How to objectively rate Investment Experts in absence of full disclosure? An approach based on a near perfect discrimination model

Introduction

- Fund managers in financial institutions often employ
  - Technical analysis
  - Statistical (time series) models
  - Regression models
  - Various types of decision models (endogenous variable is binary or categorical)

  for investment decision purposes.

Introduction

- Hedge fund managers
  - Are free to choose their Investment Strategy (IS)
  - Disclose only little or no information about their IS
- “Ordinary” fund managers
  - Are restricted in their IS
  - But don't like to provide full details about their predictions and decisions
- In both cases there is no full disclosure about the statistical decision making process

Example: market neutral IS

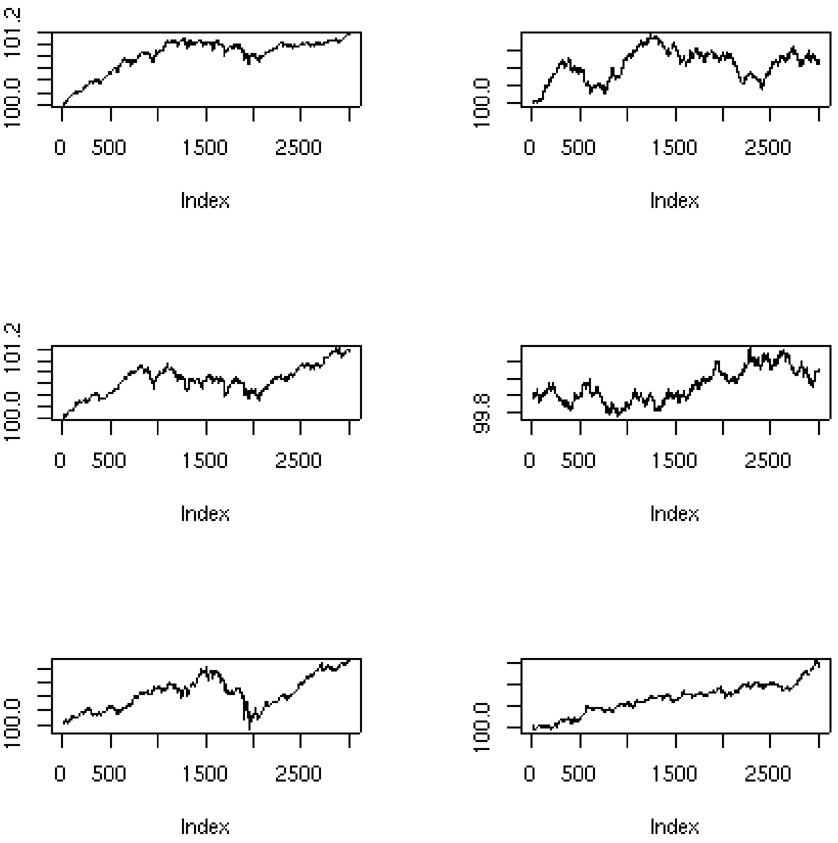
- Hedge funds often use a market neutral IS:
  - Define a Universe of tradable items (stocks)
  - Define a investment horizon (2 – 30 days)
  - Create a discrimination model that makes 3 piles:
    - Neutral pile (no position)
    - Long pile (buy stocks = long position)
    - Short pile (sell stocks = short position)
  - Hold short and long positions simultaneously during horizon period

Example: conclusion

- The hedge fund managers are unable to adequately demonstrate the qualities of their investment approach and the clients have no information about the risk they are actually taking by using the hedge fund investment vehicle.

Hypothesis

- real stock market time series exhibit fundamental, testable differences when compared to the Random-Walk (Fama-efficient)



Rating Procedure

- Rating is a function of the Expert's ability to discriminate between Real and Random-Walk time series

  This equally applies to Experts using:
  - Technical analysis
  - Statistical models
  - No model or technique at all :-)

Fundamental Problem

- This procedure does not take into account exogenous factors such as:
  - Temporal properties of the market (volatility, ...)
  - Geographical properties of the market
  - IS-related restrictions imposed by senior management or by law
- Therefore the rating can only be used for a single case (we cannot compare Experts)

Solution

- Statistical model that discriminates well (low alpha and beta errors)
- The discrimination quality of the model is used as a benchmark to create a **relative rating**
- This is possible if the model's performance is not too sensitive to external factors (time, place, ...)

Model

- Quasi Random-Walk (Airoldi, 2001)
- P. Cizeau, M. Potters, J.P. Bouchaud, *Correlation Structure of Extreme Stock Returns*, Quantitative Finance, 1, 217-222 (2001)
- Marco Airoldi, Correlation Structure and Fat Tails in Finance: a New Mechanism, Risk Management & Research, Intesa-Bci Bank, Milan, Italy (July 30, 2001)
-

Airoldi [1] formulates his model for N equities S_i for $i = 1, 2, ..., N$ that exhibit movements $\partial S_i = \pm s$ following a Quasi Random-Walk with "hopping probabilities" $P_{\partial S_i}$ that may depend on previous market returns. He continues to define the states of the market:

$$\begin{cases} h = 1 : P_{\partial S_i}\left(M^{(t-\Delta t)}\right) = \frac{1}{2} \quad + \frac{1}{2}\frac{\partial S_i^{(t)}}{s} g\left(M^{(t-\Delta t)}\right) \\ h = 0 : P_{\partial S_i}\left(M^{(t-\Delta t)}\right) = \frac{1}{2} \end{cases}$$

with $M^{(t-\Delta t)} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial S_i^{(t-\Delta t)}}{s}$, $|g(M)| \leq 1$.

Obviously, $P_{\partial S_i}$ depends on the previous market movements if $h = 1$. On the other hand, if $h = 0$ then $P_{\partial S_i}$ behaves like an ordinary Random-Walk and does not depend on market movements.

Simple Logistic Regression

- I expand on this idea and introduce the logistic relationship
  f = exp(gamma + delta X)
  where P(h=1) = f / (1+f)
  and where X is a “discriminating statistic”
- Estimation: Bias Reduced Logistic Regression:

➢ Firth, D. (1993) Bias reduction of maximum likelihood estimates. *Biometrika* 80, 27-38.

➢ Firth, D. (1992) Bias reduction, the Jeffreys prior and GLIM. In *Advances in GLIM and Statistical Modelling*, Eds. L Fahrmeir, B J Francis, R Gilchrist and G Tutz, pp91-100. New York: Springer.

➢ Heinze, G. and Schemper, M. (2002) A solution to the problem of separation in logistic regression. *Statistics in Medicine* 21, 2409-2419.

Best discriminating factor?

- Based on preliminary investigation we identified the p-value of the small sample Kurtosis

In this study we employ the following (sample) measure of kurtosis for the equities $S_i = \{x_1, x_2, ... x_n\}$ with $i = 1, 2, ..., N$;

$$K_i = \left(\frac{(n-1)n}{(n-2)(n-3)(n-4)} \sum_{j=2}^{n} \left(\frac{r_j - \overline{r}_i}{s} \right)^4 \right) - \frac{3(n-2)^2}{(n-3)(n-4)}$$

with $s = \sqrt{\frac{1}{n-2} \sum_{j=2}^{n} (r_j - \overline{r}_i)^2}$ and $r_j = \triangledown \ln x_j$ and

$$\overline{r}_i = \frac{1}{n-1} \sum_{j=2}^{n} r_j \text{ [6].}$$

The kurtosis measure K_i can be used for large and small samples. The standard error of K_i is $s_K = \sqrt{\frac{4\left((n-1)^2-1\right)s_s^2}{(n-4)(n+4)}}$ with $s_s^2 = \frac{6(n-1)n}{(n-3)n(n+2)}$. The test statistic is $z = \frac{K_i}{s_K} \leftarrow N(0,1)$ [6].

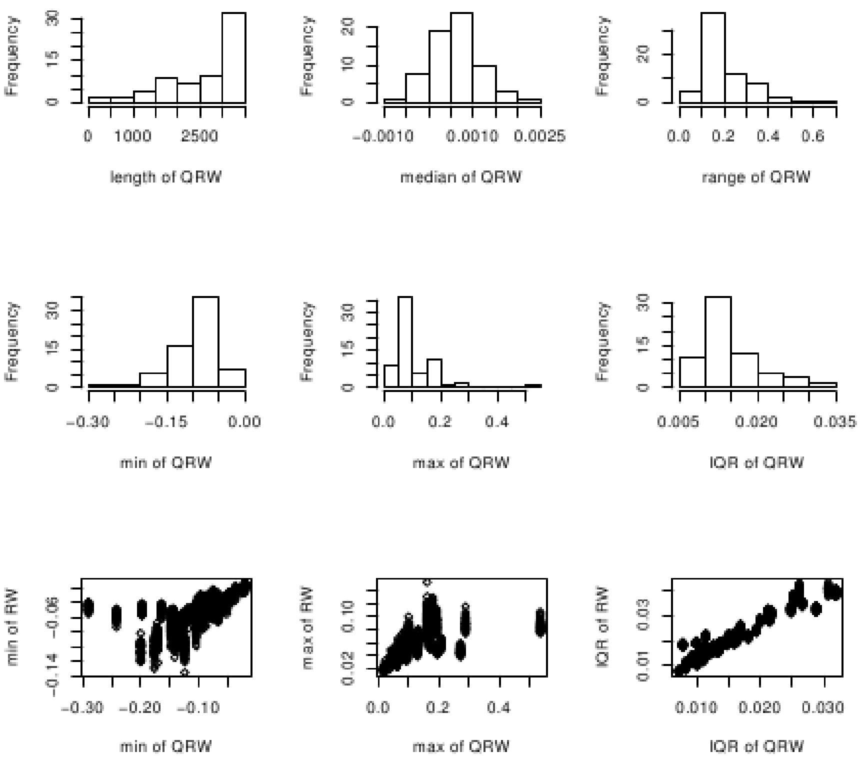
as the best factor (Vandervorst, Wessa, 2005).

Dataset

- 66 important index series including:
  - U.S. stock exchanges
  - U.S. Bonds, notes, treasury bills
  - gold, silver
  - Well-known stock exchanges in Europe and Asia
- Daily closing prices
- 1995 – 2006
- For every time series I simulate 20 R-W

Dataset

- 20 RWs were simulated for each real time series (QRW): # Series = 66 + 20*66 = 1386



Subseries

- Sequential Subseries were computed:
  - Min. length = 100 obs.
  - Max. length = 500 obs.
- Length of subseries is increased by 1 observation in each iteration
- For each subseries the small sample Kurtosis p-value was computed

Logistic Regressions

- For each length (= 100 to 500 with step 1) the logistic regression was computed:
  - Endogenous variable is binary
  - Exogenous variable is p-value of Kurtosis
  - f = exp(gamma + **delta** X)
    where P(h=1) = f / (1+f)
- All regressions have highly significant estimated **delta** parameters (T-Stats between -15.9 and -27.5)

alphas and betas

- Beta was computed for every regression, given a fixed alpha = 5%

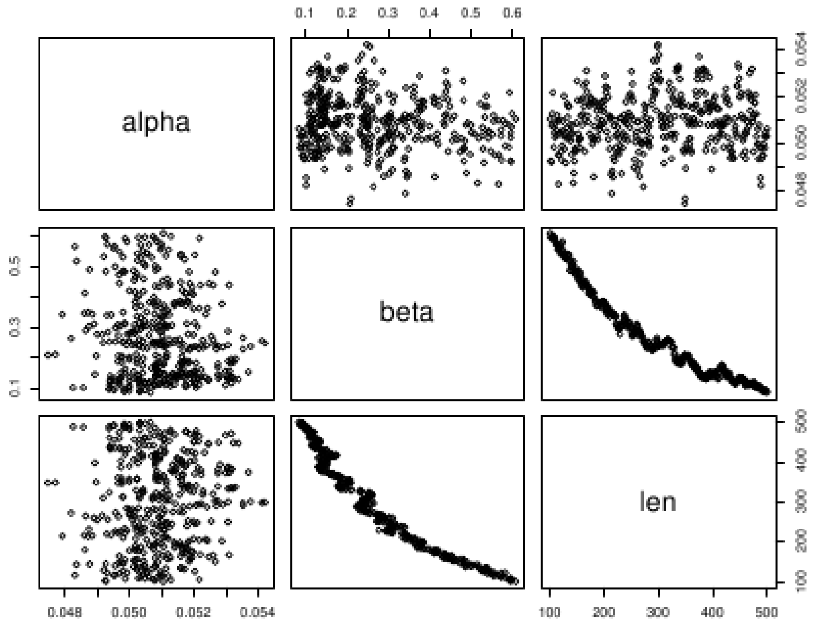
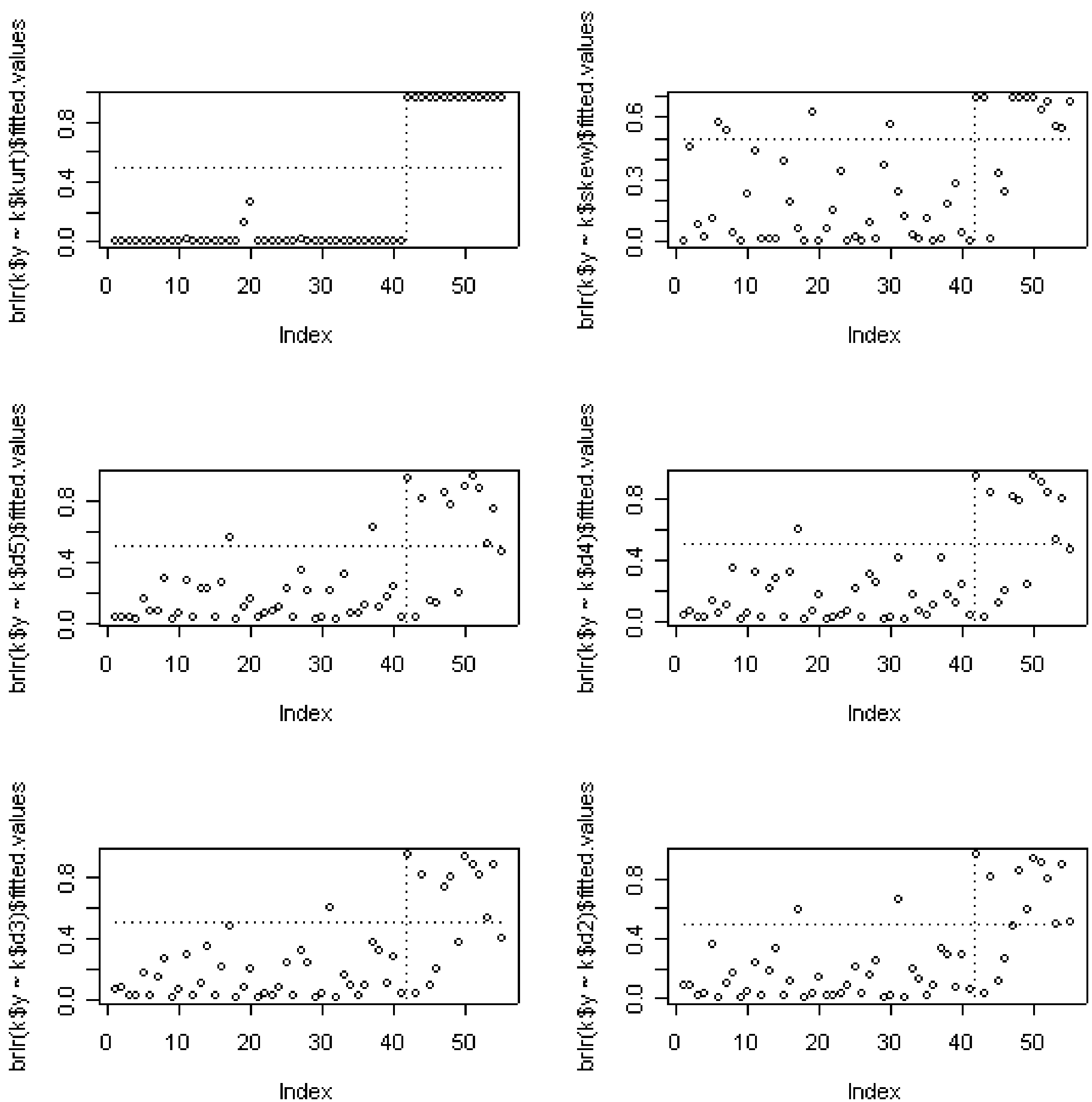


Figure 2. Type II errors depend on the time series length given a fixed type I error of 5%


brlr(k$y ~ k$kurt)$fitted.values
Index
brlr(k$y ~ k$skew)$fitted.values
Index
brlr(k$y ~ k$d5)$fitted.values
Index
brlr(k$y ~ k$d4)$fitted.values
Index
brlr(k$y ~ k$d3)$fitted.values
Index
brlr(k$y ~ k$d2)$fitted.values
Index

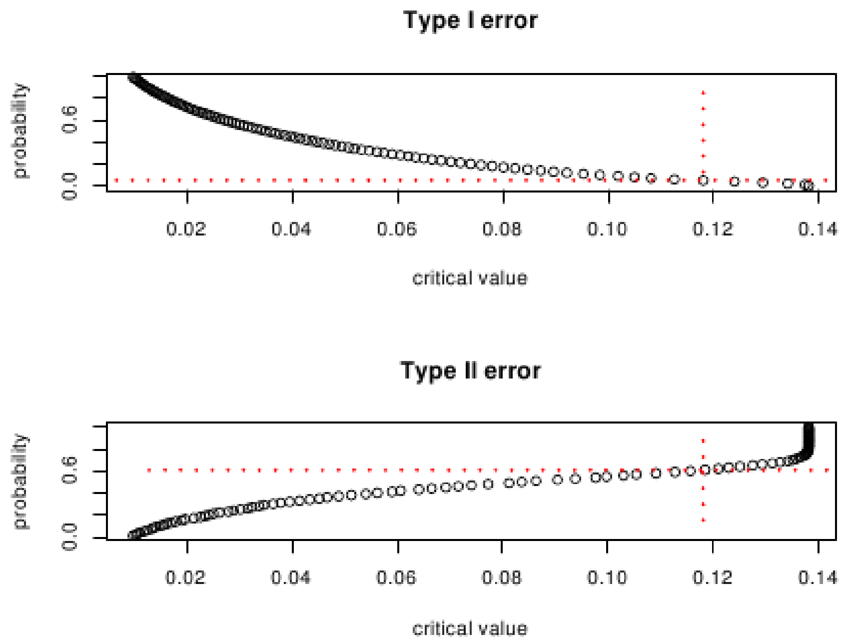


Figure 3. Type II error of kurtosis-based discrimination model in relationship with type I error (time series length = 100)

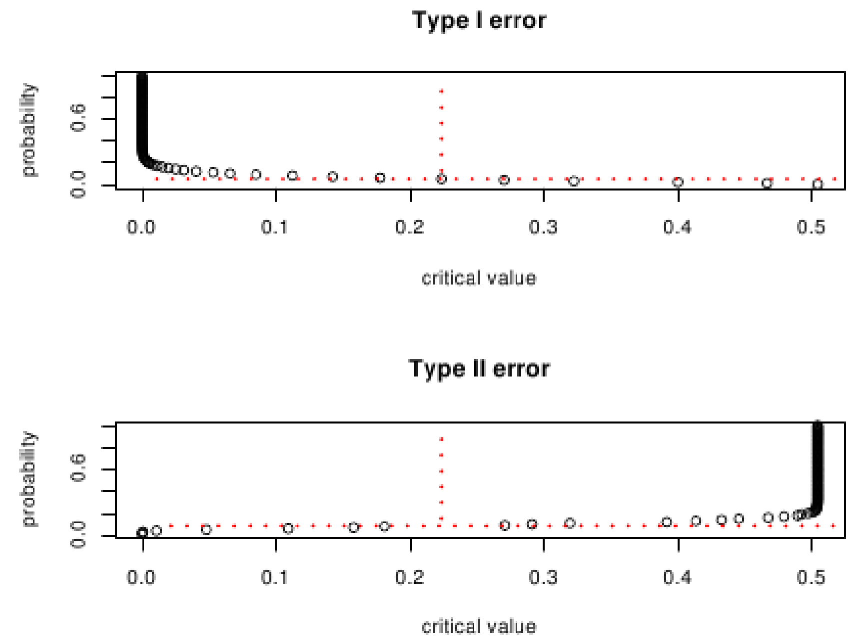


Figure 5. Type II error of kurtosis-based discrimination model in relationship with type I error (time series length = 500)

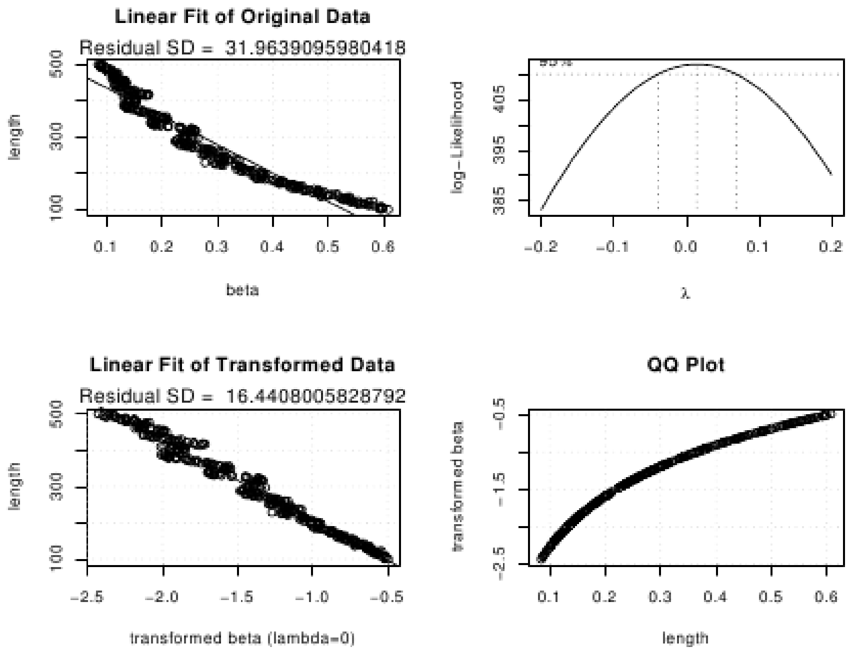


Figure 6. Relationship between required length and desired Type II error (criterion: Kurtosis p-value)

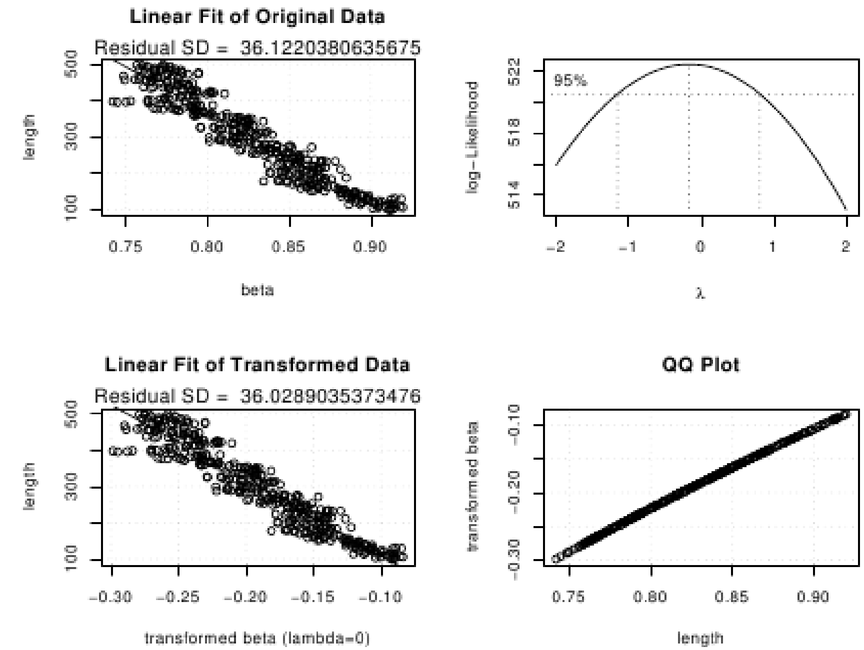


Figure 8. Relationship between required length and desired Type II error (criterion: Autocorrelation)

ally used in empirical research. For example, one might consider an autocorrelation-based discriminating statistic based on $X_{ijq} = \sum_{k=1}^{6} |\rho\left(\nabla ln\, x_t, \nabla ln\, x_{t-k}\right)|$ for $t = j\,(q-1)+1, j\,(q-1)+2, ..., j\,(q-1)+j, i = 1, 2, ..., N$, $j = n_{min}, n_{min}+1, ..., n_{max}$, and $q = 1, 2, ..., M_{ij}$. This discriminating statistic yields a power between 8.8% (for $j = 100$) and 22.73% (for $j = 500$) when used in the logistic regression instead of the kurtosis p-value. The power

Conclusions

- “Do real stock market time series exhibit fundamental, testable differences when compared to the Random-Walk?”
  -> Yes (Kurtosis p-value works great)
- Bias-Reduced Logistic Regression = non-linear transformation of probabilities
- We can use the model as a benchmark
  => it looks like we can make “fair” comparisons:
  - it only requires re-estimation of the model parameters
  - the model's performance promises to be good over time and place (and other factors?)
- The autocorrelation-based measure requires 4 times more observations to reach the same discrimination quality

Future work

- International comparison of Investment Experts
- Three categories instead of two:
(-1 = short ; 0 = neutral ; +1 = long)
- Multiple categories:
(Strong Sell, Sell, Hold, Buy, Strong Buy)
- Feel free to join...