Economic Attributes of Bankruptcy Probabilities Estimated by Maximum Likelihood

D.J. Johnstone
School of Business
University of Sydney
NSW 2006
Australia
d.johnstone@econ.usyd.edu.au

Incomplete Draft only

January, 2009

Abstract

From the perspective of a decision maker with log utility, maximum likelihood estimation [MLE] yields the best set of probabilities available from the forecaster's model, given the data in use. Of all the possible estimates that the chosen model might have generated, MLE leads to the probabilities (parameter values) that would have generated maximum financial return to a "growth-optimal" (log utility) investor, had those estimates been available and acted upon before the events in question. Decision makers with other utility functions may be similarly well served. A bootstrap experiment based on a representative set of corporate bankruptcy data suggests that although MLE estimates are not always a good proxy for probabilities estimated by maximizing another utility function, the out-of-sample economic benefits of acting upon MLE estimates are not easily improved upon by matching the estimation criterion to the user's utility function. In principle, MLE is widely justified by the proven asymptotic properties of its estimates. That MLE and other abstract statistical estimation criteria can be seen as inherently subjective (more or less suited to different forecast users) is not commonly understood.

1 Introduction

Financial decision makers obtain and condition their probabilities in ways that are only partly understood, and not necessarily supported theoretically. One of the more standard methods is to apply a probability forecasting model, such as described in econometric forecasting textbooks. This is portrayed as an impersonal or utility-free exercise, in that model estimation (fitting) relies on objective measures of "goodness of fit" and does not involve a utility function. It is possible, however, that certain estimation criteria, or rules of "best fit", have greater affinity or congruence with some users (utility functions) than with others. The clearest instance of such insidious subjectivity arises in the case of probabilities estimated by the method of maximum likelihood [henceforth MLE].

The first objective of this paper is to show that MLE-based probabilities, such as used in bankruptcy prediction and other investment decision contexts, are best suited theoretically to users with log utility. Whether in practice probabilities estimated by MLE satisfy a broader class of investors, remembering that log utility, otherwise known as "growth optimal" investment or "Kelly betting", is perceived by many as "too risky" (i.e. not sufficiently risk averse), is a more difficult question.[1] The second part of this paper describes a bootstrap exercise, using empirical bankruptcy prediction data, designed to test whether users with different utility functions (decision rules) might benefit economically from customized non-MLE goodness-of-fit criteria.

1.1 MLE-Based Bankruptcy Probabilities

What is the probability $p_i = pr(Y_i = 1|\boldsymbol{X}_i)$ of an arbitrary business entity, firm i, characterized by a vector of observables $\boldsymbol{X}_i = \{X_{i,1}, X_{i,2}, \dots, X_{i,k}\}$, obtaining a state $Y_i = 1$ of bankruptcy, insolvency or other default, prior to some given date t? This is a recurrent question, formalized by Ohlson (1980), and lately institutionalized within the financial governance framework of the Basel II capital accord (cf. Duffie et al. 2007, pp.636-7).

Empirical studies in bankruptcy, credit risk, debt ratings and related applications utilize most of the established families of probability forecasting models, including particularly logistic regression (Ohlson, 1980), hazard models (Shumway, 2001), mixed logit (Jones and Hensher, 2004) and the more recent dynamic time series model of Duffie et al. (2007). While the philosophical antecedents and specific assumptions of such roundly different models are partly contradictory, the technique by which they are most often fitted is generally uncontested. This is the method of maximum likelihood, or, in short, the rule of estimating model parameters such that the data set $\boldsymbol{Y}_i = \{Y_1, Y_2, \dots, Y_n\}$ observed is attributed maximum possible probability *ex ante* conditional on (i) the model and (ii) the explanatory variables $\boldsymbol{X}_{i,j}$ selected.

Maximum likelihood estimation was formulated and first described as such by R.A. Fisher (1921). In textbook expositions of the classical (frequentist)

[1] References include Li (1993, p. 915), MacLean et al. (1992, p.1564; 2004, pp.938,.941).

theory of point estimation, MLE is sometimes introduced as a "natural" or even axiomatic method of estimation. More conventionally, MLE-based point estimates are justified not merely by their intuitive appeal, but by their proven long-run frequentist properties, including unbiasedness or at least consistency, asymptotic efficiency and asymptotic normality (see Lehmann (1983) for further detail).[2]

When used to fit probability forecasting models in bankruptcy and related applications, the principle of MLE rarely attracts comment. It comes as a surprise, therefore, that probabilities estimated by MLE (whatever family of models is assumed) have an implicit and apparently unannounced economic rationale − namely, to best represent the purposes of a forecast user (i.e. decision maker) with log utility, $\log(W)$ of wealth W. More specifically, MLE produces the combination of probability estimates, $\boldsymbol{p} = \{p_1, p_2, \ldots, p_n\}$, over the sample firms $i \in \{1, 2, \ldots, n\}$, that, among all the possible values of $\boldsymbol{p}$ obtainable by altering one or more of the model's parameters, would have produced maximum wealth W *ex post* (and thus also maximum utility) had the various predictions $p_1, p_2, \ldots, p_n$ contained in $\boldsymbol{p}$ been acted upon *ex ante* by a rational decision maker with utility function $\log(W)$ (or any linear transform thereof).

In the context of bankruptcy or other default, and typical of nearly all portfolio optimization problems, exogenous determinants of investment performance include (i) the security prices *ex ante* and *ex post* of the firms being considered, and (ii) the set of alternative investment opportunities available to the decision maker over the time period or periods in question. Remarkably, neither makes any difference. On the contrary, taking the model and observation set $\boldsymbol{X}_i$ as given, it can be shown, under the assumptions of zero transaction costs and independent investment alternatives (multiplicative asset prices), that, regardless of (i) and (ii), the probability vector $\boldsymbol{p}$ estimated using MLE could not have been improved upon from the *ex post* perspective of a log utility investor. This presumes of course that the probabilities $p_1, p_2, \ldots, p_n$ represented by $\boldsymbol{p}$ are acted upon *ex ante* without further conditioning, as if adopted at that time as the decision maker's personal beliefs.

MLE is regarded widely as "objective" or utility-free, however the duality demonstrated in this paper between MLE and decision making under log utility, suggests otherwise. Probability estimates that suit one decision maker, with one particular utility function, can greatly disadvantage another. For instance, log utility agents are known to be severely disaffected by overstated probabilities, and relatively much less handicapped by understated probabilities (e.g. Thorp HHHH, Ziemba(?), Luenberger ???; pp.RRR; Stutzer 2003, pp.366-9).

By comparison, if we assume market betting odds in favor of bankruptcy of

[2] The history of MLE is set out by Aldrich (1997), Jaynes (2003) and Stigler (2007). For comparison of the place of maximum likelihood within Bayesian and classical inference philosophies, see Berger (1984, pp.121-41), Jaynes (2003, pp.175-7) and Howson and Urbach (1993, 293-4). Bayesian theory has no concept of MLE except in the elliptical sense that the likelihood function (the mode of which occurs at the frequentist MLE estimate) is, by corollary of Bayes' theorem, an evidentially exhaustive summary of the data, and combines with the user's prior distribution to produce a posterior probability distribution over all possible parameter values.

$\Omega = \pi_i/(1-\pi_i)$, where let us say $\Omega = 1$, a risk neutral investor is not affected by whether the estimated probability p_i is 0.51 or 1, or anywhere in between, since he wagers his entire bankroll on the same prospect either way. More generally, all that a risk neutral agent requires is a categorically correct probability estimate, meaning one on the "right side" (*ex post*) of the corresponding odds-implied market probability, π_i. The same may be true of other highly risk tolerant investors.

At the other extreme, intensely risk averse investors may wager so little when the forecast probability is near π_i that the wealth gained or lost is immaterial, irrespective of how many misclassifications the model makes. On their account, little might be gained out of a model with even zero "misclassifications". To be more profitable, such highly risk averse decision makers require "accurate" probabilities in a sense not captured by counting misclassifications. In particular, they will not be aided by a model fitted so as to minimize misclassifications when those probability estimates on the wrong side of π_i are often near 0 or 1. Such errors may prompt large, and perhaps catastrophic, losing bets.[3]

2 Background

Perhaps because of their origins in the works of predominately Bayesian theorists, see particularly de Finetti (1937, 1962, 1965, 1970, 2008), Good (1952), and Savage (1954, 1971), scoring rules are remarkably little known, even among professional statisticians. The most obvious exception to this clearly simplistic generalization occurs in meteorology, where probability scoring has a rich history dating at least to Brier (1950). In bankruptcy, explicit applications of scoring rules began with Lau (1987, p.135) but remain uncommon.[4] The more conventional ways of model evaluation are by counting numbers of misclassifications, often based on an arbitrary probability threshold such as 0.5, and by measuring an "accuracy ratio" or area under the model's ROC curve (see, for a good example, Duffie et al., 2007). In other economic forecasting environments, applications of scoring rules have become more common. References include Garratt et al. (2003), Lopez 2001 McKelvey and Page 1990, Onkal et al. 2003,

[3]This is why professional forecasters are inclined to "hedge" by reporting a probability closer to 0.5 or to the "climatological" (or "objective prior") probability than their true belief (cf. Johnstone 2007, Lichtendahl and Winkler 2007, and Ottaviani and Sorensen GGGG). "Climatological probability" is a term used generically for the observed long-run relative frequency of the event in question (Clemen and Winkler 1990, p.772). For example, in weather forecasting, the climatological probability of rain on a given day might be 9% or 0.09 for a certain location in a certain month of the year.

[4]In accounting, Scott (1979) and Gonedes and Ijiri (1974) introduced and made much of probability scoring rules. Yet, more recently, Hillegeist et al. (2004, p.19) suggested that "calibration", which is not the same property as "accuracy" and is not measured by a score function, is the one formal measure of the quality of probability assessments *ex post*. This would suggest that scoring rules remain unfamiliar in the accounting research literature. Moreover, it is well known that good calibration is not sufficient, since (perfectly) well calibrated probabilities can still be highly inaccurate (e.g DeGroot and Feinberg (1983, p.14). For example, a stock analyst who quotes 0.55 as his probability of any stock price increasing on any day will be close to perfectly calibrated over the long run.

Partington et al. (2005), Thomson (2004) et al. Muradoglu and Onkal (1994), Yates et al. (1991), Offerman et al. (2009) **etc etc**. Another fundamental development is the implementation of scoring rules as criteria for rewarding traders and setting bid and ask prices in probability prediction markets (e.g. Hanson 2003; Pennock 2004)

2.1 Probability Scoring Rules

Scoring rules are functions by which to measure the "accuracy" of a probability estimate *ex post*. Their most extensive use has been in meteorology, where predictions are often quoted in terms of probabilities (e.g. $pr(Rain) = 0.15$) rather than in more categorical or qualitative forms (e.g. "Fine with chance of showers"). Important references include Winkler (1969), Murphy and Winkler (1970, 1992) and Lindley (1982). Dawid (1986), O'Hagen (1994) and Winkler (1996) summarize the literature, and Bernardo and Smith (1994), Cover (IEEE Transactions paper), Roulston and Smith (2002), Daley and Vere-Jones (2004), Gneiting and Raftery (2007) and Jose et al. (2008) present important theoretical syntheses of scoring rules and related concepts from information theory.

The use of scoring rules is best explained by example. Consider an event Y_i with two possible outcomes $Y_i = 1$ (bankrupt) and its complement $Y_i = 0$ (not bankrupt). A model is used to predict the outcome of this event in the form of a probability $p_i = pr(Y_i = 1|X_i)$. The information $\boldsymbol{X}_{ij}$ utilized by the model is not at issue, nor is the type or form of model selected. All that matters is the accuracy of the end result, p_i, as can be measured only after the event.

Two possible probability score functions are illustrated. These are the conventional "Brier score" or quadratic score, in raw form

$$s(p_i\,,\,Y_i) = -(Y_i - p_i)^2 = \begin{cases} -(1-p_i)^2 & \text{if } Y_i = 1 \\ -p_i^2 & \text{if } Y_i = 0, \end{cases}$$

and the similarly well known logarithmic or "log score",

$$s(p_i\,,\,Y_i) = log\{Y_i\,p_i + (1-Y_i)(1-p_i)\} = \begin{cases} log(p_i) & \text{if } Y_i = 1 \\ log(1-p_i) & \text{if } Y_i = 0. \end{cases} \qquad (1)$$

Note that both functions imply a perfect score of zero, achieved when either $p_i = Y_i = 1$ or $p_i = Y_i = 0$. For less than perfect forecasts, both scores are negative. Perfectly bad scores are -1 for the Brier and $-\infty$ for the log score.

These are two of many plausible score functions (Jose et al., 2008). There is much discussion in the probability forecasting literature of what makes an appropriate score or measure of "accuracy" of a probability (e.g. Selton, 2007). Until recently, however, there has been little recognition of what different probability score functions reflect in terms of the potential (latent) economic value of the forecasts being scored (excepting Murphy, 1966).

The only generally agreed characteristic of a good scoring rule is that it is "strictly proper". This requires that the forecaster's (or model's) expected score,

$$E(s(g_i\,,\,Y_i)|p_i) = p_i\; s(g_i,1) + (1-p_i)\; s(g_i,0),$$

achieved by a nominal probability prediction g_i, conditional on an actual or "honest" probability assessment p_i, is maximized only when $g_i = p_i$. While inducement of forecast honesty seems desirable *prima facie*, its effect on the economic performance of a decision maker who relies on the reported probabilities when choosing whether, and how much, to invest is not clear, and is unlikely to be favorable under all conditions. For example, an inherently over-confident forecaster may be induced by a proper score function to report probabilities near 0 or 1, which, if often (or even occasionally) inaccurate, will be highly costly to a user who acts on them as his own.

2.2 MLE in Terms of the Log Score

Consider a probability prediction model such as the common logistic function

$$p_i = pr_{\boldsymbol{\beta}}(Y_i = 1|\boldsymbol{X}_i) = \frac{e^{\boldsymbol{\beta X}}}{1+e^{\boldsymbol{\beta X}}}, \tag{2}$$

fitted with a vector of parameters $\boldsymbol{\beta} = \{\beta_1, \beta_2,\ ...\ ,\beta_k\}$. On the usual assumption in logistic regression and similar models of independent or conditionally independent probabilities p_i, the estimated joint probability of the observed data $\boldsymbol{Y}_i = \{Y_1, Y_2,\ ...\ ,Y_n\}$, usually called the likelihood function, is

$$l(\boldsymbol{Y}|\boldsymbol{\beta}) = \prod_{i=1}^{n} p_i^{Y_i}(1-p_i)^{1-Y_i}, \tag{3}$$

with $Y_i \in \{0,1\}$ for all i. The principle of MLE is to find the parameter set $\boldsymbol{\beta_{MLE}} = \mathbf{arg\,max}\ \{l(\boldsymbol{Y}|\boldsymbol{\beta})\}$ that maximizes the product (3). Equivalently, and more conveniently, $\boldsymbol{\beta}$ is estimated by maximizing the log likelihood function

$$L(\boldsymbol{Y}|\boldsymbol{\beta}) = \log\{l(\boldsymbol{Y}|\boldsymbol{\beta})\} = \sum_{i=1}^{n} \log(\nu_i), \tag{4}$$

where ν_i is the model's *ex ante* probability of the outcome Y_i known (after the event) to have occurred in the case of firm i (for all $i = 1, 2,\ ...\ ,n$),

$$\nu_i = p_i^{Y_i}(1-p_i)^{1-Y_i} = \begin{cases} p_i & \text{if } Y_i = 1 \\ (1-p_i) & \text{if } Y_i = 0. \end{cases}$$

Note that the log likelihood function (4) is simply the sum of the log scores (1) of all the model's individual probability predictions, $p_1, p_2,\ ...\ ,p_n$. MLE can be described, therefore, as the principle of selecting the probability forecasting model that achieves the highest aggregate log score over all n sample observations.

The notion that a probability forecasting model can be fitted by optimizing an appropriate score function $s(p_i, Y_i)$, such as the Brier or log score, or any other strictly proper score function, is apparently very recent. Gneiting and Raftery (2007, p.375) call this "optimum score estimation" and note that the appeal of this estimation framework lies in "the potential adaption of the scoring rule to the problem at hand". They further observe (p.375) that "[m]aximum likelihood estimation forms the special case of optimum score estimation based on the logarithmic score" (this was first observed by Winkler 1969, pp.1076-7). In spirit with the Gneiting-Raftery proposal, this paper exposits MLE as a form of optimum score estimation tailored implicitly to a user (economic decision maker) with log utility.

3 Likelihood as a Measure of Value

Recent papers by Roulston and Smith (2002), Daley and Vere-Jones (2004), Johnstone (2007) and Jose et al. (2008) have emphasized parallels between the conventional log score, various related measures of information, entropy or "distance" between probability distributions, and the success of bets based on the probability distributions in question. It is assumed for the purposes of such comparisons that the decision maker can express his personal probabilities (or those produced by a model) in the form of bets, either at fixed market odds (as quoted by a bookmaker) or in a parimutuel betting market. The only requirement of either market is that there is no commission or bid-ask spread, meaning that the betting odds available on one possible outcome, say $Y_i = 1$, are simply the reciprocal of those available on the complementary event, $Y_i = 0$ (or, put another way, that the odds-implied probabilities of the two possible events sum to one).

In the case of bankruptcy, there is generally no associated betting market. There exists, however, a simple analogy between bets and investments, particularly investments in assets such as corporate bonds that may default and return nil of the investor's initial outlay. Once this connection is made explicit, the economic attributes of MLE-based probabilities can be described in either the succinct and picturesque language of professional gamblers, or, equivalently, in more conventional terms from finance and economics.

3.1 Betting on Bankruptcy

It is important to identify the formal correspondence between investment and betting, if for no other reason than to clarify what most traders and fund managers suspect intuitively − that investment and betting at a rational level are the same information-dependent profit-driven activity.[5]

[5] Gambling and financial markets differ in substance in essentially two parameters. The first is the long run rate ψ at which "informed" players siphon money away from "uninformed" players (cf. Asche et al., 1982). In roulette $\psi = 0$, since no one has an information advantage, whereas in horse racing, the same perhaps as in the stockmarket, ψ may be very high,

A conventional bet of amount δ on an outcome $Y_i = 0$, against the complementary outcome $Y_i = 1$, is defined as an outlay that returns a gross payoff equal to δ multiplied by a factor

$$\begin{cases} 0 & \text{if } Y_i = 1 \\ \alpha > 1 & \text{if } Y_i = 0. \end{cases}$$

For example, a bookmaker may quote a "price" of $\alpha = 1.91$ on Connors to beat Borg. A gambler who bets δ on Connors will be returned 1.91δ if indeed Connors wins and zero if Borg wins. The win multiple α is always greater than one because it includes the dollar wagered.

In conventional bookmaking terms, the gross payoff (per \$1 bet) α_1 from a successful bet on $Y_i = 1$ equals

$$\left[1 + \frac{1}{\Omega_1}\right] = 1/q_i,$$

where $\Omega_1 = q_i/(1 - q_i)$ represents the odds "in favor" of outcome $Y_i = 1$ and q_i is the odds-implied probability of $Y_i = 1$. In the absence of any bookmaker commission (bid-ask spread), the reciprocal of Ω_1 represents the odds in favor of $Y_i = 0$ (odds against $Y_i = 1$) and the odds-implied probability of $Y_i = 0$ is $1 - q_i$.[6]

Bets can be replicated with investments, and vice versa, as follows. Imagine a binary asset representing the stock in a company that will be either bankrupt ($Y_i = 1$) or not bankrupt ($Y_i = 0$) by period end. The current stock price is S_i (whether buying or selling) and period end stock price will be either $S_i^{Y_i} = S_i^1$ (*Bankrupt*) or $S_i^{Y_i} = S_i^0$ (*Not Bankrupt*), where $S_i^1 < S < S_i^0$.

Adapting the method of binary option pricing of Cox and Ross (1976) to bets, which are "contingent claims" just like options, an investment position short one unit in the underlying stock is replicated by a portfolio containing $-S_i^0/(1+r)$ (a short position) in risk free bonds together with a bet of amount $S_i^0/(1+r) - S_i = [S_i^0 - S_i(1+r)]/(1+r)$ on $Y = 1$ (bankruptcy) at payoff

$$\alpha_1 = \left[1 + \frac{1}{\Omega_1}\right](1+r) = \frac{S_i^0 - S_i^1}{S_i^0 - S_i(1+r)}(1+r) = \frac{(1+r)}{q_i}, \tag{5}$$

particularly over a sequence of "insider" trades. The other important parameter is ξ, the rate of growth in the pool of funds available for distribution *ex post* to the players (informed and uninformed). In gambling, ξ is negative since the payout pool is just the sum of all the bets minus whatever commission is taken by those conducting the market (e.g. the casino or the operators of a parimutuel betting pool). In the stockmarket, ξ is historically positive (thanks of course to the value added by managers and employees, borrowed capital, Government subsidies and so on). From the viewpoint of an uninformed player, positive ξ makes the stockmarket more appealing at a rational level than the casino (a random selection of stocks can be expected to yield a positive return if held long enough). An informed player sees no such qualitative distinction. If she has a systematic information advantage, her expected long run growth factor $(1+\psi)(1+\xi)$ may far exceed 1 (zero growth) in either marketplace.

[6] Where there is a spread, the odds-implied probabilities of $Y = 1$ and $Y = 0$ sum to $(1+\varepsilon) > 1$ where ε is called the over-round or "vig" (short for vigorish).

where $\Omega_1 = q_i/(1-q_i)$ represents the risk neutral odds in favor of $Y_i = 1$, and the corresponding risk-neutral probability of $Y_i = 1$ is

$$q_i = \frac{S_i^0 - S_i(1+r)}{S_i^0 - S_i^1}. \tag{6}$$

The gambler's net outlay is then $-S_i$, meaning that he receives amount S_i, the same as if he had sold one unit of underlying asset, and the total value of his portfolio of bonds-plus-bet at period end is $-S_i^0$ in the case of $Y_i = 0$ and $-S_i^1$ in the case of $Y_i = 1$. This exactly replicates the short sale of one unit of the underlying asset.[7]

Similarly, a position long one unit in the underlying stock is replicated by a portfolio containing $S_i^1/(1+r)$ in risk free bonds together with a bet of amount $S_i - S_i^1/(1+r) = [S_i(1+r) - S_i^1]/(1+r)$ on outcome $Y_i = 0$ at payoff

$$\alpha_0 = \left[1 + \frac{1}{\Omega_0}\right](1+r) = \frac{S_i^0 - S_i^1}{S_i(1+r) - S_i^1}(1+r) = \frac{(1+r)}{1-q_i}, \tag{7}$$

where $\Omega_0 = 1/\Omega_1 = (1-q_i)/q_i$ represents the "risk neutral" odds in favor of $Y_i = 0$ (not bankrupt). The gambler's net outlay is then S_i, and the total value of his portfolio of bonds-plus-bet at period end is S_i^0 in the case of $Y_i = 0$ and S_i^1 in the case of $Y_i = 1$, exactly the same as if he had bought one unit of the underlying stock.

An intuitive understanding of betting against a "risk-neutral probability" is to imagine that a winning bet is rewarded for its holding cost as well as for its winning. The gambler is first credited with the risk-free interest rate on the amount wagered over the holding period of the bet. If, for example, he bets on the firm's survival by purchasing one unit of the stock, then he is credited with having earned interest of $S_i r$ over the period. His winning bet is then of total amount $S_i(1+r)$, and is treated as occurring instantaneously at period end. At that moment, he receives a total payout of S_i^0, implying effective or "risk-neutral" betting odds of $\Omega_0 = (1-q_i)/q_i$, where q_i is defined by (6). For mathematical consistency, a losing bet should also be regarded as earning the risk-free rate over the holding period, but since this bet is lost so is the accrued interest.

An apparent deficiency of this depiction of investment-as-betting is that the two possible *ex post* stock prices S_i^0 and S_i^1 are unknown *ex ante*. Importantly, however, this does not prevent rational betting, particularly log-utility betting. First, it is reasonable to set $S_i^1 = 0$, on the simplifying assumption that stock in a bankrupt firm is worthless, or "out of the money", just like a losing bet. And second, as in any parimutuel betting market,[8] rational bets can be made

[7] A practical way to bet on bankruptcy (to hold the equivalent of B securities) is to buy what are known as digital or binary default swaps, that payout in the event that the named firm defaults. See for example Lando (2004, p.198).

[8] In a parimutuel betting market, the final payout α on a winning \$1 bet is given by the total betting pool divided by the number of \$1 bets on the winner (assuming no commission). The well known Kyle (1985) model in market microstructure is parimutuel.

either by assuming a probability distribution for S_i^0, or by adopting a decision rule whereby the assumed value of S_i^0 makes no difference to the amount bet. Interestingly, as shown below, a Kelly bettor or investor with log utility makes the same bet (takes on the same investment portfolio) whatever the assumed value of S_i^0. Remarkably, therefore, it makes no difference to him that S_i^0 is unknown *ex ante*. This was one of the unexpected results demonstrated by Kelly (1956).

A second possibility is that the security prices are those of unsecured bonds issued by the company in question. If these securities have say $100 face value (and no coupon) and expire at period end (thus matching the time horizon of the probability forecasting model) then their *ex post* value in the case of bankruptcy is $S_i^1 = 0$ (or at least very close to zero) and in the case of solvency, $S_i^0 = 100$. Under these simplifying assumptions, selling (buying) bonds at *ex ante* price S_i is equivalent to making fixed-odds bets on the event of a bankrupt (not bankrupt) firm at period end, where $q_i = (100 - S_i(1+r))/100$ (e.g. if the current bond price is $S_i = 75$ and $r = 0$, $q_i = 0.25$ and the effective fixed market betting odds are 3 to 1 against bankruptcy).

3.2 Log Optimal (Kelly) Betting

A Kelly (log utility) gambler with *ex ante* wealth W bets on a discrete binary event $Y_i \in \{1, 0\}$. His personal probabilities of $Y_i = 1$ and $Y_i = 0$ are p_i and $(1 - p_i)$ respectively. The risk-neutral betting odds in favor of $Y_i = 1$ are $\Omega_1 = q_i/(1 - q_i)$. There is no commission or breakage and hence the risk-neutral odds in favor of $Y_i = 0$ are $1/\Omega_1$. Following Kelly (1956, p.922), the gambler bets a fixed proportion γ of his wealth on $Y_i = 1$ and the remainder $(1 - \gamma)$ on $Y_i = 0$. Because there is no commission, these bets are partly self-cancelling, meaning that there is implicitly a proportion of the gamblers initial wealth that remains unbet and invested at the risk-free rate r.

The gambler's expected utility after trial i is

$$
\begin{aligned}
& p_i \ \log[\gamma W(\frac{1+r}{q_i})] + (1-p_i) \ \log(1-\gamma) W(\frac{1+r}{1-q_i}) \\
= \quad & \log[W(1+r)] + p_i \log\left(\frac{\gamma}{q_i}\right) + (1-p_i) \ \log\left(\frac{1-\gamma}{1-q_i}\right).
\end{aligned}
$$

Differentiating with respect to γ leads to a maximum such that

$$
\frac{p_i}{\gamma} - \frac{(1-p_i)}{1-\gamma} = 0,
$$

giving $\gamma = p_i$ and $(1 - \gamma) = (1 - p_i)$. It follows, therefore, that a Kelly gambler or log optimal investor allocates his initial wealth W, whatever its amount, to bets on $Y_i = 1$ (bankrupt) and $Y_i = 0$ (not bankrupt) in proportions matching his subjective probabilities of those events, p_i and $(1 - p_i)$, regardless of the available betting odds.[9]

[9] See Luenberger (1998, pp.419-25) for the first mainstream-in-finance textbook treatment

In principle, therefore, Kelly betting in a frictionless (zero commission) complete market is very straightforward. If there is a binary security B that pays \$1 in the case of $Y_i = 1$ and zero in the case of $Y_i = 0$, and a matching security *not-B* that pays \$1 in the case of $Y_i = 0$ and zero in the case of $Y_i = 1$, then a Kelly bettor with personal probability $p_i = pr(Y_i = 1) = 0.6$ and capital of say \$100 simply buys \$60 worth of B and \$40 worth of *not-B*. The unit volumes of these trades depend on the two security prices, which by the principle of no arbitrage must add up to 1, or $1/(1+r)$ if there is a holding period. Alternatively, if only the B securities are traded, then the Kelly strategy is to buy \$60 worth of these while at the same time selling $40/(1-P)$ units of the same security where P is the security price. In net terms, the Kelly gambler must therefore buy $60/P - 40/(1-P)$ units, meaning 250 units when for example the security price is $P = 0.2$. Similarly, if $P = 0.75$, he must buy -80 units, or, that is, sell 80 units.[10]

Note again that this strategy arises without reference to the implied betting odds. The only assumption is that the two securities, B and *not-B*, are priced according to the same risk-neutral probability distribution, so as to leave no arbitrage opportunities, or no "Dutch book" in the words used by decision theorists (e.g. Savage HHHH, Lindley RRRR, de Finetti KKKK, DeGroot YYYY). Their respective risk-neutral betting odds are then reciprocals, $\Omega_1 = 1/\Omega_0$.

3.3 Kelly Betting and Maximum Likelihood

MLE is used to estimate $\boldsymbol{p} = \{p_1, p_2, \ \dots \ , p_n\}$, where p_i represents the probability of result $Y_i = 1$ in trial i. The investor employs these probabilities to make Kelly (log utility) bets over the n trials observed, in whatever combination is optimal under that decision rule. It is common in bankruptcy prediction models, that all n trials in the sample occur simultaneously, rather than sequentially, in which case bets have to be made together (in portfolio) rather than one after the other.

Grant et al. (2008) examine the problem of simultaneous Kelly-betting and demonstrate several results relevant here. It is assumed that the market is complete, thus allowing the investor to trade an Arrow-Debreu type security that pays \$1 in the event of any chosen intersection of outcomes $\bigcap_{i \in M} y_i$, where $y_i \in \{0, 1\}$ is the gambler's designated value of Y_i, $I = \{1, 2, ..., n\}$ and $M \subseteq I$ is a subset of trials of size $m \leq n$. For instance, the investor can buy or sell a security that pays \$1 in the state of (say) firms $i = 2$ and $i = 4$ going bankrupt, and $i = 9$ not going bankrupt, $s = \{Y_i : Y_2 = 1, \ Y_4 = 1, \ Y_9 = 0\}$, and

of Kelly betting. In effect, Luenberger's text presents finance theory as an over-riding mathematical theory of gambling.

[10] The same effect could be achieved practically by buying \$60 of call options (or of stock) while also buying \$40 of put options with a very low strike price (both expiring at period end). If the firm goes bankrupt the call options are out of the money (the \$60 is lost) but the \$40 is a winning bet and pays out at the corresponding risk-neutral odds. Similarly, if the firm survives, the puts are out of the money but the \$60 wins at the reciprocal risk-neutral odds. Note that because the options are priced according to a given risk-neutral probability distribution, and the bid-ask spread is negligible (lets assume), the implied betting odds are very close to reciprocals.

pays zero in any other state. With this bet, the outcomes Y_i for $i \neq 2, 4, 9$ make no difference. Gambles of this description are called m-multis, partly for convenience but mainly because they pay only when all m designated results occur in conjunction.[11]

It is also assumed that the bookmaker or betting exchange (securities market) treats all n discrete events Y_i as independent, or conditionally independent given the available information. This assumption is implicit within the usual likelihood function (3). Assuming independence, the commission-free payout (per $1 bet) on a winning m-multi is

$$\alpha_s = (1+r)^m \frac{1}{\prod\limits_{i \in M} \pi_i} \tag{8}$$

where

$$\pi_i = q_i^{Y_i}(1-q_i)^{1-Y_i} = \begin{cases} q_i & \text{if } Y_i = 1 \\ (1-q_i) & \text{if } Y_i = 0 \end{cases}$$

is the market (risk-neutral) probability of the event Y_i realized in trial i. For example, if the market probabilities of firms $i = 2$, $i = 4$ and $i = 9$ going bankrupt are $q_2 = 0.2$ and $q_4 = 0.5$, and $q_9 = 0.2$, then the payout (per $1 bet) on a successful 3-multi on the conjunction $s = \{Y_i : Y_2 = 1,\ Y_4 = 1,\ Y_9 = 0\}$ is $\alpha_s = (1+r)^3(1/q_2))(1/q_5)(1/(1-q_9)) = 5 \times 2 \times 1.25 = 12.50$ (with $r = 0$). Note that π_i is the bookmaker's risk-neutral equivalent of the investor's probability assessment ν_i, implying that a successful bookmaker will generally have $\pi_i > \nu_i(1+r)$ over $i = 1, 2, \ldots, n$.

Based on the assumptions of independent events and no commissions, Grant et al. (2008, pp.14-16) demonstrate that it makes no difference under log utility betting whether bets are made simultaneously or sequentially, in any arbitrary sequence (or in any combination of simultaneous and sequential bets). The same finding holds for any utility function under which the optimal bet in each trial i is a fixed fraction of wealth W regardless of the amount of W. This includes not only Kelly betting, or log utility, but any of the broad family of power utility functions. It also includes so-called fractional-Kelly betting, where the gambler bets a constant preset fraction (e.g. 50%) of the full-Kelly bet in each trial.[12]

That log optimal bets made simultaneously or sequentially (at the same multiplicative odds) produce the same monetary outcome (and hence utility) is a highly convenient result. It allows the economic consequences of log optimal betting upon a set of probabilities $\boldsymbol{p}$, produced by a forecasting model, to be equated directly to the observed value of the likelihood function, and thus affords

[11] Compound bets such as this are known in Britain as "accumulators", in the US as "parlays", and in Australia as "multi-bets" (http://www.multibet.com). They are available only on trials that the bookmaker regards as independent, such as two different football games, but not on dependent events such as Connors winning both the semi-final of Wimbledon and the final.

[12] This is a standard decision rule in professional gambling; see for example Thorp XXXX and Ziemba RRRR (40% or 50% Kelly is widely advocated for its compromise between high expected growth and low volatility of returns).

the likelihood function an economic meaning that is not well known, and may not be intended or desirable.

To formalize this interpretation of the likelihood function, consider a sequence of single-event bets ("1-multis") over trials $i = 1, 2, \ldots, n$, each trial producing outcome $Y_i \in \{1, 0\}$. In making each bet, the investor acts on the basis of a probability $p_i = pr(Y_i = 1|\boldsymbol{X})$ produced by the chosen model, and estimated using MLE. The risk-neutral betting odds in trial i, in favor of outcome $Y = 1$, are $\Omega = q_i/(1 - q_i)$, and hence a Kelly-bettor multiplies his wealth in trial i by a factor

$$h(p_i,\, Y_i,\, q_i) = (1 + r)\, \frac{\nu_i}{\pi_i} = \begin{cases} (1 + r)\, p_i/q_i & \text{if } Y_i = 1 \\ (1 + r)\, (1 - p_i)/(1 - q_i) & \text{if } Y_i = 0, \end{cases} \tag{9}$$

After completing all n discrete bets, an investor with log utility increases his initial wealth W_0 by factor

$$\prod_{i=1}^{n} h(p_i,\, Y_i,\, q_i) = (1 + r)^n\, \frac{l(\boldsymbol{Y}|\boldsymbol{\beta})}{\prod_{i=1}^{n} \pi_i}, \tag{10}$$

where $l(\boldsymbol{Y}|\boldsymbol{\beta})$ is the observed sample value of the likelihood function. It follows, therefore, that the estimated model parameters $\boldsymbol{\beta}$ yielding maximum likelihood $l(\boldsymbol{Y}|\boldsymbol{\beta})$, conditional on the model and sample $\boldsymbol{X}$ in use, are those which (of all possible parameter estimates $\boldsymbol{\beta}$) would have led a log utility investor to maximize his return from betting over that set of trials based on that model's probability estimates $\boldsymbol{p} = \{p_1, p_2, \ldots, p_n\}$.

To out-predict the market in any given trial i, and hence profit from the bet after discounting at the risk-free rate r, the investor must have $\nu_i > \pi_i$. For example, if the market placed *ex ante* risk-neutral probability 0.5 on the event seen to occur, and the investor gave this event personal probability 0.4, then his wealth (discounted at rate r) would have been multiplied by 0.8 (i.e. 20% of his capital would have been lost). This seems a rather large loss given how close the market and investor were in their rival probability assessments, underlining the inherent "high risk" of the log utility betting rule, and thus again begging the question of whether a method of estimation that implicitly defers to log utility is appropriate.

There is one other conceptual aspect to the interpretation of $l(\boldsymbol{Y}|\boldsymbol{\beta})$ as a measure of the investment potential of a probability prediction model from the standpoint of a log optimal investor (Kelly bettor). In short, provided that all investment alternatives related to the Y_i events are treated by the market as mutually independent and also independent from all other unrelated investment alternatives (e.g. betting on sports or whether it will rain tomorrow), the bankruptcy prediction model with the highest likelihood $l(\boldsymbol{Y}|\boldsymbol{\beta})$, or equivalently log likelihood $L(\boldsymbol{Y}|\boldsymbol{\beta})$, is the model or parameter set $\boldsymbol{\beta}$ which a log utility investor would (in retrospect) have most wanted to bet upon. Under these assumptions, the investor's *ex post* wealth is still given by (10), but with the addition of one more multiplicative term, capturing the factor by which wealth is increased by taking up a log optimal portfolio of those further investment opportunities.

Note that the log of the trader's capital growth factor (10), representing his gain in log utility $\log(W_n) - \log(W_0)$ from Kelly-betting over trials $i = 1, 2, \ldots, n$, based on $\boldsymbol{p}$, can be written in terms of the difference between his aggregate log score and the aggregate log score of the market

$$\log \prod_{i=1}^{n} h_i = \sum_{i=1}^{n} \{\log(\nu_i) - \log(\pi_i)\} = L(\boldsymbol{Y}|\boldsymbol{\beta}) - \sum_{i=1}^{n} \log(\pi_i), \tag{11}$$

where $L(\boldsymbol{Y}|\boldsymbol{\beta})$ represents the observed log likelihood of the model (and letting $r = 0$). This further clarifies the duality between extensibly "statistical" measures of the accuracy of a set of probability forecasts, namely the log score and the log likelihood $L(\boldsymbol{Y}|\boldsymbol{\beta})$, and the economic returns accruing to a decision maker who bets in each trial i according to a log utility function.

4 Bootstrap Economic Returns from MLE

Granger and Pesaran (2000a; 2000b) distinguish between conventional measures of forecast accuracy, such as the Brier score, log score or number of misclassifications, and economic measures such as, for example, the terminal wealth of an investor who switches between bonds and stocks on the basis of a time series of probability forecasts. They argue that ultimately it is economic criteria that matter since forecasts exist not merely for their statistical accuracy in some theoretical sense, but as a basis for better decisions as proven by better outcomes.[13] This instrumentalist perspective on forecast quality is widely endorsed in financial forecasting and forecast evaluation (e.g. Leitch and Tanner 1991; 1995; Pesaran and Timmerman 1994; 1995). In the case of probability forecasts, there are demonstrable parallels between abstract notions of forecast accuracy, as captured by particular scoring rules, and (hypothetical) economic outcomes. Recent results proven by Jose et al. (2008) suggest that within the broadly inclusive class of power utility functions, there are specific families (e.g. quadratic and exponential utilities) that are the mathematical duals of known or identifiable probability scoring rules.

The issue then is whether it might assist forecast users with utility functions apart from $\log(W)$ if models were estimated using criteria other than that of maximizing the aggregate log score (4), as required by MLE. To test for this possibility empirically, a simulation study is conducted based on repeated sampling from a representative set of corporate bankruptcy data. The aim of the experiment is to examine whether hypothetical investments (bets) are more successful when based on MLE-based probabilities or on probabilities estimated using an

[13]...forecasts are made for a purpose and the relevant purpose in economics is to help decision makers improve their decisions. It follows that the correct way to evaluate forecasts is to consider and compare the realized values of different decisions made from using alternative sets of forecasts. (Granger and Pesaran 2000a, p.537).

alternative utility or score function, chosen to represent the ends of decision makers with decidedly greater risk aversion than log-utility agents (who implicitly act in the way of Kelly (1956) so as to maximize their expected long run capital growth rate, and hence risk large fractions of capital relative to typically more risk averse investors).

The results of this experiment, reported below, suggest that estimation procedures customized to suit the personal risk aversions of individual investors may not lead to better decisions, as proven by better (higher average utility) returns distributions, across a wide spectrum of risk aversions.

4.1 The Experiment

This experiment reveals that although probabilities estimated by MLE are not always a good proxy for estimates obtainable by "optimum score estimation" under utility functions other than $\log(W)$, the out-of-sample economic benefits from acting upon MLE estimates may remain. The data employed is a relatively large sample of bankruptcy-related observations drawn from firms listed on the Australian Stock Exchange (ASX) over the period 1993-2003.[14] To keep the experiment as simple as possible while also maintaining some level of practical realism, the prediction model estimated is a conventional binary logistic regression containing five regressors. For the sake of illustration, rather than to seek out the "best" possible model, the five explanators are familiar accounting variables, first advocated as a relevant composite of firm-specific factors by Altman (1968).[15] The model is thus as per the standard form (2), with explanatory variables $X_{i,1}$=working capital/total assets, $X_{i,2}$=retained earnings/total assets, $X_{i,3}$=earnings before interest and taxes/total assets, $X_{i,4}$=market value of equity/book value of total assets, and $X_{i,5}$=sales/total assets. After eliminating missing observations, the remaining sample contains 7012 firm-years of data.

4.1.1 Bootstrap Procedure

The experiment is designed to produce bootstrap (repeated sub-sampling) distributions of the utility realized by different decision makers (utility functions) acting on the probabilities estimated using different possible score functions (rules of best fit). These distributions are then compared to assess how well the competing estimation criteria perform relative to the economic ends of two different classes of decision makers.

The bootstrap routine (repeated many times) is to draw a random subsample containing $n = 500$ firm-years of data from the 7012 lines of data available, and then fit the logistic regression model to this sample according to different possible score (utility) functions. Sampling is with replacement, although

[14] This data was kindly made available by Stewart Jones, and is a subset of that used in Jones and Hensher (2004).

[15] Although Altman (1968) uses discriminate analysis rather than logistic regression, the variables employed are still of the same relevance.

this makes no difference to the results.

The first score function applied is the log score, implicit within MLE. The resulting estimates are the conventional MLE estimates, the same as those obtained with standard software packages. The second score function is one derived by Jose et al. (2008, p.????) and based on decision making under a quasi-linear exponential utility function. Quasi-linear utility functions are widely employed in the micro-economics literature as a way to represent the innate consumption-versus-investment preferences of typical decision makers (e.g. see ref).

To capture the aggregate utility achieved by a decision maker acting on a set of probability estimates $\boldsymbol{p} = \{p_1, p_2, \ ... \ , p_n\}$, a score function will generally involve not only the decision maker's probabilities p_i, but also the corresponding market probabilities $\boldsymbol{q} = \{q_1, q_2, \ ... \ , q_n\}$. The one obvious exception to this generalization arises under log utility, where, as apparent in equation (11) above, the "best" available set of estimates $\boldsymbol{p}$ is that with the highest sum of log scores (log likelihood) regardless of $\boldsymbol{q}$. This is clearly an important practical advantage of MLE.

For a binary event, $Y_i \in \{0, 1\}$, the quasi-linear exponential utility score function, taken from Jose et al. (2008), is

$$s(p_i\,, q_i, Y_i) = 1 - \left(\frac{\pi_i}{\nu_i}\right) + E_{q_i}\left[\log\left(\frac{\pi_i}{\nu_i}\right)\right] \qquad (12)$$

$$= \quad 1 - \left(\frac{q_i}{p_i}\right)^{Y_i}\left(\frac{1-q_i}{1-p_i}\right)^{1-Y_i} + q_i \log\left(\frac{q_i}{p_i}\right) + (1-q_i)\log\left(\frac{1-q_i}{1-p_i}\right),$$

where $E_{q_i}\left[\log\left(\pi_i/\nu_i\right)\right]$ represents the market's expectation in trial i of the difference between its own log score, $\log(\pi_i)$, and that of an investor whose score in that trial is $\log(\nu_i)$. To implement this score function, the market probability q_i of bankruptcy ($Y_i = 1$) for each firm i is set equal to the overall observed frequency of bankruptcy across the 7012 observations in the data set, leaving $q_i = 0.031945$ for all i. This is a convenient assumption, sufficient for the sake of exposition, however it may also be treated as an implicit reality check, since any reasonable model should yield positive betting returns to a rational investor (of whatever utility function) against such an uninformed (indiscriminate), albeit perfectly well calibrated, market or market maker.

The equivalent of $L(\boldsymbol{Y}|\boldsymbol{\beta})$ or (11) under the score function (12) is

$$S(\boldsymbol{Y}|\boldsymbol{\beta}) = n - \sum_i^n \left(\frac{\pi_i}{\nu_i}\right) + \sum_i^n E_{q_i}\left[\log\left(\frac{\pi_i}{\nu_i}\right)\right], \qquad (13)$$

This is a measure of the total utility gained by the investor from betting upon the vector of probabilities $\boldsymbol{p} = \{p_1, p_2, \ ... \ , p_n\}$, in the same way as the log likelihood $L(\boldsymbol{Y}|\boldsymbol{\beta})$ captures the aggregate increase in utility to a log utility investor. Corresponding functions could be derived for investors with other utility functions in the broad class described by Jose et al. (2008), although these may not be easy to optimize.

It will assist if we denote the probabilities estimated by maximizing (4) as $\boldsymbol{p}^{mle}$, and those obtained by maximizing (13) as $\boldsymbol{p}^{exp}$. The following results were obtained by repeating the bootstrap procedure $B = 2000$ times. In each repeat, the logistic regression model was estimated by optimizing (4) and then (13), so as to produce the model parameter estimates $\boldsymbol{\beta}^{mle}$ and $\boldsymbol{\beta}^{exp}$ that perform best according to these two criteria respectively. The optimization procedure employed is the standard Newton-Raphson method within *Mathematica*.

Before seeing these estimates, there is no guarantee that these two estimators will produce dissimilar (and thus practically distinct) probabilities. In trying to anticipate whether the estimates might differ, there are two factors to consider. The first is whether the $\boldsymbol{X}$ variables employed are jointly informative enough to separate the two groups of firm-years (bankrupt and not-bankrupt). If on the whole these are not good discriminators, then both estimation criteria will tend to produce probabilities around 0.5, whatever the score function, since the estimated parameters $\boldsymbol{\beta}$ will all be near zero.

The second factor, more related to the content of this paper, is whether the rival score functions reward "different" probabilities disparately, and, if so, what "types" of probabilities p_i (e.g. bold or conservative) will emerge within $\boldsymbol{p}^{mle}$ and $\boldsymbol{p}^{exp}$ respectively. Figure 1 compares the log and exponential probability score functions, (1) and (12).

Insert Figure 1 about here.

The obvious difference between these is that (12), with constant market probability of $q_i = 0.0321945$, implying $\pi_i = 0.0321945$ ($\pi_i = 0.968055$) for a bankrupt (not bankrupt) firm, rewards extreme probabilities - when they turn out to be "right" - relatively much more generously than the log score (1). That is, probabilities p_i extremely close to one (zero) for a bankrupt (not bankrupt) firm yield very large positive scores under (12), relative to the rewards under that score function for somewhat "less correct" probabilities, indicating that an investor with the exponential utility function underpinning (12) obtains relatively massive utility from such "recklessly bold", yet ultimately correct, predictions. Given that such predictions are of so much benefit under this utility function, it seems likely that the exponential rule-of-best-fit (13) will produce bolder (nearer 0 or 1) probabilities than the MLE rule (4), provided of course that the explanatory variables $\boldsymbol{X}$ are sufficiently informative enough to allow this level of discernment (without overly frequent prediction errors). Results displayed below confirm this suspicion.

4.1.2 Results

The first set of results presented in Figures 2 and 3 suggest that the two sets of estimates differ considerably. Figure 1 shows the difference $L(\boldsymbol{Y}|\boldsymbol{\beta}^{mle}) -$

$L(\boldsymbol{Y}|\boldsymbol{\beta}^{exp})$, and thus represents the added log utility that would have been obtained by betting (within sample) on the MLE estimates $\boldsymbol{\beta}^{mle}$ rather than on $\boldsymbol{\beta}^{exp}$. Note that, by definition, this difference must always be positive. Otherwise the MLE optimization procedure must have failed.

Figures 2 and 3 about here.

To give some idea in money terms of what is found here, a difference in the log likelihood functions of 20[100] corresponds to a wealth multiple of $\exp(20) = 4.8517E + 08$ [$\exp(100) = 2.6881E + 43$]. Thus, if the difference is 20 then the MLE parameters would have produced, after completion of 500 trials (bets), a multiple of 4.8175E+08 times the ending wealth (to a log utility investor) relative to that produced under $\boldsymbol{\beta}^{exp}$. This seems an astonishingly large return factor, but is achieved with the addition of just 4.0811% to the geometric average investment return per trial.

Figure 2 shows the same beneficial effect for the exponential utility investor. His incremental utility from betting (within-sample) on the estimator matched to his utility function, rather than on the MLE estimator, $S(\boldsymbol{Y}|\boldsymbol{\beta}^{exp}) - S(\boldsymbol{Y}|\boldsymbol{\beta}^{mle})$, is also, in general, very pronounced. Again, this difference should be positive in every bootstrap repeat.

The results above show that, within-sample, the two sets of probabilities are different, at least in the sense that they imply very different economic outcomes. To examine whether there is any obvious difference in the appearance of $\boldsymbol{p}^{mle}$ and $\boldsymbol{p}^{exp}$, Figures 4 and 5 represent frequency histograms of the associated vectors $\boldsymbol{\nu}^{mle}$ and $\boldsymbol{\nu}^{exp}$, where in each case v_i is obtained from p_i according to (5). When looking at these two distributions, remember that ideally all the v_i should equal one, as v_i is the estimated probability of the outcome (bankrupt or not bankrupt) that indeed occurred in firm-year i. Each distribution represents a histogram of $2000 \times 500 = 1,000,000$ $(T \times n)$ probabilities.

Note that the two distributions could each have been partitioned by conditioning on the result Y_i.[16] It is often informative to look at the conditional distributions of the p_i under $Y_i = 1$ and then separately under $Y_i = 0$. In our case, however, there is no conceptual difference between the two states of nature, since the loss function is the same either way, and is simply a function of the difference between v_i and π_i.[17] The role of the market probability π_i can be seen in the respective loss functions, namely (12) under log utility and (13) under exponential utility.

[16] Note that this avoids all the usual issues of conventional discrete loss functions, where it is assumed that for any p_i above some threshold or "critical" probability of bankruptcy, the decision maker bets on bankruptcy (e.g. sells the stock short) and either wins if the firm goes bankrupt, or losses if it does not (thus making a "Type 2" error, since betting on solvency produces a bigger and thus "Type 1" loss when the firm goes bankrupt).

[17] This would not be so if the decision problem was not symmetric, as for example when the market is incomplete and allows betting only against bankruptcy, not for it.

Figures 4 and 5 about here.

The difference in the shapes of the two distributions is clearly evident. Although the maximum likelihood estimates are more tightly distributed near one, the exponential utility estimates exhibit many more probabilities of almost exactly one. The weakness of this estimator is that it also generates far more probabilities well away from one, and a good many very near zero, constituting the worst possible prediction error both in terms of accuracy and economic consequence.

On balance, it is hard to say which distribution is better in any objective sense. Rather, all that can be said is that a log utility decision maker would have done better by using $\boldsymbol{\nu}^{mle}$ probabilities, and an exponential utility decision maker would have been better served by $\boldsymbol{\nu}^{exp}$. This is simply a restatement of what is necessitated by the two estimation rules.

The real issue is whether the benefits of user-oriented goodness-of-fit criteria remain when bets are made out-of-sample. That is, does a decision maker with a utility function other than log, obtain greater expected utility from future bets based on "customized probabilities" produced by the model ($\boldsymbol{\beta}$) that would have made him most money historically (by acting on his utility function)? In principle, it would seem that so long as that data (estimation sample) is sufficiently representative of future trials, then past success should be maintained out-of-sample (over future trials).

This does not appear to hold true in our experiment. Rather, the MLE estimates based on a sub-sample ("training sample") of 500 random observations, $\boldsymbol{\beta}^{mle}$, are seen to generate higher expected utility over a newly drawn random sample of 500 observations than the EXP estimates, $\boldsymbol{\beta}^{exp}$, regardless of whether the investor has log or exponential utility.

This can be seen in Figures 6 and 7. Figure 6 reveals the incremental utility that the investor gains out-of-sample by betting with $\boldsymbol{\beta}^{mle}$ rather than $\boldsymbol{\beta}^{exp}$. The average gain in either monetary or utility terms is immense and there is no indication that the log utility user should depart from MLE. In just a very small percentage of repeats, the EXP estimates, $\boldsymbol{\beta}^{exp}$, outperform $\boldsymbol{\beta}^{mle}$. This kind of reversal is to be expected sometimes, since chance will occasionally generate probabilities that happen to coincide closely with empirical outcomes, under any plausible (or even implausible) estimation rule.

Figures 6 and 7 about here.

Supporting MLE as a "general purpose" estimation rule, the results in Figure 7 show that the same great advantage of $\boldsymbol{\beta}^{mle}$ over $\boldsymbol{\beta}^{exp}$ remains even for the class of investor whose utility function is used to generate $\boldsymbol{\beta}^{exp}$. Note that

Figure 7 is truncated in that all negative net utility gains <10,000 are set in the -9900 to -10,000 bin. This is for ease of presentation.

The problem apparent in Figure 7 is that the model, when fitted according to the exponential score, seems to be "over-trained" by a utility function that almost pathologically rewards very extreme probabilities (when these are right). To generate such probabilities in-sample, the model coefficients are pushed to levels that produce many similarly bold estimates, p_i, some very near 0 or 1, out-of-sample. Many of these probabilities turn out to be categorically "wrong", in the sense that $|p_i - Y_i| \approx 1$, thus bringing about relatively frequent colossal losses that demolish the overall profitability of those estimates over the 500 trials out-of-sample.

The results shown are in fact worse than they appear. In an attempt to improve the performance of $\boldsymbol{\beta}^{exp}$, any probability estimate within $\delta = 0.0001$ of either 1 or 0 was truncated to 0.9999 or 0.0001. This caused significant improvement but nowhere near enough to overcome the advantage of $\boldsymbol{\beta}^{mle}$ over $\boldsymbol{\beta}^{exp}$, under either utility function. The same is true even with $\delta = 0.01$, which is about as large a truncation δ as seems reasonable in this application (where it would seem sensible that at least some firms in some years have less than 1% chance of going bankrupt).

To be continued.....

5 Conclusion

References

[1] Aldrich, J. (1997) R.A. Fisher and the Making of Maximum Likelihood 1912-1922. *Statistical Science*. 3: 162-176.

[2] Altman, E. (1968) Financial Ratios, Discriminate Analysis and the Prediction of Corporate bankruptcy. *Journal of Finance*. 23: 589-609.

[3] Aucamp, D.C. (1993) On the Extensive Number of Plays to Achieve Superior Performance with the Geometric Mean Strategy. *Management Science*. 39: 1163-1172.

[4] Bernardo, J.M. (1979) Expected Information as Expected Utility. *The Annals of Statistics*. 7: 686-690.

[5] Bernardo, J.M. and Smith, A.F.M. (1994). *Bayesian Theory.* New York: Wiley.

[6] Berger, J.O. and Wolpert, R.L. (1984) *The Likelihood Principle.* Hayward, CA: Institute of Mathematical Statistics.

[7] Blume, L. and Easley, D. (1992) Evolution and Market Behavior. *Journal of Economic Theory.* 58: 9-40.

[8] ———— (2002) Optimality and Natural Selection in Markets. *Journal of Economic Theory.* 107: 95-135.

[9] ———— (2006) If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets. *Econometrica.* 74: 929-966.

[10] Breiman, L. (1961) Optimal Gambling Systems for Favorable Games. In Neyman, J. and Scott, E. (eds) *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.* 1: 65-78. Berkeley: University of California Press.

[11] Brier, G.W. (1950) Verification of Forecasts Expressed in Terms of Probability. Monthly *Weather Review.* 78: 1-3.

[12] Clemen, R.T. (1986) Calibration and Aggregation of Probabilities. *Management Science.* 32: 312-314.

[13] ———— (1987) Combining Overlapping Information. *Management Science.* 33: 373-380.

[14] Clemen, R.T. and Winkler, R.L. (1990) Unanamity and Compromise Among Probability Forecasters. *Management Science.* 36: 767-779.

[15] Clements, M.P. (2004) Evaluating the Bank of England Density Forecasts of Inflation. *Economic Journal.* 114: 844-866.

[16] Cover, T. (TTTT)

[17] Daley, D.J. and Vere-Jones, D. (2004) Scoring Probability Forecasts for Point processes: The Entropy Score and Information Gain. *Journal of Applied Probability.* 41A: 297-312.

[18] Dawid, A.P. (1986) Probability Forecasting. In Kotz, S., Johnson, N.L. and Read, C.B. *Encyclopedia of Statistical Sciences.* Vol. 4. Pp.228-36. New York: Wiley.

[19] de Finetti, B. (1962) Does it Make Sense to Speak of 'Good Probability Appraisers'? In Good I.J. (ed.) *The Scientist Speculates: An Anthology of Partly Baked Ideas.* Pp.357-364. London: Heinemann.

[20] ———— (1965) Methods for Discriminating Levels of Partial Knowledge Concerning a Test Item. *The British Journal of Mathematical and Statistical Psychology.* 18: 87-123.

[21] ———— (1970) Logical Foundations and Measurement of Subjective Probability. *Acta Psychologica*. 34: 129-145.

[22] ———— (1974) *Theory of Probability*. Vol. 1. New York: Wiley.

[23] ———— (1976) Probability: Beware of Falsifications. Scientia. 111: 283-303. Reprinted in Kyburg, H.E. and Smokler, H.E. *Studies in Subjective Probability*. 2nd ed. 1980. Pp. 194-224. New York: Kreiger.

[24] ———— (2008) *Philosophical lectures on Probability: Collected, Edited and Annotated by Alberto Mura*. Springer.

[25] DeGroot, M.H. and Feinberg, S.E. (1983) The Comparison and Evaluation of Forecasters. *The Statistician*. 32, 12-22.

[26] Duffie, D., Saita, L. and Wang, K. (2007) Multi-period Corporate Default Prediction With Stochastic Covariates. *Journal of Financial Economics.* 83: 635-665.

[27] Garratt, A., Lee, K., Pesaran, M.H. and Shin, Y. (2003) Forecast Uncertainties in Macroeconomic Modelling: An Application to the UK economy. *Journal of the American Statistical Association*. 98: 829-838. **(check this ref)**

[28] Gneiting, T. and Raftery, A. (2007) Strictly Proper Scoring Rules, Prediction and Estimation. *Journal of the American Statistical Association.* 102: 359-378.

[29] Gonedes, N. and Ijiri, Y. (1974) Improving Subjective Probability Assessment for Planning and Control in Team-Like Organizations. *Journal of Accounting Research*. 12: 251-269.

[30] Good, I.J. (1952) Rational Decisions. *Journal of the Royal Statistical Society*. Series B. 14: 107-114.

[31] ———— (1976) Information, Rewards, and Quasi-Utilities. In Leach, J.J. et al. (eds) *Science, Decision and Value*. Pp.115-127. Dordrecht: D. Reidel.

[32] ———— (1983) Good Thinking: *The Foundations of Probability and Its Applications*. Minneapolis: University of Minnesota Press.

[33] Granger, C.W.J. and Pesaran, M.H. (2000) Economic and Statistical Measures of Forecast Accuracy. *Journal of Forecasting*. 19: 537-560.

[34] ______ (2000b) A Decision-Based Approach to Forecast Evaluation. In Chan, W.S., Li, W.K. and Tang, H. (eds) *Statistics and Finance: An Interface*. London: Imperial College Press.

[35] Granger, C.W.J. and Pesaran, M.H. (2000) Economic and Statistical Measures of Forecast Accuracy. *Journal of Forecasting*. 19: 537-560.

[36] Grant, A., Johnstone, D.J. and Kwon, O.K. (2008) Optimal Betting Strategies for Simultaneous Games. *Decision Analysis*. 5: 10-18.

[37] Hakansson, N.H. (1971) Capital Growth and the Mean-Variance Approach to Portfolio Selection. *Journal of Financial and Quantitative Analysis.* 6: 517-557.

[38] Hanson, R. (2003) Combinatorial Information Market Design. *Information Systems Frontiers*. 5: 107-119.

[39] Hillegeist, S.A., Keating, E.K., Cram, D.P. and Lundstedt, K.G. (2004) Assessing the Probability of Bankruptcy. *Review of Accounting Studies*. 9: 5-34.

[40] Howson, C. and Urbach, P. (1993) *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Chicago: Open Court.

[41] Jaynes, E.T. (2003) *Probability Theory: The Logic of Science*. New York, NY: Cambridge University Press.

[42] Johnstone, D.J. (2007) The Value of a Probability Forecast from Portfolio Theory. *Theory and Decision*. 63: 153-203.

[43] ———— (2008) The Parimutuel Kelly Probability Scoring Rule. *Decision Analysis*. 4: 66-75.

[44] Jones, S. and Hensher, D. (2004) Predicting Firm Financial Distress: A Mixed Logit Model. *The Accounting Review*. 79: 1011-1038.

[45] Kadane, J.B. and Winkler, R.L. (1988) Separating Probability Elicitation From Utilities. *Journal of the American Statistical Association*. 83: 357-363

[46] Kelly, J. (1956) A New Interpretation of the Information Rate. *Bell System Technical Journal.* 35: 917-926.

[47] Kilgour, D.M. and Gerchak, Y. (2004) Elicitation of Probabilities Using Competitive Scoring Rules. *Decision Analysis.* 2: 108-113.

[48] Kyle, P. (1985) Continuous Auctions and Insider Trading. *Econometrica*. 53: 1315-1335.

[49] Kelly, J. (1956) A New Interpretation of the Information Rate. *Bell System Technical Journal.* 35: 917-926.

[50] Kilgour, D.M. and Gerchak, Y. (2004) Elicitation of Probabilities Using Competitive Scoring Rules. *Decision Analysis.* 2: 108-113.

[51] Levitt, S.D. (2004) Why are Betting Markets Organized So Differently From Financial Markets? *The Economic Journal.* 114: 223-246.

[52] Lambert, N. et al. (including D.Pennock) Self-Financed Wagering Mechanisms for Forecasting. *Electronic Commerce* 2008

[53] Lando, D. (2004) *Credit Default Modeling: Theory and Applications.* Princeton: Princeton University Press.

[54] Lau, A.H.L. (1987) A Five-State Financial Distress Prediction Model. *Journal of Accounting Research.* 25: 127-138.

[55] Leitch, G. and Tanner, J.E. (1981) Economic Forecast Evaluation: Profits Versus Conventional Error Measures. *American Economic Review.* 81: 580-90.

[56] ———— (1995) Professional Economic Forecasts: Are They Worth Their Costs? *Journal of Forecasting.* 14: 143-157.

[57] Lehmann, E.L. (1983) *Theory of Point Estimation.* 2nd ed. Belmont, CA: Wadsworth.

[58] Levitt, S.D. (2004) Why are Betting Markets Organized So Differently From Financial Markets? *The Economic Journal.* 114: 223-246.

[59] Li, Y. (1993) Growth-Security Investment Strategy for Long and Short Runs. *Management Science.* 39: 915-924.

[60] Lichtendahl, K.C. and Winkler, R.L. (2007) Probability Elicitation, Scoring Rules, and Competition Among Forecasters. *Management Science.* In print.

[61] Lindley, D.V. (1982) Scoring Rules and the Inevitability of Probability. *International Statistical Review.* 50: 1-26.

[62] Lopez, J.A. (2001) Evaluating the Predictive Accuracy of Models. *Journal of Forecasting.* 20: 87-109.

[63] Luenberger, D. (1998) *Investment Science.* New York: Oxford University Press.

[64] MacLean, L.C., Sanegre, R., Zhao,Y. and Ziemba, W.T. (2004) Capital Growth With Security. *Journal of Economic Dynamics and Control.* 28: 937-954.

[65] MacLean, M.C. and Ziemba, W.T. (1999) Growth Versus Security Tradeoffs in Dynamic Investment Analysis. *Annals of Operations Research.* 85: 193-225.

[66] MacLean, L.C, Ziemba, W.T. and Blazenko, G. (1992) Growth Versus Security in Dynamic Investment Analysis. *Management Science.* 38: 1562-1585.

[67] Markowitz, H.M. (1976) Investment for the Long Run: New Evidence for an Old Rule. *Journal of Finance.* 31: 1273-1286.

[68] Muradoglu, G. and Onkal, D. (1994) An Exploratory Analysis of Portfolio Managers' Probabilistic Forecasts of Stock Prices. *Journal of Forecasting.* 13: 565-578.

[69] Murphy, A.H. (1966) A Note on the Utility of Probability Predictions and the Probability Score in the Cost-Loss Ratio Decision Situation. *Journal of Applied Meteorology.* 5: 534-537.

[70] Murphy, A.H. and Winkler, R.L. (1970) Scoring Rules in Probability Assessment and Evaluation. *Acta Psychologica.* 34: 273-286.

[71] ———— (1987) A General Framework for Forecast Evaluation. *Monthly Weather Review.* 115: 1330-1338.

[72] ———— (1992) Diagnostic Verification of Probability Forecasts. *International Journal of Forecasting.* 7: 435-455.

[73] Nau, R.F. (1985) Should Scoring Rules Be 'Effective'? *Management Science.* 31: 527-535.

[74] Offerman, T., Sonnemans,J. van de Kuilen, G and Wakker,P.P. (2009) A Truth-Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes. *Review of Economic Studies.* Forthcoming.

[75] O'Hagan, A. (1994) *Bayesian Statistics. Kendall's Advanced Theory of Statistics.* Vol. 2B. Cambridge: Cambridge University Press.

[76] Ohlson, J. (1980) Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research.* 19, 109-131.

[77] Ottaviani, M. and Sorensen, P.N. (2006) The Strategy of Professional Forecasting. *Journal of Financial Economics.* 81: 441-466.

[78] Partington, G., Stevenson, M. and Yao, J. (2005) Run Length and the Predictability of Stock Price Reversals. *Accounting and Finance.* 45: 653-671.

[79] Pennock, D. et al (2002) The Real Power of Artificial Markets. Science (fix this)

[80] ———— (2004) A Dynamic PariMutuel Market for Hedging, Wagering, and Information Aggregation. Electronic Commerce.(fix this)

[81] Pesaran, M.H. and Timmerman, A. (1994) Forecasting Stock Returns: An Examination of Stock Market Trading in the Presence of Transaction Costs. *Journal of Forecasting.* 13: 330-365.

[82] ——— (1995) The Robustness and Economic Significance of Predictability of Stock Market Returns. *Journal of Finance*. 50: 1201-1228.

[83] Poundstone, W. (2005) *Fortune's Formula: The Untold Story of the Scientific Betting System that Beat the Casinos and Wall Street*. New York: Farrar, Straus and Giroux.

[84] Roll, R. (1973) Evidence on the Growth Optimum Model. *Journal of Finance*. 28: 551-567.

[85] Roulston, M.S. and Smith, L.A. (2002) Evaluating Probability Forecasts Using Information Theory. *Monthly Weather Review*. 130: 1653-1660.

[86] Rubinstein, M. (1976) The Strong Case for the Generalized Logarithmic Utility Model as the Premier Model of Financial Markets. *Journal of Finance*. 31: 551-571.

[87] Santomero, A. and Visno, J.D. (1977) Estimating the Probability of Failure for Commercial Banks and the Banking System. *Journal of Banking and Finance*. 1, 185-215.

[88] Samuelson, P. and Ziemba, W.T. (2006) Understanding the Finite properties of Kelly Log Betting: A Tale of Five Investors. **???????**

[89] Savage, L.J. (1954) *The Foundations of Statistics* New York: Wiley.

[90] ——— (1971) Elicitation of Personal Probabilities and Expectations. *Journal of the American Statistical Association*. 66: 783-801.

Scott, W.R. (1979) Scoring Rules for Probabilistic Reporting. *Journal of Accounting Research*. 17: 156-178.

[91] Selton, R. (2007) Axiomatic Characterization of the Quadratic Scoring Rule. *Experimental Economics*. 1: 43-62.

[92] Shumway, T. (2001) Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*. 74: 101-124.

[93] Stigler, S.M. (2007) The Epic Story of Maximum Likelihood. *Statistical Science*. 22: 598-620.

[94] Stutzer, M. (2003) Portfolio Choice with Endogenous Utility: A Large Deviations Approach. *Journal of Econometrics*. 116: 365-386.

[95] Thomson et al. (2004) European J of Finance 10: 290-307.

[96] Thorp, E. (1966) *Beat the Dealer*. 2nd ed. New York: Vintage.

[97] ——— (1969) Optimal Gambling Systems for Favorable Games. *International Statistical Review*. 37: 273-293.

[98] ———— (1971) Portfolio Choice and the Kelly Criterion. *Proceedings of the Business and Economics Section of the American Statistical Association*. Pp. 215-224. (Reprinted in Ziemba, W.T. and Vickson, R.G. (1975) *Stochastic Optimization Models in Finance*. Pp. 599-619. New York: Academic Press.)

[99] ———— (2000) The Kelly Criterion in Blackjack, Sports Betting and the Stock Market. In Vancura, O., Cornelius, J. and Eadington, W.R. (eds) *Finding the Edge: Mathematical Analysis of Casino Games*. Pp.163-213. Reno, NV: Institute for the Study of Gambling and Commercial Gaming.

[100] West, M. (1984) Bayesian Aggregation. *Journal of the Royal Statistical Society. Series A*. 147: 600-607.

[101] Winkler, R.L. (1967) The Quantification of Judgment: Some Methodological Suggestions. *Journal of the American Statistical Association*. 62: 1105-1120.

[102] ———— (1969) Scoring Rules and the Evaluation of Probability Assessors. *Journal of the American Statistical Association*. 64: 1073-1078.

[103] ———— (1996) Scoring Rules and the Evaluation of Probabilities (with discussion). *Test*. 5: 1-60.

[104] Winkler, R. L. and Clemen, R. T. (2004) Multiple Experts vs. Multiple Methods: Combining Correlation Assessments. *Decision Analysis*. 1: 167-176.

[105] Yates, J.F, McDaniel, L.S. and Brown, E.S. (1991) Probabilistic Forecasts of Stock Prices and Earnings: The Hazards of Nascent Expertise. *Organizational Behavior and Human Decision Processes*. 49: 60-79.

[106] Ziemba, W.T. (2005) The Symmetric Downside-Risk Sharpe Ratio. *Journal of Portfolio Management*. **Vol.????**: 108-122.