

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/359360804>

# Beating the market with a bad predictive model

Article in *International Journal of Forecasting* · March 2022

DOI: 10.1016/j.ijforecast.2022.02.001

---

CITATIONS

3

---

READS

339

2 authors:



Ondřej Hubáček

Czech Technical University in Prague

10 PUBLICATIONS 180 CITATIONS

SEE PROFILE



Gustav Šír

Czech Technical University in Prague

24 PUBLICATIONS 272 CITATIONS

SEE PROFILE

# Beating the market with a bad predictive model

## Abstract

It is a common misconception that in order to make consistent profits as a trader, one needs to possess some extra information leading to an asset value estimation more accurate than that reflected by the current market price. While the idea makes intuitive sense and is also well substantiated by the widely popular Kelly criterion, we prove that it is generally possible to make systematic profits with a completely inferior price-predicting model. The key idea is to alter the training objective of the predictive models to explicitly decorrelate them from the market, enabling to exploit inconspicuous biases in market maker’s pricing, and profit on the inherent advantage of the market taker. We introduce the problem setting throughout the diverse domains of stock trading and sports betting to provide insights into the common underlying properties of profitable predictive models, their connections to standard portfolio optimization strategies, and the, commonly overlooked, advantage of the market taker. Consequently, we prove desirability of the decorrelation objective across common market distributions, translate the concept into a practical machine learning setting, and demonstrate its viability with real world market data.

## 1 Introduction

The attempt to predict the future is at the core of any successful trading strategy. Traders continuously try to predict future prices of assets, such as stocks or commodities, to secure profits on their future positions. Similarly, bettors try to predict discrete future outcomes of random events, such as elections or sports matches, to secure profits on their realizations. Having a high-quality future prediction, one can confidently estimate the expected returns associated with the potential decisions to allocate assets and select securities for successful investments and speculations. The quality of each prediction then naturally reflects the amount of information used for the estimation. In an efficient market, it is assumed that all available information is already reflected in the current market price, and it is thus impossible to make any consistent profits through trading with such predictions [46]. It intuitively follows that in a partially inefficient market, where the market price reflects some but not all of the information, one needs the extra information, leading to predictions better than those reflected by the current market price, in order to profit.

Indeed, much of the existing literature on profiting in an inefficient market is based on the idea of well-informed investors, whose superior knowledge of the given domain enables to predict the future prices, or estimate outcome probabilities, more closely than the market, commonly quoting some aggregated estimation of the rest of the trading participants. This in turn enables them to secure profits using common investment strategies such as those of Markowitz [47] and Kelly [32], the performance of which is tightly connected to such an information advantage. While possessing more information leading to superior price estimation indeed enables to profit in this straightforward manner, we argue that in order to make positive returns, it is generally not necessary to estimate the true price better than the market. Instead, there is another and generally easier way to exploit market inefficiencies.

The core principle we utilize in this paper is the, commonly overlooked, inherent advantage of the market taker over the market maker, as the two price estimators are eventually penalized for their estimation errors in a very different manner. Consequently, in order to make profits, the trader does not need to estimate the true price of a mispriced asset better than the market and, instead of predictive pricing accuracy, should generally strive to optimize different performance measures. Specifically, we show that trader’s profits derived from a price estimator can be systematically increased by decreasing its partial correlation with the market, i.e. by decorrelating its residual estimation error w.r.t. the market. While this might seem similar in spirit to the common idea of profiting from using “extra information”, we demonstrate that the trader’s estimates can be easily based on inferior data sources and their information-theoretic value w.r.t. true price can be considerably lower than that reflected by the market. Consequently, the respective trader’s predictive model can thus be considered as “bad” by the means of regular quality measures.

symbol	description	stock trading	two-way betting
$a_j \in \mathcal{A}$	assets being traded on the market	stocks, securities	sport matches
$\{\alpha, \beta\}; \zeta \in \mathbb{Z}$	market sides, in two-way markets $ \mathbb{Z}  = 2$	buy/sell (bid/ask)	home/away (back/lay)
$\omega_{(a_j, \tau)}^\zeta; \omega_i \in \Omega$	opportunities to trade assets at time $\tau$	e.g. buy $a_j$ now	e.g. bet $a_j$ home
$\mathcal{D}_i \subset \mathcal{D}^*$	relevant data available for opportunity $\omega_i$		
$r : \mathcal{D}_i^* \rightarrow \mathbb{R}_+$	fundamental value $r$ of an asset	true value	true probability
$\mathbf{m} : \mathcal{D}_i^b \rightarrow \mathbb{R}_+$	market maker $m$ , or simply “the market”	market maker	bookmaker
$\mathbf{t} : \mathcal{D}_i^m \rightarrow \mathbb{R}_+$	market taker $t$	trader	bettor
$P_\Omega$	distribution of market opportunities	all possible trades	all possible bets
$R : \omega_i \mapsto r_i$	r.v. denoting fundamental asset value	price $r_i \in \mathbb{R}_+$	probability $r_i \in [0, 1]$
$M : \omega_i \mapsto m_i$	r.v. denoting market maker’s pricing	price $m_i \in \mathbb{R}_+$	probability $m_i \in [0, 1]$
$T : \omega_i \mapsto t_i$	r.v. denoting market taker’s pricing estimate	price $t_i \in \mathbb{R}_+$	probability $t_i \in [0, 1]$
$P(R, M, T)$	distribution of the price estimates (values)		
$\theta \in \Theta$	model parameters		
$\tilde{\mathbf{e}} : \mathcal{D}_i \rightarrow \tilde{\mathbf{E}}_i$	posterior distribution $\tilde{\mathbf{E}}$ estimator $\tilde{\mathbf{e}}$ (e.g. $\tilde{\mathbf{t}}$ )		
$\tau$	time associated with a trade		
$\epsilon \in \mathbb{R}_+$	market maker’s spread (margin)	spread	margin
$\rho_i \in \mathbb{R}$	trading returns (profit)	rate of return	return on investment
$W \in \mathbb{R}_+$	wealth of the market taker	capital	bankroll
$\mathbf{f} \in \mathbb{R}^n$	vector of the wealth allocations	allocation	wager

Table 1: Overview of the used notation with the correspondence between the stock and betting market settings.

While the proposed concept might seem theoretical in nature, it has important implications on common investment practices based on predictive modelling and portfolio optimization. Particularly, the current systems typically assume the asset price prediction and the subsequent investment decision making as two independent tasks. Namely, at first a statistical model is optimized to predict the future price as closely as possible which, translated into an expected rate of return, forms an input to the subsequent investment strategy designed to maximize some profit-based utility. While this typical workflow has the advantage of decomposing the problem, enabling for a more explicit risk management, we argue that it is considerably suboptimal since the two optimization tasks are not aligned, i.e. the predictive model accuracy optimization is oblivious of the subsequent profit-focused optimization. Consequently, maximizing model price estimation quality will commonly lead to suboptimal final profits. In contrast, with the proposed decorrelation objective, we target the profits more directly. This enables to generate consistent positive returns even with no information advantage and completely inferior price predicting models.

The paper is structured as follows. We introduce the necessary background, ranging from market modelling to portfolio optimization, in Section 2. In Section 3, we provide insights into essential conditions for model profitability, limitations of popular investment strategies, and the inherent market taker’s advantage. Section 4 then introduces the proposed concept of increasing profits through decorrelating from the market, its connection to the Kelly strategy, and translation into a machine learning objective. We then demonstrate the concept in a practical setting in Section 5. Finally, we review the related work in Section 6 and conclude in Section 7.

## 2 Background

This background section introduces the problem setup (Section 2.1), and basic concepts from stock/betting markets, machine learning, and portfolio optimization. A reader familiar with these areas may then wish to skip directly to Section 3.

### 2.1 Problem Setup

In this paper we assume a standard market setting where two or more parties participate in an exchange of some forms of assets. For the sake of this paper it is not important to distinguish between securities or other financial instruments, and we further refer to all these jointly as *assets*. While the analysis in this paper covers a wide

variety of generic market settings, ranging from commodities to prediction markets, we will mostly consider two diverse enough examples of (i) stock and (ii) sports betting markets to keep the explanation grounded. Similarly, we do not restrict to any particular asset class, which can vary widely from financial derivatives to fantasy sports, however for clarity of explanation, we will consider a standard “bid-ask” setting for the stock, and a corresponding two-way<sup>1</sup> sports betting market, such as predicting the winner of a basketball game. Table 1 provides an overview of the joint notation used throughout the paper and across the two domains.

**Opportunity** In a two-way market, each asset  $a \in \mathcal{A}$  tradeable at time  $\tau$  corresponds to an opportunity to buy ( $\alpha$ ) and sell ( $\beta$ ) the asset, respectively. Given a certain time  $\tau$ , this covers bids and asks for a certain stock, as well as bets on the home ( $\alpha$ ) and away ( $\beta$ ) team win of a certain sport match. We will further refer to all existing market positions to trade a certain asset  $a$  at a certain time  $\tau$  jointly as *opportunities*  $\omega_{(a,\tau)}^\zeta \in \Omega$ . We further treat each opportunity  $\omega_{(a,\tau)}^\zeta$  as an independent trading unit denoted  $\omega_i$  since, for the analysis in this paper, it is not necessary to distinguish the asset or time that each opportunity is associated with.<sup>2</sup> Nevertheless, where necessary, we will distinguish the side of the market  $\zeta \in \{\alpha, \beta\}$  an opportunity  $\omega_i^\zeta$  is to be traded at.

**Price** Each opportunity  $\omega_i$  is then associated with a certain price. In a stock market, it is simply the current price set up by the market maker to trade the asset, as used in the usual sense. In the prediction markets, the price reflects the bookmaker’s perceived probability of an outcome to occur. It is also commonly referred to as “odds”<sup>3</sup> that determine the potential payouts received from a wager. While the particular calculations naturally differ between monetary prices and odds (Section 2.6), where it is not necessary to distinguish between the two, we will refer to these jointly as *prices*. We note that the market price of a buying opportunity  $\omega_i^\alpha$  may be different from the price to sell  $\omega_i^\beta$  (Section 2.3), and similarly for the odds.

**Fundamental Value** We further assume that each opportunity  $\omega_i$ , i.e. an asset being traded at a given time, has some fundamental *true value*  $r_i \in \mathbb{R}_+$ , which can be expressed as a positive real number and by those means compared to some price estimate (Section 2.5) in the same unit of measurement. For instance in a stock market, a fair market price should reflect the true expected value of the underlying corporation, given the current information context. Similarly in the betting market, fair odds would reflect the true (inverse) probability of the associated outcome to happen. Note that the true value of an asset  $a$  at time  $\tau$  is always the same for both sides ( $\alpha, \beta$ ) of the market. While the concept of a fundamental value  $r$  can be seen as somewhat speculative, since it cannot be directly measured, we note that we merely require its theoretical existence for defining market efficiency (Section 2.2) and the related concepts (Section 2.6).<sup>4</sup>

**Makers and Takers** We commonly distinguish two types of the exchange participants as (i) market *makers*  $m$ , also referred to as bookmakers, and (ii) market takers  $t$ , also referred to as *traders*. The market makers continuously quote both sides ( $\alpha, \beta$ ) of the market at certain price levels  $m_i$  resulting into trading opportunities  $\omega_i \in \Omega$ . By continuously *providing* such quotes of asset prices, the makers bring constant liquidity to the market. The market takers are then *selecting* from the existing opportunities  $\omega_i$  to issue specific buy and sell orders. We generally consider the problem analysed in this paper as a two-player game between a market maker  $m$  and taker  $t$ , where each player possesses a certain pricing policy  $\Omega \rightarrow \mathbb{R}_+$  over the market distribution of opportunities  $P_\Omega$  given by the game (world) environment.

**Beating the market** We then aim to design a strategy for the role of the market *taker*, i.e. some generic trader  $t$ , to make positive profits against some particular market maker  $m$ . For example, such strategy would allow a bettor to make consistent long-term profits while wagering against a particular bookmaker. Note that this is a zero-sum setting, where the profits of the market taker are at the direct expense of the market maker,

<sup>1</sup>While we limit the explanation to two-way markets for clarity, the main concept proposed in this paper is directly applicable to  $n$ -way markets, too.

<sup>2</sup>i.e. two opportunities  $\omega_i$  and  $\omega_j$  over the same asset at two different times will be treated the same as two opportunities  $\omega_i$  and  $\omega_j$  over two different assets. We note that there typically are dependencies between the opportunities that could theoretically be used to further improve the underlying techniques, however we do not exploit these in this paper.

<sup>3</sup>which we use more specifically later to refer to the app. inverse of the probability value (Section 5.1)

<sup>4</sup>We also commonly use the term true fundamental value interchangeably with the mean future price since, assuming the same information context, the latter can be expected to converge to the former in the partially efficient markets assumed (Section 2.2).

as the fundamental value is objectively the same for all the participants.<sup>5</sup> This is also the typical setting for a closed system of traders operating within the same time-frame, such as in intra-day trading, futures, options, or prediction markets. The important aspect here is that the only way to make positive profits in this setting is to “beat the average trader”, quoting the market price, by some margin (Section 2.3). In this paper, we then utilize the phrase “beating the market” to refer to profiting in this competitive setting.

## 2.2 Market Efficiency

In a (strongly) efficient market, the current market price  $m \in \mathbb{R}_+$  of an asset  $a$  reflects all existing information, making it impossible for any trader to make consistent profit by outsmarting the market<sup>6</sup> [16]. Particularly, the price  $m_i$  of each opportunity  $\omega_i$  being traded needs to be an *unbiased* estimate (Section 2.5) of its fundamental true value  $r_i$  [59]. Note that this does not imply the price  $m_i$  to be equal to the true value  $r_i$ , but merely that the error of the estimation  $m_i = \hat{r}_i$  is fully due to an irreducible variance which is *completely random* (Section 2.5) [63]. The inherent randomness of the error then ensures that no trader can consistently beat the market to secure systematic profits.

In real world liquid markets, profit-maximizing traders continuously search for under-priced bids or over-priced asks to secure their profits, pushing the market price to quickly converge to the true value in the process. Thanks to this self-regulating mechanism, the market inefficiencies tend to vanish quickly over time [68, 17]. We note that the idea of an efficient market is a rather theoretical one, since in a market that would become completely efficient, the traders would lose the incentive to search for inefficiencies, in turn making the market inefficient again. Any real world market can thus be hardly considered as perfectly efficient, nevertheless some measurable degree of efficiency, such as statistical unbiasedness of the mean prices (Section 3.4), is typically present in liquid markets [21]. In this paper, we will further consider the most realistic setting of a partially (in-)efficient market where the market price  $m$  is a very good estimate of the true value  $r$ , but not a perfect one.

## 2.3 Market Maker’s Advantage

Market makers are essential to trading by providing constant liquidity to both sides  $(\alpha, \beta)$  of the market, for which they are typically favored by the exchange operator in terms of fees and commissions. However, the position of a market maker is a difficult one, since she needs to constantly price the assets as accurately as possible. An estimate too low can lead to exploitation on the buy side  $\alpha$ , and an estimate too high on the sell side  $\beta$ , respectively. While the true value  $r$  is an unknown real number, it would be in principle impossible to avoid exploitation in full.

To improve her position and secure profit, the market maker incorporates a so called *spread*  $\epsilon$  (margin) into her estimates  $M$ , causing the offered bid opportunity  $\omega_i^\alpha$  price to differ from the ask  $\omega_i^\beta$  price by some  $2\epsilon$ . The spread  $\epsilon$  then works as a safety patch on the market maker’s estimation error, preventing from one-sided exploitations by the market takers aiming at the discrepancy from the true value. A safe strategy for the market maker is then to set her estimate and spread such that  $\forall i : m_i - \epsilon < r_i < m_i + \epsilon$ , making it impossible for any trader to make any profit, while securing herself an instant profit of  $2\epsilon$  for each pair of orders traded at the opposing sides of the book.

However, given that  $m = \hat{r}$  is merely an estimate, it is still possible for the true value  $r$  to fall outside the  $(m - \epsilon, m + \epsilon)$  region.<sup>7</sup> To further mitigate possible exploitation by the traders, the market maker can continuously adapt her estimate to the traders’ behavior. That is she can responsively move her estimate  $m$  once the demand of one side of the market starts to prevail, indicating expected value perceived by the traders, stemming from a possibly erroneous price estimate  $m$ . In the ideal case where she is continuously able to maintain a perfectly balanced book, so that half of the orders fall on the buy side and the other half on the sell side respectively, she is again guaranteed a unit profit of  $\epsilon$  per trade. Note that the market maker’s profit in this case is independent of the true price  $r$ .

On the other hand, the market maker can theoretically digress from this purely reactive position to actively speculate against the takers’ opinions and aim at a profit even higher than  $\epsilon$ , at the cost of involving risk

<sup>5</sup>We use this to remove trivial non-zero-sum settings, where all the participants can mutually profit from trading, from the scope. While we acknowledge that different traders may subjectively evaluate an opportunity  $\omega_i$  differently, e.g. due to distinct preferences for risk tolerance (Section 2.6), this would allow to trivially avoid the hard part of the introduced problem.

<sup>6</sup>Note that this does not imply that one needs comparably more information to beat an inefficient market.

<sup>7</sup>Naturally, this could be mitigated by increasing the  $\epsilon$ , however a margin set too large will discourage investors from trading, consequently removing the market maker’s profit, too.

stemming from her, possibly erroneous, estimate  $m$ . This effectively allows the market maker to speculate on the true price, which can lead to higher expected profits in settings where she possesses a superior price prediction model.<sup>8</sup> This is very common, for instance, with bookmakers in the sport prediction markets [52]. Naturally, the market makers can also combine the aforementioned methods.

We note we often omit the spread  $\epsilon$  from calculations and simulations further in this paper for simplicity of explanation<sup>9</sup>, since it does not affect the main principles introduced in this paper. The underlying assumption is that the spread constitutes an independent offset on the market prices, hence decreasing the resulting profits, but not interfering with the price estimation problem itself.

## 2.4 Market Modelling

Market modelling generally refers to the approach of fitting a statistical estimator  $\mathbf{e}$  to the available market-related data  $\mathcal{D}$  in order to capture its true distribution  $P$ . Possession of such a model then enables to answer all sorts of statistical queries, including the essential estimation of the real value  $R$  of market opportunities  $\Omega$ .

**Data** The market-related data  $\mathcal{D}$  may come from various external sources such as news and economic signals and indicators, as well as from the market itself (e.g. the traders' behavior). Recall that in an efficient market, all relevant data  $\mathcal{D}^*$  must be used by the market maker  $m$  for pricing of each opportunity  $\forall i : \mathbf{m}(\mathcal{D}_i^*) \mapsto m_i$ , however, in practice it is more likely that  $\mathcal{D}_i^m \subset \mathcal{D}_i^*$ . The market takers can theoretically use the same data as the market maker  $\mathcal{D}^t = \mathcal{D}^m$ , although their sources are typically more limited (e.g. Section 5.2). However, they commonly strive to gain at least some information advantage by obtaining data which are not available to  $m$ , i.e.  $\mathcal{D}^t \setminus \mathcal{D}^m \neq \emptyset$ , and thus not reflected in the market price. Generally if  $\mathcal{D}^t \subset \mathcal{D}^m$ , such information advantage is completely missing (e.g. Section 5.5), making it impossible to beat the market through superior price estimations<sup>10</sup>, unless using a superior model.

**Modelling** The models used to fit the data are generally some  $\theta$ -parameterized functions, the properties of which also influence the quality of the estimation. Ultimately, one would strive to model the whole joint distribution over the data  $\mathcal{D}$ , enabling to answer all possible probabilistic queries about the domain, consequently leading to truly optimal investment decision making. Nevertheless, the common investment strategies (Section 2.7) are typically based merely on estimates of returns from the opportunities  $\omega_i$  at hand, which restricts the task to modelling of a conditional of the true value  $P_\Omega(R|\mathcal{D}_i)$ .

For instance at the stock market, an estimate of the true value  $r_i$  at a given time can be modelled from a retrospective observation of local evolution of the price time series, i.e. one can use the model to repeatedly predict the (mean) future market price within some time interval (information context) used for trading. Similarly, an estimate of the true probability  $r_i$  in the betting market can be derived from repeated observations of the outcomes of the associated stochastic events aggregated over a large enough sample from the market.

**Real Value** Generally, the true value  $r_i$  of an asset  $a$  at time  $\tau$  can be a distribution itself, such as in the prediction markets. For instance, in an  $n$ -way betting market where there are  $n$  possible outcomes, the true value of each asset is the probability distribution over the  $n$  discrete outcomes. However, in the case of the two-way market with binary exclusive outcomes we use for demonstration, it is a simple Bernoulli distribution with a single parameter  $r_i$ . Therefore, the true value  $r_i$  of an opportunity  $\omega_i$  can still be considered as a single number  $r_i \in [0, 1]$ , since the value of the complementary opportunity is simply determined by  $1 - r_i$ . Consequently, there is a direct correspondence with the bid-ask stock market setting which we exploit for comprehensibility in this paper.<sup>11</sup> It follows that we also consider both sides of the stock market to be available, i.e. with a short selling option, for consistency.<sup>12</sup> As a result, we can consider every opportunity  $\omega_i$  to be associated with a single real value  $r_i \in \mathbb{R}_+$  in this paper.

<sup>8</sup>Since the market maker typically needs an in-depth knowledge of the market to operate, she can often reasonably expect to also possess price estimates superior to the average trader.

<sup>9</sup>except for the actual experiments with real data (Section 5) where the spread is naturally present.

<sup>10</sup>However, we note that it is still possible to make profits in such a scenario (Section 3.2).

<sup>11</sup>However the proposed decorrelation concept (Definition 4.1) can be directly extrapolated into  $n$ -way markets, too.

<sup>12</sup>although this is also not necessary for the approach to work.

**Mispricing** We assume a market that is not fully efficient (Section 2.2), i.e. there must be some mispriced opportunities which further need to be identifiable by some systematic means. The goal of the traders is then to identify such opportunities where  $r_i \notin (m_i - \epsilon, m_i + \epsilon)$  by comparing the market price  $m_i$  to their own estimate  $t_i$ . Should the market price  $m_i$  for an opportunity  $\omega_i$  actually deviate from the true value  $r_i$  in a non-random manner, such mispricing can be turned into positive returns. While the concrete calculation of the expected returns differs across the two market settings (Section 2.6), for the sake of this work it does not matter whether we try to predict future price distribution or unknown outcome probability distribution, as in both cases their correct estimation by  $t$  leads to profits of the trader  $t$  in the (partially) inefficient market setting assumed. Naturally, this depends on the qualities of the underlying price estimators.

## 2.5 Price Estimators

A price estimator  $e$  is generally a  $\theta$ -parameterized function mapping some input data  $\mathcal{D}_i$  associated with an opportunity  $\omega_i$  onto a point price estimate  $\hat{r}_i \in \mathbb{R}_+$

$$e : \mathcal{D}_i \mapsto \hat{r}_i \quad (1)$$

However, every prediction is associated with a certain level of uncertainty, which is either inherent<sup>13</sup> or stemming from the missing information at the time of making ( $\mathcal{D}_i^t \subset \mathcal{D}_i^*$ ). One might thus want to quantify the uncertainty by associating each possible estimate  $\hat{r}_i$  for an opportunity  $\omega_i$  with a probability, resulting into a posterior distribution estimation

$$\tilde{e} : \mathcal{D}_i \mapsto \tilde{R}_i \quad (2)$$

where  $\tilde{R}_i$  is the estimated price distribution  $\tilde{P}(\hat{R}_i | \Omega = \omega_i)$  for a single opportunity  $\omega_i$ . Such distribution  $\tilde{R}_i$  can be estimated, e.g., from histograms of past prices or event outcomes, marginalized over the same or “similar” conditions ( $\mathcal{D}_i$ ) [1]. When a point price estimate is needed, such as when we need to actually trade an asset at a particular price, it is common to take an expected value from  $\tilde{R}_i$ , calculated by multiplying each point estimate  $\hat{R}_i = \hat{r}_j$  with its associated probability estimation  $e_j$  (or probability density  $e(\hat{r})$ ):

$$\mathbb{E}_{\tilde{R}_i}[\tilde{e}(\mathcal{D}_i)] = \sum_j e_j \cdot \hat{r}_j \quad \text{or alternatively} \quad \mathbb{E}_{\tilde{R}_i}[\tilde{e}(\mathcal{D}_i)] = \int \hat{r} \cdot e(\hat{r}) \, d\hat{r} \quad (3)$$

We note that most of the machine learning models provide directly the point estimates, realizing some functional mapping  $\Omega \rightarrow \mathbb{R}$ .

**Point Estimates** The aforementioned (middle) market price  $m_i \in \mathbb{R}_+$  of an opportunity  $\omega_i$  can then be thought of as the market maker’s point estimate  $m_i = m(\mathcal{D}_i^m)$  of the true price  $r_i$ . Similarly, the market taker will also try to predict the actual true value, which we represent with her own point estimate  $t_i = \mathbb{E}_{\tilde{T}_i}[t(\mathcal{D}_i^t)]$ , or simply  $t_i = t(\mathcal{D}_i^t)$ . By the definition of her role, the market maker  $m$  continuously evaluates each opportunity  $\forall i : \omega_i \rightarrow m_i$ . We generally assume that the trader  $t$  also has the ability to estimate (predict) the true price  $\forall i : \omega_i \rightarrow t_i$  of each opportunity in the market.<sup>14</sup> The trader  $t$  then must possess her own estimate, i.e. be generally different from  $m$ <sup>15</sup>, and the true value  $r$  must be unknown at the time of trading, otherwise there would be no incentive to trade. Note that the true value is often unknown in principle, i.e. even retrospectively, such as the probability in the betting markets, indicating the need for statistical treatment of the problem.

**Definition 2.1.** We can now generalize the reasoning about individual opportunities and estimates to reasoning over the whole joint *market distribution*  $P_\Omega(R, M, T)$  capturing the relationships between the estimates across a whole set of opportunities  $\omega_1, \dots, \omega_n \in \Omega$  generated by the market environment. For each such opportunity  $\omega_i$ , we will thus operate with 3 distinct values we will refer to as (i) the true value  $R = r_i$ , (ii) the market maker’s estimate  $M = m_i$  and (iii) the market taker’s estimate  $T = t_i$ .

Given the uncertainty, it follows that both the  $m_i$  and  $t_i$  estimates are always going to be to some extent erroneous, where the former guarantees existence of mispriced assets. Note that this is a necessary but insufficient condition for market inefficiencies to exist (Section 2.2).

<sup>13</sup>Note that much of the information is often not just missing but principally unavailable, rendering the prediction problem inherently stochastic.

<sup>14</sup>Alternatively, a systematically selected subset can be considered instead.

<sup>15</sup>since e.g. copying the estimate from the market maker would not lead to any trading incentive.

**Optimization** Typically, one optimizes the estimator by tuning its parameters  $\theta \in \Theta$  to fit some historical market data  $\mathcal{D}^t$  relevant for the prediction of the underlying true value  $R$ . Estimation of unknown parameters  $\theta$  from empirical data  $\mathcal{D}$  is then one of the key problems in statistics [22]. There are several views on the problem. One class of approaches is to maximize probability of the observed data with methods such as maximum likelihood  $p(\mathcal{D}|\theta)$  or maximum a-posteriori  $p(\theta|\mathcal{D})$  estimation. One can also directly search for an estimator with some desired target properties, such as minimum-variance unbiased estimator [76] or best linear unbiased estimator [26]. A very common methodology is Bayesian estimation, both with or without an informative prior, where one tries to minimize (posterior) expectation of some error function [7]. We take the latter approach, while noting that there are close connections between all the approaches.

**Estimation Error** The quality of an estimator  $e$  can then be expressed through its empirical error  $err(e)$  over some set of opportunities  $\Omega$ . Since we assume the role of the trader  $t$ , we further present error measurements between the true values  $R$  and the trader’s estimates  $T$ . Some of the most popular error measures then include the mean square error ( $MSE$ ):

$$MSE_{\Omega}(R, T) = \mathbb{E}[(T - R)^2] = \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} (t_i - r_i)^2 \quad (4)$$

which is commonly used for regression tasks, such as the prediction of the true price  $r_i \in \mathbb{R}_+$ , and the mean crossentropy ( $XENT$ ):

$$XENT_{\Omega}(R, T) = \mathbb{E}_R[-\log(T)] = \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} \sum_{j \in \text{outcomes}} r_j \cdot \log(t_j) \quad (5)$$

which is commonly used for classification tasks, such as predicting one of the discrete outcomes (e.g.  $\{home, away\}$ ) of a sports match. Note that these correspond to the market sides  $\zeta$ . In the case of the two-sided markets corresponding to binary event outcomes, we can rewrite  $XENT$  as

$$XENT_{\Omega}(R, T) = \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} r_i \cdot \log(t_i) + (1 - r_i) \cdot \log(1 - t_i) \quad (6)$$

which can then also be also understood as a regression of the underlying value  $r_i$  of each opportunity  $\omega_i$ .

These particular error functions are of special interest as minimizing  $XENT$  generally corresponds to maximizing the data (log-)likelihood, while  $MSE$  corresponds to maximizing the data likelihood with a linear gaussian model [37]. The crossentropy error is then also closely linked to a common measure of “distance” between probability distributions known as Kullback-Leibler divergence [35], which is defined as

$$D_{KL}(R||T) = - \sum_i r_i \cdot \log \frac{t_i}{r_i} \quad (7)$$

since

$$XENT(R, T) = H(R) + D_{KL}(T||R) \quad (8)$$

where  $H(R)$  is the entropy of the true value distribution  $R$ . Considering that true distribution being estimated does not change, its entropy  $H(R)$  can be considered a constant, rendering the cross-entropy error  $XENT(R, T)$  minimization equivalent to minimizing the the KL-divergence  $D_{KL}(R||T)$ , sometimes also referred to as the relative entropy [7] (see Section 3.1 for further connections).

**Bias and Variance** The estimation error can be commonly thought of in terms of bias and variance, which are key concepts in the analysis of estimators’ quality [37]. Bias quantifies the expected difference between the price estimate and the true value:

$$\text{Bias}_R[T] = \mathbb{E}[T] - R \quad (9)$$

while variance is the expected squared distance from the mean estimate:

$$\text{Var}[T] = \mathbb{E}[(T - \mathbb{E}[T])^2] \quad (10)$$

While we naturally want to minimize both the quantities, this is typically hard to do as they commonly reflect opposing criteria, giving rise to the so called “bias-variance dilemma”, which is a central problem in statistical



learning [22]. A common approach in statistics is to strive for a minimum-variance unbiased estimator, giving good results in most practical settings [76]. The constraint for unbiasedness reflects the necessity for the estimator not to differ systematically from the true price, i.e. the estimation should be correct on average. This constraint is very natural in the proposed market setting, since a systematically biased market price would be easy to directly detect and exploit (Section 3.4). Note also that any minimum-variance mean-unbiased estimator minimizes the  $MSE$ , which can be alternatively expressed as:

$$MSE(T, R) = \mathbb{E}[(T - R)^2] = (\mathbb{E}[T] - R)^2 + \mathbb{E}[(T - \mathbb{E}[T])^2] = \text{Bias}_R[T]^2 + \text{Var}[T] \quad (11)$$

## 2.6 Expected Returns

Being able to predict the future price in a stock market, or estimate the true probability in a prediction market, can be directly turned into positive returns in trading. Note we do not explicitly distinguish between the mean future price and true value, as in the fairly efficient market setting assumed (Section 2.2), the market price  $M$  cannot systematically deviate from the true value  $R$  for too long. Similarly in the prediction markets, the relative frequency of the observed outcomes will approach the true probability distribution in the long run. Being able to correctly estimate either of these thus guaranties systematic profits, even though actual profits might deviate from the expectation in short term.

Since the total profit is dependent on the actual amount invested, which is yet to be determined by the investment strategy (Section 2.7), the traders commonly assess *relative* profitability of individual opportunities through measures such as rate of return or, without assuming any time period, return on investment (ROI), which we further denote as  $\rho_i$ . In general,  $\rho_i$  is simply the relative return made from a unit investment w.r.t. the market price  $m_i$  and the true value  $r_i$ .

The uncertain, stochastic nature of the prediction problem renders the value estimates  $\hat{r}_i$  for a particular  $\omega_i$  as random variables  $\hat{r}_i \sim \tilde{R}_i$  (Section 2.5). Consequently, a return  $\rho_i$  derived from such an estimate  $\hat{r}_i$  will be a random variable, too. Instead of the actual  $\rho_i$  we can thus again calculate with its expectation  $\mathbb{E}[\rho_i]$  w.r.t. the used estimator  $\mathbf{e}$ , further denoted as  $\mathbb{E}_{\mathbf{e}}[\rho_i]$ . For the role of a trader  $t$ , we can then define an expected  $\rho_i$  of an opportunity  $\omega_i$  based on estimator  $\mathbf{t}$  of the true price  $r_i$  of some underlying asset as

$$\mathbb{E}_{\mathbf{t}}[\rho_i^\alpha] = \mathbb{E}_{\tilde{T}_i} \left[ \frac{\tilde{T}_i - M(\omega_i^\alpha)}{M(\omega_i^\alpha)} \right] = \frac{t_i - m_i^\alpha}{m_i^\alpha} \quad \mathbb{E}[\rho_i^\beta] = \mathbb{E}_{\tilde{T}_i} \left[ \frac{M(\omega_i^\beta) - \tilde{T}_i}{M(\omega_i^\beta)} \right] = \frac{m_i^\beta - t_i}{m_i^\beta} \quad (12)$$

where  $m_i^\zeta$  is the market price at which the trader executes the buy ( $\zeta = \alpha$ ) or (short-)sell ( $\zeta = \beta$ ) orders, respectively. For example when we buy a stock for \$100 with the expectation to sell it later for  $t_i = \$150$  on average, our expected ROI is  $\rho_i = 0.5$  (50%).

In a prediction market, the market price (odds) directly reflect the relative returns, and so the expected  $\rho_i$  of an opportunity  $\omega_i$  based on a probability estimated by  $\mathbf{t}$  can be calculated as

$$\mathbb{E}_{\mathbf{t}}[\rho_i^\alpha] = \mathbb{E}_{\tilde{T}_i} \left[ \frac{\tilde{T}_i}{M(\omega_i^\alpha)} - 1 \right] = \frac{t_i}{m_i^\alpha} - 1 \quad \mathbb{E}[\rho_i^\beta] = \mathbb{E}_{\tilde{T}_i} \left[ \frac{\tilde{T}_i}{M(\omega_i^\beta)} - 1 \right] = \frac{1 - t_i}{m_i^\beta} - 1 \quad (13)$$

where  $m_i^\zeta$  is the probability estimated by the bookmaker for the binary outcomes of home ( $\zeta = \alpha$ ) and away ( $\zeta = \beta$ ) team wins, respectively. For example, a bet of \$100 on an outcome with estimated probability  $t_i = \frac{3}{4}$ , and associated bookmaker's (decimal) odds of  $\frac{1}{M(\omega_i)} = 2.0$ , i.e. yielding a net return of \$100 if realized, and a loss of \$ - 100 if not, will also result into a ROI of  $\rho_i = 0.5$  (50%).

Although the average returns should converge to the true expected returns in the long run, these can still be very different from the predicted expected returns since generally  $\mathbb{E}_{\mathbf{t}}(\rho_i) \neq \mathbb{E}_{\mathbf{r}}(\rho_i)$ . The discrepancy is of course conditioned by the quality of the predictor  $\mathbf{t}$  w.r.t. the true  $\mathbf{r}$ . The approach to minimize the prediction error (Section 2.5) then seems very natural, and there are also some theoretical guaranties like, for instance, in the case where  $XENT(R, T) < XENT(R, M)$ , i.e. the investor possesses a price prediction model superior to the market maker in terms of cross-entropy, we are guaranteed to make long-term profits with optimal investment routines such as the Kelly strategy (Section 2.7).

**Risk and Utility** To explicitly account for the discussed uncertainty involved in trading, the concept of *risk* assessment has been proposed. This means that instead of direct maximization of the expected returns, one should strive for a balance between the expected profit and risk, stemming from the uncertainty. The risk can then be quantified by statistical means such as the variance of the expected profit [47] or probability of a drawdown [12]. Note that apart from the quantifiable risk, there is also a structural risk stemming from the fact that, similarly to the expected returns, the assessment of risk is based on merely estimated parameters.

Not all investors then share the same preferences to balance the expected profit and risk in the same manner. To incorporate individual preferences into the decision making, the concept of a *utility function*  $u$  has been proposed to steer the investment optimization process. A utility function generally maps each alternative onto a real number, defining a total ordering over some set of alternative investments. In our case it is some monotonically growing function  $u$  transforming the net returns  $\rho$  into a new real quantity  $u(\rho)$  to be optimized.<sup>16</sup> The concept of risk is then closely connected to utility, as maximizing any concave utility function directly reflects a preference for risk aversion [3].

## 2.7 Investment Strategies

An investment strategy can be seen as the final step of the traders’s workflow. Given the expected returns from individual trading opportunities present at a time, the trader needs to decide on how to *allocate* her wealth across the available opportunities  $\Omega$  in order to optimize her utility  $u$ , i.e. some requested trade-off between expected returns and risk. Formally, the investment strategy is a function  $s$  mapping a vector of opportunities  $\omega$  associated with the estimated returns  $\rho$  onto a wealth allocation vector  $f$ :

$$s : \rho \mapsto f \quad \text{i.e.} \quad s : \mathbb{R}^n \rightarrow \mathbb{R}^n \quad (14)$$

The vector of the wealth allocations  $f$ , corresponding to portions of some current wealth  $W$ , is then often referred to as a *portfolio* over the opportunities  $\omega$ . We note that the trader might also want to leave certain fraction of the wealth  $W$  aside, which is commonly incorporated by introducing an extra opportunity (“cash option”) with a constant zero<sup>17</sup> net return. Generally, there are several approaches to the wealth allocation problem and we will briefly review some of the most popular ones [41].

### 2.7.1 Uniform Strategy

The most trivial investment strategy is a uniform, unit-staking strategy where one independently allocates the same absolute amount  $d$  on every opportunity with an assumed positive expected return:

$$s : \rho_i \mapsto \begin{cases} f_i = d, & \text{if } \rho_i > 0 \\ f_i = 0, & \text{otherwise} \end{cases} \quad (15)$$

Despite being very naive, this strategy is also most robust against estimation errors [58], since the allocation simply remains constant no matter the circumstances. Note that the individual investments  $d$  are not considered as fractions relative to the current wealth  $W$  here, and so a small enough unit  $d \ll W$  has to be chosen so that it is possible to invest into all profitable opportunities. Given that the allocated unit  $d$  is small enough, this strategy will typically also have a very conservative risk profile, nevertheless the expected portfolio profits can be way below optimal. The size of  $d$  then remains a hyperparameter the choice of which is left to the user.

### 2.7.2 Modern Portfolio Theory

A more principled approach is that of the Modern Portfolio Theory (MPT) [47] which strives to balance optimally between the expected return and risk. The general idea behind MPT is that a portfolio  $f^1$ , i.e. a vector of asset capital allocations  $f = f_1, \dots, f_n$  over some opportunities  $\omega_1, \dots, \omega_n$ , is superior to  $f^2$ , if its corresponding expected return  $\rho$  (Section 2.6) is at least as great

$$\mathbb{E}_t[\rho \cdot f^1] \geq \mathbb{E}_t[\rho \cdot f^2] \quad (16)$$

<sup>16</sup>Given the stochastic setting, we will again consider expected utility  $\mathbb{E}[u(\rho)]$  instead.

<sup>17</sup>This might also be further extended by introducing a slightly positive net return instead to emulate the option of, e.g., storing the money in a bank account with an interest.

and a given risk measure  $risk : \mathbb{R}^n \rightarrow \mathbb{R}$  of the portfolio w.r.t. the returns is no greater

$$risk_{\mathbb{E}_t[\rho]}(\mathbf{f}^1) \leq risk_{\mathbb{E}_t[\rho]}(\mathbf{f}^2) \quad (17)$$

This creates a partial ordering on the set of all possible portfolios. When combined into a joint utility, we can trade-off the expected profit vs. risk by maximizing the following

$$\underset{\mathbf{f} \in \mathbb{R}^n}{\text{maximize}} \quad (\mathbb{E}_t[\boldsymbol{\rho} \cdot \mathbf{f}] - \gamma \cdot risk_{\mathbb{E}_t[\rho]}(\mathbf{f})) \quad (18)$$

where  $\gamma$  is a hyperparameter reflecting the user's preference for risk.

In the most common setup, the *risk* of a portfolio  $\mathbf{f}$  is measured through the expected total variance of its profit  $\text{Var}[\boldsymbol{\rho} \cdot \mathbf{f}] = \mathbf{f}^T \Sigma \mathbf{f}$ , based on a given covariance matrix  $\Sigma_n^n$  of returns of the individual opportunities, which can be again estimated from historical data (Section 2.4). MPT can then be expressed as the following constrained maximization problem:

$$\begin{aligned} & \underset{\mathbf{f} \in \mathbb{R}^n}{\text{maximize}} \quad \mathbb{E}_t[\boldsymbol{\rho} \cdot \mathbf{f}] - \gamma \cdot \mathbf{f}^T \Sigma \mathbf{f} \\ & \text{subject to} \quad \sum_{i=1}^n f_i = 1 \end{aligned} \quad (19)$$

Note that the capital allocations sum up to one as they simply reflect fractions of the current bankroll  $W$  (including the fraction left in “cash”), and they can possibly be negative if short selling is enabled.

The main weakness of MPT is that the variance of profit is hardly a good measure of risk for profit distributions other than Gaussian [60]. Apart from the variance  $\text{Var}[\mathbf{w}]$  of the potential net returns  $\mathbf{w} = \boldsymbol{\rho} \cdot \mathbf{f}$ , different risk measures have been proposed [47], such as standard deviation  $\sigma(\mathbf{w}) = \sqrt{\text{Var}[\mathbf{w}]}$  and coefficient of variation  $CV(\mathbf{w}) = \frac{\sigma(\mathbf{w})}{\mathbb{E}[\mathbf{w}]}$ . Nevertheless these all share the same weakness. Generally, there is no agreed-upon measure of risk, rendering the whole concept a bit dubious. Moreover, the strategy only works with the opportunities  $\boldsymbol{\omega}$  currently at hand, and thus ignores any knowledge about the actual market distribution  $P_\Omega$ .

**Sharpe Ratio** Apart from the choice of the risk measure, the inherent degree of freedom in MPT is how to select a particular portfolio from the efficient frontier (based on the choice of  $\gamma$ ). Perhaps the most popular way to avoid the dilemma is to select a spot in the pareto-front with the highest expected profits w.r.t. the risk. For the risk measure of  $\sigma(\mathbf{w})$ , this is known as the “Sharpe ratio” [64], generally defined as

$$\frac{\mathbb{E}_t[\mathbf{w}] - r_f}{\sigma(\mathbf{w})} \quad (20)$$

where  $\mathbb{E}[\mathbf{w}]$  is the expected return of the portfolio,  $\sigma(\mathbf{w})$  is the standard deviation of the return, and  $r_f$  is a “risk-free rate”. We do not consider any risk free investment in our setting, and so we can reformulate the optimization problem as

$$\begin{aligned} & \underset{\mathbf{f} \in \mathbb{R}^n}{\text{maximize}} \quad \frac{\mathbb{E}_t[\boldsymbol{\rho} \cdot \mathbf{f}]}{\sqrt{\mathbf{f}^T \Sigma \mathbf{f}}} \\ & \text{subject to} \quad \sum_{i=1}^n f_i = 1 \end{aligned} \quad (21)$$

### 2.7.3 Kelly Criterion

The Kelly criterion [32, 71] assumes the investment problem in time<sup>18</sup>, i.e. it optimizes multi-period investments in contrast to MPT which is concerned only with single-period portfolio returns. It is based on the idea of expected multiplicative growth  $W_G$  of a continuously reinvested bankroll  $W_\tau$ . The goal is to find a portfolio  $\mathbf{f}$  such that the long-term expected value of the resulting profit  $W_{\tau \rightarrow \infty}$  is maximal, which is equivalent to maximizing the geometric growth rate of wealth defined as

$$W_G = \lim_{t \rightarrow \infty} \log \left( \frac{W_t}{W_o} \right)^{\frac{1}{t}} \quad (22)$$

<sup>18</sup>Note this is in contrast to MPT which assumes the problem in an ensemble of traders at the same time, i.e. through expectation.

For its multiplicative nature, it is also known as the geometric mean policy, emphasizing the contrast to the arithmetic mean approaches (e.g. MPT) based directly on the expected value of wealth. The two can, however, be looked at similarly with the use of a logarithmic utility function, transforming the geometric into the arithmetic mean, and the expected geometric growth rate into the expected value of wealth, respectively. The problem can then be again expressed by the standard means of maximizing the (estimated) expected utility value as

$$\begin{aligned} & \underset{\mathbf{f} \in \mathbb{R}^n}{\text{maximize}} && \mathbb{E}_t[\log(1 + \mathbf{f}^T \cdot \boldsymbol{\rho})] \\ & \text{subject to} && \sum_{i=1}^n f_i = 1 \end{aligned} \tag{23}$$

Note that, in contrast to MPT, there is no explicit term for risk here, as the notion of risk is inherently encompassed in the growth-based view of the wealth progression, i.e. the long-term value of a portfolio that is too risky will be smaller than that of a portfolio with the right risk balance (and similarly for portfolios that are too conservative). The risk is thus captured by the logarithmic (concave) utility transformation itself.

The calculated portfolio is then provably optimal, i.e. it accumulates more wealth than any other portfolio chosen by any other strategy in the limit of  $\tau \rightarrow \infty$ . However, this strong result only holds given, considerably unrealistic, assumptions [32, 71, 57]. Similarly to MPT, we assume to know the true returns while calculating merely with estimates and additionally, given the underlying growth perspective, that we are repeatedly presented with the same opportunities from  $P_\Omega$  ad infinitum, making the optimality of the growth-based risk treatment in Kelly likewise a bit dubious. Despite the fact that the given conditions are impossible to meet in practice, the Kelly strategy is very popular, particularly its various modifications to mitigate the aforementioned issues.

**Fractional Kelly** The result of the Kelly optimization problem is, for each opportunity, the ideal fraction  $\omega \mapsto f^*$  one is ought to invest to achieve the maximal long-term profits. The fraction  $f^*$  thus dictates an upper-bound on the possible profit, meaning that increasing the invested fraction further will actually decrease the long-term profit.<sup>19</sup> This is commonly known as “overbetting”. Since the true expected return  $\rho$  is unknown, however, such a situation might occur even while betting with a fraction assumed to be optimal. Intentionally decreasing the calculated fraction  $f^*$  by some ratio  $\frac{1}{d'}$  then decreases the risk of overbetting stemming from a possibly overvalued estimate. Such an approach is commonly referred to as “fractional Kelly” [45]. Ideally, one should estimate the optimal shrinkage  $d'$  as another hyperparameter [5, 74] based on backtesting performance, however, it is very common to simply choose a fixed ratio such as  $\frac{1}{2}$  of the estimated optimal Kelly fraction  $f^*$ , commonly referred to as “half Kelly” by practitioners. While there are other remedies to mitigate the risk with the Kelly criterion [12, 70], fractional Kelly is a very effective method which is widely adopted in practice due to its simplicity. In addition to mitigating the overbetting risk, it generally decreases volatility, which also tends to be considerably high with the plain Kelly criterion.

**Correspondence to MPT** While Kelly is clearly based on different principles than MPT, there is an interesting close connection between the two strategies. Following [12], let us make an assumption for a Taylor series approximation that our net profits are not too far from zero  $\boldsymbol{\rho}^T \cdot \mathbf{f} \approx 0$ , allowing us to proceed with the Taylor expansion of the optimized growth as

$$\log(1 + \boldsymbol{\rho}^T \cdot \mathbf{f}) = \boldsymbol{\rho}^T \cdot \mathbf{f} - \frac{(\boldsymbol{\rho}^T \cdot \mathbf{f})^2}{2} + \dots \tag{24}$$

Now taking only the first two terms from the series we transform the expectation of logarithm into a new problem objective as follows

$$\underset{\mathbf{f} \in \mathbb{R}^n}{\text{maximize}} \quad \mathbb{E} \left[ \boldsymbol{\rho}^T \cdot \mathbf{f} - \frac{(\boldsymbol{\rho}^T \cdot \mathbf{f})^2}{2} \right] \tag{25}$$

---

<sup>19</sup>this is due to the assumed multiplicative, growth-based view of Kelly, which is in contrast to the additive MPT, where overbetting would merely increase the risk.

Note that, interestingly, the problem can now be rewritten to

$$\begin{aligned} & \underset{\mathbf{f} \in \mathbb{R}^n}{\text{maximize}} && \mathbb{E}[\boldsymbol{\rho}^T \cdot \mathbf{f}] - \frac{1}{2} \mathbb{E}[\mathbf{f}^T (\boldsymbol{\rho} \cdot \boldsymbol{\rho}^T) \mathbf{f}] \\ & \text{subject to} && \sum_{i=1}^n f_i = 1 \end{aligned} \tag{26}$$

corresponding to the original MPT formulation from Equation 19 for the particular user choice of  $\gamma = \frac{1}{2}$ . It follows from the fact that the geometric mean is approximately the arithmetic mean minus  $\frac{1}{2}$  of variance [47], providing further insight into the connection of the two popular strategies of Kelly and Markowitz, respectively. While the solution is merely an approximation, it also tends to be more robust to estimation errors than the original Kelly, similarly to the fractional approach.

### 3 Problem Insights

We have introduced the main parts of a common workflow of a trader  $t$ , who relies on a statistical estimator  $\mathbf{t}$  to predict true value  $R$  of market opportunities  $\Omega$  based on some available relevant data  $\mathcal{D}^t$ , with the resulting estimates of the expected returns  $\mathbb{E}_{\mathbf{t}}[\rho]$  being fed into some subsequent portfolio optimization strategy  $\mathbf{s}$  to produce final wealth allocations  $\mathbf{f}$ . In this Section, we provide some key insights into the problem of profitability from the perspective of the predictive model  $\mathbf{t}$ .

Let us briefly recall the problem setup. We generally consider the problem of profiting from the trader's  $t$  perspective as a stochastic game against the market maker  $m$ . The market maker  $m$  uses an estimator  $\mathbf{m}$  to continuously price the incoming opportunities  $\Omega$ . Following some investment strategy, the trader  $t$  then takes particular  $\alpha$  and  $\beta$  actions (allocations) upon these opportunities  $\Omega$ , based on her own estimates produced by  $\mathbf{t}$ . Given some distribution of market opportunities  $P_{\Omega}$ , we can then set up the game in terms of three random variables  $R, M, T$  corresponding to the true value, market maker's, and market taker's estimates, respectively. The goal of the trader (as well as the market maker) is then to maximize her expected (long-term) profits  $W$  as measured by some utility  $u$  underlying the chosen strategy  $\mathbf{s}$ .

#### 3.1 From Accuracy to Profit

The key issue with the optimal investment strategies based on portfolio optimization (Section 2.7) is that they are inherently relying on accurate estimates of the asset returns. Their performance is then directly stemming from the quality of these estimates - the better the estimates, the higher the utility of the portfolio can be achieved in general. While this holds to an extent for most of the common portfolio optimization strategies, it is best demonstrated on the optimal investment approach of Kelly.<sup>20</sup>

For simplicity of demonstration, let us consider an idealized case of an  $n$ -way betting market with no margin (Section 2.3) on the market maker's odds. Recall that the Kelly strategy is to find wealth fractions  $\mathbf{f}$  so as to

$$\underset{\mathbf{f}}{\text{maximize}} \quad \mathbb{E}_R[\log(\mathbf{f}^T \cdot (1 + \boldsymbol{\rho}))] = \sum_{i=1}^n r_i \log(f_i \cdot \frac{1}{m_i}), \quad \text{s.t.} \quad \sum_{i=1}^n f_i = 1, f_i \geq 0 \tag{27}$$

Note that in this idealized case, we calculate the expectation of returns w.r.t. the true distribution  $P_{\Omega}(R)$ . It can then be proved [15] that the solution to this constrained optimization problem yields

$$\mathbf{f}^* = \mathbf{r} \tag{28}$$

i.e. the optimal fraction of wealth  $f_i$  to invest in each outcome (opportunity)  $\omega_i$  is directly equal to the underlying true value (probability)  $r_i$ . Interestingly, we can see that in this case, the optimal strategy for the investor is to completely ignore the market pricing  $m_i$  and focus solely on having the true values predicted correctly, in which case she is guaranteed the maximal possible long term profits. This is commonly known amongst Kelly practitioners as "betting your beliefs". Note that this strong result was derived from  $\mathbb{E}_R$  and thus it only holds if the true distribution  $P_{\Omega}(R)$  is known or, more precisely, if the error in its estimate via  $T = \hat{R}$  as measured through the Kullback-Leibler divergence (Section 2.5) is zero [15]:

<sup>20</sup>The correspondence between Kelly and MPT is shown in Section 2.7.3.

$$D_{KL}(R||T) = - \sum_{i=1}^n r_i \cdot \log \frac{t_i}{r_i} = 0 \quad (29)$$

Naturally, it is close to impossible to estimate the true distribution  $P_{\Omega}(R)$  perfectly in practice.<sup>21</sup> Let us thus extrapolate into more practical settings by relaxing the condition into a non-zero  $D_{KL}(R||T)$ . Given that the optimal fractions  $\mathbf{f}$  should be equal to the true outcome probabilities  $\mathbf{r}$ , let us substitute back into the long term growth rate of wealth which Kelly seeks to maximize as

$$W_G = \sum_{i=1}^n r_i \log(t_i \cdot \frac{1}{m_i}) \quad (30)$$

Now, following the proof from [15], this can be rewritten into

$$W_G = \sum_{i=1}^n r_i \log \frac{t_i}{r_i} + \sum_{i=1}^n r_i \log \frac{r_i}{m_i} \quad (31)$$

and consequently, using the formula for KL-divergence (Equation 7), back into

$$W_G = D_{KL}(R||M) - D_{KL}(R||T) \quad (32)$$

showing the important insight that, for Kelly, the growth of wealth of the trader is directly equal to the difference in quality of her estimates  $T$  over the market prices  $M$  in terms of KL-divergence from the true values  $R$ . Consequently, positive returns can only be achieved iff the model of the trader achieves a lower cross-entropy error than the market  $XENT_{\Omega}(R, M) < XENT_{\Omega}(R, B)$  (Section 2.5). Given the information theoretic interpretation of the relative entropy [35], this is sometimes referred to as the aforementioned “information advantage” of the trader over the market maker.

**Implications** Note that this result was derived with the assumption of seeking growth-optimal investments, and its extrapolation beyond that setting may lead to wrong conclusions. Particularly, it is true that one does need a better<sup>22</sup> model to make positive profits if committed to invest optimally with Kelly, however, this does not imply that one needs a better model to make positive profits if one does not require the growth optimality.

The constraint for better model accuracy in classic portfolio optimization techniques is then inherently connected to the notion of risk (Section 2.6), which is embedded together with expected returns into the same quantity being optimized. For instance, in the Markowitz’s model, it is easy to show that even negative return portfolio may be preferred to positive returns should the latter be associated with higher variance. From the Kelly’s perspective, overvalued positive return estimates may actually lead to negative growth due to overbetting (Table 2), and it is also commonly necessary to allocate certain amount of wealth onto opportunities (outcomes) with negative returns to achieve optimal growth [73].

Consequently, wrong assessment of the true prices and probabilities associated with either of such opportunities can lead to inappropriate (over-)investments, resulting into a negative overall profit, even in situations where positive returns could be generally achieved otherwise.

### 3.2 The Essence of Profit

While the performance of common portfolio optimization strategies is tightly bound to the accuracy of price predictions (Section 3.1), we argue that accuracy is not essential for profitability in general. This is best demonstrated by taking the, sophisticated but questionable, notions of risk out of the optimization scope, resorting back to simple strategies such as the uniform investments (Section 2.7.1). Consequently, one can simply base profitability directly on the ability to correctly detect opportunities with positive expected returns (Section 2.6). Note now that whether the expectation from an opportunity  $\omega_i$  is deemed positive depends purely on the comparison between  $t_i$  and  $m_i$ . This boils down to the renown “buy low, sell high” policy to trade mispriced<sup>23</sup> assets simply as:

<sup>21</sup>We exclude scenarios with known artificial distributions such as in casino games.

<sup>22</sup>We note we do not distinguish between  $XENT$  and other measures of model accuracy here for simplicity.

<sup>23</sup>Note we again exclude spread (Section 2.3) from calculations in this Section for simplicity.

$$m_i \neq t_i \begin{cases} m_i < t_i \implies \alpha = \text{buy } \textit{assumed} \text{ underpriced asset (bet home)} \\ m_i > t_i \implies \beta = \text{sell } \textit{assumed} \text{ overpriced asset (bet away)} \end{cases} \quad (33)$$

Naturally, to asses the actual return from a trade, the true asset value  $r_i$  needs to be accounted for. In a stock market setting, the actual return from a supposedly profitable opportunity can then be defined as

$$\mathbb{E}_R[\rho_i] = \begin{cases} \frac{r_i - m_i}{m_i}, & \text{if } m_i < t_i \quad (\text{buying asset}) \\ \frac{m_i - r_i}{m_i}, & \text{if } m_i > t_i \quad (\text{selling asset}) \end{cases} \quad (34)$$

and similarly in a prediction market as

$$\mathbb{E}_R[\rho_i] = \begin{cases} \frac{r_i}{m_i} - 1, & \text{if } m_i < t_i \quad (\text{betting home}) \\ \frac{1 - r_i}{1 - m_i} - 1, & \text{if } m_i > t_i \quad (\text{betting away}) \end{cases} \quad (35)$$

Note that the true expected return  $\mathbb{E}_R[\rho_i]$  can clearly be negative and that its absolute value is not dependent of the model's estimate  $t_i$ . However, the ability to correctly recognize the profitable opportunities through the  $m_i \leq t_i$  comparison is naturally dependent on the ordering of the  $t_i$  estimates w.r.t.  $m_i$  and  $r_i$ . Note nevertheless that this comparison-based quality is very different from the accuracy-based reasoning. Consequently, even if  $\text{err}(\mathbf{t}) > \text{err}(\mathbf{m})$ , a consistent profit can still be made, as we demonstrate through the following simple examples.

**Example 3.1.** Assume that a fundamental value of an asset is \$10, with the marker maker estimating the price at \$9, and the market taker's estimate being  $t_i = \$15$ . Clearly, the market taker's estimate is more erroneous here (e.g.  $MSE(t_i) > MSE(m_i)$ ). Since the true value is not observable, she compares her \$15 evaluation of the asset to the currently offered \$9, which she *correctly* evaluates as a profitable opportunity to *buy* the asset (since  $9 < 15$ ). By buying one unit, her actual expected return from this trade will be positive at \$1 (i.e.  $10 - 9$ ).

**Example 3.2.** Similarly, assume true probability of an outcome to be 0.6, with the bookmaker  $m$  estimating it at  $m_i = 0.5$ , with the corresponding fair odds set up to 2.0, and the bettor  $\mathbf{t}$  estimate being at  $t_i = 0.9$ . Clearly, the bettor's estimate is more erroneous here (e.g.  $XENT(t_i) > XENT(m_i)$ ). Nevertheless she has no choice but to use her estimate to asses the return on investment, which she *correctly* estimates as being *positive* ( $\mathbb{E}_{\mathbf{t}}[\rho] = \frac{0.9}{0.5} - 1 > 0$ ). Despite being very wrong numerically with her expectation of a 80% ROI, by betting a unit of wealth, she can still expect to obtain the actual positive ROI of 20%.

Note that the trader's  $t$  estimates  $t_i$  in these examples could have been set arbitrarily larger (within the respective domain), making the corresponding model  $\mathbf{t}$  arbitrarily bad by the standard error measures.

**Definition 3.1.** Followingly, let us define a more relaxed, necessary condition of *essential profitability* of a model  $\mathbf{t}$  simply as the consequent existence of one of the following market opportunities  $\omega_i$  in  $P_{\Omega}(R, M, T)$ :

1. the market undervalues the true price, and the model estimates a higher value than the market,  
i.e.  $m_i < r_i \wedge t_i > m_i$
2. the market overvalues the true price, and the model estimates a lower value than the market,  
i.e.  $m_i > r_i \wedge t_i < m_i$

Using sufficiently conservative (small) wealth allocations, investments into either of these cases will lead to systematic profits of the market taker in the long run. On the contrary, no investment strategy can lead to positive profits without such opportunities in the portfolio.

Nevertheless the overall profitability of a model  $\mathbf{t}$  will naturally depend on the relative occurrence of such  $\omega_i$ 's in the actual market distribution  $P_{\Omega}$  (Definition 2.1). Let us now generalize the essence of profitability, from reasoning about the necessary relationships between individual  $r_i, m_i, t_i$  estimates, to the properties of the whole market distribution  $P_{\Omega}(R, M, T)$ . Following on the aforementioned "buy low, sell high" strategy with uniform investments, the expected profitability from a market distribution  $P_{\Omega}$  is clearly

$$\mathbb{E}_{P_{\Omega}}[\rho] = \sum_i \mathbb{E}_R[\rho_i] \cdot P_{\Omega}(r_i, m_i, t_i) \quad (36)$$

Table 2: All possible orderings of the true value ( $r_i$ ), market maker’s ( $m_i$ ), and trader’s estimates ( $t_i$ ), with the implied trading decisions and resulting profitability. Additionally, the relative size of the implied Kelly fraction is indicated.

values ordering	decision ( $t_i \lessgtr m_i$ )	profit	proportion (in $P_{\mathcal{U}}$ )	Kelly
$r_i < t_i < m_i$	sell	$m_i > r_i$	1/6	overbet
$r_i < m_i < t_i$	buy	$r_i < m_i$	1/6	overbet
$t_i < r_i < m_i$	sell	$m_i > r_i$	1/6	underbet
$t_i < m_i < r_i$	sell	$m_i < r_i$	1/6	underbet
$m_i < t_i < r_i$	buy	$r_i > m_i$	1/6	underbet
$m_i < r_i < t_i$	buy	$r_i > m_i$	1/6	overbet

We already know that the fundamental value  $r_i$  is a principally unknown random variable and one can thus never perfectly assess the true return  $\rho_i$  from any trade in advance, for which we resort to an estimate  $\hat{\rho}_i$ . However, the market distribution  $P_{\Omega}(R, M, T)$  here is also principally unknown, for which one again needs to rely on statistics while estimating it from historical data as  $\hat{P}$ . Consequently, one can estimate the essential profitability of a model  $t$  w.r.t. market pricing  $m$  as

$$\mathbb{E}_{\hat{P}}[\hat{\rho}] = \sum_i \mathbb{E}_T[\rho_i] \cdot \hat{P}(t_i, m_i, t_i) \quad (37)$$

As with any investment strategy, the calculated expected returns can be very different from the actual return distribution  $\mathbb{E}_{\hat{P}}[\hat{\rho}] \neq \mathbb{E}_{P_{\Omega}}[\rho]$ , depending on the properties of the estimates (Section 2.6). However, profitability of the simple unit investment strategy leads to a much more relaxed and robust condition on the model quality, which is what we exploit to yield positive profits even with estimators of inferior predictive performance.

**Drawbacks** We note that the standard approach of focusing on predictive accuracy has many advantages, such as being naturally less noisy and more interpretable, and should be preferred when profitability is not the target [78]. We also acknowledge that by deflecting from the accuracy-based view and focusing merely on the essential profitability with the unit investments, we downplay the role of explicit optimization of growth and risk in the formal strategies of Kelly and Markowitz, respectively. Nevertheless, as discussed in the respective Sections 2.7.3 and 2.7.2, these formal notions are based on rather unrealistic (wrong) assumptions, which is why additional risk management practices, such as the fractioning (Section 2.7.3), need to be commonly employed with the strategies anyway [45, 44, 74]. Consequently, sacrificing formal optimality w.r.t. unrealistic objectives in order to transition from negative to positive profits does not seem that big of a sacrifice.

### 3.3 Market Taker’s Advantage

The market maker’s advantage (Section 2.3) is a well-worn concept. However, there is also an advantage of the market *taker* which is rarely discussed explicitly, but is essential to the traders profitability. While the market maker has the obligation to continuously *quote* price of both sides of the market (Section 2.5), the taker has the crucial liberty to *select* only those of the resulting opportunities deemed profitable. That is she is to decide whether and which side of the market to trade the assets once the market maker’s prices have been laid out. As the second player, the difficulty of her task is reduced from the correct price estimation to the estimation of the market price error direction. While this might seem as a similarly difficult problem, the latter is a considerably easier task.

For demonstration, consider the three values  $r_i, m_i, t_i$  of the true value, market maker’s, and trader’s estimates, respectively, to be laid out completely at random, yielding a uniform market distribution  $P_{\mathcal{U}}$  where  $R, M, T \sim \mathcal{U}^3$ . The possible situations that emerge from the ordering of  $r_i, m_i, t_i$  in such a setting are displayed in Table 2. Since neither of the estimators possesses any information w.r.t.  $R$ , both  $M$  and  $T$  clearly perform equally by the means of arbitrary statistical estimation measures (Section 2.5). While one might thus expect this to be a neutral trading setting for both the sides, interestingly, the market taker  $t$  would already be able to make a substantial profit with uniform investments by correctly identifying 2/3 of the profitable opportunities.

Intuitively, this demonstrates a simple fact that it is generally more likely to overestimate an undervalued



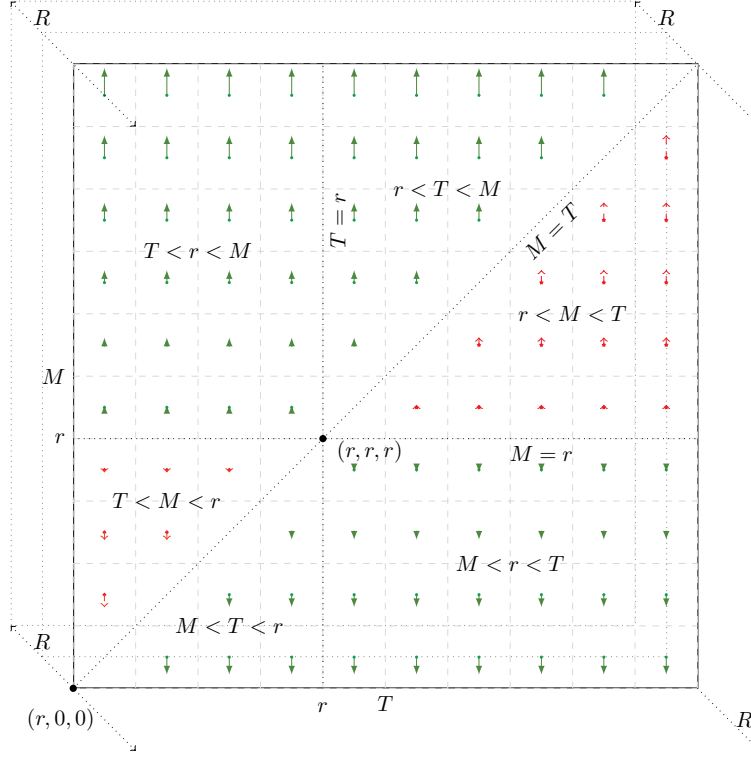


Figure 1: A 2D projection of the  $P(R, M, T)$  distribution onto  $P(M, T | R = r)$  with visualization of essential profitability (Section 3.2) of the individual point-estimates as a vector field. Green color (solid arrows) denotes positive returns and red (dashed arrows) negative returns, respectively, while the length of each vector corresponds to the ROI in an idealized two-way stock market (Section 2.6).

estimate than to further underestimate it, i.e.

$$m_i < r_i \implies P_{\mathcal{U}}(m_i < t_i) > P_{\mathcal{U}}(t_i < m_i) \quad (38)$$

and vice versa for an overvalued estimate:

$$m_i > r_i \implies P_{\mathcal{U}}(m_i < t_i) < P_{\mathcal{U}}(t_i < m_i) \quad (39)$$

Note, importantly, how this property of the completely uninformative  $P_{\mathcal{U}}$  is conveniently aligned with the essential profitability of the trader's model (Definition 3.1). The concrete proportions of the individual situations will naturally depend on the particular distribution  $P_{\Omega}$ , nevertheless the property holds very generally for unskewed distributions with unbiased estimators (Section 3.4). Note the difference from the standard model accuracy measures which would all evaluate both models equally in  $P_{\mathcal{U}}$ . Nevertheless from the perspective of profitability, the situation is very different since, as opposed to the market maker, the market taker is not penalized for estimation errors in these two situations that emerge more often than not. Consequently in  $P_{\Omega} = P_{\mathcal{U}}$ , the trader is in an inherent advantage of 2 : 1, which can be directly turned into the corresponding profits.

It is perhaps more instructive to demonstrate the concept on a particular level of true value  $r$ . Without loss of generality, let us consider all opportunities with a real value  $r$  being traded in an ideal two-way market (without spread). We can then plot a 2D projection of the  $P(R, M, T)$  market distribution by conditioning it as  $P(M, T | R = r)$ , and visualize the essential profitability (Definition 3.1) of the corresponding sub-regions of the distribution. The result is displayed in Figure 1. We can observe that the distribution of the profitable regions (green, solid arrows) is clearly in favor of the trader, and that the potential returns progressively increase with the error of the market maker, i.e. the distance of  $M = m$  from  $r$ .

Note that here we did not yet assume any particular (non-uniform) distribution of estimates, and this inherent advantage of the trader is thus completely oblivious of any information advantage as well as any other

property of the price estimators. Rather, it stems purely from the unequal roles of the market maker  $m$  and taker  $t$ . Consequently, should they, e.g., switch roles with the same estimation models ( $m \leftarrow t, t \leftarrow m$ ), the advantage would stay exactly the same on the side of the market taker  $t$ .

### 3.4 Distribution of Estimates

While demonstrating the concept of the trader's advantage, the uniform  $P_{\mathcal{U}}(R, M, T)$  distribution of estimates with completely uninformed players from Section 3.3 seems rather unlikely in practice. In real world markets, the values  $R, M, T$  are not going to be independent, but rather correlated with each other, since  $M$  and  $T$  are typically based on similar information sources and both try to model  $R$  with similar techniques (Section 2.4). Let us now review common properties of the more realistic market distributions of these estimates.

**Bias** The market makers are typically very good at being close to the real price, and we can assume their estimates  $M$  to be *unbiased* w.r.t. true value  $R$ , or, more formally:

$$\mathbb{E}_P(B) - R = 0 \quad (40)$$

which means that they are not systematically deviating when measured against the true value alone. Should a market maker be biased in this manner<sup>24</sup>, it would be extremely easy to exploit her merely by trading all opportunities on the corresponding side of the market. Moreover we can reasonably assume her to be point-wise unbiased at each particular price level  $r$ , i.e.

$$\mathbb{E}_P(B|R=r) = r \quad (41)$$

If that was not the case, the market maker would again be easily exploitable by correspondingly trading all possible assets within a certain price range  $r \pm \delta$ , i.e. by buying all assets in price ranges where the maker  $m$  systematically undervalues the assets, and vice versa for selling in overvalued price ranges. One can typically check from historical data that the market makers are not biased in this simple manner in any reasonably efficient market.<sup>25</sup> Note that the unbiasedness is only one of the conditions for a fully efficient market (Section 2.2). There is generally no reason for the market taker  $t$  to be biased in this trivial way either, unless a systematic error is present in her model  $t$ , or she reversely reflects the market maker's bias to exploit it.

**Variance** Given the assumption that the models behind  $M$  and  $T$  are both unbiased estimators of  $R$ , we can now focus merely on their (co-)variances. It is a common practice in statistical estimation of functions for one to look for an estimator with the smallest variance among the class of unbiased estimators (Section 2.5). Since we are working with function estimators, we are not interested in the total variance of the estimations  $T$ , which includes variance due to variations in the true price  $R$  itself as

$$\text{Var}[T] = \mathbb{E}_R[\text{Var}[T|R]] + \text{Var}_R\mathbb{E}[T|R] \quad (42)$$

but merely in the first term capturing the expected variance left w.r.t. predicting  $R$ . Given the assumption of the point-wise unbiased estimates, the conditional variances w.r.t.  $R$  are then equal to the covariances of the models, i.e.

$$\text{Cov}[M, R] = \text{Var}[M|R] \quad \text{and similarly} \quad \text{Cov}[T, R] = \text{Var}[T|R] \quad (43)$$

Given the unbiasedness, these co-variances then directly reflect the quality (accuracy) of the underlying models of the market maker ( $\text{Cov}[M, R]$ ) and market taker ( $\text{Cov}[T, R]$ ), respectively.

The last degree of freedom in terms of covariances in  $P$  is the relationship between  $M$  and  $T$ , i.e.  $\text{Cov}[M, T]$  ( $\text{Cov}[M, T|R]$ ). While the other two covariances have the clear introduced interpretation, the  $\text{Cov}[M, T]$  is more intriguing, but is also essential to the proposed profitability of inferior predictive models (Section 3.1). As we have seen in the case of the uniform market distribution  $P_{\mathcal{U}}$ , where both the players possess the same amount of information w.r.t. the true price, corresponding to the same accuracies of  $m$  and  $t$ , the trader  $t$  is always in advantage. However not all distributions with equally informed players are as such, as demonstrated by the following example.

<sup>24</sup>biased by more than  $\epsilon$  in practice (Section 2.3).

<sup>25</sup>as we do in experiments in Section 5.3.1

**Example 3.3.** Assume a scenario where both the trader  $t$  and market maker  $m$  possess the exact same model. Clearly, their information value, accuracy and all statistical measures will be exactly the same, just as in the case of the uniform distribution  $P_U$ . Nevertheless, the trader will not be in an advantageous position anymore. Since all her estimates coincide with the market price  $\forall i : t_i = m_i$ , it is not possible to detect any profitable opportunities where  $r \notin (b - \epsilon, b + \epsilon)$ , even if they exist in the market distribution  $P_\Omega$ . Hence, the profitability of the trader in this case is clearly zero.

From the statistical viewpoint, one can note an underlying difference between the two example distributions in the third covariance term  $\text{Cov}[M, T]$ . Whereas in the uniform distribution, the two variables were completely independent, i.e.  $\text{Cov}[M, T] = 0$ , here they are equal and thus display maximal possible covariance ( $\text{Cov}[M, T] = \sigma^2$ ). While this anecdotal reference to the connection between profitability and  $\text{Cov}[M, T]$  is rather informal, we analyse it in detail in the next Section 4.

## 4 Increasing Profit through Decorrelation

Recall that we have a two-way market with opportunities  $\Omega$  of some fundamental value  $R$ , being priced by the market maker  $m$  as  $M$  and taker  $t$  as  $T$ , resulting into some market distribution of estimates  $P_\Omega(R, M, T)$  (Definition 2.1). Let us now consider the context of the common properties (Section 3.4) of such market distributions  $P_\Omega$ , allowing assessments of model performance in terms of expected returns  $\mathbb{E}_{P_\Omega}(\rho)$  w.r.t. the distribution  $P_\Omega(R, M, T)$ . Particularly, we will explore the aforementioned statistical relationship between the market maker  $m$  and taker  $t$ . To further standardize the relationship study, i.e. to take the individual variances of  $M$  and  $T$  out of scope, we now switch from covariance  $\text{Cov}[T, M|R]$  to correlation  $\text{Corr}[T, M|R]$ .

**Definition 4.1.** We use the term *decorrelation* to refer to the concept of decreasing the partial correlation  $\text{Corr}[T, M|R]$  between a price estimator  $\mathbf{t}$  of the trader and the market maker  $\mathbf{m}$  w.r.t. the real value  $\mathbf{r}$  across opportunities  $\omega_i \in \Omega$  endowed with some market distribution  $P_\Omega$  (Definition 2.1).

The main goal of this Section 4 is to show that enforcing smaller<sup>26</sup> partial correlation  $\text{Corr}[T, M|R]$  of a price estimator  $\mathbf{t}$  generally increases its essential profitability (Definition 3.1) within common market distributions, particularly for estimators  $\mathbf{t}$  that are inferior to the market maker  $\mathbf{m}$ .

### 4.1 Unbiased Estimators

Let us first consider the common setting of unbiased price estimators, the reasoning behind which was introduced in Section 3.4.

**Theorem 4.1.** The essential profitability (Definition 3.1) of an unbiased estimator  $\mathbf{t}$  with the lowest partial correlation  $\text{Corr}[T, M|R = r] = -1$  with the market  $\mathbf{m}$  is maximal. Consequently, no deviation from such a model  $\mathbf{t}$  can thus increase the profitability further.

*Proof.* Now for  $\text{Corr}[T, M|R] = -1$  and an arbitrary  $r$ , the probability distribution  $P(T, M|R = r)$  collapses into a linear function  $(t; a, b) \mapsto m$  of the form

$$m = a \cdot t + b \quad \text{where} \quad a = \frac{\text{Cov}[M, T|r]}{\text{Var}[T|r]} \quad \text{and} \quad b = (1 - a) \cdot r \quad (44)$$

where the  $b$  is set so that the mean values  $\mathbb{E}_{P(T, M|R=r)}[M] = \mathbb{E}_{P(T, M|R=r)}[T]$  of both the marginals  $P(M|r)$  and  $P(T|r)$  are at  $r$  (unbiased). The mean of the distribution  $m = t = r$  thus lies on the line:

$$r = a \cdot r + (1 - a) \cdot r \quad (45)$$

Clearly, since  $\text{Corr}[T, M|r] < 0$ , we have also  $\text{Cov}[M, T|r] < 0$  implying a negative slope  $a < 0$  of the function line. There are consequently only 2 possible price estimate orderings (regions in Figure 1) for all  $\Omega$ , both of which are profitable, as follows:

1.  $T < r < M$  implying stock returns  $\frac{m-r}{m} > 0$ , or likewise for betting  $\frac{1-r}{1-m} - 1 > 0$

<sup>26</sup>Note that utilize the term “decorrelation” to refer to decreasing the correlation even below zero.

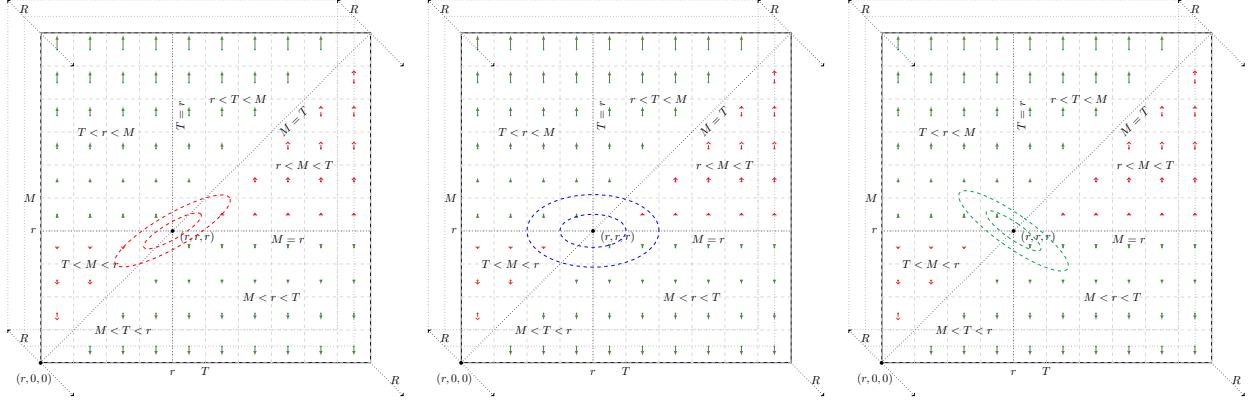


Figure 2: A sketch of the profitability regions from Figure 1 with an example distribution of estimates density for the case of unbiased estimations of  $M$  and  $T$  w.r.t  $R = r$ , where  $\text{Var}[T|R] > \text{Var}[M|R]$ , corresponding to a common situation in practice (Section 3.4, 5). The effect of  $\text{Corr}[T, M|R]$  on profitability is demonstrated on a highly correlated model (left, red), an independent model (middle, blue), and a highly negatively correlated model (right, green). By decreasing the correlation, progressively larger parts of the profitable regions of the distributions (green) are being covered, reflecting the corresponding increase in returns.

2.  $M < r < T$  implying stock returns  $\frac{r-m}{m} > 0$ , or likewise for betting  $\frac{r}{m} - 1 > 0$

Note again that the individual values of returns do not depend on the value of  $T$  but merely on the value order. From the assumed role of the trader  $t$ , the only possible changes to the distribution (and profit) can be made via changes in her model  $t$  estimates  $T$ . However, any potential deviation from this distribution (line) with  $\text{Corr}[M, T|R] = -1$  can only result into one of the following ordering (region) transitions:

1.  $T < r < M$  can change to either:

- (a)  $r < T < M$  implying *no change* in the returns  $\frac{m-r}{m} > 0$ , or likewise  $\frac{1-r}{1-m} - 1 > 0$
- (b)  $r < M < T$  implying *decrease* in the returns to  $\frac{r-m}{m} < 0$ , or likewise  $\frac{r}{m} - 1 < 0$

2.  $M < r < T$  can change to either:

- (a)  $M < T < r$  implying *no change* in the returns  $\frac{r-m}{m} > 0$ , or likewise  $\frac{r}{m} - 1 > 0$
- (b)  $T < M < r$  implying *decrease* in the returns to  $\frac{m-r}{m} < 0$ , or likewise  $\frac{1-r}{1-m} - 1 < 0$

ergo no deviation from  $\text{Corr}[M, T|R] = -1$  can increase the profitability any further.  $\square$

Note that this also means that we cannot increase the returns even via transition into a perfect estimator with both  $\text{Bias}_R[T] = \text{Var}[T|R] = 0$ , i.e. a perfect model  $P(T = r|R = r) = 1$  which always returns the correct answer in a deterministic fashion (and for which the partial correlation would be undefined). This means that for the purpose of the essential profit generation (Definition 3.1), the variance of the model no longer acts as an error since it is completely turned into our advantage. Note this is in direct contrast to a correlated investor who, given a variance higher than the market maker, is doomed to obtain completely negative returns.<sup>27</sup> A visualization of the correlation effect on the returns, with an example elliptical market distribution for the given setting, is depicted in Figure 2. An evaluation of this effect on a statistical sample from a normal distribution of estimates can also be found in Figure 9 in the Appendix.<sup>28</sup>

<sup>27</sup>We also note that the transition between the two corner cases is somewhat smooth w.r.t. the returns for common market distributions, such as the elliptical distributions used for visualization. However, we acknowledge that we do not present a formal proof of monotonicity of this property.

<sup>28</sup>We also note that the actual distributions of model estimates resemble these forms quite closely, see e.g. Figure 4.10 in [28].

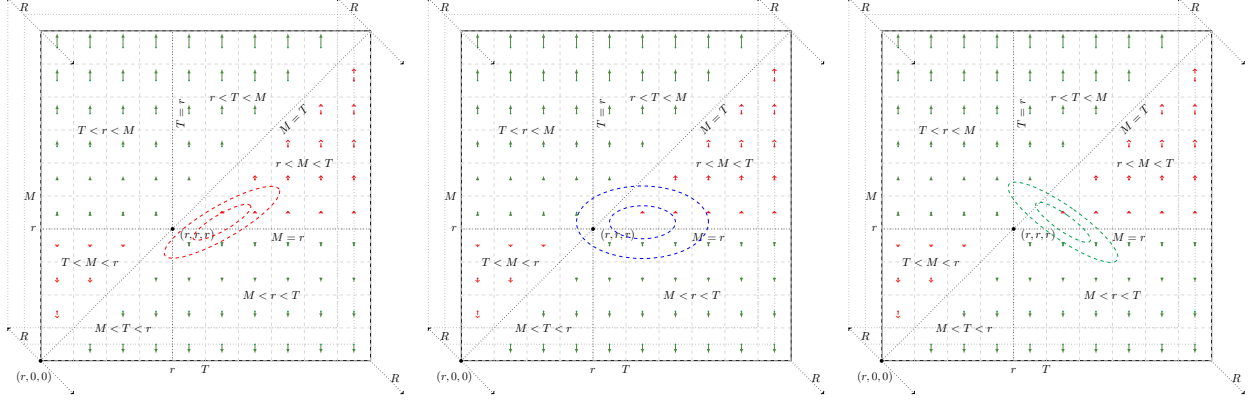


Figure 3: A sketch of the distribution of estimates density from Figure 1 for the case where the trader  $T$  is biased more than the market maker  $M$  w.r.t  $R = r$ . The decorrelation technique still commonly helps as larger parts of the profitable regions of the distribution are being covered by decreasing  $\text{Corr}[T, M|R]$ .

## 4.2 Biased Estimators

While common, the assumption of point-wise unbiased estimators might be seen as too strict in practice. The market maker is very unlikely to be overly biased, due to its constant exposure to the traders, who would likely exploit such easy opportunities (Section 3.4). Nevertheless the market takers are generally free to come up with all sorts of models. Let us briefly review the situation where the trader's model  $\mathbf{t}$ , which we seek to optimize, is more biased than  $\mathbf{m}$  w.r.t.  $R$ , i.e.  $\text{Bias}_R[T] > \text{Bias}_R[M]$ .

In this setting, one can craft counterexamples showing that  $T$  with  $\text{Corr}[T, M|R] = -1$  is not universally the most profitable model  $\mathbf{t}$  anymore.<sup>29</sup> While not the universally best possible model, a decorrelated  $T$  will still perform very well in practice here. Particularly, it will be consistently better than a highly correlated model, even if the latter has a lower variance  $\text{Var}[T|R]$ . Interestingly, it is also better than a model with zero variance (vertical line), i.e. given some bias, we are able to turn the variance into an advantage by decreasing correlation with  $M$ . Lastly, the minimal correlation will be also typically better than no correlation for common elliptical or uniform conditional distributions. A visualization of this setting, where  $\text{Bias}_R[T] > \text{Bias}_R[M]$ , is displayed in Figure 3 for an example elliptical distribution. Note that the same reasoning is also applicable to cases where the  $\text{Bias}_R[T] = \text{Bias}_R[M]$ .

## 4.3 Having a Superior Model

The primary motivation behind the concept of lowering the correlation with the market is to make profits with models of *inferior* quality, which would commonly yield negative profits otherwise. We argued that such a situation is common, since the market (maker) price tends to be a very good estimate of the true value in any fairly efficient market (Section 3.4). Nevertheless, for completeness, let us consider the opposite case of having a superior model, i.e. a situation where  $\text{err}(\mathbf{t}) < \text{err}(\mathbf{m})$ . Following the statistical decomposition of estimation errors in terms of bias and variance (Section 2.5), let us separately consider two cases of such superiority through a model with (i) lower variance and (ii) lower bias.

**Superior Variance** For a superior model  $T$  with a lower conditional variance  $\text{Var}[T|R] < \text{Var}[M|R]$ , the concept of decorrelation (Definition 4.1) no longer works as a consistent profit enhancement, even for common, realistic market distributions. Particularly, decreasing the correlation  $\text{Corr}[T, M|R]$  can actually decrease the

<sup>29</sup>For instance, consider a distribution where the model  $\mathbf{t}$  is biased w.r.t.  $R$  by some  $\delta$  as

$$T = \begin{cases} M - 1.1 \cdot \delta, & \text{if } T > r \\ -M + 1.1 \cdot \delta, & \text{if } T < r \end{cases} \quad (46)$$

This  $\delta$ -biased model  $\mathbf{t}$  is set to make maximal possible profit, and decreasing its correlation with  $\mathbf{m}$  will actually hurt its performance. However, achieving such a distribution of estimates  $P_\delta(R, M, T)$  is close to impossible in practice, as it is carefully crafted w.r.t. the unknown value of  $r$ . Consequently, this scenario is highly unstable w.r.t.  $r$  as well as variance of  $T$ , a decreasing of which will paradoxically lead to complete loss of all returns.

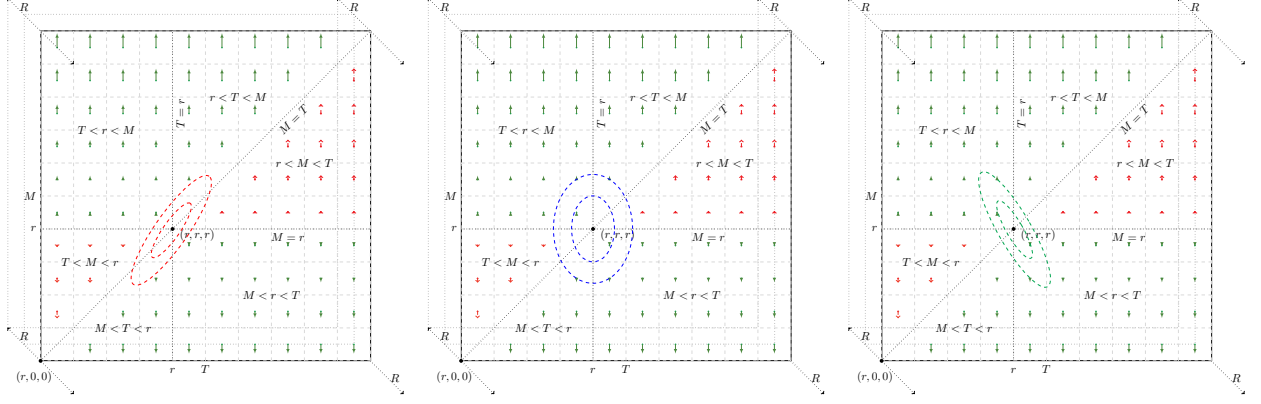


Figure 4: A sketch of the distribution of estimates density from Figure 1 for the case of unbiased estimators  $M$  and  $T$  w.r.t  $R = r$ , where  $\text{Var}[T|R] < \text{Var}[M|R]$ , corresponding to a superior estimation model  $\mathbf{t}$  of the trader. For a superior model in terms of the conditional variance, the correlation with the market does not pose such a problem as the model is already profitable, and decreasing it might actually hurt the performance, nevertheless, a model with the lowest covariance still generally provides the highest profitability.

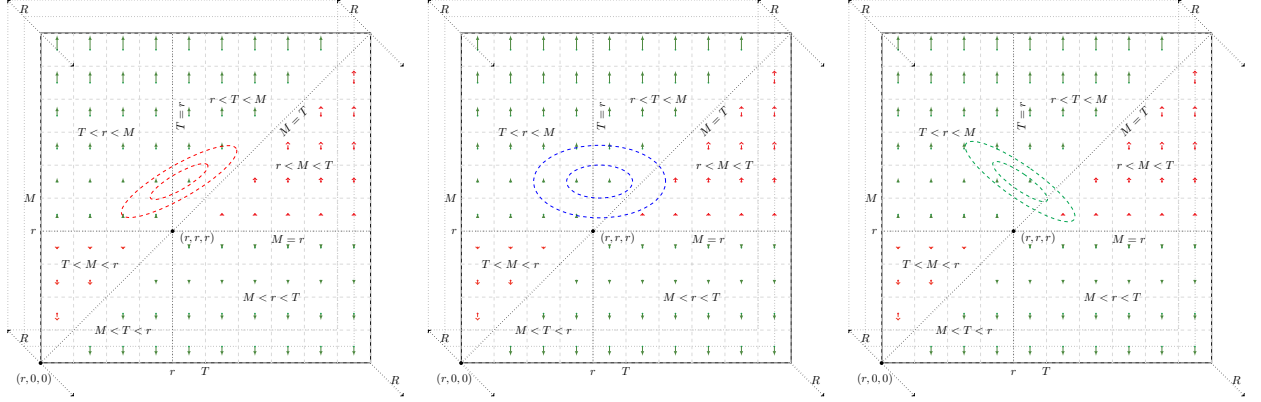


Figure 5: A sketch of the distribution of estimates density from Figure 1 for the case where the market maker  $M$  is biased more than the trader  $T$  w.r.t  $R = r$ . While having such a superior model, being correlated with the market poses no problem to profit, and decreasing the correlation will commonly hurt the performance.

returns in many scenarios. We again depict the concept on an example elliptical distribution in Figure 4. The variances of the estimators are exactly opposite to those from Figure 2. We can see that the returns with an independent model  $\text{Corr}[T, M] = 0$  are lower than those of a highly correlated model  $\text{Corr}[T, M|R] = 1$ . Note however that the decrease of returns in this case is smaller than the increase in the opposite case (i.e. shift from  $\text{Corr}[T, M|R] = 1$  to  $\text{Corr}[T, T|R] = 0$  in Figure 2). Finally, the profits of  $\text{Corr}[T, M|R] = -1$  are still maximal.

**Superior Bias** We have argued for the practical necessity of a market maker not to be systematically biased in Section 3.4, which leaves a little space for the trader to beat the market in terms of  $\text{Bias}_R[T] < \text{Bias}_R[M]$ . Nevertheless it should be acknowledged that in the unlikely case that the market maker indeed is more biased than the trader, the concept of decorrelating the estimates for increased profits breaks down severely. The situation is depicted in Figure 5. We can see that in this setting, decreasing  $\text{Corr}[T, M|R]$  can easily work in a directly counterproductive fashion by consistently decreasing the profits.

Importantly, however, it should be noted that with a model that is superior by either means, i.e. where  $\text{err}(\mathbf{t}) < \text{err}(\mathbf{m})$ , there is no need in trying to decrease  $\text{Corr}[T, M|R]$  to make profits, since such a model  $\mathbf{t}$  needs no help with that to begin with. This can be conveniently detected in advance by measuring  $\text{err}(\mathbf{t})$  and trading the model with standard investment strategies (Section 2.7) instead.

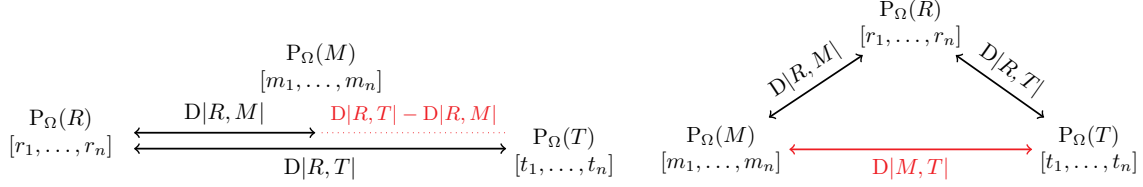


Figure 6: The difference between a typical statistical treatment of quality of the estimators, where the relationship between  $M$  and  $T$  is considered merely in the terms of their distances to  $R$  (left), and the proposed scenario, where we explicitly consider their mutual statistical relationship (right).

#### 4.4 The Problem with Kelly

The scenarios we have demonstrated so far operated with the simple uniform investment strategy (Section 2.7), which allowed us to generate profits through decorrelation (Definition 4.1), even with models of inferior accuracy w.r.t. the market. As indicated in Section 3.1, let us now explain why this cannot be done with the plain Kelly investment strategy (Section 2.7.3).

We have shown that the growth of wealth  $W_G$  with the Kelly strategy is directly equal to difference between the KL-divergence of the market maker from the true distribution  $D_{KL}(M||R)$  and the trader from the true distribution  $D_{KL}(T||R)$ , respectively, in Equation 32. It follows that a plain Kelly trader does not care about the particular relationships between  $R, M, T$  which are essential to profitability (Section 3.2), since the only thing that matters is how close are our estimates to the true distribution as compared to the bookmaker. The two principally different views of the market distribution  $P_\Omega$  properties are depicted in Figure 6.

Consequently, it thus does not matter whether an individual opportunity seems to have a positive or negative expected profit, since an optimal Kelly trader will bet an amount derived merely from the expected growth corresponding to the information advantage w.r.t.  $R$ . Given the knowledge of the true value distribution  $P_\Omega(R)$ , this provably leads to more wealth than any other strategy. It is then tempting to believe that the Kelly solution then provides an upper-bound to the amount of profit that can be made in any scenario. Nevertheless, this is not true in the real world setting where we lack the knowledge of the true probabilities  $P_\Omega(R)$ , as we demonstrate on the following example.

**Example 4.1.** Consider a simple scenario of Kelly betting on a match (asset)  $a_i$  with two equally probable exclusive outcomes  $\{home, away\}$ , and the two corresponding opportunities  $\omega_i^\alpha, \omega_i^\beta$  priced by the bookmaker  $m$  and the trader  $t$  equally as follows

$$r(\omega_i) = \begin{cases} 0.5 & \text{on } \alpha=home \\ 0.5 & \text{on } \beta=away \end{cases} \quad m(\omega_i) = \begin{cases} 0.3 & \text{on } \alpha=home \\ 0.7 & \text{on } \beta=away \end{cases} \quad t(\omega_i) = \begin{cases} 0.3 & \text{on } \alpha=home \\ 0.7 & \text{on } \beta=away \end{cases} \quad (47)$$

In this example, we again omit the bookmaker's margin for simplicity<sup>30</sup> and so, following the derivation from Section 3.1, the optimal vector of fractions  $f$  to bet on the outcomes would be

$$f = \begin{cases} 0.3 & \text{on } \alpha=home \\ 0.7 & \text{on } \beta=away \end{cases} \quad (48)$$

Since the two estimates of  $m$  and  $t$  coincide, there is clearly no information advantage of the trader and consequently zero profit to be made with Kelly. And the situation is the same for any other strategy, too, since the expected profitability of the opportunities, following the expected profit definition from Equation 13, from the perspective of the trader is simply zero

$$\mathbb{E}_t[\rho_i^\alpha] = \frac{0.3}{0.3} - 1 = 0 \quad \mathbb{E}_t[\rho_i^\beta] = \frac{0.7}{0.7} - 1 = 0 \quad (49)$$

<sup>30</sup>Note again that introducing bookmaker's margin in these examples does not change the introduced (decorrelation) principles - it would only decrease the total growth values (i.e. the true growth of Kelly would be negative, but the fractional Kelly could still remain positive), as generally outlined at the end of Section 2.3.

Now consider altering the scenario by decorrelating the estimates of the trader  $t$  as follows

$$r(\omega_i) = \begin{cases} 0.5 & \text{on } \alpha=\text{home} \\ 0.5 & \text{on } \beta=\text{away} \end{cases} \quad m(\omega_i) = \begin{cases} 0.3 & \text{on } \alpha=\text{home} \\ 0.7 & \text{on } \beta=\text{away} \end{cases} \quad t(\omega_i) = \begin{cases} 0.7 & \text{on } \alpha=\text{home} \\ 0.3 & \text{on } \beta=\text{away} \end{cases} \quad (50)$$

Despite switching the estimates, we can clearly see that both the bookmaker  $m$  and the trader  $t$  are still equally distanced from the true value  $r$  (by the means of  $D_{KL}$  as well as any other possible metric), leading again to no information advantage and, as expected (Section 3.1), to zero actual growth of wealth (Section 2.7.3):

$$W_G = \frac{1}{t} \cdot \log\left(\frac{W_t}{W_0}\right) = \sum_i r_i \cdot \log\left(\frac{f_i}{m_i}\right) = 0.5 \cdot \log\left(\frac{0.7}{0.3}\right) + 0.5 \cdot \log\left(\frac{0.3}{0.7}\right) = 0 \quad (51)$$

Nevertheless, the essential profitability of the opportunities from the perspective of the trader is now

$$\mathbb{E}_t[\rho_i^\alpha] = \frac{0.7}{0.3} - 1 = 1.33 \quad \mathbb{E}_t[\rho_i^\beta] = \frac{1 - \mathbb{E}[m(o_2)]}{1 - b(o_2)} - 1 = \frac{0.3}{0.7} - 1 = -0.57 \quad (52)$$

and betting uniformly (Section 2.7.1) some unit on the first, correctly recognized (Section 3.2), opportunity  $\alpha = \text{home}$ , the trader would make a consistent unit profit of

$$\mathbb{E}_r[\rho_i^\alpha] = \frac{r_i}{m(\omega_i^\alpha)} - 1 = \frac{0.5}{0.3} - 1 = 0.66 \quad (53)$$

despite it being different from her estimated  $\rho_i^\alpha = 1.33$ .

The inability of Kelly to make profit in such profitable scenarios follows directly from its growth-based view of optimal investments (Section 2.7.3).<sup>31</sup> While maximizing the growth  $W_G$ , less than optimal investments are just as harmful as over-investment, despite the latter leading to definite ruin while the former only leads to sub-optimal growth. To distinguish between the two arguably different types of risk, various modifications of the Kelly criterion have been proposed to reflect the natural preference for avoiding the ruin at the cost of sub-optimal growth. By giving up on the expected (theoretical) growth optimality (i.e., w.r.t.  $\mathbb{E}_t$ ), we might then actually achieve better *true* growth (i.e., w.r.t.  $\mathbb{E}_r$ ) in many real world settings, as we demonstrate below.

#### 4.4.1 Fractional Kelly

Perhaps the most common remedy to mitigate the risk stemming from the erroneous estimates is fractional Kelly (Section 2.7.3). Let us demonstrate the effect of this risk management practice on the introduced setting from Example 4.1 as follows.

**Example 4.2.** Being aware of the uncertainty in her estimates, the Kelly trader now decreases the optimal fraction by one-half, popularly referred to as “half-Kelly” betting. Considering the first scenario of coincidental estimates from Equation 47, the invested fractions are now thus decreased by half as

$$f = \begin{cases} 0.15 & \text{on } \alpha=\text{home} \\ 0.35 & \text{on } \beta=\text{away} \end{cases} \quad (54)$$

However, the growth of wealth, accounting for the half of it being held separately, stays inert at zero since

$$W_G = \frac{1}{t} \cdot \log\left(\frac{W_t}{W_0}\right) = \sum_i r_i \cdot \log\left(0.5 + \frac{f_i}{m_i}\right) = 0.5 \cdot \log\left(0.5 + \frac{0.15}{0.3}\right) + 0.5 \cdot \log\left(0.5 + \frac{0.35}{0.7}\right) = 0 \quad (55)$$

In words, the trader is still under-betting the the first opportunity  $\alpha = \text{home}$ , which is profitable, while putting a larger amount on the second opportunity  $\beta = \text{away}$ , which is loss-making. Decreasing the fractions then

<sup>31</sup>Note that the situation in this example remains the same even when allowing Kelly to keep part of the bankroll aside in the cash option (Section 2.7). This would introduce an infinite subspace of other optimal fractional solutions, all of which however, by definition (Section 2.7.3), lead to the exact same growth values (i.e. zero true growth for any optimal Kelly solution here).



cannot change the simple fact that the opportunity is not recognized correctly by the estimator  $\mathbf{t}$  w.r.t. essential profitability (Definition 3.1).

Consider now the latter scenario of the decorrelated estimates laid out in Equation 50. The invested fractions are again cut by half as

$$f = \begin{cases} 0.35 & \text{on } \alpha=\text{home} \\ 0.15 & \text{on } \beta=\text{away} \end{cases} \quad (56)$$

While the information advantage and all the properties of the estimators stay the same as in Equation 50, the growth of wealth, accounting for the half of it being held separately, is now positive, particularly

$$W_G = \frac{1}{t} \cdot \log\left(\frac{W_t}{W_0}\right) = \sum_i r_i \cdot \log\left(0.5 + \frac{f_i}{m_i}\right) = 0.5 \cdot \log\left(0.5 + \frac{0.35}{0.3}\right) + 0.5 \cdot \log\left(0.5 + \frac{0.15}{0.7}\right) = 0.087 \quad (57)$$

In words, decreasing the amount invested into the first opportunity  $\alpha = \text{home}$  from 0.7 down to 0.3, which is now below the optimal  $f_\alpha^* = r_i^\alpha = 0.5$  for the full Kelly fraction, removes the overbetting problem, while keeping the second fraction on the loss-making opportunity  $\omega_i^\beta$  comparably low.

While the full Kelly forms an interesting corner case, oblivious to the correlation of the estimator w.r.t. the market, we have demonstrated that the decorrelation concept has a positive impact on the commonly used fractional Kelly modification, where decreasing the correlation  $\text{Corr}[M, T|R]$  consistently increases profits given that other properties of the distribution stay the same. Similarly, it also improves profitability of other common portfolio optimization strategies, such as the MPT, which we demonstrate practically in experiments (Section 5).

## 4.5 Machine Learning

So far, we have analysed profitability of the price estimators  $\mathbf{t}$  w.r.t. various market distributions  $P_\Omega$  and investment strategies  $\mathbf{s}$  to derive the desired decorrelation property. Nevertheless, we have not yet discussed how to actually create such profitable estimators  $\mathbf{t}$ . As outlined in Section 2.5, the common way to create a fundamental price estimator  $\mathbf{e}$  is by optimizing its fit to historical data  $\mathcal{D}$  via minimization of some error measure  $\text{err}(\mathbf{e})$ . The standard desideratum is then to predict the real values  $R$  within  $P_\Omega$  as closely as possible.<sup>32</sup>

However, we now know that being accurate is not the only possible desideratum of the estimator, and that for the subsequent trading it might be even more important to minimize the conditional correlation with the market  $\text{Corr}[M, T|R]$ . Note that an optimal trade-off between these properties of an estimator will naturally depend on the subsequent investment strategy being used (Section 2.7). For instance, we have shown that in the corner case of the full Kelly investor, minimizing the correlation has no impact on the profits at all (Section 4.4) and to maximize the Kelly growth (Section 2.7.3) it is sufficient to resort to the plain cross-entropy error (Section 2.5), minimizing the KL-distance from the true value distribution (Section 3.1). Nevertheless for the uniform investment strategy (Section 4) as well as other practical strategies, minimizing the covariance will generally improve profitability (Section 4.4.1).

Without discussing the optimal trade-off, we can roughly conclude that it is generally advisable to strive for an unbiased estimator with low variance and low covariance with the market. One can then alter the standard machine learning objectives to reflect these new desiderata, and validate the optimal setting experimentally. The scope of the corresponding error measure would thus include not only the standard estimates  $T$  and ground truth values  $R$ , but also the market estimates  $M$ , and could then look like

$$\text{err}^*(R, M, T) = \text{Bias}[T|R] + \text{Var}[T|R] + \text{Cov}[T, M|R] \quad (58)$$

Recall the purpose of the  $MSE$  measure (Equation 4), which is to minimize the squared distance from the real value as

$$MSE_\Omega(R, T) = \mathbb{E}[(T - R)^2] = \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} (t_i - r_i)^2$$

which can be thought of as trying to jointly minimize the bias and variance of the estimator (Section 2.5). A straightforward approach to reflect the outlined desiderata is to modify the existing  $MSE$  measure with an extra term to also penalize the covariance between  $M$  and  $T$  as

<sup>32</sup>while keeping the model complexity reasonably low for good generalization.

$$MSE_{\Omega}^*(R, M, T) = \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} (t_i - r_i)^2 + \gamma \cdot (t_i - r_i)(m_i - r_i) \quad (59)$$

where the extra “decorrelation term” is weighted by a tunable hyperparameter  $\gamma > 0$ . The purpose of  $\gamma$  is then to adjust the trade-off between the decorrelation term and the standard  $MSE$ , since the two normally represent opposing objectives, as the market maker tends to be very close to the real value.

We note that altering the  $MSE$  might seem counter-intuitive from the perspective of a perfect estimator, corresponding to the minimal  $MSE$ , where the additional term will only hurt its performance by pushing it away, not only from the market price, but from the true value, too, since  $\text{argmin}[MSE] \neq \text{argmin}[MSE^*]$ . However, recall that the decorrelation only makes sense when one is not able to train such a superior model (Section 4.3). In those common cases, the model generally occupies some wider area of the error landscape around the  $MSE$  minima, where the additional term is meant to navigate away from the correlated regions (Figure 1), which may be equally distanced from the  $MSE$  minimum, but will yield inferior profits.

While we do not attempt to argue about optimality of the proposed  $MSE^*$  metric, and we acknowledge that there are likely more appropriate measures to maximize the model profitability, we will demonstrate that this simple  $MSE$ -extension already works well enough in practice to proof viability of the decorrelation concept.

## 5 Application in Sports Betting

To demonstrate viability of the concept in a concrete, practical setting, we chose the domain of sports betting. Sports betting closely follows the outlined setup of the market taker  $t$  vs. market maker  $m$  stochastic pricing game over opportunities  $\Omega$  based in some tradeable assets  $\mathcal{A}$ . The commonalities were introduced in Table 1 and throughout the paper.

Nevertheless, there are some interesting technical differences to note in betting. Firstly, we do not need to model the future price distribution, as the true value is directly determined by the probability of the traded (future) outcome to occur. On the other hand, the true probability is not directly observable, and has to be recovered from realizations of the corresponding stochastic outcomes, similarly to the observations of the future prices. Secondly, the returns are determined by the odds, reflecting the inverse of the probability estimate, in a multiplicative fashion, as opposed to the stock market setting where one profits additively on the difference from the true price (Section 2.6). Consequently, the errors in the estimates also reflect themselves differently into the estimated expected returns. Last but not least, there are lots of publicly available data from the sports betting market, which make it generally interesting as a benchmark for testing the market efficiency and the associated hypotheses on profitability of models and trading strategies. We note that much of the material discussed in this section is adapted from our previous work [29].

### 5.1 Principles of Sports Betting

Following the established trading setting (Section 2.1), sports betting can be understood as a simple 2-player stochastic game where the bettor  $t$  repeatedly allocates wagers  $\mathbf{f}$  from her current bankroll  $W_\tau \in \mathbb{R}_+$  over available opportunities  $\omega_i^\zeta \in \Omega$ , corresponding to the stochastic outcomes  $\zeta \in Z$  of sport matches  $a \in \mathcal{A}$ . In case of the established two-way betting markets (Section 2.1) there are two possible match outcomes  $Z = \{\alpha, \beta\}$ . The outcomes  $\zeta$  can then be understood as coming from a Bernoulli distribution parameterized by the corresponding true probability value  $r_i$ . Each of the possible opportunities  $\omega_i^\zeta$  is then also associated with some probability  $m_i \in [0, 1]$  by the bookmaker  $M : \omega_i \mapsto m_i$ , which is presented indirectly through the offered market odds  $o_i \approx \frac{1}{m_i \pm \epsilon}$ , which include some margin  $\epsilon$  (Section 2.3). Should the outcome  $\zeta$  associated with a particular opportunity  $\omega_i^\zeta$  be realized, a payoff  $f_i^\zeta \cdot o_i^\zeta$  from the associated odds and the waged fraction is to be received by the bettor  $t$ . In the opposite case, the bettor loses the allocated part  $f_i^\zeta$  of her bankroll  $W_\tau$  to the bookmaker. The potential net profit  $w_i$  from allocation  $f_i$  on the opportunity  $\omega_i^\zeta$  is thus

$$w_i = \begin{cases} f_i^\zeta \cdot o_i^\zeta - 1 & \text{with prob. } r_i \quad (\text{if } Z = \zeta \text{ is realized}) \\ -f_i^\zeta & \text{with prob. } 1 - r_i \quad (\text{if } Z \neq \zeta) \end{cases} \quad (60)$$

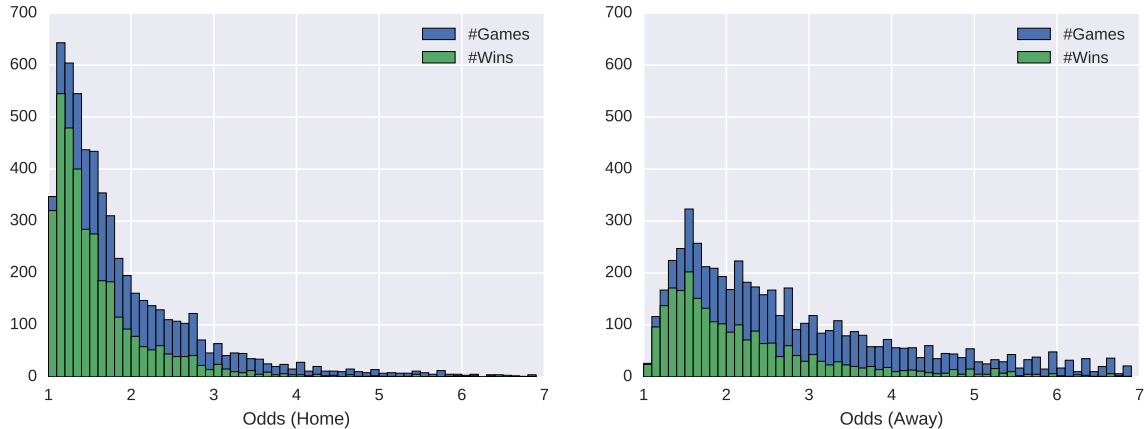


Figure 7: Raw odds histograms, displaying the progressive decrease in sample size with the increasing odds.

## 5.2 Data

For the experiments we chose the game of basketball as one of the world’s most popular sports. Conveniently, there is no draw in basketball matches, making the corresponding match outcomes ( $|Z| = 2$ ) betting market directly aligned with the two-way (home-away) market setting introduced so far. Particularly, we chose the National Basketball Association (NBA) league, which has one of the largest and most fluid betting markets, available through virtually any bookmaker. Consequently, the mainstream NBA market can be considered highly efficient and hard to beat [54, 29].

There are many data sources with a much detailed granularity, such as play-by-play transcripts or video records. Theoretically, these could enable to potentially extract information superior to that of the bookmaker (Section 2.4). For a proper demonstration of the decorrelation concept (Definition 4.1), we resort to clearly inferior data sources. For that, we suffice with basic player statistics, such as numbers of shots, passes, steals, etc. Conveniently, there are large historical datasets with comprehensive sets of such statistics publicly available for each NBA match. Particularly, we retrieved the official box score NBA data from the seasons 2000 to 2014.

**Features** The information we use for predicting the outcome of a match combines historical data relating to the players of the home and away teams, respectively. For each of the two, we aggregate the aforementioned basic quantitative measures of the players’ performances from all of the preceding matches since the beginning of the season in which each respective prediction takes place. Additionally, we also include seasonal aggregates from the preceding seasons.

Besides the historical track data considered above, the bookmaker’s odds  $o_i$  assigned to a match represent another piece of information, obviously highly relevant to the prediction of its outcome  $\zeta$ . While the odds  $o_i$  clearly present a very informative feature, their incorporation in a model naturally increases the undesirable correlation with the bookmaker (Section 4). Whether to omit or include the odds as a feature thus remains an interesting trade-off question, and so we further consider both the options for experimental evaluation.

## 5.3 Odds

For the betting odds, we used the *Pinnacle*<sup>33</sup> closing<sup>34</sup> odds for seasons 2010 – 2014. Note that, as opposed to opening odds, the used closing odds are already adjusted by the market, reflecting also new information that came from the traders, making it more difficult to beat, as it tends to be more robust against systematic errors. For the earlier seasons (2000 – 2010), we unfortunately had to collect odds data from multiple different bookmakers. Note that this also makes decorrelating against the market more difficult, as we do not aim to exploit errors of a particular market maker but an aggregate of many. Moreover, these bookmakers have

<sup>33</sup><https://www.pinnacle.com/>

<sup>34</sup>We note that, technically, these odds are impossible to place a bet upon as the market is just getting closed. However, the change in the odds at the very instant is typically negligible, and any preceding odds are commonly *easier* to beat.

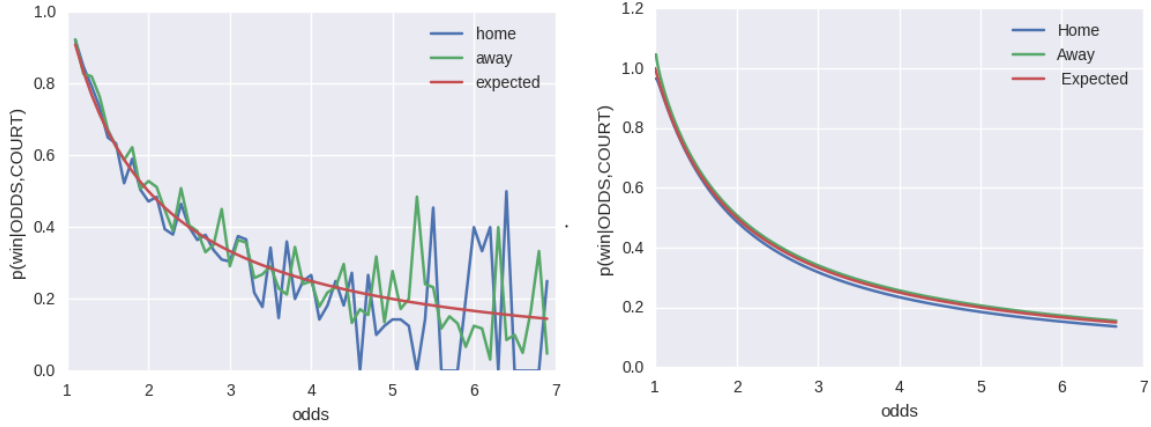


Figure 8: Raw projections of the data to the probability space may display artifacts with the increasing odds and the respective decrease in sample size (left). When smoothed out with a Beta distribution prior, the bookmakers’ odds distribution follows the true distribution very closely (right).

generally higher margins of app.  $\epsilon = 4.5\%$  as compared to the Pinnacle’s  $\epsilon = 2.5\%$ , making it more difficult to achieve returns that are positive in total (Section 2.3).

Figure 7 shows histograms of the odds distribution for the home and away teams and their winnings, respectively. Since the odds reflect the inverse of the estimated probabilities, both histograms exhibit the corresponding long-tail distributions. We can casually see that the proportion of the winning games progressively decreases with the increasing odds, and that the base winning probability of the home team is generally higher.<sup>35</sup>

### 5.3.1 Searching for the Bias

For the decorrelation concept (Definition 4.1) and betting in general, we were particularly interested in a potential statistical bias of the bookmaker. As argued in Section 3.4, should there be a statistical bias in the bookmaker’s odds (Section 2.5), it would allow to easily exploit the market in certain ranges where the odds  $o$  are systematically overestimated. For that we can simply take the histograms from Figure 7, and align the probability distribution induced by the inverted odds with the relative frequencies of the actual outcomes, the result of which is displayed in Figure 8. We can observe that for the lower odds near 1.0 the bookmaker seems to be very accurate, while the accuracy starts to fluctuate as the odds increase. While this might incite bettors to interpret as bias in the odds, as speculated in some previous works, we argue that this is simply an artefact due to the decreasing sample size with the increasing odds (Figure 7). Consequently, the odds would still be centered properly should more data be available.

To investigate further, we smoothed out the empirical data by fitting a corresponding Beta distribution, which removes the artefacts and reveals that the bookmakers probability distribution follows the true distribution very tightly.<sup>36</sup> An alternative way to reveal a systematic bias in the bookmaker’s odds is to train a statistical model to predict the outcomes based on the odds. Should there be a systematic error in the odds, such a model should be able to exploit it directly by providing more accurate predictions than the bookmaker, given enough data. However, our previous experiments [28] verified that, given a large enough out-of-sample measurement, obtaining such a model does not seem in the realm of possibility. From all our measurements, we can thus conclude that the bookmaker is *on average* precise, i.e.  $\forall r \in [0, 1] : \mathbb{E}_{P(M|r)}(M) = r$ , confirming the assumption of unbiasedness of the market maker (Section 3.4), and supporting the use of the decorrelation technique (Section 4), aiming to exploit the remaining (hopefully non-random) variance in the odds.

## 5.4 Experimental Setup

Here we describe the particular learning and investment strategies used, and the experimental protocol followed.

<sup>35</sup>commonly referred to as the home team advantage

<sup>36</sup>While there seems to be slightly more value on the home side of the market, which can be possibly attributed to a phenomenon known as the long-shot bias, there is no bias in the odds significant enough to be traded for profit.

**Learning Model** For the predictive modelling, we used a variant of a convolutional network [23] described in [28]. The input to the network are two matrices (for the home and away teams, respectively), with players in rows and all the player features in columns. Intuitively, the model may be perceived as a learnable aggregation function extracting all the player-related variables into some latent “team-level” feature space, which the subsequent fully-connected layers transform into the final home-away binary prediction. For a detailed description of the model and its hyper-parameters we refer to [28]. We note that only the decorrelation loss hyper-parameter  $\gamma$  (Equation 61) was varied for the analysis presented in this paper, with all the rest being fixed.

**Error Function** In sports betting, one cannot observe the true probability of an opportunity, even retrospectively after the match.<sup>37</sup> Instead, the label  $r_i \in \{\alpha, \beta\}$  provided with each learning sample  $\omega_i$  reflects merely the binary outcome realization  $\zeta$  endowed by the underlying probability. This poses a problem to the decorrelation term from Equation 59 which, using the binary realizations instead of the actual probability values, would degenerate severely.<sup>38</sup> Particularly, the decorrelated cases of  $m_i < r_i < t_i$  and  $t_i < r_i < m_i$  would be impossible to achieve with the binary values of  $r_i \in \{0, 1\}$ . To better accommodate the proposed  $MSE^*$  loss into this practical binary betting setting, we thus slightly alter the loss to directly decorrelate  $M$  and  $T$  instead as

$$MSE^*_\Omega(R, M, T) = \mathbb{E}[(T - R)^2 - \gamma \cdot (T - M)^2] = \frac{1}{|\Omega|} \sum_{\omega_i \in \Omega} (t_i - r_i)^2 - \gamma \cdot (t_i - m_i)^2 \quad (61)$$

where the  $m_i$  are the bookmaker’s probability estimates induced from the published odds  $o_i$ . We note that this is different from the direct penalization of the covariance of the residuals in Equation 59. Nevertheless, the motivation is that a similar effect should be achieved through the integration with the remaining  $MSE$  term, which forces the estimates  $t_i$  to be unbiased w.r.t.  $r_i$ . Hence simply penalizing model estimates that are too close to the market price should also decrease the partial covariance between the two.<sup>39</sup>

**Betting Strategy** Firstly, we evaluated the models by betting with the basic uniform investment strategy (Section 2.7.1), for which we derived the theoretical reasoning in Section 4. We further refer to this simple strategy as *unif*. Additionally, we also evaluated the classic portfolio optimization techniques of Markowitz (Section 2.7.2) and Kelly (Section 2.7.3) to investigate compatibility of the decorrelation concept with the standard investment practices. For these strategies we needed to calculate the expected returns of the opportunities. Following the probabilistic setting detailed in Equation 60, the expected profit can be defined as

$$\mathbb{E}_R[w_i] = r_i \cdot \left(\frac{f_i}{m_i} - 1\right) + (1 - r_i) \cdot (-f_i) = \left(\frac{r_i}{m_i} - 1\right) \cdot f_i \quad (62)$$

With the Kelly strategy, we had to resort to its quadratic approximation (Section 2.7.3) due to the computational complexity of numeric optimization of the logarithmic utility across a larger set of opportunities, which is commonly done with Kelly in practice [75]. We further refer to this quadratic approximation of the Kelly strategy as *qKelly*.

For the modern portfolio theory-based strategy of Markowitz, we further needed to calculate the covariances of the expected returns (Section 2.7.2). Assuming that there is no correlation between the assets, corresponding to the underlying sport matches played within a single round, it is sufficient to consider only the variances of the individual independent opportunities<sup>40</sup> instead of the whole covariance matrix  $\Sigma$ . Following from the underlying Bernoulli distribution, the variance of each opportunity can then be defined as

$$\text{Var}_R[w_i] = \mathbb{E}[w_i^2] - \mathbb{E}[w_i]^2 = (1 - r_i)r_i f_i^2 \frac{1}{m_i^2} \quad (63)$$

Naturally, we used the model estimates  $t_i$  instead of the true (unknown) values  $r_i$  in the actual calculations of both  $\mathbb{E}_T[w_i]$  and  $\text{Var}_T[w_i]$ . We then chose the particular portfolio  $\mathbf{f}$  following the Sharpe ratio criterion (Section 2.7.2), and used the algorithm of sequential quadratic programming [50] to identify its unique maximizer. We further refer to this strategy as *sharpe*.

<sup>37</sup>In contrast to the stock market setting, where fitting the model directly using the Equation 59 should work just fine.

<sup>38</sup>We note that this does not generally pose much problem to the standard  $MSE$  term, as a properly regularized model will naturally regress the underlying probability while trying to predict the discrete outcome labels over a large enough data sample.

<sup>39</sup>We also note that using other measures of (statistical) distance, such as the  $XENT$  in the loss should also work quite well, according to our practical experience.

<sup>40</sup>This is a somewhat simplifying assumption based on excluding the option to bet on both the exclusive game outcomes simultaneously which, however, works quite reasonably in practice.

**Evaluation Protocol** The sport match portfolios for the betting strategies were then formed repeatedly from the batches of app. 15 matches<sup>41</sup> played in each round of the NBA league, which were available simultaneously on the betting market. The profit from each such round was then evaluated independently, i.e. the same unit of budget was assumed for staking in each round.<sup>42</sup> Note that this repeated single-period portfolio optimization setting (Section 2.7.2) is different from the growth-based view on portfolio optimization in time, as assumed by Kelly (Section 2.7.3).<sup>43</sup>

For the model learning, the matches were further aggregated also on the level of the whole seasons. Training and evaluation of the models then followed the natural chronological order, i.e. only past data were ever used for training a model evaluated on the upcoming match. To ensure sufficient training data, the first season to be evaluated was 2006, with a training window made up of seasons 2000–2005, progressively expanding all the way to evaluation on 2014, trained on the whole preceding range of 2000–2013. In total, there were 9093 games in the evaluated seasons 2006–2014 which count towards the reported statistics. With this setting, we then evaluated 10 independent runs of the whole pipeline, including the model (re-)training, for each given decorrelation setting to obtain statistical significances of the mean returns.

## 5.5 Results

The main purpose of the experiments was to test the proposed decorrelation concept (Section 4), as materialized through the augmented loss function (Section 4.5) with machine learning from real data (Section 5.2). This is to analyse the relationships between the model accuracy, correlation and the resulting profits. Particularly, we are interested in the effects of the hyperparameter  $\gamma$  in the loss function (Equation 61), controlling the trade-off between the accuracy and decorrelation. A similar trade-off effect is also to be evaluated w.r.t. inclusion of the market odds as an input feature to the predictive model (Section 5.2). Lastly, we are interested in the interconnection of the decorrelated models with the investment strategies, and their relative performances. The results are displayed in Table 3 for the model without odds as features, and in Table 4 for the model with odds, respectively.

Firstly, as expected, we can observe from the results that models utilizing the highly informative odds feature achieve consistently higher accuracies. The accuracy (% of correctly predicted outcomes) of the bookmakers, calculated by predicting the team with the smaller odds to win, leveled over the seasons at  $69 \pm 2.5$  and was generally superior to all our models.<sup>44</sup> Similarly, the models that included the bookmakers' odds were anticipated to be more correlated with the bookmaker. This is convincingly confirmed by measurements of Pearson coefficients which stay at 0.87 for the models trained without odds as a feature, and 0.95 for the models which included them.

Finally, the results provide important insights into the relationship between decorrelation and profitability. In the analytical part of this paper, we have provided a theoretical support for lowering the correlation in order to improve model profitability (Section 4) and proposed a simple idea how to train such models from data (Section 4.5). However, it was not yet clear that such model properties can be actually achieved in practice against real market makers. From the results in Table 3, it is now clear that the proposed decorrelation loss (Equation 61), providing a flexible trade-off between accuracy and decorrelation, can be used as an effective proxy to achieve the desired decorrelation effect against a (highly-efficient) sports betting market. We can clearly see that augmenting the standard *MSE* with the decorrelation term generally helps both the basic *unif* as well as the advanced *sharpe* and *qKelly* strategies to generate more profits, while there is a sweet spot in tuning the trade-off parameter  $\gamma$ , generating the maximal returns w.r.t. each of the strategies. Additionally, we note that the number of bets placed remained basically constant<sup>45</sup> over all the strategies and levels of decorrelation, hence the reported rates of return are also directly comparable in terms of total profits generated.

<sup>41</sup>There are 30 teams in the NBA, hence one league round normally consists of 15 games.

<sup>42</sup>corresponding to the additive betting dynamics [75]

<sup>43</sup>Taking the growth-optimization approach would naturally lead to almost certain ruin of the inferior models (Section 3.1) in practice, too [75].

<sup>44</sup>We acknowledge that a model accuracy which is inferior to the (input) odds themselves might seem disappointing. However, it is exactly the point of this paper that the odds themselves, while serving as a highly accurate model, are completely useless for betting. It is only thanks to the extra features in the model that we are able to decorrelate from the odds, leading to positive returns, despite the decreased (inferior) accuracy.

<sup>45</sup>Theoretically, it might happen that the model predictions would fall inside the  $(m_i - \epsilon, m_i + \epsilon)$  spread area, resulting into neither of the two opportunities displaying positive expectation and being staked upon. However, this basically has not happened thanks to the very low margin of Pinnacle ( $\epsilon \approx 2\%$ ), resulting into basically all the games (9093) being staked upon in all of the settings.

Table 3: Accuracies and returns (in %) per round for the selected investment strategies and the prediction model *without* odds as a feature (Section 5.4), measured across different levels of decorrelation (Equation 61).

$\gamma$	$W_{\text{sharpe}}$	$W_{\text{qKelly}}$	$W_{\text{unif}}$	Accuracy
0.0	$0.38 \pm 0.10$	$0.22 \pm 0.11$	$-5.12 \pm 0.11$	$67.62 \pm 0.03$
0.2	$1.05 \pm 0.12$	$1.2 \pm 0.17$	$-3.31 \pm 0.13$	$67.47 \pm 0.03$
0.4	$1.74 \pm 0.14$	$2.33 \pm 0.3$	$-1.73 \pm 0.18$	$67.15 \pm 0.10$
0.6	$1.32 \pm 0.14$	$1.9 \pm 0.64$	$-0.61 \pm 0.28$	$66.19 \pm 0.09$
0.8	$1.10 \pm 0.29$	$-0.84 \pm 1.24$	$-0.39 \pm 0.22$	$64.93 \pm 0.35$
1.0	$-1.92 \pm 0.81$	$-11.73 \pm 2.74$	$-2.59 \pm 0.57$	$61.30 \pm 0.48$

Table 4: Accuracies and returns (in %) per round (akin to Table 3) for the model *with* odds as a feature.

$\gamma$	$W_{\text{sharpe}}$	$W_{\text{qKelly}}$	$W_{\text{unif}}$	Accuracy
0.0	$-0.12 \pm 0.24$	$-0.09 \pm 0.16$	$-3.83 \pm 0.22$	$68.80 \pm 0.06$
0.2	$0.72 \pm 0.13$	$0.59 \pm 0.17$	$-2.50 \pm 0.14$	$68.37 \pm 0.04$
0.4	$1.49 \pm 0.10$	$1.58 \pm 0.28$	$-1.30 \pm 0.12$	$67.48 \pm 0.10$
0.6	$1.02 \pm 0.20$	$2.33 \pm 0.7$	$-1.15 \pm 0.22$	$66.55 \pm 0.10$
0.8	$1.00 \pm 0.35$	$0.29 \pm 1.07$	$-0.45 \pm 0.28$	$65.19 \pm 0.27$
1.0	$-1.22 \pm 0.51$	$-9.93 \pm 1.68$	$-2.25 \pm 0.30$	$61.77 \pm 0.44$

Moreover, the benefits of the advanced investment strategies, such as the Sharpe and Kelly (Section 2.7), remain effective with the technique, as these more sophisticated strategies consistently dominate performance of the simple unif staking. Last but not least, with the **sharpe** and **qKelly** strategies, we were consistently able to generate statistically significant positive returns, despite the clearly inferior accuracy of the predictive models.

## 6 Related Work

**Portfolio optimization** The approach of splitting the trader’s workflow into the two steps of predictive modeling and investment optimization has a long tradition, and has been exploited in absolute majority of works [49, 30, 56, 51, 71, 18], with some notable exceptions [24, 38]. Extracting the parameter estimation out of the portfolio optimization problem then enabled the respective economic research to thrive in an isolated mathematical environment, giving rise to the frameworks of Markowitz [47] and Kelly [32], and their many successors [10, 77, 71, 36, 51]. While widely adopted, the optimality of the resulting portfolios is based on rather unrealistic assumptions, which has been progressively criticized by many [27, 61, 49, 57, 62, 43]. From the perspective studied in this paper, the main underlying issue is the separation from the problem of estimation of the return (price) parameters, which are simply assumed at input. The resulting issues with uncertainty in the portfolios are then typically mitigated with additional practical methods [42, 51, 31]. There are also some principled approaches to tackle the input parameter uncertainty, such as considering the portfolio optimization problem within the framework of Bayesian decision making [11, 6, 13] or distributionally robust setting [70, 8]. However, to our best knowledge, all of these methods are aimed at mitigating the additional (structural) risk, stemming from the uncertainty in the input parameters, rather than increasing the profits.

**From accuracy to profit** Since the return parameters are assumed at input, the resulting portfolio optimality is inherently based on their quality. This has been previously nicely demonstrated within the mathematically elegant framework of Kelly [32], resulting into the idea of information advantage [15]. Given the mathematical superiority of Kelly investments to any other strategy, this has further incited practitioners to focus merely

on striving for a superior predictive model, while considering the investment optimization a solved problem. This approach has been long established in the literature, where the Kelly strategy is often used to benchmark a model against a particular market maker [71]. Subsequently, the problem of beating the market has been commonly perceived as even more difficult than it actually is [66]. Nevertheless, without the information advantage required to profit with Kelly, practitioners often fell for the fractional modification [9, 4], yet without any further analysis of the implications on the relative performance of their models w.r.t. the market maker. Given the rising interest in the Kelly strategy among traders [33], the idea seems to gain further momentum, and it is not uncommon to see even researchers confuse profitability with accuracy [34, 39, 48].

**Beating the betting market** The problem of profiting on betting markets using statistical modelling has continuously attracted many researchers [72, 40]. As the trading was historically dominated by humans relying on their expert knowledge, this domain became a ground for another human versus machine battle. The idea that experts bring some additional value was questioned early by comparing magazine experts against a statistical model [65]. However, later it has been pointed out that these experts were often pressured to diversify their predictions, and showed that bookmakers under financial pressure can compete with the models [19]. As the betting industry grew, the focus shifted to comparing the models against the market makers. The market makers demonstrated the capability to outperform both experts and models [66]. In a study across multiple sports, betting lines proved to consistently outperform the statistical models [68]. Later, prediction markets, where multiple market makers compete for their positions, proved to be even more accurate than the bookmakers [67, 21]. One of the common caveats when evaluating the statistical models is the usage of datasets not large enough to draw statistically significant conclusions [25]. While there appear to be inefficiencies in the markets [53, 2], works showing positive returns on larger datasets remain scarce [29]. Selecting a particular bookmaker can then have a large effect on the profitability, as their forecasting capabilities vary considerably [69, 2].

**Profiting with inferior models** There have been a number of works experimentally demonstrating positive profits with a, presumably or verifiably, inferior predictive model [55, 4, 14, 34]. Nevertheless, as the main focus was again on the predictive modelling part, the phenomenon itself was not further analyzed in either of these works. Consequently, it was not clarified whether the idea is sound and the profits are not simply stemming from inappropriate measurements, such as the common problem of a small sample size [20, 48]. To our best knowledge, the explicit idea of exploiting a betting market with a less accurate model was first openly tackled in [28], the content of which was later presented in [29]. This is also the work upon which we directly follow in this paper with a more in-depth analysis of the decorrelation concept. A very closely related is a recent work of [78], where the authors also conclude that the relationship between model accuracy and betting returns has not yet been sufficiently studied, and clearly demonstrate that it is possible to generate positive returns in absence of superior predictive accuracy. They further analyse the underlying reasons through illustrative examples and a simulation, similarly to [29].<sup>46</sup> While their approach is not directly transferable into a practical setting, they do provide a more detailed study of the bookmaker’s margin, n-way markets, and the problem of a small sample size [25]. W.r.t. [78] and [29], in this paper we aimed to provide a more thorough analysis of the relationships between the underlying model properties, market distributions, and investment strategies.<sup>47</sup>

## 7 Conclusion

We have introduced a statistical setting in which an inferior (“bad”) predictive model can be used to generate profits in a (partially) inefficient market. Importantly, we did not restrict the properties of the market (maker) any further, and demonstrated the concepts across two different domains of stock and betting markets. After detailing an extensive background on the topic, we provided some important insights into the problem of profitability w.r.t. the properties of the predictive models to show that the relationship goes well beyond the

---

<sup>46</sup>As the authors themselves note, the simulation in [78] differs from [29] in that they directly assume correlation between prediction errors, as opposed to the predictions themselves. While this reflects the underlying principle more directly, it is also considerably unrealistic, since the error is not observable in practice, as well as the true probability required for the reported MSE calculation. Consequently, ad-hoc assumptions about the market distribution had to be made in [78], as opposed to the beta distributions (i.e. conjugate priors to the observed Bernoulli trials) fitted directly to the market data in [29], and the associated practical machine learning objective. We also note that, despite decorrelating the errors and decorrelating the estimates are different in principle, both approaches lead to the same statistical effect on profitability given the equal means (0.5) of all the estimators [29].

<sup>47</sup>including the Kelly strategy the study of which in this context was proposed to be done in [78]



basic measures of predictive accuracy. We thoroughly analysed the relationship from the perspective of the joint distribution of the price estimators, and proved that decreasing the partial correlation of the market taker with the market maker directly benefits the trader’s profitability for most common market distributions and investment strategies.

Additionally, we translated the analysis into a practical machine learning setting, and demonstrated viability of the “decorrelation” concept in practice. Particularly, we have shown on real world data from a highly efficient betting market that incorporating the decorrelation objective consistently improves the profits across diverse settings, and were even able to achieve statistically significant positive total returns.

**Future Work** There are lots of avenues to improve the presented work. For instance, the chosen approach of statistical explanation of the relationships between the estimators through the standard measure of correlation (covariance) follows mostly from convenience of description rather than appropriateness for the particular problem in scope. While the decorrelation concept coincides nicely with profitability in some particular settings assumed (Section 4), it would be generally more appropriate to base the analysis directly on an integral over (the discrete regions of) the vector field of profitability (Figure 1). While correlation is a convenient intermediate link for explanation of the phenomenon, it is certainly not suited to convey the problem insights in full.

Similarly, it is clear that the proposed *MSE*-based loss function to be optimized is rooted merely in an intuition rather than formal mathematical derivation. There is also the unfortunate discrepancy between the theoretical loss (Equation 59) and its practical form used in the betting experiments (Equation 61). Consequently, it is very likely that a more appropriate loss function can be derived. Particularly, the loss should take the context of the subsequent investment strategy and the corresponding utility function into account (such as the *XENT* for Kelly). As a general direction, instead of regression of the true asset value, it would seem more appropriate to consider the proposed scenario as a robust classification problem w.r.t.  $M$  and  $R$  from the start. This would render measures such as Hinge loss as a more viable alternative for adaptation (modification) into the proposed setting. Ultimately, it would be naturally desirable to integrate the whole workflow into a single end-to-end profit (utility) optimization problem, removing the need for the intermediate analysis and the surrogate loss function completely.

## Acknowledgments

The authors acknowledge support by a grant no. 20-29260S granted by the Czech Science Foundation.

Access to the CERIT-SC computing and storage facilities provided by the CERIT-SC Center, under the programme “Projects of Large Research, Development, and Innovations Infrastructures” (CERIT Scientific Cloud LM2015085), is greatly appreciated.

$\text{Corr}[T, R]$	$\text{Corr}[T, M]$	$\varnothing W_{\text{sharpe}}$	$\varnothing W_{\text{unif}}$	Accuracy	Consensus	Upset	Missed	Spotted
0.85	0.85	11.15	3.14	70.11	61.99	20.37	9.53	8.12
	0.90	6.14	0.52	70.05	63.60	22.04	7.91	6.45
	0.95	-1.73	-5.46	70.08	65.74	24.12	5.80	4.34
0.90	0.85	18.14	8.56	71.48	62.66	19.70	8.82	8.83
	0.90	14.05	5.74	71.48	64.36	21.35	7.17	7.12
	0.95	9.60	3.38	71.45	66.39	23.50	5.05	5.06
0.95	0.85	25.30	13.42	72.91	63.34	18.98	8.11	9.57
	0.90	22.95	12.69	72.93	65.02	20.62	6.46	7.91
	0.95	20.79	11.87	72.92	67.21	22.74	4.33	5.71

Table 5: Average profits (in %)  $W_{\text{sharpe}}$  of the *sharpe* and  $W_{\text{unif}}$  of the *unif* strategies w.r.t. to the correlations of the (estimated) probabilities. Accuracy denotes the % of correct outcome predictions by the bettor (predict win if  $t_i > 1/2$ ). The four last columns break down the proportions (in %) of different combinations of predictions by  $t$  (bettor) and  $m$  (bookmaker): *Consensus* (both right), *Upset* (both wrong), *Missed* (bettor wrong, bookmaker right), *Spotted* (bettor right, bookmaker wrong).

## A Appendix

### A.1 Simulated Data

To further demonstrate the proposed decorrelation concept (Definition 4.1), we present a simulation adopted from [29]. Here we simulated the ground truth  $R$  as well as both of the estimates  $T, M$  with various levels of their correlation, and measured the profits made by the *sharpe* and *unif* strategies for these different levels.

More precisely, we sampled triples  $R, T, M$  from a multi-variate Beta distribution. The distribution is parameterized with the marginal means and variances of the three variables and their pair-wise correlations. The mean of each of the three variables was set to  $1/2$ , reflecting the mean probability of the complementary binary outcomes. The variance of  $M$  was determined as 0.054 from real bookmaker’s data (Section 5.2), and  $T$ ’s variance copies this value. The variance of  $R$  was set to 0.08 reflecting the *glass ceiling* hypothesis.<sup>48</sup>

We let the correlations  $\text{Corr}(T, R)$ ,  $\text{Corr}(T, M)$  and  $\text{Corr}(R, M)$  range over the values  $\{0.85, 0.90, 0.95\}$ . The former two represent the independent variables of the analysis, acting as factors in Table 5, while the presented numbers average over the 3 values of  $\text{Corr}(R, M)$ . For each setting of  $\text{Corr}(T, R)$  and  $\text{Corr}(T, M)$ , we drew  $r_i, t_i, m_i$  ( $i = 1, 2, \dots, n = 30$ ) samples, to simulate one round of betting. Then we set the odds  $o_i = 1/m_i$  (the bookmaker’s margin being immaterial here) for  $1 \leq i \leq n$ , and determined the bets  $f_1, f_2, \dots, f_n$  from  $o_1, o_2, \dots, o_n$  and  $t_1, t_2, \dots, t_n$  using the *sharpe* and *unif* strategies. Finally, the match outcomes were established by a Bernoulli trial for each of the  $r_1, r_2, \dots, r_n$ . With these inputs, we calculated the cumulative profit  $W = w_1 + w_2 + \dots + w_n$  of one round. This procedure was repeated 10 000 times (rounds), averaging the  $W_{\text{sharpe}}$  and  $W_{\text{unif}}$  returns.<sup>49</sup>

Table 5 then shows the average profits as well as the accuracy of the bettor’s outcome predictions (call win if  $T > 1/2$ ), and the percentual breakdown of 4 possible combinations of bettor’s and bookmaker’s predictions. The accuracies, as well as the four latter proportions, are also averaged over all bets in all simulated rounds.

Besides the unsurprising observation that bettor’s prediction accuracy grows with  $\text{Corr}(T, R)$ , the results show that profits indeed decay systematically as the bettor’s and bookmaker’s predictions become more correlated (increasing  $\text{Corr}(T, M)$  decreases profit). An instructive observation is that the proportion of *spotted* opportunities is in all cases higher when the bookmaker’s and bettor’s predictions are less correlated. Moreover, we can see that the betting strategy is another independent factor strongly influencing the profit, with the *sharpe* strategy being superior to the *unif* strategy.

Following the projection of the space of the  $(R, M, T)$  estimates onto profitability displayed Figure 1, and analysed in Section 3.3, we also provide a direct 2D projection of a similar simulated betting scenario in Figure 9. We can again observe that decreasing the correlation is directly reflected in increase of profits.

<sup>48</sup>Articulated by [79] and expressing that sports games are predictable with a maximum of 75% accuracy at best. When  $R$  is sampled with mean  $1/2$  and variance 0.08, then with 0.75 probability the event ( $R > 1/2$ ) predicts correctly the outcome of a Bernoulli trial parameterized with  $R$ .

<sup>49</sup>Note that this is not the same (for  $W_{\text{sharpe}}$ ) as setting  $n = 30 \cdot 10000$  without repeating the procedure, as the full unit budget is supposed to be spent in each round.

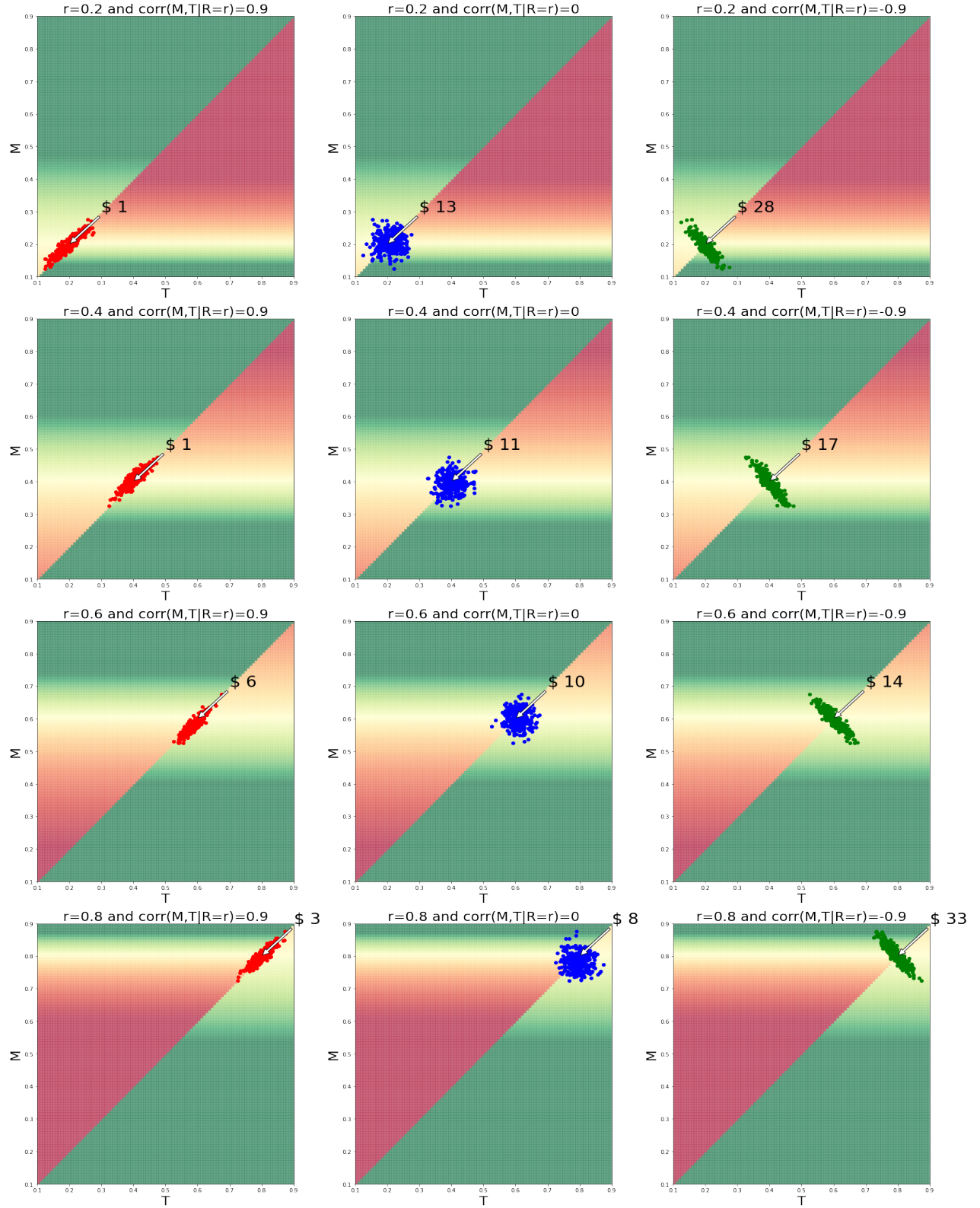


Figure 9: Particular examples of partially correlated, independent and negatively correlated betting w.r.t. varying levels of the true probability  $r$ . Measured on a random sample from a multivariate distribution with means in the respective  $(r, r, r)$  and equal variances of both the estimators  $\text{Var}[T|R] = \text{Var}[M|R]$ .

## References

- [1] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.
- [2] Giovanni Angelini and Luca De Angelis. Efficiency of online football betting markets. *International Journal of Forecasting*, 2018.
- [3] Kenneth Joseph Arrow. *Aspects of the theory of risk-bearing*. Yrjö Jahnssonin Säätiö, 1965.
- [4] Rose D Baker and Ian G McHale. Forecasting exact scores in national football league games. *International Journal of Forecasting*, 29(1):122–130, 2013.
- [5] Rose D Baker and Ian G McHale. Optimal betting under parameter uncertainty: improving the kelly criterion. *Decision Analysis*, 10(3):189–199, 2013.
- [6] Jeremy Balka, Anthony Desmond, et al. *Kelly Investing with Iteratively Updated Estimates of the Probability of Success*. PhD thesis, 2017.
- [7] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.
- [8] Jose Blanchet, Lin Chen, and Xun Yu Zhou. Distributionally robust mean-variance portfolio selection with wasserstein distances. *arXiv preprint arXiv:1802.04885*, 2018.
- [9] Georgi Boshnakov, Tarak Kharrat, and Ian G McHale. A bivariate weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2):458–466, 2017.
- [10] Leo Breiman et al. *Optimal gambling systems for favorable games*. 1961.
- [11] Sid Browne and Ward Whitt. Portfolio choice and the bayesian kelly criterion. *Advances in Applied Probability*, 28(4):1145–1176, 1996.
- [12] Enzo Busseti, Ernest K Ryu, and Stephen Boyd. Risk-constrained kelly gambling. *arXiv preprint arXiv:1603.06183*, 2016.
- [13] Dani Chu, Yifan Wu, and Tim B Swartz. Modified kelly criteria. *Journal of Quantitative Analysis in Sports*, 14(1):1–11, 2018.
- [14] Anthony C Constantinou, Norman E Fenton, and Martin Neil. pi-football: A bayesian network model for forecasting association football match outcomes. *Knowledge-Based Systems*, 36:322–339, 2012.
- [15] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.
- [16] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.
- [17] Eugene F Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of financial economics*, 49(3):283–306, 1998.
- [18] Alistair D Fitt. Markowitz portfolio theory for soccer spread betting. *IMA Journal of Management Mathematics*, 20(2):167–184, 2009.
- [19] David Forrest, John Goddard, and Robert Simmons. Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21(3):551–564, 2005.
- [20] David Forrest and Robert Simmons. Sentiment in the betting market on spanish football. *Applied Economics*, 40(1):119–126, 2008.
- [21] Egon Franck, Erwin Verbeek, and Stephan Nüesch. Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting*, 26(3):448–459, 2010.

- [22] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.
- [23] Kunihiro Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982.
- [24] Clive WJ Granger and M Hashem Pesaran. Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19(7):537–560, 2000.
- [25] Maral Haghighat, Hamid Rastegari, and Nasim Nourafza. A review of data mining techniques for result prediction in sports. *Advances in Computer Science: an International Journal*, 2(5):7–12, 2013.
- [26] Charles R Henderson. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, pages 423–447, 1975.
- [27] Lisa Holton. Is markowitz wrong? market turmoil fuels nontraditional approaches to managing investment risk. *Journal of Financial Planning*, 22(1):20, 2009.
- [28] Ondrej Hubacek. Exploiting betting market inefficiencies with machine learning. Master’s thesis, Czech Technical University, Faculty of Electrical Engineering, 2017.
- [29] Ondrej Hubacek, Gustav Sourek, and Filip Zelezny. Exploiting sports-betting market using machine learning. *International Journal of Forecasting*, 35(2):783–796, 2019.
- [30] Philippe Jorion. Portfolio optimization in practice. *Financial analysts journal*, 48(1):68–74, 1992.
- [31] Raymond Kan and Guofu Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, pages 621–656, 2007.
- [32] John L Kelly Jr. A new interpretation of information rate. *The Bell System Technical Journal*, 1956.
- [33] Gyutai Kim and Jong-Ho Shin. A comparison of the kelly criterion and a mean-variance model to portfolio selection with kospi 200. *Industrial Engineering & Management Systems*, 16(3):392–399, 2017.
- [34] Siem Jan Koopman and Rutger Lit. A dynamic bivariate poisson model for analysing and forecasting match results in the english premier league. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 167–186, 2015.
- [35] Solomon Kullback. *Information theory and statistics*. Courier Corporation, 1997.
- [36] Henry Allen Latane. Criteria for choice among risky ventures. In *The Kelly capital growth investment criterion: theory and practice*, pages 35–46. World Scientific, 2011.
- [37] Erich L Lehmann and George Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.
- [38] Gordon Leitch and J Ernest Tanner. Economic forecast evaluation: profits versus the conventional error measures. *The American Economic Review*, pages 580–590, 1991.
- [39] Stefan Lessmann, Ming-Chien Sung, and Johnnie EV Johnson. Alternative methods of predicting competitive events: An application in horserace betting markets. *International Journal of Forecasting*, 26(3):518–536, 2010.
- [40] Steven D. Levitt. Why are gambling markets organised so differently from financial markets? *Economic Journal*, 114(495):223–246, apr 2004.
- [41] Bin Li and Steven CH Hoi. Online portfolio selection: A survey. *ACM Computing Surveys (CSUR)*, 46(3):1–36, 2014.
- [42] Leonard C MacLean, Edward O Thorp, Yonggan Zhao, and William T Ziemba. Medium term simulations of the full kelly and fractional kelly investment strategies. In *The Kelly Capital Growth Investment Criterion: Theory and Practice*, pages 543–561. World Scientific, 2011.

- [43] Leonard C MacLean, Edward O Thorp, and William T Ziemba. Good and bad properties of the kelly criterion. *Risk*, 20(2):1, 2010.
- [44] Leonard C MacLean, Edward O Thorp, and William T Ziemba. *The Kelly capital growth investment criterion: Theory and practice*, volume 3. world scientific, 2011.
- [45] Leonard C MacLean, William T Ziemba, and George Blazenko. Growth versus security in dynamic investment analysis. *Management Science*, 38(11):1562–1585, 1992.
- [46] Burton G Malkiel. Efficient market hypothesis. In *Finance*, pages 127–134. Springer, 1989.
- [47] Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.
- [48] Ian McHale and Alex Morton. A bradley-terry type model for forecasting tennis match results. *International Journal of Forecasting*, 27(2):619–630, 2011.
- [49] Richard O Michaud and Robert O Michaud. *Efficient asset management: a practical guide to stock portfolio optimization and asset allocation*. Oxford University Press, 2008.
- [50] Jorge Nocedal and Stephen J Wright. Sequential quadratic programming. *Numerical optimization*, pages 529–562, 2006.
- [51] Edmund Noon, William J Knottenbelt, and Daniel Kuhn. Kelly’s fractional staking updated for betting exchanges. *IMA Journal of Management Mathematics*, 24(3):283–299, 2013.
- [52] Marco Ottaviani and Peter Norman Sørensen. The favorite-longshot bias: An overview of the main explanations. In *Handbook of Sports and Lottery markets*, pages 83–101. Elsevier, 2008.
- [53] Rodney J Paul and Andrew P Weinbach. The determinants of betting volume for sports in north america: Evidence of sports betting as consumption in the nba and nhl. *International Journal of Sport Finance*, 5(2):128, 2010.
- [54] Rodney J Paul, Andrew P Weinbach, and Mark Wilson. Efficient markets, fair bets, and profitability in nba totals 1995–96 to 2001–02. *The Quarterly Review of Economics and Finance*, 44(4):624–632, 2004.
- [55] Thomas Peeters. Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1):17–29, 2018.
- [56] Andre F Perold. Large-scale portfolio optimization. *Management science*, 30(10):1143–1160, 1984.
- [57] Ole Peters and Murray Gell-Mann. Evaluating gambles using dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(2):023103, 2016.
- [58] Georg Ch Pflug, Alois Pichler, and David Wozabal. The  $1/n$  investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417, 2012.
- [59] Keith Pilbeam. *Finance & financial markets*. Macmillan International Higher Education, 2018.
- [60] Brian M Rom and Kathleen W Ferguson. Post-modern portfolio theory comes of age. *Journal of Investing*, 3(3):11–17, 1994.
- [61] Paul A Samuelson. The “fallacy” of maximizing the geometric mean in long sequences of investing or gambling. *Proceedings of the National Academy of sciences*, 68(10):2493–2496, 1971.
- [62] Paul A Samuelson. Why we should not make mean log of wealth big though years to act are long. In *The Kelly Capital Growth Investment Criterion: Theory and Practice*, pages 491–493. World Scientific, 2011.
- [63] Paul A Samuelson. Proof that properly anticipated prices fluctuate randomly. In *The world scientific handbook of futures markets*, pages 25–38. World Scientific, 2016.
- [64] William F Sharpe. The sharpe ratio. *Journal of portfolio management*, 21(1):49–58, 1994.

- [65] Robert Simmons and David Forrest. Forecasting sport: the behaviour and performance of football tipsters. *International Journal of Forecasting*, 16(3):317–331, 2000.
- [66] ChiUng Song, Bryan L. Boulier, and Herman O. Stekler. The comparative accuracy of judgmental and model forecasts of American football games. *International Journal of Forecasting*, 23(3):405–413, 2007.
- [67] Martin Spann and Bernd Skiera. *Sports forecasting: A comparison of the forecast accuracy of prediction markets, betting odds and tipsters*, volume 28. John Wiley & Sons, Ltd., jan 2009.
- [68] H. O. Stekler, David Sendor, and Richard Verlander. Issues in sports forecasting. *International Journal of Forecasting*, 26(3):606–621, 2010.
- [69] Erik Strumbelj. On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4):934–943, 2014.
- [70] Qingyun Sun and Stephen Boyd. Distributional robust kelly gambling. *arXiv preprint arXiv:1812.10371*, 2018.
- [71] Edward O Thorp. The kelly criterion in blackjack sports betting, and the stock market. In *Handbook of asset and liability management*, pages 385–428. Elsevier, 2008.
- [72] Robert R Trippi and Duane DeSieno. Trading equity index futures with a neural network. *Journal of Portfolio Management*, 19:27–27, 1992.
- [73] Matej Uhrin. System for autonomous betting with optimal wealth allocation. Master’s thesis, Czech Technical University, Faculty of Electrical Engineering, 2018.
- [74] Matej Uhrin, Gustav Šourek, Ondrej Hubáček, and Filip Zelezný. Sports betting strategies: an experimental review. 2019.
- [75] Matej Uhrin, Gustav Šourek, Ondrej Hubáček, and Filip Zelezný. Optimal sports betting strategies in practice: an experimental review. *IMA Journal of Management Mathematics*, 2021.
- [76] Vasilii Grigorevich Voinov and Mikhail Stepanovich Nikulin. *Unbiased Estimators and Their Applications: Volume 1: Univariate Case*, volume 263. Springer Science & Business Media, 2012.
- [77] Chris Whitrow. Algorithms for optimal allocation of bets on many simultaneous events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(5):607–623, 2007.
- [78] Fabian Wunderlich and Daniel Memmert. Are betting returns a useful measure of accuracy in (sports) forecasting? *International Journal of Forecasting*, 36(2):713–722, 2020.
- [79] Albrecht Zimmermann, Sruthi Moorthy, and Zifan Shi. Predicting college basketball match outcomes using machine learning techniques: some results and lessons learned. *arXiv preprint arXiv:1310.3607*, 2013.