NONPARAMETRIC ESTIMATION AND TESTING IN PANELS OF INTERCORRELATED TIME SERIES

By Vidar Hjellvik, Rong Chen and Dag Tjøstheim

Institute of Marine Research, The University of Illinois at Chicago and University of Bergen

First Version received January 2001

Abstract. We consider nonparametric estimation and testing of linearity in a panel of intercorrelated time series. We place the emphasis on the situation where there are many time series in the panel but few observations for each of the series. The intercorrelation is described by a latent process, and a conditioning argument involving this process plays an important role in deriving the asymptotic theory. To be accurate the asymptotic distribution of the test functional of linearity requires a very large number of observations, and bootstrapping gives much better finite sample results. A number of simulation experiments and an illustration on a real data set are included.

Keywords. Heteroscedastic; intercorrelated time series; linearity testing; nonlinear; nonparametric estimation; panel.

1. INTRODUCTION

There is a vast literature on linear modelling of panel data (see e.g. Mátyas and Sevestre, 1992; Baltagi, 1995; Arellano, 2003), but surprisingly little on nonlinear models and nonparametric methods. In the univariate time series case, nonlinear and nonparametric analyses have played important roles for a long time, and in a number of practical examples a nonlinear approach has turned out to be essential to understand the data. There is no good reason to believe that the situation should be different for panels of time series data. One will surely find situations where a linear approximation is not feasible, although there is very little evidence of this in the literature. There is a considerable number of papers on curve estimation (cf. the works by Rice and Silverman, 1991; Gasser and Kneip, 1995; Ramsay and Silverman, 1997, 2002; Cardot *et al.*, 2003), but the focus is different. More close to our work is the paper by Yao *et al.* (1998).

A rather general nonlinear dynamic model for a panel of time series is given by

$$X_{(i)t} = f(X_{(i)t-1}, \ldots, X_{(i)t-p}) + \eta_t + \lambda_i + g(W_{(i)t}) + \epsilon_{(i)t}. \quad (1)$$

Here $\{X_{(i)t},\ i = 1,\ldots,n;\ t = 1,\ldots,T\}$ are the observations from n individual time series with T observations taken for each. The model (1) is the nonlinear analogue of the linear dynamical models presented in Hsiao (1986, Ch. 4). (see also

Arellano, 2003, part 2 and Greene, 2003, Ch. 13.6). The function f is assumed to be independent of i and t, but this can be relaxed if T and n, respectively, are large enough. The quantities $\{\eta_t\}$ represent effects over time influencing all of the time series, and similarly the variables $\{\lambda_i\}$ stand for individual effects not taken care of by the explanatory variables $\{W_{(i)t}\}$ that are entering the system nonlinearly through the function g. Finally, $\{\epsilon_{(i)t}\}$ are the error terms assumed to be independently identically distributed (i.i.d.) with $\text{var}(\epsilon_{(i)t}) = \sigma_\epsilon^2$ in the following.

The model (1) is far too complicated to work with at present. It is important to understand the dynamics of a simple situation first. In this paper, we therefore concentrate on the model

$$X_{(i)t} = f(X_{(i)t-1}, \ldots, X_{(i)t-p}) + \eta_t + \epsilon_{(i)t}. \tag{2}$$

The process $\{\eta_t\}$ [with $\text{var}(\eta_t) = \sigma_\eta^2$] describes the intercorrelation (or interdependence) in the panel. It is assumed to be independent of $\{\epsilon_{(i)t}\}$, and it could be thought of as an external synchronous agent influencing all the series. In the most general case it could be nonstationary or even deterministic, and actually in a typical situation of large n and small T, we allow $\{\eta_t\}$ to have an arbitrary dependence structure.

Models like (2), or rather the linear analogy, are slightly unconventional in an econometric setting where one tends to include λ_i and omit η_t. As pointed out in Hjellvik and Tjøstheim (1999a), the omission of η_t can lead to inconsistency and loss of efficiency (as can the omission of λ_i, of course; see e.g. Greene, 2003, Ch. 13, for a discussion). A practical example where this effect is present is displayed in Figure 1, which shows the logarithms of the yearly catches of grey-sided voles over a period of 31 years at 91 different locations at the island of Hokkaido (cf. Bjørnstad *et al.*, 1996). Clearly, the series are intercorrelated, and the three geographical areas have been chosen so as to minimize individual variations [measured by λ_i in (1)] from one catch site to another.

The description of the intercorrelation structure by means of the series $\{\eta_t\}$ can be looked at as a simplified one-factor model. Much more complicated factor models for describing interdependence have been proposed in the econometric literature (see e.g. Geweke, 1977; Chamberlain, 1983; Chamberlain and Rothschild, 1983). Important recent progress in dynamic factor models have been made by Forni *et al.* (2000, 2001) and Stock and Watson (2002). The capabilities of such models in handling interdependence structures are far superior to that of using a single process $\{\eta_t\}$ as in (1), but a drawback of the dynamic factor models is that they are linear. We believe that it would be of great practical and theoretical interest to introduce a nonlinear framework in a factor context, where one allows for nonlinear dependence on previous X-values and/or on a set of dynamic factors. Possibly, one could also relax the cross-sectional independence assumption on the residuals $\{\epsilon_{(i)t}\}$ in such a context. But, again, to be able to go on to this difficult task we think that it is essential first to understand how to deal with the much simpler case of (2).

For the model (2) we concentrate on two problems: (i) estimating f nonparametrically, (ii) testing whether $X_{(i)t}$ is linearly generated. In a practical

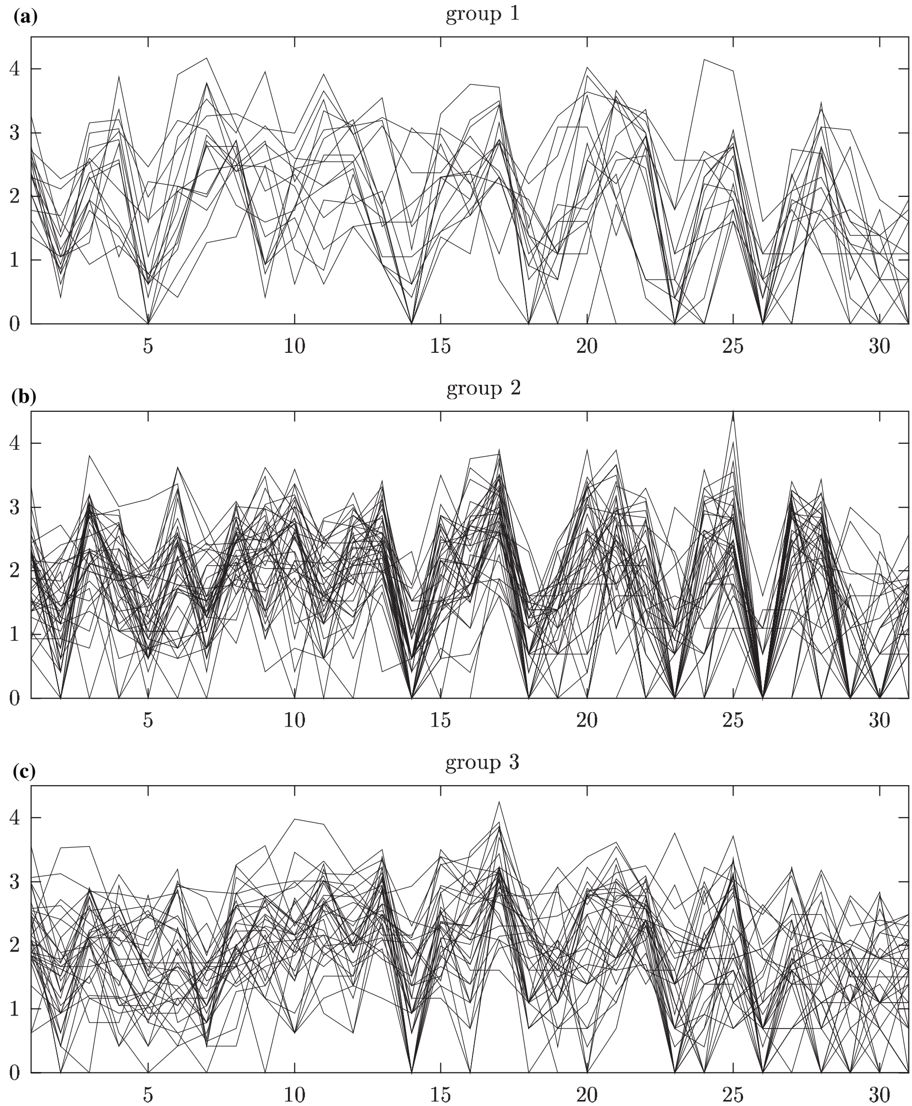

FIGURE 1. The figure shows for $n = 16,41$ and 34 for group 1, 2 and 3, respectively, $\log (X_{(i)t} + 1)$ where $\{X_{(i)t},\ i = 1,\ldots,n;\ t = 1,\ldots,31\}$ is the number of grey-sided voles trapped each year from 1961 to 1992 in n different locations in Hokkaido, Japan.

situation, the latter problem should be attacked first, as it is important to know if linearity can be used as a simplifying device. This should really form part of an exploratory analysis for a panel, but the exploratory issue has been much ignored even for linear panels (cf. Hjellvik and Tjøstheim, 1999b). We establish tests of linearity in Sections 3 and 4, whereas the estimation theory is considered in Section 2.

As for almost all nonparametric methods applied to correlated data, an essential condition, when T is large, is that the series satisfies certain mixing condition so that the whitening by window principle (Hart, 1996) is achievable. Among various mixing conditions used in the literature, α-mixing is reasonably weak and is known to be fulfilled by many nonlinear time series models under some regularity conditions. Specifically, Tjøstheim (1990) discussed the use of ergodicity and α-mixing conditions in nonlinear time series analysis. Among others Cline and Pu (1999), Fan and Li (1999), Masry and Tjøstheim (1995, 1997), Meyn and Tweedie (1994), Ling (2004), Lu (1998) have provided sets of conditions for a (univariate) nonlinear time series to be stationary and α-mixing with specified rates on the mixing coefficients, although sufficient and necessary conditions are extremely difficult to obtain. One such sufficient condition is as follows. (In addition it attains a mixing rate which is sufficient for our purposes; i.e. such that the results of Masry and Fan (1997) can be applied.)

1.1. *Geometric ergodicity condition*

For the function f in (2), there exist two linear functions

$$l_1(\mathbf{x}) = a_0 + \sum_{j=1}^{p} a_j x_j \quad \text{and} \quad l_2(\mathbf{x}) = b_0 + \sum_{j=1}^{p} b_j x_j,$$

such that $l_1(\mathbf{x}) < f(\mathbf{x}) < l_2(\mathbf{x})$ for $\mathbf{x} = (x_1, \ldots, x_p)$ outside a compact region in $\Re^p$. In addition, the roots of

$$a_0 + \sum_{j=1}^{p} a_j \lambda^j = 0 \quad \text{and} \quad b_0 + \sum_{j=1}^{p} b_j \lambda^j = 0$$

are all outside the unit circle (cf. Masry and Tjøstheim 1995).

In this paper, we will simply assume that the function f in model (2) is such that the resulting time series are stationary and α-mixing with a sufficiently fast mixing rate.

We also extend our method to a group of heteroscedastic models of the form

$$\begin{aligned} X_{(i)t} &= f(X_{(i)t-1}, \ldots, X_{(i)t-p}) + e_{(i)t}, \qquad e_{(i)t} = \{g(e_{(i)t-1})\}^{1/2} z_{(i)t}, \\ z_{(i)t} &= \eta_t + \epsilon_{(i)t}, \end{aligned} \tag{3}$$

the case where f is a linear function being emphasized.

For asymptotic behaviour of the proposed methods, we will study two cases: (i) n fixed and T large and (ii) T fixed and n large, although more space will be devoted to the latter case. This is because it is a common practical situation that there are many series with few observations (T small) for each series, and it is also theoretically challenging. Such problems have been treated in the linear case in Hjellvik and Tjøstheim (1999a), and a related question with T small ($T \geq 3$) has been dealt with by Cox and Solomon (1988). As for the univariate case, bootstrapping will play a crucial role in the linearity testing (cf. Hjellvik and Tjøstheim, 1995). We refer to Section 4 for a number of simulated examples and to Section 5 for the real data example involving the Hokkaido data depicted in Figure 1.

2. NONPARAMETRIC ESTIMATION OF PANEL TIME SERIES

2.1. *The conditional mean function: the $T \to \infty$ and n fixed case*

In this case, assuming that $\{\eta_t\}$ is an i.i.d sequence, the standard results (see e.g. Masry and Fan, 1997) are applied to each individual series $\{X_{(i)t},\ t = 1,\ldots,T\}$, given that the sequence is α-mixing or ρ-mixing. Since the extension to the panel time series case is relatively simple, while explicitly stating the conditions is tedious, we will omit the conditions and refer the readers to Masry and Fan (1997).

We first consider (2) in the first-order case $p = 1$,

$$X_{(i)t} = f(X_{(i)t-1}) + e_{(i)t}, \qquad e_{(i)t} = \eta_t + \epsilon_{(i)t}, \quad i = 1, \ldots, n, t = 1, \ldots, T \tag{4}$$

with $\mathrm{var}(e_{(i)t}) = \sigma_e^2$. In this simple case, it is feasible to use high-order local polynomial estimation.

For each series i, let $\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_\ell$ be the minimizer of

$$\sum_{t=2}^{T} \left\{ X_{(i)t} - \sum_{s=0}^{\ell} \gamma_s (X_{(i)t-1} - x)^s \right\}^2 K_h(X_{(i)t-1} - x). \tag{5}$$

Here $K_h(x) = h^{-1}K(h^{-1}x)$, K is non-negative serving as a kernel function, and h is the bandwidth controlling the size of the local neighbourhood. Define the local polynomial estimator of f from the ith panel time series as

$$\hat{f}_i(x) = \hat{\gamma}_0.$$

The combined estimator for $f(x)$ is then defined as

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}_i(x). \tag{6}$$

Note that in this case the number of series n in the panel is fixed.

We have the following theorem:

THEOREM 1. *Under conditions given in Masry and Fan* (1997), *as n is fixed and $T \to \infty$, we have*

$$\sqrt{Th}\left\{\hat{f}(x) - f(x) - \frac{f^{(\ell+1)}(x)B_0}{(\ell+1)!}h^{\ell+1}\right\} \xrightarrow{\mathcal{D}} N\left\{0, \frac{V_0\sigma_e^2}{np(x)}\right\},$$

where $p(x)$ is the stationary density function at point x for all series $\{X_{(i)t}\}$, B_0 is the first element of

$$B = S^{-1}\boldsymbol{\mu} \tag{7}$$

and V_0 is the first diagonal element of $V = S^{-1}\tilde{S}S^{-1}$ where

$$S = \begin{bmatrix} \mu_0 & \cdots & \mu_\ell \\ \mu_1 & \cdots & \mu_{\ell+1} \\ \vdots & \ddots & \vdots \\ \mu_\ell & \cdots, & \mu_{2\ell} \end{bmatrix}; \quad \tilde{S} = \begin{bmatrix} v_0 & \cdots & v_\ell \\ v_1 & \cdots & v_{\ell+1} \\ \vdots & \ddots & \vdots \\ v_\ell & \cdots, & v_{2\ell} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_{\ell+1} \\ \mu_{\ell+2} \\ \vdots \\ \mu_{2\ell+1} \end{bmatrix}, \tag{8}$$

where $\mu_i = \int u^i K(u)\mathrm{d}u$ and $v_i = \int u^i K^2(u)\mathrm{d}u$.

A sketch of the proof of the theorem is given in Appendix A.

For the general AR(p) case in (2) one encounters the curse of dimensionality in practice. Often, some dimension-reduction restrictions are required, such as the additive autoregressive models (Chen and Tsay, 1993b) or the functional coefficient autoregressive models (Chen and Tsay, 1993a; Cai *et al.*, 2000). If the estimation of the general p-dimensional function is desired, it is argued by Fan and Yao (2003) that a higher-order polynomial is rarely used. Here we present a local linear estimator for the p-dimensional case.

Let

$$\mathbf{x}_{(i)t-1} = (X_{(i)t-1}, \ldots, X_{(i)t-p}) \quad \text{and} \quad \mathbf{x} = (x_1, \ldots, x_p).$$

Define a multiplicative kernel function

$$\mathbf{K}(\mathbf{x}) = \prod_{i=1}^{p} K(x_i).$$

With a diagonal bandwidth matrix $H = \mathrm{diag}(h_1,\ldots,h_p)$, define

$$\mathbf{K}_H = (\prod_{i=1}^{p} h_i)^{-1}\mathbf{K}(H^{-1}\mathbf{x}).$$

Let $\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_p$ be the minimizer of

$$\sum_{t=p+1}^{T}\left\{X_{(i)t} - \gamma_0 - \sum_{s=1}^{p}\gamma_s(X_{(i)t-s} - x_s)\right\}^2 \mathbf{K}_H(\mathbf{X}_{(i)t-1} - \mathbf{x}),$$

and define the local linear estimator of f based on $\{X_{(i)t},\ t = 1,\ldots,T\}$ as

$$\hat{f}_i(\mathbf{x}) = \hat{\gamma}_0.$$

The overall estimator for $\hat{f}(\mathbf{x})$ is then defined as

$$\hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}_i(\mathbf{x})$$

We have the following theorem:

THEOREM 2. *Under the conditions given in Masry and Fan* (1997) *and Fan and Yao* (2003), *as n is fixed and* $T \to \infty$, *we have*

$$\left\{T\prod_{i=1}^{p} h_i\right\}^{1/2} \left\{\hat{f}(\mathbf{x}) - f(\mathbf{x}) - 0.5\,tr\{f''(\mathbf{x})HH^{\mathrm{T}}\boldsymbol{\mu}\}\right\} \xrightarrow{\mathcal{D}} N\left\{0, \frac{\nu\sigma_e^2}{np(\mathbf{x})}\right\},$$

where $f''(\mathbf{x})$ *is the Hessian matrix of* f, $\boldsymbol{\mu} = \int \mathbf{K}(\mathbf{u})\mathbf{u}\mathbf{u}^{\mathrm{T}}\,\mathrm{d}\mathbf{u}$, $\nu = \int \mathbf{K}^2(\mathbf{u})\mathrm{d}\mathbf{u}$ *and* $p(\mathbf{x})$ *is the stationary density function at the point* $\mathbf{x}$ *of* $\{X_{(i)t)1}, \ldots, X_{(i)t)p}\}$.

The proof of the theorem is a simple extension of that of Theorem 1, hence omitted.

2.2. *The conditional mean function: the* $n \to \infty$ *and T fixed case*

This case is not standard. We will use a conditioning argument with respect to $\{\eta_t\}$ to obtain asymptotic results. However, the assumptions on $\{\eta_t\}$ are relaxed to just requiring that it is a sequence of random variables. Note that it is difficult to impose any 'mixing' conditions on a panel of time series with no obvious order or distance measure between the series in the cross-sectional direction. Hence, most of the results dealing with correlated data based on the mixing conditions cannot be applied here. Fortunately, the length of the time series T is fixed and finite. It allows us to condition on the η-sequence which is the only source of correlation between the panels. The conditioning argument essentially treats the finite number of η_ts as unobservable constants. As a result the asymptotic distribution depends on the η_ts.

Again, we begin with the $p = 1$ case. We assume that the series start at $t = -T_0$ with independent observations $X_{(1)-T_0}, X_{(2)-T_0}, \ldots$ across the panel.

For each fixed time t, let $f_t(x) = f(x) + \eta_t$ and let $f_t^{(i)}(x)$ be the ith derivative of $f_t(x)$ evaluated at x. We first construct local polynomial estimators for each $f_t(x)$. Let $\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_\ell$ be the minimizer of

$$\sum_{i=1}^{n}\left\{X_{(i)t} - \sum_{s=0}^{\ell}\gamma_s(X_{(i)t-1} - x)^s\right\}^2 K_h(X_{(i)t-1} - x), \tag{9}$$

and define

$$\tilde{f}_t(x) = \hat{\gamma}_0.$$

Then we have the following theorem:

THEOREM 3. *Let* $\boldsymbol{\eta} = \{\eta_{-T_0}, \ldots, \eta_t\}$. *Under conditions* (B1)–(B6) *given in Appendix B*, *and conditioning on* $\boldsymbol{\eta}$, *we have*

$$\sqrt{nh}\left\{\tilde{f}_t(x) - f_t(x) - \frac{f^{(\ell+1)}(x)B_0}{(\ell+1)!}h^{\ell+1}\right\} \xrightarrow{\mathcal{D}} N\left\{0, \frac{V_0\sigma_\epsilon^2}{p_{t-1}(x \mid \boldsymbol{\eta})}\right\}$$

where B_0 *is the first element of* $S^{-1}\boldsymbol{\mu}$ *and* V_0 *is the first diagonal element of* $S^{-1}\tilde{S}S^{-1}$ *with* $S, \tilde{S}, \boldsymbol{\mu}$, *and* $f^{(\ell+1)}(x)$ *defined as in Theorem* 1, *and where* $p_{t-1}(x \mid \boldsymbol{\eta})$ *is the conditional density function of* $X_{(i)t-1}$ *given* $\boldsymbol{\eta}$ (*hence depending on* $\boldsymbol{\eta}$).

The proof of the theorem is given in Appendix B. The essential argument is that, under condition (B1) in Appendix B, $X_{(i)t}$ and $X_{(j)t}$ are conditionally independent of each other ($i \neq j$), given $\boldsymbol{\eta}$.

To avoid ambiguity, we assume $f(0) = 0$. In order to eliminate the random effect of η_t and to estimate $f(x)$, define

$$\tilde{f}(x) = \frac{1}{T-1}\sum_{t=2}^{T}\{\tilde{f}_t(x) - \tilde{f}_t(0)\}. \tag{10}$$

We have (suppressing the t-dependence in $\boldsymbol{\eta}$):

THEOREM 4. *Under the conditions of Theorem* 3, *given* $\boldsymbol{\eta}$, *we have*

$$\sqrt{nh}\left[\tilde{f}(x) - f(x) - \frac{\{f^{(\ell+1)}(x) - f^{(\ell+1)}(0)\}B_0}{(\ell+1)!}h^{\ell+1}\right]$$
$$\xrightarrow{\mathcal{D}} N\left[0, \frac{V_0\sigma_\epsilon^2}{(T-1)^2}\sum_{t=2}^{T}\left\{\frac{1}{p_{t-1}(x \mid \boldsymbol{\eta})} + \frac{1}{p_{t-1}(0|\boldsymbol{\eta})}\right\}\right].$$

Note that the bias term does not depend on $\boldsymbol{\eta}$. The asymptotic variance depends on $\boldsymbol{\eta}$ through the conditional densities p_{t-1}. Assuming that η_t has finite support and $\epsilon_{(i)t}$ has infinite support, one can bound the asymptotic variance with

$$\frac{V_0\sigma_\epsilon^2}{(T-1)^2}\sum_{t=2}^{T}\left\{\frac{1}{p_{t-1}(x|\boldsymbol{\eta})} + \frac{1}{p_{t-1}(0|\boldsymbol{\eta})}\right\} \leq \frac{CV_0\sigma_\epsilon^2}{T-1},$$

uniformly for all $\boldsymbol{\eta}$, where

$$C = \sup_{\eta_{-T_0+1},\ldots,\eta_T}\left\{\frac{1}{p_{t-1}(x|\boldsymbol{\eta})} + \frac{1}{p_{t-1}(0|\boldsymbol{\eta})}\right\} < \infty.$$

The convergence of $\tilde{f}(x)$ to $f(x)$ is illustrated in Figure 2 for $T = 2$ and $n = 10^2, \ldots, 10^5$ for the nonlinear exponential autoregressive model

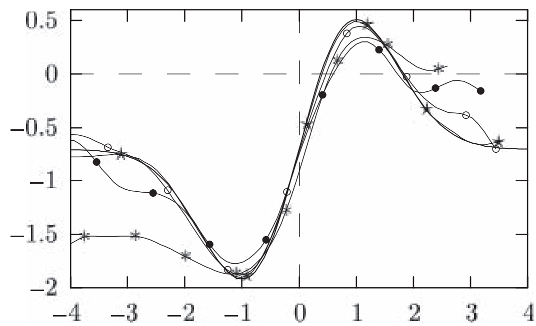

FIGURE 2. Thin lines: $\tilde{f}_2(x)$ plotted against x for four realizations of model (11) with $a = 0$, $b = 2$, $T = 2$ and $n = 100$ (asterisks), 1000 (bullets), 10,000 (open circles) and 100,000 (stars). In all cases $\eta_2 = -0.705$. Thick line: $f(x)-0.705$ plotted against x.

$$X_{(i)t} = \{a + b\, \exp(-0.5X_{(i)t-1}^2)\}X_{(i)t-1} + \eta_t + \epsilon_{(i)t}, \tag{11}$$

with $a = 0$ and $b = 2$. The bandwidth used is $h = h_n = sn^{-1/5}$, where s is the empirical standard deviation of $\{X_{(i)1},\ i = 1,\ldots,n\}$ and where we have taken $\ell = 0$ (the Nadaraya-Watson estimator) in (9). For all realizations $\eta_1 = -0.554$ and $\eta_2 = -0.705$, and $\{\eta_t\}$ and $\{\epsilon_{(i)t}\}$ are independent zero-mean Gaussian variables with $\sigma_\eta^2 = \sigma_\epsilon^2 = 1$. In the figure is shown. $\tilde{f}_2(x)$, which converges to $f(x) + \eta_2 = f(x) - 0.705$

For the p-dimensional case, again we use the local linear estimator of $f_t(\mathbf{x})$ using the cross-panel data $(\mathbf{X}_{(i)t-1}, X_{(i)t}), i = 1,\ldots,n$, with $\mathbf{X}_{(i)t-1}, \mathbf{K}, H$ defined as in Section 2.1. Specifically, let $\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_p$ be the minimizer of

$$\sum_{i=1}^{n}\left\{X_{(i)t} - \gamma_0 - \sum_{s=1}^{p}\gamma_s(X_{(i)t-s} - x_s)\right\}^2 \mathbf{K}_H(\mathbf{X}_{(i)t-1} - \mathbf{x}),$$

and define

$$\tilde{f}_t(\mathbf{x}) = \hat{\gamma}_0.$$

The overall estimator for f can then be constructed in the same way as in the $p = 1$ case.

2.3. *The conditional variance function*

It is possible to generalize Theorem 1 to an autoregressive conditionally heteroscedastic process (ARCH)-type intercorrelated panel model, with linear conditional mean function,

$$X_{(i)t} = \sum_{s=1}^{p} a_s X_{(i)t-s} + e_{(i)t}, \qquad e_{(i)t} = \{g(e_{(i)t-1})\}^{1/2} z_{(i)t}, \quad z_{(i)t} = \eta_t + \epsilon_{(i)t}. \tag{12}$$

where $\{\eta_t\}$ and $\{\epsilon_{(i)t}\}$ are independent sequences of zero-mean i.i.d. random variables, such that

$$\operatorname{var}(z_{(i)t}) = \operatorname{var}(\eta_t) + \operatorname{var}(\epsilon_{(i)t}) = \sigma_\eta^2 + \sigma_\epsilon^2 = 1.$$

Clearly, this is not the only way of defining ARCH-like structures for panels.

In the case that T is large and n is fixed, the conditional variance function g can be estimated by

$$\hat{g}(x) = \frac{(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t} - x)\tilde{e}_{(i)t+1}^2}{(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t} - x)} \tag{13}$$

where

$$\tilde{e}_{(i)t} = X_{(i)t} - \sum_{s=1}^{p} \tilde{a}_s X_{(i)t-s} \tag{14}$$

is the residual of a least squares estimation of model (12), and where the estimates $\{\tilde{a}_s, s = 1, \ldots, p\}$ are given in Hjellvik and Tjøstheim (1999a).

THEOREM 5. *Under conditions given in Masry and Tjøstheim* (1995), *for n fixed and* $T \to \infty$, *we have*

$$\sqrt{nTh}\{\hat{g}(x) - g(x) - h^2\mu_2[g'(x)p_e'(x)/p_e(x) + 0.5\,g''(x)]\} \xrightarrow{\mathcal{D}} N\left\{0, \frac{V_0 g^2(x)\sigma^2}{p_e(x)}\right\}$$

where

$$\sigma^2 = \operatorname{var}(\epsilon_{(i)t}^2) + \operatorname{var}(\eta_t^2) + 4\sigma_\epsilon^2\sigma_\eta^2, \quad \mu_2 = \int u^2 K(u)\mathrm{d}u$$

and $p_e(x)$ *is the stationary density function of* $e_{(i)t}$.

The proof of the theorem is given in Appendix C.

Note that, in practice, it may be beneficial to use $\hat{g}(x) - \hat{f}^2(x)$, where $\hat{f}(x)$ is obtained by replacing $\tilde{e}_{(i)t}^2$ with $\tilde{e}_{(i)t}$ in (13),although asymptotically they are the same.

For the case that T is small but n is large, it is difficult to define an estimator analogous to $\tilde{f}_t(x)$ and $\tilde{f}(x)$ of (9) and (10) because we do not have the additive structure $f_t(x) = f(x) + \eta_t$. If we modify model (12) to be

$$X_{(i)t} = \sum_{s=1}^{p} a_s X_{(i)t-s} + \eta_t + e_{(i)t}, \quad e_{(i)t} = \{g(e_{(i)t-1})\}^{1/2}\epsilon_{(i)t}, \tag{15}$$

it is simpler. A first-stage least-square estimation, treating η_ts as unknown constants, provides consistent estimates of $e_{(i)t}$. Then a conditioning argument warrants that the estimator (13) has the same asymptotic distribution as in Theorem 5, except that $\sigma^2 = \sigma_\epsilon^2$. However, for an ARCH-type model with

$a_1 = \cdots = a_p = 0$, whether (15) is the most appropriate generalization remains to be discussed. Indeed, at present, it is not entirely clear as to what is the best generalization of the ARCH-concept to the intercorrelated panel case. If the problem is to *test* for the presence of a nonconstant conditional variance, this is in a sense easier than the problem of estimation, since under the null hypothesis of $g(x) \equiv$ constant, both (12) and (15) reduce to the same model. As in Section 3 we have decided to use essentially the estimate $\hat{g}(x)$ of (13) in this situation, although the properties of this estimate are unknown under the alternative hypothesis of a nonconstant g for T small and n large. Some examples of $\hat{g}(x)$ are given in Figure 5, but we refer to Section 3.3 for a closer description.

3. TESTING OF LINEARITY

We are interested in testing the hypothesis that the process is linear, i.e. that the function f in (1) is linear and that g in (12) or (15) is constant so that each individual series follows a linear AR(p) model. We consider both (i) n fixed and $T \to \infty$ and (ii) T fixed and $n \to \infty$ with emphasis on the latter.

Linearity tests in the univariate case have been studied by many authors. For example, Keenan (1985), Tsay (1986), Luukkonen *et al.* (1988) and Granger and Teräsvirta (1993, Ch.6) proposed different forms of Lagrange multiplier tests. These tests and the tests by Petruccelli and Davis (1986), Chan and Tong (1986) and Tsay (1989) are designed for testing against specific types of nonlinearity.

Hjellvik and Tjøstheim (1995, 1996), Hjellvik *et al.* (1998) and Dette and Spreckelsen (2003) considered nonparametric tests of linearity in univariate time series. They suggested comparing nonparametric estimates of the lagged conditional mean $M_k(x) = E(X_t|X_{t-k} = x)$ and least-squares estimates of the linear lag-k predictor $\theta_k x$ where $\theta_k = \{\mathrm{var}(X_{t-k})\}^{-1}\mathrm{cov}(X_t,X_{t-k})$, assuming the process has zero mean. If the process is linear, then these two predictors are the same for all lags. The concept of linearity used here is discussed in Hjellvik *et al.* (1998, p. 297), and we take the same pragmatic view in this paper. The tests have been shown to be able to detect general nonlinearities. In this section we try to establish similar tests for intercorrelated panel time series. This is a nontrivial task.

We present three versions of the tests. In Section 3.1 we exploit the conditioning argument of Section 2 and we construct an asymptotic theory for test functionals based on the conditional expectation $M_{k,t}(x) = E(X_{(i)t}|X_{(i)t-k},\boldsymbol{\eta})$. This test is designed to handle the cases where T is small, while the number of series n is large. Modifications, where we do not condition on $\boldsymbol{\eta}$, are given in Section 3.2. These modifications are more difficult to handle theoretically, but they work better in certain situations, primarily when n is small and T large.

3.1. *Test functionals for the conditional mean*

For our panel time series model (2), for fixed $t, t = 1, \ldots, T$, define

$$M_{k,t}(x) = E(X_{(i)t} | X_{(i)t-k} = x, \boldsymbol{\eta}).$$

Corresponding to the minimization in (9), let $\tilde{M}_{k,t}^{(v)}$ be the ℓth order ($\ell \geq v$) local polynomial estimator of the vth derivative of $M_{k,t}(\cdot)$. Specifically,

$$\begin{aligned}
&\left\{\tilde{M}_{k,t}(x), \tilde{M}_{k,t}^{(1)}(x), \ldots, \tilde{M}_{k,t}^{(\ell)}(x)\right\} \\
&\quad = \arg\min_{\gamma_0,\ldots,\gamma_\ell} \sum_{i=1}^{n} \left\{X_{(i)t} - \sum_{s=0}^{\ell} \frac{\gamma_s}{s!}(X_{(i)t-k} - x)^s\right\}^2 K\left(\frac{X_{(i)t-k} - x}{h}\right), \qquad (16)
\end{aligned}$$

where again K is a non-negative kernel function and h is the bandwidth. Let $\tilde{a}_{k,t}$ and $\tilde{b}_{k,t}$ be the estimated intercept and slope, respectively, of the global linear regression of $X_{(i)t}$ on $X_{(i)t-k}, i=1, \ldots, n$, i.e.

$$(\tilde{a}_{k,t}, \tilde{b}_{k,t}) = \arg\min_{a,b} \sum_{i=1}^{n} (X_{(i)t} - a - bX_{(i)t-k})^2.$$

We define, with a slight abuse of notation, the following statistics for testing linearity of the process. Let

$$\begin{aligned}
L_1(M_k) &= \frac{1}{T-k} \sum_{t=k+1}^{T} L_1(M_{k,t}), \\
L_1(M_{k,t}) &= \frac{1}{n} \sum_{i=1}^{n} \left\{\tilde{M}_{k,t}(X_{(i)t-k}) - \tilde{a}_{k,t} - \tilde{b}_{k,t} X_{(i)t-k}\right\}^2 w(X_{(i)t-k}), \\
L_1(M_k^{(1)}) &= \frac{1}{T-k} \sum_{t=k+1}^{T} L_1(M_{k,t}^{(1)}), \\
L_1(M_{k,t}^{(1)}) &= \frac{1}{n} \sum_{i=1}^{n} \left\{\tilde{M}_{k,t}^{(1)}(X_{(i)t-k}) - \tilde{b}_{k,t}\right\}^2 w(X_{(i)t-k}), \\
L_1(M_k^{(2)}) &= \frac{1}{T-k} \sum_{t=k+1}^{T} L_1(M_{k,t}^{(2)}), \\
L_1(M_{k,t}^{(2)}) &= \frac{1}{n} \sum_{i=1}^{n} \left\{\tilde{M}_{k,t}^{(2)}(X_{(i)t-k})\right\}^2 w(X_{(i)t-k}),
\end{aligned}$$

where w is a weight function. Here and elsewhere we have used the trapezium (cf. Hjellvik *et al.*, 1998)

$$w(x) = 1(|x - \bar{x}_0| \leq 2s_0) + (3 - |x - \bar{x}_0|/s_0)1(2s_0 < |x - \bar{x}_0| \leq 3s_0),$$

where $\bar{x}_0$ and s_0 are the empirical mean and standard deviation, respectively, of $\{X_{(i)t}, i = 1, \ldots, n, \ t = k+1, \ldots, T\}$. Although, for a fixed i, the empirical mean of data generated from a zero-mean model may be far from zero for T small, $X_{..} = n^{-1}T^{-1}\sum_i\sum_t X_{(i)t}$ is still the best estimate of $E(X_{(i)t})$, and, following Hjellvik *et al.* (1998) we therefore subtract $X_{..}$ from the data before performing the linearity test.

Theorems 6 and 7 then give the asymptotic behaviour of the test statistics under the null hypothesis and the alternative hypotheses, respectively. For reasons of simplicity (and space) we only state and prove the results for $\ell = 2$ in (16).

THEOREM 6. *Consider $\ell = 2$. Under assumptions* (B1)–(B4) *and* (D1)–(D5) *given in Appendix* B *and* D, *and the null hypothesis that $M_{k,t}(\cdot)$ is linear and the processes $\{X_{(i)t}\}$ are stationary, we have*

$$(i)\ L_1(M_k) \sim N\{B_{0,k}/(nh), s_{0,k}^2/(n^2h)\},$$
$$(ii)\ L_1(M_k^{(1)}) \sim N\{B_{1,k}/(nh^3), s_{1,k}^2/(n^2h^5)\},$$
$$(iii)\ L_1(M_k^{(2)}) \sim N\{B_{2,k}/(nh^5), s_{2,k}^2/(n^2h^9)\},$$

where

$$B_{0,k} = \frac{1}{(\mu_4 - \mu_2^2)^2}\sigma_k^2 \int (\mu_4 - \mu_2 u^2)^2 K^2(u)\mathrm{d}u \int w(x)\mathrm{d}x,$$
$$B_{1,k} = \frac{1}{\mu_2^2}\sigma_k^2 \int u^2 K^2(u)\mathrm{d}u \int w(x)\mathrm{d}x,$$
$$B_{2,k} = \frac{4}{(\mu_4 - \mu_2^2)^2}\sigma_k^2 \int (u^2 - \mu_2)^2 K^2(u)\mathrm{d}u \int w(x)\mathrm{d}x,$$
$$s_{0,k}^2 = \frac{2}{(T-k)(\mu_4 - \mu_2^2)^4}\sigma_k^4 \int w^2(x)\mathrm{d}x \int (\mu_4 - \mu_2 u^2)(\mu_4 - \mu_2 v^2)\{\mu_4 - \mu_2(u-z)^2\}$$
$$\times \{\mu_4 - \mu_2(v-z)^2\}K(u)K(v)K(u-z)K(u-v)\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}z,$$
$$s_{1,k}^2 = \frac{2}{(T-k)\mu_2^2}\sigma_k^4 \int w^2(x)\mathrm{d}x \int uv(u-z)(v-z)K(u)K(v)K(u-z)K(v-z)\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}z,$$
$$s_{2,k}^2 = \frac{32}{(T-k)(\mu_4 - \mu_2^2)^4}\sigma_k^4 \int w^2(x)\mathrm{d}x \int (u^2 - \mu_2)(v^2 - \mu_2)\{(u-z)^2 - \mu_2\}$$
$$\times \{(v-z)^2 - \mu_2\}K(u)K(v)K(u-z)K(v-z)\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}z,$$

where $\mu_i = \int u^i K(u)\mathrm{d}u$ and $\sigma_k^2 = \mathrm{var}(X_{(i)t} \mid X_{(i)t-k} = x, \boldsymbol{\eta})$, which is independent of x and $\boldsymbol{\eta}$.

Although a conditioning argument is used to prove Theorem 6 it should be noted that the asymptotic distribution does not depend on $\boldsymbol{\eta}$. Let Ω be the interval defined by Lemma B.2 in Appendix B. Again, let

$$M_{k,t} = E(X_{(i)t} \mid X_{(i)t-k} = x, \boldsymbol{\eta})$$

and let

$$b_{k,t} = \operatorname{var}(X_{(i)t-k} \mid \boldsymbol{\eta})^{-1} \operatorname{cov}(X_{(i)t}, X_{(i)t-k} \mid \boldsymbol{\eta})$$

and

$$a_{k,t} = E(X_{(i)t} \mid \boldsymbol{\eta}) - b_{k,t} E(X_{(i)t-k} \mid \boldsymbol{\eta}).$$

We consider the following local alternative hypotheses:

$$\begin{aligned}
H_a^{(0)}(k, \alpha_0): &\quad \int_\Omega \left\{ M_{k,t}(x) - a_{k,t} - b_{k,t} x \right\}^2 w(x) \mathrm{d}x > c n^{-\alpha_0}, \\
H_a^{(1)}(k, \alpha_1): &\quad \int_\Omega \left\{ M_{k,t}^{(1)}(x) - b_{k,t} \right\}^2 w(x) \mathrm{d}x > c n^{-\alpha_1}, \\
H_a^{(2)}(k, \alpha_2): &\quad \int_\Omega \left\{ M_{k,t}^{(2)}(x) \right\}^2 w(x) \mathrm{d}x > c n^{-\alpha_2},
\end{aligned}$$

uniformly for all sequence of $\{\eta_{-T_0}, \ldots, \eta_T\}$.

THEOREM 7. *Under assumptions* (B1)–(B4) *and* (D1)–(D5) *given in Appendix* B *and* D, *and the alternative hypotheses* $H_a^{(i)}(k, \alpha_i), i = 0, 1, 2,$ *if* $n^{\alpha_0} = o(nh)$, $n^{\alpha_1} = o(nh^3)$, $n^{\alpha_2} = o(nh^5)$, *we have*

$$\begin{aligned}
nhL_1(M_k) &\to \infty \quad \textit{in probability} \\
nh^3 L_1(M_k^{(1)}) &\to \infty \quad \textit{in probability} \\
nh^5 L_1(M_k^{(2)}) &\to \infty \quad \textit{in probability.}
\end{aligned}$$

Proofs of Theorems 6 and 7 are given in Appendix D.

3.2. *Two modifications*

Corresponding to (5) but minimizing over both n and t, define the local polynomial minimizer

$$\begin{aligned}
&\left\{ \hat{M}_k(x), \ldots, \hat{M}_k^{(\ell)}(x) \right\} \\
&\quad = \arg \min_{\gamma_0, \ldots, \gamma_\ell} \sum_{i=1}^{n} \sum_{t=k+1}^{T} \left\{ X_{(i)t} - \sum_{s=0}^{\ell} \frac{\gamma_s}{s!} (X_{(i)t-k} - x)^s \right\}^2 K_h(X_{(i)t-k} - x). \quad (17)
\end{aligned}$$

Since $M_k(x)$ is linear under the null hypothesis, the most natural choice of ℓ is perhaps $\ell = 1$. But since we need to evaluate $\hat{M}_k^{(\ell)}(x)$ for $\ell \leq 2$, and since $\ell = 2$ is not very natural under a linear null hypothesis, we have taken $\ell = 3$ in our simulations and for the real data. Taking $\ell = 3$ means that by letting $h \to \infty$ the local estimate will approach a global estimate of a linear function. Similarly, we define,

$$(\hat{a}_k, \hat{b}_k) = \arg\min_{a,b} \sum_{i=1}^{n} \sum_{t=k+1}^{T} \left(X_{(i)t} - a - bX_{(i)t-k}\right)^2,$$

obtaining the test functional

$$L_2(M_k) = \frac{1}{n(T-k)} \sum_{i=1}^{n} \sum_{t=k+1}^{T} \left\{\hat{M}_k(X_{(i)t}) - \hat{a}_k - \hat{b}_k X_{(i)t-k}\right\}^2 w(X_{(i)t}).$$

We can define $L_2(M_k^{(1)})$ and $L_2(M_k^{(2)})$ in the same way.

The functional $L_2(M_k)$ could be compared with the corresponding $L(M_k)$-functionals in Hjellvik *et al.* (1998), but it should be carefully noted that $\hat{M}_k$ is not in general a consistent estimator of $E(X_{(i)t}|X_{(i)t-k} = x)$. However, we have $\hat{M}_k(x) - \hat{a}_k - \hat{b}_k x \to 0$ in probability under fairly weak regularity conditions under the null hypothesis of a linear model, which serves as a justification for using it as a test functional.

A more direct modification of $L_1(M_k)$ of Section 3.1 is obtained by replacing $\tilde{M}_{k,t}(x)$ by a 'zero-mean intercept' alternative $\tilde{M}_{k,t}(x) - \tilde{a}_{k,t}$ and then taking a density-weighted average to obtain

$$\tilde{M}_k(x) = \frac{1}{T-k} \sum_{t=k+1}^{T} \{\tilde{M}_{k,t}(x) - \tilde{a}_{k,t}\} u_t(x)/u(x),$$

where

$$u_t(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h_t}(x - X_{(i)t}), \quad u(x) = \sum_{t=k+1}^{T} u_t(x), \qquad h_t = s_t h^{-1/5},$$

where s_t is the empirical standard deviation of $\{X_{(i)t}, i = 1,\ldots,n\}$. The motivation for taking a weighted average is that the estimates $\tilde{M}_{k,t}(x)$ may be concentrated at rather different x-values as t varies because of the different values of η_t being realized. To avoid very large variance in an x-region with few observations it seems reasonable to use a density-weighted average. We compare it with the average zero-intercept linear approximation

$$\tilde{b}_k x = (T-k)^{-1} \sum_{t=k+1}^{T} \tilde{b}_{k,t} x$$

to obtain the functional

$$L_3(M_k) = \frac{1}{n(T-k)} \sum_{i=1}^{n} \sum_{t=k+1}^{T} \{\tilde{M}_k(X_{(i)t}) - \tilde{b}_k X_{(i)t}\}^2 w(X_{(i)t}).$$

Stages in the construction of $\tilde{M}_1(x)$ and $\tilde{b}_1 x$ are shown in Figure 3 for an intercorrelated ($\sigma_\eta^2 = \sigma_\epsilon^2 = 0.5$) linear model [$a = 0.5$, $b = 0$ in (11)] and a corresponding nonlinear model [$a = 0.5$, $b = 2$ in (11)] with $T = 4$. The differences in vertical level for the three curves/lines in plots (a) and (b) are due

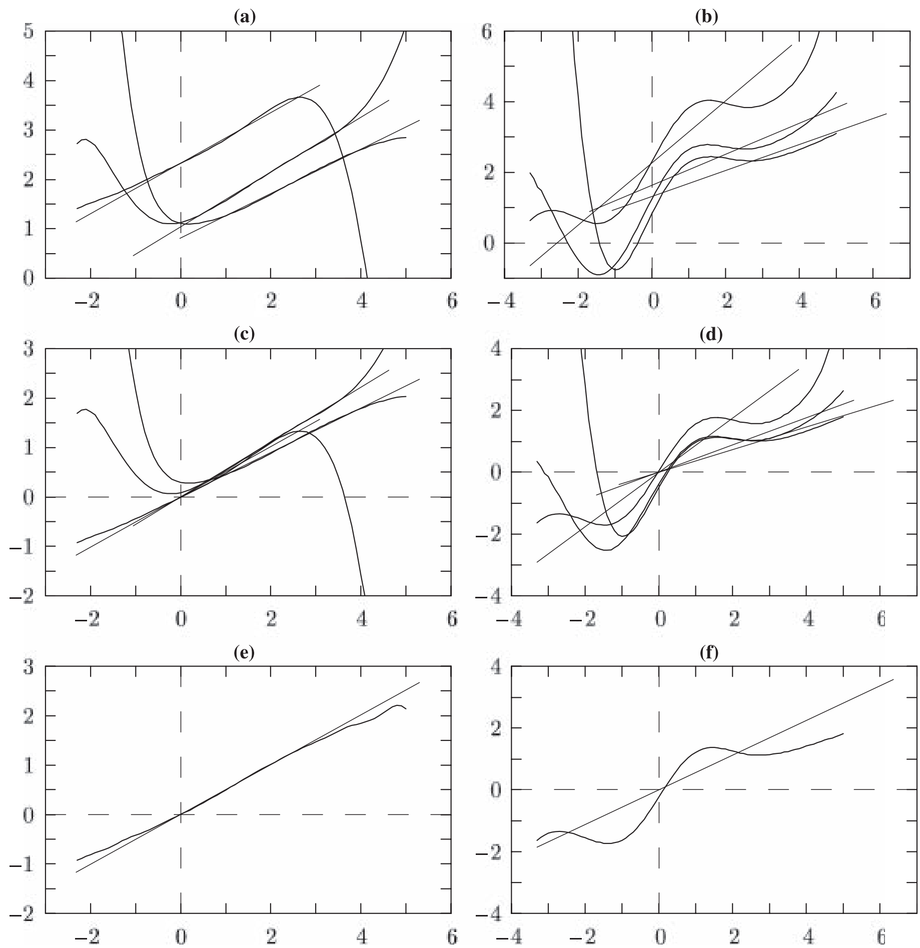

FIGURE 3. (a) $\tilde{M}_{1,t}(x), t = 1, 2, 3$ (thick lines) and $\tilde{a}_{1,t} + \tilde{b}_{1,t}x, t = 1, 2, 3$ (thin lines), plotted against x for one intercorrelated realization of model (11) with $T = 4$, $n = 1024$, $a = 0.5$ and $b = 0$. (c) $\tilde{M}_{1,t}(x) - \tilde{a}_{1,t}$ and $\tilde{b}_{1,t}x$. (e) Thick line: the density-weighted average $\tilde{M}_1(x)$ of the thick lines in (c). Thin line: the average $\tilde{b}_k x$ of the thin lines in (c). (b), (d), and (f). The same as (a), (c) and (e) for model (11) with $a = 0.5$ and $b = 2$.

to differences in η_2, η_3 and η_4. In Figure 4, $\tilde{M}_1(x)$ is plotted for various combinations of n and T for linear and nonlinear panels of intercorrelated series generated by (11).

All the functionals considered so far are for a specified lag. As in Hjellvik *et al.* (1998), functionals accumulated up to lag k can easily be constructed, e.g.

$$L_{s,\sup}(M_k^{(r)}) = \sup_{1 \le j \le k} L_s(M_j^{(r)}) \quad \text{and} \quad L_{s,\text{ave}}(M_k^{(r)}) = \frac{1}{k}\sum_{j=1}^{k} L_s(M_j^{(r)}),$$

for $s = 1, 2, 3$ and $r = 0, 1, 2$.

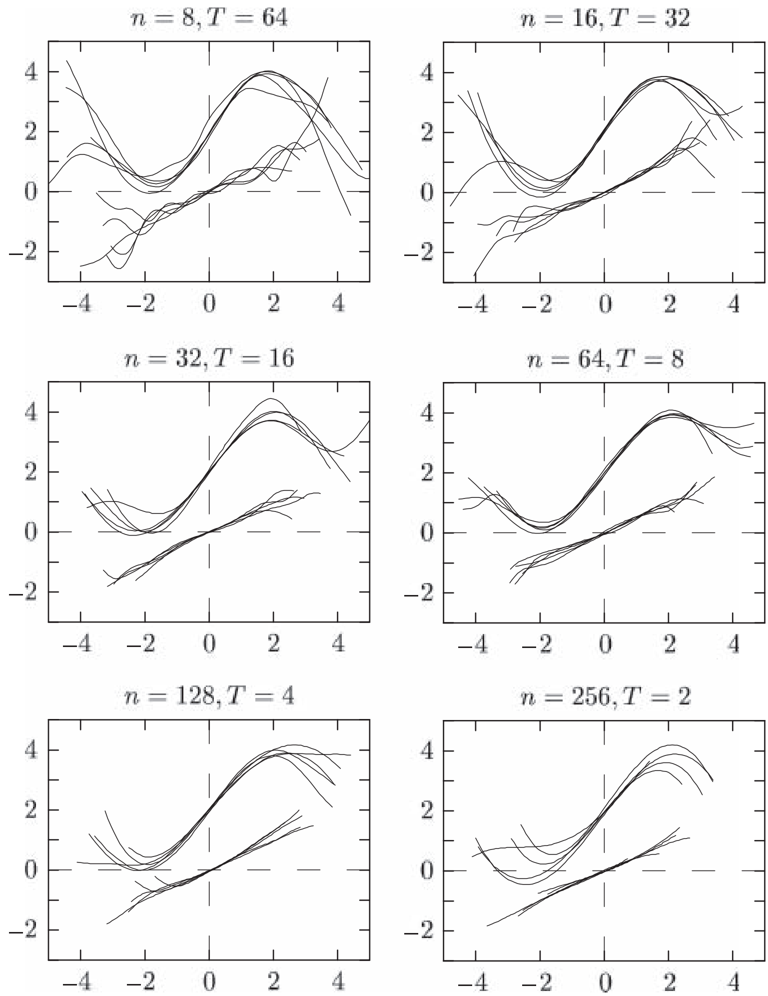

FIGURE 4. $\tilde{M}_1(x)$ plotted against x for five independent realizations of (11) with $a = 0.5$ and $b = 0$ and 2, and $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.5$. For $b = 2$, the curves are raised vertically by 2 units.

3.3. *The conditional variance*

When it comes to testing for homoscedasticity (i.e. constant conditional variance) for the ARCH-type models of Section 2.2, as in Hjellvik and Tjøstheim (1995, 1996) and Hjellvik *et al.* (1998), it is the conditional variance of the residual process from the best linear autoregressive fit to the data we are interested in, rather than of the series $\{X_{(i)t}\}$ itself; i.e. we assume that the conditional mean function is linear. To find the order p of the best autoregressive fit, we follow the procedure described in Hjellvik and Tjøstheim (1999a, 1999b). We assume an upper limit L of p, where $L = 1$ for $T \leq 4$ and $L = \min(10, T/2)$ for $T > 4$. We then choose the order p which minimizes the Final Prediction Error (FPE)-type criteria $\mathrm{FPE}_{\tilde{e},p}$ and $\mathrm{FPE}_{\tilde{\epsilon},p}$ defined by (5.2) and (5.3) in Hjellvik and Tjøstheim (1999b). For $T > 8$ we always use $\mathrm{FPE}_{\tilde{\epsilon},p}$, but for $T \leq 8$ we use $\mathrm{FPE}_{\tilde{e},p}$ if the intercorrelation ρ is estimated to be less than ρ_0 given by (4.9) in Hjellvik and Tjøstheim (1999b). The intercorrelation $\rho = \mathrm{corr}(X_{(i)t}, X_{(j)t})$ is estimated by

$$\hat{\rho} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{t=1}^{T} (X_{(i)t} - X_{(i)\cdot})(X_{(j)t} - X_{(j)\cdot})}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{ \sum_{t=1}^{T} (X_{(i)t} - X_{(i)\cdot})^2 \sum_{t=1}^{T} (X_{(j)t} - X_{(j)\cdot})^2 \right\}^{1/2}}$$

A conditional variance estimator can now be based on the local polynomial estimator defined in Sections 2.2.

For testing against heteroscedasticity, we define functionals similar to those of Sections 3.1 and 3.2, using $\tilde{e}^2_{(i)t}$, instead of $X_{(i)t}$. Here, $\tilde{e}_{(i)t}$ is defined by (14). Corresponding to (13), the $T \to \infty$, n fixed case, we define the conditional variance estimator

$$\hat{V}_k(x) = \frac{\sum_{i=1}^{n} \sum_{t=k+1}^{T} \tilde{e}^2_{(i)t} K_h(\tilde{e}_{(i)t-k} - x)}{\sum_{i=1}^{n} \sum_{t=k+1}^{T} K_h(\tilde{e}_{(i)t-k} - x)} - \left\{ \frac{\sum_{i=1}^{n} \sum_{t=k+1}^{T} \tilde{e}_{(i)t} K_h(\tilde{e}_{(i)t-k} - x)}{\sum_{i=1}^{n} \sum_{t=k+1}^{T} K_h(\tilde{e}_{(i)t-k} - x)} \right\}^2.$$

Using the reasoning of Section 2.2 we have that under the hypothesis of homoscedasticity, i.e. $g \equiv c$ in (3), if we have a linear AR(p) model, and if $\hat{p} = p$, then under relatively weak regularity conditions $\hat{V}_k(x) \to c$ in probability as $T \to \infty$ with n fixed.

Plots of $\hat{V}_1(x)$ for the AR(1) process obtained by taking $a = 0.5$ and $b = 0$ in (11) are shown in Figure 5. To get an idea of the power against ARCH-type alternatives, $\hat{V}_1(x)$ for the process

$$X_{(i)t} = 0.5X_{(i)t-1} + e_{(i)t}, \quad e_{(i)t} = z_{(i)t}(0.2 + 0.8e_{(i)t-1})^{1/2}, \qquad z_{(i)t} = \eta_t + \epsilon_{(i)t} \quad (18)$$

is included in the same figure. For both (11) and (18) we have taken $\sigma^2_\epsilon = \sigma^2_\eta = 0.5$. The bandwidth is $h = 2s\{n(T - \hat{p} - 1)\}^{-1/5}$, where s is the empirical standard deviation of $\{\tilde{e}_{(i)t}\}$. The estimated order of the AR approximation is greater than or equal to 1 in all cases. As can be expected (cf. Section 2.2), the behaviour of $\hat{V}_1(x)$ deteriorates as T decreases.

Homoscedasticity tests for the conditional variance can now be constructed in analogy with the conditional mean case: corresponding to $L_2(M_k)$ we introduce the test functional

$$L_2(V_k) = \frac{1}{n(T - \hat{p} - k)} \sum_{i=1}^{n} \sum_{t=1}^{T-\hat{p}-k} \left\{ \hat{V}_k(\tilde{e}_{(i)t}) - \tilde{\sigma}^2_{e,\hat{p}} \right\}^2 w(\tilde{e}_{(i)t}),$$

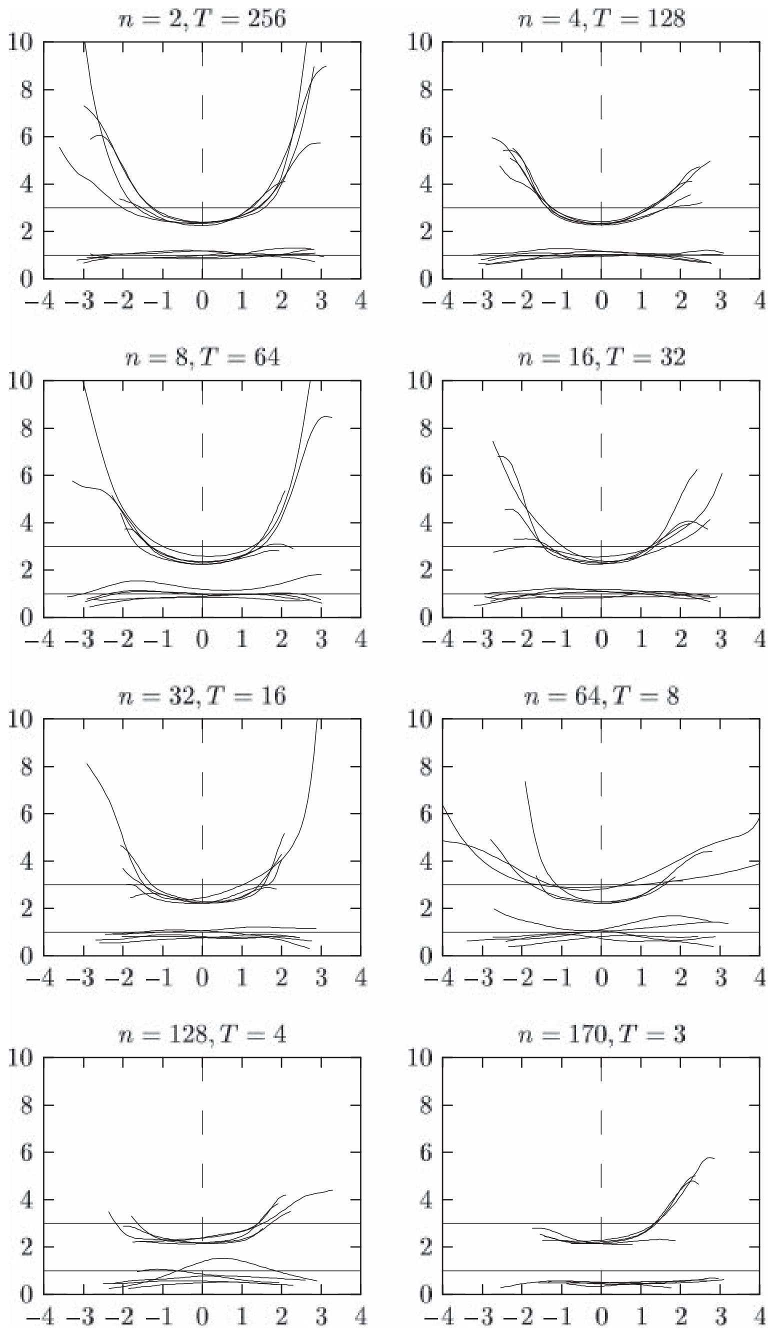

FIGURE 5. $\hat{V}_1(x)$ plotted against x for five independent realizations of (18) and (11) with $a = 0.5$ and $b = 0$. The straight lines indicate $\sigma_e^2 = 1$ in the homoscedastic model. For model (18) the curves are raised vertically by 2 units.

where $\hat{p}$ is the estimated order and $\tilde{\sigma}^2_{e,p}$ is the estimate defined in Hjellvik and Tjøstheim (1999b) of σ^2_e.

It is also possible to construct an analogue of $L_3(M_k)$, suitable for the $n \to \infty$ T-fixed case, by using the weighted average

$$\tilde{V}_k(x) = \frac{1}{T-k} \sum_{t=k+1}^{T} \tilde{V}_{k,t}(x) \frac{u_t(x)}{u(x)},$$

where

$$\tilde{V}_{k,t}(x) = \frac{\sum_{i=1}^{n} \tilde{e}^2_{(i)t} K_h(\tilde{e}_{(i)t-k} - x)}{\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t-k} - x)} - \left\{ \frac{\sum_{i=1}^{n} \tilde{e}_{(i)t} K_h(\tilde{e}_{(i)t-k} - x)}{\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t-k} - x)} \right\}^2.$$

It is easy to show that under the hypothesis of homoscedasticity, i.e. $g \equiv 1$ in (12), in the limit as $n \to \infty$, η_t drops out and

$$\tilde{V}_k(x) \to \sigma^2_\epsilon$$

in probability as $n \to \infty$, and corresponding to $L_3(M_k)$ one can introduce the functional

$$L_3(V_k) = \frac{1}{n(T-\hat{p}-k)} \sum_{i=1}^{n} \sum_{t=1}^{T-\hat{p}-k} \left\{ \tilde{V}_k(\tilde{e}_{(i)t}) - \tilde{\sigma}^2_{\epsilon,\hat{p}} \right\}^2 w(\tilde{e}_{(i)t}),$$

with $\tilde{\sigma}^2_{\epsilon,\hat{p}}$ as estimated in Hjellvik and Tjøstheim (1999b).

The more general test statistics $L_{2,\text{sup}}(V_k)$, $L_{3,\text{sup}}(V_k)$, $L_{2,\text{ave}}(V_k)$ and $L_{3,\text{ave}}(V_k)$ are constructed as for the conditional mean case.

4. BOOTSTRAPPING THE NULL DISTRIBUTION

The asymptotic theory of Section 3.1 requires very large sample sizes for it to be accurate. This is consistent with the results of Hjellvik and Tjøstheim (1995, 1996) and Hjellvik *et al.* (1998). We have therefore developed a bootstrap procedure which works much better for small and moderate sample sizes.

There are two possibilities for bootstrapping according to whether $\{\eta_t\}$ is considered to consist of i.i.d. random variables or not. If dependence is allowed for $\{\eta_t\}$, which is covered in our theory and which could well be of interest in practical situations, then we do not bootstrap η_t; i.e. one may use the following scheme: fit an autoregressive model

$$X_{(i)t} = \sum_{j=1}^{p} a_j X_{(i)t-j} + \eta_t + \epsilon_{(i)t} \tag{19}$$

to the data using the FPE (cf. Hjellvik and Tjøstheim, 1999b) to determine the order and the method developed in Hjellvik and Tjøstheim (1999a) to estimate the AR coefficients, Calculate

$$\tilde{\eta}_t = \frac{1}{n}\sum_{i=1}^{n}\tilde{e}_{(i)t} = \frac{1}{n}\sum_{i=1}^{n}\left(X_{(i)t} - \sum_{j=1}^{\hat{p}}\tilde{a}_j X_{(i)t-j}\right), \quad t = \hat{p}+1,\ldots,T \tag{20}$$

and

$$\tilde{\epsilon}_{(i)t} = X_{(i)t} - \sum_{j=1}^{\hat{p}}\tilde{a}_j X_{(i)t-j} - \tilde{\eta}_t.$$

We then bootstrap $\{\tilde{\epsilon}_{(i)t}\}$ to get a version $\{\epsilon^{\star}_{(i)t}\}$, and get a version of $X_{(i)t}$ by

$$X^{*}_{(i)t} = \sum_{j=1}^{\hat{p}}\tilde{a}_j X^{\star}_{(i)t-j} + \tilde{\eta}_t + \epsilon^{\star}_{(i)t}, \quad i = 1,\ldots,n, \qquad t = 1,\ldots,T,$$

without bootstrapping the $\tilde{\eta}_t$s.

Alternatively, if the η_ts are supposed to be i.i.d., then $\{\eta_t\}$ and $\{\epsilon_{(i)t}\}$ may be bootstrapped separately. We are generally able to recreate the intercorrelation structure with $\rho = \sigma^2_\eta/(\sigma^2_\epsilon + \sigma^2_\eta)$ (cf. Hjellvik and Tjøstheim, 1999a), but for n small and T small, some adjustments are needed. We have shown (Hjellvik and Tjøstheim, 1999b) that $\mathrm{var}(\tilde{\eta}_t) \approx \sigma^2_\eta + n^{-1}\sigma^2_\epsilon$ and $\mathrm{var}(\tilde{\epsilon}_{(i)t}) \approx \sigma^2_\epsilon(n-1)/n$ (if we replace $\tilde{e}_{(i)t}$ by $e_{(i)t}$ in the definitions of $\tilde{\eta}_t$ and $\tilde{\epsilon}_{(i)t}$, the variance expressions are exact), and this implies that the inter-individual correlation of the bootstrap replicas becomes too large when n is small. We therefore define

$$\begin{aligned} \tilde{\eta}'_t &= \begin{cases} \tilde{\eta}_t \frac{\tilde{\sigma}_{\eta,\hat{p}}}{s_\eta} & n < 10 \\ \tilde{\eta}_t, & n \geq 10 \end{cases} \\ \tilde{\epsilon}'_{(i)t} &= \tilde{\epsilon}_{(i)t}\left(\frac{n}{n-1}\right)^{1/2}, \quad t = \hat{p}+1,\ldots,T, \end{aligned} \tag{21}$$

where

$$s^2_\eta = (T-\hat{p})^{-1}\sum_t(\tilde{\eta}_t - \tilde{\eta}_.)^2, \quad \tilde{\eta}_. = (T-\hat{p})^{-1}\sum_t \eta_t, \quad \text{and} \quad \tilde{\sigma}^2_{\eta,\hat{p}} = s^2_\eta - n^{-1}\tilde{\sigma}^2_{\epsilon,\hat{p}}$$

as defined in Hjellvik and Tjøstheim (1999b). Then $\mathrm{var}(\tilde{\eta}'_t) \approx \sigma^2_\eta$ and $\mathrm{var}(\tilde{\epsilon}'_{(i)t}) \approx \sigma^2_\epsilon$ and the inter-individual correlation should be approximately the same for the bootstrap replicas as for the mother realization. This is confirmed for n small in simulation studies. For T small, a problem is that we only have $T - \hat{p}$ distinct $\tilde{\eta}_t$s. Especially for $T = 2$ and $\hat{p} = 1$, this makes it difficult to recreate the inter-individual correlation! But for a zero-mean process, $E(\eta_1 + \cdots + \eta_T) = 0$, and since we have subtracted the average $X_{..}$ to get zero-mean observations, we may, for $\hat{p} = 1$, utilize this to create one 'extra' $\tilde{\eta}_t$ by estimating η_1 by $\tilde{\eta}_1 = -(\tilde{\eta}_2 + \cdots + \tilde{\eta}_T)$. This leads to some improvement, but, as may be expected, such a small T is still a difficult matter.

5. EVALUATION OF FINITE SAMPLE PROPERTIES OF THE TESTS

We will examine the finite sample properties of the tests by simulation of the five linear models

(1) $X_{(i)t} = -0.5X_{(i)t-1} + e_{(i)t}$

(2) $X_{(i)t} = e_{(i)t}$

(3) $X_{(i)t} = 0.5X_{(i)t-1} + e_{(i)t}$

(4) $X_{(i)t} = 1.189X_{(i)t-1} - 0.249X_{(i)t-2} + 0.029X_{(i)t-3} - 0.137X_{(i)t-4} + e_{(i)t}$

(5) $X_{(i)t} = 1.063X_{(i)t-1} - 0.638X_{(i)t-2} + 0.218X_{(i)t-3} - 0.151X_{(i)t-4} - 0.014X_{(i)t-5} + 0.033X_{(i)t-6} - 0.241X_{(i)t-7} + 0.344X_{(i)t-8} + e_{(i)t}$

and the six nonlinear models

(6) $X_{(i)t} = \{0.5 + b\,\exp(-0.5X^2_{(i)t-1})\}X_{(i)t-1} + e_{(i)t}$

(7) $X_{(i)t} = (0.5 + be_{(i)t-1})X_{(i)t-1} + e_{(i)t}$

(8) $X_{(i)t} = bX_{(i)t-1}1(X_{(i)t-1} \leq 1) + 0.3X_{(i)t-1}1(X_{(i)t-1} > 1) + e_{(i)t}$

(9) $X_{(i)t} = (-bX_{(i)t-6} + bX_{(i)t-10})1(X_{(i)t-6} \leq 0) + 0.8X_{(i)t-10}1(X_{(i)t-6} > 0) + e_{(i)t}$

(10) $X_{(i)t} = \{0.4 + b\,\exp(-0.5X^2_{(i)t-6})\}X_{(i)t-6} + \{0.5 - 0.5\exp(-0.5X^2_{(i)t-10})\}X_{(i)t-10} + e_{(i)t}$

(11) $X_{(i)t} = e_{(i)t}\left(1 + bX^2_{t-1}\right)^{1/2}, b \geq 0$

where $e_{(i)t} = \eta_t + \epsilon_{(i)t}$ and $\{\eta_t\}$ and $\{\epsilon_{(i)t}\}$ are i.i.d. Gaussian zero-mean variables with variance σ^2_η and σ^2_ϵ, respectively. We have based our results for each example on 500 realizations and 40 bootstrap replicas for each of the realizations.

We have conducted separate simulation studies for the test functional L_1 of Section 3.1 and the alternative test functionals L_2 and L_3 of Section 3.2.

5.1. *The test functional* $L_1(M_k^{(\ell)})$

The bootstrap algorithm allowing for dependence in $\{\eta_t\}$ has been used. We study the size and the power of the test using models (1)–(3) and (6)–(8). The first lag conditional mean M_1, and variances $\sigma^2_\epsilon = \sigma^2_\eta = 0.5$ are used. For each model, the median of the optimal cross-validation bandwidth of 10 simulated series from the model is obtained and used for calculating the test statistics. Under the null hypothesis, we fit the best AR model using the FPE (Hjellvik and Tjøstheim, 1999b), thus obtaining the residuals $\{\tilde{\epsilon}_{(i)t}\}$ and the common random effects $\{\tilde{\eta}_t\}$. As mentioned in Section 4, only the residuals $\{\tilde{\epsilon}_{(i)t}\}$ are bootstrapped in generating bootstrap samples. The estimated $\tilde{\eta}_t$s are treated as fixed values. We have generated Gaussian zero-mean i.i.d. η_ts to facilitate the comparison with the other bootstrap technique, but it should be realized that the bootstrap described in this sub-section can be carried through for a dependent sequence $\{\eta_t\}$. The nominal size is fixed at 0.05.

Table I shows the empirical size of the tests $L_1(M_1)$, $L_1(M_1^{(1)})$ and $L_1(M_1^{(2)})$ for models (1)–(3) with different sample sizes. As can be seen, the sizes of the tests are reasonable. Tables II–IV show the empirical power of the tests $L_1(M_1)$, $L_1(M_1^{(1)})$ and $L_1(M_1^{(2)})$ for models (6)–(8) with different sample sizes and different values of the coefficient b, which indicates how strong the nonlinearity is. We can see from Tables II–IV, that, in general, increasing n is more effective than increasing T. Note that $(n,T) = (256,3)$ gives the same effective sample size as $(n,T) = (128,5)$ and $(n,T) = (64,9)$. This is expected since the tests are designed for T small and n large. Moreover, as expected, as $|b|$ increases, the models exhibit stronger nonlinearity. Consequently, the tests have more power, except $L_1(M_1^{(1)})$ and $L_1(M_2^{(2)})$ for model (8), when $b = 0.4$. It is also noticed that for model (7) all of the tests have very little power for $T = 2$, no matter how large n is.

5.2. *The test functionals $L_2(M_k)$ and $L_3(M_k)$*

Here we bootstrap $\{\tilde{\eta}_t\}$ and $\{\tilde{\epsilon}_{(i)t}\}$ separately as in (21). The nominal size is 0.05 everywhere. The cubic spline algorithm described in Hjellvik and Tjøstheim (1995, Sect. 4, Remark 3) is used to increase computational speed.

Table V shows the empirical size of $L_2(M_1)$ and $L_3(M_1)$ (in parentheses) for various combinations of n and T. In the uncorrelated case ($\sigma_\eta^2 = 0$), the empirical size is lower than 0.060 in 107 of 115 cases. The highest empirical size is 0.068, obtained for model (1) with $(n,T) = (256,2)$. The clearly lowest empirical size in the uncorrelated case is 0.002 (0.004) obtained for the 8th order model (5) with $(n,T) = (128,4)$. Difficulties should be expected here since the highest order allowed in the linear approximation on which the null distribution is based, is 1 in this case. In the correlated case the average empirical size is a bit higher than in the uncorrelated case and higher than 0.060 in 26 cases. Here the 8th order model

TABLE I

THE EMPIRICAL SIZE OF $L_1(M_k)$ (LEFT), $L_1(M_k^{(1)})$ (MIDDLE), AND $L_1(M_k^{(2)})$ (RIGHT) FOR MODELS (1)–(3)

n	T	Model 1			Model 2			Model 3		
64	2	0.080	0.067	0.050	0.067	0.070	0.080	0.060	0.070	0.043
64	3	0.070	0.080	0.070	0.040	0.033	0.040	0.053	0.057	0.053
64	5	0.057	0.047	0.047	0.053	0.057	0.070	0.047	0.047	0.060
64	9	0.077	0.060	0.073	0.050	0.047	0.027	0.040	0.060	0.067
128	2	0.063	0.063	0.063	0.073	0.063	0.067	0.050	0.043	0.053
128	3	0.063	0.053	0.043	0.047	0.047	0.050	0.037	0.037	0.043
128	5	0.037	0.030	0.030	0.060	0.060	0.063	0.073	0.080	0.083
128	9	0.060	0.047	0.050	0.060	0.067	0.063	0.040	0.060	0.043
256	2	0.063	0.063	0.060	0.037	0.050	0.040	0.033	0.073	0.040
256	3	0.053	0.053	0.040	0.060	0.050	0.050	0.057	0.057	0.070
256	5	0.057	0.057	0.070	0.047	0.037	0.037	0.060	0.050	0.083
256	9	0.043	0.060	0.050	0.060	0.053	0.060	0.033	0.060	0.057

TABLE II

The Empirical Power of $L_1(M_k)$ (left), $L_1(M_k^{(1)})$ (middle), and $L_1(M_k^{(2)})$ (right) for Model (6), with Different Values of b

		Model 6											
n	T	$b = 0.2$			$b = 0.4$			$b = 0.6$			$b = 0.8$		
64	2	0.073	0.080	0.090	0.107	0.093	0.110	0.233	0.200	0.253	0.413	0.203	0.303
64	3	0.080	0.083	0.080	0.163	0.113	0.150	0.260	0.213	0.320	0.527	0.263	0.367
64	5	0.087	0.093	0.087	0.200	0.130	0.180	0.440	0.370	0.460	0.667	0.293	0.527
64	9	0.083	0.057	0.063	0.267	0.143	0.187	0.600	0.563	0.717	0.760	0.430	0.690
128	2	0.083	0.050	0.057	0.217	0.167	0.250	0.417	0.217	0.290	0.670	0.380	0.537
128	3	0.110	0.110	0.113	0.287	0.237	0.347	0.567	0.277	0.413	0.813	0.517	0.703
128	5	0.103	0.073	0.090	0.317	0.273	0.457	0.777	0.370	0.597	0.890	0.680	0.887
128	9	0.137	0.093	0.117	0.523	0.480	0.660	0.863	0.537	0.780	0.923	0.857	0.943
256	2	0.150	0.143	0.143	0.327	0.347	0.433	0.657	0.563	0.717	0.903	0.717	0.810
256	3	0.130	0.130	0.133	0.453	0.460	0.607	0.887	0.793	0.927	0.957	0.863	0.930
256	5	0.210	0.227	0.263	0.653	0.673	0.793	0.950	0.917	0.983	0.963	0.940	0.977
256	9	0.280	0.313	0.327	0.820	0.823	0.947	0.963	0.957	0.987	0.980	0.957	0.990

TABLE III

The Empirical Power of $L_1(M_k)$ (left), $L_1(M_k^{(1)})$ (middle), and $L_1(M_k^{(2)})$ (right) for Model (7), with Different Values of b

		Model 7											
n	T	$b = 0.1$			$b = 0.2$			$b = 0.3$			$b = 0.4$		
64	2	0.063	0.053	0.043	0.053	0.077	0.043	0.083	0.083	0.070	0.120	0.070	0.050
64	3	0.103	0.093	0.093	0.323	0.163	0.170	0.470	0.290	0.347	0.603	0.420	0.450
64	5	0.230	0.190	0.177	0.567	0.327	0.350	0.833	0.757	0.780	0.877	0.773	0.787
64	9	0.317	0.287	0.270	0.753	0.477	0.503	0.937	0.930	0.957	0.967	0.867	0.903
128	2	0.073	0.057	0.060	0.097	0.097	0.090	0.077	0.063	0.060	0.110	0.083	0.063
128	3	0.177	0.160	0.163	0.597	0.400	0.440	0.827	0.830	0.820	0.787	0.663	0.750
128	5	0.407	0.293	0.337	0.900	0.783	0.840	0.987	0.987	0.997	0.977	0.953	0.973
128	9	0.583	0.453	0.577	0.967	0.960	0.983	0.997	1.00	1.00	0.993	0.983	0.990
256	2	0.070	0.070	0.067	0.073	0.077	0.080	0.107	0.100	0.080	0.130	0.087	0.047
256	3	0.437	0.390	0.410	0.887	0.863	0.907	0.953	0.923	0.943	0.947	0.913	0.927
256	5	0.790	0.770	0.790	0.997	1.00	1.00	0.993	0.997	0.997	0.990	0.977	0.987
256	9	0.953	0.970	0.980	1.00	1.00	1.00	0.997	0.997	1.00	1.00	1.00	1.00

(5) represents both the highest (0.164) and the lowest (0.014) empirical size. The lowest occurs as in the uncorrelated case for $T = 4$, whereas the highest occurs for $T = 8$ where the highest order allowed in the linear approximation is 4. Similar comments can be made for model (4). For $T = 2$, however, the empirical size is quite close to the nominal size for models (4) and (5) both in the correlated and the uncorrelated case. One possible explanation is the definition of $\tilde{\eta}_1$ for $\hat{p} = 1$ in Section 4, which yields $\tilde{\eta}_1 = -\tilde{\eta}_2$ for $T = 2$.

Table VI shows the empirical power for $L_2(M_1)$ and $L_3(M_1)$ for models (6)–(8) and for $L_2(M_6)$ and $L_3(M_6)$ for models (9)–(10). The nonlinearity parameter b is chosen to get roughly the same empirical power for the different models and sample sizes. For $nT = 512, b = 0.6, 0.1, -0.1, 0.5$ and -0.7 and for $nT = 128, b = 1.3, 0.5, -0.4$, 0.7 and -2.0 for models (6), (7), (8), (9), (10) respectively. We make the following observations: In most cases: (i) $L_2(\cdot)$ yields higher power in the

TABLE IV

THE EMPIRICAL POWER OF $L_1(M_k)$ (LEFT), $L_1(M_k^{(1)})$ (MIDDLE), AND $L_1(M_k^{(2)})$ (RIGHT) FOR MODEL (8), WITH DIFFERENT VALUES OF b

		Model 8											
n	T	$b = -0.2$			$b = -0.4$			$b = -0.6$			$b = -0.8$		
64	2	0.207	0.093	0.130	0.340	0.123	0.227	0.497	0.347	0.517	0.673	0.220	0.283
64	3	0.247	0.113	0.133	0.390	0.150	0.177	0.670	0.437	0.627	0.807	0.170	0.270
64	5	0.260	0.107	0.150	0.513	0.150	0.187	0.787	0.547	0.813	0.860	0.260	0.377
64	9	0.257	0.070	0.130	0.610	0.163	0.280	0.863	0.693	0.937	0.850	0.253	0.427
128	2	0.313	0.193	0.263	0.507	0.287	0.430	0.657	0.173	0.217	0.773	0.350	0.400
128	3	0.403	0.217	0.330	0.613	0.337	0.507	0.840	0.190	0.187	0.933	0.417	0.470
128	5	0.547	0.273	0.467	0.827	0.473	0.800	0.937	0.217	0.233	0.957	0.500	0.573
128	9	0.727	0.420	0.660	0.870	0.657	0.877	0.890	0.240	0.273	0.953	0.617	0.673
256	2	0.520	0.420	0.553	0.647	0.487	0.580	0.727	0.417	0.413	0.793	0.273	0.180
256	3	0.623	0.523	0.677	0.840	0.653	0.800	0.903	0.517	0.513	0.953	0.223	0.123
256	5	0.837	0.743	0.900	0.933	0.800	0.930	0.967	0.607	0.580	0.977	0.180	0.067
256	9	0.913	0.883	0.987	0.980	0.943	0.987	0.963	0.740	0.733	0.960	0.233	0.093

uncorrelated case than in the correlated case, (ii) $L_3(\cdot)$ yields higher power in the correlated case than in the uncorrelated case, and (iii) $L_3(\cdot)$ yields higher power than $L_2(\cdot)$ in the correlated case and lower power than $L_2(\cdot)$ in the uncorrelated case. As expected, $L_2(\cdot)$ performs best for a large T (cf. Section 3.2). As a result of the double sum in (17) it is also fairly stable over a wide range of different combinations of n and T, whereas both $L_1(\cdot)$ and $L_3(\cdot)$ are unstable for a small n since then the minimization in (16) is based on a smaller number of terms.

Figures 6, 7 show the empirical size and power of $L_{2,\sup}(M_k)$ and $L_{3,\sup}(M_k)$ for $k = 1,\ldots,\min(10,T-1)$ and $nT = 128$. As can be seen, the size is acceptable for models (1)–(3) both in the uncorrelated and the correlated case, whereas for models (4) and (5) it is more variable, especially in the correlated case. For $T = 8$ and $T = 16$, the empirical size of $L_{3,\sup}(M_k)$ is closer to 0.05 than that of $L_{2,\sup}(M_k)$ in most cases.

The models (9) and (10) have their nonlinearities defined in terms of lag 6 and 10, and the rise in power for lag 6 for these models is consistent with this.

The empirical power drops quicker as k increases for n large than for n small. This is expected because as we increase k with one unit we get n observations less to calculate $\hat{M}_k(x)$ from.

The statistics $L_{2,\mathrm{ave}}(M_k)$ and $L_{3,\mathrm{ave}}(M_k)$ behave approximately as $L_{2,\sup}(M_k)$ and $L_{3,\sup}(M_k)$ for the linear models, but, as should be expected, for models (7)–(10) the power of $L_{2,\mathrm{ave}}(M_k)$ and $L_{3,\mathrm{ave}}(M_k)$ drops faster than for $L_{2,\sup}(M_k)$ and $L_{3,\sup}(M_k)$, as k increases.

5.3. *The test functionals $L_2(V_k)$ and $L_3(V_k)$*

Table VII shows the empirical size and power of $L_2(V_1)$ and $L_3(V_1)$. Since the tests are based on the residual process, we have not included models (1) and (2)

TABLE V

THE EMPIRICAL SIZE OF $L_2(M_1)$ AND $L_3(M_1)$ (IN PARENTHESIS) FOR MODELS (1)–(5) OF SECTION 5.1

		Models				
n	T	(1)	(2)	(3)	(4)	(5)
Linear models, $\sigma_\epsilon^2 = 1, \sigma_\eta^2 = 0$						
1	512	0.048	0.038	0.046	0.044	0.062
2	256	0.058	0.044	0.044	0.032	0.058
4	128	0.040	0.058	0.050	0.052	0.042
8	64	0.042	0.030	0.052	0.050	0.056
16	32	0.036 (0.046)	0.048 (0.042)	0.044 (0.058)	0.054 (0.032)	0.038 (0.044)
32	16	0.050 (0.046)	0.038 (0.064)	0.058 (0.062)	0.036 (0.052)	0.054 (0.050)
64	8	0.046 (0.044)	0.048 (0.054)	0.054 (0.042)	0.044 (0.048)	0.046 (0.040)
128	4	0.042 (0.046)	0.052 (0.058)	0.062 (0.054)	0.038 (0.048)	0.002 (0.004)
256	2	0.068	0.062	0.050	0.046	0.056
1	128	0.038	0.044	0.050	0.046	0.058
2	64	0.052	0.048	0.040	0.042	0.046
4	32	0.038	0.030	0.050	0.034	0.054
8	16	0.040 (0.046)	0.034 (0.042)	0.052 (0.050)	0.042 (0.036)	0.034 (0.058)
16	8	0.040 (0.066)	0.030 (0.040)	0.038 (0.048)	0.042 (0.040)	0.050 (0.048)
32	4	0.046 (0.060)	0.038 (0.046)	0.054 (0.044)	0.044 (0.050)	0.010 (0.020)
64	2	0.042	0.038	0.042	0.034	0.054
Linear models, $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.5$						
2	256	0.046	0.050	0.042	0.042	0.050
4	128	0.042	0.052	0.042	0.052	0.048
8	64	0.058	0.044	0.058	0.048	0.064
16	32	0.030 (0.060)	0.036 (0.048)	0.038 (0.056)	0.036 (0.046)	0.054 (0.044)
32	16	0.054 (0.064)	0.044 (0.052)	0.052 (0.052)	0.070 (0.056)	0.090 (0.062)
64	8	0.060 (0.074)	0.066 (0.056)	0.058 (0.064)	0.096 (0.078)	0.164 (0.126)
128	4	0.048 (0.064)	0.080 (0.072)	0.054 (0.080)	0.028 (0.056)	0.014 (0.040)
256	2	0.072	0.048	0.046	0.044	0.050
2	64	0.038	0.044	0.058	0.036	0.040
4	32	0.044	0.052	0.036	0.048	0.028
8	16	0.048 (0.058)	0.048 (0.032)	0.060 (0.040)	0.066 (0.070)	0.076 (0.070)
16	8	0.024 (0.038)	0.050 (0.058)	0.034 (0.038)	0.068 (0.074)	0.116 (0.086)
32	4	0.050 (0.060)	0.058 (0.058)	0.038 (0.046)	0.018 (0.048)	0.016 (0.040)
64	2	0.042	0.048	0.038	0.040	0.048

which would yield results similar to those of model (3). For the power results the nonlinear models have $b = 0.3$ for both model (7) and (11) with $nT = 512$, and for $nT = 128$, $b = 0.5$ and 0.7 for model (7) and (11), respectively. As can be seen, the empirical size is too high for the 4th-and the 8th-order model with $T = 4$, but in all the other cases the size is acceptable. There is power against model (7) and (11) for a fairly wide range of combinations of n and T.

6. A REAL-DATA EXAMPLE

We end by taking a look at a real-data example discussed in Bjørnstad *et al.* (1996) and Stenseth *et al.* (1996). The data set contains the logarithms of the yearly catch of grey-sided voles over a period of $T = 31$ years at 91 different

TABLE VI

THE EMPIRICAL SIZE OF $L_2(M_1)$ AND $L_3(M_1)$ (IN PARENTHESIS) FOR MODELS (6)–(8) OF SECTION 5.1 AND OF $L_2(M_6)$ AND $L_3(M_6)$ (IN PARENTHESIS) FOR MODELS (9)–(10) OF SECTION 5.1

		Models				
n	T	(6)	(7)	(8)	(9)	(10)
Nonlinear models, $\sigma_\epsilon^2 = 1, \sigma_\eta^2 = 0$						
1	512	0.892	0.772	0.836	0.762	0.810
2	256	0.876	0.792	0.842	0.784	0.792
4	128	0.876	0.784	0.882	0.760	0.788
8	64	0.868	0.774	0.876	0.764	0.778
16	32	0.826 (0.770)	0.746 (0.322)	0.850 (0.626)	0.716 (0.458)	0.718 (0.618)
32	16	0.888 (0.836)	0.750 (0.508)	0.838 (0.756)	0.772 (0.620)	0.622 (0.660)
64	8	0.780 (0.786)	0.690 (0.580)	0.782 (0.770)	0.358 (0.324)	0.282 (0.288)
128	4	0.776 (0.768)	0.600 (0.538)	0.688 (0.678)		
256	2	0.534	0.426	0.526		
1	128	0.890	0.956	0.724	0.482	0.950
2	64	0.836	0.942	0.682	0.494	0.918
4	32	0.832	0.924	0.692	0.458	0.872
8	16	0.802 (0.408)	0.880 (0.132)	0.646 (0.138)	0.682 (0.102)	0.840 (0.266)
16	8	0.804 (0.774)	0.872 (0.598)	0.632 (0.452)	0.182 (0.096)	0.440 (0.310)
32	4	0.776 (0.768)	0.832 (0.740)	0.538 (0.508)		
64	2	0.630	0.676	0.350		
Nonlinear models, $\sigma_\epsilon^2 = \sigma_\eta^2 = 0.5$						
2	256	0.870	0.730	0.810	0.692	0.778
4	128	0.804	0.654	0.762	0.640	0.700
8	64	0.794	0.568	0.672	0.526	0.638
16	32	0.610 (0.918)	0.412 (0.638)	0.544 (0.898)	0.330 (0.630)	0.438 (0.812)
32	16	0.560 (0.956)	0.362 (0.716)	0.450 (0.934)	0.456 (0.810)	0.382 (0.804)
64	8	0.510 (0.952)	0.272 (0.650)	0.372 (0.886)	0.240 (0.346)	0.338 (0.422)
128	4	0.582 (0.922)	0.278 (0.508)	0.430 (0.746)		
256	2	0.738	0.236	0.410		
2	64	0.832	0.902	0.632	0.430	0.896
4	32	0.830	0.884	0.614	0.380	0.850
8	16	0.764 (0.568)	0.804 (0.350)	0.570 (0.244)	0.576 (0.220)	0.822 (0.410)
16	8	0.710 (0.922)	0.718 (0.720)	0.466 (0.664)	0.148 (0.170)	0.534 (0.464)
32	4	0.748 (0.942)	0.636 (0.726)	0.468 (0.686)		
64	2	0.814	0.478	0.368		

locations of the island of Hokkaido (cf. Figure 1). They are distributed in three panels (groups), with $n = 16,41$ and 34 in group 1, 2 and 3, respectively. The intercorrelation is estimated to be $\hat{\rho} = 0.430, 0.437$ and 0.305, in group 1, 2 and 3, respectively. A more detailed presentation of parts of the data is given in Section 7 of Hjellvik and Tjøstheim (1999a) and in Section 6 of Hjellvik and Tjøstheim (1999b).

At this stage, we use this data set merely as an illustration. We are primarily interested in checking whether the data are nonlinearly generated or not. We do not attempt any biological or ecological interpretation of our results. For readers interested in some of the biological/ecological aspects of the data we refer to the two references mentioned in the beginning of this section and to the special issue of *Researches of Population Ecology* (vol. 40, no. 1, 1998) on the population ecology of the vole *Clethrionomys rufocanus*.

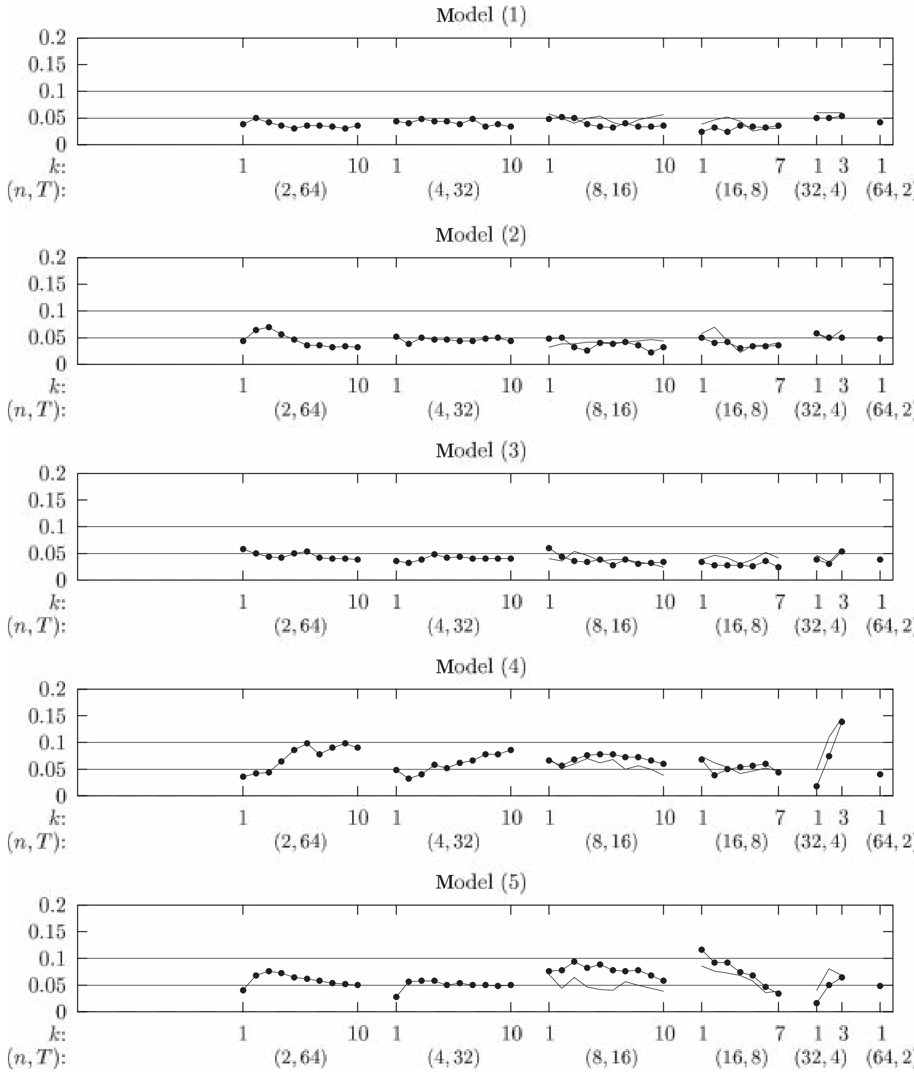

FIGURE 6. Empirical size of $L_{2,\sup}(M_k)$(lines with bullets) and $L_{3,\sup}(M_k)$ (clean lines) for models (1)–(5) of Section 5.1 with $\rho = 0.5$.

Figure 8a–f shows the p-values of $L_2(M_k)$, $L_3(M_k)$ (independence of $\{\eta_t\}$ cannot be ruled out; cf. Hjellvik and Tjøstheim, 1999a) and the corresponding cumulative statistics for $k = 1,\ldots,10$. We have used 500 bootstrap replicas and an adaptive kernel estimate with $\alpha = 0.5$ (cf. Silverman, 1986, p. 100 ff and Section 6.4 of Hjellvik *et al.*, 1998) to estimate the nulldistribution. For group 2, there are clear indications of a nonlinearity. The lowest p-value obtained for this group is 0.00042 for $L_{3,\text{ave}}(M_k)$. For group 1, the lowest p-value obtained for the

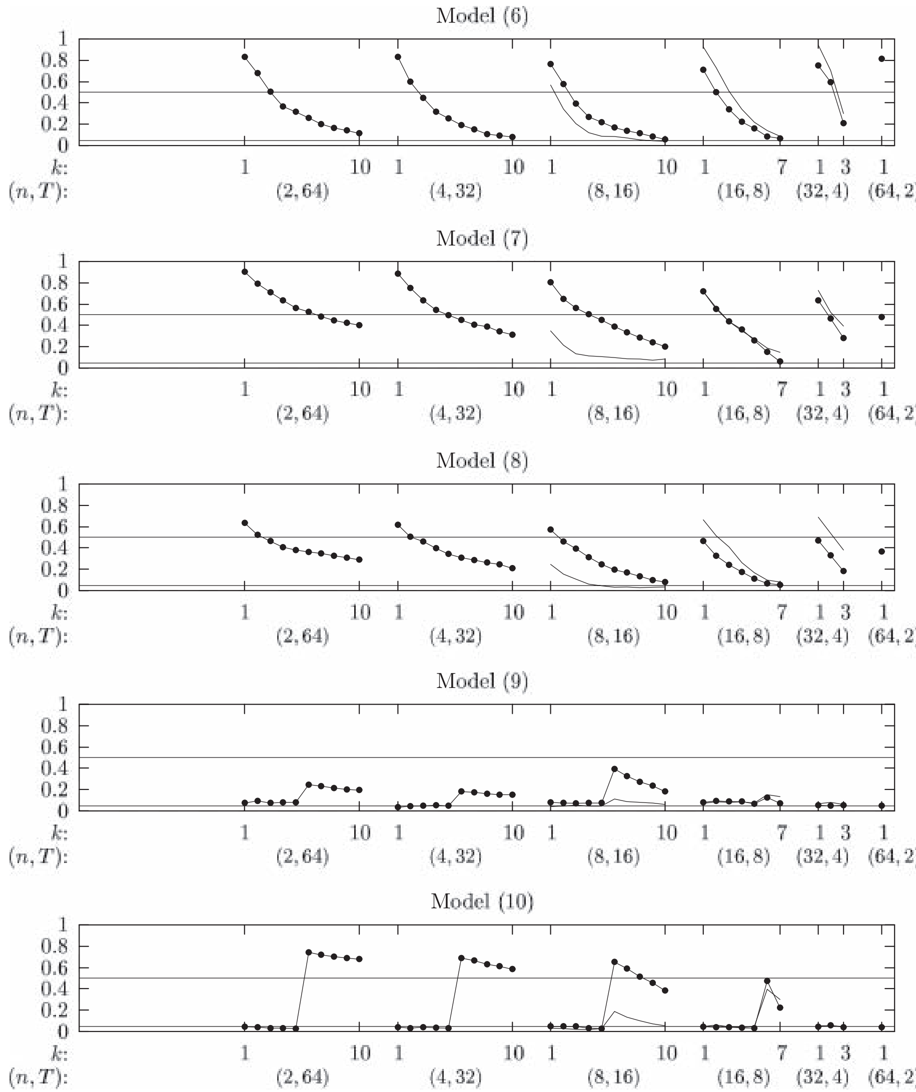

FIGURE 7. Empirical power of $L_{2,\sup}(M_k)$(lines with bullets) and $L_{3,\sup}(M_k)$ (clean lines) for models (6)–(10) of Section 5.1 with $\rho = 0.5$.

cumulative statistics is 0.048 for $L_{3,\text{ave}}(M_k)$, and for group 3, it is 0.020 for $L_{3,\sup}(M_k)$. It could be suspected that the difference in nonlinearity between the three groups is due to the varying numbers of observations, but if we split group 2 into three subgroups with $n = 14$, 14 and 13, respectively, the lowest p-values for the cumulative statistics are for the three subgroups 0.00051, 0.019 and 0.0010, respectively.

TABLE VII

THE EMPIRICAL SIZE AND POWER OF $L_2(V_1)$ AND $L_3(V_1)$ (IN PARENTHESIS) FOR MODELS (3)–(5), (7) AND (11) OF SECTION 5.1

		Models				
n	T	(3)	(4)	(5)	(7)	(11)
$\sigma_\epsilon^2 = 1, \sigma_\eta^2 = 0$						
1	512	0.052	0.064	0.048	0.762	0.964
2	256	0.062	0.064	0.052	0.766	0.952
4	128	0.040	0.036	0.058	0.752	0.962
8	64	0.060	0.050	0.048	0.738	0.954
16	32	0.036 (0.034)	0.042 (0.044)	0.028 (0.040)	0.654 (0.852)	0.958 (0.940)
32	16	0.044 (0.046)	0.050 (0.044)	0.042 (0.052)	0.650 (0.810)	0.930 (0.916)
64	8	0.048 (0.046)	0.044 (0.054)	0.040 (0.050)	0.520 (0.598)	0.892 (0.856)
128	4	0.062 (0.068)	0.218 (0.258)	0.470 (0.524)	0.354 (0.390)	0.824 (0.794)
170	3	(0.038)	(0.122)	(0.238)	(0.202)	(0.698)
1	128	0.060	0.066	0.058	0.844	0.888
2	64	0.044	0.052	0.052	0.802	0.912
4	32	0.046	0.050	0.050	0.796	0.878
8	16	0.040 (0.042)	0.036 (0.064)	0.050 (0.042)	0.690 (0.528)	0.832 (0.638)
16	8	0.046 (0.068)	0.052 (0.072)	0.030 (0.044)	0.480 (0.478)	0.718 (0.592)
32	4	0.042 (0.066)	0.070 (0.116)	0.092 (0.130)	0.332 (0.324)	0.610 (0.544)
42	3	(0.050)	(0.058)	(0.054)	(0.140)	(0.392)
$\sigma_\epsilon^2 = \sigma_\eta^2 = 0.5$						
2	256	0.036	0.030	0.040	0.754	0.958
4	128	0.046	0.048	0.040	0.712	0.950
8	64	0.056	0.050	0.048	0.548	0.858
16	32	0.046 (0.044)	0.040 (0.044)	0.030 (0.042)	0.414 (0.948)	0.712 (0.922)
32	16	0.030 (0.044)	0.038 (0.036)	0.028 (0.044)	0.310 (0.880)	0.518 (0.916)
64	8	0.034 (0.050)	0.036 (0.040)	0.028 (0.052)	0.218 (0.674)	0.312 (0.868)
128	4	0.036 (0.058)	0.124 (0.234)	0.242 (0.438)	0.162 (0.354)	0.220 (0.782)
170	3	(0.048)	(0.116)	(0.230)	(0.222)	(0.388)
2	64	0.044	0.058	0.050	0.816	0.890
4	32	0.042	0.052	0.054	0.742	0.866
8	16	0.050 (0.040)	0.052 (0.056)	0.030 (0.040)	0.576 (0.602)	0.746 (0.672)
16	8	0.032 (0.056)	0.038 (0.062)	0.024 (0.042)	0.420 (0.532)	0.494 (0.588)
32	4	0.028 (0.052)	0.060 (0.082)	0.086 (0.116)	0.238 (0.342)	0.330 (0.522)
42	3	(0.052)	(0.060)	(0.056)	(0.172)	(0.334)

Figure 9 shows the same as Figure 8 for the conditional variance. Using the cumulative statistics, group 2 is the only group for which the null hypothesis of homoscedasticity is rejected even with a size as high as 0.10. However, the rejection for this group may be due to nonlinearity for the conditional mean in this case.

APPENDIX A

PROOF OF THEOREM 1. The proof of the theorem follows that of Masry and Fan (1997) very closely. First, for a fixed series i, the solution $\hat{\boldsymbol{\gamma}}_i = \{\hat{\gamma}_0, \ldots, \hat{\gamma}_\ell\}^{\mathrm{T}}$ to (5) is

$$\hat{\boldsymbol{\gamma}}_i = (\mathbf{X}_i^{\mathrm{T}} \mathbf{W}_i \mathbf{X}_i)^{-1} \mathbf{X}_i^{\mathrm{T}} \mathbf{W}_i \mathbf{y}_i$$

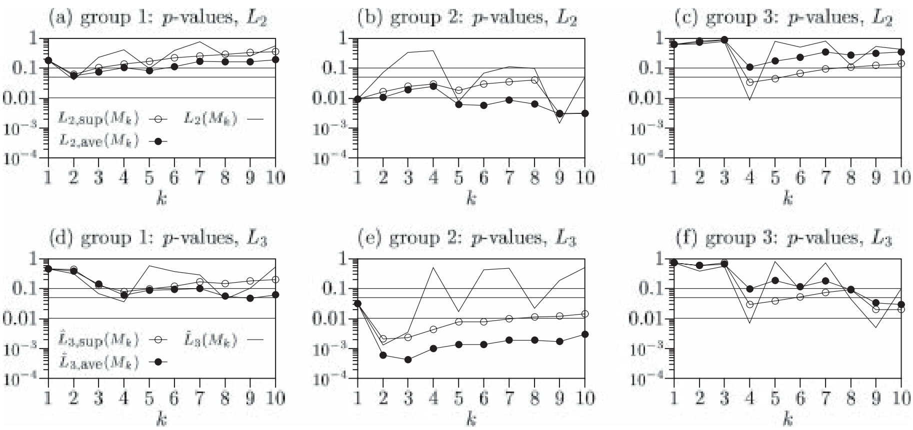

FIGURE 8. p-values of various tests based on the conditional mean for the grey-sided voles data.

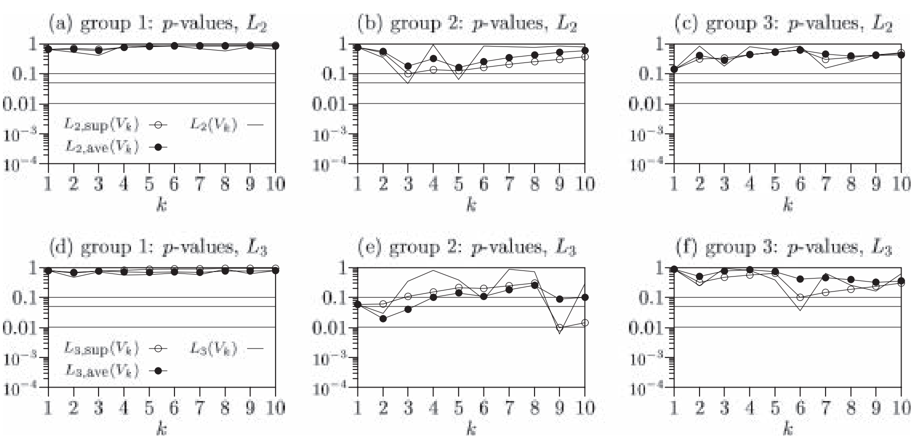

FIGURE 9. p-values of various tests based on the conditional variance for the grey sided voles data.

where $\mathbf{y}_i=\{X_{(i)2},\ldots,X_{(i)T}\}^{\mathrm{T}}, \mathbf{W}_i$ is the diagonal matrix the tth element of which is $K_h(X_{(i)t}-x)$, and

$$\mathbf{X}_i = \begin{pmatrix} 1 & (X_{(i)1}-x) & \ldots & (X_{(i)1}-x)^{\ell} \\ \vdots & \vdots & & \vdots \\ 1 & (X_{(i)T-1}-x) & \ldots & (X_{(i)T-1}-x)^{\ell} \end{pmatrix}.$$

Let

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}(x) = \{f(x),\ldots,f^{(\ell)}(x)/\ell!\}^{\mathrm{T}} \quad \text{and} \quad \mathbf{m}_i = \{f(X_{(i)1}),\ldots,f(X_{(i)T-1})\}^{\mathrm{T}}.$$

Following Masry and Fan (1997), write

$$\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma} = (\mathbf{X}_i^{\mathrm{T}}\mathbf{W}_i\mathbf{X}_i)^{-1}\mathbf{X}_i^{\mathrm{T}}\mathbf{W}_i\{\mathbf{m}_i - \mathbf{X}_i\boldsymbol{\gamma}(x)\} + (\mathbf{X}_i^{\mathrm{T}}\mathbf{W}_i\mathbf{X}_i)^{-1}\mathbf{X}_i^{\mathrm{T}}\mathbf{W}_i\{\mathbf{y}_i - \mathbf{m}_i\}$$
$$= \mathbf{b}_i + \mathbf{t}_i.$$

As $T \to \infty$,

$$\mathbf{b}_i = \frac{1}{(\ell+1)!} f^{(\ell+1)}(x) H^{-1} B h^{\ell+1}\{1 + o_p(1)\}$$

with $H = \mathrm{diag}(1,h,\ldots,h^\ell)$ and B as defined in (7), and

$$\mathbf{t}_i = p^{-1}(x) H^{-1} S^{-1}\mathbf{u}_i\{1 + o_p(1)\}$$

where $\mathbf{u}_i = (T-1)^{-1}H^{-1}\mathbf{X}_i^{\mathrm{T}}\mathbf{W}_i(\mathbf{y}_i - \mathbf{m}_i)$. Hence,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\gamma}}_i(x) - \boldsymbol{\gamma}(x)) &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{b}_i + \frac{1}{n}\sum_{i=1}^{n}\mathbf{t}_i \\ &= \frac{f^{(\ell+1)}(x)}{(\ell+1)!}H^{-1}Bh^{\ell+1}\{1 + o_p(1)\} \\ &\quad + p^{-1}(x)H^{-1}S^{-1}\mathbf{u}\{1 + o_p(1)\}\end{aligned}$$

where $\mathbf{u} = n^{-1}\sum_{i=1}^{n}\mathbf{u}_i$.

Again, following Masry and Fan (1997), we can show the asymptotic normality of $\mathbf{u}$, by considering an arbitrary linear combination $\mathbf{c}^{\mathrm{T}}\mathbf{u}$ with $\mathbf{c}^{\mathrm{T}} = [c_0, \ldots, c_\ell]$ and the representation

$$\mathbf{c}^{\mathrm{T}}\mathbf{u} = \frac{1}{T-1}\sum_{t=1}^{T-1} Z_t$$

where

$$Z_t = n^{-1}\sum_{i=1}^{n}\{X_{(i)t+1} - f(X_{(i)t})\}C_h(X_{(i)t} - x)$$

and $C(v) = \sum_{j=0}^{\ell} c_j v^j K(v)$ and $C_h(v) = C(v/h)/h$. The rest of the proof follows Masry and Fan (1997) except in calculating $\mathrm{var}(Z_t)$.

Specifically,

$$\begin{aligned}\mathrm{var}(Z_t) &= n^{-2}\,\mathrm{var}\left[\sum_{i=1}^{n}\epsilon_{(i)t}C_h(X_{(i)t} - x) + \eta_t\sum_{i=1}^{n}C_h(X_{(i)t} - x)\right] \\ &= n^{-1}\sigma_\epsilon^2\,\mathrm{var}[C_h(X_{(i)t} - x)] + n^{-2}\sigma_\eta^2\mathrm{var}[\sum_{i=1}^{n}C_h(X_{(i)t} - x)].\end{aligned}$$

It is known that

$$\mathrm{var}[C_h(X_{(i)t} - x)] = h^{-1}(p(x)\mathbf{c}^{\mathrm{T}}\tilde{S}\mathbf{c} + o(1)),$$

where $\tilde{S}$ is defined in (8). It is easy to show that $\mathrm{cov}[C_h(X_{(i)t}-x), C_h(X_{(j)t}-x)] = o(1)$. Note that n is a fixed finite number, hence we have

$$\operatorname{var}(Z_t) = \frac{1}{nh}\{(\sigma_\epsilon^2 + \sigma_\eta^2)p(x)\mathbf{c}^{\mathrm{T}}\tilde{S}\mathbf{c} + o(1)\}.$$

The rest of the proof follows exactly in the same line as that of Masry and Fan (1997), utilizing the mixing condition of the panel time series. QED

APPENDIX B

The following assumptions are used for Theorems 3 and 4.

(B1) The kernel function K is a symmetric density function with a bounded support in $\Re$, and $|K(x_1)-K(x_2)| < c|x_1-x_2|$ for all x_1 and x_2 in its support.
(B2) The function $f^{(\ell+1)}(x)$ exists and is continuous in the neighbourhood of x.
(B3) $h = O(n^{-1/(2p+3)})$.
(B4) For some small $\delta > 0, E(X_{(i)t}^{2+\delta}) < \infty$ for all t.
(B5) The distribution of η_t has bounded support and the distribution of $\epsilon_{(i)t}$ has infinite support.
(B6) The series starts at $t = -T_0$ with independent observations $X_{(1)-T_0}, X_{(2)-T_0}, \ldots$ across the panel.

Lemma B.1. *Under condition* (B6), *for fixed t and conditioning on* $\boldsymbol{\eta}$, $\{X_{(i)t}\}$ *and* $\{X_{(j)t}\}$ *are conditionally independent for* $i \neq j$.

Proof. By induction. QED

Lemma B.2. *Under assumptions* (B5) *and* (B6), *for any* $t < m$, (m *fixed*), *there is a compact interval in* $\Re$ *such that the conditional distribution* $p_t(x) = p_t(x \mid \boldsymbol{\eta})$ *of* $X_{(i)t}$ *given* $\boldsymbol{\eta}$ *is bounded below uniformly for x in the compact interval for any sequence of* $\eta_{-T_0}, \ldots, \eta_T$ *and all* $-T_0 \leq t \leq T$.

Proof. By construction. QED

Proof of Theorem 3. By Lemma B.1, we have that, conditioning on $\boldsymbol{\eta}, X_{(i)t}, i = 1, \ldots, n$ are conditionally independent for fixed t. Then by standard results (on independent observations) (Fan and Gijbels, 1996) and the fact that $f_t^{(\ell+1)}(x) = f^{(\ell+1)}(x)$, the theorem follows. QED

Proof of Theorem 4. First, let $\mathbf{y}_t = \{X_{(1)t}, \ldots, X_{(n)t}\}^{\mathrm{T}}$, let $\mathbf{W}_t(x)$ be the diagonal matrix the ith element of which is $K_h(X_{(i)t-1} - x)$, and

$$\mathbf{X}_t(x) = \begin{pmatrix} 1 & (X_{(1)t-1} - x) & \ldots & (X_{(1)t-1} - x)^\ell \\ \vdots & \vdots & & \vdots \\ 1 & (X_{(n)t-1} - x) & \ldots & (X_{(n)t-1} - x)^\ell \end{pmatrix}.$$

Then for fixed t, the solution $\hat{\boldsymbol{\gamma}}_t = \{\hat{\gamma}_0, \ldots, \hat{\gamma}_\ell\}^{\mathrm{T}}$ minimizing (9) is

$$\hat{\boldsymbol{\gamma}}_t(x) = (\mathbf{X}_t^{\mathrm{T}}(x)\mathbf{W}_t(x)\mathbf{X}_t(x))^{-1}\mathbf{X}_t^{\mathrm{T}}(x)\mathbf{W}_t(x)\mathbf{y_t}.$$

Let

$$\mathbf{m}_t = \{f_t(X_{(1)t}), \ldots, f_t(X_{(n)t})\}^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{\gamma}_t(x) = \{f_t(x), \ldots, f_t^{(\ell)}(x)/\ell!\}^{\mathrm{T}}.$$

Write

$$\begin{aligned}\hat{\boldsymbol{\gamma}}_t(x) - \boldsymbol{\gamma}_t(x) &= (\mathbf{X}_t^{\mathrm{T}}(x)\mathbf{W}_t(x)\mathbf{X}_t(x))^{-1}\mathbf{X}_t^{\mathrm{T}}(x)\mathbf{W}_t(x)\{\mathbf{m}_t - \mathbf{X}_t(x)\boldsymbol{\gamma}_t(x)\} \\ &\quad + (\mathbf{X}_t^{\mathrm{T}}(x)\mathbf{W}_t(x)\mathbf{X}_t(x))^{-1}\mathbf{X}_t^{\mathrm{T}}(x)\mathbf{W}_t(x)\{\mathbf{y}_t - \mathbf{m}_t\} \\ &= \mathbf{b}_t(x) + \mathbf{t}_t(x).\end{aligned}$$

As $n \to \infty$, following Masry and Fan (1997), and using that $f_t^{(s)}(x) = f^{(s)}(x)$ for $s > 0$, we have

$$\mathbf{b}_t(x) = \frac{1}{(\ell+1)!} f^{(\ell+1)}(x)H^{-1}Bh^{\ell+1}\{1 + o_p(1)\}$$

with $H = \mathrm{diag}(1,h,\ldots,h^\ell)$ and B as defined in (7). Moreover, writing $p_{t-1}(x)$ for $p_{t-1}(x|\boldsymbol{\eta})$,

$$\mathbf{t}_t(x) = p_{t-1}^{-1}(x)H^{-1}S^{-1}\mathbf{u}_t(x)\{1 + o_p(1)\}$$

where $\mathbf{u}_t(x) = n^{-1}H^{-1}\mathbf{X}_t^{\mathrm{T}}(x)\mathbf{W}_t(x)(\mathbf{y}_t - \mathbf{m}_t)$ and S is defined in (8).

Hence, letting $\mathbf{f}(x) = \{f(x), \ldots, f^{(\ell)}(x)/\ell!\}^{\mathrm{T}}$ and

$$\tilde{\mathbf{f}}(x) = \frac{1}{T-1}\sum_{t=2}^{T}(\hat{\boldsymbol{\gamma}}_t(x) - \hat{\boldsymbol{\gamma}}_t(0)),$$

$$\begin{aligned}\tilde{\mathbf{f}}(x) - \mathbf{f}(x) &= \frac{1}{T-1}\sum_{t=2}^{T}[\hat{\boldsymbol{\gamma}}_t(x) - \boldsymbol{\gamma}_t(x) - \{\hat{\boldsymbol{\gamma}}_t(0) - \boldsymbol{\gamma}_t(0)\}] \\ &= \frac{1}{T-1}\sum_{t=2}^{T}[\mathbf{b}_t(x) - \mathbf{b}_t(0)] + \frac{1}{T-1}\sum_{t=2}^{T}[\mathbf{t}_t(x) - \mathbf{t}_t(0)] \\ &= \frac{f^{\ell+1}(x) - f^{\ell+1}(0)}{(\ell+1)!}H^{-1}Bh^{\ell+1}\{1 + o_p(1)\} \\ &\quad + H^{-1}S^{-1}\mathbf{u}\{1 + o_p(1)\}\end{aligned}$$

where

$$\mathbf{u} = (T-1)^{-1}\sum_{t=2}^{T}\{\mathbf{u}_t(x)/p_{t-1}(x) - \mathbf{u}_t(0)/p_{t-1}(0)\}.$$

Reasoning as in the proof of Theorem 1 for an arbitrary linear combination $\mathbf{c}^{\mathrm{T}}\mathbf{u}$, and again, following Masry and Fan (1997), we can show the asymptotic normality of $\mathbf{u}$, with only the exception in calculation of $\mathrm{var}(Z_i|\boldsymbol{\eta})$, where

$$
\begin{aligned}
Z_i &= (T-1)^{-1}\sum_{t=2}^{T} p_{t-1}^{-1}(x)\{X_{(i)t} - f(X_{(i)t-1})\}C_h(X_{(i)t-1} - x) \\
&\quad - (T-1)^{-1}\sum_{t=2}^{T} p_{t-1}^{-1}(0)\{X_{(i)t} - f(X_{(i)t-1})\}C_h(X_{(i)t-1} - 0) \\
&= (T-1)^{-1}\sum_{t=2}^{T} \varepsilon_{(i)t}[p_{t-1}^{-1}(x)C_h(X_{(i)t-1} - x) - p_{t-1}^{-1}(0)C_h(X_{(i)t-1})]
\end{aligned}
$$

where

$$
C(v) = \sum_{j=0}^{\ell} c_j v^j K(v) \text{ and } C_h(v) = C(v/h)/h.
$$

Specifically,

$$
\begin{aligned}
\mathrm{var}(Z_i \mid \boldsymbol{\eta}) &= (T-1)^{-2}\mathrm{var}\left\{\sum_{t=2}^{T} \epsilon_{(i)t}[p_{t-1}^{-1}(x)C_h(X_{(i)t-1} - x) - p_{t-1}^{-1}(0)C_h(X_{(i)t-1} - 0)]\right\} \\
&= (T-1)^{-2}\sigma_\epsilon^2 \sum_{t=2}^{T} \mathrm{var}[p_{t-1}^{-1}(x)C_h(X_{(i)t-1} - x) - p_{t-1}^{-1}(0)C_h(X_{(i)t-1})].
\end{aligned}
$$

It is known that

$$
\mathrm{var}[C_h(X_{(i)t-1} - x)] = h^{-1}(p_{t-1}(x)\mathbf{c}^{\mathrm{T}}\tilde{S}\mathbf{c} + o(1))
$$

where $\tilde{S}$ is as in (8), and it is easy to show

$$
\mathrm{cov}[C_h(X_{(i)t-1} - x), C_h(X_{(i)t-1} - 0)] = o(1),
$$

under the condition that $C(v \to 0$ as $v \to \infty$. Note that T is a fixed finite number, hence we have

$$
\mathrm{var}(Z_i|\boldsymbol{\eta}) = \frac{1}{(T-1)^2 h}\left[\sigma_\epsilon^2 \sum_{t=1}^{T-1}\left\{\frac{1}{p_{t-1}(x)} + \frac{1}{p_{t-1}(0)}\right\}\mathbf{c}^{\mathrm{T}}\tilde{S}\mathbf{c} + o(1)\right].
$$

The rest of the proof follows exact the same line as that of Masry and Fan (1997). QED

APPENDIX C

Proof of Theorem 5. Since $\tilde{e}_{(i)t}$ is the residual of the least squares estimation of model (12), we have $\tilde{e}_{(i)t} = e_{(i)t} + \mathbf{X}_{(i)t-1}(\mathbf{a} - \tilde{\mathbf{a}})$ with $\mathbf{a}$ estimated as in Hjellvik and Tjøstheim (1999a) and where $\mathbf{X}_{(i)t} = (X_{(i)t},\ldots,X_{(i)t-p+1})$ and $\mathbf{a} = (a_1,\ldots,a_p)^{\mathrm{T}}$ and $\tilde{\mathbf{a}} - \mathbf{a} = O_p(T^{-1/2})$. (Note that n is fixed and $T \to \infty$.)

Then we have

$$
\hat{g}(x) - g(x) = \frac{(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t} - x)\{\tilde{e}_{(i)t+1}^2 - g(x)\}}{(nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t} - x)} \triangleq \frac{\mathrm{I}}{\mathrm{II}},
$$

where

$$
\begin{aligned}
\mathrm{II} &= (nT)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x) \\
&= \frac{1}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n} K\left\{\frac{e_{(i)t}-x}{h}+\frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right\} \\
&= \frac{1}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[K\left(\frac{e_{(i)t}-x}{h}\right)+K'(e^*_{(i)t})\left\{\frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right\}\right] \\
&\triangleq \mathrm{II}_1+\mathrm{II}_2,
\end{aligned}
$$

where K' is the first derivative of the kernel function K and $e^*_{(i)t}$ is between $(e_{(i)t}-x)/h$ and $(e_{(i)t}-x)/h+\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})/h$. By standard results, $\mathrm{II}_1 \to p_e(x)$ in probability under a strong mixing condition (cf. introductory remarks on geometric mixing and see e.g. Masry and Tjøstheim, 1995) and, with K' bounded,

$$
|\mathrm{II}_2| \le \frac{1}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n}|K'(e^*)|\frac{|\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})|}{h} \le O_p\left(\frac{|\mathbf{a}-\tilde{\mathbf{a}}|}{nTh^2}\right) = o_p(1).
$$

Hence $\mathrm{II} \to p_e(x)$ in probability.

For the numerator I, we have

$$
\begin{aligned}
\mathrm{I} &= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x)\left[\left\{e_{(i)t+1}+\mathbf{X}_{(i)t}(\mathbf{a}-\tilde{\mathbf{a}})\right\}^2 - g(x)\right] \\
&= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x)\{g(e_{(i)t})-g(x)\} \\
&\quad + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x)g(e_{(i)t})(\epsilon^2_{(i)t+1}-\sigma^2_\epsilon) \\
&\quad + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x)g(e_{(i)t})(\eta^2_{t+1}-\sigma^2_\eta) \\
&\quad + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x)g(e_{(i)t})(2\epsilon_{(i)t+1}\eta_{t+1}) \\
&\quad + \frac{2}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x)g^{1/2}(e_{(i)t})\epsilon_{(i)t+1}\mathbf{X}_{(i)t}(\mathbf{a}-\tilde{\mathbf{a}}) \\
&\quad + \frac{2}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x)g^{1/2}(e_{(i)t})\eta_{t+1}\mathbf{X}_{(i)t}(\mathbf{a}-\tilde{\mathbf{a}}) \\
&\quad + \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_h(\tilde{e}_{(i)t}-x)\{\mathbf{X}_{(i)t}(\mathbf{a}-\tilde{\mathbf{a}})\}^2 \\
&\triangleq \mathrm{I}_1+\mathrm{I}_2+\mathrm{I}_3+\mathrm{I}_4+\mathrm{I}_5+\mathrm{I}_6+\mathrm{I}_7.
\end{aligned}
$$

Now we study each I_i separately. First,

$$\begin{aligned}
\mathrm{I}_1 &= \frac{1}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n} K\left\{\frac{e_{(i)t}-x}{h}+\frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right\}\{g(e_{(i)t})-g(x)\} \\
&= \frac{1}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[K\left(\frac{e_{(i)t}-x}{h}\right)+K'\left(\frac{e_{(i)t}-x}{h}\right)\frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right. \\
&\quad \left.+\frac{1}{2}K''\left(\frac{e_{(i)t}-x}{h}\right)\frac{\{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})\}^2}{h^2}+\frac{1}{6}K'''(e^*_{(i)t})\frac{\{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})\}^3}{h^3}\right]\{g(e_{(i)t})-g(x)\} \\
&= \mathrm{I}_{11}+\mathrm{I}_{12}+\mathrm{I}_{13}+\mathrm{I}_{14},
\end{aligned}$$

where $e^*_{(i)t}$ lies between $(e_{(i)t}-x)/h$ and $(e_{(i)t}-x)/h+\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})/h$. By standard calculation,

$$\mathrm{I}_{11} = h^2B(x) + o_p(h^2),$$

where $B(x) = \mu_2\{0.5g''(x)p_e(x) + g'(x)p'_e(x)\}$. Under standard assumptions,

$$\begin{aligned}
n\mathrm{I}_{12} &\to \frac{1}{h}\left[\int K'(u)\mathbf{y}\{g(x+hu)-g(x)\}p_{e_{(i)t},X_{(i)t-1}}(x+hu,\mathbf{y})\mathrm{d}u\,\mathrm{d}y\right](\tilde{\mathbf{a}}-\mathbf{a}) \\
&= \left\{\int K'(u)u\,\mathrm{d}u\right\}g'(x)p_{e_{(i)t}}(x)E(\mathbf{X}_{(i)t-1}|e_{(i)t}=x)(\tilde{\mathbf{a}}-\mathbf{a}) \\
&= O_p(T^{-1/2}) = o_p(h^2).
\end{aligned}$$

Similarly, we can show that $\mathrm{I}_{13} = O_p(T^{-1}h^{-1}) = o_p(h^2)$. For I_{14}, we have

$$\begin{aligned}
|\mathrm{I}_{14}| &\le \frac{1}{6nTh}\sum_{t=1}^{T}\sum_{i=1}^{n}|K'''(e^*_{(i)t})|\left|\frac{\{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})\}^3}{h^3}\right| \\
&= O_p(T^{-3/2}h^{-4}) \\
&= o_p(h^2).
\end{aligned}$$

Hence, $\mathrm{I}_1 = h^2B(x)+o_p(h^2)$. Similarly,

$$\begin{aligned}
\mathrm{I}_2 &= \frac{1}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n} K\left\{\frac{e_{(i)t}-x}{h}+\frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right\}g(e_{(i)t})(\epsilon^2_{(i)t+1}-\sigma^2_\epsilon) \\
&= \frac{1}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n}\left[K\left(\frac{e_{(i)t}-x}{h}\right)+K'\left(\frac{e_{(i)t}-x}{h}\right)\frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right. \\
&\quad \left.+\frac{1}{2}K''\left(\frac{e_{(i)t}-x}{h}\right)\frac{\{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})\}^2}{h^2}+\frac{1}{6}K'''(e^*_{(i)t})\frac{\{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})\}^3}{h^3}\right] \\
&\quad \times g(e_{(i)t})(\epsilon^2_{(i)t+1}-\sigma^2_\epsilon) \\
&= \mathrm{I}_{21}+\mathrm{I}_{22}(\mathbf{a}-\tilde{\mathbf{a}})+(\mathbf{a}-\tilde{\mathbf{a}})'\mathrm{I}_{23}(\mathbf{a}-\tilde{\mathbf{a}})+\mathrm{I}_{24}.
\end{aligned}$$

Standard results yield that $\sqrt{nTh}\mathrm{I}_{21} \to N(0, s_1^2)$, where $s_1^2 = V_0g^2(x)\mathrm{var}(\epsilon^2_{(i)t})$. For I_{22}, we note that $E(\mathrm{I}_{22}) = 0$ and, because of the fact that $\epsilon_{(i)t+1}$ is independent of $X_{(i)s}$ and $\epsilon_{(i)s}$ for all $s \le t$, we have

$$\begin{aligned}
\operatorname{var}(\mathrm{I}_{22}) &= \frac{\operatorname{var}(\epsilon_{(i)t}^2)}{n^2T^2h^4} E\left\{\sum_{t=1}^{T}\sum_{i=1}^{n} K'^2\left(\frac{e_{(i)t}-x}{h}\right)\mathbf{X}_{(i)t-1}^{\mathrm{T}}\mathbf{X}_{(i)t-1} g^2(e_{(i)t})\right\} \\
&\to \frac{\operatorname{var}(\epsilon_{(i)t}^2)}{nTh^3}\int K'^2(u)g^2(x+hu)\mathbf{y}^{\mathrm{T}}\mathbf{y}p_{e_{(i)t},\mathbf{x}_{(i)t-1}}(x+hu,\mathbf{y})\mathrm{d}u\,\mathrm{d}\mathbf{y} \\
&= O(T^{-1}h^{-3}).
\end{aligned}$$

Hence $\mathrm{I}_{22}(\mathbf{a}-\tilde{\mathbf{a}}) = O_p(T^{-1}h^{-3/2}) = o_p(T^{-1/2}h^{1/2})$. Similarly, we can show that $(\mathbf{a}-\tilde{\mathbf{a}})'\mathrm{I}_{23}(\mathbf{a}-\tilde{\mathbf{a}}) = o_p(T^{-1/2}h^{1/2})$. As for I_{14}, we can show that $|\mathrm{I}_{24}| = o_p(h^2)$. Hence, $(nTh)^{1/2}\mathrm{I}_2 \to N(0, s_1^2)$. Analogous to the calculation of I_2 and that in the proof of Theorem 1, we obtain

$$\sqrt{nTh}\mathrm{I}_3 \to N(0, s_2^2) \quad \text{and} \quad \sqrt{nTh}\mathrm{I}_4 \to N(0, s_3^2),$$

where $s_2^2 = V_0 g^2(x)\operatorname{var}(\eta_t^2)$ and $s_3^2 = 4V_0 g^2(x)\sigma_\eta^2\sigma_\epsilon^2$. Moreover, it is easy to show that

$$\operatorname{cov}(\mathrm{I}_2,\mathrm{I}_3) = \operatorname{cov}(\mathrm{I}_2,\mathrm{I}_4) = \operatorname{cov}(\mathrm{I}_3,\mathrm{I}_4) = 0.$$

Hence,

$$\sqrt{nTh}(\mathrm{I}_2+\mathrm{I}_3+\mathrm{I}_4) \to N(0, s_1^2+s_2^2+s_3^2).$$

For I_5, we have

$$\begin{aligned}
\mathrm{I}_5 &= \frac{2}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n} K\left\{\frac{e_{(i)t}-x}{h} + \frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right\} g^{1/2}(e_{(i)t})\epsilon_{(i)t+1}\mathbf{X}_{(i)t}(\mathbf{a}-\tilde{\mathbf{a}}) \\
&= \frac{2}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n}\left\{K\left(\frac{e_{(i)t}-x}{h}\right) + K'(e_{(i)t}^*)\frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right\} \\
&\quad \times g^{1/2}(e_{(i)t})\epsilon_{(i)t+1}\mathbf{X}_{(i)t}(\mathbf{a}-\tilde{\mathbf{a}}) \\
&\stackrel{\triangle}{=} \mathrm{I}_{51}(\mathbf{a}-\tilde{\mathbf{a}}) + \mathrm{I}_{52},
\end{aligned}$$

where $E(\mathrm{I}_{51}) = 0$ and

$$\begin{aligned}
\operatorname{var}(\mathrm{I}_{51}) &= \frac{4\sigma_\epsilon^2}{n^2T^2h^2} E\left\{\sum_{t=1}^{T}\sum_{i=1}^{n} K^2\left(\frac{e_{(i)t}-x}{h}\right) g(e_{(i)t})\mathbf{X}_{(i)t}^{\mathrm{T}}\mathbf{X}_{(i)t}\right\} \\
&= O(T^{-1}h^{-2}).
\end{aligned}$$

Hence, $\mathrm{I}_{51}(\mathbf{a}-\tilde{\mathbf{a}}) = o_p(T^{-1/2}h^{1/2})$, and

$$|\mathrm{I}_{52}| \le \frac{2}{nTh}\sum_{t=1}^{T}\sum_{i=1}^{n}\left|K'(e_{(i)t}^*)\frac{\mathbf{X}_{(i)t-1}(\mathbf{a}-\tilde{\mathbf{a}})}{h}\right|\left|g^{1/2}(e_{(i)t})\epsilon_{(i)t+1}\mathbf{X}_{(i)t}(\mathbf{a}-\tilde{\mathbf{a}})\right| = O_p\left\{\frac{(\mathbf{a}-\tilde{\mathbf{a}})^2}{h^2}\right\} = o_p(h^2).$$

Hence, $\mathrm{I}_5 = o_p\{(Th)^{-1/2}\} + o_p(h^2)$. Similarly, $\mathrm{I}_6 = o_p\{(Th)^{-1/2}\} + o_p(h^2)$, and $|\mathrm{I}_7| = O_p\{(\mathbf{a}-\tilde{\mathbf{a}})^2h^{-1}\} = o_p(h^2)$. The theorem follows. QED

APPENDIX D

(D1) The series starts at $t = -T_0$. The starting observations $(X_{(i)-T_0},\ldots,X_{(i)-T_0+p})$ and $(X_{(j)-T_0},\ldots,X_{(j)-T_0+p})$ are independent for $i \neq j$; that is, the series are started independently.
(D2) For all t, $E(X_{(i)t}^{8+\delta}) < \infty$ for some small $\delta > 0$, and $E(X_{(i)t}^2 \mid X_{(i)t-k} = x)$ is a bounded function for all $k > 0$.
(D3) The joint density of $(X_{(i)t},X_{(i)t-k})$ is a bounded function for all $k > 0$.
(D4) As $n \to \infty$, then $h \to 0$ and $nh^{(2+4\delta)/(1+\delta)}/\log(n) \to \infty$ for some $\delta > 0$.
(D5) The weight function $w(\cdot)$ is continuous and with compact support in $\{x{:}p_t(x) > 0 \text{ for all } t\}$.

Lemma D.1. *Let X_t be a stationary and reversible linear AR process*

$$X_t = c + a_1X_{t-1} + \cdots + a_pX_{t-p} + \eta_t + \epsilon_t$$

where $\{\epsilon_t\}$ and $\{\eta_t\}$ are independent sequences, each consisting of independent variables with zero mean and constant variances. Then the conditional variance function $\mathrm{var}(X_t \mid X_{t)k},\boldsymbol{\eta})$ *is a constant only depending on k and the model coefficients, not on* $\boldsymbol{\eta}$.

Proof. Conditioning on $\boldsymbol{\eta}$, it is easy to show that X_t can be written as $X_t = Y_t + b_t$ where b_t is a constant depending on $\boldsymbol{\eta}$ and the coefficients of the model, and Y_t follows the linear AR model $Y_t = a_1Y_{t-1} + \cdots + a_{pY_{t-p}} + \epsilon_t$, which is a zero-mean stationary process. Using the fact that the process is reversible and a Yale-Walker equation-type of argument, it can be shown that $\mathrm{var}(Y_t \mid Y_{t-k},\boldsymbol{\eta})$ is a constant only depending on k and the coefficients. Since $\mathrm{var}(X_t \mid X_{t-k},\boldsymbol{\eta}) = \mathrm{var}(Y_t + b_t \mid Y_{t-k} + b_{t-k},\boldsymbol{\eta})$, the lemma follows. QED

Proof of Theorem 6. From Lemma D.1, it follows that σ_k^2 is independent of x and $\boldsymbol{\eta}$. We will only look at the case of $L_1(M_k^{(1)})$, the proof of $L_1(M_k)$ and $L_1(M_k^{(2)})$ being similar. From the proof of Theorem 3.2 (ii) in Hjellvik *et al.* (1996), we have, conditioning on $\boldsymbol{\eta}$,

$$nh^{5/2}L_1(M_{k,t}^{(1)}) = \mathrm{I}_{1,k,t} + \mathrm{I}_{2,k,t} + o_p(1),$$

where $\mathrm{I}_{1,k,t} = h^{-1/2}B_{1,k,t} + o_p(1)$ and $I_{2,k,t} \sim N(0, s_{1,k,t}^2)$, where

$$B_{1,k,t} = \frac{1}{\mu_2^2}\int u^2K^2(u)\mathrm{d}u \int \sigma_{k,t}^2(x)w(x)\mathrm{d}x,$$

and

$$s_{1,k,t}^2 = \frac{2}{\mu_2^2}\int \sigma_{k,t}^4(x)w^2(x)\mathrm{d}x \int uv(u-z)(v-z)K(u)K(v)K(u-z)K(v-z)\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}z,$$

where $\sigma_{k,t}^2(x) = \mathrm{var}(X_{(i)t} \mid X_{(i)t-k}, \boldsymbol{\eta})$. By Lemma D.1, we have $B_{1,k,t} = B_{1,k}$ and $s_{1,k,t} = (T-k)s_{1,k}$.

Note that the above result was proved under weaker conditions in Hjellvik *et al.* (1996, 1998) where they assumed that the observations come from an absolutely regular process. Here, conditioned on $\boldsymbol{\eta}$, our $(X_{(i)t-k},X_{(i)t})$ are independent for $i = 1,\ldots,n$.

In addition, let $\xi_{t,i} = (X_{(i)t},e_{(i)t})$. Then, $I_{2,k,t}$ can be expressed as

$$I_{2,k,t} = \frac{1}{n^2h^4\mu_2^2}\sum_{i\neq j}\phi(\xi_{t,i}\xi_{t,j}),$$

where

$$\phi(\xi_{t,i},\xi_{t,j})=e_{(i)t}e_{(j)t}\int\frac{w(x)}{p_{t-k}(x)}(X_{(i)t-k}-x)(X_{(j)t-k}-x)K_h(X_{(i)t-k}-x)K_h(X_{(j)t-k}-x)\mathrm{d}x,$$

where $K_h(x) = K(x/h)/h$, $e_{(i)t} = X_{(i)t} - E(X_{(i)t}| X_{(i)t-k},\boldsymbol{\eta})$ and $p_{t-k}(x)$ is the conditional density of $X_{(i)t-k}$ given $\boldsymbol{\eta}$.

To prove the theorem, we now only need to show that

$$E(nh^{5/2}I_{2,k,t_1}nh^{5/2}I_{2,k,t_2}) = o(1)$$

for $t_1 \neq t_2$. Note that

$$\begin{aligned}E(I_{2,k,t_1}I_{2,k,t_2}) &= \sum_{i_1\neq j_1}\sum_{i_2\neq j_2}E\{\phi(\xi_{t_1,i_1},\xi_{t_1,j_1})\phi(\xi_{t_2,i_2},\xi_{t_2,j_2})\}\\ &= \sum_{i\neq j}E\{\phi(\xi_{t_1,i},\xi_{t_1,j})\phi(\xi_{t_2,i},\xi_{t_2,j})\}.\end{aligned}$$

The last equality is due to the fact that for $i \neq j$, $X_{(i)t}$ and $X_{(j)s}$ are conditionally independent given $\boldsymbol{\eta}$ for all $s \leq t$. Furthermore,

$$E\{\phi(\xi_{t_1,i},\xi_{t_1,j})\phi(\xi_{t_2,i},\xi_{t_2,j})\} = \int\frac{w(x)w(y)}{p_{t_1-k}(x)p_{t_2-k}(y)}A^2(x,y)\mathrm{d}x\,\mathrm{d}y,$$

where

$$\begin{aligned}A(x,y) &= E\big\{e_{(i)t_1}e_{(i)t_2}(X_{(i)t_1-k}-x)(X_{(i)t_2-k}-y)K_h(X_{(i)t_1-k}-x)K_h(X_{(i)t_2-k}-y)\big\}\\ &\leq \big\{E(|e_{(i)t_1}e_{(i)t_2}|^q)\big\}^{1/q}\\ &\quad\times\big[E\{(X_{(i)t_1-k}-x)(X_{(i)t_2-k}-y)K_h(X_{(i)t_1-k}-x)K_h(X_{(i)t_2-k}-y)\}^p\big]^{1/p}\\ &= h^{2/p}\left\{\int u^p v^p K^p(u)K^p(v)\mathrm{d}u\,\mathrm{d}v\right\}^{1/p}\big\{E\big(|e_{(i)t_1}e_{(i)t_2}|^q\big)\big\}^{1/q}\\ &= O(h^{2/p}),\end{aligned}$$

where $q = 4+\delta/2$ and $p < 4/3$. The last equality is due to fact that

$$E(|e_{(i)t_1}e_{(i)t_2}|^q) \leq \{E(e^{2q}_{(i)t_1}E(e^{2q}_{(i)t_2})\}^{1/2} \leq \{E(X^{8+\delta}_{(i)t_1})E(X^{8+\delta}_{(i)t_2})\}^{1/2} = O(1)$$

by assumption (D2).

Hence,

$$E\{\phi(\xi_{t_1,i},\xi_{t_1,j})\phi(\xi_{t_2,i},\xi_{t_2,j})\} = O(h^{4/p}),$$

and consequently $E(nh^{5/2}I_{2,k,t_1}nh^{5/2}I_{2,k,t_2}) = O(h^{-3+4/p}) = o(1)$. Finally, note that $B_{1,k}$ and $s_{1,k}$ do not depend on $\boldsymbol{\eta}$. QED

Proof of Theorem 7. Following Theorem 3.3 of Hjellvik *et al.* (1998), given $\boldsymbol{\eta}$,

$$E\{L_1(M_{k,t})\} = E\Big[\{M_{k,t}(X_{(i)t-k}) - a_{k,t} - b_{k,t}X_{(i)t-k}\}^2 w(X_{(i)t-k}) \mid \boldsymbol{\eta}\Big].$$

By Lemma B.2 and the hypothesis $H_a^{(0)}$, $E\{L_1(M_{k,t})\} > c$ where c is a constant which does not depend on $\boldsymbol{\eta}$. The functionals $L_1(M_k^{(1)})$ and $L_1(M_k^{(2)})$ have similar properties. The theorem follows. QED

ACKNOWLEDGEMENTS

This work has received support from the Norwegian Supercomputer Committee through a grant of computing time. We are grateful to an anonymous referee for a number of useful comments.

REFERENCES

Arellano, M. (2003) *Panel Data Econometrics*. Oxford: Oxford University Press.

Baltagi, B. H. (1995) *Econometric Analysis of Panel Data*. New York: Wiley.

Bjørnstad, O. N., Champley, S., Stenseth, N. C. and Saitoh, T. (1996) Cyclicity and stability of grey-sided voles, *Clethrionomys rufocanus*, of Hokkaido: spectral and principal components analyses. *Philosophical Transactions Royal Society London B* 351, 867–75.

Cai, Z., Fan, J. and Yao, Q. (2000) Functional coefficient regression models for nonlinear time series. *Journal of American Statistical Association* 95, 941–56.

Cardot, H., Ferraty, F. and Sarda, P. (2004) Spline estimators for the functional linear model. *Statistica Sinica* (forthcoming).

Chamberlain, G. (1983) Funds, factors and diversification in arbitrage pricing models. *Econometrica* 51, 1281–1304.

Chamberlain, G. and Rotschild, M. (1983) Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* 51, 1305–24.

Chan, W. S. and Tong, H. (1986) On tests for non-linearity in time series analysis. *Journal of Forecasting* 5, 217–28.

Chen, R. and Tsay, R. S. (1993a) Functional coefficient autoregressive models. *Journal of American Statistical Association* 88, 298–308.

Chen, R. and Tsay, R. S. (1993b) Nonlinear additive ARX models. *Journal of American Statistical Association* 88, 955–67.

Cline, D. B. H. and Pu, H. H. (1999) Stability of nonlinear AR(1) time series with delay. *Stochastic Processes and their Applications* 82, 307–33.

Cox, D. R. and Solomon, P. J. (1988) On testing for serial correlation in large numbers of small samples. *Biometrika* 75, 145–8.

Dette, H. and Spreckelsen, I. (2004) Some comments on specification tests in nonparametric absolutely regular processes. *Journal of Time Series Analysis* 25, 159–72.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and its Applications*. London: Chapman and Hall.

Fan, Y. and Li, Q. (1999) Root-N-consistent estimation of partially linear time series models. *Journal of Nonparametric Statistics* 10, 251–69.

Fan, J. and Yao, Q. (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000) The generalized dynamic factor model: identification and estimation. *Review of Economics and Statistics* 82, 540–54.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2001) Coincident and leading indicators for the EURO area. *Economic Journal* 111, C62–85.

Gasser, T. and Kneip A. (1995) Searching for structures in curve samples. *Journal of American Statistical Association* 90 1179–88.

Geweke, J. (1977) The dynamic factor analysis of economic time series. In *Latent Variables in Socio-Economic Models* (eds D. I. Aigner and A. S. Goldberger). Amsterdam: North Holland, Ch. 19.

Granger, C. W. J. and Teräsvirta, T. (1993) *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Greene, W. H. (2003) *Econometric Analysis* (5th edn). Upper Saddle River, NJ: Prentice Hall.

Hart, J. D. (1996) Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics* 6, 115–42.

Hjellvik, V. and Tjøstheim, D. (1995) Nonparametric tests of linearity for time series. *Biometrika* 82 351–68.

HJELLVIK, V. and TJØSTHEIM, D. (1996) Nonparametric statistics for testing of linearity and serial independence. *Journal of Nonparametric Statistics* 6, 223–51.
HJELLVIK, V. and TJØSTHEIM, D. (1999a) Modelling panels of intercorrelated autoregressive time series. *Biometrika* 86, 573–90.
HJELLVIK, V. and TJØSTHEIM, D. (1999b) Order determination and residual variance estimates in panels of intercorrelated autoregressive time series. In *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger* (invited paper) (eds R. F. ENGLE and H. WHITE). Oxford: Oxford University Press, pp. 385–409.
HJELLVIK, V., YAO, Q. and TJØSTHEIM, D. (1996) Linearity testing using local polynomial approximation. Discussion Paper 60, 1996, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.
HJELLVIK, V., YAO, Q. and TJØSTHEIM, D. (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.
HSIAO, C. (1986) *Analysis of Panel Data*. Cambridge: Cambridge University Press.
KEENAN, D. M. (1985) A Tukey non-additivity type test for time series nonlinearity. *Biometrika* 72, 39–44.
LING, S. (2004) Estimating and testing stationarity for double autoregressive models. *Journal Royal Statistical Society, Series* B 66, 63–78.
LU, Z. (1998) On the ergodicity of non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica* 8, 1205–17.
LUUKKONEN, R., SAIKKONEN, P. and TERÄSVIRTA, T. (1988) Testing linearity in univariate time series. *Scandinavian Journal of Statistics* 15, 161–75.
MASRY, E. and FAN, J. (1997) Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* 24, 165–80.
MASRY, E. and TJØSTHEIM (1995) Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality. *Econometric Theory* 11, 258–89.
MASRY, E. and TJØSTHEIM (1997) Additive nonlinear ARX time series and projection estimates. *Econometric Theory* 13, 214–52.
MÁTYAS, L. and SEVESTRE, P. (1992) *The Econometrics of Panel Data*. Amsterdam: Kluwer Academic Publishers.
MEYN, S. P. and TWEEDIE, R. L. (1994) *Markov Chains and Stochastic Stability*. New York: Springer.
PETRUCCELLI, J. D. and DAVIS, N. (1986) A portmanteu test for self-exciting threshold autoregressive-type non-linearity in time series. *Biometrika* 73, 687–94.
RAMSAY, J. O. and SILVERMAN, B. W. (1997) *Functional Data Analysis*. New York: Springer.
RAMSAY, J. O. and SILVERMAN, B. W. (2002) *Applied Functional Data Analysis. Methods and Case Studies*. New York: Springer.
RICE, J. A. and SILVERMAN, B. W. (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal Royal Statistical Society, Series B* 53, 233–43.
SILVERMAN, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
STENSETH, N. C., BJØRNSTAD, O. N. and SAITOH, T. (1996) A gradient from stable to cyclic populations of *Clethrionomys rufocanus* in Hokkaido, Japan. *Proceedings Royal Society London B* 263, 1117–26.
STOCK, J. H. and WATSON, M. H. (2002) Forecasting using principal components from a large number of predictors. *Journal of American Statistical Association* 97, 1167–79.
TJØSTHEIM, D. (1990) Nonlinear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.
TSAY, R. (1986) Nonlinearity tests for time series. *Biometrika* 73, 461–66.
TSAY, R. (1989) Testing and modelling threshold autoregressive processes. *Journal of American Statistical Association* 84, 231–40.
YAO, Q., TONG, H., FINKENSTÄDT, B. and STENSETH, N. C. (1998) Common structure in panels of short time series. Preprint.