Economics Bulletin

Volume 32, Issue 4

# Multivariate Local Polynomial Regression With Autocorrelated Errors

Ke Yang
*University of Hartford*

## Abstract

We propose a three-step local polynomial procedure for a multivariate nonparametric regression in which the errors are autocorrelated. The proposed estimator uses all sample points to estimate m(x), the regression function evaluated at point x, but the contributions from all non-local points are used only through their residuals. Our proposed estimator exhibits good finite sample performance in a Monte Carlo simulation study.

**Citation:** Ke Yang, (2012) "Multivariate Local Polynomial Regression With Autocorrelated Errors", *Economics Bulletin*, Vol. 32 No. 4 pp. 3298-3305.
**Contact:** Ke Yang - kyang@hartford.edu.
**Submitted:** October 02, 2012. **Published:** December 03, 2012.

# 1 Introduction

Many recent research have studied the nonparametric regression models with autocorrelated errors, which is not uncommon in times series economic data. Consider the regression model

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

where the stationary residual process $u_t$ is autocorrelated but satisfies $E(u_t|X_1, \cdots, X_T) = 0$ almost surely. The function $m(\cdot)$ is assumed to be smooth but unknown otherwise and is the object of central interest. Many studies, for example, Altman (1990), Marsry (1996) and Martins-Filho and Yao (2009), showed that the standard kernel estimator on model (1), which ignores the error structure, is consistent but not efficient. Recent research have focused on improving the estimation efficiency of the kernel smoothing by correcting standard errors for serial correlation. One promising procedure, based on prewhitening the errors to make them look like white noise, has been analyzed by, among others, Xiao et al. (2003), Su and Ullah (2006) and Martins-Filho and Yao (2009). They show that the prewhittening estimator can improve the first order properties of kernel method. In this paper, we propose a new local polynomial procedure for $m(x)$ in the time series regression model (1). With this proposed estimator, we first estimate the error autocorrelations through residuals from a preliminary regression. Then in the final local polynomial type regression, all sample points are used to estimate $m(x)$ but contributions from all points but those within local neighborhood of $x$ are used only through their residuals. We show in a Monte Carlo study that the proposed estimator not only improves estimation accuracy over standard local polynomial estimator but also outperforms the prewhitening procedure in all DGP designs under study.

# 2 Estimation

Let $\{Y_t, X_t, u_t\}_{t=1}^T$, where $X_t \in \mathbb{R}^d$ and $Y_t \in \mathbb{R}$, be a sample from regression model (1). We assume the residual process $u_t$ has a invertible linear process representation

$$u_t = \gamma(L)\epsilon_t = \sum_{j=0}^{\infty} \gamma_j L^j \epsilon_t = \sum_{j=0}^{\infty} \gamma_j \epsilon_{t-j}, \tag{2}$$

where $\epsilon_t$ is independent identically distributed with mean 0 and variance $\sigma_\epsilon^2$, $L$ is the usual lag operator. The error process $\{u_t\}_{t=1}^T$ is assumed to be independent of the process $\{X_t\}_{t=1}^T$. The coefficients $\{\gamma_j\}_{j=0}^\infty$ sequence is unknown except it is absolutely summable, i.e. $\sum_{j=1}^\infty |\gamma_j| < \infty$. The autocorrelation matrix of $\{u_t\}_{t=1}^T$ will be denoted by $R$. The process allows $u_t$ to be any finite-order ARMA$(p, q)$ process. We invert $\gamma(L)$ to obtain an autoregressive representation of $u_t$ of potentially infinite order,

$$\gamma(L)^{-1} = \alpha(L) = \alpha_0 - \alpha_1 L - \cdots - \alpha_j L^j - \cdots = \alpha_0 - \sum_{j=1}^{\infty} \alpha_j L^j \tag{3}$$

where we define $\alpha_0 = 1$ without loss of generality. So we have $\alpha(L)u_t = \epsilon_t$.

Let $K$ denote a symmetric kernel function on $\mathbb{R}^d$ and $H = diag(h_1, \cdots, h_d)$ a matrix of bandwidth sequences that degenerate with sample size $T$. Define $K_H(z) = |H|^{-1}K(H^{-1}z)$. Following the notations in Marsry (1996), in the standard multivariate local polynomial regression of order $q$ one solves

$$\min_{b_l(x)} \quad \sum_{t=1}^{T}[Y_t - \sum_{0\leq|l|\leq p} b_l(x_0)(X_t - x_0)^l]^2 K_H(X_t - x_0) \tag{4}$$

where $l = (l_1, \cdots, l_d)'$, $l! = l_1! \cdot l_2! \cdots l_d!)'$, $|l| = \sum_{i=1}^{l} l_i$, $x^l = x_1^{l_1} \times x_2^{l_2} \times x_d^{l_d}$, $\sum_{0\leq|l|\leq q} = \sum_{j=0}\sum_{\substack{l_1=0 \\ l_1+\cdots+l_d=j}}^{j}\cdots\sum_{l_d=0}^{j}$, and $b_l(x)$ is the estimate for polynomial parameter $\frac{1}{l!}(D^l m)(x)$ with $(D^l m)(x) = \frac{\partial^l m(x)}{\partial x_1^{l_1}\cdots\partial x_d^{l_d}}$. Let $N_v = \binom{v+d-1}{d-1}$ be the number of distinct $d$-tuples $l$ with $|l| = v$. Arrange these $N_v$ $d$-tuples as a sequence in a lexicographical order with highest priority to last position so that $(0, \cdots, 0, v)$ is the first element and let $g_{|l|}^{-1}$ denote this one-to-one map from the sequence to its index. Arrange the $N_{|l|}$ values of $(H^{-1}(X_t - x))^l$ in a column vector $G_{t,|l|}(x)$ according to the above order. Then $[G_{t,|l|}(x)]_k = (H^{-1}(X_t - x))^{g_{|l|}(k)}$ for $k = 0, \cdots, |l|$. Define $G_{t,\leq q}(x_0) = \{G_{t,0}(x)', G_{t,1}(x)', \cdots, G_{t,q}(x)'\}'$, a column vector of dimension $N = \sum_{v=0}^{q} N_v \times 1$. Then define $G_{\leq q}(x) = \{G_{1,\leq q}(x), G_{2,\leq q}(x), \cdots, G_{T,\leq q}(x)\}'$, a matrix of dimension $T \times N$. Similarly for $0 \leq |l| \leq q$ arrange the distinct values of $h^l b_l(x)$ as a column vector $b_{|l|}(x)$ of dimension $N_{|l|} \times 1$, in the same lexicographical order. Then define $b(x) = \{b_0(x)', b_1(x)', \cdots, b_q(x)'\}'$. With these notations the proposed procedure is as follows.

1. Calculate a preliminary consistent estimate of $m$ by local polynomial (of order $q_1$) smoothing $Y_t$ on $X_t$ with corresponding kernel $K_{H_0}(\cdot)$ and bandwidth $H_0$. Denote the preliminary estimate as $\hat{m}(X_t)$ and calculate the estimated residuals

$$\hat{u}_t = Y_t - \hat{m}(X_t)$$

2. Let $\tau = \tau(T)$ be some truncation parameter suitably small relative to the sample size $T$ but large enough to avoid serious bias. Conduct a $\tau^{th}$ order autoregression of $\hat{u}_t$,

$$\hat{u}_t = \hat{\alpha}_1\hat{u}_{t-1} + \cdots + \hat{\alpha}_\tau\hat{u}_{t-\tau} + residual.$$

Define the estimate $\hat{A}_\tau = (\hat{\alpha}_1, \cdots, \hat{\alpha}_\tau)'$ of $A_\tau = (\alpha_1, \cdots, \alpha_\tau)'$, as

$$\hat{A}_\tau = (\hat{U}_\tau'\hat{U}_\tau)^{-1}\hat{U}_\tau'\hat{u}_t,$$

where $\hat{u} = (\hat{u}_\tau, \cdots, \hat{u}_T)'$ and $\hat{U}_\tau$ is the $(T-\tau) \times \tau$ matrix of regressors with typical element $\hat{u}_{t-\tau}$. Then we construct the estimated autocorrelation matrix $\hat{R}$ using the Yule-Walker equations from AR($\tau$) process:

$$\rho_j = \hat{\alpha}_1\rho_{j-1} + \hat{\alpha}_2\rho_{j-2} + \cdots + \hat{\alpha}_p\rho_{j-p} \qquad \text{for} \quad j = 1, 2, \cdots$$

3. Let $G_{.t}(x) = E_{.t}G_{\leq q}(x)$ be the $T \times N$ matrix where $E_t$ denotes the $T \times T$ indicator matrix with the $[t,t]^{th}$ entry equal 1, and 0 elsewhere. The $m(x)$ is estimated by $\tilde{m}(x) = b_0(x)$, where $b(x)$ solves the following estimating equation of order $q_2$

$$0 = \frac{1}{T}\sum_{t=1}^{T} K_H(x_t - x)(G_{.t}(x))'(\hat{R})^{-1}[\vec{Y} - m_{.t}\{\vec{X}, b(x), \hat{m}(\vec{X})\}], \tag{5}$$

where $\vec{Y} = \{Y_1, \cdots, Y_T\}'$, $\vec{X} = \{X_1, \cdots, X_T\}'$, $\hat{m}(\vec{X}) = \{\hat{m}(X_1), \cdots, \hat{m}(X_T)\}'$ and the $l^{th}$ element of $m_{.t}\{\vec{X}, \hat{b}, \hat{m}(\vec{X})\}$ is $G_{t,\leq q}(x)'b(x)$, when $l = t$; and is $\hat{m}(X_l)$, when $l \neq t$.

The truncation parameter $\tau$ needs to satisfy $\tau(T) = \kappa \log T$ for some $\kappa > 0$ in order to avoid serious bias is estimating the autocorrelations (Xiao et al. 2003, pg. 983, assumption 6). In practice, an optimal $\tau$ can be selected based on selection criteria as AIC and BIC (see Xiao et al. 2003 for details). The last step of the procedure is an extension of the marginal kernel method in Wang (2003) to the time series model. The idea is to use all sample points in estimating $m(x_0)$ in order to reduce the variance, but points lie outside of the local neighborhood contribute to the estimation only through their residuals. For an optimal bandwidth $H*$, one can use cross-validation to select a global bandwidth.

# 3 A Simulation Study

In this simulation We investigate the numerical performance of the proposed estimator (denoted by PE), $\tilde{m}(\cdot)$, and compare it to the standard local polynomial estimator (denoted by LL) and the prewhitening estimator (denoted by PW) as in Xiao et al. (2003). The model we consider is as follows. We take $m(X_t) = sin(2X_t)$, where $X_t$ is generated from a uniform distribution on $[-2, 2]$. The error process $u_t$ is various special cases of the ARMA(1,1) and AR(2), AR(4) processes

$$u_t = \alpha_1 u_{t-1} + \epsilon_t + \gamma_1 \epsilon_{t-1}, \tag{6}$$

$$u_t = \alpha_1 u_{t-1} + \cdots + \alpha_p u_{t-p} + \epsilon_t, \quad \text{for} \quad p = 2, 4 \tag{7}$$

where $\epsilon_t$ is iid $N(0, \sigma_\epsilon^2)$. Various values for $\alpha_p$ $(p = 1, 2, 4)$, $\gamma_1$ and $\sigma_\epsilon^2$ are considered to reflect various level of autocorrelations and overall scale of noise. In (7), instead of using different values for $\alpha_p$ to generate error process, we use different values for roots of the characteristic function $\alpha(L) = 0$. [1] We use two sample sizes $T = 100$ and $T = 500$ with 500 replications being generated for each model specification. In each replication, the target function $m(\cdot)$ is estimated at 40 fixed equally-spaced grid points within the range of $X$.

In implementation of the estimators, we chose the same kernel function and bandwidth in all three estimators, i.e. $H_o = H_1 = H$. Specifically, Gaussian density function, and a

[1] For the convenience of experiment design, we choose to use equal roots, i.e. $G_1 = \cdots = G_p = G_0$ to generate the autoregressive processes, where $G_0^{-1}$ is the single root of the characteristic function $\alpha(L) = 0$.

"rule-of-thumb" bandwidth $h = 1.06\sigma_X T^{-1/5}$ are used throughout the simulation. In both the PE and the PW estimators, we use local cubic regression in the first step, use local linear regression in the third step, use the truncation parameter $\tau = 2$.

In this simulation study, the average squared errors, denoted by ASE [2] for each estimator are obtained. A relative efficiency based on the ratio of average squared errors are calculated, with numerators obtained from either the standard local linear estimator (denoted by RE1) or prewhitenning estimator (denoted by RE2) and denominators obtained from the proposed estimator. An efficiency value above 1 indicates that our proposed estimator outperforms its competitor and the higher the efficiency value the better is the proposed estimator. Tables 1, 2 and 3 correspond to three error structures, i.e. ARMA(1,1), AR(2) and AR(4).

Here we summarize some general findings from the simulation studies:

1. The results show that increasing sample sizes reduces ASE for all three estimators under evaluation, which is in comply with the asymptotic results in previous research. This also suggests that our proposed estimator is likely to be asymptotically consistent.

2. When the underlying process has a nontrivial serial correlation, our proposed estimator achieves efficiency gains over both the standard local linear estimator and the prewhitening estimator. The more serial correlation, the larger efficiency gains achieved by our procedure over the local linear estimator. As sample size increases from 100 to 500, this efficiency gains over local linear estimator becomes more significant. Furthermore, in both AR(2) and AR(4) models, the efficiency gains over the local linear estimator increases with overall scale of noise in the error process ($\sigma_\epsilon^2$).

# References

Altman, N.S. (1990), "Kernel smoothing of data with correlated errors." *Journal of the Acoustical Society of America*, 85, 749–759.

Marsry, W. (1996), "Multivariate regression estimation: local polynomial fitting for time series." *Stochastic Processes and Their Applications*, 65, 81–101.

Martins-Filho, C. and F. Yao (2009), "Nonparametric regression estimation with general parametric error covariance." *Journal of Multivriate Analysis*, 100, 309–333.

Su, L. and A. Ullah (2006), "More efficient estimation in nonparametric regression with nonparametric autocorrelated errors." *Econometric Theory*, 22, 98–126.

Wang, N. (2003), "Marginal nonparametric kernel regression accounting for within subject correlation." *Biometrika*, 90, 43–52.

[2] The ASE values in the tables are the original values multiplied by 10.

Xiao, Z., O.B. Linton, R.J. Carroll, and E. Mammen (2003), "More efficient local polynomial estimation in nonparametric regression with autocorrelated errors." *Journal of the American Statistical Association*, 98, 980–992.

Table 1: ASE Comparison on model with ARMA(1,1) errors

| | | n=100 | | | | | n=500 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | | ASE | | | Efficiency | | ASE | | | Efficiency | |
| $\alpha$ | $\gamma$ | LL | PE | PW | RE1 | RE2 | LL | PE | PW | RE1 | RE2 |
| 0 | 0 | 0.428 | 0.434 | 0.428 | 0.987 | 0.986 | 0.030 | 0.031 | 0.030 | 0.995 | 0.997 |
| 0 | 0.25 | 0.445 | 0.445 | 0.444 | 1.000 | 0.998 | 0.029 | 0.029 | 0.029 | 1.009 | 1.003 |
| 0 | 0.75 | 0.488 | 0.443 | 0.472 | 1.100 | 1.065 | 0.033 | 0.029 | 0.031 | 1.127 | 1.062 |
| 0.25 | 0 | 0.440 | 0.438 | 0.441 | 1.004 | 1.007 | 0.031 | 0.030 | 0.031 | 1.017 | 1.015 |
| 0.75 | 0 | 0.655 | 0.591 | 0.614 | 1.108 | 1.040 | 0.047 | 0.040 | 0.042 | 1.166 | 1.043 |
| 0.25 | 0.25 | 0.484 | 0.459 | 0.470 | 1.053 | 1.022 | 0.031 | 0.029 | 0.030 | 1.046 | 1.037 |
| 0.25 | 0.75 | 0.573 | 0.485 | 0.526 | 1.183 | 1.085 | 0.036 | 0.030 | 0.032 | 1.207 | 1.085 |
| 0.75 | 0.25 | 1.024 | 0.820 | 0.887 | 1.249 | 1.082 | 0.052 | 0.041 | 0.043 | 1.263 | 1.058 |
| 0.75 | 0.75 | 1.689 | 1.263 | 1.337 | 1.337 | 1.058 | 0.079 | 0.058 | 0.061 | 1.374 | 1.054 |

Table 2: ASE Comparison on model with AR(2) errors

| | | n=100 | | | | | n=500 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | | ASE | | | Efficiency | | ASE | | | Efficiency | |
| $G_0$ | $\sigma_\epsilon^2$ | LL | PE | PW | RE1 | RE2 | LL | PE | PW | RE1 | RE2 |
| 0.4 | 0.11 | 0.400 | 0.382 | 0.388 | 1.047 | 1.015 | 0.029 | 0.027 | 0.028 | 1.054 | 1.011 |
| 0.4 | 0.25 | 0.490 | 0.454 | 0.470 | 1.079 | 1.036 | 0.031 | 0.029 | 0.029 | 1.083 | 1.019 |
| 0.8 | 0.11 | 4.117 | 3.546 | 3.696 | 1.161 | 1.042 | 0.178 | 0.147 | 0.149 | 1.215 | 1.019 |
| 0.8 | 0.25 | 6.312 | 5.010 | 5.157 | 1.260 | 1.029 | 0.436 | 0.365 | 0.369 | 1.194 | 1.010 |

Table 3: ASE Comparison on model with AR(4) errors

| | | n=100 | | | | | n=500 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | | ASE | | | Efficiency | | ASE | | | Efficiency | |
| $G_0$ | $\sigma_\epsilon^2$ | LL | PE | PW | RE1 | RE2 | LL | PE | PW | RE1 | RE2 |
| 0.2 | 0.11 | 0.380 | 0.371 | 0.371 | 1.024 | 0.999 | 0.028 | 0.026 | 0.027 | 1.045 | 1.019 |
| 0.2 | 0.25 | 0.446 | 0.424 | 0.435 | 1.051 | 1.026 | 0.031 | 0.029 | 0.029 | 1.070 | 1.014 |
| 0.5 | 0.11 | 1.908 | 1.472 | 1.539 | 1.296 | 1.045 | 0.096 | 0.070 | 0.071 | 1.373 | 1.013 |
| 0.5 | 0.25 | 4.159 | 3.183 | 3.321 | 1.307 | 1.044 | 0.181 | 0.125 | 0.127 | 1.446 | 1.018 |