Kernel estimation under linear-exponential loss

by

Stanislav Anatolyev*
New Economic School, Moscow

Abstract

We consider nonparametric estimation of optimal predictors when the loss function is linear-exponential (Linex). We derive asymptotic distributions of the local constant and local linear kernel estimators under Linex, and discuss the rules for the optimal bandwidth.

Key Words and Phrases: Linear-exponential loss, Kernel methods, Optimal bandwidth. **JEL classification codes:** C13, C14.

*Address for correspondence: Stanislav Anatolyev, New Economic School, Nakhimovsky Prospekt, 47, Moscow, 117418 Russia. Phone: +7 (095) 129-1700. Fax: +7 (095) 129-3722. E-mail: sanatoly@nes.ru.

1 Optimal prediction under Linex

The symmetric quadratic (Quad) loss function

$$Q(u) = u^2 \tag{1}$$

is prevailing in econometrics by the virtue of its convenience and tractability, which are consequences of linearity of the derivative of $Q(u)$ with respect to its argument. In particular, this leads to a simple form of optimal predictor of y given x, which is

$$q(x) = E\left[y|x\right]. \tag{2}$$

The use of Quad loss, however, often contradicts reality where economic agents put different weights to overprediction and underprediction (e.g., Stockman, 1987; Tversky and Khaneman, 1991; West, Edison and Cho, 1993). A tractable example of an asymmetric loss function is linear exponential (Linex) which has become a workhorse in the literature on asymmetric loss. It has the form

$$L(u) = \exp\left(\alpha u\right) - \alpha u - 1, \tag{3}$$

where the known parameter α indexes the degree of asymmetry. When $\alpha > 0$, the loss is nearly exponential for positive errors, and nearly linear for negative errors; thus the loss is smaller for overprediction than for underprediction. The Linex loss function was initially introduced by Varian (1974) in the context of real estate assessment; estimation under the Linex loss from the Bayesian perspective was studied by Zellner (1986).

The existing literature on econometric analysis under asymmetric Linex loss is limited to parametric inference; see, for example, Weiss (1996), Christoffersen and Diebold (1997), Batchelor and Peel (1998), Hwang, Knight and Satchell (2001), among others. In this paper we instead take a nonparametric approach and present a set of results related to kernel estimation under the Linex loss. We take as a starting point the result that under Linex, the optimal predictor of y given x is (e.g., Zellner, 1986)

$$g(x) = \alpha^{-1} \log E\left[\exp\left(\alpha y\right)|x\right]. \tag{4}$$

Let us use the following terminology. We call the predictor $q(x)$ *Quad-optimal*, the predictor $g(x)$ *Linex-optimal*, the function $u \mapsto \exp\left(\alpha u\right)$ the *Linex transformation*, and the function

$u \mapsto \alpha^{-1} \log(u)$ the *antiLinex transformation*. Then, the Linex-optimal predictor of y given x is equal to the antiLinex-transformed Quad-optimal predictor of the Linex-transformed y given x. Let us also denote this Quad-optimal predictor by $h(x)$:

$$h(x) = E\left[\exp\left(\alpha y\right)|x\right].$$

2 Kernel estimators of Linex-optimal predictor

Suppose we are given a series of n independent and identically distributed pairs $(x_1, y_1), \cdots,$ (x_n, y_n). Let $K(\cdot)$ be a kernel function, and b be a bandwidth.

First, we modify the well-known Nadaraya–Watson kernel estimator to the case of Linex loss. The locally constant predictor $\hat{g}(x)$ at x is set to solve the following problem of minimization of the average kernel-weighted Linex loss:

$$\begin{aligned}
\hat{g}(x) &= \arg\min_{\beta_0} n^{-1} \sum_{i=1}^{n} L\left(y_i - \beta_0\right) K\left(\frac{x_i - x}{b}\right) \\
&= \arg\min_{\beta_0} n^{-1} \sum_{i=1}^{n} \left(\frac{\exp\left(\alpha y_i\right)}{\exp\left(\alpha \beta_0\right)} + \alpha \beta_0\right) K\left(\frac{x_i - x}{b}\right).
\end{aligned}$$

This problem has a unique closed-form solution

$$\hat{g}(x) = \alpha^{-1} \log \frac{\sum_{i=1}^{n} \exp\left(\alpha y_i\right) K\left(\frac{x_i - x}{b}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{b}\right)}. \tag{5}$$

The estimator so constructed is just the antiLinex-transformation of the usual Nadaraya–Watson estimator of $h(x)$. This makes sense because we are looking for a locally constant estimator.

Second, we modify the classic local linear estimator. The local linear predictor $\hat{g}(x)$ at x and its first derivative $\hat{g}'(x)$ are set to solve the following problem of minimization of the average kernel-weighted Linex loss:

$$\begin{aligned}
\begin{pmatrix} \hat{g}(x) \\ \hat{g}'(x) \end{pmatrix} &= \arg\min_{\binom{\beta_0}{\beta_1}} n^{-1} \sum_{i=1}^{n} L\left(y_i - \beta_0 - \beta_1 \left(x_i - x\right)\right) K\left(\frac{x_i - x}{b}\right) \\
&= \arg\min_{\binom{\beta_0}{\beta_1}} n^{-1} \sum_{i=1}^{n} \left(\frac{\exp\left(\alpha y_i\right)}{\exp\left(\alpha\left(\beta_0 + \beta_1\left(x_i - x\right)\right)\right)} + \alpha\left(\beta_0 + \beta_1\left(x_i - x\right)\right)\right) K\left(\frac{x_i - x}{b}\right).
\end{aligned}$$

This problem, unfortunately, does not have a closed-form solution, so the Linex-optimal predictor should be found using numerical optimization techniques. Note however that the objective function is strictly convex with respect to the parameters, so the solution is unique and can easily be obtained numerically.

Asymptotic results for such estimators may be easily deduced using the statistical literature on the so-called local quasi-likelihood estimation (see, for example, Staniswalis (1989) and Fan, Heckman and Wand (1995)). In particular, we have the following result.

Proposition 1 *Let the kernel K be a symmetric density with support $[-1,1]$, the density $f(x)$ be continuously differentiable, the function $g(x)$ be three times continuously differentiable, the conditional variance* $\mathrm{var}\left(\exp\left(\alpha y_t\right)|x\right)$ *exist and be twice continuously differentiable. Let x be a fixed point isolated from the boundaries of its support, and* $\mathrm{var}\left(\exp\left(\alpha y_t\right)|x\right)$ *be nonzero. Then, provided that $b \to 0$ and $nb^3 \to \infty$ as $n \to \infty$,*

$$\sqrt{nb}\left(\frac{\mathrm{var}\left(\exp\left(\alpha y_t\right)|x\right)}{\alpha^2 h(x)^2 f(x)} R_K\right)^{-1/2}\left(\hat{g}(x) - g(x) - b^2\frac{B(x)}{\alpha h(x)}\sigma_K^2\right) \xrightarrow{d} N(0,1),$$

where $R_K \equiv \int K(u)^2 du$, $\sigma_K^2 \equiv \int u^2 K(u) du$, and $B(x) \equiv h''(x)/2 + h'(x)f'(x)/f(x)$ when the local constant estimator is used, and $B(x) \equiv h''(x)/2$ when the local linear estimator is used.

Proof. The results follow from application of Theorems 1a and 1b in Fan, Heckman and Wand (1995). ■

These asymptotic results are very similar to those obtained under Quad loss, and the difference reveals itself in two instances. First, the "dependent variable" is $\exp\left(\alpha y_t\right)$, the Linex-transformed y_t rather than the original y_t. Second, additional divisors $\alpha h(x)$ and $\left(\alpha h(x)\right)^2$ are present in the asymptotic bias and variance expressions due to the antiLinex-transformation. In fact, the result of Proposition 1 concerning the local constant estimator follows straightforwardly from the asymptotics of the Nadaraya–Watson estimator under Quad loss, the closed-form formula (5) for $\hat{g}(x)$, and the delta method.

The presented asymptotic results also hold in time series contexts when data are stationary and mixing (Robinson, 1983).

3 Optimal bandwidth

In this section we derive the formula for the optimal bandwidth that may be used in the plug-in method. While under Quad loss the measure of performance yielding optimal bandwidths is taken to be the (pointwise or integrated) mean squared error $E[Q(\cdot)]$, under Linex loss it is more reasonable as the measure of performance to take the (pointwise or integrated) expected Linex value $E[L(\cdot)]$. It is interesting that both criteria result in the same expression for the optimal bandwidth.

Proposition 2 *Suppose the conditions of Proposition 1 hold. Then the local optimal bandwidth rate in the sense of minimizing either the asymptotic expected Linex loss or the asymptotic mean squared error is*

$$b^*(x) = \left(\frac{R_K}{4\sigma_K^4} \frac{\mathrm{var}\left(\exp\left(\alpha y_t\right)|x\right)}{B(x)^2 f(x)} \right)^{1/5} n^{-1/5}.$$

Similarly, the global optimal bandwidth rate is

$$b^* = \left(\frac{R_K}{4\sigma_K^4} \frac{\int \mathrm{var}\left(\exp\left(\alpha y_t\right)|x\right) h(x)^{-2} f(x)^{-1} w(x) dx}{\int B(x)^2 h(x)^{-2} w(x) dx} \right)^{1/5} n^{-1/5},$$

where $w(x)$ is the chosen weight function.

Proof. Let $\zeta_n \sim N(\mu_n, \omega_n)$ represent the asymptotic distribution of $g(x) - \hat{g}(x)$. Under the asymptotic mean squared error criterion, the formula for $b^*(x)$ is obtained in the standard way by minimizing the expression $\mu_n^2 + \omega_n$, and the formula for b^* – by applying the same technique to its integrated analog.

Now when $b^*(x)$ and b^* are proportional to $n^{-1/5}$, μ_n^2 and ω_n are proportional to $n^{-4/5}$, and the difference between $g(x) - \hat{g}(x)$ and ζ_n is $o(n^{-2/5})$. Then

$$\begin{aligned}
E\left[L(g(x) - \hat{g}(x))\right] &= \exp\left(\alpha\mu_n + \frac{1}{2}\alpha^2\omega_n\right) - \alpha\mu_n - 1 + o(n^{-2/5}) \\
&= 1 + \left(\alpha\mu_n + \frac{1}{2}\alpha^2\omega_n\right) + \frac{1}{2}\left(\alpha\mu_n + \frac{1}{2}\alpha^2\omega_n\right)^2 + o\left(n^{-4/5}\right) \\
&\quad -\alpha\mu_n - 1 + o(n^{-2/5}) \\
&= \frac{1}{2}\alpha^2\left(\mu_n^2 + \omega_n\right) + o(n^{-2/5}) \\
&= \frac{1}{2}\alpha^2 E\left[Q(g(x) - \hat{g}(x))\right] + o(n^{-2/5}).
\end{aligned}$$

Hence, both the Linex and Quad criteria yield the same optimal bandwidths. ■

References

Batchelor, R. and D.A. Peel, 1998, Rationality testing under asymmetric loss. *Economics Letters* 6, 49–54.

Christoffersen, P.F. and F.X. Diebold, 1997, Optimal prediction under asymmetric loss. *Econometric Theory* 13, 808–817.

Fan, J., N.E. Heckman, and M.P. Wand, 1995, Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of American Statistical Association* 90, 141–150.

Hwang, S., J. Knight, and S.E. Satchell, 2001, Forecasting nonlinear functions of returns using LINEX loss function. *Annals of Economics and Finance* 2, 187–213.

Robinson, P.M., 1983, Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.

Staniswalis, J.G., 1989, The kernel estimate of a regression function in likelihood-based models. *Journal of American Statistical Association* 84, 276–283.

Stockman, A.C., 1987, Economic theory and exchange rate forecasts. *International Journal of Forecasting* 3, 3–15.

Tversky, A. and D. Khaneman, 1991, Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* 106, 1039–1061.

Varian, H.R., 1974, A Bayesian approach to real estate assessment. In: *Studies in Bayesian econometrics and statistics in honor of Leonard J, Savage* (ed. S.E. Fienberg and A. Zellner), North Holland, 195–208.

Weiss, A., 1996, Estimating time series models using the relevant cost function. *Journal of Applied Econometrics* 11, 539–560.

West, K.D., H.J. Edison, and D. Cho, 1993, A utility based comparison of some models of exchange rate volatility. *Journal of International Economics* 35, 23–46.

Zellner, A., 1986, Bayesian estimation and prediction under asymmetric loss functions. *Journal of American Statistical Association* 81, 446–451.