Nonparametric Estimation of Generalized Impulse Response Functions

ROLF TSCHERNIG* and LIJIAN YANG
Humboldt-Universität zu Berlin, Michigan State University

January 2000
preliminary

Abstract

We derive a local linear estimator of generalized impulse response (GIR) functions for nonlinear conditional heteroskedastic autoregressive processes and show its asymptotic normality. We suggest a plug-in bandwidth based on the derived asymptotically optimal bandwidth. A local linear estimator for the conditional variance function is proposed which has simpler bias than the standard estimator. This is achieved by appropriately eliminating the conditional mean. Alternatively to the direct local linear estimators of the k-step prediction functions which enter the GIR estimator we suggest to use multi-stage prediction techniques. In a small simulation experiment the latter estimator is found to perform best.

KEY WORDS: Confidence intervals; heteroskedasticity; local polynomial; multistage predictor; nonlinear autoregression; plug-in bandwidth.

1 INTRODUCTION

Recent advances in statistical theory and computer technology have made it possible to use nonparametric techniques for nonlinear time series analysis. Consider the nonlinear conditional heteroskedastic autoregressive process $\{Y_t\}_{t\geq 0}$

$$Y_t = f(\mathbf{X}_{t-1}) + \sigma(\mathbf{X}_{t-1})U_t, \quad t = m, m+1, ... \tag{1}$$

where $\mathbf{X}_{t-1} = (Y_{t-1}, ..., Y_{t-m})^T$, $t = m, m+1, ...$ denotes the vector of lagged observations up to lag m, and f and σ denote the conditional mean and conditional standard deviation, respectively. The series $\{U_t\}_{t\geq m}$ represents i.i.d. random variables with $E(U_t) = 0$, $E(U_t^2) = 1$, $E(U_t^3) = m_3$, $E(U_t^4) = m_4 < +\infty$ and which are independent of $\mathbf{X}_{t-1}$. Masry and Tjøstheim (1995) showed asymptotic normality of the Nadaraya-Watson estimator for estimating the conditional mean function f under the condition that the process is α-mixing. Härdle, Tsybakov and Yang (1998) proved asymptotic normality for the local

*Address for Correspondence: Institut für Statistik und Ökonometrie, Wirtschaftswissenschaftliche Fakultät, Humboldt-Universität zu Berlin, Spandauer Str.1, D-10178 Berlin, Germany, email: rolf@wiwi.hu-berlin.de.

linear estimator of f. For selecting the order m one may use the nonparametric procedures suggested by Tjøstheim and Auestad (1994) and Tschernig and Yang (2000) which are based on local constant and local linear estimators of the final prediction error, respectively. Alternatively one may use cross-validation, see Yao and Tong (1994). For further references the reader is referred to the surveys of Tjøstheim (1994) or Härdle, Lütkepohl and Chen (1997).

An important goal of nonlinear time series modelling is the understanding of the underlying dynamics. As is well known from linear time series analysis it is not sufficient for this task to estimate the conditional mean function. This is even more so if the conditional mean function is a nonlinear function of lagged observations. One appropriate tool that allows to study the dynamics of processes like (1) are generalized impulse response functions.

In this paper we propose nonparametric estimators for generalized impulse response (GIR) functions for nonlinear conditional heteroskedastic autoregressive processes (1) and derive their asymptotic properties. Here, we follow Koop, Pesaran and Potter (1996) and define the generalized impulse response GIR_k for horizon k as the quantity by which a prespecified shock u in period t changes the k-step ahead prediction based on information up to period $t-1$ only. Formally, one has

$$\begin{aligned} GIR_k(\mathbf{x},u) &= E(Y_{t+k-1}|\mathbf{X}_{t-1}=\mathbf{x},U_t=u) - E(Y_{t+k-1}|\mathbf{X}_{t-1}=\mathbf{x}) \\ &= E(Y_{t+k-1}|Y_t=f(\mathbf{x})+\sigma(\mathbf{x})u, Y_{t-1}=x_1,...,Y_{t-m+1}=x_{m-1}) \\ &\quad -E(Y_{t+k-1}|Y_{t-1}=x_1,...,Y_{t-m}=x_m). \end{aligned} \tag{2}$$

In general, the GIR_k depends on the condition $\mathbf{x}$ as well as the size and sign of the shock u. An alternative definition of nonlinear impulse response functions is given by Gallant, Rossi and Tauchen (1993).

We propose local linear estimators for the prediction functions which are contained in GIR_k and derive the asymptotic properties of the resulting GIR_k estimator. This also delivers an asymptotically optimal bandwidth allowing to compute a plug-in bandwidth. The estimation of GIR_k also requires to estimate the conditional standard deviation σ which can be done e.g. with the local linear volatility estimator suggested by Härdle and Tsybakov (1997). In this paper we propose an alternative local linear estimator that exhibits a simpler asymptotic bias. For estimating the prediction functions, we alternatively suggest to apply multi-stage prediction techniques which were recently analysed by Chen, Yang and Hafner (1999). An initial evaluation of the performance of both local linear GIR_k estimators is provided by a small Monte Carlo study where we compare the mean squared errors of nonparametric and parametric GIR_{10} estimators for a logistic autoregressive process of order one. Higher order processes are currently analyzed.

The paper is organized as follows. In Section 2 we define local linear estimators for the generalized impulse response function and investigate its asymptotic properties. The alternative estimator for the conditional standard deviation is introduced in Section 3. In Section 4 a GIR estimator based on multi-stage prediction is proposed. Issues of implementation are discussed in Section 5. The results of the small Monte Carlo study are summarized in Section 6.

2 AN ESTIMATOR FOR THE GIR FUNCTION

To facilitate the presentation, we use the following notation. Denote for any $k \geq 1$ the k-step ahead prediction function by

$$f_k(\mathbf{x}) = E(Y_{t+k-1}|\mathbf{X}_{t-1} = \mathbf{x}) \tag{3}$$

and write

$$Y_{t+k-1} = f_k(\mathbf{X}_{t-1}) + \sigma_k(\mathbf{X}_{t-1})U_{t,k} \tag{4}$$

where

$$\sigma_k^2(\mathbf{x}) = Var(Y_{t+k-1}|\mathbf{X}_{t-1} = \mathbf{x}) \tag{5}$$

and where the $U_{t,k}$'s are martingale differences since $E(U_{t,k}|\mathbf{X}_{t-1}) = E(U_{t,k}|Y_{t-1}, ...) = 0$, $E(U_{t,k}^2|\mathbf{X}_{t-1}) = E(U_{t,k}^2|Y_{t-1}, ...) = 1$, $t = m, m+1, ...$ Apparently, $f_1 = f$, $\sigma_1 = \sigma$. One also denotes

$$\sigma_{k',k}(\mathbf{x}) = Cov\left\{(Y_{t+k'-1}, Y_{t+k-1})|\mathbf{X}_{t-1} = \mathbf{x}\right\}, \tag{6}$$

$$\sigma_{k'k',k}(\mathbf{x}) = Cov\left\{\left[\left\{Y_{t+k'-1} - f_{k'}(\mathbf{X}_{t-1})\right\}^2, Y_{t+k-1} - f_k(\mathbf{X}_{t-1})\right]|\mathbf{X}_{t-1} = \mathbf{x}\right\}. \tag{7}$$

One can now write the generalized impulse response (GIR_k) function defined in (2) more compactly as

$$GIR_k(\mathbf{x}, u) = f_{k-1}\left\{f(\mathbf{x}) + \sigma(\mathbf{x})u, \mathbf{x}'\right\} - f_k(\mathbf{x}) = f_{k-1}(\mathbf{x}_u) - f_k(\mathbf{x}) \tag{8}$$

where $\mathbf{x}' = (x_1, ..., x_{m-1})$ and $\mathbf{x}_u = \{f(\mathbf{x}) + \sigma(\mathbf{x})u, \mathbf{x}'\}$.

The estimated GIR_k function is then

$$\widehat{GIR}_k(\mathbf{x}, u) = \widehat{f}_{k-1}(\widehat{\mathbf{x}}_u) - \widehat{f}_k(\mathbf{x}) \tag{9}$$

where all unkown functions are replaced by local linear estimates. The estimator of $\mathbf{x}_u$ is $\widehat{\mathbf{x}}_u = \left\{\widehat{f}(\mathbf{x}) + \widehat{\sigma}(\mathbf{x})u, \mathbf{x}'\right\}$. For defining the local linear estimators, $K : I\!R^1 \longrightarrow I\!R^1$ denotes a kernel function which is assumed to be a continuous, symmetric and compactly supported probability density and

$$K_h(\mathbf{x}) = 1/h^m \prod_{j=1}^{m} K(x_j/h)$$

defines the product kernel for $\mathbf{x} \in I\!R^m$ and the bandwidth $h = \beta n^{-1/(m+4)}$. Define further the matrices

$$e = (1, 0_{1\times m})^T, \quad \mathbf{Z}_k = \begin{pmatrix} 1 & \cdots & 1 \\ \mathbf{X}_{m-1} - \mathbf{x} & \cdots & \mathbf{X}_{n-k} - \mathbf{x} \end{pmatrix}^T$$

$$W_k = \text{diag}\left\{K_h(\mathbf{X}_{i-1} - \mathbf{x})/n\right\}_{i=m}^{n-k+1}, \quad \mathbf{Y}_k = \begin{pmatrix} Y_{m+k-1} & \cdots & Y_n \end{pmatrix}^T.$$

Then the local linear estimator $\widehat{f}_k(\mathbf{x})$ of the k-step ahead prediction function $f_k(\mathbf{x})$ can be written as

$$\widehat{f}_k(\mathbf{x}) = e^T \left(\mathbf{Z}_k^T W_k \mathbf{Z}_k\right)^{-1} \mathbf{Z}_k^T W_k \mathbf{Y}_k. \tag{10}$$

The local linear estimate $\widehat{\sigma}_k(\mathbf{x})$ of the conditional k-step ahead standard deviation is defined by

$$\widehat{\sigma}_k(\mathbf{x}) = \left\{ e^T \left(\mathbf{Z}_k^T W_k \mathbf{Z}_k \right)^{-1} \mathbf{Z}_k^T W_k \mathbf{Y}_k^2 - \widehat{f}_k^2(\mathbf{x}) \right\}^{1/2}. \tag{11}$$

For simplicity, we write $\widehat{f}(\mathbf{x}) = \widehat{f}_1(\mathbf{x})$, $\widehat{\sigma}(\mathbf{x}) = \widehat{\sigma}_1(\mathbf{x})$.

In the following theorem we show the asymptotic normality of the local linear GIR_k estimator (9) based on (10) and (11). The theorem also states the asymptotically optimal bandwidth. We denote $\|K\|_2^2 = \int K^2(u)du$, $\sigma_K^2 = \int K(u)u^2 du$.

Theorem 1 *Define the asymptotic variance*

$$\sigma_{GIR,k}^2(\mathbf{x}, u) = \frac{\|K\|_2^{2m} \sigma^2(\mathbf{x})}{\mu(\mathbf{x})} \left[\frac{\sigma_{k-1}^2(\mathbf{x}_u)\mu(\mathbf{x})}{\mu(\mathbf{x}_u)\sigma^2(\mathbf{x})} + \frac{\sigma_k^2(\mathbf{x})}{\sigma^2(\mathbf{x})} + \right.$$

$$\left. \left\{ \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \right\}^2 \left\{ 1 + u m_3 + \frac{u^2(m_4 - 1)}{4} \right\} - \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \left\{ \frac{2\sigma_{1k}(\mathbf{x})}{\sigma^2(\mathbf{x})} + u \frac{\sigma_{11,k}(\mathbf{x})}{\sigma^3(\mathbf{x})} \right\} \right]$$

$$- \frac{\|K\|_2^{2m}}{\mu(\mathbf{x})} I(\mathbf{x} = \mathbf{x}_u) \left\{ 2\sigma_{k-1,k}(\mathbf{x}) - 2 \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \sigma_{1,k-1}(\mathbf{x}) + u \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \frac{\sigma_{11,k-1}(\mathbf{x})}{\sigma(\mathbf{x})} \right\} \tag{12}$$

and the asymptotic bias

$$b_{GIR,k}(\mathbf{x}, u) = b_{f,k-1}(\mathbf{x}_u) - b_{f,k}(\mathbf{x}) + \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \{ b_f(\mathbf{x}) + b_\sigma(\mathbf{x}) u \} \tag{13}$$

where

$$\begin{aligned} b_{f,k}(\mathbf{x}) &= \sigma_K^2 \operatorname{Tr} \{ \nabla^2 f_k(\mathbf{x}) \} / 2 \\ b_{\sigma,k}(\mathbf{x}) &= \sigma_K^2 \left[\operatorname{Tr} \nabla^2 \{ f_k^2(\mathbf{x}) + \sigma_k^2(\mathbf{x}) \} - 2 f_k(\mathbf{x}) \operatorname{Tr} \nabla^2 \{ f_k(\mathbf{x}) \} \right] / \{ 4\sigma_k(\mathbf{x}) \}. \end{aligned} \tag{14}$$

$\operatorname{Tr}\{\nabla^2 f_k(\mathbf{x})\}$ *denotes the Laplacian operator, and one abbreviates* $b_{f,1}(\mathbf{x}), b_{\sigma,1}(\mathbf{x})$ *simply as* $b_f(\mathbf{x}), b_\sigma(\mathbf{x})$. *Then under assumptions (A1)-(A3) given in the Appendix, one has*

$$\sqrt{nh^m} \left\{ \widehat{GIR}_k(\mathbf{x}, u) - GIR_k(\mathbf{x}, u) - b_{GIR,k}(\mathbf{x}, u) h^2 \right\} \to N \left\{ 0, \sigma_{GIR,k}^2(\mathbf{x}, u) \right\} \tag{15}$$

and so the optimal bandwidth for estimating $GIR_k(\mathbf{x}, u)$ *is*

$$h_{opt}(\mathbf{x}, u) = \left\{ \frac{m \sigma_{GIR,k}^2(\mathbf{x}, u)}{4 b_{GIR,k}^2(\mathbf{x}, u) n} \right\}^{1/(m+4)}. \tag{16}$$

In practice, some quantities in the asymptotically optimal bandwidth (16) are unknown. In Section 5 we discuss estimators for those quantities in order to obtain a plug-in bandwidth. This plug-in bandwidth is then used in the small Monte Carlo experiment presented in Section 6.

Koop, Pesaran and Potter (1996) consider various definitions of generalized impulse response functions. For example, one alternative to (2) is to allow the condition to be a compact set. Denoting by $\mathbf{C}_{\mathbf{x}}$ and C_u compact subsets of R^m and R, respectively, the generalized impulse response function over these compact sets is defined by

$$GIR_k(\mathbf{C}_{\mathbf{x}}, C_u) = E\left\{GIR_k(\mathbf{X}_{i-1}, U_i) | \mathbf{X}_{i-1} \in \mathbf{C}_{\mathbf{x}}, U_i \in C_u\right\}. \tag{17}$$

For its estimation, we consider its empirical version

$$\widehat{GIR}_k(\mathbf{C}_{\mathbf{x}}, C_u) = \frac{1}{n\widehat{P}\left(\mathbf{C}_{\mathbf{x}}, C_u\right)} \sum_{i=m}^{n-k+1} \widehat{GIR}_k(\mathbf{X}_{i-1}, U_i) I\left(\mathbf{X}_{i-1} \in \mathbf{C}_{\mathbf{x}}, U_i \in C_u\right) \tag{18}$$

where

$$\widehat{P}\left(\mathbf{C}_{\mathbf{x}}, C_u\right) = \frac{1}{n} \sum_{i=m}^{n-k+1} I\left(\mathbf{X}_{i-1} \in \mathbf{C}_{\mathbf{x}}, U_i \in C_u\right).$$

The asymptotic properties of the estimator (18) for generalized impulse response functions over compact sets $(\mathbf{C}_{\mathbf{x}}, C_u)$ are summarized in the next theorem.

Theorem 2 *Under assumptions (A1)-(A3) given in the Appendix*

$$\widehat{GIR}_k(\mathbf{C}_{\mathbf{x}}, C_u) - GIR_k(\mathbf{C}_{\mathbf{x}}, C_u) = b_{GIR,k}(\mathbf{C}_{\mathbf{x}}, C_u)h^2 + o_p(h^2) \tag{19}$$

where

$$b_{GIR,k}(\mathbf{C}_{\mathbf{x}}, C_u) = E\left\{b_{GIR,k}(\mathbf{X}_{i-1}, U_i) | \mathbf{X}_{i-1} \in \mathbf{C}_{\mathbf{x}}, U_i \in C_u\right\}.$$

Theorem 2 shows that for the generalized impulse response functions over compact sets there does not exist the usual bias-variance trade-off. Within the constraint of $h = \beta n^{-1/(m+4)}$ it is better to use a smaller h. This, of course, has to be qualified for finite samples.

While the estimator for GIR_k proposed in this section has reasonable asymptotic properties, it may cause problems in finite samples. In the next section we discuss the problem in more detail and present an improved estimator.

3 AN ALTERNATIVE LOCAL LINEAR ESTIMATOR OF THE CONDITIONAL VOLATILITY

The GIR_k estimator (9) is based on the standard estimator (11) for the conditional volatility. This local linear estimator $\widehat{\sigma}^2(\mathbf{x})$, however, may produce negative values for $\sigma(\mathbf{x})$ if f^2 is estimated badly and is then not usable. This problem can also occur for other auxiliary functions such as $\widehat{\sigma}_k(\mathbf{x}), \widehat{\sigma}_{1,k}(\mathbf{x}), \widehat{\sigma}_{11,k}(\mathbf{x})$, etc., which will be needed for computing the plug-in bandwidth based on formula (16). In this section we present an alternative local linear estimator for the conditional standard deviation that cannot become negative due to a badly estimated f^2. The proposed method can also be used for estimating the covariance functions $\sigma_k^2(\mathbf{x}), \sigma_{1,k}(\mathbf{x}), \sigma_{11,k}(\mathbf{x})$.

The idea for estimating $\sigma_k^2(\mathbf{x})$ is to base the estimator on the estimated residuals and use

$$\widetilde{\sigma}_k^2(\mathbf{x}) = e^T \left(\mathbf{Z}_k^T W_k \mathbf{Z}_k\right)^{-1} \mathbf{Z}_k^T W_k \mathbf{V}_k \tag{20}$$

where $\mathbf{V}_k = \left(\left\{Y_{m+k-1} - \widehat{f}_k(\mathbf{X}_{m-1})\right\}^2 \quad \cdots \quad \left\{Y_n - \widehat{f}_k(\mathbf{X}_{n-k})\right\}^2 \right)^T$. In the next lemma it is shown that this approach is indeed useful.

Lemma 1 *Under assumptions (A1)-(A3) in the Appendix, one has*

$$\widetilde{\sigma}_k^2(\mathbf{x}) - \sigma_k^2(\mathbf{x}) = \widetilde{b}_{\sigma,k}(\mathbf{x})h^2 + \frac{1}{n\mu(\mathbf{x})}\sum_{j=m}^{n} K_h(\mathbf{X}_{j-1}-\mathbf{x})\sigma_k^2(\mathbf{X}_{j-1})(U_{j,k}^2 - 1) + o_p(h^2) \quad (21)$$

where

$$\widetilde{b}_{\sigma,k}(\mathbf{x}) = \frac{\sigma_K^2}{2}\,\mathrm{Tr}\,\nabla^2\left\{\sigma_k^2(\mathbf{x})\right\} \quad (22)$$

and

$$\sqrt{nh^m}\left\{\widetilde{\sigma}_k^2(\mathbf{x}) - \sigma_k^2(\mathbf{x}) - \widetilde{b}_{\sigma,k}(\mathbf{x})h^2\right\} \to N\left\{0, \sigma_{\sigma,k}^2(\mathbf{x})\right\}$$

with

$$\sigma_{\sigma,k}^2(\mathbf{x}) = \frac{\|K\|_2^{2m}\,\sigma_k^4(\mathbf{x})}{\mu(\mathbf{x})}(m_{4,k}-1)$$

where $m_{4,k} = E(U_{j,k}^4)$.

This lemma basically says that by de-meaning one can estimate $\sigma_k^2(\mathbf{x})$ as well as if one knew the true k-step regression function f_k. As one would expect, the noise level is the same for both $\widehat{\sigma}_k^2(\mathbf{x})$ and $\widetilde{\sigma}_k^2(\mathbf{x})$ which can be seen from (21) and (28). However, from comparing $b_{\sigma,k}$ and $\widetilde{b}_{\sigma,k}$ given by (14) and (22), it can be seen that $\widetilde{\sigma}_k^2(\mathbf{x})$ has a simpler bias which does not depend on f_k.

In a similar way one can define estimators for the quantities (6) and (7). The following lemma states their asymptotic properties.

Corollary 1 *Under assumptions (A1)-(A3) in the Appendix, one can also estimate* $\sigma_{11,k}(\mathbf{x})$ *as*

$$\widetilde{\sigma}_{11,k}(\mathbf{x}) = e^T\left(\mathbf{Z}_k^T W_k \mathbf{Z}_k\right)^{-1}\mathbf{Z}_k^T W_k \mathbf{V}_{11,k}$$

where

$$\mathbf{V}_{11,k} = \left(\ \left\{Y_m - \widehat{f}(\mathbf{X}_{m-1})\right\}^2\left\{Y_{m+k-1} - \widehat{f}_k(\mathbf{X}_{m-1})\right\} \ \cdots \ \left\{Y_{n-k+1} - \widehat{f}(\mathbf{X}_{n-k})\right\}^2\left\{Y_n - \widehat{f}_k(\mathbf{X}_{n-k})\right\}\ \right)$$

and likewise $\sigma_{1,k}(\mathbf{x})$. *The respective estimators have similar properties as* $\widetilde{\sigma}_k^2(\mathbf{x})$.

The fact that $\widetilde{\sigma}_k(\mathbf{x})$ has a simpler bias facilitates the computation of the plug-in bandwidth since the asymptotic bias term in the asymptotically optimal bandwidth (16) becomes simpler as well. For this reason we use from now on in the GIR_k estimator (9) the new estimator (20) instead of (11) for estimating conditional volatilities. We note that in some cases e.g. if the bandwidth is not appropriate and $\mathbf{x}$ is outside the range of the observed data, $\widetilde{\sigma}_k(\mathbf{x})$ can lead to negative estimates for the conditional variance. Then one replaces in (20) the local linear by the local constant estimator which always produces positive estimates.

4 GIR ESTIMATION USING MULTI-STAGE PREDICTION

The main ingredient of the GIR_k estimator (9) are the direct local linear predictors $\widehat{f}_k$ and $\widehat{f}_{k-1}$. While they are simple to implement, they may contain too much noise which has accumulated over the k prediction periods.

To estimate $f_k(\mathbf{x})$ more efficiently, we therefore propose to use instead the multi-stage method. It was analyzed in detail by Chen, Yang and Hafner (1999). To describe the procedure, one starts with $Y_t^{(0)} = Y_t$, and repeats the following stage for $j = 1, \ldots, k-1$. For an easy presentation, we use here the Nadaraya-Watson form.

Stage j: Estimate

$$\widetilde{f}_j(\mathbf{x}) = \frac{\sum_{t=m-1}^{n-k} K_{h_j}(\mathbf{X}_t - \mathbf{x}) Y_{t+j}^{(j-1)}}{\sum_{t=m-1}^{n-k} K_{h_j}(\mathbf{X}_t - \mathbf{x})},$$

and obtain the j-th smoothed version of Y_{t+j} by $Y_{t+j}^{(j)} = \hat{f}_j(\mathbf{X}_t)$.

Then, the conditional mean function $f_k(\mathbf{x})$ is estimated by

$$\widetilde{f}_k(\mathbf{x}) = \frac{\sum_{t=m-1}^{n-k} K_{h_k}(\mathbf{X}_t - \mathbf{x}) Y_{t+k}^{(k-1)}}{\sum_{t=m-1}^{n-k} K_{h_k}(\mathbf{X}_t - \mathbf{x})}. \tag{23}$$

Graphically, the above recursive method can be presented as

$$Y_{t+k} \overset{(Y_{t+k},\mathbf{X}_{t+k-1})}{\Longrightarrow} Y_{t+k}^{(1)} \overset{(Y_{t+k}^{(1)},\mathbf{X}_{t+k-2})}{\Longrightarrow} Y_{t+k}^{(2)} \overset{(Y_{t+k}^{(2)},\mathbf{X}_{t+k-3})}{\Longrightarrow} \cdots \overset{(Y_{t+k}^{(k-2)},\mathbf{X}_{t+1})}{\Longrightarrow} Y_{t+k}^{(k-1)} \overset{(Y_{t+k}^{(k-1)},\mathbf{X}_{t})}{\Longrightarrow} \widetilde{f}_k(\mathbf{x}).$$

The following theorem is shown in Chen, Yang and Hafner (1999).

Theorem 3 *Under conditions (A1)-(A3) in the Appendix, if* $h_j = o(h_k), nh_j^m \to \infty$ *for* $j = 1, \ldots, k-1$, *and* $h_k = \beta n^{-1/(m+4)}$ *for some* $\beta > 0$, *and if the estimators* $\widetilde{f}_j(\mathbf{x})$ *are all obtained local linearly, then*

$$\sqrt{nh_k^m}\left\{\widetilde{f}_k(\mathbf{x}) - f_k(\mathbf{x}) - b_{f,k}(\mathbf{x})h_k^2\right\} \longrightarrow N\left\{0, \frac{\|K\|_2^{2m}\, s_k^2(\mathbf{x})}{\mu(\mathbf{x})}\right\}$$

where

$$s_k^2(\mathbf{x}) = Var\left\{\hat{f}_{k-1}(\mathbf{X}_t)|\mathbf{X}_{t-1} = \mathbf{x}\right\}.$$

The local linear GIR_k estimator based on multi-stage prediction is therefore given by

$$\widetilde{GIR}_k(\mathbf{x}, u) = \widetilde{f}_{k-1}(\widetilde{\mathbf{x}}_u) - \widetilde{f}_k(\mathbf{x}) \tag{24}$$

with the multi-stage predictor $\widetilde{f}_k(\mathbf{x})$ and the alternative estimator for the conditional standard deviation $\widetilde{\sigma}_k(\mathbf{x})$ given by (23) and (20), respectively. In the next section we turn to issues of implementation.

5 IMPLEMENTATION

Computing the direct or multi-stage GIR estimators (9) or (24) requires suitable bandwidth estimates. Both estimators were implemented in GAUSS and use the Gaussian kernel. We first discuss how to obtain a plug-in bandwidth by estimating the unknown quantities in the asymptotically bandwidth (16) where (14) is replaced by (22) since (20) is used. For estimating the densities $\mu(\mathbf{x})$ and $\mu(\mathbf{x}_u)$ in (12) we use a kernel density estimator with the Silverman's (1986) rule-of-thumb bandwidth $h_\mu = h\left(m+2, \sqrt{\widehat{Var}(\mathbf{X})}\right)$ where

$$h_S(k,\sigma) = \sigma\,(4/k)^{1/(k+2)}\,n^{-1/(k+2)} \tag{25}$$

and where $\widehat{Var}(\mathbf{X})$ denotes the geometric mean of the variances for each regressor. The bandwidth h_μ is also used for estimating all other unknown quantities and is of the correct order except for estimating the second order direct derivatives in (13). For the latter quantities we use a partial quadratic estimator which is a simplified version of the partial cubic estimator presented in Yang and Tschernig (1999) and for which they show that $h_{sd} = h_S\left(m+4, 3\sqrt{\widehat{Var}(\mathbf{X})}\right)$ has the correct order.

For the multi-stage GIR_k estimator (24) there does not exist a scalar optimal bandwidth. According to Chen, Yang and Hafner (1999) the optimal bandwidth for the first $j \leq k-1$ predictions $\widetilde{f}_j(\mathbf{x})$ has a different rate. In their simulations they find $h_{MS,k-1} = \widehat{h}_{opt} n^{-4/(m+4)^2}/5$ to work quite well. For the kth-step we use $\widehat{h}_{opt}$. If the multi-stage predictor is used for computing the plug-in bandwidth, $\widehat{h}_{opt}$ is replaced by h_μ.

6 A SMALL SIMULATION STUDY

In this section we investigate the performance of the proposed GIR_k estimators based on 500 observations of the logistic autoregressive process

$$Y_t = 0.9Y_{t-1} - 0.7Y_{t-1}\frac{1}{1+\exp(-3Y_{t-1})} + U_t, \quad U_t \sim i.i.d.N(0,1). \tag{26}$$

One realization of the process is shown in Figure 1a). In the following we present results for estimating $GIR_k(\mathbf{x},u)$ for $k=10$, a unit shock $u=1$ and $\mathbf{x}$ taking values from -5 to 1 in steps of 1. Figure 1b) displays the true $f_k(\mathbf{x})$ and $f_{k-1}(\mathbf{x}_u)$ functions which were computed by simulation.

Next we conducted 100 simulations of this process and estimated $GIR_k(\mathbf{x},u)$ by (9) with (10) and (20) as well as by the alternative estimator based on the multi-stage predictor (23) and (20). We also fitted a linear AR(1) model and computed the corresponding impulse responses. Finally, we estimated the impulse responses on the estimated parameters of the correct logistic AR model. Figure 2 displays the various estimates for the 54th simulation. The multi-stage based GIR estimate (short dashes) seems to be closest to the true GIR function while using the one-stage predictors (long dashes) perform worse for negative values of $\mathbf{x}$. The parametric estimate of the impulse response (short dots at the top of the plot) based on the true model is the worst. This can be attributed to the difficulties in estimating the parameter in the exponential function. The linear impulse response (dots) also misses the GIR by construction. This observations are indeed

Table 1: Mean squared errors of various estimates of the generalized impulse responses for $k = 10$ and $u = 1$

Estimator \ $\mathbf{x}$	-5	-4	-3	-2	-1	0	1
linear IR	0.0118	0.0092	0.0085	0.0114	0.0238	0.0475	0.0527
local linear one-stage	0.1657	0.0838	0.0468	0.0257	0.0182	0.0125	0.0099
local linear multi-stage	0.0289	0.0211	0.0114	0.0061	0.0024	0.0042	0.0027
GIR with est. par. of (26)	0.3593	0.3960	0.4493	0.5694	0.5954	0.1860	0.1273

Table 2: Mean integrated squared errors of various estimates of the generalized impulse responses for $k = 10$ and $u = 1$

Estimators	
linear IR	0.0272
local linear one-stage	0.0257
local linear multi-stage	0.0059
GIR with est. par. of (26)	0.3767

representative. Table 1 displays the mean squared error of each estimator for each $\mathbf{x}$. If one is interested in further aggregating these performance measures, one can consider the mean integrated squared error. It is obtained by the weighted sum of the MSE's where the weights are given by the density of $\mathbf{x}$. Inspecting the MISE's in Table 2 confirms the superiority of the multi-stage local linear estimator for the generalized impulse responses $GIR_{10}(\mathbf{x}, 1)$.

From this little simulation study we conclude that the proposed multi-stage estimator may be useful in practice although much more Monte Carlo experiments are needed for assessing the empirical applicability of the proposed methods. This is particularly true for nonlinear autoregressive processes of higher order. In any case, these methods have standard asymptotic properties.

APPENDIX

With regard to the process (1) we assume the following:

(A1) The vector process $\mathbf{X}_{t-1} = (Y_{t-1}, ..., Y_{t-m})^T$ is strictly stationary and geometrically β-mixing: $\beta(n) \le c_0 \rho^{-n}$ for some $0 < \rho < 1$, $c_0 > 0$. Here

$$\beta(n) = E \sup \left\{ \left| P(A|\mathcal{F}_m^k) - P(A) \right| : A \in \mathcal{F}_{n+k}^{\infty} \right\}$$

where $\mathcal{F}_t^{t'}$ is the σ-algebra generated by $\mathbf{X}_t, \mathbf{X}_{t+1}, ..., \mathbf{X}_{t'}$.

(A2) The stationary distribution of the process $\mathbf{X}_{t-1}$ has a density $\mu(\mathbf{x})$, $\mathbf{x} \in I\!R^m$, which is continuous. If the Nadaraya-Watson estimator is used, $\mu(\cdot)$ has to be continuously differentiable.

(A3) The function $f(\cdot)$ is twice continuously differentiable while $\sigma(\cdot)$ is continuous and positive on the support of $\mu(\cdot)$.

A discussion of these assumptions can be found e.g. in Tschernig and Yang (2000).

For proving Theorem 1 it is necessary to derive some auxiliary results first and decompose the GIR_k estimator in several terms. By Härdle, Tsybakov and Yang (1998), we have

$$\widehat{f}_k(\mathbf{x}) = f_k(\mathbf{x}) + b_{f,k}(\mathbf{x})h^2 + \frac{1}{n\mu(\mathbf{x})} \sum_{i=m}^{n-k+1} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma_k(\mathbf{X}_{i-1})U_{i,k} + o_p(h^2) \tag{27}$$

$$\widehat{\sigma}_k(\mathbf{x}) = \sigma_k(\mathbf{x}) + b_{\sigma,k}(\mathbf{x})h^2$$
$$+\frac{1}{2n\mu(\mathbf{x})\sigma_k(\mathbf{x})} \sum_{i=m}^{n-k+1} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma_k^2(\mathbf{X}_{i-1})(U_{i,k}^2 - 1) + o_p(h^2) \tag{28}$$

Now the estimated GIR function is

$$\widehat{GIR}_k(\mathbf{x}, u) = \widehat{f}_{k-1}(\widehat{\mathbf{x}}_u) - \widehat{f}_k(\mathbf{x})$$
$$= f_{k-1}(\widehat{\mathbf{x}}_u) - f_k(\mathbf{x}) + \left\{b_{f,k-1}(\widehat{\mathbf{x}}_u) - b_{f,k}(\mathbf{x})\right\} h^2 +$$
$$\frac{1}{n\mu(\widehat{\mathbf{x}}_u)} \sum_{i=m}^{n-k+2} K_h(\mathbf{X}_{i-1} - \widehat{\mathbf{x}}_u)\sigma_{k-1}(\mathbf{X}_{i-1})U_{i,k-1}$$
$$-\frac{1}{n\mu(\mathbf{x})} \sum_{i=m}^{n-k+1} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma_k(\mathbf{X}_{i-1})U_{i,k} + o_p(h^2)$$
$$= f_{k-1}(\mathbf{x}_u) - f_k(\mathbf{x}) + \left[b_{f,k-1}(\mathbf{x}_u) - b_{f,k}(\mathbf{x})\right] h^2 +$$
$$\frac{1}{n\mu(\mathbf{x}_u)} \sum_{i=m}^{n-k+2} K_h(\mathbf{X}_{i-1} - \mathbf{x}_u)\sigma_{k-1}(\mathbf{X}_{i-1})U_{i,k-1}$$
$$-\frac{1}{n\mu(\mathbf{x})} \sum_{i=m}^{n-k+1} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma_k(\mathbf{X}_{i-1})U_{i,k} +$$
$$\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \left\{\widehat{f}(\mathbf{x}) - f(\mathbf{x}) + \widehat{\sigma}(\mathbf{x})u - \sigma(\mathbf{x})u\right\} + o_p(h^2)$$
$$= GIR_k(\mathbf{x}, u) + b_{GIR,k}(\mathbf{x}, u)h^2 + T_1 + T_2 + T_3 + T_4 + o_p(h^2) \tag{29}$$

where $b_{GIR,k}(\mathbf{x}, u)$ is as defined in (13) while

$$T_1 = \frac{1}{n\mu(\mathbf{x}_u)} \sum_{i=m}^{n-k+2} K_h(\mathbf{X}_{i-1} - \mathbf{x}_u)\sigma_{k-1}(\mathbf{X}_{i-1})U_{i,k-1}$$

$$T_2 = -\frac{1}{n\mu(\mathbf{x})} \sum_{i=m}^{n-k+1} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma_k(\mathbf{X}_{i-1})U_{i,k}$$

$$T_3 = \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \frac{1}{n\mu(\mathbf{x})} \sum_{i=m}^{n} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma(\mathbf{X}_{i-1})U_i$$

$$T_4 = \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \frac{u}{2n\mu(\mathbf{x})\sigma(\mathbf{x})} \sum_{i=m}^{n} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma^2(\mathbf{X}_{i-1})(U_i^2 - 1) \tag{30a}$$

by Härdle, Tsybakov and Yang (1998). We now consider the expectations of all products T_iT_j, $i, j = 1, \ldots, 4$ which are needed to compute the asymptotic variance. First, one has the following five equations

$$E(T_1^2) = \|K\|_2^{2m} \frac{\sigma_{k-1}^2(\mathbf{x}_u)}{nh^m\mu(\mathbf{x}_u)} + o\left(n^{-1}h^{-m}\right)$$

$$E(T_2^2) = \|K\|_2^{2m} \frac{\sigma_k^2(\mathbf{x})}{nh^m\mu(\mathbf{x})} + o\left(n^{-1}h^{-m}\right)$$

$$E(T_3^2) = \left\{\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\right\}^2 \|K\|_2^{2m} \frac{\sigma^2(\mathbf{x})}{nh^m\mu(\mathbf{x})} + o\left(n^{-1}h^{-m}\right)$$

$$E(T_4^2) = \left\{\frac{u}{2}\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\right\}^2 \|K\|_2^{2m} \frac{\sigma^2(\mathbf{x})(m_4 - 1)}{nh^m\mu(\mathbf{x})} + o\left(n^{-1}h^{-m}\right)$$

$$E(T_3T_4) = \frac{u}{2}\left\{\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\right\}^2 \|K\|_2^{2m} \frac{\sigma^2(\mathbf{x})}{nh^m\mu(\mathbf{x})} m_3 + o\left(n^{-1}h^{-m}\right) \tag{31}$$

Lemma 2

$$E(T_1T_2) = -\frac{\sigma_{k-1,k}(\mathbf{x})I(\mathbf{x} = \mathbf{x}_u)}{nh^m\mu(\mathbf{x})} \|K\|_2^{2m} + o\left(n^{-1}h^{-m}\right)$$

$$E(T_1T_3) = \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \frac{\sigma_{1,k-1}(\mathbf{x})I(\mathbf{x} = \mathbf{x}_u)}{nh^m\mu(\mathbf{x})} \|K\|_2^{2m} + o\left(n^{-1}h^{-m}\right)$$

$$E(T_1T_4) = -\frac{u}{2}\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \frac{\sigma_{11,k-1}(\mathbf{x})I(\mathbf{x} = \mathbf{x}_u)}{nh^m\mu(\mathbf{x})\sigma(\mathbf{x})} \|K\|_2^{2m} + o\left(n^{-1}h^{-m}\right) \tag{32}$$

Proof: We take $i = 3$ as an illustration. By the definitions in (30a)

$$E(T_1T_3) =$$

$$\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \frac{1}{n^2\mu(\mathbf{x})\mu(\mathbf{x}_u)} \sum_{i=m}^{n} \sum_{j=m}^{n-k+2} E\left\{K_h(\mathbf{X}_{i-1} - \mathbf{x})K_h(\mathbf{X}_{j-1} - \mathbf{x}_u)\sigma(\mathbf{X}_{i-1})\sigma_{k-1}(\mathbf{X}_{j-1})U_iU_{j,k-1}\right\}.$$

Take a typical term from the double sum

$$E\left\{K_h(\mathbf{X}_{i-1} - \mathbf{x})K_h(\mathbf{X}_{j-1} - \mathbf{x}_u)\sigma(\mathbf{X}_{i-1})\sigma_{k-1}(\mathbf{X}_{j-1})U_iU_{j,k-1}\right\}$$

and apply change of the random variable $\mathbf{X}_{i-1} = \mathbf{x}+h\mathbf{Z}$, the term becomes

$$\frac{1}{h^m}E\left\{K(\mathbf{Z})K\left(\frac{\mathbf{X}_{j-1}-\mathbf{x}_u}{h}\right)\sigma(\mathbf{x}+h\mathbf{Z})\sigma_{k-1}(\mathbf{X}_{j-1})U_iU_{j,k-1}\right\}.$$

If $i \neq j$, then $\mathbf{X}_{j-1} = (Y_{j-1}, ..., Y_{j-m})^T$ contains variables that are not in $\mathbf{X}_{i-1}$ and so further changes of variable will make the above term of order $O(h^{-m+1})$. If $i < j$, then both $\mathbf{X}_{i-1}$ and U_i are predictable from $Y_{j-1}, ..., Y_{j-m}, ...$ and so by the marthingale property of $U_{j,k-1}$ the above term equals 0. Similarly the term equals 0 if $i > j + k - 2$. Hence, the only nonzero terms satisfy $0 \leq i - j \leq k - 2$, and there are only $O(n)$ such terms. Furthermore, these nonzero terms are of order $O(h^{-m+1})$ unless $i = j$. So one has

$$E(T_1T_3) = O(n^{-1}h^{-m+1})+$$

$$\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu(\mathbf{x})\mu(\mathbf{x}_u)}\sum_{i=m}^{n-k+2}E\left\{K_h(\mathbf{X}_{i-1}-\mathbf{x})K_h(\mathbf{X}_{i-1}-\mathbf{x}_u)\sigma(\mathbf{X}_{i-1})\sigma_{k-1}(\mathbf{X}_{i-1})U_iU_{i,k-1}\right\}.$$

If $\mathbf{x} = \mathbf{x}_u$ then by definition of $\sigma_{1k}(\mathbf{x})$

$$E\left\{\sigma(\mathbf{X}_{i-1})\sigma_{k-1}(\mathbf{X}_{i-1})U_iU_{i,k-1}|\mathbf{X}_{i-1}\right\} = \sigma_{1,k-1}(\mathbf{X}_{i-1})$$

and so

$$\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu(\mathbf{x})\mu(\mathbf{x}_u)}\sum_{i=m}^{n-k+2}E\left\{K_h^2(\mathbf{X}_{i-1}-\mathbf{x})\sigma(\mathbf{X}_{i-1})\sigma_{k-1}(\mathbf{X}_{i-1})U_iU_{i,k-1}\right\}$$

$$=\frac{\partial f_{k-1}(\mathbf{x})}{\partial x_1}\frac{1}{n^2\mu^2(\mathbf{x})}\sum_{i=m}^{n-k+2}E\left\{K_h^2(\mathbf{X}_{i-1}-\mathbf{x})\sigma_{1,k-1}(\mathbf{X}_{i-1})\right\}$$

$$=\frac{\partial f_{k-1}(\mathbf{x})}{\partial x_1}\frac{\|K\|_2^{2m}\sigma_{1,k-1}(\mathbf{x})}{nh^m\mu^2(\mathbf{x})}+o(n^{-1}h^{-m}).$$

If $\mathbf{x} \neq \mathbf{x}_u$, use the same change of variable $\mathbf{X}_{i-1} = \mathbf{x}+h\mathbf{Z}$, one gets

$$\frac{1}{h^{2m}}E\left\{K\left(\frac{\mathbf{X}_{i-1}-\mathbf{x}}{h}\right)K\left(\frac{\mathbf{X}_{i-1}-\mathbf{x}_u}{h}\right)\sigma(\mathbf{X}_{i-1})\sigma_{k-1}(\mathbf{X}_{i-1})U_iU_{i,k-1}\right\}=$$

$$\frac{1}{h^m}E\left\{K(\mathbf{Z})K\left(\frac{\mathbf{x}-\mathbf{x}_u}{h}+\mathbf{Z}\right)\sigma(\mathbf{x}+h\mathbf{Z})\sigma_{k-1}(\mathbf{x}+h\mathbf{Z})U_iU_{i,k-1}\right\}$$

which is of order $o(h^{-m})$ as

$$\sup_{\mathbf{z}\in R^m}K(\mathbf{z})K\left(\frac{\mathbf{x}-\mathbf{x}_u}{h}+\mathbf{z}\right)\to 0$$

The latter follows from the fact that $\mathbf{x} \neq \mathbf{x}_u$ makes the maximum of $\|\mathbf{z}\|$ and $\left\|\frac{\mathbf{x}-\mathbf{x}_u}{h}+\mathbf{z}\right\|$ go to zero uniformly for all $\mathbf{z} \in R^m$, the boundedness of K and that $\lim_{\mathbf{z}\to\infty}K(\mathbf{z}) = 0$. Hence, now one has

$$E(T_1T_3) = O(n^{-1}h^{-m+1}) + o(n^{-1}h^{-m}).$$

Lemma 3

$$E(T_2T_3) = -\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{\sigma_{1k}(\mathbf{x})}{nh^m\mu(\mathbf{x})}\|K\|_2^{2m} + o\left(n^{-1}h^{-m}\right) \tag{33}$$

$$E(T_2T_4) = -\frac{u}{2}\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{\sigma_{11,k}(\mathbf{x})}{nh^m\mu(\mathbf{x})\sigma(\mathbf{x})}\|K\|_2^{2m} + o\left(n^{-1}h^{-m}\right) \tag{34}$$

Proof: We prove (33) as an illustration. By the definitions in (30a)

$$E(T_2T_3) = -\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu^2(\mathbf{x})}\sum_{i=m}^{n}\sum_{j=m}^{n-k+1}E\left\{K_h(\mathbf{X}_{i-1}-\mathbf{x})K_h(\mathbf{X}_{j-1}-\mathbf{x})\sigma(\mathbf{X}_{i-1})\sigma_k(\mathbf{X}_{j-1})U_iU_{j,k}\right\}$$

and by the same reasoning as in Lemma 2, one has

$$E(T_2T_3) = -\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu^2(\mathbf{x})}\sum_{i=m}^{n-k+1}E\left\{K_h^2(\mathbf{X}_{i-1}-\mathbf{x})\sigma(\mathbf{X}_{i-1})\sigma_k(\mathbf{X}_{i-1})U_iU_{i,k}\right\}+o\left(n^{-1}h^{-m}\right)$$

Note that by definition of $\sigma_{1k}(\mathbf{x})$

$$E\left\{\sigma(\mathbf{X}_{i-1})\sigma_k(\mathbf{X}_{i-1})U_iU_{i,k}|\mathbf{X}_{i-1}\right\} = \sigma_{1k}(\mathbf{X}_{i-1})$$

and so

$$\begin{aligned} E(T_2T_3) &= -\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu^2(\mathbf{x})}\sum_{i=m}^{n-k+1}E\left\{K_h^2(\mathbf{X}_{i-1}-\mathbf{x})\sigma_{1k}(\mathbf{X}_{i-1})\right\} + o\left(n^{-1}h^{-m}\right) \\ &= -\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{nh^m\mu(\mathbf{x})}\|K\|_2^{2m}\sigma_{1k}(\mathbf{x}) + o\left(n^{-1}h^{-m}\right) \end{aligned}$$

which is (33).

Lemma 4

$$E(T_1+T_2+T_3+T_4)^2 = n^{-1}h^{-m}\sigma^2_{GIR,k}(\mathbf{x},u) + o\left(n^{-1}h^{-m}\right)$$

where $\sigma^2_{GIR,k}(\mathbf{x},u)$ is as defined in (12).

Proof: This follows from equations (31), (32), (33) and (34), together with

$$E(T_1+T_2+T_3+T_4)^2 = \sum_{i=1}^{4}ET_i^2 + 2\sum_{1\le i<j\le 4}E(T_iT_j).$$

Proof of Theorem 1.

Note that all the four terms T_1, T_2, T_3, T_4 and their linear combinations can be written as sample mean of martingale differences, and so one can apply Corollary 6 of Liptser and Shirjaev (1980). Then using Lemma 4, the asymptotic normal distribution is established.

Proof of Lemma 1.

Note that by definition

$$\left\{Y_{j+k-1}-\widehat{f}_k(\mathbf{X}_{j-1})\right\}^2 = \left\{Y_{j+k-1}-f_k(\mathbf{X}_{j-1})\right\}^2 + \left\{f_k(\mathbf{X}_{j-1})-\widehat{f}_k(\mathbf{X}_{j-1})\right\}^2$$

$$+2\left\{Y_{j+k-1}-f_k(\mathbf{X}_{j-1})\right\}\left\{f_k(\mathbf{X}_{j-1})-\widehat{f}_k(\mathbf{X}_{j-1})\right\} \tag{35}$$

and that

$$\sup_{\mathbf{x}\in\mathbf{C}_\mathbf{X}}\left\{f_k(\mathbf{x})-\widehat{f}_k(\mathbf{x})\right\}^2=o_p(h^2)$$

and so one can drop the second term when smoothing $\mathbf{V}_k$ in the decomposition (35). Since

$$Y_{j+k-1}-f_k(\mathbf{X}_{j-1})=\sigma_k(\mathbf{X}_{j-1})U_{j,k}$$

so instead of $\mathbf{V}_k$, one smoothes local linearly a vector whose terms are

$$\sigma_k^2(\mathbf{X}_{j-1})U_{j,k}^2+2\sigma_k(\mathbf{X}_{j-1})U_{j,k}\left\{f_k(\mathbf{X}_{j-1})-\widehat{f}_k(\mathbf{X}_{j-1})\right\}=$$

$$\sigma_k^2(\mathbf{X}_{j-1})U_{j,k}^2+2\sigma_k(\mathbf{X}_{j-1})U_{j,k}\left\{b_{f,k}(\mathbf{X}_{j-1})h^2+\frac{1}{n\mu(\mathbf{X}_{j-1})}\sum_{i=m}^{n}K_h(\mathbf{X}_{i-1}-\mathbf{X}_{j-1})\sigma_k(\mathbf{X}_{i-1})U_{i,k}\right\}+o_p(h^2)$$

Now obviously

$$\frac{2h^2}{n\mu(\mathbf{x})}\sum_{j=m}^{n}K_h(\mathbf{X}_{j-1}-\mathbf{x})b_{f,k}(\mathbf{X}_{j-1})\sigma_k(\mathbf{X}_{j-1})U_{j,k}=o_p(h^2)$$

so one only needs to smooth the following term local linearly on $\mathbf{X}_{j-1}=\mathbf{x}$:

$$\sigma_k^2(\mathbf{X}_{j-1})U_{j,k}^2+\frac{2\sigma_k(\mathbf{X}_{j-1})U_{j,k}}{n\mu(\mathbf{X}_{j-1})}\sum_{i=m}^{n}K_h(\mathbf{X}_{i-1}-\mathbf{X}_{j-1})2\sigma_k(\mathbf{X}_{i-1})U_{i,k}.$$

By using the geometric mixing conditions as in Härdle, Tsybakov and Yang (1998), local linear smoothing of $\sigma_k^2(\mathbf{X}_{j-1})U_{j,k}^2$ gives the two terms on the right hand side of (21) except the higher order term, so it remains to show that local linear smoothing of the following term is $o_p(h^2)$:

$$\frac{2\sigma_k(\mathbf{X}_{j-1})U_{j,k}}{n\mu(\mathbf{X}_{j-1})}\sum_{i=m}^{n}K_h(\mathbf{X}_{i-1}-\mathbf{X}_{j-1})2\sigma_k(\mathbf{X}_{i-1})U_{i,k}.$$

Writing explicitly the local linear smoothing, one needs to show that

$$\frac{2}{n^2\mu(\mathbf{x})}\sum_{m\le i,j\le n}T_{ij}=\sum_{\gamma=1}^{2}S_\gamma=o_p(h^2)$$

where

$$T_{ij}=\left\{\frac{K_h(\mathbf{X}_{j-1}-\mathbf{x})}{\mu(\mathbf{X}_{j-1})}+\frac{K_h(\mathbf{X}_{i-1}-\mathbf{x})}{\mu(\mathbf{X}_{i-1})}\right\}K_h(\mathbf{X}_{i-1}-\mathbf{X}_{j-1})\sigma_k(\mathbf{X}_{i-1})\sigma_k(\mathbf{X}_{j-1})U_{i,k}U_{j,k}$$

$$S_1=\frac{2}{n^2\mu(\mathbf{x})}\sum_{m\le i\le n}T_{ii}=\frac{2}{n^2\mu(\mathbf{x})}\sum_{j=m}^{n}\frac{1}{\mu(\mathbf{X}_{j-1})}K_h(\mathbf{X}_{j-1}-\mathbf{x})K_h(\mathbf{0})\sigma_k^2(\mathbf{X}_{j-1})U_{j,k}^2$$

$$S_2=\frac{2}{n^2\mu(\mathbf{x})}\sum_{m\le i<j\le n}T_{ij}$$

It is easy to verify that $S_1 = O(n^{-1}h^{-m})$ by Corollary 6 of Liptser and Shirjaev (1980). It is also clear that $E(T_{ij}T_{i'j'}) = 0$ for all $m \leq i < j \leq n, m \leq i' < j' \leq n, j \neq j$. Thus

$$ES_2^2 = \frac{4}{n^4\mu^2(\mathbf{x})} \sum_{m\leq i<j\leq n} E(T_{ij}^2) + \frac{8}{n^4\mu^2(\mathbf{x})} \sum_{m\leq i<i'<j\leq n} E(T_{ij}T_{i'j})$$

Now let $k_n = [c\ln n]$ be such that $\beta(k_n) \leq n^{-4}$, then

$$\frac{4}{n^4\mu^2(\mathbf{x})} \sum_{m\leq i<j\leq n} E(T_{ij}^2) = \frac{4}{n^4\mu^2(\mathbf{x})} \left(\sum_{m\leq i<j-k_n<j\leq n} + \sum_{m\leq j-k_n\leq i<j\leq n} \right) E(T_{ij}^2)$$

$$\leq \frac{4}{n^4\mu^2(\mathbf{x})} \sum_{m\leq i<j-k_n<j\leq n} C\frac{h^{2m}}{h^{4m}} + \frac{4}{n^4\mu^2(\mathbf{x})} \sum_{m\leq j-k_n\leq i<j\leq n} C\frac{h^{m+1}}{h^{4m}}$$

$$= O(n^{-2}h^{-2m} + n^{-3}k_n h^{1-3m}) = O(n^{-1}h^{-m}) = o(h^4). \tag{36}$$

Meanwhile $\sum_{m\leq i<i'<j\leq n} E(T_{ij}T_{i'j})$ is decomposed into also two parts: part 1 consists of those terms with $\max(i'-i, j-i') > k_n$ while part 2 those terms with $\max(i'-i, j-i') \leq k_n$. Then it is clear that terms in part 1 can be treated as if $U_{i,k}$ or $U_{i',k}$ is independent of the other variables index around j or j', with negligible errors, so part 1 is of smaller order than n^4h^4. Part 2 has at most $O(nk_n^2)$ terms, so it is at most $O(nk_n^2h^{1-3m}) = o(n^4h^4)$. Hence we have proved that

$$\frac{8}{n^4\mu^2(\mathbf{x})} \sum_{m\leq i<i'<j\leq n} E(T_{ij}T_{i'j}) = o_p(h^4). \tag{37}$$

Combining (36) and (37), we have shown that

$$S_1 + S_2 = o_p(h^2)$$

and thus also the lemma.

7 Acknowledgements

Both authors received financial support from Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse", Humboldt-Universität zu Berlin. Lijian Yang's research was also partially supported by NSF grant DMS 9971186.

References

Chen, R., Yang, L. and Hafner, C. (1999) Nonparametric multi-step ahead prediction in time series analysis, preprint.

Gallant, A.R., Rossi, P.E. and Tauchen, G. (1993) Nonlinear dynamic structure, *Econometrica* 61, 871-908.

Härdle, W. and Tsybakov, A. (1997), Local polynomial estiomators of the volatility function in nonparametric autoregression, *Journal of Econometrics* 81, 223–242.

Härdle, W., Lütkepohl, H. and Chen, R. (1997) A review of nonparametric time series analysis. *International Statistical Review* 65, 49-72.

Härdle, W., Tsybakov, A. B. and Yang, L. (1998) Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* 68, 221-245.

Koop, G., Pesaran, M.H. and Potter, S.M. (1996) Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* 74, 119 - 147.

Liptser, R. Sh. and Shirjaev, A. N. (1980) A functional central limit theorem for martingales. *Theory Probab. Appl.* 25, 667-688.

Silverman, B. (1986), *Density estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Masry, E. and Tjøstheim, D. (1995) Nonparametric estimation and identification of nonlinear ARCH time series, *Econometric Theory* 11, 258–289.

Tschernig, R. and Yang, L. (2000) Nonparametric lag selection for time series, *Journal of Time Series Analysis*, in press.

Tjøstheim, D. (1994) Non-linear time series analysis: a selective review. *Scandinavian Journal of Statistics* 21, 97-130.

Tjøstheim, D. and Auestad, B. (1994) Nonparametric identification of nonlinear time series: selecting significant lags. *Journal of the American Statistical Association* 89, 1410-1419.

Yao, Q. and Tong, H. (1994) On subset selection in non-parametric stochastic regression. *Statistica Sinica* 4, 51-70.

Yang, L. and Tschernig, R. (1999) Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society, Series B* 61, 793 - 815.

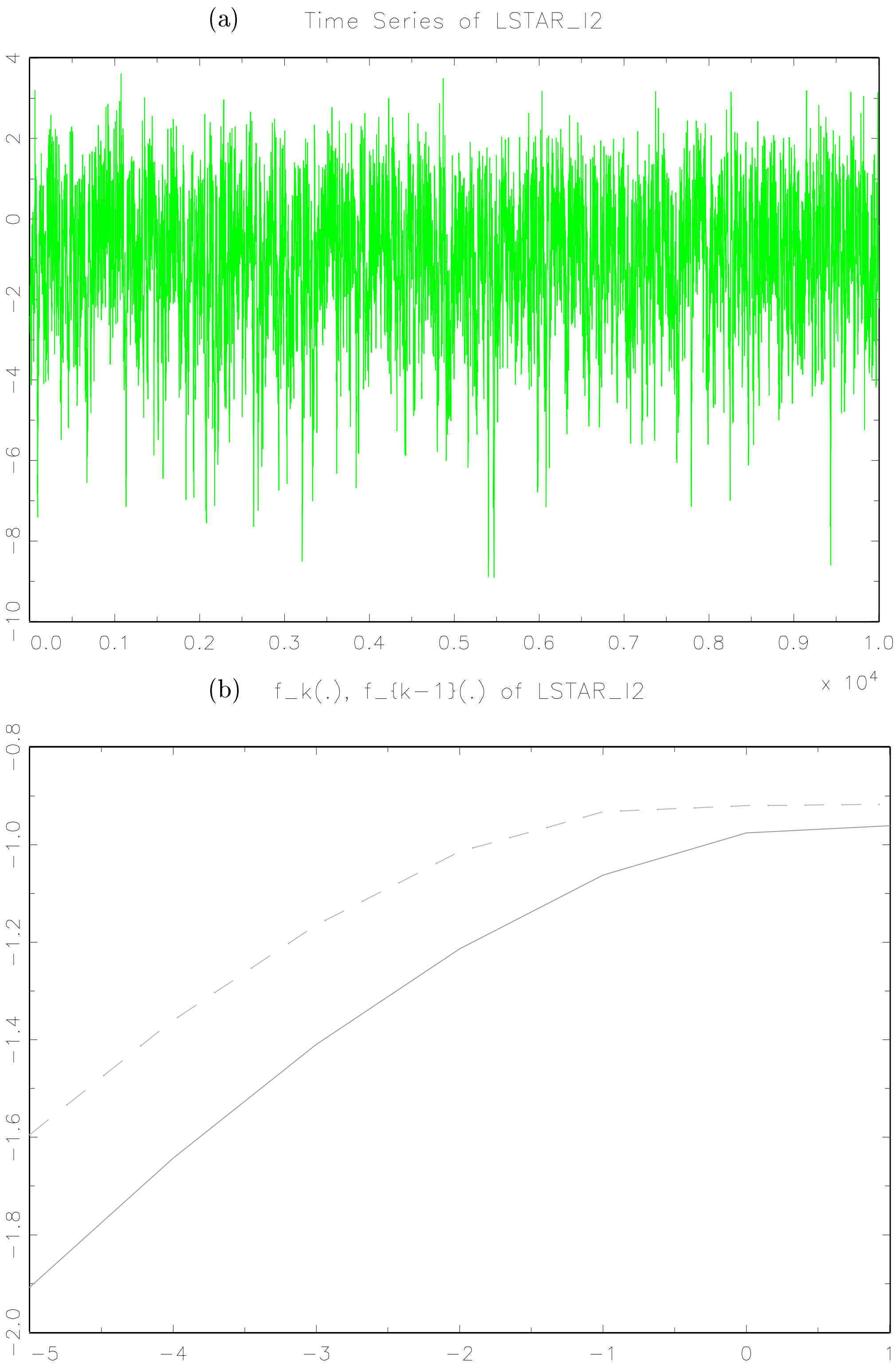

Figure 1: a) Realization of 500 observations of the logistic autoregressive process; b) True k-step and $k-1$-step ahead prediction functions for various $\mathbf{x}$ for the logistic autoregressive process.

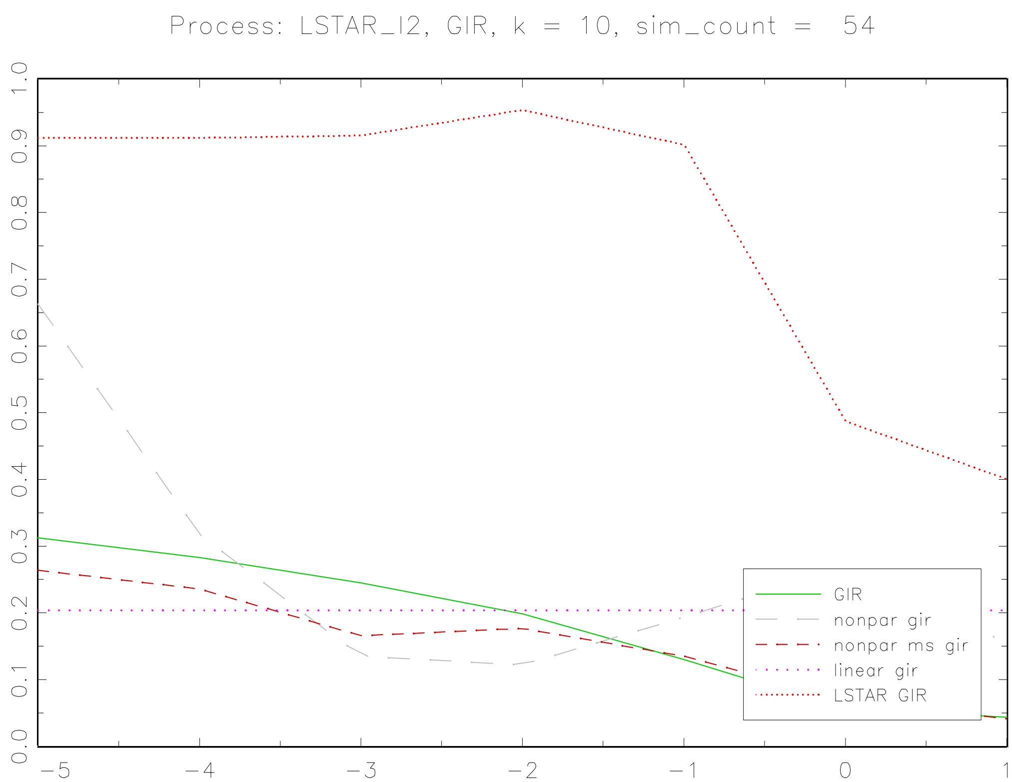

Figure 2: Various estimates of generalized impulse response functions for the logistic autoregressive process for 10 periods ahead, $u = 1$ and various $\mathbf{x}$: through line: true GIR, long dashes: local linear estimator, short dashes: local linear estimator using multi-stage prediction, dotted line: estimated impulse responses of linear AR model, short dotted line: GIR based on estimated correctly specified logistic autoregressive model.