Sports betting: A source for empirical Bayes

Research practice 1: Final presentation

Mateo Graciano-Londoño
Mathematical Engineering
Student

Andrés Ramírez-Hassan
Department of Economics
Tutor

EAFIT University, Medellín
Colombia

November 24^{th}, 2015

Previous works

Sports forecasting is a topic which has been widely studied in the literature [Stekler et al., 2010], where betting odds have played a remarkable role in this aim. Some methodologies have been based on a frequentist approach [Leitner et al., 2010], whereas others on a Bayesian framework [Baio and Blangiardo, 2010].

Previous works

Spann and Skiera [2009] conclude on their work that betting odds are the best source for sport forecasting but we can see at Štrumbelj [2014] that there are different ways to obtain the associated probability from a betting odd and they conclude that the Shin [1993] methodology is the most accurate to calculate the probabilities among the different methods that had been used before.

Obtaining the probabilities

The idea is to use a Dirichlet categorized model, so we obtain the following posterior [Ramírez and Cardona 2014]:

$$\pi(\boldsymbol{p}|Data) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i + c_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i + c_i)} \prod_{i=1}^{k} p_i^{\alpha_i + c_i - 1}$$

where k=number of possible outcomes, c_i=number of times the event i had happened, p_i=probability for the event i and α is the vector of hyper-parameters of the prior.

Betting strategy

For the betting strategy we will proceed since we already have a good estimation of the probability ($\boldsymbol{p}$) of each outcome from the posterior. So now, given the odds (θ_i) for each outcome, how are we suppose to bet?

- Bet for the most likely event.
- Bet if $\theta_i p_i > 1$.
- Bet for the greatest $\theta_i p_i$ for each bookmarker.

Betting strategy

It is easy to show that the first (which is the most common idea), is statistically the worst, but with the second and third one the gambler is supposed to win.

The question is how much to bet? For that matter is a good idea to use the Kelly [1956] criterion which is based on the following equation:

$$f^* = \frac{\theta p - 1}{\theta - 1} \tag{1}$$

From (1) we calculate which fraction of the bankroll we are supposed to bet.

Betting strategy

Consider only a fraction of the full Kelly criterion:

$$\lambda f^* = \frac{\theta p - 1}{\theta - 1}$$

where $\lambda \in [0, 1]$.

Betting strategy

To determine which λ is the more profitable we will use three different approaches:

- Stopping loss in a matchday
- Fixed expected win with simulation each matchday
- Maximize the expected win with simulation each matchday

Dirichlet distribution

- Random numbers generation from the Dirichlet distribution.
- Estimation of the Dirichlet distribution parameters with the moments method for the tennis case.
- Estimation of the Dirichlet distribution parameters with maximum likelihood estimation method for the football case.

Posterior and simulation exercises

Once we obtained the posterior distributions of the parameters, we simulated our model 500.000 times for each game, and then, we calculate the mean for our estimation of $\boldsymbol{p}$.

Forecast power on BPL

		Model			
Season	Measure	Posterior	Shin	Normalization	Naive
2013	MSE	0.6362	0.6325	0.6367	0.8164
-2014	MAE	1.0256	1.0180	1.0265	1.3333
2014	MSE	0.7170	0.7207	0.7055	0.8164
-2015	MAE	1.1514	1.1598	1.1297	1.3333

Table: Comparing different methodologies.

Simple betting

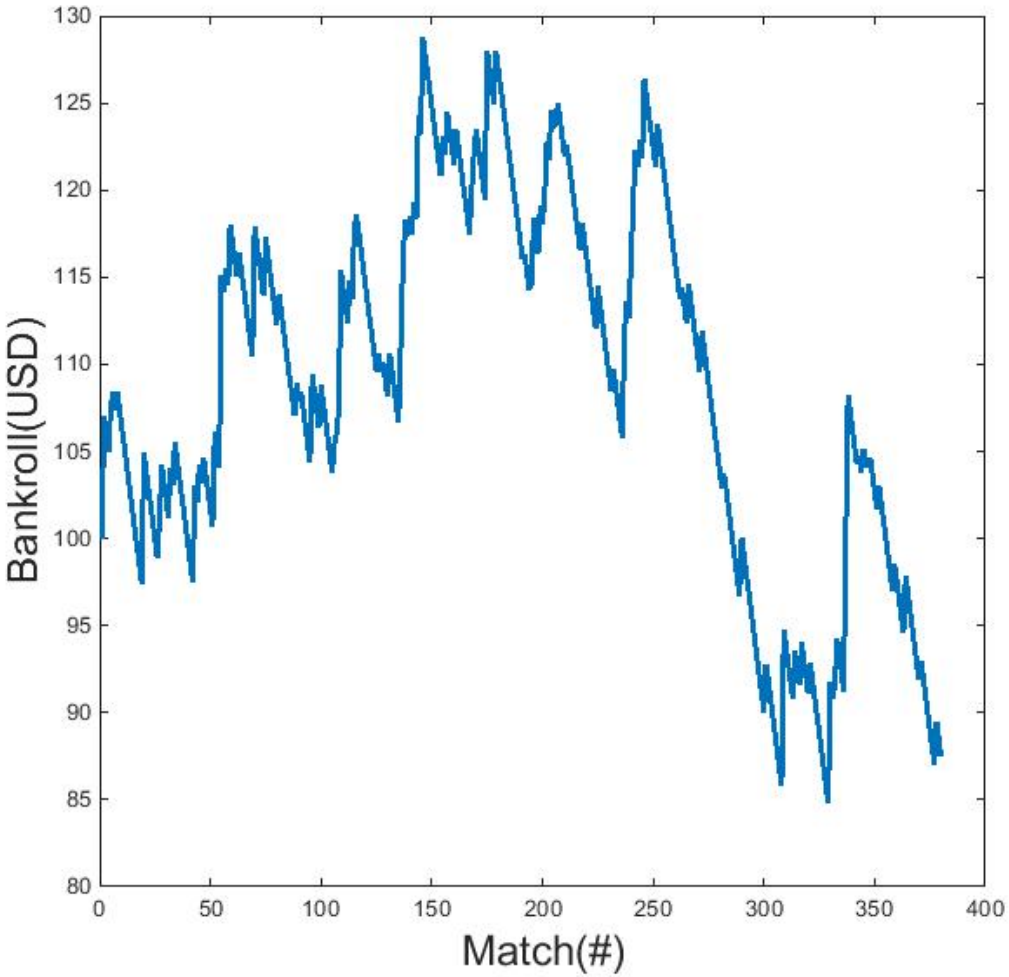

Figure: Betting for the greatest odd in the 13-14 season.

Simple betting

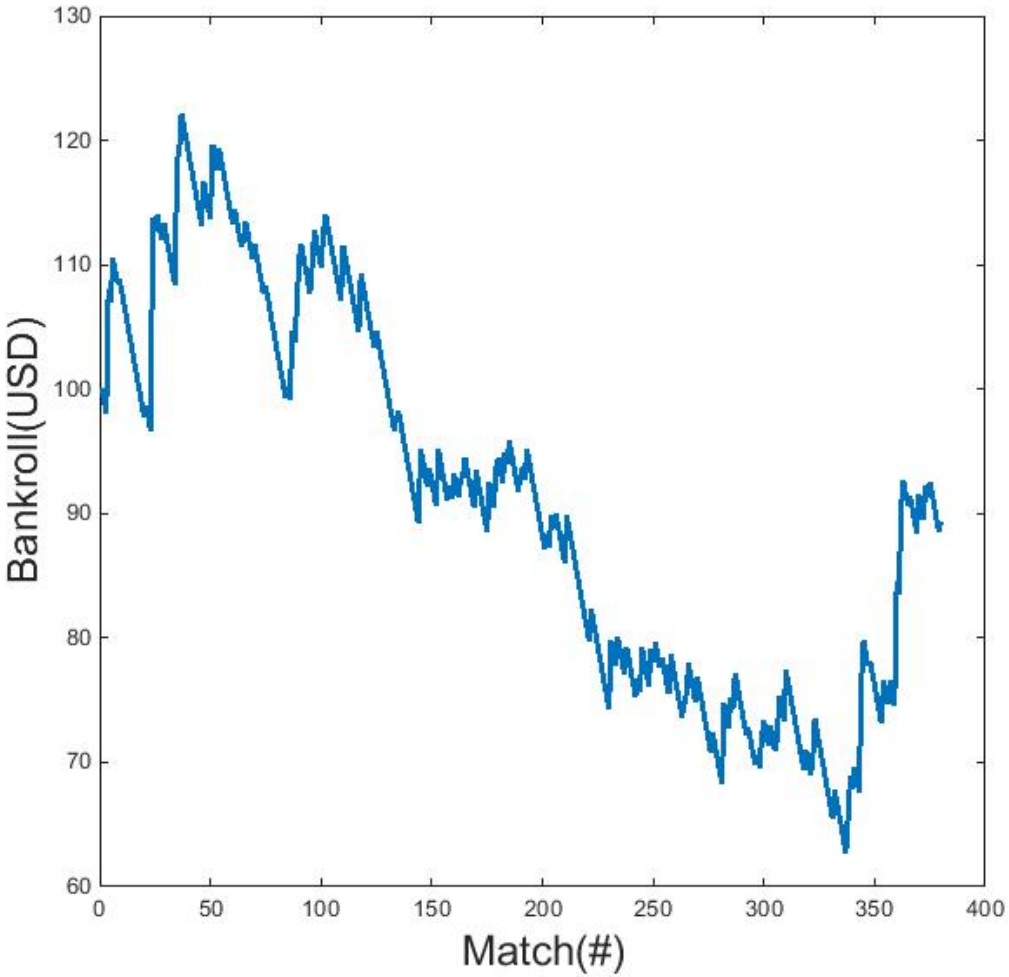

Figure: Betting for the greatest odd in the 14-15 season.

Simple betting

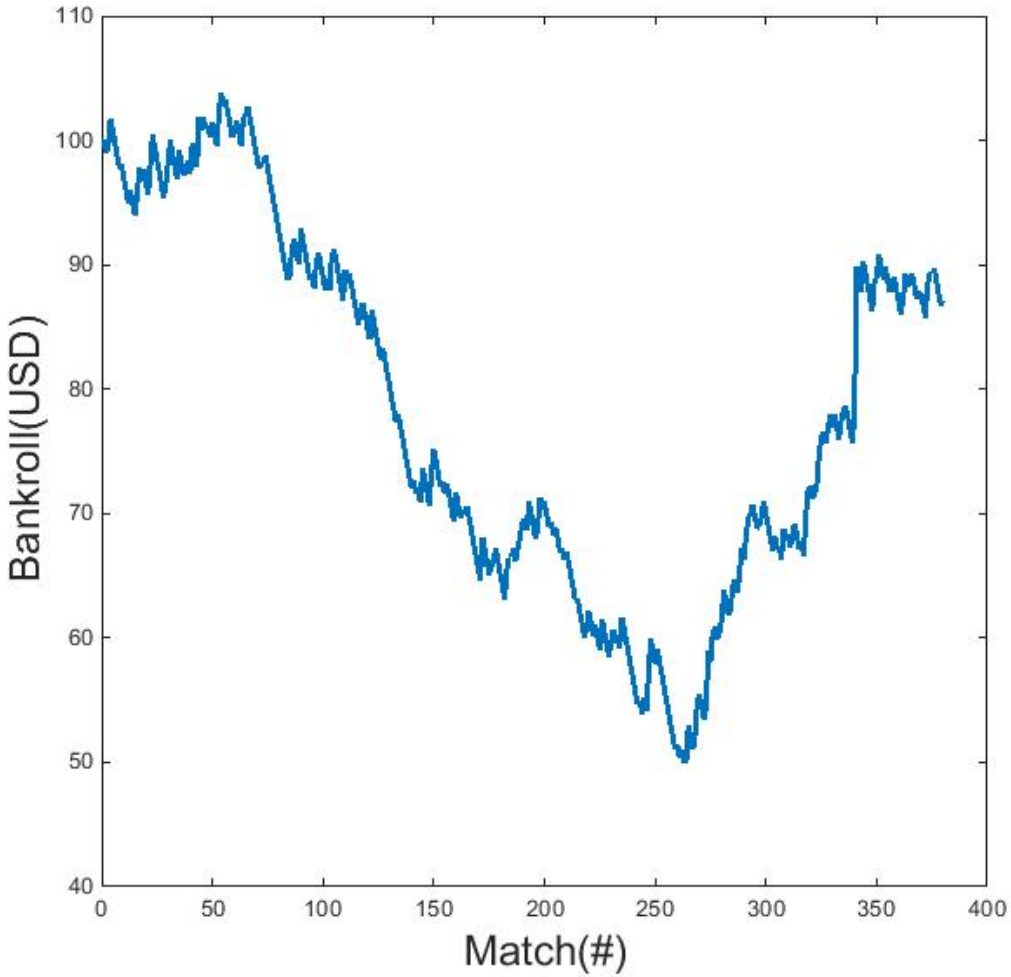

Figure: Betting for the lowest odd in the 13-14 season.

Simple betting

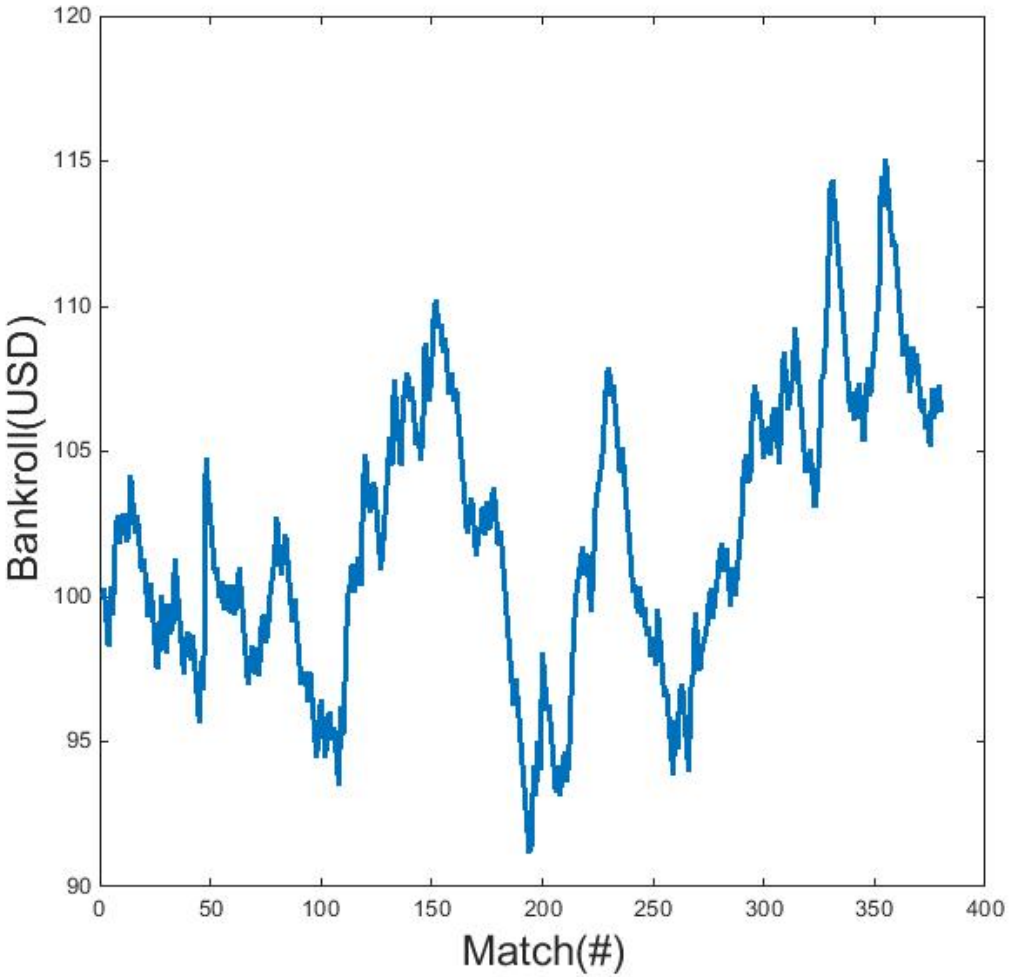

Figure: Betting for the lowest odd in the 14-15 season.

Better than Shin?

Model	Mean	Standard deviation	Skewness	Kurtosis
Posterior	124.9	17.4	0.17	1.85
Shin	98.2	3.9	0.52	2.54
Normalization	100.7	10.9	1.3	4.22

Table: Betting performance using Kelly criterion 13-14 season.

Better than Shin?

Model	Mean	Standart deviation	Skewness	Kurtosis
Posterior	103.2	17.5	0.31	2,09
Shin	92.97	2.31	1.62	5.59
Normalization	62.78	22.9	0.56	2.31

Table: Betting performance using Kelly criterion 14-15 season.

Checking performance

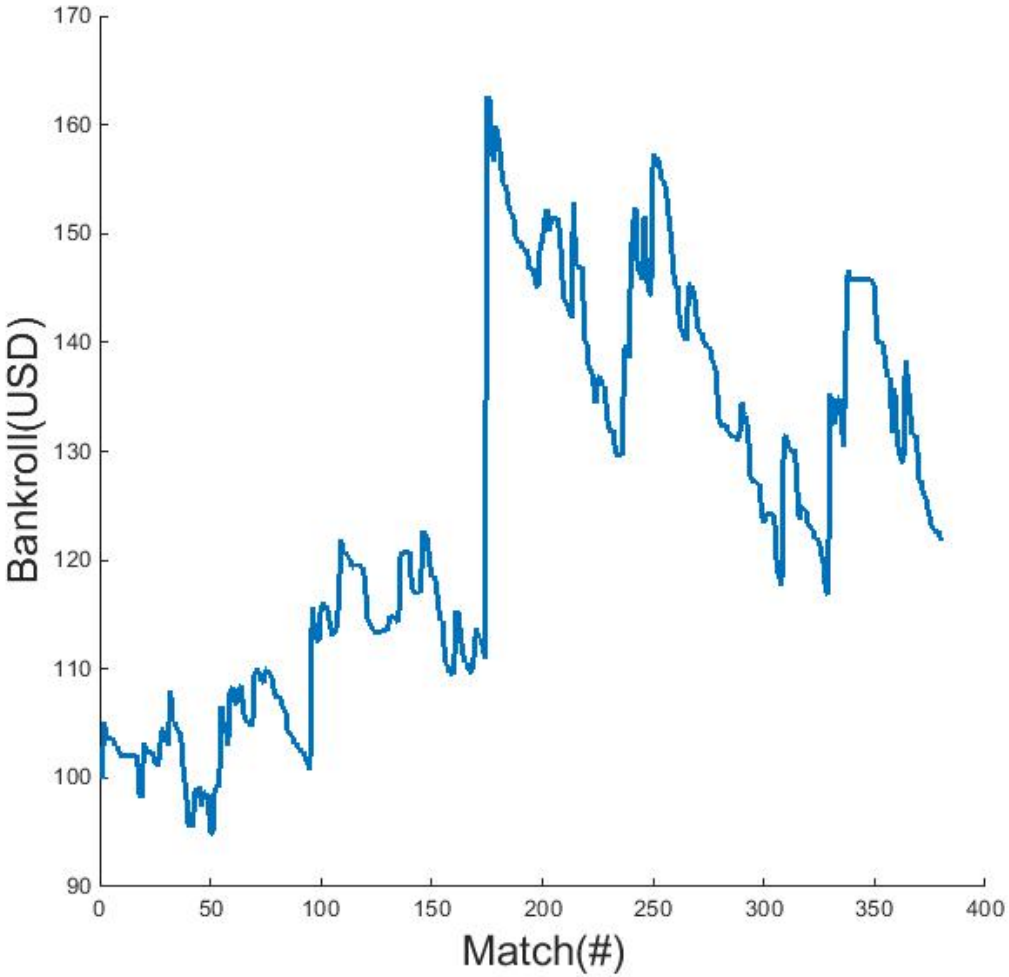

Figure: Betting with our process in the 13-14 season.

Checking performance (on BPL)

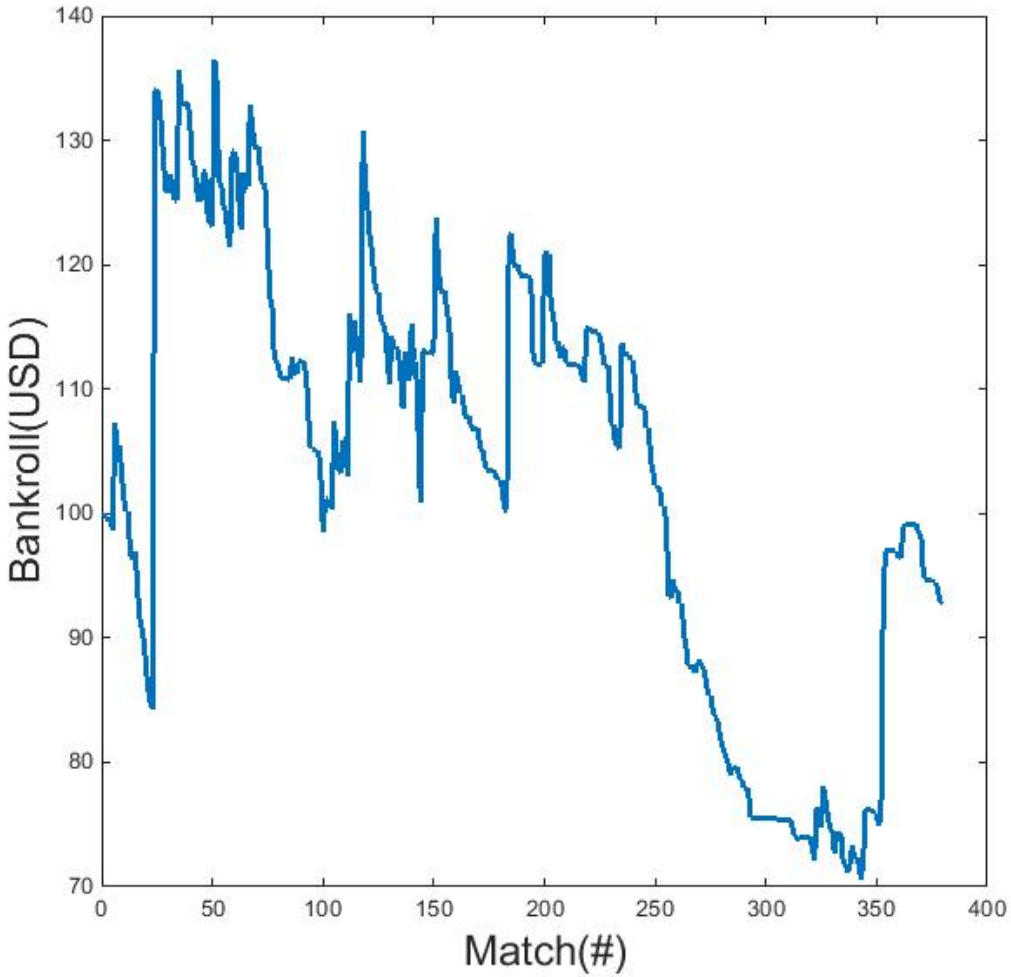

Figure: Betting with our process in the 14-15 season.

Checking performance

For both, the 13-14 and 14-15 season the results showed that they would had been growing in a good rate until the half season. [Forrest et al., 2005; Peter F. Pope, 1989] conclude that as the season goes the bookmarker's predictions (or the expert's) improves, so it is really plausible that as the season goes it is more difficult to win against market.

Improving performance

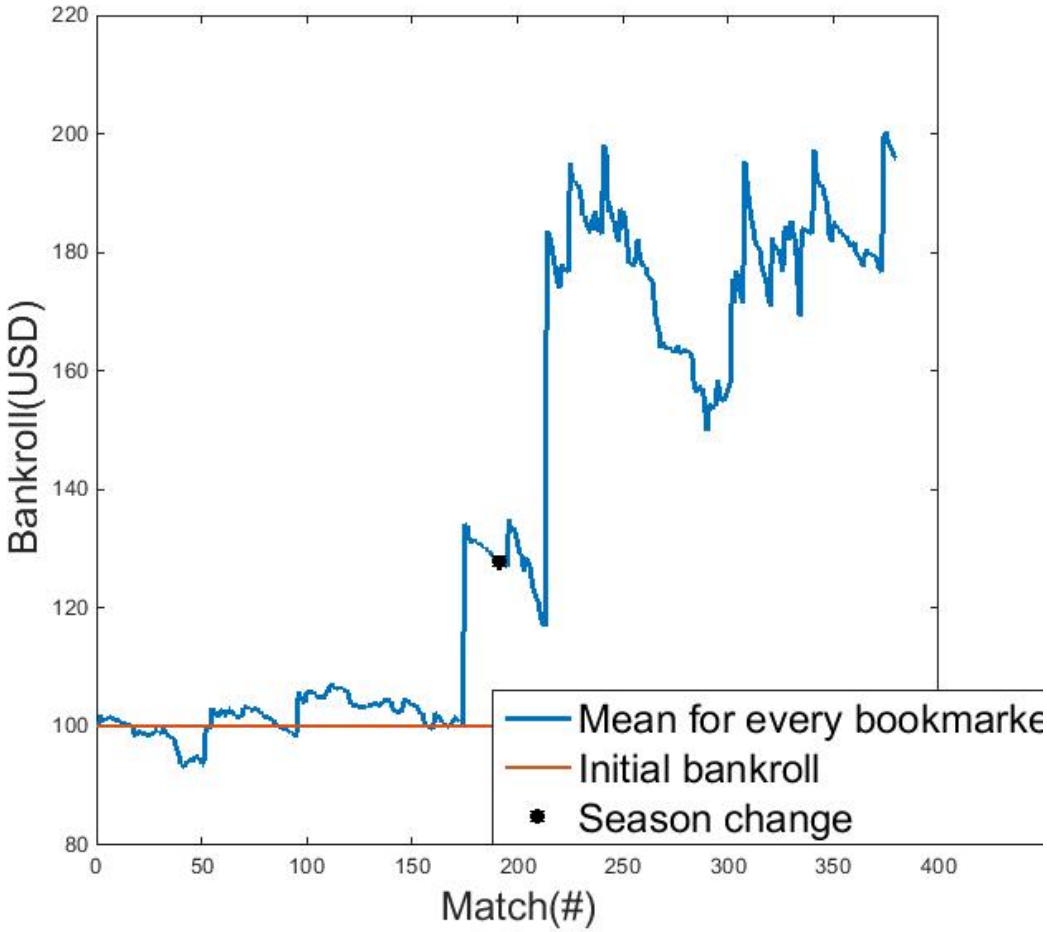

Figure: Betting until the half of both seasons.

λ-Kelly

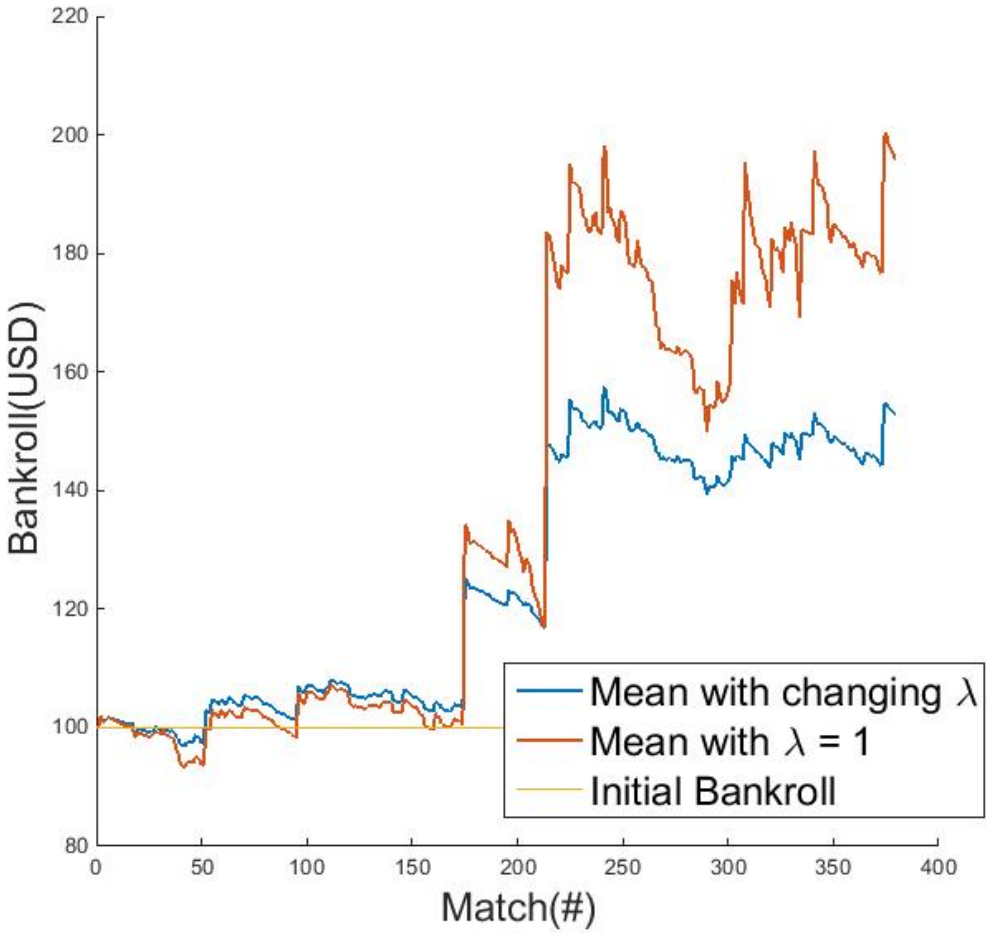

Figure: Stopping loss 5%.

λ-Kelly

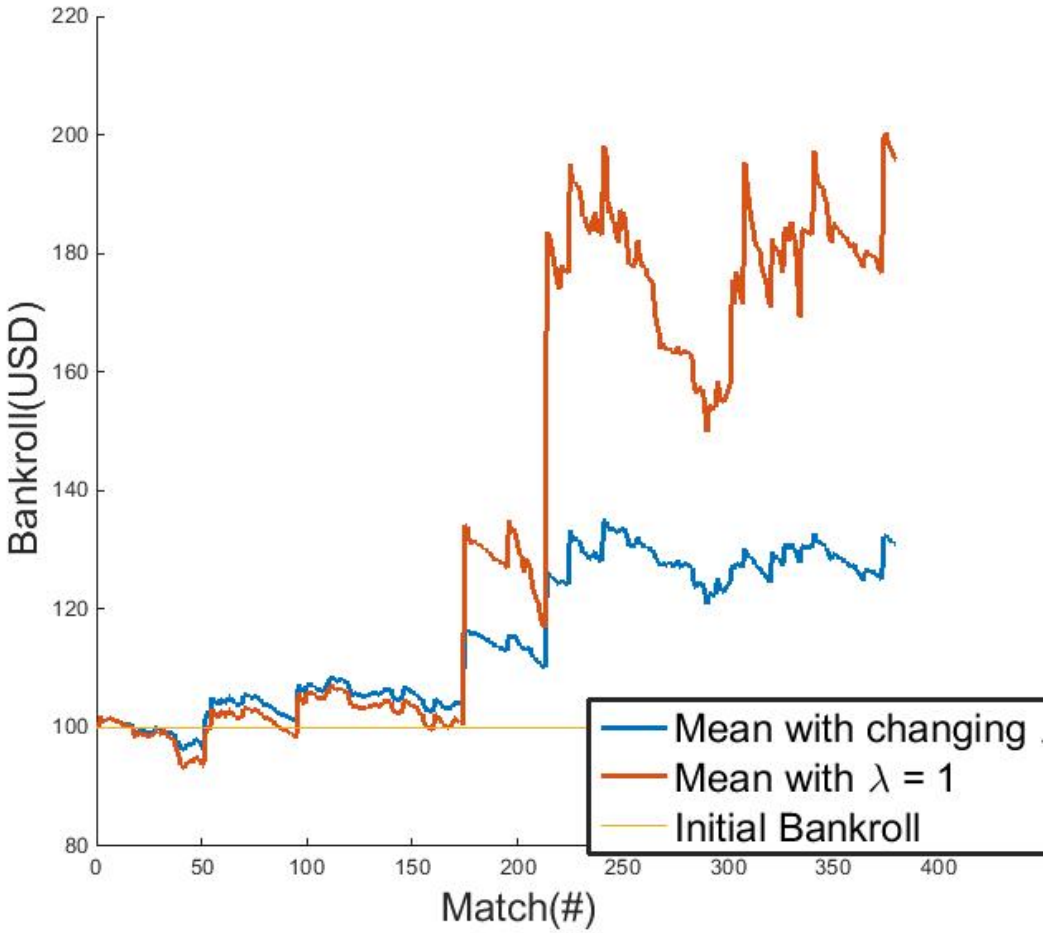

Figure: Expecting profit about 20%.

λ-Kelly

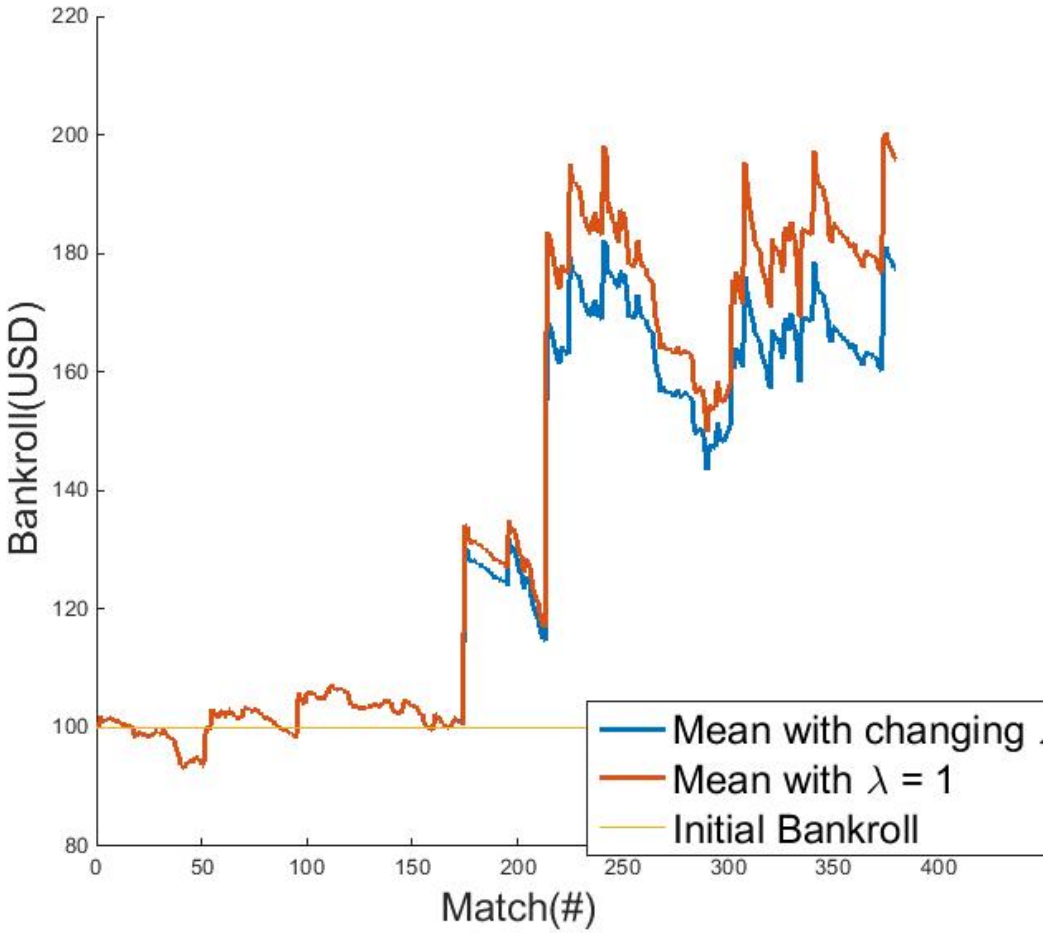

Figure: Expecting to double initial bankroll.

λ-Kelly

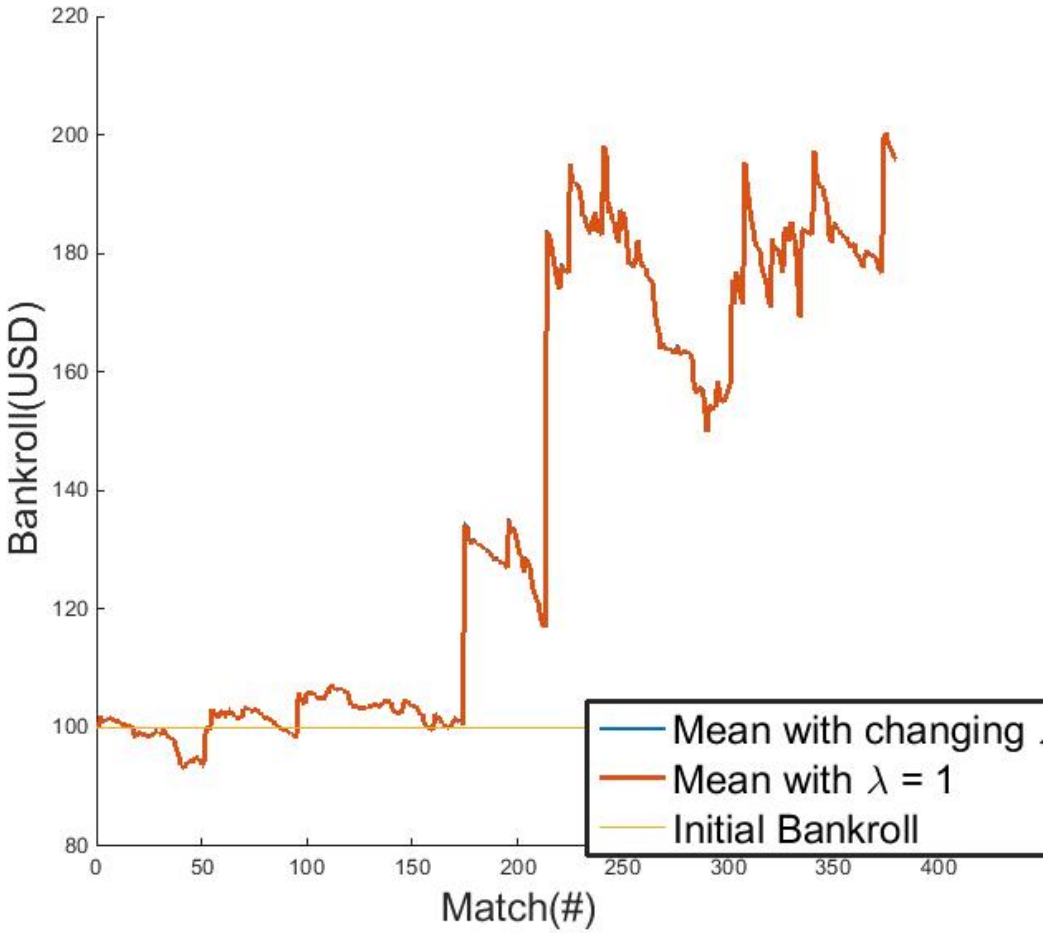

Figure: Maximise expected win.

Further work

Why do we consider to stop betting at the half of the season?

- Define what do we call a "Black swan".
- Stop betting when a "Black swan" appears.

References I

Baio, G. and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. *Journal of Applied Statistics*, 37(2):253–264.

Forrest, D., Goddard, J., and Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, 21:551–564.

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, 35(4):917–926.

Leitner, C., Zeileis, A., and Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. *International Journal of Forecasting*, 26(3):471–481.

References II

Peter F. Pope, D. A. P. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 56(223):323–341.

Ramírez, A. and Cardona, J. (2014). Which Team Will Win the 2014 FIFA World Cup? A Bayesian Approach for Dummies. `http://ssrn.com/abstract=2396259`. [Online; accessed 18-August -2015].

Shin, H. (1993). Measuring the incidence of insiders trading in a market for state-contingent claims. *The Economic Journal*, 103(420):1141–1153.

Spann, M. and Skiera, B. (2009). Sports forecasting: a comparison of the forecast accuracy of prediction markets, betting odds and tipsters. *Journal of Forecasting*, 28(1):55–72.

References III

Stekler, H., Sendor, D., and Verlander, R. (2010). Issues in sports forecasting. *International Journal of Forecasting*, 26(3):606–621.

Štrumbelj, E. (2014). On determining probability forecasts from betting odds. *International Journal of Forecasting*, 30(4):934–943.

Any questions?