

Dynamic modeling under linear-exponential loss

by

Stanislav Anatolyev^{*}
New Economic School, Moscow

Abstract

We develop a methodology of parametric modeling of time series dynamics when the underlying loss function is linear-exponential (Linex). We propose to directly model the dynamics of the conditional expectation that determines the optimal predictor. The procedure hinges on the exponential quasi maximum likelihood interpretation of the Linex loss and nicely fits the multiple error modeling framework. Many conclusions relating to estimation, inference and forecasting follow from results already available in the econometric literature. The methodology is illustrated using data on United States GNP growth and Treasury bill returns.

Key words: Linear-exponential loss, optimal predictor, quasi maximum likelihood, multiple error model, autoregressive conditional durations.

JEL codes: C22, C51, C52

^{*}Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Pr., 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru. I thank the editor, two anonymous referees, participants of the 2004 Econometric Society European meeting in Leganés, Spain, and the New Economic School faculty.

1 Introduction

The symmetric quadratic (Quad) loss function

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

is prevailing in econometrics because of its convenience and tractability. These are consequences of the linearity of the first derivative of $Q(u)$. In particular, this leads to a simple form of optimal predictor of variable y_t given its past $I_{t-1} = \{y_{t-1}, y_{t-2}, \dots, y_1\}$, which is

$$q_t = E[y_t | I_{t-1}]. \tag{2}$$

The use of Quad loss, however, often contradicts reality, where economic agents put different weights on overprediction and underprediction. For example, macroeconomic forecasters often prefer to underpredict GNP growth because inventory holding costs exceed stockout costs (Elliott, Komunjer, and Timmermann, 2008); the IMF, OECD, and EC tend to systematically overpredict government budget deficits (Artis and Marcellino, 2001); and the Fed has systematically underpredicted or overpredicted inflation during different historical periods (Capistrán-Carmona, 2005). In financial markets, volatility underestimation leads to lower expected utility than its overestimation in making asset allocation decisions (West, Edison, and Cho, 1993); that investors are more sensitive to losses than to gains may help explain the equity premium puzzle (Benartzi and Thaler, 1995).

A tractable example of an asymmetric loss function is linear exponential (Linex), which has the form

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

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, and the reverse is true when $\alpha < 0$. The convenience of the Linex loss function is primarily due to its everywhere differentiability and continuity of its derivative, properties that are not shared by the Linlin $((1 - \theta) \mathbb{I}_{\{u < 0\}} + \theta \mathbb{I}_{\{u > 0\}}) |u|$ and Quadquad $((1 - \varphi) \mathbb{I}_{\{u < 0\}} + \varphi \mathbb{I}_{\{u > 0\}}) u^2$ loss functions, although, of course, it is not as convenient as the Quad loss function because of nonlinearity of the first derivative.

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). Subsequently, 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, conventional, and intermediate predictors. Batchelor and Peel (1998) developed a valid test for unbiasedness of forecasts under the Linex loss. Hwang, Knight, and Satchell (2001) derived optimal forecasts for some conventional volatility models under Linex, and Patton and Timmermann (2007) used as an example the Linex loss coupled with the Markov Switching DGP to derive some interesting properties of optimal forecasts. Anatolyev (2006) discussed nonparametric estimation under Linex loss.

In this paper, we propose a “direct” approach of parsimonious modeling under Linex. We directly model the dynamics of the conditional expectation that determines the optimal predictor, hence avoiding having to specify the whole conditional distribution. We

emphasize that the proposed models are not aimed at fitting the data better than the conventional time series models. Rather, they are better suited to construct optimal predictions when the underlying loss function is Linex.¹ A methodology close in spirit to ours has been encountered in the value-at-risk literature: while main parametric methodologies concentrate on modeling the dynamics and shape of the conditional distribution, the CAViaR model of Engle and Manganelli (2004) is based on directly modeling the dynamics of conditional quantiles.

Our method rests on the exponential quasi-maximum likelihood interpretation of the Linex loss function. As a result, it is convenient to decompose a suitably transformed original variable into a product of its conditional mean and a multiplicative disturbance, with the conditional mean following an autoregressive process. This nicely fits into the multiplicative error model (MEM) framework of Engle (2002), where the class of autoregressive conditional duration (ACD) models originated in Engle and Russell (1998) is a leading example. Thus, we are able to draw many conclusions directly from results already available in the MEM, ACD, and other related literatures. In particular, they suggest various forms of dynamic specifications and asymptotic results for estimators of parameters in these specifications.

The paper is structured as follows. In section 2 the modeling framework is presented, dynamic models are described, and forecasting is briefly discussed. In section 3 we tackle estimation issues, including asymptotic theory and the possibilities of improving estimation efficiency. In section 4, the proposed methodology is illustrated using data on US GNP growth and Treasury bill returns. We conclude in section 5.

2 Modeling under Linex

We take as a starting point the well-known result that under the Linex loss, the optimal predictor of stationary series y_t given I_{t-1} is (e.g., Zellner, 1986)

$$g_t = \alpha^{-1} \log E [\exp(\alpha y_t) | I_{t-1}]. \quad (4)$$

Note that for g_t to be well-defined, $E[\exp(\alpha y_t) | I_{t-1}]$ should exist for (almost) any realization of the history. This requires the conditional distribution of y_t to be sufficiently slim-tailed. If y_t is conditionally normal, g_t exists for any α ; if y_t is conditionally leptokurtic, g_t may be well-defined only for small values of α , while for larger α optimal prediction under Linex may be unavailable. When y_t is a log-difference of some variable x_t (for example, when it represents a growth rate or log-return), g_t is well-defined when $\sup_t E[(x_t/x_{t-1})^\alpha | I_{t-1}]$ is finite.

We will use the following terminology throughout. We call the predictor q_t defined in (2) *Quad-optimal* for y_t , the predictor g_t *Linex-optimal* for y_t , and the function $u \mapsto \exp(\alpha u)$ the *Linex transformation*. Let us also denote

$$h_t = E[\exp(\alpha y_t) | I_{t-1}].$$

At this point a researcher may estimate this conditional expectation nonparametrically and form a feasible Linex-optimal predictor. This nonparametric route is discussed in

¹In effect, our approach lies within the “decisionmetrics” paradigm of Skouras (2007), where an econometric model is developed so that it serves a particular purpose rather than being used in a variety of contexts.

Anatolyev (2006). In this paper we take a parametric approach, but instead of postulating the dynamics of the conditional distribution of y_t given I_{t-1} , or of its first and/or second conditional moments, we intend to model the dynamics of h_t directly, because it is the description of *its* dynamics that is sufficient for formulating the Linex objective function.

Suppose that an econometrician is equipped with the Quad loss, and the dynamics of q_t is parameterized as some function of a finite-dimensional parameter vector. The expected minimal Quad loss corresponding to one observation equals

$$E [Q (y_t - q_t)] = E [(y_t - q_t)^2],$$

whose minimization is equivalent to the maximization of expected conditional log-density of a normal distribution with mean q_t and fixed variance. Hence, estimation under Quad can be interpreted as quasi-maximum likelihood (QML) estimation based on the normal density. Similarly, estimation under absolute loss is associated with QML estimation based on the Laplace density.

Let us consider the Linex loss. Denote by β the true value of the $k \times 1$ parameter b in a dynamic parameterization $h_t(b)$ of h_t constrained to yield positive values for all allowable b . The expected minimal Linex loss corresponding to one observation equals, using (3) and (4),

$$\begin{aligned} E [L (y_t - g_t)] &= E [\exp (\alpha (y_t - g_t)) - \alpha (y_t - g_t) - 1] \\ &= E \left[\frac{\exp (\alpha y_t)}{h_t} + \log h_t \right] + C, \end{aligned}$$

where $C = -E [\alpha y_t + 1]$ does not depend on b . Therefore, the minimization problem

$$\min_b E [L (y_t - g_t)]$$

is equivalent to the maximization problem

$$\max_b E \left[-\frac{\exp (\alpha y_t)}{h_t(b)} - \log h_t(b) \right].$$

The object under the expectation sign is a conditional log-density of the random variable $\exp (\alpha y_t)$ distributed exponentially with mean $h_t(b)$. Thus, minimization of the expected Linex loss is equivalent to, or can be interpreted as, QML estimation on the basis of exponential density for the Linex-transformed variable $\exp (\alpha y_t)$.

This key observation leads to the following implication. In contrast to the usual decomposition of the dependent variable into a sum of a regression function and an additive regression error

$$y_t = q_t + e_t \tag{5}$$

with the property $E [e_t | I_{t-1}] = 0$, which is convenient under Quad loss, under Linex it is more convenient to handle the model in the multiplicative form

$$\exp (\alpha y_t) = h_t \eta_t, \tag{6}$$

where η_t is the multiplicative *Linex-regression error* with the property $E [\eta_t | I_{t-1}] = 1$. This alternative decomposition is a particular case of a “multiplicative error model” (MEM) of

Engle (2002, section 4). The MEM approach is a convenient tool for building models of non-negative valued time series, and here the series $\exp(\alpha y_t)$ is positive by construction. According to Engle (2002), MEM modeling allows one to avoid the dependence of the range of the disturbance on the mean that is a feature of a corresponding additive model², and it is more convenient than alternatives in terms of estimation.

The leading application of the MEM approach in the econometrics literature is to durations between successive transactions in a high-frequency financial market.³ Such durations typically exhibit clustering of high and low values in time. The temporal clustering property of durations has led to the class of *autoregressive conditional durations* (ACD) models proposed by Engle and Russell (1998)⁴, which spurred a vast literature in modeling duration dynamics. Many important results and ideas may be borrowed from it, in particular regarding the dynamic parameterization of h_t . It is noticed that ACD modeling turns out to be surprisingly isomorphic to ARCH modeling (Engle and Russell, 1998; Engle, 2002), and thus the broad ARCH literature may also be utilized.

The classical ACD specification is $\text{ACD}_1(q, p)$ (Engle and Russell, 1998)

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

Here, the forcing variables are lagged values of the modelled variable, $\exp(\alpha y_t)$ in our case. In a simple modification, these lagged values are standardized by the conditional mean so that the forcing variables are lagged multiplicative errors:

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

This specification will be referred to as $\text{ACD}_2(q, p)$. It is called ‘‘Additive ACD’’ in Hautsch (2004).

A disadvantage of specifications (7) and (8) is a need to impose positivity constraints on the coefficients in order to ensure positivity on h_t for all t . This may be restrictive if the original variable y_t is only weakly or negatively autocorrelated. Two logarithmic modifications proposed in Bauwens and Giot (2000) do not require restrictions on coefficients. Like the ACD_1 and ACD_2 models, these differ from each other by what is set to be a forcing variable:

$$\log h_t = \omega + \alpha \sum_{j=1}^p \phi_j y_{t-j} + \sum_{j=1}^q \psi_j \log h_{t-j}, \quad (9)$$

$$\log h_t = \omega + \sum_{j=1}^p \chi_j \eta_{t-j} + \sum_{j=1}^q \psi_j \log h_{t-j}. \quad (10)$$

²The additive model for $\exp(\alpha y_t)$ is $\exp(\alpha y_t) = \mu_t + \omega_t$, where $\mu_t = E[\exp(\alpha y_t) | I_{t-1}]$ and $E[\omega_t | I_{t-1}] = 0$. Indeed, the range of ω_t has to be $[-\mu_t, +\infty)$, i.e., time dependent, with unpleasant consequences for maximum likelihood estimation.

³There are other applications of the MEM approach. Engle (2002) illustrates the MEM methodology using exchange rate realized volatilities. Chou (2005) models a high/low range of asset prices in the MEM framework. Engle and Gallo (2006) analyze the dynamics of three volatility indexes using a multivariate version of the MEM.

⁴Hence, exponential QML estimation of ACD models for durations implicitly minimizes the Linex loss with $\alpha = 1$ for log-durations.

We call these specifications $\text{LACD}_1(q, p)$ and $\text{LACD}_2(q, p)$, respectively. *A priori*, LACD specifications seem more logical to use given our exponential transformation of the original variable.⁵

Modifications of specifications (7)–(10) are possible, of course. Consider, for example, the following flexible specification:

$$h_t = \omega + \tau \exp \left(\alpha \sum_{j=1}^p \lambda_j y_{t-j} \right) + \sum_{j=1}^q \psi_j h_{t-j}. \quad (11)$$

When the original variable y_t follows a linear homoskedastic normal autoregression of order p , the true model for h_t is (11) with zero restrictions put on some coefficients, and the Linex errors η_t are IID lognormal. If the autoregression is conditionally heteroskedastic, the true model for h_t is the same, but the Linex errors are not IID. The same is true of the LACD_1 model (9). Other possible specifications can be found in Hautsch (2004), Fernandes and Grammig (2006), and Meitz and Teräsvirta (2006); which model is most sensible depends on the data at hand. To make sure that the mean equation is correctly specified, one may use the LM-type test for no remaining ACD described in Meitz and Teräsvirta (2006, cor. 9 and proc. 3). The Meitz–Teräsvirta test is an asymptotically valid analog of the Ljung–Box test for no autocorrelation applicable to ACD-type models, and it is robust to misspecification of conditional density.

Of course, the right hand side of any specification may also contain exogenous factors that are able to predict the variable of interest (or more precisely, its Linex transformation). In this case the information set I_{t-1} is expanded to include the history of exogenous factors. Naturally, deterministic regressors such as seasonal components can also be included in the right hand side.

Now we will briefly discuss forecasting. When the autoregressive dynamics of h_t is parameterized, one-step Linex-optimal forecasts can be generated straightforwardly.⁶ Using (4),

$$\hat{y}_{t+1|t}^L = \alpha^{-1} \log E [\exp(\alpha y_t) | I_{t-1}] = \alpha^{-1} \log (h_{t+1} E [\eta_{t+1} | I_{t-1}]) = \alpha^{-1} \log h_{t+1},$$

because $\exp(\alpha y_{t+1}) = h_{t+1} \eta_{t+1}$ and $E [\eta_{t+1} | I_t] = 1$. Generation of multistep Linex-optimal forecasts generally requires the knowledge of the unknown multistep transition distribution; hence, simulation-based methods are called for (e.g., Franses and van Dijk, 2000, section 3.5). However, in case the dynamic model for h_t is linear, which includes ACD_1 and ACD_2 specifications (7)–(8), the multistep point forecasts can be deduced by iterating the MEM model. For example, for $\text{ACD}_1(0, 1)$, the point forecast is

$$\hat{y}_{t+s|t}^L = \alpha^{-1} \log \left(\omega \frac{1 - \phi_1^s}{1 - \phi_1} + \phi_1^s \exp(\alpha y_t) \right),$$

⁵Bauwens and Giot (2000) discover that empirically the LACD_2 equation describes typical duration data better than the ACD_1 and LACD_1 models.

⁶Alternatively, in addition to modeling the whole conditional distribution as in Hwang, Knight, and Satchell (2001) and Patton and Timmermann (2007), various (computationally tedious) approximations to the optimal forecast are suggested in Weiss (1996) and Christoffersen and Diebold (1996, 1997).

whereas in the $\text{ACD}_2(0, 1)$ case,

$$\hat{y}_{t+s|t}^L = \alpha^{-1} \begin{cases} \log \left(\omega + \chi_1 \frac{\exp(\alpha y_t)}{h_t} \right) & \text{if } s = 1, \\ \log(\omega + \chi_1) & \text{if } s > 1, \end{cases}$$

where $s > 1$ is the horizon (of course, in practice consistent estimates are used in place of true parameters). This is a phenomenon familiar from Quad-optimal forecasting: when the model for conditional mean is linear ARMA, multistep point forecasts can be deduced analytically (e.g., Franses and van Dijk, 2000, section 2.2).

Concluding this section, we note that the modeling of h_t under Linex is analogous to the modeling of conditional θ -quantiles $\inf \{u : \Pr[y_t \leq u | I_{t-1}] \geq \theta\}$ under the Linlin loss function $((1 - \theta) \mathbb{I}_{\{u < 0\}} + \theta \mathbb{I}_{\{u > 0\}}) |u|$. It is well known that the conditional quantile function is the optimal predictor under Linlin. Whereas earlier parametric methodologies of evaluating conditional quantiles, and hence of value-at-risk (VaR) measures popular in finance, concentrated on modeling the dynamics and shape of the conditional distribution or of its tails, the recently proposed CAViaR methodology (Engle and Manganelli, 2004) is based on modeling the dynamics of conditional quantiles directly in the GARCH/ACD spirit.

3 Estimation under Linex

Let us denote by β the vector of parameters entering the dynamic specification. For example, in the $\text{ACD}_1(q, p)$ case, $\beta = (\omega, \phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q)'$.

The exponential density belongs to the linear exponential family of distributions. The theory of quasi (or pseudo) maximum likelihood estimation (e.g., Gouriéroux, Monfort and Trognon, 1984) implies that under suitable conditions $\hat{\beta}^L$ is consistent for β and asymptotically normal. Exact suitable conditions under which this asymptotics occurs are established in the ACD literature. Most general is the result in Theorem 1 in Engle (2000); see also Wooldridge (1994, section 4) and Hautsch (2004). Let $h_{\beta t} = \partial h_t / \partial b$, and denote by \hat{h}_t and $\hat{h}_{\beta t}$ the functions h_t and $h_{\beta t}$, respectively, evaluated at β . Construct a robust asymptotic variance estimator as

$$\hat{V}^L = n \left(\sum_{t=1}^n \frac{\hat{h}_{\beta t} \hat{h}'_{\beta t}}{\hat{h}_t^2} \right)^{-1} \left(\sum_{t=1}^n \left(\frac{\exp(\alpha y_t)}{\hat{h}_t} - 1 \right)^2 \frac{\hat{h}_{\beta t} \hat{h}'_{\beta t}}{\hat{h}_t^2} \right) \left(\sum_{t=1}^n \frac{\hat{h}_{\beta t} \hat{h}'_{\beta t}}{\hat{h}_t^2} \right)^{-1},$$

where n is the sample size. Under suitable conditions, $\hat{\beta}^L$ will be consistent and asymptotically normal, and \hat{V}^L will consistently estimate the asymptotic variance of $\hat{\beta}^L$. Note that in the special case when η_t is conditionally (on I_{t-1}) homoskedastic, the asymptotic variance simplifies, and may be estimated by

$$\hat{V}^L = \left(\sum_{t=1}^n \frac{\hat{h}_{\beta t} \hat{h}'_{\beta t}}{\hat{h}_t^2} \right)^{-1} \sum_{t=1}^n \left(\frac{\exp(\alpha y_t)}{\hat{h}_t} - 1 \right)^2.$$

The suitable conditions referenced above are usually hard to verify. Moment restrictions implicit in them may or may not tighten the requirement of sufficient slim-tailedness

of the conditional distribution of y_t . Primitive conditions can be established on a case-by-case basis for specific ACD models. Engle and Russell (1998) and Engle (2002) give a set of sufficient conditions for consistency and asymptotic normality for the stationary ACD₁(1,1) model. More generally, the aforementioned isomorphism between ACD and ARCH structures allows one to utilize asymptotic results on normal QML estimation from the ARCH literature. In particular, conditions for consistency and asymptotic normality of QML estimates are established by Francq and Zakoian (2004) for a linear GARCH(q, p) model (an analog of our ACD(q, p)) and by Ling and McAleer (2000) for the EGARCH(1, 1) model (including an analog of our LACD₂(1, 1)), among others. Most asymptotic results do not constrain the shape of the density of standardized errors but restrict them to be IID; many authors note, however, that this assumption is not critical. The moment conditions are specific for each parameterization: in the ACD(1, 1) case, for example, they require that $\sup_t E[\eta_t^{2+\delta}|I_{t-1}] < \infty$ for some $\delta > 0$ and $\sup_t E[\log(\psi_1 + \phi_1\eta_t)|I_{t-1}] < 0$ (Engle, 2002); these conditions do not rule out though even integratedness of $\exp(\alpha y_t)$.

The Linex estimator is a true maximum likelihood estimator if $\exp(\alpha y_t)$ is conditionally exponentially distributed, which is equivalent to $y_t \text{sign}(\alpha)$ being conditionally distributed as Gumbel with location parameter $|\alpha|^{-1} \log h_t$ and scale parameter $|\alpha|^{-1}$. If this condition does not hold, and it is not likely to hold in reality, the Linex estimator is not asymptotically efficient; in the same way least squares estimation is not asymptotically efficient under conditional non-normality. It is possible to increase the efficiency of estimation by using maximum likelihood estimation based on the true distribution. Of course, as it happens under Quad loss, a wrong specification of the density may lead to inconsistent estimation.

Natural candidates for the distributional assumption for η_t are the Weibull density as in Engle and Russell (1998) with the standard exponential as a special case, and the Burr density as in Grammig and Maurer (2000) with Weibull as a limiting case; other choices are also possible. When the Weibull distribution is chosen, the conditional density of η_t given I_{t-1} is

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

When $\varsigma = 1$, the Weibull distribution reduces to the standard exponential. When the Burr distribution is chosen, the conditional density of η_t given I_{t-1} is

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

When $\varrho \rightarrow 0$, the Burr distribution reduces to Weibull, and when $\varrho = 1$, it reduces to log-logistic. It is also possible to use positive distributions not containing the exponential. One option is a lognormal distribution

$$f(\eta; \kappa) = \frac{\varsigma}{\eta \sqrt{2\pi\kappa}} \exp\left(-\frac{(\log \eta + \kappa/2)^2}{2\kappa}\right), \quad \kappa > 0,$$

which is a true distribution for η if the original variable y_t is conditionally homoskedastic normal. The densities on the previous three displays are normalized to have the expectation of unity.

If the conditional density $f(\eta; q)$ is correctly specified, where q contains additional shape parameters whose true value is θ , the estimate

$$\begin{pmatrix} \hat{\beta}^f \\ \hat{\theta}^f \end{pmatrix} = \arg \max_{(b, q) \in \mathbb{B} \times \Theta} \sum_{t=1}^n \log \frac{f_t}{h_t},$$

where

$$f_t = \log f \left(\frac{\exp(\alpha y_t)}{h_t}; q \right),$$

will be consistent, asymptotically normal, and asymptotically efficient under suitable conditions (see Wooldridge, 1994, section 5). Moreover, the matrix

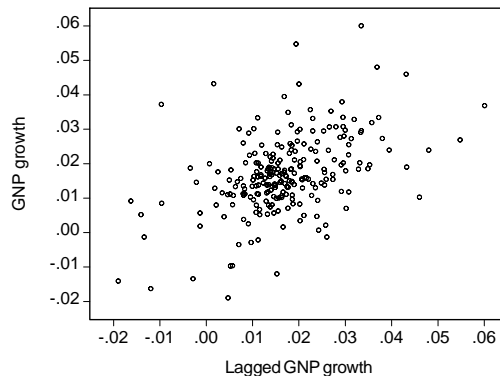
$$n \left(\sum_{t=1}^n \hat{s}_t \hat{s}_t' \right)^{-1},$$

where \hat{s}_t is $s_t = \partial \log(h_t^{-1} f_t) / \partial (b' q)'$ evaluated at $(\hat{\beta}^{f'} \hat{\theta}^{f'})'$, will consistently estimate the asymptotic variance of $(\hat{\beta}^{f'} \hat{\theta}^{f'})'$.

4 Empirical illustrations

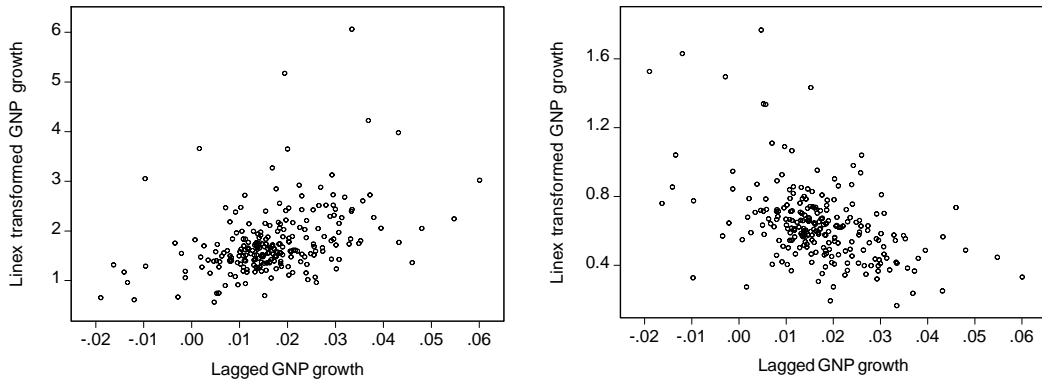
In this section, we illustrate the proposed methodology using data on two time series, GNP growth and Treasury bill returns. There is empirical evidence that macroeconomic forecasters often prefer to underpredict GNP growth (Elliott, Komunjer, and Timmermann, 2008), because additional inventory holding costs resulting from GNP overprediction exceed stockout costs resulting from GNP underprediction. This corresponds to negative α . On the other hand, financial investors may tend to overpredict the Treasury bill rate when evaluating projects because they are more sensitive to losses than to gains (Benartzi and Thaler, 1995); monetary authorities may also find it more beneficial to overpredict the T-bill rate considering future debt repayments and facing a restricted budget. This corresponds to positive α . In the illustrative empirical analysis, however, for the sake of symmetry we consider α of both signs for both series.

Figure 1: Scatterplot of GNP growth against its first lag.



The GNP growth series is a log-difference of seasonally adjusted quarterly US GNP⁷ from Q2:1953 to Q3:2005 totaling 210 observations. Figure 1 presents a scatterplot of GNP growth against its lagged value. The T-bill returns series is a difference of weekly 3-month US Treasury bill rates⁸ from Jan:1954 to May:1992 totaling 1,586 observations. For the latter series, we use the first 600 observations for modeling, and the rest for forecasting purposes. The loss function parameter α is set in absolute value to 30 for GNP growth and to 3 for T-bill returns. This is convenient for illustrative purposes because such values make the asymmetry of the Linex loss function pronounced yet not extremely severe. Such $|\alpha|$ imply values of 1.25 in the case of GNP growth and 1.27 in the case of T-bill returns for the ratio of Linex losses for positive and negative prediction errors of one standard deviation.⁹ Figure 2 presents scatterplots of Linex-transformed GNP growth against its lagged value. These diagrams illustrate how the Linex-optimal predictor puts different weights on errors of different sign after the Linex transformation inflates or deflates data points differently (cf. Figure 1).

Figure 2: Scatterplots of Linex transformed GNP growth against its first lag, $\alpha > 0$ (left panel) and $\alpha < 0$ (right panel).



Tables 1a and 1b present estimation results of fitting exponential ACD models (7)–(10) to GNP growth and T-bill returns, accompanied by various diagnostic statistics. The latter are: average loglikelihood values labelled “LogL”, values of the Meitz–Teräsvirta test statistic labelled “Q(3)”, values of the excess dispersion test statistic labelled “Disp”, and values of the Pearson goodness-of-fit test statistic labelled “Pear”. The Meitz–Teräsvirta test statistic is asymptotically distributed as $\chi^2_{(3)}$ under the null of correct mean specification, where 3 is the number of additional lags specified under the alternative. The excess dispersion test for exponential distribution is described in Engle and Russell (1998, p. 1144). The corresponding statistic is asymptotically distributed as $N(0, 1)$ under the null

⁷All data used in the paper is downloadable from the FRED[®] database of St. Louis Fed. GNP data is available at <http://research.stlouisfed.org/fred2/series/GNP>.

⁸Available at <http://research.stlouisfed.org/fred2/series/WTB3MS>.

⁹Such economic considerations can also be used by an econometrician to heuristically choose the value of α in empirical practice. A more technical way to fix α is to estimate it from some auxiliary data specific to the same economic agents (using methods similar to those in Elliott, Komunjer, and Timmermann, 2005). Experimentation shows that as long as α is set within reasonable limits, regression results are not very sensitive to its exact value. It is the sign of α , in addition to a wish to acknowledge the asymmetry in decision making and deviate from the Quad loss, that is important in such modeling.

of exponential distribution. The Pearson goodness-of-fit test (e.g., Kendall and Stuart, 1973, chapter 30) compares the multinomial distribution induced by Linex residuals and that implied by the standard exponential distribution. We set the number of equiprobable classes to 10, so the corresponding statistic is asymptotically distributed as $\chi^2_{(9)}$ under the null of correct distributional specification.

Table 1a. Results of fitting exponential ACD models to Linex-transformed GNP growth

	ACD ₁	ACD ₂	LACD ₁	LACD ₂
$\alpha > 0$				
ω	0.903 (0.147)	0.843 (0.186)	0.308 (0.048)	0.071 (0.103)
ϕ_1	0.333 (0.092)		0.340 (0.088)	
ϕ_2	0.150 (0.069)		0.130 (0.082)	
χ_1		0.552 (0.170)		0.290 (0.083)
χ_2		0.340 (0.102)		0.184 (0.052)
LogL	-1.5450*	-1.5460	-1.5450	-1.5462
Q(3)	2.76	1.70	0.44	3.13
Disp	-4.64	-4.63	-4.64	-4.63
Pear	442.4	448.5	432.9	444.3
$\alpha < 0$				
ω	0.317 (0.058)	0.276 (0.071)	-0.230 (0.049)	-1.036 (0.111)
ϕ_1	0.335 (0.090)		0.321 (0.086)	
ϕ_2	0.155 (0.095)		0.166 (0.082)	
χ_1		0.210 (0.062)		0.337 (0.095)
χ_2		0.141 (0.045)		0.227 (0.078)
LogL	-0.5248	-0.5257	-0.5248*	-0.5256
Q(3)	0.59	0.66	0.36	0.97
Disp	-4.74	-4.73	-4.74	-4.73
Pear	499.0	477.6	503.3	478.8

Note: robust standard errors are in parentheses.

Table 1b. Results of fitting exponential ACD models to Linex-transformed T-bill returns

	ACD ₁	ACD ₂	LACD ₁	LACD ₂
$\alpha > 0$				
ω	0.772 (0.065)	0.735 (0.064)	0.074 (0.015)	-0.257 (0.059)
ϕ_1	0.290 (0.062)		0.257 (0.063)	
χ_1		0.352 (0.066)		0.334 (0.059)
LogL	-1.0787	-1.0780	-1.0798	-1.0775*
Q(3)	6.34	5.19	8.33	4.67
Disp	-7.57	-7.58	-7.50	-7.60
Pear	1042.1	1093.8	1032.0	1103.8
$\alpha < 0$				
ω	0.733 (0.073)	0.769 (0.065)	0.044 (0.017)	-0.166 (0.060)
ϕ_1	0.299 (0.077)		0.299 (0.061)	
χ_1		0.275 (0.069)		0.204 (0.060)
LogL	-1.0374	-1.0381	-1.0370*	-1.0388
Q(3)	3.09	5.96	1.19	4.43
Disp	-7.26	-7.20	-7.25	-7.18
Pear	1077.8	1117.2	1065.8	1155.7

Note: robust standard errors are in parentheses.

The lag polynomial orders of ACD models are selected so that coefficients for more lagged values would yield insignificant estimates. As a result, for both series the order q equals 0, and the order p equals 2 for GNP growth and 1 for T-bill returns. The adequacy of selected orders is confirmed by the Meitz–Teräsvirta test for no remaining ACD (at least for best models in terms of LogL). The Linex residuals exhibit severe underdispersion (in contrast to overdispersion in ACD applications to duration data) and do not conform to the exponential distribution in all models, in a larger degree in the case of T-bill returns. A comparison of loglikelihood values allows one to determine which dynamic model fits better to the data; the corresponding LogL values are marked with an asterisk. In two out of four cases the LACD₁ model fares best, whereas the ACD₂ model wins in none¹⁰. Along with the four ACD models (7)–(10), we have fitted the data to the flexible model (11) and also (7)–(10) augmented with threshold-type asymmetries. None of these models has proved to be more successful: the additional parameters are statistically insignificant, and their presence pushes information criteria such as AIC upward.

Because diagnostic tests signal misspecification of the conditional density, we try to use the Weibull, Burr, and lognormal distributions in place of the exponential. In doing this, we use those models for the conditional mean that showed to advantage under the exponential specification. The results are presented in Tables 2a and 2b in columns “Weib” (standing for “Weibull”), “Burr”, and “Lnorm” (standing for “lognormal”); columns

¹⁰The ranking of specifications, however, is quite sensitive to which dates are set to be the beginning and end of the sample.

“Exp” (standing for “exponential”) reproduce corresponding results from Tables 1a and 1b. Because the distributions differ in the degree of parameterization, we also show values of the Akaike information criterion (line “AIC”). The excess dispersion test (line “Disp”) and Pearson goodness-of-fit test (line “Pear”) are adapted to the use of Weibull, Burr, and lognormal densities. In particular, the null distribution of the Pearson statistic is bounded between $\chi^2_{(9)}$ and $\chi^2_{(8)}$ in case of Weibull and lognormal, and between $\chi^2_{(9)}$ and $\chi^2_{(7)}$ in case of Burr, because of the presence of additional shape parameters (Kendall and Stuart, 1973, section 30.11–30.19).

Table 2a. Results of fitting best ACD models with various distributions to Linex-transformed GNP growth

	ACD ₁ , $\alpha > 0$				LACD ₁ , $\alpha < 0$			
	Exp	Weib	Burr	Lnorm	Exp	Weib	Burr	Lnorm
ω	0.903 (0.147)	0.939 (0.204)	0.890 (0.136)	0.894 (0.147)	−0.230 (0.049)	−0.223 (0.062)	−0.228 (0.045)	−0.233 (0.046)
ϕ_1	0.333 (0.092)	0.336 (0.128)	0.277 (0.079)	0.319 (0.088)	0.321 (0.086)	0.324 (0.109)	0.303 (0.073)	0.330 (0.083)
ϕ_2	0.150 (0.069)	0.114 (0.080)	0.212 (0.069)	0.167 (0.069)	0.166 (0.082)	0.192 (0.109)	0.188 (0.077)	0.149 (0.080)
ς		3.130 (0.325)	8.061 (0.988)			3.758 (0.248)	6.899 (0.714)	
ϱ			1.309 (0.314)				0.864 (0.206)	
κ				0.069 (0.010)				0.069 (0.010)
LogL	−1.545	−0.794	−0.543	−0.592	−0.525	0.350	0.472	0.428
AIC	3.119	1.627	1.133	1.222	1.078	−0.661	−0.895	−0.816
Disp	−4.64	−1.48	0.58	0.72	−4.74	−1.69	0.37	−0.24
Pear	442.4	66.5	6.0	19.2	503.3	38.0	7.1	16.2

Note: robust standard errors are in parentheses.

In all cases, the additional parameter ς in the Weibull distribution shows statistically significant departures from the value of 1 implied by the exponential distribution. The values of ς are close to (or slightly larger than) those in the CARR model of Chou (2005) for high/low ranges of asset prices; the values of ς in ACD applications are typically smaller than 1. The dispersion test signals no underdispersion, but the Pearson test still indicates inappropriateness of the Weibull density. The Burr conditional density improves the fit, even more boosting the likelihood and appreciably reducing the AIC. The Pearson test statistic also goes down significantly, so that in the case of GNP growth the Burr density is appropriate. In the case of T-bill returns, the Burr distribution is still rejected. Interestingly, in all cases the estimates of the additional shape parameter ϱ are about or closer than one standard deviation from the value of 1 implied by the log-logistic distribution. The use of the lognormal conditional density leads to a better fit than the use of Weibull, but it does not do as good a job as the Burr density.

Table 2b. Results of fitting best ACD models with various distributions to Linex-transformed T-bill returns

	LACD ₂ , $\alpha > 0$				LACD ₁ , $\alpha < 0$			
	Exp	Weib	Burr	Lnorm	Exp	Weib	Burr	Lnorm
ω	-0.257 (0.059)	-0.283 (0.067)	-0.231 (0.070)	-0.235 (0.060)	0.044 (0.017)	0.038 (0.019)	0.037 (0.015)	0.040 (0.058)
ϕ_1					0.299 (0.061)	0.383 (0.065)	0.228 (0.063)	0.268 (0.060)
χ_1	0.334 (0.059)	0.353 (0.065)	0.307 (0.070)	0.314 (0.060)				
ς		2.841 (0.127)	6.092 (0.446)			2.413 (0.184)	6.042 (0.377)	
ϱ			1.078 (0.138)				1.085 (0.136)	
κ				0.112 (0.010)				0.114 (0.010)
LogL	-1.078	-0.446	-0.278	-0.345	-1.037	-0.513	-0.243	-0.308
AIC	2.162	0.903	0.570	0.700	2.081	1.037	0.499	0.627
Disp	7.60	-1.67	0.90	0.22	-7.25	-0.95	1.47	1.21
Pear	1103.8	251.0	63.8	133.4	1065.8	319.7	82.7	142.6

Note: robust standard errors are in parentheses.

Next we examine that exogenous factors can be used as right hand side variables in describing the dynamics of Linex-transformed variables. We add “-X” to the acronym of a basic autoregressive specification to indicate the presence of exogenous variables. Empirical studies find that interest rate term spread has forecasting power for GDP growth (e.g., Ang, Piazzesi, and Wei, 2006). Using this practice, we add to the equation for linex-transformed GNP growth the variables TB (the 3-month Treasury bill rate¹¹) and YS (yield spread equaling the difference between 5-year bond yield¹² and 3-month Treasury bill rate), instead of the second lag of the dependent variable. The resulting specification is an analog of the predictive regression in Ang, Piazzesi, and Wei (2006, eqn. 10 with $k = 1$), where the authors find that both current interest and yield spread have forecasting power. In the case of T-bill returns, we add the variables FP (interest forward premium equaling the difference between 6-month¹³ and 3-month Treasury bill rates) and BY (standing for 5-year bond yield¹⁴). The resulting specification is an analog of a series of predictive regressions in Cochrane and Piazzesi (2005), but for T-bill rates rather than bond rates. For simplicity, we use the exponential distribution, and the dynamic specification corresponds to the best among the four in Table 1. All exogenous variables are dated one period earlier than the left hand side variable. The results are presented in Table 3.

¹¹ Available at <http://research.stlouisfed.org/fred2/series/TB3MS>.

¹² Available at <http://research.stlouisfed.org/fred2/series/GS5>.

¹³ Available at <http://research.stlouisfed.org/fred2/series/WTB6MS>.

¹⁴ Available at <http://research.stlouisfed.org/fred2/series/WGS5YR>.

Table 3. Results of fitting exponential ACD-X models to Linex-transformed GNP growth and T-bill returns

	GNP growth		T-bill returns	
	ACD ₁ -X, $\alpha > 0$	LACD ₁ -X, $\alpha < 0$	LACD ₂ -X, $\alpha > 0$	LACD ₁ -X, $\alpha < 0$
ω	0.959 (0.150)	-0.220 (0.059)	-0.391 (0.094)	-0.171 (0.064)
ϕ_1	0.375 (0.086)	0.394 (0.071)		0.275 (0.059)
χ_1			0.318 (0.059)	
TB	0.017 (0.014)	-0.002 (0.008)		
YS	0.034 (0.034)	-0.037 (0.018)		
FP			0.647 (0.141)	-0.476 (0.145)
BY			0.005 (0.016)	0.056 (0.015)
LogL	-1.546	-0.525	-1.075	-1.033
AIC	3.130	1.089	2.163	2.080

Note: robust standard errors are in parentheses.

One can see that in the determination of GNP growth, both exogenous factors are statistically insignificant at conventional levels when $\alpha > 0$, but the yield spread turns out to be a statistically significant (at the 5% level) determinant when $\alpha < 0$. In the case of T-bill returns, however, both factors are important predictors when $\alpha < 0$ but only the interest forward premium is when $\alpha > 0$. The fact that the structural coefficient estimates tend to have different signs depending on the sign of α is natural given that the Linex transformation $u \mapsto \exp(\alpha u)$ is increasing when α is positive and decreasing when α is negative. It also comes as no surprise that some factors that are statistically insignificant predictors when $\alpha > 0$ which implies underprediction become statistically significant when $\alpha < 0$ and there is overprediction.

Finally, we turn to forecasting. Table 4 contains values of the average Linex prediction error for s -step forecasting of T-bill returns, where $s = 1, 2$, and 4, from several series of forecasts that avoid modeling the whole conditional distribution. The rolling scheme is used, with a window of width 600; the number of out-of-sample forecasts is then $985 - s$. Estimates from each Linex estimation round are used as initial values for minimization of the Linex criterion in the next estimation round. Two of the four series of forecasts are computed using a linear AR(1) Quad-autoregression: one forecast labelled “AR” is set equal to the conditional mean, while the other, labelled “AR+bias”, is set equal to the conditional mean plus a constant bias term (namely, $\frac{1}{2}\alpha$ times the estimate of innovation’s unconditional variance) that makes the forecast Linex-optimal under conditional homoskedasticity (Granger, 1969; such a forecast is called “pseudo-optimal” in Christoffersen and Diebold, 1997). The other two of the four series of forecasts are generated from the ACD₁(0, 1) models: one is pure ACD₁-autoregression, and the other in addition includes the two exogenous factors as predictors; of course, in the latter case only one-step predictions (i.e. $s = 1$) can be generated.

Table 4. Values of average Linex prediction error for T-bill returns

Model	$\alpha > 0$			$\alpha < 0$		
	$s = 1$	$s = 2$	$s = 4$	$s = 1$	$s = 2$	$s = 4$
AR	0.781	1.022	1.051	1.136	1.505	1.441
AR+bias	0.648	0.812	0.837	0.843	1.100	1.058
ACD ₁	0.559	0.761	0.876	0.760	1.012	1.229
ACD ₁ -X	0.509			0.648		

The ACD-type models generate on average much better forecasts in terms of the Linex criterion than a mean model. The use of exogenous factors appreciably improves forecasts further, especially in the case $\alpha < 0$ when both exogenous predictors are important in-sample (see Table 3). For the longer horizons (when the strength of conditional heteroskedasticity is lower), however, the pseudo-optimal predictor may turn out to be more advantageous. It is quite possible that careful modeling of the conditional distribution would provide better forecasts than ACD models do; similarly, the CAViaR models (Engle and Manganelli, 2004) are inferior to fully parametric models in terms of forecasting ability (Kuester, Mittnik and Paoletta, 2006). However, such an investigation is outside the scope of the present paper.

5 Conclusion

We have proposed a convenient framework to model the dynamics and construct optimal forecasts when the underlying loss function is linear-exponential. Most components of this framework are borrowed from various places of existing econometric literature. The procedure hinges on the exponential quasi maximum likelihood interpretation of the Linex loss and the multiple error modeling framework. The discovered parallelism between the conventional (under Quad) and proposed (under Linex) time series analyses may potentially be extended, for example, to modeling volatility in the two frameworks. Such extensions are a subject of future research.

References

- Anatolyev, S. (2006): “Kernel estimation under linear-exponential loss,” *Economics Letters*, 91, 39–43.
- Ang, A., M. Piazzesi, and M. Wei (2006): “What does the yield curve tell us about GDP growth?,” *Journal of Econometrics*, 31, 359–403.
- Artis, M., and M. Marcellino (2001): “Fiscal forecasting: the track record of the IMF, OECD and EC,” *Econometrics Journal*, 4, S20–S36.
- Batchelor, R., and D.A. Peel (1998): “Rationality testing under asymmetric loss,” *Economics Letters*, 6, 49–54.
- 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.
- Benartzi, S., and R.H. Thaler (1995): “Myopic loss aversion and the equity premium puzzle,” *Quarterly Journal of Economics*, 110, 73–92.
- Capistrán-Carmona, C. (2005): “Bias in federal reserve inflation forecasts: is the federal reserve irrational or just cautious?” *Computing in Economics and Finance 2005*, 127, Society for Computational Economics.
- Chou, R.Y. (2005): “Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model,” *Journal of Money, Credit and Banking*, 37, 561–582.
- Christoffersen, P.F., and F.X. Diebold (1996): “Further results on forecasting and model selection under asymmetric loss,” *Journal of Applied Econometrics*, 11, 561–571.
- Christoffersen, P.F., and F.X. Diebold (1997): “Optimal prediction under asymmetric loss,” *Econometric Theory*, 13, 808–817.
- Cochrane, J.H and M. Piazzesi (2005): “Bond risk premia,” *American Economic Review*, 95, 138–160.
- Elliott, G., I. Komunjer, and A. Timmermann (2005): “Estimation and testing of forecast rationality under flexible loss,” *Review of Economic Studies*, 72, 1107–1125.
- Elliott, G., I. Komunjer, and A. Timmermann (2008): “Biases in macroeconomic forecasts: irrationality or asymmetric loss?” *Journal of the European Economic Association*, 6, 122–157.
- Engle, R.F. (2000): “The econometrics of ultra-high frequency data,” *Econometrica*, 68, 1–22.
- Engle, R.F. (2002): “New frontiers for ARCH models,” *Journal of Applied Econometrics*, 17, 425–446.
- Engle, R.F., and G.M. Gallo (2006): “A multiple indicators model for volatility using intra-daily data,” *Journal of Econometrics*, 131, 3–27.

- Engle, R.F., and S. Manganelli (2004): “CAViaR: conditional autoregressive value at risk by regression quantiles,” *Journal of Business & Economic Statistics*, 22, 367–381.
- Engle, R.F., and J.R. Russell (1998): “Autoregressive conditional duration: a new model for irregularly spaced transaction data,” *Econometrica*, 66, 1127–1162.
- Fernandes, M., and J. Grammig (2006): “A family of autoregressive conditional duration models,” *Journal of Econometrics*, 130, 1–23.
- Francq, C., and J.-M. Zakoïan (2004): “Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes,” *Bernoulli*, 10, 605–637.
- Franses, P., and D. van Dijk (2000): *Nonlinear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.
- Gourieroux, C., A. Monfort, and A. Trognon (1984): “Pseudo maximum likelihood methods: theory,” *Econometrica*, 52, 681–700.
- Grammig, J., and K.-O. Maurer (2000): “Non-monotonic hazard functions and the autoregressive conditional duration model,” *Econometrics Journal*, 3, 16–38.
- Granger, C.W.J. (1969): “Prediction with a generalized cost of error function,” *Operational Research Quarterly*, 20, 199–207, reprinted in the collected papers of C.W.J. Granger, Cambridge: Cambridge University Press.
- Hautsch, N. (2004): *Modelling Irregularly Spaced Financial Data – Theory and Practice of Dynamic Duration Models*. Berlin: Springer-Verlag.
- Hwang, S., J. Knight, and S. Satchell (2001): “Forecasting nonlinear functions of returns using LINEX loss function,” *Annals of Economics and Finance*, 2, 187–213.
- Kendall, M.G., and A. Stuart (1973): *The Advanced Theory of Statistics*. London: Griffin.
- Kuester, K., S. Mittnik, and M.S. Paolella (2006): “Value-at-risk prediction: a comparison of alternative strategies,” *Journal of Financial Econometrics*, 4, 53–89.
- Ling, S., and M. McAleer (2000): “Testing GARCH versus EGARCH,” in W.-S. Chan, W.K. Li and H. Tong, eds., *Statistics and Finance: An Interface*, London: Imperial College Press, 226–242.
- Meitz, M., and T. Teräsvirta (2006): “Evaluating models of autoregressive conditional duration,” *Journal of Business & Economic Statistics*, 24, 104–124.
- Patton, A.J., and A. Timmermann (2007): “Properties of optimal forecasts under asymmetric loss and nonlinearity,” *Journal of Econometrics*, 140, 884–918.
- Skouras, S.P. (2007): “Decisionmetrics: a decision-based approach to econometric modelling,” *Journal of Econometrics*, 137, 414–440.

Varian, H.R. (1974): “A Bayesian approach to real estate assessment,” in S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, 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.

Wooldridge, J. (1994): “Estimation and inference for dependent processes,” in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, ch.45, vol.IV, Elsevier, 2639–2738.

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