Modeling, estimation, inference and prediction under linear-exponential loss

by

Stanislav Anatolyev*
New Economic School, Moscow

Abstract

The quadratic loss function is prevailing in econometrics due to its convenience and tractability. Its use, however, often contradicts the reality where economic agents put different weights to overprediction and underprediction. The existing econometric literature on asymmetric loss does not divert radically from the analysis under quadratic loss, hence the progress in this part of econometric theory is not impressive. In this paper we take the linear-exponential (LinEx) loss function and demonstrate that by turning from conventional econometric concepts specific to the quadratic loss function, to analogs dictated by the LinEx, one may go much further in constructing optimal predictions than by sticking to conventional concepts. A parametric regression turns out to be conveniently represented in a multiplicative, rather than additive, form, which is a consequence of the exponential PML interpretation of the optimal LinEx loss. We introduce the notion of a LinEx-volatility, a counterpart to the conditional variance under quadratic loss. Among other things, we also consider nonparametric kernel estimation under the LinEx loss, and derive some asymptotic results. The methodology is illustrated using the series of US interest rates, stock market returns and GNP growth.

Key Words and Phrases: Linear-exponential loss, Nonparametric methods, Parametric ACD models, Forecasting. **JEL codes:** C22, C51, C52.

*Address for correspondence: Stanislav Anatolyev, New Economic School, Nakhimovsky Prospekt, 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru.

1 Introduction

The quadratic loss function is prevailing in econometrics by the virtue of its convenience and tractability. Its use, however, often contradicts the reality where economic agents put different weights to overprediction and underprediction (e.g., Stockman 1987, Tversky and Khaneman 1991, West, Edison and Cho 1993). There exists limited literature on econometric analysis under asymmetric loss. Newey and Powell (1987) carefully analyze estimation and testing under asymmetric least squares. Christoffersen and Diebold (1997) compare optimal, conventional, and intermediate predictors under asymmetric loss, and propose an approximation to the optimal predictor. Christoffersen and Diebold (1996) suggest using a piecewise-linear approximation to the loss function to arrive at another approximate solution of the prediction problem. Weiss (1996) suggests yet other alternative approximations to optimal forecasts. Patton and Timmermann (2003) construct a general theory of optimal forecasts under asymmetric loss, and derive some interesting properties of them. Elliott and Timmermann (2003) characterize the weights of optimal combinations of forecasts in the context of an asymmetric loss. Batchelor and Peel (1998) show that conventional tests for rational expectations yield biased results when agents are guided by asymmetric loss, and Elliott, Komunjer and Timmermann (2003a, 2003b) estimate a parametrization of an asymmetric loss function from the data basing on the rationality condition. The given references nearly exhaust the research on econometrics under (differentiable) asymmetric loss.

A tractable example of an asymmetric loss is the linear exponential (LinEx) function. It has the form

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

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), and estimation under LinEx loss was studied to some extent by Zellner (1986). Subsequently, due to its tractability, the LinEx loss became a workhorse in the literature on asymmetric loss. Christoffersen and Diebold (1997) used LinEx as an example of asymmetric loss for comparison of optimal and conventional predictors. Batchelor and Peel (1998) developed a valid test for unbiasedness of forecasts under LinEx loss. Hwang, Knight and Satchell (2001) derived optimal forecasts for some conventional volatility models under LinEx loss, and Knight, Satchell and Wang (2002) modified the value-at-risk methodology to the case of LinEx loss. Patton and Timmermann (2003) used as an example the LinEx loss coupled with the Markov Switching DGP to derive some interesting properties of optimal forecasts under asymmetric loss.

Under LinEx loss, the optimal predictor of y given x is (e.g., Zellner (1986))

$$g(x) = \alpha^{-1} \log E\left[\exp\left(\alpha y\right)|x\right]. \tag{2}$$

Starting from this point, one often observes the following scenario in the literature on prediction under asymmetric loss. Some familiar model and DGP are taken (e.g., conditionally normal with conditional variance following a certain volatility process as in Weiss (1996), Hwang, Knight and Satchell (2001), or a simple Markov Switching DGP as in Patton and Timmermann (2003)), the optimal predictor under LinEx loss is developed, and the discrepancies between the optimal and conditional mean predictors and between the respective loss values are analyzed. The following trivial example may help illustrate this strategy. Let the truth be $y_t \sim IIDN(0, \sigma^2)$, so the LS-true mean model is $y_t = \varepsilon_t$, $\varepsilon_t \sim WN$. Then the LS-best forecast is $\hat{y}_{t+1|t} = 0$, and the MSE loss is $E\left[\varepsilon_t^2\right] = \sigma^2$, the LINEX loss is $E\left[\exp\left(\alpha\varepsilon_t\right) - \alpha\varepsilon_t - 1\right] = \exp\left(\alpha^2\sigma^2/2\right) - 1$. The LinEx-true mean model is $\exp\left(\alpha y_t\right) = \exp\left(\alpha^2\sigma^2/2\right) + e_t$, $e_t \sim WN$. The LinEx-best forecast is $\hat{y}_{t+1|t} = \alpha^{-1}\log E\left[\exp\left(\alpha y_{t+1}\right)\right] = \alpha\sigma^2/2$, the forecast error in terms of y_t is $u_{t+1} = \varepsilon_{t+1} - \alpha\sigma^2/2$, the MSE loss is $E\left[u_{t+1}^2\right] = \sigma^2 + \left(\alpha\sigma^2/2\right)^2$, and the LINEX loss is $E\left[\exp\left(\alpha u_{t+1}\right) - \alpha u_{t+1} - 1\right] = \alpha^2\sigma^2/2$. This value is strictly smaller than the LINEX loss under LS forecasting.

The outlined strategy, however,

(a) leads at best to a rather involved analytic solution for the optimal predictor (e.g., Hwang, Knight and Satchell 2001), or an approximation to it, with cumbersome computations including simulation methods (e.g., Christoffersen and Diebold 1996, 1997);

(b) leads to results that are specific to a particular DGP, which should be parameterized up to the form of conditional density (e.g., Patton and Timmermann 2003, Hwang, Knight and Satchell 2001);

(c) in an autoregressive context, does not allow one to handle multiperiod predictions as "easily" as one-period ones (e.g., Christoffersen and Diebold 1996, Hwang, Knight and Satchell 2001).

These limitations discourage use by applied researchers of econometrics based on asymptotic loss.

It is recognized in the literature that consideration of alternative to quadratic loss functions requires reappraisal of some habitual notions. Granger (1969) and Christoffersen and

Diebold (1997) demonstrate unconditional and conditional biasedness of LinEx-optimal forecasts; Weiss (1996) stresses that "... for many non-quadratic forecast CFs [cost functions], the mean of the forecast errors is expected to be non-zero... A researcher who cannot say that this is a reasonable outcome should be probably not be using such a CF." (p. 540). Batchelor and Peel (1998) show that conventional rationality tests yield misleading results when facing the LinEx loss. However, such reappraisal have not gone much farther. As the aforecited literature attests, conventional notions of conditional mean and variance of original series still prevail in the analysis, together with the habitual attachment to unbiasedness, conditional normality and other concepts specific to the symmetric quadratic loss.

In this paper, we propose more drastic changes to the analysis by changing the principles of *modeling*. We introduce a notion of a *LinEx-regression*, which is free of the aforementioned shortcomings, and rests on explicit modeling of the main ingredient in (2), namely, the conditional mean of the transformed series, $E[\exp(\alpha y)|x]$. In brief, we consider convenient modeling the conditional mean of $\exp(\alpha y)$ directly, which obviously differs from habitual modeling of the conditional mean and (possibly) variance of y. A parametric regression turns out to be conveniently represented in a multiplicative, rather than additive, form, a consequence of the exponential PML interpretation of the optimal LinEx loss. We also consider nonparametric kernel estimation of optimal predictors, and derive some asymptotic results. In the autoregressive time series context, we introduce the notion of a LinEx-volatility, a counterpart to the conditional variance under quadratic loss, and show how one- and multi-step prediction can be performed.

Throughout, we illustrate the methodology with the experiments with autoregressive models using the following data.

1. T-bill returns: the differenced 3-Month Treasury Bill, secondary market rate. Frequency: weekly. Date range: January 1954 to December 2003, totaling to 2605 returns. The first 1000 are used in modeling, the rest – for forecasting exercises. Source: Board of Governors of the Federal Reserve System. The parameter α equals 3.

2. S&P500 returns: the differenced logarithm of the S&P500 index. Frequency: weekly. Date range: January 1950 to May 2003, totaling to 2783 returns. The first 1000 are used in modeling, the rest – for forecasting exercises. Source: `finance.yahoo.com`. The parameter α equals 30.

3. GNP growth: the differenced logarithm of seasonally adjusted annual US GNP in $bln. Frequency: quarterly. Date Range: Q1 1959 to Q3 2003. Source: U.S. Department of Commerce, Bureau of Economic Analysis. The parameter α equals 3.

Figure 1 presents scatterplots of each of these series, say y_t, and of their *LinEx-transformations*, $\exp(\alpha y_t)$, against the lagged value y_{t-1}. The plots clearly illustrate the idea that the LinEx-optimal predictor puts higher weights on positive errors than on negative, because the LinEx-transformation inflates the former so that they have greater influence upon the predictor when transformed errors are averaged.

The rest of the paper is organized as follows. Section 2 describes nonparametric estimation under LinEx loss using kernel methods. We then turn in Section 3 to parametric modeling, and consider a convenient family of models implicit in the LinEx loss. In Section 4 we introduce the notion of LinEx-volatility, a counterpart to the conditional variance in the case of quadratic loss. Section 5 contains discussions of one- and multi-period-ahead forecasting. Section 6 concludes.

2 Nonparametric LinEx-regression

In the absence of a parametric model, the optimal predictor may be evaluated nonparametrically. Suppose we want to modify the Nadaraya–Watson kernel estimator to the case of the LinEx loss. The locally constant predictor $\hat{g}(x)$ at x solves the following problem of minimization of the average kernel-weighted LinEx loss:

$$\begin{aligned}\hat{g}(x) &= \arg\min_{\beta_0} n^{-1}\sum_{t=1}^{n} L(y_t-\beta_0)\,K\left(\frac{\Delta(x_t,x)}{b}\right)\\ &= \arg\min_{\beta_0} n^{-1}\sum_{t=1}^{n}\left(\frac{\exp(\alpha y_t)}{\exp(\alpha\beta_0)}+\alpha\beta_0\right)K\left(\frac{\Delta(x_t,x)}{b}\right),\end{aligned}$$

where $K(\cdot)$ is a kernel function, b is a bandwidth, $\Delta(\cdot,\cdot)$ is some distance measure. The closed-form solution is

$$\hat{g}(x) = \alpha^{-1}\log\frac{\sum_{t=1}^{n}\exp(\alpha y_t)\,K\left(\frac{\Delta(x_t,x)}{b}\right)}{\sum_{t=1}^{n}K\left(\frac{\Delta(x_t,x)}{b}\right)},$$

i.e., the LinEx-transformation of the Nadaraya–Watson estimator of

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

This makes sense, as we are looking for a locally constant estimator. There is choice to be made, however, what to take as a distance $\Delta(x_t,x)$ between x_t and x even in the scalar regressor case – the difference between them, or the difference between the transformations, $\exp(\alpha x_t)$ and $\exp(\alpha x)$. It appears that the former way is more reasonable as the initial observations tend to be more evenly spaced than the transformed ones (see Figure 1). A generalization to the nearest neighbors regression is straightforward, once the distance measure is decided upon.

Suppose now that we want to arrive at the locally linear estimator. The locally linear predictor $\hat{g}(x)$ at x solves the following problem of minimization of the average kernel-weighted LinEx loss:

$$\begin{aligned}\begin{pmatrix}\hat{g}(x)\\ \hat{g}'(x)\end{pmatrix} &= \arg\min_{\beta_0,\beta_1} n^{-1}\sum_{t=1}^{n} L\left(y_t-\beta_0-\beta_1\left(x_t-x\right)\right) K\left(\frac{\Delta\left(x_t,x\right)}{b}\right)\\ &= \arg\min_{\beta_0,\beta_1} n^{-1}\sum_{t=1}^{n}\left(\frac{\exp\left(\alpha y_t\right)}{\exp\left(\alpha\left(\beta_0+\beta_1\left(x_t-x\right)\right)\right)}+\alpha\left(\beta_0+\beta_1\left(x_t-x\right)\right)\right) K\left(\frac{\Delta\left(x_t,x\right)}{b}\right),\end{aligned}$$

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

Asymptotic results on such estimators can be obtained from 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

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 $\operatorname{var}\left(\exp\left(\alpha y_t\right)|x\right)$ be twice continuously differentiable. Let x be isolated from boundaries of the support, and $\operatorname{var}\left(\exp\left(\alpha y_t\right)|x\right)$ be nonzero. Then under IID sampling,*

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

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 Nadaraya–Watson estimator is used, and $B(x) \equiv h''(x)/2$ when the locally linear estimator is used.

As one can see, such asymptotic results are very similar to those obtained under usual LS loss. The difference reveals itself in two instances: first, the "dependent variable" is $\exp\left(\alpha y_t\right)$ rather than y_t due to the LinEx-transformation, and second, additional divisors $\alpha h(x)$ and $\left(\alpha h(x)\right)^2$ are present in the bias and variance due to the "anti-LinEx-transformation". In fact, the above result for the Nadaraya–Watson estimator follows straightforwardly from asymptotics under LS loss, the closed-form formula for $\hat{g}(x)$, and the delta method. Analogous results can be established for optimal bandwidths. While under LS loss the objective function yielding optimal bandwidths is taken to be the (integrated) mean squared error, it is more reasonable under LinEx loss that the objective function be the (integrated) LINEX value. It is interesting that both criteria result in the same expression for the optimal bandwidth.

Corollary 1 *The optimal bandwidth rate in the sense of minimizing the asymptotic integrated LINEX loss or the asymptotic integrated MSE is*

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

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

Heuristically, the reason of this invariance is the fact that

$$\begin{aligned} E\left[L(N(\mu,\omega))\right] &= \exp\left(\alpha\mu + \frac{1}{2}\alpha^2\omega\right) - \alpha\mu - 1 \\ &\approx 1 + \left(\alpha\mu + \frac{1}{2}\alpha^2\omega\right) + \frac{1}{2}\left(\alpha\mu + \frac{1}{2}\alpha^2\omega\right)^2 - \alpha\mu - 1 \\ &\approx \frac{1}{2}\alpha^2\left(\mu^2 + \omega\right), \end{aligned}$$

as other terms are asymptotically negligible when μ^2 and ω tend to zero in a balanced way.

Asymptotic results are likely to also hold in time series contexts when data are stationary and mixing, as they do under LS loss (Robinson 1983).

Figure 2 depicts curves of Nadaraya–Watson LS- and LinEx-autoregression of first order, for T-bill returns ($\alpha = 3$) and S&P500 returns ($\alpha = 30$), with $n = 1000$ observations, and bandwidths 0.06 and 0.012. One can clearly see that the fitted LinEx-regression values lie uniformly above those for the LS-regression, which is consistent with the results of Christoffersen and Diebold (1997). The non-constant spread between the two curves supports the presence of conditional heteroskedasticity (for more on this, see section 4). Figure 3 depicts curves of local linear LS- and LinEx-autoregression of first order, for the same two series using the same bandwidths. The `co` (constrained optimization) Gauss package was used for minimization; for starting values we used OLS estimates.

Table 1 contains loss values under both loss functions. As expected, all four 2×2 matrices of loss values are "diagonal", i.e. minimal loss values in columns are attained on the main diagonal where the loss function matches the regression type.

3 Parametric estimation and inference

The object of intermediate interest is $h(x) \equiv E[\exp(\alpha y)|x]$. One may specify the LinEx-regression

$$E[\exp(\alpha y)|x] = h(x,\beta)$$

for some known up to β function $h(\cdot,\cdot)$, and explore what the LinEx loss suggests as an estimator of β (recall that the LS loss implies the OLS estimate). Observe that the minimal LinEx loss is

$$\begin{aligned} L(y-g(x)) &= \exp(\alpha(y-g(x))) - \alpha(y-g(x)) - 1 \\ &= \exp\left(\alpha\left(y-\alpha^{-1}\log h(x,\beta)\right)\right) - \alpha\left(y-\alpha^{-1}\log h(x,\beta)\right) - 1 \\ &= A(y) + \frac{\exp(\alpha y)}{h(x,\beta)} + \log h(x,\beta), \end{aligned}$$

where $A(y)$ does not depend on β. The sample mean loss is

$$n^{-1}\sum_{t=1}^{n} L(y_t - g(x_t)) = n^{-1}\sum_{t=1}^{n} A(y_t) + n^{-1}\sum_{t=1}^{n}\left(\frac{\exp(\alpha y_t)}{h(x_t,\beta)} + \log h(x_t,\beta)\right),$$

Now note that the sample mean loss is, up to a term independent of β, the minus conditional loglikelihood of a conditionally exponential random variable $\exp(\alpha y)$ with parameter $h(x,\beta)$ (which translates to $-y_t$ distributed as extreme value with parameters 0 and $\alpha^{-1}h(x_t,\beta)^{-1}$ conditionally on x). This implies that rather than using the usual decomposition of the dependent variable into a sum of regression function and regression error, the model is more convenient to handle in the multiplicative form

$$\exp(\alpha y) = h(x,\beta)\cdot\eta, \tag{3}$$

where η is the multiplicative *LinEx-regression error*, a random variable having the mean 1 conditionally on x. The natural distribution implied by the LinEx loss for η is standard exponential, similar to how the natural distribution for the additive LS-regression error implied by the LS loss is centered normal. The use of the standard exponential distribution implicit in the LinEx loss makes the LinEx estimate of β, $\hat{\beta}_{LE}$, consistent and asymptotically normal even if the true distribution of η is different. This is because the standard exponential distribution belongs to the linear exponential family and the LinEx estimation is in fact the exponential Pseudo-Maximum Likelihood (PML) estimation, which was considered in the context of random sampling by Gourieroux, Monfort and Trognon (1984). Hence, we have

Proposition 2 *Let $\beta \in int(\mathbb{B})$, where $\mathbb{B} \subseteq \mathbb{R}^k$, and $h(x,\beta)$ be Borel measurable for all $\beta \in \mathbb{B}$ and twice continuously differentiable in β for all $\beta \in \mathbb{B}$ for all x in its support. Then under IID sampling,*

$$\hat{\beta}_{LE} \xrightarrow{p} \beta$$

and

$$\sqrt{n}\left(\hat{\beta}_{LE} - \beta\right) \xrightarrow{d} N\left(0, V_{LE}\right),$$

where

$$V_{LE} = E\left[\frac{h_\beta(x,\beta)h_\beta(x,\beta)'}{h(x,\beta)^2}\right]^{-1} E\left[(\eta-1)^2 \frac{h_\beta(x,\beta)h_\beta(x,\beta)'}{h(x,\beta)^2}\right] E\left[\frac{h_\beta(x,\beta)h_\beta(x,\beta)'}{h(x,\beta)^2}\right]^{-1}.$$

When η is conditionally on x homoskedastic, the asymptotic variance simplifies to

$$V_{LE} = E\left[\frac{h_\beta(x,\beta)h_\beta(x,\beta)'}{h(x,\beta)^2}\right]^{-1} \mathrm{var}\left(\eta\right).$$

Evidently, the estimates $\hat{\beta}_{LE}$ should be obtained by numerical optimization even when a linear or exp-linear model is postulated for $h(x,\beta)$. The asymptotic variance may be consistently estimated in a straightforward way:

$$\begin{aligned}\hat{V}_{LE} &= n\left(\sum_{t=1}^{n} \frac{h_\beta(x_t,\hat{\beta}_{LE})h_\beta(x_t,\hat{\beta}_{LE})'}{h(x_t,\hat{\beta}_{LE})^2}\right)^{-1} \sum_{t=1}^{n}\left(\frac{\exp\left(\alpha y_t\right)}{h(x_t,\hat{\beta}_{LE})}-1\right)^2 \frac{h_\beta(x_t,\hat{\beta}_{LE})h_\beta(x_t,\hat{\beta}_{LE})'}{h(x_t,\hat{\beta}_{LE})^2} \\ &\times \left(\sum_{t=1}^{n} \frac{h_\beta(x_t,\hat{\beta}_{LE})h_\beta(x_t,\hat{\beta}_{LE})'}{h(x_t,\hat{\beta}_{LE})^2}\right)^{-1}.\end{aligned}$$

In the time series autoregressive context, the autoregressive version of (3) has the form

$$\exp\left(\alpha y_t\right) = h_t \eta_t, \tag{4}$$

where $h_t = h_t(\beta)$ is a function of past realizations of y_t and the finite-dimensional parameter β. This equation is reminiscent of an ACD (Autoregressive Conditional Durations) model of Engle and Russell (1998), with the variable y_t being α^{-1} times the logarithm of intertrade durations[1]. The literature on econometrics of ultra-high frequency finance can be useful to parametrize the evolution of h_t and conditional distribution of η_t. An up-to-date survey of parameterizations of the ACD model is contained in Hautsch (2002). The classic is the ACD(q,p) (Engle and Russell 1998) specification; modifications and extensions include

[1] As a by-product, we obtain the following interpretation of the estimation of ACD models: the estimation is carried out to minimize the LinEx loss with $\alpha = 1$ for log durations.

AACD(q,p) (Additive ACD, Hautsch 2002), and two LACD(q,p) (Bauwens and Giot 2000) specifications that after change of variables result in the following equations:

$$h_t = \omega + \sum_{j=1}^{p} \phi_j \exp\left(\alpha y_{t-j}\right) + \sum_{j=1}^{q} \psi_j h_{t-j}, \qquad (5)$$

$$h_t = \omega + \sum_{j=1}^{p} \chi_j \eta_{t-j} + \sum_{j=1}^{q} \psi_j h_{t-j} \qquad (6)$$

$$\log h_t = \omega + \alpha \sum_{j=1}^{p} \phi_j y_{t-j} + \sum_{j=1}^{q} \psi_j \log h_{t-j}, \qquad (7)$$

$$\log h_t = \omega + \sum_{j=1}^{p} \chi_j \eta_{t-j} + \sum_{j=1}^{q} \psi_j \log h_{t-j}. \qquad (8)$$

The LACD specification seems more logical to use given our exponential transformation of the original variable. In addition, the ACD and AACD models require unpleasant parameter constraints to guarantee positiveness of h_t, while the LACD models do not. Empirically, *ceteris paribus*, the LACD equation (8) seems to better fit typical duration data (Bauwens and Giot 2000). An analog of the previous proposition when y_t is dependent follows from Engle (2000, Theorem 1).

Table 3 presents the PML estimation results based on the exponential distribution implicit in the LinEx loss function. For all three series, the lagged values of h_t or $\log h_t$ are insignificant so that the order q equals 0. The order of the other part equals 1 for the T-bill and S&P500 returns (with the first lag often being non-significant or marginally significant in the case of S&P500), and 2 for the GNP growth. The additive ACD model fits better the T-bill returns, the usual ACD model – the S&P500 returns, and the LACD$_1$ model – the GNP growth.

The PML estimates are not asymptotically efficient if the true distribution of η_t is not exponential (recall that this exponentiality is equivalent to $-y_t$ being distributed as extreme value with parameters 0 and $\alpha^{-1}h(x_t,\beta)^{-1}$ conditionally on x, which may not hold in the data, just like conditional normality may not hold under LS loss). In this case it is possible to increase efficiency of estimation of β by using ML estimation basing on the true distribution. It is natural to consider distributions that encompass the standard exponential, such as Weibull (Engle and Russell 1998), Generalized Gamma (Tsay 2002) and Burr (Grammig and Maurer 2000) distributions. Here we try the Weibull distribution whose density is normalized to have the expectation of unity:

$$f\left(\varepsilon;\varsigma\right) = \frac{\varsigma}{\chi^{\varsigma}} \varepsilon^{\varsigma-1} \exp\left(-\left(\frac{\varepsilon}{\chi}\right)^{\varsigma}\right), \quad \chi = \Gamma\left(1+\varsigma^{-1}\right)^{-1}, \quad \varsigma > 0$$

For the three series, Table 4 contain the results. The parameter ς varies from 2.5 to 3.1 showing significant departures from the exponential distribution, to which corresponds the value $\varsigma = 1$.

4 LinEx-volatility

In this section we introduce a notion of volatility specific for the LinEx loss. Note that the conventional conditional volatility measure, $\text{var}_{t-1}(y_t)$, is poorly suited in the context of asymmetric loss as it neglects differences between positive and negative deviations from the optimal predictor, and the conditional mean is a non-optimal predictor in the LinEx context. A switch to $\text{var}_{t-1}(\exp(\alpha y_t))$ is also poorly motivated. A proper volatility notion should not only be tied to the degree of mismatch between the variable of interest and its predictor, but also this measure, together with the predictor, should be dictated by the adopted loss function. In the LS case, the expected loss function $MSE = E\left[(y_t - E_{t-1}y_t)^2\right]$ dictates the following LS-volatility measure, a conditional contribution of t's observation to the MSE,

$$\sigma_t^2 = E_{t-1}\left[(y_t - E_{t-1}y_t)^2\right] = \text{var}_{t-1}(y_t).$$

Similarly, in the LinEx case, the expected loss function under LinEx-optimal prediction,

$$\begin{aligned} LINEX &= E\left[\exp\left(\alpha\left(y_t - \alpha^{-1}\log E_{t-1}\left[\exp\left(\alpha y_t\right)\right]\right)\right) - \alpha\left(y_t - \alpha^{-1}\log E_{t-1}\left[\exp\left(\alpha y_t\right)\right]\right) - 1\right] \\ &= E\left[\exp\left(\alpha y_t\right)/E_{t-1}\left[\exp\left(\alpha y_t\right)\right]\right] - \alpha E\left[\left(y_t - \alpha^{-1}\log E_{t-1}\left[\exp\left(\alpha y_t\right)\right]\right)\right] - 1 \\ &= -\alpha E\left[\left(y_t - \alpha^{-1}\log E_{t-1}\left[\exp\left(\alpha y_t\right)\right]\right)\right] \end{aligned}$$

dictates the *LinEx-volatility* measure, a conditional contribution of t's observation to the LINEX,

$$\delta_t^2 = -\alpha E_{t-1}\left[\left(y_t - \alpha^{-1}\log E_{t-1}\left[\exp\left(\alpha y_t\right)\right]\right)\right],$$

which is interpreted as a measure of discrepancy between the variable of interest, y_t, and its optimal predictor in the sense of LinEx loss. In particular, in the simple case when y_t follows a conditionally heteroskedastic normal LS-autoregression,

$$y_t = E_{t-1}\left[y_t\right] + \varepsilon_t, \quad \varepsilon_t | I_{t-1} \sim N(0, \sigma_t^2),$$

we have

$$\delta_t^2 = \frac{\alpha^2}{2}\sigma_t^2,$$

so that the LinEx-volatility δ_t^2 is proportional to the conventional LS-volatility σ_t^2. In a more interesting cases, this correspondence breaks down.

An even more appealing representation of the LinEx-volatility is

$$\delta_t^2 = -E_{t-1}\left[\log \frac{\exp(\alpha y_t)}{E_{t-1}[\exp(\alpha y_t)]}\right],$$

which clearly has a positive sign by the conditional Jensen inequality (except that $\delta_t^2 = 0$ when y_t is perfectly predictable at $t-1$). Using the multiplicative representation (3) of the LinEx-regression, we have

$$\delta_t^2 = -E_{t-1}[\log \eta_t],$$

where η_t is the multiplicative error. When η_t is distributed independently from the past of y_t, the LinEx-volatility measure is a constant.

Figure 4 depicts curves of nonparametric regressions of $(y_t - \hat{g}(y_t))^2$ and $-\log \hat{\eta}_t$, where $\hat{g}(y_t)$ and $\hat{\eta}_t$ are obtained nonparametrically (see Section 2), on y_{t-1}, which give an idea of the LS- and LinEx-volatility dependence, for T-bills and S&P500 returns, with $n = 1000$ observations and bandwidths 0.06 and 0.012 (The LS-volatility measure is inflated by 8 and 200 times, respectively, to make the units comparable). One can clearly see that there is significant LS-heteroskedasticity in both series, and the LinEx-volatility also greatly depends on the history. There is much less comovement between the two volatility measures, however, unlike there was one between the two nonparametric regressions. Note that the spread between the nonparametric Nadaraya–Watson LS- and LinEx-regressions is roughly proportional to the LS-volatility pretty closely, which is again in line with the result of Christoffersen and Diebold (1997), assuming little departures from conditional normality.

Let us turn to parametric inference. In the previous section the parameter ς of the Weibull distribution was assumed constant, so that the series η_t was IID. This results in the LinEx-volatility function being constant. To explore the opposite possibility, we parameterize the parameter ς as some function of the past (cf. ARCH-type parameterizations of conditional variance under quadratic loss). In particular, we try the following LinEx-analogs of GARCH processes:

$$\varsigma_t^{-1} = \left(\varsigma_0 + \sum_{j=1}^{p} \varsigma_{1,j}\eta_{t-p}\right)^{-1} + \sum_{j=1}^{q} \varsigma_{2,j}\varsigma_{t-j}^{-1},$$

$$\varsigma_t^{-1} = \left(\varsigma_0 + \sum_{j=1}^{p} \varsigma_{1,j}\log \eta_{t-j}\right)^{-1} + \sum_{j=1}^{q} \varsigma_{2,j}\varsigma_{t-j}^{-1},$$

We name such models $ARCD(q,p)$ and $ARCD\text{-}l(q,p)$ (autoregressive conditional density), following Hansen (1994) who analogously parameterized additional density parameters (such as degrees of freedom of the Student's t distribution) in the GARCH-t framework. The reason for postulating the dynamics of ς_t^{-1} rather than of ς_t is the following. From the properties of the Weibull distribution it follows that $-\log \eta_t$ is distributed, conditionally on the past of y_t, as extreme value with parameters $\log \Gamma\left(1+\varsigma_t^{-1}\right)$ and ς_t^{-1} having the conditional mean of

$$\log \Gamma\left(1+\varsigma_t^{-1}\right) + \gamma \varsigma_t^{-1},$$

where $\gamma \approx .5772$ is Euler's constant. The conditional mean is nearly linear in ς_t^{-1}, hence it is natural to parametrize its evolution in a form close to linear. Of course, other parameterizations are also possible.

The results for the S&P500 return series are contained in Table 5. The GNP growth series does not exhibit time-varying LinEx-volatility, which is also understandable from the viewpoint of the quadratic loss analysis; the T-bill return exhibits rather weak LinEx-volatility clustering. Figure 5 depicts the evolution of δ_t^2 for the S&P500 series basing on the Weibull–ACD$(0,0)$–ARCD$(1,1)$ model, which attains the largest loglikelihood value. The clustering in this variable is apparent.

5 Prediction under LinEx-modeling

In this section we show that the outlined methodology with linear LinEx regressions allows one to easily make one- and multiperiod forecasts (almost) just like in the case of quadratic loss. Recall that the h-step-ahead LinEx-optimal predictor of y is

$$\hat{y}_{t+h|t} = \alpha^{-1} \log E_t\left[h_{t+h}\eta_{t+h}\right].$$

However, $E_t\left[h_{t+h}\eta_{t+h}\right] = E_t\left[E_{t+h-1}\left[h_{t+h}\eta_{t+h}\right]\right] = E_t\left[h_{t+h}E_{t+h-1}\left[\eta_{t+h}\right]\right] = E_t\left[h_{t+h}\right]$, as $E_{t+h-1}\left[\eta_{t+h}\right] = 1$. Now, if the dynamic model for h_{t+h} linear, for example in (5)–(6), the LinEx-optimal point forecast may be expressed as a function of consistently estimable (with no distributional assumptions placed on η_t) parameters. For example, in the ACD(0,1) case, the point forecast is $\hat{y}_{t+h|t} = \alpha^{-1}\log\left(\omega(1-\phi_1^h)/(1-\phi_1) + \phi_1^h \exp\left(\alpha y_t\right)\right)$; in the AACD(0,1) case, $\hat{y}_{t+h|t} = \alpha^{-1}\log\left(\omega(1-\chi_1^h)/(1-\chi_1) + \chi_1^h \exp\left(\alpha y_t\right)/h_t\right)$.

Note that multiperiod point predictions in the linear model are handled as "easily" as one-period ones, and no knowledge of the form of conditional density is needed; only that the conditional mean is correctly specified suffices (or, in other words, that the conditional mean of the multiplicative regression errors is unity). Recall a similar situation in linear autoregressive models under quadratic loss where one also needs to know that the conditional mean is correctly specified (or, in other words, that the conditional mean of the additive regression errors is zero), and no knowledge of the form of conditional density is needed.

Table 6 presents results of one-step-ahead prediction of the T-bill returns and S&P500 returns using the rolling scheme and two estimation methods: parametric, based on the

AR(1) LS-autoregression and ACD(0, 1) LinEx-autoregression, and nonparametric based on the Nadaraya–Watson first order autoregression. The rolling window width is $R = 1,000$; the number of out-of-sample predictions are $P = 1605$ in the T-bills case and $P = 1783$ in the S&P500 case. for both series and for both methods, the LinEx-optimal prediction outperforms the LS-optimal prediction in terms of the LINEX loss, and vice versa. Note that sometimes nonparametric methods can outperform simple parametric ones if the smoothing parameter is tuned carefully. It is worth mentioning that other parametric ACD specifications and nonparametric prediction based on the local linear regression fare worse.

The interval predictions can be computed if the from of conditional distribution if known. For example, if the natural exponential conditional distribution if assumed for $\exp(\alpha y_t)$ (recall again that this is equivalent to $-y_t$ being distributed as extreme value with parameters 0 and $\alpha^{-1}h(x_t, \beta)^{-1}$ conditionally on x), then the symmetric one-step-ahead interval forecast for y_{t+1} with significance level q is

$$\left[\alpha^{-1}\log h_t + \alpha^{-1}\log\log\frac{2}{2-q},\ \alpha^{-1}\log h_t + \alpha^{-1}\log\log\frac{2}{q}\right].$$

Many-step-ahead interval forecasts, one-step-ahead interval forecasts in case of unknown conditional distribution, as well as point forecasts with nonlinear mean dynamics may be computed using Monte–Carlo methods and bootstrapping in standard ways (see, e.g., Franses and van Dijk, 2000, pp. 117–125).

6 Conclusion

In this paper we took the linear-exponential loss function and demonstrated that by turning from conventional econometric concepts specific to the quadratic loss function, to analogs dictated by this loss function, one may go much further in constructing optimal predictions than the literature does by sticking to conventional concepts. The outlined theory touches upon only basics of the LinEx-counterpart of traditional LS-based econometrics, and when developed further, may become a convenient tool when a problem at hand requires consideration of asymmetric loss.

The tractability of the LinEx example hinges on the exponential PML interpretation of the LinEx loss, and many conclusions are actually derived from already existing results. Future research will show if such rethinking may be helpful with other examples of asymmetric loss functions and whether a general theory can be constructed.

References

1. Batchelor, R. and Peel, D.A. (1998) Rationality Testing under Asymmetric Loss. *Economics Letters* 6, 49–54.

2. Bauwens, L. and P. Giot (2000) The logarithmic ACD model: an application to the bid/ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique* 60, 117–149.

3. Christoffersen, P.F. and Diebold, F.X. (1996) Further Results on Forecasting and Model Selection under Asymmetric Loss. *Journal of Applied Econometrics*, 11, 561–571.

4. Christoffersen, P.F. and Diebold, F.X. (1997) Optimal Prediction Under Asymmetric Loss. *Econometric Theory* 13, 808–817.

5. Dufour, A. and Engle, R.F. (2000) Time and the price impact of a trade. *Journal of Finance* 55, 2467–2498.

6. Elliott, G. and Timmermann, A. (2003) Optimal forecast combinations under general loss function and forecast error distribution. Forthcoming in *Journal of Econometrics.*

7. Elliott, G., Komunjer I. and Timmermann, A. (2003a) Estimating Loss Function Parameters. Working Paper, UCSD.

8. Elliott, G., Komunjer I. and Timmermann, A. (2003b) Biases in Macroeconomic Forecasts: Irrationality of Asymmetric Loss? Working Paper, UCSD.

9. Engle, R.F. (2000) The econometrics of ultra-high frequency data. *Econometrica* 68, 1–22.

10. Engle, R.F. and J.R. Russell (1998) Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.

11. Fan, J., Heckman, N.E. 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.

12. Franses, P. and D. van Dijk (2000) Nonlinear Time Series Models in Empirical Finance. Cambridge University Press.

13. Gourieroux, C., Monfort, A. and A. Trognon (1984) Pseudo Maximum Likelihood Methods: Theory *Econometrica* 52, 681–700.

14. Grammig, J. and K.-O. Maurer (2000) Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal* 3, 16–38.

15. Granger, C.W.J. (1969) Prediction with a generalized cost of error function. *Operational Research Quarterly* 20, 199–207.

16. Hansen, B. (1994) Autoregressive conditional density estimation. *International Economic Review* 35, 705–730.

17. Hautsch, N. (2002) Testing the conditional mean function of autoregressive conditional duration models. Working paper, University of Konstanz.

18. Hwang, S., Knight, J. and Satchell, S.E. (2001) Forecasting Nonlinear Functions of Returns Using LINEX Loss Function. *Annals of Economics and Finance* 2, 187–213.

19. Knight, J., Satchell, S.E. and Wang, G. (2002) Value At Risk Linear Exponent Forecasts. Working paper, University of Western Ontario.

20. Newey, W.K. and Powell, J.L. (1987) Asymmetric least squares estimation and testing. *Econometrica* 55, 819–847.

21. Patton, A.J. and Timmermann, A. (2003) Properties of Optimal Forecasts. Working Paper, UCSD.

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

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

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

25. Tsay, R. (2002) *Analysis of Financial Time Series*, chapter 5. John Wiley & Sons.

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

27. 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.

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

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

30. Zellner, A. (1986) Bayesian Estimation and Prediction Under Asymmetric Loss Functions. *Journal of American Statistical Association* 81, 446–451.

Series	T-bills		S&P500	
	MSE	LINEX	MSE	LINEX
	Nadaraya–Watson estimator			
LS-regression	0.012713	1.0730	0.00023788	1.1524
LinEx-regression	0.013106	1.0709	0.00024977	1.1456
	Local linear estimator			
LS-regression	0.012550	1.0722	0.00023508	1.1512
LinEx-regression	0.013034	1.0700	0.00025139	1.1448

Table 1. Loss values from nonparametric kernel estimation.

	ACD	AACD	$LACD_1$	$LACD_2$
		T-bill returns		
ω	0.7834	0.7726	0.0700	−0.1971
	(0.0522)	(0.0467)	(0.0120)	(0.0500)
ϕ_1	0.2756		0.2485	
	(0.0501)		(0.0479)	
χ_1		0.3081		0.2699
		(0.0472)		(0.0494)
$LINEX$	−1.0728	−1.0725	−1.0734	−1.0727
		S&P500 return		
ω	1.0941	1.0968	0.1543	0.0954
	(0.0541)	(0.0516)	(0.0171)	(0.0422)
ϕ_1	0.0635		0.0267	
	(0.0391)		(0.0502)	
χ_1		0.0715		0.0597
		(0.0430)		(0.0340)
$LINEX$	−1.1551	−1.1551	−1.1555	−1.1551
		GNP growth		
ω	0.9215	0.9166	0.3023	0.1138
	(0.1636)	(0.1949)	(0.0549)	(0.1075)
ϕ_1	0.3009		0.3093	
	(0.0960)		(0.0939)	
ϕ_2	0.1776		0.1844	
	(0.0736)		(0.0842)	
χ_1		0.4997		0.2635
		(0.1742)		(0.0869)
χ_2		0.3411		0.1818
		(0.0989)		(0.0504)
$LINEX$	−1.5593	−1.5602	−1.5593	−1.5603

Table 3. Results of fitting the conditional exponential distribution.

	Weibull AACD	Weibull ACD	Weibull $LACD_1$
	T-bill returns	S&P500 returns	GNP growth
ω	0.7811	1.1692	0.3067
	(0.0464)	(0.0193)	(0.0762)
ϕ_1			0.3007
			(0.1390)
ϕ_2			0.1642
			(0.1102)
χ_1	0.2859		
	(0.0455)		
ς	2.555	2.0708	3.0701
	(0.2000)	(0.0870)	(0.3214)
LL	−0.5052	−0.7975	−0.8239

Table 4. Results of fitting the conditional Weibull distribution.

	Weibull ACD(0,0)–ARCD			Weibull ACD(0,0)—ARCD-l		
ω	1.1692	1.1792	1.1505	1.1692	1.1759	1.1543
	(0.0193)	(0.0198)	(0.0155)	(0.0193)	(0.0180)	(0.0163)
ς_0	2.0708	1.7199	−1.2618	2.0708	2.2733	7.820
	(0.0870)	(0.1650)	(0.9570)	(0.0870)	(0.0833)	(2.409)
ς_1		0.4191	22.23		0.5890	3.278
		(0.1744)	(8.60)		(0.1177)	(1.013)
ς_2			0.8576			0.6788
			(0.0494)			(0.0964)
LL	−0.7975	−0.7805	−0.6969	−0.7975	−0.7584	−0.7203

Table 5. Results of fitting the conditional Weibull distribution with time-varying parameter to the S&P500 returns.

Series				
	LS loss		LinEx loss	
Model	MSE	LINEX	MSE	LINEX
		T-bill returns		
AR(1)/ACD$(0,1)$	0.05523	0.4802	0.07884	0.3490
NW, $b = 0.3$	0.04658	0.8910	0.05836	0.5687
NW, $b = 0.5$	0.05326	0.5920	0.06971	0.3237
		S&P500 returns		
AR(1)/ACD$(0,1)$	0.0004911	0.2609	0.0005315	0.2318
NW, $b = 0.3$	0.0004913	0.2598	0.0005320	0.2320
NW, $b = 0.5$	0.0004912	0.2598	0.0005317	0.2318

Table 6. Forecasting quality measures in the prediction exercise.

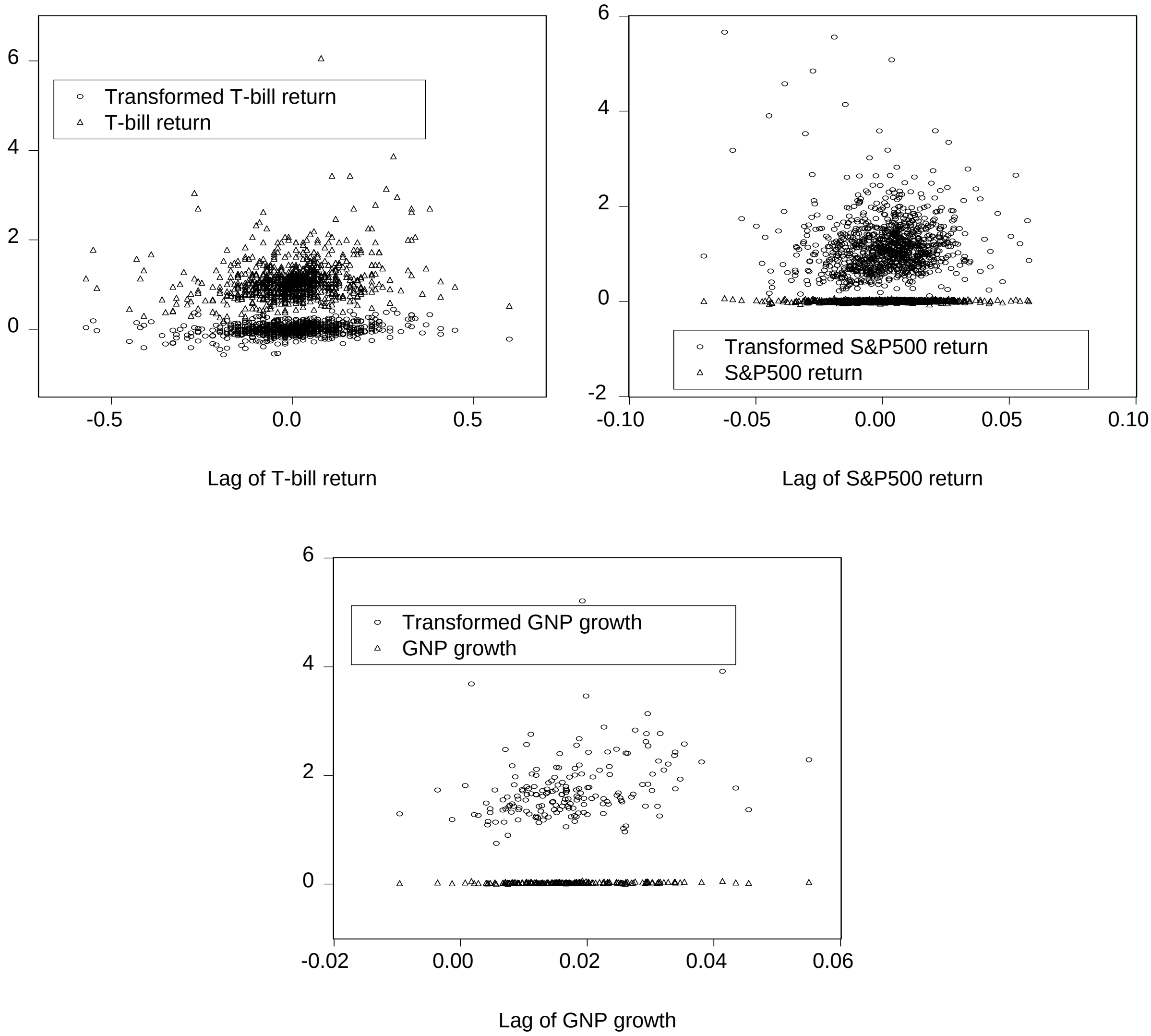

Figure 1. Scatter diagrams of raw and transformed series.

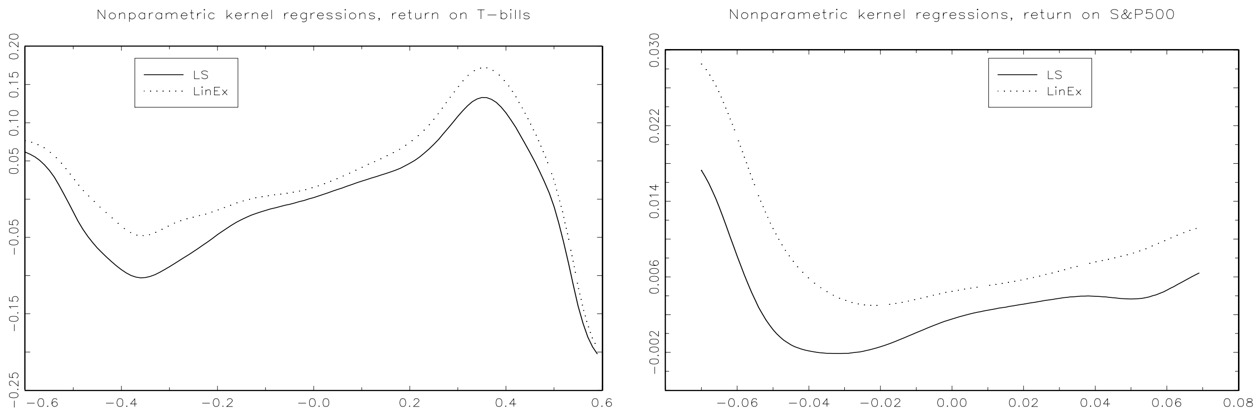

Figure 2. Nadaraya–Watson LS- and LinEx-regressions.

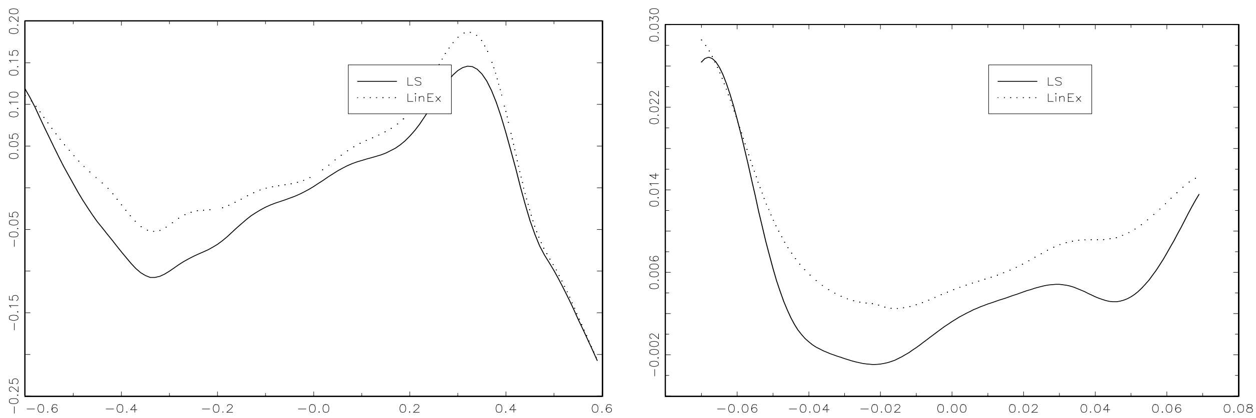

Figure 3. Local linear LS- and LinEx-regressions.

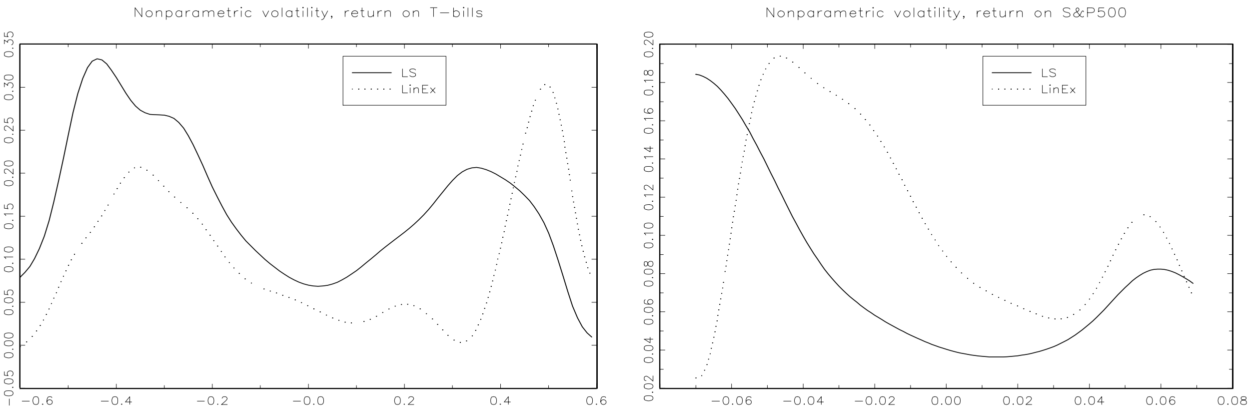

Figure 4. Nadaraya–Watson estimates of LS- and LinEx-volatility functions

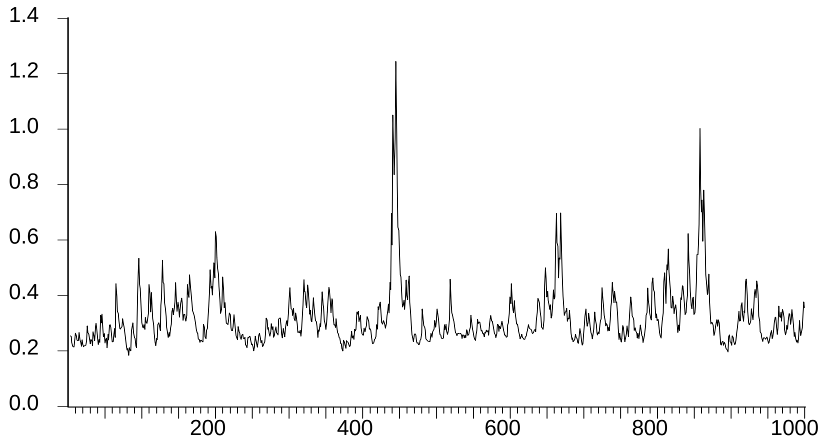

Figure 5. LinEx-volatility evolution for the S&P500 return series, from the Weibull model.