Nonparametric Estimation of Generalized Impulse Response Functions

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

October 2000

Abstract

A local linear estimator of generalized impulse response (GIR) functions for nonlinear conditional heteroskedastic autoregressive processes is derived and shown to be asymptotically normal. A plug-in bandwidth is obtained that minimizes the asymptotical mean squared error of the GIR estimator. 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 the use of multi-stage prediction techniques is suggested. Simulation experiments show the latter estimator to perform best. For quarterly data of the West German real GNP it is found that the size of generalized impulse response functions varies across different histories, a feature which cannot be captured by linear models.

KEY WORDS: Confidence intervals; general impulse response function; heteroskedasticity; local polynomial; multi-stage 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 conditional heteroskedastic autoregressive nonlinear (CHARN) 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 of

*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.

the conditional mean function f under an α-mixing assumption. Härdle, Tsybakov and Yang (1998) proved asymptotic normality for the local 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 on nonparametric time series analysis, see 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 only the conditional mean function. This is even more so if the conditional mean function is nonlinear. One appropriate tool that allows to study the dynamics of processes like (1) is the generalized impulse response function.

In this paper we propose nonparametric estimators for generalized impulse response (GIR) functions for CHARN 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

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

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 multi-step ahead prediction functions which are contained in GIR_k and derive the asymptotic properties of the resulting plug-in estimator of GIR_k. This also delivers an asymptotically optimal bandwidth allowing to compute a plug-in bandwidth. The conditional standard deviation σ in GIR_k can be estimated with the local linear volatility estimator of Härdle and Tsybakov (1997). Alternatively, we propose a simpler local linear estimator based on a "de-meaning" idea that is asymptotically as efficient as if the true conditional mean function is known. The prediction functions can be estimated either directly or via the multi-stage prediction techniques of Chen, Yang and Hafner (2000). For two CHARN processes we compare the performance of the nonparametric direct and multi-stage GIR_k estimators with parametric ones using Monte Carlo simulations. The multi-stage GIR_k estimator with improved volatility estimation shows the best overall performance. It is then used to investigate the dynamics of the West German real GNP.

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 Monte Carlo study are summarized in Section 6. Section 7 presents the empirical analysis and Section 8 concludes.

2 DIRECT GIR ESTIMATION

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

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

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

The plug-in estimate of the GIR_k function in (8) is then

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

where all unknown functions are replaced by local linear estimates. The estimator of $\mathbf{x}_u$ is $\hat{\mathbf{x}}_u = \left\{\hat{f}(\mathbf{x}) + \hat{\sigma}(\mathbf{x})u, \mathbf{x}'\right\}$. 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 $h = \beta n^{-1/(m+4)}, \beta > 0$. Define further the matrices

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

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

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

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

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). Closed formulae for optimal bandwidth are derived in the corollary. 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*

$$\sigma_{GIR,k}^2(\mathbf{x}, u) = \frac{\|K\|_2^{2m} \sigma^2(\mathbf{x})}{\mu(\mathbf{x})} \left[\frac{\sigma_{k-1}^2(\mathbf{x}_u)\mu(\mathbf{x})}{\mu(\mathbf{x}_u)\sigma^2(\mathbf{x})} + \frac{\sigma_k^2(\mathbf{x})}{\sigma^2(\mathbf{x})} + \right.$$
$$\left. \left\{ \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \right\}^2 \left\{ 1 + u m_3 + \frac{u^2(m_4 - 1)}{4} \right\} - \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \left\{ \frac{2\sigma_{1,k}(\mathbf{x})}{\sigma^2(\mathbf{x})} + u \frac{\sigma_{11,k}(\mathbf{x})}{\sigma^3(\mathbf{x})} \right\} \right]$$
$$- \frac{\|K\|_2^{2m}}{\mu(\mathbf{x})} I(\mathbf{x} = \mathbf{x}_u) \left\{ 2\sigma_{k-1,k}(\mathbf{x}) - 2\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \sigma_{1,k-1}(\mathbf{x}) + u \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1} \frac{\sigma_{11,k-1}(\mathbf{x})}{\sigma(\mathbf{x})} \right\} \tag{12}$$

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

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

$\operatorname{Tr}\left\{ \nabla^2 f_k(\mathbf{x}) \right\}$ *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*

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

Corollary 1 *The optimal bandwidth for estimating* $GIR_k(\mathbf{x}, u)$ *is*

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

which asymptotically minimizes the mean squared error (MSE)

$$MSE_k(\mathbf{x}, u; h) = E\left\{ \widehat{GIR}_k(\mathbf{x}, u) - GIR_k(\mathbf{x}, u) \right\}^2.$$

For any compact subset $\mathbf{C}_\mathbf{x}$ *of* R^m*, the minimizer of the mean integrated squared error (MISE)*

$$MISE_k(\mathbf{C}_\mathbf{x}, u; h) = \int_{\mathbf{C}_\mathbf{x}} E\left\{ \widehat{GIR}_k(\mathbf{x}, u) - GIR_k(\mathbf{x}, u) \right\}^2 \mu(\mathbf{x}) d\mathbf{x}$$

is asymptotically

$$h_{opt}(\mathbf{C}_\mathbf{x}, u) = \left\{ \frac{m \int_{\mathbf{C}_\mathbf{x}} \sigma_{GIR,k}^2(\mathbf{x}, u)\mu(\mathbf{x}) d\mathbf{x}}{4n \int_{\mathbf{C}_\mathbf{x}} b_{GIR,k}^2(\mathbf{x}, u)\mu(\mathbf{x}) d\mathbf{x}} \right\}^{1/(m+4)}. \tag{17}$$

Obviously, each of the optimal bandwidth formulas (16) and (17) contains unknown quantities. In Section 5 we discuss estimators for those quantities in order to obtain a plug-in version of the optimal bandwidth (16). This plug-in bandwidth is then used in the Monte Carlo experiments and in the empirical analysis presented in Sections 6 and 7, respectively.

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 $\mathbf{C_x}$. Denoting by C_u a compact subset of R, the generalized impulse response function over these compact sets is defined by

$$GIR_k(\mathbf{C_x}, C_u) = E\left\{GIR_k(\mathbf{X}_{i-1}, U_i)|\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u\right\}. \tag{18}$$

For its estimation, we consider its empirical version

$$\widehat{GIR}_k(\mathbf{C_x}, C_u) = \frac{1}{n\widehat{P}\left(\mathbf{C_x}, C_u\right)} \sum_{i=m}^{n-k+1} \widehat{GIR}_k(\mathbf{X}_{i-1}, U_i) I\left(\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u\right) \tag{19}$$

where

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

The asymptotic properties of the estimator (19) for generalized impulse response functions over compact sets $(\mathbf{C_x}, C_u)$ are summarized in the next theorem.

Theorem 2 *Under assumptions (A1)-(A4) given in the Appendix*

$$\widehat{GIR}_k(\mathbf{C_x}, C_u) - GIR_k(\mathbf{C_x}, C_u) = b_{GIR,k}(\mathbf{C_x}, C_u)h^2 + o_p(h^2) \tag{20}$$

where

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

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, its application may be problematic in finite samples. In the next two sections we discuss this problem in more detail and present an improved estimator.

3 EFFICIENT VOLATILITY ESTIMATION

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, has an asymptotic bias involving the mean function f, as seen in (14), and hence may perform poorly due to the influence of f. This problem can also occur for other auxiliary functions such as $\widehat{\sigma}_k(\mathbf{x})$ and $\widehat{\sigma}_{1,k}(\mathbf{x})$,etc., which will be needed for computing the plug-in bandwidth based on formulas (16) or (17). In this section we present an alternative local linear estimator for the conditional standard deviation that is asymptotically as accurate as if the true function f is known. 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

$$\widetilde{\sigma}_k^2(\mathbf{x}) = e^T \left(\mathbf{Z}_k^T W_k \mathbf{Z}_k\right)^{-1} \mathbf{Z}_k^T W_k \mathbf{V}_k \tag{21}$$

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 theorem it is shown that this approach is indeed useful.

Theorem 3 *Under assumptions (A1)-(A4) in the Appendix, one has*

$$\widetilde{\sigma}_k^2(\mathbf{x}) - \sigma_k^2(\mathbf{x}) = \widetilde{b}_{\sigma,k}(\mathbf{x})h^2 + \frac{1}{n\mu(\mathbf{x})} \sum_{j=m}^{n} K_h(\mathbf{X}_{j-1} - \mathbf{x})\sigma_k^2(\mathbf{X}_{j-1})(U_{j,k}^2 - 1) + o_p(h^2) \tag{22}$$

where

$$\widetilde{b}_{\sigma,k}(\mathbf{x}) = \frac{\sigma_K^2}{2} \operatorname{Tr} \nabla^2 \left\{\sigma_k^2(\mathbf{x})\right\} \tag{23}$$

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 theorem basically says that by "de-meaning" one can estimate $\sigma_k^2(\mathbf{x})$ as well as if one knew the true k-step prediction 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 (22) and (29). However, from comparing $b_{\sigma,k}$ and $\widetilde{b}_{\sigma,k}$ given by (14) and (23), 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). For example, one can 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

$$\mathbf{V}_{11,k} =$$

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

and likewise $\sigma_{1,k}(\mathbf{x})$. Under assumptions (A1)-(A3) in the Appendix, these respective estimators have similar properties as $\widetilde{\sigma}_k^2(\mathbf{x})$.

The fact that $\widetilde{\sigma}_k(\mathbf{x})$ has a simpler bias that does not involve f_k facilitates the computation of the plug-in bandwidth since the asymptotic bias term in the optimal bandwidth (16) becomes much simpler as well. For this reason and also the fact that it is more efficient, we use from now on in the GIR_k estimator (9) the new estimator (21) instead of (11) for estimating conditional volatilities. We note that despite of the fact that $\widetilde{\sigma}_k^2(\mathbf{x})$ is obtained by smoothing positive quantities, it still can take negative values in finite samples. In such cases, one replaces in (21) the local linear by either the local constant (Nadaraya-Watson) estimator or even a homoskedastic estimator which always produces positive estimates.

4 MULTI-STAGE GIR ESTIMATION

The main ingredients 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 of Chen, Yang and Hafner (2000). 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 $Y_{t+j}^{(j)} = \widetilde{f}_j(\mathbf{X}_t)$.

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

$$\widetilde{f}_k(\mathbf{x}) = \frac{\sum_{t=m-1}^{n-k} K_{h_k}(\mathbf{X}_t - \mathbf{x}) Y_{t+k}^{(k-1)}}{\sum_{t=m-1}^{n-k} K_{h_k}(\mathbf{X}_t - \mathbf{x})}. \tag{24}$$

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 (2000).

Theorem 4 *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

$$\widetilde{GIR}_k(\mathbf{x}, u) = \widetilde{f}_{k-1}(\widetilde{\mathbf{x}}_u) - \widetilde{f}_k(\mathbf{x}) \tag{25}$$

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 (24) and (21), respectively. In the next section we turn to issues of implementation.

5 IMPLEMENTATION

Computing the direct or multi-stage GIR estimators (9) or (25) requires suitable bandwidth estimates. Both estimators use the Gaussian kernel and a plug-in bandwidth $\widehat{h}_{opt}$ which is obtained by consistently estimating the unknown quantities in the asymptotically optimal bandwidth (16). Since the efficient volatility estimator (21) is used instead of the standard estimator (11), the bias term (14) in the optimal bandwidth (16) is replaced by (23). For estimating the second order direct derivatives in the bias (13) we use a partial quadratic estimator with bandwidth $h\left\{m+4, 3\sqrt{\widehat{Var}(\mathbf{X})}\right\}$ with $\widehat{Var}(\mathbf{X})$ denoting the geometric mean of the variances for each regressor and

$$h(k,\sigma) = \sigma\{4/k\}^{1/(k+2)} n^{-1/(k+2)}.$$

The partial quadratic estimator is a simplified version of the partial cubic estimator presented in Yang and Tschernig (1999). For estimating all other unknown functions in $\sigma^2_{GIR,k}$ and $b_{GIR,k}$ we employ a plug-in bandwidth for estimating the conditional mean function, see Tschernig and Yang (2000, Section 5) for details of implementation. The main consideration for these bandwidth choices is that they are of the right order.

For the multi-stage GIR_k estimator (25) there does not exist a scalar optimal bandwidth. According to Chen, Yang and Hafner (2000) 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. However, in simulations the application of $h_{MS,k-1}$ was often found to produce too small bandwidths and thus causing singular matrices in the local linear estimator. The plug-in bandwidth $\widehat{h}_{opt}$ may be smaller than the optimal bandwidth for exclusively estimating f_k since it serves several estimation purposes simultaneously. We therefore use $\widehat{h}_{opt}$ for all steps.

These quantities are also used for computing confidence intervals based on (15). All computations are carried out in GAUSS.

6 MONTE CARLO STUDY

In this section we investigate the performance of the proposed GIR_k estimators for two conditional heteroskedastic autoregressive nonlinear (CHARN) processes (1), each with two lags, 300 observations and i.i.d. standard normal errors U_t:

CHARN1

$$Y_t = -0.4\frac{3-Y_{t-1}^2}{1+Y_{t-1}^2} + 0.6\frac{3-(Y_{t-2}-0.5)^3}{1+(Y_{t-2}-0.5)^4} + 0.1\,\sigma(Y_{t-1})U_t,$$

with

$$\sigma^2(y) = 0.8 + 0.4\left(0.1 + 2\frac{1}{1+\exp(5y)}\right)y^2 + 0.4\left(0.1 + 2\Phi(-5y)\right)y^2$$

and Φ denoting the c.d.f. of the standard normal distribution.

CHARN2

$$Y_t = 0.7Y_{t-1} - 0.2Y_{t-2} + (-0.3Y_{t-1} + 0.7Y_{t-2})\frac{1}{1+\exp\{-10(-Y_{t-1} - 0.02)\}} + 0.5\sigma(Y_{t-1})U_t,$$

with

$$\sigma^2(y) = 0.25 + 0.75\frac{y^2}{1+y^2}.$$

A plot of one realization of the **CHARN1** process, of $\sigma(y)$ and of the density on the range of the realizations are shown in Figure 1(a) to (c). For the **CHARN2** process the corresponding plots are displayed in Figure 1(d) to (f).

For illustration we computed $GIR_4(\mathbf{x}, 1)$ functions on a two-dimensional grid for $\mathbf{x}$ using the simulation method described in Koop, Pesaran and Potter (1996). Figure 2(a) displays the resulting surface for the **CHARN1** process where all grid points outside the range of one realization of 300 observations are ignored. Figure 2(b) shows the surface of the multi-stage GIR estimates for the 389 relevant grid points. For the **CHARN2** process the corresponding plots are shown in Figure 2(c) and (d).

While these estimates seem to be encouraging, we conducted a simulation study to obtain a more precise evaluation of the suggested methods. To save computation time, we only considered about 50 different histories $\mathbf{x}$. They were obtained by taking subsequent observations of one realization for each process and discarding 5% of those observations for which the density is lowest. The density is estimated using 10000 observations. Based on 100 replications we computed the mean squared errors for various estimators of the generalized impulse response function $GIR_k(\mathbf{x}, u)$ with $k = 4, 7$ and 10 and $u = -1, 1$. We considered both the one-stage estimator (9) with (10) and (21) and the multi-stage estimator (25) based on (24) and (21). We also fitted a linear homoskedastic AR(2) model and computed the corresponding generalized impulse responses. Finally, for the **CHARN2** process we computed the generalized impulse responses based on the estimated parameters of the correct parameterization of the **CHARN2** mean function. This last exercise is not possible for the **CHARN1** process due to identifiability problems of the parameters.

Since presenting the mean squared errors of the $GIR_k(\mathbf{x}, u)$ for each $k, \mathbf{x}, u$ of one process involves 300 numbers, we decided to summarize this information into the mean integrated squared error $\sum_j GIR_k(\mathbf{x}^j, u)$. Inspecting the mean integrated squared errors for the **CHARN1** process in Table 1 indicates that both nonparametric estimators perform substantially better than the linear estimator based on a misspecified homoskedastic AR(2) model. Overall, the multi-stage GIR estimator shows the smallest mean integrated squared error. The superiority of the multi-stage GIR estimator over the direct GIR and the linear estimator is also found for the **CHARN2** process although it is now less dramatic with respect to the linear estimator, see Table 2. Note that the multi-stage GIR estimator also outperforms the nonlinear parametric GIR estimator except for $k = 7$ and $u = -1$.

The results of this simulation study suggest that the proposed multi-stage estimator can be useful in practice if one expects the underlying process to exhibit substantial nonlinearities or heteroskedasticity or both and the functional form is unknown. In the next section it will be applied to a typical macroeconomic time series problem.

Table 1: Mean integrated squared errors ($*10^{-3}$) of generalized impulse response estimators for the **CHARN1** process

horizon k	4		7		10	
Estimator \ shock u	-1	1	-1	1	-1	1
multi-stage GIR	5.945	3.754	5.397	8.443	3.952	6.381
direct GIR	8.557	5.382	8.388	8.967	8.918	7.902
linear AR(2)	28.297	28.337	16.931	16.880	7.008	8.013

Table 2: Mean integrated squared errors ($*10^{-3}$) of generalized impulse response estimators for the **CHARN2** process

horizon k	4		7		10	
Estimator \ shock u	-1	1	-1	1	-1	1
multi-stage GIR	1.606	1.495	2.273	1.808	1.280	1.267
direct GIR	3.085	2.730	3.671	3.421	3.836	3.490
linear AR(2)	2.261	1.800	2.883	2.695	1.960	1.909
nonlinear AR(2)	1.700	1.631	1.784	1.818	1.823	1.850

7 An empirical application

For the analysis of business cycles linear models may be inadequate if the dynamics vary with the stage of the cycle. Potential explanations include asymmetric adjustment costs of labor (see Hamermesh and Pfann (1996) for a survey) or recessions as cleansing periods (see, for example, Caballero and Hammour (1994)) or the insider-outsider theory (Lindbeck and Snower (1988)).

In general, the empirical analysis of relevant macroeconomic time series is based on parametric models such as the smooth transition model (e.g. Skalin and Teräsvirta (1998)) which incorporates seasonal features of macroeconomic time series. On the other hand, it is not easy to choose an appropriate class of nonlinear models. For the latter reason, Yang and Tschernig (1998) extend the CHARN model (1) by deterministic seasonal components. Let S denote the number of seasons. The simplest seasonal model which they propose for a seasonal process V_t is the seasonal shift model

$$\begin{aligned} V_{s+\tau S} - \delta_s &= f\left(V_{s+\tau S-1} - \delta_{\{s-1\}}, \dots, V_{s+\tau S-m} - \delta_{\{s-m\}}\right) \\ &+ \sigma\left(V_{s+\tau S-1} - \delta_{\{s-1\}}, \dots, V_{s+\tau S-m} - \delta_{\{s-m\}}\right) U_{s+\tau S} \end{aligned}$$

where the time index t is replaced by $s+\tau S$, $s = 0, 1, \dots, S-1$ and $\tau = 0, 1, \dots$, and the δ_s denote seasonal mean shifts. For any integer a we define $\{a\}$ as the unique integer between 0 and $S-1$ that is in the same congruence class as a modulo S. For identifiability one requires $\delta_0 = 0$. In the following we estimate the GIR function for the CHARN process $Y_{s+\tau S} = V_{s+\tau S} - \delta_s$. Yang and Tschernig (1998) show that for the purpose of nonparametric inference the deseasonalized Y_t can be obtained by subtracting the estimated δ_s's.

We use this model for the analysis of the quarterly (seasonally non-adjusted) West German real GNP from 1960:1 to 1990:4 compiled by Wolters (1992, p. 424, note 4). Based on seasonal unit root testing by Franses (1996) we take the first differences of the logs. By subtracting the estimated seasonal means $\hat{\delta}_1$ to $\hat{\delta}_3$, the deseasonalized Y_t's are growth rates with respect to the spring season. In order to avoid the dependence on a specific season, we ignore the identifiability issue and subtract all four means $\hat{\delta}_0 = 0.0386$, $\hat{\delta}_1 = 0.0518$, $\hat{\delta}_2 = 0.0089$ and $\hat{\delta}_3 = -0.0673$.

To keep the model parsimonious we employ a CHARN model with all lags up to four. Since we have more than two lags, we can no longer investigate the generalized impulse response function on a grid as it was done in the previous section. Instead, we pick six distinct histories: $\mathbf{x}^1 = (-0.02, -0.01, 0, 0.01)^T$, $\mathbf{x}^2 = (-0.01, 0, 0.01, 0.02)^T$, $\mathbf{x}^3 = (0, 0.01, 0.02, 0.01)^T$, $\mathbf{x}^4 = (0.01, 0, -0.01, -0.02)^T$, $\mathbf{x}^5 = (0.02, 0.01, 0, -0.01)^T$, $\mathbf{x}^6 = (0.03, 0.02, 0.01, 0)^T$ which represent various stages of the business cycle ranging from a substantial downswing to a considerable upswing. These growth rates correspond to the deseasonalized Y_t process.

In Figure 3 we display for each history and a positive unit shock the multi-stage $GIR_k(\mathbf{x}, 1)$ estimator (solid line) plus 95% confidence intervals (dots and dashes) and the $GIR_k(\mathbf{x}, 1)$ estimator based on a homoskedastic linear model (dashed line) for $k = 3, \ldots, 20$ or up to five years. Note that the unit shock is multiplied with the estimated conditional standard deviation which for the given histories $\mathbf{x}^i$, $i = 1, \ldots, 6$ takes values in the range from 0.008 to 0.016. Both models exhibit strong seasonal dynamics. Subtracting seasonal means cannot remove all seasonal effects. The overall dynamics implied by both estimators are qualitatively similar. For history $\mathbf{x}^4$, they are basically identical, see Figure 3(d). For histories that include a 2% growth of the deseasonalized real GNP, Figures 3(b), (c), (e) and (f) indicate that for the first two years the linear model would overestimate the generalized impulse responses of a unit shock. Taking absolute values, these results also hold for a negative unit shock as can be seen from Figure 4. Such differences cannot be revealed by GIR estimates based on linear models.

8 Conclusion

Impulse responses have proven important to study the dynamics of linear time series processes. For conditional heteroskedastic autoregressive nonlinear processes we provide local linear estimators of generalized impulse response functions as defined by Koop, Pesaran and Potter (1996). Asymptotic normal distributions are derived for the proposed nonparametric estimators without prior choice of the parametric forms of the process. An efficient new estimator of the conditional variance function is proposed based on a "de-meaning" idea. Plug-in optimal bandwidths are obtained and implemented, and multi-stage prediction techniques are used for enhanced performance.

In a simulation study we compare the direct and multi-stage GIR estimators with a linear parametric estimator for two conditional heteroskedastic autoregressive nonlinear processes of order two and find the multi-stage GIR estimator to perform best in terms of its mean integrated squared error.

Finally we investigate quarterly data of the West German real GNP using the multi-stage GIR estimator and a GIR estimator based on a linear model after subtracting seasonal means. For six distinct histories it is found that the magnitude of the nonparametrically estimated generalized impulse response functions differ across histories, a feature that is completely missed by linear models. The responses are smaller if there was considerable growth in at least one quarter. Based on the confidence intervals which are computed using the asymptotic distribution, these differences are significant.

In sum, we may conclude from the results of the Monte Carlo study and the empirical analysis that the proposed nonparametric multi-stage estimator of generalized impulse response functions can be a useful tool for studying nonlinear phenomena in economics and other fields.

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: $\beta(n) \leq c_0 \rho^{-n}$ 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.

(A3) The functions f and σ have bounded continuous derivatives up to order 4 and σ is positive on the support of μ.

(A4) There exists constants $a, r > 0$ such that $E \exp\{a |Y_0|^r\} < +\infty$.

A discussion of assumptions (A1) to (A3) can be found e.g. in Tschernig and Yang (2000). Assumption (A3) is needed for the functions f_k, σ_k^2 to be 4-th order smooth, as shown in the lemma that follows. The 4-th order smoothness is needed for using the plug-in bandwidths of Yang and Tschernig (1999).

Lemma 1 *Under assumptions (A1)-(A3), one has for* $t \geq m$ *and* $k \geq 2$

$$f_k(\mathbf{x}) = E f_{k-1}\{f(\mathbf{x}) + \sigma(\mathbf{x})U_t, \mathbf{x}'\}, \quad (26)$$

$$\sigma_k^2(\mathbf{x}) = E f_{k-1}^2\{f(\mathbf{x}) + \sigma(\mathbf{x})U_t, \mathbf{x}'\} - f_k^2(\mathbf{x}) + E\sigma_{k-1}^2\{f(\mathbf{x}) + \sigma(\mathbf{x})U_t, \mathbf{x}'\}. \quad (27)$$

Moreover, all functions $f_k, \sigma_k^2, k = 2, 3, ...$ *have continuous derivatives up to order* 4.

Proof. By the definitions in (3), (5) and (4), for any $t \geq m$, one has

$$Y_{t+k-1} = f_{k-1}(\mathbf{X}_t) + \sigma_{k-1}(\mathbf{X}_t)U_{t+1,k-1}$$

and hence

$$f_k(\mathbf{x}) = E(Y_{t+k-1}|\mathbf{X}_{t-1} = \mathbf{x}) = E\left[f_{k-1}(\mathbf{X}_t)\,|\mathbf{X}_t = \{f(\mathbf{x}) + \sigma(\mathbf{x})U_t, \mathbf{x}'\}\right]$$

which is the same as in (26). Likewise, using the martingale property of $U_{t+1,k-1}$, one has

$$\sigma_k^2(\mathbf{x}) = Var(Y_{t+k-1}|\mathbf{X}_{t-1} = \mathbf{x}) = Var\left\{f_{k-1}(\mathbf{X}_t) + \sigma_{k-1}(\mathbf{X}_t)U_{t+1,k-1}|\mathbf{X}_{t-1} = \mathbf{x}\right\}$$

$$= Var\left\{f_{k-1}(\mathbf{X}_t)|\mathbf{X}_{t-1} = \mathbf{x}\right\} + Var\left\{\sigma_{k-1}(\mathbf{X}_t)U_{t+1,k-1}|\mathbf{X}_{t-1} = \mathbf{x}\right\}$$

$$= E\left\{f_{k-1}^2(\mathbf{X}_t)|\mathbf{X}_{t-1} = \mathbf{x}\right\} - \left[E\left\{f_{k-1}(\mathbf{X}_t)|\mathbf{X}_{t-1} = \mathbf{x}\right\}\right]^2 + E\left\{\sigma_{k-1}^2(\mathbf{X}_t)U_{t+1,k-1}^2|\mathbf{X}_{t-1} = \mathbf{x}\right\}$$

$$= E\left[f_{k-1}^2(\mathbf{X}_t)|\mathbf{X}_t = \{f(\mathbf{x}) + \sigma(\mathbf{x})U_t, \mathbf{x}'\}\right] - f_k^2(\mathbf{x}) + E\left[\sigma_{k-1}^2(\mathbf{X}_t)|\mathbf{X}_t = \{f(\mathbf{x}) + \sigma(\mathbf{x})U_t, \mathbf{x}'\}\right]$$

which is the same as in (27). Now the recursive formulae (26) and (27), assumption (A3) plus the fact that the shock variable U_t has finite 4-th moment yield the smoothness results.

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

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

$$\widehat{\sigma}_k^2(\mathbf{x}) = \sigma_k^2(\mathbf{x}) + h^2\sigma_K^2\left[\mathrm{Tr}\,\nabla^2\left\{f_k^2(\mathbf{x}) + \sigma_k^2(\mathbf{x})\right\} - 2f_k(\mathbf{x})\,\mathrm{Tr}\,\nabla^2\left\{f_k(\mathbf{x})\right\}\right]/2$$
$$+\frac{1}{n\mu(\mathbf{x})}\sum_{i=m}^{n-k+1} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma_k^2(\mathbf{X}_{i-1})(U_{i,k}^2 - 1) + o_p(h^2) \quad (29)$$

which entails that

$$\widehat{\sigma}_k(\mathbf{x}) = \sigma_k(\mathbf{x}) + b_{\sigma,k}(\mathbf{x})h^2$$
$$+\frac{1}{2n\mu(\mathbf{x})\sigma_k(\mathbf{x})}\sum_{i=m}^{n-k+1} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma_k^2(\mathbf{X}_{i-1})(U_{i,k}^2 - 1) + o_p(h^2). \quad (30)$$

Now (28) and (30) imply that the estimated GIR function is

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

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

$$= f_{k-1}(\mathbf{x}_u) - f_k(\mathbf{x}) + \left[b_{f,k-1}(\mathbf{x}_u) - b_{f,k}(\mathbf{x})\right]h^2 +$$
$$\frac{1}{n\mu(\mathbf{x}_u)}\sum_{i=m}^{n-k+2} K_h(\mathbf{X}_{i-1} - \mathbf{x}_u)\sigma_{k-1}(\mathbf{X}_{i-1})U_{i,k-1}$$
$$-\frac{1}{n\mu(\mathbf{x})}\sum_{i=m}^{n-k+1} K_h(\mathbf{X}_{i-1} - \mathbf{x})\sigma_k(\mathbf{X}_{i-1})U_{i,k} +$$

$$\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\left\{\widehat{f}(\mathbf{x})-f(\mathbf{x})+\widehat{\sigma}(\mathbf{x})u-\sigma(\mathbf{x})u\right\}+o_p(h^2)$$

hence

$$\widehat{GIR}_k(\mathbf{x},u)=GIR_k(\mathbf{x},u)+b_{GIR,k}(\mathbf{x},u)h^2+T_1+T_2+T_3+T_4+o_p(h^2) \tag{31}$$

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

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

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

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

$$T_4=\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{u}{2n\mu(\mathbf{x})\sigma(\mathbf{x})}\sum_{i=m}^{n}K_h(\mathbf{X}_{i-1}-\mathbf{x})\sigma^2(\mathbf{X}_{i-1})(U_i^2-1) \tag{32}$$

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

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

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

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

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

$$E(T_3T_4)=\frac{u}{2}\left\{\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\right\}^2\|K\|_2^{2m}\frac{\sigma^2(\mathbf{x})}{nh^m\mu(\mathbf{x})}m_3+o\left(n^{-1}h^{-m}\right). \tag{33}$$

Lemma 2

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

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

$$E(T_1T_4)=-\frac{u}{2}\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{\sigma_{11,k-1}(\mathbf{x})I(\mathbf{x}=\mathbf{x}_u)}{nh^m\mu(\mathbf{x})\sigma(\mathbf{x})}\|K\|_2^{2m}+o\left(n^{-1}h^{-m}\right). \tag{34}$$

Proof. We take $i = 3$ as an illustration. By the definitions in (32)

$$E(T_1T_3) = \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu(\mathbf{x})\mu(\mathbf{x}_u)} \times$$

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

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 martingale 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

$$E(T_1T_3) = O(n^{-1}h^{-m+1}) + \frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu(\mathbf{x})\mu(\mathbf{x}_u)} \times$$

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

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 $\mathbf{x} \neq \mathbf{x}_u$, using 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

$$E(T_2T_3)=-\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{\sigma_{1k}(\mathbf{x})}{nh^m\mu(\mathbf{x})}\|K\|_2^{2m}+o\left(n^{-1}h^{-m}\right), \tag{35}$$

$$E(T_2T_4)=-\frac{u}{2}\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{\sigma_{11,k}(\mathbf{x})}{nh^m\mu(\mathbf{x})\sigma(\mathbf{x})}\|K\|_2^{2m}+o\left(n^{-1}h^{-m}\right). \tag{36}$$

Proof. We prove (35) as an illustration. By the definitions in (32)

$$E(T_2T_3)=-\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu^2(\mathbf{x})}\times$$

$$\sum_{i=m}^{n}\sum_{j=m}^{n-k+1}E\left\{K_h(\mathbf{X}_{i-1}-\mathbf{x})K_h(\mathbf{X}_{j-1}-\mathbf{x})\sigma(\mathbf{X}_{i-1})\sigma_k(\mathbf{X}_{j-1})U_iU_{j,k}\right\}$$

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

$$E(T_2T_3)=-\frac{\partial f_{k-1}(\mathbf{x}_u)}{\partial x_1}\frac{1}{n^2\mu^2(\mathbf{x})}\times$$

$$\sum_{i=m}^{n-k+1}E\left\{K_h^2(\mathbf{X}_{i-1}-\mathbf{x})\sigma(\mathbf{X}_{i-1})\sigma_k(\mathbf{X}_{i-1})U_iU_{i,k}\right\}+o\left(n^{-1}h^{-m}\right).$$

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

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

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

which is (35).

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 (33), (34), (35) and (36), 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 means 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 Theorem 3.

Note that by definition

$$\left\{Y_{j+k-1} - \widehat{f}_k(\mathbf{X}_{j-1})\right\}^2 = \left\{Y_{j+k-1} - f_k(\mathbf{X}_{j-1})\right\}^2 + \left\{f_k(\mathbf{X}_{j-1}) - \widehat{f}_k(\mathbf{X}_{j-1})\right\}^2$$
$$+2\left\{Y_{j+k-1} - f_k(\mathbf{X}_{j-1})\right\}\left\{f_k(\mathbf{X}_{j-1}) - \widehat{f}_k(\mathbf{X}_{j-1})\right\} \tag{37}$$

and that Theorem 3.2 of Bosq (1998) entails

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

and so one can drop the second term when smoothing $\mathbf{V}_k$ in the decomposition (37). 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 (22) 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 $m \le i < j \le n, m \le i' < j' \le n, j \ne j$. Thus

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

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

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

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

$$= O(n^{-2}h^{-2m} + n^{-3}k_n h^{1-3m}) = O(n^{-1}h^{-m}) = o(h^4). \tag{38}$$

Meanwhile $\sum_{m\le i<i'<j\le n} E(T_{ij}T_{i'j})$ 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 $\max(i'-i, j-i') \le k_n$. 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

$$\frac{8}{n^4\mu^2(\mathbf{x})}\sum_{m\le i<i'<j\le n} E(T_{ij}T_{i'j}) = o_p(h^4). \tag{39}$$

Combining (38) and (39), we have shown that

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

and thus also the theorem.

Proof of Theorem 2.

Applying the uniform almost sure convergence results of Theorems 2.2 and 3.2 of Bosq (1998), the decomposition of $\widehat{GIR}_k(\mathbf{x}, u)$ in (31) holds uniformly for $\mathbf{x} \in \mathbf{C_x}, u \in C_u$ under assumptions (A1)-(A4). Hence, by (18) and (19)

$$\widehat{GIR}_k(\mathbf{C_x}, C_u) - GIR_k(\mathbf{C_x}, C_u) =$$

$$\frac{1}{n\widehat{P}(\mathbf{C_x}, C_u)} \sum_{i=m}^{n-k+1} \widehat{GIR}_k(\mathbf{X}_{i-1}, U_i) I(\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u) - GIR_k(\mathbf{C_x}, C_u)$$

$$= \frac{1}{n\widehat{P}(\mathbf{C_x}, C_u)} \sum_{i=m}^{n-k+1} GIR_k(\mathbf{X}_{i-1}, U_i) I(\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u) - GIR_k(\mathbf{C_x}, C_u) +$$

$$\frac{1}{n\widehat{P}(\mathbf{C_x}, C_u)} \sum_{i=m}^{n-k+1} b_{GIR,k}(\mathbf{X}_{i-1}, U_i) I(\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u) h^2 +$$

$$\frac{1}{n\widehat{P}(\mathbf{C_x}, C_u)} \sum_{i=m}^{n-k+1} (T_1 + T_2 + T_3 + T_4)(\mathbf{X}_{i-1}, U_i) I(\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u) + o_p(h^2)$$

$$= b_{GIR,k}(\mathbf{C_x}, C_u) h^2 + o_p(h^2) + \sum_{s=1}^{4} \frac{1}{n\widehat{P}(\mathbf{C_x}, C_u)} \sum_{i=m}^{n-k+1} T_s(\mathbf{X}_{i-1}, U_i) I(\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u).$$

To show that all the terms containing T_s are $o_p(h^2)$, we take $s = 1$ as an example.

$$\sum_{i=m}^{n-k+1} T_1(\mathbf{X}_{i-1}, U_i) I(\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u) =$$

$$\sum_{i=m}^{n-k+1} \sum_{j=m}^{n-k+2} \frac{1}{n\mu(\mathbf{X}_{i-1,U_i})} K_h(\mathbf{X}_{j-1} - \mathbf{X}_{i-1,U_i}) \sigma_{k-1}(\mathbf{X}_{j-1}) U_{j,k-1} I(\mathbf{X}_{i-1} \in \mathbf{C_x}, U_i \in C_u)$$

which has second moment less than h^4 by the same technique used in the proof of Theorem 3. We have completed the proof of (20) and hence of the theorem.

9 Acknowledgements

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

References

BOSQ, D. (1998) *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics 110, Springer-Verlag, New York.

CABALLERO, R. and HAMMOUR, M. (1994) The cleansing effect of recessions, *American Economic Review* 84, 1350–1368.

CHEN, R., YANG, L. and HAFNER, C. (2000) 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.

FRANSES, H. F. (1996) *Periodicity and Stochastic Trends in Economic Time Series*, Oxford University Press, Oxford.

HÄRDLE, W. and TSYBAKOV, A. (1997), Local polynomial estimators 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.

HAMERMESH, D.S. and PFANN, G.A. (1996) Adjustment costs in factor demand, *Journal of Economic Literature* 34, 1264–1292.

KOOP, G., PESARAN, M.H. and POTTER, S.M. (1996) Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* 74, 119–147.

LINDBECK, A. and SNOWER, D. (1988) *The Insider-Outsider Theory of Employment and Unemployment*, MIT Press, Cambridge, Mass.

LIPTSER, R. SH. and SHIRJAEV, A. N. (1980) A functional central limit theorem for martingales, *Theory of Probability and its Applications* 25, 667–688.

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

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

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.

TSCHERNIG, R. and YANG, L. (2000) Nonparametric lag selection for time series, *Journal of Time Series Analysis* 21, 457–487.

WOLTERS, J. (1992) Persistence and seasonality in output and employment of the Federal Republic of Germany, *Rechereches Economiques de Louvain* 58, 421–39.

YANG, L. and TSCHERNIG, R. (1998) Non- and semiparametric identification of seasonal nonlinear autoregression models, *Discussion Paper 114*, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

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

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

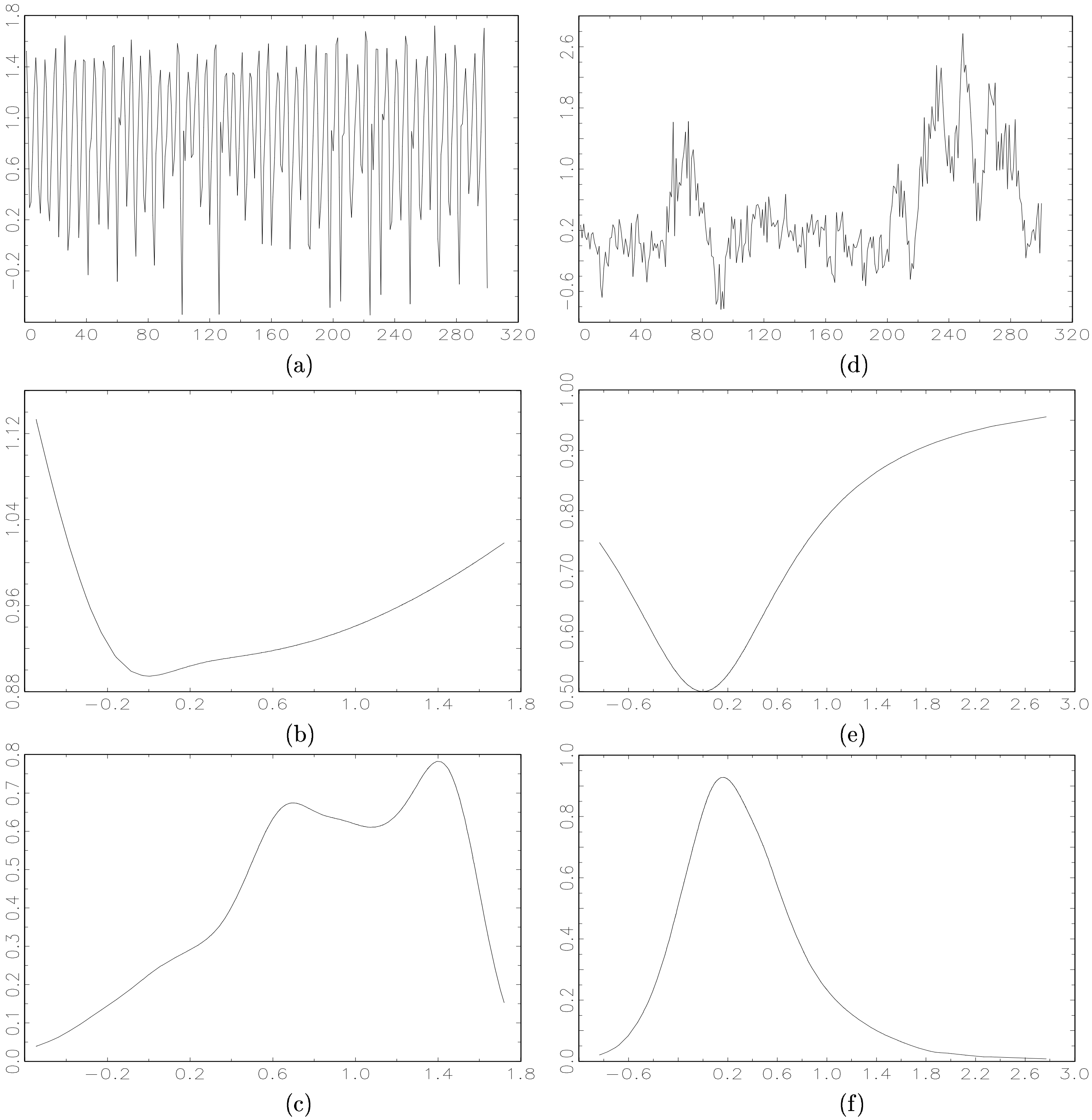

Figure 1: **CHARN1** process: (a) realisation of 300 observations, (b) conditional standard deviation, (c) marginal density; **CHARN2** process: (d) realisation of 300 observations, (e) conditional standard deviation, (f) marginal density;

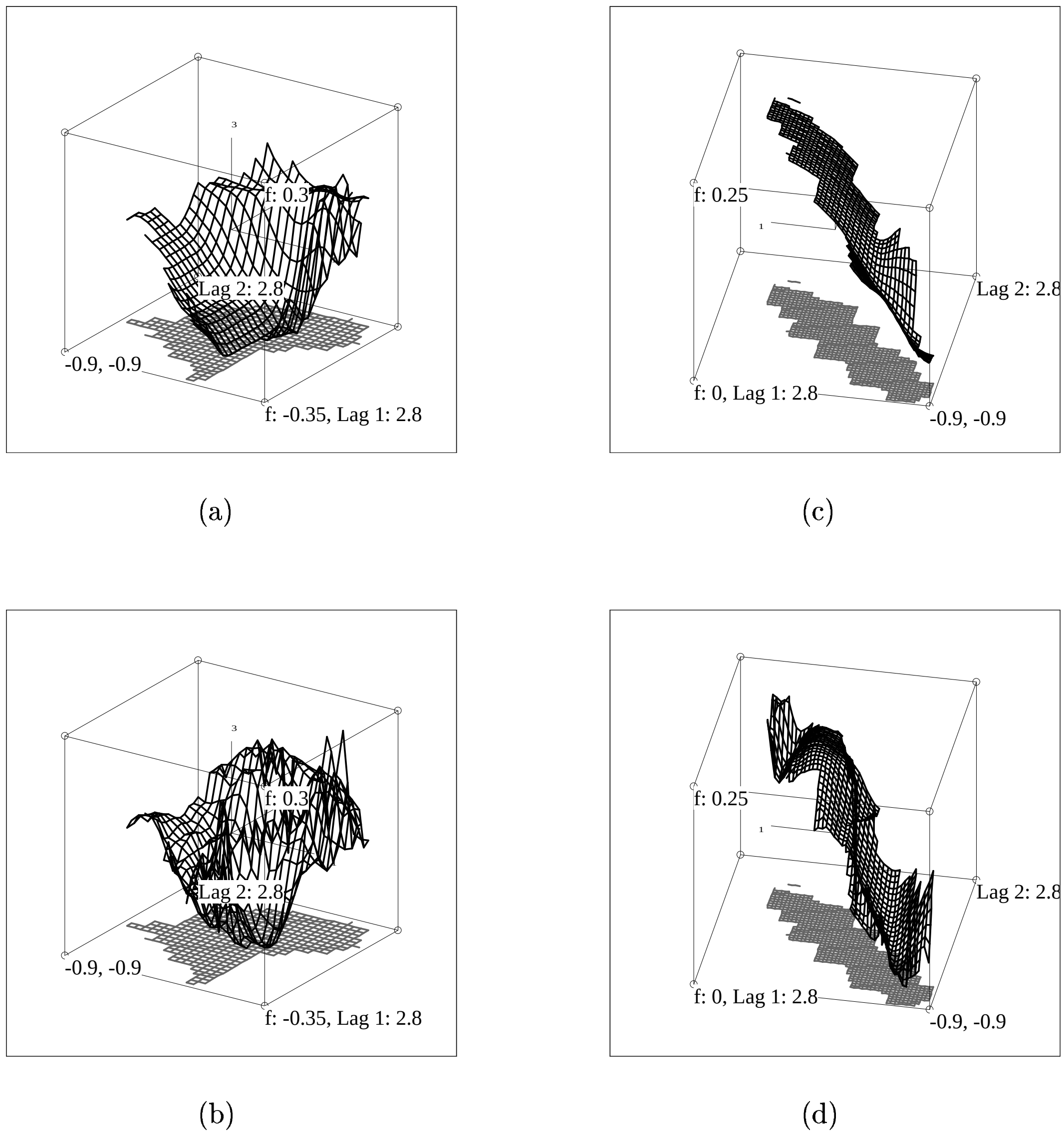

Figure 2: **CHARN1** process: (a) true generalized impulse response functions $GIR_4(\mathbf{x}, 1)$, (b) multi-stage estimates of $GIR_4(\mathbf{x}, 1)$ based on 300 observations; **CHARN2** process: (c) true generalized impulse response functions $GIR_4(\mathbf{x}, 1)$, (d) multi-stage estimates of $GIR_4(\mathbf{x}, 1)$ based on 300 observations

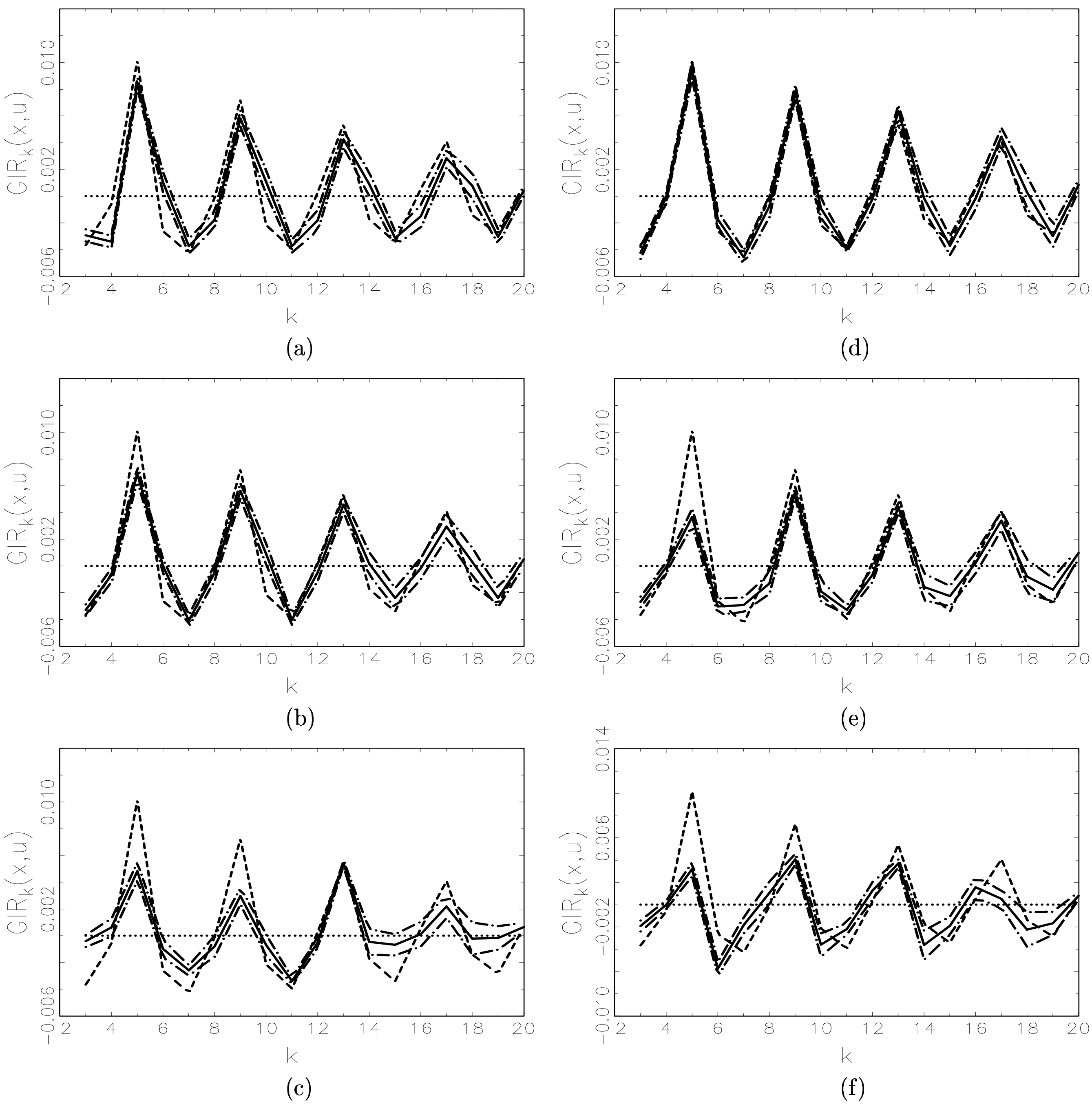

Figure 3: Growth rates of seasonally demeaned West German real GNP: estimated generalized impulse responses for a positive unit shock over $k = 3, \dots, 20$: (a) $\mathbf{x}^1 = (-0.02, -0.01, 0, 0.01)^T$; (b) $\mathbf{x}^2 = (-0.01, 0, 0.01, 0.02)^T$; (c) $\mathbf{x}^3 = (0, 0.01, 0.02, 0.01)^T$; (d) $\mathbf{x}^4 = (0.01, 0, -0.01, -0.02)^T$; (e) $\mathbf{x}^5 = (0.02, 0.01, 0, -0.01)^T$; (f) $\mathbf{x}^6 = (0.03, 0.02, 0.01, 0)^T$ – multi-stage GIR estimates (solid line) and 95% confidence intervals (dots and dashes), GIR estimates based on homoskedastic linear model (dashed line)

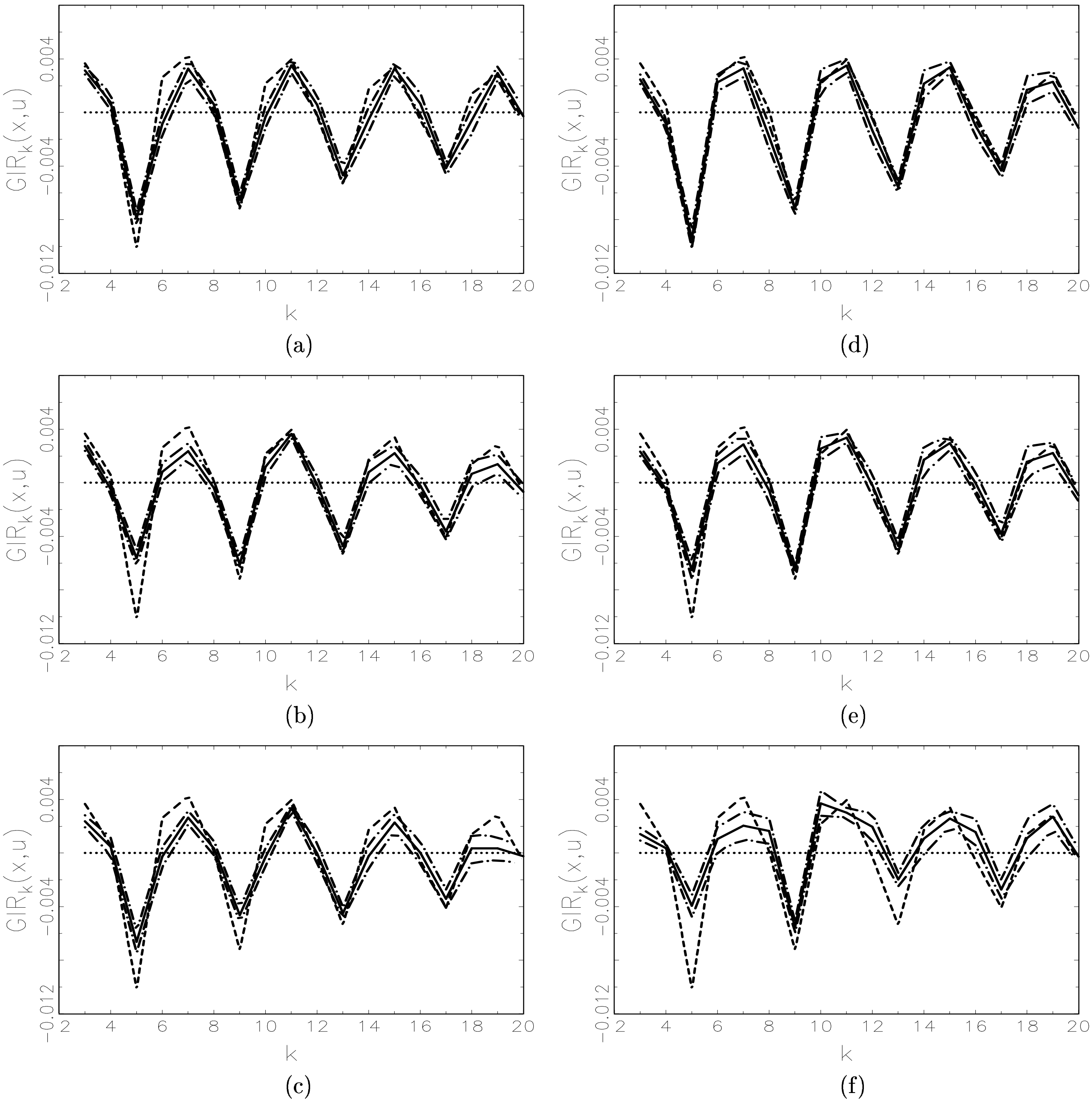

Figure 4: Growth rates of seasonally demeaned West German real GNP: estimated generalized impulse responses for a negative unit shock over $k = 3, \ldots, 20$: (a) $\mathbf{x}^1 = (-0.02, -0.01, 0, 0.01)^T$; (b) $\mathbf{x}^2 = (-0.01, 0, 0.01, 0.02)^T$; (c) $\mathbf{x}^3 = (0, 0.01, 0.02, 0.01)^T$; (d) $\mathbf{x}^4 = (0.01, 0, -0.01, -0.02)^T$; (e) $\mathbf{x}^5 = (0.02, 0.01, 0, -0.01)^T$; (f) $\mathbf{x}^6 = (0.03, 0.02, 0.01, 0)^T$ – multi-stage GIR estimates (solid line) and 95% confidence intervals (dots and dashes), GIR estimates based on homoskedastic linear model (dashed line)