Local Instrumental Variable (LIVE) Method For The Generalized Additive-Interactive Nonlinear Volatility Model

Michael Levine, Jinguang (Tony) Li *†‡

Abstract

In this article we consider a new separable nonparametric volatility model related to the GANARCH model of Kim and Linton (2004) [24]. Unlike the GANARCH model, it does not assume the known link function but includes second-order interaction terms in both mean and variance functions instead. The assumption of the known link function implies knowing the distribution of the data. The exact data distribution is often not so easy to verify, especially in the multivariate case; thus, it can be said that our model imposes fewer difficult to verify assumptions. Motivated by the local instrumental variable estimation method introduced by Kim and Linton (2004)[24], we propose an instrumental variable-based estimation method of both additive and interactive mean and variance component functions for this model. This method is computationally more effective than most other nonparametric estimation methods that can potentially be used to estimate components of such a model. Asymptotic behavior of the resulting estimators is investigated and their asymptotic normality is established. Explicit expressions for asymptotic means and variances of these estimators are also obtained. Simulation experiments provide strong evidence that these estimators are well-behaved in finite samples as well.

1 Introduction

Volatility modeling has been one of the most active research areas in empirical finance and time series econometrics in the past two decades. In practice, most empirical finance data reveal relatively little correlation in mean but a lot of correlation in conditional variance;

*Keywords and phrases: nonparametric volatility model, second-order interaction, time series, instrumental variable

†The work of Jinguang(Tony) Li was partially supported by the Purdue Research Foundation Grant #6904320

‡Both authors would like to thank Prof. Lijian Yang and Prof. Stefan Sperlich for helpful discussions

thus, modeling of the conditional variance (volatility) function is rather important. Pricing of financial derivatives, such as options, is dependent on the estimation of the underlying asset volatility (note that the classical Black-Scholes pricing formula assumes constant volatility of the underlying asset, which is usually untenable in practice); in order to assess measures of risks of traded assets, such as value at risk, a good volatility model is also needed. The most common pattern volatilities of financial assets exhibit is that of phases of relative tranquility followed by periods of high volatility (volatility clustering). Being able to model this feature has long been an important benchmark of a successful volatility model.

Numerous parametric volatility models were proposed following the seminal work of Engle (1982)[12] that introduced the classical ARCH model. Bollerslev (1986)[7] provided the first generalization of ARCH by introducing GARCH (generalized ARCH) modeling framework that made more parsimonious volatility modeling possible. Since models that fall into the (G)ARCH framework suffer from their inherent symmetry with respect to the sign of the process and thus are incapable of incorporating such an important phenomenon as the leverage effect (Black (1976)[6]), additional models had to be introduced explicitly for that purpose. Two classical examples are exponential GARCH (EGARCH) (Nelson, 1991[33]) and threshold GARCH (TGARCH) (Glosten, Jagannathan and Runkle (1993)[16] and Zakoian (1994)[49]). Numerous other models have also been introduced over the years. Two classical review articles that provide thorough catalogues of numerous volatility models are Bollerslev, Chou and Kroner (1992)[8] and Shephard (1996)[40]; the survey would also be incomplete without mentioning a recent review of multivariate GARCH-type models by Bauwens, Laurent and Rombouts (2006)[5] and the more specific survey of ARCH(∞) models in Giraitis, Leipus and Surgailis (2007)[11]. All of the models mentioned so far are parametric: values of stochastic volatility function are explicit nonlinear functions of the past values of asset price and/or past values of the volatility function itself. As such, all of them tend to suffer from the common problem of all parametric models - misspecification, especially when there is no theoretical reason to prefer one over another.

This suggests switching to nonparametric modeling as a possible alternative. Let us first consider the model

$$y_t = m(y_{t-1}) + v^{1/2}(y_{t-1})\epsilon_t \tag{1}$$

where the error process ϵ_t is either iid or a martingale difference sequence with the unit scale, that is, $E[\epsilon_t|\mathcal{F}_t] = 0$, $E[\epsilon_t^2|\mathcal{F}_t] = 1$ and $\mathcal{F}_t$ is the σ-algebra of events generated by $\{y_k\}_{k=-\infty}^{t}$. Both $m(\cdot)$ and $v(\cdot)$ belong in some functional smoothness class but are otherwise unknown and $y_t \in \mathcal{R}$. This model is commonly known as CHARN (conditionally heteroscedastic autoregressive model). The kernel estimation of the mean function $m(\cdot)$ in (1) has been studied by Robinson (1983)[39], Auestad and Tjostheim (1990)[3] and Härdle and Vieu (1992)[17]. The model with zero mean where the main interest lies in estimating the volatility function had been considered in Pagan and Schwert (1990)[36] and Pagan and Hong (1991)[37]; they also studied the multi-lag version od (1) where, in particular, $v \equiv v(y_{t-1}, y_{t-2}, \cdots, y_{t-d})$ for some $d > 0$. For the same model (1), Masry and Tjostheim (1995)[30] estimated the mean and variance function $m(\cdot)$ and $v(\cdot)$ jointly using the Nadaraya-Watson Kernel estimator while Härdle and Tsybakov (1997)[19] applied local linear fit to estimate the same model; they also derived the asymptotic properties of the joint estimator. Härdle, Tsybakov and Yang (1996)[18] gave the multivariate extension of the above problem. Fan and Yao (1998)[14] and Ziegelmann (2002)[50] proposed local linear least square estimators for the volatility function while Avramidis (2002)[4] gave an extension based on the local linear maximum likelihood function.

Flexibility of a nonparametric volatility model in comparison with the parametric one comes at a cost. In practice it is often necessary to include many lagged variables to obtain a good fit. The resulting problem of nonparametric multivariate function estimation suffers from the well-known "curse of dimensionality" whereby the rate of convergence quickly decreases as the number of dimensions increases. This problem has been first clearly elucidated in Silverman(1986)[41]. It is also hard to describe and interpret the multidimensional surface of dimensionality more than 2.

Generalized additive models (GAM) offer an intermediate position between the complete generality of nonparametric models and the restrictiveness of parametric ones. The classical GAM model assumes that the conditional variance function $v(\cdot) \equiv const$ and considers

$$m(y_{t-1}, \ldots, y_{t-d}) = C_m + \sum_{j=1}^{d} m_j(y_{t-j}).$$

The best achievable rate of convergence for an estimate of m_α, $\alpha = 1, \ldots, d$ is the same as in one-dimensional nonparametric regression. [See Stone (1985[43],1986[44])]; as an example, any twice continuously differentiable mean function $m_\alpha(\cdot)$ can be estimated at the rate of $n^{-2/5}$, regardless of d. The defining monograph on the generalized additive models data is Hastie and Tibshirani (1990)[21]. It is entirely natural to make the next step and introduce the generalized additive structure in the variance function as well, assuming that

$$v(y_{t-1},\dots,y_{t-d}) = C_v + \sum_{j=1}^{d} v_j(y_{t-j}).$$

It is also known that the best achievable rate of convergence for components of v is the same as the one for the variance function in the one-dimensional model (1). (again, see Stone (1985[43],1986[44])). Yang, Härdle and Nielsen (1999)[48] introduced the nonparametric volatility model with additive mean structure but multiplicative volatility; they argued that this is rather natural since volatility function must be presumed to be non-negative. Kim and Linton (2004)[24] provided further generalization by introducing the generalized additive nonlinear ARCH model (GANARCH)

$$\begin{aligned} y_t &= m(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) + u_t \\ u_t &= v^{1/2}(y_{t-1}, y_{t-2}, \cdots, y_{t-d})\varepsilon_t \end{aligned} \tag{2}$$

where

$$\begin{aligned} m(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) &= F_m(C_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha})) \\ v(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) &= F_v(C_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha})) \end{aligned} \tag{3}$$

where $m_\alpha(\cdot)$ and $v_\alpha(\cdot)$ are any smooth but unknown functions, while $F_m(\cdot)$ and $F_v(\cdot)$ are known monotone transformations. It is quite clear that the model of Yang, Härdle and Nielsen (1999)[48] is the special case of (2) when F_m is an identity function and F_v is an exponent. The GANARCH model, although quite interesting from the theoretical viewpoint, limits somewhat our horizon because of the known link function assumption which implies that some information about the distribution of data is available. It is often hard to tell which functional transformation is right for the given data, particularly so in multidimensional settings. Horowitz (2001)[22] considered a model similar to (2)-(3) with unknown link functions but only in cross-sectional context. Later, Horowitz and Mammen (2006)[23] considered an even more general model that contains the generalized additive model with unknown link function as a special case and is also a natural generalization of the neural-network models; however, it is also meant for cross-sectional setting only. Direct generalization to the time-series case seems to be rather difficult and this point was noticed in Kim and Linton (2004)[24].

In this paper, we propose a different way to generalize the GAM framework by considering nonparametric "interactions" $m_{\alpha\beta}(\cdot,\cdot)$, $v_{\alpha\beta}(\cdot,\cdot)$ for any pair of lagged variables

(y_α, y_β), $1 \le \alpha < \beta \le d$ in both mean and variance functions. Such a model is related to the model of Yang, Härdle and Nielsen (1999) [48] but utilizes the additive, instead of multiplicative, structure for its volatility function; it also provides a reasonable amount of flexibility while retaining relative simplicity of interpretation. It does not require us to guess the unknown link function but rather presents a more data-driven approach; if the asymptotic distribution of the interaction estimators is known, a large-sample test can be easily constructed that would allow us to select as many interaction terms as the data itself dictates. Of course, conceptually it is entirely feasible to consider three-way nonparametric "interactions" and so on; note, however, that the "curse of dimensionality" and difficulty of interpretation gradually return as we increase the dimensionality of "interaction terms".

The rest of the paper is organized as follows. In section 2 we introduce our model formally. Section 3 describes the main estimation idea and defines the local instrumental variable(LIVE) estimators. In section 4 we show the main results, including the asymptotic normality of our LIVE estimators. Section 5 presents results of the Monte-Carlo simulations with both uniformly and normally distributed innovations. The article ends with the mathematical appendix that contains proofs of all main results.

2 The Model

In this section, our model and some basic assumptions are introduced. We consider the following additive-interactive nonlinear ARCH model:

$$
\begin{aligned}
& y_t = m(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) + v^{1/2}(y_{t-1}, y_{t-2}, \cdots, y_{t-d})\varepsilon_t \qquad (4) \\
& m(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = C_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha}) \\
& + \sum_{1\le\alpha<\beta\le d} m_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta}) \\
& v(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = C_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha}) \\
& + \sum_{1\le\alpha<\beta\le d} v_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta})
\end{aligned}
$$

where $m_\alpha(\cdot)$ and $v_\alpha(\cdot)$ are smooth but unknown univariate functions while $m_{\alpha\beta}(\cdot)$ and $v_{\alpha\beta}(\cdot)$ are also smooth but unknown bivariate functions. The error process ε_t is assumed to be a martingale difference sequence with unit scale. Such a model can be viewed as a nonlinear autoregressive time series model.

Under some weak assumptions, the general nonlinear autoregressive time series model can be shown to be stationary and strongly mixing with mixing coefficients decaying exponentially fast. Auestad and Tjostheim (1990)[3] used α-mixing or geometric ergodicity to identify the nonlinear time series model. Ango Nze (1992)[1] studied the L_1 geometric ergodicity of the multivariate generalization of the model (4)

$$X_t = f(X_{t-1}, \ldots, X_{t-p}) + H(X_{t-1}, \ldots, X_{t-q})\epsilon_t \tag{5}$$

where X_t and ϵ_t are two sequences of m-dimensional random variables defined on a common probability space and ϵ_t is an m-dimensional white noise process. Ango Nze (1992)[1] gave, probably, the first in the literature sufficient condition that ensures L_1 geometric ergodicity of the general model (5). Lu and Jiang (2001)[28] derived another sufficient condition that also ensures L_1 geometric ergodicity of the model (5) but that is much less restrictive. In this paper we impose constraints from Lu and Jiang (2001)[28] and, in doing so, assume that the conditions for strict stationarity and strong mixing property of the process $\{y_t\}_{t=1}^n$ in (4) are met.

The issue of identifiability arises naturally. First, note that one can add a constant to any of the components $m_\alpha(\cdot)$, $v_\alpha(\cdot)$, $m_{\alpha\beta}(\cdot)$ or $v_{\alpha\beta}(\cdot)$ and subtract the constant from another component without changing the model. It is also possible to add an arbitrary function $f(\cdot)$ to the additive component ($m_\alpha(\cdot)$ or $v_\alpha(\cdot)$) and then subtract it from the interactive component ($m_{\alpha\beta}(\cdot)$ or $v_{\alpha\beta}(\cdot)$). Of course, this presents identification problems when trying to estimate the model components. To prevent ambiguity, the following identifiability conditions are imposed:

$$E[m_\alpha(Y_{t-\alpha})] = 0\,, \alpha = 1, \cdots, d \tag{6}$$

$$E[v_\alpha(Y_{t-\alpha})] = 0\,, \alpha = 1, \cdots, d \tag{7}$$

and

$$E[m_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\alpha} = y_\alpha] = E[m_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\beta} = y_\beta] = 0 \tag{8}$$

$$E[v_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\alpha} = y_\alpha] = E[v_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\beta} = y_\beta] = 0 \tag{9}$$

where $1 \leq \alpha < \beta \leq d$.

Remark: Similar conditions were imposed in Sperlich, Tjöstheim and Yang (2002) for the model that considers only the interactions in the mean function. As is the case in their paper, if our representation as given in (4) doesn't satisfy conditions (6) and (8), one can easily change it to ensure that it does. To achieve this goal, one can:

1. Replace all $\{m_{\alpha\beta}(y_\alpha, y_\beta)\}_{1\le\alpha<\beta\le d}$ by $\{m_{\alpha\beta}(y_\alpha, y_\beta) - m_{\alpha,\alpha\beta}(y_\alpha) - m_{\beta,\alpha\beta}(y_\beta) + C_{0,\alpha\beta}\}_{1\le\alpha<\beta\le d}$ where

$$\begin{aligned} m_{\alpha,\alpha\beta}(y_\alpha) &= \int m_{\alpha\beta}(y_\alpha, v) p_\beta(v) dv, \\ m_{\beta,\alpha\beta}(y_\beta) &= \int m_{\alpha\beta}(u, y_\alpha) p_\alpha(u) du, \\ C_{0,\alpha\beta} &= \int m_{\alpha\beta}(u, v) p_{\alpha\beta}(u, v) du dv \end{aligned}$$

   with $p_\alpha(\cdot)$, $p_\beta(\cdot)$ the marginal densities of $Y_{t-\alpha}$ and $Y_{t-\beta}$ respectively, and $p_{\alpha\beta}(\cdot)$ the joint density of $(Y_{t-\alpha}, Y_{t-\beta})$.

2. Adjust the $\{m_\alpha(y_\alpha)\}_{\alpha=1}^d$ and the constant term C_m accordingly so that $m(\cdot)$ remains unchanged.

3. Replace the $\{m_\alpha(y_\alpha)\}_{\alpha=1}^d$ by $\{m_\alpha(y_\alpha) - C_{0,\alpha}\}_{\alpha=1}^d$, where $C_{0,\alpha} = \int m_\alpha(u) p_\alpha(u) du$.

4. Adjust the constant term C_m accordingly so that $m(\cdot)$ remains unchanged.

The analogous procedure can be followed if the model (4) doesn't satisfy (7) and (9). As a result, the set of identifiability conditions (6)-(9) does not really impose any additional constraints on the model (4).

3 The Local Instrumental Variable Estimators

3.1 Estimation Methods

The main objective of this paper is the estimation of the additive and interactive components of both the mean and volatility functions in (4). Allowing two-way nonparametric interactions in addition to additive terms makes the model much more flexible compared to the traditional GAM model, while still greatly alleviating the "curse of dimensionality". It is known (see, for example, Stone (1994)[45]) that the interactive terms can be estimated at the optimal rate $O(n^{-q/(2q+2)})$ while for the additive ones this rate is $O(n^{-q/(2q+1)})$ whenever the function to be estimated is q-smooth in the sense of Stone (1994)[45]. In other words, estimating the additive effect is as hard as one-dimensional nonparametric smoothing and the interactive effect as hard as a two-dimensional nonparametric smoothing.

The GAM literature suggests several possible approaches that can be conceptually extended to our model. The first one is the so-called backfitting algorithm of Breiman and Friedman (1985)[9]; see also Hastie and Tibshirani (1987[20], 1990[21]) as well as Buja, Hastie and Tibshirani (1989)[10]. The classical backfitting algorithm essentially consists of repeated iterative application of some one-dimensional smoother (e.g. local polynomial regression) until convergence. As a consequence, it would be more precise to say that the backfitting is not just a method but a collection of methods depending on the particular nonparametric smoother (local polynomial regression, smoothing spline etc) being used. Both classical backfitting and the modified backfitting approach (one-step efficient estimator) of Linton (1997)[27] can be conceptually extended to our model. A significant disadvantage of the backfitting method is, however, the difficulty of its theoretical analysis. The asymptotic properties of backfitting estimators in the generalized additive model case had not been established until the pioneering work of Opsomer and Ruppert (1997)[35] and Mammen, Linton and Nielsen (1999)[29]. The latter group of authors investigated the modified (one-step) backfitting estimator while the first one worked with the original definition. Both papers derived asymptotic properties of the estimators considered, establishing the geometric rate of convergence under some regularity conditions. A lot of their analysis relies on the projector theory which is hard to extend to analysis of interaction terms.

Another approach is marginal integration, which was introduced independently by Newey (1994)[34], Tjostheim and Auestad (1994)[46], and Linton and Nielsen (1995)[26]. Marginal integration approach has been extended to fit the nonparametric model with interactions in Sperlich, Tjostheim and Yang (2002)[42], albeit in the cross-sectional setting only. One advantage of the marginal integration method is that its statistical properties are relatively easy to describe; specifically, one can easily prove central limit theorems for the resulting estimators as well as give explicit expressions for their asymptotic biases and variances. However, marginal integration is relatively computationally expensive and another method based on the local instrumental variable (LIVE) approach outperforms it.

In this paper, we suggest an alternative approach that uses the above mentioned LIVE idea and represents a generalization of the approach from Kim and Linton (2004)[24]. One of the most important advantages of this approach is that it reduces the number of smoothings required to estimate a model component by a factor of n; for example, it takes only $O(n^2)$ smoothings to estimate an additive component in the example we give later in the next section as opposed to $O(n^3)$ when using marginal integration. This reduction in the number of computations was noticed earlier in Kim and Linton (2004)[24] in the context of GANARCH model. Also, the asymptotic properties of the LIVE estimator are fairly easy to derive and, in this regard, it is much more tractable than the backfitting. The rest of this section is dedicated to the description of the LIVE approach in our context.

3.2 Instrumental Variable Approach

Instrumental variable approach is widely used in econometric modeling. [See Angrist, Imbens and Rubin (1996)[2], and the accompanying discussion of the econometric concept of instrument and some background literature.] In the following we explain the basic idea behind the instrumental variable method and describe the estimation procedure. For ease of exposition, we use a simple example ($d = 3$) with i.i.d. data. In section 4 we extend the procedure to the general case of (4). Consider the following model for i.i.d. data (Y, X_1, X_2, X_3)

$$Y = m(X_1, X_2, X_3) + v^{1/2}(X_1, X_2, X_3)\varepsilon \tag{10}$$

$$m(X_1, X_2, X_3) = C_m + m_1(X_1) + m_2(X_2) + m_3(X_3) \tag{11}$$

$$+m_{12}(X_1, X_2) + m_{13}(X_1, X_3) + m_{23}(X_2, X_3)$$

$$v(X_1, X_2, X_3) = C_v + v_1(X_1) + v_2(X_2) + v_3(X_3) \tag{12}$$

$$+v_{12}(X_1, X_2) + v_{13}(X_1, X_3) + v_{23}(X_2, X_3)$$

and assume that the identifiability conditions (6)-(9) are satisfied. Let us denote $\mathbf{X} = (X_1, X_2, X_3)$. Let $p_\alpha(X_\alpha)$ be a marginal density of X_α, $\alpha = 1, 2, 3$, $p_{\alpha\beta}(X_{\alpha\beta})$ be a joint density of (X_α, X_β), $\alpha, \beta = 1, 2, 3$ and $p(\mathbf{X}) = p(X_1, X_2, X_3)$ be the joint density of (X_1, X_2, X_3). For us, an instrumental variable (instrument for short) is defined as a random variable W such that $E[W|F] \neq 0$ but $E[W\eta|F] = 0$ where F is a σ-algebra generated by any of $\{X_1, X_2, X_3\}$ and η is a missing "regression variable"; for example, (10) can be expressed as $Y = m_1(X_1) + \eta + v^{1/2}(X_1, X_2, X_3)\epsilon$ where η includes all of the additive mean components except for the "main effect" $m_1(X_1)$. In this case, we note that

$$m_1(x_1) = \frac{E(WY|X_1 = x_1)}{E(W|X_1 = x_1)}. \tag{13}$$

As suggested before in Kim and Linton (2004), it is now possible to use local smoothers in both the numerator and denominator of (13) to estimate the function $m_1(\cdot)$. However, it is best to note that any choice of an instrument needed to define the function $m_1(\cdot)$ is unique only up to a measurable function of X_1 as a factor. It is clear that we can choose

$W = \frac{p_{23}(X_2,X_3)}{p(\mathbf{X})}$ as an instrument in (13); however, if we choose $W = \frac{p_1(X_1)p_{23}(X_2,X_3)}{p(\mathbf{X})}$, we have another instrument that does not only satisfy the definition of the instrumental variable but for whom $E(W|X_1) = 1$, and, therefore,

$$m_1(x_1) = E(WY|X_1 = x_1). \tag{14}$$

(14) allows us to use one smoother instead of two to estimate the value of $m_1(x_1)$. This is the definition we will be using in the rest of this paper.

As a next step, we describe how to obtain an identity for each of the additive and interactive components of both mean function $m(\mathbf{X})$ and variance function $V(\mathbf{X})$ that can be used for estimation purposes.

1. Due to identifiability conditions above, $C_m = E[Y]$

2. Define $Y_m = Y - C_m$ and an instrumental variable $W_1 = \frac{p_1(X_1)p_{23}(X_2,X_3)}{p(\mathbf{X})}$. It's easy to verify that $E[W_1|X_1] = 1$, $E[W_1 m_i|X_1] = 0$, $i = 2,3$, and $E[W_1 m_{ij}|X_1] = 0$, $i,j = 1,2,3$. Therefore, multiplying both sides of (10) by W_1 and taking an expectation conditionally on $X_1 = x_1$ we obtain

$$m_1(x_1) = E[W_1 Y_m|X_1 = x_1]$$

   Similarly, we can define analogous instrumental variables $W_2 = \frac{p_2(X_2)p_{13}(X_1,X_3)}{p(\mathbf{X})}$ and $W_3 = \frac{p_3(X_3)p_{12}(X_1,X_2)}{p(\mathbf{X})}$; they, in turn, produce identities $m_i(x_i) = E[W_i Y_m|X_i = x_i]$, $i = 2,3$

3. Now, define the residual $\tilde{Y}_m = Y_m - [m_1(X_1) + m_2(X_2) + m_3(X_3)]$ and a new instrumental variable $W_{12} = \frac{p_{12}(X_1,X_2)p_3(X_3)}{p(\mathbf{X})}$.
   It's easy to verify that $E[W_{12}|X_1,X_2] = 1$ but $E[W_{12}m_{13}|X_1,X_2] = 0$ and $E[W_{12}m_{23}|X_1,X_2] = 0$ which leads us to the identity

$$m_{12}(x_1,x_2) = E[W_{12}\tilde{Y}_m|X_1 = x_1, X_2 = x_2]$$

   In an analogous way, we can define instruments $W_{13} = \frac{p_{13}(X_1,X_3)p_2(X_2)}{p(\mathbf{X})}$ and $W_{23} = \frac{p_{23}(X_2,X_3)p_1(X_1)}{p(\mathbf{X})}$ and obtain identities $m_{i3}(x_i,x_3) = E[W_{i3}\tilde{Y}_m|X_i = x_i, X_3 = x_3]$, $i = 1,2$

4. To identify components of the conditional variance functions, assume that the mean function $m(X_1,X_2,X_3)$ is known. Then, we can center the data to define $Y^* = [Y - m(X_1,X_2,X_3)]^2$ and note that $C_v = E[Y^*]$.

5. Define $Y_v^* = Y^* - C_v$ and note that $v_i(x_i) = E[W_i Y_v^* | X_i = x_i]$, $i = 1, 2, 3$.

6. Again, define $\tilde{Y}_v^* = Y_v^* - [v_1(X_1) + v_2(X_2) + v_3(X_3)]$ and we have $v_{ij}(x_i, x_j) = E[W_{ij} \tilde{Y}_v^* | X_i = x_i, X_j = x_j]$, $i, j = 1, 2, 3$.

Of course, in practice the mean function $m(X_1, X_2, X_3)$ will not be known and all of the conditional expectations need to be estimated. Thus, several additional steps are needed before we can arrive at the working estimation algorithm. In the next subsection we describe this estimation algorithm in details in the general setting (4).

3.3 LIVE algorithm for the additive-interactive model (4)

We begin with introducing the notation that will be used repeatedly throughout the paper. We denote

$$
\begin{aligned}
\mathbf{y}_t &= (y_{t-1}, \ldots, y_{t-d}) \\
\mathbf{y} &= (y_1, \ldots, y_d) \\
\mathbf{y}_{t,\underline{\alpha}} &= (y_{t-1}, \ldots, y_{t-\alpha+1}, y_{t-\alpha-1}, \ldots, y_{t-d}) \\
\mathbf{y}_{\underline{\alpha}} &= (y_1, \ldots, y_{\alpha-1}, y_{\alpha+1}, \ldots, y_d) \\
\mathbf{y}_{t,\alpha\beta} &= (y_{t-\alpha}, y_{t-\beta}) \\
\mathbf{y}_{\alpha\beta} &= (y_\alpha, y_\beta) \\
\mathbf{y}_{t,\underline{\alpha\beta}} &= (y_{t-1}, \cdots, y_{t-\alpha+1}, y_{t-\alpha-1}, \ldots, y_{t-\beta+1}, y_{t-\beta-1}, \ldots, y_{t-d}) \\
\mathbf{y}_{\underline{\alpha\beta}} &= (y_1, \ldots, y_{\alpha-1}, y_{\alpha+1}, \ldots, y_{\beta-1}, y_{\beta+1}, \ldots, y_d)
\end{aligned}
$$

The underscore in the above means that a particular direction α or directions α and β have been omitted; boldface is used for all multidimensional quantities. Let $p_\alpha(y_\alpha)$ be the marginal density of $y_{t-\alpha}$ while $p_{\underline{\alpha}}(\mathbf{y}_{\underline{\alpha}})$, $p_{\alpha\beta}(y_\alpha, y_\beta)$, $p_{\underline{\alpha\beta}}(\mathbf{y}_{\underline{\alpha\beta}})$ and $p(\mathbf{y})$ are joint densities of $\mathbf{y}_{t,\underline{\alpha}}$, $\mathbf{y}_{t,\alpha\beta}$, $\mathbf{y}_{t,\underline{\alpha\beta}}$ and $\mathbf{y}_t$, respectively.

1. **<u>Preliminary density estimation</u>**

   As we mentioned before, we use regular product kernel density estimators. Specifically, we estimate the marginal density $p_\alpha(\cdot)$ as

$$
\hat{p}_\alpha(y_\alpha) = \frac{1}{ng} \sum_{t=1}^{n} L\left(\frac{y_{t-\alpha} - y_\alpha}{g}\right), \alpha = 1, 2, \cdots, d
$$

and the joint densities $p_{\alpha\beta}$, $p_{\underline{\alpha\beta}}$, $p_{\underline{\alpha}}$ and $p(\mathbf{y})$ as

$$\begin{aligned}
\hat{p}_{\alpha\beta}(y_\alpha, y_\beta) &= \frac{1}{ng^2}\sum_{t=1}^{n} L\left(\frac{y_{t-\alpha}-y_\alpha}{g}\right) L\left(\frac{y_{t-\beta}-y_\beta}{g}\right), 1 \leq \alpha < \beta \leq d \\
\hat{p}_{\underline{\alpha\beta}}(y_{\underline{\alpha}}, y_{\underline{\beta}}) &= \frac{1}{ng^{d-2}}\sum_{t=1}^{n} \prod_{\substack{\lambda=1 \\ \lambda\notin\{\alpha,\beta\}}}^{d} L\left(\frac{y_{t-\lambda}-y_\lambda}{g}\right), 1 \leq \alpha < \beta \leq d \\
\hat{p}_{\underline{\alpha}}(y_{\underline{\alpha}}) &= \frac{1}{ng^{d-1}}\sum_{t=1}^{n} \prod_{\substack{\lambda=1 \\ \lambda\neq\alpha}}^{d} L\left(\frac{y_{t-\lambda}-y_\lambda}{g}\right), \alpha = 1, 2, \cdots, d \\
\hat{p}(\mathbf{y}) &= \frac{1}{ng^{d}}\sum_{t=1}^{n} \prod_{\alpha=1}^{d} L\left(\frac{y_{t-\alpha}-y_\alpha}{g}\right).
\end{aligned}$$

In the above, $g = g(n)$ is the bandwidth and $L(\cdot)$ is the unimodal one-dimensional symmetric kernel function.

Remark Of course, multivariate product kernels are not the only possibility we could have considered. In general, two popular ways of constructing multivariate kernels are usually considered. The product kernel is the first while the second is the so-called spherically symmetric multivariate kernel. In general, multivariate product kernel based estimators are slightly less efficient than those based on spherically symmetric kernels (for details, see e.g. Wand and Jones (1995)[47]). However, since the observed loss of efficiency is rather minor, we prefer to use the product kernel which implies an easy and straightforward notation.

2. **Estimation of the constant component of the mean C_m**

C_m can be directly estimated as

$$\hat{C}_m = \frac{1}{n}\sum_{t=1}^{n} y_t.$$

3. **Estimation of the additive components of the mean $m_\alpha(\cdot)$**

Define the instrumental variable

$$\hat{W}_\alpha(\mathbf{y}_t) = \frac{\hat{p}_\alpha(y_{t-\alpha})\hat{p}_{\underline{\alpha}}(\mathbf{y}_{t,\underline{\alpha}})}{\hat{p}(\mathbf{y}_t)}, \alpha = 1, 2, \cdots, d$$

denote $\tilde{y}_t = y_t - \hat{C}_m$ and use it to estimate $m_\alpha(y_\alpha)$ as

$$\hat{m}_\alpha(y_\alpha) = E[\hat{W}_\alpha(\mathbf{y}_t)\tilde{y}_t | y_{t-\alpha} = y_\alpha]\,, \alpha = 1, 2, \cdots, d$$

Of course, the conditional expectation above needs to be estimated itself. In practice, $\hat{m}_\alpha(y_\alpha)$ in the above is determined as the minimizer a_α of the kernel-weighted least squares problem

$$\min_{a_\alpha, b_\alpha} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha)\{\hat{W}_\alpha(\mathbf{y}_t)\tilde{y}_t - a_\alpha - b_\alpha(y_{t-\alpha} - y_\alpha)\}^2.$$

which is equivalent to smoothing $\hat{W}_\alpha(\mathbf{y}_t)\tilde{y}_t$ using the local linear regression. The use of local linear regression is adopted to avoid the lack of design adaptivity and increased bias associated with using simpler kernel regression. (See Fan and Gijbels (1996)[13] for more details concerning the right choice of the local polynomial regression order under varying circumstances.)

4. **Estimation of the interactive components of the mean $m_{\alpha\beta}(\cdot)$**
   Let us denote

$$\bar{y}_t = y_t - \left[\hat{C}_m + \sum_{\alpha=1}^{d} \hat{m}_\alpha(y_{t-\alpha})\right]$$

   Define the instrumental variable

$$\hat{W}_{\alpha\beta}(\mathbf{y}_t) = \frac{\hat{p}_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta})\hat{p}_{\underline{\alpha\beta}}(\mathbf{y}_{t,\underline{\alpha\beta}})}{\hat{p}(\mathbf{y}_t)}, 1 \leq \alpha < \beta \leq d$$

   and estimate the interactive component $m_{\alpha\beta}$ by means of the minimizer $a_{\alpha\beta}$ of the two-dimensional kernel-weighted least squares problem

$$\min_{a_{\alpha\beta}, \mathbf{b}_{\alpha\beta}} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha)K_h(y_{t-\beta} - y_\beta) \times$$
$$\times\{\hat{W}_{\alpha\beta}(\mathbf{y}_t)\bar{y}_t - a_{\alpha\beta} - b_{\alpha\beta,\alpha}(y_{t-\alpha} - y_\alpha) - b_{\alpha\beta,\beta}(y_{t-\beta} - y_\beta)\}^2.$$

   In the above, the vector "slope" $\mathbf{b}_{\alpha\beta} = (b_{\alpha\beta,\alpha}, b_{\alpha\beta,\beta})'$.

5. **Estimation of the constant component of the variance C_v**
   Denote the squared residuals from the mean estimation

$$y_t^* = \left(\bar{y}_t - \sum_{1 \leq \alpha < \beta \leq d} \hat{m}_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta})\right)^2$$

   and estimate C_v as $\hat{C}_v = \frac{1}{n}\sum_{t=1}^{n} y_t^*$.

6. **Estimation of the additive components of the variance $v_\alpha(\cdot)$**

   Using the instrumental variables defined in step (3) we can estimate $v_\alpha(\cdot)$ as the minimizer of the localized least squares problem

$$\min_{a_\alpha, b_\alpha} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha)\{\hat{W}_\alpha(\mathbf{y}_t) y_t^* - a_\alpha - b_\alpha(y_{t-\alpha} - y_\alpha)\}^2.$$

7. **Estimation of the interactive components of the variance$v_{\alpha\beta}(\cdot)$**
   Denote

$$\tilde{y}_t^* = y_t^* - \left[\hat{C}_v + \sum_{\alpha=1}^{d} \hat{v}_\alpha(y_{t-\alpha})\right]$$

   and estimate interactive components $v_{\alpha\beta}(\cdot)$ as

$$\min_{a_{\alpha\beta}, \mathbf{b}_{\alpha\beta}} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha) K_h(y_{t-\beta} - y_\beta) \times$$
$$\times \{\hat{W}_{\alpha\beta}(\mathbf{y}_t)\tilde{y}_t^* - a_{\alpha\beta} - b_{\alpha\beta,\alpha}(y_{t-\alpha} - y_\alpha) - b_{\alpha\beta,\beta}(y_{t-\beta} - y_\beta)\}^2.$$

4 Main Results

In this section we state the main results for estimation in our additive-interactive nonlinear ARCH model. For this, we need the following definitions and assumptions. Let $\mathcal{F}_a^b$ be the σ-algebra generated by $\{y_t\}_{t=a}^{t=b}$ and $\alpha(k)$ the strong mixing coefficient of $\{y_t\}$ defined by

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

1. $\{y_t\}_{t=1}^\infty$ is a stationary and strongly mixing process generated by (1)-(3), with a mixing coefficient $\alpha(k)$ such that $\sum_{k=0}^\infty k^a \{\alpha(k)\}^{1-2/\nu} < \infty$, for some $\nu > 2$ and $a > (1 - 2/\nu)$. For simplicity, we assume that the process $\{y_t\}_{t=1}^\infty$ has a compact support.

2. The functions $m_\alpha(\cdot), m_{\alpha\beta}(\cdot), v_\alpha(\cdot), v_{\alpha\beta}(\cdot)$, $1 \leq \alpha, \beta \leq d$, are continuous and twice differentiable with bounded (partial) derivatives on the compact support

3. The joint and marginal density functions, $p(\cdot), p_\alpha(\cdot), p_{\underline{\alpha}}(\cdot), p_{\alpha\beta}(\cdot)$ and $p_{\underline{\alpha\beta}}(\cdot)$ are twice continuously differentiable and has bounded third derivatives. All of the density functions above are also bounded away from zero on the compact support.

4. The kernels $L(\cdot)$ and $K(\cdot)$ are bounded, nonnegative, symmetric (around zero), compactly supported, Lipschitz continuous, and satisfying

$$\int L(\underline{u})d\underline{u} = 1, \quad \int \underline{u}L(\underline{u})d\underline{u} = 0,$$
$$\int K(\underline{u})d\underline{u} = 1, \quad \int \underline{u}K(\underline{u})d\underline{u} = 0.$$

Furthermore we assume their moments of order higher than 2 are not equal to zero and $\| \underline{u} \|^2 L(\underline{u}) \in L_1$, $\| \underline{u} \|^4 K(\underline{u}) \in L_1$ and $\| \underline{u} \|^{(2\nu+d)} K(\underline{u}) \to 0$ as $\| \underline{u} \| \to \infty$.

5. (a) As $g \to 0$ and $n \to \infty$, $ng^d \to \infty$
   (b) As $h \to 0$ and $n \to \infty$, $nh \to \infty$
   (c) As $g \to 0$ and $h \to 0$, $\frac{g^{2\nu}}{h} \to 0$
   (d) i. There exists a sequence of positive integers satisfying $t(n) \to \infty$ and $t(n) = o(\sqrt{nh})$ such that $\sqrt{\frac{n}{h}}\alpha(t(n)) \to 0$
       ii. $\sqrt{\frac{\log n}{nh}} \to 0$ as $n \to \infty$, $h \to 0$ and $nh \to \infty$

6. (a) As $g \to 0$ and $n \to \infty$, $ng^d \to \infty$
   (b) As $h \to 0$ and $n \to \infty$, $nh^2 \to \infty$
   (c) As $g \to 0$ and $h \to 0$, $\frac{g^{2\nu}}{h^2} \to 0$

       i. There exists a sequence of positive integers satisfying $t(n) \to \infty$ and $t(n) = o(\sqrt{nh^2})$ such that $\sqrt{\frac{n}{h^2}}\alpha(t(n)) \to 0$
       ii. $\sqrt{\frac{\log n}{nh^2}} \to 0$ as $n \to \infty$, $h \to 0$ and $nh^2 \to \infty$.

Theorem 1 *Let y_α be in the interior of the support of $p_\alpha(\cdot)$. Then under conditions* (1) *through* (5)*, we have*

$$\sqrt{nh}[\hat{\phi}_\alpha(y_\alpha) - \phi_\alpha(y_\alpha) - B_\alpha(y_\alpha)] \xrightarrow{d} N[0, \Sigma^*_\alpha(y_\alpha)]$$

where

$$\hat{\phi}_\alpha(y_\alpha) = \begin{pmatrix} \hat{m}_\alpha(y_\alpha) + \hat{C}_m \\ \hat{v}_\alpha(y_\alpha) + \hat{C}_v \end{pmatrix}, \qquad \phi_\alpha(y_\alpha) = \begin{pmatrix} m_\alpha(y_\alpha) + C_m \\ v_\alpha(y_\alpha) + C_v \end{pmatrix}$$
$$B_\alpha(y_\alpha) = \begin{pmatrix} b^m_\alpha(y_\alpha) \\ b^v_\alpha(y_\alpha) \end{pmatrix}, \quad \Sigma^*_\alpha(y_\alpha) = \begin{pmatrix} \sigma^m_\alpha(y_\alpha) & \sigma^{mv}_\alpha(y_\alpha) \\ \sigma^{mv}_\alpha(y_\alpha) & \sigma^v_\alpha(y_\alpha) \end{pmatrix}$$

with

$$
\begin{aligned}
& b_\alpha^m(y_\alpha) = \frac{1}{2}h^2\mu_K^2 m_\alpha^{(2)}(y_\alpha) + \frac{1}{2}g^2\mu_L^2 \int \Big[p_{\underline{\alpha}}^{(2)}(z_{\underline{\alpha}}) \\
+ \quad & \frac{p_\alpha^{(2)}(y_\alpha)}{p_\alpha(y_\alpha)} p_{\underline{\alpha}}(z_{\underline{\alpha}}) - \frac{p_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} p^{(2)}(y_\alpha, z_{\underline{\alpha}})\Big] m(y_\alpha, z_{\underline{\alpha}}) dz_{\underline{\alpha}} \\
& b_\alpha^v(y_\alpha) = \frac{1}{2}h^2\mu_K^2 v_\alpha^{(2)}(y_\alpha) + \frac{1}{2}g^2\mu_L^2 \int \Big[p_{\underline{\alpha}}^{(2)}(z_{\underline{\alpha}}) \\
+ \quad & \frac{p_\alpha^{(2)}(y_\alpha)}{p_\alpha(y_\alpha)} p_{\underline{\alpha}}(z_{\underline{\alpha}}) - \frac{p_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} p^{(2)}(y_\alpha, z_{\underline{\alpha}})\Big] v(y_\alpha, z_{\underline{\alpha}}) dz_{\underline{\alpha}} \\
& \sigma_\alpha^m(y_\alpha) = \| K \|_2^2 \int \left\{ \frac{p_{\underline{\alpha}}^2(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} v(y_\alpha, z_{\underline{\alpha}}) + \left[p_{\underline{\alpha}}(z_{\underline{\alpha}}) m(y_\alpha, z_{\underline{\alpha}}) - p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha) m_\alpha(y_\alpha)\right]^2 \right\} dz_{\underline{\alpha}} \\
& \sigma_\alpha^v(y_\alpha) = \| K \|_2^2 \int \left\{ \frac{p_{\underline{\alpha}}^2(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} v^2(y_\alpha, z_{\underline{\alpha}})\kappa_4(y_\alpha, z_{\underline{\alpha}}) + \left[p_{\underline{\alpha}}(z_{\underline{\alpha}}) v(y_\alpha, z_{\underline{\alpha}}) - p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha) v_\alpha(y_\alpha)\right]^2 \right\} dz_{\underline{\alpha}} \\
& \sigma_\alpha^{mv}(y_\alpha) = \| K \|_2^2 \int \left[p_{\underline{\alpha}}(z_{\underline{\alpha}}) m(y_\alpha, z_{\underline{\alpha}}) - p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha) m_\alpha(y_\alpha)\right] \left[p_{\underline{\alpha}}(z_{\underline{\alpha}}) v(y_\alpha, z_{\underline{\alpha}}) - p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha) v_\alpha(y_\alpha)\right] \\
+ \quad & \frac{p_{\underline{\alpha}}^2(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} v^{3/2}(y_\alpha, z_{\underline{\alpha}})\kappa_3(y_\alpha, z_{\underline{\alpha}}) dz_{\underline{\alpha}}
\end{aligned}
$$

and

$$
\begin{aligned}
& \mu_K^l \equiv \int K(u) u^l du, \; l = 2, 3 \\
& \mu_L^l \equiv \int L(u) u^l du, \; l = 2, 3 \\
& \| K \|_2^2 \equiv \int K^2(u) du, \\
& \kappa_3(y_\alpha, z_{\underline{\alpha}}) \equiv E\left[\varepsilon_k^3 | (y_{t-\alpha}, \boldsymbol{y}_{t,\underline{\alpha}}) = (y_\alpha, z_{\underline{\alpha}})\right], \\
& \kappa_4(y_\alpha, z_{\underline{\alpha}}) \equiv E\left[(\varepsilon_k^2 - 1)^2 | (y_{t-\alpha}, \boldsymbol{y}_{t,\underline{\alpha}}) = (y_\alpha, z_{\underline{\alpha}})\right]
\end{aligned}
$$

In the above, $p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha)$ *is the conditional density of* $\mathbf{y}_{\underline{\alpha}}$ *given* y_α.

Theorem 2 *Let* (y_α, y_β) *be in the interior of the support of* $p_{\alpha\beta}(\cdot)$*. Then under conditions* (1) *through* (4) *and* (6)*, we have that*

$$
\sqrt{nh^2}[\hat{\phi}_{\alpha\beta}(y_\alpha, y_\beta) - \phi_{\alpha\beta}(y_\alpha, y_\beta) - B_{\alpha\beta}(y_\alpha, y_\beta)] \xrightarrow{d} N[0, \Sigma_{\alpha\beta}^*(y_\alpha, y_\beta)]
$$

where

$$
\hat{\phi}_{\alpha\beta}(y_\alpha, y_\beta) = \begin{pmatrix} \hat{m}_{\alpha\beta}(y_\alpha, y_\beta) + \hat{C}_m \\ \hat{v}_{\alpha\beta}(y_\alpha, y_\beta) + \hat{C}_v \end{pmatrix}, \qquad \phi_{\alpha\beta}(y_\alpha, y_\beta) = \begin{pmatrix} m_{\alpha\beta}(y_\alpha, y_\beta) + C_m \\ v_{\alpha\beta}(y_\alpha, y_\beta) + C_v \end{pmatrix}
$$

$$
B_{\alpha\beta}(y_\alpha, y_\beta) = \begin{pmatrix} b_{\alpha\beta}^m(y_\alpha, y_\beta) \\ b_{\alpha\beta}^v(y_\alpha, y_\beta) \end{pmatrix}, \quad \Sigma_{\alpha\beta}^*(y_\alpha, y_\beta) = \begin{pmatrix} \sigma_{\alpha\beta}^m(y_\alpha, y_\beta) & \sigma_{\alpha\beta}^{mv}(y_\alpha, y_\beta) \\ \sigma_{\alpha\beta}^{mv}(y_\alpha, y_\beta) & \sigma_{\alpha\beta}^v(y_\alpha, y_\beta) \end{pmatrix}
$$

with

$$B_{\alpha\beta}(y_\alpha, y_\beta)$$
$$= \frac{h^2}{2}\mu_K^2 \left[\frac{\partial^2 \phi_{\alpha\beta}}{\partial y_{t-\alpha}^2}(y_\alpha, y_\beta) + \frac{\partial^2 \phi_{\alpha\beta}}{\partial y_{t-\beta}^2}(y_\alpha, y_\beta)\right]$$
$$+ \frac{g^2}{2}\mu_L^2 \int \left[p^{(2)}_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) + \frac{p^{(2)}_{\alpha\beta}(y_\alpha, y_\beta)}{p_{\alpha\beta}(y_\alpha, y_\beta)} p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) - \frac{p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{p(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} p^{(2)}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})\right] \Delta_{\underline{\alpha\beta}}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) dz_{\underline{\alpha\beta}}$$
$$+ \int p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) \Delta_{\alpha\beta}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) dz_{\underline{\alpha\beta}}$$

$$\sigma^m_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Big\{ \frac{p^2_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{p(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} v(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})$$
$$+ \Big[p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) m(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) - p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta) m_{\alpha\beta}(y_\alpha, y_\beta)\Big]^2 \Big\} dz_{\underline{\alpha\beta}}$$

$$\sigma^v_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Big\{ \frac{p^2_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{p(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} v^2(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) \kappa_4(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})$$
$$+ \Big[p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) v(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) - p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta) v_{\alpha\beta}(y_\alpha, y_\beta)\Big]^2 \Big\} dz_{\underline{\alpha\beta}}$$

$$\sigma^{mv}_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Big\{ \Big[p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) m(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) - p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta) m_{\alpha\beta}(y_\alpha, y_\beta)\Big]$$
$$\Big[p_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}}) v(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) - p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta) v_{\alpha\beta}(y_\alpha, y_\beta)\Big]$$
$$+ \frac{p^2_{\underline{\alpha\beta}}(z_{\underline{\alpha\beta}})}{p(y_\alpha, y_\beta, z_{\underline{\alpha\beta}})} v^{3/2}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) \kappa_3(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) \Big\} dz_{\underline{\alpha\beta}}$$

and

$$\Delta_{\alpha\beta}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) = \begin{pmatrix} b^m_\alpha(y_\alpha) + b^m_\beta(y_\beta) + \sum_{\lambda\neq\alpha,\beta}^d b^m_\lambda(z_\lambda) \\ b^v_\alpha(y_\alpha) + b^v_\beta(y_\beta) + \sum_{\lambda\neq\alpha,\beta} b^v_\lambda(z_\lambda) \end{pmatrix}$$
$$\Delta_{\underline{\alpha\beta}}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) = \begin{pmatrix} m_{\alpha\beta}(y_\alpha, y_\beta) - b^m_\alpha(y_\alpha) - b^m_\beta(y_\beta) + \sum_{\lambda,\theta\neq\alpha,\beta;\lambda<\theta} m_{\lambda\theta}(z_\lambda, z_\theta) - \sum_{\lambda\neq\alpha,\beta} b^m_\lambda(z_\lambda) \\ v_{\alpha\beta}(y_\alpha, y_\beta) - b^v_\alpha(y_\alpha) - b^v_\beta(y_\beta) + \sum_{\lambda,\theta\neq\alpha,\beta;\lambda<\theta} v_{\lambda\theta}(z_\lambda, z_\theta) - \sum_{\lambda\neq\alpha,\beta} b^v_\lambda(z_\lambda) \end{pmatrix}$$

In the above, $p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta)$ is the conditional density of $y_{\underline{\alpha\beta}}$ given y_α and y_β. Proofs of these results are rather demanding technically are omitted here. They are available upon request.

Remark 3 *For both additive and interactive component estimators, the first term in the bias has the form standard for all local linear regression estimators. The second term in each case appears because we do not know true marginal and joint densities of the process y_t and effectively the "penalty" we pay for not knowing them. If these densities are known, this second term goes away in each case.*

Remark 4 *Note that constants C_m and C_v are estimated by $\hat{C}_m$ and $\hat{C}_v$ with the degree of precision of $O_p\left(\frac{1}{\sqrt{n}}\right)$ and, therefore, the individual additive and interactive components $m_\alpha(y_\alpha)$, $m_{\alpha\beta}(y_\alpha, y_\beta)$, $v_\alpha(y_\alpha)$ and $v_{\alpha\beta}(y_\alpha, y_\beta)$ have the same asymptotic bias and variance as $\phi_\alpha(y_\alpha)$ and $\phi_{\alpha\beta}(y_\alpha, y_\beta)$, respectively.*

5 Simulation

This chapter is dedicated to a simulation study that illustrates the finite-sample behavior of the LIVE estimators $\hat{v}_\alpha(y_\alpha)$ and $\hat{v}_{\alpha\beta}(y_\alpha, y_\beta)$. We follow in the footsteps of Kim and Linton (2004), considering examples with zero conditional mean function only for now.

Example 1 *The design in this example is additive-interactive nonlinear ARCH(3) with uniformly distributed innovations. We define the process*

$$y_t = \sqrt{4 + v_1(y_{t-1}) + v_2(y_{t-2}) + v_3(y_{t-3}) + v_{12}(y_{t-1}, y_{t-2}) + v_{13}(y_{t-1}, y_{t-3}) + v_{23}(y_{t-2}, y_{t-3})}\varepsilon_t$$

where $v_1(u) = v_2(u) = -v_3(u) = 0.5sin(u)$,
$v_{12}(u,v) = v_{13}(u,v) = v_{23}(u,v) = 0.5arctan(u)arctan(v)$, ε_t *independent of* $\mathcal{F}_{t-1}$ *and* $\varepsilon_t \sim Uniform(-\sqrt{3}, \sqrt{3})$.

The components of the volatility function are selected to satisfy conditions of Lu and Jiang (2001) to ensure the geometric ergodicity (and, therefore, the strict stationarity) of the process y_t. Condition (B1) of Lu and Jiang (2001) reduces in the one-dimensional case to the requirement that the growth rate in each coordinate must not exceed the linear one; apparently, the product of arctan functions satisfies this condition. Based on this model, we simulate 500 samples with sample size $n = 500$. For each realization of the ARCH process, we apply the instrumental variable estimation procedure from Section (3.3) to obtain estimates of $v_\alpha(\cdot)$ and $v_{\alpha\beta}(\cdot)$, $1 \leq \alpha, \beta \leq 3$. Gaussian kernels are used for all of the nonparametric estimates. We use one-dimensional Gaussian and product Gaussian kernel to estimate density functions as well as additive and interactive components of the mean and variance functions. Bandwidth g is selected according to the Gaussian rule of thumb as $g = cn^{-1/(l+4)}$ where $c = (\frac{4}{l+2})^{1/(l+4)}$ and l is the number of dimensions of the function to be estimated. For example, $l = 1$ for any one-dimensional marginal density, $l = 2$ for $\hat{p}_{\alpha\beta}(y_\alpha, y_\beta)$, $l = d$ for $\hat{p}(\mathbf{y})$, etc. (see, for example, Wand and Jones (1995)[47] for details). The constant c is selected to ensure that the bandwidth g is asymptotically optimal in the mean squared error sense under the assumption that the true density is Gaussian. The same rule is used to select the second bandwidth h where $l = 1$ or $l = 2$ for additive and

Table 1: AVERAGED MSE AND MAE FOR SIX VOLATILITY ESTIMATORS IN THE UNIFORM DISTRIBUTION CASE

	v_1	v_2	v_3	v_{12}	v_{13}	v_{23}
MSE	0.33	0.31	0.30	0.54	0.56	0.53
MAE	0.25	0.24	0.24	0.36	0.37	0.36

interactive components, respectively. To evaluate the performance of the estimators, the mean squared error (MSE) and the mean absolute deviation error (MAE) are computed for each simulated sample. They are defined as

$$MSE(v_\alpha) = \left\{ \frac{1}{101} \sum_{j=1}^{101} [v_\alpha(x_j) - \hat{v}_\alpha(x_j)]^2 \right\}^{1/2},$$

$$MSE(v_{\alpha\beta}) = \left\{ \frac{1}{2601} \sum_{j=1}^{2601} [v_{\alpha\beta}(x_{j,1}, x_{j,2}) - \hat{v}_{\alpha\beta}(x_{j,1}, x_{j,2})]^2 \right\}^{1/2},$$

$$MAE(v_\alpha) = \frac{1}{101} \sum_{j=1}^{101} |v_\alpha(x_j) - \hat{v}_\alpha(x_j)|,$$

$$MAE(v_{\alpha\beta}) = \frac{1}{2601} \sum_{j=1}^{2601} |v_{\alpha\beta}(x_{j,1}, x_{j,2}) - \hat{v}_{\alpha\beta}(x_{j,1}, x_{j,2})|,$$

$1 \leq \alpha < \beta \leq 3$. In the above, $\{x_j\}_{j=1}^{101}$ is an equispaced grid on $[-4, 4]$, and $\{x_{j,1}\}_{j=1}^{51} \times \{x_{j,1}\}_{j=1}^{51}$ is an equispaced grid on $[-4, 4] \times [-4, 4]$. The grid range covers more than 99% of all observations in both one- and two- dimensions so very little information is lost. Table (1) shows averages of MSE's and MAE's for all six components from 500 repetitions. Figure (1) shows the averaged estimates of three additive components as well as the true functions; the solid lines in red are the true curves and the dotted ones in black are the estimates averaged over 500 repetitions. Figures (2) through (5) show the true surface of the interactive components next to the estimates averaged over 500 simulations. In general, the results show a very good fit for both additive and interactive components; among interactive components, $v_{12}(\cdot)$ and $v_{23}(\cdot)$ seem to have been fit particularly well. It can be clearly seen that the use of local polynomial regression eliminated boundary effects to a great extent in both additive and interactive component estimation.

Example 2 *Again, consider the model*

$$y_t = \sqrt{4 + v_1(y_{t-1}) + v_2(y_{t-2}) + v_3(y_{t-3}) + v_{12}(y_{t-1}, y_{t-2}) + v_{13}(y_{t-1}, y_{t-3}) + v_{23}(y_{t-2}, y_{t-3})}\varepsilon_t$$

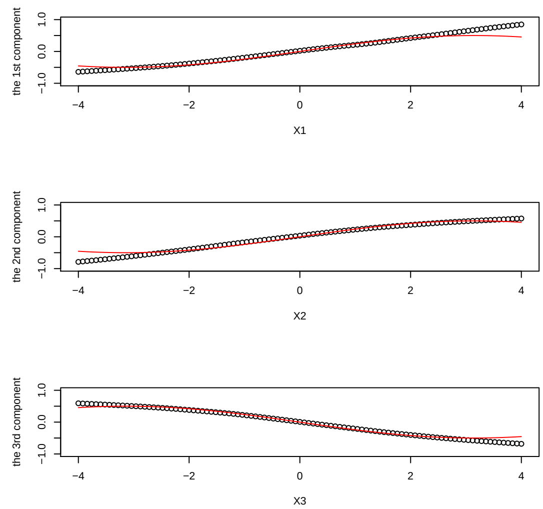

Figure 1: Estimates of three additive components in example 1

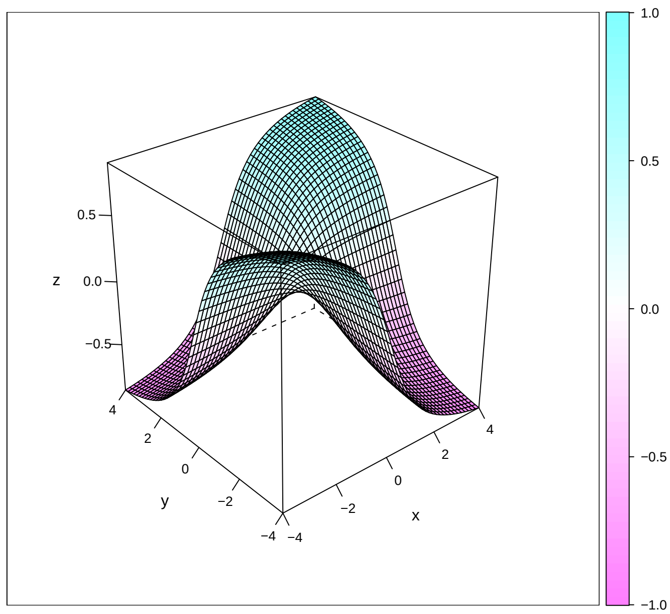

Figure 2: True surface of the interactive components in example 1

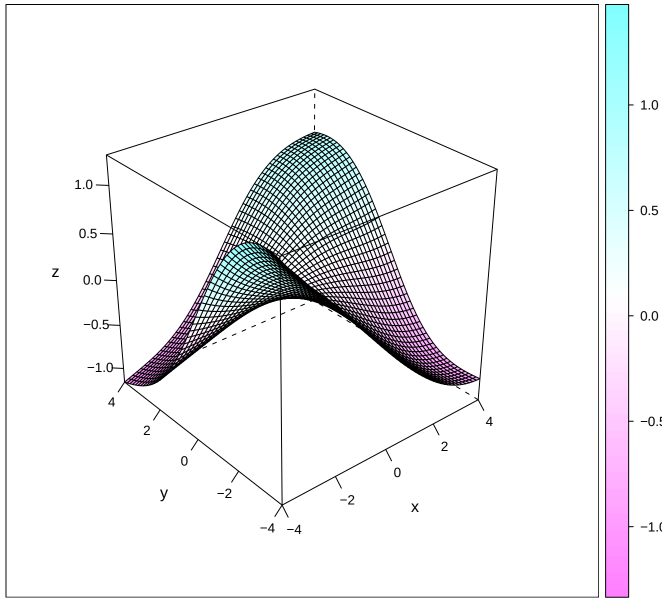

Figure 3: Estimated surface of $v_{12}(\cdot)$ in example 1

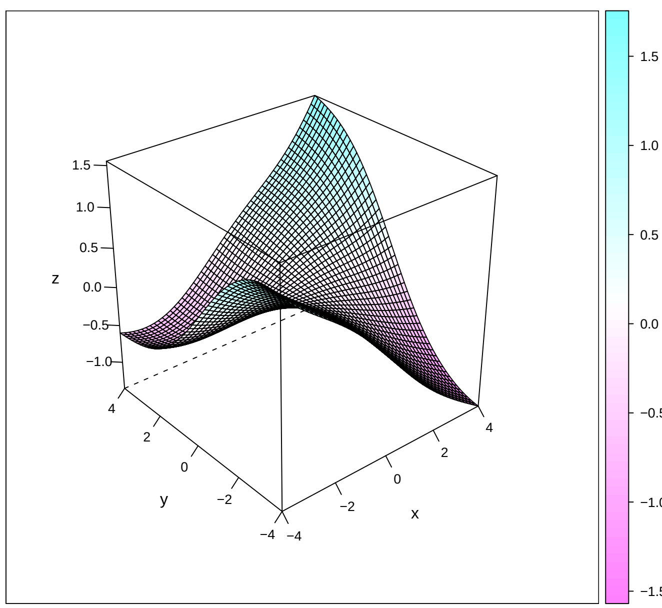

Figure 4: Estimated surface of $v_{13}(\cdot)$ in example 1

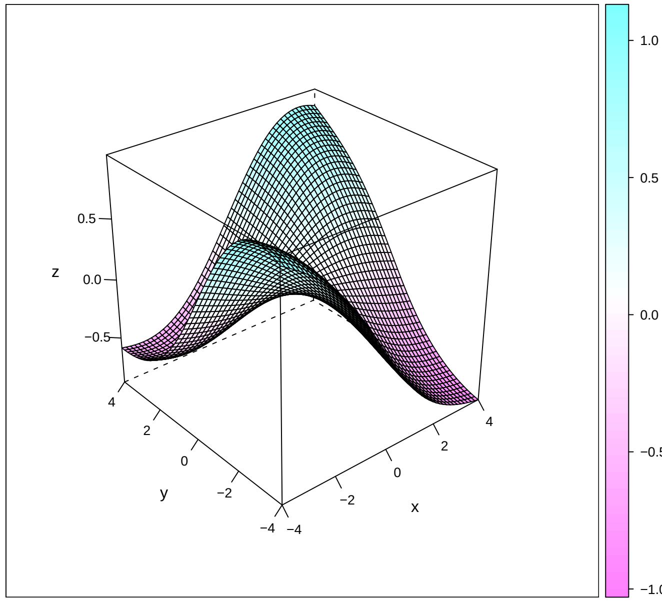

Figure 5: Estimated surface of $v_{23}(\cdot)$ in example 1

where
$v_1(u) = v_2(u) = -v_3(u) = 0.5sin(u)$
$v_{12}(u,v) = v_{13}(u,v) = v_{23}(u,v) = 0.5arctan(u)arctan(v)$
ε_t *independent of* $\mathcal{F}_{t-1}$ *and* $\varepsilon_t \sim N(0,1)$

The previous example has the finitely supported error(innovation) distribution and this may not be realistic enough in practice. Therefore, we are interested in testing the performance of the method in case where the innovation distribution does not have a compact support. Obviously, the most intuitive choice is the standard normal distribution. The grid ranges we chose are $[-3.2, 3.2]$ in one-dimensional regressions and $[-3.2, 3.2] \times [-3.2, 3.2]$ in two-dimensional ones. These ranges cover approximately 90% and 80% of all observations, respectively. The averages of MSE's and MAE's for all the six components from 500 repetitions are shown in table (2). Note that the performance of the method does not seem to be any worse compared to the previous example. While additive components seem to be estimated with slightly less precision, the opposite is true when it comes to interactive components for either choice of the loss function. The averaged estimates of six volatility components as well as the true ones are presented in

Table 2: AVERAGED MSE AND MAE FOR SIX VOLATILITY ESTIMATORS IN THE NORMAL DISTRIBUTION CASE

	v_1	v_2	v_3	v_{12}	v_{13}	v_{23}
MSE	0.40	0.40	0.39	0.41	0.40	0.42
MAE	0.32	0.32	0.32	0.31	0.30	0.31

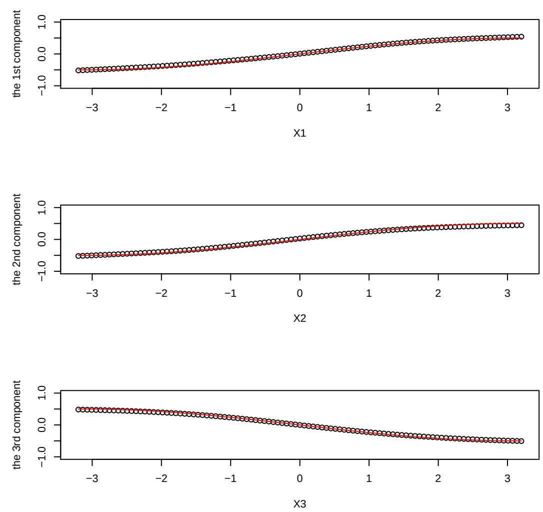

Figure 6: Estimates of three additive components in example 2

figures (6) through (10).

6 Discussion

The additive-interactive model (4) represents a further step in the development of the nonparametric volatility model theory. The article provides the instrumental variable based algorithm that can be easily used to fit such a model. The algorithm is computationally efficient and easy to implement. At the same time, central limit theorems for the estimators of the individual components are obtained and closed form expressions for asymptotic biases and variances of these estimators are given.

Among several interesting questions that remain unanswered for now in the context of the model(4) is the question of testing the statistical significance of individual additive and interactive components. This is the question of obvious practical interest. It has some

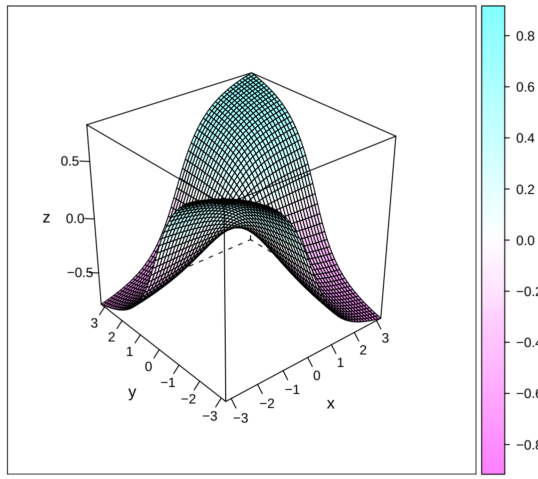

Figure 7: True surface of the interactive components in example 2

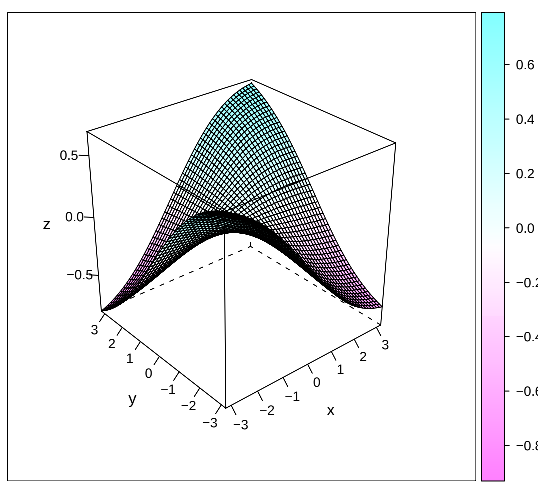

Figure 8: Estimated surface of $v_{12}(\cdot)$ in example 2

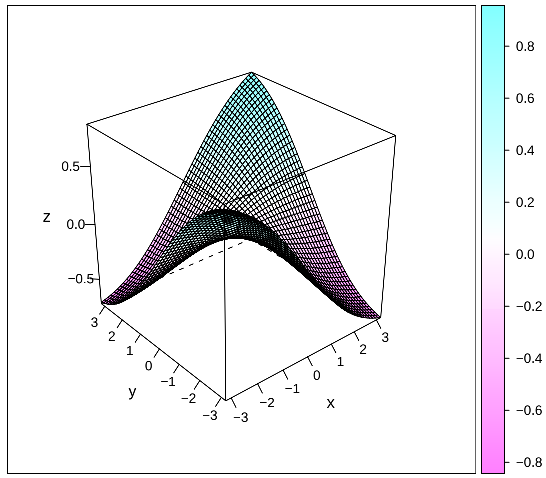

Figure 9: Estimated surface of $v_{13}(\cdot)$ in example 2

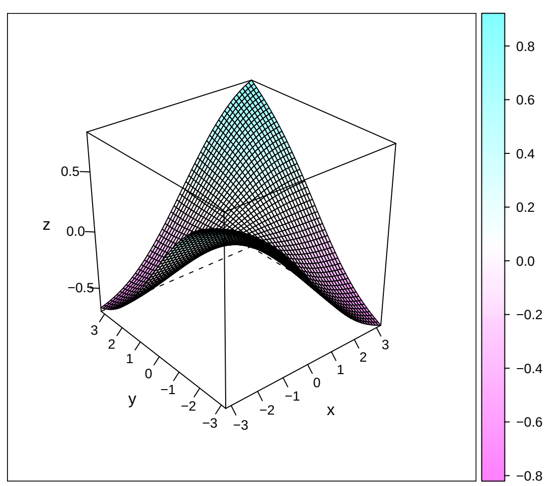

Figure 10: Estimated surface of $v_{23}(\cdot)$ in example 2

prior history in the cross-sectional context. Specifically, a test that can handle the separability hypothesis in the mean function under a specific alternative (inclusion of second order interactions) for cross-sectional data had been proposed in Sperlich, Tjostheim and Yang (2002)[42]. Consistent specification tests for nonparametric/semiparametric models proposed in Li, Hsiao and Zinn (2003)[25] are designed for null models that may include, among other possible nonparametric components, second order interactions. However, not much is known about similar testing problems in the time series context. Note that many of the modern applications are concerned with situations where the number of lags d considered can be quite large. Even in the cross-sectional context, multicollinearity among many different explanatory variables is very much a commonplace; in the time series context, it is always the case. Therefore, multiple hypotheses testing is, probably, much more important under these circumstances. For example, to test the separability hypothesis in the mean(variance) function for model (4), it is necessary to test $m_{\alpha\beta} \equiv 0$, $1 \le \alpha < \beta \le d$ ($v_{\alpha\beta} \equiv 0$, $1 \le \alpha < \beta \le d$, respectively). It may also be of interest to test the null hypothesis that includes both additive and interactive components. Thus, the design of the F-type tests here seems to be an important issue.

References

[1] Ango Nze, P. (1992): "Criteres d'ergodicite de quelques modeles a representation markovenne," C. R. Acad. Sci. Paris 315, ser 1, 1301-1304.

[2] Angrist, J.D., Imbens, G.W. & Rubin, D.B. (1996): "Identification of Causal Effects Using Instrumental Variables (with discussion)," Journal of the American Statistical Association, 91, 444-473.

[3] Auestad, B. & Tjostheim, D. (1990): "Identification of nonlinear time series: First order characterization and order estimation," Biometrika, 77, 669-687.

[4] Avramidis, P. (2002): "Local maximum likelihood estimation of volatility function," Manuscript, LSE.

[5] Bauwens, L., Laurent S. & Rombouts J. (2006): "Multivariate GARCH models: a survey," Journal of Applied Econometrics, 21, 2006, 79-109.

[6] Black, F. (1976): "Studies in stock price volatility changes," Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association, 177-181.

[7] Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," Journal of Econometrics, 31, 307-327.

[8] Bollerslev, T., Chou, R. & Kroner, K. (1992): "ARCH modeling in finance," Journal of Econometrics, 52, 5-59.

[9] Breiman, L. & Friedman, J.H. (1985): "Estimating optimal transformations for multiple regression and correlation (with discussion)," Journal of the American Statistical Association, 80, 580-619.

[10] Buja, A., Hastie, T. & Tibshirani, R. (1989): "Linear smoothers and additive models (with discussion)," Annals of Statistics, 17, 453-555.

[11] Giraitis, L., Leipus, R., and Surgailis, D. (2007) "ARCH(∞) models and long memory properties",in:Richard Davis and Jens-Peter Kreis and Thomas Mikosch (editors). Springer.

[12] Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," Econometrica, Vol. 50, No. 4 (Jul., 1982), 987-1007

[13] Fan, J. and Gijbels, I. (1996): Local Polynomial Modelling and Its Applications. Chapman & Hall, London.

[14] Fan, J. & Yao, Q. (1998): "Efficient estimation of conditional variance functions in stochastic regression," Biometrika, 85, 645-660.

[15] Fan, J. and Yao, Q. (2003): Nonlinear time series: nonparametric and parametric methods. Springer.

[16] Glosten, L.R., Jagannathan, R. & Runkle, D.E. (1993): "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," The Journal of Finance, Vol. 48, No. 5. (Dec., 1993), 1779-1801.

[17] Härdle, W. & Vieu, P. (1992): "Kernel Regression Smoothing of Time Series," Journal of Time Series Analysis, 13, 209-232.

[18] Härdle, W., Tsybakov, A.B. & Yang, L. (1996): "Nonparametric vector autoregression," Journal of Statistical Planning and Inference, 68, 221-245.

[19] Härdle, W. & Tsybakov, A.B. (1997): "Locally polynomial estimators of the volatility function," Journal of Econometrics, 81, 223-242.

[20] Hastie, T. & Tibshirani, R. (1987): "Generalized additive models: Some applications," Journal of the American Statistical Association, 82, 371-386.

[21] Hastie, T. & Tibshirani, R. (1990): Generalized Additive Models, Chapman & Hall, London.

[22] Horowitz, J. (2001): "Nonparametric estimation of a generalized additive model with an unknown link function," Econometrica, 69, 499-513.

[23] Horowitz, J. & Mammen, E. (2006): "Nonparametric estimation of a generalized additive model with an unknown link function," preprint.

[24] Kim, W. & Linton, O.B. (2004): "The live method for generalized additive volatility models," Econometric theory, 20(6), 1094-1139.

[25] Li, Q., Hsiao, C., and Zinn, J. (2003):"Consistent specification test for semiparametric/nonparametric models based on series estimation methods", Journal of Econometrics, 112, 295-325

[26] Linton, O.B. & Nielsen, J. (1995): "A kernel method of estimating structured nonparametric regression based on marginal integration," Biometrika, 82, 93-100.

[27] Linton, O.B. (1997): "Efficient estimation of additive nonparametric regression models," Biometrika, 84, 469-474.

[28] Lu Z. & Jiang, Z. (2001): "L1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term," Statistics & Probability Letters, 51, 121-130.

[29] Mammen, E., Linton, O.B. & Nielsen, J. (1999): "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions," Annals of Statistics, 27, 1443-1490.

[30] Masry, E. & Tjostheim, D. (1995): "Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality," Econometric Theory, 11, 258-289.

[31] Masry, E. (November 1996): "Multivariate local polynomial regression for time series: Uniform strong consistency and rates," Journal of Time Series Analysis, 17, 571-599.

[32] Masry, E. (December 1996): "Multivariate regression estimation: Local polynomial fitting for time series," Stochastic Processes and Their Applics., vol. 65, 81-101.

[33] Nelson, D.B. (1991): "Conditional heteroskedasticity in asset returns: A new approach," Econometrica, 59, 347-370.

[34] Newey, W.K. (1994): "Kernel estimation of partial means," Econometric Theory, 10, 233-253.

[35] Opsomer, J.D. & Ruppert, D. (1997): "Fitting a bivariate additive model by local polynomial regression," Annals of Statistics, 25, 186-211.

[36] Pagan, A.R. & Schwert, G.W. (1990): "Alternative models for conditional stock volatility," Journal of Econometrics, 45, 267-290.

[37] Pagan, A.R. & Hong, Y.S. (1991): "Nonparametric Estimation and the Risk Premium," in W. Barnett, J. Powell, and G.E. Tauchen (eds.) Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge University Press.

[38] Rio, Emmanuel (1993): "Covariance inequalities for strongly mixing processes," Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques, 29, no. 4, 587-597.

[39] Robinson, P. (1983): "Non-parametric estimation for time series models," Journal of Time Series Analysis, 4, 185-208.

[40] Shephard, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility," Time Series Models in Econometrics, Finance and other Fields, ed. by D. Cox, D. Hinkley, and O. Barndorff-Nielsen, 1-55, London. Chapman & Hall.

[41] Silverman, B.W. (1986): Density Estimation for Statistics and Data Analysis, Chapman & Hall.

[42] Sperlich, S., Tjostheim, D. & Yang L. (2002): "Nonparametric Estimation and Testing of Interaction in Additive Models," Econometric Theory, 18, 197-251.

[43] Stone, C.J. (1985): "Additive regression and other nonparametric models," Annals of Statistics, 13, 685-705.

[44] Stone, C.J. (1986): "The dimensionality reduction principle for generalized additive models," Annals of Statistics, 14, 592-606.

[45] Stone, C.J. (1994): “The use of polynomial splines and their tensor products in multivariate function estimation (with discussion),” The Annals of Statistics, 22, 118-184.

[46] Tjostheim, D. & Auestad, B. (1994): “Nonparametric identification of nonlinear time series: Projections,” Journal of the American Statistical Association, 89, 1398-1409.

[47] Wand, M.P. and Jones, M.C. (1995): Kernel Smoothing. Chapman & Hall, London.

[48] Yang, L., Härdle W. & Nielsen, J. (1999): “Nonparametric autoregression with multiplicative volatility and additive mean,” Journal of Time Series Analysis, 20, 579-604.

[49] Zakoian, J. (1994): “Threshold heteroskedastic functions,” Journal of Economic Dynamics and Control, 18, 931-955.

[50] Ziegelmann, F. (2002): “Nonparametric estimation of volatility functions: The local exponential estimator,” Econometric Theory, 18, 985-991.