数理解析研究所講究録
第 1703 巻 2010 年 151-158

非線形時系列モデルのイノベーション密度の推定について

広島大学大学院理学研究科　加藤賢悟 (Kengo Kato)
Graduate School of Science, Hiroshima University

1 Introduction

本稿では非線形自己回帰モデルのイノベーション密度の推定を考え，いくつかの応用例を与える．なお，本稿は著者の論文 "Estimation of the innovation density of nonlinear autoregressive models with applications", Hiroshima Statistical Research Group Techincal Report TR09-05 の要約 (résumé) である．証明および詳細な議論に関しては，上記のテクニカルリポートを参照されたい．p 次の自己回帰 (AR) モデルは

$$X_t = m(X_{t-1},\dots,X_{t-p};\boldsymbol{\theta}) + e_t,\ t = 0,\pm1,\pm2,\dots, \tag{1.1}$$

で定義される．ここで，$\boldsymbol{\theta} = (\theta_1,\dots,\theta_q)$ は未知パラメータ，$\Theta \subset \mathbb{R}^q$ はパラメータ空間（$\mathbb{R}^q$ のボレル可測集合），$m : \mathbb{R}^p \times \Theta \to \mathbb{R}$ は自己回帰関数，$\{e_t\}$ は i.i.d. の確率変数列で，すべての t に関して e_t は $\{X_{t-k}, k \geq 1\}$ と独立とする．考える問題は e_t の密度 f の推定である．

まず，e_t が観測可能であるなら，問題は簡単であり，良く知られているようにカーネル推定量

$$f_n(u_0) = \frac{1}{n}\sum_{t=1}^{n} K_{h_n}(e_t - u_0)$$

を用いればよい．ここで，$u_0 \in \mathbb{R}$ は任意の標本点、$K(\cdot)$ カーネル関数 (確率密度関数)，$h_n > 0$ はバンド幅で $K_{h_n}(u) = h_n^{-1}K(u/h_n)$ である．観測可能なデータに対するカーネル推定量の性質は良く知られている．Fan and Yao (2005) の 5 章などが良い参考となるだろう．もちろん，実際は e_t は観測できないので，f を推定する一つの自然な方法として，e_t の代わりに残差を用いることが考えられるだろう．いま，$\boldsymbol{\theta}$ の $\sqrt{n}$ 一致推定量 $\hat{\boldsymbol{\theta}}$ が与えられたとする．以下では，標本は $\{X_{-p+1},\dots,X_n\}$ とする．このとき，モデル (1.1) の残差は $\hat{e}_t = X_t - m(\mathbf{X}_{t-1};\hat{\boldsymbol{\theta}})$，ただし $\mathbf{X}_{t-1} = (X_{t-1},\dots,X_{t-p})$，で与えられる．残差に基づく f のカーネル推定量は

$$\hat{f}_n(u_0) = \frac{1}{n}\sum_{t=1}^{n} K_{h_n}(\hat{e}_t - u_0)$$

で与えられる．

ところで，いま $\hat{e}_t$ は e_t の代わり (proxy) として用いているのでだから，$\hat{f}_n$ は f_n に何らかの意味で "近い" ことが予想される．そこで，本稿では $\hat{f}_n$ と f_n の間の一様距離 $\|\hat{f}_n - f_n\|_\infty := \sup_{u_0\in\mathbb{R}} |\hat{f}_n(u_0) - f_n(u_0)|$ の収束レート考える．類似の問題に取り組んだ論文として，Liebscher (1999) と Müller et al. (2005) がある．前者は，適当な条件のもとで，$\mathbb{R}$ の任意のコンパクト集合上での一様収束レートが $o_{a.s.}\{(nh_n)^{-1/2}\}$ となるこ

とを示した．一方，後者は $\hat{f}_n$ と f_n の間の重み付 L_1 ノルムの収束レートを考察している． ただし，本稿で扱われている問題はいずれの論文においても含まれない．

本稿の目的は $\|\hat{f}_n - f_n\|_\infty$ の収束レートに関するより sharp な結果を導くことであり，本稿の結果は Liebscher (1999) で得られた結果をいくつかの観点から改良している．応用として，$\hat{f}_n$ の exact な一様収束レート，pointwise な漸近正規性，及び，残差に基づく Bickel-Rosenblatt 統計量の漸近分布の導出を扱う．

2 主結果

次の仮定を考える．以下，パラメータの真値 $\boldsymbol{\theta} \in \Theta$ は固定とする．

(A1) $\{X_t\}$ は狭義定常で指数 β ミキシング．

(A2) 真値 $\boldsymbol{\theta}$ はパラメータ空間 Θ の内点．

(A3) $(\mathbf{x}, \boldsymbol{\vartheta}) \mapsto m(\mathbf{x}; \boldsymbol{\vartheta})$ はボレル可測; ボレル可測な関数 $M(\mathbf{x})$ が存在して，$\boldsymbol{\theta}$ の近傍の点 $\boldsymbol{\vartheta}$ に対して，$|m(\mathbf{x}; \boldsymbol{\vartheta}) - m(\mathbf{x}; \boldsymbol{\theta})| \leq M(\mathbf{x})\|\boldsymbol{\vartheta} - \boldsymbol{\theta}\|$ が成り立つ．また，$\mathrm{E}[M^2(\mathbf{X}_{t-1})] < \infty$ である．

(A4) イノベーション密度 f は有界で，λ 次 Hölder 連続である．ただし，$\lambda \in (0, 1]$．

(A5) カーネル関数 $K(\cdot)$ はリプシッツ連続な密度関数であって，

$$\int_{-\infty}^{\infty} |uK'(u)| du < \infty. \tag{2.1}$$

を満たす．

(A6) $\hat{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}$ の $\sqrt{n}$ 一致推定量である．

さまざまなミキシングの概念とその基本的な性質に関しては，Fan and Yao (2005) の 2.6 節が参考になる．実際は，指数 β ミキシングという仮定は必要以上に強いが，簡単のためこの仮定をおく．なお，Liebscher (1999) も同様の仮定を置いている．条件 (A5) を満たすカーネルの例として，Gaussian，Epanchnikovka カーネルがある．コンパクトなサポートを持つリプシッツ連続なカーネルなら条件 (A5) を満たすことに注意する．また，リプシッツ連続であることと，(2.1) より $K(\cdot)$ は $\mathbb{R}$ 上で有界変動であることに注意．非線形 AR モデルのパラメータ推定に関する論文はたくさんあるが，例えば Klimko and Nelson (1978)，Tjøstheim (1986)，Koul (1996) などが参考になる．

次の定理が本稿の主結果である．

Theorem 2.1. 条件 *(A1)-(A6)* を仮定する．また，$\{r_n\}, \{h_n\}$ を

$$h_n \to 0,\ r_n \to \infty,\ \frac{n^{3/2} h_n^2}{r_n^2 \log n} \to \infty \tag{2.2}$$

を満たす数列とする．このとき，$\|\hat{f}_n - f_n\|_\infty = o_p(r_n^{-1}) + O_p(n^{-1/2} h_n^{\lambda - 1} \wedge 1)$ が成り立つ．

Remark 2.1. 条件 (A6) を次の条件

$$\limsup_{n\to\infty}\sqrt{\frac{n}{\log\log n}}\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|\le \text{const., a.s.} \tag{2.3}$$

に取り換えれば，上記の定理の概収束バージョンが得られる．Liebscher (1999) は (2.3) を条件として仮定している．また，Liebscher (2003) は (2.3) が成り立つための十分条件を調べている．

Theorem 2.1 の証明の概略を述べる．$\Delta(\mathbf{X}_{t-1};\boldsymbol{\vartheta})=m(\mathbf{X}_{t-1};\boldsymbol{\vartheta})-m(\mathbf{X}_{t-1};\boldsymbol{\theta})$ とおく．$f_n(u_0,\boldsymbol{\vartheta})=n^{-1}\sum_{t=1}^{n}K_{h_n}(e_t-u_0-\Delta(\mathbf{X}_{t-1},\boldsymbol{\vartheta}))$ と定めると，差 $\hat{f}_n(u_0)-f_n(u_0)$ は

$$\begin{aligned}\hat{f}_n(u_0)-f_n(u_0)=[\{f_n(u_0,\hat{\boldsymbol{\vartheta}})-\mathrm{E}[f_n(u_0,\boldsymbol{\vartheta})]|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\theta}}}\}-\{f_n(u_0,\boldsymbol{\theta})-\mathrm{E}[f_n(u_0,\boldsymbol{\theta})]\}]\\+\{\mathrm{E}[f_n(u_0,\boldsymbol{\vartheta})]|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\theta}}}-\mathrm{E}[f_n(u_0,\boldsymbol{\theta})]\}\end{aligned}$$

と分解される．$\hat{\boldsymbol{\theta}}$ が $\sqrt{n}$ 一致であることに注意して，第一項を経験過程の手法を使って評価する．ただし，データが i.i.d. ではないので，Yu (1993, 1994) のように，Eberlein (1984) の補題を用いて，もともとのデータ系列を独立なブロック系列に置き換える．ブロック系列に対して，対称化と “squre root trick” を用いて確率を評価する，ということを行う．詳細はテクニカルリポートを参照のこと．第二項は非確率的な写像 $\boldsymbol{\vartheta}\mapsto\mathrm{E}[f_n(u_0,\boldsymbol{\vartheta})]$ にのみ依存しているので評価は難しくない．

3 応用

本節では，Theorem 2.1 のいくつかの応用例を考察する．

Exactな一様収束レート：観測可能なデータに対するカーネル推定量の exact な一様収束レートに関しては Silverman (1978)，Stute (1984)，Deheuvels (2000)，Einmahl and Mason (2000, 2005)，Giné and Guillou (2002) などを参照．ここでは，Giné and Guillou (2002) の結果を紹介する．いま，$K(\cdot)$ は有界でコンパクトなサポートを持ち，関数族 $\mathcal{K}=\{e\mapsto K((e-u_0)/h):u_0\in\mathbb{R},h>0\}$ は pointwise measurable (van der Vaart & Wellner (1996) を参照) であって，Euclidean (Nolan and Pollard (1987) を参照) とする．また，f は $\mathbb{R}$ 上有界で一様連続であるとする．このとき，Giné and Guillou (2002) の Theorem 3.3 より，

$$h_n\downarrow 0,\ nh_n\uparrow\infty,\ \frac{nh_n}{|\log h_n|}\to\infty,\ \frac{|\log h_n|}{\log\log n}\to\infty,$$

であるなら，

$$\lim_{n\to\infty}\sqrt{\frac{nh_n}{2\log h_n^{-1}}}\|f_n-\mathrm{E}[f_n(\cdot)]\|_\infty=\|K\|_2\|f\|_\infty^{1/2}\ \text{a.s.}, \tag{3.1}$$

となることが分かる．ただし，$\|K\|_2^2=\int_{-\infty}^{\infty}K^2(u)du$ である．この結果を鑑みて，次の

$$\operatorname*{plim}_{n\to\infty}\sqrt{\frac{nh_n}{2\log h_n^{-1}}}\|\hat{f}_n-\mathrm{E}[f_n(\cdot)]\|_\infty=\|K\|_2\|f\|_\infty^{1/2} \tag{3.2}$$

が成り立つような条件を考えよう．いま，$\|\hat{f}_n - f_n\|_\infty$ が無視できるようにするためには，$r_n = \sqrt{nh_n/\log h_n^{-1}}$ となればよい．このとき，(2.2) の最後の条件は，$n^{1/2}h_n \log h_n^{-1}/\log n \to \infty$ であれば満たされる．また，$K(\cdot)$ が連続で有界変動であるなら，$\mathcal{K}$ に関する条件が成り立つことも分かるから，次の命題を得る．

Proposition 3.1. 条件 *(A1)-(A6)* を仮定する．ただし，条件 *(A4)* で $\lambda \in [1/2, 1]$ とする．$K(\cdot)$ がコンパクトなサポートを持ち，

$$h_n \downarrow 0,\ nh_n \uparrow \infty,\ \frac{|\log h_n|}{\log\log n} \to \infty,\ \frac{n^{1/2}h_n \log h_n^{-1}}{\log n} \to \infty,$$

であるならば，*(3.2)* が成り立つ．

Pointwise な漸近正規性：Pointwise な漸近正規性が成り立つための条件も，上記のような手順で見つけることができる．ここでは，結果のみ述べる．

Proposition 3.2. $u_0 \in \mathbb{R}$ を固定する．条件 *(A1)-(A6)* を仮定する．ただし，条件 *(A4)* で $\lambda \in (1/2, 1]$ とする．このとき，$h_n \to 0$ かつ $n^{1/2}h_n/\log n \to \infty$ ならば，

$$\sqrt{nh_n}\{\hat{f}_n(u_0) - \mathrm{E}[f_n(u_0)]\} \xrightarrow{d} N\{0, f(u_0)\|K\|_2^2\}$$

が成り立つ．

残差に基づく Bickel-Rosenblatt 統計量の漸近分布：最後の応用例として，残差に基づく Bickel-Rosenblatt 統計量の漸近分布を考える．いま，f は $\mathbb{R}$ 上で正かつ連続とする．次の統計量を考える：

$$\hat{M}_n = (nh_n)^{1/2} \sup_{u_0 \in [0,1]} |\hat{f}_n(u_0) - \mathrm{E}[f_n(u_0)]|/\sqrt{f(u_0)},$$
$$M_n = (nh_n)^{1/2} \sup_{u_0 \in [0,1]} |f_n(u_0) - \mathrm{E}[f_n(u_0)]|/\sqrt{f(u_0)}.$$

Bickel and Rosenblatt (1973) は Theorem 3.1 において，彼らの条件 A1-(b), A2，A3 のもとで，$h_n = \text{const.} \times n^{-\delta}$，$0 < \delta < 1/2$ なら，

$$\mathrm{P}\left((-2\log h_n)^{1/2}\left(\frac{M_n}{\|K\|_2} - d_n\right) < x\right) \to \exp\{-2\exp(-x)\}, \tag{3.3}$$

where

$$d_n = (-2\log h_n)^{1/2} + \frac{1}{(-2\log h_n)^{1/2}}\left(\log \frac{1}{2\pi}\frac{\|K'\|_2}{\|K\|_2}\right)$$

となることを示した．$\hat{M}_n - M_n = o_p\{(-\log h_n)^{-1/2}\}$ とするには，Theorem 2.1 において $r_n = (-nh_n \log h_n)^{1/2}$ とすればよい．このとき，(2.2) の最後の条件は $h_n = \text{const.} \times n^{-\delta}$ with $0 < \delta < 1/2$ に対して満たされる．条件 (A5) は Bickel and Rosenblatt (1973) の条件 A1-(b) を含意する．Bickel and Rosenblatt (1973) の条件 A3 の最初の部分は f がリプシッツ連続で $f'/f^{1/2}$ が有界なら満たされる．従って，次の命題を得る．

Proposition 3.3. 条件 *(A1)-(A6)* を仮定する．ただし，条件 *(A4)* は $\lambda = 1$ に対して成り立つとする（すなわち，f はリプシッツ連続とする）．さらに，f は正であって，$f'/f^{1/2}$ は有界，$u^2K(u), u^2K'(u)$ は可積分とする．このとき，$h_n = const. \times n^{-\delta}$ *with* $0 < \delta < 1/2$ であるならば，*(3.3)* において M_n を $\hat{M}_n$ に取り換えたものが，すべての x に対して成り立つ．

4 例

本節ではいくつかの具体的なモデルに対して，Theorem 2.1 の条件を確かめる．具体的には線形な AR モデル，閾値自己回帰モデル（TAR モデル）(Tong and Lim, 1980)，指数 AR モデル（EXPAR モデル） (Ozaki, 1980; Haggan and Ozaki, 1981) および対数変換された ARCH モデル (Engle, 1982) を扱う．

Example 4.1 (線形 AR モデル)．　次の線形 AR モデルを考える：

$$X_t = \theta_1 X_{t-1} + \cdots + \theta_p X_{t-p} + e_t.$$

ここで，$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$，$\{e_t\}$ は平均ゼロ，分散有限の i.i.d. の確率変数列とする．このとき，$q = p$，$m(\mathbf{x}; \boldsymbol{\vartheta}) = \vartheta_1 x_1 + \cdots + \vartheta_p x_p$ である．AR 過程の定常 (因果性) 条件は Brockwell and Davies (1991) の Theorem 3.1.1 で述べられている：特性多項式 $\theta(z) = 1 - \theta_1 z - \cdots - \theta_p z^p$ が単位円内 $\{z \in \mathbb{C} : |z| \leq 1\}$ で零点を持たないなら，AR 方程式は定常解 $X_t = \sum_{j=0}^{\infty} a_j e_{t-j}$ をもち，係数 a_j は指数的に減衰する．もし，さらに e_t が密度を持つなら，$\{X_t\}$ は指数 β ミキシングとなることが知られている．条件 (A3) は $M(\mathbf{x}) = \|\mathbf{x}\|$ とすれば満たされる．条件 (A6) に関しては，Brockwell and Davies (1991) などを参照.

Example 4.2 (TAR モデル)．閾値が既知の次の TAR モデルを考える：

$$X_t = \theta_1 X_{t-1} I(X_{t-1} \leq 0) + \theta_2 X_{t-1} I(X_{t-1} > 0) + e_t.$$

ここで，$\boldsymbol{\theta} = (\theta_1, \theta_2)$ は $\theta_1 \neq \theta_2$ を満たし，$\{e_t\}$ は平均ゼロ，分散有限の i.i.d. の確率変数列とする．このとき，$p = 1, q = 2$，$m(x; \boldsymbol{\vartheta}) = \vartheta_1 x I(x \leq 0) + \vartheta_2 x I(x > 0)$ である．過程 $\{X_t\}$ は $\theta_1 < 1, \theta_2 < 1, \theta_1\theta_2 < 1$ かつ e_t が正の密度を持つとき条件 (A1) をみたすことが知られている（An and Huang (1996) の Example3.1 を参照)．条件 (A3) は $M(x) = |x|$ とすれば満たされる．条件 (A6) に関しては，適当な正則条件のもとで，例えば条件付最小二乗推定量が $\sqrt{n}$ 一致推定量となる．

Example 4.3 (EXPAR モデル)．次の EXPAR モデルを考える：

$$X_t = \{\theta_1 + \theta_2 \exp(-\theta_3 X_{t-1}^2)\} X_{t-1} + e_t.$$

ここで，$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$，$\theta_3 > 0$ であって，$\{e_t\}$ は平均ゼロ，分散有限の i.i.d. の確率変数列とする．このとき，$p = 1, q = 3$，$m(x; \boldsymbol{\vartheta}) = \{\vartheta_1 + \vartheta_2 \exp(-\vartheta_3 x^2)\} x$ である．過程 $\{X_t\}$ は，$|\theta_1| < 1$ かつ e_t が正の密度を持つとき条件 (A1) を満たすことが知られている（An and Huang (1996) の Example 3.2 を参照)．条件 (A3) は適当な定数 $C > 0$ に対して，$M(x) = C|x|$ とすれば満たされる．条件 (A6) に関しては，Tjøstheim (1986) が上記の条件のほかに $\mathrm{E}[e_t^6] < \infty$ を仮定すれば，条件付最小二乗推定量が $\sqrt{n}$ 一致となることを示している．

Example 4.4 (対数変換された ARCH 過程). 最後の例は，GARCH モデルに対する LAD 推定を提案した Peng and Yao (2003) が動機となっている．次数 p の ARCH は

$$Y_t = \sigma_t \epsilon_t,\ \sigma_t^2 = \theta_0 + \sum_{j=1}^{p} \theta_j Y_{t-j}^2$$

で与えられる．ただし，$\theta_j > 0$ $(j = 0, \ldots, p)$ であり，$\boldsymbol{\theta} = (\theta_0, \ldots, \theta_p)$, $\{\epsilon_t\}$ は平均ゼロ，分散有限の i.i.d. の確率変数列とする．通常は，ϵ_t を $\mathrm{E}[\epsilon_t^2] = 1$ と基準化して，ガウシアン擬最尤推定量 (QMLE) を用いることが多い．ガウシアン QMLE の漸近的な性質は，Weiss (1986)，Hall and Yao (2003) などによって研究されている．一方，Peng and Yao (2003) は，ϵ_t^2 のメディアンが 1 となるような基準化を考えた．この基準化の下で，彼らは LAD 型推定量

$$\hat{\boldsymbol{\theta}}_{\mathrm{LAD}} = \arg\min_{\boldsymbol{\vartheta}} \sum_{t=1}^{n} |\log Y_t^2 - \log(\vartheta_0 + \textstyle\sum_{j=1}^{p} \vartheta_j Y_{t-j}^2)|$$

を提案した．この LAD 推定量のガウシアン QMLE に対する利点は，LAD 推定量が ϵ_t に関するより弱いモーメント条件のもとで漸近正規性が成り立つことがあげられる．実際，LAD 推定量は $\mathrm{E}[\epsilon_t^2] < \infty$ なる条件のもとで漸近正規であるのに対して，ガウシアン QMLE は $\mathrm{E}[|\epsilon_t|^d] = \infty$ with $2 < d < 4$ なるとき漸近正規とはならず収束レートも $\sqrt{n}$ よりも遅いことが知られている（Hall and Yao (2003) を参照）．正則条件のもとで，$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{LAD}} - \boldsymbol{\theta})$ の漸近分布は

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{LAD}} - \boldsymbol{\theta}) \xrightarrow{d} N[\mathbf{0}, \Sigma/\{4f^2(0)\}]$$

となることが示される．ここで，f は $\log \epsilon_t^2$ の密度であり，Σ に関しては Peng and Yao (2003) を参照．従って，LAD に基づいて統計的推測を行う場合は $f(0)$ を推定する必要がある．また，LAD 推定量は $\log \epsilon_t^2$ がラプラス分布を持つときの最尤推定量に対応しているので，$\log \epsilon_t^2$ の分布がラプラス分布に近いかどうかの検定は興味のあるところである（Huang et al. (2008) を参照）．

ところで，$X_t = \log Y_t^2$ と変換すれば，$\{X_t\}$ は $e_t = \log \epsilon_t^2$ をイノベーションに，$m(\mathbf{x}; \boldsymbol{\vartheta}) = \log\{\vartheta_0 + \sum_{j=1}^{p} \vartheta_j \exp(x_j)\}$ を自己回帰関数に持つ非線形 AR モデルに従うことが分かる．従って，前節までの結果を f の統計的推測に用いることが可能である．もともとの過程 $\{Y_t\}$ は $(\mathrm{E}[\epsilon_t^2])^{1/2} \sum_{j=1}^{p} \theta_j < 1$ であって ϵ_t が 0 を含む区間上で正の密度を持てば，定常かつ指数 β ミキシングとなるとなることが知られている（Carrasco and Chen (2002) を参照）ので，同じ条件のもとで，過程 $\{X_t\}$ は条件 (A1) を満たす．条件 (A3) は $M(\mathbf{x}) = \mathrm{const}$ で満たされる．条件 (A6) に関しては，すでにいくつか紹介した．

References

An, H.Z. and Huang, F.C. (1996). The geometric ergodicity of nonlinear autoregressive models. *Statist. Sinica* **6**, 943-956.

Bickel, P.J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071-1095. [Corrections: **3**, (1975) 1370.]

Brockwell, P.J. and Davies, R.A. (1991). *Time Series: Theory and Methods*, 2nd eds. Springer-Verlag, New York.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18**, 17-39.

Deheuvels, P. (2000). Uniform limit laws for kernel density estimators on possibly unbounded intervals. In *Recent Advaces in Reliability Theory: Methodology, Practice and Inference* (eds. N. Limnios and M. Nikulin), 477-492, Birkhauser, Boston.

Eberlein, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.* **2**, 291-293.

Einmahl, U. and Mason, D.M. (2000). An empirical process approach to the uniform consistency of kernel type function estimators. *J. Theoret. Probab.* **13**, 1-37.

Einmahl, U. and Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33**, 1380-1403.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**, 987-1008.

Fan, J. and Yao, Q. (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag, New York.

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.* **38**, 907-921.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189-196.

Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**, 285-317.

Huang, D., Wang, H. and Yao, Q. (2008). Estimating GARCH models: when to use what? *Econometrics J.* **11**, 27-38.

Klimko, L.A. and Nelson, P.L. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 629-643.

Koul, H.L. (1996). Asymptotics of some estimators and sequential residual empiricals in nonlinear autoregressive models. *Ann. Statist.* **24**, 380-404.

Liebscher, E. (1999). Estimating the density of the residuals in autoregressive models. *Stat. Inference Stoch. Process.* **2**, 105-117.

Liebscher, E. (2003). Strong convergence of estimators in nonlinear autoregressive models. *J. Multivariate Anal.* **84**, 247-261.

Müller, U.U., Schick, A. and Wefelmeyer, W. (2005). Weighted residual-based density estimators for nonlinear autoregressive models. *Statist. Sinica* **15**, 177-195.

Nolan, D. and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.* **15**, 780-799.

Ozaki, T. (1980). Non-linear time series models for nonlinear random variables. *J. Appl. Probab.* **17**, 84-93.

Peng, L. and Yao, Q. (2003). Least absolute deviations for ARCH and GARCH models. *Biometrika* **90**, 967-975.

Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6**, 177-184.

Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case. *Ann. Probab.* **22**, 361-379.

Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stochastic Process. Appl.*. **21**, 251-273.

Tong, H. and Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **42**, 245-292.

van der Vaart, A.W. & Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag, New York.

Weiss, A.A. (1986). Asymptotic theory of ARCH models: estimation and testing. *Econometric Theory* **2**, 107-131.

Yu, B. (1993). Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. *Ann. Statist.* **21**, 711-735.

Yu, B. (1994). Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.* **22**, 94-116.