Biased updating creates a gambling trap: Evidence from sports betting in Kenya

Joshua E. Blumenstock*    Matthew Olckers†

January 8, 2020

Preliminary version. Please do not cite or distribute.

Abstract

We develop a model to show how biased updating can lead to persistent overconfidence in one’s ability, and highlight the negative welfare implications of this overconfidence. We validate key assumptions and predictions of this model using a unique dataset that captures rich details on the gambling decisions of over 50,000 Kenyans. The data show that gamblers react asymmetrically to (exogenous variation in) wins and losses. The bias in the learning process causes gamblers to increase betting expenditures over time. Exogenous increases in betting expenditures cause gamblers to take out high-interest loans, thus creating scope for persistent debt traps.


We thank Milo Bianchi and Jonathan Guryan for helpful and detailed feedback. The project started while Matthew was a PhD student at the Paris School of Economics and a visiting student researcher at UC Berkeley. We are grateful to researchers from both institutions for helpful comments. In particular, we thank Francis Bloch, Margherita Comola, Fabrice Etilé, Simon Gleyze, Sylvie Lambert, Liam-Wren Lewis, Karen Macours and David Margolis.

*U.C. Berkeley
†Monash University
1 Introduction

The popularity of gambling poses a challenge to expected utility theory, the benchmark framework of choice under risk. The theory predicts a risk averse person would not accept a fair gamble — let alone place wagers with negative expected returns (Tversky and Wakker, 1995). So, why do people gamble at all?

This question has troubled policymakers in developed economies like the United States, where gambling losses account for roughly 2% of average household consumption (Kearney, 2005), and where gambling has been associated with increased indebtedness, crime, and addiction (Grote and Matheson, 2013). It is also of increasing relevance in developing economies, where the popularity of online sports betting has exploded in recent years. Enabled by the widespread adoption of mobile money, which makes is possible for people to place small bets via their mobile phones, recent surveys suggest that between 37% and 60% of urban youth have recently gambled online.1

This paper focuses on the potential for one particular type of cognitive bias — overconfidence — to increase demand for gambling.2 We begin by developing a simple model of gambling behavior to explain why people may start betting, and how betting behavior adapts to experience. The model illustrates how biased updating, whereby a gambler updates his belief in his own ability asymmetrically in response to wins and losses, can generate overconfidence, which in turn leads to persistent gambling.

We then exploit a novel empirical setting to look for evidence of biased updating and overconfidence, and to test the model’s predictions. Specifically, we analyze a large administrative dataset that contains the betting histories and mobile money transactions of tens of thousands of Kenyan mobile phone owners. In this sample of primarily urban, primarily male individuals, roughly 62% use one or more sports betting services. We also exploit a discontinuity in the gambling system — whereby gamblers win only if their correct predictions exceed a specific threshold — to sep-

---

1A survey of urban Kenyan youth found that 60 percent of respondents had placed bets on football matches in the past and a further 16 percent had tried other forms of betting (GeoPoll, 2017). Ahaibwe et al. (2016) estimates that 37% of adult males in the capital of Uganda, Kampala, have placed bets in the past year. More generally, the most popular internet search query in Kenya in 2018 was SportPesa, Kenya’s leading sports betting operator. SportPesa received four times more search queries than Facebook, the most popular search query worldwide.

2We consider overconfidence to be the belief that one’s ability is greater than it actually is. See ? for a recent review of the literature.
arately identify the effect of successful gambles (which cause biased updating) from the income effect of winning (which may independently induce future gambles). This separation is critical to our model, and is not possible in traditional settings, where income effects cannot be differentiated from gambling success.

Three main empirical results stand out in this analysis. First, and consistent with recent legal evidence on sports betting, we find evidence of substantial heterogeneity in gambling ability (??). Put simply, some gamblers consistently predict the outcome of sports games better than others. Second, we find strong evidence of biased updating about one’s own ability. When people correctly predict the outcomes of the previous week’s matches — but do not have sufficient correct predictions to actually win any money — they are significantly more likely to bet in subsequent weeks. By contrast, when their past week’s predictions are unusually poor, it has very little effect on future behavior. In other words, an asymmetry exists whereby gamblers react strongly to past success but not to past failure.

Our third main result highlights the negative effects that overconfidence in gambling has on welfare. In addition to detailed information on sports betting transactions, our data includes information on mobile loan repayments and defaults. We use these data to study how sports betting impacts loan repayments, using the gambling performance in the prior week as an instrument for subsequent betting expenditure. We find suggestive evidence that sports betting increases the likelihood of loan default.

These results provide a link between the economics literature on the causes and consequences of gambling, and a separate literature on overconfidence. In the former literature, common explanations for the popularity of gambling include appending an entertainment value of gambling to the expected utility framework (Conlisk, 1993), overweighting the probability of extreme events (Tversky and Kahneman, 1992; Barberis, 2012), the addictive nature of gambling (Becker and Murphy, 1988; Guryan and Kearney, 2010), and possible convexities in the utility function over wealth (Friedman and Savage, 1948). In work most closely related to our own, Herskowitz (2019) argue

---

3The rise of sports betting in Kenya has occurred concurrently with the mass expansion of consumer loans that are provided via mobile phones. See ?.

4The exclusion restriction requires that past gambling performance can only impact future loan default through current gambling expenditures. As we discuss in detail below, we believe this assumption is reasonable, after conditioning on past expenditures and several other possible confounding variables, and after eliminating any possible income effects of past gambling (recall that income effects are eliminated by the structure of the gambling scheme).
that sports betting in Uganda may be partially motivated by the need to finance large, lumpy expenditures.\footnote{Large indivisible purchases such as a car (Ng, 1965) or the transaction costs of changing commitment goods such as housing (Chetty and Szeidl, 2007) induce convexity into the utility function over wealth. Gambling could fund such large purchases, especially for people who lack access to credit. (Crossley, Low, and Smith, 2016) also finds empirical support for this theory in the United Kingdom.} Through a randomized control trial and two lab-in-the-field experiments, Herskowitz (2019) shows that Ugandan men gamble to raise funds for large lump sum purchases. He suggests that limited access to credit and savings products makes gambling an attractive method of raising funds.

The economics literature on overconfidence dates back, per \footnote{The \textit{Journal of Economic Perspectives} (Vol. 29, No. 4) contains a recent symposium on overconfidence. See also Hoffrage (2004) for a review of research on overconfidence in psychology.}, to Adam Smith’s \textit{Wealth of Nations}.\footnote{The \textit{Journal of Economic Perspectives} (Vol. 29, No. 4) contains a recent symposium on overconfidence. See also Hoffrage (2004) for a review of research on overconfidence in psychology.} Overconfidence has received particular attention in the finance literature, as a way explain why individuals trade excessively on financial markets (Barber and Odean, 2001; Barber et al., 2008, 2018; Daniel and Hirshleifer, 2015). For instance, an early study of American football betting found overconfidence fit the data better than assuming bettors are risk-loving (Golec and Tamarkin, 1995) and experimental tests suggest people are overconfident when predicting the results of soccer matches (Erceg and Galić, 2014). Since sports betting and trading on financial markets are similarly structured, the same cognitive biases could drive behavior in both activities. In this way, sports betting markets serve as simplified financial markets so we can draw insights from sports betting behavior which may too difficult to study in financial markets (Sauer, 1998). In our case, we can separate the positive signals of successful predictions from the income effect of winning, which would be far more difficult in financial markets.

Finally, this paper relates to a recent literature on the welfare effects of mobile money and other digital financial services (Jack and Suri, 2014; Suri and Jack, 2016; Blumenstock, Callen, and Ghani, 2018). In this literature, Kenya has been celebrated for the country’s early and widespread adoption of mobile money, and yet mobile money allows people who do not have a bank account or credit card to gamble online. With mobile money technology, placing a bet is a simple as sending a text message. As mobile money adoption increases in other developing countries, we may see similar growth in online sports betting.

The findings in this paper can also provide some guidance to policymakers interested in reducing the negative consequences of gambling. We show that biased
updating creates overconfidence, which in turn leads to persistent gambling — and that this gambling eventually leads to consumer default. In this sense, we highlight the role that cognitive biases play in persistent and detrimental behavior (Guryan and Kearney, 2008). If gambling is driven by cognitive biases like overconfidence, one way to reduce the bias is to make people aware of the bias. Indeed, recent experiments suggest that helping an individual to update her believe about her own probability of winning the lottery can decrease demand for gambling (Abel, Cole, and Zia, 2015). In contrast to lotteries, sports betting has some role for skill rather than luck. If people can believe they are luckier than average then they can be sure to believe they are more skillful than average. We suggest that overconfidence in one’s own gambling ability can help to explain why people bet on sports.

2 Model of gambling behavior

We provide a simple model of gambling behavior to explain why people may start betting and how betting behavior adapts to experience.

A gambling operator offers a bet on a match with two outcomes, either the home team wins or the home team loses. For simplicity, we ignore draws. Also for simplicity, we assume the teams are evenly matched so the gambling operator offers even odds.

A gambler can predict the outcome of the match with probability \( p \), which represents the gambler’s ability. Sports betting differs from other gambles such as lotteries in that \( p \) can differ between gamblers. In sports betting, some gamblers may have better knowledge or more skill in determining the outcome of a match whereas, in a lottery, the probability of winning is identical for all gamblers.

The gambler does not know his ability. He has a prior belief \( \Pr[p] \sim \text{Beta}(\alpha, \beta) \).

---

7Abel, Cole, and Zia (2015) conducted a field experiment in South Africa, where researchers asked participants to roll dice and count the number of sixes they rolled (as a way to demonstrate the tiny probability of winning the national lottery). Participants who took many rolls to get a six decreased demand for lottery tickets relative to the control group whereas those who took a few rolls increased demand. A similar experiment in Thailand demonstrated the probability of winning the national lottery with a poster containing one million dots representing the number of tickets sold and several pins representing number of winning tickets per prize category (Zenker, Vollmer, and Wagener, 2018). The demonstration lead to improved knowledge but no change in the willingness to pay for lottery tickets. To reconcile the results of these two experiments, perhaps an individual’s belief in his own probability of winning can differ from his belief about the population average. Abel, Cole, and Zia’s (2015) experiment, which found changes in behavior, focused on the individual whereas Zenker, Vollmer, and Wagener’s (2018) experiment, which found no change in behavior, focused on the population average.
Using the Beta distribution to specify the gambler’s prior allows for an intuitive interpretation. Suppose, before starting to gamble, the gambler watches some matches to test how many he can predict correctly. His prior belief of $p$ has the distribution $\text{Beta}(\alpha, \beta)$ if the gambler predicts the outcome of $\alpha - 1$ matches correctly and $\beta - 1$ matches incorrectly.

The expectation of the Beta distribution is $E[p] = \frac{\alpha}{\alpha + \beta}$ and the variance is $\text{Var}[p] = \frac{\alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$. Since $\text{Var}(p)$ is decreasing in $\alpha + \beta$, the more matches the gambler watches, the more precise his prior belief of $p$ becomes.

The gambler is risk neutral and chooses whether to bet and how much to bet as a function of $E[p]$, his expected ability. The gambler only bets if $E[p] \geq c$, where $c$ is a cutoff probability. For $E[p] \geq c$, the gamblers bets an amount given by the function $f(E[p])$, where $f'(E[p]) \geq 0$.

Once the gambler starts betting, he uses Bayes Rule to update his belief on $p$. Let $s$ be the number of successful bets in $m$ matches, $s \sim \text{Binomial}(m, p)$. Since the Beta distribution is a conjugate of the Binomial distribution, the posterior belief also has a Beta distribution.

$$\Pr[p|s] \sim \text{Beta}(s + \alpha, m - s + \beta)$$

$$E[p|s] = \frac{s + \alpha}{m + \alpha + \beta}$$

This relationship between the posterior and prior beliefs has an intuitive interpretation. The prior simply adds $\alpha - 1$ successful bets and $\beta - 1$ unsuccessful bets to the gambler’s record of $s$ successful bets and $m - s$ unsuccessful bets.

The gambler’s prior belief will affect how quickly he learns $p$ up to a certain level of precision. For example, if the gambler has the uninformative prior of $\alpha = \beta = 1$, his prior is given by the uniform distribution and he believes that any value of $p$ is equally likely. As $\alpha + \beta$ increases, he has more precise beliefs about his ability $p$. Provided the outcome of the bets match the gambler’s prior, his posterior belief will be more precise if his prior was more precise.

The outcomes of the bets will also affect how quickly the gambler learns $p$. For example, assume the gambler has ability $p = \frac{1}{2}$. If he has lucky streaky of many successful bets, his posterior belief $\Pr[p|s]$ will be pulled away from $p$. The gambler will have to bet on many matches for $\Pr[p|s]$ to converge to $p$.

Even if the gambler’s ability $p$ is below his cutoff $c$, he may gamble because he is
overconfident in his ability. We can define overconfidence as:

\[ r \equiv \mathbb{E}[p|s] - p. \]

What can cause overconfidence? As mentioned above, a lucky streak could cause \( \mathbb{E}[p|s] \) to diverge from \( p \), but this divergence will likely be short-lived. How might the overconfidence persist? We consider two channels that may cause persistent overconfidence—a biased prior belief and biased updating.

2.1 Biased prior generates overconfidence

The gambler could form a biased prior for \( p \) if, instead of forming his belief using \( \alpha \) and \( \beta \), he uses \( \alpha' \) and \( \beta' \) with

\[ \frac{\alpha'}{\beta'} > \frac{\alpha}{\beta}. \]

The biased parameters \( \alpha' \) and \( \beta' \) may arise for several reasons.

The potential gambler may have biased recall on the “simulated bets” he uses to form his prior. If he does not systematically record his predictions of matches, he may think he predicted more matches correctly than he actually did. Without committing to a prediction ex-ante, the gambler’s memory of his prediction could be swayed by the outcome.

2.2 Biased updating generates overconfidence

Once a potential gambler starts gambling, he can learn his ability \( p \) by observing his betting outcomes. If he uses Bayesian updating correctly, he will eventually learn his ability—even if his prior is biased. However, the updating process could also be a source of persistent overconfidence.

Evidence from psychology experiments suggest people can be biased in their interpretation of past success relative to past failure. People attribute success to their own skill and explain away failure, even when the task has no element of skill—such as flipping a coin (Langer and Roth, 1975). A series of experiments focusing on sports betting showed the gamblers tend to explain away incorrect predictions as a near miss but accept correct predictions as an expected outcome (Gilovich, 1983).
Research in behavioral finance has used the insights from psychology to model biased updating (Daniel, Hirshleifer, and Subrahmanyam, 1998; Gervais and Odean, 2001). We take a slightly different approach, which is closer in spirit to the study on sports betting by Gilovich (1983). When the gambler makes an unsuccessful prediction, he counts a fraction $\gamma$ of his failures as success, where $0 \leq \gamma \leq 1$. For $\gamma = 0$, the gambler processes information in a rational manner. For $\gamma > 0$, he is biased, with the bias increasing in $\gamma$. At the extreme, when $\gamma = 1$, the gambler perceives all failure as success. The gambler’s posterior belief becomes:

$$
\Pr[p|s, \gamma] \sim \text{Beta}(s + \gamma(m - s) + \alpha, (1 - \gamma)(m - s) + \beta)
$$

$$
\mathbb{E}[p|s, \gamma] = \frac{s + \gamma(m - s) + \alpha}{m + \alpha + \beta}
$$

One implication of biased updating is the gambler with $\gamma > 0$ will be less responsive to extreme negative outcomes, such as a long sequence of unsuccessful predictions, than a rational gambler with $\gamma = 0$. We can test this prediction with our data on gambling transactions. Before describing the data and empirical results, we provide details on the context.

3 Context: Mobile money and sports betting

In 2007, the Kenyan telecom company Safaricom launched M-Pesa, a digital system which allows users to conduct basic financial transactions over the mobile phone network. Since 2007, mobile money has proliferated in the developing world and today there are more than half a billion registered mobile money accounts across 270 mobile money services in at least 90 countries (GSMA, 2016). While bank accounts are much more common than mobile money in most regions, this is not true in Africa. Even by 2014, mobile money ownership exceeded bank account ownership in many African countries (and this gap has surely grown by now).

The introduction of mobile money has been associated with important welfare effects. In Kenya, mobile money has been linked to improved risk-coping (Jack and Suri, 2014) and is estimated to have lifted as many as 194 000 Kenyans out of poverty (Suri and Jack, 2016). Businesses also benefit from mobile phone adoption as payments for goods and services can be collected with mobile money. Consequently, many in the policy and aid communities view mobile financial services as the future
to improving financial access in poor countries (GSMA, 2016; Lauer and Lyman, 2015).

Yet one aspect of this “digital finance revolution” that has received little scrutiny is that of mobile phone-based gambling. Gambling operators have leveraged the M-Pesa mobile money network to collect bets on soccer matches and other sports. Since the transaction cost of collecting bets and disbursing winnings has been lowered by mobile phone technology, the companies offer bets for as little as 1 KSH (0.01 USD), an amount most Kenyans can afford.

Many Kenyans have strong interest in sport, especially football. Gambling operators have tapped into this interest and sports betting has been widely adopted in a short period of time. In just five years of operation, SportPesa, the leading sports betting operator, has gathered enough revenue to sponsor two English premiership football teams, deals estimated at 4 and 12 million dollars per year.

To emphasize the scale of sports betting, we use data on search queries on Google using the Google Trends tool. In Figure 1 we plot the relative popularity of SportPesa against Facebook as measured by Google search queries. Facebook serves as a good benchmark as it is the most popular search query worldwide. Users of online services typically search for the name of service rather than type out the web address so the index of search queries serves as a good proxy for the relative number of users. For example, a user may type in “facebook” in the search bar rather than typing out “www.facebook.com”. Figure 1 shows a clear pattern. Since 2014, sports betting has grown rapidly in popularity.

---

8Visit this link to find the updated graph on Google Trends.
4 Data

We use a dataset of mobile money and betting transactions sourced from a financial services provider. The provider collects this data to make lending decisions and shared an anonymized sample with us. The sample includes users who registered with the financial services provider between April and June 2016 and in July 2017.

For a typical user, we have a log of mobile money transactions and betting transactions. Importantly, the mobile money transactions show the amount and the name of the company the money is transferred to. Since sports betting in Kenya operates almost exclusively via mobile money, we can track when a user tops up or withdraws from their betting account. We also observe transaction confirmations sent from the betting company so we can track the results of certain bets.

Table 1 provides descriptive statistics. Our sample consists of mostly young men. Among the subsample of gamblers, 78.2 percent are men and 75 percent are 35 years of age or younger. Gamblers spend 525.24 KSH (approximately 5 USD) and receive 386.06 KSH (approximately 4 USD) income from gambling on average per week.

The means mask extremely fat tails in the distribution of average weekly betting expenditure and income so we also report the 25th and 75th percentile in brackets. The mean weekly expenditure is larger than the 75th percentile because a small group of gamblers spends a very large amount per week. The top 20 largest spenders have an average expenditure of KSH 84 349.41 (approximately 843 USD).
Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Gamblers</th>
<th>Non-Gamblers</th>
<th>Full sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of individuals</td>
<td>54 771</td>
<td>33 863</td>
<td>88 634</td>
</tr>
<tr>
<td>Male</td>
<td>78.2 %</td>
<td>50.5 %</td>
<td>67.7 %</td>
</tr>
<tr>
<td>Age</td>
<td>30.66</td>
<td>31.92</td>
<td>31.13</td>
</tr>
<tr>
<td>[25, 34]</td>
<td>[25, 36]</td>
<td>[25, 35]</td>
<td></td>
</tr>
<tr>
<td>Betting expenditure (weekly)</td>
<td>525.24</td>
<td>0</td>
<td>324.57</td>
</tr>
<tr>
<td>[23.52, 350.00]</td>
<td>[0, 0]</td>
<td>[0, 156.88]</td>
<td></td>
</tr>
<tr>
<td>Betting income (weekly)</td>
<td>386.06</td>
<td>0</td>
<td>238.56</td>
</tr>
<tr>
<td>[0, 192.62]</td>
<td>[0, 0]</td>
<td>[0, 50.00]</td>
<td></td>
</tr>
</tbody>
</table>

Notes: We report means and the 25th and 75th percentile in brackets. We define gamblers as individuals with at least one mobile money transfer to or from a betting company. We exclude users with less than three weeks of mobile money transactions from the sample.

has fat tails because the largest prizes are awarded for bets with extreme odds. Also, betting income only measures withdrawals from the betting account so small wins which are used to fund subsequent bets will not appear in this table.

We can divide the 54 771 gamblers according to the ratio of their betting income to their betting expenditure. We define three groups: gamblers who have taken zero withdrawals from their betting accounts, gamblers who have withdrawn less than they have spent and gamblers who have withdrawn more they have spent. We show the split between these groups in Figure 2 as a function of the number of weeks of observed gambling activity. (We observe a variable number of weeks because we have an unbalanced panel of mobile money transactions.) The percentage of gamblers who have made a profit from gambling ranges from between 10 and 20 percent depending on the number of weeks of observed gambling.
The proportion of individuals who have withdrawn zero, less than they spent or more than they spent from their betting accounts.

4.1 Jackpots as a standardized measure of gambling ability

One challenge to compare gamblers is to define a standardized measure of betting ability. Sports betting companies offer a range of different bets. For example, users can bet on single matches or they can chain matches together to increase the payout (and decrease the probability of winning). Also, individual matches differ in odds. It is easier to predict the outcome of a match with a clear favorite than to predict the outcome of a match with similar strength teams. We solve this challenge by studying the number of correct predictions on the weekly jackpots offered by SportPesa, Kenya’s leading sports betting operator.

Each week, SportPesa selects 13 matches for a midweek jackpot and 17 matches for a weekend jackpot (called the MegaJackpot). The jackpot is awarded if the outcome of all the pre-selected matches are predicted correctly. Bonuses are awarded for 10 or more correct predictions on the midweek jackpot and 12 or more correct predictions on the weekend jackpot. The amount of the bonus increases exponentially with the number of correct predictions. For example, the bonuses for the weekend jackpot on 31 March 2019 were approximately 460, 2190, 8140 and 64300 USD for 12, 13, 14 and 15 correct predictions.

The jackpots provide a standardized measure of betting ability, both between
gamblers and across time. Since SportPesa selects the matches each week we need not worry about how the gambler picks the matches. The types of matches SportPesa select each week are also very similar. SportPesa has an incentive to select matches with no clear favorite, where the odds are balanced across the three possible outcomes of home team wins, away team wins or a draw. The more difficult to predict the outcome of a single match in the jackpot, the less likely a gambler will predict all the matches correctly.

The jackpots are also ideal because they are extremely popular. In our data, we observe midweek or weekend jackpot bets placed by 17 776 individuals, which represents 30.7 percent of all individuals for whom we observe some type of betting transaction.

5 Empirical results

We provide empirical results to distinguish between rational and behavioral sports betting behavior. First, we provide evidence of differences in gambling ability by showing that certain gamblers consistently predict more matches correctly than others. Second, we show that gamblers react to signals of success but not to signals of failure, which provides support for an overconfidence bias and against rational betting behavior.

5.1 Heterogeneity in gambling ability

Our model of gambling allows for heterogeneity in gambling ability. High ability gamblers have a greater probability of predicting the outcome of a sports match than low ability gamblers. To support this assumption, we show that certain gamblers consistently predict more matches correctly than others.

There are two challenges to detect differences in gambling ability. First, we should only compare gamblers making predictions on the same or similar matches. Second, we need to separate ability from luck. Random guessing will generate variance in the number of correct predictions and this variance should not be mistaken for heterogeneity in ability.

To ensure we only compare gamblers betting on the same or similar matches, we focus on the matches selected by SportPesa for the jackpot. As detailed in Section
4.1, the types of matches selected for the jackpot are very similar, both within and across weeks, so these matches provide a standardized format to compare gamblers.

To separate ability from luck, we study gamblers’ bets across time. If all gamblers are identical and predict the outcome of matches with some fixed probability, there will be no correlation in the number of correct predictions across time. If a gambler who predicts a high number of matches correctly this week is more likely to predict a high number of matches correctly next week, this suggests heterogeneity in gambling ability.

To test for correlation in the number of correct predictions across time, we use the following specification:

\[
Correct\ predictions_{it} = \sum_{l=1}^{L} \beta_l Correct\ predictions_{i(t-l)} + \theta_t + \epsilon_{it}
\]

The model checks if the number of correct prediction in week \( t \) by gambler \( i \) is positively correlated with the number of correct predictions in previous weeks up to a lag of \( L \) weeks. We include week fixed effects, \( \theta_t \), so the estimates use only within week variation to compare gamblers.

We report the results in Table 2. In all specifications, the number of correct predictions on jackpot matches is positively correlated with the number of correct predictions in previous weeks. If all gamblers had identical chances of success, this correlation would not be present. The positive correlation can only be explained by heterogeneity in the chance of success between gamblers, which we interpret in our model as heterogeneity in gambling ability.
Table 2: Testing for heterogeneity in betting ability

<table>
<thead>
<tr>
<th>Correct predictions in week:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t - 1 )</td>
<td>0.0125** (0.0051)</td>
<td>0.0164*** (0.0057)</td>
<td>0.0185*** (0.0062)</td>
<td>0.0234*** (0.0067)</td>
</tr>
<tr>
<td>( t - 2 )</td>
<td>0.0161*** (0.0056)</td>
<td>0.0173*** (0.0062)</td>
<td>0.0169** (0.0067)</td>
<td></td>
</tr>
<tr>
<td>( t - 3 )</td>
<td>0.0151** (0.0060)</td>
<td>0.0203*** (0.0065)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( t - 4 )</td>
<td>0.0128** (0.0062)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Individuals: 6,255, 4,345, 3,326, 2,649
Weeks: 82, 81, 80, 79
Individual fixed effects: No, No, No, No
Week fixed effects: Yes, Yes, Yes, Yes

Notes: Dependent variable is number of correct predictions on the weekend jackpot in week \( t \). Standard errors are robust to heteroskedasticity. P-value <0.10 *, <0.05 **, <0.01 ***

5.2 Bias in the learning process

Our model allows for gamblers to learn rationally by applying Bayes rule or to learn with a bias by treating a proportion of failure as success. In support of the biased learning model, we show gamblers have stronger reactions to success than to failure relative to average outcomes.

To compare how a gambler responds to success and failure, we need to separate the positive signal of success from the income effect of a winning bet. SportPesa’s midweek and weekend jackpots provides the ideal scenario to separate these effects. The midweek jackpot awards prizes for 10 or more correct predictions out of 13 matches and the weekend jackpots awards prizes for 12 or more correct predictions out of 17 matches. Any variation in the number of correct predictions below 10 for the midweek jackpot and below 12 for the weekend jackpot has no income effect. We can isolate the effect of signals by focusing on this range of the jackpot results.

The probability of winning a prize for the jackpots is very low, so focusing on the range of results below the cutoff for prizes still provides us with ample variation in
signals of ability. If a gambler selected the favorite (the team most likely to win) for every match in the jackpot, he would have a less than one percent chance of winning a prize. In our sample, 0.68 percent of midweek jackpot bets and 0.94 percent of weekend jackpot bets won a prize.

We study gamblers’ betting behavior in response to the number of correct predictions on the jackpot in the previous week. We use the following specification

\[
Betting_{it} = \mu \text{Jackpot}_{i(t-1)} + \\
\beta_1 \text{Count of jackpot bets}_{i(t-1)} + \beta_2 \text{Won other bet}_{i(t-1)} + \\
v_i + \theta_t + \epsilon_{it}
\]

where \(v_i\) and \(\theta_t\) are individual and week fixed effects. We divide the jackpot result into four categories: low, base, high and prize. The prize category is for results above the prize cutoff to separate the income effects. We use the low and high categories as signals of failure and success.

We define the low and high categories to occur with low probability so that they provide strong signals of ability. For the midweek jackpot we set low as 0 or 1 correct predictions (1.5 percent of bets) and high as 8 or 9 correct predictions (9 percent of bets) out of 13 matches. For the weekend jackpot we set low as 0 or 1 correct predictions (0.4 percent of bets) and high as 10 or 11 correct predictions (7.34 percent of bets). Although both occur with low probability, the low outcome occurs much less frequently than the high outcome and so should provide a much stronger signal of ability.

We report the results in Table 3. We use two measures of betting behavior as dependent variables. First, we use an indicator for placing a jackpot bet the following week. Second, we use the top up amount in the following week, which we measure as \(\log(1 + x)\) where \(x\) is the sum of all mobile money transfers to betting companies in that week.

A signal of success on the jackpot does impact future betting behavior. A gambler who predicts the outcome correctly of a high number of matches on the midweek jackpot is 6.32 percentage points more likely to play the jackpot the following week and spends approximately 14 percent more on topping up his betting account in comparison to weeks in which he predicts a moderate number of matches correctly. Results are similar for the weekend jackpot. A gambler is 3.84 percentage points more
Table 3: Responses to the previous week’s jackpot

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Midweek jackpot</th>
<th>Weekend jackpot</th>
<th>Topup amount</th>
<th>Topup amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jackpot results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.0041</td>
<td>-0.0198</td>
<td>-0.2219</td>
<td>0.2311</td>
</tr>
<tr>
<td></td>
<td>(0.0287)</td>
<td>(0.0486)</td>
<td>(0.1350)</td>
<td>(0.2137)</td>
</tr>
<tr>
<td>High</td>
<td>0.0632***</td>
<td>0.0384***</td>
<td>0.1490***</td>
<td>0.0777*</td>
</tr>
<tr>
<td></td>
<td>(0.0117)</td>
<td>(0.0103)</td>
<td>(0.0484)</td>
<td>(0.0462)</td>
</tr>
<tr>
<td>Prize</td>
<td>0.1250***</td>
<td>0.1775***</td>
<td>0.0243</td>
<td>-0.3479**</td>
</tr>
<tr>
<td></td>
<td>(0.0375)</td>
<td>(0.0232)</td>
<td>(0.1760)</td>
<td>(0.1450)</td>
</tr>
<tr>
<td>Type</td>
<td>Midweek</td>
<td>Weekend</td>
<td>Midweek</td>
<td>Weekend</td>
</tr>
<tr>
<td>Observations</td>
<td>25 314</td>
<td>36 822</td>
<td>23 361</td>
<td>33 936</td>
</tr>
<tr>
<td>Individuals</td>
<td>4 621</td>
<td>6 101</td>
<td>4 436</td>
<td>5 898</td>
</tr>
<tr>
<td>Weeks</td>
<td>92</td>
<td>82</td>
<td>92</td>
<td>82</td>
</tr>
<tr>
<td>Individual fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Week fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: The dependent variables for ‘Midweek jackpot’ and ‘Weekend jackpot’ are indicators for whether the individual placed a jackpot bet in week t. ‘Topup amount’ is log(1+betting expenditure) where betting expenditure is the total mobile money transfers to gambling companies in the week. We control for the sum of jackpot bets placed and an indicator for winning another type of bet in the same week. If the gambler places multiple jackpot bets, we use the maximum number of correct predictions on a single bet to measure the number of correct predictions. Standard errors are clustered at the individual level. P-value <0.10 *, <0.05 **, <0.01 ***s

likely to play the weekend jackpot the following week and spends approximately 7 percent more on topping up his betting account in weeks in which he predicts a high number of matches correctly in comparison to weeks in which he predicts a moderate number of matches correctly.

In contrast, a signal of failure on the jackpot does not impact future betting behavior. A gambler who predicts one or zero of the matches correctly does not behave differently relative to base category of 2-7 correct predictions on the midweek jackpot and 2-9 correct predictions on the weekend jackpot. These results suggest an asymmetry in learning. Gamblers react strongly to past success but are less responsive to past failure.

To illustrate our results, we compare the real gamblers to simulated gamblers.
who learn their ability using Bayesian updating. We draw a population of 5,000
gamblers with heterogeneous ability \( p \in \{0.1, 0.2, 0.3, 0.4, 0.5\} \) to predict the outcome of a
match correctly. For 10 weeks, each simulated gambler \( i \) bets if his expected ability
given his past betting results, \( E[p_i|s_i] \), is greater than his week-specific threshold, \( c_t \).
The cost \( c_t \) is normally distributed with mean \( p_i \) and a 0.1 standard deviation. All
simulated gamblers start with a uniform prior and only bet on one type of bet, the
weekend jackpot of 17 matches. The simulated gamblers learn their ability rationally,
as in our baseline model with \( \gamma = 0 \).

For the real gamblers, we combine the weekend and midweek jackpot data to
generate more statistical power. For both the simulated and real data, we run the
following regression

\[
Bet_{it} = \mu \text{Count of correct predictions}_{i(t-1)} + v_i + \epsilon_{it}
\]

where \( v_i \) are individual fixed effects and \( \text{Count of correct predictions} \) is a set of indicator
variables for the number of correct predictions the gambler made the previous week. The reference category is the modal outcome.

We summarize the regression results from the real and simulated data in Figure
3. Each point on the graph shows the regression coefficient and the shaded area
shows the 95% confidence interval. In the bottom panel, the simulated gamblers
react symmetrically to failure and success. The simulated gambler is more likely to
bet following a high number of correct predictions and less likely to bet following a
low number of correct predictions. As is clear from the top panel of the figure, we do
not find symmetry in the real data. The real gamblers are more responsive to success
than to failure.
(a) Real data

(b) Simulation with Bayesian updating

Figure 3: Comparison between the model and the empirical results
6 Welfare impact: Betting and loan defaults

Given the rapid growth of sports betting in East Africa, governments and policymakers have turned their attention to the sports betting industry. In Kenya, the government has introduced taxes on revenue and winnings and the gambling regulator has introduced bans on outdoor advertising and celebrity endorsements. In Tanzania, religious leaders have pushed for sports betting to be banned altogether. An outright ban has started to take force in Uganda. The Ugandan government has suspended all new gambling licenses and pledged not to renew existing licenses.

Despite sports betting becoming a central policy concern, little is known about the welfare impacts. Our data allows us to shed light on one aspect of welfare—the ability to repay loans. The rise of sports betting has occurred concurrently with the rise in consumer loans disbursed using mobile phones. A recent article in The Economist noted high default rates and suggested sports betting may be a cause:

Thanks to M-Pesa, its largest mobile-money service, with over 20m users, Kenya has been a pioneer in both mobile money and mobile financial services, such as lending. Anecdotal evidence is mounting of abuses—most notoriously of young Kenyans borrowing to splurge on online betting sites.

Our data includes both sports betting and loan transactions, which allows us to assess if sports betting is a cause of the high rates of loan default. In the following sections, we investigate how sports betting impacts loan repayment. Persistent gamblers, those who bet in more than 90 percent of the observed weeks of mobile money transactions, have higher default rates than people who gamble less frequently. An instrumental variables approach provides suggestive evidence that sports betting causes more defaults.

6.1 Comparing gamblers and non-gamblers

We provide descriptive statistics of loan repayment in Table 4. For each individual, we observe a certain number of weeks of mobile money transactions. We separate the sample according to the share of the observed weeks in which the individual made a mobile money transfer to a sports betting company. We have four categories: zero (non-gamblers), 1-50 percent (occasional), 51-90 percent (frequent) and 91-100 percent (persistent).
Table 4: Loan repayments by frequency of gambling activity

<table>
<thead>
<tr>
<th>Share of weeks gambling</th>
<th>Non-gamblers</th>
<th>Occasional 1 - 50 %</th>
<th>Frequent 51 - 90 %</th>
<th>Persistent 91 - 100 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals</td>
<td>34,292</td>
<td>27,924</td>
<td>16,385</td>
<td>10,033</td>
</tr>
<tr>
<td>Granted loans</td>
<td>54.52%</td>
<td>63.67%</td>
<td>59.99%</td>
<td>48.64%</td>
</tr>
</tbody>
</table>

Among the subsample who have taken one or more loans:

<table>
<thead>
<tr>
<th></th>
<th>Default rate</th>
<th>Count of loans repaid</th>
<th>Value of loans repaid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.57%</td>
<td>4.37%</td>
<td>5.55%</td>
</tr>
<tr>
<td></td>
<td>4.37%</td>
<td>5.43%</td>
<td>5.19%</td>
</tr>
<tr>
<td></td>
<td>5.55%</td>
<td>4.46%</td>
<td>8539.73</td>
</tr>
<tr>
<td></td>
<td>6.56%</td>
<td></td>
<td>11028.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9244.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8212.27</td>
</tr>
</tbody>
</table>

Notes: For each individual, we observe a certain number of weeks of mobile money transactions. We divide the table according to the share of those weeks in which the individual made a transfer to a sports betting company. We exclude users with less than three weeks of mobile money transactions.

The loan data is sourced from the administrative records of the financial services provider who also provided the other data for this project. Thus, the loan data does not cover all loans the individuals may have used. Still, the data provides significant coverage with 57.75 percent of the sample having taken one or more loans with the provider. The remaining individuals in the sample either had loan applications rejected or did not complete their application.

Table 4 reveals several patterns. Occasional and frequent gamblers are more likely to have taken a loan than non-gamblers. Among the subsample who have taken loans, the frequent and non-gamblers have similar default rates. The default rate is calculated as the value of loans marked as defaulted as a share of the value of all loans. In contrast, the persistent gamblers, who displayed gambling activity in 91-100 percent of their weeks of observed mobile money transactions, diverge from the other groups. Persistent gamblers are less likely to be granted a loan and the default rate of 6.56 percent is higher than in the other groups.

People who choose to gamble every week and those who choose not to may differ in many dimensions. The difference in default rates cannot be attributed to gambling without more careful analysis. In the following section, we use an instrumental variables approach to test how an exogenous increase in betting activity impacts loan repayment.
6.2 Instrumental variables approach

We are interested in estimating the causal effect of betting increases on loan default. Since betting expenditure, in general, is not random, we use an instrumental variables strategy. For an instrument to be valid in this context, it must meet two conditions. First, it must be relevant—in this case, correlated with betting. Second, it must satisfy the exclusion restriction that it should be related to loan default only through the endogenous variable.

We use the number of correct jackpot predictions in week \( t - 1 \) as an instrument for betting expenditure in week \( t \). In Section 5.2, we demonstrated the relevance of this instrument by showing that an increase in the number of correct predictions on SportPesa’s jackpot increased betting expenditure in the following week. For the exclusion restriction, we assume that the number of correct predictions on the jackpot is random conditional on the gambler’s ability and the number of jackpot tickets purchased. We believe this is warranted given that it is difficult to imagine how loan behavior may be impacted by the previous week’s jackpot results other than through the current week’s betting behavior.

We therefore estimate the following model:

\[
\text{Loan behavior}_{it} = \beta \log(1 + \text{Betting expenditure}_{it}) + v_i + \epsilon_{it}
\]

where \( v_i \) are individual fixed effects and betting expenditures are instrumented by the number of correct jackpot predictions in the prior week. We use two measures of \( \text{Loan behavior}_{it} \). First, we use an indicator which is one if the individual fails to make a loan repayment in the week the repayment was due. Either the payment is paid late or not at all. We only use the subsample of observations when payments were due. Second, we use an indicator for loan default which is one in weeks when a repayment on the loan was due and that loan was eventually defaulted.

We report the results in Table 5. The first stage results repeat the regression specifications used in Table 3 but with a smaller subsample. This subsample only includes individuals who have taken at least one loan with the digital credit provider and have a repayment due in the week following a jackpot bet. The first stage results show that an individual who predicts 10 or 11 matches correctly spends approximately 14 percent more on betting the following week compared to when he predicts 9 or fewer matches correctly. Unfortunately, the F-statistic of 5.62 suggests the instrument
Table 5: Instrumental variables regression results

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Late repayment</th>
<th>Defaulted</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(1+ Betting expenditure)</td>
<td>0.1000</td>
<td>0.0539</td>
</tr>
<tr>
<td></td>
<td>(0.1247)</td>
<td>(0.0454)</td>
</tr>
</tbody>
</table>

First stage

| Weekend jackpot Close (10 - 11 correct) | 0.1460**       |
|                                         | (0.0680)       |

Observations 12,600
Individuals 3,231
Individual fixed effects Yes
F-statistic 5.62

Notes: Late repayment is an indicator for not making a loan repayment that was due in the week and Defaulted is an indicator for not making a loan repayment on a loan that was eventually categorized as defaulted by the financial services company. We control for the number of jackpot bets placed. Standard errors are clustered at the individual level. P-value <0.10 *, <0.05 **, <0.01 ***

is weak (Stock and Yogo, 2005).

The results suggest that increased betting expenditure increases the probability of late loan repayment and default. A one percent increase in betting expenditure causes a 10 percent increase in the likelihood of late payment and a 5.39 percent increase in the likelihood of default. However, the estimates are not precise. The 95 percent confidence interval for late repayment estimate is between -14.44 and 34.44 percent and for the default estimate is between -3.51 and 14.29 percent. These results should be treated with caution.

7 Conclusion

To separate the positive signal of a successful bet from the income effect of winning a prize, we used jackpot bets, which only award a prize above a cut-off number of successful predictions. This strategy provides clean comparison of the positive signal of predicting many matches correctly and the negative signal of predicting none or
just one of the matches correctly.

Given the strong relationship between the jackpot bets and the following weeks betting expenditure, we used this variable in an instrumental variables approach to investigate the causal impact of sports betting on loan repayments. Our estimates are too noisy to provide a clear answer, but the point estimates provide suggestive evidence that sports betting increases the likelihood of late repayment and loan default.

References

Abel, Martin, Shawn Cole, and Bilal Zia. 2015. “(De-)biasing on a roll: Changing gambling behavior through experiential learning.”, Working Paper.


