

Cost of Capital and Regulator's Preferences:
Investigation into a new method of
estimating regulatory beta

Alessio Sancetta and Steve Satchell

July 2004

CWPE 0441

Not to be quoted without permission

Cost of Capital and the Regulator's Preferences: An Investigation into a New Method of Estimating Regulatory Beta

Alessio Sancetta* and Steve Satchell**

* Department of Applied Economics, Cambridge

** Faculty of Economics and Politics, Cambridge

July 21, 2004

1 Introduction

The optimal regulatory framework of utilities seeks to resolve an inherent conflict between the interests of consumers and investors. On the one hand, to attract investment, the regulator has to assure the utility that the sunk cost of capital will be rewarded appropriately, and prices will be set to cover both short run as well as the long run costs. On the other hand, utilities are natural monopolies and as such are prone to earn excessive profits if unregulated. Thus, acting in the public interest the regulator should seek to transfer to consumers the lower production costs resulting from technological innovation. In the UK this conflict of interest is resolved via price-cap regulation. This paper assumes that the regulator has a reasonable estimate of the profit function of the company, but is faced with the problem of evaluating the cost of capital that the company faces.

The purpose of this paper is to investigate a new estimator for computing the cost of capital in regulated industries. The vast majority of calculations by regulators involve the use of the capital asset pricing model (CAPM). The beta of the CAPM is estimated dynamically, or otherwise based on least squares criteria or some time varying variation (see Buckland and Fraser, 1999, for a discussion). This methodology was thought to be quite robust. However, in practice, this has been a very unstable measure. This problem has been documented since the early 1970 (e.g. Blume, 1971, 1975, and Baesel, 1974)

We thank Dharma Holding SA for kindly providing the data used in this research. Financial support from the ESRC research award 000-23-0400 is gratefully acknowledged.

untill quite recent (e.g. Bos and Newbold, 1984, and Black et al., 1992). In particular, Buckland and Fraser (1999) document this problem in the case of regulated listed companies. Therefore, it seems reasonable to investigate alternative procedures consistent with the CAPM, but more likely to lead to more satisfactory answers than least squares methods have provided. Such a methodology can be criticized for treating equally the overvaluation of the cost of capital with the undervaluation. As discussed in Newey and Powell (1987), interpretation of the estimated coefficients is dependent on the method used as soon as we depart from the classical case where the dependent variable is a linear function of explanatory variables plus iid noise

Suppose, S_t is the price of the stock listed in the market. Let P_t be the price that the company charges to the consumers for its services. Suppose that $P_t \mapsto \pi_t$, where π_t is the profit at time t . Then, in terms of long-run equilibrium

$$S_t = \mathbb{E}_t \left(\sum_{k=1}^K \frac{\pi_{t+k} (P_{t+k})}{(1 + \tilde{r})^k} \right),$$

where \tilde{r} is the appropriate discount rate based on the cost of capital, and the right hand side can be thought of as the fair price of the stock. If the company distribute its earnings in the form of dividends, then π_t has to be interpreted as dividends payments (or expected dividends payments). We are concerned with the estimation of \tilde{r} when the regulator decides that the capital asset pricing model holds. In this case, let r_f be the interest free rate over one period, and S_t^m be the market price at time t . Then,

$$\tilde{r} = \mathbb{E}_t \left(\frac{S_t}{S_{t-1}} - 1 \right) - r_f = Y_t = \beta \mu_m \quad (1)$$

$$\mu_m := \mathbb{E}_t \left(\frac{S_t^m}{S_{t-1}^m} - 1 \right) - r_f.$$

It follows that the regulator should consider the estimation of β in such a way that is consistent with its mission: to protect the consumers whilst giving an incentive for future investment.

We propose a method to do this, namely estimation via Linex loss functions. Linex loss functions are a class of asymmetric loss functions which appear to have been used by Varian (1975) in the first place in the context of real estate. It has also been applied in other areas, such as to Bayesian estimation problems in statistics (Zellner, 1986), and to construct optimal forecasts (e.g. Hwang et al., 1997, Elliott and Timmermann, 2003, and references therein). While Linex has been used as a loss function for constructing forecasts, it has been rarely used for in-sample parameter estimation. When the errors in the classical linear regression context are asymmetric or do not satisfy some of the OLS assumptions, asymmetric estimation has been advocated by several authors (e.g. Newey and Powell, 1987, and references therein). In these

cases, the interpretation of the estimated coefficients rely heavily on the loss function used. This is particularly so in the case of misspecification when the optimal misspecified model (within the class of chosen models) is the closest to the true (usually unknown) model relatively to the chosen loss function. In this case, preferences must determine the characteristics of the loss function. Linex allows for this flexible approach, as it depends on one parameter that has direct interpretation in terms of risk aversion.

The plan for the paper is as follows. In Section 2 we discuss the population properties of Linex estimation. We restrict attention to Linex best linear prediction. These tools allows us to define the population value of beta, which we shall describe how to estimate in the remainder of the paper. In Section 3, we consider the sample properties together with some further discussion and a small simulation to compare the Linex estimator to the least square. In Section 4 we apply our methodology to both UK water and electricity companies. We find substantial differences in the cost of capital when computed by our method versus the traditional one. Conclusions follow in Section 5. An appendix provides some information on the companies used in our empirical study.

2 Asymmetric Prediction and Estimation

Suppose Y is a real random variable. Let h be our predictor of Y and write $e = Y - h$ for the prediction error. Suppose Θ is a compact parameter space, and X is a random variable in S . With increasing level of generality, we can restrict h to be an element into the following classes of functions: 1. the constant functions

$$\mathcal{M}_1 := \{h : h \in \mathbb{R}\},$$

2. linear maps of X (clearly, X may include a constant term)

$$\mathcal{M}_2(S, X, \Theta) := \left\{h : h = \beta^T X, \beta \in \Theta, X \in S\right\} \quad (2)$$

3. arbitrary functions of X indexed in Θ

$$\mathcal{M}_3(S, X, \Theta) := \{h : h = \mu_\theta(X), \mu_\theta : S \times \Theta \rightarrow \mathbb{R}\}.$$

For each of these categories, the predictor $h \in \mathcal{M}_i$ ($i = 1, 2, 3$) will depend on the loss function we wish to minimize, i.e. the optimal predictor is the predictor which minimizes the expectation of the loss function $L(e)$ with respect to h , $L : \mathbb{R} \rightarrow \mathbb{R}_+$

$$\min_{h \in \mathcal{M}_i} \mathbb{E}L(Y - h),$$

for a fixed $i = 1, 2, 3$.

2.1 Linear Predictors and Projections

As above h is our predictor of Y and write $e = Y - h$ for the prediction error. We restrict $h \in \mathcal{M}_2(S, \Theta)$. In this case, if $\mathbb{E}e^2 < \infty$, it is natural to consider the best predictor in the space of square integrable linear functions. Therefore, it is natural to equip this space with an inner product, so that we have an Hilbert space, say \mathcal{H} . In this space, the best predictor $h \in \mathcal{M}_2(S, \Theta)$ for Y is equivalent to finding the projection of Y onto \mathcal{H} , where $Y \in \mathcal{H}$ and $h \in \mathcal{M}_2(S, \Theta) \cap \mathcal{H}$. This projection is given by $X^T \tilde{\beta}$, where

$$\tilde{\beta} := \arg \min_{\beta \in \Theta} \mathbb{E} \left(Y - \theta^T X \right)^2$$

is the solution of the mean square error problem, i.e. the closest element in $\mathcal{M}_2(S, X, \Theta) \cap \mathcal{H}$ to Y under the L_2 norm $\|\dots\|_2$.

While this is a natural approach, we do not need to restrict our attention to $\mathcal{M}_2(S, \Theta) \cap \mathcal{H}$ (i.e. we do not need to equip our space with $\|\dots\|_2$) hence $X^T \tilde{\beta}$ does not need to be the best choice for our purposes: the class of best predictions in $\mathcal{M}_2(S, X, \Theta)$ is much larger than the one in $\mathcal{M}_2(S, X, \Theta) \cap \mathcal{H}$.

Our motivation for considering the prediction problem in a more general setup stems from the fact that we are restricted to $\mathcal{M}_2(S, X, \Theta)$, but this is just a convenient parametric approximation. In this case, choosing the best predictor under $\|\dots\|_2$ is adequate only if we have symmetric beliefs. The following is an example to illustrate the point.

Example 1 *Suppose $Y = a + Zb + \varepsilon$, where Z is a conditioning variable and ε is a random variables such that $\mathbb{E}\varepsilon = 0$, $\mathbb{E}\varepsilon^2 < \infty$. Assume we do not know the true generating process for Y , and we choose a predictor $h \in \mathcal{M}_2(S, X = Z, \Theta \subset \mathbb{R})$ i.e. we choose $h = Z\beta$. The second step it to decide the metric under which the predictor is optimal. Suppose we choose the L_2 norm:*

$$\tilde{\beta} = (\mathbb{E}Z)^{-2} \mathbb{E}ZY.$$

Substituting for Y ,

$$\mathbb{E}ZY = a\mathbb{E}Z + \mathbb{E}Z^2b + \mathbb{E}Z\varepsilon,$$

so that

$$\tilde{\beta} = b + \mathbb{E}Za.$$

If we knew that $Z > 0$ and $a > 0$, our predictor would be biased upwards. In this simple case, the problem would be solved by choosing

$$h \in \mathcal{M}_2(S, X = (1, Z), \Theta \subset \mathbb{R}^2).$$

However, in other cases, we will have to stick to a specific class irrespective either of previous knowledge or preference.

2.2 Linex Loss Function

The previous example shows that if we are forced to restrict attention to a specific class of predictors, than choosing a quadratic risk function may not be the best choice. For example, in the previous case, we may decide to penalize more when the expectation of the predictor error is negative instead of being positive. For this reason, we propose a different approach via Linex loss functions. As mentioned in the Introduction, Linex is not commonly used for in-sample parameter estimation.

When L is a Linex loss function, we have

$$L(e) = \exp\{\alpha e\} - \alpha e - 1. \quad (3)$$

For small $|\alpha| < 1$ the function is nearly symmetric (just take a Taylor expansion), while asymmetry becomes more pronounced as $|\alpha|$ increases (Zellner, 1986, for further details). In particular, for $\alpha > 0$, we penalize more (i.e. $L(e)$ is larger) when e is positive. In the context of the regulation problem, we would punish a value of beta that would lead to an under fitting of the company returns in excess of what the CAPM might predict. Thus beta is being chosen such that positive "alpha" in the company is being penalized. When $h \in \mathcal{M}_i$ ($i = 1, 2$) the Linex loss function always has a unique minimum and this follows by convexity (e.g. Zellner, 1986).

Example 2 Consider $h \in \mathcal{M}_1$ (recall \mathcal{M}_1 is the class of constant functions). The optimal h , say h^* , will depend on the loss function $L : \mathbb{R} \rightarrow \mathbb{R}$. If L is Linex, we have that

$$\begin{aligned} h^* &= \frac{1}{\alpha} \ln \mathbb{E} \exp\{\alpha Y\} \\ &= \mathbb{E}Y + \alpha \mathbb{E}(Y - \mathbb{E}Y)^2 + O(\alpha^2), \end{aligned} \quad (4)$$

using the fact that $\ln \mathbb{E} \exp\{\alpha Y\}$ is the cumulative generating function of Y (e.g. McCullagh, 1987, for details). In our case, it is of interest to look at $\alpha > 0$, as we want to penalize gains more than losses. For example, under normality, this implies that for $\alpha > 0$, the Linex estimator is larger than the mean. On the other hand, if $L(e) := e^2$ (which is least squares, a popular symmetric loss function), then $h^* = \mathbb{E}Y$ is the mean. Of course our problem differs from this in that it is the beta that we wish to estimate.

2.2.1 Linex Solution via Moment Condition

The Linex solution to the problem $\mathbb{E}L(e)$ for L as in (3), where we confine attention to $h \in \mathcal{M}_2(S, X, \Theta)$ as in (2), can be characterized via moment restriction. We include an intercept term, and for convenience we write $h = \beta_0 + X^T \beta_1$. Clearly, Y does not need to have linear conditional expectation

function with respect to X . For example $Y = f(X, \eta)$, for some measurable function f and some stochastic term η .

$$\mathbb{E}L(\alpha(Y - \beta_0 - X^T \beta_1)) = \mathbb{E} \exp(\alpha(Y - \beta_0 - X^T \beta_1)) - \alpha \mathbb{E}(Y - \beta_0 - X^T \beta_1) - 1,$$

where expectation is taken over the joint distribution of (X, Y) which can be conditional, steady state et cetera. Under regularity conditions, setting $\beta = (\beta_0, \beta_1)$,

$$\frac{\partial \mathbb{E}L(\alpha(Y - \beta_0 - X^T \beta_1))}{\partial \beta} = \frac{\mathbb{E} \partial L(\alpha(Y - \beta_0 - X^T \beta_1))}{\partial \beta}.$$

Hence

$$\frac{\mathbb{E} \partial L(\alpha(Y - \beta_0 - X^T \beta_1))}{\partial \beta_1} = -\mathbb{E} X \exp(\alpha(Y - \beta_0 - X^T \beta_1)) + \mu_X = 0$$

$$\frac{\partial \mathbb{E}L(\alpha(Y - \beta_0 - X^T \beta_1))}{\partial \beta_0} = -\mathbb{E} \exp(\alpha(Y - \beta_0 - X^T \beta_1)) + 1 = 0,$$

so that, under Linex, the first order conditions can be written as

$$\begin{aligned} \text{cov}(X, \exp(\alpha e)) &= \mathbb{E} X \exp(\alpha e) - \mathbb{E} X \mathbb{E} \exp(\alpha e) \\ &= \mathbb{E} X \exp(\alpha e) - \mu_X = 0. \end{aligned} \quad (5)$$

We see that (5) can be seen as the analogue to the conditions in OLS, where the regressors are orthogonal to the error term.

Computing the Optimal Beta Value using Linex. We intend to use Linex to compute the population value of β in (1). The first thing to notice is that it is a function of α in (3), that is the loss function of the regulator determines the "true" value of the discount factor. Define $\varphi_{Y,X}(s, t) := \mathbb{E} \exp\{sY + tX\}$ to be the moment generating function of (X, Y) , and

$$\varphi_{Y,X}^{i,j}(s, t) := \frac{\partial^{i+j} \varphi_{Y,X}(u, v)}{\partial u^i \partial v^j} \Big|_{u=s, v=t}.$$

Notice that

$$\mathbb{E} \exp(\alpha e) = \varphi_{Y,X}(\alpha, -\alpha \beta_1) \exp\{-\alpha \beta_0\}.$$

Then, from (5), β is indirectly defined as the solution of the system of equations

$$\begin{aligned} \varphi_{Y,X}^{0,1}(\alpha, -\alpha \beta_1) &= \mu_X \exp\{\alpha \beta_0\}, \\ \varphi_{Y,X}(\alpha, -\alpha \beta_1) &= \exp\{\alpha \beta_0\}, \end{aligned} \quad (6)$$

and β_1 is the solution of

$$\frac{\varphi_{Y,X}^{0,1}(\alpha, -\alpha\beta_1)}{\varphi_{Y,X}(\alpha, -\alpha\beta_1)} = \mu_X, \quad (7)$$

while β_0 is proportional to the cumulant generating function of (Y, X) at $(\alpha, -\alpha\beta_1)$, i.e.

$$\beta_0 = \frac{1}{\alpha} \ln \varphi_{Y,X}(\alpha, -\alpha\beta_1).$$

This shows that under linex, β_0 is usually different from $\mu_Y - \beta_1\mu_X$. Therefore, the restriction on the intercept used to test for the CAPM is not meaningful in the general case.

In the case of bivariate normality, we have the following.

Proposition 1 *Suppose (X, Y) is jointly normal with $\sigma_X^2 = \text{var}(X)$, $\sigma_{XY} = \text{cov}(X, Y)$. Then, the linex optimal linear predictor $h \in \mathcal{M}_2(S, (1, X), \Theta \in \mathbb{R}^2)$ is given by the linear projection*

$$\tilde{\beta}_1 = \frac{\sigma_{XY}}{\sigma_{XX}},$$

and

$$\tilde{\beta}_0 = \mu_Y - \beta_1\mu_X + \frac{\alpha}{2} \left(\sigma_{YY} - \frac{\sigma_{XY}^2}{\sigma_{XX}} \right).$$

Proof. The moment generating function of Gaussian vector (X, Y) is given by

$$\varphi_{Y,X}(s, t) = \exp \left\{ s\mu_Y + t\mu_X + \frac{s^2\sigma_{YY} + 2st\sigma_{XY} + t^2\sigma_{XX}}{2} \right\}.$$

Then, substitute in (7) with $s = \alpha$, $t = -\alpha\beta_1$, and solve. ■

The proposition shows that under joint normality, not only $\tilde{\beta}_1$ does not depend on α , but it is also equal to the linear projection onto the space spanned by the explanatory variable. On the other hand, β_0 does depend on α : the larger α , the lower β_0 . Furthermore, in this special case, the validity of the CAPM ($\mu_Y = \beta\mu_X$) implies that $\tilde{\beta}_0 = \frac{\alpha}{2} \left(\sigma_{YY} - \frac{\sigma_{XY}^2}{\sigma_{XX}} \right)$.

The result that β_1 is independent of α may hold in some other cases, when X is Gaussian and Y conditionally Gaussian.

Example 3 *Suppose Y is a measurable random variable with the following stochastic representation in terms of a random variable $X \sim N(\mu_X, \sigma_{XX})$*

$$Y = \gamma + \delta X + \theta X^2 + \eta,$$

where $\eta \sim N(0, \sigma_\eta^2)$. Then,

$$\mu_Y = \gamma + \delta\mu_X + \theta(\sigma_{XX} + \mu_X^2).$$

This model has been proposed by Treynor and Mazuy (1996) to test for market timing (however, we have imposed some further distributional assumptions). Under the linear forecast model

$$Y = \beta_0 + \beta_1 X + e,$$

use Linex to solve the following

$$\min_{\beta} \mathbb{E}L(\alpha(Y - \beta_0 - \beta_1 X)).$$

Now,

$$\begin{aligned} & \mathbb{E}L(\alpha(\gamma + \delta X + \theta X^2 + \eta - \beta_0 - \beta_1 X)) \\ = & \mathbb{E} \exp\{\alpha(\gamma + \delta X + \theta X^2 + \eta - \beta_0 - \beta_1 X)\} \\ & - \alpha \mathbb{E}(\gamma + \delta X + \theta X^2 - \eta - \beta_0 - \beta_1 X) - 1 \\ = & \exp\{\alpha(\gamma - \beta_0 + \alpha\sigma_{\eta}^2/2)\} \mathbb{E} \exp\{\alpha(\theta X^2 + (\delta - \beta_1)X)\} \\ & - \alpha(\gamma - \beta_0 + \theta(\sigma_{XX} + \mu_X^2) + (\delta - \beta_1)\mu_X) - 1 \\ = & A(\alpha, \beta_1) \exp\{-\alpha\beta_0\} + B(\alpha, \beta_1) - \alpha\beta_0. \end{aligned}$$

By Lemma 1 in the appendix, with $s = \alpha(\delta - \beta_1)$ and $t = \alpha\theta$,

$$\mathbb{E} \exp\{\alpha(\delta - \beta_1)X + \alpha\theta X^2\} = \exp\left\{-\frac{\mu_X^2 - \psi(\beta_1)^2}{2\sigma_{XX}}\right\} (1 - 2\alpha\theta\sigma_{XX})^{-1/2},$$

where

$$\psi(\beta_1) = (\mu_X + \alpha(\delta - \beta_1)\sigma_{XX})(1 - 2\alpha\theta\sigma_{XX})^{-1/2}.$$

Differentiating with respect to β_1

$$A(\alpha, \beta_1) \exp\{-\alpha\beta_0\} \alpha \left(\frac{\mu_X + \alpha(\delta - \beta_1)\sigma_{XX}}{(2\alpha\theta\sigma_{XX} - 1)} \right) + \alpha\mu_X = 0,$$

and with respect to β_0 ,

$$A(\alpha, \beta_1) \exp\{-\alpha\beta_0\} = 1.$$

The solution is given by

$$\begin{aligned} \alpha(\delta - \beta_1)\sigma_{XX} &= -(2\alpha\theta\sigma_{XX} - 1)\mu_X - \mu_X \\ &= -2\alpha\theta\sigma_{XX}\mu_X \end{aligned}$$

$$\beta_1 = \delta + 2\theta\mu_X,$$

for $\theta < 1/(2\alpha\sigma_{XX})$, and

$$\beta_0 = \gamma - \mu_X^2 + \alpha \frac{\sigma_{\eta}^2}{2} - \frac{1}{2\alpha} \ln(1 - 2\alpha\theta\sigma_{XX}).$$

In this case, β_1 is independent of α , while β_0 does depend on α .

However, this is not always the case.

Example 4 Suppose Y is as in the previous example, but the error term is now heteroscedastic, $\eta \sim N(0, \sigma_\eta^2 X^2)$. We can solve the same problem. Using conditional expectations

$$\begin{aligned}
& \mathbb{E}L\alpha(\gamma + \delta X + \theta X^2 + \eta - \beta_0 - \beta_1 X) \\
&= \mathbb{E} \exp \{ \alpha(\gamma + \delta X + \theta X^2 + \eta - \beta_0 - \beta_1 X) \} \\
&\quad - \alpha \mathbb{E}(\gamma + \delta X + \theta X^2 - \eta - \beta_0 - \beta_1 X) - 1 \\
&= \mathbb{E} \exp \{ \alpha(\gamma - \beta_0 + \alpha \sigma_\eta^2 X^2 / 2) \} \mathbb{E} \exp \{ \alpha(\theta X^2 + (\delta - \beta_1) X) \} \\
&\quad - \alpha(\gamma - \beta_0 + \theta(\sigma_{XX} + \mu_X^2) + (\delta - \beta_1)\mu_X) - 1 \\
&= \exp \{ \alpha(\gamma - \beta_0) \} \mathbb{E} \exp \{ \alpha((\theta + \alpha \sigma_\eta^2 / 2) X^2 + (\delta - \beta_1) X) \} \\
&\quad - \alpha(\gamma - \beta_0 + \theta(\sigma_{XX} + \mu_X^2) + (\delta - \beta_1)\mu_X) - 1.
\end{aligned}$$

Then calculations similar to the previous example, give

$$\beta_1 = \delta + 2(\theta + \alpha \sigma_\eta^2 / 2) \mu_X,$$

for $(\theta + \alpha \sigma_\eta^2 / 2) < 1 / (2\alpha \sigma_{XX})$. Here β_1 directly depends on α : the larger α , the larger β_1 .

2.2.2 Expected Utility Interpretation of Linex

Linex estimation is given by

$$\min_{h \in \mathcal{M}_i} \mathbb{E}L(Y - h) := \min_{h \in \mathcal{M}_i} \mathbb{E} \exp \{ \alpha(Y - h) \} - \alpha(Y - h) - 1,$$

such that $\mathbb{E}(Y - h) = 0$. This problem is equivalent to

$$\max_{h \in \mathcal{M}_i} \mathbb{E}[-L(Y - h)] = \max_{h \in \mathcal{M}_i} \mathbb{E}[\alpha(Y - h) - \exp \{ \alpha(Y - h) \} + 1].$$

Then we have the natural interpretation of the Linex optimization in terms of expected utility optimization of an agent with linear plus Bernoulli utility function. A similar interpretation has been provided by Knight et al. (2003). The only difference is that we are considering a mapping from $e \in \mathbb{R}$, whereas in a standard utility context the argument of objective function (i.e. wealth) is a non-negative random variable. If we are to retain this interpretation, then $\alpha(Y - h) \rightarrow \infty$ implies a higher utility and $\alpha(Y - h) \rightarrow -\infty$ lower utility. For this reason $sign(\alpha)$ will have to be chosen consistently with this. Considering our context, we see that $|\alpha|$ is chosen so to reflect the level of risk aversion of the regulator, which both in theory and good practice leads to a coefficient of risk aversion independent of wealth (Bell and Fishburn, 2001). In fact, the linear plus Bernoulli utility function is one of the only two Bernoulli utility functions that are consistent with the notion that gambles may be ordered by riskiness and that this rank is independent of the decision maker's wealth (Bell and Fishburn, 2001).

3 In Sample Properties of Linex Estimators

3.1 Generalized Linear Estimation

Suppose $(Y_i)_{i \in \mathbb{Z}}$ is a sequence of stationary random variables with values in the probability space $(\mathbb{R}, \mathcal{F}, \mathbb{P})$. Suppose $(X_i)_{i \in \mathbb{Z}}$ is a sequence of \mathbb{R}^K ($K \geq 1$) stationary random variables. Consider $h \in \mathcal{M}_2$, i.e.

$$h = \beta^T X, \tag{8}$$

for $\beta \in \Theta$ (Θ is some compact separable space, e.g. $\subset \mathbb{R}^K$), where for simplicity β may contain the intercept. In the previous section we phrased the problem in terms of the true measure \mathbb{P} . However this is not always known. In this case, it needs to be replaced by the empirical measure. Suppose we want to choose the best linear forecast with respect to the Linex loss function using the empirical measure. For any sample $\{x_1, \dots, x_n\}$ from X , define $\mathbb{P}_n := n^{-1} \sum_i^n \delta_{x_i}$, where δ_{x_i} is the Dirac measure (i.e. \mathbb{P}_n is the empirical measure which assign mass n^{-1} at each observation). Setting $e_\beta := Y - \beta^T X$, we define the following M-estimator (which we may call the empirical linex estimator),

$$\min_{h \in \mathcal{M}_2} \mathbb{P}_n L(Y - h) = \min_{\beta \in \Theta} \mathbb{P}_n L(Y - \beta^T X) = \min_{\beta \in \Theta} \mathbb{P}_n (\exp \{ \alpha e_\beta \} - \alpha e_\beta - 1),$$

and

$$\hat{\beta}(n) := \arg \min_{\beta \in \Theta} \mathbb{P}_n L(Y - \beta^T X)$$

is its solution. Under suitable conditions, $\mathbb{P}_n L(Y - \beta^T X)$ converges to a non-random limit. In this case, $\mathbb{P} L(Y - \beta^T X) = \mathbb{E} L(Y - \beta^T X)$ is the "asymptotic" version of $\mathbb{P}_n L(Y - \beta^T X)$, i.e. its limit with respect to n . This limit is the population value considered in terms of X and Y . Then,

$$\tilde{\beta} := \arg \min_{\beta \in \Theta} \mathbb{E} L(Y - \beta^T X),$$

and

$$L_{\tilde{\beta}} := \left[\frac{\partial L(Y - \beta^T X)}{\partial \beta^T} \right]_{\beta = \tilde{\beta}}, \tag{9}$$

$$L_{\tilde{\beta}\tilde{\beta}} := \left[\frac{\partial^2 L(Y - \beta^T X)}{\partial \beta \partial \beta^T} \right]_{\beta = \tilde{\beta}}.$$

Under suitable conditions on e_β , $\hat{\beta}(n) \rightarrow \tilde{\beta}$, and

$$\sqrt{n} \left(\hat{\beta}(n) - \tilde{\beta} \right) \rightsquigarrow \mathcal{N} \left(0, \left(\mathbb{E} L_{\tilde{\beta}\tilde{\beta}} \right)^{-1} \mathbb{E} L_{\tilde{\beta}} L_{\tilde{\beta}}^T \left(\mathbb{E} L_{\tilde{\beta}\tilde{\beta}} \right)^{-1} \right),$$

where \rightsquigarrow stands for weak convergence and \mathcal{N} for the Gaussian distribution. In particular,

$$L_{\tilde{\beta}} = \alpha X \left(1 - \exp \left\{ \alpha e_{\tilde{\beta}} \right\} \right),$$

and

$$L_{\tilde{\beta}\tilde{\beta}} = \alpha^2 X X^T \exp \left\{ \alpha e_{\tilde{\beta}} \right\},$$

so that

$$\begin{aligned} & \sqrt{n} \text{cov} \left(\hat{\beta}(n) - \tilde{\beta} \right) \\ = & \alpha^{-2} \left[\mathbb{E} X X^T \exp \left\{ \alpha e_{\tilde{\beta}} \right\} \right]^{-1} \mathbb{E} \left[X X^T \left(1 - \exp \left\{ \alpha e_{\tilde{\beta}} \right\} \right)^2 \right] \left[\mathbb{E} X X^T \exp \left\{ \alpha e_{\tilde{\beta}} \right\} \right]^{-1} \end{aligned}$$

The distribution of the Least Squares Estimator is exactly Gaussian when the regressors are non-stochastic and the errors are normal. Suppose that x_1, \dots, x_n , $x_i \in \mathbb{R}^K$ are non-stochastic and we want to use $h \in \mathcal{M}_2$ as best linear predictor for Y , i.e. we want to estimate

$$Y = \beta^T x + \varepsilon$$

using Linex.

Proposition 2 *Under regularity conditions, $(\mathbb{P}_n - \mathbb{P}) L_{\tilde{\beta}} = \mathbb{P}_n L_{\tilde{\beta}} \rightarrow 0$, and*

$$\Pr \left(\hat{\beta}(n) \leq \tilde{\beta} \right) = \Pr \left(\sqrt{n} \mathbb{P}_n L_{\hat{\beta}(n)} \leq \mathbb{P} L_{\tilde{\beta}} \right) = \Pr \left(\sqrt{n} \mathbb{P}_n L_{\hat{\beta}(n)} \leq 0 \right),$$

where, under stationarity,

$$\sqrt{n} \mathbb{P}_n L_{\beta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \alpha x_i^T - \frac{1}{\sqrt{n}} \sum_{i=1}^n \alpha x_i^T \exp \left\{ -\alpha \beta^T x_i \right\} \exp \left\{ \alpha Y_i \right\}$$

has distribution given by the weighted sum of $\exp \left\{ \alpha Y \right\}$.

Example 5 *Suppose, $Y_i \sim \mathcal{N}(\theta x_i, \sigma_Y)$. Then, $Z_i = \exp \left\{ \alpha Y_i \right\}$ is lognormally distributed with $\mathbb{E} Z_i = \exp \left\{ \alpha \theta x_i + \alpha^2 \sigma_Y^2 / 2 \right\}$. Therefore, $\hat{\beta}(n)$ solves*

$$\frac{1}{n} \sum_{i=1}^n x_i \exp \left\{ \frac{\alpha^2 \sigma_Y^2}{2} \right\} = \frac{1}{n} \sum_{i=1}^n x_i \exp \left\{ \alpha \left(\theta - \tilde{\beta} \right) x_i \right\},$$

and $\Pr \left(\hat{\beta}(n) \leq \tilde{\beta} \right)$ is equal to the centered weighted sum of Lognormal random variables. This is a function of the x_i 's, θ , σ_Y^2 , and α . As discussed in the previous section, it is not necessarily true that $\tilde{\beta} = \theta$.

3.2 Linex in a Time Series One Step Ahead Prediction Environment

Under suitable modification, consistency of the estimator is guaranteed in a time series context under possible misspecification. We apply the approach presented in Skouras and Dawid (2000) to our Linex forecasting problem. Suppose $(Y_i)_{i \in \mathbb{Z}}$ is a sequence of possibly non-stationary random variables with values in the probability space $(\mathbb{R}, \mathcal{F}, \mathbb{P})$. We define $(\mathcal{F}_t)_{t \in \mathbb{Z}_+}$ (with \mathcal{F}_0 trivial, if variables are not defined on a product probability space) to be a filtration of \mathcal{F} . Let $(h_t)_{t \in \mathbb{Z}_+}$ be an \mathcal{F}_{t-1} -measurable sequence of forecasts for $(Y_t)_{t \in \mathbb{Z}_+}$ with values in $(\mathbb{R}, \mathcal{F}, \mathbb{P})$. Notice that $(h_t)_{t \in \mathbb{Z}_+}$ is a non-misspecified forecast if $(Y_t - h_t)_{t \in \mathbb{Z}_+}$ is a martingale difference. Clearly, we cannot restrict ourselves to non-misspecified forecasts.

As before, we shall restrict h_t to belong to the class of linear forecasts, i.e.

$$h_t = \beta^T X_{t-1},$$

for $\beta \in \Theta$ (Θ is some compact separable space, e.g. $\subset \mathbb{R}^k$) where X_t ($k \times 1$) is measurable with respect to \mathcal{F}_t (i.e. it is known at time t). Therefore, our \mathcal{F}_{t-1} conditional forecast for time t is given by

$$\mathbb{E}(h_t | \mathcal{F}_{t-1}) = \beta^T X_{t-1},$$

which is possibly different from

$$\mathbb{E}(Y_t | \mathcal{F}_{t-1}),$$

due to misspecification. Suppose we want to choose the best linear forecast with respect to the loss function L . Setting $e_{\beta|t} := Y_t - \beta^T X_{t-1}$, we define the following problem

$$\min_{\beta \in \Theta} L_T(\beta) := \min_{\beta \in \Theta} \frac{1}{T} \sum_{t=1}^T L(Y_t - \beta^T X_{t-1}) = \min_{\beta \in \Theta} \frac{1}{T} \sum_{t=1}^T (\exp\{\alpha e_{\beta|t}\} - \alpha e_{\beta|t} - 1),$$

and consider

$$L_T^*(\beta) := \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left[L(Y_t - \beta^T X_{t-1}) | \mathcal{F}_{t-1} \right]$$

to be the natural asymptotic version of $L_T(\beta)$. Therefore, we define

$$\tilde{\beta}^* := \min_{\beta \in \Theta} L_T^*(\beta),$$

where in many cases, but not all (e.g. see Skouras and Dawid, 2000), $\tilde{\beta}^* = \tilde{\beta}$, as defined previously. We could choose a previsible sequence $(A_T)_{T \geq 1}$ as alternative denominator T (e.g. Skouras and Dawid). Clearly, $L_T(\beta) - L_T^*(\beta) \xrightarrow{a.s.}$

0 by the reverse Martingale convergence theorem whenever $L_T(\beta) - L_T^*(\beta)$ is uniformly integrable (Rogers and Williams, 2000). In the case of LinEx, it can be shown that under suitable conditions on e_β , $\hat{\beta}(n) \rightarrow \tilde{\beta}$, and (e.g. Hall and Heyde, 1980)

$$\sqrt{n} \left(\hat{\beta}(n) - \tilde{\beta} \right) \rightsquigarrow \mathcal{N} \left(0, \Sigma \left(\tilde{\beta} \right) \right),$$

where

$$\Sigma \left(\tilde{\beta} \right) := \left[\frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(L_{\tilde{\beta}\tilde{\beta}} | \mathcal{F}_{t-1} \right) \right]^{-1} \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(L_{\tilde{\beta}} L_{\tilde{\beta}}^T | \mathcal{F}_{t-1} \right) \left[\frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(L_{\tilde{\beta}\tilde{\beta}} | \mathcal{F}_{t-1} \right) \right]^{-1}.$$

3.3 Efficiency of the Estimator: A Short Simulation Study

As shown in the previous section, the linear coefficient under Linex estimation is asymptotically equivalent to the least square one. However, it would be interesting to know, at least under simple conditions, something about the efficiency of the estimator itself. To this end, we consider the following small Monte Carlo exercise. We generate B samples of size n from the same bivariate distribution. Hence, we estimate the following

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t.$$

For comparison reasons we also estimate the restricted version ($\beta_0 = 0$) under OLS. In particular, we choose $B = 1000$ and $n = 40, 120, \alpha = -.8, -.2, .2, .8$. For $(X_t, Y_t)_{t \in \mathbb{N}}$, we generate samples from the following bivariate distributions:

1. a Gaussian, with mean and covariance matrix,

$$\mu^{(1)} = (2, (2/4) 2); \text{vec} \left(\Sigma^{(1)} \right) = (4, 2, 2.5),$$

respectively;

2. a Gaussian with

$$\mu^{(2)} = (2, (4.8/4) 2); \text{vec} \left(\Sigma^{(2)} \right) = (4, 4.8, 12);$$

3. a skewed non-elliptic distribution with fat tails given by the Kimeldorf-Sampson copula (e.g. Joe, 1997, for details) with shifted double exponential marginals:

$$C(u, v) = \left(u^{-\delta} + v^{-\delta} - 1 \right)^{\frac{1}{\delta}},$$

$$\begin{aligned} u & : = F_X(x) = (1 - q_x) \exp \left\{ \lambda_x^- (x - \mu_x) \right\} I_{\{x < \mu_x\}} \\ & + (1 - q_x \exp \left\{ \lambda_x^+ (x - \mu_x) \right\}) I_{\{x \geq \mu_x\}}, \end{aligned}$$

$$v \quad : \quad = F_Y(y) = (1 - q_y) \exp \{ \lambda_y^- (y - \mu_y) \} I_{\{y < \mu_y\}} \\ + (1 - q_y \exp \{ \lambda_y^+ (y - \mu_y) \}) I_{\{y \geq \mu_y\}},$$

where the parameters (rounded to nearest third digit)

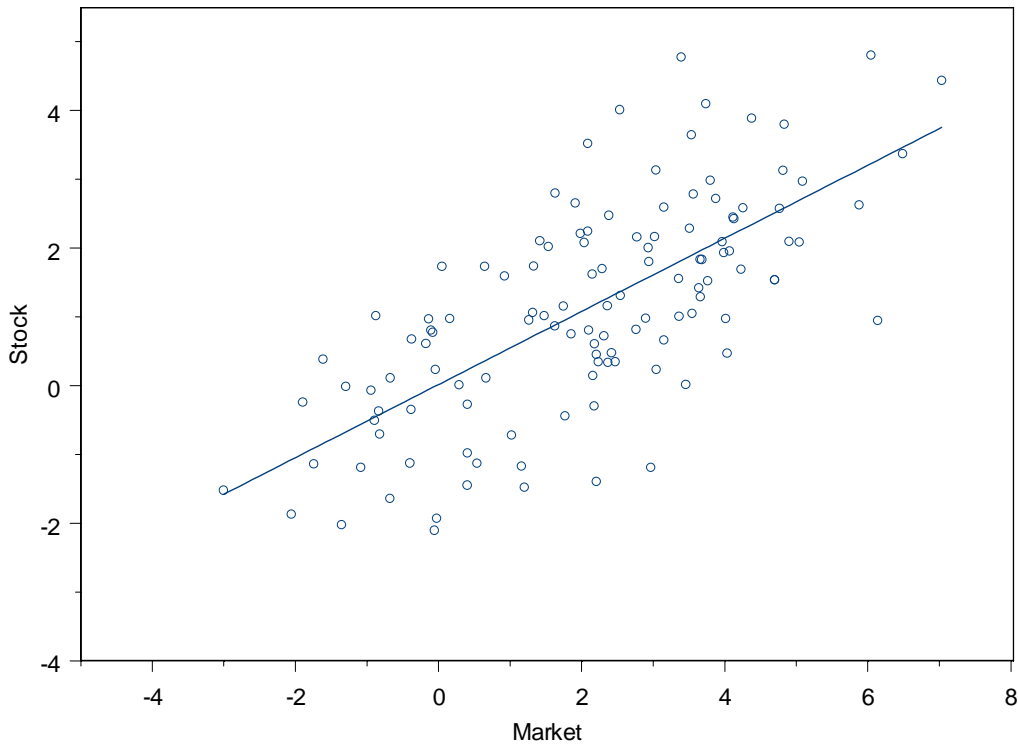
$$\begin{pmatrix} \lambda_x^- \\ \lambda_x^+ \\ \mu_x \\ q_x \\ \lambda_y^- \\ \lambda_y^+ \\ \mu_y \\ q_y \\ \delta \end{pmatrix} = \begin{pmatrix} 0.6 \\ 0.8 \\ 3 \\ 0.229 \\ 0.3 \\ 0.4 \\ 5.549 \\ 0.032 \\ 1.5 \end{pmatrix},$$

are chosen to match the mean and the covariance matrix of distribution 2.

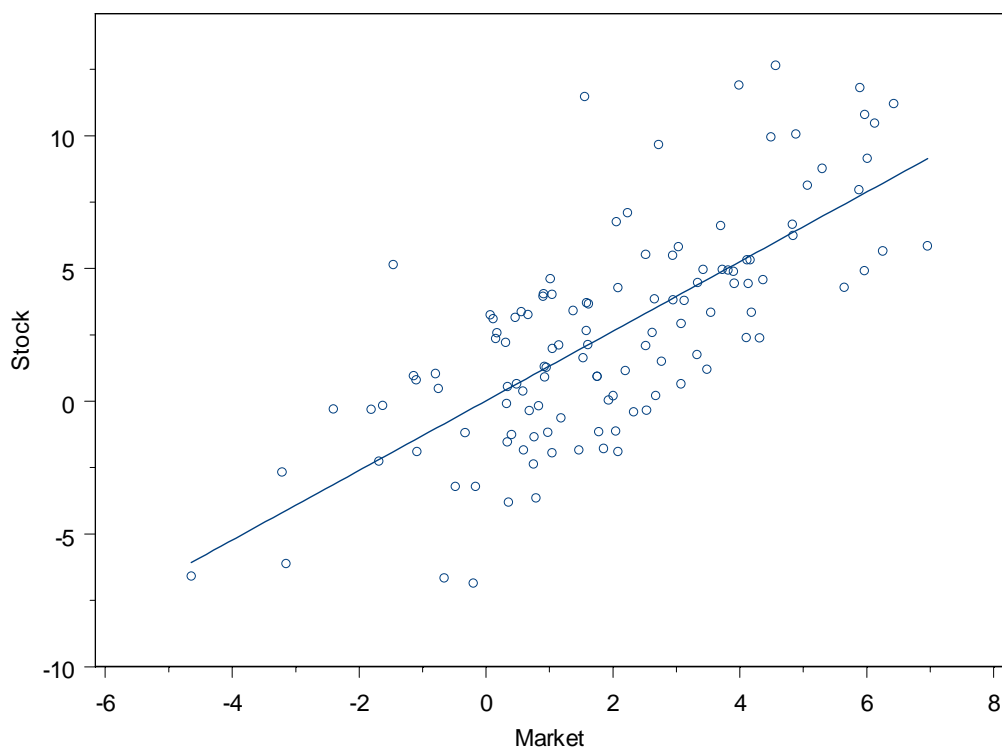
Notice that for the above distributions $\mu_Y = \beta_1 \mu_X$. The motivation for the first two models is that they are consistent with the CAPM: the first gives a low beta, the second a high beta. For the third distribution, the CAPM does not apply without assumptions on the utility function, i.e. to justify the CAPM we would need a quadratic Bernoulli utility function.

Figure I plots the cross plot for a sample of observations from the three distributions. Both figures include a linear fit with intercept.

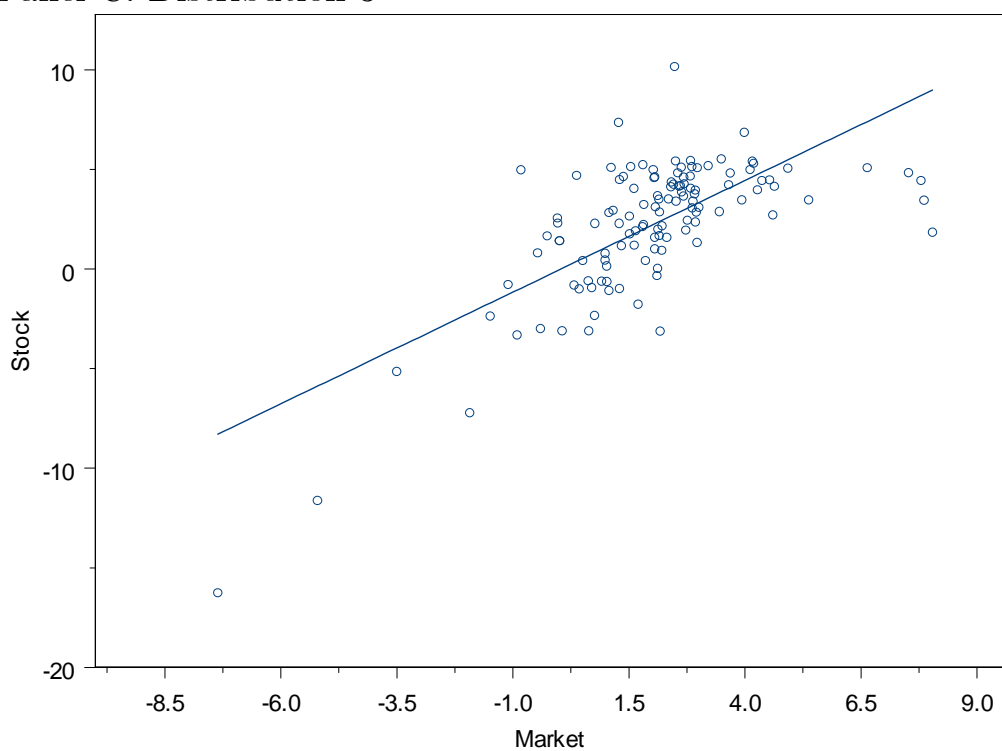
Figure I. Simulated Cross Plot of Market with Stock, $n = 120$
Panel A. Distribution 1



Panel B. Distribution 2



Panel C. Distribution 3



Unfortunately, the situation in Panel C is more often the rule than the exception. Hence, an asymmetric loss function appears to be more adequate. From our Monte Carlo simulation we compute the salient summary statistics

for the beta estimator. Results are reported in Table I.

Table I. Beta Estimator
 Panel A. Using Data simulated from Distribution 1

	OLS			alpha=							
	restricted slope	unrestricted intercept	unrestricted slope	-0.8		-0.2		0.2		0.8	
n=40				intercept	slope	intercept	slope	intercept	slope	intercept	slope
Mean	0.499	-0.010	0.501	-0.555	0.499	-0.150	0.501	0.130	0.502	0.536	0.505
Var.	0.005	0.080	0.010	0.113	0.015	0.081	0.011	0.082	0.011	0.116	0.015
Skewness	-0.026	0.132	-0.043	0.031	-0.124	0.135	-0.067	0.141	-0.018	0.320	0.057
Kurtosis	-0.127	0.261	0.120	0.153	0.484	0.298	0.159	0.233	0.136	0.537	0.342
Min.	0.298	-0.877	0.143	-1.545	0.076	-1.019	0.131	-0.730	0.154	-0.404	0.076
1st Qu.	0.452	-0.195	0.430	-0.779	0.425	-0.345	0.429	-0.060	0.436	0.299	0.424
Median	0.500	-0.022	0.502	-0.556	0.499	-0.161	0.505	0.119	0.504	0.532	0.503
3rd Qu.	0.551	0.181	0.570	-0.324	0.581	0.045	0.572	0.325	0.572	0.749	0.586
Max.	0.734	1.240	0.876	0.896	0.947	1.158	0.893	1.322	0.863	1.833	0.946
n=120				intercept	slope	intercept	slope	intercept	slope	intercept	slope
Mean	0.501	-0.004	0.502	-0.586	0.502	-0.151	0.502	0.143	0.503	0.576	0.504
Var.	0.002	0.024	0.003	0.037	0.005	0.024	0.003	0.025	0.003	0.040	0.005
Skewness	0.017	-0.154	0.178	-0.123	-0.002	-0.113	0.146	-0.158	0.179	0.093	0.021
Kurtosis	0.113	0.108	0.067	0.082	0.054	0.035	0.015	0.214	0.065	0.494	-0.028
Min.	0.375	-0.612	0.339	-1.270	0.270	-0.746	0.337	-0.482	0.341	-0.125	0.287
1st Qu.	0.475	-0.104	0.465	-0.716	0.456	-0.252	0.465	0.041	0.464	0.437	0.455
Median	0.503	0.001	0.501	-0.579	0.502	-0.146	0.502	0.149	0.501	0.579	0.502
3rd Qu.	0.527	0.102	0.537	-0.456	0.544	-0.042	0.537	0.253	0.541	0.713	0.551
Max.	0.640	0.476	0.712	0.024	0.736	0.355	0.709	0.747	0.712	1.480	0.716

Panel B. Using Data simulated from Distribution 2

	OLS			alpha=							
	restricted slope	unrestricted intercept	unrestricted slope	-0.8		-0.2		0.2		0.8	
n=40				intercept	slope	intercept	slope	intercept	slope	intercept	slope
Mean	1.200	-0.015	1.204	-2.128	1.208	-0.613	1.205	0.586	1.202	2.118	1.200
Var.	0.020	0.322	0.042	0.983	0.112	0.366	0.047	0.361	0.047	0.950	0.113
Skewness	-0.002	-0.042	-0.018	-0.659	-0.076	-0.014	-0.082	0.050	-0.027	0.749	-0.119
Kurtosis	-0.031	0.097	-0.165	1.326	1.480	0.094	-0.051	0.247	-0.207	1.608	0.488
Min.	0.743	-1.899	0.572	-7.259	-0.445	-2.632	0.375	-1.420	0.534	-0.560	-0.130
1st Qu.	1.104	-0.382	1.060	-2.704	1.002	-1.003	1.061	0.194	1.048	1.448	0.984
Median	1.197	0.002	1.206	-2.029	1.203	-0.609	1.207	0.577	1.206	2.037	1.203
3rd Qu.	1.300	0.354	1.347	-1.438	1.426	-0.205	1.358	0.998	1.352	2.659	1.416
Max.	1.738	1.954	1.874	0.991	2.523	1.485	1.829	2.634	1.910	7.583	2.198
n=120				intercept	slope	intercept	slope	intercept	slope	intercept	slope
Mean	1.201	0.005	1.200	-2.291	1.197	-0.607	1.200	0.619	1.199	2.325	1.199
Var.	0.007	0.108	0.014	0.424	0.048	0.121	0.015	0.122	0.015	0.404	0.044
Skewness	0.097	0.128	-0.076	-0.918	-0.042	0.040	-0.055	0.112	-0.038	0.594	0.055
Kurtosis	0.326	0.082	-0.008	2.561	0.660	0.072	-0.052	0.071	-0.147	0.976	0.359
Min.	0.924	-0.956	0.837	-6.182	0.210	-1.888	0.839	-0.511	0.845	0.655	0.422
1st Qu.	1.145	-0.220	1.123	-2.667	1.058	-0.844	1.115	0.403	1.114	1.875	1.058
Median	1.203	0.012	1.203	-2.255	1.200	-0.611	1.204	0.628	1.202	2.289	1.201
3rd Qu.	1.252	0.214	1.277	-1.861	1.334	-0.371	1.284	0.837	1.283	2.691	1.339
Max.	1.519	1.108	1.577	-0.693	2.070	0.486	1.584	1.719	1.608	5.121	1.978

Panel C. Using Data simulated from Distribution 3

	OLS			alpha=							
	restricted		unrestricted	-0.8		-0.2		0.2		0.8	
n=40	slope	intercept	slope	intercept	slope	intercept	slope	intercept	slope	intercept	slope
Mean	1.213	0.078	1.184	-2.673	1.324	-0.658	1.246	0.766	1.118	2.347	1.007
Var.	0.029	0.637	0.096	2.320	0.181	0.813	0.107	0.732	0.112	2.675	0.417
Skewness	0.255	0.280	-0.016	-1.253	0.463	0.027	0.059	1.083	0.147	4.110	0.505
Kurtosis	0.202	0.246	-0.120	5.353	1.666	0.626	-0.072	6.503	0.923	31.318	4.444
Min.	0.797	-2.453	0.257	-14.677	-0.036	-4.401	0.281	-1.709	-0.480	-0.403	-2.443
1st Qu.	1.100	-0.456	0.973	-3.498	1.041	-1.265	1.017	0.208	0.888	1.505	0.630
Median	1.208	0.031	1.185	-2.483	1.308	-0.690	1.242	0.703	1.115	2.124	0.932
3rd Qu.	1.320	0.555	1.392	-1.687	1.582	-0.102	1.473	1.257	1.331	2.808	1.295
Max.	1.917	2.962	2.165	1.107	3.521	2.527	2.408	8.090	2.708	21.065	4.512
n=120	restricted		unrestricted	-0.8		-0.2		0.2		0.8	
	slope	intercept	slope	intercept	slope	intercept	slope	intercept	slope	intercept	slope
Mean	1.209	0.022	1.199	-3.312	1.331	-0.777	1.265	0.786	1.120	2.898	1.087
Var.	0.010	0.214	0.033	1.448	0.088	0.284	0.040	0.254	0.041	3.255	0.455
Skewness	0.205	0.213	-0.099	-1.043	0.364	0.121	-0.054	0.355	0.360	3.884	0.162
Kurtosis	0.115	0.324	-0.010	2.266	1.801	0.514	0.085	0.785	2.031	24.041	3.023
Min.	0.902	-1.468	0.611	-9.558	0.202	-2.831	0.584	-0.670	0.426	0.203	-2.296
1st Qu.	1.140	-0.271	1.080	-3.935	1.141	-1.126	1.129	0.461	0.989	1.988	0.724
Median	1.205	0.007	1.204	-3.126	1.323	-0.786	1.265	0.770	1.114	2.443	0.989
3rd Qu.	1.275	0.310	1.320	-2.517	1.508	-0.449	1.398	1.101	1.251	3.174	1.414
Max.	1.545	1.712	1.767	-0.277	2.677	1.132	1.875	3.094	2.519	20.715	3.977

Table I shows that in the case of normality, the Linex slope estimator provides very similar answers to the least square estimator and it is basically independent of α . As noticed above, the intercept is not invariant with respect to α , but exhibit a symmetric behaviour. Moreover, as α is set closer to zero, not only the intercept becomes closer to the OLS intercept, but also the standard errors for both estimators under linex and OLS tend to become similar. On the other hand, results are very different for the third case, where Gaussianity does not hold anymore. In this case, standard errors appear to increase as $|\alpha|$ increases, but in a very different way. By construction, distribution 3 exhibits lower tails dependence, i.e. joint large negative values of Y and X are more likely to occur than joint large positive values of Y and X . For $\alpha < 0$, we penalize losses more than gains. However, by lower tail dependence, we can expect a higher concentration of observations in the extremes of the negative quadrant than in the positive one, hence estimates are more precise there.

4 Application to Regulated UK Companies

In this section we report the results we obtained by computing the beta coefficient for many UK regulated companies which are listed in the FTSE all share. Our time series sample comprises weekly prices on shares and the FTSE all share together with the three month-guild rate. The time series goes back to as much as 27/11/1992 to as short as just a few weeks, depending on the stock, up to 14/11/2003. We include the starting date in parenthesis for each stock. In particular, we consider the following companies identified by their Bloomberg ticker (details are in the appendix):

Electricity:

IPR LN Equity (13/10/2000), SSE LN Equity (27/11/1992), SPW LN Equity (27/11/1992), VRD LN Equity (02/07/1993)

Utilities:

CNA LN Equity (21/02/1997), IEG LN Equity (16/06/1995), KEL LN Equity (27/11/1992), NGT LN Equity (01/12/1995), PNN LN Equity (27/11/1992), SVT LN Equity (27/11/1992), UU LN Equity (27/11/1992), UUA LN Equity (26/09/2003).

We construct the excess returns on the companies and the market by subtracting the risk free interest rate over a week using the rule $7/365 \times$ (three-month risk free rate). We estimate the beta coefficient using the OLS estimator and the constrained and unconstrained Linex estimator for $\alpha = (-1 : 1) \cdot (.1)$. Table II provides summary statistics for the excess returns on all time series.

The beta estimates are reported in Table III.

Table II. Summary Statistics

	Mean	Var.	Skewness	Kurtosis	Min.	1st Qu.	Median	3rd Qu.	Max.
FT	0.01	4.20	-0.19	1.81	-8.11	-1.14	0.08	1.24	9.89
IPR	-0.52	32.01	-0.39	1.89	-22.94	-3.62	-0.91	2.55	14.74
SSE	0.12	12.48	1.13	10.27	-14.38	-1.96	-0.02	2.01	29.54
SPW	0.21	25.78	4.69	58.30	-17.06	-2.30	-0.10	2.31	68.17
VRD	0.24	16.38	3.84	46.66	-15.80	-1.68	-0.05	1.78	51.04
CAN	0.26	20.44	0.49	1.89	-12.34	-2.19	0.14	2.53	18.71
IEG	-0.02	5.04	1.46	11.37	-9.77	-0.66	-0.11	0.20	14.89
KEL	-0.02	21.04	-1.57	21.23	-47.25	-2.11	-0.08	2.16	24.07
NGT	0.09	12.00	0.39	0.82	-8.39	-2.01	-0.11	2.03	12.19
PNN	0.00	13.13	-0.49	17.67	-33.02	-1.72	0.04	1.58	23.58
SVT	0.03	12.54	-0.03	1.95	-18.10	-2.12	-0.03	2.17	14.38
UU	-0.05	12.62	0.13	2.07	-15.91	-2.04	-0.02	1.89	15.42
UUA	0.51	10.36	-0.05	-0.27	-4.78	-1.07	0.14	2.30	4.95

Table III. Beta Estimates

	OLS				alpha=											
	restricted		unrestricted		-0.8		-0.4		-0.2		0.2		0.4		0.8	
	slope	intercept	slope	intercept	slope	intercept	slope	intercept	slope	intercept	slope	intercept	slope	intercept	slope	
IPR	1.15	-0.22	1.14	-8.26	2.13	-2.68	1.53	-1.40	1.30	0.92	1.08	2.05	1.06	7.26	1.00	
SSE	0.39	0.12	0.39	-7.09	-0.17	-1.04	0.33	-0.44	0.37	0.91	0.39	3.28	0.55	20.41	3.19	
SPW	-0.17	0.21	-0.17	-9.41	0.42	-1.63	-0.18	-0.72	-0.18	9.21	2.23	30.43	5.80	54.80	7.83	
VRD	0.18	0.24	0.18	-8.11	0.40	-1.04	0.23	-0.39	0.24	2.64	-1.29	9.73	-2.94	28.02	-3.35	
CAN	0.69	0.31	0.70	-5.26	0.79	-1.31	0.72	-0.52	0.70	1.30	0.71	2.64	0.72	11.15	0.61	
IEG	0.02	-0.02	0.02	-2.98	0.23	-0.48	0.08	-0.25	0.05	0.28	-0.02	0.73	-0.08	8.02	-0.58	
KEL	0.27	-0.02	0.27	-32.10	-3.67	-12.15	-2.65	-2.40	-0.44	1.00	0.15	2.51	-0.19	13.00	-0.71	
NGT	0.54	0.11	0.54	-3.60	0.20	-0.90	0.43	-0.40	0.49	0.66	0.57	1.29	0.58	5.88	0.56	
PNN	0.17	0.00	0.17	-24.72	3.18	-4.59	0.40	-0.89	0.23	0.73	0.05	2.11	-0.26	14.54	-1.51	
SVT	0.27	0.03	0.27	-10.09	0.97	-1.33	0.31	-0.60	0.29	0.65	0.20	1.35	0.09	5.93	-0.47	
UU	0.46	-0.05	0.46	-6.40	0.91	-1.27	0.49	-0.64	0.47	0.56	0.47	1.29	0.47	8.40	0.35	
UUA	-0.41	0.66	-0.47	-1.32	-0.73	-0.06	-0.61	0.28	-0.55	1.05	-0.38	1.45	-0.29	3.12	-0.06	

Both summary statistics and the the results in Table III emphasize a large degree of nonlinearity for the data and the possibility of misspecification under OLS.

5 Conclusion

Our paper presents a theory of best linear prediction based on linex loss functions. The extra parameter involved reflects the regulator's preferences in the context of estimating the cost of capital. In the case of bivariate

normality, our answers do not differ substantially from the actual ones given by OLS. However, for arbitrary bivariate distributions, the answers are quite different and seem to reflect the degree of nonlinearity in the data. More work is needed to understand general conditions when β_1 , the slope parameter and "beta" of the stock, may be independent of α , and when it is not. We hope to report further results in later research.

A Brief Description of the Companies

Electricity:

IPR LN Equity:

International Power plc generates and sells electricity internationally. The Group has operating facilities in some 13 countries, including Australia, the Czech Republic, Malaysia, Pakistan, Portugal, Spain, Turkey, the United States and the United Kingdom.

SSE LN Equity: Scottish & Southern Energy Plc

Scottish and Southern Energy plc generates, transmits, distributes and supplies electricity to industrial, commercial and domestic customers in England, Wales and Scotland. The Group also provides electrical and utility contracting services, environmental control systems for the pharmaceutical and manufacturing sectors, and supplies natural gas.

SPW LN Equity: Scottish Power Plc

Scottish Power plc is an integrated power and energy group, which generates, trades and supplies electricity, in addition to providing electrical power systems in the UK and USA.

VRD LN Equity: Viridian Group Plc

Viridian Group PLC operates a utility company which procures, transmits, distributes and supplies electricity. The Group's main subsidiary, Northern Ireland Electricity plc, buys energy in bulk from independent power generating companies and distributes it to other retail suppliers and customers in Northern Ireland. Sx3 is the Group's information technology and business support services group.

Utilities:

CNA LN Equity: Centrica Plc

Centrica plc, through various subsidiaries, provides gas and energy related products and services to residential and business customers throughout Great Britain. The Group has a home services business, which allows customers to protect themselves against problems with their plumbing, heating and kitchen appliances. They also offer roadside assistance and indirect telecom services.

IEG LN Equity: International Energy Group Plc

International Energy Group Ltd. produces, distributes and transports gas. The Company also sells and distributes petroleum products. The Group's

main operations are based in Guernsey, with a number of operating subsidiaries in the United Kingdom, Guernsey, the Isle of Man, Portugal and Jersey.

KEL LN Equity: Kelda Group Plc

Kelda Group PLC, through its wholly-owned subsidiary, Yorkshire Water Services Ltd., provides drinking water and waste water services to the Yorkshire region. The Group's other core activities include treatment of liquid and solid wastes, collection and disposal of medical waste. Aquarion offers water services in the US. The company also has interests in water engineering and property management.

NGT LN Equity: National Grid Transco PLC

National Grid Transco PLC owns, operates and develops electricity and gas networks. The Group's electricity transmission and gas distribution networks are located throughout the United Kingdom and in the north-eastern section of the United States. They also own liquefied natural gas storage facilities in Britain and provide infrastructure services to the mobile telecom industry.

PNN LN Equity: Pennon Group Plc

Pennon Group Plc operates and invests primarily in the areas of water and sewerage services and waste management. Their principal subsidiary, SouthWest Water Limited, holds the water and sewerage appointments for Devon, Cornwall and parts of Somerset and Dorset. Viridor Waste Limited operates a waste treatment and disposal businesses in the United Kingdom.

SVT LN Equity: Severn Trent plc

Severn Trent plc supplies water, waste and utility services throughout the UK, Europe and the USA. The Group offers a range of water purification, sewage treatment and disposal, and recycling services. They also provide utility companies with a range of IT services and software solutions, as well as engineering consultancy and project management services.

UU/ LN Equity: United Utilities PLC

United Utilities plc is an international multi-utility business whose activities include water supply and distribution, wastewater services, facilities management, electricity distribution, business process outsourcing and telecommunications solutions. The Company provides services to approximately seven million people, primarily in the North west of England.

UU/A LN Equity: United Utilities PLC

United Utilities plc is an international multi-utility business whose activities include water supply and distribution, wastewater services, facilities management, electricity distribution, business process outsourcing and telecommunications solutions. The Company provides services to approximately seven million people, primarily in the North west of England.

B Lemma

Lemma 1 Suppose $X \sim N(\mu, \sigma^2)$. If $t < 1/(2\sigma^2)$, then

$$\mathbb{E} \exp \{sX + tX^2\} = \exp \left\{ -\frac{\mu^2 - \psi^2}{2\sigma^2} \right\} (1 - 2t\sigma^2)^{-1/2},$$

with

$$\psi = (\mu + s\sigma^2) (1 - 2t\sigma^2)^{-1/2}.$$

Proof.

$$\begin{aligned} \mathbb{E} \exp \{sX + tX^2\} &= \int_{\mathbb{R}} \frac{\exp \{sx + tx^2\}}{\sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{\sigma^2} \right\} dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2(1 - 2t\sigma^2) - 2x(\mu + s\sigma^2) + \mu^2}{2\sigma^2} \right\} dx. \end{aligned}$$

Make the substitution $y = x(1 - 2t\sigma^2)^{1/2}$ with Jacobian $dy/dx = (1 - 2t\sigma^2)^{1/2}$ and set

$$\psi = (\mu + s\sigma^2) (1 - 2t\sigma^2)^{-1/2}.$$

Then the last display equals

$$\begin{aligned} &\int_{\mathbb{R}} \frac{(1 - 2t\sigma^2)^{-1/2}}{\sqrt{2\pi}} \exp \left\{ -\frac{y^2 - 2\psi y + \mu^2}{2\sigma^2} \right\} dx \\ &= \exp \left\{ -\frac{\mu^2 - \psi^2}{2\sigma^2} \right\} \int_{\mathbb{R}} \frac{(1 - 2t\sigma^2)^{-1/2}}{\sqrt{2\pi}} \exp \left\{ -\frac{(y - \psi)^2}{2\sigma^2} \right\} dx \\ &= \exp \left\{ -\frac{\mu^2 - \psi^2}{2\sigma^2} \right\} (1 - 2t\sigma^2)^{-1/2}. \end{aligned}$$

■

References

- [1] Baesel, J.B. (1974) On the Assessment of Risk: Some Further Considerations. *Journal of Finance* 29, 1491-1494.
- [2] Bell D.E. and P.C. Fishburn (2001) Strong One-Switch Utility. *Management Science* 47, 601-604.
- [3] Black, A., P. Fraser and D. Power (1992) Unit Trust Performance 1980-1989: A Passive Time-Varying Approach. *Journal of Banking and Finance* 16, 1015-1033.

- [4] Blume, M.E. (1975) Betas and Their Regression Tendencies. *Journal of Finance* 30, 785-795.
- [5] Blume, M.E. (1971) On the Assessment of Risk. *Journal of Finance* 26, 1-10.
- [6] Bos, T. and P. Newbold (1984) An Empirical Investigation of the Possibility of Stochastic Systematic Risk in the Market Model. *Journal of Business* 57, 34-41.
- [7] Buckland, R. and P. Fraser (1999) Political and Regulatory Risk in Water Utilities: Beta Sensitivity in the United Kingdom. *Aberdeen Papers in Accountancy, Finance and Management*.
- [8] Elliott, G. and A. Timmermann (2003) Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions. Forthcoming in *Journal of Econometrics*.
- [9] Hall, P. and C.C. Heyde (1980) *Martingale Limit Theory and its Application*. New York: Academic Press.
- [10] Joe, H. (1997) *Multivariate Models and Dependence Concepts*. London: Chapman & Hall Ltd.
- [11] Knight, J., S. Satchell and G. Wang (2003) Value at Risk Linear Exponent Forecasts. *Quantitative Finance* 3, 332-344.
- [12] McCullagh, P. (1987) *Tensor Methods in Statistics*. London: Chapman and Hall Ltd.
- [13] Newey W.K. and J.L. Powell (1987) Asymmetric Least Squares Estimation and Testing. *Econometrica* 55, 819-847.
- [14] Rogers, L.C.G. and D. Williams (2000) *Diffusions, Markov Processes and Martingales*. Cambridge: Cambridge University Press.
- [15] Skouras, K. and A.P. Dawid. Consistency in Misspecified Models. Research Report, Department of Statistical Science, University College London.
- [16] Treynor, J. and F. Mazuy (1966) Can Mutual Funds Outguess the Market?. *Harvard Business Review* 44, 347-368.
- [17] 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*, p.195-205. North Holland.
- [18] Zellner, A. (1986) A Bayesian Estimation and Prediction Under Asymmetric Loss Functions. *Journal of the American Statistical Association* 81, 446-451.