Kernel Estimation of Functional Coefficients in Nonparametric ARX Time Series Models*

Woocheol Kim[†]

Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin

Abstract

This paper suggests a general functional-coefficient regression model in a form of ARX time series model. Contrast to the common threshold variable in the previous works, our model allows each coefficient to possess a different threshold variable and can cover a wide range of nonlinear dynamic processes. The estimation procedure consists of two steps; local linear smoothing and marginal integration. The asymptotic normality of the proposed estimator is derived with the explicit form of bias and variance.

1 Introduction

While the classical ARMA models (Box and Jenkins, 1970) are popular among practitioners due to easy interpretation and simple implmentability, some nonstandard dynamic features of economic data often require a more sophisticated approach than a linear scheme. For example, motivated by a model of noisy trading and arbitrage, the recent studies on exchange rates (Taylor, Peel and Sarno, 2001; Kilian and Taylor, 2001) show that the spot exchange rate adjusts toward economic fundamentals in a nonlinear fashion. Both works used smooth transition models of Granger and Teräsvirta (1993) and Teräsvirta (1994) to capture the nonlinear mean reversion in the deviation of nominal exchange rate from equilibrium. As another example, Lütkepohl, Teräsvirta and Wolters (1998) augmented a standard error correction model with a smooth transition function to investigate the stability and linearity of the German money demand. Once one goes beyond linear domain, there are infinitely many nonlinear form of functions. Thus, unless supported by economic theory, a given parametric form is prone to misspecification. Without *a priori* knowledge on model structure, a data-driven method would be desirable to explore a proper functional specification. The general flexibility of nonparametric method, however, is delimited by a severe problem of slower convergence rate which is a cost of high-dimensional smoothing; see Silverman (1986). To avoid the curse of dimensionality, the variety of restrictive structures have been incorporated with general nonparametric models. Additive models of Breiman and Friedman (1985) is one well-known example.

*The author acknowledges support by the Deutsche Forschungsgemeinschaft via SFB 373 at Humboldt University, Berlin.

[†]Institute for Statistics and Econometrics, Humboldt University of Berlin, Spandauer Str. 1, 10178, Berlin, Germany. E-mail address: `woocheol@wiwi.hu-berlin.de`

This paper follows an alternative line, assuming varying-coefficient models where the coefficients are a unknown function of random variables. In the previous works, the varying-coefficient models mostly focus on i.i.d. case of cross-section data. We consider a general class of autoregressive time series models with exogenous variables (ARX) which are of interest for economic data. More importantly, multiple threshold variables are allowed; that is, each coefficient can be a function of a different threshold. Let $\{y_t, X_t, W_t\}_{i=-\infty}^{\infty}$ be jointly strictly stationary with X_t taking values in $\mathbb{R}^q$and W_t in $\mathbb{R}^{p+q}$. The specific model in the paper takes the form of

$$y_t = \sum_{j=1}^{q} \beta_j(W_{jt}) y_{t-j} + \sum_{k=q+1}^{p+q} \beta_k(W_{kt}) X_{kt} + \varepsilon_t, \tag{1.1}$$

where ε_t is iid $(0, \sigma_\varepsilon^2)$, $\beta_j(\cdot)$ is a (unknown) smooth function, and $\{X_k\}$ are exogenous variables. By letting $X_1 \equiv 1$, a varying intercept term can be included in (1.1). Since, unlike the parametric linear models, the coefficients β_j's now are allowed to depend on W_j, our model can cover a wide range of nonlinear time series processes, significantly reducing modeling bias. At the same time, the curse of dimensionality can be avoided by estimating only one-dimensional functions. The main theorem in section 3 shows that each unknown functional coefficient can be consistently estimated at the one-dimensional optimal rate. Also, the results from estimating (1.1) are now easily interpreted, providing how the regression coefficients change over different economic states denoted by W. When W_j is some lagged variable of y_t and $X_{kt} = 0$, for all $k = q{+}1, .., p{+}q$, the smooth transition models used in the aforementioned empirical works fall within our model with $\beta_j(\cdot)$ being specified as an exponential function. In the early time series literature, (1.1) was studied by several authors, at least, in a parametric framework. The threshold autoregressive model (TAR) of Tong (1990) assumed $\beta_j(\cdot)$ to be a step function of W, while the exponential autoregressive model (EXPAR) of Haggan and Ozaki (1981) used a more complicated form of nonlinear continuous functions for AR coefficients. In a nonparametric context, the study of varying-coefficient models was initiated by the seminal works of Hastie and Tibshirani (1993) and Chen and Tsay (1993) where the unknown functions were estimated by the data-driven methods of smoothing spline or running window procedure. More close to our approach is the recent work of Fan and Zhang (1999) and Cai, Fan and Yao (2000) which suggested a kernel smoothing method to estimate the coefficient functions. For i.i.d. samples, Fan and Zhang (1999) developed a new two-step method of local linear regression and Cai, Fan and Yao (2000) extended the similar technique to the case of autoregressive time series models. However, the nonparametric models assumed in those works are more restrictive than (1.1) in the sense that the varying coefficients are determined by only a single set of variables. In real applications, it often arises that one needs to weaken the somewhat strong assumption of 'common' threshold variables, in order to allow for more rich dynamic relationship between economic variables. For example, when considering the relation between the growth rates of nominal income and money stock, it is possible that the effects of the lagged income and the money stock on the current growth rate of income are governed by different variables, say, another lagged income and interest rates, respectively. The varying-coefficient model in (1.1) can easily handle such case and provide an interesting way of policy evaluation. Although allowing for multiple threshold variables can increase the flexibility of dynamic modeling, it necessarily complicates the estimation procedure. This paper proposes a two-step kernel smoothing method to estimate $\beta_j(\cdot)$'s in (1.1). The first step is to obtain consistent estimates for $\beta_j(\cdot)$, via the local linear smoothing as in Fan and Zhang (1999) or Cai, Fan and Yao (2000). In the second step, the preliminary estimates are marginally integrated to achieve the optimal convergence rate. Note that all the asymptotic results developed in the paper can be applied to cross-section case, since i.i.d. random samples are a special case of stationary processes.

The rest of the paper is organized as follows. Section 2 defines the two-step kernel estimator for (1.1). In section 3, we derive the main results including the asymptotic normality of our estimators. Section 4 concludes and the proofs are contained in Section 5.

2 Estimation

Local linear fitting has been widely used in nonparametric regression due to its simple procedure and nice properties, since the basic idea was originated by Cleveland, Grosse and Shyu (1991). Some of its advantages, among others, include the high asymptotic minimax efficiency (Fan, 1993) and automatic correction of edge effects (Fan and Gijbels, 1992). Fan and Zhang (1999) and Cai, Fan and Yao (2000) showed the asymptotic properties of local linear smoothing in varying-coefficient models based on i.i.d. and time series data, respectively. To estimate $\beta_j(\cdot)$, we define the two-step kernel method based on local linear regression. Assuming that the second order derivative of $\beta_j(W_{jt})$ exists and is continuous at W_{jt}, we can approximate $\beta_j(W_{jt})$ locally by a linear function at point w_j :

$$\beta_j(W_{jt}) \approx \beta_{j0} + (W_{jt} - w_j)\beta_{j1},$$

where $\beta_{j0} \equiv \beta_j(w_j)$ and $\beta_{j1} \equiv \beta_j'(w_j)$. Let $K(\cdot)$ and $L(\cdot)$ be a kernel function on $\mathbb{R}^1$ and $\mathbb{R}^{d-1}$, respectively, with $d = p + q$ and h, g be a bandwidth parameter. In what follows, we adopt the notation $W_t = (W_{jt}, W_{-jt})$ to highlight a direction of interest W_{jt} for all $1 \le j \le d$, while W_{-jt} is the $(d-1)$ dimensional vector that consists of all the rest W_{kt}'s $1 \le k \le d$, $k \ne j$. Also, for a compact expression, we define $X_t \equiv (X_{1t}, .., X_{pt})^T$; $Y_t \equiv (y_{t-1}, .., y_{t-q})^T$; $Z_t \equiv (X_t^T, Y_t^T)^T$. Given the observation $\{y_t, X_t, W_t\}_{t=1}^n$, we get the pilot estimator of $\beta_j(w_j)$ by considering the following multivariate weighted least squares,

$$\min_{\{b_{j0}, b_{j1}\}} \sum_{t=p+1}^{n+p} K_h\left(W_{jt} - w_j\right) L_g\left(W_{-jt} - w_{-j}\right) \{y_t - \sum_{j=1}^{q} [b_{j0} + b_{j1}\left(W_{jt} - w_j\right)] y_{t-j}$$
$$- \sum_{k=1+q}^{p+q} [b_{k0} + b_{k1}\left(W_{kt} - w_k\right)] X_{kt}\}^2, \tag{2.2}$$

where $K_h(w_j) = K(w_j/h)/h$ and $L_g(w_{-j}) = L(w_{-j}/h)/h^{d-1}$. Minimizing (2.2) with respect to b_{j0}'s and b_{j1}'s, one obtains the preliminary estimates of β_{j0}'s and β_{j1}'s, respectively;

$$\begin{aligned} \widehat{\beta}_0(w) &= (\widehat{\beta}_{10}(w_1), \widehat{\beta}_{20}(w_2), .., \widehat{\beta}_{(p+q)0}(w_{p+q}))^T, \\ \widehat{\beta}_1(w) &= (\widehat{\beta}_{11}(w_1), \widehat{\beta}_{21}(w_2), .., \widehat{\beta}_{(p+q)1}(w_{p+q}))^T. \end{aligned}$$

In a more compact form, the solution to minimizing (2.2) can be written as

$$\widehat{\boldsymbol{\beta}}(w) = \begin{pmatrix} \widehat{\beta}_0(w) \\ \widehat{\beta}_1(w) \end{pmatrix} = (\mathbf{Z}^T \Omega \mathbf{Z})^{-1} \mathbf{Z}^T \Omega \mathbf{y}, \tag{2.3}$$

where

$$
\begin{aligned}
Z^T &= (Z_{p+1}, .., Z_{n+p}),\ \widetilde{W}_t = W_t - w,\ \ \widetilde{W}^T = (\widetilde{W}_{p+1}, .., \widetilde{W}_{n+p}) \\
Z^W &= Z \odot \widetilde{W},\ \ \mathbf{Z} = (Z, Z_W),\ \ \mathbf{y} = (y_{p+1}, .., y_{n+p})^T,
\end{aligned}
$$

and

$$\Omega = diag\{K_h\left(W_{j(p+1)} - w_j\right) L_g\left(W_{-j(p+1)} - w_{-j}\right), .., K_h\left(W_{j(n+1)} - w_j\right) L_g\left(W_{-j(n+p)} - w_{-j}\right)\}.$$

Now, focusing only on the level estimates, we define $\widehat{\beta}_j(w) \equiv \widehat{\beta}_{j0}(w)$.

Note that $\widehat{\beta}_j(w)$ depends on w, although the true $\beta_j(\cdot)$ is a function of only w_j. This is because in the kernel regression in (2.2) we simply smooth the data around at the point of w. That is, our pilot estimates are obtained by high-dimensional smoothing and their convergence rate is not optimal. To attain the optimal rate, we, in the second step, marginally integrate the pilot estimates $\widehat{\beta}(w_j, W_{-jt})$ over W_{-jt};

$$\widehat{\beta}_j^*(w_j) = \frac{1}{n} \sum_{t=p+1}^{n+p} \widehat{\beta}_j(w_j, W_{-jt}). \tag{2.4}$$

A similar idea has been used extensively in the study of additive models, called *marginal integration*; see Linton and Nielsen (1995).

3 Main Results

Before deriving the asymptotic properties of the two-step estimators, we introduce the necessary technical conditions.

3.1 Conditions

Let $\alpha(k)$ be the strong mixing coefficient of the joint process $\{y_t, X_t, W_t\}_{t=-\infty}^{\infty}$, defined by

$$\alpha(k) \equiv \sup_{A \in \mathcal{F}_{-\infty}^0,\ B \in \mathcal{F}_k^{\infty}} |P(A \cap B) - P(A) P(B)|,$$

where $\mathcal{F}_b^a$ be the σ-algebra of events generated by $\{y_k, X_k, W_k\}_{k=a}^b$. If $\alpha(k) \to 0$ as $k \to \infty$, the process $\{y_t, X_t, W_t\}_{t=-\infty}^{\infty}$ is strongly mixing.

A1. $\{y_t, X_t, W_t\}_{t=-\infty}^{\infty}$ is stationary and strongly mixing, and ε_t is i.i.d. with $E(\varepsilon_t | Z_t, W_t) =$ 0 and
$E\left(\varepsilon_t^2 | Z_t = z, W_t = w\right) = \sigma^2(z, w) < \infty$.

A2. The functions $\beta_j(\cdot)$'s have bounded Lipschitz continuous second order derivatives for all $1 \le j \le d$.

A3. The stationary distribution function $F(\cdot)$ has a density $p(\cdot)$. The function $p(\cdot)$,together with the densities $p_j(\cdot)$ of $F_j(\cdot)$ and $p_{-j}(\cdot)$ of $F_{-j}(\cdot)$ are all uniformly bounded away from zero and infinity and have bounded Lipschitz continuous second order derivatives, for all $1 \le j \le d$.

A4. The matrix $E\left(W^T W|Z\right)$ is of full rank, and $E(W^T W|Z_s)^{-1}$ is bounded element-wise.

A5. Both kernels $K(\cdot)$ and $L(\cdot)$ are symmetric, bounded, compactly supported, and Lipschitz continuous with $\int K\left(w_j\right)dw_j = \int L\left(w_{-j}\right)dw_{-j} = 1$.

A6. $g, h \downarrow 0$ and $nhg^{d-1} \uparrow \infty$.

A7. $g^{d-1}/h^2 \to \infty$, $nhg^{2(d-1)}/\ln^2 n \to \infty$, $g^q/h^2 \to 0$, and $h = h_0 n^{\frac{-1}{5}}$.

Most of the assumptions above are standard in kernel regression. The additional bandwidth condition in A7 is assumed to apply the approximation argument of Lemma 4.2 and 4.3 in Yang, Härdle, and Nielsen (1999).

3.2 Asymptotic Properties

To facilitate the asymptotic derivation, we begin with rewriting the estimates in terms of second moments. For notational convenience, define

$$\begin{aligned} \widetilde{W}_t^h &= \left(\frac{W_{jt}-w_j}{h}, \frac{W_{-jt}-w_{-j}}{g}\right), \\ Z_{Wt}^h &= Z_t \odot \widetilde{W}_t^h, \end{aligned}$$

and

$$Q = diag\{\iota_{p+q}^T, g, .., g, h, g.., g\}.$$

Then, the pilot estimates in (2.3) can be expressed as

$$\widehat{\boldsymbol{\beta}}(w) = \begin{pmatrix} \widehat{\beta}_0(w) \\ \widehat{\beta}_1(w) \end{pmatrix} = Q^{-1} S_n^{-1}(w) t_n(w), \tag{3.5}$$

where

$$S_n = S_n(w) = \begin{bmatrix} S_{0n}(w) & S_{1n}^T(w) \\ S_{1n}(w) & S_{2n}(w) \end{bmatrix}, \quad t_n(w) = \begin{bmatrix} t_{0n}(w) \\ t_{1n}(w) \end{bmatrix}$$

with

$$\begin{bmatrix} S_{0n}(w) \\ S_{1n}(w) \\ S_{2n}(w) \\ t_{0n}(w) \\ t_{1n}(w) \end{bmatrix} = \frac{1}{n} \sum_{t=p+1}^{n+p} K_h\left(W_{jt}-w_j\right) L_g\left(W_{-jt}-w_{-j}\right) \begin{bmatrix} Z_t Z_t^T \\ \widetilde{Z}_{Wt} Z_t^T \\ \widetilde{Z}_{Wt} \widetilde{Z}_{Wt}^T \\ Z_t y_t \\ \widetilde{Z}_{Wt} y_t \end{bmatrix},$$

Since we are only interested in the level estimates, we can separate them out by premultiplying $\widehat{\boldsymbol{\beta}}(w)$ with $e_j^T = [0, .., 0, 1, 0, .., 0]$. From $e_j^T Q^{-1} = e_j^T$, the local linear estimator of $\beta_j(w_j)$ is

$$\widehat{\beta}_j(w) \equiv \widehat{\beta}_{j0}(w) = e_j^T \widehat{\boldsymbol{\beta}}(w) = e_j^T S_n^{-1}(w) t_n(w). \tag{3.6}$$

Our first lemma concerns the decomposition of the estimation errors into two parts of bias and stochastic terms. For the following lemma, we only need the differentiability of coefficient functions. See section 5 for proof.

Lemma 1 Under A.2,

$$\text{i) } \widehat{\beta}_j(w_j, w_{-j}) - \beta_j(w_j) = \widetilde{\tau}_n(w) + \frac{h^2}{2} B_n(w) + o_p(h^2), \tag{3.7}$$

where $\widetilde{\tau}_n(w) = e_j^T S_n^{-1}(w)\tau_n(w)$ and $B_n(w) = e_j^T S_n^{-1}(w) B_n'$ with

$$\begin{aligned} \tau_n(w) &= \frac{1}{n} \sum_{t=p+1}^{n+p} K_h(W_{jt} - w_j) L_g(W_{-jt} - w_{-j}) (Z_t^T, Z_{Wt}^{hT})^T \varepsilon_t, \\ B_n'(w) &= \frac{1}{n} \sum_{t=p+1}^{n+p} K_h(W_{jt} - w_j) L_g(W_{-jt} - w_{-j}) (Z_t^T Z_{jt}, Z_{Wt}^{hT} Z_{jt})^T (\frac{W_{jt} - w_j}{h})^2 \beta_j''(w_j). \end{aligned}$$

$$\text{ii) } \widehat{\beta}_j^*(w_j) - \beta_j(w_j) = \widetilde{\tau}_n^*(w_j) + \frac{h^2}{2} B_n^*(w_j) + o_p(h^2),$$

where

$$\begin{aligned} \widetilde{\tau}_n^*(w_j) &= \frac{1}{n} \sum_{l=p+1}^{n+p} e_j^T S_n^{-1}(w_j, W_{-jl}) \tau_n(w_j, W_{-jl}), \\ B_n^*(w_j) &= \frac{1}{n} \sum_{l=p+1}^{n+p} e_j^T S_n^{-1}(w_j, W_{-jl}) B_n'(w_j, W_{-jl}). \end{aligned}$$

Now, applying the mean square convergence for the bias term, $B_n^*(w_j)$ and the asymptotic normality of the stochastic term, $\widetilde{\tau}_n^*(w_j)$, in Lemma 5.1 and 5.3 of section 5, we obtain the main theorem. Henceforth, let $\Gamma(w_j, W_{-jl}) \equiv E_{|W=(w_j, W_{-jl})}(Z_1 Z_1^T)$, with $\Gamma_j(w_j, W_{-jl})$ being the j-th row of $\Gamma(w_j, W_{-jl})$, and $\sigma_\varepsilon^2(w, z) \equiv E(\varepsilon^2 | W = w, Z = z)$.

Theorem 2. Assume that A.1 through A.2 hold. Then,

$$\sqrt{nh}[\widehat{\beta}_j^*(w_j) - \beta_j(w_j) - Bias] \xrightarrow{d} N(0, ||K||^2 \gamma_j(w_j)),$$

where

$$\begin{aligned} Bias &= \frac{h^2}{2} \beta_j''(w_j) \int K(u) u^2 du, \\ \gamma_j(w_j) &= \int \Gamma_j^{-1}(w_j, s) E_{|W=(w_j,s)}[\sigma_\varepsilon^2(w_j, s, Z_1) Z_1 Z_1^T] \Gamma_j^{-1}(w_j, s)^T \frac{p_{W_{-j}}^2(s)}{p_W(w_j, s)} ds. \end{aligned}$$

Theorem 2 shows that the two-step estimator achieves the one-dimensional optimal convergence rate, and, as in Fan (1992), its asymptotic bias is of a simple form, consisting of only the second order derivative. If we further assume that $E(\varepsilon^2 | W, Z) = \sigma_\varepsilon^2(W)$, the asymptotic variance reduces to $||K||_2^2 \int \Gamma_j^{-1}(w_j, s) \sigma_\varepsilon^2(w_j, s) \frac{p_{W_{-j}}^2(s)}{p_W(w_j, s)} ds$. In this case, Theorem 2 coincides with the results on the marginal integration estimator for additive models, see Linton and Nielsen (1995).

4 Conclusion

Varying-coefficient models are a useful tool which allows the flexibility of nonparametric method, while still preserving the estimability and interpretability of the model. In this paper, the previous results on varying-coefficient models are generalized in two directions. Instead of commonality condition on the threshold variables, the regression coefficients in our model possibly possess a different threshold variable. Also, in order to capture the variety of nonlinear dynamic relations, we assume a very general form of data generating process based on a stationary autoregressive model with exogenous variable. To estimate the unknown coefficient functions, a new two-step procedure is proposed by combining the local linear fit and marginal integration method. The new procedure is shown to attain the optimal convergence rate. Although the model in the paper confines the interest to estimating conditional mean relation, we may extend the idea of functional-coefficient models into the study of conditional volatility. For example, some empirical findings such as Schwert(1989) indicate that stock market volatility rises sharply during recessions and drops during expansions. To study such an issue, one can think of replacing the ARX setting with a functional-coefficient ARCH model.

5 Proof

Proof of Lemma 1. We first separate out the leading stochastic term of the estimation errors;

$$\widehat{\beta}_j(w_j, w_{-j}) - \beta_j(w_j) = e_j^T S_n^{-1}(w)\tau_n(w) + \left[e_j^T S_n^{-1}(w)t_n^*(w) - \beta_j(w_j)\right], \qquad (5.8)$$

where $t_n^*(w) = [t_{0n}^T(w), t_{1n}^T(w)]^T$ with

$$\begin{aligned} t_{0n}^*(w) &= \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt}-w_j\right) L_g\left(W_{-jt}-w_{-j}\right) Z_t[\sum_{j=1}^{q}\beta_j(W_{jt})y_{t-j} + \sum_{k=q+1}^{p+q}\beta_k(W_{kt})X_{kt}], \\ t_{1n}^*(w) &= \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt}-w_j\right) L_g\left(W_{-jt}-w_{-j}\right) Z_{Wt}^h[\sum_{j=1}^{q}\beta_j(W_{jt})y_{t-j} + \sum_{k=q+1}^{p+q}\beta_k(W_{kt})X_{kt}]. \end{aligned}$$

Observe that

$$\begin{aligned} \beta_j(w_j) &= e_j^T Q^{-1}(Q^{-1}\mathbf{Z}^T\Omega\mathbf{Z}Q^{-1})^{-1}Q^{-1}\mathbf{Z}^T\Omega\mathbf{Z}\overline{\boldsymbol{\beta}} \\ &= e_j^T S_n^{-1}(w)Q^{-1}\mathbf{Z}^T\Omega[Z\overline{\boldsymbol{\beta}}_1 + hZ_W^h\overline{\boldsymbol{\beta}}_2], \end{aligned}$$

using the identity $\beta_j(w_j) = e_j(\mathbf{Z}^T\Omega\mathbf{Z})^{-1}\mathbf{Z}^T\Omega\mathbf{Z}\overline{\boldsymbol{\beta}}$, where $\overline{\boldsymbol{\beta}} = (\overline{\boldsymbol{\beta}}_1^T, \overline{\boldsymbol{\beta}}_2^T)^T$, $\overline{\boldsymbol{\beta}}_1 = [\beta_j\left(w_j\right), \beta_{-j}\left(W_{-jt}\right)]^T$, $\overline{\boldsymbol{\beta}}_2 = [0, ., \beta_j'(w_j), ., 0]^T$. Under the twice differentiability of $\beta_j(\cdot)$ in A.2, we use Taylor expansions to get

$$\begin{aligned} \beta_j(W_{jt}) &= \beta_j(w_j) + h(W_{jt}-w_j)/h\beta_j'(w_j) \\ &\quad + (h^2/2)((W_{jt}-w_j)/h)^2\beta_j''(w_j) + o_p(h^2). \end{aligned}$$

If we plug in this in the expression for $t_n^*(w)$,

$$t_n^*(w) = Q^{-1}\mathbf{Z}^T\Omega[Z\overline{\boldsymbol{\beta}}_1 + hZ_W^h\overline{\boldsymbol{\beta}}_2 + (h^2/2)((W_{jt}-w_j)/h)^2\beta_j''(w_j)Ze_j + o_p(h^2)].$$

This implies that the second term in (5.8) is

$$
\begin{aligned}
& e_j^T S_n^{-1}(w) t_n^*(w) - \beta_j(w_j) \\
= \; & e_j^T S_n^{-1}(w)\{t_n^*(w) - Q^{-1}\mathbf{Z}^T \Omega [Z\overline{\boldsymbol{\beta}}_1 + hZ_W^h \overline{\boldsymbol{\beta}}_2]\} \\
= \; & \frac{h^2}{2} e_j^T S_n^{-1}(w) Q^{-1}\mathbf{Z}^T \Omega Z e_j [(\frac{W_{jt} - w_j}{h})^2 \beta_j''(w_j) + o_p(h^2)] \\
= \; & \frac{h^2}{2} e_j^T S_n^{-1}(w) B_n'(w) + o_p(h^2),
\end{aligned}
$$

where

$$
B_n'(w) = \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt} - w_j\right) L_g\left(W_{-jt} - w_{-j}\right)(Z_t^T Z_{jt}, Z_{Wt}^{hT} Z_{jt})^T (\frac{W_{jt} - w_j}{h})^2 \beta_j''(w_j).
$$

The decomposition result for a two-step estimator now follows directly by (3.6). ■

Lemma 5.1.

(i) Under A.1 through A.5,

$$
S_n(w) \xrightarrow{p} S(w),
$$

where $S(w) \equiv p_W(w) \begin{bmatrix} E_{|W=w}(Z_t Z_t^T) & O_{(p+q)\times(p+q)} \\ O_{(p+q)\times(p+q)} & [E_{|W=w}(Z_t Z_t^T) \odot diag_{(p+q)}\{\mu_{KL}^2\}] \end{bmatrix}$.

(ii) Under A.1 through A.6,

$$
B_n^*(w_j) \xrightarrow{p} \beta_j''(w_j)\mu_K^2.
$$

Proof . By the standard law of large numbers on stationary time series, we get

$$
S_{0n}(w) \xrightarrow{p} E(K_h\left(W_{jt} - w_j\right) L_g\left(W_{-jt} - w_{-j}\right) Z_t Z_t^T),
$$

and, applying the integration with substitution of variable and Taylor expansion, we can show the expectation term is equal to

$$
\begin{aligned}
& \int K_h\left(u_j - w_j\right) L_g\left(u_{-j} - w_{-j}\right) vv^T p_{W,Z}(u,v) du dv \\
= \; & \int K\left(s_j\right) L_g\left(s_{-j}\right) vv^T p_{W,Z}(w + hs, v) ds dv \\
\simeq \; & p_W(w) \int vv^T p_{Z|W}(w,v) ds dv = p_W(w) E_{|W=w}(Z_t Z_t^T).
\end{aligned}
$$

Using the same argument, it follows that

$$
\begin{aligned}
& S_{1n}(w) \xrightarrow{p} E(K_h\left(W_{jt} - w_j\right) L_g\left(W_{-jt} - w_{-j}\right) \widetilde{Z}_{Wt} Z_t^T) \\
= \; & \int K_h\left(u_j - w_j\right) L_g\left(u_{-j} - w_{-j}\right) [v \odot (\frac{W_{jt} - w_j}{h}, \frac{W_{-jt} - w_{-j}}{g})] v^T p_{W,Z}(u,v) du dv \\
\simeq \; & \int K\left(s_j\right) L\left(s_{-j}\right) [v \odot s] v^T p_{W,Z}(w,v) ds dv \\
= \; & \int [v \odot \int K\left(s_j\right) L\left(s_{-j}\right) s ds] v^T p_{W,Z}(w,v) dv = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
& S_{2n}(w) \xrightarrow{p} E(K_h\left(W_{jt}-w_j\right) L_g\left(W_{-jt}-w_{-j}\right) \widetilde{Z}_{Wt}\widetilde{Z}_{Wt}^T) \\
= & \int K_h\left(u_j-w_j\right) L_g\left(u_{-j}-w_{-j}\right) [v \odot (\frac{W_{jt}-w_j}{h}, \frac{W_{-jt}-w_{-j}}{g})] \times \\
& [v \odot (\frac{W_{jt}-w_j}{h}, \frac{W_{-jt}-w_{-j}}{g})]^T p_{W,Z}(u,v) du dv \\
\simeq & \int K\left(s_j\right) L\left(s_{-j}\right) [v \odot s][v \odot s]^T p_{W,Z}(w,v) ds dv \\
= & \; p_W(w) \int (vv^T \odot \int K\left(s_j\right) L\left(s_{-j}\right) ss^T ds) p_{Z|W}(w,v) dv \\
= & \; p_W(w)[E_{|W=w}(Z_t Z_t^T) \odot diag_{(p+q)}\{\mu_{KL}^2\}],
\end{aligned}
$$

where μ_{KL}^2 is a $(p+q)\times 1$ column vector whose j-th element is μ_K^2 with others being μ_L^2 and $\mu_K^2 = \int K(u)u^2 du$.

(ii) By the strong convergence theorems for stationary time series (with strong mixing condition) in Masory (1996), it holds that, uniformly in (w_j, W_{-jl}),

$$cov(D_l, D_k) = \rho^{|l-k|}\{O_p(h + \ln n/\sqrt{nhg^{d-1}})\}^2,$$

where

$$D_l = S_n^{-1}(w_j, W_{-jl}) - S^{-1}(w_j, W_{-jl}).$$

Let

$$I_{tl,1} = K_h\left(W_{jt}-w_j\right) L_g\left(W_{-jt}-W_{-jl}\right) (Z_t^T Z_{jt}, Z_{Wt}^{hT} Z_{jt})^T (\frac{W_{jt}-w_j}{h})^2 \beta_j''(w_j).$$

Then, it follows from Lemma 4.2 in Yang at al (1999) that

$$E(I_{tl,1} I_{sk,1}) = \rho^{\min(|t-s|,|l-k|)} O(1/hg^{d-1}),$$

uniformly, for s, t, l, and k. Thus, under the bandwidth conditions in A.6 and A.7,

$$
\begin{aligned}
& \frac{1}{n}\sum_{l=p+1}^{n+p} e_j^T D_n \frac{1}{n}\sum_{l=p+1}^{n+p} I_{tl,1} \\
= & \frac{1}{\sqrt{nhg^{d-1}}} O_p(h + \ln n/\sqrt{nhg^{d-1}}) \\
= & O_p[\frac{h}{\sqrt{nhg^{d-1}}} + \frac{\ln n}{nhg^{d-1}}] = O_p[\frac{1}{\sqrt{nh}}(\sqrt{\frac{h^2}{g^{d-1}}} + \frac{\ln n}{\sqrt{nhg^{2(d-1)}}})] \\
= & o_p(\frac{1}{\sqrt{nh}}),
\end{aligned}
$$

which implies

$$B_n^*(w_j) = \frac{1}{n}\sum_{l=p+1}^{n+p} e_j^T S^{-1}(w_j, W_{-jl}) B_n'(w_j, W_{-jl}) + o_p(\frac{1}{\sqrt{nh}}).$$

Now, it is straightforward to calculate the probability limit of bias term.

$$
\begin{aligned}
B_n^*(w_j) &\simeq \frac{1}{n}\sum_{l=p+1}^{n+p} p_W^{-1}(w_j, W_{-jl})[\Gamma_j^{-1}(w_j, W_{-jl})\ O_{1\times(p+q)}]B_n'(w_j, W_{-jl}) \\
&= \frac{1}{n}\sum_{l=p+1}^{n+p} p_W^{-1}(w_j, W_{-jl})\Gamma_j^{-1}(w_j, W_{-jl}) \times \\
&\quad \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt}-w_j\right) L_g\left(W_{-jt}-W_{-jl}\right) Z_t Z_{jt}\left(\frac{W_{jt}-w_j}{h}\right)^2 \beta_j''(w_j) \\
&= \beta_j''(w_j)\frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt}-w_j\right)\left[\frac{1}{n}\sum_{l=p+1}^{n+p} p_W^{-1}(w_j, W_{-jl})\Gamma_j^{-1}(w_j, W_{-jl})L_g\left(W_{-jt}-W_{-jl}\right)\right] \\
&\quad \times Z_t Z_{jt}\left(\frac{W_{jt}-w_j}{h}\right)^2 \\
&\simeq \beta_j''(w_j)\frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt}-w_j\right) p_W^{-1}(w_j, W_{-jt})p_{W_{-j}}(W_{-jt})\Gamma_j^{-1}(w_j, W_{-jt})Z_t Z_{jt}\left(\frac{W_{jt}-w_j}{h}\right)^2 \\
&\xrightarrow{p} \beta_j''(w_j)E\left\{K_h\left(W_{jt}-w_j\right) p_W^{-1}(w_j, W_{-jt})p_{W_{-j}}(W_{-jt})\Gamma_j^{-1}(w_j, W_{-jt})Z_t Z_{jt}\left(\frac{W_{jt}-w_j}{h}\right)^2\right\} \\
&\simeq \beta_j''(w_j)\mu_K^2\int \Gamma_j^{-1}(w_j, W_{-jt})\left\{\int Z_t Z_{jt}\frac{p_{W,Z}(w_j, W_{-jt}, Z_t)}{p_W(w_j, W_{-jt})}dZ_t\right\}p_{W_{-j}}(W_{-jt})dW_{-jt} \\
&= \beta_j''(w_j)\mu_K^2\int \Gamma_j^{-1}(w_j, W_{-jt})[\Gamma_j(w_j, W_{-jt})]^T p_{W_{-j}}(W_{-jt})dW_{-jt} \\
&= \beta_j''(w_j)\mu_K^2.
\end{aligned}
$$

The last equality comes from the fact that $\Gamma(w_j, W_{-jt})$ is symmetric and $[\Gamma_j(w_j, W_{-jt})]^T = \Gamma^j(w_j, W_{-jt})$, where Γ^j is jth column of Γ. ■

Next we proceed to asymptotic normality of the main stochastic term. Our proof for asymptotic normality is based on the central limit theorem of Lipster and Shirjaev (1980).

Lemma 5.2. (CLT for martingale differences: Lipster and Shirjaev, 1980, Corollary 6) *Let, for every $n > 0$, the sequence $\eta^n = (\eta_{nk,} F_k)$ be a square integrable martingale difference, i.e.,*

$$E(\eta_{nk}|\mathcal{F}_{k-1}) = 0,\ E(\eta_{nk}^2) < \infty,\ 1 \le k \le n \tag{5.9}$$

and let

$$\sum_{k=1}^{n} E(\eta_{nk}^2) = 1,\ \forall\, n \ge n_0 > 0. \tag{5.10}$$

The conditions

$$\sum_{k=1}^{n} E(\eta_{nk}^2|\mathcal{F}_{k-1}) \xrightarrow{p} 1, \ \ as\ n \ \ \rightarrow \ \ \infty, \tag{5.11}$$

$$\sum_{k=1}^{n} E(\eta_{nk}^2 I\left[|\eta_{nk}| > \varepsilon)|\mathcal{F}_{k-1}\right]) \xrightarrow{p} 0, as\ n \ \ \rightarrow \ \ \infty,\ \forall\, \varepsilon > 0, \tag{5.12}$$

are sufficient for convergence

$$\sum_{k=1}^{n} \eta_{nk} \xrightarrow{D} \ N(0,1), as\ n \rightarrow \infty.$$

Lemma 5.3. Assume that A.1 through A.2 hold. Then,

$$\sqrt{nh}\widetilde{\tau}_n^*(w_j) \xrightarrow{d} N(0, ||K||^2 \gamma_j(w_j)),$$

where

$$\gamma_j(w_j) = \int \Gamma_j^{-1}(w_j, s) E_{|W=(w_j,s)}[\sigma_\varepsilon^2(w_j, s, Z_1) Z_1 Z_1^T] \Gamma_j^{-1}(w_j, s)^T \frac{p_{W_{-j}}^2(s)}{p_W(w_j, s)} ds.$$

Proof. Using the same argument in (ii) of Lemma 5.1, it follows that

$$\begin{aligned}
\widetilde{\tau}_n^*(w_j) &= \frac{1}{n}\sum_{l=p+1}^{n+p} e_j^T S^{-1}(w_j, W_{-jl}) \tau_n(w_j, W_{-jl}) + o_p(\frac{1}{\sqrt{nh}}) \\
&\simeq \frac{1}{n}\sum_{l=p+1}^{n+p} p_W^{-1}(w) E_{|W=(w_j,W_{-jl})}^{-1}(Z_{j1} Z_1^T) \tau_n(w_j, W_{-jl}) \\
&= \frac{1}{n}\sum_{l=p+1}^{n+p} p_W^{-1}(w_j, W_{-jl}) [\Gamma_j^{-1}(w_j, W_{-jl}) \ O_{1\times(p+q)}] \tau_n(w_j, W_{-jl}) \\
&= \frac{1}{n}\sum_{l=p+1}^{n+p} p_W^{-1}(w_j, W_{-jl}) \Gamma_j^{-1}(w_j, W_{-jl}) \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt} - w_j\right) L_g\left(W_{-jt} - W_{-jl}\right) Z_t \varepsilon_t \\
&= \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt} - w_j\right) \{\frac{1}{n}\sum_{l=p+1}^{n+p} p_W^{-1}(w_j, W_{-jl}) \Gamma_j^{-1}(w_j, W_{-jl}) L_g\left(W_{-jt} - W_{-jl}\right)\} Z_t \varepsilon_t \\
&\simeq \frac{1}{n}\sum_{t=p+1}^{n+p} K_h\left(W_{jt} - w_j\right) p_W^{-1}(w_j, W_{-jt}) p_{W_{-j}}(W_{-jt}) \Gamma_j^{-1}(w_j, W_{-jt}) Z_t \varepsilon_t
\end{aligned}$$

Let

$$\begin{aligned}
\widetilde{\gamma}_{jt}(w_j, W_{-jt}, Z_t) &= \Gamma_j^{-1}(w_j, W_{-jt}) \sigma_\varepsilon^2(W_t, Z_t) Z_t Z_t^T \Gamma_j^{-1}(w_j, W_{-jt})^T p_W^{-2}(w_j, W_{-jt}) p_{W_{-j}}^2(W_{-jt}), \\
\gamma_j(w_j) &= E[\widetilde{\gamma}_{jt}(w_j, W_{-jt}, Z_t)],
\end{aligned}$$

and

$$\eta_t = \frac{\sqrt{h}}{\sqrt{n}} \frac{1}{\sqrt{||K||^2 \gamma_j(w_j)}} K_h \left(W_{jt} - w_j\right) p_W^{-1}(w_j, W_{-jt}) p_{W_{-j}}(W_{-jt}) \Gamma_j^{-1}(w_j, W_{-jt}) Z_t \varepsilon_t.$$

Note that

$$\begin{aligned}
\gamma_j(w_j) &= \int \Gamma_j^{-1}(w_j, s) \sigma_\varepsilon^2(W_t, Z_t) Z_t Z_t^T \Gamma_j^{-1}(w_j, s)^T p_{W_{-j}}^2(s) p_W^{-2}(w_j, s) p_{W,Z}(w_j, s, Z_t) dZ_t ds \\
&= \int \Gamma_j^{-1}(w_j, s) [\int \sigma_\varepsilon^2(W_t, Z_t) Z_t Z_t^T \frac{p_{W,Z}(w_j, W_{-jt}, Z_t)}{p_W(w_j, s)} dZ_t] \Gamma_j^{-1}(w_j, s)^T p_{W_{-j}}^2(s) p_W^{-1}(w_j, s) ds \\
&= \int \Gamma_j^{-1}(w_j, s) E_{|W=(w_j, s)}(\sigma_\varepsilon^2(W_t, Z_t) Z_1 Z_1^T) \Gamma_j^{-1}(w_j, s) \frac{p_{W_{-j}}^2(s)}{p_W(w_j, s)} ds.
\end{aligned}$$

Let $\mathcal{G}_t \equiv \sigma(v_t, v_{t-1}, .., v_0)$ is the σ-algebra generated by $\{v_t\}_{t=1}^n$ with $v_t = (y_t, X_{t+1}, W_{t+1})$. In the following, we show the validity of (5.9) through (5.12). The equality in (5.9) is obvious, since $E\left(\varepsilon_t | W_t, Z_t\right) = 0$.

From

$$\begin{aligned}
& E_{|\mathcal{G}_{t-1}}[\sqrt{h} K_h \left(W_{jt} - w_j\right) p_W^{-1}(w_j, W_{-jt}) p_{W_{-j}}(W_{-jt}) \Gamma_j^{-1}(w_j, W_{-jt}) Z_t \varepsilon_t]^2 \\
= \; & \frac{1}{h} K^2 \left(W_{jt} - w_j/h\right) \widetilde{\gamma}_{jt}(w_j, W_{-jt}),
\end{aligned}$$

it is easy to check that

$$\sum_{t=1}^n E(\eta_{nt}^2 | \mathcal{G}_{t-1}) = \frac{1}{nh} \sum_{t=1}^n \frac{K^2 \left(W_{jt} - w_j/h\right) \widetilde{\gamma}_j(w_j, W_{-jt})}{||K||^2 \gamma_j(w_j)]}, \tag{5.13}$$

and

$$\begin{aligned}
& E[\frac{1}{h} K^2 \left(W_{jt} - w_j/h\right) \widetilde{\gamma}_j(w_j, W_{-jt})] \\
= \; & E\{\frac{1}{h} K^2 \left(W_{jt} - w_j/h\right) p_W^{-2}(w_j, W_{-jt}) p_{W_{-j}}^2(W_{-jt}) \Gamma_j^{-1}(w_j, W_{-jt}) \sigma_t^2 Z_t Z_t^T \Gamma_j^{-1}(w_j, W_{-jt})^T\} \\
= \; & ||K||^2 E\{p_W^{-2}(w_j, W_{-jt}) p_{W_{-j}}^2(W_{-jt}) \Gamma_j^{-1}(w_j, W_{-jt}) [\sigma_t^2 Z_t Z_t^T] \Gamma_j^{-1}(w_j, W_{-jt})^T\} \\
= \; & ||K||^2 \gamma_j(w_j).
\end{aligned}$$

Hence, by the law of iterations and the stationarity condition, (5.10) holds;

$$\sum_{t=1}^n E(\eta_{nt}^2) = \sum_{t=1}^n E(E(\eta_{nt}^2 | \mathcal{G}_{t-1})) = 1.$$

To show (5.11), we just consider (5.13) and

$$\begin{aligned}
& \frac{1}{nh} \sum_{t=1}^n K^2 \left(W_{jt} - w_j/h\right) \widetilde{\gamma}_j(w_j, W_{-jt}) \\
& \xrightarrow{p} E[\frac{1}{h} K^2 \left(W_{jt} - w_j/h\right) \widetilde{\gamma}_j(w_j, W_{-jt})].
\end{aligned}$$

It remains to show (5.12). Using the conditions that $K(\cdot)$ is bounded and compactly supported, $p_W(\cdot)$ is bounded below from 0, and $E^{-1}(Z_t Z_t^T | W = w)$ exists for all w, we can show that for $n \geq n_0$,

$$\begin{aligned} \eta_t^2 &\leq C_1 \frac{1}{nh} K^2\left(\frac{W_{jt} - w_j}{h}\right) p_W^{-2}(w_j, W_{-jt}) \Gamma_j^{-1}(w_j, W_{-jt}) Z_t Z_t^T \Gamma_j^{-1}(w_j, W_{-jt})^T \varepsilon_t^2 \\ &\leq C_2 \frac{1}{nh} K\left(\frac{W_{jt} - w_j}{h}\right) {u_t}^2, \end{aligned}$$

where $C_i > 0$ is a constant and $u_t = (c^T Z_t)\varepsilon_t$. Hence

$$\begin{aligned} & E\left[\eta_{nt}^2 I(|\eta_{nt}| \geq \tau) | \mathcal{G}_{t-1}\right] \\ \leq\ & C_2 \frac{1}{nh} K\left(\frac{W_{jt} - w_j}{h}\right) c^T Z_t Z_t^T c E\left[\varepsilon_t^2 I(\left|c^T Z_t \varepsilon_t\right| \geq \tau nh C_2^{-1} ||K||_\infty^{-1/2}) | \mathcal{G}_{t-1}\right] \\ =\ & C_2 \frac{1}{nh} K\left(\frac{W_{jt} - w_j}{h}\right) (c^T Z_t Z_t^T c) \times o(1), \text{ as } n \to \infty, \end{aligned}$$

from $E\left(Z_{it} Z_{jt} \varepsilon_t^2\right) \leq C_3 E\left(Z_{it} Z_{jt}\right) < \infty$, where $o(1)$ does not depend on t. This implies

$$\begin{aligned} & \sum_{t=1}^n E\left[\eta_{nt}^2 I(|\eta_{nt}| > \tau) | \mathcal{G}_{t-1}\right] \\ \leq\ & o(1) \frac{1}{nh} \sum_{t=1}^n K\left(\frac{W_{jt} - w_j}{h}\right) (c^T Z_t Z_t^T c) \\ =\ & o(1) \int K(s) ds c^T E(Z_t Z_t^T) c, \text{ as } n \to \infty, \forall\, \tau > 0, \end{aligned}$$

which builds up the validity of all necessary conditions for

$$\sum_{t=1}^n \eta_{nt} \xrightarrow{D} N(0,1) \text{ as } n \to \infty.$$

■

References

[1] Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis, Forecasting, and Control. Holden Day, San Fransisco.

[2] Breiman, L. and J. H. Friedman (1985) Estimating optimal transformations for multiple regression and correlation (with discussion). J. Amer. Statist. Assoc. 80, 580-619.

[3] Cai, Z., J. Fan, and Q. Yao, (2000). Functional-coefficient regression models for nonlinear time series. J. Ameri. Statist. Assoc. 2000, 95, 941-956.

[4] Chen, R. and R.S. Tsay (1993) Functional-coefficient autoregressive models. J. Amer. Statist. Assoc. 88, 298-308.

[5] Cleveland, W.S., E. Grosse, and W.M. Shyu (1991) Local linera regression. In Statistical Models in S (Chambers, J.M. and Hastie, T.J., eds.), 309-376. Wadsworth & Brooks, Pacific Grove

[6] FAN, J. (1993). Local linear regresion smoothers and their minimax. The Annals of Statistics, 21, 196-216.

[7] FAN, J. AND I. GIJBELS (1992). Variable bandwidth and local linear regression smoothers. The Annals of Statistics, 20, 2008-2036.

[8] FAN, J. AND W. ZHANG, (1999). Statistical estimation in varying coefficient models The Annals of Statistics, 27, 1999, 1491-1518

[9] GRANGER, C.W.J. AND T. TERÄSVIRTA (1993). Modeling Nonlinear Economic Relationships, Oxford University Press, Oxford.

[10] HAGGAN, V. AND T. OZAKI (1981) Modeling nonlinear variations using an amplitude-dependent autoregressive time series model. Biometrica, 68, 189-196.

[11] HASTIE, T.J. AND R.J. TIBSHIRANI (1993). Varying-coefficients models (with discussion), J. of the Royal Statistical Society, Series B, 55, 757-796.

[12] KILIAN, L. AND M. TAYLOR (2001) Why is it so difficult to beat the random walk forecast of exchange rates ?, mimeo.

[13] LINTON, O.B. AND NIELSEN, J.B. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-100.

[14] LÜTKEPOHL, H., T. TERÄSVIRTA, AND J. WOLTERS (1998). Investigating stability and linearity of a German M1 money demand function, J. of Applied Economics 14, 511-525

[15] SCHWERT, G.W. (1989) Why does stock market volatility change over time ?, Journal of Finance, 44, 1115-1153.

[16] SILVERMAN, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall.

[17] TAYLOR, M.P. AND D.A. PEEL (2000). Nonlinear adjustment, long run equilibrium and exchange rate fundamentals, J. of International money and Finance, 11, 304-314.

[18] TAYLOR, M.P., D.A. PEEL AND L. SARNO (2001). Nonlinear adjustment in real exchange rates:towards a solution to the purchasing power parity puzzles, forth coming in Internatinal Economic Review

[19] TERÄSVIRTA, T. (1994), Specification, estimation, and evaluation of smooth autoregessive models, J. of the American Statistical Association, 89, 208-218

[20] TONG, H. (1990) Nonlinear Time Series: A Dynamic System Approach. Oxford University

[21] YANG, L., W. HÄRDLE, AND J. NIELSEN, (1999). Nonparametric autoregression with multiplicative volatility and additive mean. Journal of Time Series Analysis. **20**(5): 579-604.