MORE EFFICIENT KERNEL ESTIMATION IN NONPARAMETRIC REGRESSION WITH AUTOCORRELATED ERRORS*

Raymond J Carroll
Department of Statistics, Texas A&M University

Oliver B Linton
Department of Economics, London School of Economics and Political Science

Enno Mammen
Institut für Angewandte Mathematik, Universität Heidelberg

Zhijie Xiao
Department of Economics, University of Illinois, Urbana-Champaign

Contents:

Discussion Paper
No. EM/02/435
June 2002

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 020-7955 6698

* We would like to thank M Francisco-Fernandez, Jean Opsomer and Michael Schimek for interesting discussions. We would like to thank the Cowles Foundation, the National Cancer Institute, the National Institute of Environmental Health Sciences, the National Science Foundation, the Economics and Social Research Council of Great Britain, and the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373, for financial support.

Abstract

We propose a modification of kernel time series regression estimators that improves efficiency when the innovation process is autocorrelated. The procedure is based on a pre-whitening transformation of the dependent variable that has to be estimated from the data. We establish the asymptotic distribution of our estimator under weak dependence conditions. It is shown that the proposed estimation procedure is more efficient than the conventional kernel method. We also provide simulation evidence to suggest that gains can be achieved in moderate sized samples.

Keywords: Backfitting, efficiency, kernel estimation, time series.
JEL Nos.: C13, C14.

Contact addresses:
Professor Oliver B Linton, Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK Email: o.linton@lse.ac.uk

Professor Raymond J Carroll, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA. Email: carroll@stat.tamu.edu

Professor Enno Mammen, Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany. Email: mammen@statlab.uni-heidelberg.de

Professor Zhijie Xiao, Department of Economics, University of Illinois at Urbana-Champaign, 1206 S. Sixth Street, Champaign, IL 61820, USA. Email: zxiao@uiuc.edu

1 INTRODUCTION

Consider the following regression model

$$Y_t = m(X_t) + u_t,\ t = 1, \dots, T, \tag{1}$$

where the stationary residual process u_t is autocorrelated, but satisfies $E(u_t|X_1, \dots, X_T) = 0$ almost surely. The function $m(.)$ is assumed to be unknown but smooth, and is the object of central interest. There are two leading sampling schemes with regard to the process $\{X_t\}$. First, the 'fixed design' case where X_t is time or some smooth function thereof, i.e., $X_t = f(t/T)$ for some smooth f; and second, the 'random design' case where X_t is a stationary stochastic process itself with a nondegenerate marginal distribution.[1] In the former case, both the standard least squares parametric and kernel nonparametric estimator have variances proportional to the long run variance [i.e., the spectral density at frequency zero] of the process $\{u_t\}$. However, adjusting for serial correlation brings no advantage in terms of estimator variance in either parametric or nonparametric method. Specifically, when the regressors are polynomials in time OLS=GLS, see for example Andersen (1971, p581). Much methodological work in nonparametric statistics has focussed on this sampling scheme, especially with regard to bandwidth selection, see Hart (1991) for references.

The focus of this paper is the second sampling scheme where X_t is a non-degenerate stochastic process. This setting arises in many applications, because time itself is often not the only relevant covariate. Indeed, in the 1970's the linear regression model with autocorrelated disturbances was one of the central models of interest and numerous procedures were created to deal with the estimation and testing issues that ensued, including: Cochrane-Orcutt, Hildreth-Lu, Prais-Winsten, and Durbin-Watson. As is well known, when the regression function is parametric the variance of the parameter estimators is proportional to the long run variance of the process $\{X_t u_t\}$ and least squares standard errors that ignore this fact are inconsistent and need to be modified in a non-trivial way. Also, one can generally improve efficiency of least squares estimators by using a GLS weighting scheme that reflects the error autocorrelation function. Compare this with the case where $m(.)$ is nonparametric, which has been analyzed in Robinson (1983), Masry (1996ab) for example. In this case, standard kernel regression smoothers do not take account of the correlation structure in X_t or u_t and estimate the regression function in the same way as if these processes were independent. Furthermore, the variance of such estimators is proportional to the short run variance of u_t, $\sigma_u^2 = \mathrm{var}(u_t)$ and does not depend on the regressor or error covariance functions $\gamma_X(j) = \mathrm{cov}(X_t, X_{t-j})$, $\gamma_u(j) = \mathrm{cov}(u_t, u_{t-j})$, $j \neq 0$. Practitioners accustomed to correcting standard errors for dependence believe that the standard errors in nonparametric regression are therefore suspect. As Conley, Hansen, Luttmer, and Scheinkman (1997) say: "Although theoretically correct the practice of ignoring serial correlation is

[1] Opsomer, Wang, and Yang (2001) have discussed the related case where the regressors are multivariate and random, but the error covariance is a smooth function of the regressors. This case is more like the 'fixed design' in some respects.

not likely to work well for the temporal dependence present in our short-term interest rate data". The purpose of this paper is to show that the autocorrelation function of the error process has useful information to provide for improving estimators of the regression function. As a by-product one might hope to obtain more accurate standard errors, given that the resulting error process is purged of all correlation.

There is a related literature on estimating nonparametric regression with longitudinal or panel data. For example: Severini and Staniswalis (1994), Zeger and Diggle (1994), Wild and Yee (1996), and Wu, Chiang and Hoover (1998), among others. The first authors estimate the covariance matrix of the correlated observations and use this in their kernel construction of the nonparametric regression estimate. The other papers effectively ignore the correlation structure entirely and "pretend" that the data are really independent, this being the so–called "working independence" method. Ruckstuhl, Welsh and Carroll (2000) and Lin and Carroll (2000) provided theoretical evidence in support of the working independence method. In fact, they showed that for many situations and different methods of kernel estimation, the working independence method is most efficient in terms of mean squared error. That is, for the kernel methods proposed in the literature, it is generally better to ignore the correlation structure entirely. Carroll et al. (2001) construct a kernel–type method that can take advantage of the correlations among the data. The method is a simple modification, and generalization to an arbitrary covariance matrix, of a method proposed by Ruckstuhl, Welsh and Carroll (2000). The resulting estimator is asymptotically more efficient than the working independence estimator.

In this paper, we propose a new kernel-based procedure for estimating $m(x)$ in the time series regression model (1) that takes account of the correlation structure of the error terms and is asymptotically more efficient than the usual methods. The basic idea of the proposed estimation is to transform or "prewhiten" the original regression model so that the filtered regression has a residual term that is uncorrelated. However, because of the nonlinear feature of the regression function $m(\cdot)$, the transformation depends on both the function $m(.)$ and on the parameters of the autoregressive representation of u. We therefore first estimate these quantities and then construct a feasible transformation of the dependent variable Y_t. The resulting estimator we show to be asymptotically normal and to be more efficient than the conventional kernel estimator. We allow for an error correlation structure of unknown form, i.e., the autoregressive representation of the process need not be of finite order.

The rest of the paper is organized as follows. We introduce the proposed estimation method in Section 2. Regularity assumptions and the limiting distribution of the estimator are given in Section 3. Section 4 discusses model selection and bandwidth choice. Section 5 proposes an even more efficient estimator. In Section 6 we report some numerical results on simulated data and on stock index return data. Section 7 concludes. All proofs are given in the Appendix. For notation, we define $\mu_q(K) = \int u^q K(u) du$, and for $f_X : \mathbb{R}^d \to \mathbb{R}$, we denote

$$f_X^{(r)}(x) = \sum_{r_1+\cdots+r_d=r} \frac{\partial^r f_X(x)}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}}.$$

2 ESTIMATION METHOD

2.1 Motivation and An Infeasible Estimator

Suppose that we have a sample $\{(X_1, Y_1), \ldots, (X_T, Y_T)\}$, where $X_t \in \mathbb{R}^d$ and $Y_t \in \mathbb{R}$, from the nonparametric regression model (1). We assume that the residual process u_t is stationary, mean zero, and has an invertible linear process representation

$$u_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \tag{2}$$

where ε_t are independent identically distributed with mean zero and variance σ_ε^2. Without loss of generality, $c_0 = 1$. For convenience we shall temporarily assume that the process u_t is independent of the process X_t, but relax this assumption below. The coefficients $\{c_j\}_{j=0}^{\infty}$ and the regression function $m(\cdot)$ are unknown except that $m(.)$ is a smooth function and the coefficients c_j satisfy certain summability conditions [e.g., the process is short memory] as specified later in our assumptions. Our assumptions permit u_t to be any finite order ARMA(p,q) process but we allow for the full class of linear processes as is common in much literature on estimating linear regression with correlated errors. The objective is to estimate $m(x)$ at some interior point x and to provide confidence intervals for such estimates.

Let $c(L) = \sum_{j=0}^{\infty} c_j L^j$, where L is the usual lag operator. Inverting $c(L)$ we obtain an autoregressive representation of u_t of potentially infinite order. Let

$$c(L)^{-1} = a(L) = a_0 - a_1 L - \cdots - a_j L^j - \cdots = a_0 - \sum_{j=1}^{\infty} a_j L^j \tag{3}$$

be the inverse, and define that $a_0 = 1$ without loss of generality, we have

$$a(L) u_t = \varepsilon_t.$$

Applying $a(L)$ to regression (1), we obtain

$$a(L) Y_t = a(L) m(X_t) + \varepsilon_t. \tag{4}$$

The error term in this transformed model is now uncorrelated; however, the immediate usefulness of this is unclear because m is nonlinear and so does not commute with the operator $a(L)$ as would be the case with a linear model.

We rewrite equation (4) as

$$\underline{Y}_t = m(X_t) + \varepsilon_t, \tag{5}$$

where $\underline{Y}_t$ is the filtered series

$$\underline{Y}_t = Y_t - \sum_{j=1}^{\infty} a_j \left(Y_{t-j} - m(X_{t-j})\right).$$

The transformed model (5) is also a valid regression equation since ε_t is independent of X_t. If $\underline{Y}_t$ were known, as shown by the following Theorem, a nonparametric kernel regression of $\underline{Y}_t$ on X_t would be more efficient than the conventional kernel estimation. In this paper, we give asymptotic analysis based on the Nadaraya-Watson procedure and make comparison for the corresponding estimators. However, the same idea can be applied to other types of estimators, like local polynomials. This leads to a difference in the bias expression but the same variance for comparable implementations. Let $\breve{m}(x)$ be the nonparametric estimator based on kernel regression of Y_t on X_t and let $\overline{m}(x)$ be the estimator based on kernel regression of $\underline{Y}_t$ on X_t:

$$\breve{m}(x) = \frac{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right) Y_t}{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)} \qquad ; \qquad \overline{m}(x) = \frac{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right) \underline{Y}_t}{\sum_{t=1}^{T} K\left(\frac{x-X_t}{h}\right)},$$

where

$$K\left(\frac{x-X_i}{h}\right) = \prod_{j=1}^{d} k\left(\frac{x_j - X_{ij}}{h}\right), \tag{6}$$

with k being the corresponding kernel function and h being the bandwidth in the preliminary estimation, Theorem 1 below gives the asymptotic distribution of $\overline{m}(x)$ and show that it is asymptotically more efficient than $\breve{m}(x)$.

Theorem 1 *Suppose that the assumptions given in Section 3 hold. Then,*

$$\sqrt{Th^d}[\overline{m}(x) - m(x) - h^q \mu_q(K)\mathcal{B}(x)] \Longrightarrow N\left(0, \frac{\sigma_\varepsilon^2 \|K\|^2}{f_X(x)}\right),$$

where $\mathcal{B}(x)$ is a bias term that equals

$$\sum_{p+r=q, 1\le p\le q, 0\le r\le q} \frac{1}{p!r!} m^{(p)}(x) \frac{f_X^{(r)}(x)}{f_X(x)}.$$

Theorem 1 shows that the bias term of the estimator $\overline{m}(x)$ is the same as that of the conventional kernel estimator $\breve{m}(x)$. In the case with a quadratic kernel, $q = 2$, and the bias term is simply $\frac{1}{2}\mu_2(K_1)[m''(x) + 2m'(x)\frac{f'}{f}(x)]$. The smoother $\overline{m}(x)$ has a variance proportional to σ_ε^2 and hence is more efficient than the traditional kernel estimator $\breve{m}(x)$, which has variance proportional to

$$\sigma_u^2 = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} c_j^2 \ge \sigma_\varepsilon^2.$$

For example, when $u_t = au_{t-1} + \varepsilon_t$, we have $\sigma_u^2 = \sigma_\varepsilon^2/(1-a^2)$, which strictly exceeds σ_ε^2 except when $a = 0$. In fact, the efficiency gain of $\overline{m}(x)$ can be arbitrarily large in this case because $1/(1-a^2)$ is unbounded as a function of a.

2.2 The Estimator

In practice, $\underline{Y}_t$ is unknown. Thus the regression (5) and $\overline{m}(x)$ are infeasible. We propose in this section a feasible estimator of regression (5) by replacing the left hand side of this equation by an approximation of $\underline{Y}_t$ based on estimates of the coefficients a_j and a truncation of the infinite sum to a finite but large order sum. The proposed estimation procedure is as follows:

1. First obtain a preliminary consistent estimate of m by conventional kernel smoothing Y_t on X_t with corresponding kernel K_0 and bandwidth h_0. Denote the preliminary estimates as $\widehat{m}(X_t)$ [see more discussions of our preliminary estimators in later sections] and calculate the estimated residuals
$$\widehat{u}_t = Y_t - \widehat{m}(X_t).$$

2. Let $\tau = \tau(T)$ be some truncation parameter suitably small relative to the sample size T but large enough to avoid serious bias [see Assumption 6 in Section 3]. We conduct a τ-th order autoregression of $\widehat{u}_t$:
$$\widehat{u}_t = \widehat{a}_1 \widehat{u}_{t-1} + \cdots + \widehat{a}_\tau \widehat{u}_{t-\tau} + \text{residual}. \tag{7}$$
Define the estimate $\widehat{A}_\tau = (\widehat{a}_1, \ldots, \widehat{a}_\tau)'$ of $A_\tau = (a_1, \ldots, a_\tau)'$, where
$$\widehat{A}_\tau = (\widehat{U}_\tau' \widehat{U}_\tau)^{-1} \widehat{U}_\tau' \widehat{u},$$
where $\widehat{u} = (\widehat{u}_\tau, \ldots, \widehat{u}_T)'$ and $\widehat{U}_\tau$ is the $(T-\tau) \times \tau$ matrix of regressors with typical element $\widehat{u}_{t-j}$.

3. Construct an approximation of $\underline{Y}_t$ by
$$\widehat{\underline{Y}}_t = Y_t - \sum_{j=1}^{\tau} \widehat{a}_j \left(Y_{t-j} - \widehat{m}(X_{t-j}) \right),$$
the proposed estimator of $m(x)$ is then obtained from kernel smoothing $\widehat{\underline{Y}}_t$ on X_t, calling the resulting estimator $\widetilde{m}(x)$, i.e.,
$$\widetilde{m}(x) = \frac{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right) \widehat{\underline{Y}}_t}{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)}, \tag{8}$$
where $K_1\left(\frac{x-X_i}{h_1}\right)$ is defined by the same formula as (6) with the corresponding kernel and bandwidth replaced by k_1 and h_1.

The above procedures may be iterated to achieve better finite sample performance in practice. Also, in estimating the coefficients $(\widehat{a}_1, \cdots, \widehat{a}_\tau)$, for reasons of parsimony, it may be advantageous to 'model' the residual process u_t by some parametric ARMA process $A(L)u_t = B(L)\varepsilon_t$; estimates of a_j may be obtained from inverting $B(L)$.

We show in Section 3 that, under appropriate assumptions, the proposed estimator $\widetilde{m}(x)$ is asymptotically equivalent to the infeasible estimator $\overline{m}(x)$, which is more efficient than the conventional kernel estimation. In fact, the transformation we propose is also effective in parametric models [although not as effective as a full GLS transform], see Kristensen and Linton (2001).

Recently, Vilar-Fernandez and Francisco-Fernandez (2001) have analyzed an alternative modification of standard local polynomial regression. They included a 'GLS-weighting' for autocorrelation in the criterion function. The resulting estimator involves transformation of both Y and X processes by a matrix P, which is the square root of the inverse covariance matrix of $(u_1, \ldots, u_T)$. This transformation does not improve the first order properties of the estimator although they have shown in simulations that it can improve the finite sample MSE.

2.3 Estimation of the Residuals

An important input in our procedure is the estimated residual $\widehat{u}_t = Y_t - \widehat{m}(X_t)$, whose construction presupposes an estimate of $m(X_t)$. For the choice of $\widehat{m}(X_t)$, natural candidates include the conventional Nadaraya-Watson estimator and the widely used local polynomial estimator or sieve estimators. When the ordinary kernel estimator is used, additional trimming is usually needed to remove the boundary bias because if we use all observations in estimating the error density, we are pushed into the boundary. To avoid introducing another trimming on $\widehat{m}(X_t)$, we use local polynomials instead of ordinary kernel estimators in the construction of residuals $\widehat{u}_t$. See Fan (1992), and Fan and Gijbels (1996) for discussion on the attractive properties of local polynomials. Given observations $\{Y_t, X_t\}_{t=1}^n$, the preliminary estimate of the regression function $m(x)$ can be obtained using the multivariate weighted least squares criterion

$$\sum_{t=1}^{n} \left[Y_t - \sum_{0 \le |k| \le p} b_{\mathbf{k}} \cdot (X_t - x)^{\mathbf{k}} \right]^2 K_0\left(\frac{X_t - x}{h_0} \right), \tag{9}$$

where $K_0(u)$ is a nonnegative weight function on $\mathbb{R}^d$ and h_0 is a bandwidth parameter, while p is an integer with $p \geq 2$. Let $\widehat{m}(x) = \widehat{b}_0$, where $\widehat{b}_0$ is the minimizing intercept in (9). We compute this estimator for each sample point and use it to construct the residuals $\widehat{u}_t = Y_t - \widehat{m}(X_t)$, which are the key input to the density estimate. Again, for convenience of comparison, we choose $p = q - 1$ so that the bias and variance of the preliminary estimator are of the same orders of magnitude as the final estimator. We give more discussion about the technical details of the local polynomial estimator in the appendix.

3 MAIN RESULT

In this section we shall assume that the error process $\{u_t\}$ is independent of the process $\{X_t\}$. To proceed, we assume that $\{X_t\}$ is a α-mixing process. Let $\mathcal{F}_a^b$ be the σ-algebra of events generated by the random variables $\{X_t; \quad a \leq j \leq b\}$. The stationary processes $\{X_t\}$ is called strongly mixing [Rosenblatt (1956)] if

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |\Pr(A \cap B) - \Pr(A)\Pr(B)| \equiv \alpha(k) \to 0 \quad \text{as } k \to \infty. \tag{10}$$

To facilitate the asymptotic analysis, we make the following assumptions on the residuals and regressors, the kernel function $k(\cdot)$, and the bandwidth parameters h_0 and h_1. In practice, even when some of these conditions do not hold, if the residuals are autocorrelated, efficiency gain over the conventional kernel estimator may still be found in the proposed estimator.

ASSUMPTION A

1. *The kernels $k = k_j$, $j = 0, 1$ are bounded, have compact support $[-1, 1]$, are symmetric about zero, and are Lipschitz continuous, i.e., there exists a positive finite constant C such that $|k(u) - k(v)| \leq C|u - v|$. They also satisfy the property that $\int k(u)du = 1$. For k_1, there exists an even positive integer q such that*

$$\int u^j k_1(u)du = 0, \; j = 1, \ldots, q-1, \text{ and } \int u^q k_1(u)du \neq 0.$$

   The functions $H_j(u) = u^j K_0(u)$ for all j with $0 \leq |j| \leq 2p+1$, where K_0 is defined by (6), are Lipschitz continuous, i.e., there exists finite C_1 such that $|H_j(u) - H_j(v)| \leq C||u - v||$.

2. *The process $\{X_t\}$ is strongly mixing with $\sum_{i=1}^\infty i^\delta \{\alpha(i)\}^{1-2/\nu} < \infty$ for some $2 < \nu \leq \theta$ and $\delta > 1 - 2/\nu$. The density f_X of X_t and the joint densities of $(X_t, X_{t+\ell})$, $(X_t, X_{t+\ell}, X_{t+j})$, $(X_t, X_{t+\ell}, X_{t+j}, X_{t+s})$ are uniformly bounded and are bounded away from zero on their supports.*

3. *For some $\theta > 2$, $E(|u_t|^\theta) < \infty$.*

4. *The function $m(\cdot)$ is q times partially differentiable and the q^{th} order partial derivatives are Lipschitz continuous on $\mathcal{X}$. The partial derivatives of f_X exist and are continuous on $\mathcal{X}$.*

5. *The process $\{u_t\}$ is a stationary invertible linear process representable in the form of (2), and has inverse (3). In addition, there exists some $\lambda \in (0, 1)$ such that the linear process coefficients $|a_j|$ are bounded by a constant multiple of λ^j.*

6. *The truncation parameter τ satisfies $\tau(T) = \kappa \log T$ for some $\kappa > 0$.*

7. *Bandwidths h_0 and h_1 satisfy that $h_0/h_1 \to 0$, $T^{1/2}h_1^{d/2}h_0^{2q}(\log T) \to 0$,* and *$T^{-1/2}h_0^{-d}h_1^{d/2}(\log T) \to 0$.*

The stationarity condition rules out examples like $X_t = f(t/T)$ for smooth f. Assumption 1 is a standard assumption for kernel functions in nonparametric estimation. Under the mixing conditions of Assumption 2, the temporal dependence among $\{X_t\}$ decreases fast enough as the time distance increases, and thus is asymptotically ignorable. In particular, strong law of large numbers and central limiting theorems continuous to hold for standardized summations and uniform convergence results on the kernel smooth quantities still hold. The differentiability of Assumption 4 ensures a Taylor expansion to appropriate order. While Assumption 5 is stronger than the summability conditions in, say Phillips and Solo (1992), the dominance requirement that $|a_j|$ are bounded by a constant multiple of λ^j is general enough to include leading cases like the widely considered stationary invertible ARMA process. This dominance condition is useful in our technical development and, in particular, provides a sufficient condition for controlling the order of magnitude of various summations involving c_j. No doubt this condition could be weakened, but we do not attempt to do so or to find minimal conditions under which our results hold. The expansion rate of the truncation parameter given in Assumption 6 is also for convenience and our results hold for a much wider range of τ. In fact, from the proof in the Appendix we can see that as long as the tail summation $(\sum_{j=\tau+1}^{\infty} a_j)$ of the sequence a_j is controlled under appropriate order, our results still hold. Assumption 7 assumes that we undersmooth in the preliminary estimation stage so that the bias term coming from preliminary estimation will be smaller than the leading bias term. Consequently, the feasible estimator has the same asymptotic mean squared error (MSE) as the infeasible estimator $\overline{m}$. Note that if we take $h_1 = O(T^{-1/(2q+d)})$ then Assumption 7 is satisfied for all q, d and many sequences $h_0(T)$.

Theorem 2 *Suppose that Assumptions 1 to 7 hold. Then,*

$$\sqrt{Th_1^d}[\widetilde{m}(x) - m(x) - h_1^q \mu_q(K_1)\mathcal{B}(x)] \Longrightarrow N\left(0, \frac{\sigma_\varepsilon^2 \|K_1\|^2}{f_X(x)}\right).$$

We have a sort of 'oracle' property here: the feasible estimator $\widetilde{m}(x)$ is asymptotically equivalent to $\overline{m}(x)$ and hence is more efficient than $\widehat{m}(x)$. By undersmoothing the pilot estimator $\widehat{m}(x)$ we can make the bias of $\widetilde{m}(x)$ the same as the bias of $\widehat{m}(x)$. Therefore, $\widetilde{m}(x)$ should be preferred to $\widehat{m}(x)$. A similar result applies to the procedure defined throughout with local polynomials of given order under appropriate smoothness conditions, except that the bias function is different.

The asymptotic normal distribution given by Theorem 2 can be used to calculate pointwise confidence intervals for estimators described here. To do this we require an estimate of the asymptotic variance. Let

$$\widetilde{v}(x) = \frac{\sum_t K\left(\frac{x-X_t}{h}\right)^2 \widetilde{\varepsilon}_t^2}{\left[\sum_t K\left(\frac{x-X_t}{h}\right)\right]^2},$$

where $\widetilde{\varepsilon}_t = \underline{\widehat{Y}}_t - \widetilde{m}(X_t)$. Then,

$$\widetilde{m}(x) \pm z_{\alpha/2}\sqrt{\widetilde{v}(x)}, \tag{11}$$

where z_α are the standard normal critical values, provide valid two sided pointwise confidence intervals provided the estimation is undersmoothed, i.e., $h_1 = o(T^{-1/(2q+d)})$. By definition ε_t is supposed to be an uncorrelated sequence so that we might expect these standard errors to be more accurate than those for $\widehat{m}(x)$.

One may substitute different smoothers like local polynomials and one may employ a different estimation scheme to obtain the $\widehat{a}_j's$. One can also expect some improvement by iterating the process. Specifically, define again

$$\widetilde{\underline{Y}}_t = Y_t - \sum_{j=1}^{\tau} \widetilde{a}_j \left(Y_{t-j} - \widetilde{m}(X_{t-j})\right),$$

where $(\widetilde{a}_1, \ldots, \widetilde{a}_\tau)'$ are obtained from the least squares regression of $Y_t - \widetilde{m}(X_t)$ on $(Y_{t-1} - \widetilde{m}(X_{t-1}), \ldots, Y_{t-\tau} - \widetilde{m}(X_{t-\tau}))'$, and kernel smooth $\widetilde{\underline{Y}}_t$ against X_t.

Finally, we can weaken our assumption of independence of X from u. For example, suppose that $u_t = \sigma(X_t)v_t$ with $\sigma(X_t)$ a smooth function bounded away from zero and $E(v_t|X_1, \ldots, X_T) = 0$ and $\text{cov}(v_s, v_t|X_1, \ldots, X_T) = \gamma_v(|s-t|)$ for some covariance function γ_v. We will also need further conditions like Masry (1996ab) on the dependence of the joint process (Y_t, X_t). Under such conditions it can be shown that our main result continues to hold, and indeed (11) is still valid as stated.

4 MODEL SELECTION

In practice, it is important to choose good values of the bandwidths as well as the truncation parameter τ. For the bandwidth h_0 in preliminary estimation, we may simply choose h_0 to be, say, h_1^δ for some $\delta > 1$ for convenience. A more complicate choice might be derived from higher order expansions of the estimator. If we look at the higher order effects, the leading bias and variance terms in $\widetilde{m}(x)$ are of order h_1^q and $T^{-1/2}h_1^{-d/2}$, and the second order terms are of orders h_0^q, $T^{-1}h_0^{-d/2}h_1^{-d/2}$, and $T^{-1/2}h_0^{-d/2}h_1^q$. Balancing the leading terms gives us an optimal order of $T^{-1/(2q+d)}$ for h_1 (see formula below). Given h_1, we may choose h_0 to balance the second order terms, giving order of $T^{-1/(2q+d)}h_1^{2q/(2q+d)}$ for h_0.

For the truncation parameter τ, in practice we may use various selection criteria such as AIC and BIC in autoregression (7). If we consider an autoregression on the true u_t, in the case where u_t is actually generated by a finite order autoregression, the order selection based on the BIC criterion is consistent and thus might be preferred. However, if the true model is not a finite order autoregression, AIC may be preferred since it leads to asymptotically efficient choice of optimal order in the class of some projected infinite order autoregressive processes. Let $RSS_T(\tau)$ be the residual sum of squares of the autoregression (7), then if we use the Akaike criterion, we choose τ that minimizes

$$\log \frac{RSS_T(\tau)}{T} + \frac{2\tau}{T}.$$

Or, if we consider the BIC criterion, we choose τ that minimizes

$$\log \frac{RSS_T(\tau)}{T} + \frac{\tau \log T}{T}.$$

For bandwidth h_1, if our object is to find a point estimate we may choose h_1 to minimize the mean squared error. From our analysis we know that the leading terms in $E\left[\widetilde{m}(x) - m(x)\right]^2$ are $h_1^{2q}\mu_q(K)^2\mathcal{B}(x)^2$ and $T^{-1}h_1^{-d}\sigma_\varepsilon^2 \left\|K\right\|^2 / f_X(x)$. Minimizing the leading mean squared error us the conventional optimal bandwidth choice of order $T^{-1/(2q+d)}$:

$$h_1^{opt} = \left[\frac{d}{2q}\frac{\|K\|^2}{\mu_q(K)^2}\frac{\sigma_\varepsilon^2}{f_X(x)\mathcal{B}(x)^2}\right]^{-1/(2q+d)} T^{-1/(2q+d)}. \qquad (12)$$

This formula is identical to the formula for the conventional kernel estimators except that the smaller variance σ_ε^2 replaces the usual σ_u^2, so that any plug-in method defined for the usual estimators can be easily applied here with a simple modification. For example, a nonparametric plug-in method can then be applied to estimate $\mathcal{B}(x)$ and $f_X(x)$. Alternatively, a 'rule-of-thumb' approach as that defined in Fan and Gijbels (1996, p111) would appear to be attractive in practice.

Another convenient approach to global bandwidth choice is cross-validation. Denoting the residual sum of squares corresponding to bandwidth h_1 as

$$p(h_1) = \frac{1}{T}\sum_{t=1}^{T}\left[\underline{\widehat{Y}}_t - \widetilde{m}_{h_1}(X_t)\right]^2 \pi(X_t),$$

where $\pi(X_t)$ is a weight function introduced to allow elimination (or reduction) of boundary effects, we multiply $p(h_1)$ by a correction factor $\Xi(T^{-1}h_1^{-d}K_1(0)/\widehat{f}_X(X_t))$, which penalizes values of h_1 too low. Thus, we may select h_1 based on minimizing the following generalized cross-validation :

$$G(h_1) = \frac{1}{T}\sum_{t=1}^{T}\left[\underline{\widehat{Y}}_t - \widetilde{m}_{h_1}(X_t)\right]^2 \Xi(T^{-1}h_1^{-d}K_1(0)/\widehat{f}_X(X_t))\pi(X_t).$$

For candidates of the correction function Ξ, see, e.g., Härdle (1990). If, say, we choose the Akaike's information criterion (Akaike 1974), $\Xi(u) = \exp(2u)$.

5 EFFICIENT ESTIMATION

We now discuss how we can improve the efficiency of our estimator even more and to approach a sort of GLS bound. There are two ways of doing this. The first approach is based on the backfitting type of methodology. Recall that

$$a(L)Y_t = a(L)m(X_t) + \varepsilon_t,$$

where ε_t is an uncorrelated sequence. Suppose that the coefficients $a(L)$ are known so we can define the variable $a(L)Y_t$. Then we have an infinite order additive regression on the right hand side with certain restrictions on the terms. From this representation we can in principle apply the 'backfitting' methodology of Linton and Mammen (2002) and proceed to estimation of m by an iterative smoother. Consider the special case where the error process is AR(1), i.e.,

$$u_t = au_{t-1} + \varepsilon_t,$$

where ε_t are i.i.d. mean zero and finite variance. Then, letting $Z_t(a) = Y_t - aY_{t-1}$ we have

$$Z_t(a) = m(X_t) - am(X_{t-1}) + \varepsilon_t.$$

For each given a this is an additive model, Hastie and Tibshirani (1991), with a specific restriction on the component functions that their ratio is proportional to a. Linton and Mammen (2002) analyzes a similar problem and proposes a method of estimation based on backfitting and then profiled likelihood to obtain estimates of a. This method works quite nicely in simple models but is less satisfactory when the error process is a general $ARMA(p,q)$ because the many unknown parameters in $a(L)$ make the algorithm with the estimated parameters numerically unstable.

It turns out that the following alternative yet more convenient approach is just as efficient. Notice that for each j where $a_j \neq 0$, we can rewrite (4) as follows

$$\underline{Y}_t^j = m(X_{t-j}) + \frac{1}{a_j}\varepsilon_t, \tag{13}$$

where

$$\underline{Y}_t^j = \frac{1}{a_j}\left[a(L)Y_t - \sum_{k\neq j}^{\infty} a_k m(X_{t-k})\right].$$

Given some estimate of $\underline{Y}_t^j$, denoted $\widehat{\underline{Y}}_t^j$, we can now smooth this against X_{t-j}, call the resulting estimator $\widetilde{m}_j(x)$. Then we have under the same conditions as above that $\widetilde{m}_j(x)$ has asymptotic variance $\sigma_\varepsilon^2/a_j^2$ for any j where $a_j \neq 0$. Furthermore $\widetilde{m}_j(x), \widetilde{m}_k(x)$ will be asymptotically independent. By combining the estimators we can improve efficiency: specifically, let

$$\widetilde{m}_{eff}(x) = \sum_{j=0}^{\tau} \omega_j \widetilde{m}_j(x),$$

where

$$\omega_j = \frac{a_j^2}{\sum_{j=0}^{\tau} a_j^2}.$$

In practice, one has to use estimated weights, i.e., replace a_j by $\widetilde{a}_j$.[2] It can be shown that

$$\sqrt{Th^d}[\widetilde{m}_{eff}(x) - m(x) - h^q \mu_q(K)\mathcal{B}(x)] \Longrightarrow N\left(0, \frac{\sigma_\varepsilon^2}{\sum_{j=0}^{\infty} a_j^2} \frac{\|K\|^2}{f_X(x)}\right).$$

[2] See Chen and Linton (2001) for a discussion of this approach to efficiency. In parametric models, this would be called minimum distance or minimum chi-squared.

Therefore, because $a_0, c_0 = 1$ we have

$$\frac{\text{avar}[\widetilde{m}_{eff}(x)]}{\text{avar}[\widehat{m}(x)]} = \frac{1}{\sum_{j=0}^{\infty} a_j^2 \sum_{j=0}^{\infty} c_j^2} \leq \frac{\text{avar}[\widetilde{m}(x)]}{\text{avar}[\widehat{m}(x)]} = \frac{1}{\sum_{j=0}^{\infty} c_j^2} \leq 1.$$

We expect that avar$[\widetilde{m}_{eff}(x)]$ provides a lower bound achievable by this sort of method. In the AR(1) case, the asymptotic variance of $\widetilde{m}(x)$ is $(\|K\|^2 / f_X(x))\sigma_\varepsilon^2/(1-a^2)$, while that of $\widetilde{m}_{eff}(x)$ is $(\|K\|^2 / f_X(x))\sigma_\varepsilon^2/(1+a^2)$. Compare this with the linear regression model $Y_t = \beta X_t + u_t$, where X_t is an i.i.d. process with zero mean. The variance of the OLS estimator of βx is $(x^2/\sigma_X^2)\sigma_\varepsilon^2/(1-a^2)$ and of the GLS estimator of βx is $(x^2/\sigma_X^2)\sigma_\varepsilon^2/(1+a^2)$.[3] This is suggestive that our efficient estimator is like GLS and can't be beaten on these terms.

In practice, the gain of $\widetilde{m}_{eff}(x)$ over $\widetilde{m}(x)$ may not be so great in comparison with the gain of $\widetilde{m}(x)$ over $\widehat{m}(x)$. For example, in the AR(1) case, the improvement of $\widetilde{m}(x)$ over the usual kernel smoother $\widehat{m}(x)$ can be arbitrarily large, but $\widetilde{m}_{eff}(x)$ can only have at best half the variance of $\widetilde{m}(x)$. Therefore, it may be that in practice the benefit from computing $\widetilde{m}_{eff}(x)$ may be exceeded by its small sample cost. We investigate this in the simulation experiments below.

One final comment on the relative advantage of our 'ad hoc' approach to efficiency relative to the 'backfitting' method of Mammen, Linton, and Nielsen (1999) and Linton and Mammen (2002). In the two different situations of these cited papers, there is either no alternative estimator, or the alternative estimator requires higher dimensional smoothing operations [e.g., the marginal integration approach of Linton and Nielsen (1995)]. In the setting of our paper, there exist many consistent estimators of m, and all of the proposed estimators, including our own, rely on smoothing operations with the same number of covariates. Therefore, the backfitting methodology has no particular advantage here.

6 NUMERICAL RESULTS

6.1 Simulations

We investigate the performance of our procedure on simulated data. We have not tried to optimize the performance of either the conventional kernel estimator or our own more efficient modifications. Rather, we have taken what are fairly common choices, in real applications, of bandwidth etc., and demonstrate that even with these implementations there are finite sample gains to be made.

In the design we consider a wide range of time series specifications for the residual process u_t, including AR(1), AR(2), MA(1), MA(2), and ARMA(1,1) processes with different parameter values.

[3] There are some differences though. First, the variance of the nonparametric estimators depend on the covariate density at the point of interest [and the kernel and bandwidth of course]. Second, the nonparametric estimators have variance that does not depend on the correlation properties of the covariate process and the variance of the standard kernel procedure doesn't even depend on the correlation of the error process, although our modified estimators do depend on this quantity indirectly. Interestingly, the effect on the estimator variance is through the sum of squared coefficients $\sum_j c_j^2$ and $\sum_j a_j^2$ rather than through the covariance function of u_t, which is proportional to $\sum_k c_j c_{j+k}$.

For convenience, we write the residual process in the form of an $ARMA(p,q)$ process with p and q less than or equal to 2:

$$u_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-1} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2}$$

with ε_t i.i.d. $N(0,1)$. We examined the time series for various combinations of different parameter values that specified in the tables below.

For the regression function, in the first design we took $m(x) = 0$ throughout, X_t i.i.d. $U[-1,1]$. In our efficient estimator we consider both AR(1) and AR(2) prewhitening. The AR parameters in the prewhitening process are estimated by least squares. We considered four sample sizes: $T = 100, 200, 500, 1000$. The number of replications is 200.

We investigate the proposed efficient estimator $\widetilde{m}(x)$ given by (8), as well as the estimator $\widetilde{m}_{eff}(x)$ considered in Section 6. We compare these estimators with the conventional kernel estimator $\breve{m}(x)$. We chose exactly the same kernel and bandwidth in all these three estimators. In particular, we use the fourth order kernel $K(u) = 15(7u^4 - 10u^2 + 3)_+/32$ and bandwidth $h = 1.06 s_X T^{-1/5}$, where s_X is the sample standard deviation of $X_1, \ldots, X_T$. [other kernels are also tried and qualitatively similar results were obtained]. For the preliminary estimation (to obtain the residuals), we use a local polynomial estimation of order 3. Below we report the relative efficiency [the ratio of average squared errors over the 200 replications] for different sample sizes and ARMA parameters. We consider estimation at the point $x = 0$.

Tables 1-4 (corresponding to different sample sizes) report the relative efficiency (the ratio of average squared errors) for the case that an AR(2) prewhitening was used (lag length was set at 2). Various combinations of parameter values were examined. In these tables, Column "RE1" reports the Relative Efficiency of the proposed efficient estimator $\widetilde{m}(x)$ over the conventional estimator $\breve{m}(x)$. Column "RE2" reports the Relative Efficiency of the efficient estimator $\widetilde{m}_{eff}(x)$ over the conventional estimator $\breve{m}(x)$. For comparison purpose, we also provide the infeasible theoretical asymptotic relative efficiency calculated based on the asymptotic variances of $\widetilde{m}(x)$ $(\sigma_\varepsilon^2 \|K\|^2 / f_X(x))$ and $\breve{m}(x)$ $(\sigma_u^2 \|K\|^2 / f_X(x))$, this is reported as "RE0".

We also considered an AR(1) prewhitening and reported the results in Tables 5-8. Note that when the underlying process has a nontrivial MA part, our method is likely to be quite far from matching the true autocorrelation structure in the errors. Nevertheless, even in those cases there are positive results.

In the second design we took $m(x) = x$, where X_t are again i.i.d. $U[-1,1]$, and considered estimation of a range of x. The same sample sizes and number of replications as in the first design were used.

The proposed estimation can be applied to other smoothing procedures such as local polynomial method. In this case, the proposed efficient estimator $\widetilde{m}(x)$ is given by a local polynomial regression of $\widehat{\underline{Y}}_t$ on X_t. Similarly, we can apply local polynomial smoothing to construct $\widetilde{m}_{eff}(x)$ in Section 6. In our second design, we compare these estimators (using AR(2) prewhitening) with the conventional local polynomial estimator $\breve{m}(x)$. Local linear smoothing was used in our experiments. Again, we chose the same kernel and bandwidth in these three estimators. In particular, we use the Gaussian kernel and bandwidth $h = 1.06 s_X T^{-1/5}$. We consider estimation of $m(\cdot)$ at the sample points $X_1, \ldots, X_T$. In Tables 9 to 12, we report the relative efficiency for different sample sizes and ARMA parameters. The relative efficiency reported in Tables 9-12 are calculated based on the ratio of average squared errors over all x's and the 200 replications. Summation of squared errors (denoted as ISE) are also reported. In particular, ISE0, ISE1 and ISE2 give the sum of squared errors of the conventional local linear estimator $\breve{m}(x)$, the proposed efficient estimators $\widetilde{m}(x)$ and $\widetilde{m}_{eff}(x)$.

Some general conclusion can be found from the simulation experiments:

(1). The results show that the relative efficiency improves with sample size - there is likely a considerable small sample effect that is dominating in this range of parameters, and this requires a very large sample indeed before the asymptotic predictions become reality. Nevertheless, in most cases apart from i.i.d. (all parameters are zeros) our estimator improves on the standard kernel procedure.

(2). In general, the more serial correlation, the larger efficiency gain is achieved from our prewhitening procedure. However, consider the AR(1) case for example, note that the relative efficiency first improves as the AR coefficient increases and then disimproves as it approaches one. This is partly due to the large downward bias in estimating α in this region. We could perhaps improve the relative efficiency by taking a larger bandwidth in the second step as would be permitted by our theory.

(3). Both $\widetilde{m}(x)$ and $\widetilde{m}_{eff}(x)$ improves the estimation in the presence of serial correlation, especially for large sample sizes, but none of them dominates the other. It seems that $\widetilde{m}(x)$ performs slightly better than $\widetilde{m}_{eff}(x)$ when the true error process is actually an AR process. This is intuitive because an AR prewhitening was used. But different results were obtained when the error terms are MA processes.

6.2 Application

We apply the proposed estimation procedure to stock return data on cross-market feedback effect. There have been some studies of the effect of one market on another, specially the impact of North American markets on the markets of other countries. In this application, we investigate the effect of returns on the S&P500 index on the subsequent volatility of the FTSE100 index. We estimate the following model

$$r_{UK,t}^2 = m(r_{US,t-1}) + u_{UK,t} \tag{14}$$

on both daily and weekly data. With this frequency of data the means of $r_{UK,t}, r_{US,t-1}$ are small and not modelling them does not make much difference to the results. The function m describes the response of UK volatility to the returns on the US market in the day before. We might expect an asymmetric response whereby negative returns in the US raise the volatility of the UK market by more than positive returns, following work of Nelson (1991).

Our data sets are as follows: the weekly data are from April 2, 1984 to April 8, 2002, with 942 observations in total. The daily data starts from April, 2, 1984, and ends at April 17, 2002, with 4624 observations. We first estimated (14) by the standard kernel estimator; the correlograms in Figures 1a and 1b show that there is quite a bit of structure left in the error terms, more so in the daily data for sure.

We then fitted an AR(p) model to the residuals where p was chosen by BIC criterion and then computed our prewhitened estimator. We report the autoregression estimates $\widehat{a}_j$ and the choices of truncation parameter τ in Table 13 for the case $h = 1.06 s_X T^{-1/5}$. For the weekly data, Figures 2a, 2b, 2c and 2d show the prewhitened estimators in comparison with the standard estimator. From Figure 2a to 2d, we used the following bandwidth choices $h_j = \delta_j s_X T^{-1/5}$ with $h_j = 0.66$, 1.66, 2.66, and 3.66, for $j = 1, 2, 3, 4$. Thus, these graphics provide estimates of the impact function $m(\cdot)$ from the case of under smoothing to the case of oversmoothing. In each figure, we give the conventional and prewhitened estimates using the same bandwidth (h_j) and the prewhitened estimate using a smaller bandwidth $h_{jb} = \left[1 - \sum_{j=1}^{\tau} \widehat{a}_j^2\right]^{1/5} \delta_j s_X T^{-1/5}$.

We also show in figures 3a, 3b, 3c and 3d our prewhitened estimator for the weekly data along with 95% confidence bands using the formula (11). Again, the bandwidth choices are the same as those in figures 2a, 2b, 2c and 2d.

The daily data gave qualitatively similar results, and we report the prewhitened estimators in comparison with the standard estimator in Figures 4a to 4d, and the prewhitened estimator with 95% confidence bands in Figures 5a to 5d, where the bandwidth choices are parallel to those in figures 2a to 2d.

The basic shape of the function m is certainly asymmetric. As expected, negative US returns are generally associated with upward revisions of the conditional volatility in the UK market, while positive US returns are associated with smaller revisions in the UK market. The presence of asymmetric cross-market feedback effect on volatility is most apparent during a market crisis when large declines in stock prices in the US market are associated with a significant increase in the UK market volatility. From these graphics, we see that for small return shocks in the US market, the UK volatility does not change very much. However, as the magnitude of a negative US shock increases, the impact on the UK volatility increases dramatically.

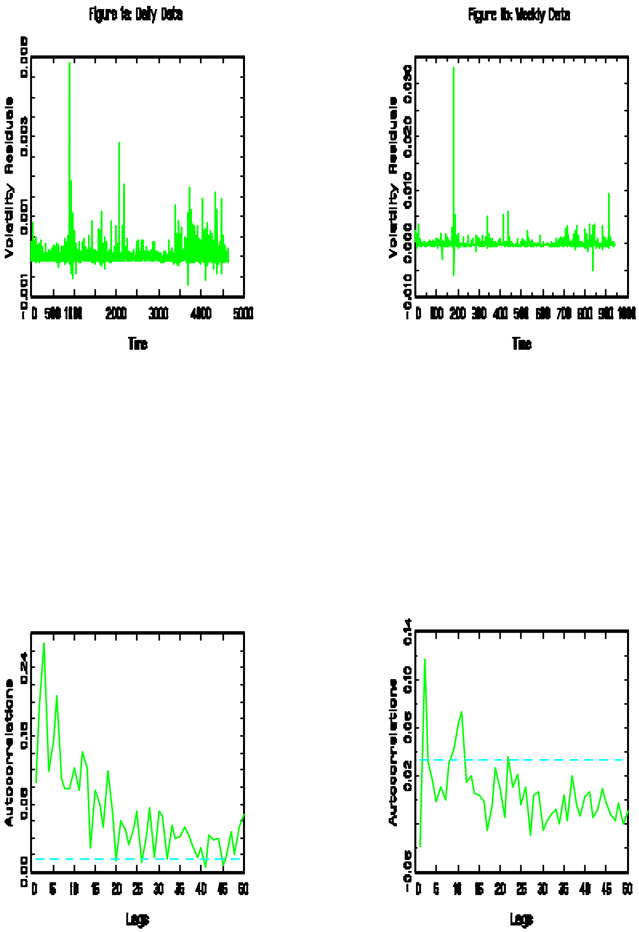
Figure 1a: Daily Data
Figure 1b: Weekly Data
Volatility Residuals
Time
Autocorrelations
Lags

Figure 1:

7 CONCLUSIONS AND GENERALIZATIONS

We expect that the numerical performance of our method can be improved in small samples. There are a number of things to work on. First, better bandwidth choice should make a big difference to the goodness of fit of our method. Second, it may be that iterating the procedure can confer benefits through more accurate estimates of the autoregressive coefficients. Along this line, it may be that recentering the residuals and using quasi likelihood methods might also bring improvements.

The proposed estimation procedure may also be generalized to semiparametric models like partial linear regression models or single index models in which there is interest in estimating the nonparametric function in the presence of serial correlation. Typically, the parametric estimates do not affect the distribution of the nonparametric functions, so the procedures and results are rather obvious to state.

A Proof of Theorems

We use $\|\bullet\|$ to denote the Euclidean norm of $\bullet$, C to signify a generic positive constant whose exact value may vary from case to case. We denote $\phi(x, y, z, \ldots)$ as a general function whose exact form may change from case to case. For two random variables X_T, Y_T, we say that $X_T \simeq Y_T$ whenever $X_T = Y_T(1 + o_p(1))$ as $T \to \infty$.

Preliminaries The asymptotic properties of local polynomial estimator have been well developed and documented, see, e.g., Fan and Gijbels (1996) and Masry (1996ab) and the references therein. For convenience, we first give some general definitions for our local polynomial kernel nonparametric regression estimators. Let $N_\ell = \begin{pmatrix} \ell + d - 1 \\ d - 1 \end{pmatrix}$ be the number of distinct d-tuples j with $|j| = \ell$. Arrange these N_ℓ d-tuples as a sequence in a lexicographical order (with highest priority to last position so that $(0, \ldots, 0, \ell)$ is the first element in the sequence and $(\ell, \underbrace{0, \ldots,}0)$ the last element) and let ϕ_ℓ^{-1} denote this one-to-one map. Arrange the distinct values of $(D^{\mathbf{k}})(m)$, $0 \leq |\mathbf{k}| \leq p$, as a column vector of dimension $N \times 1$, where $N = \sum_{\ell=0}^{p} N_\ell \times 1$, where the i^{th} element of that vector is obtained by the following relation $i = \phi_{|j|}^{-1}(j) + \sum_{k=0}^{|j|-1} N_k$. Similarly, arrange the vector $(D^{\mathbf{k}})(m)$. For each j with $0 \leq |j| \leq 2p$, let

$$\mu_j(K_0) = \int_{\mathbb{R}^d} u^j K_0(u) du, \ \nu_j(\mathcal{K}) = \int_{\mathbb{R}^d} u^j K_0^2(u) du,$$

and define the $N \times N$ dimensional matrices M and Γ and $N \times 1$ vector B by

$$M = \begin{bmatrix} M_{0,0} & M_{0,1} & \cdots & M_{0,p} \\ M_{1,0} & M_{1,1} & \cdots & M_{1,p} \\ \vdots & & & \vdots \\ M_{p,0} & M_{p,1} & \cdots & M_{p,p} \end{bmatrix}, \ \Gamma = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & \Gamma_{0,p} \\ \Gamma_{1,0} & \Gamma_{1,1} & \cdots & \Gamma_{1,p} \\ \vdots & & & \vdots \\ \Gamma_{p,0} & \Gamma_{p,1} & \cdots & \Gamma_{p,p} \end{bmatrix}, \quad B = \begin{bmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{bmatrix}, \qquad (15)$$

where $M_{i,j}$ and $\Gamma_{i,j}$ are $N_i \times N_j$ dimensional matrices whose (ℓ, m) element are, respectively, $\mu_{\phi_i(\ell)+\phi_j(m)}$ and $\nu_{\phi_i(\ell)+\phi_j(m)}$. Note that the elements of the matrices M and Γ are simply multivariate moments of the kernel K_0 and K_0^2, respectively. Define also we denote

$$M^{-1} = \begin{bmatrix} M^{0,0} & M^{0,1} & \cdots & M^{0,p} \\ M^{1,0} & M^{1,1} & \cdots & M^{1,p} \\ \vdots & & & \vdots \\ M^{p,0} & M^{p,1} & \cdots & M^{p,p} \end{bmatrix}.$$

Finally, arrange the N_{p+1} elements of the derivatives $(1/j!)(D^j m)(x)$ for $|j| = p+1$ as a column vector $\mathcal{D}_{p+1}(x; m)$ using the lexicographical order introduced earlier.

Minimizing (9) with respect to $b_{\mathbf{k}}$ gives an estimate $\hat{b}_{\mathbf{k}}(x)$ and $\widetilde{m}(x) = \hat{b}_0(x) = e_1' M_T^{-1} \Psi_n$, where $e_1 = (1, 0, \ldots, 0)'$ is the vector with the one in the first position, $M_T(x)$ and $\Psi_T(x)$ are symmetric $N \times N$ $(N = \sum_{\ell=0}^{p} N_\ell \times 1)$ matrix and $N \times 1$ dimensional column vector respectively and are defined as

$$M_T(x) = \begin{bmatrix} M_{T,0,0}(x) & M_{T,0,1}(x) & \cdots & M_{T,0,p}(x) \\ \vdots & M_{T,1,1}(x) & \cdots & M_{T,1,p}(x) \\ \vdots & & \ddots & \vdots \\ M_{T,p,0}(x) & \cdots & \cdots & M_{T,p,p}(x) \end{bmatrix}, \ \Psi_T(x) = \begin{bmatrix} \Psi_{T,0}(x) \\ \Psi_{T,1}(x) \\ \vdots \\ \Psi_{T,p}(x) \end{bmatrix},$$

where $M_{T,|j|,|k|}(x)$ is a $N_{|j|} \times N_{|k|}$ dimensional submatrix with the (l, r) element given by

$$\left[M_{T,|j|,|k|}\right]_{l,r} = \frac{1}{Th_0^d} \sum_{i=1}^{T} \left(\frac{x - X_i}{h_0}\right)^{\phi_{|j|}(l) + \phi_{|k|}(r)} K_0\left(\frac{x - X_i}{h_0}\right),$$

and $\Psi_{T,|j|}(x)$ is a $N_{|j|}$ dimensional subvector whose r-th element is given by

$$\left[\Psi_{T,|j|}\right]_r = \frac{1}{Th_0^d} \sum_{i=1}^{T} \left(\frac{x - X_i}{h_0}\right)^{\phi_{|j|}(r)} K_0\left(\frac{x - X_i}{h_0}\right) Y_i.$$

The estimate of $m(x)$ is given by $\widetilde{m}(x) = e_1 M_T^{-1} \Psi_T$ and its bias and variance effects can be written as $\widetilde{m}(x) - m(x) = e_1' M_T^{-1}(x) U_T(x) + e_1' M_T^{-1}(x) B_T(x)$. The stochastic term $U_T(x)$ and the bias term $B_T(x)$ are $N \times 1$ vectors

$$U_T(x) = \begin{bmatrix} U_{T,0}(x) \\ U_{T,1}(x) \\ \vdots \\ U_{T,p}(x) \end{bmatrix}, \ B_T(x) = \begin{bmatrix} B_{T,0}(x) \\ B_{T,1}(x) \\ \vdots \\ B_{T,d}(x) \end{bmatrix},$$

where $U_{T,l}(x)$ and $B_{T,l}(x)$ are defined similarly as $\Psi_{T,l}(x)$ so that $U_{T,|j|}(x)$ and $B_{T,|j|}(x)$ are a $N_{|j|}$ dimensional subvectors whose r-th elements are given by

$$\left[U_{T,|j|}\right]_r = \frac{1}{Th_0^d} \sum_{i=1}^{n} \left(\frac{x - X_i}{h_0}\right)^{\phi_{|j|}(r)} K_0\left(\frac{x - X_i}{h_0}\right) u_i$$

and

$$\left[B_{T,|j|}\right]_r = \frac{1}{Th_0^d}\sum_{i=1}^{n}\left(\frac{x-X_i}{h_0}\right)^{\phi_{|j|}(r)} K_0\left(\frac{x-X_i}{h_0}\right)\Delta_i(x),$$

where $\Delta_i(x) = m(X_i) - \frac{1}{\mathbf{k}!}\sum_{0\le|\mathbf{k}|\le p}(D^{\mathbf{k}}m)(x)(X_i-x)^{\mathbf{k}}$.

Under our assumptions given in the paper, we have the following uniform convergence results:

$$\sup_{x\in\mathcal{X}}|M_T(x) - f(x)M| = O_p(h_0 + T^{-1/2}h_0^{-d/2}\log T)$$

$$\sup_{x\in\mathcal{X}}|\widetilde{m}(x) - m(x)| = O_p(h_0^{p+1} + T^{-1/2}h_0^{-d/2}\log T), \tag{16}$$

which follow from the results of Masry (1996ab).

Proof of Theorem 1 To be comparable with notation in the feasible estimator $\widetilde{m}$ and Theorem 2, we conduct our proof using the notation K_1 for the kernel and h_1 for the bandwidth. Write

$$\begin{aligned}\overline{m}(x) &= m(x) + \frac{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)[m(X_t)-m(x)]}{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)} + \frac{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)\varepsilon_t}{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)} \\ &\equiv m(x) + \overline{B}_x + \overline{V}_x.\end{aligned}$$

First note that

$$\overline{V}_x = \frac{\frac{1}{Th_1^d}\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)\varepsilon_t}{\widehat{f}_X^1(x)} = V_x(1+o_p(1)),$$

where

$$V_x = \frac{1}{Th_1^d}\sum_{t=1}^{T}\frac{K_1\left(\frac{x-X_t}{h_1}\right)\varepsilon_t}{f_X(x)},$$

by the law of large numbers applied to $T^{-1}h_1^{-d}\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)$. Since $f_X(x)>0$, we can apply the central limit theorem to V_x:

$$\frac{1}{f_X(x)}\frac{1}{T^{1/2}h_1^{d/2}}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\varepsilon_t \xrightarrow{d} N\left(0, \frac{\sigma_\varepsilon^2\|K_1\|^2}{f_X(x)}\right).$$

Similarly,

$$\overline{B}_x = B_x(1+o_p(1)),$$

where

$$\begin{aligned}B_x &= \frac{1}{Th_1^d}\sum_{t=1}^{T}\frac{K_1\left(\frac{x-X_t}{h_1}\right)[m(X_t)-m(x)]}{f_X(x)} \\ &\simeq h_1^q\mu_q(K_1)\sum_{p+r=q,1\le p\le q,0\le r\le q}\frac{1}{p!r!}m^{(p)}(x)\frac{f_X^{(r)}(x)}{f_X(x)} \\ &= h_1^q\mu_q(K_1)\mathcal{B}(x).\end{aligned}$$

For conventional quadratic kernel, $q = 2$, and the bias term is simply $\frac{1}{2}\mu_2(K_1)[m''(x) + 2m'(x)\frac{f'}{f}(x)]$.

Thus,

$$\sqrt{Th_1^d}[\overline{m}(x) - m(x) - h_1^q \mu_q(K_1)\mathcal{B}(x)] \Longrightarrow N\left(0, \frac{\sigma_\varepsilon^2 \|K_1\|^2}{f_X(x)}\right).$$

■

Proof of Theorem 2 We decompose $\widetilde{m}(x)$ into $\overline{m}(x)$ plus error terms coming from the preliminary estimation and the truncation, and show that these terms are small order terms. First we write

$$\begin{aligned}
\underline{\widehat{Y}}_t &= Y_t - \sum_{j=1}^{\tau} \widehat{a}_j \left(Y_{t-j} - \widehat{m}(X_{t-j})\right) \\
&= Y_t - \sum_{j=1}^{\infty} a_j u_{t-j} + \sum_{j=\tau+1}^{\infty} a_j u_{t-j} - \sum_{j=1}^{\tau} (\widehat{a}_j - a_j) u_{t-j} \\
&\quad + \sum_{j=1}^{\tau} a_j \left(\widehat{m}(X_{t-j}) - m(X_{t-j})\right) + \sum_{j=1}^{\tau} (\widehat{a}_j - a_j) \left(\widehat{m}(X_{t-j}) - m(X_{t-j})\right).
\end{aligned}$$

Substituting the above expression into (8), we have

$$\begin{aligned}
\widetilde{m}(x) &= \overline{m}(x) + \frac{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right) \sum_{j=\tau+1}^{\infty} a_j u_{t-j}}{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)} - \frac{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right) \sum_{j=1}^{\tau} (\widehat{a}_j - a_j) u_{t-j}}{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)} \\
&\quad + \frac{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right) \sum_{j=1}^{\tau} a_j \left(\widehat{m}(X_{t-j}) - m(X_{t-j})\right)}{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)} \\
&\quad + \frac{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right) \sum_{j=1}^{\tau} (\widehat{a}_j - a_j) \left(\widehat{m}(X_{t-j}) - m(X_{t-j})\right)}{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)} \\
&= \overline{m}(x) + Q_{T1} - Q_{T2} + Q_{T3} + Q_{T4}.
\end{aligned}$$

We analyze the asymptotic properties of Q_{Tj}, $j = 1, ..., 4$, in Lemmas A1 to A4, which are key results for the proof of the Theorem.

Lemma A1. *Under Assumptions 1 to 7*

$$Q_{T1} = o_p(T^{-1/2} h_1^{-d/2}).$$

Proof of Lemma A1. Q_{T1} is of smaller order because of the tail properties of the summable sequence a_j. Specifically,

$$Q_{T1} = \frac{\frac{1}{Th_1^d} \sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right) \sum_{j=\tau+1}^{\infty} a_j u_{t-j}}{\widehat{f}_X^1(x)},$$

where

$$\widehat{f}_X^1(x) = \frac{1}{Th_1^d} \sum_{t=1}^{T} K_1\left(\frac{x - X_t}{h_1}\right)$$

is the conventional nonparametric density estimator that is uniformly consistent. First note that

$$Q_{T1} = \frac{\frac{1}{Th_1^d} \sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right) \sum_{j=\tau+1}^{\infty} a_j u_{t-j}}{f_X(x)} (1 + o_p(1)),$$

by the law of large numbers applied to $T^{-1}h_1^{-d}\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)$.

Since $f_X(x) > 0$, we only need to verify the order of

$$\frac{1}{Th_1^d} \sum_{t=1}^{T} K_1\left(\frac{x - X_t}{h_1}\right) \sum_{j=\tau+1}^{\infty} a_j u_{t-j}.$$

Notice that it has mean zero and

$$\begin{aligned}
& \operatorname{var}\left[\frac{1}{Th_1^d} \sum_{t=1}^{T} K_1\left(\frac{x - X_t}{h_1}\right) \sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right] \\
= & \left(\frac{1}{Th_1^d}\right)^2 \mathrm{E}\left\{\sum_{t=1}^{T} K_1\left(\frac{x - X_t}{h_1}\right) \sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right\}\left\{\sum_{s=1}^{T} K_1\left(\frac{x - X_s}{h_1}\right) \sum_{i=\tau+1}^{\infty} a_i u_{s-i}\right\} \\
= & \left(\frac{1}{Th_1^d}\right)^2 \mathrm{E}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T} K_1\left(\frac{x - X_t}{h_1}\right) K_1\left(\frac{x - X_s}{h_1}\right) \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_j a_i u_{s-i} u_{t-j}\right\} \\
= & \left(\frac{1}{Th_1^d}\right)^2 \mathrm{E}\left\{\sum_{t=s=1}^{T} K_1\left(\frac{x - X_t}{h_1}\right)^2 \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_j a_i u_{t-i} u_{t-j}\right\} + \\
& \left(\frac{1}{Th_1^d}\right)^2 \mathrm{E}\left\{\sum_{t=1}^{T}\sum_{s\neq t, s=1}^{T} K_1\left(\frac{x - X_t}{h_1}\right) K_1\left(\frac{x - X_s}{h_1}\right) \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_j a_i u_{s-i} u_{t-j}\right\} \\
= & \left(\frac{1}{Th_1^d}\right)^2 \left\{\sum_{t=s=1}^{T} \mathrm{E}K_1\left(\frac{x - X_t}{h_1}\right)^2 \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_i a_j \gamma_u(|i-j|)\right\} + \\
& \left(\frac{1}{Th_1^d}\right)^2 \left\{\sum_{t=1}^{T}\sum_{s\neq t, s=1}^{T} \mathrm{E}K_1\left(\frac{x - X_t}{h_1}\right) K_1\left(\frac{x - X_s}{h_1}\right) \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_j a_i \gamma_u(|t-s+i-j|)\right\}
\end{aligned}$$

The first term is $o(T^{-1}h_1^{-d})$ because:

$$\begin{aligned}
& \left(\frac{1}{Th_1^d}\right)^2 \left\{\sum_{t=s=1}^{T} \mathrm{E}K_1\left(\frac{x - X_t}{h_1}\right)^2 \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_i a_j \gamma_u(|i-j|)\right\} \\
\leq & \left(\frac{1}{Th_1^d}\right)^2 T \cdot \mathrm{E}\left\{K_1\left(\frac{x - X_1}{h_1}\right)^2\right\}\left\{\sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_i a_j\right\} \sup_{0\leq i,j\leq\infty} |\gamma_u(|i-j|)|
\end{aligned}$$

and

(1). $\sup_{0\le j,l<\infty} |\gamma_u(|j-l|)| < \infty$, by stationarity/mixing property of u;

(2). $T \cdot \mathrm{E}K_1\left(\frac{x-X_1}{h_1}\right)^2 = O(Th_1^d)$, by a direct calculation of expectation; and

(3). $\sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_i a_j = o(1)$ as $\tau \to \infty$, by summability of $\{a_j\}_{j=1}^{\infty}$.

The second term is $o(T^{-1}h_1^{-d})$:

$$
\begin{aligned}
& \left(\frac{1}{T}\right)^2 \sum_{t=1}^{T} \sum_{s\neq t,s=1}^{T} \mathrm{E}\left[\frac{1}{h_1^{2d}} K_1\left(\frac{x-X_t}{h_1}\right) K_1\left(\frac{x-X_s}{h_1}\right)\right] \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_j a_i \gamma_u(|t-s+i-j|) \\
= & \left(\frac{1}{T}\right)^2 \sum_{t=1}^{T} \sum_{s\neq t,s=1}^{T} \left[\int \frac{1}{h_1^{2d}} K_1\left(\frac{x-y}{h_1}\right) K_1\left(\frac{x-z}{h_1}\right) f_{X,|t-s|}(y,z)dydz\right] \\
& \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_j a_i \gamma_u(|t-s+i-j|) \\
= & \left(\frac{1}{T}\right)^2 \sum_{t=1}^{T} \sum_{s\neq t,s=1}^{T} \left[\int K_1(u) K_1(v) f_{X,|t-s|}(x-uh_1, y-vh_1)dudv\right] \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_j a_i \gamma_u(|t-s+i-j|) \\
\le & C\left(\frac{1}{T}\right)^2 \sum_{t=1}^{T} \sum_{s\neq t,s=1}^{T} \sum_{i=\tau+1}^{\infty}\sum_{j=\tau+1}^{\infty} a_j a_i \gamma_u(|t-s+i-j|),
\end{aligned}
$$

where the last inequality follows from the boundedness assumption of the density and joint densities and the fact that

$$
\sup_{0\le i,j\le\infty} \left| \sum_{s\neq t,t=1}^{T} \sum_{s=1}^{T} \gamma_u(|t-s+i-j|) \right| = O(T), \tag{17}
$$

where, again, the result (17) comes from the stationarity/mixing property of u. Thus

$$
\mathrm{var}\left[\frac{1}{Th_1}\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right) \sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right] = o(T^{-1}h_1^{-d}).
$$

Therefore, the magnitude of Q_{T1} is as stated. ■

Lemma A2. *Under Assumptions 1 to 7*

$$
Q_{T2} = o_p(T^{-1/2}h_1^{-d/2}).
$$

Proof of Lemma A2. We denote

$$
\overline{A}_\tau = (U_\tau' U_\tau)^{-1} U_\tau' u = (\overline{a}_1, \ldots, \overline{a}_\tau)',
$$

where $u = (u_{\tau+1}, \ldots, u_T)'$ and U_τ is like $\widehat{U}_\tau$ with $\widehat{u}_t$ replaced by u_t, and write

$$
\widehat{a}_j - a_j = (\widehat{a}_j - \overline{a}_j) + (\overline{a}_j - a_j),
$$

i.e.

$$\widehat{A}_\tau - A_\tau = \left(\widehat{A}_\tau - \overline{A}_\tau\right) + \left(\overline{A}_\tau - A_\tau\right).$$

We first show that

$$\frac{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}(\overline{a}_j - a_j)u_{t-j}}{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)} = o_p(T^{-1/2}h_1^{-d/2}). \tag{18}$$

Denote that

$$U_{\tau t} = (u_{t-1}, \ldots, u_{t-\tau})'$$

and define the $\tau \times \tau$ matrices

$$\begin{aligned} G_\tau &= \frac{1}{T}U_\tau' U_\tau = \frac{1}{T}\sum_t U_{\tau t}U_{\tau t}' = \left(\frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}u_{t-l}\right)_{j,l} \\ \Gamma_\tau &= \frac{1}{T}E(U_\tau' U_\tau) = \frac{1}{T}\sum_t EU_{\tau t}U_{\tau t}' = \left(E(u_{t-j}u_{t-l})\right)_{j,l}. \end{aligned}$$

Then, there exists a $c > 0$ such that

$$\lambda_{\min}(\Gamma_\tau) \geq c\tau^{-\alpha}$$

for some $\alpha > 0$. Therefore,

$$\left\|\Gamma_\tau^{-1}\right\| \leq c^{-1}\tau^{\alpha},$$

and

$$\|G_\tau - \Gamma_\tau\| = O_p(Q_T), \tag{19}$$

where

$$Q_T = \sqrt{\frac{\log\log T}{T}},$$

provided $\tau \leq (\log T)^\kappa$ for some $\kappa > 0$. [Hannan and Deistler (1988, §5.3)]. Notice that

$$\overline{A}_\tau - A_\tau = G_\tau^{-1}\left[\frac{1}{T}\sum_t U_{\tau t}\left(\varepsilon_t + \sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right)\right],$$

we verify the magnitude of

$$\frac{1}{T}\sum_t U_{\tau t}\varepsilon_t, \tag{20}$$

and

$$\frac{1}{T}\sum_t U_{\tau t}\left(\sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right).$$

For the first component,

$$E\left\|\frac{1}{T}\sum_t U_{\tau t}\varepsilon_t\right\|^2 = \frac{1}{T^2}\sum_{i=1}^{\tau} E\left[\sum_t u_{t-i}\varepsilon_t\right]^2 = \frac{\tau}{T}\gamma_u(0)\sigma_\varepsilon^2 = O\left(\frac{\tau}{T}\right), \tag{21}$$

thus (20) is of order $O_p(T^{-1/2}\tau^{1/2})$. For the second component, notice that u_t is a stationary invertible process whose linear process coefficients satisfy the given summability assumption,

$$\begin{aligned} & E\left\|\frac{1}{T}\sum_t U_{\tau t}\left(\sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right)\right\|^2 \\ = & \frac{1}{T^2}\sum_{i=1}^{\tau} E\left[\sum_t u_{t-i}\left(\sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right)\right]^2 \\ = & \frac{1}{T^2}\sum_{i=1}^{\tau} E\left[\sum_{j=\tau+1}^{\infty}\sum_{l=\tau+1}^{\infty} a_j a_l \sum_t\sum_s u_{t-i}u_{t-j}u_{s-i}u_{s-l}\right]. \end{aligned}$$

Using the linear process representation of u_t, we obtain

$$\begin{aligned} & E\left[\sum_{j=\tau+1}^{\infty}\sum_{l=\tau+1}^{\infty} a_j a_l \sum_t\sum_s u_{t-i}u_{t-j}u_{s-i}u_{s-l}\right] \qquad (22) \\ = & E\left[\sum_{j=\tau+1}^{\infty}\sum_{l=\tau+1}^{\infty} a_j a_l \sum_t\sum_s \left(\sum_{r=0}^{\infty} c_r\varepsilon_{t-i-r}\right)\left(\sum_{p=0}^{\infty} c_p\varepsilon_{t-j-p}\right)\left(\sum_{g=0}^{\infty} c_g\varepsilon_{s-i-g}\right)\left(\sum_{h=0}^{\infty} c_h\varepsilon_{s-l-h}\right)\right] \\ = & \sum_{j=\tau+1}^{\infty}\sum_{l=\tau+1}^{\infty} a_j a_l \sum_t\sum_s\left(\sum_{r=0}^{\infty}\sum_{p=0}^{\infty}\sum_{g=0}^{\infty}\sum_{h=0}^{\infty} c_r c_p c_g c_h E\left[\varepsilon_{t-i-r}\varepsilon_{t-j-p}\varepsilon_{s-i-g}\varepsilon_{s-l-h}\right]\right). \end{aligned}$$

Notice that ε_i are i.i.d. with mean zero, the expectation $E\left[\varepsilon_{t-i-r}\varepsilon_{t-j-p}\varepsilon_{s-i-g}\varepsilon_{s-l-h}\right]$ is non-zero when (i) $s-i-g = s-l-h$ and $t-i-r = t-j-p$; or (ii) $s-i-g = t-i-r$ and $t-j-p = s-l-h$; or (iii) $s-i-g = t-j-p$ and $t-i-r = s-l-h$; or (iv) $s-i-g = s-l-h = t-i-r = t-j-p$. By the summability condition of $\{c_i\}_{i=0}^{\infty}$, direct calculations show that

$$E\left\|\frac{1}{T}\sum_t U_{\tau t}\left(\sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right)\right\|^2 = O\left(\tau\left[\sum_{j=\tau+1}^{\infty} a_j^2\right]\right).$$

Under Assumption 5, there exists some $0 < \lambda < 1$ such that $|a_j|$ is bounded by a constant multiple of λ^j, we have

$$\sum_{j=\tau+1}^{\infty} a_j^2 = O(\lambda^\tau).$$

Thus, under Assumption 6 that $\tau = \kappa \log T$, with appropriately chosen κ (say, $\kappa = -\ln\lambda > 0$), $\sum_{j=\tau+1}^{\infty} a_j^2 = O(T)$. Thus, combining the result of (21),

$$\left\|\frac{1}{T}\sum_t U_{\tau t}\left(\varepsilon_t + \sum_{j=\tau+1}^{\infty} a_j u_{t-j}\right)\right\| = O_p(T^{-1/2}\tau^{1/2}).$$

Giving our choice of τ, we have,for any small $\nu > 0$,

$$\left\|\overline{A}_\tau - A_\tau\right\| = o_p(T^{-1/2+\nu})$$

This concludes the first part.

Next, we show that

$$\frac{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}(\widehat{a}_j - \overline{a}_j)u_{t-j}}{\sum_{t=1}^{T} K_1\left(\frac{x-X_t}{h_1}\right)} = o_p(T^{-1/2}h_1^{-d/2}). \tag{23}$$

We have

$$\begin{aligned}
\widehat{A}_\tau - \overline{A}_\tau &= \widehat{G}_\tau^{-1}\widehat{g}_\tau - G_\tau^{-1}g_\tau \\
&= -G_\tau^{-1}[\widehat{G}_\tau - G_\tau]G_\tau^{-1}g_\tau + G_\tau^{-1}[\widehat{g}_\tau - g_\tau] \\
&\quad +\widehat{G}_\tau^{-1}[\widehat{G}_\tau - G_\tau]G_\tau^{-1}[\widehat{G}_\tau - G_\tau]G_\tau^{-1}g_\tau - \widehat{G}_\tau^{-1}[\widehat{G}_\tau - G_\tau]G_\tau^{-1}[\widehat{g}_\tau - g_\tau],
\end{aligned}$$

where

$$\begin{aligned}
\widehat{G}_\tau &= \frac{1}{T}\widehat{U}_\tau'\widehat{U}_\tau = \left(\frac{1}{T}\sum_{t=\tau+1}^{T}\widehat{u}_{t-j}\widehat{u}_{t-l}\right)_{j,l}, \\
\widehat{g}_\tau &= \frac{1}{T}\widehat{U}_\tau'\widehat{u} = \left(\frac{1}{T}\sum_{t=\tau+1}^{T}\widehat{u}_{t-j}\widehat{u}_{t}\right)_{j}, \\
g_\tau &= \frac{1}{T}U_\tau'u = \left(\frac{1}{T}\sum_{t=\tau+1}^{T}u_{t-j}u_{t}\right)_{j}.
\end{aligned}$$

Further define the $\tau \times 1$ vector

$$\gamma_\tau = \frac{1}{T}E(U_\tau'u) = (E(u_{t-j}u_t))_j\,.$$

Then,

$$\|g_\tau - \gamma_\tau\| = O_p(Q_T). \tag{24}$$

Notice that

$$\left(\widehat{G}_\tau - G_\tau\right)_{j,l} = \frac{1}{T}\sum_{t=\tau+1}^{T}(\widehat{u}_{t-j}\widehat{u}_{t-l} - u_{t-j}u_{t-l})$$

and

$$(\widehat{g}_\tau - g_\tau)_j = \frac{1}{T}\sum_{t=\tau+1}^{T}(\widehat{u}_{t-j}\widehat{u}_{t} - u_{t-j}u_{t}).$$

Now write

$$\widehat{u}_t = u_t - \widehat{V}_t - \widehat{B}_t,$$

where

$$\widehat{B}_t = e_1' M_T^{-1}(X_t) B_n(X_t),\ \widehat{V}_t = e_1' M_T^{-1}(X_t) U_n(X_t), \tag{25}$$

for short. Then

$$\begin{aligned}\widehat{u}_{t-j}\widehat{u}_{t-l} - u_{t-j}u_{t-l} &= -u_{t-j}\widehat{V}_{t-l} - u_{t-j}\widehat{B}_{t-l} - u_{t-l}\widehat{V}_{t-j} - u_{t-l}\widehat{B}_{t-j} \\ &\quad + \widehat{V}_{t-l}\widehat{V}_{t-j} + \widehat{B}_{t-j}\widehat{B}_{t-l} + \widehat{V}_{t-l}\widehat{B}_{t-j} + \widehat{B}_{t-j}\widehat{V}_{t-l}.\end{aligned}$$

Clearly,

$$\begin{aligned}&\left|\frac{1}{T}\sum_{t=\tau+1}^{T}\left(\widehat{V}_{t-l}\widehat{V}_{t-j} + \widehat{B}_{t-j}\widehat{B}_{t-l} + \widehat{V}_{t-l}\widehat{B}_{t-j} + \widehat{B}_{t-j}\widehat{V}_{t-l}\right)\right| \qquad (26)\\ \leq\ & \frac{1}{T}\sum_{t=\tau+1}^{T}\left(\left|\widehat{V}_{t-l}\right|\left|\widehat{V}_{t-j}\right| + \left|\widehat{B}_{t-j}\right|\left|\widehat{B}_{t-l}\right| + \left|\widehat{V}_{t-l}\right|\left|\widehat{B}_{t-j}\right| + \left|\widehat{B}_{t-j}\right|\left|\widehat{V}_{t-l}\right|\right) \\ \leq\ & \frac{1}{T}\sum_{t=\tau+1}^{T}\left(\left(\sup_s\left|\widehat{V}_s\right|\right)^2 + \left(\sup_s\left|\widehat{B}_s\right|\right)^2 + 2\sup_s\left|\widehat{V}_s\right|\sup_s\left|\widehat{B}_s\right|\right) \\ =\ & O_p((\log T)T^{-1}h_0^{-d} + h_0^{2q})\end{aligned}$$

by virtue of the uniform rate of convergence of the terms $\widehat{V}_s, \widehat{B}_s$ over s.

The cross-product terms require more detailed analysis. Notice that

$$\begin{aligned}&\frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\widehat{V}_{t-l} \\ =\ & \frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left[e_1' M_T^{-1}(X_{t-l})U_n(X_{t-l})\right] \simeq \frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left[e_1'\left[Mf_X(X_{t-l})\right]^{-1}U_n(X_{t-l})\right],\end{aligned}$$

and $M^{0,m}$ are $1 \times N_m$ row vectors, we have

$$\begin{aligned}&\frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left[e_1'\left[Mf_X(X_{t-l})\right]^{-1}U_n(X_{t-l})\right] \\ =\ & \frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}f_X(X_{t-l})^{-1}\sum_{m=0}^{p} M^{0,m}U_{n,m}(X_{t-l}) \\ =\ & \frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}f_X(X_{t-l})^{-1}\sum_{\kappa}\omega^{0,\kappa}\left(\frac{1}{Th_0^d}\sum_{r=1}^{T}\left(\frac{X_{t-l}-X_r}{h_0}\right)^{\kappa}K_0\left(\frac{X_{t-l}-X_r}{h_0}\right)u_r\right) \\ =\ & \sum_{\kappa}\omega^{0,\kappa}\frac{1}{T}\sum_{t=\tau+1}^{T}\sum_{r=1}^{T}\frac{1}{Th_0^d}f_X(X_{t-l})^{-1}\left(\frac{X_{t-l}-X_r}{h_0}\right)^{\kappa}K_0\left(\frac{X_{t-l}-X_r}{h_0}\right)u_{t-j}u_r,\end{aligned}$$

where $\omega^{0,\kappa}$ are elements in the first row of M^{-1} and the sum over κ is over a finite index set. Thus, notice that u_r has linear process representation $u_t = \sum_{j=0}^{\infty} c_j\varepsilon_{t-j}$, denoting

$$\frac{1}{Th_0^d}f_X(X_{t-l})^{-1}\left(\frac{X_{t-l}-X_r}{h_0}\right)^{\kappa}K_0\left(\frac{X_{t-l}-X_r}{h_0}\right)$$

as $w_{\kappa,t-l,r}$, we have

$$\begin{aligned} & \frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left[e_1'\left[Mf_X(X_{t-l})\right]^{-1}U_n(X_{t-l})\right] \\ = & \sum_{\kappa}\omega^{0,\kappa}\frac{1}{T}\sum_{t=\tau+1}^{T}\sum_{r=1}^{T} w_{\kappa,t-l,r}\left(\sum_{s=0}^{\infty}c_s\varepsilon_{t-j-s}\right)\left(\sum_{b=0}^{\infty}c_b\varepsilon_{r-b}\right) \\ = & \sum_{\kappa}\omega^{0,\kappa}\varphi_{\kappa,T,j,l}, \end{aligned}$$

where

$$\varphi_{\kappa,T,j,l} = \frac{1}{T}\sum_{t=\tau+1}^{T}\sum_{r=1}^{T} w_{\kappa,t-l,r}\left(\sum_{s=0}^{\infty}c_s\varepsilon_{t-j-s}\right)\left(\sum_{b=0}^{\infty}c_b\varepsilon_{r-b}\right).$$

In addition, notice that X and ε are independent, thus,

$$\begin{aligned} & E\left|\varphi_{\kappa,T,j,l}\right|^2 \\ = & \frac{1}{T^2}\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{g=0}^{\infty}\sum_{s=0}^{\infty}\sum_{t=\tau+1}^{T}\sum_{p=\tau+1}^{T}\sum_{r=1}^{T}\sum_{h=1}^{T} c_a c_b c_g c_s \mathrm{E}\left(w_{\kappa,t-l,r}w_{\kappa,p-l,h}\right)\mathrm{E}\left(\varepsilon_{t-j-s}\varepsilon_{p-j-g}\varepsilon_{r-b}\varepsilon_{h-a}\right). \end{aligned}$$

Since ε's are i.i.d., the above expectation is non-zero when (i) $r-b=h-a$ and $t-s=p-g$; or, (ii) $r-b=t-j-s$ and $h-a=p-j-g$; or, (iii) $r-b=p-j-g$ and $h-a=t-j-s$; or, (iv) $h-a=r-b=t-j-s=p-j-g$. Simple calculations show that

$$\varphi_{\kappa,T,j,l} = \frac{1}{T}\sum_{t=\tau+1}^{T}\sum_{r=1}^{T} w_{\kappa,t-l,r}\left(\sum_{s=0}^{\infty}c_s\varepsilon_{t-j-s}\right)\left(\sum_{b=0}^{\infty}c_b\varepsilon_{r-b}\right) = O_p\left(\frac{1}{T}\right). \tag{27}$$

For example, if $r-b=h-a$ and $t-s=p-g$, we have the corresponding expectation

$$\begin{aligned} & \frac{1}{T^2}\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{g=0}^{\infty}\sum_{s=0}^{\infty}\sum_{t=\tau+1}^{T}\sum_{r=1}^{T} c_a c_b c_g c_s \mathrm{E}\left(w_{\kappa,t-l,r}w_{\kappa,t-s-g-l,h}\right)\mathrm{E}\left(\varepsilon_{t-j-s}^2\varepsilon_{r-b}^2\right) \\ = & \frac{1}{T^2}\sigma_\varepsilon^4\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\sum_{g=0}^{\infty}\sum_{s=0}^{\infty}\sum_{t=\tau+1}^{T}\sum_{r=1}^{T} c_a c_b c_g c_s \mathrm{E}\left(w_{\kappa,t-l,r}w_{\kappa,t-s-g-l,h}\right) \\ = & O\left(\frac{1}{T^2}\right). \end{aligned}$$

by summability condition of c_a and calculation of expectation that

$$\begin{aligned} & \mathrm{E}\left(w_{\kappa,t-l,r}w_{\kappa,t-s-g-l,h}\right) \\ = & \left(\frac{1}{Th_0^d}\right)^2\int\frac{1}{f_X(x)f_X(z)}\left(\frac{x-y}{h_0}\right)^{\kappa}\left(\frac{z-w}{h_0}\right)^{\kappa}K_0\left(\frac{x-y}{h_0}\right)K_0\left(\frac{z-w}{h_0}\right) \\ & \times f_{X,|t-l-r|,|g+s|,|t-l-h|}(x,y,z,w)dxdydzdw \\ = & \left(\frac{1}{T}\right)^2\int\frac{1}{f_X(x)f_X(z)}u^{\kappa}v^{\kappa}K_0(u)K_0(v)f_{X,|t-l-r|,|g+s|,|t-l-h|}(x,x-uh_0,z,z-vh_0)dxdudzdv \\ = & O(T^{-2}). \end{aligned}$$

For the term with bias effects,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\widehat{B}_{t-l} \\
=\ & \frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left(e_1' M_T^{-1}(X_{t-l})B_n(X_{t-l})\right) \simeq \frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left(e_1'[Mf_X(X_{t-l})]^{-1}B_n(X_{t-l})\right) \\
\simeq\ & \frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left(\sum_{s\neq t-l}\sum_{\kappa}\omega^{0,\kappa}\frac{1}{Th_0^d}f_X(X_{t-l})^{-1}K_0\left(\frac{X_{t-l}-X_s}{h_0}\right)\left(\frac{X_{t-l}-X_s}{h_0}\right)^{q+\kappa-1}h^q m^{(q)}(X_{t-l})\right) \\
=\ & \sum_{\kappa}\omega^{0,\kappa}\frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left(\frac{1}{Th_0^d}\sum_{s\neq t-l} f_X(X_{t-l})^{-1}K_0\left(\frac{X_{t-l}-X_s}{h_0}\right)\left(\frac{X_{t-l}-X_s}{h_0}\right)^{q+\kappa-1}h^q m^{(q)}(X_{t-l})\right).
\end{aligned}$$

By verifications of moments, we show that $\frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\widehat{B}_{t-l} = O_p(h^q)$. In particular,

$$\begin{aligned}
&\mathrm{E}\left|\frac{1}{T}\sum_{t=\tau+1}^{T} u_{t-j}\left(\frac{1}{Th_0^d}\sum_{s\neq t-l} f_X(X_{t-l})^{-1}K_0\left(\frac{X_{t-l}-X_s}{h_0}\right)\left(\frac{X_{t-l}-X_s}{h_0}\right)^{q+\kappa-1}h^q m^{(q)}(X_{t-l})\right)\right|^2 \\
=\ & \frac{1}{T^2}\sum_{t=\tau+1}^{T}\sum_{p=\tau+1}^{T}\mathrm{E}u_{t-j}u_{p-j}\left[\frac{1}{Th_0^d}\right]^2 h_0^{2q}\sum_s\sum_r f_X(X_{t-l})^{-1}f_X(X_{p-l})^{-1} \\
& K_0\left(\frac{X_{t-l}-X_s}{h_0}\right)K_0\left(\frac{X_{p-l}-X_r}{h_0}\right)\left(\frac{X_{t-l}-X_s}{h_0}\right)^{q+\kappa-1}\left(\frac{X_{p-l}-X_r}{h_0}\right)^{q+\kappa-1}m^{(q)}(X_{t-l})m^{(q)}(X_{p-l}) \\
=\ & \frac{h_0^{2q}}{T^2}\sum_{t=\tau+1}^{T}\sum_{p=\tau+1}^{T}\gamma_u(|t-p|)\left(\frac{1}{Th_0^d}\right)^2\sum_s\sum_r \mathrm{E}f_X(X_{t-l})^{-1}f_X(X_{p-l})^{-1} \\
& K_0\left(\frac{X_{t-l}-X_s}{h_0}\right)K_0\left(\frac{X_{p-l}-X_r}{h_0}\right)\left(\frac{X_{t-l}-X_s}{h_0}\right)^{q+\kappa-1}\left(\frac{X_{p-l}-X_r}{h_0}\right)^{q+\kappa-1}m^{(q)}(X_{t-l})m^{(q)}(X_{p-l}) \\
=\ & O\left(\frac{h_0^{2q}}{T}\right), \qquad (28)
\end{aligned}$$

by the summability of $\sum_h \gamma_u(h)$ implied by the stationarity/mixing property of u_t, and calculation of expectations. The other terms follow by symmetric arguments. Therefore, by (26)-(28) we have

$$\left\|\widehat{G}_\tau - G_\tau\right\| = O_p(T^{-1/2}h_0^q + (\log T)T^{-1}h_0^{-d} + h_0^{2q})\tau. \qquad (29)$$

Similarly, we have

$$\|\widehat{g}_\tau - g_\tau\| = O_p(T^{-1/2}h_0^q + (\log T)T^{-1}h_0^{-d} + h_0^{2q})\tau. \qquad (30)$$

Notice that

$$\begin{aligned}
\widehat{A}_\tau - \overline{A}_\tau &= -G_\tau^{-1}[\widehat{G}_\tau - G_\tau]G_\tau^{-1}g_\tau + G_\tau^{-1}[\widehat{g}_\tau - g_\tau] \\
&\quad + \widehat{G}_\tau^{-1}[\widehat{G}_\tau - G_\tau]G_\tau^{-1}[\widehat{G}_\tau - G_\tau]G_\tau^{-1}g_\tau - \widehat{G}_\tau^{-1}[\widehat{G}_\tau - G_\tau]G_\tau^{-1}[\widehat{g}_\tau - g_\tau].
\end{aligned}$$

Furthermore, we can substitute Γ_τ^{-1} and γ_τ for G_τ^{-1} and g_τ. Using (29), (30), and (19) and (24), we obtain

$$\left\|\widehat{A}_\tau - \overline{A}_\tau + \Gamma_\tau^{-1}[\widehat{G}_\tau - G_\tau]\Gamma_\tau^{-1}\gamma_\tau - \Gamma_\tau^{-1}[\widehat{g}_\tau - g_\tau]\right\| = O_p(\Delta_n^2), \tag{31}$$

where $\Delta_n = \left((\log T)T^{-1}h_0^{-d} + h_0^{2q}\right)\tau$.

We can then write

$$\frac{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^\tau(\widehat{a}_j - \overline{a}_j)u_{t-j}}{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)} \tag{32}$$

$$\begin{aligned}
&\simeq \frac{1}{f_X(x)}\frac{1}{Th_1^d}\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^\tau(\widehat{a}_j - \overline{a}_j)u_{t-j} \\
&\simeq \frac{1}{f_X(x)}\frac{1}{Th_1^d}\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)U_{\tau t}'\left[\Gamma_\tau^{-1}[\widehat{G}_\tau - G_\tau]\Gamma_\tau^{-1}\gamma_\tau - \Gamma_\tau^{-1}[\widehat{g}_\tau - g_\tau]\right],
\end{aligned}$$

notice that

$$\begin{aligned}
&\left\|\frac{1}{Th_1^d}\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)U_{\tau t}'\left[\Gamma_\tau^{-1}[\widehat{G}_\tau - G_\tau]\Gamma_\tau^{-1}\gamma_\tau - \Gamma_\tau^{-1}[\widehat{g}_\tau - g_\tau]\right]\right\| \\
\leq\ &\frac{1}{T}\sum_{t=1}^T\left\|\frac{1}{h_1^d}K_1\left(\frac{x-X_t}{h_1}\right)\right\|\|U_{\tau t}'\|\left[\left\|\Gamma_\tau^{-1}\right\|\left\|\widehat{G}_\tau - G_\tau\right\|\left\|\Gamma_\tau^{-1}\gamma_\tau\right\| + \left\|\Gamma_\tau^{-1}\right\|\|\widehat{g}_\tau - g_\tau\|\right].
\end{aligned}$$

Thus (32) is of order $O_p\left((\log T)T^{-1}h_0^{-d} + h_0^{2q}\right)\tau^c$, where c is a constant. Under Assumption 6, (32) is $o_p(T^{-1/2}h_1^{-d/2})$, which finishes the proof for the second part. ■

Lemma A3. *Under Assumptions 1 to 9*

$$Q_{T3} = O_p(h_0^q) + o_p(T^{-1/2}h_1^{-d/2}).$$

Proof of Lemma A3. First note that $\widehat{m}(X_t) - m(X_t) = \widehat{V}_t + \widehat{B}_t$, where $\widehat{B}_t$ and $\widehat{V}_t$ are defined as (25). Denote

$$\widehat{f}_X^1(x) = \frac{1}{Th_1^d}\sum_{s=1}^T K_1\left(\frac{x-X_s}{h_1}\right).$$

We have

$$Q_{T3} = \frac{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^\tau a_j\left(\widehat{m}(X_{t-j}) - m(X_{t-j})\right)}{\sum_{t=1}^T K_1\left(\frac{x-X_t}{h_1}\right)}$$

$$
\begin{aligned}
&= \frac{T^{-1}h_1^{-d}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{V}_{t-j}}{T^{-1}h_1^{-d}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)} + \frac{T^{-1}h_1^{-d}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{B}_{t-j}}{T^{-1}h_1^{-d}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)} \\
&= \frac{1}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{V}_{t-j} + \frac{1}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{B}_{t-j} \\
&\quad + \frac{1}{Th_1^d}\frac{\widehat{f}_X^1(x)-f_X(x)}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{V}_{t-j} \\
&\quad + \frac{1}{Th_1^d}\frac{\widehat{f}_X^1(x)-f_X(x)}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{B}_{t-j}.
\end{aligned}
$$

We start with the first term, which can be written as

$$
\begin{aligned}
&\frac{1}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{V}_{t-j} \\
\simeq\ &\frac{1}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\left[e_1'\left[Mf_X(X_{t-j})\right]^{-1}U_n(X_{t-j})\right] \\
=\ &\sum_{\kappa}\omega^{0,\kappa}\frac{1}{T}\sum_{r=1}^{T}u_r\sum_{j=1}^{\tau}a_j w_{\kappa,T,j,r},
\end{aligned}
$$

where

$$
w_{\kappa,T,j,r} = \frac{1}{Th_1^d h_0^d}\sum_{t=1}^{T}\frac{1}{f_X(X_{t-j})f_X(x)}K_1\left(\frac{x-X_t}{h_1}\right)K_0\left(\frac{X_{t-j}-X_r}{h_0}\right)\left(\frac{X_{t-j}-X_r}{h_0}\right)^{\kappa}.
$$

We need to verify the boundedness of $E(|w_{\kappa,T,j,r}|)$:

$$
\begin{aligned}
E\left|w_{\kappa,T,j,r}\right| &= \frac{1}{Th_1^d h_0^d}\sum_{t=1}^{T}\mathrm{E}\frac{1}{f_X(X_{t-j})f_X(x)}K_1\left(\frac{x-X_t}{h_1}\right)K_0\left(\frac{X_{t-j}-X_r}{h_0}\right)\left(\frac{X_{t-j}-X_r}{h_0}\right)^{\kappa} \\
&= \frac{1}{T}\sum_{t=1}^{T}\int\left[\frac{1}{h_1^d h_0^d}\frac{f_{X,j,t-r}(y,z,w)}{f_X(x)f_X(z)}K_1\left(\frac{x-y}{h_1}\right)K_0\left(\frac{z-w}{h_0}\right)\left(\frac{z-w}{h_0}\right)^{\kappa}dydzdw\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\int\frac{f_{X,j,t-r}(x-uh_1,z,z-vh_0)}{f_X(x)f_X(z)}K_1(u)\,K_0(v)\,v^{\kappa}dudzdv.
\end{aligned}
$$

Again, under assumption 2 that the densities are bounded, $E\left|w_{\kappa,T,j,r}\right|$ is uniformly bounded over all j and r. Since w_{Tjs} only depends on $X_1,\ldots,X_T$, and u,X are mutually independent, we have

$$
\operatorname{var}\left[\frac{1}{T}\sum_{r=1}^{T}u_r\sum_{j=1}^{\tau}a_j w_{\kappa,T,j,r}\right] = \frac{1}{T^2}\sum_{r=1}^{T}\sum_{s=1}^{T}E(u_s u_r)\left[E\left(\sum_{j=1}^{\tau}a_j w_{\kappa,T,j,r}\right)\left(\sum_{i=1}^{\tau}a_i w_{\kappa,T,i,s}\right)\right]
$$

$$
\begin{aligned}
&= \frac{1}{T^2}\sum_{r=1}^{T}\sum_{s=1}^{T}\gamma_u(|s-r|)\sum_{j=1}^{\tau}\sum_{i=1}^{\tau}a_j a_i\left(Ew_{\kappa,T,j,r}w_{\kappa,T,i,s}\right)\\
&\leq \frac{1}{T}\left(\gamma_u(0)+2\sum_{j=1}^{\infty}\gamma_u(j)\right)\left(\sum_{j=1}^{\infty}|a_j|\right)^2\left(\sup_{j,r,T}E(|w_{\kappa,T,j,r}|)\right)^2\\
&= O(T^{-1})
\end{aligned}
$$

by the summability of the a_j and $\gamma_u(j)$, and the boundedness of $E(|w_{\kappa,T,j,r}|)$. Thus

$$
\frac{1}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{V}_{t-j}=O_p(T^{-1/2}).
$$

We now turn to the leading bias term, notice that

$$
\begin{aligned}
&\frac{1}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{B}_{t-j}\\
\simeq\ &\sum_{\kappa}\omega^{0,\kappa}\frac{h^q}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\\
&\sum_{j=1}^{\tau}a_j\frac{1}{Th_0^d}\sum_{s\neq t-j}K_0\left(\frac{X_{t-j}-X_s}{h_0}\right)\left(\frac{X_{t-j}-X_s}{h_0}\right)^{q+\kappa-1}\frac{m^{(q)}(X_{t-j})}{f_X(X_{t-j})}.
\end{aligned}
$$

Conditional on X_{t-j}, for each κ,

$$
\begin{aligned}
&\frac{1}{T}\sum_{s\neq t-j}\frac{1}{h_0^d}K_0\left(\frac{X_{t-j}-X_s}{h_0}\right)\left(\frac{X_{t-j}-X_s}{h_0}\right)^{q+\kappa-1}\frac{m^{(q)}(X_{t-j})}{f_X(X_{t-j})}\\
\simeq\ &\mathrm{E}_{t-j}\left[\frac{1}{h_0^d}K_0\left(\frac{X_{t-j}-X_s}{h_0}\right)\left(\frac{X_{t-j}-X_s}{h_0}\right)^{q+\kappa-1}\frac{m^{(q)}(X_{t-j})}{f_X(X_{t-j})}\right]\\
=\ &\frac{1}{h_0^d}\int K_0\left(\frac{X_{t-j}-X_s}{h_0}\right)\left(\frac{X_{t-j}-X_s}{h_0}\right)^{q+\kappa-1}\frac{m^{(q)}(X_{t-j})}{f_X(X_{t-j})}f_{X,t-j-s}(X_s|X_{t-j})dX_s\\
=\ &\int K_0(u)\,u^{q+\kappa-1}\frac{m^{(q)}(X_{t-j})}{f_X(X_{t-j})}f_{X,t-j-s}(X_{t-j}-uh_0|X_{t-j})du\\
\simeq\ &m^{(q)}(X_{t-j})\frac{f_{X,t-j-s}(X_{t-j})}{f_X(X_{t-j})}\int K_0(u)\,u^{q+\kappa-1}du.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\frac{1}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{B}_{t-j}\\
\simeq\ &h^q\sum_{\kappa}\omega^{0,\kappa}\mu_{q+\kappa-1}(K_0)\sum_{j=1}^{\tau}a_j\frac{1}{Th_1^d}\sum_{t=1}^{T}\frac{1}{f_X(x)}K_1\left(\frac{x-X_t}{h_1}\right)m^{(q)}(X_{t-j})\frac{f_{X,t-j-s}(X_{t-j})}{f_X(X_{t-j})}
\end{aligned}
$$

$$\begin{aligned}
&\frac{1}{Th_1^d}\sum_{t=1}^{T}\frac{1}{f_X(x)}K_1\left(\frac{x-X_t}{h_1}\right)m^{(q)}(X_{t-j})\frac{f_{X,t-j-s}(X_{t-j})}{f_X(X_{t-j})} \\
\simeq\ & \mathrm{E}\left[\frac{1}{h_1^d}\frac{1}{f_X(x)}K_1\left(\frac{x-X_t}{h_1}\right)m^{(q)}(X_{t-j})\frac{f_{X,t-j-s}(X_{t-j})}{f_X(X_{t-j})}\right].
\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{1}{Th_1^d}\frac{1}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{B}_{t-j} \\
\simeq\ & h_0^q\sum_{\kappa}\omega^{0,\kappa}\mu_{q+\kappa-1}(K_0)\sum_{j=1}^{\tau}a_j\frac{1}{Th_1^d}\sum_{t=1}^{T}\frac{1}{f_X(x)}K_1\left(\frac{x-X_t}{h_1}\right)m^{(q)}(X_{t-j})\frac{f_{X,t-j-s}(X_{t-j})}{f_X(X_{t-j})} \\
\simeq\ & h_0^q\sum_{\kappa}\omega^{0,\kappa}\mu_{q+\kappa-1}(K_0)\sum_{j=1}^{\tau}a_j\mathrm{E}\left[\frac{1}{h_1^d}\frac{1}{f_X(x)}K_1\left(\frac{x-X_t}{h_1}\right)m^{(q)}(X_{t-j})\frac{f_{X,t-j-s}(X_{t-j})}{f_X(X_{t-j})}\right] \\
=\ & O_p(h_0^q),
\end{aligned}$$

since

$$\sum_{j=1}^{\infty}|a_j|<\infty$$

and

$$\mathrm{E}\left[\frac{1}{h_1^d}\frac{1}{f_X(x)}K_1\left(\frac{x-X_t}{h_1}\right)m^{(q)}(X_{t-j})\frac{f_{X,t-j-s}(X_{t-j})}{f_X(X_{t-j})}\right]=O(1).$$

Finally we turn to the remainder terms

$$\begin{aligned}
&\frac{1}{Th_1^d}\frac{\widehat{f}_X^1(x)-f_X(x)}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{V}_{t-j},\ \text{and} \\
&\frac{1}{Th_1^d}\frac{\widehat{f}_X^1(x)-f_X(x)}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{B}_{t-j}.
\end{aligned}$$

Notice that

$$\sup_{x\in\mathcal{X}}\left|\widehat{f}_X^1(x)-f_X(x)\right|=O_p(h_1^q)+O_P(T^{-1/2}h_1^{-d/2}(\log T)^{1/2}), \tag{33}$$

$$\sup_{1\le t\le T}\left|\widehat{V}_t\right|=O_P(T^{-1/2}h_0^{-d/2}(\log T)^{1/2}), \tag{34}$$

and

$$\sup_{1\le t\le T}\left|\widehat{B}_t\right|=O_p(h_0^q). \tag{35}$$

Under Assumption 2, $f_X(x)$ is bounded away from zero, we have

$$\left|\frac{1}{Th_1^d}\frac{\widehat{f}_X^1(x)-f_X(x)}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{V}_{t-j}\right|$$

$$
\begin{aligned}
&\leq \left(\sum_{j=1}^{\infty}|a_j|\right)\sup_{x\in\mathcal{X}}\left|\widehat{f}_X^1(x)-f_X(x)\right|\sup_t\left|\widehat{V}_t\right|\left|\frac{1}{Th_1^d}\sum_{t=1}^{T}\frac{1}{f_X(x)}\left|K_1\left(\frac{x-X_t}{h_1}\right)\right|\right| \\
&= O_p(h_1^q+T^{-1/2}h_1^{-d/2}(\log T)^{1/2})O_P(T^{-1/2}h_0^{-d/2}(\log T)^{1/2})
\end{aligned}
$$

$$
\begin{aligned}
&\left|\frac{1}{Th_1^d}\frac{\widehat{f}_X^1(x)-f_X(x)}{f_X(x)}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}a_j\widehat{B}_{t-j}\right| \\
&\leq \left(\sum_{j=1}^{\infty}|a_j|\right)\sup_{x\in\mathcal{X}}\left|\widehat{f}_X^1(x)-f_X(x)\right|\sup_t\left|\widehat{B}_t\right|\left|\frac{1}{Th_1^d}\sum_{t=1}^{T}\frac{1}{f_X(x)}\left|K_1\left(\frac{x-X_t}{h_1}\right)\right|\right| \\
&= O_p(h_1^q+T^{-1/2}h_1^{-d/2}(\log T)^{1/2})O_P(h_0^q),
\end{aligned}
$$

noticing that

$$
\begin{aligned}
\frac{1}{Th_1^d}\sum_{t=1}^{T}\left|K_1\left(\frac{x-X_t}{h_1}\right)\right| &\to \mathrm{E}\left\{\frac{1}{h_1^d}\left|K_1\left(\frac{x-X_t}{h_1}\right)\right|\right\} \\
&= \int|K_1(u)|\,f_X(x-uh_1)du \simeq f_X(x)\int|K_1(u)|\,du.
\end{aligned}
$$

■

Lemma A4. *Under Assumptions 1 to 7*

$$
Q_{T4}=o_p(T^{-1/2}h_1^{-d/2}).
$$

Proof of Lemma A4. We have

$$
\begin{aligned}
&\left|\frac{1}{Th_1^d}\sum_{t=1}^{T}K_1\left(\frac{x-X_t}{h_1}\right)\sum_{j=1}^{\tau}(\widehat{a}_j-a_j)\left(\widehat{m}(X_{t-j})-m(X_{t-j})\right)\right| \\
&\leq \frac{1}{Th_1^d}\sum_{t=1}^{T}\left|K_1\left(\frac{x-X_t}{h_1}\right)\right|\left\|\widehat{A}_\tau-A_\tau\right\|\left[\sum_{j=1}^{\tau}\left(\widehat{m}(X_{t-j})-m(X_{t-j})\right)^2\right]^{1/2} \\
&\leq \frac{1}{Th_1^d}\sum_{t=1}^{T}\left|K_1\left(\frac{x-X_t}{h_1}\right)\right|\left\|\widehat{A}_\tau-A_\tau\right\|\cdot\tau\max_s\left|\widehat{m}(X_s)-m(X_s)\right|.
\end{aligned}
$$

Notice that

$$
\left\|\widehat{A}_\tau-A_\tau\right\|\leq\left\|\widehat{A}_\tau-\overline{A}_\tau\right\|+\left\|\overline{A}_\tau-A_\tau\right\|,
$$

and, from the proof of Lemma 2, we have

$$
\left\|\widehat{A}_\tau-\overline{A}_\tau\right\|=o_p((\log T)T^{-1/2}h_0^{-d/2}+h_0^q),
$$

and

$$\|\overline{A}_\tau - A_\tau\| = O_p(T^{-1/2}\tau^{3/2}).$$

In addition,

$$\max_{1\le s\le T} |\widehat{m}(X_s) - m(X_s)| = O_p(h_0^q + T^{-1/2}h_0^{-d/2}(\log T)^{1/2}),$$

thus

$$|Q_{T4}| = o_p(T^{-1/2}h_1^{-d/2}).$$

■

Proof of Theorem 3 The steps of proving Theorem 3 is similar to that for Theorem 2. ■

REFERENCES

AKAIKE, H. (1974). A new look at the statistical model identification. IEEE Transactions in Automatic Control. AC-19, 716-723.

ANDERSEN, T.W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley and Sons.

CHEN, X., AND O. LINTON (2001). An Alternative way of computing efficient semiparametric instrumental variables estimators. Web page:
`http://econ.lse.ac.uk/staff/olinton/research/cliv.pdf`

CONLEY, T.G., L.P. HANSEN, E.G.J. LUTTMER, AND J.A. SCHEINKMAN (1997). Short-Term Interest Rates as Subordinated Diffusions," The Review of Financial Studies 10, 525-577.

FAN, J. (1992). Design-Adaptive Nonparametric Regression. Journal of the American Statistical Association, 87 , 998-1004.

FAN, J. AND I. GIJBELS (1996). Local Polynomial Modelling and Its Applications, Champman & Hall.

FAN, J, E. MAMMEN, AND W. HÄRDLE (1998). Direct estimation of low dimensional components in additive models. *Ann. Statist.*, **26**, 943-971.

HÄRDLE, W. (1991). *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

HANNAN, E.J., AND M. DEISTLER (1988). *The Statistical Theory of Linear Systems.* John Wiley; New York.

HART, J.D. (1991). Kernel Regression Estimation with Time Series Errrors. *Journal of the Royal Statistical Society.* 53, 173-187.

Hastie, T. and R. Tibshirani (1991). *Generalized Additive Models*. Chapman and Hall, London.

Kristensen, D. and O. Linton (2001): An Alternative GLS-like Transformation in Regression Models with AR(1) errors. *Problem Corner, Econometric Theory*.

Lin, X. and Carroll, R. J. (2000), "Nonparametric Function Estimation for Clustered Data When the Predictor is Measured Without/With Error," Journal of the American Statistical Association, 95, 520-534.

Linton, O.B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika*, **84**, 469-474.

Linton, O.B. and E. Mammen. (2002). Estimating an ARCH(∞) model by kernel regression methods. Manuscript, LSE.

Linton, O.B. and J.P. Nielsen. (1995). Estimating structured nonparametric regression by the kernel method. *Biometrika* **82**, 93-101.

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

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

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

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59, 347-370.

Opsomer, J., Y. Wang, and Y. Yang (2001). Nonparametric Regression with Correlated Errors. Manuscript.

Phillips, P.C.B., and V. Solo (1992). Asymptotics for Linear Processes. *Annals of Statistics* 20, 971-1001.

Robinson, P.M. (1983). Nonparametric Estimators for Time Series. Journal of Time Series Analysis 4, 185-207.

Rosenblatt, M. (1956). A central limit theorem and strong mixing conditions, *Proc. Nat. Acad. Sci.* **4**, 43-47.

Ruckstuhl, A., Welsh, A. H. and Carroll, R. J. (2000), "Nonparametric Function Estimation of the Relationship Between Two Repeatedly Measured Variables," Statistica Sinica, 10, 51–71.

Ruppert, D., and M. Wand (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22**, 1346-1370.

Severini, T. A. and Staniswalis, J. G. (1994), "Quasilikelihood Estimation in Semiparametric Models," Journal of the American Statistical Association, 89, 501–511.

Vilar-Fernandez, J.M. and M. Francisco-Fernandez (2001). Local polynomial regression smoothers with AR-error structure. Forthcoming in TEST.

Wild, C. J. and Yee, T. W. (1996), "Additive Extensions to Generalized Estimating Equation Methods," Journal of the Royal Statistical Society, Series B, 58, 711-725.

Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998), "Asymptotic Confidence Regions for Kernel Smoothing of a Varying Coefficient Model with Longitudinal Data," Journal of the American Statistical Association, 93, 1388–1402.

Zeger, S. L. and Diggle, P. J. (1994), "Semiparametric Models for Longitudinal Data With Application to CD4 Cell Numbers in HIV Seroconverters," Biometrics, 50, 689-699.

B Tables and Figures

TABLE 1: Relative Efficiency (RE), $n = 100$

(AR(2) Prewhitening, Estimating $m(x)$ at $x = 0$)

ARMA Parameters				Relative Efficiency		
α_1	α_2	γ_1	γ_2	RE1	RE2	RE0
0	0	0	0	1.070	1.048	1.0000
		0.1		1.046	1.012	0.9901
		0.2		1.022	0.986	0.9615
		0.5		0.981	0.962	0.8000
		0.7		0.962	0.942	0.6711
		0.9		0.963	0.933	0.5524
0.1				1.042	1.039	0.9900
0.2				1.024	0.994	0.9600
0.5				0.980	0.971	0.7500
0.7				0.961	0.982	0.5100
0.9				0.980	0.990	0.1900
0.1		0.1		1.026	0.999	0.9612
0.1		0.2		1.008	0.986	0.9166
0.2		0.2		0.991	0.983	0.8571
0.5		0.2		0.959	0.972	0.6048
0.7		0.2		0.950	0.962	0.3864
0.9		0.2		0.970	0.986	0.1357
0.1		0.5		0.971	0.954	0.7333
0.2		0.5		0.960	0.950	0.6621
0.5		0.5		0.942	0.941	0.4286
0.7		0.5		0.940	0.943	0.2615
0.9		0.5		0.976	0.980	0.0884
		0.5	0.2	0.973	0.976	0.7752
		0.2	0.2	1.016	0.997	0.9259
		0.2	0.7	0.986	0.987	0.6536
0.2	0.2			0.990	0.988	0.9000
0.5	0.2			0.964	0.979	0.5850
0.7	0.2			0.980	0.990	0.2250

TABLE 2: Relative Efficiency (RE), $n = 200$

(AR(2) Prewhitening)

ARMA Parameters				Relative Efficiency	
α_1	α_2	γ_1	γ_2	RE1	RE2
0	0	0	0	1.058	1.046
		0.1		1.024	1.021
		0.2		1.007	0.996
		0.5		0.975	0.960
		0.7		0.966	0.938
		0.9		0.964	0.929
0.1				1.012	1.026
0.2				1.001	0.996
0.5				0.938	0.953
0.7				0.908	0.919
0.9				0.933	0.940
0.1		0.1		1.004	1.001
0.1		0.2		0.988	0.986
0.2		0.2		0.970	0.972
0.5		0.2		0.919	0.919
0.7		0.2		0.896	0.894
0.9		0.2		0.929	0.929
0.1		0.5		0.961	0.945
0.2		0.5		0.947	0.929
0.5		0.5		0.907	0.887
0.7		0.5		0.889	0.874
0.9		0.5		0.926	0.922
		0.5	0.2	0.946	0.952
		0.2	0.2	0.970	0.990
		0.2	0.7	0.956	0.962
0.2	0.1			0.979	0.987
0.2	0.2			0.958	0.971
0.5	0.2			0.933	0.948
0.7	0.2			0.936	0.951

TABLE 3: Relative Efficiency (RE), $n = 500$

(AR(2) Prewhitening)

α_1	α_2	γ_1	γ_2	RE1	RE2
0	0	0	0	1.032	1.026
		0.1		1.003	1.001
		0.2		0.982	0.992
		0.5		0.951	0.950
		0.7		0.940	0.926
		0.9		0.938	0.915
0.1				0.999	0.997
0.2				0.970	0.990
0.5				0.905	0.930
0.7				0.867	0.881
0.9				0.861	0.866
0.1		0.1		0.978	0.992
0.1		0.2		0.961	0.979
0.2		0.2		0.942	0.961
0.5		0.2		0.887	0.896
0.7		0.2		0.857	0.858
0.9		0.2		0.857	0.855
0.1		0.5		0.935	0.932
0.2		0.5		0.920	0.914
0.5		0.5		0.876	0.865
0.7		0.5		0.850	0.841
0.9		0.5		0.854	0.848
		0.5	0.2	0.918	0.937
		0.2	0.2	0.941	0.978
		0.2	0.7	0.897	0.921
0.2	0.1			0.949	0.976
0.2	0.2			0.926	0.955
0.5	0.2			0.867	0.881
0.7	0.2			0.861	0.866

TABLE 4

Relative Efficiency (RE), $n = 1000$

α_1	α_2	γ_1	γ_2	RE1	RE2
0	0	0	0	1.016	1.012
		0.1		0.994	0.996
		0.2		0.971	0.988
		0.5		0.949	0.948
		0.7		0.943	0.928
		0.9		0.940	0.921
0.1				0.988	0.997
0.2				0.967	0.987
0.5				0.912	0.931
0.7				0.880	0.890
0.9				0.871	0.886
0.1		0.1		0.971	0.989
0.1		0.2		0.958	0.976
0.2		0.2		0.942	0.959
0.5		0.2		0.899	0.903
0.7		0.2		0.875	0.874
0.9		0.2		0.876	0.875
0.1		0.5		0.937	0.933
0.2		0.5		0.925	0.918
0.5		0.5		0.890	0.881
0.7		0.5		0.871	0.862
0.9		0.5		0.875	0.871
		0.5	0.2	0.924	0.937
		0.2	0.2	0.941	0.975
		0.2	0.7	0.900	0.927
0.2	0.2			0.929	0.953
0.5	0.2			0.883	0.891
0.7	0.2			0.878	0.880

TABLE 5: Relative Efficiency (RE), $n = 100$

(AR(1) Prewhitening)

α_1	α_2	γ_1	γ_2	RE1	RE2
0	0	0	0	1.043	1.029
		0.1		1.024	1.016
		0.2		1.006	1.005
		0.5		0.966	0.991
		0.7		0.953	0.979
		0.9		0.948	0.974
0.1				1.024	1.138
0.2				1.007	1.009
0.5				0.966	0.989
0.7				0.952	0.976
0.9				0.971	0.990
0.2		0.2		0.976	0.996

TABLE 6: Relative Efficiency (RE), $n = 200$

(AR(1) Prewhitening)

α_1	α_2	γ_1	γ_2	RE1	RE2
		0.1		1.000	1.001
		0.2		0.980	0.993
		0.5		0.935	0.960
		0.7		0.920	0.944
		0.9		0.914	0.937
0.1				0.999	0.999
0.2				0.979	0.993
0.5				0.923	0.948
0.7				0.897	0.916
0.9				0.928	0.939
0.2		0.2		0.943	0.968

TABLE 7: Relative Efficiency (RE), $n = 500$

(AR(1) Prewhitening)

α_1	α_2	γ_1	γ_2	RE1	RE2
		0.1		0.995	1.000
		0.2		0.967	0.992
		0.5		0.917	0.951
		0.7		0.901	0.931
		0.9		0.894	0.922
0.1				0.990	0.999
0.2				0.966	0.992
0.5				0.900	0.930
0.7				0.864	0.881
0.9				0.861	0.866
0.2		0.2		0.925	0.959

TABLE 8: Relative Efficiency (RE), $n = 1000$

(AR(1) Prewhitening)

α_1	α_2	γ_1	γ_2	RE1	RE2
		0.1		0.983	0.996
		0.2		0.964	0.988
		0.5		0.925	0.950
		0.7		0.912	0.934
		0.9		0.908	0.927
0.1				0.983	0.996
0.2				0.960	0.987
0.5				0.909	0.930
0.7				0.880	0.890
0.9				0.877	0.881
0.2		0.2		0.930	0.950

TABLE 9: Relative Efficiency (RE), $n = 100$

(AR(2) Prewhitening, $m(x) = x$, estimate all points)

ARMA Parameters		Integrated Squared Errors			Relative Efficiency		
α_1	γ_1	ISE0	ISE1	ISE2	RE1	RE2	RE0
0	0	0.0641	0.0658	0.0661	1.027	1.031	1.0000
	0.1	0.0770	0.0776	0.0778	1.007	1.010	0.9901
	0.2	0.0912	0.0905	0.0901	0.992	0.988	0.9615
	0.5	0.1413	0.1363	0.1331	0.964	0.942	0.8000
	0.7	0.1809	0.1733	0.1660	0.957	0.916	0.6711
	0.9	0.2255	0.2157	0.2044	0.956	0.906	0.5524
0.1		0.0784	0.0789	0.0791	1.006	1.009	0.9900
0.2		0.0983	0.0968	0.0970	0.984	0.986	0.9600
0.5		0.2393	0.2216	0.2227	0.926	0.931	0.7500
0.7		0.6073	0.5409	0.5375	0.891	0.885	0.5100
0.9		3.6257	3.1559	3.1297	0.870	0.863	0.1900
0.2	0.2	0.1402	0.1343	0.1341	0.957	0.956	0.8571

TABLE 10: Relative Efficiency (RE), $n = 200$

(AR(2) Prewhitening, $m(x) = x$, at all points)

ARMA Parameters		Integrated Squared Errors			Relative Efficiency		
α_1	γ_1	ISE0	ISE1	ISE2	RE1	RE2	RE0
0	0	0.03488	0.03525	0.03531	1.0110	1.0120	1.0000
	0.1	0.04205	0.04168	0.04171	0.9910	0.9918	0.9901
	0.2	0.04991	0.04871	0.04907	0.9761	0.9834	0.9615
	0.5	0.07756	0.07379	0.07287	0.9514	0.9396	0.8000
	0.7	0.09941	0.09397	0.09118	0.9453	0.9173	0.6711
	0.9	0.12400	0.11690	0.11260	0.9433	0.9076	0.5524
0.1		0.04287	0.04241	0.04252	0.9890	0.9918	0.9900
0.2		0.05398	0.05228	0.05302	0.9684	0.9822	0.9600
0.5		0.13480	0.12280	0.12410	0.9107	0.9197	0.7500
0.7		0.35850	0.31430	0.31330	0.8768	0.8736	0.5100
0.9		2.52290	2.14610	2.12810	0.8506	0.8435	0.1900
0.2	0.2	0.07732	0.07294	0.07349	0.9434	0.9505	0.8571

TABLE 11: Relative Efficiency (RE), $n = 500$

(AR(2) Prewhitening, $m(x) = x$, at all points)

ARMA Parameters		Integrated Squared Errors			Relative Efficiency		
α_1	γ_1	ISE0	ISE1	ISE2	RE1	RE2	RE0
0	0	0.01463	0.01498	0.01487	1.024	1.016	1.0000
	0.1	0.01802	0.01775	0.01792	0.985	0.994	0.9901
	0.2	0.02143	0.02076	0.02110	0.968	0.984	0.9615
	0.5	0.03345	0.03150	0.03140	0.942	0.938	0.8000
	0.7	0.04295	0.04018	0.03940	0.935	0.917	0.6711
	0.9	0.05360	0.05006	0.04870	0.933	0.908	0.5524
0.1		0.01838	0.01806	0.01827	0.982	0.994	0.9900
0.2		0.02323	0.02231	0.02281	0.960	0.982	0.9600
0.5		0.05895	0.05318	0.05399	0.902	0.916	0.7500
0.7		0.16130	0.14000	0.14021	0.868	0.869	0.5100
0.9		1.32120	1.10900	1.10240	0.839	0.834	0.1900
0.2	0.2	0.03340	0.03122	0.03168	0.934	0.948	0.8571

TABLE 12: Relative Efficiency (RE), $n = 1000$

(AR(2) Prewhitening, $m(x) = x$, at all points)

ARMA Parameters		Integrated Squared Errors			Relative Efficiency		
α_1	γ_1	ISE0	ISE1	ISE2	RE1	RE2	RE0
0	0	0.00794	0.00797	0.00801	1.0032	1.0084	1.0000
	0.1	0.00961	0.00943	0.00956	0.9808	0.9954	0.9901
	0.2	0.01142	0.01101	0.01125	0.9636	0.9844	0.9615
	0.5	0.01783	0.01666	0.01667	0.9347	0.9348	0.8000
	0.7	0.02289	0.02122	0.02086	0.9269	0.9110	0.6711
	0.9	0.02857	0.02641	0.02575	0.9241	0.9009	0.5524
0.1		0.00980	0.00959	0.00975	0.9784	0.9951	0.9900
0.2		0.01239	0.01183	0.01216	0.9547	0.9818	0.9600
0.5		0.03154	0.02808	0.02868	0.8929	0.8580	0.7500
0.7		0.08695	0.07416	0.07460	0.8529	0.8580	0.5100
0.9		0.75080	0.61500	0.61340	0.8192	0.8170	0.1900
0.2	0.2	0.01781	0.01650	0.01684	0.9264	0.9452	0.8571

TABLE 13: AR Coefficients in the Residuals

α_1	$\widehat{a}_1$	$\widehat{a}_2$	$\widehat{a}_3$	$\widehat{a}_4$	$\widehat{a}_5$	$\widehat{a}_6$	$\widehat{a}_7$	$\widehat{a}_8$	τ (chosen by BIC)
Weekly Data	0.136	0.169	-0.022	-0.005	0.011	0.017	0.062	0.029	$\tau = 2$
Daily Data	0.012	0.085	0.132	0.042	0.078	0.162	0.021	0.009	$\tau = 6$

The residuals are estimated from the conventional procedure with $h = 1.06 s_X T^{-1/5}$.

Figure 2a: Weekly data, h1

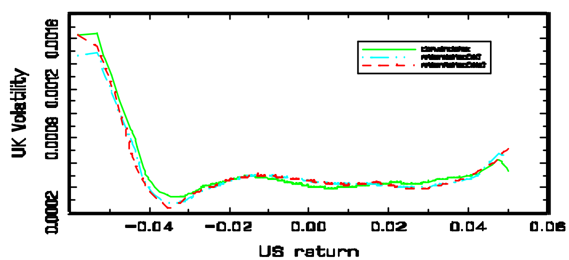

Figure 2b: Weekly data, h2

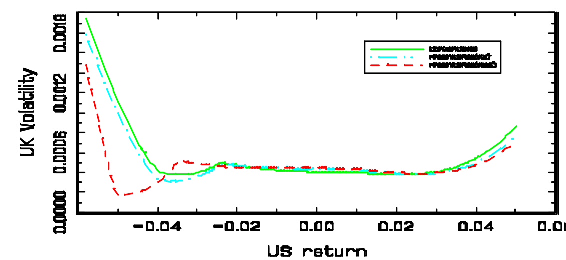

Figure 2c: Weekly data, h3

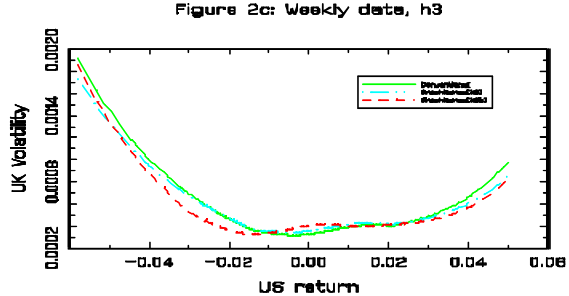

Figure 2d: Weekly data, h4

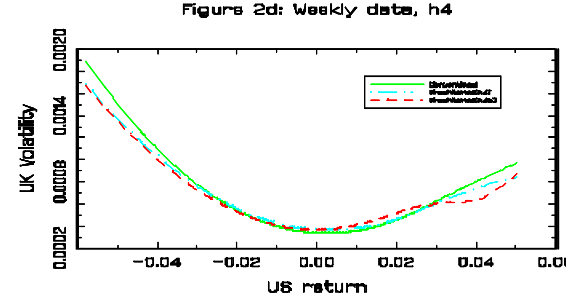

Figure 3a: Weekly data, The Prewhitened Estimator with Confidence Band, h1

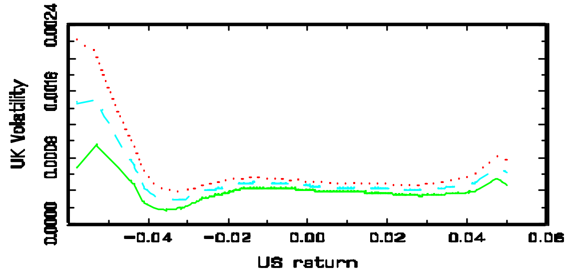

Figure 3b: Weekly data, The Prewhitened Estimator with Confidence Band, h2

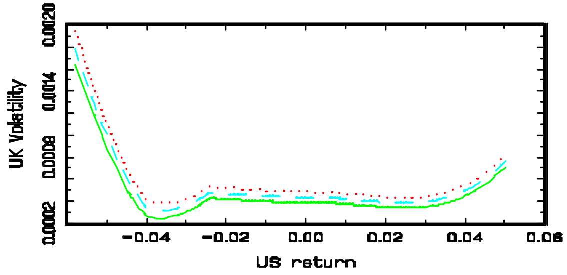

Figure 3c: Weekly data, The Prewhitened Estimator with Confidence Band, h3

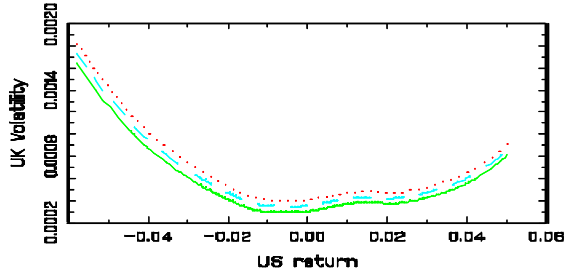

Figure 3d: Weekly data, The Prewhitened Estimator with Confidence Band, h4

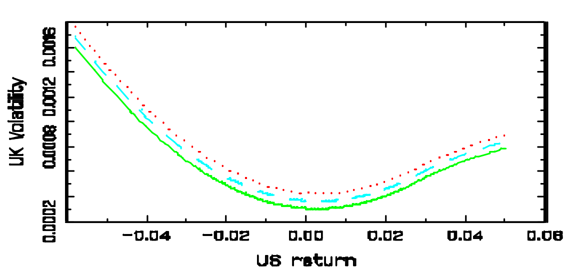

Figure 4a: Daily data, h1

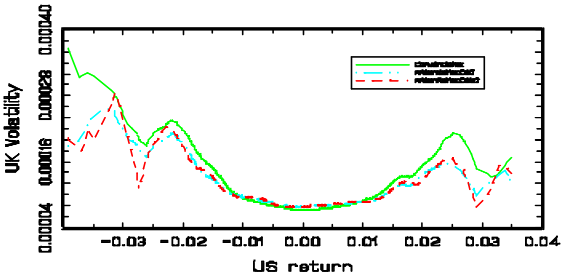

Figure 4b: Daily data, h2

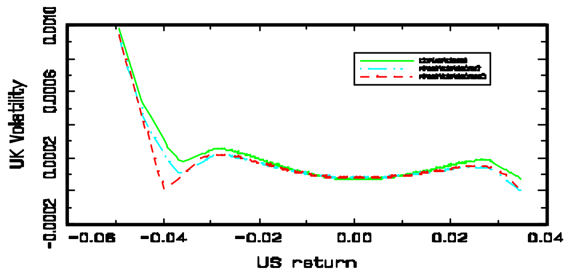

Figure 4c: Daily data, h3

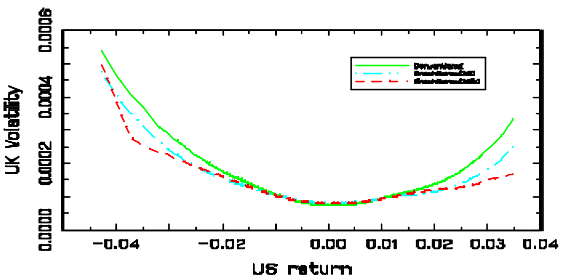

Figure 4d: Daily data, h4

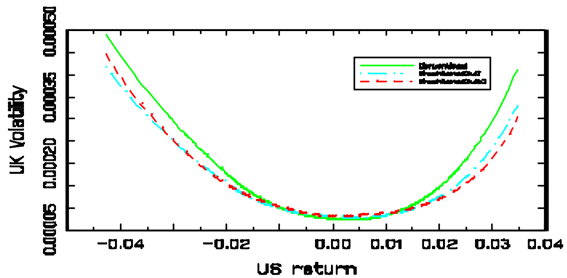

Figure 5a: Daily data, The Prewhitened Estimator with Confidence Band, h1

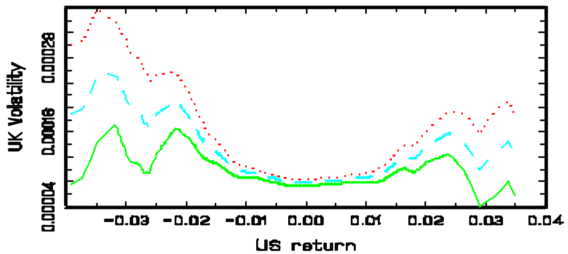

Figure 5b: Daily data, The Prewhitened Estimator with Confidence Band, h2

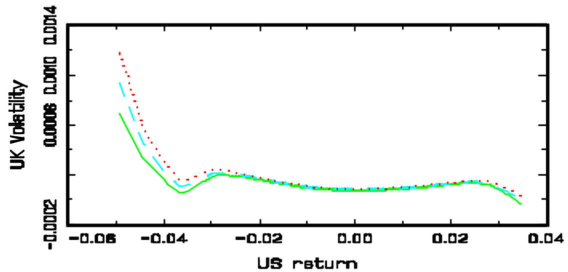

Figure 5c: Daily data, The Prewhitened Estimator with Confidence Band, h3

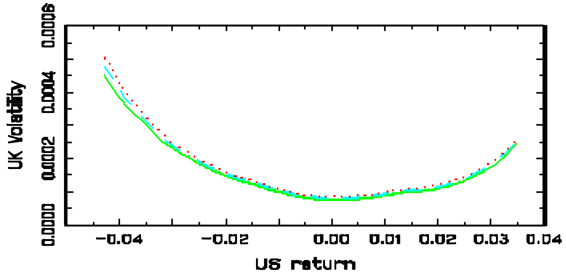

Figure 5d: Daily data, The Prewhitened Estimator with Confidence Band, h4