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

Michael Levine *and Jinguang (Tony) Li †

**Abstract**

In this article we consider a new separable nonparametric volatility model that includes second-order interaction terms in both mean and conditional variance functions. This is a very flexible non-parametric ARCH model that can potentially explain the behavior of the wide variety of financial assets. The model is estimated using the generalized version of the Local Instrumental Variable Estimation method (LIVE) first introduced in Kim and Linton (2004). 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.

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

## 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. Given the importance of predicting volatility in a number of asset-pricing and portfolio management problems, this is hardly surprising. The most popular class of volatility models historically has been the class of

*Department of Statistics, Purdue University 250 N. University St. West Lafayette, IN 47907-2066 e-mail mlevins@purdue.edu

†Canadian Pension Fund (CPP)

ARCH/GARCH models, originally introduced by Engle (1982) and generalized to GARCH by Bollerslev (1986).

The parametrization employed in the classical ARCH and GARCH models cannot capture some of the salient features of financial data such as, for example, the leverage effect first documented in Black (1976). In the parametric setting, one possible way to capture leverage effect is to use models with cross-product terms in the conditional variance function. One of the most important models of that kind that generalizes the standard ARCH (GARCH) paradigm is the QARCH (Quadratic ARCH) model of Sentana (1995) that considers the conditional variance function to be a quadratic polynomial of the past $q$ values of the process $Y_t$.

That model encompasses a number of earlier considered volatility models proposed in the literature, such as the classical ARCH model of Engle (1982), Augmented ARCH(AARCH) of Bera and Lee (1990) and others. Another example of an earlier model that has interaction terms in the conditional variance and is capable of capturing the leverage effect is the Nonlinear Asymmetric GARCH (NGARCH) model of Engle and Ng (1993).

There are a few reasons why parametric models are sometimes inadequate. First, there is often very little reason to choose one parametric specification over the other; as a result, parametric models are often subject to misspecification. Nonparametric specification, on the contrary, imposes only some basic smoothness constraints on the conditional variance function. In practice, in order to make the model tractable, additional assumptions (such as additivity) often need to be imposed; however, even in this case, it is a much more general approach to modeling than the parametric one. Second, many features present in the data (such as nonnormality, asymmetric cycles, nonlinearity between lagged variables) require nonlinear models to describe the law that generates the data. However, the number of such models is infinitely large and time series analysts cannot explore all of them. Therefore, a more general nonparametric setting can be explored as an alternative.

The general nonparametric volatility model assumes that the conditional variance function depends on a given number of past lags of the process $Y_t$; in other words, the conditional variance function $v \equiv v(Y_{t-1}, \ldots, Y_{t-d})$ for some integer $d > 0$. In general, the unstructured version of this model suffers from the well-known "curse of dimensionality" problem. This means that the best possible rate of convergence for an estimator of $v$ quickly decreases as the number of dimensions $d$ increases. Such a model, under assumption of zero mean, was considered in Pagan and Schwert (1990) and Pagan and Hong (1991). Historically, the first way to get around the curse of dimensionality has been to assume the generalized additive structure of the function to be estimated. Thus, a simple generalized additive model for the conditional variance function is $v(Y_{t-1}, \ldots, Y_{t-d}) = v_1(Y_{t-1}^2) + \ldots + v(Y_{t-p}^2)$; such a model has been discussed in, for example, Fan and Yao (2003). This model is, clearly, a generalization of ARCH($p$) model. In many situations, however, generalized additive structure is not sufficiently flexible and other alternatives have to be considered.

Yang, Härdle and Nielsen (1999) 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. A rather general nonparametric volatility model was considered in Kim and Linton (2004) who called it GANARCH (generalized additive nonlinear ARCH). That model defines both conditional mean and conditional variance as $m(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = F_m(C_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha}))$ and $v(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = F_v(C_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha}))$ 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. Note also that the generalized additive model in both mean and variance is also a special case of the GANARCH model when both link functions are identically equal to 1. The GANARCH model assumes that the link function is known and, therefore, some information about the distribution of data is available. Such information is often very hard to obtain, particularly so in multidimensional settings.

We are interested in studying a model that avoids the assumption of the known distribution of the data that is needed for the GANARCH model of Kim and Linton (2004) does. Because of that, we do not consider link functions for mean and variance. We would also like to avoid the fairly rigid assumption of additivity of the mean and variance function; therefore, we take a step ahead and introduce a model that contains nonparametric "interactions" in the volatility function as well as in the conditional mean function. The resulting model is indirectly related to the model of Kim and Linton (2004) and can be represented in the nonlinear ARCH form as

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

$$m(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = C_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha})$$

$$+ \sum_{1 \leq \alpha < \beta \leq d} m_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta}) \tag{2}$$

$$v(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = C_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha})$$

$$+ \sum_{1 \leq \alpha < \beta \leq d} v_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta}) \tag{3}$$

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. Such a model can be viewed as a generalization of the QARCH model of Sentana (1995)and, indeed, nests the QARCH model of Sentana (1995) directly with the appropriate choice of functions $v_\alpha$ and $v_{\alpha\beta}$. As is the case with QARCH, this model allows us to consider the possibility of two lags of the process $y_t$ with the same sign influencing the volatility more than the regular ARCH model, or even a generalized additive model for conditional variance, would allow. Due to the nonparametric specification of the conditional variance function, this influence can take on many different functional forms and is, thus, more flexible than QARCH. As is the case with QARCH model, the model (1)-(3) also allows for a dynamic asymmetric effect of positive and negative values of $y_t$ on the conditional variance; thus, it is capable of capturing the leverage effect that may have an arbitrary functional form. In this way it is also more flexible than the QARCH model.

In terms of estimation note that, according to Stone (1994), interactive components in such a model can be estimated at the same rate as a regular two-dimensional non-parametric smoothing problem; more specifically, the interactive terms can be estimated at the optimal rate $O(n^{-q/(2q+2)})$ whenever the function to be estimated is q-smooth in the sense of Stone (1994). In other words, the interactive effect estimation is as hard a problem as a two-dimensional nonparametric smoothing. Note also that the expansion of functions $m$ and $v$ used in (1) can be interpreted as a functional ANOVA representation for general functions $m(y_{t-1}, \ldots, y_{t-d})$ and $v(y_{t-1}, \ldots, y_{t-d})$.

The specific contribution of this paper consists in proposing an easy to implement and computationally effective estimation method for the model (1) that represents a non-trivial extension of the instrumental variable method of Kim and Linton (2004). Moreover, we prove asymptotic normality of the resulting estimators under a fairly mild set of regularity conditions and provide explicit expressions of asymptotic mean and variance for all of the functional components' estimators.

The rest of the paper is organized as follows. In Section 2 we make some comments on stationarity and strong mixing property of our model. Section 3 describes the main estimation idea and defines the local instrumental variable (LIVE) estimators. In Section 4 we describe our main theoretical results and then give the proof of the first one in Section 5. The last Section contains discussion of the results. The Appendix with simulation results is attached.

## 2 Stationarity, Strong Mixing and Identifiability Properties of the Model (1)-(3)

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) used $\alpha$-mixing or geometric ergodicity to identify the nonlinear time series model. Ango Nze (1992) studied the $L_1$ geometric ergodicity of the multivariate model $X_t = f(X_{t-1}, \ldots, X_{t-p}) + H(X_{t-1}, \ldots, X_{t-q})\epsilon_t$ where $X_t$ and $\epsilon_t$ are two sequences of $m$-dimensional random variables defined on a common probability space and $\epsilon_t$ is an $m$-dimensional white noise process. Ango Nze (1992) gave, probably, the first in the literature sufficient condition that ensures $L_1$ geometric ergodicity of such a model. Lu and Jiang (2001) derived another sufficient condition that also ensures $L_1$ geometric ergodicity of the same model but that is much less restrictive. In this paper we impose constraints from Lu and Jiang (2001) and, in doing so, assume that the conditions for strict stationarity and strong mixing property of the process $\{y_t\}_{t=1}^n$ in (1)-(3) are met.

The model (1)-(3) is not identifiable unless specific conditions are met. These conditions are similar to those of Kim and Linton (2004); they can be summarized as

$$
\begin{aligned}
E[m_\alpha(Y_{t-\alpha})] &= 0\,, \alpha = 1, \cdots, d && (4)\\
E[v_\alpha(Y_{t-\alpha})] &= 0\,, \alpha = 1, \cdots, d && (5)
\end{aligned}
$$

and

$$
\begin{aligned}
&E[m_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\alpha} = y_\alpha] = E[m_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\beta} = y_\beta] = 0 && (6)\\
&E[v_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\alpha} = y_\alpha] = E[v_{\alpha\beta}(Y_{t-\alpha}, Y_{t-\beta})|Y_{t-\beta} = y_\beta] = 0 && (7)
\end{aligned}
$$

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

# 3 The Local Instrumental Variable Estimators

## 3.1 Estimation Methods

The generalized additive model 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); see also Hastie and Tibshirani (1987, 1990) as well as Buja, Hastie and Tibshirani (1989). It can be extended to our model but it is rather hard to analyze theoretically. Another possibility is to extend the marginal integration method that was introduced independently by Newey (1994), Tjostheim and Auestad (1994), and Linton and Nielsen (1995). However, marginal integration is relatively computationally expensive even in case of generalized additive model and this shortcoming is only going to become more dramatic if it is used to fit our proposed model.

In this paper, we suggest an alternative approach that is a generalization of the LIVE idea suggested first in Kim and Linton (2004). There, instrumental variables were used to estimate each additive component of the mean and/or variance function separately. We extend this approach by suggesting an extra set of instruments that can be used to estimate the interactive components of both functions as well. 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 $n$, compared to the similar (and related) marginal integration method; 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. 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 LIVE algorithm for the additive-interactive model (1)-(3)

We begin with introducing the notation that will be used repeatedly throughout the paper. We denote $\mathbf{y}_t = (y_{t-1}, \dots, y_{t-d})$, $\mathbf{y} = (y_1, \dots, y_d)$. We use underscore to imply that a particular direction $\alpha$ or directions $\alpha$ and $\beta$ have been omitted; boldface is used for all multidimensional quantities. Thus, $\mathbf{y}_{t,\underline{\alpha}} = (y_{t-1}, \dots, y_{t-\alpha+1}, y_{t-\alpha-1}, \dots, y_{t-d})$ and $\mathbf{y}_{\underline{\alpha}} = (y_1, \dots, y_{\alpha-1}, y_{\alpha+1}, \dots, y_d)$. Analogously, $\mathbf{y}_{t,\alpha\beta} = (y_{t-\alpha}, y_{t-\beta})$, $\mathbf{y}_{\alpha\beta} = (y_\alpha, y_\beta)$ while $\mathbf{y}_{t,\underline{\alpha\beta}}$ and $\mathbf{y}_{\underline{\alpha\beta}}$ are defined analogously to $\mathbf{y}_{t,\underline{\alpha}}$ and $\mathbf{y}_{\underline{\alpha}}$. 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. **Preliminary density estimation**

   The instrumental variables we are going to define depend on several marginal and joint densities. Thus, these density functions need to be estimated first. As we mentioned before, we use regular product kernel density estimators. Specifically, we estimate the marginal density $p_\alpha(\cdot)$ as

$$\hat{p}_\alpha(y_\alpha) = \frac{1}{ng_1} \sum_{t=1}^{n} L\left(\frac{y_{t-\alpha} - y_\alpha}{g_1}\right), \alpha = 1, 2, \cdots, d$$

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

$$\begin{aligned}
\hat{p}_{\alpha\beta}(y_\alpha, y_\beta) &= \frac{1}{ng_\alpha g_\beta} \sum_{t=1}^{n} L\left(\frac{y_{t-\alpha} - y_\alpha}{g_\alpha}\right) L\left(\frac{y_{t-\beta} - y_\beta}{g_\beta}\right), 1 \le \alpha < \beta \le d \\
\hat{p}_{\underline{\alpha\beta}}(y_{\underline{\alpha}}, y_{\underline{\beta}}) &= \frac{1}{n \prod\limits_{\substack{\lambda=1 \\ \lambda \notin \{\alpha,\beta\}}}^{d} g_\lambda} \sum_{t=1}^{n} \prod_{\substack{\lambda=1 \\ \lambda \notin \{\alpha,\beta\}}}^{d} L\left(\frac{y_{t-\lambda} - y_\lambda}{g_\lambda}\right), 1 \le \alpha < \beta \le d
\end{aligned}$$

$\hat{p}_{\underline{\alpha}}(y_{\underline{\alpha}}) = \frac{1}{n \prod\limits_{\substack{\lambda=1 \\ \lambda \neq \alpha}} g_\lambda} \sum_{t=1}^{n} \prod\limits_{\substack{\lambda=1 \\ \lambda \neq \alpha}}^{d} L\left(\frac{y_{t-\lambda} - y_\lambda}{g_\lambda}\right), \alpha = 1, 2, \cdots, d$ and

$\hat{p}(\mathbf{y}) = \frac{1}{n \prod\limits_{\alpha=1}^{d} g_\alpha} \sum_{t=1}^{n} \prod\limits_{\alpha=1}^{d} L\left(\frac{y_{t-\alpha} - y_\alpha}{g_\alpha}\right)$. In the above, $g_i = g_i(n)$ are the bandwidths, $i = 1, \ldots, d$ and $L(\cdot)$ is the unimodal one-dimensional symmetric kernel function.

**Remark 1** 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)). 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

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

The conditional expectation above is estimated using the local linear regression; in other words, $\hat{m}_\alpha(y_\alpha)$ in the above is the minimizer $a_\alpha$ of the kernel-weighted least squares problem $\min_{a_\alpha, b_\alpha} \sum_{t=d+1}^{d+n} K_h(y_{t-\alpha} - y_\alpha)\{\hat{W}_\alpha(\mathbf{y}_t)\tilde{y}_t - a_\alpha - b_\alpha(y_{t-\alpha} - y_\alpha)\}^2$. Here and further we use $h$ to denote the bandwidth used to estimate functional components of the conditional mean and variance functions. Also, from this point on we denote the kernel function $K(\cdot)$ to suggest that it is not necessarily the same as the kernel $L(\cdot)$ used to estimate marginal and joint densities earlier.

4. <u>**Estimation of the interactive components of the mean** $m_{\alpha\beta}(\cdot)$</u>

   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. <u>**Estimation of the constant component of the variance** $C_v$</u>

   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. <u>**Estimation of the additive components of the variance** $v_\alpha(\cdot)$</u>

   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. <u>**Estimation of the interactive components of the variance**$v_{\alpha\beta}(\cdot)$</u>

   Analogously to estimation of the interactive components of the mean, 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 $\sigma$-algebra generated by $\{y_t\}_{t=a}^{t=b}$ and $\alpha(k)$ the strong mixing coefficient of $\{y_t\}$ defined as $\alpha(k) = sup_{A\in\mathcal{F}_{-\infty}^0, B\in\mathcal{F}_k^\infty}|P(A\cap B) - P(A)P(B)|$.

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 $m_\alpha(\cdot), m_{\alpha\beta}(\cdot), v_\alpha(\cdot), v_{\alpha\beta}(\cdot)$, $1 \le \alpha, \beta \le d$, are continuous and twice differentiable with bounded (partial) derivatives on the compact support. Also, 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 have bounded third derivatives. All of the density functions above are also bounded away from zero on the compact support.

3. The functions $L(\cdot)$ and $K(\cdot)$ are bounded, nonnegative, symmetric around zero, compactly supported, Lipschitz continuous first-order kernels. 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 $\| \underline{u} \| \to \infty$.

4. As $g \to 0$, $h \to 0$ and $n \to \infty$ we have $ng^d \to \infty$, $nh^s \to \infty$ and $\frac{g^{2\nu}}{h^s} \to 0$. Also, there exists a sequence of positive integers satisfying $t(n) \to \infty$ and $t(n) = o(\sqrt{nh^s})$ such that $\sqrt{\frac{n}{h^s}}\alpha(t(n)) \to 0$. Finally, $\sqrt{\frac{\log n}{nh^s}} \to 0$ as $n \to \infty$, $h \to 0$ and $nh^s \to \infty$, where $s \in \{1, 2\}$.

Condition (1) implies that strong mixing coefficients decay at the polynomial rate. This is milder than the usual assumption for a strongly mixing process where the rate is exponential. This assumption is similar to the assumption A1 in Kim and Linton (2004). Conditions (2)-(3) are standard in kernel and local polynomial estimation. Set of conditions (4) on the bandwidths is needed to show asymptotic negligibility of the stochastic error terms resulting from the preliminary estimation of marginal densities $p_\alpha(\cdot)$, $p_{\alpha\beta}(\cdot,\cdot)$ and the functions $m(\cdot)$ and $v(\cdot)$. It is also needed to take into account the effect of dependence of the mixing process on the asymptotic results. Before presenting the main result, some extra notation has to be introduced. Let us denote $\mu_K^l \equiv \int K(u)u^l du$ and $\mu_L^l \equiv \int L(u)u^l du$ for $l = 2, 3$. Also, $\| K \|_2^2 \equiv \int K^2(u)du$, $\kappa_3(y_\alpha, z_{\underline{\alpha}}) \equiv E\left[\varepsilon_k^3 | (y_{t-\alpha}, \mathbf{y}_{t,\underline{\alpha}}) = (y_\alpha, z_{\underline{\alpha}})\right]$, and $\kappa_4(y_\alpha, z_{\underline{\alpha}}) \equiv E\left[(\varepsilon_k^2 - 1)^2 | (y_{t-\alpha}, \mathbf{y}_{t,\underline{\alpha}}) = (y_\alpha, z_{\underline{\alpha}})\right]$. Let $p_{\underline{\alpha}|\alpha}(z_{\underline{\alpha}}|y_\alpha)$ be the conditional density of $\mathbf{y}_{\underline{\alpha}}$ given $y_\alpha$. Define $\hat{\phi}_\alpha(y_\alpha) = \begin{pmatrix} \hat{m}_\alpha(y_\alpha) + \hat{C}_m \\ \hat{v}_\alpha(y_\alpha) + \hat{C}_v \end{pmatrix}$, $\phi_\alpha(y_\alpha) = \begin{pmatrix} m_\alpha(y_\alpha) + C_m \\ v_\alpha(y_\alpha) + C_v \end{pmatrix}$, $B_\alpha(y_\alpha) = \begin{pmatrix} b_\alpha^m(y_\alpha) \\ b_\alpha^v(y_\alpha) \end{pmatrix}$, and $\Sigma_\alpha^*(y_\alpha) = \begin{pmatrix} \sigma_\alpha^m(y_\alpha) & \sigma_\alpha^{mv}(y_\alpha) \\ \sigma_\alpha^{mv}(y_\alpha) & \sigma_\alpha^v(y_\alpha) \end{pmatrix}$ where

$$
\begin{aligned}
b_\alpha^m(y_\alpha) &= \frac{1}{2}h^2\mu_K^2 m_\alpha^{(2)}(y_\alpha) + \frac{1}{2}g^2\mu_L^2 \int \Big[p_{\underline{\alpha}}^{(2)}(z_{\underline{\alpha}}) \\
&+ \frac{p_\alpha^{(2)}(y_\alpha)}{p_\alpha(y_\alpha)} p_{\underline{\alpha}}(z_{\underline{\alpha}}) - \frac{p_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} p^{(2)}(y_\alpha, z_{\underline{\alpha}})\Big] m(y_\alpha, z_{\underline{\alpha}}) dz_{\underline{\alpha}}
\end{aligned}
$$

and $b_\alpha^v(y_\alpha)$ is the same as above with $v$ substituted for $m$. Let

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

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

and

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

**Theorem 1** *Let $y_\alpha$ be in the interior of the support of $p_\alpha(\cdot)$. Then under conditions* (1) *through* (4) *with $s = 1$, we have*

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

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

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

$$
\begin{aligned}
&\sigma^m_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Bigg\{ \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}}) \\
&+ \left[ 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) \right]^2 \Bigg\} dz_{\underline{\alpha\beta}} \\
&\sigma^v_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Bigg\{ \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}}) \\
&+ \left[ 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) \right]^2 \Bigg\} dz_{\underline{\alpha\beta}} \\
&\sigma^{mv}_{\alpha\beta}(y_\alpha, y_\beta) = \| K \|_2^2 \int \Bigg\{ \left[ 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) \right] \\
&\left[ 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) \right] \\
&\qquad\qquad + \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}}) \Bigg\} dz_{\underline{\alpha\beta}}
\end{aligned}
$$

and

$$
\Delta_{\alpha\beta}(y_\alpha, y_\beta, z_{\underline{\alpha\beta}}) = \begin{pmatrix} b^m_\alpha(y_\alpha) + b^m_\beta(y_\beta) + \sum^d_{\lambda \neq \alpha,\beta} b^m_\lambda(z_\lambda) \\ b^v_\alpha(y_\alpha) + b^v_\beta(y_\beta) + \sum_{\lambda \neq \alpha,\beta} b^v_\lambda(z_\lambda) \end{pmatrix}
$$

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

Define $p_{\underline{\alpha\beta}|\alpha,\beta}(z_{\underline{\alpha\beta}}|y_\alpha, y_\beta)$ as the conditional density of $y_{\underline{\alpha\beta}}$ given $y_\alpha$ and $y_\beta$.

**Theorem 2** *Let $(y_\alpha, y_\beta)$ be in the interior of the support of $p_{\alpha\beta}(\cdot)$. Then under conditions*

(1) *through* (4) *with* $s = 2$, *we have that*

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

**Remark 2** 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.

**Remark 3** The asymptotic bias expressions $b^m_\alpha(y_\alpha)$ and $b^v_\alpha(y_\alpha)$ consist of two main terms. The first term is of the order $O(h^2)$ which is the same as in the ordinary local linear regression. This would have been the order of the asymptotic bias if the densities used in constructing instrumental variables were known. The second term in both cases has the order of $O(g^2)$ and is the penalty we pay for not knowing these densities.

**Remark 4** The method that is most relevant for comparison purposes here is probably the traditional version of the marginal integration method. As described in Linton and Nielsen (1995), in the two-dimensional case it requires the practitioner to choose two different bandwidths $h_1$ and $h_2$ for the two directions. The resulting bias is then of the order $O(h_1^2+h_2^2)$. This is the same situation one encounters in LIVE algorithm since, for an additive component that is twice continuously differentiable and is defined on the compact support, the resulting bias is then $O(h^2+g^2)$ which is conceptually the same as in standard marginal integration. As we noted already earlier, the main benefit of the LIVE method is its speed of computation; whereas marginal integration requires $O(n^3)$ smoothings to estimate a linear component the LIVE method only takes $O(n^2)$ smoothings.

**Remark 5** It is also fairly clear that there is some room for improvement when it comes to asymptotic variances of LIVE estimators. When compared to the regular marginal integration estimators of Linton and Nielsen (1995), one can see that, for example, the asymptotic variance of the additive mean component for LIVE estimator is larger than the one for the MI estimator. More precisely, if one assumes constant conditional variance function $v(\cdot)$, the first term of $\sigma^m_\alpha(y_\alpha)$, which is $\| K \|_2^2 \int \left\{ \frac{p^2_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} v(y_\alpha, z_{\underline{\alpha}}) \right\} dz_{\underline{\alpha}}$, coincides with the asymptotic variance of the MI estimator. This is related to the fact that the LIVE estimator is based on the omitted variable regression idea that treats everything except the additive component of interest as the error term. Such an error term is not conditionally independent of the component of interest. Finally, as in Kim and Linton (2004), a number of possible efficiency standards can be considered in order to construct possible oracle estimators. This is one of the topics we intend to consider as part of the future research.

**Remark 6** Note that one of the conditions of Theorems 1 and 2 is that $\frac{g^{2\nu}}{h} \to 0$ as $g \to 0$ and $h \to 0$. Recall that $\nu > 2$ is the strong mixing index that characterizes the strength of dependence of the process $y_t$. In practice, this means that the bandwidth $g$ can be selected much smaller than $h$. In other words, undersmoothing of the densities that comprise instrumental variables is possible. If $g$ is thus selected to be of the smaller order of magnitude then $h$, the penalty for not knowing true marginal and joint densities of the process $y_t$ becomes of the order $o(h^2)$ and thus can be disregarded asymptotically.

# 5 Proof of Theorem 1

The proof of Theorem 1 contains four steps: first, we decompose the estimation error into the main stochastic term and bias; second, we approximate each component, then we compute the asymptotic bias, and, finally, we establish the asymptotic normality. Without loss of generality we assume that $\alpha = 1$.

For expositional convenience we use the subscript 2 to denote the nuisance directions. For example, $p_2(\mathbf{y}_{t,\underline{1}}) = p_{\underline{1}}(\mathbf{y}_{t,\underline{1}})$ in the case of density function, and $m_2(\mathbf{y}_{t,\underline{1}}) = m_{\underline{1}}(\mathbf{y}_{t,\underline{1}})$, $v_2(\mathbf{y}_{t,\underline{1}}) = v_{\underline{1}}(\mathbf{y}_{t,\underline{1}})$ in the case of component functions. We use $X_n \simeq Y_n$ to signify $X_n = Y_n\{1 + o_p(1)\}$. Let $vec(X)$ denote the vector of matrix $X$ arranged by columns.

*Step I. Decompositions*

The LIVE estimator of $\phi_1(y_1)$, as any local linear regression estimator, can be expressed in the closed form:

$$\hat{\phi}_1(y_1)^T = e_1^T(\mathbf{Y}_-^T\mathbf{K}\mathbf{Y}_-)^{-1}\mathbf{Y}_-^T\mathbf{K}\mathbf{W}\tilde{\mathbf{R}} \tag{8}$$

where

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{Y}_{\text{-}} = \begin{pmatrix} 1 & y_{d+1-1} - y_1 \\ 1 & y_{d+2-1} - y_1 \\ \vdots & \vdots \\ 1 & y_{d+n-1} - y_1 \end{pmatrix}, \quad \tilde{\mathbf{R}} = \begin{pmatrix} \tilde{r}_1 \\ \tilde{r}_2 \\ \vdots \\ \tilde{r}_n \end{pmatrix} = \begin{pmatrix} y_{d+1} & \tilde{\varepsilon}^2_{d+1} \\ y_{d+2} & \tilde{\varepsilon}^2_{d+2} \\ \vdots & \vdots \\ y_{d+n} & \tilde{\varepsilon}^2_{d+n} \end{pmatrix},$$

$\mathbf{K} = diag\{K_h(y_{d+l-1} - y_1)\}$, $l = 1, \ldots, n$, $\mathbf{W} = diag\{\hat{W}_{d+k}\}$, $k = 1, \ldots, n$ and $\tilde{\varepsilon}^2_t = (y_t - \tilde{m}(\mathbf{y}_t))^2$. The preliminary density estimators used are $\hat{p}_1(y_1) = \frac{1}{ng}\sum_{t=d+1}^{d+n} L_g(y_{t-1} - y_1)$, $\hat{p}_2(\mathbf{y}_{\underline{1}}) = \frac{1}{ng^{d-1}}\sum_{t=d+1}^{d+n}\prod_{\alpha=2}^{d} L_g(y_{t-\alpha} - y_\alpha)$ and $\hat{p}(\mathbf{y}) = \frac{1}{ng^d}\sum_{t=d+1}^{d+n}\prod_{\alpha=1}^{d} L_g(y_{t-\alpha} - y_\alpha)$ while $\hat{W}_t = \frac{\hat{p}_1(y_{t-1})\hat{p}_2(\mathbf{y}_{t,\underline{1}})}{\hat{p}(\mathbf{y}_t)}$.

Using standard properties of the vec operator, it is easy to verify that (8) can be written as $\hat{\phi}_1(y_1) = [I_2 \otimes e_1^T(\mathbf{Y}_{\text{-}}^T\mathbf{K}\mathbf{Y}_{\text{-}})^{-1}][I_2 \otimes \mathbf{Y}_{\text{-}}^T\mathbf{K}]vec[\mathbf{W}\tilde{\mathbf{R}}]$ while the true function $\phi_1(y_1)$ satisfies the identity

$$\phi_1(y_1) = [I_2 \otimes e_1^T(\mathbf{Y}_{\text{-}}^T\mathbf{K}\mathbf{Y}_{\text{-}})^{-1}][I_2 \otimes \mathbf{Y}_{\text{-}}^T\mathbf{K}]vec[l\phi_1^T(y_1) + Y_{\text{-}} \bigtriangledown \phi_1^T(y_1)] \tag{9}$$

where

$$l = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad Y_{\text{-}} = \begin{pmatrix} y_{d+1-1} - y_1 \\ y_{d+2-1} - y_1 \\ \vdots \\ y_{d+n-1} - y_1 \end{pmatrix}$$

It follows easily from the above results that the estimation error can be represented as

$$\hat{\phi}_1(y_1) - \phi_1(y_1) = [I_2 \otimes e_1^T Q_n^{-1}]\tau_n \tag{10}$$

where $Q_n = \frac{1}{n}D_h^{-1}\mathbf{Y}_{\text{-}}^T\mathbf{K}\mathbf{Y}_{\text{-}}D_h^{-1}$ and $\tau_n = \frac{1}{n}[I_2 \otimes D_h^{-1}\mathbf{Y}_{\text{-}}^T\mathbf{K}]vec[\mathbf{W}\tilde{\mathbf{R}} - l\phi_1^T(y_1) - Y_{\text{-}}\bigtriangledown\phi_1^T(y_1)]$ with $D_h = diag\{1,\ h\}$

*Step II. Approximations*

We rewrite $\tau_n$ in (5) as

$$\tau_n = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1} - y_1)[\hat{W}_k\tilde{r}_k - \phi_1(y_1) - (y_{k-1} - y_1) \bigtriangledown \phi_1(y_1)] \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T\right\}$$

and expand it next by adding and subtracting $\phi_1(y_{k-1})$ within the inner factor enclosed in brackets as

$$\begin{aligned}
\tau_n &= \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)[\hat{W}_k\tilde{r}_k-\phi_1(y_1)-(y_{k-1}-y_1)\nabla\phi_1(y_1)]\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\} \quad (11)\\
&= \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)[\hat{W}_k\tilde{r}_k-\phi_1(y_{k-1})]\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}\\
&+ \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)[\phi_1(y_{k-1})-\phi_1(y_1)-(y_{k-1}-y_1)\nabla\phi_1(y_1)]\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}
\end{aligned}$$

At this point, we introduce the notation $r_k=[m(\mathbf{y}_k),\ v(\mathbf{y}_k)]^T$ and $\tilde{r}_k^*=\begin{pmatrix} v^{1/2}(\mathbf{y}_k)\varepsilon_k \\ v(\mathbf{y}_k)(\varepsilon_k^2-1)\end{pmatrix}$ that will prove useful in deriving approximations of the two terms in the last formula. By plugging in the estimated instruments and bounding the difference between the true and estimated values of instruments $W_k,\ k=1,\dots,n$, we find in a way similar to Kim and Linton (2004) that the bias can be viewed as consisting of the four components

$$\begin{aligned}
T_{1n} &= \frac{h^2}{2n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^2 D^2\phi_1(y_1)\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}\\
T_{2n} &= \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left[\frac{\hat{p}_1(y_{k-1})\hat{p}_2(\mathbf{y}_{k,\underline{1}})}{\hat{p}(\mathbf{y}_k)}-\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\tilde{r}_k\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}\\
T_{3n} &= \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\frac{p_1(y_{k-1})p_2\left(\mathbf{y}_{k,\underline{1}}\right)}{p(\mathbf{y}_k)}\tilde{r}_k^*\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}\\
T_{4n} &= \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left[\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}r_k-\phi_1(y_{k-1})\right]\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\},
\end{aligned}$$

*Step III. Asymptotic Bias*

As a reminder, we assume that the sufficient conditions of Lu and Jiang (2001) are satisfied, and, therefore, the process $y_t$ is geometrically ergodic. This implies strict stationarity and $\beta$-mixing (therefore, also $\alpha$-mixing).

Since measurable function of strictly stationary and $\alpha$-mixing process is again strictly stationary and $\alpha$-mixing, we can apply the ergodic theorem for $\alpha$-mixing stationary processes to $T_{1n}$ (see e.g. Fan and Yao (2003) for details). As a result, using the Taylor expansion of the density function $p_1(\cdot)$, we obtain:

$$
\begin{aligned}
T_{1n} &= \frac{h^2}{2n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^2 D^2\phi_1(y_1)\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}\\
&\xrightarrow{p} \frac{h^2}{2}E\left[K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^2 D^2\phi_1(y_1)\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right]\\
&= \frac{h^2}{2}p_1(y_1)D^2\phi_1(y_1)\otimes(\mu_K^2,\ \mu_K^3)^T + o(h^2)
\end{aligned}
$$

Next, we define marginal expectations of estimated density functions $\hat{p}_1(\cdot)$, $\hat{p}_2(\cdot)$ and $\hat{p}(\cdot)$ as $\bar{p}_1(y_{k-1}) = \int L_g(z_1 - y_{k-1})p_1(z_1)dz_1$, $\bar{p}_2(\mathbf{y}_{k,\underline{1}}) = \int L_g(z_2 - \mathbf{y}_{k,\underline{1}})p_2(z_2)dz_2$ and $\bar{p}(\mathbf{y}_k) = \int L_g(z_1 - y_{k-1})L_g(z_2 - \mathbf{y}_{k,\underline{1}})p(z_1,z_2)dz_1dz_2$;

it can be shown analogously to Kim and Linton (2004) that $T_{2n}$ can be approximated as

$$
\begin{aligned}
T_{2n} &= \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left[\frac{\bar{p}_1(y_{k-1})\bar{p}_2(\mathbf{y}_{k,\underline{1}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\tilde{r}_k\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right\}\\
&+ o_p(1/\sqrt{nh}).
\end{aligned}
$$

Applying the ergodic theorem for $\alpha$-mixing stationary process and using the fact that $E\left[K_h(y_{k-1}-y_1)[\frac{\bar{p}_1(y_{k-1})\bar{p}_2(\mathbf{y}_{k,\underline{1}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}]\,[\tilde{m}(\mathbf{y}_k)-m(\mathbf{y}_k)]^2\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right] = o(1/\sqrt{nh})$, we have

$$
\begin{aligned}
T_{2n} &= \frac{1}{n}\sum_{k=d+1}^{d+n}\{K_h(y_{k-1}-y_1)[\frac{\bar{p}_1(y_{k-1})\bar{p}_2(\mathbf{y}_{k,\underline{1}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}]\tilde{r}_k\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\} + o_p(1/\sqrt{nh})\\
&\xrightarrow{p} E\left[K_h(y_{k-1}-y_1)\left[\frac{\bar{p}_1(y_{k-1})\bar{p}_2(\mathbf{y}_{k,\underline{1}})}{\bar{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\tilde{r}_k\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right]\\
&\simeq \frac{g^2}{2}p_1(y_1)\int\left[p_2^{(2)}(z_2) + \frac{p_1^{(2)}(y_1)}{p_1(y_1)}p_2(z_2) - \frac{p_2(z_2)}{p(y_1,z_2)}p^{(2)}(y_1,z_2)\right]\phi(y_1,z_2)dz_2\otimes(\mu_L^2,0)^T
\end{aligned}
$$

As a next step, we note that the terms $T_{3n}$ and $T_{4n}$ consist of are $\mathcal{F}_{k-1}$-measurable for any $k$, and, therefore, we can argue that, by ergodic theorem for $\alpha$-mixing stationary process, both $T_{3n}$ and $T_{4n}$ converge in probability to zero: $T_{mn}\xrightarrow{p} 0$, for $m=3$ and $m=4$.

Finally, for the probability limit of $[I_2\otimes e_1^T Q_n^{-1}]$, we have $Q_n = \frac{1}{n}D_h^{-1}\mathbf{Y}_-^T\mathbf{K}\mathbf{Y}_-D_h^{-1} = \begin{pmatrix} q_{n0} & q_{n1}\\ q_{n1} & q_{n2}\end{pmatrix}$ where $q_{ni} = \frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^i$, $i = 0,1,2$. Using the

Taylor expansion of $p_1(y_1+uh)$, it is easy to obtain $q_{ni} \xrightarrow{p} q_i$ where $q_0 = p_1(y_1)\mu_K^0 = p_1(y_1)$, $q_1 = p_1(y_1)\mu_K^1 = 0$, $q_2 = p_1(y_1)\mu_K^2$.

Clearly, $e_1^T Q_n^{-1} \xrightarrow{p} \frac{1}{p_1(y_1)} e_1^T$. By Slutzky theorem, we conclude that the bias of the estimator $\hat{\phi}_1(y_1)$ is

$$
\begin{aligned}
&B_{1n}(y_1) = [I_2 \otimes e_1^T Q_n^{-1}](T_{1n} + T_{2n} + T_{3n} + T_{4n}) \\
&\xrightarrow{p} \begin{pmatrix} \frac{1}{2}h^2\mu_K^2 m_1^{(2)}(y_1) + \frac{1}{2}g^2\mu_L^2 \int \left[ p_2^{(2)}(z_2) + \frac{p_1^{(2)}(y_1)}{p_1(y_1)} p_2(z_2) - \frac{p_2(z_2)}{p(y_1,z_2)} p^{(2)}(y_1,z_2) \right] m(y_1,z_2)dz_2 \\ \frac{1}{2}h^2\mu_K^2 v_1^{(2)}(y_1) + \frac{1}{2}g^2\mu_L^2 \int \left[ p_2^{(2)}(z_2) + \frac{p_1^{(2)}(y_1)}{p_1(y_1)} p_2(z_2) - \frac{p_2(z_2)}{p(y_1,z_2)} p^{(2)}(y_1,z_2) \right] v(y_1,z_2)dz_2 \end{pmatrix} \\
&\equiv B_1(y_1)
\end{aligned}
$$

*Step IV. Asymptotic Normality*

As a final step, we derive the asymptotic distribution of $\sqrt{nh}[I_2 \otimes e_1^T Q_n^{-1}](T_{1n} + T_{2n} + T_{3n} + T_{4n})$.

First, we rewrite $T_{1n}$ as

$$
\begin{aligned}
T_{1n} &= \frac{h^2}{2n} \sum_{k=d+1}^{d+n} \{ K_h(y_{k-1} - y_1) \left( \frac{y_{k-1} - y_1}{h} \right)^2 D^2\phi_1(y_1) \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T \} \\
&\equiv \frac{h^2}{2} (r_{1n},\ r_{2n},\ r_{3n},\ r_{4n})^T
\end{aligned}
$$

where $r_{qn} = \frac{1}{n}\sum_{k=d+1}^{d+n} \{K_h(y_{k-1} - y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^{q+1} m_1^{(2)}(y_1)\}$, $q = 1, 2$ and $r_{qn} = \frac{1}{n}\sum_{k=d+1}^{d+n} \{K_h(y_{k-1} - y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^{q-1} v_1^{(2)}(y_1)\}$, $q = 3, 4$. Applying Theorem 1 in Masry (November 1996) directly, we can easily find that each of the $nh \cdot var(r_{in})$, $i = 1, \dots, 4$ is uniformly bounded: $\sup_{y_1 \in R} nh \cdot var(r_{1n}) \le sup_{y_1 \in R} m_1^{(2)}(y_1) \cdot C_2\nu^4(1 + o(1))$, $\sup_{y_1 \in R} nh \cdot var(r_{2n}) \le sup_{y_1 \in R} m_1^{(2)}(y_1) \cdot C_2\nu^6(1 + o(1))$, $\sup_{y_1 \in R} nh \cdot var(r_{3n}) \le sup_{y_1 \in R} v_1^{(2)}(y_1) \cdot C_2\nu^4(1+o(1))$ and $sup_{y_1 \in R} nh \cdot var(r_{4n}) \le sup_{y_1 \in R} v_1^{(2)}(y_1) \cdot C_2\nu^6(1+o(1))$ where $\nu_i \equiv \int K^2(u)u^i du,\ \ i = 1, \cdots, 6$ and $C_2$ is the upper bound of the density function $p_1(\cdot)$; recall that functions $m_1$ and $v_1(\cdot)$ have bounded second derivatives by Condition 2 and observe that, as a result, we have $var(\sqrt{nh}T_{1n}) = O(h^4)$ uniformly over $y_1 \in R$. Note that the orders of the covariances are $O(h^4)$ due to Cauchy-Schwartz Inequality, i.e., $cov(\sqrt{nh}r_{in},\ \sqrt{nh}r_{jn}) \le \sqrt{var(\sqrt{nh}r_{in}) \cdot var(\sqrt{nh}r_{jn})}$, for $1 \le i < j \le 4$. It can be similarly shown that $var(\sqrt{nh}T_{2n})$ is asymptotically negligible as well. To see this, rewrite $T_{2n}$ as

$$T_{2n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\left[\frac{\hat{p}_1(y_{k-1})\hat{p}_2(\mathbf{y}_{k,\underline{1}})}{\hat{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]\tilde{r}_k \otimes \left(1, \frac{y_{k-1}-y_1}{h}\right)^T\right\}$$
$$\equiv (x_{1n},\ x_{2n},\ x_{3n},\ x_{4n})^T$$

where

$$\begin{aligned}
x_{1n} &= \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K\left(\frac{y_{k-1}-y_1}{h}\right)\left[\frac{\hat{p}_1(y_{k-1})\hat{p}_2(\mathbf{y}_{k,\underline{1}})}{\hat{p}(\mathbf{y}_k)} - \frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\right]y_k\right\} \\
&= \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\xi_{1k} \\
&\equiv \frac{1}{h}\frac{1}{n}\xi_1,
\end{aligned}$$

$x_{3n} = \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\xi_{3k} \equiv \frac{1}{h}\frac{1}{n}\xi_3$ with $\xi_{3k}$ being the same as $\xi_1$ except that $\tilde{\epsilon}_t^2$ is substituted for $y_t$, and $x_{2n} = \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\xi_{2k} \equiv \frac{1}{h}\frac{1}{n}\xi_2$, with $\xi_{2k} = \xi_{1k}\left(\frac{y_{k-1}-y_1}{h}\right)$, $x_{4n} = \frac{1}{h}\frac{1}{n}\sum_{k=d+1}^{d+n}\xi_{4k} \equiv \frac{1}{h}\frac{1}{n}\xi_4$ with $\xi_{4k} = \xi_{3k}\left(\frac{y_{k-1}-y_1}{h}\right)$ Applying Davydov's lemma (see Theorem 1.0 in Rio (1993)), we obtain $\frac{1}{n}var(\xi_1) \le 2CM_{\nu,1}\sum_{i\in[0,n+d-1]}\alpha(i)^{1-2/\nu}$ where $M_{\nu,1} = E[|\xi_{1k}|^\nu] = O(g^{2\nu})$ and $C$ is a nonnegative constant. Thus, we find that $var(\sqrt{nh}x_{1n}) = O\left(\frac{g^{2\nu}}{h}\right) \to 0$, as $n \to \infty$; it can be shown analogously that $var(\sqrt{nh}x_{in}) \to 0$, as $n \to \infty$, for $i = 2, 3, 4$. By Cauchy-Schwartz inequality, $cov(\sqrt{nh}x_{in},\ \sqrt{nh}x_{jn}) \to 0$ as $n \to \infty$, for $1 \le i < j \le 4$ as well.

Third, we rewrite $T_{3n}$ as

$$T_{3n} = \frac{1}{n}\sum_{k=d+1}^{d+n}\left\{K_h(y_{k-1}-y_1)\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}\tilde{r}_k^* \otimes \left(1, \frac{y_{k-1}-y_1}{h}\right)^T\right\}$$
$$\equiv (s_{0n},\ s_{1n},\ t_{0n},\ t_{1n})^T$$

where

$$\begin{aligned}
s_{ln} &= \frac{1}{n}\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^l\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}v^{1/2}(\mathbf{y}_k)\varepsilon_k \\
t_{mn} &= \frac{1}{n}\sum_{k=d+1}^{d+n}K_h\left(y_{k-1}-y_1\right)\left(\frac{y_{k-1}-y_1}{h}\right)^m\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}v(\mathbf{y}_k)(\varepsilon_k^2-1),
\end{aligned}$$

$l = 0, 1$ and $m = 0, 1$. We first examine the term $s_{0n}$. It's easy to see that $E\left[s_{0n}|\mathcal{F}_{n+d}\right] = 0$ and

$$\begin{aligned} var\left[s_{0n}|\mathcal{F}_{n+d}\right] &= E\left[s_{0n}^2|\mathcal{F}_{n+d}\right] \\ &= \frac{1}{nh}\left\{\frac{1}{n}\sum_{k=d+1}^{d+n}\frac{1}{h}K^2\left(\frac{y_{k-1}-y_1}{h}\right)\frac{p_1^2(y_{k-1})p_2^2(\mathbf{y}_{k,\underline{1}})}{p^2(\mathbf{y}_k)}v(\mathbf{y}_k)\right\} \end{aligned}$$

Applying the ergodic theorem for $\alpha$-mixing processes and changing variables, we find that

$$\begin{aligned} var\left[\sqrt{nh}s_{0n}\right] &= E\left[nh\cdot var\left[s_{0n}|\mathcal{F}_{n+d}\right]\right] \\ &= E\left[\frac{1}{h}K^2\left(\frac{y_{k-1}-y_1}{h}\right)\frac{p_1^2(y_{k-1})p_2^2(\mathbf{y}_{k,\underline{1}})}{p^2(\mathbf{y}_k)}v(\mathbf{y}_k)\right] \\ &= \parallel K \parallel_2^2 p_1^2(y_1)\int\frac{p_2^2(z_2)}{p(y_1,z_2)}v(y_1,z_2)dz_2[1+o(1)] \end{aligned}$$

Applying the central limit theorem (Theorem 2.22) in Fan and Yao (2003), we have that

$$\sqrt{nh}s_{0n}\xrightarrow{d}N[0,\ \sigma_{0,s}^m(y_1)]$$

where

$$\sigma_{0,s}^m(y_1) = \parallel K \parallel_2^2 p_1^2(y_1)\int\frac{p_2^2(z_2)}{p(y_1,z_2)}v(y_1,z_2)dz_2 \tag{12}$$

It is equally easy to conclude that $E\left[t_{0n}|\mathcal{F}_{n+d}\right] = 0$ and

$$\begin{aligned} var\left[t_{0n}|\mathcal{F}_{n+d}\right] &= E\left[t_{0n}^2|\mathcal{F}_{n+d}\right] \\ &= \frac{1}{n^2}E\left[\left(\sum_{k=d+1}^{d+n}K_h(y_{k-1}-y_1)\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)}v(\mathbf{y}_k)(\varepsilon_k^2-1)\right)^2|\mathcal{F}_{n+d}\right] \\ &= \frac{1}{n^2}\sum_{k=d+1}^{d+n}K_h^2(y_{k-1}-y_1)\frac{p_1^2(y_{k-1})p_2^2(\mathbf{y}_{k,\underline{1}})}{p^2(\mathbf{y}_k)}v^2(\mathbf{y}_k)\kappa_4(\mathbf{y}_k) \end{aligned}$$

In much the same way as before for $s_{0n}$, applying the ergodic theorem for $\alpha$-mixing process, changing variables and using the central limit theorem in Fan and Yao (2003), we have that

$$\sqrt{nh}t_{0n}\xrightarrow{d}N[0,\ \sigma_{0,s}^v(y_1)] \tag{13}$$

$$\text{where}\quad \sigma_{0,s}^v(y_1) = \parallel K \parallel_2^2 p_1^2(y_1)\int\frac{p_2^2(z_2)}{p(y_1,z_2)}v^2(y_1,z_2)\kappa_4(y_1,z_2)dz_2 \tag{14}$$

Note that $cov(\sqrt{nh}s_{0n}, \sqrt{nh}t_{0n}) = E(nhs_{0n}t_{0n})$ since $E(s_{0n}) = E(t_{0n}) = 0$. Using the same approach as before, we can easily show that

$$\begin{aligned}
&cov(\sqrt{nh}s_{0n}, \sqrt{nh}t_{0n}) = E\left[E\left[nhs_{0n}t_{0n}|\mathcal{F}_{n+d}\right]\right] \\
&=\| K \|_2^2 \, p_1^2(y_1) \int \frac{p_2^2(z_2)}{p(y_1, z_2)} v^{3/2}(y_1, z_2)\kappa_3(y_1, z_2)dz_2[1 + o(1)]
\end{aligned}$$

Combining the results for $s_{0n}$ and $t_{0n}$, we have the asymptotic normality for $T_{3n}$ :

$$\sqrt{nh}T_{3n} \xrightarrow{d} N(0, \Sigma_{1,T3}) \tag{15}$$

where

$$\Sigma_{1,T3} = \begin{pmatrix} \sigma_{0,s}^m(y_1) & * & \sigma_{0,s}^{mv}(y_1) & * \\ * & * & * & * \\ \sigma_{0,s}^{mv}(y_1) & * & \sigma_{0,s}^v(y_1) & * \\ * & * & * & * \end{pmatrix}, \tag{16}$$

$\sigma_{0,s}^m(y_1)$ and $\sigma_{0,s}^v(y_1)$ defined as in (12) and (14), while

$$\sigma_{1,s}^{mv}(y_1) \equiv \| K \|_2^2 \, p_1^2(y_1) \int \frac{p_2^2(z_2)}{p(y_1, z_2)} v^{3/2}(y_1, z_2)\kappa_3(y_1, z_2)dz_2 \tag{17}$$

The rest of the elements of the matrix $\Sigma_{1,T3}$ are left blank since they are not needed to formulate the final result.

Next, we rewrite $T_{4n}$ as

$$\begin{aligned}
T_{4n} &= \frac{1}{n}\sum_{k=d+1}^{d+n} \left\{ K_h(y_{k-1} - y_1)\left[\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)} r_k - \phi_1(y_{k-1})\right] \otimes \left(1, \frac{y_{k-1} - y_1}{h}\right)^T \right\} \\
&\equiv (u_{0n},\ u_{1n},\ w_{0n},\ w_{1n})^T
\end{aligned}$$

where

$$\begin{aligned}
u_{ln} &= \frac{1}{n}\sum_{k=d+1}^{d+n} K_h(y_{k-1} - y_1)\left(\frac{y_{k-1} - y_1}{h}\right)^l \left[\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)} m(\mathbf{y}_k) - m_1(y_{k-1})\right] \\
w_{mn} &= \frac{1}{n}\sum_{k=d+1}^{d+n} K_h(y_{k-1} - y_1)\left(\frac{y_{k-1} - y_1}{h}\right)^m \left[\frac{p_1(y_{k-1})p_2(\mathbf{y}_{k,\underline{1}})}{p(\mathbf{y}_k)} v(\mathbf{y}_k) - v_1(y_{k-1})\right],
\end{aligned}$$

for $l = 0, 1$ and $m = 0, 1$. Corollary 2 in Masry (December 1996) provides us with the tool we need to establish the asymptotic normality for $T_{4n}$ :

$$\sqrt{nh}T_{4n} \xrightarrow{d} N(0, \Sigma_{1,T4}) \tag{18}$$

where

$$\Sigma_{1,T4} = \begin{pmatrix} \sigma^{m}_{0,u}(y_1) & * & \sigma^{mv}_{0,uw}(y_1) & * \\ * & * & * & * \\ \sigma^{mv}_{0,uw}(y_1) & * & \sigma^{v}_{0,w}(y_1) & * \\ * & * & * & * \end{pmatrix} \tag{19}$$

with

$$\begin{aligned} \sigma^{m}_{0,u}(y_1) &= \| K \|_2^2 \, p_1^2(y_1) \int \left[ p_2(z_2) m(y_1, z_2) - p_{2|1}(z_2|y_1) m_1(y_1) \right]^2 dz_2, \\ \sigma^{v}_{0,w}(y_1) &= \| K \|_2^2 \, p_1^2(y_1) \int \left[ p_2(z_2) v(y_1, z_2) - p_{2|1}(z_2|y_1) v_1(y_1) \right]^2 dz_2, \\ \sigma^{mv}_{0,uw}(y_1) &= \| K \|_2^2 \, p_1^2(y_1) \int \left[ p_2(z_2) m(y_1, z_2) - p_{2|1}(z_2|y_1) m_1(y_1) \right] \left[ p_2(z_2) v(y_1, z_2) - p_{2|1}(z_2|y_1) v_1(y_1) \right] dz_2 \end{aligned}$$

Finally, it is easy to notice that $cov(T_{3n}, T_{4n}) = E[T_{3n}T_{4n}] = E[E[T_{3n}T_{4n}|\mathcal{F}_{n+d}]] = 0$. Now, the asymptotic normality of an additive component is almost at hand. $e_1^T Q_n^{-1} \xrightarrow{p} \frac{1}{p_1(y_1)} e_1^T$ implies that $I_2 \otimes e_1^T Q_n^{-1} \xrightarrow{p} I_2 \otimes \frac{1}{p_1(y_1)} e_1^T$. Applying the Slutzky theorem, we find that

$$\sqrt{nh}[I_2 \otimes e_1^T Q_n^{-1}](T_{1n} + T_{2n} + T_{3n} + T_{4n}) \xrightarrow{d} N(0, \Sigma_1^*(y_1))$$

and

$$\sqrt{nh}[\hat{\phi}_1(y_1) - \phi_1(y_1) - B_1(y_1)] \xrightarrow{d} N[0, \Sigma_1^*(y_1)]$$

where

$$B_1(y_1) = \begin{pmatrix} b_1^m(y_1) \\ b_1^v(y_1) \end{pmatrix}, \quad \Sigma_1^*(y_1) = \begin{pmatrix} \sigma_1^m(y_1) & \sigma_1^{mv}(y_1) \\ \sigma_1^{mv}(y_1) & \sigma_1^v(y_1) \end{pmatrix}$$

with the components as in the statement of Theorem 1 for $\alpha = 1$.

# 6 Discussion

The additive-interactive model (1) 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 (1) 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 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). Consistent specification tests for nonparametric/semiparametric models proposed in Li, Hsiao and Zinn (2003) 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 (1), 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


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

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

Bera, A.K., and Lee, S. (1990) On the Formulation of a General Structure for Conditional Heteroskedasticity (mimeo, University of Illinois)

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.

Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

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.

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

Engle, R. F. (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, **50(4)**, 987-1007

Engle, R.F. and Ng, V. K. (1993) Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, **48**, no.5, 1749-1778

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

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

Hastie, T. & Tibshirani, R. (1990) *Generalized Additive Models*, Chapman & Hall, London.

Kim, W. & Linton, O.B. (2004) The live method for generalized additive volatility models. *Econometric Theory*, **20(6)**, 1094-1139.

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

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

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

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.

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

Masry, E. (December 1996) Multivariate regression estimation: Local polynomial fitting for time series *Stochastic Processes and Their Applications*, **65**, 81-101.

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

Newey, W.K. (1994) Kernel estimation of partial means. *Econometric Theory*, **10**, 233-253.

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

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.

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

Sentana, E. (1995) Quadratic ARCH Models *Review of Economic Studies*, **62**, 639-661.

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

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.

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

Wand, M.P. and Jones, M.C. (1995) *Kernel Smoothing*. Chapman & Hall, London.