0)\big(\varepsilon_t(\theta_n)\big) \Big\} + \Lambda_n^s,$$

where

$$l(\zeta)(\cdot) = \log \sigma^{-1} g((\cdot - \mu)/\sigma),$$

$$\Lambda_n^s = \log \frac{dP_{\tilde{\theta}_n g}(X_n)}{dP_{\theta_n g}(X_n)} = \log \frac{dP_{\theta_0 g}\big(X\big(X_n^*, \tilde{\theta}_n\big)\big)}{dP_{\theta_0 g}\big(X(X_n^*, \theta_n)\big)}.$$

It is clear that an appropriate choice for W_t as required in Assumption D is given by

$$(4.4)\qquad W_t = \sigma_t(\theta_n)^{-1} \frac{\partial}{\partial \theta'}\big(\mu_t(\theta), \sigma_t(\theta)\big)\big|_{\theta = \theta_n}.$$

Note that W_{nt} and W_t are both measurable with respect to $\mathscr{F}_{t-1}^n$.

To be able to apply the results of Sections 2 and 3 we need to verify Assumptions A–H. Throughout we assume that the starting conditions have a negligible effect; for example, the distribution of X_n is assumed to be independent of θ or to vary smoothly with respect to θ. More precisely stated, we assume the following.

A′. $\Lambda_n^s \to_P 0$, under $P_{\theta_n g}$.

Moreover, we suppose that the family of $\mathscr{G}$ of densities g (with respect to Lebesgue measure) is such that location-scale models generated by elements of this family are regular; the following is a more precise formulation.

B′. The density g is absolutely continuous with derivative g' such that the Fisher information for location $I_l(g) = \int (g'/g)^2 g(x)\, dx$ and the Fisher information for scale $I_s(g) = \int (1 + xg'/g(x))^2 g(x)\, dx$ are finite.

By the technique of pages 211–212 and page 214 of Hájek and Šidák (1967), it can be proved that finiteness of $I_l(g)$ yields regularity of the location model and finiteness of $I_s(g)$ regularity of the scale model, respectively. Similarly it can be shown that finiteness of both $I_l(g)$ and $I_s(g)$, implies regularity of the location-scale model. The corresponding score is given by

$$\psi(x) = -(g'/g(x), 1 + xg'/g(x))'.$$

The definition of W_{nt} and Conditions A′ and B′ yield Assumptions A–C; E is merely notation and F is fulfilled by B′ and Proposition 7.8.1, page 400 of BKRW. So we have to construct an initial $\sqrt{n}$-consistent estimator as required in Assumption G and it suffices to verify (2.3), (2.4) and (3.3) with $C = 0_q$ under all sequences $P_{\theta_n g}$. As is shown in the Appendix, these convergences can be obtained by proving the set of sufficient convergence relations (2.3′), (2.4′), (3.3′) and (A.1) under $P_{\theta_0 g}$. Usually these latter relations will be obtained from smoothness and an ergodicity argument.

One easily verifies that the arguments above also apply to either pure location models or pure scale models. In such situations the conditions may be relaxed somewhat. For example, in pure location models, define W_{nt} by deleting the second component at the right-hand side of (4.2), define W_t similarly, let ζ be a location parameter, write (4.3) around $\zeta_0 = 0$, and require that $I_l(g)$ is finite (scale is unimportant and may be absorbed into the infinite dimensional nuisance parameter). The pure scale situation can be treated analogously.

Finally, we will verify the additional requirement (3.6). As said before, for general group models in time series, the parameter θ can usually be split up into two components, say $\theta = (\nu', \eta')'$, where the second component η has the same interpretation as ζ. In the location-scale context considered here, this means that $\eta = (\mu, \sigma)'$. Generally μ (respectively σ) will serve as an additive (multiplicative) parameter in the time dependent location (scale) parameter while this component does not appear at other places; that is, $\mu_t(\theta) = \mu + \mu_t^*(\nu)$ $[\sigma_t(\theta) = \sigma\sigma_t^*(\nu)]$. In such cases the last two components of (4.4) do not depend on t and one obtains (3.6) along the lines of Example 3.1.

We will apply our results to some specific classes of models satisfying (4.1). To be precise, we will consider the linear regression model, with possibly dependent regressor sequences, and the ARMA, TAR and ARCH model. The regression model has been studied in an i.i.d. context in Bickel (1982). Semiparametric analysis of the ARMA model is discussed in Kreiss (1987a, 1987b). Semiparametric ARCH type models have been discussed by Engle and Gonzalez-Rivera (1991), Linton (1993) and Drost and Klaassen (1997). All these models are easily analyzed using our general approach based on time-series group models. Adaptive estimation in semiparametric TAR models has, to the best of our knowledge, not been studied before. The regression, ARMA and TAR examples are pure location models while the ARCH model is a pure scale model. By the arguments above it suffices to verify the set of alternative conditions (2.3′), (2.4′), (3.3′) and (A.1) given in the Appendix and to construct an initial $\sqrt{n}$-consistent estimator.

EXAMPLE 4.1 (Linear regression). Let X be a function generating square-integrable regressors $X_n = (Z_1', \dots, Z_n')' = X(X_n^*, \theta)$ and consider the linear regression model

$$Y_t = \mu + Z_t'\beta + \varepsilon_t. \tag{4.5}$$

Assume A′, B′ (location version, i.e., $I_l(g) < \infty$) and

$$n^{-1} \sum_{t=1}^{n} (Z_t', 1)' I_l(g)(Z_t', 1) \to_P I_g(\theta_0),$$

$$\forall\, \delta > 0\colon n^{-1} \sum_{t=1}^{n} |Z_t|^2 1_{\{n^{-1/2}|Z_t| > \delta\}} \to_P 0,$$

where the convergences are under $P_{\theta_0 g}$. The conditions include quite general, possibly heavily dependent regressor sequences. Of course they are satisfied in the classical situation, where regressors are i.i.d. If the regressors are purely exogenous, then $\Lambda_n^s = 0$. This model fits into the general framework of Section 2 with $\theta = (\beta', \mu)'$. Moreover, it is a specialization of (4.1) with $\mu_t(\theta) = \mu + Z_t'\beta$ and $\sigma_t(\theta) = 1$.

The assumptions stated in the Appendix will be verified with $W_{nt} = W_t = (Z_t', 1)'$ and W_0 the last column of $I_g(\theta_0)/I_l(g)$. The relations (2.3′) and (3.3′) are immediate from the corresponding ones about Z_t. The smoothness conditions (A.1) and (2.4′) are trivially fulfilled. In addition, assume $E_g\varepsilon = 0$ and $E_g\varepsilon^2 < \infty$. Then the conditions about Z_t ensure that $n^{-1/2}\sum_{t=1}^{n}(Z_t', 1)'\varepsilon_t$ converges in distribution and the ordinary least squares estimator is a $\sqrt{n}$-consistent estimator of θ. Discretization yields an estimator satisfying Assumption G. From Example 3.1 we conclude that β is adaptively estimable with respect to g. If the average of the regressors tends to zero, β will be also adaptively estimable with respect to μ. Observe that we do not require that the densities g be symmetric. Therefore, there is loss of information for the whole parameter θ due to the nonadaptive location parameter μ.

If it is known at the outset that the densities are symmetric, then there exist estimators of ψ such that the bias is zero. With $C = 1$ we may conclude from Theorem 3.1 that θ can be estimated fully adaptively. Of course, the estimation problem of μ in the nonsymmetric case reduces to the i.i.d. location problem with a completely unknown distribution.

The semiparametric linear regression model has been considered previously by Bickel (1982), Examples 2 and 3. Our conditions allow more general regressor sequences. In the symmetric case Bickel [(1982), Example 2] does not need the additional assumption $E_g\varepsilon^2 < \infty$ to obtain an initial estimator. Therefore, we can also dispense with the condition of finite second moments here.

EXAMPLE 4.2 (ARMA). Let X be a function generating square-integrable starting values $X_n = (Y_0, \dots, Y_{1-p}, \varepsilon_0, \dots, \varepsilon_{1-q})' = X(X_n^*, \theta)$ and consider the ARMA (p, q) process defined by

$$Y_t = \rho_1 Y_{t-1} + \cdots + \rho_p Y_{t-p} + \varepsilon_t + \phi_1\varepsilon_{t-1} + \cdots + \phi_q\varepsilon_{t-q}. \tag{4.6}$$

As usual, we assume that the orders of the process are known. Assume $E_g \varepsilon = 0$, $E_g \varepsilon^2 < \infty$, A′, B′ (location version) and suppose that the parameters are such that (4.6) permits a causal, invertible solution and that the AR and MA polynomial have no common roots. Write

$$V_t(\theta) = \big(Y_{t-1}, \dots, Y_{t-p}, \varepsilon_{t-1}(\theta), \dots, \varepsilon_{t-q}(\theta)\big)'.$$

Then it is clear that the ARMA model fits into the general framework with $\theta = (\rho', \phi')' = (\rho_1, \dots, \rho_p, \phi_1, \dots, \phi_q)'$, $\mu_t(\theta) = V_t(\theta)'\theta$, $\sigma_t(\theta) = 1$ and

$$W_t(\theta) = V_t(\theta) + D_t(\theta)'\theta,$$

where $D_t(\theta)$ denotes the derivative of $V_t(\theta)$ with respect to θ. For ζ we may take a location parameter and the log-likelihood is written around $\zeta_0 = 0$.

To verify the conditions, define $A(\theta), B(\theta) \in \mathbb{R}^{(p+q)\times(p+q)}$ by

$$A(\theta) = \begin{pmatrix} 0'_{p+q} \\ I_{p-1}, 0_{(p-1)\times(q+1)} \\ -\theta' \\ 0_{(q-1)\times p}, I_{q-1}, 0_{q-1} \end{pmatrix}, \qquad B(\theta) = \begin{pmatrix} \theta' \\ I_{p-1}, 0_{(p-1)\times(q+1)} \\ 0'_{p+q} \\ 0_{(q-1)\times p}, I_{q-1}, 0_{q-1} \end{pmatrix},$$

and observe that

$$\begin{aligned} V_t(\theta) &= A(\theta)V_{t-1}(\theta) + \big(Y_{t-1}, 0'_{p-1}, Y_{t-1}, 0'_{q-1}\big)' \\ &= B(\theta)V_{t-1}(\theta) + \big(\varepsilon_{t-1}(\theta), 0'_{p-1}, \varepsilon_{t-1}(\theta), 0'_{q-1}\big)', \\ D_t(\theta) &= A(\theta)D_{t-1}(\theta) - \big(0_{(p+q)\times p}, V_{t-1}(\theta), 0_{(p+q)\times(q-1)}\big)'. \end{aligned}$$

The invertibility condition on the MA polynomial guarantees that $A(\theta)$ has all eigenvalues within the unit circle. Similarly, the eigenvalues of $B(\theta)$ are within the unit circle by the causality condition on the AR polynomial. This ensures an ergodic solution of $V_t(\theta)$ and $D_t(\theta)$ and, hence, (2.3′) and (3.3′) are fulfilled. Since $W_0 = 0$ this also verifies (3.6) of Theorem 3.1. To verify (2.4) and (A.1), recall that the calculated residual $\varepsilon_t(\theta_n)$ is equal to the true innovation ε_t under $P_{\theta_n g}$ and observe that the following relations hold true:

$$\begin{aligned} V_t(\tilde\theta_n) - V_t(\theta_n) &= A(\tilde\theta_n)\big(V_{t-1}(\tilde\theta_n) - V_{t-1}(\theta_n)\big) \\ &\quad - \big(0'_p, (\tilde\theta_n - \theta_n)'V_{t-1}(\theta_n), 0'_{q-1}\big)', \\ D_t(\tilde\theta_n) - D_t(\theta_n) &= A(\tilde\theta_n)\big(D_{t-1}(\tilde\theta_n) - D_{t-1}(\theta_n)\big) \\ &\quad - \big(0_{(p+q)\times p}, V_{t-1}(\tilde\theta_n) - V_{t-1}(\theta_n) \\ &\qquad + D_{t-1}(\theta_n)'(\tilde\theta_n - \theta_n), 0_{(p+q)\times(q-1)}\big)', \end{aligned}$$

$$\begin{aligned} &V_t(\tilde\theta_n) - V_t(\theta_n) - D_t(\theta_n)(\tilde\theta_n - \theta_n) \\ &\quad = A(\theta_n)\big(V_{t-1}(\tilde\theta_n) - V_{t-1}(\theta_n) - D_{t-1}(\theta_n)(\tilde\theta_n - \theta_n)\big) \\ &\qquad - \big(0'_p, (\tilde\theta_n - \theta_n)'\big(V_{t-1}(\tilde\theta_n) - V_{t-1}(\theta_n)\big), 0'_{q-1}\big)'. \end{aligned}$$

Hence, using Lemma A.2, continuity of $A(\theta)$ and $B(\theta)$ and $\sqrt{n}(\tilde{\theta}_n - \theta_n) = O(1)$, we obtain subsequently, under $P_{\theta_n g}$,

$$n^{-1} \sum_{t=1}^{n} |V_t(\theta_n)|^2 = O_P(1), \qquad n^{-1} \sum_{t=1}^{n} |D_t(\theta_n)|^2 = O_P(1),$$

$$\sum_{t=1}^{n} |V_t(\tilde{\theta}_n) - V_t(\theta_n)|^2 = O_P(1), \qquad \sum_{t=1}^{n} |D_t(\tilde{\theta}_n) - D_t(\theta_n)|^2 = O_P(1)$$

and

$$\sum_{t=1}^{n} |V_t(\tilde{\theta}_n) - V_t(\theta_n) - D_t(\theta_n)(\tilde{\theta}_n - \theta_n)|^2 = o_P(1).$$

Equations (2.4) and (A.1) are immediate from these results and from

$$\begin{aligned}
&\sum_{t=1}^{n} \left|(W_{nt} - W_t(\theta_n))'(\tilde{\theta}_n - \theta_n)\right|^2 \\
&\quad = \sum_{t=1}^{n} \Big|\big(V_t(\tilde{\theta}_n) - V_t(\theta_n) - D_t(\theta_n)(\tilde{\theta}_n - \theta_n)\big)'\tilde{\theta}_n \\
&\qquad\qquad + (\tilde{\theta}_n - \theta_n)' D_t(\theta_n)(\tilde{\theta}_n - \theta_n)\Big|^2, \\
&n^{-1} \sum_{t=1}^{n} |W_t(\theta_n) - W_t(\theta_0)|^2 \\
&\quad = n^{-1} \sum_{t=1}^{n} \big|V_t(\theta_n) - V_t(\theta_0) \\
&\qquad\qquad + (D_t(\theta_n) - D_t(\theta_0))'\theta_n + D_t(\theta_0)'(\theta_n - \theta_0)\big|^2.
\end{aligned}$$

An initial $\sqrt{n}$-consistent estimator of θ is easily obtained from the first $p + q + 1$ autocovariances. From Lemma A.1 we conclude that θ is adaptively estimable with respect to g.

In the proof above, symmetry of the densities is not necessary. If symmetry is given in advance we may obtain the result in another way by taking an appropriate symmetrized estimator of ψ with zero bias. Then knowledge of $W_0 = 0$ is not necessary to verify (3.6). Moreover, this implies that replacement of $C = 0$ by $C = 1$ does not influence the adaptiveness of the estimator $\hat{\nu}_n$. Note, however, that the use of $\sqrt{n}$-unbiased estimates of ψ does not improve the (first order) asymptotics. To guard against possible nonsymmetries it seems better to use nonsymmetrized estimators of ψ and $C = 0$.

The semiparametric ARMA has been considered previously by Kriess (1987a) (symmetric error distribution) and Kreiss (1987b) (nonsymmetric AR case). Our conditions are somewhat more general and do not need the strict positiveness of g on $\mathbb{R}$.

EXAMPLE 4.3 (TAR). Let $X_n = Y_0 = X(X_n^*, \theta)$ denote some starting value and consider the threshold autoregressive model

$$Y_t = \sum_{j=1}^{k} (\mu_j + \rho_j Y_{t-1}) I_{A_j}(Y_{t-1}) + \varepsilon_t, \tag{4.7}$$

where $A_1, \dots, A_k$ is a measurable partition of $\mathbb{R}$. See, for example, Tong and Lim (1980) for a discussion of the parametric model. Assume $E_g \varepsilon = 0$, $E_g \varepsilon^2 < \infty$, A′, B′ (location version) and suppose that (4.7) admits an ergodic, stationary solution with finite second moments [cf. Chan, Petruccelli, Tong and Woolford (1985) for a set of sufficient conditions]. The TAR model is a special case of (4.1) with $\theta = (\rho', \mu')' = (\rho_1, \dots, \rho_k, \mu_1, \dots, \mu_k)'$,

$$\mu_t(\theta) = \sum_{j=1}^{k} (\mu_j + \rho_j Y_{t-1}) I_{A_j}(Y_{t-1}),$$

$\sigma_t(\theta) = 1$ and $W_{nt} = W_t = (I_{A_1}(Y_{t-1})Y_{t-1}, \dots, I_{A_k}(Y_{t-1})Y_{t-1}, I_{A_1}(Y_{t-1}), \dots, I_{A_k}(Y_{t-1}))'$.

The assumptions stated in the Appendix are immediate. The smoothness conditions are trivial since $W_{nt} = W_t$ and (2.3′) and (3.3′) follow from ergodicity. An initial estimator is derived in Chan, Petruccelli, Tong and Woolford (1985). Put $\bar{\mu} = k^{-1}\sum_{j=1}^{k} \mu_j$ and obtain from Corollary 3.1 [with $A = I_{2k} - k^{-1}(0'_k, 1'_k)'(0'_k, 1'_k)$ and $B = (0'_k, 1'_k)'$] that $(\rho', \mu' - \bar{\mu}1'_k)'$ is adaptively estimable with respect to g.

Just as in the regression example, adaptive estimation of the whole parameter is possible if the densities are known to be symmetric. In that case one needs to use a symmetrized estimator of ψ with zero bias. In the nonsymmetric case, the estimation problem of the remaining nonadaptive parameter is equivalent to the i.i.d. location problem with a completely unknown distribution.

EXAMPLE 4.4 (ARCH). Let $X_n = (Y_0, \dots, Y_{1-p})' = X(X_n^*, \theta)$ denote some starting values and consider Engle's (1982) autoregressive conditional heteroskedastic model

$$\begin{aligned} Y_t &= \sigma_t(\theta)\varepsilon_t, \\ \sigma_t^2(\theta) &= \sigma^2\big(1 + \alpha_1 Y_{t-1}^2 + \cdots + \alpha_p Y_{t-p}^2\big). \end{aligned} \tag{4.8}$$

Assume $\sigma, \alpha_1, \dots, \alpha_p > 0$, $E_g \varepsilon = 0$, $E_g \varepsilon^2 = 1$, A′ and B′ (scale version, i.e., $I_s(g) < \infty$). Suppose also that (4.8) admits a stationary solution, that is, assume the necessary and sufficient condition $E_g \ln \varepsilon^2 + \ln \sigma^2 + \ln \max_{j=1,\dots,p} \alpha_j < 0$ [see Nelson (1990)]. Under this assumption, the solution will also be ergodic. The ARCH model is a special case of (4.1) with $\theta = (\alpha', \sigma)' = (\alpha_1, \dots, \alpha_p, \sigma)'$, $\mu_t(\theta) = 0$ and

$$W_t = \left(\frac{\sigma_n^2 Y_{t-1}^2}{2\sigma_t^2(\theta_n)}, \dots, \frac{\sigma_n^2 Y_{t-p}^2}{2\sigma_t^2(\theta_n)}, \frac{1}{\sigma_n} \right)'.$$

To be able to apply Lemma A.1 we will verify the set of sufficient equations (2.3′), (2.4), (3.3′) and (A.1). Since $|W_t|^2$ is bounded, (2.3′) and (3.3′) are direct consequences of the ergodic theorem. The smoothness conditions (2.4) and (A.1) follow from

$$\begin{aligned}
&\sum_{t=1}^{n} \left|(W_{nt} - W_t)'\left(\tilde{\theta}_n - \theta_n\right)\right|^2 \\
&\quad = \sum_{t=1}^{n} \left| \sigma_t(\theta_n)^{-1}\left\{\sigma_t\left(\tilde{\theta}_n\right) - \sigma_t(\theta_n) - \frac{\partial}{\partial\theta}\sigma_t(\theta)|_{\theta=\theta_n}\left(\tilde{\theta}_n - \theta_n\right)\right\}\right|^2 \\
&\quad = \sum_{t=1}^{n} \left| \frac{\sigma_t^2\left(\tilde{\theta}_n\right)/\tilde{\sigma}_n^2 - \sigma_t^2(\theta_n)/\sigma_n^2}{2\sigma_t^2(\theta_n)/\sigma_n^2} \right. \\
&\qquad\qquad \left. \times \left\{ \left(\frac{\tilde{\sigma}_n}{\sigma_n} - 1\right) \frac{2\sigma_t(\theta_n)/\sigma_n}{\sigma_t\left(\tilde{\theta}_n\right)/\tilde{\sigma}_n + \sigma_t(\theta_n)/\sigma_n} - \frac{\sigma_t\left(\tilde{\theta}_n\right)/\tilde{\sigma}_n - \sigma_t(\theta_n)/\sigma_n}{\sigma_t\left(\tilde{\theta}_n\right)/\tilde{\sigma}_n + \sigma_t(\theta_n)/\sigma_n} \right\} \right|^2 \\
&\quad \le n\left(\sum_{j=1}^{p} \frac{|\tilde{\alpha}_{nj} - \alpha_{nj}|}{2\alpha_{nj}}\right)^2 \left(2\left|\frac{\tilde{\sigma}_n}{\sigma_n} - 1\right| + \sum_{j=1}^{p}\frac{|\tilde{\alpha}_{nj} - \alpha_{nj}|}{\tilde{\alpha}_{nj} + \alpha_{nj}}\right)^2 \to 0
\end{aligned}$$

and

$$\begin{aligned}
&n^{-1}\sum_{t=1}^{n} |W_t(\theta_n) - W_t(\theta_0)|^2 \\
&\quad = \left(\sigma_n^{-1} - \sigma_0^{-1}\right)^2 + n^{-1}\sum_{t=1}^{n}\sum_{j=1}^{p}\frac{1}{4}Y_{t-j}^4\left(\frac{\sigma_n^2}{\sigma_t^2(\theta_n)} - \frac{\sigma_0^2}{\sigma_t^2(\theta_0)}\right)^2 \\
&\quad \le \left(\sigma_n^{-1} - \sigma_0^{-1}\right)^2 + \left(\sum_{j=1}^{p}\frac{1}{4\alpha_{0j}^2}\right)\left(\sum_{j=1}^{p}\frac{|\alpha_{nj} - \alpha_{0j}|}{\alpha_{nj}}\right)^2 \to 0.
\end{aligned}$$

An initial estimator is derived in Weiss (1986). From Example 3.1 we conclude that α is adaptively estimable with respect to g. In this example it is not possible to obtain a $\sqrt{n}$-unbiased estimator of ψ. In this model, the additional parameter σ yields the required orthogonality property (3.6) to obtain adaptation. Given the estimation results for α, the estimation problem for σ reduces to estimation of scale in i.i.d. models.

Previous results for the semiparametric ARCH model are obtained in Engle and González-Rivera (1991) and Linton (1993). They obtained the same result in a subset of our parameter space. Their proofs need the existence of, for example, second moments of Y_t implying $\sigma^2(\alpha_1 + \cdots + \alpha_p) < 1$, which is clearly stronger than the condition leading to stationarity. Drost and Klaassen (1997) derive weak adaptiveness conditions for the more parsimonious generalized ARCH model of Bollerslev (1986).

As a further example of semiparametric time-series group models we will discuss elliptic models. For a discussion of these models in an i.i.d. context, see BKRW, Example 4.2.3, pages 89, 90, 96–99. Elliptic models are important in the analysis of financial markets since for these models mean-variance analysis is fully compatible with expected utility maximization.

EXAMPLE 4.5 (Elliptic model). The elliptic model is specified by

$$Y_t = \mu_t + S_t \varepsilon_t,$$

where the random d-vector μ_t and the random $d \times d$-matrix S_t are predictable and ε_t has some elliptic distribution, that is, a Lebesgue absolutely continuous distribution on $\mathbb{R}^d$ invariant under orthogonal transformations. In i.i.d. elliptic models adaptive estimation of both the location parameter μ and the parameter $\Sigma = SS'/\text{trace}(SS')$ is generally possible [see BKRW, Example 4.2.3, page 96]. Using Theorem 3.1 with C the identity matrix, this implies that adaptive estimation in a time-series context is possible as well, provided that the time dependent location and scale μ_t and $S_t S_t'$ satisfy conditions like those in previous examples. The details under which these conditions are satisfied are easily obtained by the reader. Comparing the elliptic model with general location-scale models, we see that on the one hand elliptic models allow for less general error distributions, but on the other hand adaptive estimation of most of the covariance structure becomes feasible.

APPENDIX

We show that the convergences (2.3), (2.4) and (3.3) which are needed under all sequences $P_{\theta_n g}$ may be replaced by the following sufficient set of equations:

$$(2.3') \qquad n^{-1} \sum_{t=1}^{n} W_t(\theta_0) J W_t(\theta_0)' \to_P I(\theta_0) > 0,$$
$$\forall \delta > 0\colon n^{-1} \sum_{t=1}^{n} |W_t(\theta_0)|^2 1_{\{n^{-1/2}|W_t(\theta_0)| > \delta\}} \to_P 0,$$

$$(2.4') \qquad \sum_{t=1}^{n} \left| \left(W_t(\theta_n, \tilde{\theta}_n) - W_t(\theta_n) \right)' \left(\tilde{\theta}_n - \theta_n \right) \right|^2 \to_P 0,$$

$$(3.3') \qquad n^{-1} \sum_{t=1}^{n} W_t(\theta_0) \to_P W_0,$$

$$(\text{A.1}) \qquad n^{-1} \sum_{t=1}^{n} |W_t(\theta_n) - W_t(\theta_0)|^2 \to_P 0,$$

where the convergences are under $P_{\theta_0 g}$. Since we consider convergence under a fixed probability measure, (2.3′) and (3.3′) will usually follow from a weak law of large numbers or from an ergodicity argument. Equation (2.4′) is a relaxation of (2.4) and relation (A.1) is an additional smoothness condition. The precise statement is given in the following lemma.

LEMMA A.1. *Assume that Assumptions* A–E *hold for all sequences* $\theta_n = \theta_0 + O(n^{-1/2})$, $\tilde{\theta}_n = \theta_0 + O(n^{-1/2})$ *but replace* (2.3) *by* (2.3′) *and* (A.1) *and/or replace* (2.4) *by* (2.4′). *Then Assumption* D *is satisfied and hence the conclusion of Theorem* 2.1 *remains valid*.

Additionally, assume F–H *but with* (3.3) *replaced by* (3.3′) *and* (A.1). *Then Assumption* H *is satisfied and hence Theorem* 3.1 *is still applicable*.

PROOF. Apply Theorem 2.1 with $\theta_n = \theta_0$, $\tilde{\theta}_n = \theta_n$ and establish contiguity of $P_{\theta_0 g}$ and $P_{\theta_n g}$. Now, by (A.1), we obtain (2.3) from (2.3′). Relation (2.4) is obtained in an analogous manner from contiguity and (2.4′). The final assertion follows similarly. □

We conclude the Appendix with a technical lemma which will be handy in time-series models with autoregressive parts.

LEMMA A.2. *Let* $\rho(A)$ *denote the spectral radius of a matrix* A. *Suppose that* X_{nt}, Y_{nt} *are arrays satisfying*

$$X_{nt} = A_n X_{n,t-1} + Y_{nt},\ t = 1, \dots, n, \tag{A.2}$$

where A_n *is a sequence of square matrices with* $\limsup_{n \to \infty} \rho(A_n) < \alpha < 1$. *Then, for* n *sufficiently large*,

$$\sum_{t=1}^{n} |X_{nt}|^2 \le (1-\alpha)^{-2} \sum_{t=1}^{n} |Y_{nt}|^2 + \alpha(1-\alpha)^{-1} |X_{n0}|^2. \tag{A.3}$$

PROOF. Fix $N > 0$ such that $n \ge N$ implies $\rho(A_n) \le \alpha$. Then we have, for $n \ge N$,

$$|X_{nt}| \le \alpha |X_{n,t-1}| + |Y_{nt}|$$

and hence

$$|X_{nt}|^2 \le \alpha |X_{n,t-1}|^2 + \frac{1}{1-\alpha} |Y_{nt}|^2.$$

Straightforward calculations complete the proof. □

Acknowledgments. We thank Peter Bickel and Benedikt Pötscher for some useful comments, and Editor John Rice and an Associate Editor for helpful suggestions about the presentation.

REFERENCES

BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press.

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327.

CHAN, K. S., PETRUCCELLI, J. D., TONG, H. and WOOLFORD, S. W. (1985). A multiple-threshold AR(1) model. *J. Appl. Probab.* **22** 267–279.

DROST, F. C. and KLAASSEN, C. A. J. (1997). Efficient estimation in semiparametric GARCH models. *J. Econometrics*. To appear.

DROST, F. C., KLAASSEN, C. A. J. and WERKER, B. J. M. (1994). Adaptiveness in time-series models. *Asymptotic Statistics* (P. Mandl and M. Hušková, eds.) 203–211. Physica, New York.

ENGLE, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50** 987–1008.

ENGLE, R. F. and GONZÁLEZ-RIVERA, G. (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics* **9** 345–359.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **14** 323–330.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academia, Prague.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11** 818–887.

KLAASSEN, C. A. J. (1987). Consistent estimation of the influence function of locally asymptotically linear estimates. *Ann. Statist.* **15** 1548–1562.

KOUL, H. L. and SCHICK, A. (1995). Efficient estimation in nonlinear time series models. Working paper, Michigan State Univ. and State Univ. New York, Binghamton.

KREISS, J. P. (1987a). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.

KREISS, J. P. (1987b). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5** 112–133.

LINTON, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory* **9** 539–569.

NELSON, D. B. (1990). Stationarity and persistence in the GARCH(1, 1) model. *Econometric Theory* **6** 318–334.

NEWEY, W. K. (1990). Semiparametric efficiency bounds. *J. Applied Econometrics* **5** 99–135.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

PÖTSCHER, B. M. (1995). Comment on "Adaptive estimation in time series regression models" by D. G. Steigerwald. *J. Econometrics* **66** 123–129.

ROBINSON, P. M. (1988). Semiparametric econometrics: a survey. *J. Applied Econometrics* **3** 35–51.

SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151.

SCHICK, A. (1987). A note on the construction of asymptotically linear estimates. *J. Statist. Plann. Inference* **16** 89–105.

STOKER, T. M. (1991). *Lectures on Semiparametric Econometrics*. CORE Foundation, Louvain la Neuve, Belgium.

STEIGERWALD, D. G. (1992). Adaptive estimation in time-series regression models. *J. Econometrics* **54** 251–275.

STEIGERWALD, D. G. (1995). Reply to B. M. Pötscher's comment on "Adaptive estimation in time series regression models." *J. Econometrics* **66** 131–132.

TONG, H. and LIM, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Statist. Soc. Ser. B* **42** 242–292.

WEFELMEYER, W. (1994). Improving maximum quasi-likelihood estimators. *Asymptotic Statistics* (P. Mandl and M. Hušková, eds.) 467–474. Physica, New York.

WEFELMEYER, W. (1996). Quasi-likelihood models and optimal inference. *Ann. Statist.* **24** 405–422.

WEISS, A. A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* **2** 107–131.

FEIKE C. DROST
BAS J. M. WERKER
CENTER AND DEPARTMENT OF ECONOMETRICS
TILBURG UNIVERSITY
P.O. BOX 90153
NL-5000 LE TILBURG
THE NETHERLANDS
E-MAIL: f.c.drost@kub.nl
b.j.m.werker@kub.nl

CHRIS A. J. KLAASSEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF AMSTERDAM
PLANTAGE MUIDERGRACHT 24
NL-1018 TV AMSTERDAM
THE NETHERLANDS
E-MAIL: chrisk@wins.uva.nl