Extensions of the Elo Rating System for Margin of Victory

MathSport International 2019 - Athens, Greece

Stephanie Kovalchik

Revolutionising sport through science.

The Game Insight Group (GIG), formed by Tennis Australia in partnership with Victoria University, is a team of experts revolutionising tennis through science.

With an immediate focus on Tennis, this endeavour is supported by a team of leading data scientists, computer engineers and sports scientists, aiming to innovate the way the fans, coaches and players enjoy and think about the game.

By connecting experts, science and data more than ever before, GIG is challenging the status quo and helping players, coaches, policy makers, fans and the media take their understanding and enjoyment of the game to new heights.

Our Partners

tennis

VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

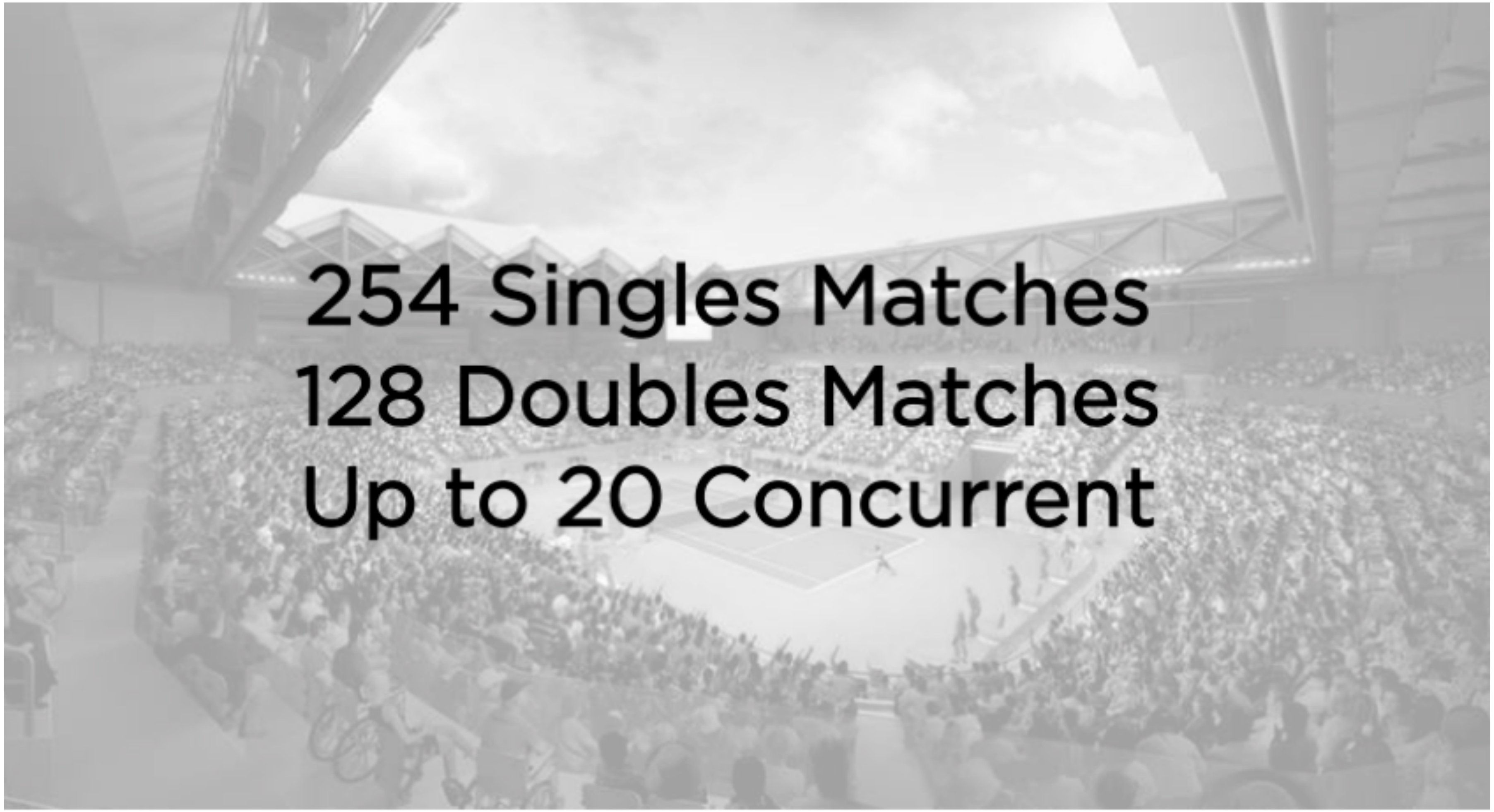
Media

@TennisAusGIG
@tennisausgig
giginfo@tennis.com.au


57%
ROU
HALEP
SIMONA
WIN
PREDICTOR
2018
43%
DEN
CAROLINE
WOZNIACKI
KIA
1 HALEP 0 15
2 WOZNIACKI 2 15


WIN PREDICTOR
43%
PRE-MATCH
57%
89%
SET 2
11%
+46%
SWING
CAROLINE WOZNIACKI
SIMONA HALEP
1 HALEP 6 40
2 WOZNIACKI 7 0
KIA
Fly Emirates


{prediction api}


254 Singles Matches
128 Doubles Matches
Up to 20 Concurrent

How Do We Make Match Forecasts?

It Starts with Player Ratings

Assume the the ith player has some true ability θ_i. Models of player abilities assume game outcomes are a function of the difference in abilities

$$Prob(W_{ij} = 1) = F(\theta_i - \theta_j)$$

Paired Comparison Models

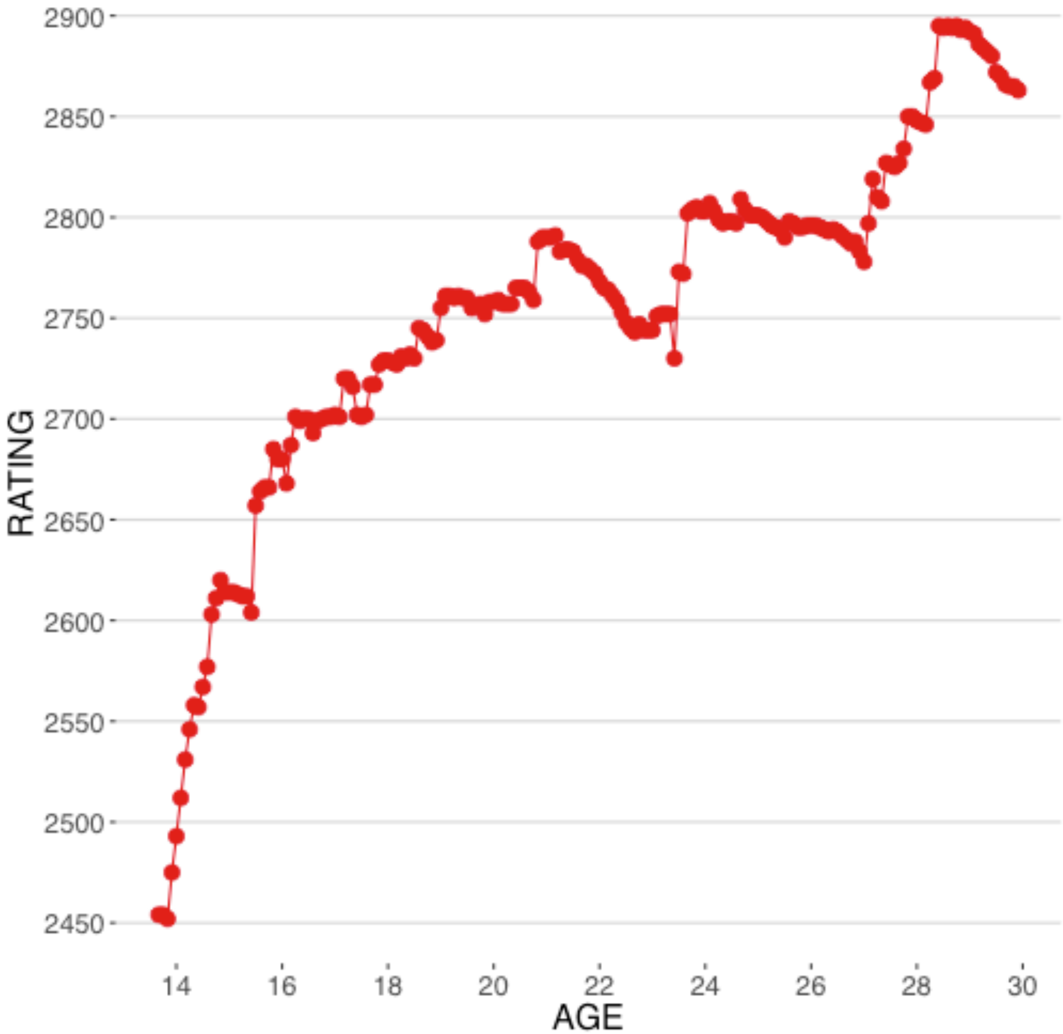
Bradley-Terry models are a general class of paired comparison models of latent abilities with a logistic function for win probabilities.

$$F(\theta_i - \theta_j) = \frac{1}{1 + \alpha^{-(\theta_i - \theta_j)}}$$

With BT, player abilities are treated as fixed in time which is unrealistic in most cases.

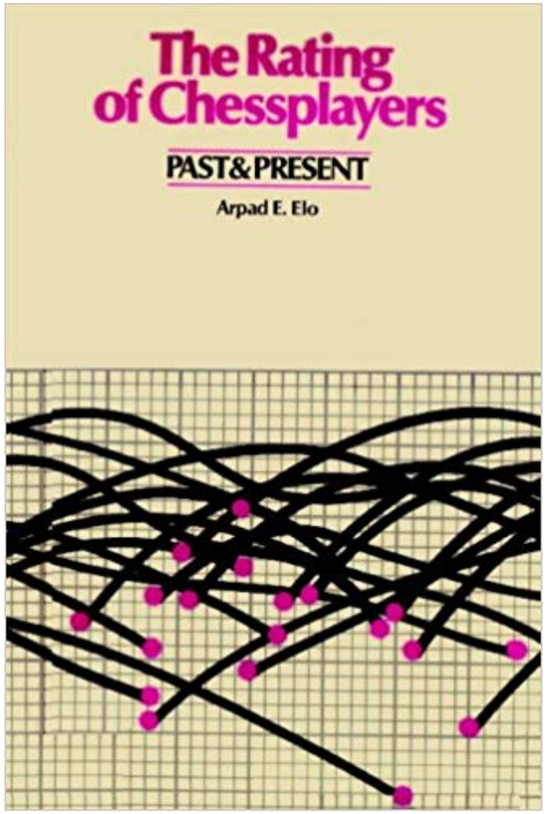
Bobby Fischer

Fischer's Meteoric Rise

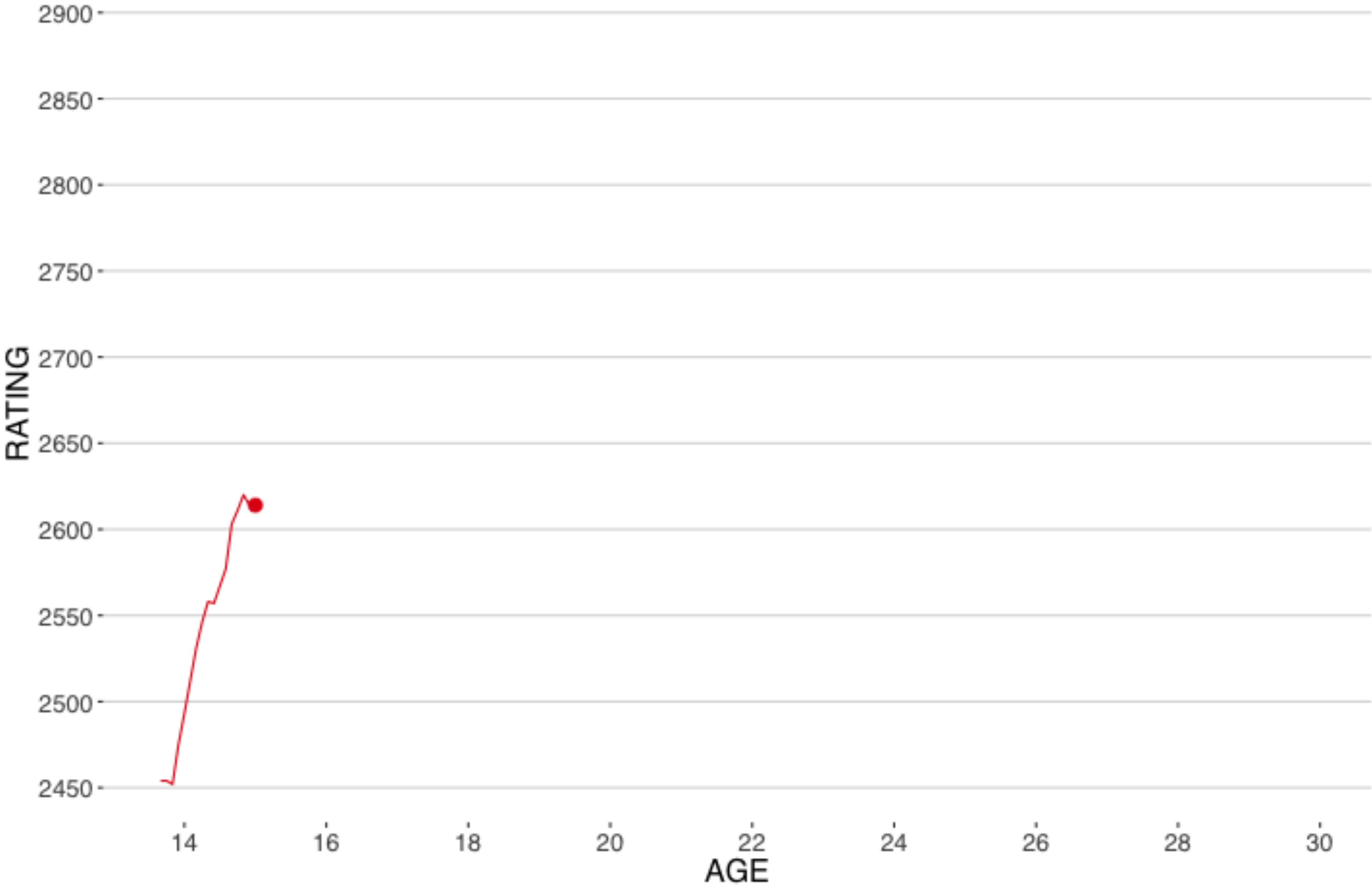


Arpad Elo

In His Own Words



Ability is a Moving Target



Standard Elo

Can be broken down into two steps:

1. Estimate (E-Step)
2. Update (U-Step)

Standard Elo E-Step

For tth match of player i against player j, the chance that player i wins is estimated as,

$$\hat{W}_{ijt} = \frac{1}{1 + 10^{-(R_{it} - R_{jt})/\sigma}}$$

Elo Derivation

Elo supposed that the ratings of any two competitors were independent and normal with shared standard deviation δ. Given this, he likened the chance of a win to the chance of observing a difference in ratings for ratings drawn from the same distribution,

$$R_{it} - R_{jt} \sim N(0, 2\delta^2)$$

which leads to,

$$P(R_{it} - R_{jt} > 0) = \Phi(\frac{R_{it} - R_{jt}}{\sqrt{2}\delta})$$

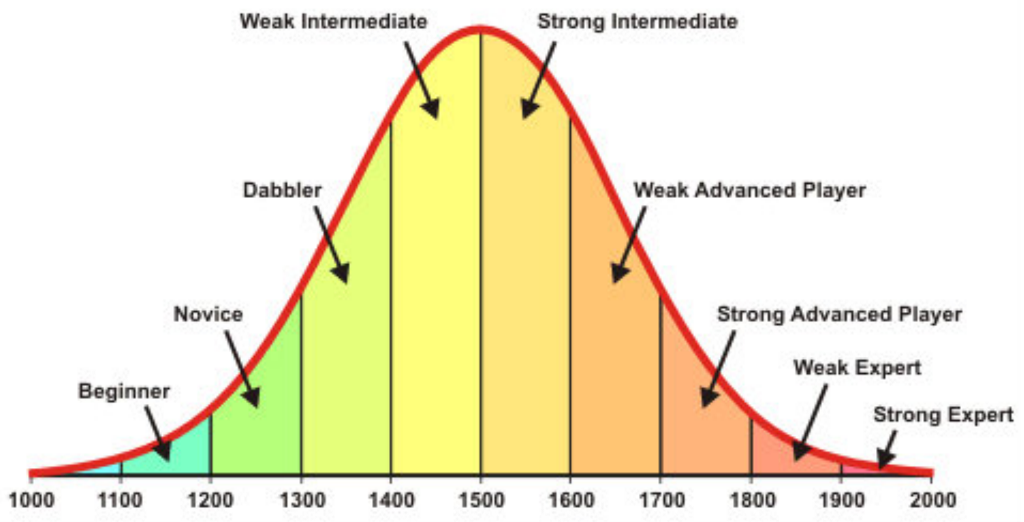
and

$$\approx 1/(1 + 10^{-(R_{it} - R_{jt})/2\delta})$$

Elo's formula was just a hack for the cumulative normal density function.

Choice of σ

Was based on the standard deviation of chessplayer ratings when Elo made the system, which was $SD \approx 200$. Thus $\sigma = 400$ in Elo's system.



Source: chess-site.com

Standard Elo U-Step

For a binary result W_{ijt}, the update to the ith player rating is,

$$R_{i(t+1)} = R_{it} + K(W_{ijt} - \hat{W}_{ijt})$$

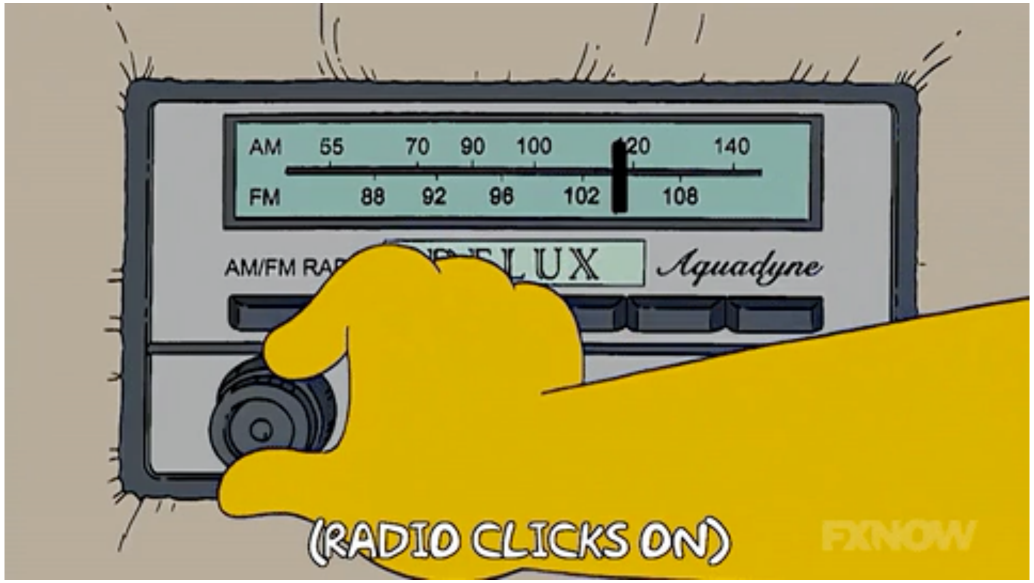
Standard Elo U-Step

For a binary result W_{ijt}, the update to the ith player rating is,

$$R_{i(t+1)} = R_{it} + K(W_{ijt} - \hat{W}_{ijt})$$

This adjusts according to the win *residual* and maximum possible gain (loss) of K.

Choice of K



Elo would vary K depending on the tournament type but 32 was one value he often used.

Elo's Model-Based Connections

State-space representation

$$P(W_{ij} = 1|\theta_i, \theta_j) = \frac{1}{1 + 10^{-(\theta_i - \theta_j)/400}}$$

Abilities are assumed to follow a normal distribution over a rating period τ

$$\theta_i^{t+\tau}|\theta_i^t, \nu^2, t \sim N(\theta_i^t, \nu^2 t)$$

Glicko (1999) is a Bayesian version, Fahrmeir and Tutz (1994) used Empirical Bayes

Elo's Model-Based Connections

Glickman showed that the Elo model is a special case of a state-space paired comparison model that assumes

1. The same prior knowledge about a player's strength throughout time
2. The strengths of opponents are known constants

Elo's Model-Based Connections

Glickman showed that the Elo model is a special case of a state-space paired comparison model that assumes

1. The same prior knowledge about a player's strength throughout time
2. The strengths of opponents are known constants

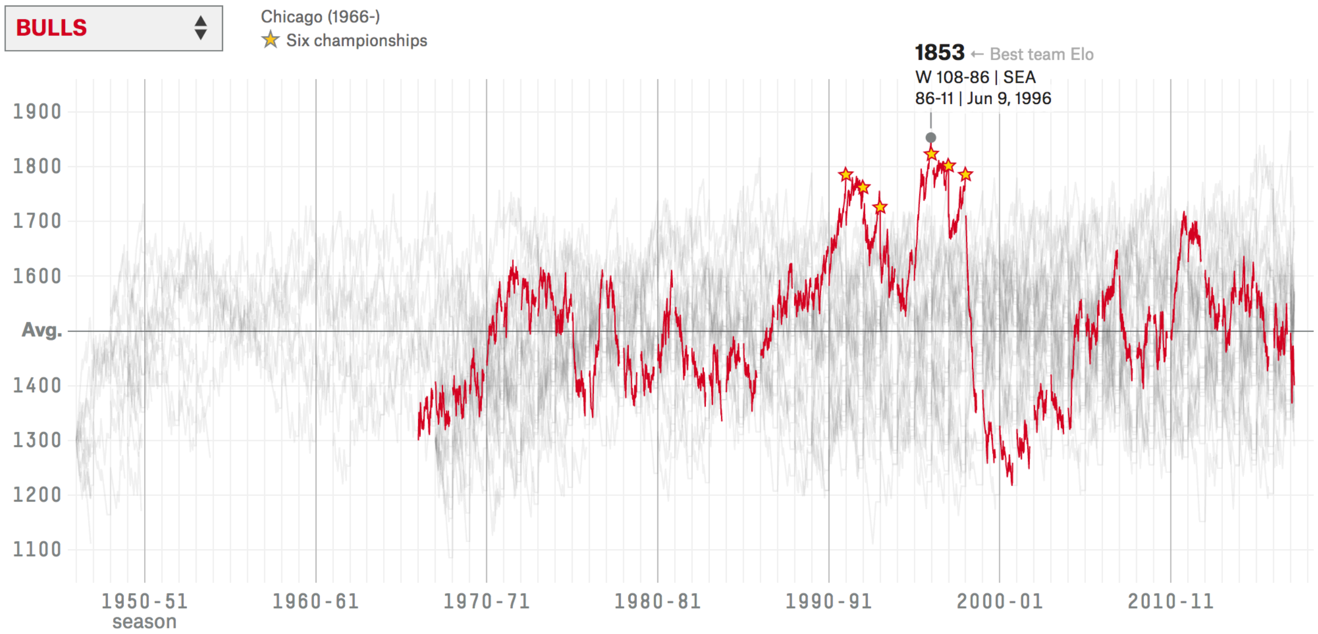
Thus, we can consider Elo as a pared down version of Glicko.

Simplicity Works

The Complete History Of The NBA

Every franchise's relative strength after every game. How this works »

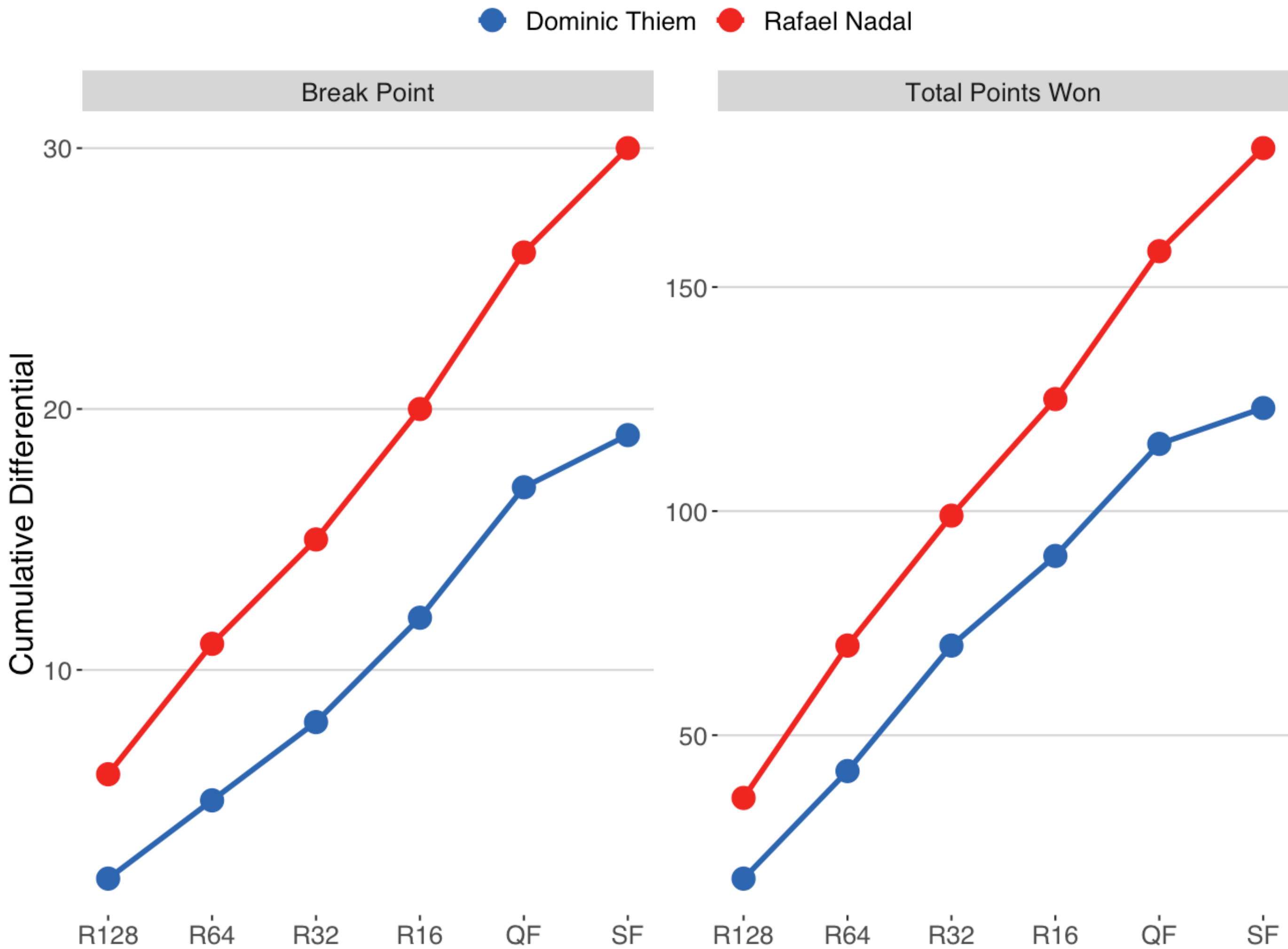
More NBA: Predictions for the 2017-18 season



Can Elo be Simple But Better?

Men's 2019 French Open Final

Can Elo be Simple But Better?



Margin Of Victory Modelling Principles

Margin Of Victory Modelling Principles

- Consider two-step 'estimate then update' algorithms

Margin Of Victory Modelling Principles

- Consider two-step 'estimate then update' algorithms
- Targets of estimation must be functions of relative ratings

Margin Of Victory Modelling Principles

- Consider two-step 'estimate then update' algorithms
- Targets of estimation must be functions of relative ratings
- Ratings updates are functions of residuals

Margin Of Victory Modelling Principles

- Consider two-step 'estimate then update' algorithms
- Targets of estimation must be functions of relative ratings
- Ratings updates are functions of residuals
- The MOV is incorporated into estimation, updating, or both

MOV Models

- Linear
- Joint Additive
- Multiplicative
- Logistic

Linear

E-Step

$$\hat{M}_{ijt} = \frac{R_{it} - R_{jt}}{\sigma}$$

Linear

E-Step

$$\hat{M}_{ijt} = \frac{R_{it} - R_{jt}}{\sigma}$$

U-Step

$$R_{i(t+1)} = R_{it} + K(M_{ijt} - \hat{M}_{ijt})$$

Joint Additive

E-Step

$$\hat{M}_{ijt} = \frac{R_{it} - R_{jt}}{\sigma_1}, \ \hat{W}_{ijt} = \frac{1}{1 + 10^{-(R_{it} - R_{jt})/\sigma_2}}$$

Joint Additive

E-Step

$$\hat{M}_{ijt} = \frac{R_{it} - R_{jt}}{\sigma_1},\ \hat{W}_{ijt} = \frac{1}{1 + 10^{-(R_{it} - R_{jt})/\sigma_2}}$$

U-Step

$$R_{i(t+1)} = R_{it} + K_1(M_{ijt} - \hat{M}_{ijt}) + K_2(W_{ijt} - \hat{W}_{ijt})$$

Multiplicative

E-Step

$$\hat{W}_{ijt} = \frac{1}{1 + 10^{-(R_{it} - R_{jt})/\sigma_2}}$$

Multiplicative

E-Step

$$\hat{W}_{ijt} = \frac{1}{1 + 10^{-(R_{it} - R_{jt})/\sigma_2}}$$

U-Step

$$R_{i(t+1)} = R_{it} + K(1 + |M_{ijt}/\sigma_1|)^{\alpha}(W_{ijt} - \hat{W}_{ijt})$$

$\alpha > 0$

When $\sigma_1 = 1$ this is the same Elo goal-based model of Hvattum and Arntzen (2010)

Logistic

E-Step

$$\hat{W}_{ijt} = L(\frac{R_{it} - R_{jt}}{\sigma_2})$$

Logistic

E-Step

$$\hat{W}_{ijt} = L(\frac{R_{it} - R_{jt}}{\sigma_2})$$

U-Step

$$R_{i(t+1)} = R_{it} + K[L(\frac{M_{ijt}}{\sigma_1}) - L(\frac{R_{it} - R_{jt}}{\sigma_2})]$$

where $L(x) = 1/(1 + \alpha^{-x})$ is a generalized logistic function.

Kinetic Model for Elo Asymptotics

Jabin and Junca (2015) propose a continuous kinetic model based on density $f(t, r, \theta)$, for players with rating r, true ability θ at time t,

$$\frac{\partial}{\partial t} f + \frac{\partial}{\partial r}(a[f]\ f) = 0$$

where $a[f]$ is a scalar vector field,

$$a[f] = \int_{\mathfrak{R}^2} w(r - r')(b(\theta - \theta') - b(r - r'))f(t, r', \theta')d\theta' dr'$$

- $w(.)$ describes the probability of interactions between players of different ratings
- $b(.)$ is the update function, describing how ratings change after a new result

Validity Conditions

Condition 1: Stationarity

When players have reached their true rating, the expected change in ratings should be zero.

Validity Conditions

Condition 1: Stationarity

When players have reached their true rating, the expected change in ratings should be zero.

Condition 2: Convergence

The rating system should converge to player true strengths. Under the kinetic model, Jabin and Junca showed that any Elo system with update function $b(.)$ *that meets the stationarity property and is Lipschitz continuous and strictly increasing satisfies this condition.*

Validity

Validity

- The **linear** model update, $(M_{ijt} - \hat{M}_{ijt})$ meets the stationarity and convergence conditions when $E[M_{ijt}] = \hat{M}_{ijt}$. That is, when we have correctly specified the expectation for the margin.

Validity

- The **linear** model update, $(M_{ijt} - \hat{M}_{ijt})$ meets the stationarity and convergence conditions when $E[M_{ijt}] = \hat{M}_{ijt}$. That is, when we have correctly specified the expectation for the margin.
- The **joint additive** is the sum of the linear and standard Elo updates, so it's validity depends on the same conditions as the **linear** model.

Validity

- The **linear** model update, $(M_{ijt} - \hat{M}_{ijt})$ meets the stationarity and convergence conditions when $E[M_{ijt}] = \hat{M}_{ijt}$. That is, when we have correctly specified the expectation for the margin.
- The **joint additive** is the sum of the linear and standard Elo updates, so it's validity depends on the same conditions as the **linear** model.
- The **multiplicative** model's validity is established by showing that its update function can be reparameterized as standard Elo with a modified K'.

Validity

- The **linear** model update, $(M_{ijt} - \hat{M}_{ijt})$ meets the stationarity and convergence conditions when $E[M_{ijt}] = \hat{M}_{ijt}$. That is, when we have correctly specified the expectation for the margin.
- The **joint additive** is the sum of the linear and standard Elo updates, so it's validity depends on the same conditions as the **linear** model.
- The **multiplicative** model's validity is established by showing that its update function can be reparameterized as standard Elo with a modified K'.
- The **logistic** model needs the strongest set of conditions as it's update, $L(M_{ijt}/\sigma_1) - L((R_{it} - R_{jt})/\sigma_2)$, is not a standard residual.

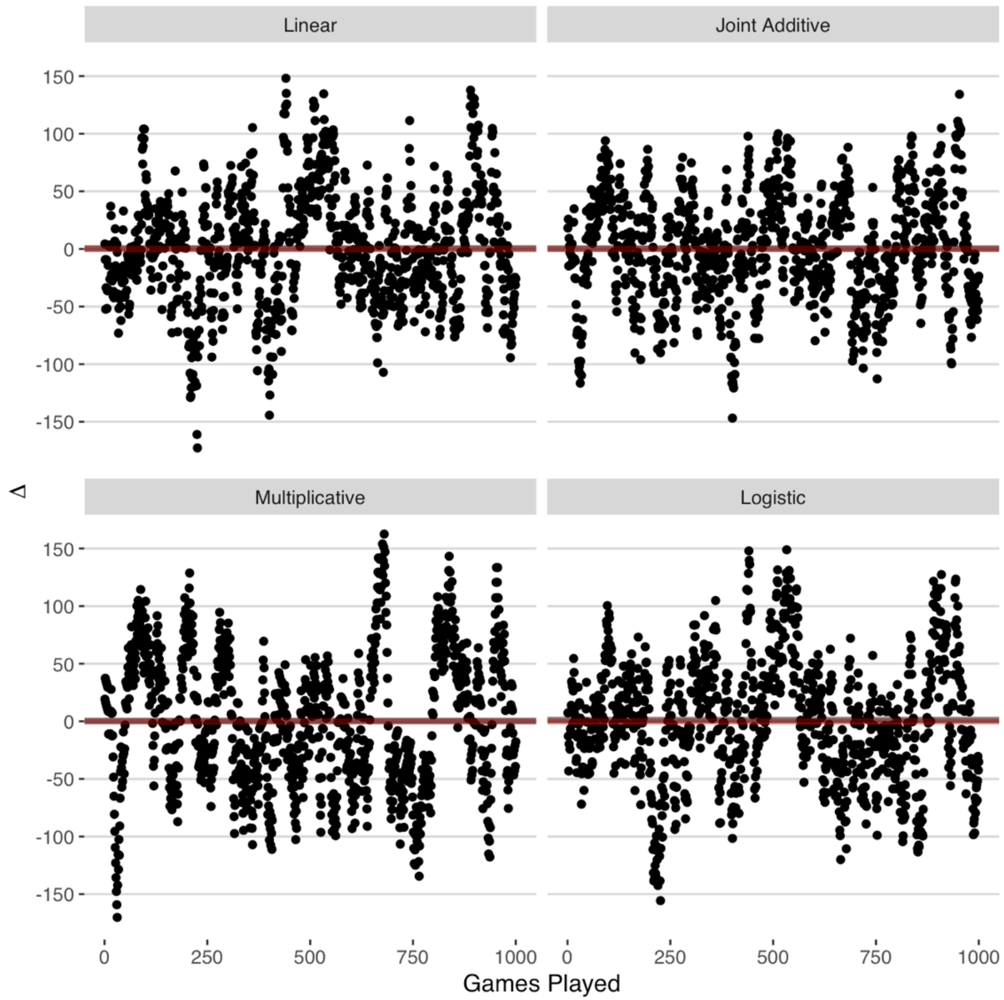
Simulation Study

For $N = 1000$,

$$R_{in} - R_{jn} \sim N(0, 50)$$

$$MOV_{ijn} | (R_{in} - R_{jn}) \sim N((R_{in} - R_{jn})/200, 1)$$

$$W_{ijn} | MOV_{ijn} \sim Bernoulli(1/(1 + 10^{-MOV_{ijn}/2}))$$


Linear
Joint Additive
Multiplicative
Logistic
150
100
50
0
-50
-100
-150
Δ
0
250
500
750
1000
Games Played

Application Study

ATP Dataset, Tuning 2000-2015, Testing 2016-2018

Margin Of Victory	Median	IQR	% Positive for Winner
SETS WON	2	1	100
GAMES WON	5	4	95
BREAK POINTS WON	2	2	90
TOTAL POINTS WON	14	10	94
SERVE PERCENTAGE WON	10	12	93

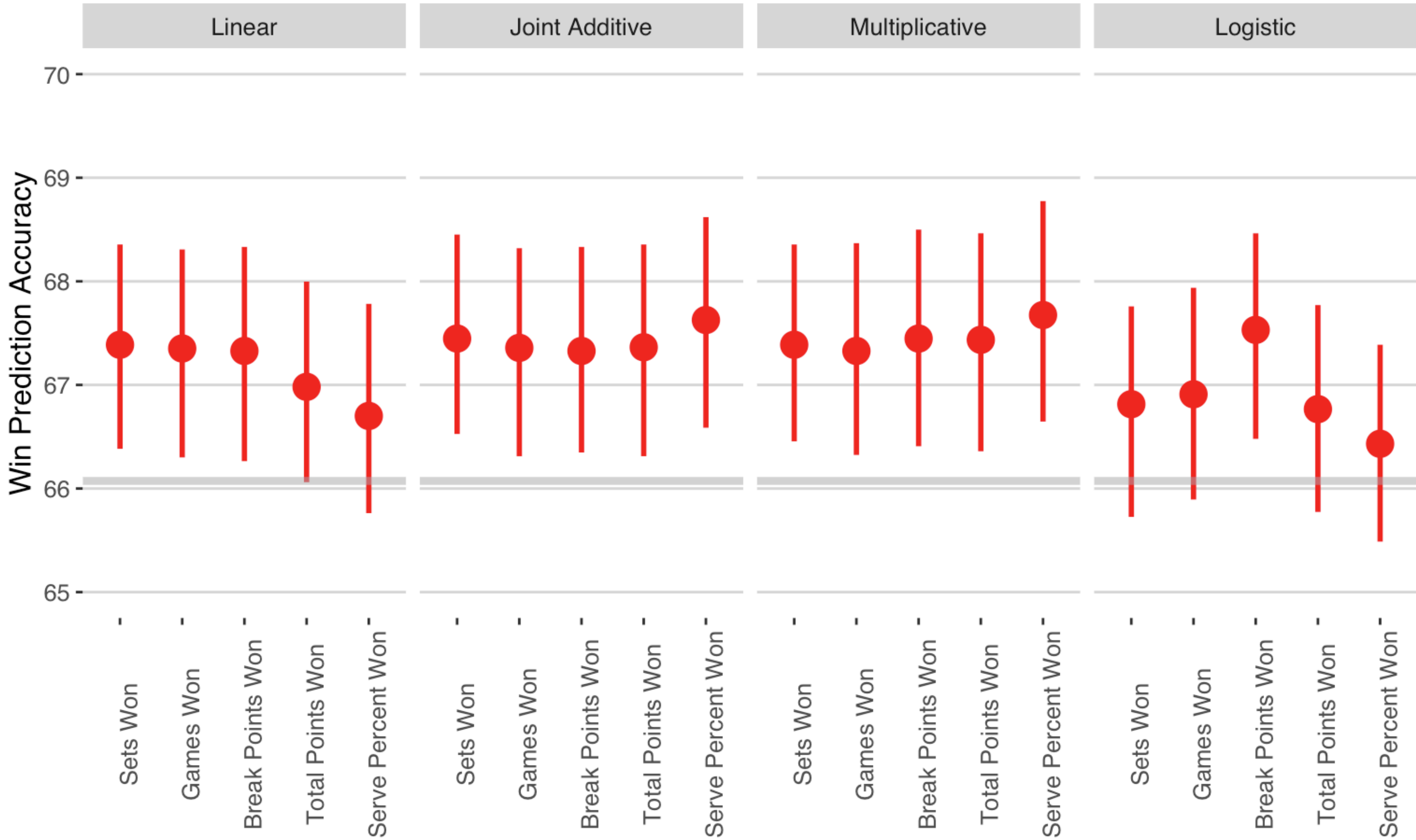
Model Tuning

Optimization with loss function that combines RMSE of MOV and log-loss of win predictions,

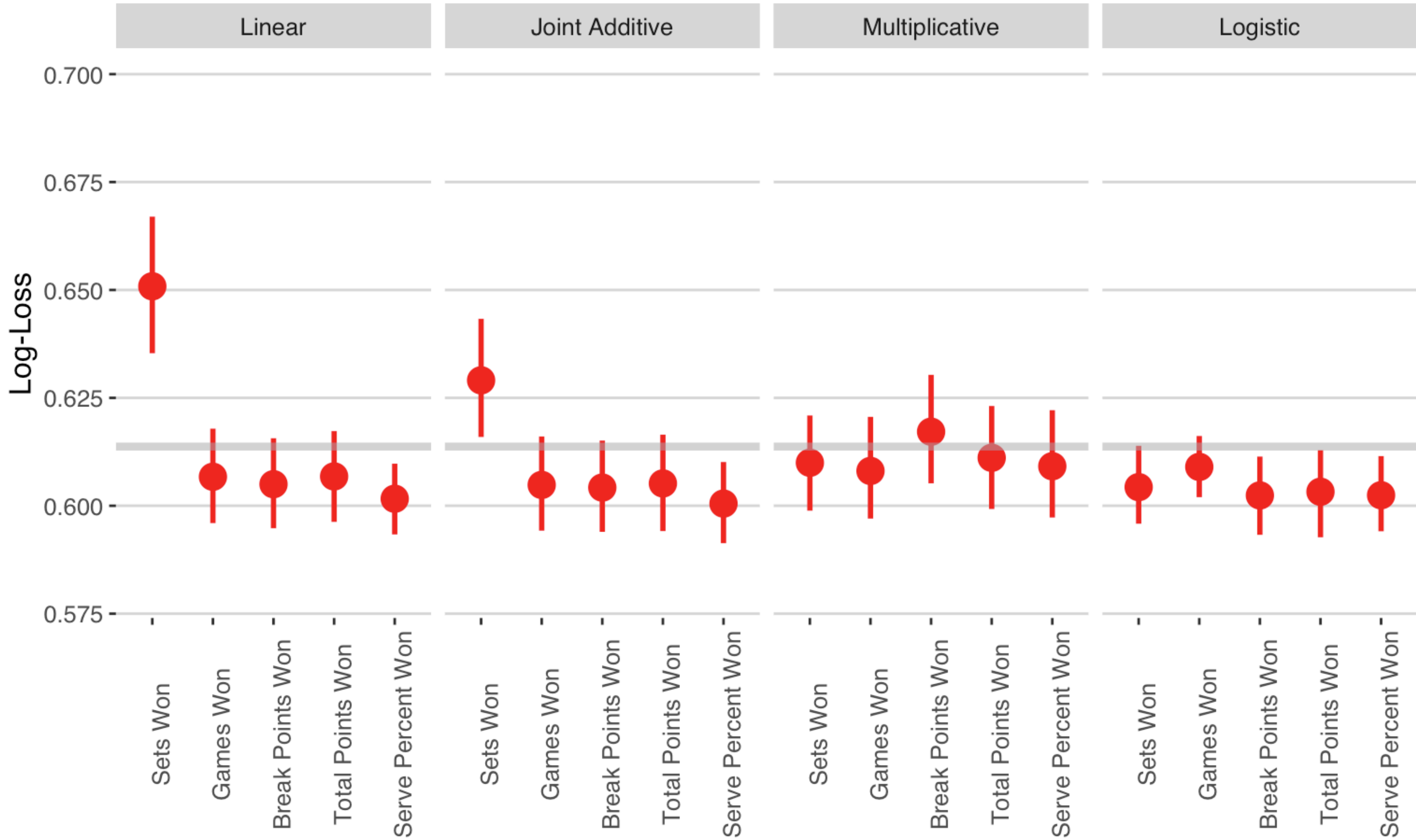
$$\mathcal{L}(\theta) = 1/N \left[\frac{\sqrt{\sum_{i,j,t} (\hat{M}_{ijt}(\theta) - M_{ijt})^2}}{3SD} - \sum_{i,j,t} log(\hat{P}_{ijt}(\theta)) \right.$$

Initial values:

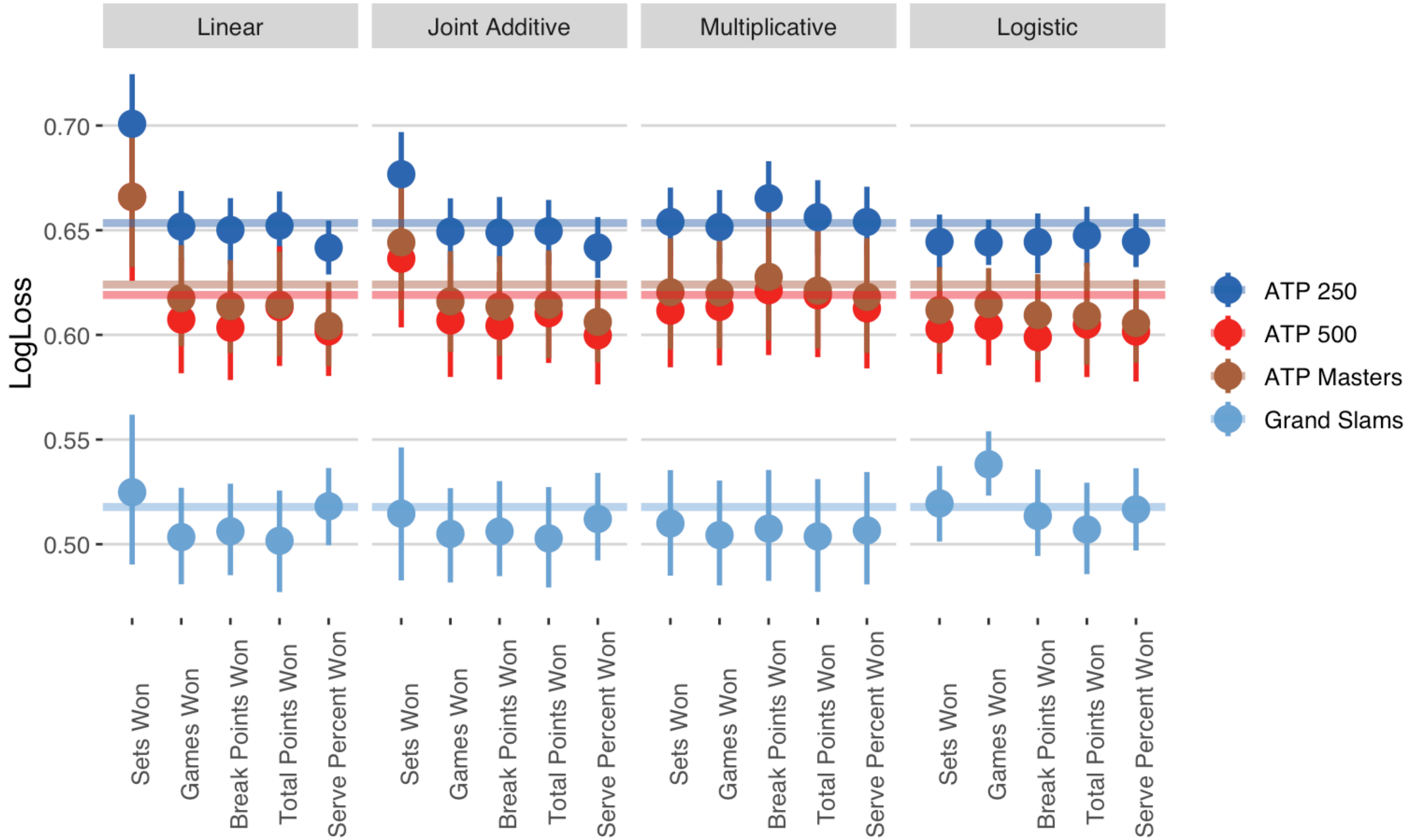
- Scaling rating difference to MOV $200/SD_{MOV}$
- Scaling learning rate to MOV residual $32/3SD_{MOV}$



Source: Horizontal line is standard Elo



Source: Horizontal line is standard Elo



Source: Horizontal lines are standard Elo

Takeaways

Takeaways

- Modellers have several valid options for incorporating MOV into their player ratings whether wins or the MOV are the target of interest

Takeaways

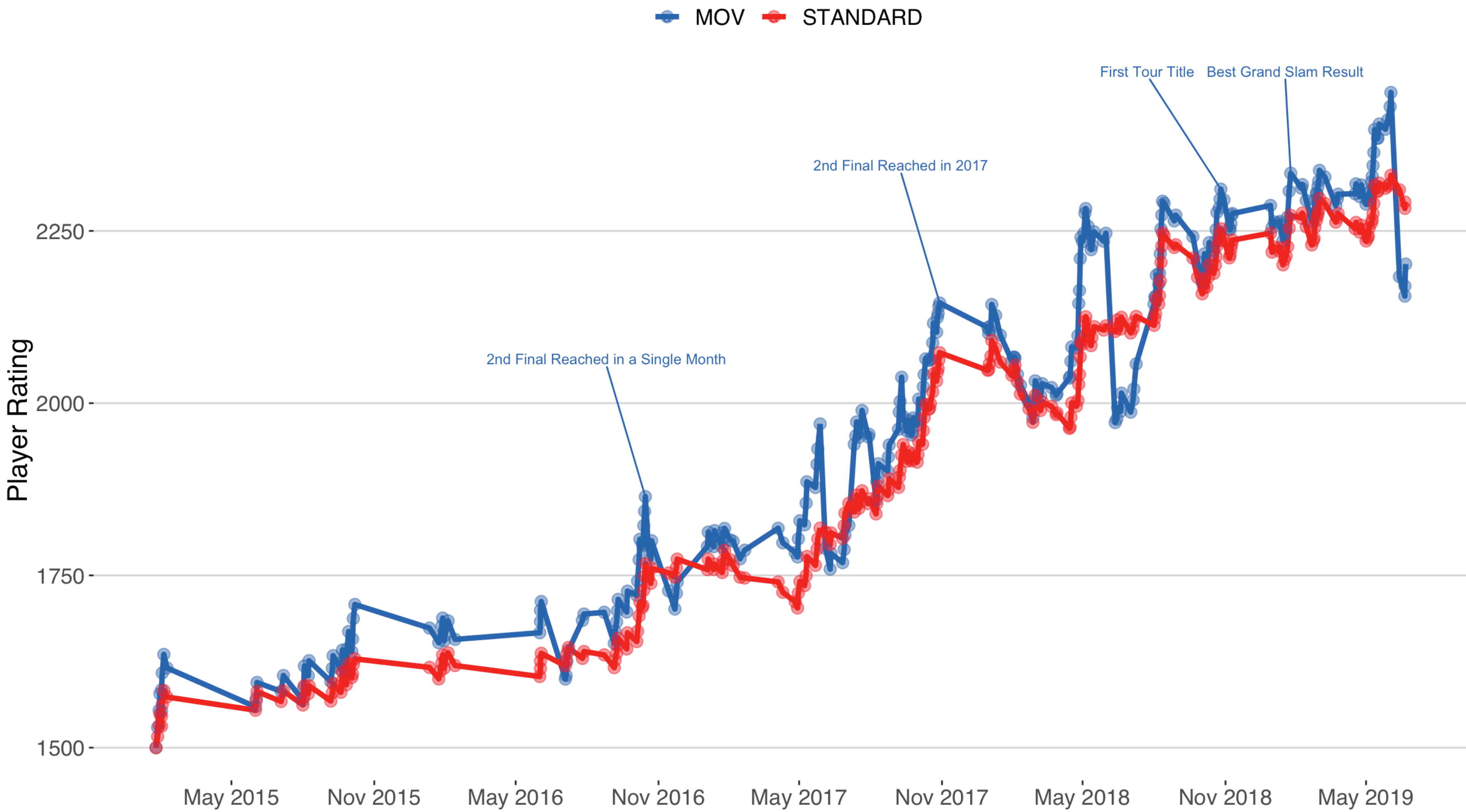
- Modellers have several valid options for incorporating MOV into their player ratings whether wins or the MOV are the target of interest

- When applied to men's tennis, MOV models improve predictive performance over standard Elo, the differences in gains depending more on the choice of MOV than model type

Takeaways

- Modellers have several valid options for incorporating MOV into their player ratings whether wins or the MOV are the target of interest
- When applied to men's tennis, MOV models improve predictive performance over standard Elo, the differences in gains depending more on the choice of MOV than model type
- State-space analogs to these models would allow for inference but aren't expected to improve predictive performance

The Rise of Tsitsipas



Wimbledon Prospects

Player	Grass Adjusted MOV Elo
Novak Djokovic	2562
Rafael Nadal	2539
Roger Federer	2478
Dominic Thiem	2279
David Goffin	2250
Kei Nishikori	2248
Gael Monfils	2244
John Isner	2238
Marin Cilic	2211
Roberto Bautista Agut	2207
Matteo Berrettini	2205
Alexander Zverev	2182
Milos Raonic	2178
Daniil Medvedev	2169
Stefanos Tsitsipas	2168

References

- Fahrmeir, L., Tutz, G., 1994. Dynamic stochastic models for time-dependent ordered paired comparison systems. Journal of the American Statistical Association 89 (428), 1438--1449.
- Glickman, M. E., 1999. Parameter estimation in large dynamic paired comparison experiments. Journal of the Royal Statistical Society: Series C (Applied Statistics) 48 (3), 377--394.
- Hvattum, L. M., Arntzen, H., 2010. Using Elo ratings for match result prediction in association football. International Journal of forecasting 26 (3), 460--470.
- Jabin, P.-E., Junca, S., 2015. A continuous model for ratings. SIAM Journal on Applied Mathematics 75 (2), 420--442.
- R package for Elo MOV github.com/GIGTennis/elomov