Econometric Theory, **18**, 2002, 985–991. Printed in the United States of America.
DOI: 10.1017.S026646660218409X

NONPARAMETRIC ESTIMATION OF VOLATILITY FUNCTIONS: THE LOCAL EXPONENTIAL ESTIMATOR

FLAVIO A. ZIEGELMANN
University of Kent at Canterbury and
Universidade Federal do Rio Grande do Sul

Kernel smoothing techniques free the traditional parametric estimators of volatility from the constraints related to their specific models. In this paper the nonparametric local exponential estimator is applied to estimate conditional volatility functions, ensuring its nonnegativity. Its asymptotic properties are established and compared with those for the local linear estimator. It theoretically enables us to determine when the exponential is expected to be superior to the linear estimator. A very strong and novel result is achieved: the exponential estimator is asymptotically fully adaptive to unknown conditional mean functions. Also, our simulation study shows superior performance of the exponential estimator.

1. INTRODUCTION

Nonparametric methods can offer flexible ways of estimating volatility functions. In a nonparametric context, kernel smoothing techniques provide estimators with nice asymptotic properties. Silverman (1986), Wand and Jones (1995), and Fan and Gijbels (1996) give comprehensive coverage of these techniques, whereas Simonoff (1996) and Bowman and Azzalini (1997) present a more practical approach.

Regarding the specific estimation of variance functions, Härdle and Tsybakov (1997) propose an improved version of the direct estimator, whose idea is to separately estimate $I\!E(Y^2|X = x)$ and $I\!E(Y|X = x)$ and then combine them through $\mathrm{Var}(Y|X = x) = I\!E(Y^2|X = x) - I\!E(Y|X = x)^2$. Fan and Yao (1998) suggest a novel approach, asymptotically fully adaptive to the unknown conditional mean, estimating first the conditional mean and in a second step the conditional variance. In both steps the local linear estimator is used. Our idea in

The author's research was supported by a grant from CAPES (Brazil). The helpful comments of Dr. Qiwei Yao are very gratefully acknowledged. I am also grateful to Professor Phil Brown and Dr. Michael Cooper, who kindly revised my manuscripts, and to the co-editor and two referees of this journal for their comments. Address correspondence to: Flavio A. Ziegelmann, University of Kent at Canterbury, Institute of Mathematics and Statistics, Canterbury (Kent), UK CT2 7NF; e-mail: faziegelmann@yahoo.com.

© 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S026646660218409X Published online by Cambridge University Press

this paper is to introduce an exponential estimator of the conditional variance in the second step, ensuring the nonnegativity of the estimator of the variance. In addition, as a novel result, we prove the complete asymptotic adaptiveness of our estimator to the unknown conditional mean. Also, our estimator is presented as a particular case of a much broader class, which we call the "functional" class (studied by Cheng, Hall, and Turlach, 1999). It is straightforward to see that the local linear estimator is also a special case under the functional approach. The idea of this functional estimator is not new in kernel smoothing, although the application to volatility functions seems to be. Related work has been done specially in the context of estimating density and distribution probability functions. Hyndman and Yao (1999) use the exponential function to estimate conditional densities. Hall, Wolff, and Yao (1999) study a logistic function to estimate conditional distributions. Moreover, this functional approach with the exponential function is equivalent to using local likelihood estimation (Tibshirani and Hastie, 1987; Copas, 1995; Hjort and Jones, 1996) for the regression of $r^2 = \{Y - I\!E(Y|X = x)\}^2$ against X with a Gaussian likelihood for r^2 and logarithm link function.

In the next section our model is presented and the exponential estimator properties are highlighted, comparing them to those of the local linear estimator. In Section 3 we present a simulation study, comparing both estimators. Finally, in Section 4, a theoretical proof of the asymptotic distribution of the exponential estimator is given along with the statement of some regularity conditions.

2. THE LOCAL EXPONENTIAL ESTIMATOR

Let $\{(Y_t, X_t)\}$ be a two-dimensional strictly stationary process having the same marginal distribution as (Y, X). Let $m(x) = I\!E(Y|X = x)$ be the conditional mean function and $\sigma^2(x) = \mathrm{Var}(Y|X = x) \neq 0$ the conditional volatility function. We can now write our model as

$$Y_t = m(X_t) + \sigma(X_t)\varepsilon_t, \tag{1}$$

where $I\!E(\varepsilon_t|X_t) = 0$ and $\mathrm{Var}(\varepsilon_t|X_t) = 1$. In our problem let us define $X_t \equiv Y_{t-1}$. The local linear estimator of $m(x)$ is $\hat{m}(x) = \hat{a}$, where $\hat{a}$ is defined as

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{t=1}^{n} \{Y_t - a - b(X_t - x)\}^2 W_{h_1}(X_t - x), \tag{2}$$

where $W(.)$ is a symmetric density function on $\Re$, $W_h(u) = (1/h)W(u/h)$, and $h_1 > 0$ is a bandwidth, which is responsible for the smoothing level of the estimators. In the numerical implementation of this work the bandwidth is chosen empirically, being proportional to the optimal bandwidth in the case of estimating density functions when both $W(.)$ and the density to be estimated, $f(.)$, are Gaussian. It is well known from standard least squares theory that this es-

https://doi.org/10.1017/S026646660218409X Published online by Cambridge University Press

timator has a closed form. Fan and Gijbels (1996, p. 220) establish its asymptotic properties.

Our main interest in this work is estimating the conditional volatility function $\sigma^2(.)$, in (1). Our approach is similar to that of Fan and Yao (1998); i.e., we estimate $\sigma^2(x)$ in a second step, using the squared residuals obtained from estimating the conditional mean $m(x)$ in the first step. From our model,

$$r_t^2 = (Y_t - m(X_t))^2 = \sigma^2(X_t)\varepsilon_t^2. \tag{3}$$

Looking at equation (3) and taking the conditional expectation of the squared residuals, we obtain $\mathbb{E}(r_t^2|X_t) = \sigma^2(X_t)$. Therefore it is natural to estimate $\sigma^2(x)$ from a regression of our estimated squared residuals $\hat{r}_t^2$ on X_t. Hence the local functional estimator of the conditional variance is obtained from the solution of the following minimization problem:

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{\alpha,\beta} \sum_{t=1}^{n} \{\hat{r}_t^2 - \psi\{\alpha + \beta(X_t - x)\}\}^2 K_{h_2}(X_t - x), \tag{4}$$

where $h_2 > 0$ is a second bandwidth. Note that when $\psi(x) \equiv x$, we have the linear estimator of the variance, $\hat{\sigma}_l^2(x) = \hat{\alpha}$. In the case of our exponential estimator, the functional has the form $\psi(x) \equiv \exp(x)$ and we define the variance estimator as $\hat{\sigma}_e^2(x) = \exp\hat{\alpha}$. We will prove that this estimator is asymptotically fully adaptive to the unknown conditional mean. To simplify notation, consider $L(x,\theta) \equiv \psi\{\alpha + \beta x\} \equiv \exp\{\alpha + \beta x\}$, where $\theta = (\alpha, \beta)$. Define $\kappa_2 = \int u^2 K(u)\, du$ and $c_0 = \mathbb{E}\{\varepsilon_i^4\} > 1$. Also define $\sigma^{2(i)}(x) \equiv (\partial/\partial x)^i \sigma^2(x)$ and $L^{(i)}(x,\theta) \equiv (\partial/\partial x)^i L(x,\theta)$, for $i = 0,1,2$.

THEOREM 2.1. *Under certain regularity conditions stated in Section 4,* $(nh)^{1/2}\{\hat{\sigma}_e^2(x) - \sigma^2(x) - \theta_n\}$ *is asymptotically normal with mean 0 and variance* $f(x)^{-1}\sigma^4(x)(c_0 - 1)\int K^2(u)\, du$, *where* $\theta_n = h_2^2\mu_2(x) + o_p\{h_2^2 + h_1^2\}$, $\mu_2(x) = \frac{1}{2}\{\sigma^{2(2)}(x) - L^{(2)}(0,\theta_0)\}\kappa_2$, *and* $\theta_0 = (\alpha_0, \beta_0)$ *is uniquely defined by* $\sigma^{2(i)}(x) = L^{(i)}(0,\theta_0)$, $i = 0,1$.

If we compare the preceding result to that in Theorem 1 of Fan and Yao (1998) (regarding the linear estimator), we can see that the two estimators have exactly the same asymptotic variance (resulting in the same efficiency). However, if we consider the coefficient for the bias, for the exponential we have $[\sigma^{2(2)}(x) - L^{(2)}(0,\theta_0)]\kappa_2$, whereas for the linear it is equal to $\sigma^{2(2)}(x)\kappa_2$. Because $L^{(2)}(0,\theta_0) = \beta_0^2 e^{\alpha_0}$ is a nonnegative quantity, we can conclude that the bias of the exponential will be smaller if and only if $\sigma^{2(2)}(x)$ is a positive quantity greater than $\beta_0^2 e^{\alpha_0}/2$, i.e., at x where the second derivative of the conditional volatility is greater than $\beta_0^2 e^{\alpha_0}/2$.

3. SIMULATIONS

Consider the model $X_{t+1} = 0.235X_t(16 - X_t) + \varepsilon_t$, where ε_t's are $IIDN(0,0.3^2)$ random variables. This model was used previously in Yao and Tong (1994).

https://doi.org/10.1017/S026646660218409X Published online by Cambridge University Press

Our simulated series have 500 observations, and we generate 200 of them. We estimate the conditional mean and the conditional variance considering the two-step-ahead prediction, because the variance is constant for the one-step-ahead case. We have 131 equidistant grid points in the interval [2,15]. Also, for each replicate we compute the mean absolute deviations given by $\delta_i = \frac{1}{41}\sum_{j=1}^{41}|\hat{\sigma}^2(x_j) - \sigma^2(x_j)|$, $i = 1,2,\ldots,100$, where x_j is the jth grid point. To perform the practical estimations for the exponential estimator we use the "downhill simplex" algorithm (see Press, Teukolsky, Vetterling, and Flannery, 1992, Sect. 10.4).

Parts (a)–(c) of Figure 1 show the true functions, the estimated functions, and the two-standard-deviation confidence intervals. All functions at time t are plotted against the grid points, which represent X_{t-2}. In part (a) we have the local linear estimator of the conditional mean. Comparing the estimators of the variance in parts (b) and (c), we notice better behavior for the exponential on the right-hand-end side of the grid. In part (d) the box plots of the mean absolute deviations confirm our findings and show remarkable superiority of the exponential estimator. They also seem to agree with the theoretical result in the sense that the estimators suggest similar variances and in this case smaller bias for the exponential.

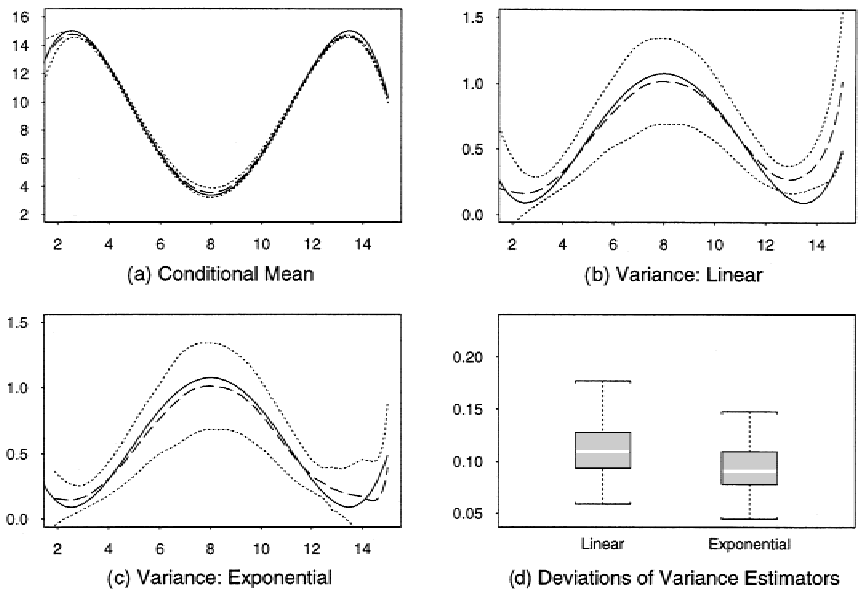


FIGURE 1. (a) Local linear estimator of the conditional mean. (b) and (c) Conditional variance estimators. The true functions, estimated functions, and confidence intervals are represented by continuous, dashed, and point lines, respectively. (d) Box plots of the mean absolute deviations.

https://doi.org/10.1017/S026646660218409X Published online by Cambridge University Press

4. PROOFS

The following regularity conditions are assumed.

(C1) For fixed x, $f(x) > 0$, $\sigma^2(x) > 0$, f is continuous at x and $\sigma^2(.)$ has two continuous derivatives in a neighborhood of x. Moreover $\mathbb{E}\{Y^k|X = x\}$ is continuous at x for $k = 3,4$ and $\mathbb{E}\{Y^{4(1+\delta)}\} < \infty$ for $\delta \in [0,1)$.

(C2) The kernels W and K are symmetric, compactly supported densities satisfying $|W(x_1) - W(x_2)| \leq C|x_1 - x_2|$ and $|K(x_1) - K(x_2)| \leq C|x_1 - x_2|$ for $x_1, x_2 \in \Re$.

(C3) The process $\{(X_t, Y_t)\}$ is absolutely regular; i.e., as $j \to \infty$,

$$\beta(j) \equiv \sup_{i \geq 1} \mathbb{E}\left\{ \sup_{A \in \mathcal{F}_{i+j}^{\infty}} |P(A|\mathcal{F}_1^i) - P(A)| \right\} \to 0,$$

where $\mathcal{F}_i^j$ denotes the σ-field generated by $\{(X_k, Y_k) : i \leq k \leq j\}$. Further, $\sum_{j \geq 1} j^2 \beta(j)^{\delta/(1+\delta)} < \infty$ for the same δ as in condition (C1).

(C4) As $n \to \infty$, $h_i \to 0$ and $\liminf_{n \to \infty} nh_i^4 > 0$, $i = 1,2$.

Now to prove Theorem 2.1, consider a Taylor expansion of $L(X_t - x, \theta) = \exp\{\alpha - \beta(X_t - x)\}$ in equation (4). Hence, we can write

$$R(\theta; x) = \sum_{t=1}^{n} \left[\hat{r}_t^2 - \sum_{j=0}^{1} (j!)^{-1} L^{(j)}(0, \theta)(X_t - x)^j - \frac{1}{2} L^{(2)}\{c_t(X_t - x), \theta\}(X_t - x)^2 \right]^2 K_{h_2}(X_t - x), \tag{5}$$

where $c_t \in [0,1]$. Define $R^*(\theta; x)$ as $R(\theta; x)$ with θ replaced by the functional estimator $\hat{\theta}_l$ in the last term $L^{(2)}\{c_t(X_t - x), \theta\}$. Let now $\hat{\theta}^*$ denote the minimizer of $R^*(\theta; x)$ and put $\hat{\sigma}_*^2(x) = L(0, \hat{\theta}^*)$. Therefore, to derive Theorem 2.1 we can show that the result holds for $\hat{\sigma}_*^2(x)$ and that

$$\hat{\sigma}_e^2(x) = \hat{\sigma}_*^2(x) + o_p(h_2^2). \tag{6}$$

Using the same arguments as in the proof of Theorem 1 in Hall et al. (1999), we can show that the ratio of $\hat{\sigma}_*^2 - \sigma^2(x)$ and the quantity

$$(nh_2)^{-1} f(x)^{-1} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_2}\right)\{\xi_t + R_t(X_t - x)^2\} \tag{7}$$

is equal to $1 + o_p(1)$, where $\xi_t = \hat{r}_t^2 - \sigma^2(X_t)$ and

$$R_t = \frac{1}{2}[\sigma^{2(2)}\{x + c_t'(X_t - x)\} - L^{(2)}\{c_t(X_t - x), \hat{\theta}_l\}],$$

https://doi.org/10.1017/S026646660218409X Published online by Cambridge University Press

with $c_t' \in [0,1]$. Now, looking at (7), we can analyze this expression for each of its two terms. By the ergodic theorem, h_2^{-2} times the second term equals $\mu_2(x) + o_p(1)$. Regarding its first term, it can be decomposed as

$$T_1 = (nh_2)^{-1} f(x)^{-1} \sum_{t=1}^{n} K\left(\frac{X_t - x}{h_2}\right) \times [\sigma^2(X_t)\varepsilon_t^2 + 2\sigma(X_t)\varepsilon_t\{m(X_t) - \hat{m}(X_t)\} + \{m(X_t) - \hat{m}(X_t)\}^2 - \sigma^2(X_t)],$$

because $\hat{r}_t^2 = \{\sigma(X_t)\varepsilon_t + m(X_t) - \hat{m}(X_t)\}^2$. Looking at each of the four terms of the preceding expression and using the results in Appendix 2 of Fan and Yao (1998), we can show that

$$(nh_2)^{1/2}(T_1 + \zeta) \to N\left(0, f(x)^{-1}\sigma(x)^4(c_0 - 1)\int K(t)^2\,dt\right),$$

where $\zeta = o(h_2^2 + h_1^2)$, completing the first part of our proof.

Now we have to prove (6) and combine both results to get the whole proof. For that, we prove the consistency of our estimator, which comes directly from the proof of Proposition 1 in Hall et al. (1999) (not in the published version), in our case with

$$D_n(x,\theta) = (nh_2)^{-1} \sum_{t=1}^{n} \left(1, \frac{X_t - x}{h_2}\right)' \{\hat{r}_t^2 - L(X_t - x, \theta)\} \times L(X_t - x, \theta) K\left(\frac{X_t - x}{h_2}\right)$$

and

$$D(x,\theta,h_2) = \frac{f(x)}{h_2} \int (1,t)' L(0,\theta) K(t) \times \{\sigma^2(x) - L(0,\theta) + \sigma^{2(1)}(x) - L^{(1)}(0,\theta)\}\,dt.$$

Therefore, to prove (6) and complete our proof, it is enough to use the same arguments that Hall et al. (1999) give in the second part of their proof of Theorem 1. ■

REFERENCES

Bowman, A.W. & A. Azzalini (1997) *Applied Smoothing Techniques for Data Analysis*. New York: Oxford University Press.

Cheng, M.-Y., P. Hall, & B.A. Turlach (1999) High-derivative parametric enhancement of nonparametric curve estimators. *Biometrika* 86, 417–428.

Copas, J.B. (1995) Local likelihood based on kernel censoring. *Journal of the Royal Statistical Society, Series B* 57, 221–235.

https://doi.org/10.1017/S026646660218409X Published online by Cambridge University Press

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

Fan, J. & Q. Yao (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* 85, 645–660.

Hall, P., R.C.L. Wolff, & Q. Yao (1999) Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* 94, 154–163.

Härdle, W. & A. Tsybakov (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics* 81, 233–242.

Hjort, N.L. & M.C. Jones (1996) Locally parametric nonparametric density estimation. *Annals of Statistics* 24, 1619–1647.

Hyndman, R.J. & Q. Yao (1999) Nonparametric Estimation and Symmetry Tests for Conditional Density Functions. Technical Report, University of Kent at Canterbury UKC/IMS/99/04.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, & B.P. Flannery (1992) *Numerical Recipes in C*. Cambridge: Cambridge University Press.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simonoff, J.S. (1996) *Smoothing Methods in Statistics*. New York: Springer.

Tibshirani, R. & T. Hastie (1987) Local likelihood estimation. *Journal of the American Statistical Association* 82, 559–567.

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

Yao, Q. & H. Tong (1994) Quantifying the influence of initial values on nonlinear prediction. *Journal of the Royal Statistical Society, Series B* 56, 701–725.

https://doi.org/10.1017/S026646660218409X Published online by Cambridge University Press