Kim, Woocheol; Linton, Oliver

Working Paper

A local instrumental estimation method for generalized additive volatility models

SFB 373 Discussion Paper, No. 2000,86

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Kim, Woocheol; Linton, Oliver (2000) : A local instrumental estimation method for generalized additive volatility models, SFB 373 Discussion Paper, No. 2000,86, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10048128

This Version is available at:
https://hdl.handle.net/10419/62247

A LOCAL INSTRUMENTAL VARIABLE ESTIMATION METHOD FOR GENERALIZED ADDITIVE VOLATILITY MODELS

Woocheol Kim
Institute for Statistics and Econometrics, Humboldt University of Berlin, Spandauer Str. 1, 10178, Berlin, Germany

Oliver Linton
Department of Economics, The London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom

Abstract

We investigate a new separable nonparametric model for time series, which includes many ARCH models and AR models already discussed in the literature. We also propose a new estimation procedure based on a localization of the econometric method of instrumental variables. Our method has considerable computational advantages over the competing marginal integration or projection method.

This paper is based on Chapter 2 of the first author's PhD dissertation from Yale University. We would like to thank Wolfgang Härdle, Joel Horowitz, Peter Phillips, and Dag Tjøstheim for helpful discussions. We also acknowledge support by the Deutsche Forschungsgemeinschaft via SFB 373 at Humboldt University, Berlin.

1. Introduction

Stochastic volatility models are of considerable interest in empirical finance. There are many types of parametric volatility model following the seminal work of Engle (1982); these models are typically nonlinear, which poses difficulties both in computation and in deriving useful tools for statistical inference. Parametric models are prone to misspecification, especially when there is no theoretical reason to prefer one specification over another. Nonparametric models can provide greater flexibility. However, the greater generality of these models comes at a cost - including a large number of lags requires estimation of a high dimensional smooth, which is known to behave very badly [Silverman (1986)]. The curse of dimensionality puts severe limits on the dynamic flexibility of nonparametric models. Separable models offer an intermediate position between the complete generality of nonparametric models, and the restrictiveness of parametric ones. These models have been investigated in cross-sectional settings as well as time series ones.

In this paper, we investigate a *generalized additive nonlinear* ARCH model (GANARCH);

$$y_t = m\left(y_{t-1}, y_{t-2}, \ldots, y_{t-d}\right) + v^{1/2}\left(y_{t-1}, y_{t-2}, \ldots, y_{t-d}\right)\varepsilon_t, \tag{1.1.1}$$

$$m\left(y_{t-1}, y_{t-2}, \ldots, y_{t-d}\right) = F_m\left(c_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha})\right), \tag{1.1.2}$$

$$v\left(y_{t-1}, y_{t-2}, \ldots, y_{t-d}\right) = F_v\left(c_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha})\right), \tag{1.1.3}$$

where $m_\alpha(\cdot)$ and $v_\alpha(\cdot)$ are any smooth but unknown function, while $F_m(\cdot)$ and $F_v(\cdot)$ are known monotone transformations [whose inverses are $G_m(\cdot)$ and $G_v(\cdot)$, respectively].* The error process, $\{\varepsilon_t\}$, is assumed to be a martingale difference with unit scale, i.e., $E(\varepsilon_t|\mathcal{F}_{\sqcup-\infty}) = 1$ and $E(\varepsilon_t^2|\mathcal{F}_{\sqcup-\infty}) = \infty$, where $\mathcal{F}_\sqcup$ is the σ-algebra of events generated by $\{y_k\}_{k=-\infty}^{t}$. Under some weak assumptions, the time series of nonlinear autoregressive models can be shown to be stationary and strongly mixing with mixing coefficients decaying exponentially fast. Auestadt and Tjøstheim (1990) used α-mixing or geometric ergodicity to identify the nonlinear time series model. Similar results were obtained for the additive nonlinear ARCH process by Masry and Tjøstheim (1997), see also Cai and Masry (2000). We follow the same argument as Masry and Tjøstheim (1997), and will assume all the necessary conditions for stationarity and mixing property of the process $\{y_t\}_{t=1}^{n}$ in (1.1.1). The standard identification for the components of the mean and variance is made by

$$E\left[m_\alpha\left(y_{t-\alpha}\right)\right] = 0 \quad \text{and} \quad E\left[v_\alpha\left(y_{t-\alpha}\right)\right] = 0 \tag{1.1.4}$$

for all $\alpha = 1, \ldots, d$. The notable aspect of the model is additivity via known links for conditional mean and volatility functions. As will be shown below, (1.1.1)-(1.1.3) includes a wide variety of time series models in the literature.

In a much simpler univariate setup, Robinson (1983), Auestad and Tjøstheim (1990), and Härdle and Vieu (1992) studied the kernel estimation of conditional mean function, $m(\cdot)$ in (1.1.1). The so-called CHARN (Conditionally Heteroscedastic Autoregressive Nonlinear) is the same as (1.1.1) except that $m(\cdot)$ and $v(\cdot)$ are univariate functions of y_{t-1}. Masry and Tjøstheim (1995) and Härdle and Tsybakov (1997) applied the Nadaraya-Watson and local linear smoothing methods, respectively, to jointly estimate $v(\cdot)$ together with $m(\cdot)$. Also, in a nonlinear VAR context, Härdle, Tsybakov

*The extension to allow the F transformations to be of unknown functional form is considerably more complicated, but see Horowitz (1999).

and Yang (1996) dealt with the estimation of conditional mean in a multilagged extension similar to (1.1.1). Unfortunately, however, introducing more lags in nonparametric time series models has unpleasant consequences, more so than in the parametric approach. As is well known, smoothing method in high dimensions suffers from a slower convergence rate - the "curse of dimensionality". Under twice differentiability of $m(\cdot)$, the rate is $n^{2/(4+d)}$, which gets rapidly worse with dimension. It remains also a problem to find a proper (geometric) tool for interpreting estimation results.

Additive structure has been proposed as a useful way to circumvent these problems in multivariate smoothing. By assuming the target function to be a sum of functions of covariates, say, $m(y_{t-1}, y_{t-2}, \ldots, y_{t-d}) = c_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha})$, we can effectively reduce the dimensionality of a regression problem and improve the implementability of multivariate smoothing up to that of the one-dimensional case. Stone (1985,1986) showed that it is possible to estimate $m_\alpha(\cdot)$ and $m(\cdot)$ with the one-dimensional optimal rate of convergence - e.g., $n^{2/5}$ for twice differentiable functions - regardless of d. The estimates are now easily illustrated and interpreted. For these reasons, since the eighties, additive models have been fundamental to nonparametric regression among both econometricians and statisticians. Regarding the estimation method for achieving the one-dimensional optimal rate, the literature suggests two different approaches: *backfitting* and *marginal integration*. The former, originally suggested by Breiman and Friedman (1985), Buja, Hastie and Tibshirani (1989), and Hastie and Tibshirani (1987,1991) is to execute iterative calculations of one-dimensional smoothing, until some convergence criterion is satisfied. Though appealing to our intuition, the statistical properties of backfitting algorithm were not clearly understood until the very recent works by Opsomer and Ruppert (1997) and Mammen, Linton, and Nielsen (1999). Both papers developed specific backfitting procedures and addressed some statistical efficiency as well as the algorithmic properties on the existence and uniqueness of their estimators. However, one disadvantage of these procedures is the time consuming iterations required for implementation. The latter approach, marginal integration (MI), is theoretically more manipulable and its statistical properties are easy to derive, since it simply uses averaging of multivariate kernel estimates. Developed independently by Newey (1994), Tjøstheim and Auestadt (1994a), and Linton and Nielsen (1995), its advantage of theoretical convenience inspired the subsequent applications such as Linton, Wang, Chen, and Härdle (1997) for transformation models and Linton, Nielsen, and van de Geer (1999) for hazard models with censoring. In the time series models that are special cases of (1.1.1) and (1.1.2) with F_m being identity, Chen and Tsay (1993 a,b) and Masry and Tjøstheim (1997) applied backfitting and MI, respectively, to estimate the conditional mean function. Mammen, Linton, and Nielsen (1999) provided useful results for the same type of models, by improving the previous backfitting method with some modification and successfully deriving the asymptotic properties under weak conditions. The separability assumption was also used in volatility estimation by Yang, Härdle, and Nielsen (1999), where the nonlinear ARCH model is of additive mean and multiplicative volatility in the form of

$$y_t = c_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha}) + \left(c_v \prod_{\alpha=1}^{d} v_\alpha(y_{t-\alpha}) \right)^{1/2} \varepsilon_t. \tag{1.1.5}$$

To estimate (1.1.5), they relied on marginal integration with local linear fits as a pilot estimate, and derived asymptotic properties.

This paper features two contributions to the additive literature. The first concerns theoretical development of a new estimation tool called local instrumental variable method for additive models. The novelty of the procedure lies in the simple definition of the estimator based on univariate smoothing combined with new kernel weights. That is, adjusting kernel weights via

conditional density of the covariate enables an univariate kernel smoother to estimate consistently the corresponding additive component function. In many respects, the new estimator preserves the good properties of univariate smoothers. The instrumental variable method is analytically tractable for asymptotic theory and can be easily shown to attain the optimal one-dimensional rate as required. Furthermore, it is computationally more efficient than the two existing methods (backfitting and MI), in the sense that it reduces the computations up to a factor of n smoothings. The other contribution relates to the general coverage of the model we work with. The model in (1.1.1) through (1.1.3) extends ARCH models to a generalized additive framework where both the mean and variance functions are additive after some known transformation [see Hastie and Tibshirani (1990)]. All the time series models in our discussion above are regarded as a subclass of the data generating process for $\{y_t\}$ in (1.1.1) through (1.1.3). For example, setting G_m to be an identity and G_v a logarithmic function reduces our model to (1.1.5). Similar efforts to apply transformation were made in a parametric ARCH models. Nelson (1991) considered a model for the log of the conditional variance - the Exponential (G)ARCH class, to embody the multiplicative effects of volatility. It was also argued to use the Box-Cox transformation for volatility which is intermediate between linear and logarithm. Since it is hard to tell *a priori* which structure of volatility is more realistic and it should be determined by real data, our generalized additive model provides useful flexible specifications for empirical works. Additionally, from the perspective of potential misspecfication problems, the transformation used here alleviates the restriction imposed by additivity assumption, which increases the approximating power of our model. Note that when the lagged variables in (1.1.1) through (1.1.3) are replaced by different covariates and the observations are i.i.d., the model becomes the cross sectional additive model studied by Linton and Härdle (1996).

The rest of the paper is organized as follows. Section 2 describes the main estimation idea in a simple setting. In section 3, we define the estimator for the full model. In section 4 we give our main results including the asymptotic normality of our estimators. Section 5 discusses prediction. Section 6 gives some Monte Carlo and some empirical applications. The proofs are contained in the appendix.

2. Nonparametric Instrumental Variables: The Main Idea

This section explains the basic idea behind the instrumental variable method and defines the estimation procedure. For ease of exposition, this will be carried out using an example of simple additive models. We then extend the definition to the generalized additive ARCH case in (1.1.1) through (1.1.3).

Consider a bivariate additive regression model for i.i.d. data,

$$y = m_1(X_1) + m_2(X_2) + \varepsilon,$$

where $E(\varepsilon|X) = 0$ with $X = (X_1, X_2)$, and the components satisfy the identification conditions $E[m_\alpha(X_\alpha)] = 0$, for $\alpha = 1, 2$ [the constant term is assumed to be zero, for simplicity]. Let $p(\cdot)$, $p_1(\cdot)$, and $p_2(\cdot)$ be the density functions of the covariates X, X_1, and X_2, respectively. Letting $\eta = m_2(X_2) + \varepsilon$, we rewrite the model as

$$y = m_1(X_1) + \eta, \tag{2.2.1}$$

which is a classical example of "omitted variable" regression. That is, although (2.2.1) appears to take the form of a univariate nonparametric regression model, smoothing y on X_1 will incur a bias due to the omitted variable η, because η contains X_2, which in general depends on X_1. One solution

to this is suggested by the classical econometric notion of instrumental variable. That is, we look for an instrument W, usually just a function of X, such that

$$E\left(W|X_1\right) \neq 0 \quad ; \quad E\left(W\eta|X_1\right) = 0 \tag{2.2.2}$$

with probability one. If such a random variable exists, we can write

$$m_1\left(x_1\right) = \frac{E(Wy|X_1 = x_1)}{E(W|X_1 = x_1)}. \tag{2.2.3}$$

This suggests that we estimate the function $m_1\left(\cdot\right)$ by nonparametric smoothing of both Wy and W on X_1. In parametric models the choice of instrument is usually not obvious and requires some caution. However, our additive model has a natural class of instruments – any measurable function of X_1 times $p_2\left(X_2\right)/p\left(X\right)$ will do.† In fact, we will take

$$W(X) = \frac{p_2\left(X_2\right)}{p\left(X\right)} \tag{2.2.4}$$

throughout. Note that

$$\begin{aligned} \frac{E(Wy|X_1)}{E(W|X_1)} &= \frac{\int W(X)m(X)\frac{p(X)}{p_1(X_1)}dX_2}{\int W(X)\frac{p(X)}{p_1(X_1)}dX_2} \\ &= \frac{\int W(X)m(X)p(X)dX_2}{\int W(X)p(X)dX_2}, \end{aligned}$$

which is to say that the marginal density $p_1(X_1)$ cancels out. This is useful when we come to construct estimators. This formula also shows what the instrumental variable estimator is estimating when m is not additive. It is estimating just a ratio of weighted averages of the function. When $W(X) = p_2\left(X_2\right)/p\left(X\right)$, the target is exactly the same as the target of the marginal integration estimator.

Up to now, it was implicitly assumed that the distributions of the covariates are known *a priori.* In practice, this is rarely true, and we have to rely on estimates of these quantities. Let $\widehat{p}\left(\cdot\right), \widehat{p}_1\left(\cdot\right)$, and $\widehat{p}_2\left(\cdot\right)$ be kernel estimates of the densities $p\left(\cdot\right), p_1\left(\cdot\right)$, and $p_2\left(\cdot\right)$, respectively. Then, the feasible procedure is defined with a replacement of instrumental variable, $\widehat{W} = \widehat{p}_2\left(X_2\right)/\widehat{p}\left(X\right)$. Section 3 provides rigorous statistical treatment for feasible instrumental variable estimators based on local linear estimation. See Kim, Linton, and Hengartner (1999), for a slightly different approach.

Note the contrast with the marginal integration or projection method. In this approach one defines m_1 by some unconditional expectation

$$m_1(x_1) = E[m(x_1, X_2)W(X_2)]$$

for some weighting function W that depends only on X_2 and which satisfies

$$E[W(X_2)] = 1 \quad ; \quad E[W(X_2)m_2(X_2)] = 0.$$

†Indeed, suppose we take

$$W(X) = \frac{p_1(X_1)p_2\left(X_2\right)}{p\left(X\right)}.$$

This satisfies $E\left(W|X_1\right) = 1$ and $E(W\eta|X_1) = 0$.

Next, we come to the main advantage that the local instrumental variable method has. This is in terms of the computational cost. The marginal integration method actually needs n^2 regression smoothings evaluated at the pairs (X_{1i}, X_{2j}), for $i, j = 1, \ldots, n$, while the backfitting method requires nr operations-where r is the number of iterations to achieve convergence. The instrumental variable procedure, in contrast, takes at most $2n$ operations of kernel smoothings in a preliminary step for estimating instrumental variable, and another n operations for regressions. Thus, it can be easily combined with bootstrap method whose computational costs often becomes prohibitive in the case of marginal integration.

Finally, we show how the instrumental variable approach can be applied to generalized additive models. Let $F(\cdot)$ be the inverse of a known link function $G(\cdot)$ and let $m(X) = E(y|X)$. The model is defined as

$$y = F(m_1(X_1) + m_2(X_2)) + \varepsilon, \tag{2.2.5}$$

or equivalently $G(m(X)) = m_1(X_1) + m_2(X_2)$. We maintain the same identification condition, $E[m_\alpha(X_\alpha)] = 0$. Unlike in the simple additive model, there is no direct way to relate Wy to $m_1(X_1)$, here, but, nevertheless

$$m_1(X_1) = \frac{E[WG(m(X))|X_1]}{E[W|X_1]}.$$

Since $m(\cdot)$ is unknown, we need consistent estimates of $m(X)$ in a preliminary step.

3. Instrumental Variable Procedure for GANARCH

We start with some simplifying notations that will be used repeatedly throughout the paper. Let x_t be the vector of d lagged variables until $t-1$, that is, $x_t = (y_{t-1}, \ldots, y_{t-d})$, or concisely, $x_t = (y_{t-\alpha}, \underline{y}_{t-\alpha})$, where $\underline{y}_{t-\alpha} = (y_{t-1}, \ldots, y_{t-\alpha-1}, y_{t-\alpha+1}, \ldots, y_{t-d})$. Defining $m_{\underline{\alpha}}(\underline{y}_{t-\alpha}) = \sum_{\beta=1,\neq\alpha}^{d} m_\beta(y_{t-\beta})$ and $v_{\underline{\alpha}}(\underline{y}_{t-\alpha}) = \sum_{\beta=1,\neq\alpha}^{d} v_\beta(y_{t-\beta})$, we can reformulate (1.1.1) through (1.1.3) with a focus on the αth components of mean and variance as

$$\begin{aligned} y_t &= m(x_t) + v^{1/2}(x_t)\,\varepsilon_t, \\ m(x_t) &= F_m\left(c_m + m_\alpha(y_{t-\alpha}) + m_{\underline{\alpha}}(\underline{y}_{t-\alpha})\right), \\ v(x_t) &= F_v\left(c_v + v_\alpha(y_{t-\alpha}) + v_{\underline{\alpha}}(\underline{y}_{t-\alpha})\right). \end{aligned}$$

To save space we will use the following abbreviations for functions to be estimated:

$$\begin{aligned} H_\alpha(y_{t-\alpha}) &\equiv [m_\alpha(y_{t-\alpha}), v_\alpha(y_{t-\alpha})]^T,\ H_{\underline{\alpha}}(\underline{y}_{t-\alpha}) \equiv \left[m_{\underline{\alpha}}(\underline{y}_{t-\alpha}), v_{\underline{\alpha}}(\underline{y}_{t-\alpha})\right]^T, \\ c &\equiv [c_m, c_v]^T,\ z_t \equiv H(x_t) = [G_m(m(x_t)), G_v(v(x_t))]^T \\ \varphi_\alpha(y_\alpha) &\equiv [M_\alpha(y_\alpha), V_\alpha(y_\alpha)]^T = c + H_\alpha(y_\alpha). \end{aligned}$$

Note that the components $[m_\alpha(\cdot), v_\alpha(\cdot)]^T$ are identified, up to constant, c, by $\varphi_\alpha(\cdot)$, which will be our major interest in estimation. Below, we examine some details in each relevant step for computing the feasible nonparametric instrumental variable estimator of $\varphi_\alpha(\cdot)$. The set of observations is given by $\mathcal{Y} = \{y_t\}_{t=1}^{n'}$, where $n' = n + d$.

3.1. **Step I: Preliminary Estimation:** $z_t = H(x_t)$**.** Since z_t is unknown, we start with computing the pilot estimates of the regression surface by a local linear smoother. Let $\widetilde{m}(x)$ be the first component of $(\widetilde{a}, \widetilde{b})$ that solves

$$\min_{a,b} \sum_{t=d+1}^{n'} K_h(x_t - x) \{y_t - a - b(x_t - x)\}^2, \tag{3.3.1}$$

where $K_h(x) = \Pi_{i=1}^d K(x_i/h)/h^d$ and K is a one-dimensional kernel function and $h = h(n)$ is a bandwidth sequence. In a similar way, we get the estimate of the volatility surface, $\widetilde{v}(\cdot)$, from (3.3.1) by replacing y_t with the squared residuals, $\widetilde{\varepsilon}_t^2 = (y_t - \widetilde{m}(x_t))^2$. Then, transforming $\widetilde{m}$ and $\widetilde{v}$ by the known links will leads to consistent estimates of $\widetilde{z}_t$,

$$\widetilde{z}_t = \widetilde{H}(x_t) = [G_m(\widetilde{m}(x_t)), G_v(\widetilde{v}(x_t))]^T.$$

3.2. **Step II: Instrumental Variable Estimation of Additive Components.** This step involves the estimation of $\varphi_\alpha(\cdot)$, which is equivalent to $[m_\alpha(\cdot), v_\alpha(\cdot)]^T$, up to the constant c. Let $p(\cdot)$ and $p_{\underline{\alpha}}(\cdot)$ denote the density functions of the random variables $(y_{t-\alpha}, \underline{y}_{t-\alpha})$ and $\underline{y}_{t-\alpha}$, respectively. Define the feasible instrument as

$$\widehat{W}_t = \frac{\widehat{p}_{\underline{\alpha}}(\underline{y}_{t-\alpha})}{\widehat{p}(y_{t-\alpha}, \underline{y}_{t-\alpha})},$$

where $\widehat{p}_{\underline{\alpha}}(\cdot)$ and $\widehat{p}(\cdot)$ are computed using the kernel function $L(\cdot)$, e.g., $\widehat{p}(x) = \sum_{t=1}^n \prod_{i=1}^d L_g(x_{it} - x_i)/n$ with $L_g(\cdot) \equiv L(\cdot/g)/g$. The instrumental variable local linear estimates $\widehat{\varphi}_\alpha(y_\alpha)$ are given as $(a_1, a_2)^T$ through minimizing the localized squared errors elementwise

$$\min_{a_j, b_j} \sum_{t=d+1}^{n'} K_h(y_{t-\alpha} - y_\alpha) \widehat{W}_t \{\widetilde{z}_{jt} - a_j - b_j(y_{t-\alpha} - y_\alpha)\}^2, \tag{3.3.2}$$

where $\widetilde{z}_{jt}$ is the j-th element of $\widetilde{z}_t$. The closed form of the solution is

$$\widehat{\varphi}_\alpha(y_\alpha)^T = e_1^T \left(\mathbf{Y}_-^\mathbf{T} \mathbf{K} \mathbf{Y}_-\right)^{-1} \mathbf{Y}_-^\mathbf{T} \mathbf{K} \widetilde{\mathbf{Z}}, \tag{3.3.3}$$

where $e_1 = (1, 0)^T$, $\mathbf{Y}_- = [\iota, \mathbf{Y}_-]$, $\mathbf{K} = \mathrm{diag}[\mathrm{K_h}(\mathrm{y_{d+1-\alpha}} - \mathrm{y_\alpha})\widehat{\mathrm{W}}_{\mathrm{d+1}}, \ldots, \mathrm{K_h}(\mathrm{y_{n'-\alpha}} - \mathrm{y_\alpha})\widehat{\mathrm{W}}_{\mathrm{n'}}]$, and $\widetilde{\mathbf{Z}} = (\widetilde{z}_{d+1}, \ldots, \widetilde{z}_{n'})^T$, with $\iota = (1, \ldots, 1)^T$ and $Y_- = (y_{d+1-\alpha} - y_\alpha, \ldots, y_{n'-\alpha} - y_\alpha)^T$.

4. Main Results

Let $\mathcal{F}_{\rfloor}^{\dashv}$ be the σ-algebra of events generated by $\{y_t\}_a^b$ and $\alpha(k)$ the strong mixing coefficient of $\{y_t\}$ which is defined by

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

Throughout the paper, we assume

C1. $\{y_t\}_{t=1}^\infty$ is stationary and strongly mixing with a mixing coefficient, $\alpha(k) = \rho^{-\beta k}$, for some $\beta > 0$.

C.1 is a standard mixing condition with a geometrically decreasing rate. However, the asymptotic theory for the instrumental variable estimator is developed based on a milder condition on the mixing coefficient - as was pointed out by Masry and Tjøstheim (1997), $\sum_{k=0}^{\infty} k^a \{\alpha(k)\}^{1-2/\nu} < \infty$, for some $\nu > 2$ and $0 < a < (1-2/\nu)$. It is easy to verify that this condition holds under C.1. Some technical conditions for regularity are stated.

C2. The additive component functions, $m_\alpha(\cdot)$, and $v_\alpha(\cdot)$, for $\alpha = 1, \ldots, d$, are continuous and twice differentiable on their compact supports.

C.3 The link functions, G_m and G_v, have bounded continuous second order derivatives over any compact interval.

C4. The joint and marginal density functions, $p(\cdot)$, $p_{\underline{\alpha}}(\cdot)$, and $p_\alpha(\cdot)$, for $\alpha = 1, \ldots, d$, are continuous, twice differentiable with bounded (partial) derivatives, and bounded away from zero on the compact support.

C5. The kernel functions, $K(\cdot)$ and $L(\cdot)$, are a real bounded nonnegative symmetric function on compact support satisfying $\int K(u)\,du = \int L(u)\,du = 1$, $\int uK(u)\,du = \int uL(u)\,du = 0$. Also, assume that the kernel functions are Lipschitz-continuous, $|K(u) - K(v)| \leq C|u - v|$.

C6. (i) $g \to 0$, $ng^d \to \infty$, and (ii) $h \to 0$, $nh \to \infty$. (iii) The bandwidth satisfies $\sqrt{\frac{n}{h}}\alpha(t(n)) \to 0$,where $\{t(n)\}$ be a sequence of positive integers, $t(n) \to \infty$ such that $t(n) = o(\sqrt{nh})$.

Conditions C.2 through C.5 are standard in kernel estimation. The continuity assumption in C2 and C4, together with the compact support, implies that the functions are bounded. The additional bandwidth condition in C.6(iii) is necessary to control the effects from the dependence of mixing processes in showing the asymptotic normality of instrumental variable estimates. The proof of consistency, however, does not require this condition for bandwidths. Define $D^2 f(x_1, \ldots, x_d) = \sum_{l=1}^{d} \partial^2 f(x_l)/\partial^2 x$ and $[\nabla G_m(t), \nabla G_v(t)] = [dG_m(t)/dt, dG_v(t)/dt]$. Let $(K*K)_i(u) = \int K(w)K(w+u)w^i dw$, a convolution of kernel functions, and $\mu^2_{K*K} = \int (K*K)_0(u)u^2 du$, while $||K||_2^2$ denotes $\int K^2(u)\,du$. The asymptotic properties of the feasible instrumental variable estimates in (3.3.3) are summarized in the following theorem whose proof is in the Appendix. Let $\kappa_3(y_\alpha, z_{\underline{\alpha}}) = E[\varepsilon_t^3 | x_t = (y_\alpha, z_{\underline{\alpha}})]$, and $\kappa_4(y_\alpha, z_{\underline{\alpha}}) = E[(\varepsilon_t^2 - 1)^2 | x_t = (y_\alpha, z_{\underline{\alpha}})]$. $A \odot B$ denotes the matrix Hadamard product.

Theorem 1. Assume that conditions C.1 through C.6 hold. Then,

$$\sqrt{nh}[\widehat{\varphi}_\alpha(y_\alpha) - \varphi_\alpha(y_\alpha) - B_\alpha] \xrightarrow{d} N[0, \Sigma^*_\alpha(y_\alpha)],$$

where

$$\begin{aligned} B_\alpha(y_\alpha) &= \frac{h^2}{2}\mu_K^2 D^2\varphi_\alpha(y_\alpha) \\ &\quad + \frac{h^2}{2}\int [\mu^2_{K*K} D^2\varphi_\alpha(y_\alpha) + \mu_K^2 D^2\varphi_{\underline{\alpha}}(z_{\underline{\alpha}})] \odot [\nabla G_m(m(y_\alpha, z_{\underline{\alpha}})), \nabla G_v(v(y_\alpha, z_{\underline{\alpha}}))]^T p_{\underline{\alpha}}(z_{\underline{\alpha}}) dz_{\underline{\alpha}} \\ &\quad + \frac{g^2}{2}\mu_K^2 \int [D^2 p_{\underline{\alpha}}(z_{\underline{\alpha}}) - \frac{p_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} D^2 p(y_\alpha, z_{\underline{\alpha}})] H_{\underline{\alpha}}(z_{\underline{\alpha}}) dz_{\underline{\alpha}}, \end{aligned}$$

$$\Sigma^*_\alpha(y_\alpha) = ||K||_2^2 \int \frac{p^2_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} \begin{bmatrix} m^2_{\underline{\alpha}}(z_{\underline{\alpha}}) & m_{\underline{\alpha}}(z_{\underline{\alpha}})v_{\underline{\alpha}}(z_{\underline{\alpha}}) \\ m_{\underline{\alpha}}(z_{\underline{\alpha}})v_{\underline{\alpha}}(z_{\underline{\alpha}}) & v^2_{\underline{\alpha}}(z_{\underline{\alpha}}) \end{bmatrix} dz_{\underline{\alpha}}$$

$$+||(K*K)_0||_2^2 \int \frac{p^2_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} \begin{bmatrix} \nabla G_m(m)^2 \cdot v & (\nabla G_m \cdot \nabla G_v)(\kappa_3 \cdot v^{3/2}) \\ (\nabla G_m \cdot \nabla G_v)(\kappa_3 \cdot v^{3/2}) & \nabla G_v(v)^2 \cdot \kappa_4 \cdot v^2 \end{bmatrix} (y_\alpha, z_{\underline{\alpha}}) dz_{\underline{\alpha}}.$$

REMARKS. 1. To estimate $[m_\alpha(y_\alpha), v_\alpha(y_\alpha)]^T$, we can use the following recentered estimates, $\widehat{\varphi}_\alpha(y_\alpha) - \widehat{c}$, where $\widehat{c} = [\widehat{c}_m, \widehat{c}_v] = \frac{1}{n}[\sum_t y_t, \sum_t \widetilde{\varepsilon}^2_t]^T$ and $\widetilde{\varepsilon}_t = y_t - \widetilde{m}(x_t)$. Since $\widehat{c} = c + O_p(1/\sqrt{n})$, the bias and variance of $[\widehat{m}_\alpha(y_\alpha), \widehat{v}_\alpha(y_\alpha)]^T$ are the same as those of $\widehat{\varphi}_\alpha(y_\alpha)$. For $y = (y_1, \ldots, y_d)$, the estimates for the conditional mean and volatility are defined by

$$[\widehat{m}(y), \widehat{v}(y)] \equiv \left[F_m[-(d-1)\,\widehat{c}_m + \sum_{\alpha=1}^d \widehat{\varphi}_{\alpha 1}(y_\alpha)],\ F_v[-(d-1)\,\widehat{c}_v + \sum_{\alpha=1}^d \widehat{\varphi}_{\alpha 2}(y_\alpha)] \right].$$

Let $\nabla F(y) \equiv [\nabla F_m(m(y)), \nabla F_v(v(y))]^T$. Then, by Theorem 1 and the Delta method, their asymptotic distribution satisfies

$$\sqrt{nh}\,[\widehat{m}(y) - m(y) - b_m(y), \widehat{v}(y) - v(y) - b_v(y)]^T \xrightarrow{d} N[0, \Sigma^*(y)],$$

where $[b_m(y), b_v(y)]^T = \nabla F(y) \odot \sum_{\alpha=1}^d B_\alpha(y_\alpha)$, and $\Sigma^*(y) = [\nabla F(y) \nabla F(y)^T] \odot [\Sigma^*_1(y_1) + \ldots + \Sigma^*_d(y_d)]$. It is easy to see that $\widehat{\varphi}_\alpha(y_\alpha)$ and $\widehat{\varphi}_\beta(y_\beta)$ are asymptotically uncorrelated for any α and β, and the asymptotic variance of their sum is also the sum of the variances of $\widehat{\varphi}_\alpha(y_\alpha)$ and $\widehat{\varphi}_\beta(y_\beta)$.

2. The first term of the bias is of the standard form, depending only on the second derivatives as in other local linear smoothing. The last term reflects the biases from using estimates for density functions to construct the feasible instrumental variable, $\widehat{p}_{\underline{\alpha}}(\underline{y}_{t-\alpha})/\widehat{p}(x_t)$. When the instrument consisting of known density functions, $p_{\underline{\alpha}}(\underline{y}_{t-\alpha})/p(x_t)$, is used in (3.3.2), the asymptotic properties of IV estimates are the same as those from Theorem 1 except that the new asymptotic bias now includes only the first two terms of $B_\alpha(y_\alpha)$.

3. The convolution kernel $(K*K)(\cdot)$ is the legacy of double smoothing in the instrumental variable estimation of 'generalized' additive models, since we smooth $[G_m(\widetilde{m}(\cdot)), G_v(\widetilde{v}(\cdot))]$ with $\widetilde{m}(\cdot)$ and $\widetilde{v}(\cdot)$ given by (multivariate) local linear fits. When $G_m(\cdot)$ is the identity, we can directly smooth y instead of $G_m(\widetilde{m}(x_t))$ to estimate the components of the conditional mean function. Then, as the following theorem shows, the second term of the bias of B_α does not arise, and the convolution kernel in the variance is replaced by a usual kernel function.

Suppose that $F_m(t) = F_v(t) = t$ in (1.1.2) and (1.1.3). The instrumental variable estimate of the α-th component, $[\widehat{M}_\alpha(y_\alpha), \widehat{V}_\alpha(y_\alpha)]$, is now the solution to the adjusted-kernel least squares in (3.3.2) with a modification that the (2×1) vector $\widetilde{z}_t$ is replaced by $[y_t, \widetilde{\varepsilon}^2_t]^T$ with $\widetilde{\varepsilon}_t$ defined in step I of section 2.2. Theorem 2 shows the asymptotic normality of these instrumental variable estimates. The proof is almost the same as that of Theorem 1 and is thus omitted.

Theorem 2. Under the same conditions as Theorem 1,

$$i)\ \sqrt{nh}[\widehat{M}_\alpha(y_\alpha) - M_\alpha(y_\alpha) - b^m_\alpha] \xrightarrow{d} N[0, \sigma^m_\alpha(y_\alpha)],$$

where

$$b_\alpha^m(y_\alpha) = \frac{h^2}{2}\mu_K^2 D^2 m_\alpha(y_\alpha) + \frac{g^2}{2}\mu_K^2 \int [D^2 p_{\underline{\alpha}}(z_{\underline{\alpha}}) - \frac{p_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} D^2 p(y_\alpha, z_{\underline{\alpha}})] m_{\underline{\alpha}}(z_{\underline{\alpha}}) dz_{\underline{\alpha}},$$

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

and

$$ii)\ \sqrt{nh}[\widehat{V}_\alpha(y_\alpha) - V_\alpha(y_\alpha) - b_\alpha^v] \xrightarrow{d} N[0, \sigma_\alpha^v(y_\alpha)],$$

where

$$b_\alpha^v(y_\alpha) = \frac{h^2}{2}\mu_K^2 D^2 v_\alpha(y_\alpha) + \frac{g^2}{2}\mu_K^2 \int [D^2 p_{\underline{\alpha}}(z_{\underline{\alpha}}) - \frac{p_{\underline{\alpha}}(z_{\underline{\alpha}})}{p(y_\alpha, z_{\underline{\alpha}})} D^2 p(y_\alpha, z_{\underline{\alpha}})] v_{\underline{\alpha}}(z_{\underline{\alpha}}) dz_{\underline{\alpha}},$$

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

Although the instrumental variable estimators achieve the one-dimensional optimal convergence rate, there is room for improvement in terms of variance. For example, compared to the marginal integration estimators of Linton and Härdle (1996) or Linton and Nielsen (1995), the asymptotic variances of the instrumental variable estimates for $m_1(\cdot)$ in Theorem 1 and 2 include an additional factor of $m_2^2(\cdot)$. This is because the instrumental variable approach treats $\eta = m_2(X_2) + \varepsilon$ in (2.2.1) as if it were the error term of the regression equation for $m_1(\cdot)$. Note that the asymptotic covariance in Theorem 1 is the same as that in Yang, Härdle, and Nielsen (1999), where they only considered the case with additive mean and multiplicative volatility functions. The issue of efficiency in estimating an additive component was first addressed by Linton (1996) based on 'oracle efficiency' bounds of infeasible estimators under the knowledge of other components. According to this, both instrumental variable and marginal integration estimators are inefficient, but they can attain the efficiency bounds through one simple additional step, following Linton (1996, 2000) and Kim, Linton, and Hengartner (1999).

5. Prediction

Suppose that the time series $\{Y_i\}_{i=1}^\infty$ is a Markov process of order d. Given $\{Y_i\}_{i=1}^t$, the best (nonlinear) predictor of the future value, Y_{t+l} for $l \geq 1$ is $E[Y_{t+l}|Y_t, \ldots, Y_{t-d+1}]$. In a linear model, the l-step predictor is linear in the variables for any l [although not in the parameters], this structure does not carry over to nonparametric models, since each l-step ahead conditional expectation, in general, takes a different form of a function of lagged variables depending on l. In other words, the nonlinear prediction surface of a given process varies with the number of steps ahead. This implies that the estimated regression function in an autoregressive model, for example, is not able to give information for prediction, except of the one-step ahead kind. Instead, we have to resmooth Y_{t+l} on the lagged variables at a given point of prediction.

The curse of dimensionality might again interfere with reliable prediction when the projection subspace is of a multiple dimension. It is natural to retain additive restrictions for the purpose of approximating the prediction function by a more estimable one. Compared to the prediction error of $O_p(1)$ due to the disturbance term, such approximation error may not be so serious and is often preferable on the grounds of the faster convergence rate. Thus, to predict Y_{t+l}, we will use an additive

approximation of $E[Y_{t+l}|Y_t,\ldots,Y_{t-d+1}]$, and the best nonlinear "additive" l-step ahead predictor of Y_{t+l} is defined by

$$E_{AP}[Y_{t+l}|Y_t,\ldots,Y_{t-d+1}] = c_l + \sum_{\alpha=1}^{d} m_\alpha^l(Y_{t+1-\alpha}),$$

with an identification condition of $E\left[m_\alpha^l(Y_{t+1-\alpha})\right]=0$, for all $\alpha=1,\ldots,d$, and $l=1,2\ldots$. Now, if we apply the instrumental variable method in section 3, the estimates of additive components are given by

$$\widehat{M^l}_\alpha(y_\alpha) = e_1^T\left(\mathbf{Y_-^T K Y_-}\right)^{-1}\mathbf{Y_-^T K}\tilde{\mathbf{Z}},$$

where $\mathbf{Y_-}=[\iota,\mathbf{Y_-}]$, $\mathbf{K}=\mathrm{diag}\left[\mathrm{K_h(Y_{d+1-\alpha}-y_\alpha)\widehat{W}_{d+l},..,K_h(Y_{t-(l-1)-\alpha}-y_\alpha)\widehat{W}_t}\right]$, and $\tilde{\mathbf{Z}}=(Y_{d+l},\ldots,Y_t)^T$, and $Y_-=(Y_{d+1-\alpha}-y_\alpha,..,Y_{t-(l-1)-\alpha}-y_\alpha)^T$. The predicted value of Y_{t+l} is estimated by

$$\widehat{Y}_{t+l} = -(d-1)\hat{c}_l + \sum_{\alpha=1}^{d}\widehat{M^l}_\alpha(Y_{t+1-\alpha}), \tag{5.5.1}$$

where $\hat{c}_l=\frac{1}{d}\sum_{\alpha=1}^{d}(\frac{1}{n_l}\sum_{i=d}^{t-l}\widehat{M^l}_\alpha(Y_{i+1-\alpha}))$,with $n_l=t-l-(d-1)$. The asymptotic properties of the predictor (5.5.1) are already shown in the theorems in the previous section. Although it is a consistent estimate of $E_{AP}[Y_{t+l}|Y_t,\ldots,Y_{t-d+1}]$, some alternative way can reduce the prediction error. Using the information contained in $(Y_{t+l-1},\ldots,Y_{t+l-d})$, Chen (1996) developed a multistage smoother whose mean squared error is smaller than the direct predictor in (5.5.1). In what follows, we integrate the method of multistage smoothers with the additivity-based prediction above. First, observe that, by means of the property of a Markov process, we get, under additivity assumption,

$$\begin{aligned}
E(Y_{t+l}|Y_t,\ldots,Y_{t-d+1}) &= E\left[E(Y_{t+l}|Y_{t+l-1},\ldots,Y_{t+l-d},Y_t,\ldots,Y_{t-d+1})\,|Y_t,\ldots,Y_{t-d+1}\right]\\
&= E\left[E(Y_{t+l}|Y_{t+l-1},\ldots,Y_{t+l-d})\,|Y_t,\ldots,Y_{t-d+1}\right]\\
&= E\left[c_l+\sum_{\alpha=1}^{d}m_\alpha^l(Y_{t+l-\alpha})\,|Y_t,\ldots,Y_{t-d+1}\right].
\end{aligned}$$

The inequality

$$\mathrm{var}\left[\mathrm{c_l}+\sum_{\alpha=1}^{\mathrm{d}}\mathrm{m_\alpha^l(Y_{t+l-\alpha})}|\mathrm{Y_t},\ldots,\mathrm{Y_{t-d+1}}\right]\le\mathrm{var}\left[\mathrm{Y_{t+l}}|\mathrm{Y_t},\ldots,\mathrm{Y_{t-d+1}}\right]$$

implies that it will be ideal to use the pairs $\left[c_l+\sum_{\alpha=1}^{d}m_\alpha^l(Y_{k+l-\alpha}),Y_k,\ldots,Y_{k-d+1}\right]$, $k=d,d+1,\ldots,n-l+d$, in estimating $E(Y_{t+l}|Y_t,\ldots,Y_{t-d+1})$, if we knew the true regression function. Since the direct predictor is using $\{Y_{k+l}\}_{k=d}^{t-l}$ that equals the noisy representative of $\{E(Y_{k+l}|Y_k,\ldots,Y_{k-d+1})\}_{k=d}^{t-l}$ with an error of $O_p(1)$, we can still hope for improvement in the efficiency of prediction through its consistent estimate, $\hat{c}^l+\sum_{\alpha=1}^{d}\widehat{m}_\alpha^l(Y_{k+l-\alpha})$ which is $E(Y_{k+l}|Y_k,\ldots,Y_{k-d+1})$ plus an error of $o_p(1)$. Letting $\widehat{m}(\cdot)$ be the estimated regression surface by instrumental variable method, our multistage smoother is defined as

$$\widehat{Y}_{t+l}^* = -(d-1)\hat{c}_l^* + \sum_{\alpha=1}^{d}\widehat{M}_\alpha^{*l}(Y_{t+1-\alpha}), \tag{5.5.2}$$

where $\widehat{M}_\alpha^{*l}(Y_{t+1-\alpha})$ has the same form as $\widehat{M}_\alpha^{l}(Y_{t+1-\alpha})$ except that $\widetilde{\mathbf{Z}} = (Y_{d+l}, \ldots, Y_t)^T$ is now replaced by $[\widehat{m}(Y_{d+l-1}, \ldots, Y_l), \ldots, \widehat{m}(Y_{t-1}, \ldots, Y_{t-d})]^T$, and $\widehat{c}_l^* = \frac{1}{d}\sum_{\alpha=1}^{d}\left(\frac{1}{n_l}\sum_{i=d}^{t-l}\widehat{M}_\alpha^{*l}(Y_{i+1-\alpha})\right)$.

6. Empirical Examples

A Monte Carlo simulation is carried out to investigate the finite sample properties of instrumental variable estimates. The design in our experiment is Additive Nonlinear ARCH(2):

$$y_t = [0.2 + v_1(y_{t-1}) + v_2(y_{t-2})]\varepsilon_t,$$

$$\begin{aligned} v_1(y) &= 0.4[0.1 + 2\Psi_L(-5y)]y^2, \\ v_2(y) &= 0.4[0.1 + 2\Phi_N(-5y)]y^2, \end{aligned}$$

where $\Psi_L(\cdot)$ and $\Phi_N(\cdot)$ are the cdf of logistic distribution and standard normal, respectively, and ε_t is i.i.d. with $N(0,1)$. Fig.1(solid lines) depicts the asymmetric shape of the volatility component functions defined by $v_1(\cdot)$ and $v_2(\cdot)$. For each realization of the ARCH process we apply the instrumental variable procedure in (3.3.2) with $\widetilde{z}_t = y_t^2$ to estimate the volatility functions. The sample size $n = 500$ and total number of repetitions is 1000. The quartic kernel is used for both (3.3.2) and density estimates which constitute the instruments, W; i.e., $K(x) = L(x) = (15/16)I(|x| \leq 1)(1 - x^2)$. For the kernel weights of (3.3.2), we allow for different bandwidths according to the rule of thumb (Härdle, 1990), $h = c_h std(y_t) n^{-1/5}$ by taking various bandwidth constants, c_h, where $std(y_t)$ is the standard deviation of y_t. A similar rule determines the bandwidths for marginal and joint density estimates but with a fixed bandwidth constant of 2.

Table 1 here

Table 1 compares the nonparametric instrumental variable to a parametric method based on quadratic specification via the average and median of their MSE. For various bandwidth choices, the mean squared error of the nonparametric estimator are much smaller than those of the parametric approach. Noting that the median is far below the average due to a few worst cases, we also report the *mse* of both estimates after removing 24 (2.4%) extreme values - see the last two columns. In sum, our simulation shows that the new method performs well in finite sample cases, although we did not optimize the choice bandwidths. Fig. 1 gives the instrumental variable estimates (with $c_h = 4$) of volatility functions for four typical (consecutive) realizations of ARCH processes.

Fig. 1 here

We continue to illustrate our methodology in an empirical study for conditional heteroscedasticity of stock returns. The data of our interest is the monthly return series ($n = 744$) of the S&P 500 index from the New York Stock Exchange during the period of 1926.01 - 1987.12. An AR-ARCH(2,2) model is considered to characterize the stock returns, y_t,

$$y_t = c_m + m_1(y_{t-1}) + m_2(y_{t-2}) + [c_v + v_1(y_{t-1}) + v_2(y_{t-2})]^{1/2}\varepsilon_t,$$

where ε_t is i.i.d. with mean zero and variance $\sigma_\varepsilon^2 = 1$. This model is a special case of (1.1.1) - (1.1.3) with F_m and F_v being the identity. The instrumental variable procedure in Section 2.2. is first applied to estimate the conditional mean and then the volatility function based on the squared residuals. The quartic kernel is used with a bandwidth of $h = 3std(y_t)n^{-1/5}$ for instrumental variable

estimators and $g = 2std(y_t)n^{-1/(4+d)}$ for the marginal and joint density estimates. Fig. 2 gives the estimates for each component function of conditional mean and volatility with 95% pointwise (asymptotic) confidence bands.

*** Fig. 2 here ***

As Fig. 2-1 shows, the confidence intervals of both $m_1(\cdot)$ and $m_2(\cdot)$ include zero for most values of lagged variables. This implies that the lagged effects on conditional mean are not significant, partly supporting the efficient market hypothesis. However, the effect on the volatility does seem to be significant-the negative values of volatility in Fig. 2-2 are due to recentering. It also strongly suggests that the usual quadratic specification may lack empirical evidence. More importantly, Fig. 2-2 shows that the future risk is affected in a different way by lagged returns, reflecting the asymmetric behavior of volatility with respect to the changes of past returns. Such asymmetry has been found in other empirical studies on the "leverage effect" of stock returns, see Bollerslev, Engle, and Nelson (1994).

Appendix A.

The proof of Theorem 1 consists of three steps. Without loss of generality we deal with the case $\alpha = 1$; below we will use the subscript '2', for expositional convenience, to denote the nuisance direction. That is, $p_2(\underline{y}_{k-1}) = p_{\underline{1}}(\underline{y}_{k-1})$ in the case of density function. For component functions, $m_2(\underline{y}_{k-1})$, $v_2(\underline{y}_{k-1})$, and $H_2(\underline{y}_{k-1})$ will be used instead of $m_{\underline{1}}(\underline{y}_{k-1})$, $v_{\underline{1}}(\underline{y}_{k-1})$, and $H_{\underline{1}}(\underline{y}_{k-1})$, respectively. We start by decomposing the estimation errors, $\widehat{\varphi}_1(y_1) - \varphi_1(y_1)$, into the main stochastic term and bias. Use $X_n \simeq Y_n$ to mean $X_n = Y_n\{1 + o_p(1)\}$ in the following. Let $vec(X)$ denote the vectorization of the elements of the matrix X along with columns.

Step I: Decompositions and Approximations

Since $\widehat{\varphi}_1(y_1)$ is a column vector, the vectorization of eq. (3.3.3) gives

$$\widehat{\varphi}_1(y_1) = [I_2 \otimes e_1^T \left(\mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\mathbf{Y}_-\right)^{-1}] \left(I_2 \otimes \mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\right) \mathrm{vec}\left(\widetilde{\mathbf{Z}}\right).$$

A similar form is obtained for the true function, $\varphi_1(y_1)$,

$$[I_2 \otimes e_1^T \left(\mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\mathbf{Y}_-\right)^{-1}] \left(I_2 \otimes \mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\right) \mathrm{vec}\left(\iota\varphi_1^{\mathrm{T}}(\mathrm{y}_1) + \mathrm{Y}_-\nabla\varphi_1^{\mathrm{T}}(\mathrm{y}_1)\right),$$

by the identity,

$$\varphi_1(y_1) = \mathrm{vec}\{\mathrm{e}_1^{\mathrm{T}} \left(\mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\mathbf{Y}_-\right)^{-1} \mathbf{Y}_-^{\mathbf{T}}\mathbf{K}[\iota\varphi_{\mathbf{1}}^{\mathbf{T}}(\mathbf{y_1}) + \mathbf{Y}_-\nabla\varphi_{\mathbf{1}}^{\mathbf{T}}(\mathbf{y_1})]\},$$

since

$$e_1^T \left(\mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\mathbf{Y}_-\right)^{-1} \mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\iota = \mathbf{1},\ \mathbf{e_1^T} \left(\mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\mathbf{Y}_-\right)^{-\mathbf{1}} \mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\mathbf{Y}_- = \mathbf{0}.$$

By defining $D_h = \mathrm{diag}(1, \mathrm{h})$ and $Q_n = D_h^{-1}\mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\mathbf{Y}_-\mathbf{D}_{\mathbf{h}}^{-\mathbf{1}}$, the estimation errors are

$$\widehat{\varphi}_1(y_1) - \varphi_1(y_1) = [I_2 \otimes e_1^T Q_n^{-1}]\tau_n,$$

where

$$\tau_n = \left(I_2 \otimes D_h^{-1}\mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\right) \mathrm{vec}[\widetilde{\mathbf{Z}} - \iota\varphi_1^{\mathrm{T}}(\mathrm{y}_1) - \mathbf{Y}_-\nabla\varphi_{\mathbf{1}}^{\mathbf{T}}(\mathbf{y_1})].$$

Observing

$$\tau_n = \frac{1}{n}\sum_{k=d+1}^{n'} K_h^{\widehat{W}_k}(y_{k-1}-y_1)\left[\widetilde{z}_k - \varphi_1(y_1) - (y_{k-1}-y_1)\nabla\varphi_1(y_1)\right]\otimes(1,\frac{y_{k-1}-y_1}{h})^T,$$

where $K_h^{\widehat{W}_k}(y) = K_h(y)\widehat{W}_k$, it follows by adding and subtracting $z_k = \varphi_1(y_{k-1}) + H_2(\underline{y}_{k-1})$ that

$$\begin{aligned}\tau_n &= \frac{1}{n}\sum_{k=d+1}^{n'} K_h^{\widehat{W}_k}(y_{k-1}-y_1)\left[\widetilde{z}_k - z_k + H_2\left(\underline{y}_{k-1}\right)\right]\otimes(1,\frac{y_{k-1}-y_1}{h})^T \\ &+\frac{1}{n}\sum_{k=d+1}^{n'} K_h^{\widehat{W}_k}(y_{k-1}-y_1)\left[\varphi_1(y_{k-1}) - \varphi_1(y_1) - (y_{k-1}-y_1)\nabla\varphi_1(y_1)\right]\otimes(1,\frac{y_{k-1}-y_1}{h})^T.\end{aligned}$$

Due to the boundedness condition in C.2, the Taylor expansion applied to $[G_m(\widetilde{m}(x_k)),\quad G_v(\widetilde{v}(x_k))]$ at $[m(x_k), v(x_k)]$ yields the first term of τ_n as

$$\widetilde{\tau}_n \equiv \frac{1}{n}\sum_{k=d+1}^{n'} K_h^{\widehat{W}_k}(y_{k-1}-y_1)\left[\widetilde{u}_k\otimes(1,\frac{y_{k-1}-y_1}{h})^T\right],$$

where $\widetilde{u}_k \equiv \widetilde{z}_k^1 + \widetilde{z}_k^2 + H_2(\underline{y}_{k-1})$,

$$\begin{aligned}\widetilde{z}_k^1 &\equiv \left\{\nabla G_m(m(x_k))\left[\widetilde{m}(x_k) - m(x_k)\right], \nabla G_v(v(x_k))\left[\widetilde{v}(x_k) - v(x_k)\right]\right\}^T \\ \widetilde{z}_k^2 &\equiv \frac{1}{2}\left\{D^2G_m(m^*(x_k))\left[\widetilde{m}(x_k) - m(x_k)\right]^2, D^2G_v(v^*(x_k))\left[\widetilde{v}(x_k) - v(x_k)\right]^2\right\}^T,\end{aligned}$$

and $m^*(x_k)\,[v^*(x_k)]$ is between $\widetilde{m}(x_k)\,[\widetilde{v}(x_k)]$ and $m(x_k)[v(x_k)$, respectively]. In a similar way, the Taylor expansion of $\varphi_1(y_{k-1})$ at y_1 gives the second term of τ_n as

$$s_{0n} = \frac{h^2}{2}\frac{1}{n}\sum_{k=d+1}^{n'} K_h^{\widehat{W}_k}(y_{k-1}-y_1)\,(\frac{y_{k-1}-y_1}{h})^2[D^2\varphi_1(y_1)\otimes(1,\frac{y_{k-1}-y_1}{h})^T](1+o_p(1)).$$

$\widetilde{\tau}_n$ continues to be simplified by some further approximations. Define the marginal expectation of estimated density functions, $\widehat{p}_2(\cdot)$ and $\widehat{p}(\cdot)$ as follows

$$\begin{aligned}\overline{p}(y_{k-1},\underline{y}_{k-2}) &\equiv \int L_g(z_1-y_{k-1})L_g(z_2-\underline{y}_{k-2})p(z_1,z_2)dz_1dz_2, \\ \overline{p}_2(\underline{y}_{k-2}) &\equiv \int L_g(z_2-\underline{y}_{k-2})p_2(z_2)dz_2.\end{aligned}$$

In the first approximation, we replace the estimated instrument, $\widehat{W}$, by the ratio of the expectations of the kernel density estimates, $\overline{p}_2(\underline{y}_{k-1})/\overline{p}(x_k)$, and deal with the linear terms in the Taylor expansions. That is, $\widetilde{\tau}_n$ is approximated with an error of $o_p\left(1/\sqrt{nh}\right)$ by $t_{1n}+t_{2n}$:

$$\begin{aligned}t_{1n} &\equiv \frac{1}{n}\sum_{k=d+1}^{n'} K_h(y_{k-1}-y_1)\frac{\overline{p}_2(\underline{y}_{k-1})}{\overline{p}(x_k)}[\widetilde{z}_k^1\otimes(1,\frac{y_{k-1}-y_1}{h})^T], \\ t_{2n} &\equiv \frac{1}{n}\sum_{k=d+1}^{n'} K_h(y_{k-1}-y_1)\frac{\overline{p}_2(\underline{y}_{k-1})}{\overline{p}(x_k)}[H_2(\underline{y}_{k-1})\otimes(1,\frac{y_{k-1}-y_1}{h})^T],\end{aligned}$$

based on the following results:

$(i)\frac{1}{n}\sum_{k=d+1}^{n'} K_h\left(y_{k-1}-y_1\right)\frac{\widehat{p}_2(\underline{y}_{k-1})}{\widehat{p}(x_k)}[\widetilde{z}_k^2\otimes(1,\frac{y_{k-1}-y_1}{h})^T]=o_p\left(\frac{1}{\sqrt{nh}}\right)$,

$(ii)\ \frac{1}{n}\sum_{k=d+1}^{n'} K_h\left(y_{k-1}-y_1\right)[\frac{\widehat{p}_2(\underline{y}_{k-1})}{\widehat{p}(x_k)}-\frac{\overline{p}_2(\underline{y}_{k-1})}{\overline{p}(x_k)}][H_2(\underline{y}_{k-1})\otimes(1,\frac{y_{k-1}-y_1}{h})^T]=o_p\left(\frac{1}{\sqrt{nh}}\right)$,

$(iii)\ \frac{1}{n}\sum_{k=d+1}^{n'} K_h\left(y_{k-1}-y_1\right)[\frac{\widehat{p}_2(\underline{y}_{k-1})}{\widehat{p}(x_k)}-\frac{\overline{p}_2(\underline{y}_{k-1})}{\overline{p}(x_k)}][\widetilde{z}_k^1\otimes(1,\frac{y_{k-1}-y_1}{h})^T]=o_p\left(\frac{1}{\sqrt{nh}}\right)$.

To show (i), consider the first two elements of the term, for example, which are bounded elementwise by

$$\begin{aligned}&\sup_k\left|\widetilde{m}\left(x_k\right)-m\left(x_k\right)\right|^2\times\\&\frac{1}{2}\frac{1}{n}\sum_k K_h\left(y_{k-1}-y_1\right)\frac{\widehat{p}_2(\underline{y}_{k-1})}{\widehat{p}(x_k)}D^2G_m\left(m\left(x_k\right)\right)(1,\frac{y_{k-1}-y_1}{h})^T\\=\ &o_p\left(1/\sqrt{nh}\right).\end{aligned}$$

The last equality is direct from the uniform convergence theorems in Masry (1992) that

$$\sup_t\left|\widetilde{m}\left(x_t\right)-m\left(x_t\right)\right|=O_p\left(\log n/\sqrt{nh^d}\right), \tag{A.A.1}$$

and $\frac{1}{n}\sum_k K_h\left(y_{k-1}-y_1\right)\frac{\widehat{p}_2(\underline{y}_{k-1})}{\widehat{p}(x_k)}D^2G_m\left(m\left(x_k\right)\right)(1,\frac{y_{k-1}-y_1}{h})^T=O_p\left(1\right)$. The proof for (ii) is given in Lemma A.1. The negligibility of (iii) follows in a similar way from (ii), considering (). While the asymptotic properties of s_{0n} and t_{2n} are relatively easy to derive, additional approximation is necessary to make t_{1n} more tractable. Note that the estimation errors of local linear fits, $\widetilde{m}\left(x_k\right)-m\left(x_k\right)$ of $\widetilde{z}_k^1$, are decomposed into

$$\frac{1}{n}\sum_l\frac{K_h\left(x_l-x_k\right)}{p(x_l)}v^{1/2}\left(x_l\right)\varepsilon_l+\text{ the remaining bias}$$

from the approximation results for local linear smoother in Jones, Davies and Park(1994). A similar expression holds for volatility estimates, $\widetilde{v}\left(x_k\right)-v\left(x_k\right)$, with a stochastic term of $\frac{1}{n}\sum_l\frac{K_h(x_l-x_k)}{p(x_l)}v\left(x_l\right)\left(\varepsilon_l^2-1\right)$. Define

$$\begin{aligned}&J_{k,n}\left(x_l\right)\\\equiv\ &\frac{1}{nh^d}\sum_k\frac{K\left(y_{k-1}-y_1/h\right)K\left(x_l-x_k/h\right)}{p(x_l)}\frac{p_2(\underline{y}_{k-1})}{p(x_k)}[\mathrm{diag}(\nabla\mathrm{G_m},\nabla\mathrm{G_v})\otimes(1,\frac{\mathrm{y_{k-1}-y_1}}{\mathrm{h}})^{\mathrm{T}}],\end{aligned}$$

and let $\overline{J}\left(x_l\right)$ denote the marginal expectation of $J_{k,n}$ w.r.t. x_k. Then, the stochastic term of t_{1n}, after rearranging its the double sums, is approximated by

$$\widetilde{t}_{1n}=\frac{1}{nh}\sum_l\overline{J}\left(x_l\right)[(v^{1/2}\left(x_l\right)\varepsilon_l,v\left(x_l\right)\left(\varepsilon_l^2-1\right))^T\otimes I_2],$$

since the approximation errors from $\overline{J}\left(X_l\right)$ is negligible, i.e.,

$$\frac{1}{nh}\sum_l(J_{k,n}-\overline{J})[(v^{1/2}\left(X_l\right)\varepsilon_l,v\left(X_l\right)\left(\varepsilon_l^2-1\right))^T\otimes I_2]^T=o_p\left(1/\sqrt{nh}\right),$$

applying the same method as in Lemma A.1. A straightforward calculation gives

$$\begin{aligned}
\overline{J}(X_l) &\simeq \frac{1}{h}\int K(u_1 - y_1/h)\, K(u_1 - y_{l-1}/h) \int \frac{1}{h^{d-1}} K\left(\underline{y}_{l-1} - u_2/h\right) \frac{p_2(u_2)}{p(x_l)} \times \\
&\quad [\mathrm{diag}(\nabla \mathrm{G_m(u)}, \nabla \mathrm{G_v(u)}) \otimes (1, \frac{\mathrm{u_1 - y_1}}{\mathrm{h}})^{\mathrm{T}}] \mathrm{du_2 du_1} \\
&\simeq \frac{1}{h}\int K(u_1 - y_1/h)\, K(u_1 - y_{l-1}/h) \frac{p_2(\underline{y}_{l-1})}{p(x_l)} \times \\
&\quad [\mathrm{diag}(\nabla \mathrm{G_m(u_1, \underline{y}_{l-1})}, \nabla \mathrm{G_v(u_1, \underline{y}_{l-1})}) \otimes (1, \frac{\mathrm{u_1 - y_1}}{\mathrm{h}})^{\mathrm{T}}] \mathrm{du_1} \\
&\simeq \frac{p_2(\underline{y}_{l-1})}{p(x_l)} [\mathrm{diag}(\nabla \mathrm{G_m(y_1, \underline{y}_{l-1})}, \nabla \mathrm{G_v(y_1, \underline{y}_{l-1})}) \\
&\quad \otimes ((K*K)_0\left(\frac{y_{l-1} - y_1}{h}\right), (K*K)_1\left(\frac{y_{l-1} - y_1}{h}\right))^T],
\end{aligned}$$

where

$$(K*K)_i\left(\frac{y_{l-1} - y_1}{h}\right) = \int w_1^i K(w_1)\, K\left(w_1 + \frac{y_{l-1} - y_1}{h}\right) dw.$$

Observe that $(K*K)_i\left(\frac{y_{l-1}-y_1}{h}\right)$ in $\overline{J}(X_l)$ is actually a convolution kernel and behaves just like a one dimensional kernel function of y_{l-1}. This means that the standard method (CLT, or LLN) for univariate kernel estimates can be applied to show the asymptotics of

$$\widetilde{t}_{1n} = \frac{1}{nh}\sum_l \frac{p_2(\underline{y}_{l-1})}{p(x_l)} \left\{ \begin{bmatrix} \nabla G_m(y_1, \underline{y}_{l-1}) v^{1/2}(X_l)\, \varepsilon_l \\ \nabla G_v(y_1, \underline{y}_{l-1}) v(X_l)\, (\varepsilon_l^2 - 1)) \end{bmatrix} \otimes \begin{bmatrix} (K*K)_0\left(\frac{y_{l-1}-y_1}{h}\right) \\ (K*K)_1\left(\frac{y_{l-1}-y_1}{h}\right) \end{bmatrix} \right\}.$$

If we define $\widetilde{s}_{1n}$ as the remaining bias term of t_{1n}, the estimation errors of $\widehat{\varphi}_1(y_1) - \varphi_1(y_1)$, consist of two stochastic terms, $[I_2 \otimes e_1^T Q_n^{-1}]\left(\widetilde{t}_{1n} + \widetilde{t}_{2n}\right)$, and three bias terms, $[I_2 \otimes e_1^T Q_n^{-1}](s_{0n} + s_{1n} + s_{2n})$, where

$$\begin{aligned}
\widetilde{t}_{2n} &= \frac{1}{n}\sum_{k=d+1}^{n'} K_h(y_{k-1} - y_1) \frac{p_2(\underline{y}_{k-1})}{p(X_k)} [H_2(\underline{y}_{k-1}) \otimes (1, \frac{Y_{k-1} - y_1}{h})^T], \\
s_{2n} &= t_{2n} - \widetilde{t}_{2n}.
\end{aligned}$$

Step II: Computation of Variance and Bias

We strat with showing the order of the main stochastic term,

$$\widetilde{t}_n^* = \widetilde{t}_{1n} + \widetilde{t}_{2n} = \frac{1}{n}\sum_k \xi_k,$$

where $\xi_k = \xi_{1k} + \xi_{2k}$,

$$\begin{aligned}
\xi_{1k} &= \frac{p_2(\underline{y}_{k-1})}{p(y_{k-1}, \underline{y}_{k-1})} \left\{ \begin{bmatrix} \nabla G_m(y_1, \underline{y}_{k-1}) v^{1/2}(X_k)\, \varepsilon_k \\ \nabla G_v(y_1, \underline{y}_{k-1}) v(X_k)\, (\varepsilon_k^2 - 1)) \end{bmatrix} \otimes \begin{bmatrix} \frac{1}{h}(K * K)_0 \left(\frac{y_{k-1} - y_1}{h}\right) \\ 0 \end{bmatrix} \right\} \\
\xi_{2k} &= \frac{p_2(\underline{y}_{k-1})}{p(y_{k-1}, \underline{y}_{k-1})} \left\{ \begin{bmatrix} m_2\left(\underline{y}_{k-1}\right) \\ v_2\left(\underline{y}_{k-1}\right) \end{bmatrix} \otimes \begin{bmatrix} \frac{1}{h} K\left(\frac{y_{k-1} - y_1}{h}\right) \\ \frac{1}{h} K\left(\frac{y_{k-1} - y_1}{h}\right)\left(\frac{y_{l-1} - y_1}{h}\right) \end{bmatrix} \right\},
\end{aligned}$$

by calculating its asymptotic variance. Dividing a normalized variance of $\widetilde{t}_n^*$ into the sums of variances and covariances gives

$$\begin{aligned}
\mathrm{var}\left(\sqrt{\mathrm{nh}}\widetilde{\mathrm{t}}_{\mathrm{n}}^*\right) &= \mathrm{var}\left(\frac{\sqrt{\mathrm{h}}}{\sqrt{\mathrm{n}}} \sum_{\mathrm{k}} \xi_{\mathrm{k}}\right) = \frac{\mathrm{h}}{\mathrm{n}} \sum_{\mathrm{k}} \mathrm{var}\left(\xi_{\mathrm{k}}\right) + \frac{\mathrm{h}}{\mathrm{n}} \mathrm{k} \neq \mathrm{l} \sum \sum \mathrm{cov}\left(\xi_{\mathrm{k}}, \xi_{\mathrm{l}}\right) \\
&= h\mathrm{var}\left(\widetilde{\xi}_{\mathrm{k}}\right) + \sum_{\mathrm{k}} \left[\frac{\mathrm{n} - \mathrm{k}}{\mathrm{n}}\right] \mathrm{h}\left[\mathrm{cov}\left(\xi_{\mathrm{d}}, \xi_{\mathrm{d+k}}\right)\right],
\end{aligned}$$

where the last equality comes from the stationarity assumption.

We claim that

(a) $h\mathrm{var}\left(\xi_{\mathrm{k}}\right) \longrightarrow \Sigma_1(\mathrm{y}_1)$,

(b) $\sum_k \left[1 - \frac{k}{n}\right] h\mathrm{cov}\left(\xi_{\mathrm{d}}, \xi_{\mathrm{d+k}}\right) = \mathrm{o}(1)$, and

(c) $nh\mathrm{var}\left(\widetilde{\mathrm{t}}_{\mathrm{n}}^*\right) \longrightarrow \Sigma_1(\mathrm{y}_1)$,

where

$$\begin{aligned}
\Sigma_1(y_1) &= \{\int \frac{p_2^2(z_2)}{p(y_1, z_2)} \begin{bmatrix} \nabla G_m(y_1, z_2)^2 v(y_1, z_2) & (\nabla G_m \cdot \nabla G_v)(\kappa_3 \cdot v^{3/2})(y_1, z_2) \\ (\nabla G_m \cdot \nabla G_v)(\kappa_3 \cdot v^{3/2})(y_1, z_2) & \nabla G_v(y_1, z_2)^2 \kappa_4(y_1, z_2) v^2(y_1, z_2) \end{bmatrix} dz_2 \\
&\quad \otimes \begin{bmatrix} ||(K * K)_0||_2^2 & 0 \\ 0 & 0 \end{bmatrix}\} \\
&\quad + \int \frac{p_2^2(z_2)}{p(y_1, z_2)} H_2(z_2)\, H_2^T(z_2)\, dz_2 \otimes \begin{bmatrix} ||K||_2^2 & 0 \\ 0 & \int K^2(u)\, u^2 du \end{bmatrix}
\end{aligned}$$

Proof of (a). Noting $E\left(\xi_{1k}\right) = E\left(\xi_{2k}\right) = 0_{4\times 1}$ and $E\left(\xi_{1k}\xi_{2k}^T\right) = 0_{4\times 4}$,

$$h\mathrm{var}\left(\xi_{\mathrm{k}}\right) = \mathrm{hE}\left(\xi_{1\mathrm{k}}\xi_{1\mathrm{k}}^{\mathrm{T}}\right) + \mathrm{hE}\left(\xi_{2\mathrm{k}}\xi_{2\mathrm{k}}^{\mathrm{T}}\right),$$

by the stationarity assumption. Applying the integration with substitution of variable and Taylor expansion, the expectation term is

$$\begin{aligned}
hE\left(\xi_{1k}\xi_{1k}^T\right) &= \{\int \frac{p_2^2(z_2)}{p(y_1, z_2)} \begin{bmatrix} \nabla G_m(y_1, z_2)^2 v(y_1, z_2) & (\nabla G_m \cdot \nabla G_v)(\kappa_3 \cdot v^{3/2})(y_1, z_2) \\ (\nabla G_m \cdot \nabla G_v)(\kappa_3 \cdot v^{3/2})(y_1, z_2) & \nabla G_v(y_1, z_2)^2 \kappa_4(y_1, z_2) v^2(y_1, z_2) \end{bmatrix} dz_2 \\
&\quad \otimes \begin{bmatrix} ||(K * K)_0||_2^2 & 0 \\ 0 & 0 \end{bmatrix}\},
\end{aligned}$$

and

$$hE\left(\xi_{2k}\xi_{2k}^{T}\right)=\int\frac{p_{2}^{2}(z_{2})}{p(y_{1},z)}\left[\begin{array}{cc} m_{2}^{2}\left(z_{2}\right)dz_{2} & m_{2}\left(z_{2}\right)v_{2}\left(z_{2}\right)dz_{2} \\ m_{2}\left(z_{2}\right)v_{2}\left(z_{2}\right)dz_{2} & v_{2}^{2}\left(z_{2}\right)dz_{2}\end{array}\right]\otimes\left[\begin{array}{cc} ||K||_{2}^{2} & 0 \\ 0 & \int K^{2}\left(u\right)u^{2}du\end{array}\right]\}+o\left(1\right),$$

where $\kappa_{3}(y_{1},z_{2})=E[\varepsilon_{t}^{3}|x_{t}=(y_{1},z_{2})]$ and $\kappa_{4}(y_{1},z_{2})=E[(\varepsilon_{t}^{2}-1)^{2}|x_{t}=(y_{1},z_{2})]$.

Proof of (b). Since $E\left(\xi_{1k}\xi_{1j}^{T}\right)|_{j\neq k}=E\left(\xi_{1k}\xi_{2j}^{T}\right)|_{j\neq k}=0$, $\mathrm{cov}\left(\xi_{\mathrm{d}+1},\xi_{\mathrm{d}+1+\mathrm{k}}\right)=\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right)$. By setting $c(n)h\rightarrow0$, as $n\rightarrow\infty$, we separate the covariance terms into two parts:

$$\sum_{k=1}^{c(n)}\left[1-\frac{k}{n}\right]h\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right)+\sum_{\mathrm{k=c(n)+1}}^{\mathrm{n}'}\left[1-\frac{\mathrm{k}}{\mathrm{n}}\right]\mathrm{hcov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right).$$

To show the negligibility of the first part of covariances, consider that the dominated convergence theorem used after Taylor expansion and the integration with substitution of variables gives

$$\begin{aligned}&\left|\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right)\right| \\ \simeq\ &\left|\int H_{2}\left(\underline{y}_{d}\right)H_{2}^{T}\left(\underline{y}_{d+k}\right)\frac{p(y_{1},\underline{y}_{d},y_{1},\underline{y}_{d+k})}{p_{1|2}(y_{1}|\underline{y}_{d})p_{1|2}(y_{1}|\underline{y}_{d+k})}d(\underline{y}_{d},\underline{y}_{d+k})\right|\otimes\left[\begin{array}{cc}1 & 0\\ 0 & 0\end{array}\right].\end{aligned}$$

Therefore, it follows from the assumption on the boundedness condition in C.2 that

$$\begin{aligned}\left|\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right)\right| &\leq E\left|H_{2}\left(\underline{y}_{d}\right)\right|E\left|H_{2}^{T}\left(\underline{y}_{d+k}\right)\right|\int\frac{p(y_{1},\underline{y}_{d},y_{1},\underline{y}_{d+k})}{p_{1|2}(y_{1}|\underline{y}_{d})p_{1|2}(y_{1}|\underline{y}_{d+k})}d(\underline{y}_{d},\underline{y}_{d+k})\otimes\left[\begin{array}{cc}1 & 0\\ 0 & 0\end{array}\right] \\ &\equiv A^{*}.\end{aligned}$$

where '$A\leq B$' mean $a_{ij}\leq b_{ij}$, for all element of matrices A and B. By the construction of $c(n)$,

$$\begin{aligned}&\sum_{k=1}^{c(n)}\left[1-\frac{k}{n}\right]h\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right) \\ \leq\ &2c(n)\left|h\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right)\right|\leq2c(n)hA^{*}\longrightarrow0,\ \text{as } n\rightarrow\infty.\end{aligned}$$

Next, we turn to the negligibility of the second part of the covariances,

$$\sum_{k=c(n)+1}^{n'}\left[1-\frac{k}{n}\right]h\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right).$$

Let ξ_{2k}^{i} be i-the element of ξ_{2k}, for $i=1,\ldots,4$. Using Davydov's lemma (in Hall and Heyde 1980, Theorem A.5), we obtain

$$\left|h\mathrm{cov}\left(\xi_{2\mathrm{d}+1}^{\mathrm{i}},\xi_{2\mathrm{d}+1+\mathrm{k}}^{\mathrm{j}}\right)\right|=\left|\mathrm{cov}\left(\sqrt{\mathrm{h}}\xi_{2\mathrm{d}+1}^{\mathrm{i}},\sqrt{\mathrm{h}}\xi_{2\mathrm{d}+1+\mathrm{k}}^{\mathrm{j}}\right)\right|\leq8\left[\alpha(k)^{1-2/v}\right]\left[\max_{i=1,\ldots,4}E\left(\sqrt{h}\left|\xi_{2k}^{i}\right|^{v}\right)\right]^{2/v},$$

for some $v>2$. The boundedness of $E(\sqrt{h}\left|\xi_{2k}^{1}\right|^{v})$, for example, is evident from the direct calculation that

$$\xi_{2k}=\frac{p_{2}(\underline{y}_{k})}{p(x_{k})}\left\{\left[\begin{array}{c}m_{2}\left(\underline{y}_{d}\right)\\ v_{2}\left(\underline{y}_{d}\right)\end{array}\right]\otimes\left[\begin{array}{c}\frac{1}{h}K\left(\frac{y_{k-1}-y_{1}}{h}\right)\\ \frac{1}{h}K\left(\frac{y_{k-1}-y_{1}}{h}\right)\left(\frac{y_{k-1}-y_{1}}{h}\right)\end{array}\right]\right\}$$

$$E\left(\left|\sqrt{h}\xi_{2k}^{1}\right|^{v}\right) \simeq \frac{h^{v/2}}{h^{v-1}}\int \frac{p_2^v(z_2)}{p^{v-1}(y_1,z_2)}\left|m_2^v(z_2)\right|dz_2$$
$$= O(\frac{h^{v/2}}{h^{v-1}}) = O(\frac{1}{h^{v/2-1}}).$$

Thus, the covariance is bounded by

$$|h\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right)| \leq C\left[\frac{1}{h^{v/2-1}}\right]^{2/v}\left[\alpha(k)^{1-2/v}\right].$$

This implies

$$\sum_{k=c(n)+1}^{n'}\left[1-\frac{k}{n}\right]h\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right)$$
$$\leq 2\sum_{k=c(n)+1}^{\infty}|h\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right)| \leq C'\left[\frac{1}{h^{1-2/v}}\right]\sum_{k=c(n)+1}^{\infty}\left[\alpha(k)^{1-2/v}\right]$$
$$= C'\sum_{k=c(n)+1}^{\infty}\left[\frac{1}{h^{1-2/v}}\right]\left[\alpha(k)^{1-2/v}\right] \leq C'\sum_{k=c(n)+1}^{\infty}k^a\left[\alpha(k)^{1-2/v}\right],$$

if a is such that

$$k^a \geq (c(n)+1)^a \geq c(n)^a = \frac{1}{h^{1-2/v}},$$

for example,

$$c(n)^a h^{1-2/v} = 1,$$

which implies

$$c(n) \to \infty.$$

If we further restrict a such that

$$0 < a < 1 - \frac{2}{v},$$

then,

$$c(n)^a h^{1-2/v} = 1 \text{ implies}$$
$$c(n)^a h^{1-2/v} = [\,c(n)h]^{1-2/v}\,c(n)^{-\delta} = 1, \quad \text{for } \delta > 0.$$

Thus,

$$c(n)h \to 0$$

as required. Therefore,

$$\sum_{k=c(n)+1}^{n'}\left[1-\frac{k}{n}\right]h\mathrm{cov}\left(\xi_{2\mathrm{d}+1},\xi_{2\mathrm{d}+1+\mathrm{k}}\right) \leq \mathrm{C}'\sum_{\mathrm{k=c(n)+1}}^{\infty}\mathrm{k^a}\left[\alpha(\mathrm{k})^{1-2/\mathrm{v}}\right] \to 0,$$

as n goes to ∞.

The proof of (c) is immediate from (a) and (b).

Next, we consider the asymptotic bias. Using the standard result on kernel weighted sum of stationary series, we first get,

$$s_{0n} \xrightarrow{p} \frac{h^2}{2}[D^2\varphi_1(y_1) \otimes (\mu_K^2, 0)^T],$$

since

$$\begin{aligned}
& \frac{1}{n}\sum_{k=d+1}^{n'} K_h^{\widehat{W}}(y_{k-1}-y_1)\left(\frac{y_{k-1}-y_1}{h}\right)^2[D^2\varphi_1(y_1)\otimes(1,\frac{y_{k-1}-y_1}{h})^T] \\
\to & \ {}^p\int K_h^{\widehat{W}}(z_1-y_1)\left(\frac{z_1-y_1}{h}\right)^2[D^2\varphi_1(y_1)\otimes(1,\frac{z_1-y_1}{h})^T]p(z)\,dz \\
\simeq & \int K_h(z_1-y_1)\,p_2(z_2)\left(\frac{z_1-y_1}{h}\right)^2[D^2\varphi_1(y_1)\otimes(1,\frac{z_1-y_1}{h})^T]dz \\
= & \int K_h(z_1-y_1)\left(\frac{z_1-y_1}{h}\right)^2[D^2\varphi_1(y_1)\otimes(1,\frac{z_1-y_1}{h})^T]dz_1 \\
= & [D^2\varphi_1(y_1)\otimes\int K_h(z_1-y_1)\left(\frac{z_1-y_1}{h}\right)^2(1,\frac{z_1-y_1}{h})^T dz_1] \\
= & [D^2\varphi_1(y_1)\otimes(\mu_K^2,0)^T].
\end{aligned}$$

For the asymptotic bias of $\widetilde{s}_{1n}$, we again use the approximation results in Jones, Davies and Park(1994). Then, the first component of $\widetilde{s}_{1n}$, for example, is

$$\frac{1}{n}\sum_k K_h(y_{k-1}-y_1)\frac{p_2(\underline{y}_{k-1})}{p(x_k)}\nabla G_m(m(x_k))\{\frac{1}{2}\frac{1}{n}\sum_l \frac{K_h(x_l-x_k)}{p(x_l)}\sum_{\alpha=1}^{d}(y_{l-\alpha}-y_{k-\alpha})^2\frac{\partial^2 m(x_k)}{\partial y_{k-\alpha}^2}\},$$

and converges to

$$\frac{h^2}{2}\int p_2(z_2)\nabla G_m(m(y_1,z_2))[\mu_{K*K}^2 D^2 m_1(y_1)+\mu_K^2 D^2 m_2(z_2)]dz_2,$$

based on the argument for convolution kernel in the above. A convolution of symmetric kernels is symmetric, so that $\int (K*K)_0(u)\,udu = 0$, and $\int (K*K)_1(u)\,u^2du = \int\int wK(w)K(w+u)\,u^2dwdu = 0$. This implies that

$$\widetilde{s}_{1n} \xrightarrow{p} \frac{h^2}{2}\int p_2(z_2)\{[\nabla G_m(m(y_1,z_2)),\nabla G_v(v(y_1,z_2))]^T\odot[\mu_{K*K}^2 D^2\varphi_1(y_1)+\mu_K^2 D^2\varphi_2(z_2)]\}\otimes(1,0)^T dz_2.$$

To calculate $\widetilde{s}_{2n}$, we use the Taylor series expansion of $\frac{\overline{p}_2(\underline{y}_{k-1})}{\overline{p}(X_k)}$:

$$\begin{aligned}
& \left[\overline{p}_2(\underline{y}_{k-1})-\frac{p_2(\underline{y}_{k-1})\overline{p}(X_k)}{p(X_k)}\right]\frac{1}{\overline{p}(X_k)} \\
= & \left[\overline{p}_2(\underline{y}_{k-1})-\frac{p_2(\underline{y}_{k-1})\overline{p}(X_k)}{p(X_k)}\right]\frac{1}{p(X_k)}\times\left[1-\frac{\overline{p}(X_k)-p(X_k)}{p^2(X_k)}+\ldots\right] \\
= & \frac{\overline{p}_2(\underline{y}_{k-1})}{p(X_k)}-\frac{p_2(\underline{y}_{k-1})\overline{p}(X_k)}{p^2(X_k)}+o_p(1).
\end{aligned}$$

Thus,

$$\widetilde{s}_{2n} = \frac{1}{n}\sum_{k=d+1}^{n'} K_h\left(y_{k-1}-y_1\right)\left[\frac{\overline{p}_2(\underline{y}_{k-1})}{\overline{p}(X_k)} - \frac{p_2(\underline{y}_{k-1})}{p(X_k)}\right]\left[H_2(\underline{y}_{k-1})\otimes\left(1,\frac{y_{k-1}-y_1}{h}\right)^T\right]$$

$$\begin{aligned}
&\xrightarrow{p} \int K_h\left(z_1-y_1\right)\left[\frac{\overline{p}_2(z_2)}{\overline{p}(z)} - \frac{p_2(z_2)}{p(z)}\right]\left[H_2(z_2)\otimes\left(1,\frac{z_1-y_1}{h}\right)^T\right]p\left(z\right)dz \\
\simeq\ & \int K_h\left(z_1-y_1\right)\left[\frac{\overline{p}_2(z_2)}{p(z)} - \frac{p_2(z_2)\overline{p}(z)}{p^2\left(z\right)}\right]\left[H_2(z_2)\otimes\left(1,\frac{z_1-y_1}{h}\right)^T\right]p\left(z\right)dz \\
=\ & \int K_h\left(z_1-y_1\right)\left[\frac{\overline{p}_2(z_2)}{p(z)} - \frac{p_2(z_2)}{p\left(z\right)}\right]\left[H_2(z_2)\otimes\left(1,\frac{z_1-y_1}{h}\right)^T\right]p\left(z\right)dz \\
&+\int K_h\left(z_1-y_1\right)\left[\frac{p_2(z_2)p(z)}{p^2(z)} - \frac{p_2(z_2)\overline{p}(z)}{p^2\left(z\right)}\right]\left[H_2(z_2)\otimes\left(1,\frac{z_1-y_1}{h}\right)^T\right]p\left(z\right)dz \\
\simeq\ & \frac{g^2}{2}\left[\int D^2 p_2\left(z_2\right)H_2(z_2)dz_2\otimes\left(\mu_K^2,0\right)^T\right] \\
&-\frac{g^2}{2}\left[\int \frac{p_2(z_2)}{p\left(y_1,z_2\right)}D^2 p(y_1,z_2)H_2(z_2)dz_2\otimes\left(\mu_K^2,0\right)^T\right].
\end{aligned}$$

Finally, for the probability limit of $[I_2\otimes e_1^T Q_n^{-1}]$, we note that $Q_n = D_h^{-1}\mathbf{Y}_-^{\mathbf{T}}\mathbf{K}\mathbf{Y}_-\mathbf{D}_\mathbf{h}^{-\mathbf{1}} = [\widehat{\mathbf{q}}_{\mathbf{ni+j-2}}(\mathbf{y_1};\mathbf{h})]_{\mathbf{(i,j)=1,2}}$ with $\widehat{q}_{ni} = \frac{1}{n}\sum_{k=d}^{n} K_h^{\widehat{W}}\left(Y_{k-1}-y_1\right)\left(\frac{y_{k-1}-y_1}{h}\right)^i$, for $i=0,1,2$, and

$$\begin{aligned}
\widehat{q}_{ni} \xrightarrow{p} \int K_h\left(z_1-y_1\right)\left(\frac{z_1-y_1}{h}\right)^i p_2\left(z_2\right)dz &= \int K\left(u_1\right)u_1^i du_1\int p_2\left(z_2\right)dz_2 \\
&= \int K\left(u_1\right)u_1^i du_1 \equiv q_i,
\end{aligned}$$

where $q_0=1$, $q_1=0$ and $q_2=\mu_K^2$.

Thus, $Q_n \to \begin{bmatrix} 1 & 0 \\ 0 & \mu_K^2 \end{bmatrix}$, $Q_n^{-1}\to \frac{1}{\mu_K^2}\begin{bmatrix} \mu_K^2 & 0 \\ 0 & 1 \end{bmatrix}$, and $e_1^T Q_n^{-1}\to e_1^T$. Therefore,

$$\begin{aligned}
B_{1n}(y_1) &= \left[I_2\otimes e_1^T Q_n^{-1}\right]\left(s_{0n}+s_{1n}+s_{2n}\right) \\
&= \frac{h^2}{2}\mu_K^2 D^2\varphi_1(y_1) \\
&\quad \frac{h^2}{2}\int\left[\mu_{K*K}^2 D^2\varphi_1(y_1)+\mu_K^2 D^2\varphi_2(z_2)\right]\odot\left[\nabla G_m(m(y_1,z_2)),\nabla G_v(v(y_1,z_2))\right]^T p_2(z_2)dz_2 \\
&\quad +\frac{g^2}{2}\mu_K^2\int Dp_2\left(z_2\right)H_2(z_2)dz_2 - \frac{g^2}{2}\mu_K^2\int\frac{p_2(z_2)}{p\left(y_1,z_2\right)}D^2 p(y_1,z_2)H_2(z_2)dz_2 \\
&\quad +o_p\left(h^2\right)+o_p\left(g^2\right).
\end{aligned}$$

Step III: Asymptotic Normality of $\widetilde{t}_n^*$

Applying the Cramer-Wold device, it is sufficient to show

$$D_n \equiv \frac{1}{\sqrt{n}} \sum_k \sqrt{h}\widetilde{\xi}_k \xrightarrow{\mathcal{D}} N\left(0, \beta^T \Sigma_1 \beta\right),$$

for all $\beta \in \mathbb{R}^{\not 2}$, where $\widetilde{\xi}_k = \beta^T \xi_k$. We use the small block-large block argument-see Masry and Tjøstheim (1997). Partition the set $\{d, d+1, \ldots, n\}$ into $2k+1$ subsets with large blocks of size $r = r_n$ and small blocks of size $s = s_n$ where

$$k = \left[\frac{n_1}{r_n + s_n}\right]$$

and $[x]$ denotes the integer part of x. Define

$$\begin{aligned} \eta_j &= \sum_{t=j(r+s)}^{j(r+s)+r-1} \sqrt{h}\widetilde{\xi}_t, \quad \omega_j = \sum_{t=j(r+s)+r}^{(j+1)(r+s)-1} \sqrt{h}\widetilde{\xi}_t, \quad 0 \le j \le k-1, \\ \varsigma_k &= \sum_{t=k(r+s)}^{n} \sqrt{h}\widetilde{\xi}_t, \end{aligned}$$

then,

$$D_n = \frac{1}{\sqrt{n}} \left(\sum_{j=0}^{k-1} \eta_j + \sum_{j=0}^{k-1} \omega_j + \varsigma_k \right) \equiv \frac{1}{\sqrt{n}} \left(S_n' + S_n'' + S_n''' \right).$$

Due to C.6., there exist a sequence $a_n \to \infty$ such that

$$a_n s_n = o\left(\sqrt{nh}\right) \text{ and } a_n \sqrt{n/h}\alpha\left(s_n\right) \to 0, \; as \; n \to \infty, \tag{A.A.2}$$

and define the large block size as

$$r_n = \left[\frac{\sqrt{nh}}{a_n}\right]. \tag{A.A.3}$$

It is easy to show by (A.A.2) and (A.A.3) that as $n \to \infty$:

$$\frac{r_n}{n} \to 0, \; \frac{s_n}{r_n} \to 0, \; \frac{r_n}{\sqrt{nh}} \to 0, \tag{A.A.4}$$

and

$$\frac{n}{r_n}\alpha\left(s_n\right) \to 0.$$

We first show that S_n'' and S_n''' are asymptotically negligible. The same argument used in Step II yields

$$\begin{aligned} \mathrm{var}\left(\omega_j\right) &= s \times \mathrm{var}\left(\sqrt{h}\widetilde{\xi}_t\right) + 2s \sum_{k=1,}^{s-1} \left(1 - \frac{k}{s}\right) \mathrm{cov}\left(\sqrt{h}\widetilde{\xi}_{d+1}, \sqrt{h}\widetilde{\xi}_{d+1+k}\right) \qquad \text{(A.A.5)} \\ &= s\beta^T \Sigma_1 \beta \left(1 + o\left(1\right)\right), \end{aligned}$$

which implies

$$\sum_{j=0}^{k-1} \mathrm{var}\left(\omega_j\right) = O\left(ks\right) \sim \frac{ns_n}{r_n + s_n} \sim \frac{ns_n}{r_n} = o\left(n\right),$$

from the condition (A.A.4). Next, consider

$$\sum_{\substack{i,j=0,\\ i\neq j}}^{k-1} \operatorname{cov}\left(\omega_{\mathrm{i}},\omega_{\mathrm{j}}\right)=\sum_{\substack{i,j=0,\\ i\neq j}}^{\mathrm{k}-1}\sum_{\mathrm{k}_1=1}^{\mathrm{s}}\sum_{\mathrm{k}_2=1}^{\mathrm{s}}\operatorname{cov}\left(\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{N_i}+\mathrm{k}_1},\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{N_j}+\mathrm{k}_2}\right),$$

where $N_j=j(r+s)+r$. Since $|N_i-N_j+k_1-k_2|\geq r$, for $i\neq j$, the covariance term is bounded by

$$\begin{aligned}&2\sum_{k_1=1}^{n-r}\sum_{k_2=k_1+r}^{n}\left|\operatorname{cov}\left(\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{k}_1},\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{k}_2}\right)\right|\\ \leq\ & 2n\sum_{j=r+1}^{n}\left|\operatorname{cov}\left(\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{d}+1},\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{d}+1+\mathrm{j}}\right)\right|=o\left(n\right).\end{aligned}$$

The last equality also follows from Step II. Hence, $\frac{1}{n}E\left\{(S_n'')^2\right\}\to 0$, as $n\to\infty$. Repeating a similar argument for S_n''', we get

$$\begin{aligned}\frac{1}{n}E\left\{(S_n''')^2\right\} &\leq \frac{1}{n}\left[n-k\left(r+s\right)\right]\operatorname{var}\left(\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{d}+1}\right)\\ &\quad +2\frac{n-k\left(r+s\right)}{n}\sum_{j=1}^{n-k(r+s)}\operatorname{cov}\left(\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{d}+1},\sqrt{\mathrm{h}}\widetilde{\xi}_{\mathrm{d}+1+\mathrm{j}}\right)\\ &\leq \frac{r_n+s_n}{n}\beta^T\Sigma_1\beta+o\left(1\right)\\ &\to 0,\quad \text{as } n\to\infty.\end{aligned}$$

Now, it remains to show $\frac{1}{\sqrt{n}}S_n'=\frac{1}{\sqrt{n}}\sum_{j=0}^{k-1}\eta_j\xrightarrow{\mathcal{D}}N\left(0,\beta^T\Sigma_1\beta\right)$.

Since η_j is a function of $\left\{\widetilde{\xi}_t,\right\}_{t=j(r+s)+1}^{j(r+s)+r-1}$ which is $\mathcal{F}_{\lfloor(\nabla+\int)+\infty-\lceil}^{\lfloor(\nabla+\int)+\nabla-\infty}$-measurable, the Volkonskii and Rozanov's lemma in the appendix of Masry and Tjøstheim(1997) implies that, with $\widetilde{s}_n=s_n-d+1$,

$$\begin{aligned}&\left|E[\exp(it\frac{1}{\sqrt{n}}\sum_{j=0}^{k-1}\eta_j)]-\prod_{j=0}^{k-1}E\left(\exp\left(it\eta_j\right)\right)\right|\\ \leq\ & 16k\alpha\left(\widetilde{s}_n-d+1\right)\simeq\frac{n}{r_n+s_n}\alpha\left(\widetilde{s}_n\right)\simeq\frac{n}{r_n}\alpha\left(\widetilde{s}_n\right)\simeq o\left(1\right),\end{aligned}$$

where the last two equalities follows from hold (A.A.4). Thus, the summands $\{\eta_j\}$ in S_n' are asymptotically independent. Since the similar operation to (A.A.5) yields

$$\operatorname{var}\left(\eta_{\mathrm{j}}\right)=\mathrm{r_n}\beta^{\mathrm{T}}\Sigma_1\beta\left(1+\mathrm{o}\left(1\right)\right),$$

and hence

$$\operatorname{var}(\frac{1}{\sqrt{\mathrm{n}}}\mathrm{S_n'})=\frac{1}{\mathrm{n}}\sum_{\mathrm{j}=0}^{\mathrm{k}-1}\mathrm{E}\left(\eta_{\mathrm{j}}^2\right)=\frac{\mathrm{k_n r_n}}{\mathrm{n}}\beta^{\mathrm{T}}\Sigma_1\beta\left(1+\mathrm{o}\left(1\right)\right)\to\beta'\Sigma^*\beta.$$

Finally, due to the boundedness of density and kernel functions, the Lindeberg-Feller condition for the asymptotic normality of S_n' holds,

$$\frac{1}{n}\sum_{j=0}^{k-1}E\left[\eta_j^2I\left\{|\eta_j|>\sqrt{n}\delta\sqrt{\beta^T\Sigma_1\beta}\right\}\right]\to 0,$$

for every $\delta > 0$. This completes the proof of Step III.

From $e_1^T Q_n^{-1} \xrightarrow{p} e_1^T$, the Slutzky Theorem implies $\sqrt{nh}[I_2 \otimes e_1^T Q_n^{-1}]\widetilde{t}_n^* \xrightarrow{d} N(0, \Sigma_1^*)$, where $\Sigma_1^* = [I_2 \otimes e_1^T]\Sigma_1[I_2 \otimes e_1]$. In sum, $\sqrt{nh}(\widehat{\varphi}_1(y_1) - \varphi_1(y_1) - B_n) \xrightarrow{d} N(0, \Sigma_1^*)$.

$$
\begin{aligned}
&\Sigma_1^*(y_1) \\
= &\int \frac{p_2^2(z_2)}{p(y_1, z_2)} ||(K * K)_0||_2^2 \begin{bmatrix} \nabla G_m(y_1, z_2)^2 v(y_1, z_2) & (\nabla G_m \cdot \nabla G_v)(\kappa_3 \cdot v^{3/2})(y_1, z_2) \\ (\nabla G_m \cdot \nabla G_v)(\kappa_3 \cdot v^{3/2})(y_1, z_2) & \nabla G_v(y_1, z_2)^2 \kappa_4(y_1, z_2) v^2(y_1, z_2) \end{bmatrix} dz_2 \\
&+ \int \frac{p_2^2(z_2)}{p(y_1, z_2)} ||K||_2^2 H_2(z_2) H_2^T(z_2)\, dz_2.
\end{aligned}
$$

Lemma A.1 *Assume the conditions in C.1, C.4. through C.6. For a bounded function, $F(\cdot)$, it holds that*

$$
\begin{aligned}
(a)\ r_{1n} &= \frac{\sqrt{h}}{\sqrt{n}} \sum_{k=d}^{n} K_h(y_{k-1} - y_1)(\widehat{p}_2(\underline{y}_{k-2}) - \overline{p}_2(\underline{y}_{k-2}))F(x_k) = o_p(1), \\
(b)\ r_{2n} &= \frac{\sqrt{h}}{\sqrt{n}} \sum_{k=d}^{n} K_h(y_{k-1} - y_1)(\widehat{p}(x_k) - p(x_k))F(x_k) = o_p(1)
\end{aligned}
$$

PROOF. The proof of (b) is almost the same as (a). So, we only show (a). By adding and subtracting $\overline{L}_{l|k}(y_{l-2}|y_{k-2})$, the conditional expectation of $L_g\left(\underline{y}_{l-2} - \underline{y}_{k-2}\right)$ given $\underline{y}_{k-2}$ in r_{1n}, we get $r_{1n} = \xi_{1n} + \xi_{2n}$, where

$$
\begin{aligned}
\xi_{1n} &= \frac{1}{n^2} \sum_{k=d}^{n} \sum_{l=d}^{n} K_h(y_{k-1} - y_1) F(x_k) [L_g\left(\underline{y}_{l-2} - \underline{y}_{k-2}\right) - \overline{L}_{l|k}(y_{l-2}|y_{k-2})] \\
\xi_{2n} &= \frac{1}{n^2} \sum_{k} \sum_{l} K_h(y_{k-1} - y_1) F(x_k) [\overline{L}_{l|k}(y_{l-2}|y_{k-2}) - \overline{p}_2(\underline{y}_{k-2})]
\end{aligned}
$$

Rewrite ξ_{2n} as

$$
\begin{aligned}
&\frac{1}{n^2} \sum_{k} \sum_{s<k^*(n)} K_h(y_{k-1} - y_1) F(x_k) [\overline{L}_{k+s|k}(y_{k+s-2}|y_{k-2}) - \overline{p}_2(\underline{y}_{k-2})] \\
&+ \frac{1}{n^2} \sum_{k} \sum_{s \geq k^*(n)} K_h(y_{k-1} - y_1) F(x_k) [\overline{L}_{k+s|k}(y_{k+s-2}|y_{k-2}) - \overline{p}_2(\underline{y}_{k-2})],
\end{aligned}
$$

where $k^*(n)$ is increasing to infinity as $n \to \infty$. Let

$$
B = E\{K_h(y_{k-1} - y_1) F(x_k) [\overline{L}_{k+s|k}(y_{k+s-2}|y_{k-2}) - \overline{p}_2(\underline{y}_{k-2})]\},
$$

which exists due to the boundedness of $F(x_k)$. Then, for a large n, the first part of ξ_{2n} is asymptotically equivalent to $\frac{1}{n}k^*(n)B$. The second part of ξ_{2n} is bounded by

$$\begin{aligned} & \sup_{s\geq k^*(n)} |p_{k+s|k}(y_{k+s-2}|y_{k-2}) - p(y_{k-2})| \frac{1}{n}\sum_{k}^{n} K_h(y_{k-1}-y_1)|F(x_k)| \\ \leq\ & \rho^{k(n)}O_p(1). \end{aligned}$$

Therefore, $\sqrt{nh}\xi_{2n} \leq O_p(\frac{\sqrt{h}}{\sqrt{n}}k^*(n)) + O_p\left(\rho^{-k^*(n)}\sqrt{nh}\right) = o_p(1)$, for $k(n) = \log n$, for example.

It remains to show $\xi_{1n} = o_p\left(\frac{1}{\sqrt{nh}}\right)$. Since $E(\xi_{1n}) = 0$ from the law of iteration, we just compute

$$\begin{aligned} E\left(\xi_{1n}^2\right) =\ & \frac{1}{n^4}\sum_{k\neq l}^{n}\sum^{n}\sum_{i\neq j}^{n}\sum^{n} E\{K_h(y_{k-1}-y)K_h(y_{i-1}-y)F(x_k) \\ & F(X_l)\left[L_g\left(\underline{y}_{l-2}-\underline{y}_{k-2}\right) - \overline{L}_{l|k}(\underline{y}_{k-2})\right]\left[L_h\left(\underline{y}_{j-2}-\underline{y}_{i-2}\right) - \overline{L}_{j|i}(\underline{y}_{i-2})\right]\}. \end{aligned}$$

(1) Consider the case $k = i$ and $l \neq j$.

$$\begin{aligned} & \frac{1}{n^4}\sum_{k}^{n}\sum^{n}\sum_{l\neq j}^{n} E\{K_h^2(y_{k-1}-y)F^2(x_k) \\ & \left[L_g\left(\underline{y}_{l-2}-\underline{y}_{k-2}\right) - \overline{L}_{l|k}(\underline{y}_{k-2})\right]\left[L_g\left(\underline{y}_{j-2}-\underline{y}_{k-2}\right) - \overline{L}_{j|k}(\underline{y}_{k-2})\right]\} \\ =\ & 0, \end{aligned}$$

since, by the law of iteration and the definition of $\overline{L}_{j|k}(\underline{y}_{k-2})$,

$$\begin{aligned} & E_{|k,l}\left[L_g\left(\underline{y}_{j-2}-\underline{y}_{k-2}\right) - \overline{L}_{j|k}(\underline{y}_{k-2})\right] \\ =\ & E_{|k}\left[L_g\left(\underline{y}_{j-2}-\underline{y}_{k-2}\right) - \overline{L}_{j|k}(\underline{y}_{k-2})\right] = E_{|k}\left[L_g\left(\underline{y}_{j-2}-\underline{y}_{k-2}\right)\right] - \overline{L}_{j|k}(\underline{y}_{k-2}) = 0 \end{aligned}$$

(2) Consider the case $l = j$ and $k \neq i$.

$$\begin{aligned} & \frac{1}{n^4}\sum_{k\neq i}^{n}\sum^{n}\sum_{l}^{n} E\{K_h(y_{k-1}-y)K_h(y_{i-1}-y)F(x_k)F(x_i) \times \\ & \left[L_g\left(\underline{y}_{l-2}-\underline{y}_{k-2}\right) - \overline{L}_{l|k}(\underline{y}_{k-2})\right]\left[L_g\left(\underline{y}_{l-2}-\underline{y}_{i-2}\right) - \overline{L}_{l|i}(\underline{y}_{i-2})\right]\} \end{aligned}$$

We only calculate

$$\frac{1}{n^4}\sum_{k\neq i}^{n}\sum^{n}\sum_{l}^{n} E\{K_h(y_{k-1}-y)K_h(y_{i-1}-y)L_g\left(\underline{y}_{l-2}-\underline{y}_{k-2}\right)L_g\left(\underline{y}_{l-2}-\underline{y}_{i-2}\right)F(x_k)F(x_i)\} \tag{A.A.6}$$

since the rest of the triple sum consist of expectations of standard kernel estimates and are $O(1/n)$. Note that

$$\begin{aligned} & E_{|(i,k)}L_g\left(\underline{y}_{l-2}-\underline{y}_{k-2}\right)L_g\left(\underline{y}_{l-2}-\underline{y}_{i-2}\right) \\ \simeq\ & (L*L)_g\left(\underline{y}_{k-2}-\underline{y}_{i-2}\right)p_{l|(k,i)}\left(\underline{y}_{k-2}|\underline{y}_{k-2},\underline{y}_{i-2}\right), \end{aligned}$$

where $(L*L)_g(\cdot) = (1/g)\int L(u)\,L(u+\cdot/g)$ is a convolution kernel. Thus, (A.A.6) is

$$\frac{1}{n^4}\sum_{k\neq i}^{n}\sum^{n}\sum_{l}^{n} E[K_h(y_{k-1}-y)\,K_h(y_{i-1}-y)\,(L*L)_g\left(\underline{y}_{k-2}-\underline{y}_{i-2}\right)\times F(x_k)F(x_i)p_{l|(k,i)}\left(\underline{y}_{k-2}|\underline{y}_{k-2},\underline{y}_{i-2}\right)$$

$$= O\left(\frac{1}{n}\right),$$

(3) Consider the case with $i=k,\ j=m$.

$$\frac{1}{n^4}\sum_{k\neq l}^{n}\sum^{n} E\{K_h^2(y_{k-1}-y)\,F^2(x_k)\left[L_g(y_{l-2}-y_{k-2})-\overline{L}_{l|k}(\underline{y}_{k-2})\right]^2$$

$$= O\left(\frac{1}{n^2hg}\right) = o\left(\frac{1}{nh}\right)$$

(4) Consider the case $k\neq i,\ l\neq j$.

$$\frac{1}{n^4}\sum_{k\neq l}^{n}\sum^{n}\sum_{i\neq j}^{n}\sum^{n} E\{K_h(y_{k-1}-y)\,K_h(y_{i-1}-y)\,F(x_k)\,F(x_i)$$
$$\left[L_g\left(\underline{y}_{l-2}-\underline{y}_{k-2}\right)-\overline{L}_{l|k}(\underline{y}_{k-2})\right]\left[L_g\left(\underline{y}_{j-2}-\underline{y}_{i-2}\right)-\overline{L}_{j|i}(\underline{y}_{i-2})\right]\}$$
$$= 0,$$

for the same reason as in (1).

References

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

[2] Auestad, B. and D. Tjøstheim, (1991). Functional identification in nonlinear time series. In *Nonparametric Functional Estimation and Related Topics*, ed. G. Roussas, Kluwer Academic: Amsterdam. pp 493–507.

[3] Black, F. (1989). Studies of stock price volatility changes. *proceedings from the American Association, Business and Economics Section,* 177-181.

[4] Buja, A., T. Hastie, and R. Tibshirani, (1989). Linear smoothers and additive models (with discussion). *Ann. Statist.* **17**, 453-555.

[5] Cai, Z., and E. Masry (2000). Nonparametric Estimation of Additive Nonlinear ARX Time Series: Local Linear Fitting and Projections. *Econometric Theory* 16, 465-501.

[6] Chen, R. (1996). A nonparametric multi-step prediction estimator in Markovian structures, *Statistical Sinica **6**, 603-615.*

[7] Chen, R., and R.S. Tsay, (1993a). Nonlinear additive ARX models, *Journal of the American Statistical Association* **88**, 955-967.

[8] Chen, R., and R.S. Tsay, (1993b). Functional-coefficient autoregressive models, *Journal of the American Statistical Association* **88**, 298-308.

[9] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, *Econometrica* **50**: 987-1008.

[10] Fan, J. (1992). Design-adaptive nonparametric regression. *J. Am. Statist Soc.* **82**, 998-1004.

[11] Granger, C. and T. Teräsvirta, (1993). *Modeling Nonlinear Economic Relationships,* Oxford University Press, Oxford.

[12] Hall, P. and C. Heyde (1980). *Martingale Limit Theory and Its Application.* New York: Academic Press.

[13] Härdle, W. (1990). *Applied Nonparametric Regression.* Econometric Monograph Series 19. Cambridge University Press.

[14] Härdle, W. and A.B. Tsybakov, (1997). Locally polynomial estimators of the volatility function. *Journal of Econometrics* , **81**, 223-242.

[15] Härdle, W., A.B. Tsybakov, and L. Yang, (1998). Nonparametric vector autoregression . *Journal of Statistical Planning and Inference,* **68**(2), 221-245

[16] Härdle, W. and P. Vieu, (1991). Kernel regression smoothing of time seires. *Journal of Time Series Analysis, **13, 209-232.***

[17] Hastie, T. and R. Tibshirani, (1990). *Generalized Additive Models.* Chapman and Hall, London.

[18] Hastie, T. and R. Tibshirani, (1987) Generalized additive models: Some applications. *J. Am. Statist Soc.* **82** 371- 386.

[19] Horowitz, J., (1999). Estimating generalized additive models. Forthcoming in Econometrica.

[20] Jones, M.C., S.J. Davies, and B.U. Park, (1994). Versions of kernel-type regression estimators. *J. Am. Statist Soc.* **89**, 825-832.

[21] Kim, w. and O. Linton, and N. Hengartner, (1999). A Computationally Efficient Oracle Estimator of Additive Nonparametric Regression with Bootstrap Confidence Intervals. *Journal of Computational and Graphical Statistics* 8, 1-20.

[22] Linton, O.B. (1996). Efficient estimation of additive nonparametric regression models. *Biometrika **84,*** 469-474.

[23] Linton, O.B. (2000). Efficient estimation of generalized additive nonparametric regression models. *Econometric Theory* **16,** 502-523.

[24] Linton, O.B. and W. Härdle, (1996). Estimating additive regression models with known links. *Biometrika* **83**, 529-540.

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

[26] Linton, O.B., Nielsen, J.P., and S. van de Geer, (1999). Estimating Multiplicative and Additive marker dependent hazard functions by Backfitting with the Assistance of Marginal Integration. Manuscript, LSE.

[27] Linton, O.B., N. Wang, R. Chen, and W. Härdle, (1995). An analysis of transformation for additive nonparametric regression. *Journal of the American Statistical Association **92,*** 1512-21.

[28] Mammen, E., O.B. Linton, and J. Nielsen, (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics,* **27**(5), 1443-1490.

[29] Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *J. Time Ser. Anal.* **17**, 571-599.

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

[31] Masry, E., and D. Tjøstheim, (1997). Additive nonlinear ARX time series and projection estimates. *Econometric Theory* 13, 214-252.

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

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

[34] Opsomer, J. D., and D. Ruppert, (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics, **25**, 186-211*

[35] Robinson, P.M. (1983). Nonparametric estimation for time series models, *Journal of Time Series Analysis, 4, 185-208.*

[36] Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall.

[37] Stone, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 685-705.

[38] Stone, C.J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.* **14**, 592-606.

[39] Teräsvirta, T., D. Tjøstheim, and C.W.J. Granger, (1994). Aspects of Modelling Nonlinear Time Series in *The Handbook of Econometrics*, vol. IV, eds. D.L. McFadden and R.F. Engle, 2919-2960, Amsterdam: Elsevier.

[40] Tjøstheim, D., and B. Auestad, (1994). Nonparametric identification of nonlinear time series: projections. *J. Am. Stat. Assoc.* **89**, 1398-1409.

[41] Tong, H. (1990). *Nonlinear Time Series Analysis: A dynamic Approach*, Oxford University Press, Oxford.

[42] Volkoniskii and Y.U. Rozanov, (1959). Some limit theorems for random functions. *Theory of Probability and Applications,* **4**, 178-197.

[43] Yang, L., W. Härdle, and J. Nielsen, (1999). Nonparametric autoregression with multiplicative volatility and additive mean. Journal of Time Series Analysis. **20**(5): 579-604.

TABLE 1: MSE FROM NONPARAMETRIC IV ESTIMATION

	Ave. MSE		Med. MSE		Ave. MSE^{TR}	
c_h	v_1	v_2	v_1	v_2	v_1	v_2
0.5	.0672	.0598	.0196	0171	.0497	.0524
1	.0343	.0322	.0122	0116	.0268	.0272
2	.0189	.0190	.0085	0078	.0156	.0163
3	.0146	.0165	.0069	0071	.0121	.0142
4	.0139	.0181	.0069	0075	.0110	.0158
5	.0166	.0233	.0083	0091	.0123	.0203
6	.0236	.0319	.0103	0119	.0161	.0272
7	.0348	.0439	.0131	0147	.0227	.0360
QF	.0524	.0687	.0426	0507	.0509	.0666

*Note: c_h is a bandwidth constant in $h = c_h \times$stnd.dev.$(y_t) \times n^{-1/5}$. Ave. MSE and Med. MSE denote the average and median of the mean squarred errors, respectively. Ave. MSE^{TR} in the last two columns means the average of MSE after removing 24 (less than 2.5%) worst extremes out of 1000. QF means the case with parametric quadratic fits.

Figure 1−1(a)
IV estimation of volatility: the first lag

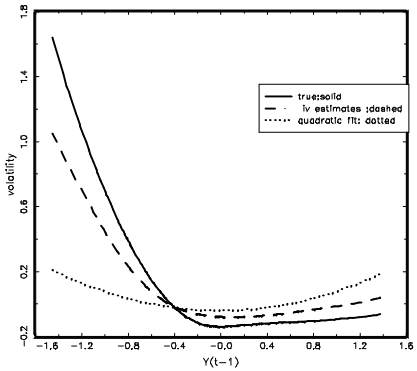

Figure 1−1(b)
IV estimation of volatility: the second lag

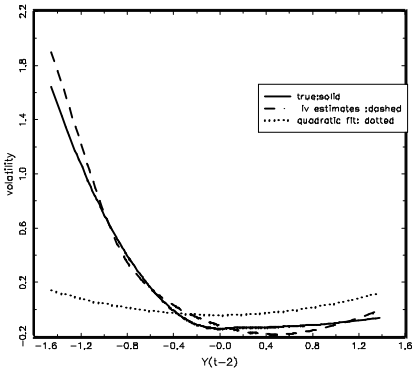

Figure 1−2(a)
IV estimation of volatility: the first lag

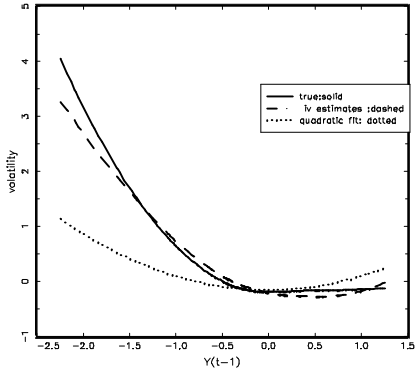

Figure 1−2(b)
IV estimation of volatility: the second lag

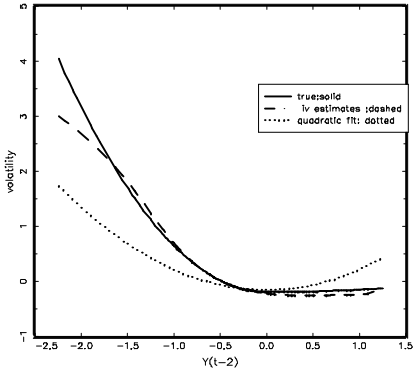

Figure 1−3(a)
IV estimation of volatility: the first lag

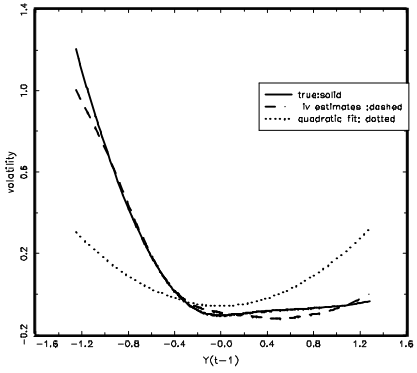

Figure 1−3(b)
IV estimation of volatility: the second lag

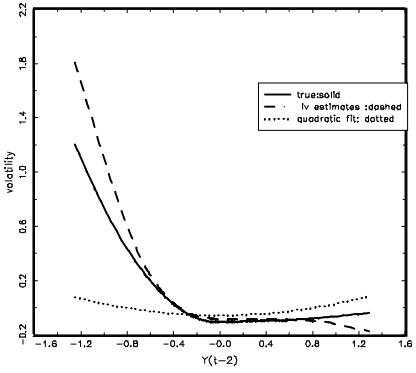

Figure 1−4(a)
IV estimation of volatility: the first lag

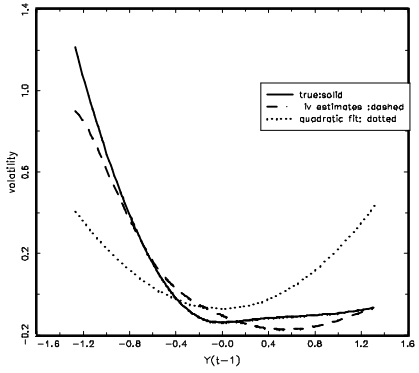

Figure 1−4(b)
IV estimation of volatility: the second lag

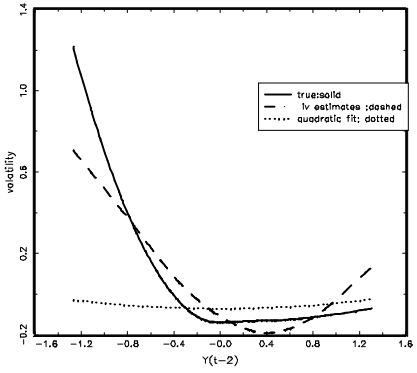

Figure 2−1(a)
IV estimates of Conditinal Mean: the first lag

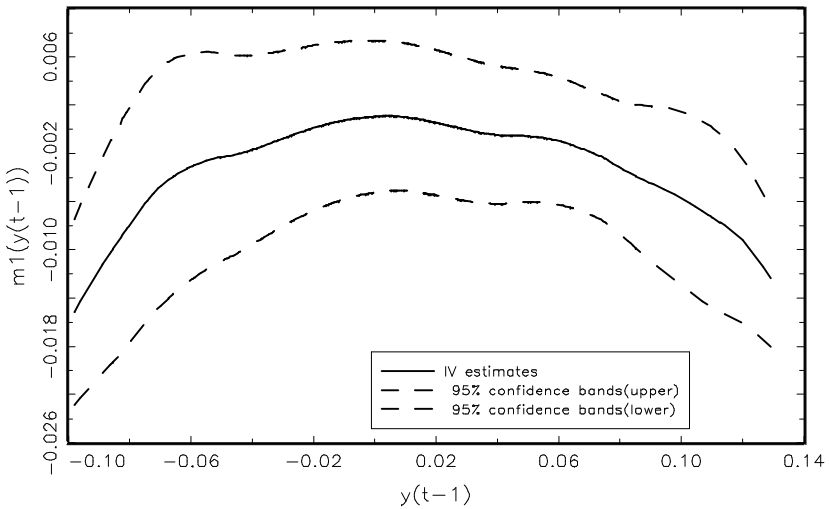

Figure 2−1(b)
IV estimates of Conditinal Mean: the second lag

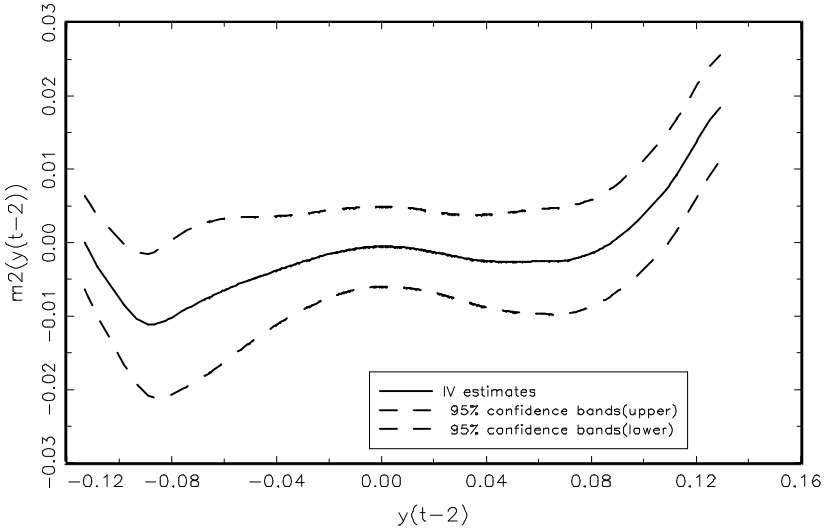

Figure 2−2(a)
IV estimates of Volatility: the first lag

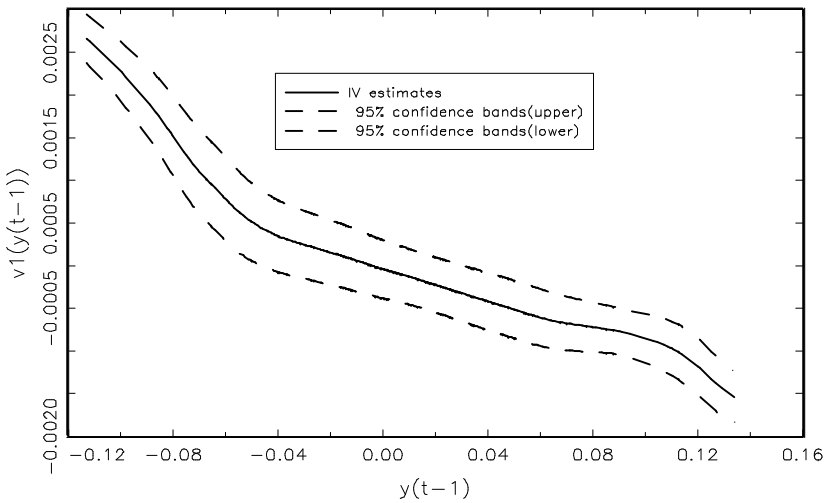

Figure 2−2(b)
IV estimates of Volatility: the second lag