Design-weighted Regression Adjusted Plus-Minus

Schuckers, Im, Macdonald, McNulty

August 3, 2018

A better title

Design-weighted Regression Adjusted Plus-Minus

A better title

Design-weighted Regression Adjusted Plus-Minus : A First Pass

A better title

Design-weighted Regression Adjusted Plus-Minus : A First Pass (and an incompletion)

Regression Adjusted Plus-Minus (RAPM)

[Rosenbaum, 2004] basketball . . . Sagarin and Winston
[Ilardi, 2007] results for NBA 2006-7 season
[Ilardi and Barzilai, 2007] - separate Offensive and Defensive Impact
[Sill, 2010] - regularization and out of sample
[Engelman, 2017] & [Engelman, 2015] - single event credit

Applications outside basketball:
Hockey: [Macdonald, 2011] & [Schuckers and Curro, 2013]
Soccer: [scaryice, 2009],
[Hamilton, 2010b, Hamilton, 2010a, Hamilton, 2014]
Football: ????

RAPM

Two-way RAPM (Rosenbaum) + Sill

$$x_{ij} = \begin{cases} -1 & \text{if player } j \text{ is on the field for the Away team for event } i, \\ 0 & \text{if player } j \text{ is not on the field for event } i, \textit{and} \\ 1 & \text{if player } j \text{ is on the field for the Home team for event } i. \end{cases} \quad (1)$$

Augment design matrix, $\mathbf{X}$, with other predictors, and choose appropriate response, $\mathbf{Y}$.

$$\mathbf{Y} = \mathbf{X}\beta \quad (2)$$

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{Y} \quad (3)$$

Goals

- Generalize RAPM
- Account for player impact
- Reduce multicollinearity
- Allow for wider application

Generalization of the design matrix

Two-way RAPM

$$\gamma_{ij} = \begin{cases} -d_{ij} & \text{if player } j \text{ is on the field for the Away team for event } i, \\ 0 & \text{if player } j \text{ is not on the field for event } i, \textit{and} \\ d_{ij} & \text{if player } j \text{ is on the field for the Home team for event } i. \end{cases} \quad (4)$$

Augment design matrix, $\mathbf{\Gamma}$, with other predictors, and choose appropriate response, $\mathbf{Y}$.

$$\mathbf{Y} = \mathbf{\Gamma}\beta \quad (5)$$

$$\hat{\beta} = (\mathbf{\Gamma}^T\mathbf{\Gamma} + \lambda\mathbf{I})^{-1}\mathbf{\Gamma}^T\mathbf{Y} \quad (6)$$

How to choose d_{ij} as measure of impact of the player j on event i?

Choosing d_{ij}

Schuckers Intuition: larger $d_{ij} \rightarrow$ larger impact on coefficient for player i
B-Mac Intuition: smaller $d_{ij} \rightarrow$ larger impact on coefficient for player i

Choosing d_{ij}

Schuckers Intuition: larger $d_{ij} \rightarrow$ larger impact on coefficient for player i
B-Mac Intuition: smaller $d_{ij} \rightarrow$ larger impact on coefficient for player i

B-Mac wins

Choosing d_{ij}

Schuckers Intuition: larger $d_{ij} \rightarrow$ larger impact on coefficient for player i
B-Mac Intuition: smaller $d_{ij} \rightarrow$ larger impact on coefficient for player i

B-Mac wins and we can prove it

Choosing d_{ij}

Schuckers Intuition: larger $d_{ij} \rightarrow$ larger impact on coefficient for player i
B-Mac Intuition: smaller $d_{ij} \rightarrow$ larger impact on coefficient for player i

B-Mac wins and we can prove it but not today.

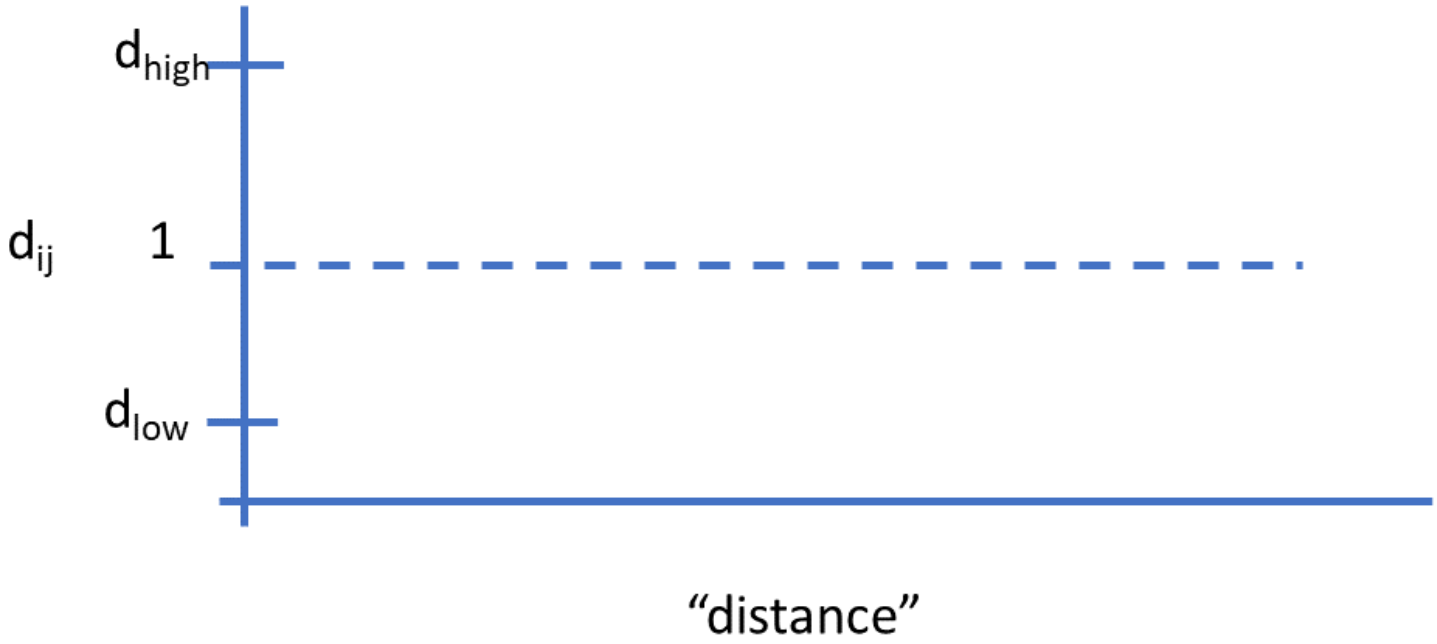
Choosing d_{ij}

Schuckers Intuition: larger $d_{ij} \rightarrow$ larger impact on coefficient for player i
B-Mac Intuition: smaller $d_{ij} \rightarrow$ larger impact on coefficient for player i

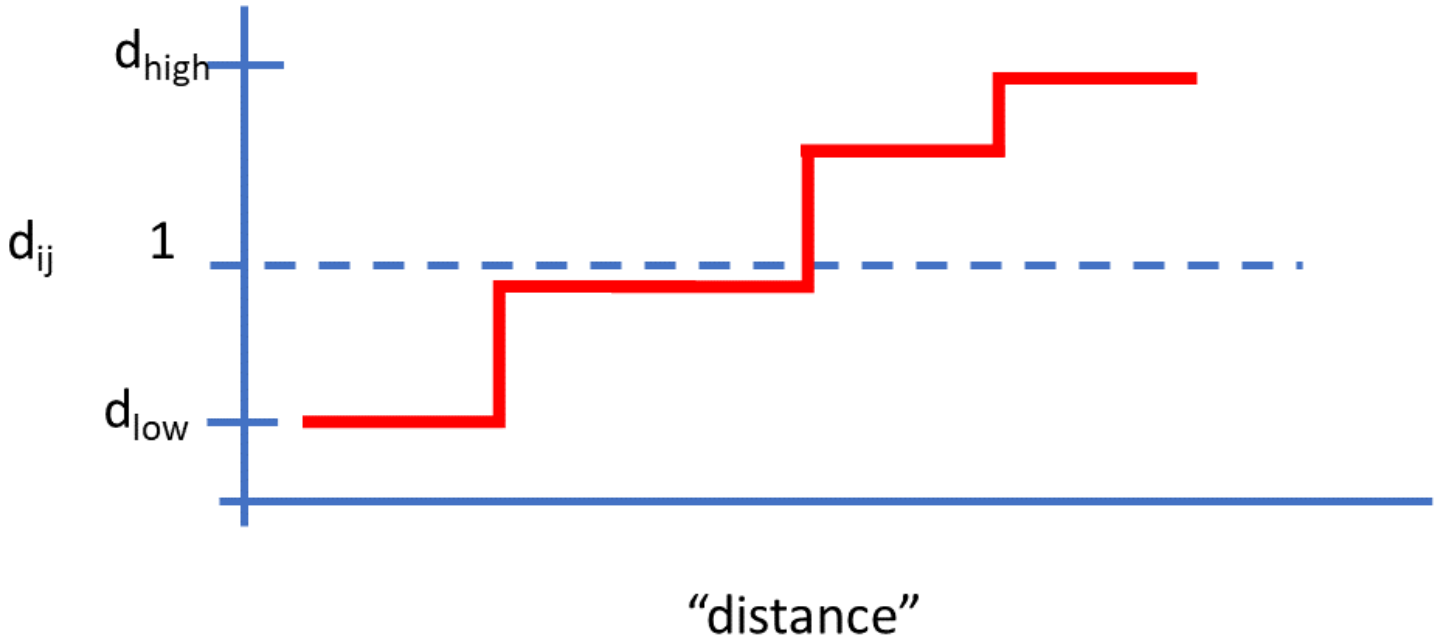
B-Mac wins and we can prove it but not today.

$\hat{\beta}_j$ is slope $= \Delta Y_i / \Delta d_{ij}$.
Think of d_{ij} as signed 'distance' metric from event i.

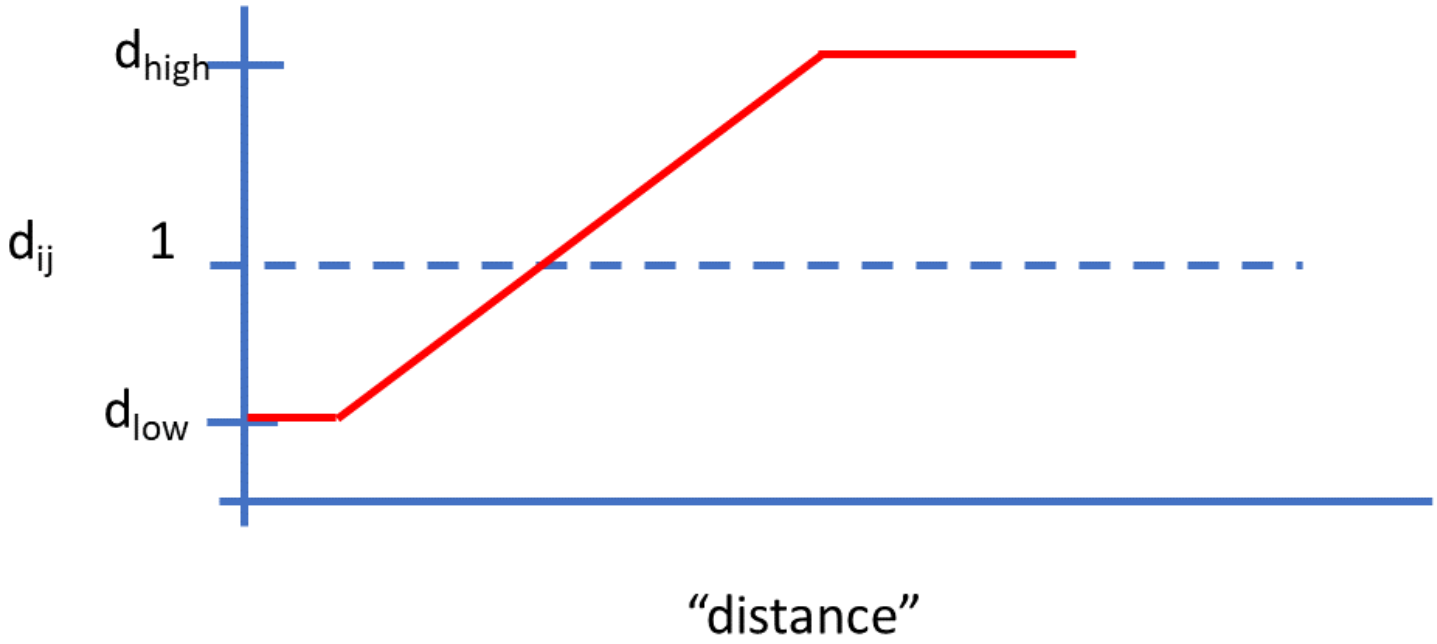
Choosing d_{ij}



Choosing d_{ij}



Choosing d_{ij}



Choosing d_{ij}

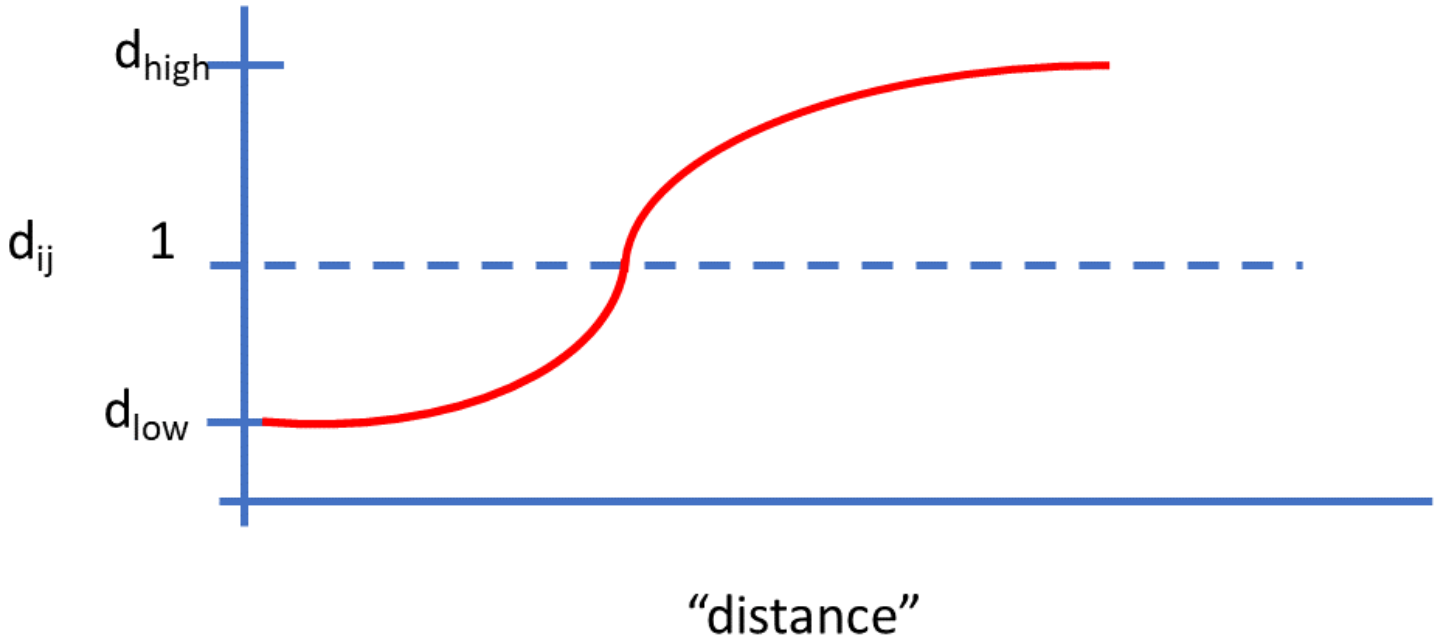
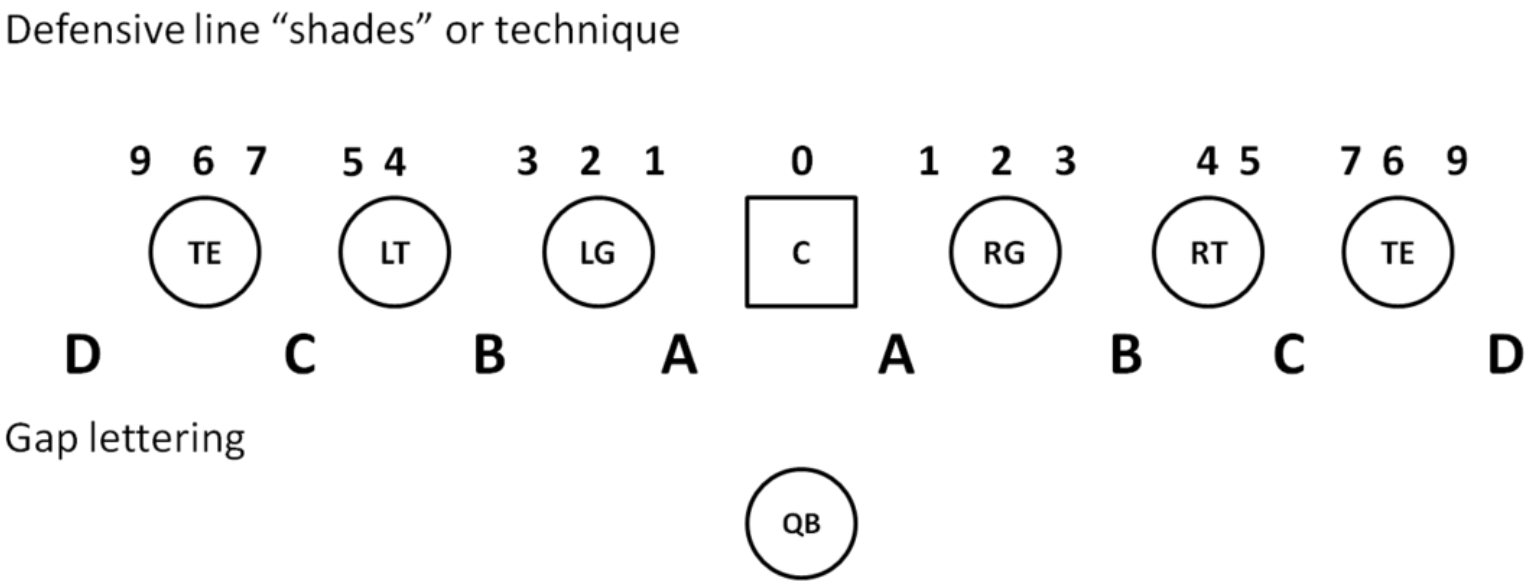


Illustration: Football

Application to Football data (or not)

Focus below on OLine to start, serious thoughts on formation, QB's, alignment of WR's, etc.

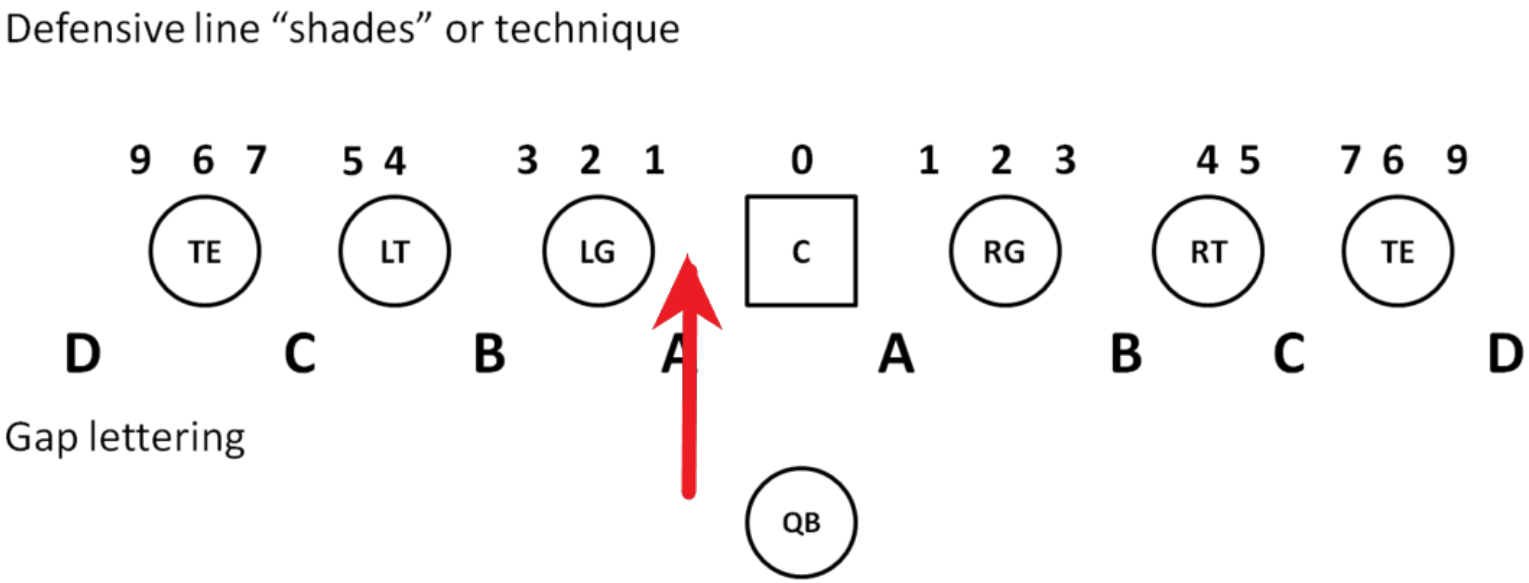
Illustration: Football



d_{ij} weights dependent on location of the run from this formation
d_{LTE}, d_{LT}, d_{LG}, d_C, d_{RG}, d_{RT}, d_{RTE}

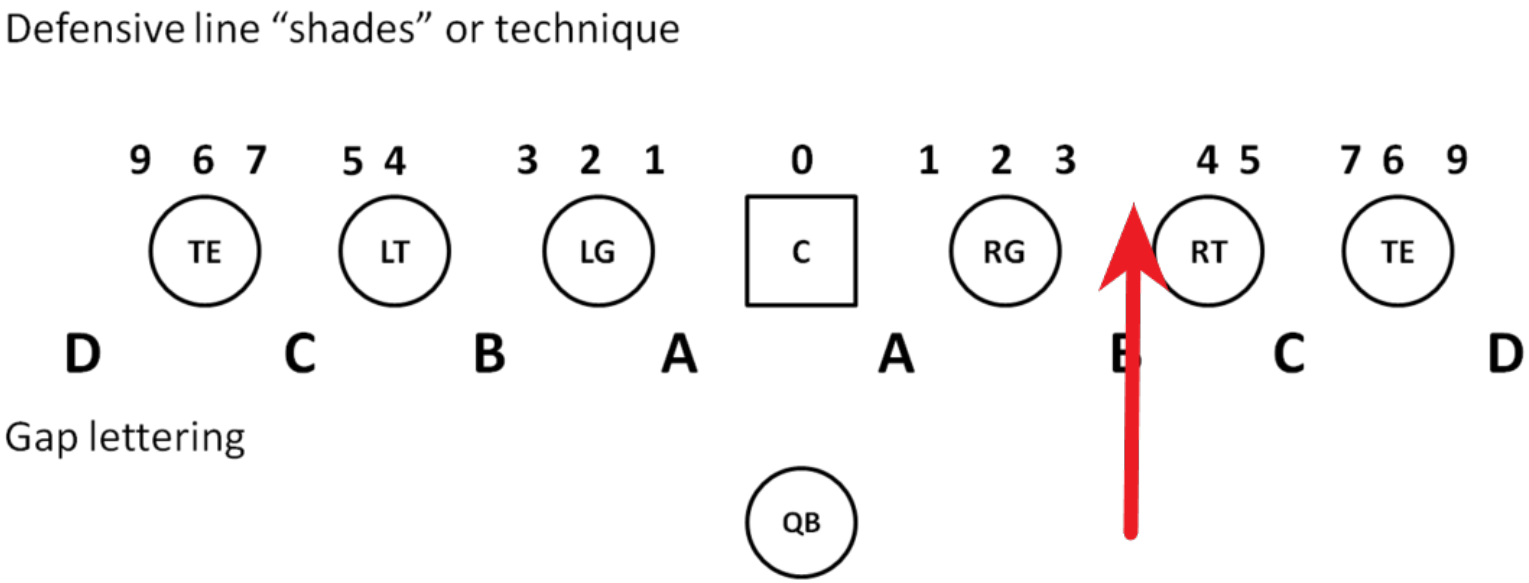
dropping i from notation for now
Source for image: `https://www.cougcenter.com/2013/3/28/4093000/air-raid-playbook-pass-protection-schemes`

Illustration: Football



$d_{LG} = 0.8,\ d_C = 0.8$
$d_{RG} = 1,\ d_{LT} = 1$
$d_{LTE} = 1.1,\ d_{RT} = 1.1$
$d_{RTE} = 1.15$

Illustration: Football



$d_{RG} = 0.8$, $d_{RT} = 0.8$
$d_{RTE} = 1$, $d_C = 1$
$d_{LG} = 1.1$, $d_{LT} = 1.15$
$d_{LTE} = 1.2$

Illustration: Hockey

Going to use THoR structure (Schuckers & Curro, 2013)

- Rosenbaum approach, single parameter Off & Def per player
- Response =NP20, ΔxG, home xG minus away xG in next 20 seconds
- Events (HITs, SHOTs, BLOCKs, . . .)
- Even Strength
- Covariates for team/goalie, rink, zone starts, Score effects, Score effects $\times$ time remaining in 3rd period

Evaluation on RMSE, Predicted OOS RMSE, Correlation player coeff.

Illustration: Hockey

For a given play i, use $d_{ij} = 1$ for goalies and the following for F and D:

	F	D
Off. Zone	a	(5-3a)/2
Neu. Zone	1	1
Def. Zone	(5-2b)/3	b

Consider values for $a, b = (0.7, 0.9, 0.95, 0.97, 1, 1.2)$ Evaluation on RMSE, Predicted OOS RMSE, Correlation player coeff.

Illustration: Hockey

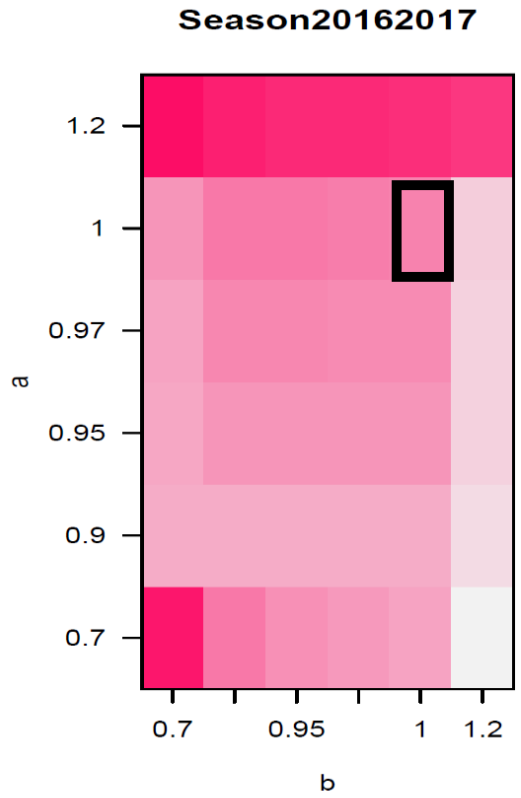
For a given play i, use $d_{ij} = 1$ for goalies and the following for F and D:

	F	D
Off. Zone	0.9	1.15
Neu. Zone	1	1
Def. Zone	1.2	0.7

Consider values for $a, b = (0.7, 0.9, 0.95, 0.97, 1, 1.2)$ Evaluation on RMSE, Predicted OOS RMSE, Correlation player coeff.

Illustration: THoR

2016-17 Season, $\lambda = 400$, Rows =a, Columns=b
Ratio of OOS RMSE (a,b)/ OOS RMSE (a=1,b=1)



White is low (good), Red is high (low)

Illustration: THoR

Columns=Seasons, Rows= λ, inside rows =a, inside columns=b

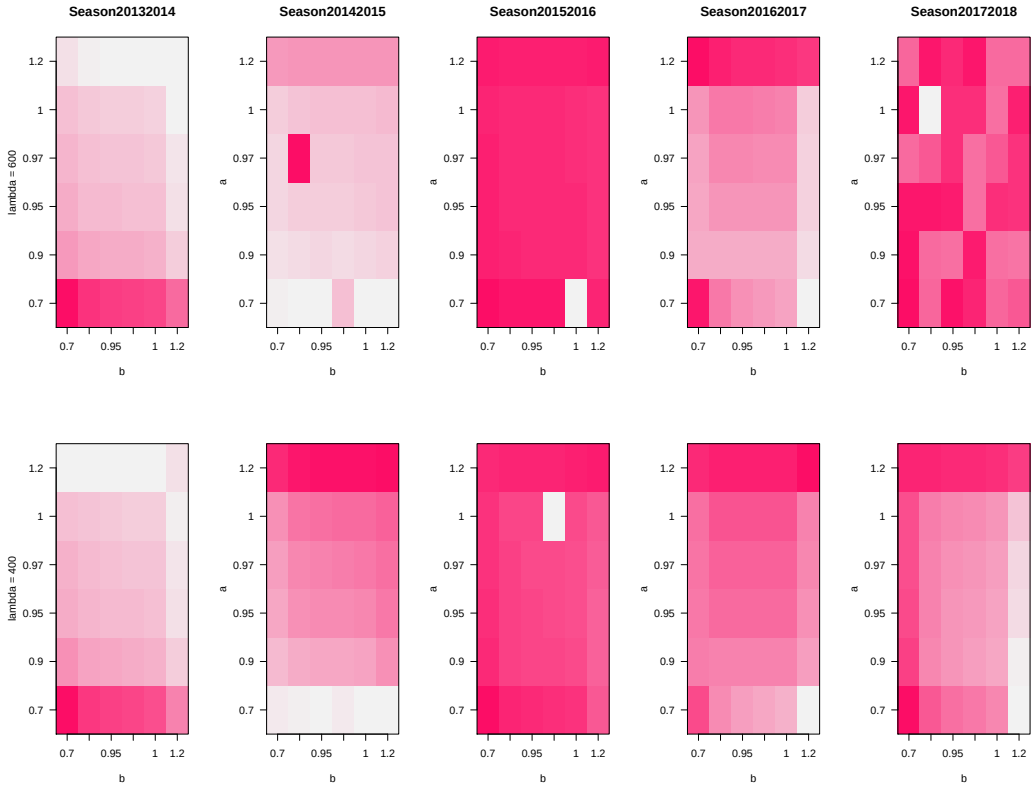
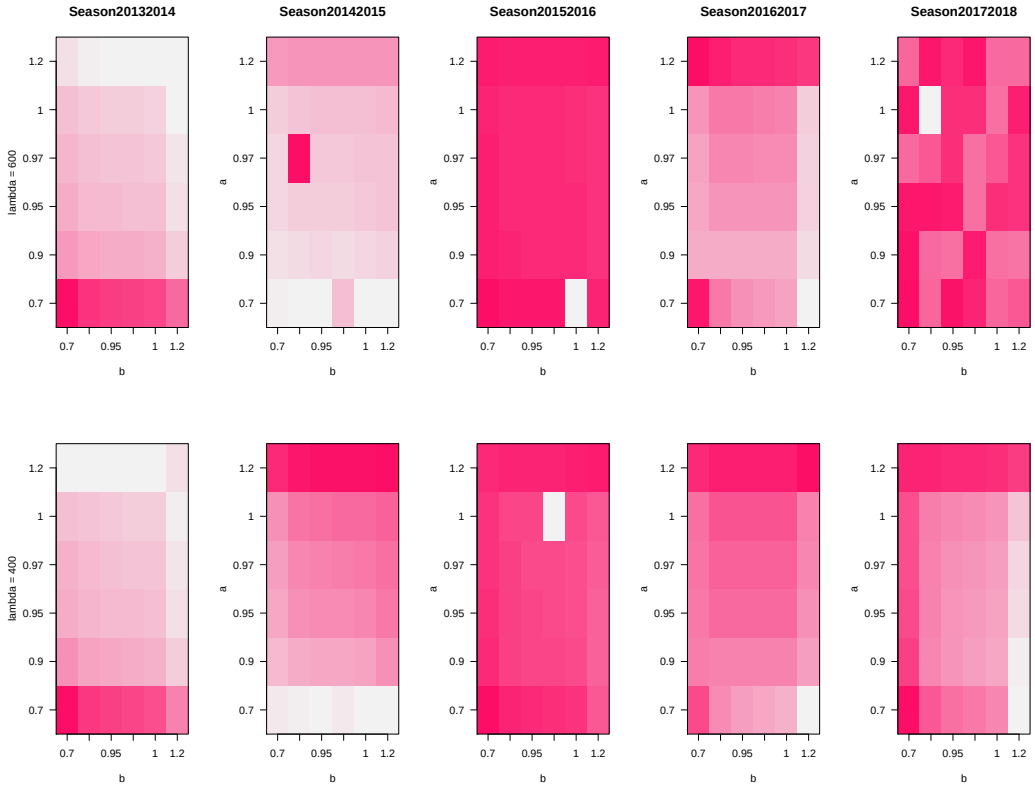


Illustration: THoR

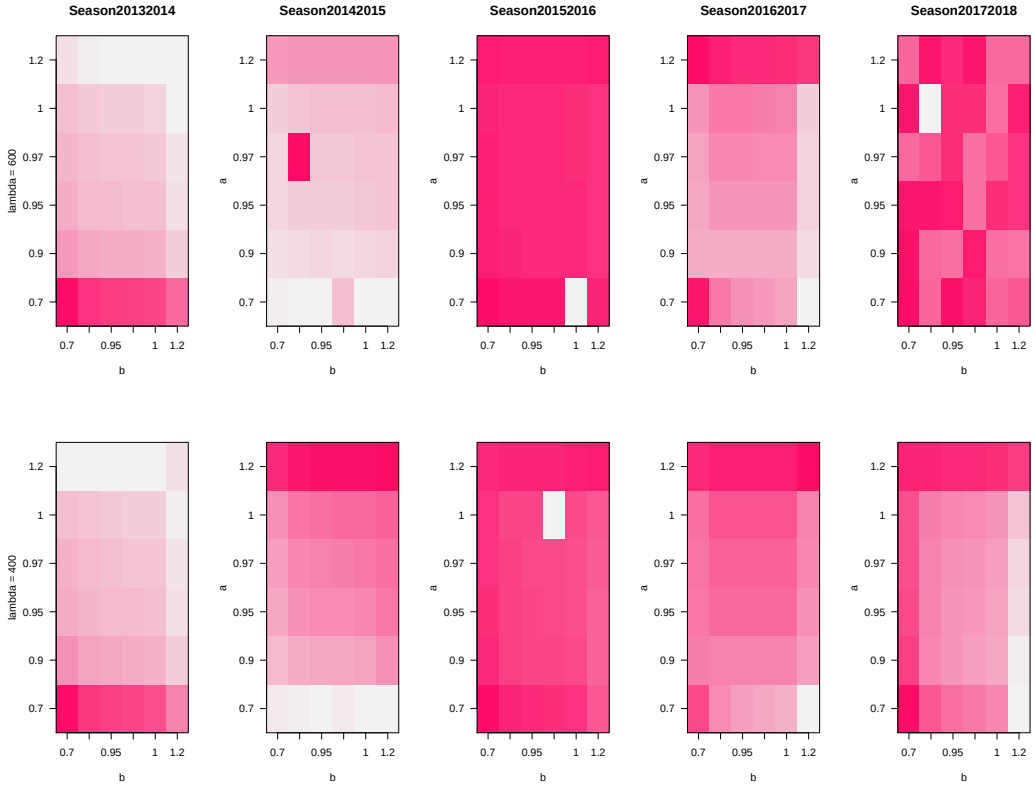
Columns=Seasons, Rows= λ, inside rows =a, inside columns=b



White is low, Red is high

Illustration: THoR

Columns=Seasons, Rows= λ, inside rows =a, inside columns=b



White is 0.999, Red is 1.001

Illustration: THoR

Summary of results:

- Some cases for some seasons do better than $a = 1, b = 1$.

Illustration: THoR

Summary of results:

- Some cases for some seasons do better than $a = 1, b = 1$.
- Generally very little improvement over RAPM

Illustration: THoR

Summary of results:

- Some cases for some seasons do better than $a = 1, b = 1$.
- Generally very little improvement over RAPM
- Same results for Season to Season correlation

Illustration: THoR

Summary of results:

- Some cases for some seasons do better than $a = 1, b = 1$.
- Generally very little improvement over RAPM
- Same results for Season to Season correlation
- Same results for within sample correlation

Illustration: THoR

Summary of results:

- Some cases for some seasons do better than $a = 1, b = 1$.
- Generally very little improvement over RAPM
- Same results for Season to Season correlation
- Same results for within sample correlation
- Same results for different set of events and response (Fenwick)

Illustration: THoR

Summary of results:

- Some cases for some seasons do better than $a = 1, b = 1$.
- Generally very little improvement over RAPM
- Same results for Season to Season correlation
- Same results for within sample correlation
- Same results for different set of events and response (Fenwick)
- Two way nature of hockey? Two way nature of this model

Illustration: THoR

Summary of results:

- Some cases for some seasons do better than $a = 1, b = 1$.
- Generally very little improvement over RAPM
- Same results for Season to Season correlation
- Same results for within sample correlation
- Same results for different set of events and response (Fenwick)
- Two way nature of hockey? Two way nature of this model
- What I wouldn't give for tracking data?

Next Steps

- Pick myself up
- Dig into coefficients (who changes)
- Rosenbaum (2-way) to Ilardi & Barzalai (Off & Def ratings)
- $\sum_i \mid d_{ij} \mid = k$ vs. $\sum_i \mid d_{ij} \mid^2 = k$ and $tr()$
- Metrics of 'distance'
- Other sports?

2018 Ottawa Hockey Analytics Conference
Sept 14 (social and workshops) & 15,
#OTTHAC18
Carleton University, Ottawa, ON
www.statsportsconsulting.com/otthac18

Thanks

schuckers@stlawu.edu

Bibliography


Engelman, J. (2015).
Estimating a player's influence on his teammates' boxscore statistics using a modified rapm framework.
https://www.youtube.com/watch?v=0uC0YZTADcE.
Accessed April 18, 2018.


Engelman, J. (2017).
Possession-based player performance analysis in basketball (adjusted +/- and related concepts).
In Albert, J., Glickman, M. E., Swartz, T. B., and Koning, R. H., editors, *Handbook of Statistical Methods and Analyses in Sports*. Chapman and Hall/CRC.


Hamilton, H. (2010a).
Some starting documents on the plus/minus problem.
http://www.soccermetrics.net/player-performance/some-starting-documents-on-the-plusminus-problem-2.