A RISK-MANAGEMENT FRAMEWORK FOR BINARY INVESTMENTS

ELLIOT CHRISTOPHERS

Master's Thesis

Stockholm School of Economics

2025

A Risk-Management Framework for Binary Investments

Abstract:

In this paper, we introduce three separate risk-management models for binary-investment markets: speculatory opportunities where the investor loses their entire position in the negative state, and gains some prespecified profit multiple in the positive state. The first is a model that maximises capital growth, when the investor's current capital is diminished due to existing unresolved investments. Second, a Value-at-Risk model, where the investor's stakes in a repeated investment scenario depend on their chosen risk tolerance. Third, a hypothesis test for probability calibration, where the investor can ensure that their forecast-generating model is suitable. We discuss more broadly the complementary nature of these models, and suggest specific conventional financial markets where this framework can be applied.

Keywords:

Risk Management, Kelly Criterion, Brier Score, Position Sizing, Hypothesis Testing

Authors:

Elliot Christophers (25487)

Tutor:

Riccardo Sabbatucci, Associate Professor, Department of Finance

Examiner:

Gualtiero Azzalini, Assistant Professor, Department of Finance

Master’s Thesis
Master’s Program in Finance
Stockholm School of Economics

1 Introduction

Investors in conventional financial markets dedicate their forecasting resources to the future development of prices; conversely, in markets of, what we call, binary investments, prices are prespecified, with forecasting instead devoted to the probabilities of the two possible states respectively occurring. With what probability will the bond issuer default? How likely is it that the Fed lowers rates? What is the likelihood the the stock's price will exceed a certain threshold? These are the types of questions that binary investments pose, when the returns in the two states are given.

The Kelly criterion (Kelly, 1956), in its most famous form, is a risk-management and stake-sizing solution, applicable to the binary-investment situation where the positive-state return and probability are known, with the entire stake being lost in the negative state. It is an attractive investment philosophy, maximising logarithmic utility and hence long-term capital development. However, in its basic form, it is limited to the aforementioned scenario only. In this paper, therefore, we develop a pair of related risk-management models: Kelly-based stake-sizing equations for two distinct and unique binary-investment situations. The first relates to scenarios where existing investments have not yet resolved, meaning that the capital which the Kelly criterion seeks to maximise is unrepresentative. The second is a Value-at-Risk model, allowing the investor to manage downside risk in repeated binary-investment scenarios.

With binary investments, the returns to each state are known, and the investor demands an accurate probabilistic forecast for the two states occurring. Therefore, although the forecaster's particular probability-generating model will vary with context, we develop a model-agnostic hypothesis test, based on the Brier score (Brier, 1950), to evaluate whether the model provides calibrated forecasts. We evaluate the extent to which the test is accurate, and consider the criteria that make it applicable.

The paper is structured as follows. We begin with a short introduction of the Kelly criterion, also presenting the broader Kelly-criterion literature, with contributions which are similar in nature to the two stake-sizing models we derive. In the subsequent two sections, we develop and evaluate first the unresolved-investment model, and then the Value-at-Risk model. After this, we present and test the hypothesis test for the calibration of probabilistic forecasts. Finally, we discuss areas of conventional finance in which our results may be applied, while justifying the wagering-framed presentation of our results throughout.

2 The Kelly Criterion and Extended Kelly Frameworks

The Kelly criterion is a famous result in the investment literature, providing, in general, a framework for optimal wealth growth, and, specifically in gambling contexts, a closed-form expression for the optimal stake to place on favourable wagers. Suppose that the probability of an event occurring is $0 < p < 1$ and that a bettor is offered a wager whereby if the event occurs they receive a profit of their stake multiplied by b, while they lose the entire stake if the event does not occur. The Kelly criterion states that the bettor should stake a fraction f

$$f = \frac{p(1+b) - 1}{b} \tag{1}$$

of their investable capital on this opportunity. The growth rate of capital is given by

$$G = (1 + fb)^p (1 - f)^{1-p} \tag{2}$$

and the fraction in (1) maximises the logarithmic growth rate

$$\ln(G) = p\ln(1+fb) + (1-p)\ln(1-f) \tag{3}$$

The derivation is rather straight-forward, differentiating (3) in f and equating with zero

$$\frac{\delta \ln(G)}{\delta f} = 0 = \frac{pb}{1+fb} - \frac{1-p}{1-f} \tag{4}$$

rearranges for f as given in (1). This is a standard result, and is, in general, a principle more than a practical tool: it is seldom that a betting scenario arises where the criterion can be applied as specified without amendment. There is therefore extensive research on extensions to the Kelly criterion, expanding it to more complicated and involved wagers. Below we briefly discuss a sample of these models, both to contextualise the Kelly criterion in the broader investment space, and to better position our contribution to this literature.

Baker and McHale (2013) consider a scenario where there is uncertainty in the probability of a wager winning, with it being distributed according to the Beta distribution (Krzysztofowicz and Long, 1991), with a mean of p and variance of σ^2. (1) assumes that p is known exactly, but if there is uncertainty in its estimation, there is risk for over-staking when using the mean estimate as input. The authors find that the optimal stake for such a scenario is given by

$$f = \frac{(p(1+b)-1)^3}{b((p(1+b)-1)^2 + (1+b)^2\sigma^2)} \tag{5}$$

where, if $\sigma = 0$, (5) simplifies to (1). Due to the concave nature of the Kelly criterion, uncertainty leads to a lower stake, with over-staking more detrimental to returns than under-staking.

Busseti et al. (2016), present a framework for a single event, with K different outcomes, where N different bets are made. In this case, $\boldsymbol{f}$ is an N-length vector of fractions for the different wagers, $\boldsymbol{p}$ is a K-length vector of probabilities for the different outcomes, and $\boldsymbol{b}$ is a $K \times N$ matrix of returns for the N wagers in each of the K outcomes. The authors offer a numerical solution to maximising capital growth in this scenario, subject to a convex risk-constraint, avoiding low-wealth outcomes, formalised as

$$\begin{aligned} \text{maximise} \quad & \sum_{i=1}^{K} \boldsymbol{p}_i \log(\boldsymbol{b}_i^T \boldsymbol{f}) \\ \text{subject to} \quad & \mathbf{1}^T \boldsymbol{f} = 1, \boldsymbol{f} \geq 0 \\ & \sum_{i=1}^{K} \boldsymbol{p}_i (\boldsymbol{b}_i^T \boldsymbol{f})^{-\lambda} \leq 1 \end{aligned} \tag{6}$$

where the first constraint is a no-short-selling constraint, and the latter is the downside-protection constraint, controlled by the risk-aversion parameter λ, where larger values for this parameter lead to stronger aversion to poor outcomes and thus more moderate staking. This framework considers a portfolio of bets (rather than merely one bet, as in (1) and (5)), optimising staking for the entire portfolio across the single event.

Sun and Boyd (2018) conceptually combine the probability-uncertainty of Baker and McHale with the multi-outcome-multi-wager with down-side protection framework of Buesseti et al. The difference here, compared to Baker and McHale, is that while the latter modeled p as following a beta distribution, the authors here instead consider the K different outcome probabilities as being drawn from a set of plausible probability distributions $\boldsymbol{P}$. With $\boldsymbol{f}$, $\boldsymbol{p}$ and $\boldsymbol{b}$ as in

Busseti, the authors produce numerical solutions for

$$\max_{\boldsymbol{f}} \min_{\boldsymbol{p} \in \boldsymbol{P}} \sum_{i=1}^{K} \boldsymbol{p}_i \log(\boldsymbol{b}_i^T \boldsymbol{f}) \tag{7}$$

maximising returns in the scenario with the worst-case probability distribution. These numerical solutions are based on the nature of $\boldsymbol{P}$. For instance, there is one suggested numerical solution for the box distribution set

$$\boldsymbol{P} = \left\{\boldsymbol{p} \mid p_i \in [\hat{p}_i - \epsilon_i, \hat{p}_i + \epsilon_i] \; \forall \; p_i \in \boldsymbol{p}, \mathbf{1}^T \boldsymbol{p} = 1\right\} \tag{8}$$

where, for any of the individual $\boldsymbol{p} \in \boldsymbol{P}$, each of the K $p_i \in \boldsymbol{p}$ are within some small value ϵ_i from their point estimate $\hat{p}_i$. A different numerical solution exists for the polyhedral set

$$\boldsymbol{P} = \left\{\boldsymbol{p} \mid \boldsymbol{A}\boldsymbol{p} \leq \boldsymbol{B}, \boldsymbol{p} \geq 0, \mathbf{1}^T \boldsymbol{p} = 1\right\} \tag{9}$$

where A is an $m \times K$ matrix and B an m-length vector, with $\boldsymbol{A}\boldsymbol{p} \leq \boldsymbol{B}$ representing m different linear constraints, which are combinations of the K different probabilities. The authors discuss further types of sets, with numerical solutions similarly provided for these.

The above three papers are all general in scope. However, the literature also offers extremely specific frameworks. Smoczynski and Tomkins (2010), for instance, derive a solution with a very narrow scope: parimutuel horse races. In conventional gambling games, as those discussed above, the returns to different outcomes b are known exactly; however, in parimutuel (or pooled) betting, the entire bet amount by all gamblers across the different outcomes is pooled, with the organisers taking their commission, and the remainder then shared by the winning bettors. The authors offer a closed-form solution for a race between n horses, where fractions f_i of existing capital are to be wagered on them, respectively:

$$f_i = p_i - \beta_i \left(\frac{\sum_{j \in S} p_j}{D - \sum_{j \in S} \beta_j} \right), i \in S \tag{10}$$

where S is the set of horses that it is optimal to bet on ($f_i = 0 \; \forall \; i \notin S$). Here, p_i is the true probability of horse i winning, β_i is the public betting proportion on horse i, and D is the dividend ratio, equaling one minus the track take (that is, the amount of the total bet pool paid back to investors after the commission). The authors provide an algorithm for whether each horse i should be included in S or not, which is iterative, first ranking the horses according to expected contribution to revenues, then adding the marginal horse to the set if it increases total profit. This paper is not as generally applicable as the others discussed here, but demonstrates the ability of the Kelly criterion (which is essentially a principle of capital-growth optimisation) to be applied to detailed, practical scenarios.

Kim (2024), in a more recent paper, considers a framework where the bettor makes wagers in good and bad states (denoted 1 and 0, respectively) with different success rates. If there are N_1 good states and N_0 bad states, with the probability of winning being p_1 in good states and p_0 in bad states, and returns to wins and losses being b_1 and b_0 in both states (that is, a won bet in both good and bad states returns b_1), then the optimal fraction of current capital is given by

$$f = \frac{p_1 N_1}{(1 + b_0)(p_1 N_1 + (1 - p_0) N_0)} - \frac{(1 - p_0) N_0}{b_1 (p_1 N_1 + (1 - p_0) N_0)} \tag{11}$$

This framework helps bettors manage risk when their model's edge is not constant: perhaps in N_1 out of $N_0 + N_1$ cases, the model has an edge ($p_1 > p_0$), while in the remaining N_0 instances it does not.

Xu (2022) considers the Kelly criterion for stock markets, in a framework which maximises the probability of reaching a certain wealth target sooner rather than later, under the worst plausible market scenario:

$$f = \frac{\mu - r}{\sigma^2}\left(1 - \frac{1}{2r}\left(\frac{(\mu - r)^2}{2\sigma^2} + \lambda + r - \sqrt{\left(\frac{(\mu - r)^2}{2\sigma^2} + \lambda - r\right)^2 + \frac{2r(\mu - r)^2}{\sigma^2}}\right)\right) \tag{12}$$

where μ is the asset drift, σ is its volatility, r is the risk-free rate, and λ is a discount factor. Adapting the Kelly criterion to the stock market is not a new endeavour (Merton, 1969; Thorp, 2008), but Xu considers an alternative property of Kelly, minimising time to reach a certain goal (see Breiman et al. (1961) below) as opposed to purely maximising capital growth.

Wu et al. (2025) is a more modern approach to Kelly-criterion investing, applying it in a machine-learning environment. The authors propose an LSTM-based architecture that predicts the daily win probability of a futures day-trading strategy. They then use this as input to the classical Kelly criterion expression, to determine the optimal stake size. Because futures require large margin desposits, traders cannot generally implement the full Kelly fraction, which leads the authors to convert their strategy into options trading with a high degree of contract granularity. Although there is no closed-form extension to the Kelly criterion, the paper is in the cutting-edge technically, using advanced machine-learning techniques, and can still leverage the basic insights of Kelly investing to generate impressive returns.

The extensions to, and broad application of, the Kelly criterion emphasise its core insights: generally, an investor should maximise capital growth. But their position-sizing should be reduced when outcome probabilities are subject to uncertainty. Similarly, when several wagers are placed on the same event, these should be optimised jointly in a portfolio, particularly when down-side risk is being managed. Detailed solutions for extremely specific scenarios can be derived, while stock markets and other asset classes also lie within the scope of the Kelly criterion. In the next two sections of this paper, we develop two additional Kelly-extension models.

3 Unresolved Wagers

As presented above, Kelly investing is done relative to an existing capital level. Consider the scenario, however, when an investment strategy finds two or more lucrative wagers in short succession, rendering the capital impact of the preceding wagers uncertain for the nth wager. In this situation, with capital reduced due to these existing stakes, regular Kelly investing would recommend a stake that is much lower than what one would reasonably place with capital fully reflecting the existing wagers. In this section, we develop and evaluate a solution to this problem, providing a contribution with similar practical relevance for bettors to the other elaborated Kelly models presented above, where bettors can adjust their staking to maximise capital growth despite their effective capital being unknown.

3.1 One Unresolved Wager

Consider the case where the current capital has been diminished because there exists an outstanding, unresolved wager: we staked s_1 on a bet with probability p_1 and returns to a win of b_1. In this case, staking the regular Kelly fraction of our current capital will be misleading, as our effective capital (accounting for some expected return from the outstanding bet) is higher. Therefore, we consider the growth rate of our current capital w with respect to the existing bet as well as the new bet with p_2 and b_2, where we stake a fraction f of w: that is, $s_2 = fw$. We have

$$G = \left(1 + fb_2 + \frac{s_1}{w}(1+b_1)\right)^{p_1 p_2} (1+fb_2)^{(1-p_1)p_2} \left(1 - f + \frac{s_1}{w}(1+b_1)\right)^{p_1(1-p_2)} (1-f)^{(1-p_1)(1-p_2)} \tag{13}$$

Simplifying the notation by substituting $u = \frac{s_1}{w}(1+b_1)$, we can take the logarithm

$$\ln(G) = p_1 p_2 \ln(1+fb_2+u) + (1-p_1)p_2 \ln(1+fb_2) + p_1(1-p_2)\ln(1-f+u) + (1-p_1)(1-p_2)\ln(1-f) \tag{14}$$

and then differentiate with respect to f

$$\begin{aligned}
&\frac{d\ln(G)}{df} = 0 = \frac{p_1 p_2 b_2}{1+fb_2+u} + \frac{(1-p_1)p_2 b_2}{1+fb_2} - \frac{p_1(1-p_2)}{1-f+u} - \frac{(1-p_1)(1-p_2)}{1-f} = \\
&p_1 p_2 b_2 (1+fb_2)(1-f+u)(1-f) + \\
&(1-p_1)p_2 b_2 (1+fb_2+u)(1-f+u)(1-f) - \\
&p_1(1-p_2)(1+fb_2+u)(1+fb_2)(1-f) - \\
&(1-p_1)(1-p_2)(1+fb_2+u)(1+fb_2)(1-f+u)
\end{aligned} \tag{15}$$

which rearranges for a cubic equation in f

$$\begin{aligned}
&\frac{d\ln(G)}{df} = 0 = Af^3 + Bf^2 + Cf + D \implies \\
&f = \varepsilon^2 \sqrt[3]{-\frac{Q}{2} + \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3}} + \varepsilon \sqrt[3]{-\frac{Q}{2} - \sqrt{\left(\frac{Q}{2}\right)^2 + \left(\frac{P}{3}\right)^3}} - \frac{B}{3A} \\
&\text{where} \\
&P = \frac{3AC - B^2}{3A^2} \\
&Q = \frac{2B^3 - 9ABC + 27A^2 D}{27A^3} \\
&\varepsilon = \frac{-1 + i\sqrt{3}}{2} \\
&A = b_2^2 \\
&B = b_2^2 \left(u(p_1 - p_1 p_2 - 1) - p_2 - 1\right) + b_2 \left(u(1 - p_1 p_2) - p_2 + 2\right) \\
&C = b_2^2 p_2 (u+1) + b_2 \left(u(p_1 u + 2p_1 - u - 3) - 2\right) + u(1-p_2) - p_2 + 1 \\
&D = u^2 \left(-b_2 p_1 p_2 + b_2 p_2 - p_1 p_2 + p_1 + p_2 - 1\right) + \\
&u\left(-b_2 p_1 p_2 + 2b_2 p_2 - p_1 p_2 + p_1 + 2p_2 - 2\right) + p_2(b_2+1) - 1 \\
&u = \frac{s_1}{w}(1+b_1)
\end{aligned} \tag{16}$$

where the chosen combination of ε is that which selects the correct root (among the three conventionally available roots in the qubic equation).

To evaluate the accuracy of the formula, we perform a simulation where we over 100 iterations draw p_1 and p_2 between 0.1 and 0.9, and two values for $d = \ln\left(\frac{1}{1-p}\right) - \ln\left(\frac{1}{1-\frac{1}{1+b}}\right)$ (a distance metric proposed as suitable in Law and Peel (2002)) between 0.001 and 0.025, from which b_1 and b_2 can be derived. With, $s_1 = \frac{p_1(1+b_1)-1}{b_1}$, we have $w = 1 - s_1$. We thus calculate

f using (16). Then, we simulate over 500 different values of f, where, for each one, the outcome of the two bets is simulated 500,000 times. We take the geometric mean of the growth of w across these 500,000 iterations, thus getting 500 (f, G) points. We fit a cubic to these 500 points (a quadratic would likely suffice, but the cubic provides more flexibility to fit the strictly concave function) and select the value of f which maximises this continuous function, denoting this optimal parameter $\hat{f}$. Thus, for each of the 100 iterations, we have the calculated value for f and the simulated value $\hat{f}$. If our expression for f is correct, then these two values will be strongly correlated. The upper panel of Figure 1 shows this relation. Similarly, we also find the optimal value for f, denoted f^*, using numerical maximisation of (13). The lower panel in Figure 1 demonstrates that the fit of our solution in (16) is perfectly calibrated to the numerical optimum, with the relationship between f and $\hat{f}$ less strong. Indeed, the correlations (Pearson, 1895) $\text{corr}(f, \hat{f}) = 0.9635$ and $\text{corr}(f, f^*) = 1.0000$ suggest that the simulation method, while indeed highly calibrated in its own right, is not as attuned as is the optimisation, or even our solution f, with the 500,000 iterations insufficient to find the correct value for $\hat{f}$.

3.2 Multiple Unresolved Wagers

If there are $n-1$ outstanding bets, and we want to optimise the Kelly fraction f for the nth, there will then be 2^n possible outcomes, with each bet either winning or losing. For $n = 2$, we obtained a third-degree polynomial for f. In general, for $n \geq 2$, the polynomial will be of degree $2^n - 1$. Therefore, we create an approximate formula for any $n \geq 3$ (for $n = 2$, obviously use the above formula). Thus, enumerate the 2^n combinations with indices i. For each of the n bets, enumerated j, we can create indicator functions $I_{i,j}(u_1, u_2)$, returning some value u_1 if the jth bet wins in the ith combination and some other value u_2 if it loses. Each bet j has stake s_j (other than the nth bet), probability p_j and profit to a won bet of b_j. Current capital is w and we want to optimise for the fraction f of this capital which we will stake on the nth bet. With respect to w, our growth rate G is

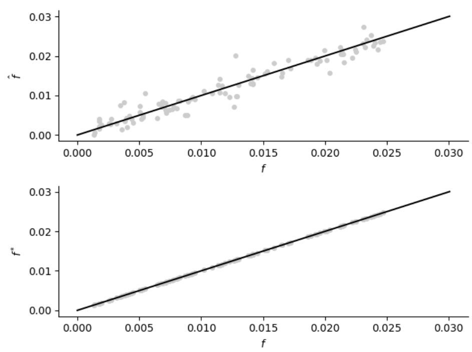

Figure 1: Our Kelly fraction f versus the simulated value $\hat{f}$ as estimated in a Monte Carlo simulation (upper panel) and the optimal value f^* (lower panel). For the betting scenario where $n = 2$.

$$G = \prod_{i=1}^{2^n} \left(1 + f I_{i,n}(b_n, -1) + \sum_{j=1}^{n-1} I_{i,j} \left(\frac{s_j}{w}(1 + b_j), 0 \right) \right)^{\prod_{j=1}^{n} I_{i,j}(p_j, 1-p_j)} \tag{17}$$

the geometric mean return across the growth of capital in the 2^n different combinations of bet outcomes. We can take the logarithm to get

$$\ln(G) = \sum_{i=1}^{2^n} \left(\ln \left(1 + f I_{i,n}(b_n, -1) + \sum_{j=1}^{n-1} I_{i,j} \left(\frac{s_j}{w}(1 + b_j), 0 \right) \right) \prod_{j=1}^{n} I_{i,j}(p_j, 1 - p_j) \right) \tag{18}$$

and if we differentiate this with respect to f, we get

$$\frac{d\ln(G)}{df} = 0 = \sum_{i=1}^{2^n} \frac{I_{i,n}(b_n,-1)\prod_{j=1}^{n} I_{i,j}(p_j,1-p_j)}{1+fI_{i,n}(b_n,-1)+\sum_{j=1}^{n-1} I_{i,j}\left(\frac{s_j}{w}(1+b_j),0\right)} =$$
$$\sum_{i=1}^{2^n} I_{i,n}(b_n,-1)\prod_{j=1}^{n}\left[I_{i,j}(p_j,1-p_j)\right]\prod_{k\neq i}^{2^n}\left[1+fI_{k,n}(b_n,-1)+\sum_{j=1}^{n-1} I_{k,j}\left(\frac{s_j}{w}(1+b_j),0\right)\right] \tag{19}$$

by multiplying up the denominators. We can approximate this product term across $k \neq i$ by expanding the Taylor series to the second-degree term of $\ln(1+a_{1,k}f+a_{2,k})$ about $f = f_0$, where f_0 is some value which we believe will be close to the solution, for instance, $f_0 = \frac{p_n(1+b_n)-1}{b_n}$, the regular Kelly criterion.

$$\prod_{k\neq i}^{2^n}\left[1+fI_{k,n}(b_n,-1)+\sum_{j=1}^{n-1} I_{k,j}\left(\frac{s_j}{w}(1+b_j),0\right)\right] =$$
$$\exp\left(\sum_{k\neq i}^{2^n}\ln\left(1+a_{1,k}f+a_{2,k}\right)\right) \approx$$
$$\exp\left(\sum_{k\neq i}^{2^n}\left[\ln(1+a_{1,k}f_0+a_{2,k}) - \frac{a_{1,k}f_0}{1+a_{1,k}f_0+a_{2,k}} - \frac{a_{1,k}^2 f_0^2}{2(1+a_{1,k}f_0+a_{2,k})^2} + \right.\right.$$
$$\left.\left.\left(\frac{a_{1,k}}{1+a_{1,k}f_0+a_{2,k}} + \frac{a_{1,k}^2 f_0}{(1+a_{1,k}f_0+a_{2,k})^2}\right)f - \frac{a_{1,k}^2}{2(1+a_{1,k}f_0+a_{2,k})^2}f^2\right]\right)$$
$$= \exp\left(\sum_{k\neq i}^{2^n}\ln(1+a_{1,k}f_0+a_{2,k}) - \sum_{k\neq i}^{2^n}\left[\frac{a_{1,k}f_0}{1+a_{1,k}f_0+a_{2,k}} + \frac{a_{1,k}^2 f_0^2}{2(1+a_{1,k}f_0+a_{2,k})^2}\right] + \right.$$
$$\left. f\sum_{k\neq i}^{2^n}\left[\frac{a_{1,k}}{1+a_{1,k}f_0+a_{2,k}} + \frac{a_{1,k}^2 f_0}{(1+a_{1,k}f_0+a_{2,k})^2}\right] - \frac{f^2}{2}\sum_{k\neq i}^{2^n}\frac{a_{1,k}^2}{(1+a_{1,k}f_0+a_{2,k})^2}\right) =$$
$$\prod_{k\neq i}^{2^n}\left[\frac{1+a_{1,k}f_0+a_{2,k}}{\exp\left(\frac{a_{1,k}f_0}{1+a_{1,k}f_0+a_{2,k}} + \frac{a_{1,k}^2 f_0^2}{2(1+a_{1,k}f_0+a_{2,k})^2}\right)}\right]\exp\left(a_{3,i}f - a_{4,i}f^2\right)$$

with

$$a_{1,k} = I_{k,n}(b_n,-1),\ a_{2,k} = \sum_{j=1}^{n-1} I_{k,j}\left(\frac{s_j}{w}(1+b_j),0\right),$$
$$a_{3,i} = \sum_{k\neq i}^{2^n}\left[\frac{a_{1,k}}{1+a_{1,k}f_0+a_{2,k}} + \frac{a_{1,k}^2 f_0}{(1+a_{1,k}f_0+a_{2,k})^2}\right],\ a_{4,i} = \frac{1}{2}\sum_{k\neq i}^{2^n}\frac{a_{1,k}^2}{(1+a_{1,k}f_0+a_{2,k})^2} \tag{20}$$

We can similarly further expand $\exp\left(a_{3,i} f - a_{4,i} f^2\right)$ about $f = f_0$, finding

$$\begin{aligned}
&\exp\left(a_{3,i} f - a_{4,i} f^2\right) \approx \\
&\exp\left(a_{3,i} f_0 - a_{4,i} f_0^2\right) \left[1 + (f - f_0)(a_{3,i} - 2a_{4,i} f_0) + \frac{(f-f_0)^2}{2}\left((a_{3,i} - 2a_{4,i} f_0)^2 - 2a_{4,i}\right)\right] = \\
&\exp\left(a_{3,i} f_0 - a_{4,i} f_0^2\right) \left[1 - a_{5,i} f_0 + a_{6,i} f_0^2 + (a_{5,i} - 2a_{6,i} f_0) f + a_{6,i} f^2\right]
\end{aligned}$$

with

$$a_{5,i} = a_{3,i} - 2a_{4,i} f_0,\ a_{6,i} = \frac{1}{2} a_{5,i}^2 - a_{4,i} \tag{21}$$

Putting everything together, then

$$\begin{aligned}
&\frac{d\ln(G)}{df} = 0 = \sum_{i=1}^{2^n} I_{i,n}(b_n, -1) \prod_{j=1}^{n} \left[I_{i,j}(p_j, 1-p_j)\right] \prod_{k\neq i}^{2^n} \left[1 + f I_{k,n}(b_n, -1) + \sum_{j=1}^{n-1} I_{k,j}\left(\frac{s_j}{w}(1+b_j), 0\right)\right] \approx \\
&\sum_{i=1}^{2^n} I_{i,n}(b_n, -1) \prod_{j=1}^{n} \left[I_{i,j}(p_j, 1-p_j)\right] \prod_{k\neq i}^{2^n} \left[\frac{1 + a_{1,k} f_0 + a_{2,k}}{\exp\left(\frac{a_{1,k} f_0}{1 + a_{1,k} f_0 + a_{2,k}} + \frac{a_{1,k}^2 f_0^2}{2(1 + a_{1,k} f_0 + a_{2,k})^2}\right)}\right] \cdot \\
&\exp\left(a_{3,i} f_0 - a_{4,i} f_0^2\right) \left[1 - a_{5,i} f_0 + a_{6,i} f_0^2 + (a_{5,i} - 2a_{6,i} f_0) f + a_{6,i} f^2\right]
\end{aligned} \tag{22}$$

We have a quadratic equation in f,

$$\sum_{l=0}^{2} A_l f^l \approx 0 \implies f \approx f(p_1, \ldots, p_n, b_1, \ldots, b_n, s_1, \ldots, s_{n-1}, w, f_0) = \frac{-A_1 - \sqrt{A_1^2 - 4A_2 A_0}}{2A_2}$$

with

$$\begin{aligned}
&A_l = \sum_{i=1}^{2^n} \left[B_{i,l} \exp\left(a_{3,i} f_0 - a_{4,i} f_0^2\right) I_{i,n}(b_n, -1) \prod_{j=1}^{n} \left[I_{i,j}(p_j, 1-p_j)\right] \cdot \right. \\
&\left. \prod_{k\neq i}^{2^n} \left[\frac{1 + a_{1,k} f_0 + a_{2,k}}{\exp\left(\frac{a_{1,k} f_0}{1 + a_{1,k} f_0 + a_{2,k}} + \frac{a_{1,k}^2 f_0^2}{2(1 + a_{1,k} f_0 + a_{2,k})^2}\right)}\right]\right] \\
&B_{i,0} = 1 - a_{5,i} f_0 + a_{6,i} f_0^2,\ B_{i,1} = a_{5,i} - 2a_{6,i} f_0,\ B_{i,2} = a_{6,i} \\
&a_{1,k} = I_{k,n}(b_n, -1),\ a_{2,k} = \sum_{j=1}^{n-1} I_{k,j}\left(\frac{s_j}{w}(1+b_j), 0\right), \\
&a_{3,i} = \sum_{k\neq i}^{2^n} \left[\frac{a_{1,k}}{1 + a_{1,k} f_0 + a_{2,k}} + \frac{a_{1,k}^2 f_0}{(1 + a_{1,k} f_0 + a_{2,k})^2}\right],\ a_{4,i} = \frac{1}{2} \sum_{k\neq i}^{2^n} \frac{a_{1,k}^2}{(1 + a_{1,k} f_0 + a_{2,k})^2}, \\
&a_{5,i} = a_{3,i} - 2a_{4,i} f_0,\ a_{6,i} = \frac{1}{2} a_{5,i}^2 - a_{4,i} \\
&I_{i,j}(u_1, u_2) = u_1 x_{i,j} + u_2(1 - x_{i,j}),\ \{x_{i,j} : \{x_i : i = 1, 2, \ldots, 2^n\} = \{0,1\}^n\} \\
&f_0 = \frac{p_n(1+b_n) - 1}{b_n} \cdot \frac{w + \sum_{j=1}^{n-1} s_j p_j (1+b_j)}{w}
\end{aligned} \tag{23}$$

where the conventional $\pm$ is replaced by $-$ to give the correct root. We have found that the given expression for f_0 is useful, multiplying the regular Kelly criterion with the expected arithmetic growth of wealth from the existing bets.

Repeating the simulation in Figure 1, only now drawing n as a random integer from three to seven, inclusive, Figure 2 shows, again, the strong correlation between the analytical value for f from (23) and the numerical value $\hat{f}$ and optimal value f^*, respectively, with coefficients 0.9813 and 1.0000. The values for f^* come from, again numerically, maximising (17).

The two main properties which make the Kelly criterion an optimal investment criterion were established by Breiman et al. (1961) (with extended discussion in Thorp (1975)). If we let the random variable $W_m(f)$ be the wealth after m repeated gambles with investment strategy f, and $M_w(f)$ be the number of trials m for $W_m(f) > w$ (which is also a random variable, since different investment histories will give different W_m), then if f^* is the optimal strategy in (23) and f' is some "essentially different" strategy such that $G(f^*) > G(f')$, where G is the capital growth rate from (2), (13) and (17) (depending on scenario), then

$$\begin{aligned} &\lim_{m\to\infty} \frac{W_m(f^*)}{W_m(f')} = \infty \\ &\lim_{w\to\infty} \Pr(M_w(f^*) < M_w(f')) = 1 \end{aligned} \tag{24}$$

that is, f^* maximises the rate of growth of wealth, and it minimises the expected time required to reach a certain level of wealth.

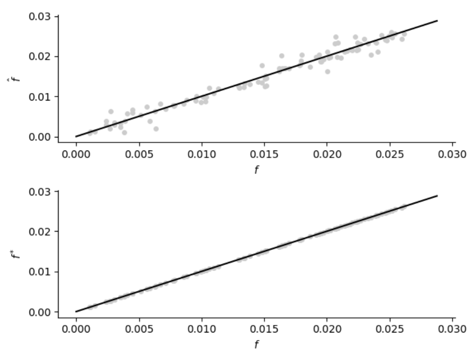

Figure 2: Our Kelly fraction f versus the simulated value $\hat{f}$ as estimated in a Monte Carlo simulation (upper panel) and the optimal value f^* (lower panel). For the betting scenario where $n \in \{3, 4, 5, 6, 7\}$, chosen randomly for each of the 100 iterations.

To evaluate whether our f^* solution in (23) is better than the regular Kelly criterion f', we perform the following simulation. For three distinct values of n ($3, 5, 7$), in turn, we perform 100 iterations of 10,000 wagers. For each of the 10,000 wagers in each iteration, we create $n-1$ outstanding bets, drawing $p \sim U(0.1, 0.9)$ and $d \sim U(0.001, 0.025)$, from which we can rearrange for b in $d = \ln\left(\frac{1}{1-p}\right) - \ln\left(\frac{1}{1-\frac{1}{1-b}}\right)$. We calculate the value for f which give growth rates $G(f) = 1, f \neq 0$, with G here from (2). There does not exist an algebraic solution, but we find this root using the solution in Brent (2013). On these $n-1$ unresolved bets, we stake these generated values for f, meaning that on average the wealth will not grow due to these outstanding bets, meaning their average impact on wealth is isolated to the capital effect in calculating f^* and f'. Then with these $n-1$ bets outstanding, we similarly draw p and b, but instead calculate f^* from (23) and f' from (1). Throughout the iteration, we keep a running track of capital w_{f^*} and $w_{f'}$, with the only difference between the two being the choice of f (that is, the outstanding wagers relative to wealth and outcomes to wagers are the same in both cases). We consider the values 100, 1,000 and 10,000 for m and 1.1, 2 and 10 for w (representing 10%, 100% and 1,000% capital growth from the initial unit wealth), and whenever the wager count passes one of these values for m, or either running wealth passes w, we note the relevant quantity. At the end of each of the 100 iterations, we take the median wealth ratio for each value of m for $\frac{W_m(f^*)}{W_m(f')}$ and similarly calculate $\Pr(M_w(f^*) < M_w(f'))$. Table 1 displays these results.

	$\frac{W_m(f^*)}{W_m(f')}$				$\Pr(M_w(f^*) < M_w(f'))$			
n	m	100	1,000	10,000	w	1.1	2	10
3		1.09	2.42	7,035.10		0.94	1.00	1.00
5		1.15	4.17	1,786,187.00		0.94	1.00	1.00
7		1.21	7.26	356,674,317.96		0.86	1.00	1.00

Table 1: Comparison of the optimal strategy f^* from (23) to the general Kelly-criterion solution f', for three different values of n. The ratios of W_m are medians across 100 iterations.

We see that, for all values of n, when m increases, so does the ratio $\frac{W_m(f^*)}{W_m(f')}$, confirming the first Breiman property, that the optimal solution maximises the growth of wealth (compared to the benchmark standard Kelly criterion). The relationship in n is interesting, with the ratio increasing in n, with m held constant. This suggests that, with more wagers unresolved, the optimal solution has an incrementally stronger edge over the benchmark method, essentially wagering higher stakes. Similarly, as w increases, so does $\Pr(M_w(f^*) < M_w(f'))$, demonstrating the second Breiman property, that the optimal solution minimises the time taken to achieve a certain wealth threshold (again compared to the benchmark). As n increases, with $w = 1.1$, the proportion of iterations where the optimal strategy reaches the wealth threshold first decreases slightly. With six outstanding wagers ($n = 7$), and with the threshold rather low (10% cumulative returns), one could imagine that in certain iterations the wealth differences up to the threshold due to the strategies varying is minimal. That the optimal strategy, then, still reaches the threshold first in 94 or 86 percent of iterations illustrates the stark overall outperformance, even in this smaller-sample scenario.

3.3 Computational Note

In evaluating (23), we have limited ourselves to $n \leq 7$. This is due to computational issues when n becomes large, with, for instance, the product

$$\prod_{k \neq i}^{2^n} \left[\frac{1 + a_{1,k} f_0 + a_{2,k}}{\exp\left(\frac{a_{1,k} f_0}{1 + a_{1,k} f_0 + a_{2,k}} + \frac{a_{1,k}^2 f_0^2}{2(1 + a_{1,k} f_0 + a_{2,k})^2} \right)} \right] \tag{25}$$

easily diverging in n. There are cases when the expression works well even for $n = 10$, but in general $n \leq 7$ is the safest range. This does not necessarily mean that (23) does not work, mathematically, for larger n, merely that our computing capabilities were not able to cope. The prospective user can readily evaluate the extent to which they can handle these larger terms.

3.4 Conclusion

In this section, we have solved a Kelly investment problem for games with unresolved wagers, optimising stakes for the nth wager when the outcome of $n-1$ existing bets are yet to be decided. Kelly investment is made relative to the investor's existing capital, meaning regular Kelly stakes calculated relative to the capital, which excludes these existing wagers, are understated. We first calculated a closed-form expression for the case where $n = 2$ (that is, one outstanding wager), demonstrating its calibration. We then generalised this expression for any value of n, which we showed remained calibrated for moderate values. Finally, we evaluated the investing capabilities of our expressions, compared to regular Kelly staking, by considering Brieman's conditions, finding that our solutions indeed maximise wealth and minimise the time to achieve a certain wealth objective, compared to the benchmark strategy.

4 Value-at-Risk

Value-at-Risk (Jorion, 2007) is an immensely popular risk-management framework amongst financial practitioners (Diebold et al., 2010), whereby an investment is made such that there is a certain theoretical probability of losses exceeding a predefined threshold over a given investment horizon. Used primarily with respect to conventional financial investments, in this section we develop an expression for Value-at-Risk staking for a repeated gambling game. A bettor can wager a certain amount of their capital on each wager, which is repeated a given number of times, and they seek to maximise growth of capital subject to their risk tolerance, in the form of Value-at-Risk. We evaluate our expression, demonstrating some practical limitations, and then expand the analysis to consider a related problem, where a gambler seeks to maintain a certain Value-at-Risk over a collection of distinct wagers, rather than a repetition of the same gamble. We finally discuss the inherent riskiness of binary securities and the suitability of Value-at-Risk as an associated metric.

4.1 Derivation

We have a gamble on an event with probability p of occurring, offering profit to a winning bet of b. If we place this bet n times, winning w times, our return β is

$$\beta = (1 + fb)^w (1 - f)^{n-w} - 1 \tag{26}$$

where f is the Kelly fraction: the proportion of current capital staked on each bet. We want to model, after these n bets have been placed, the distribution of the return β. Since bets are independent and p is constant, we get a binomial distribution for the number of wins, $w \sim \mathrm{B}(n, p)$. We can rearrange (26) to obtain

$$w_\beta = \frac{\ln(1 + \beta) - n \ln(1 - f)}{\ln(1 + fb) - \ln(1 - f)} \tag{27}$$

For a given bet with Kelly fraction f, probability p and profit b, made n times, given a specified return of β, equation (27) gives the required number of won bets to achieve this return, w_β. Thus, to calculate the probability of cumulative returns less than or equal to β, we have that $P(B \leq \beta) = P(w \leq w_\beta)$, where B is the stochastic variable for returns after these n bets, and w is the binomially-distributed stochastic variable. Hence,

$$P(B \leq \beta) = \sum_{i=0}^{\lfloor w_\beta \rfloor} \binom{n}{i} p^i (1 - p)^{n-i} \tag{28}$$

where $\lfloor w_\beta \rfloor$ is the floor of w_β, the greatest integer less than or equal to it. To avoid this rounding of w_β, assuming that the number of bets n under consideration is large, we can use a normal approximation to the binomial, giving

$$P(B \leq \beta) = \phi \left(\frac{w_\beta - np}{\sqrt{np(1 - p)}} \right) \tag{29}$$

where ϕ is the cumulative distribution function of the standard normal distribution, and np and $\sqrt{np(1-p)}$ are the mean and standard deviation, respectively, of w. Since w_β is continuous, there is no need for a continuity correction. In general, the cumulative distribution function integral is

$$\phi(x) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{30}$$

where erf is the error function

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{31}$$

and, in our case,

$$\frac{x}{\sqrt{2}} = \frac{w_\beta - np}{\sqrt{2np(1-p)}} = u \tag{32}$$

with u a simplifying substitution. Our optimisation problem is to find a Kelly fraction f such that we have a probability of at most α of experiencing cumulative returns of β or worse. Thus,

$$P(B \leq \beta) = \frac{1}{2}\left(1 + \text{erf}(u)\right) = \alpha \implies \text{erf}(u) = 2\alpha - 1 \tag{33}$$

A standard approximation for erf(u) (Winitzki, 2003; Zeng and Chen, 2015) is

$$\text{erf}(u) \approx \text{sgn}(u)\sqrt{1 - \exp\left(-u^2 \frac{\frac{4}{\pi} + au^2}{1 + au^2}\right)} \tag{34}$$

where $a = \frac{8(\pi-3)}{3\pi(4-\pi)}$. Substituting in $2\alpha - 1$ from equation (33), equation (34) can be rearranged to a biquartic equation in u, as

$$\begin{aligned}
&Au^4 + Bu^2 + C = 0 \\
&\text{with} \\
&A = a \\
&B = \frac{4}{\pi} + AC \\
&C = \ln(1 - (2\alpha - 1)^2) \\
&\text{giving} \\
&u = \pm\sqrt{\frac{-B \pm \sqrt{B^2 - 4AC}}{2A}}
\end{aligned} \tag{35}$$

which depends entirely on α as an input parameter. We, thus, can find the number of wins w_α equivalent to the α-th percentile of the distribution of w. Given the domain $\alpha \in [0, 1]$, we can reason about the $\pm$ in the expression for u. Over this domain, C is negative, the discriminant $\Delta = B^2 - 4AC$ is strictly positive, and $\sqrt{\Delta} \geq B$. Therefore, the inner $\pm$ must be $+$, given that we want real solutions. For the outer $\pm$, from (32), assuming that this risk-management system is applied for minimising the probability of losses, we get that $u < 0$, meaning that the outer $\pm$ is $-$. Thus,

$$\begin{aligned}
&u = \frac{w_\alpha - np}{\sqrt{2np(1-p)}} = -\sqrt{\frac{-B + \sqrt{B^2 - 4AC}}{2A}} \\
&\implies \\
&w_\alpha = np - \sqrt{np(1-p)\frac{-B + \sqrt{B^2 - 4AC}}{A}} = \frac{\ln(1+\beta) - n\ln(1-f)}{\ln(1+fb) - \ln(1-f)}
\end{aligned} \tag{36}$$

with the second equality coming from equating w_α with w_β from (27).

We now want to isolate f for a closed-form expression for the Kelly fraction given a desire to have at most a probability of α of seeing losses of β or worse through n repeated bets. This cannot be done directly; however, using the fact that for small x, $\ln(1+x) \approx x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots$, and $\ln(1-x) \approx -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots$, via Maclaurin series, we can approximate $\ln(1-f) \approx -f - \frac{f^2}{2}$ and $\ln(1+fb) \approx fb - \frac{f^2b^2}{2}$, taking only the second-degree expansion. We thus get

$$w_\alpha \approx \frac{\ln(1+\beta) + nf(1+\frac{f}{2})}{fb(1-\frac{fb}{2}) + f(1+\frac{f}{2})} \tag{37}$$

which rearranges to the quadratic in f,

$$\begin{aligned}
&Af^2 + Bf + C = 0\\
&\text{with}\\
&A = \frac{n}{2} + \frac{w_\alpha}{2}(b^2-1)\\
&B = n - w_\alpha(1+b)\\
&C = \ln(1+\beta)\\
&\text{giving}\\
&f = \frac{w_\alpha(1+b) - n \pm \sqrt{(n - w_\alpha(1+b))^2 - 2\ln(1+\beta)(n + w_\alpha(b^2-1))}}{n + w_\alpha(b^2-1)}
\end{aligned} \tag{38}$$

where, given that $n > w_\alpha(1+b)$ in most cases and that $f > 0$, we can replace the $\pm$ with $+$.

Thus, we get the final closed-form expression for f:

$$\begin{aligned}
&f = \frac{w_\alpha(1+b) - n + \sqrt{(n - w_\alpha(1+b))^2 - 2\ln(1+\beta)(n + w_\alpha(b^2-1))}}{n + w_\alpha(b^2-1)}\\
&w_\alpha = np - \sqrt{np(1-p)\frac{-B + \sqrt{B^2 - 4AC}}{A}}\\
&A = \frac{8(\pi-3)}{3\pi(4-\pi)}\\
&B = \frac{4}{\pi} + AC\\
&C = \ln(1 - (2\alpha-1)^2)
\end{aligned} \tag{39}$$

which takes arguments p, b, n, α, β.

4.2 Single-Gamble Evaluation

To evaluate the accuracy of our formula, we perform simulations. For specified values of the input parameters, we calculate f and then create a $100,000 \times n$ matrix of random probabilities $q \sim \mathrm{U}(0,1)$. For each element in this matrix, we compare q to p, and in the corresponding cell in another $100,000 \times n$ matrix, we insert $1 + fb$ if $q \leq p$ and $1 - f$ otherwise. We then calculate the cumulative product along the rows of this second matrix, giving 100,000 final returns from the final column, finding the proportion $\hat{\alpha}$ of which are lower than β and comparing this value to α. We do this 1,000 times, drawing $p \sim \mathrm{U}(0.1, 0.9)$; the bet expected value $EV \sim \mathrm{U}(0.005, 0.05)$, giving b from $EV = p(1+b) - 1$; and $n = 10^u$, where $u \sim \mathrm{U}(2,3)$, and n is rounded to the nearest integer. We similarly have $\alpha \sim \mathrm{U}(0.01, 0.1)$ and $\beta \sim \mathrm{U}(-0.1, -0.01)$.

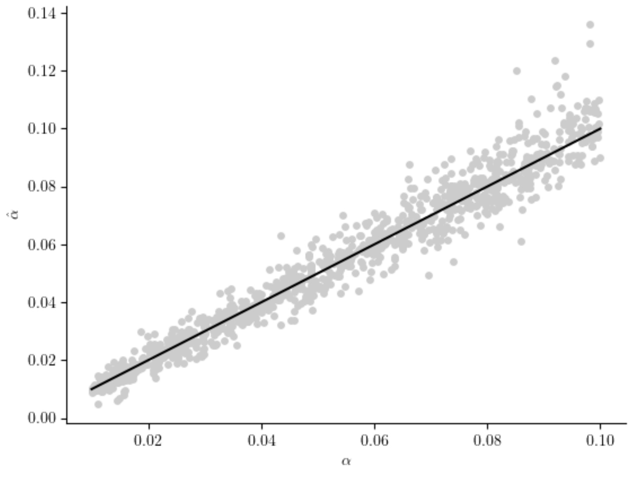

Figure 3: For a gamble with parameterisation n, p, b, α, β, we compare α to the proportion of runs with final returns less than or equal to β. The black line is $\hat{\alpha} = \alpha$, not a trend line, with our expression therefore an excellent fit.

Worthy of note is the discrete nature of $\hat{\alpha}$. In (39), we use the analytical value for the constant $A = \frac{8(\pi-3)}{3\pi(4-\pi)}$. If we instead investigate how $\hat{\alpha}$ varies for a fixed set of parameters (p, b, n, α, β) when we change the constant A, we find that $\hat{\alpha}$ does not respond to small changes in A. It instead displays a discontinuous stepwise relationship, shifting to a lower level when A increases sufficiently. This makes intuitive sense. The number of bets won in an iteration is determined solely by p and n in a binomial distribution. Since the cumulative multiplication process of Kelly capital is commutative, the sequencing of won and lost bets does not matter. Therefore, for a given value of f, we can imagine the range of won bets $w \in [0, \lfloor w_\beta \rfloor]$ which gives returns less than β, and then the range of won bets $[\lceil w_\beta \rceil, n]$ which does not. With this distribution of wins unchanged, small movements in f occasioned by nudges to A will change cumulative returns, but not sufficiently to shift the value of w_β into the domain between the next pair of integers. Only when this occurs, when one more or less won bet is required to fall below β, does $\hat{\alpha}$ shift. We can therefore conclude that $\hat{\alpha}$ can never necessarily equal α exactly, as it is binary not continuous. In the simulation studies, then, we cannot expect these $\hat{\alpha}$ to fit the α perfectly as the inputs vary, but instead hope to see a strong relationship between them, sensitive to changes in all inputs. As such, the results in Figure 3 are particularly impressive given this property.

n	100	250	500	1000	
α	0.01	0.025	0.05	0.1	
β	-0.01	-0.025	-0.05	-0.1	
(p, b)	(0.1,9.5)	(0.25,3.15)	(0.5, 1.05)	(0.75, 0.35)	(0.9,0.115)

Table 2: Parameter values for n, α, β, (p, b) in the simulation, giving $4 \cdot 4 \cdot 4 \cdot 5 = 320$ combinations.

We perform the simulation outlined above for all combinations of the parameters shown in Table 2. Across these 320 iterations, corr$(\alpha, \hat{\alpha}) = 0.981$, with a mean absolute percentage error of $\hat{\alpha}$ with respect to α of 13.1%, impressive given that $\hat{\alpha}$ is discrete. For $\alpha = 0.01$, the mean absolute error was 0.00203, with values of 0.00332, 0.00598 and 0.00668 for $\alpha = 0.025$, 0.05, and 0.1 respectively. Additionally, 48.75% of iterations saw $\hat{\alpha} < \alpha$, close enough to the expected 50%. Dissecting further, if we use the notation $P(\hat{\alpha} < \alpha \mid n)$, for instance, to represent this proportion for those iterations with a specific value of n, we find that $P(\hat{\alpha} < \alpha \mid n) = P(\hat{\alpha} < \alpha)\ \forall\ n$, $P(\hat{\alpha} < \alpha \mid \alpha) = P(\hat{\alpha} < \alpha)\ \forall\ \alpha$ and $P(\hat{\alpha} < \alpha \mid \beta) = P(\hat{\alpha} < \alpha)\ \forall\ \beta$, that is, there are no significant trends in the data suggesting that the levels of n, α and β, respectively, influence the calibration of the calculation of f to the desired risk level.

However, while the correlation between $\hat{\alpha}$ and α remains almost perfect when controlling for p; at low probabilities, $\hat{\alpha}$ is lower than α, while the opposite is true at high values of p. This suggests that f is exaggerated for large probabilities and understated for low-probability bets. However, this is unsurprising. As outlined above, with $\hat{\alpha}$ being discrete, when considering the ranges $[0, \lfloor w_\beta \rfloor]$ and $[\lceil w_\beta \rceil, n]$ of won bets w giving returns less than respectively greater than β, when probabilities are extreme, the difference between α and the cumulative probability of the floor of w_β increases. From (36), we see that w_β depends on n, p and α as input. We therefore take the values $p = 0.1, 0.5, 0.9$ and for each of the combinations of these with the four values of n and α in Table 2, we calculate the difference between α and the cumulative probability of $\lfloor w_\beta \rfloor$, from the Binomial with n and p as parameters. In this simulation, the mean value (note, not mean absolute value) of $P(w \leq \lfloor w_\beta \rfloor) - \alpha$ was -0.0027, -0.00026 and 0.0050, respectively, for the three values of p. In our equation for f, we calculate the cumulative probability of w_β by approximating it as continuous; however, in the simulations, it is discrete. At low levels of p, the discrete probability is lower than α, which is why we see a large proportion of $\hat{\alpha} < \alpha$ at low probabilities, and vice versa for high probabilities. Thus, the equation remains calibrated in p as well, the distorted values of $P(\hat{\alpha} < \alpha \mid p)$ simply a result of the discrete nature of the simulations.

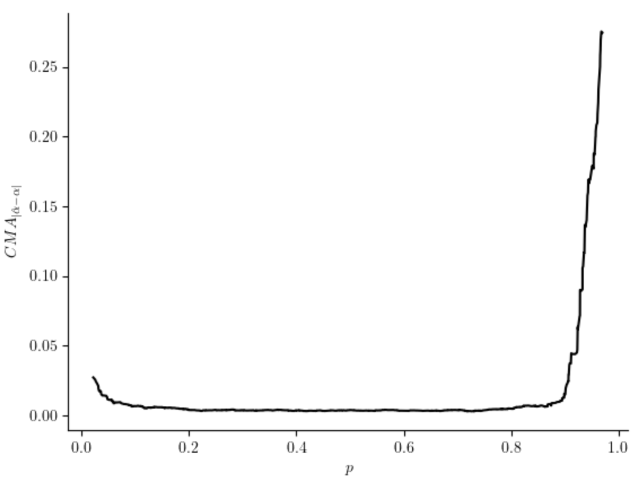

Figure 4: Central moving average across 25 preceding and succeeding points of the absolute difference between $\hat{\alpha}$ and α. Calibration decreases when probabilities become extreme.

We have limited ourselves to calculations where $p \in [0.1, 0.9]$. This is due to the deterioration in the calibration of $\hat{\alpha}$ to α as the probabilities move towards 0 or 1. Figure 4 plots the absolute difference between the two versus p, where b and f are calculated as in the simulations above. Plotted is a central moving average, at each point p the average absolute difference between $\hat{\alpha}$ and α across the previous and immediately succeeding twenty-five points. As p becomes increasingly extreme, the error increases, especially for large probabilities, meaning, on average, that the calibration of our risk-management is superior for more moderate probabilities. The points 0.1 and 0.9 used above are largely arbitrary, but seem to be roughly where the rate of deterioration increases drastically.

4.3 Taylor Expansion

The approximation in (37) uses the Maclaurin series of $\ln(x)$, estimating the function close to the point $x \approx 0$. If we instead consider a point f_0 as an estimate for f, then we can use the Taylor series instead, expanding the logarithmic terms including f around f_0 instead of 0. This derivation is more elaborate, but we find that $\ln(1-f) \approx \ln(1-f_0) - \frac{1}{1-f_0}(f-f_0) - \frac{1}{2(1-f_0)^2}(f-f_0)^2$ and that $\ln(1+fb) \approx \ln(1+f_0 b) + \frac{b}{1+f_0 b}(f-f_0) - \frac{b^2}{2(1+f_0 b)^2}(f-f_0)^2$, both to the second degree. We can similarly substitute into (37), and then rearrange for a quadratic in f, which has the following solution:

$$
\begin{aligned}
f &= \frac{-b_f - \sqrt{b_f^2 - 4a_f c_f}}{2a_f} \\
a_f &= w_\alpha \cdot \frac{1}{2}\left(\frac{1}{(1-f_0)^2} - \frac{b^2}{(1+f_0 b)^2}\right) - \frac{n}{2(1-f_0)^2} \\
b_f &= w_\alpha \left(\frac{b}{1+f_0 b} + \frac{1}{1-f_0}\right) - \frac{n}{1-f_0} - 2f_0 a_f \\
c_f &= w_\alpha \cdot \ln\left(\frac{1+f_0 b}{1-f_0}\right) - \ln(1+\beta) + n\ln(1-f_0) + f_0^2 a_f - f_0 b_f
\end{aligned}
\tag{40}
$$

which takes arguments p, b, n, α, β, f_0, and has w_α, A, B and C calculated as in (39).

We can evaluate the solution in (40) compared to (39) during the same simulation as above, iterating across the 320 parameter combinations outlined in Table 2. Our estimate for f_0 in (40) is the value for f from (39), the rationale being that the Maclaurin approximation is good for values of f close to zero, and the Taylor approximation being more accurate for values of f close to f_0, meaning (39) is used to identify the correct range of f, and (40) the correct value.

Across the 320 iterations, then, we find that $\text{corr}(f_0, f) = 0.999$, suggesting that (40) only makes minor adjustments to (39). However, where the correlation between α and $\hat{\alpha}_{f_0}$ above was 0.981, now we have $\text{corr}(\alpha, \hat{\alpha}_f) = 0.970$, the Taylor-series method giving an inferior fit to α. Indeed, $P\left(\mid \alpha - \hat{\alpha}_{f_0} \mid > \mid \alpha - \hat{\alpha}_f \mid\right) = 0.103$, with the original method in (39) being worse in just 10.3% of iterations. If we let I_i be the indicator variable such that $I_i = 1$ if in iteration i, $\mid \alpha_i - \hat{\alpha}_{i,f_0} \mid > \mid \alpha_i - \hat{\alpha}_{i,f} \mid$ and 0 otherwise, then we can regress the linear probability model

$$
I_i = \delta_0 + \delta_1 p_i + \delta_2 n_i + \delta_3 \alpha_i + \delta_4 \beta_i + \epsilon_i \tag{41}
$$

to see whether any of the input variables (omitting b due to collinearity issues with p) can inform whether (39) or (40) is more suitable for a particular case. We find that the test statistics from estimating (41) on the 320-iteration simulation for δ_1 and δ_3, respectively, are 3.184 and -2.869, with coefficient estimates of $\hat{\delta}_1 = 0.3301$ and $\hat{\delta}_3 = -1.3554$, suggesting that (40) is better for high-probability, low-α cases. However, for the 16 iterations with $p_i = 0.9$ and $\alpha_i = 0.01$, they have the same $\hat{\alpha}$ in 12 iterations, with (40) better in just 4. We saw above that (39) generally produces values for f that give $\hat{\alpha} > \alpha$ when p is large, which we explained was to be expected due to the discrete nature of $\hat{\alpha}$. (40) only producing four instances of improvement over (39) despite this is therefore not sufficient evidence to suggest that it is a superior alternative for these particular cases, given its inferiority otherwise across the input parameter space. The conclusion of this section, then, is that the Maclaurin-based solution is more calibrated than the Taylor-based solution, and preferable especially given the lower complexity.

4.4 Elaboration for Portfolios of Distinct Gambles

Above we have seen that the equation for f is accurate across a broad range of the input parameters. The analysis presented a single bet with probability p and profit b, and optimised the Kelly fraction such that, over n repeated wagers, the probability of experiencing returns worse than β equals α. In many practical scenarios, however, a wager is available as a one-time opportunity, not being able to be repeated arbitrarily. We therefore explore the possibilities of using our calculation for f to create a general risk-management strategy, in the sense that, if we take n separate wagers, each optimised according to the same Value-at-Risk profile defined by n, α and β,

(p_l, p_u)	$(0.1, 0.2)$	$(0.2, 0.3)$	$(0.3, 0.4)$	$(0.4, 0.5)$	$(0.5, 0.6)$	$(0.6, 0.7)$	$(0.7, 0.8)$	$(0.8, 0.9)$
0	0.04777	0.0487	0.05101	0.05119	0.0528	0.0534	0.05705	0.07518
$\boldsymbol{\theta}$	0.05275	0.04928	0.0479	0.04586	0.04504	0.04486	0.04608	0.05799

Table 3: For eight ranges of (p_l, p_u), a comparison of $\hat{\alpha}$ for the median-error case between the optimised and unoptimised models.

then for these n distinct bets, combined, they will achieve the same Value-at-Risk. The intuition is simple. Take the two wagers $(p, b) = (0.25, 3.15), (0.75, 0.35)$, calculating f optimised over the same n, α and β. Repeating each bet $n-1$ times, adding the n-th bet, in each case, gives a probability α of experiencing returns less than or equal to β. Therefore, it seems reasonable that if we instead take the $n-1$ bets of each wager, but then add the n-th bet from the other, the same Value-at-Risk risk profile will be achieved. We can then use this reasoning inductively (despite it not necessarily being robust mathematically), with the commutativity of the Kelly gambling returns, to create a portfolio of n separate wagers, hopefully with a well-defined Value-at-Risk.

Therefore, we seek to optimise f with respect to p. We do so by adjusting the constant A

$$A = \frac{8(\pi - 3)}{3\pi(4 - \pi)} + \theta_0 + \theta_1 p + \theta_2 p^2 + \theta_3(p(1 + b) - 1) \tag{42}$$

where the θ are parameters to be optimised, as an adjustment to the intercept, the probability and its quadratic term, and the wager expected value. The simulation performed is as follows, for a chosen risk profile n, α and β, we for a given set of parameters $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)$, for each of 100,000 runs (as above), draw n values $EV \sim \mathrm{U}(0, 0.025)$ and $p \sim \mathrm{U}(p_l, p_u)$, where we have eight pairs $(p_l, p_u) = (0.1, 0.2), (0.2, 0.3), \ldots (0.8, 0.9)$. We calculate the adjusted f with the input parameters, and hence get a value for $\hat{\alpha}$, as above. For each pair (p_l, p_u) we repeat this process five times, obtaining five values for $\hat{\alpha}$. For each pair, we select the median value of $\mid \hat{\alpha} - \alpha \mid$, and then across the eight medians, return the maximum value. We thus seek the parameters in $\boldsymbol{\theta}$ which minimises this maximum median value. The rationale is that we want the model to be calibrated across the entire domain of p, and hence split it into these segments of width 0.1, while considering only the worst value.

For the Value-at-Risk profile $(n, \alpha, \beta) = (500, 0.05, -0.05)$, in our simulation we found that $\boldsymbol{\theta} = (-0.4, 0.933, 1.55, 0.3833)$ gives the best fit, with a representative distribution of $\hat{\alpha}$ values corresponding to the median value of $\mid \hat{\alpha} - \alpha \mid$ for each of the eight segments being displayed in Table 3, compared to the same values without optimisation (with parameters from the zero vector **0**). A couple of comments are that, first, here the model error is related to p, in a convex quadratic effect. For moderate values of p, the expression value for f will be too conservative, while it will be exaggerated for extreme values. Second, we have a rather conservative range of expected values, these ranging from 0 to 0.025. When the upper limit of this range is increased, the model calibration deteriorates rapidly (also true for the single-gamble case above). This is a clear limitation, with our expression only applicable for wagers which are only slightly favourable; of course, if a bettor can systematically identify extremely favourable betting opportunities, then their need for a conservative risk-management system will be decreased. Furthermore, these parameters are results which hold only for $(n, \alpha, \beta) = (500, 0.05, -0.05)$. When attempting to optimise $\boldsymbol{\theta}$ while also varying n, α and β, the calibration again exploded. Finally, in the majority of segments, the basic model actually performs reasonably well, sometimes even better than the optimised model; however, it is evident that, when the parameters are fit, the worst case is significantly better. As such, for this particular Value-at-Risk specification, and for the given ranges of probabilities and expected values, we get the final closed-form expression for f:

$$\begin{aligned}
f &= \frac{w_\beta(1+b) - n + \sqrt{(n - w_\beta(1+b))^2 - 2\ln(1+\beta)(n + w_\beta(b^2-1))}}{n + w_\beta(b^2-1)} \\
w_\beta &= np - \sqrt{np(1-p)\frac{-B + \sqrt{B^2 - 4AC}}{A}} \\
A &= \frac{8(\pi - 3)}{3\pi(4-\pi)} + \theta_0 + \theta_1 p + \theta_2 p^2 + \theta_3(p(1+b) - 1) \\
B &= \frac{4}{\pi} + AC \\
C &= \ln(1 - (2\alpha - 1)^2)
\end{aligned} \tag{43}$$

which takes arguments p, b, n, α, β, and $\boldsymbol{\theta}$, with $(\theta_0, \theta_1, \theta_2, \theta_3) = (-0.4, 0.933, 1.55, 0.3833)$ for $(n, \alpha, \beta) = (500, 0.05, -0.05)$. Of course, if an individual bettor has a certain class of wagers, in terms of the typical values for p and b that they can systematically identify, then they can customise $\boldsymbol{\theta}$ accordingly. We have here merely demonstrated, in a thoroughly non-exhaustive manner, that it is possible to create a reasonably well-calibrated model for the broadest range of parameters possible.

4.5 General Risk Discussion

We have discussed a Value-at-Risk model for wagers with binary outcomes. One might think that such wagers are extremely risky, given that a losing wager loses the investor their entire stake. Similarly, VaR models have received substantial criticism due to their unsuitability for financial markets with heavy tails (Sollis, 2009). Given the apparent inherent riskiness of wagers, and the deficiencies of VaR models, the combination does not intuitively seem to be the best risk-management solution.

To evaluate this, we perform a similar simulation to the above. We keep $\alpha = 0.05$ and $\beta = -0.05$ fixed, and consider $p \in \{0.25, 0.5, 0.75\}$ and $n \in \{100, 500, 1{,}000\}$, with a 1% edge giving b, meaning we can use (39) as our model for calculating f. For each of the nine combinations, we calculate mean returns and volatility, giving the Sharpe ratio (Sharpe, 1966), as well as skewness and kurtosis. The Sharpe ratio is standardised to the 1,000-wager, level, multiplying the Sharpe ratio for a 100-wager iteration by $\sqrt{10}$ and that for a 500-wager iteration by $\sqrt{2}$. Table 4 shows the results.

The first concern listed above is the perceived riskiness embedded in binary-outcome wagers. This riskiness should be reflected in high volatility, but in all cases across p, the Sharpe ratio is good or even excellent, ranging from 0.175-0.18 for $p = 0.25$ to 0.53-0.54 for $p = 0.75$ (note that our normalisation mechanism seems to be sensible, given that the Sharpe ratios for different values of n converge for a certain value of p). It is reasonable that lower-probability events should have a lower Sharpe ratio than high-probability events, since fewer wagers will be won. Regarding the second concern, that of VaR being a generally flawed metric, we can see here that the kurtosis fluctuates around 3, as one would expect for a normally-distributed random variable. Similarly, the skewness is slightly positive, meaning that we have more tail events in the positive tail than the negative tail. Since VaR only cares about the negative tail, this too is not an issue: if anything, our models will be overly conservative. The skewness, converesely to Sharpe ratios, decreases in p, which is reasonable for similar reasons: just a few extra won wagers will have an inordinately large impact on returns, which is more likely to happen when fewer wagers are won (due to low p), all else being equal.

p	n	Mean	Volatility	Sharpe Ratio	Skewness	Kurtosis
	100	0.0018	0.0320	0.175	0.203	3.046
0.25	500	0.0042	0.0336	0.175	0.157	3.059
	1,000	0.0063	0.0350	0.180	0.140	3.035
	100	0.0031	0.0330	0.300	0.096	2.981
0.50	500	0.0079	0.0360	0.310	0.108	3.029
	1,000	0.0120	0.0384	0.311	0.126	3.029
	100	0.0059	0.0346	0.539	-0.007	2.972
0.75	500	0.0153	0.0407	0.531	0.072	2.996
	1,000	0.0248	0.0467	0.531	0.090	2.991

Table 4: Returns simulation for different values of p and n, with $\alpha = 0.05, \beta = -0.05$ and a 1% edge. Sharpe ratios are standardised to the 1,000-wager level. Mean, volatility, skewness and kurtosis relate to the distribution of returns across iterations.

We can confidently say, then, that binary-outcome investments can have attractive risk profiles when risk is managed sensibly, and that VaR is a very useful tool for curtailing such risk.

4.6 Conclusion

In this section, we have demonstrated that our closed-form expression for the optimal Kelly fraction, subject to a desired Value-at-Risk, is well-calibrated across a broad range of gambling scenarios. We have also seen that a bettor can use an adjusted version of our expression when they stake on n distinct wagers, rather than the same wager repeated n times, to similarly control their downside. Although not applicable in all situations, our framework is useful for practitioners, who want to maximise their capital growth in the spirit of Kelly-criterion investing, while simultaneously managing risk.

5 Brier Score Test

The Brier score (Brier, 1950) is a strictly proper scoring rule for quantifying the strength of a vector of probabilistic forecasts, given their outcomes. Bradley et al. (2008) developed – and Wilks (2010) discussed further – an expression for the variance of Brier scores, incorporating both the forecasts and their outcomes. These two authors demonstrate the accuracy of the expression, illustrating with Monte Carlo simulation that the randomly sampled variances match the predicted variances across a broad range of forecasts. Although useful, for instance, in defining a confidence interval for the forecast-generating model's Brier skill score, what the Bradley et al. framework fails to accomplish is testing whether the model is calibrated or not, in the absence of some target or benchmark Brier score. We supplement this deficiency by providing estimates for the Brier score mean and variance using only the probabilistic forecasts, allowing us to use the observed Brier score as a test statistic to compare against this estimated distribution. In this section, we provide these estimates for the mean and variance of the true Brier score under the null hypothesis of perfect calibration, and proceed to evaluate the accuracy and suitability of these expressions using Monte Carlo simulation.

5.1 Brier Score Hypothesis Testing Framework

We have an n-tuple of probability forecasts $\boldsymbol{f} = (f_1, \ldots, f_n)$ for n events, and the n-tuple of outcomes for these events $\boldsymbol{x} = (x_1, \ldots, x_n)$, where each x_i is a Bernoulli trial with unknown parameter, $x_i \in \{0, 1\} \; \forall \; i$. We can calculate $\hat{S}$, an estimate for the Brier score S (Brier, 1950) of these forecasts, using

$$\hat{S} = \frac{1}{n} \sum_{i=1}^{n} (f_i - x_i)^2 \tag{44}$$

where the Brier score is a strictly proper scoring rule (Gneiting and Raftery, 2007), incentivising the forecaster to provide their true belief f_i for each forecast i. The Brier score is a loss function, ranging from 0 to 1, with 0 representing perfect forecasting.

We want to develop a hypothesis test for whether our model, from which $\boldsymbol{f}$ is generated, is calibrated. We define this null hypothesis as

$$H_0 : E[x_i] = f_i \; \forall \; i \tag{45}$$

that is, the expected value of the i-th outcome variable x_i equals the model forecast: $x_i \sim \text{Bernoulli}(f_i)$. In this framework, we assume that $\boldsymbol{f}$ is calibrated, and hence calculate the distribution of sample Brier scores that would be observed under the null hypothesis. We calculate our observed Brier score using (44) by combining $\boldsymbol{f}$ with $\boldsymbol{x}$. We then compare this test statistic to the null distribution, and evaluate (45) at a given significance level.

The expected Brier score $E[S]$ is calculated

$$\begin{gathered} E[S] = E\left[\frac{1}{n}\sum_{i=1}^{n}(f_i - x_i)^2\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[(f_i - x_i)^2\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[f_i^2 - 2f_i x_i + x_i^2\right] = \\ \frac{1}{n}\sum_{i=1}^{n}\left(E\left[f_i^2\right] + E\left[-2f_i x_i\right] + E\left[x_i^2\right]\right) = \frac{1}{n}\sum_{i=1}^{n}(f_i^2 - 2f_i^2 + f_i) = \frac{1}{n}\sum_{i=1}^{n}(f_i - f_i^2) \end{gathered} \tag{46}$$

using that, under the null, the f_i are constant, and that, with the x_i being Bernoulli random variables, $x_i^2 = x_i$.

We can similarly calculate the variance of the Brier score, $Var[S]$, as

$$Var[S] = E[S^2] - E[S]^2 = E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(f_i - x_i)^2\right)^2\right] - \left(\frac{1}{n}\sum_{i=1}^{n}(f_i - f_i^2)\right)^2 \tag{47}$$

Expanding, $\left(\frac{1}{n}\sum_{i=1}^{n}(f_i - x_i)^2\right)^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n}(f_i - x_i)^2\right)\left(\sum_{j=1}^{n}(f_j - x_j)^2\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}(f_i - x_i)^4 + \sum_{i\neq j}(f_i - x_i)^2(f_j - x_j)^2\right)$. Hence, $E[S^2] = \frac{1}{n^2}\left(\sum_{i=1}^{n} E\left[(f_i - x_i)^4\right] + \sum_{i\neq j} E\left[(f_i - x_i)^2(f_j - x_j)^2\right]\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n} E\left[(f_i - x_i)^4\right] + \sum_{i\neq j} E\left[(f_i - x_i)^2\right] E\left[(f_j - x_j)^2\right]\right)$. We have that $E\left[(f_i - x_i)^4\right] = f_i(1 - f_i)^4 + (1 - f_i)f_i^4$ and that $E\left[(f_i - x_i)^2\right] = f_i - f_i^2$, which gives $E[S^2] = \frac{1}{n^2}\left(\sum_{i=1}^{n}\left[f_i(1 - f_i)^4 + (1 - f_i)f_i^4\right] + \sum_{i\neq j}\left[(f_i - f_i^2)(f_j - f_j^2)\right]\right)$, and finally

$$Var[S] = \frac{1}{n^2}\left(\sum_{i=1}^{n}\left[f_i(1 - f_i)^4 + (1 - f_i)f_i^4\right] + \sum_{i\neq j}\left[(f_i - f_i^2)(f_j - f_j^2)\right]\right) - \left(\frac{1}{n}\sum_{i=1}^{n}(f_i - f_i^2)\right)^2 \tag{48}$$

Thus, given $\boldsymbol{f}$, we can use (46) and (48) to calculate the mean and variance of the distribution of Brier scores, under the null hypothesis in (45). As desired, these expressions are independent of $\boldsymbol{x}$, which only enters into the calculation of our test statistic from (44). With the Brier score bounded by 0 and 1, it seems reasonable to assume that these sample Brier scores are beta distributed $S \sim \text{Beta}(v, w)$ (Bayes, 1763; Krzysztofowicz and Long, 1991), with probability density function

$$\frac{S^{v-1}(1-S)^{w-1}}{B(v,w)} \tag{49}$$

where B is the beta function and v and w are parameters such that if a beta distribution has mean μ and variance σ^2, then

$$\begin{aligned} v &= \mu\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right) \\ w &= \frac{1-\mu}{\mu}v \end{aligned} \tag{50}$$

Thus, with $\mu = E[S]$ from (46) and $\sigma^2 = Var[S]$ from (48), using (50) we can calculate the v_S and w_S which govern the distribution of Brier scores under the null hypothesis in (45). For our test statistic $\hat{S}$ from (44) using $\boldsymbol{f}$ and $\boldsymbol{x}$, we obtain a p-value

$$p_S = 1 - I_{\hat{S}}(v_S, w_S) \tag{51}$$

where $I_{\hat{S}}$ is the regularised incomplete beta function, which is the cumulative distribution function of the beta distribution, evaluated at $\hat{S}$ (Peizer and Pratt, 1968). Since the Brier score is a strictly proper scoring rule, our hypothesis test only considers the positive tail, as this is where the scores from uncalibrated models would appear: the negative tail consists of calibrated models which have experienced fortunate sequences of outcomes, with respect to their forecasts.

We now consider the accuracy of our expressions in (46) and (48), the validity of the assumption that the Brier scores are beta distributed, and the performance of (51) as a solution to (45).

5.2 Analysis

To evaluate the accuracy of our equations for $E[S]$ and $Var[S]$, as well as the beta assumption, we run a Monte Carlo simulation. In each of $m = 10,000$ iterations, we draw $n = 10^u$ where $u \sim U(\log_{10}(50), 3)$ – that is, n is drawn on a logarithmic scale between 50 and 1,000. We also draw parameters for a beta distribution $v_f, w_f \sim U(0.5, 5)$, giving $\boldsymbol{f} = (f_1, \ldots, f_n \mid f_i \sim \text{Beta}(v_f, w_f) \ \forall \ i)$, from which we calculate $E[S]$ and $Var[S]$ using

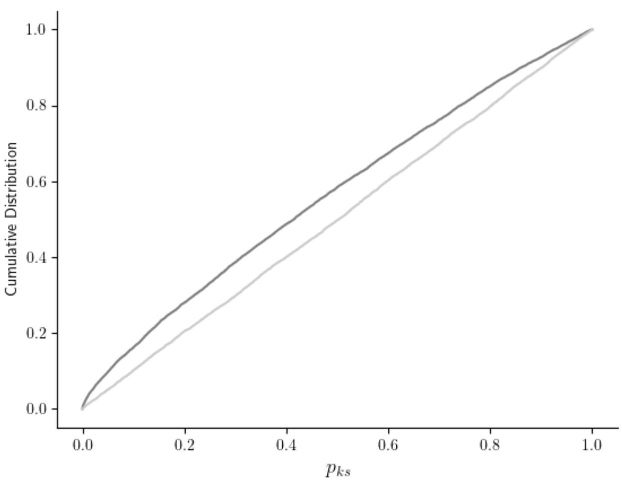

Figure 5: The cdf of Kolmogorov-Smirnov p-values across 10,000 iterations. The darker line corresponds to our expressions for $E[S]$ and $Var[S]$, while the lighter line tests data drawn from the specified distribution.

(46) and (48), respectively. The bounds for the v_f and w_f are chosen to give a relatively broad range of distributions of forecasts, to mimic the breadth of possible forecasting scenarios which may occur in practice. Then, with $\boldsymbol{f}$, we generate 10,000 n-length vectors $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj} \mid x_{ij} \sim \text{Bernoulli}(f_i)\ \forall\ i)$, and for each of these $\boldsymbol{x}_j$ calculate the corresponding Brier score $\hat{S}_j$ using (44). Thus, for each iteration's $\boldsymbol{f}$, we have 10,000 Brier-scores, which follow the distribution described under the null hypothesis in (45).

We want to evaluate, first, whether $E[S] \approx \bar{S}$ and $Var[S] \approx s_S^2$, that is, whether our expressions for the mean and variance match the sample values across the $\hat{S}_j$; and second, whether the $\hat{S}_j \sim \text{Beta}(v_S, w_S)$, where v_S and w_S are calculated from (50) using $E[S]$ and $Var[S]$ (note, not using $\bar{S}$ and s_S^2: we want to ascertain that the $\hat{S}_j$ are beta distributed with our expressions for the mean and variance, not merely with the observed ones). To accomplish the former evaluation, we simply calculate the correlation across the 10,000 $(E[S], \bar{S})$ and $(Var[S], s_S^2)$ pairs, respectively. For the latter, we use the one-sample Kolmogorov-Smirnov (KS) test (An, 1933; N. Smirnov, 1948), a non-parametric test which gives the probability of a sample being drawn from a certain continuous probability distribution with specified parameterisation, with the alternative hypothesis being that the data is drawn from some other distribution.

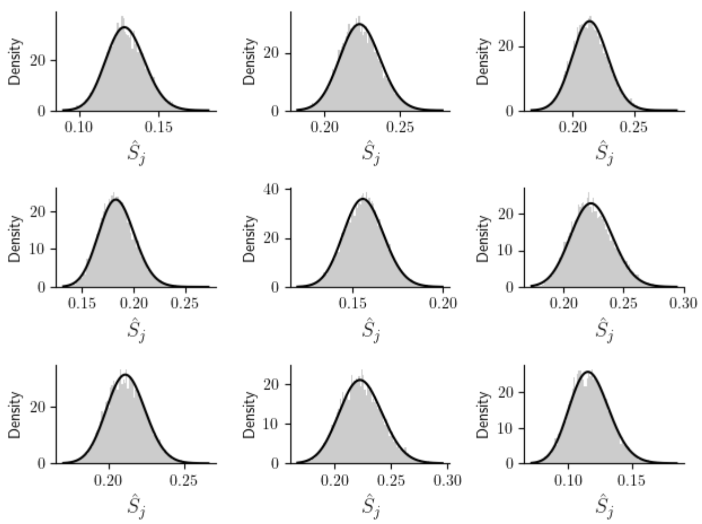

Figure 6: Nine sample cases of predicted (dark) and realised distributions of $\hat{S}_j$ where the Kolmogorov-Smirnov p-value is $p_{ks} < 0.001$. Even for these poorly-fit cases, the visual match is very good.

We find that $\text{Corr}(E[S], \bar{S}) = 0.999993$ and $\text{Corr}(Var[S], s_S^2) = 0.999753$, with a practically perfect fit between the predicted values from our equations (46) and (48). Indeed, when these are plotted, the resulting figures are barely distinguishable from a straight line through the origin with unit slope. The results from the KS tests are also convincing, with the null hypothesis being sustained in 90.3% and 96.8% of the $m = 10,000$ iterations, at 5% and 1% significance levels, respectively. The darker line in Figure 5 shows the cumulative distribution of the p-values p_{ks} from the 10,000 iterations. The lighter line is the cumulative distribution of p-values from a KS test when the $\hat{S}_j$ values are drawn immediately from the beta distributions with v_S and w_S, and gives the cumulative uniform distribution which p-values follow under the null, as expected. Hence, the concavity of the dark line demonstrates that there is a greater proportion of smaller p-values than one would expect if the $\hat{S}_j$ were indeed beta distributed with parameters v_S, w_s, suggesting that our expressions for the mean and variance do not give perfect distributional fits. However, for our purposes, testing the hypothesis in (45), it should be sufficient. Indeed, Figure 6 plots nine sample distributions of 10,000 $\hat{S}_j$ as well as the beta distribution (dark line) parameterised by v_S, w_S, for cases where $p_{ks} < 0.001$, which occurred in only 0.68% of iterations. The visual fit to the general shape of the distribution is perfect, with the disparity arising from arbitrary spikes at certain values of $\hat{S}_j$. Given our use case, where we will compare an observed Brier score $\hat{S}$ to its distribution under the null hypothesis in (45) according to (51), this approximate fit is more than adequate.

We now evaluate whether the KS p-value depends on the distribution of the $\boldsymbol{f}$ or its number of elements n. As described above, in each iteration we randomly draw n probabilities $f_i \sim \text{Beta}(v_f, w_f)$. We use beta regression (Kieschnick and McCullough, 2003; Ferrari and Cribari-Neto, 2004; Geissinger et al., 2022) to regress the p_{ks} on linear combinations of n, v_f, w_f and their higher polynomials. Essentially, beta regression takes a conventional multi-linear regression model and transforms it using, in this case, a logit link function, whereby the model predictions are bounded by 0 and 1. We can thus evaluate the extent to which n, v_f and w_f contribute to p_{ks}. We use the Bayesian information criterion (Schwarz, 1978), given by

$$\text{BIC} = k \ln(m) - 2 \ln(L) \tag{52}$$

where k is the number of model parameters, m is the data sample size, and L is the maximised likelihood of the model, upon optimisation. The BIC provides a nuanced comparison of different models, penalising unnecessary parameters (unduly high k) while rewarding improved model performance (higher L). Relatively, the more negative is BIC, the better is the model. For a base model with only a constant, we find that BIC $= -724.93$. A second model with a constant as well as the linear and quadratic terms of v_f and w_f has BIC $= -753.76$, suggesting that these two parameters contribute valuable information in attempting to predict p_{ks}. Adding the linear and quadratic terms of n to this model, however, gives BIC $= -1105.42$, a vast outperformance, demonstrating that the majority of variation in p_{ks} is explained by n, the number of forecasts made.

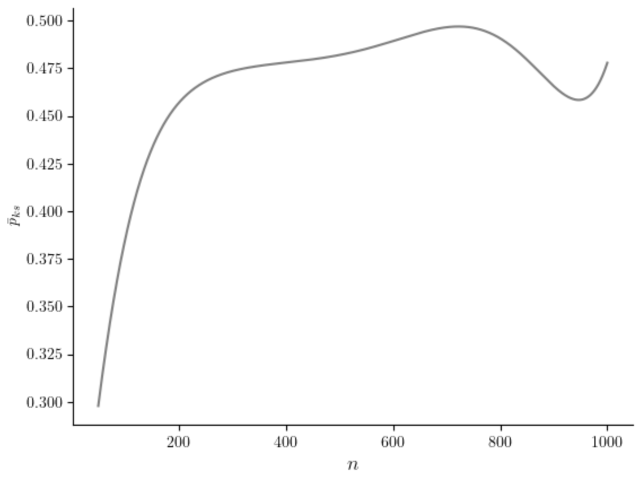

Figure 7: Using Beta regression, we model the Kolmogorov-Smirnov p-values p_{ks} as a function of n and its higher-order polynomial terms.

Figure 7 plots p_{ks} as a function merely of n and its polynomials up to degree seven. Obviously this model is overfit, but our intention is to explain the observed data to the utmost extent, not to make a-priori predictions. We observe that, above $n \approx 200$, the p-values are practically constant, with a mean of 0.475 or so, while there is a sharp increase from $n = 50$, suggesting that the majority of cases where our expressions for $E[S]$ and $Var[S]$ did not fit the observed data occurred when $n < 200$. Of course, even at $n = 50$, the mean value $\bar{p}_{ks}$ is roughly 0.3, so on average our calculations are valid, and indeed, Figure 6 suggests that even when the fit is bad, per se, it is still more than serviceable for our purposes.

We hence conclude that for two n-tuples of forecasts $\boldsymbol{f}$ and realised outcomes $\boldsymbol{x}$, the observed Brier score $\hat{S}$ from (44) follows a beta distribution with mean given by (46) and variance given by (48) under the null hypothesis in (45) of the forecast-generating model being perfectly calibrated. We can then use (51) to evaluate whether the model is indeed calibrated, at a chosen significance level. To demonstrate the test, we perform a second Monte Carlo simulation. In each of 1,000 iterations, we draw $v_f, w_f \sim U(0.5, 5)$. Then, in each iteration, we draw n as above 10,000 times and for each generate $\boldsymbol{f} = (f_1, \ldots, f_n \mid f_i \sim \text{Beta}(v_f, w_f)\ \forall\ i)$. For each $\boldsymbol{f}$, we calculate $E[S]$ and $Var[S]$. To generate $\boldsymbol{x}$ we use the linear model of Murphy and Wilks (1998), adopted by Bradley et al. (2008), given by

$$E[x_i] = (1 - \Delta) f_i + \Delta \frac{v_f}{v_f + w_f} \tag{53}$$

where $\Delta \geq 0$ is a parameter and $\frac{v_f}{v_f+w_f}$ is the mean of the beta distribution of the f_i. If $\Delta = 0$, then (53) gives $E[x_i]$ as (45) would demand, while if $\Delta > 0$, then the forecasts are less individually calibrated, increasingly becoming the base-rate frequency, the climatological probability. For different values of Δ, we generate $\boldsymbol{x}$ corresponding to $\boldsymbol{f}$ according to (53), with the Bernoulli parameter being $E[x_i]$. For different values of the significance level α, we then evaluate the null hypothesis in (45) using (51). Each of the 1,000 iterations will then have, for each (α, Δ)-pair, 10,000 observations of whether the null hypothesis in (45) is rejected or not. For each iteration, we take the proportion of cases in which the null is sustained. Table 5 gives $1 - \bar{p}_{\alpha\Delta}$ the mean of these proportions across the iterations, for each (α, Δ)-pair.

Considering first the case where $\Delta = 0$, we would expect to see a proportion α of cases leading to rejections, since we are under the null hypothesis. With a slight bias, this is indeed what we observe. The test seems to be somewhat over-sensitive, rejecting a proportion slightly greater than α. But clearly it is very well-calibrated overall. Indeed, as Δ increases, the proportion or rejections increases at each significance level, as expected, since model calibration decreases in Δ.

This relationship is actually extremely predictable: denoting $1 - \bar{p}_{\alpha\Delta}$ as the mean proportion of rejections across the 1,000 iterations, with the given values of Δ and α, if we regress $\bar{p}_{\alpha\Delta} = \beta_0 + \beta_1\Delta + \beta_2\alpha + \varepsilon_{\alpha\Delta}$ by using ordinary least squares, we get an R-squared metric of 0.956. This, of course, is merely across the nine observations shown in Table 5, but if we instead index the 1,000 iterations by i and then denote $1 - p_{i\alpha\Delta}$ as the proportion of rejections for pair $(\alpha\Delta)$ in the i-th iteration and regress $p_{i\alpha\Delta} = \beta_0 + \beta_1\Delta_i + \beta_2\alpha_i + \varepsilon_{i\alpha\Delta}$, we instead have 9,000 observations and obtain an R-squared of 0.775: not quite as perfect, but representative of the strong relationship between the three variables nonetheless, especially given that there are only nine pairs of values to regress on.

		α		
		0.01	0.05	0.1
	0	0.989	0.949	0.899
Δ	0.125	0.898	0.759	0.652
	0.25	0.695	0.512	0.407

Table 5: $1 - \bar{p}_{\alpha\Delta}$: the mean proportion of cases where the null hypothesis in (45) is sustained using (51), as a function of Δ and α.

We conclude this discussion by stating, then, that our hypothesis test developed above is robust and calibrated across a large range of values for the number of forecasts n as well as the distribution of these forecasts $\boldsymbol{f}$. A forecaster can thus, after making a moderate number of predictions and observing the outcomes, evaluate whether their model is calibrated or not.

5.3 Test Eligibility

We have seen above that the beta fit is good even for $n = 50$, but this was for distributions of probabilities f_i ranging from $v_f, w_f \in [0.5, 5]$. In this section we suggest a rule for determining whether the test is applicable. The criteria for a normal approximation to the beta distribution being accurate relate, in general, to the approximate equality of the parameters v and w, or to their size (Wise, 1960). Here, however, we will imagine that a related criterion is applicable, namely, that $E[S] - z\sqrt{Var[S]} \geq 0$, which comes from the rule for a normal approximation being appropriate for the binomial distribution, where $z = 3$. We can rearrange to find $\frac{E[S]}{\sqrt{Var[S]}} \geq z$.

Figure 8 plots the usefulness of this rule at different values of z, for a Monte Carlo simulation very similar to that above, the only difference being that we draw n on a logarithmic scale between 5 and 1,000, and v_f and w_f on a logarithmic scale between 0.1 and 100, to give a broader range of possible forecasting scenarios. In the three panels, the darker line considers those iterations which gave $\frac{E[S]}{\sqrt{Var[S]}} \geq z$, while the lighter line shows the complement. The upper panel gives the mean KS p-value for these two groups. The second panel gives the proportion of p-values above and below 0.05, while the lower panel does the same but for the significance level 0.01. The horizontal lines in each are drawn at 0.475, 0.903 and 0.968. These numbers were mentioned above as the representative values when the mean and variance estimates were good and the beta distribution behaved well. We can see that this normalcy is attained roughly for $z = 10$, demonstrating the ability of this rule to filter out the cases where the test is not quite applicable.

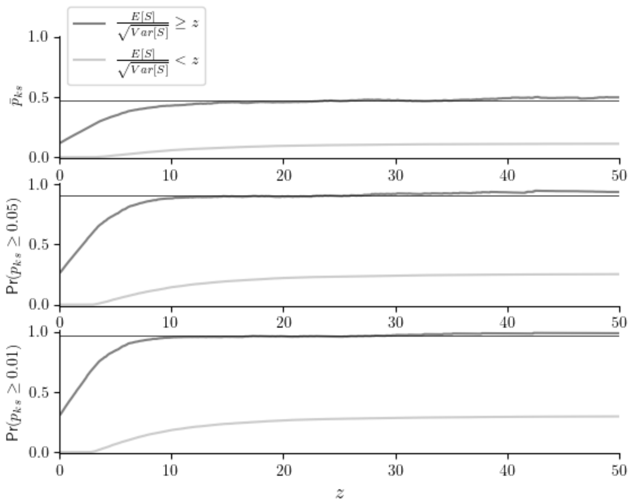

Figure 8: Filtering by $\frac{E[S]}{\sqrt{Var[S]}}$ (dark lines aggregate above z), displaying the mean p-value (upper panel) and the probability of sustaining the KS null hypothesis. The horizontal lines denote usual behaviour, attained at $z \approx 10$.

We want to extract n from this relation, giving a threshold sample size. We can express the numerator as $E[S] = E[f_i - f_i^2]$. Separating the individual components of $Var[S]$, we find $\frac{1}{n^2}\sum_{i=1}^{n}\left[f_i(1-f_i)^4 + (1-f_i)f_i^4\right] = \frac{1}{n}E\left[f_i(1-f_i)^4 + (1-f_i)f_i^4\right]$ and $\frac{1}{n^2}\sum_{i\neq j}\left[(f_i - f_i^2)(f_j - f_j^2)\right] \approx \frac{n-1}{n}\left(E[f_i - f_i^2]\right)^2$ giving $Var[S] \approx \frac{1}{n}E\left[f_i(1-f_i)^4 + (1-f_i)f_i^4\right] + \frac{n-1}{n}\left(E[f_i - f_i^2]\right)^2 - \left(E[f_i - f_i^2]\right)^2 = \frac{1}{n}\left(E\left[f_i(1-f_i)^4 + (1-f_i)f_i^4\right] - \left(E[f_i - f_i^2]\right)^2\right)$. Thus,

$$\frac{E[S]}{\sqrt{Var[S]}} \approx \frac{E[f_i - f_i^2]}{\sqrt{\frac{1}{n}\left(E\left[f_i(1-f_i)^4 + (1-f_i)f_i^4\right] - \left(E[f_i - f_i^2]\right)^2\right)}} \tag{54}$$

If we assume that the distribution of $\boldsymbol{f}$ remains constant, this implies that $\frac{E[S]}{\sqrt{Var[S]}} \mathrel{\underset{\sim}{\propto}} \sqrt{n}$. We thus obtain the heuristic $n_\tau = z^2\frac{Var[S]}{E[S]^2}$, where n_τ is the threshold value of n which allows us to apply the hypothesis test. However, when this rule was used instead of that in Figure 8, the output was surprisingly dissimilar, despite the correlation between the left and right hand sides in (54) being 0.9998 in the simulation. Indeed, for the cases where $\frac{E[S]}{\sqrt{Var[S]}} < 10$, 54.5% had $n \geq n_\tau$. We suggest, then, that the applicability of our test has more to do with the distribution of the forecast probabilities and the ensuing distribution of Brier scores than it does the sample size, and that if our rule for eligibility is not fulfilled, simply gathering more forecasts will not ensure that it will be applicable eventually. This is not contradictory to the evidence in Figure 7, as we merely showed there that larger n improved the fit for cases which were already applicable.

5.4 Conclusion

In this section, we have derived a framework complementary to that of Bradley et al. (2008), where highly accurate estimates of the mean Brier score and its variance can be calculated using only the vector of forecasts $\boldsymbol{f}$, where Bradley et al. required the outcome vector $\boldsymbol{x}$ as well. This development allows a forecaster to evaluate the hypothesis that their model is calibrated, in the sense that the expected value of the outcome variable x_i for the i-th forecast equals the probability f_i provided for this event, $E[x_i] = f_i$. After a reasonable number of predictions, the forecaster can calculate the observed Brier score $\hat{S}$ and compare it to the null distribution of Brier scores, possibly deciding, at a given significance level, that $\hat{S}$ is significantly above the mean $E[S]$, thus rejecting the null hypothesis that the model is calibrated, and indicating that the forecaster would be well-advised to return to the model design and consider why forecasts are thus inaccurate.

6 Financial Applications

The paper in entirety has been framed in the diction of wagering and fixed-odds gambling games. The Kelly criterion (as well as the Brier Score, albeit to a more limited extent) belongs more naturally to this environment. However, the results derived in this paper are certainly applicable in more conventional financial markets as well.

The main characteristics of Kelly wagers is their binary nature – either the wager wins or it does not, in states with probability p and $1 - p$, respectively – and the predefined payout structure: a profit multiple of b in the winning state and -1 in the losing state. This structure is clearly not representative for stock markets, although the Kelly-criterion style of investing can indeed be applied there, see Rotando and Thorp (1992), among others (note also that, for instance, the binomial tree model for stocks (Cox et al., 1979) has a strong resemblance to the underlying mechanisms in the Value-at-Risk model proposed below, so there is certainly scope for further application here).

However, several other markets have this structure, or can at least be approximated by it. Binary options (S. Smirnov et al., 2024; Venter and De Jongh, 2022) have this exact structure: bets traded on financial markets of whether an asset's price will exceed some strike price or not. 0DTE (zero-days-to-expiry) options, which are traded for at most one day, also approach this structure. Event-based contracts (Wolfers and Zitzewitz, 2006; Bossaerts et al., 2024) – bets on economical, political, cultural and more general events – are of this nature as well. Polymarket, Kalshi, PredictIt, Betfair Exchange, among others, are seeing large increases in liquidity and trading volumes, increasing the scope for hedge funds and other shrewd financial investors to generate abnormal returns in unconventional markets.

It is also possible to express more general assets in this manner. Zero-coupon bonds, for instance, when held to maturity, pay a specified rate (which gives b), as long as the issuer does not default (the payment occurs with probability p). For investment-grade debt, p will be very close to 1, while junk bonds, on the other hand, will have lower p and correspondingly higher b. Conversely, credit default swaps, paying a fixed premium in case of default (again, b) use the same estimate p, paying off with probability $1 - p$.

All this to say, despite the discussions being framed in a wagering context, the results are wholly relevant for financial markets: a broad range of assets can be considered in the presented framework, given that their payoff is fixed up-front and is not guaranteed.

7 Conclusion

In this paper, we have presented three distinct risk-management frameworks for binary investments: investment scenarios where there are two possible outcomes with prespecified returns. We developed an extension of the Kelly criterion, optimising the stake size for an investment when several unresolved investments exist, which thus obscure the true reference capital. Remaining in the Kelly-criterion scope, we found that our separate expression for Value-at-Risk stake-sizing was equally calibrated. Finally, we derived a hypothesis test based on the Brier score for evaluating whether the binary investor's probability-generating model is accurate. In the two position-sizing frameworks, prices are given, and any metric for value derives from the probability p. The hypothesis test, then, is highly complementary, where the binary investor can manage their risk using either of the frameworks where relevant, while ensuring that the key input, p, is generally correct.

8 References

An, Kolmogorov. "Sulla determinazione empirica di una legge didistribuzione". *Giorn Dell'inst Ital Degli Att*, vol. 4, 1933, pp. 89–91.

Baker, Rose D, and Ian G McHale. "Optimal betting under parameter uncertainty: Improving the Kelly criterion". *Decision Analysis*, vol. 10, no. 3, 2013, pp. 189–99.

Bayes, Thomas. "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S". *Philosophical transactions of the Royal Society of London*, no. 53, 1763, pp. 370–418.

Bossaerts, Frederik, et al. "Price formation in field prediction markets: The wisdom in the crowd". *Journal of Financial Markets*, vol. 68, 2024, p. 100881.

Bradley, A., et al. "Sampling Uncertainty and Confidence Intervals for the Brier Score and Brier Skill Score". *Weather and Forecasting*, vol. 23, 2008, pp. 992–1006. https://doi.org/10.1175/2007WAF2007049.1.

Breiman, Leo, et al. "Optimal gambling systems for favorable games". *The Kelly Capital Growth Investment Criterion*, 1961, pp. 47–60.

Brent, Richard P. *Algorithms for minimization without derivatives*. Courier, 2013.

Brier, Glenn W. "Verification of forecasts expressed in terms of probability". *Monthly weather review*, vol. 78, no. 1, 1950, pp. 1–3.

Busseti, Enzo, et al. "Risk-constrained Kelly gambling". *arXiv preprint arXiv:1603.06183*, 2016.

Cox, John C, et al. "Option pricing: A simplified approach". *Journal of financial Economics*, vol. 7, no. 3, 1979, pp. 229–63.

Diebold, F., et al. "The Known, the Unknown, and the Unknowable in Financial Risk Management: Measurement and Theory Advancing Practice". 2010. https://doi.org/10.1515/9781400835287.

Ferrari, Silvia, and Francisco Cribari-Neto. "Beta regression for modelling rates and proportions". *Journal of applied statistics*, vol. 31, no. 7, 2004, pp. 799–815.

Geissinger, Emilie A, et al. "A case for beta regression in the natural sciences". *Ecosphere*, vol. 13, no. 2, 2022, e3940.

Gneiting, Tilmann, and Adrian E Raftery. "Strictly proper scoring rules, prediction, and estimation". *Journal of the American statistical Association*, vol. 102, no. 477, 2007, pp. 359–78.

Jorion, Philippe. *Value at risk: the new benchmark for managing financial risk*. McGraw-Hill, 2007.

Kelly, John L. "A new interpretation of information rate". *the bell system technical journal*, vol. 35, no. 4, 1956, pp. 917–26.

Kieschnick, Robert, and Bruce D McCullough. "Regression analysis of variates observed on (0, 1): percentages, proportions and fractions". *Statistical modelling*, vol. 3, no. 3, 2003, pp. 193–213.

Kim, Song-Kyoo. "Kelly Criterion Extension: Advanced Gambling Strategy". *Mathematics*, vol. 12, no. 11, 2024, p. 1725.

Krzysztofowicz, Roman, and Dou Long. "Beta likelihood models of probabilistic forecasts". *International Journal of Forecasting*, vol. 7, no. 1, 1991, pp. 47–55.

Law, David, and David A Peel. "Insider trading, herding behaviour and market plungers in the British horse–race betting market". *Economica*, vol. 69, no. 274, 2002, pp. 327–38.

Merton, Robert C. "Lifetime portfolio selection under uncertainty: The continuous-time case". *The review of Economics and Statistics*, 1969, pp. 247–57.

Murphy, Allan H, and Daniel S Wilks. “A case study of the use of statistical models in forecast verification: Precipitation probability forecasts”. *Weather and Forecasting*, vol. 13, no. 3, 1998, pp. 795–810.

Pearson, K. Notes on regression and inheritance in the case of two parents proceedings of the royal society of London, Vol. 58. 1895.

Peizer, David B, and John W Pratt. “A normal approximation for binomial, F, beta, and other common, related tail probabilities, I”. *Journal of the American Statistical Association*, vol. 63, no. 324, 1968, pp. 1416–56.

Rotando, Louis M, and Edward O Thorp. “The Kelly criterion and the stock market”. *The American Mathematical Monthly*, vol. 99, no. 10, 1992, pp. 922–31.

Schwarz, Gideon. “Estimating the dimension of a model”. *The annals of statistics*, 1978, pp. 461–64.

Sharpe, William F. “Mutual fund performance”. *The Journal of business*, vol. 39, no. 1, 1966, pp. 119–38.

Smirnov, Nickolay. “Table for estimating the goodness of fit of empirical distributions”. *The annals of mathematical statistics*, vol. 19, no. 2, 1948, pp. 279–81.

Smirnov, Sergey, et al. “Approximation and asymptotics in the superhedging problem for binary options”. *Annals of Finance*, vol. 20, no. 4, 2024, pp. 421–58.

Smoczynski, Peter, and Dave Tomkins. “An explicit solution to the problem of optimizing the allocations of a bettor’s wealth when wagering on horse races”. *Mathematical Scientist*, vol. 35, no. 1, 2010, pp. 10–17.

Sollis, Robert. “Value at risk: a critical overview”. *Journal of Financial Regulation and Compliance*, vol. 17, no. 4, 2009, pp. 398–414.

Sun, Qingyun, and Stephen Boyd. “Distributional robust Kelly gambling”. *arXiv preprint arXiv:1812.10371*, 2018.

Thorp, Edward O. “Portfolio choice and the Kelly criterion”. *Stochastic optimization models in finance*, Elsevier, 1975, pp. 599–619.

———. “The Kelly criterion in blackjack sports betting, and the stock market”. *Handbook of asset and liability management*, Elsevier, 2008, pp. 385–428.

Venter, Johannes Hendrik, and Pieter Juriaan De Jongh. “Trading binary options using expected profit and loss metrics”. *Risks*, vol. 10, no. 11, 2022, p. 212.

Wilks, Daniel S. “Sampling distributions of the Brier score and Brier skill score under serial dependence”. *Quarterly Journal of the Royal Meteorological Society*, vol. 136, no. 653, 2010, pp. 2109–18.

Winitzki, Serge. “Uniform approximations for transcendental functions”. *Computational Science and Its Applications—ICCSA 2003: International Conference Montreal, Canada, May 18–21, 2003 Proceedings, Part I 3*. Springer, 2003, pp. 780–89.

Wise, ME. “On normalizing the incomplete beta-function for fitting to dose-response curves”. *Biometrika*, vol. 47, nos. 1–2, 1960, pp. 173–75.

Wolfers, Justin, and Eric Zitzewitz. Prediction markets in theory and practice. 2006.

Wu, Jimmy Ming-Tai, et al. “Convert index trading to option strategies via LSTM architecture”. *Neural Computing and Applications*, vol. 37, no. 28, 2025, pp. 23047–64.

Xu, Yuhong. “Optimal growth under model uncertainty”. *North American Journal of Economics and Finance*, vol. 60, 2022, p. 101634.

Zeng, Caibin, and Yang Quan Chen. “Global Padé approximations of the generalized Mittag-Leffler function and its inverse”. *Fractional Calculus and Applied Analysis*, vol. 18, 2015, pp. 1492–506.

9 Appendix: Artificial Intelligence

For this thesis, generative AI has played a supportive role. The exclusive application has been to sanity-check our ideas with ChatGPT. Essentially, after writing a technical or interpretive section, or producing a next step in a mathematical deriviation (for instance, going from (15) to (16)), we verified with the AI tool that the analysis is both sound and reasonable. As such, it has been a quality-filtration solution, in the sense that it has helped ensure the consistent correctness of our own ideas, rather than having it generate our ideas. In the instances where our ideas were misaligned, the AI tool elucidated this, and after a back-and-forth session of prompting, so that we could understood our mistake, we ourselves rephrased the offending section or mathematical expression. Of course, we did not blindly trust the feedback that the AI provided: if in the back-and-forth we did not agree with its advice, we ignored it, but only after rephrasing our original prompt and still receiving the same response, which we believed was less correct than our original content. AI tools, then, have increased the quality of the thesis by minimising the probability of incorrect content: the ideas and writing are all original.