Estimating Player Contribution in Hockey with Regularized Logistic Regression

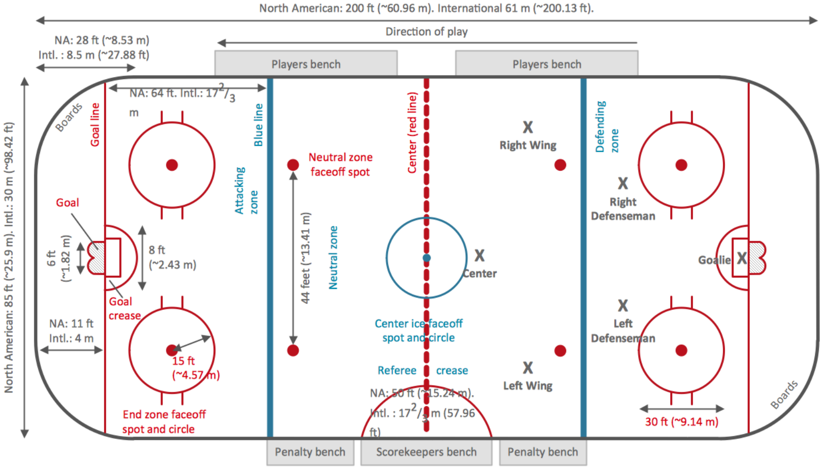
Robert B. Gramacy   Shane T. Jensen   Matt Taddy

March 24, 2017

Sports Statistics, Spring 2017

Alex Zajichek

Background


North American: 200 ft (~60.96 m). International 61 m (~200.13 ft).
Direction of play
NA: 28 ft (~8.53 m)
Intl. : 8.5 m (~27.88 ft)
Players bench
Players bench
NA: 64 ft. Intl.: 17 2/3 m
Boards
Goal line
Blue line
Center (red line)
Defending zone
Attacking zone
Neutral zone faceoff spot
Right Wing
Right Defenseman
Goal
6 ft (~1.82 m)
8 ft (~2.43 m)
44 feet (~13.41 m)
Neutral zone
Center
Goalie
North American: 85 ft (~25.9 m). Intl.: 30 m (~98.42 ft)
Goal crease
NA: 11 ft
Intl.: 4 m
Center ice faceoff spot and circle
Left Defenseman
15 ft (~4.57 m)
Referee crease
Left Wing
NA: 50 ft (~15.24 m).
Intl. : 17 2/3 m (57.96 ft)
End zone faceoff spot and circle
30 ft (~9.14 m)
Boards
Penalty bench
Scorekeepers bench
Penalty bench

Introduction

Goal: Develop metric to evaluate individual player performance

Introduction

Goal: Develop metric to evaluate individual player performance

- Traditional measure is the *plus-minus* value

Plus-Minus

Plus-Minus

For the i^{th} player in the NHL,

Plus-Minus

For the i^{th} player in the NHL,

$$PM_i = \sum_{j=1}^{N_i} x_{ij}$$

where

Plus-Minus

For the i^{th} player in the NHL,

$$PM_i = \sum_{j=1}^{N_i} x_{ij}$$

where

$$N_i = \# \text{ of goals player } i \text{ is on the ice for}$$

and

Plus-Minus

For the i^{th} player in the NHL,

$$PM_i = \sum_{j=1}^{N_i} x_{ij}$$

where

$$N_i = \# \text{ of goals player } i \text{ is on the ice for}$$

and

$$x_{ij} = \begin{cases} 1 & \text{if player } i \text{ is on the ice for his team's goal} \\ -1 & \text{if player } i \text{ is on the ice for opponent's goal} \end{cases}$$

Introduction

Goal: Develop metric to evaluate individual player performance

- Traditional measure is the *plus-minus* value

Introduction

Goal: Develop metric to evaluate individual player performance

- Traditional measure is the *plus-minus* value
    - Most popular metric; easy to calculate and understand

Introduction

Goal: Develop metric to evaluate individual player performance

- Traditional measure is the *plus-minus* value
  - Most popular metric; easy to calculate and understand
  - Depends heavily on teammate and opponent ability (marginal effect)

Introduction

Goal: Develop metric to evaluate individual player performance

- Traditional measure is the *plus-minus* value
  - Most popular metric; easy to calculate and understand
  - Depends heavily on teammate and opponent ability (marginal effect)
  - No control for sample size

Introduction

Goal: Develop metric to evaluate individual player performance

- Traditional measure is the *plus-minus* value
  - Most popular metric; easy to calculate and understand
  - Depends heavily on teammate and opponent ability (marginal effect)
  - No control for sample size
- Other metrics have been proposed to take into account team effect, hits, faceoffs, etc. (adj. plus-minus, Corsi)

Introduction

Goal: Develop metric to evaluate individual player performance

- Traditional measure is the *plus-minus* value
  - Most popular metric; easy to calculate and understand
  - Depends heavily on teammate and opponent ability (marginal effect)
  - No control for sample size
- Other metrics have been proposed to take into account team effect, hits, faceoffs, etc. (adj. plus-minus, Corsi)
- Proposed a logistic regression model to estimate *partial effects* for players

Preliminary modeling framework

Preliminary modeling framework

Let

q_i = probability that a given goal i was scored by the home team

Preliminary modeling framework

Let

$q_i = \text{probability that a given goal } i \text{ was scored by the home team}$

then

$$log\left(\frac{q_i}{1-q_i}\right) = \alpha_i + \beta_{h_{i1}} + \ldots + \beta_{h_{i6}} - \beta_{a_{i1}} - \ldots - \beta_{a_{i6}} \qquad (1)$$

Preliminary modeling framework

Let

$q_i = \text{probability that a given goal } i \text{ was scored by the home team}$

then

$$log\left(\frac{q_i}{1-q_i}\right) = \alpha_i + \beta_{h_{i1}} + ... + \beta_{h_{i6}} - \beta_{a_{i1}} - ... - \beta_{a_{i6}} \qquad (1)$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_{n_p})$ is the vector of *partial plus-minus effects* for all n_p players in the analysis, and $\{h_{i1}...h_{i6}\}, \{a_{i1}...a_{i6}\}$ are the indices of $\boldsymbol{\beta}$ corresponding to home and away players on the ice for goal i, respectively. α_i may depend upon additional information such as team effect.

Data

Data

- Player and goalie data for every *even strength* regular season goal from 2007-2011 seasons

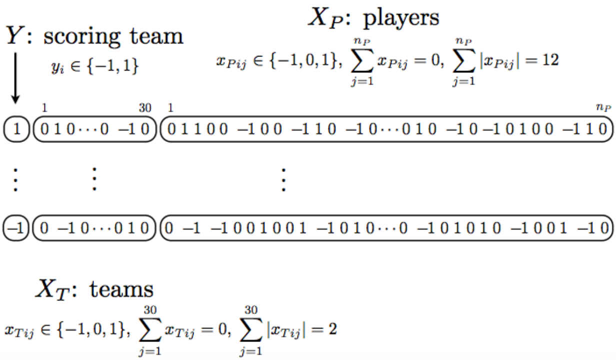
Data

- Player and goalie data for every *even strength* regular season goal from 2007-2011 seasons
- There were $n_p = 1467$ players involved in $n_g = 18154$ goals

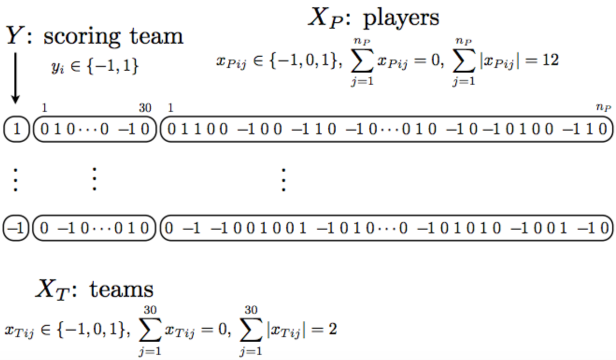
Data

- Player and goalie data for every *even strength* regular season goal from 2007-2011 seasons
- There were $n_p = 1467$ players involved in $n_g = 18154$ goals
- Player effect is treated as constant over the range of the seasons considered

Data

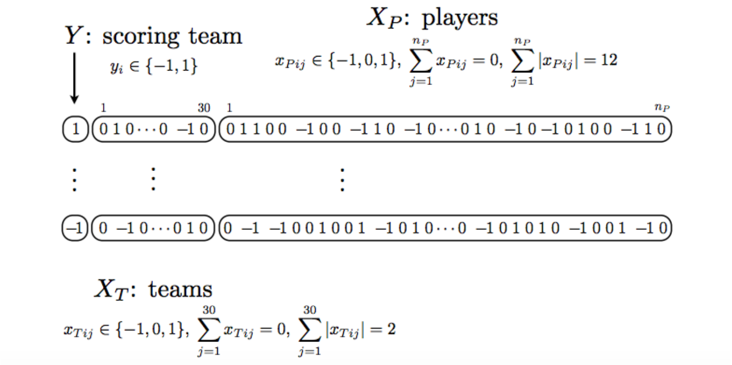

X_P: players
Y: scoring team
$y_i \in \{-1, 1\}$
$x_{Pij} \in \{-1, 0, 1\}, \sum_{j=1}^{n_P} x_{Pij} = 0, \sum_{j=1}^{n_P} |x_{Pij}| = 12$
1 30 1 n_P
1 | 0 1 0 ··· 0 –1 0 | 0 1 1 0 0 –1 0 0 –1 1 0 –1 0 ··· 0 1 0 –1 0 –1 0 1 0 0 –1 1 0
–1 | 0 –1 0 ··· 0 1 0 | 0 –1 –1 0 0 1 0 0 1 –1 0 1 0 ··· 0 –1 0 1 0 1 0 –1 0 0 1 –1 0
X_T: teams
$x_{Tij} \in \{-1, 0, 1\}, \sum_{j=1}^{30} x_{Tij} = 0, \sum_{j=1}^{30} |x_{Tij}| = 2$

Data



- Over 99% sparsity in the design matrix

Data



- Over 99% sparsity in the design matrix
- Highly imbalanced; only 27000 of the over 1 million possible player pairs are observed for a goal

Logistic Likelihood Model

Logistic Likelihood Model

A more specific reformulation of equation (1) is

Logistic Likelihood Model

A more specific reformulation of equation (1) is

$$log\left(\frac{q_i}{1-q_i}\right) = \boldsymbol{x}'_{T_i}\boldsymbol{\alpha} + \boldsymbol{x}'_{P_i}\boldsymbol{\beta} \tag{2}$$

Logistic Likelihood Model

A more specific reformulation of equation (1) is

$$log\left(\frac{q_i}{1-q_i}\right) = \boldsymbol{x}'_{T_i}\boldsymbol{\alpha} + \boldsymbol{x}'_{P_i}\boldsymbol{\beta} \quad (2)$$

where $\boldsymbol{\alpha}$ is the 30×1 vector of team effects, and $\boldsymbol{\beta}$ is the $n_p \times 1$ vector of player effects.

Logistic Likelihood Model

A more specific reformulation of equation (1) is

$$log\left(\frac{q_i}{1-q_i}\right) = \boldsymbol{x}'_{T_i}\boldsymbol{\alpha} + \boldsymbol{x}'_{P_i}\boldsymbol{\beta} \qquad (2)$$

where $\boldsymbol{\alpha}$ is the 30×1 vector of team effects, and $\boldsymbol{\beta}$ is the $n_p \times 1$ vector of player effects.

- Again, the $\boldsymbol{\alpha}$ vector can be extended to incorporate effects of different game situations

Bayesian approach and prior regularization

Bayesian approach and prior regularization

- Impose shrinkage on estimated coefficients by putting a zero-centered prior distribution on the parameters

Bayesian approach and prior regularization

- Impose shrinkage on estimated coefficients by putting a zero-centered prior distribution on the parameters
- Use maximum *a posteriori* (*MAP*) estimates of the unknown parameters

Bayesian approach and prior regularization

- Impose shrinkage on estimated coefficients by putting a zero-centered prior distribution on the parameters
- Use maximum *a posteriori* (*MAP*) estimates of the unknown parameters
- Regularization is needed to protect against overfitting and stability of estimates

Bayesian approach and prior regularization

- Impose shrinkage on estimated coefficients by putting a zero-centered prior distribution on the parameters
- Use maximum *a posteriori* (MAP) estimates of the unknown parameters
- Regularization is needed to protect against overfitting and stability of estimates
- Allows the model to "pick out" the most influential players by shrinking unimportant parameters toward zero

Bayesian approach and prior regularization

The following is the joint prior distribution used for α and β:

Bayesian approach and prior regularization

The following is the joint prior distribution used for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{30} N(\alpha_t | 0, \sigma_t^2) \prod_{j=1}^{n_p} Laplace(\beta_j | \lambda_j) \tag{3}$$

Bayesian approach and prior regularization

The following is the joint prior distribution used for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{30} N(\alpha_t|0, \sigma_t^2) \prod_{j=1}^{n_p} Laplace(\beta_j|\lambda_j) \tag{3}$$

Under certain settings, with *MAP* estimation...

Bayesian approach and prior regularization

The following is the joint prior distribution used for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{30} N(\alpha_t | 0, \sigma_t^2) \prod_{j=1}^{n_p} Laplace(\beta_j | \lambda_j) \tag{3}$$

Under certain settings, with *MAP* estimation...

- a Normal prior is equivalent to $L2$ regularized regression (Ridge)

Bayesian approach and prior regularization

The following is the joint prior distribution used for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{30} N(\alpha_t | 0, \sigma_t^2) \prod_{j=1}^{n_p} Laplace(\beta_j | \lambda_j) \tag{3}$$

Under certain settings, with *MAP* estimation...

- a Normal prior is equivalent to $L2$ regularized regression (Ridge)
- a Laplace prior is equivalent to $L1$ regularized regression (LASSO)

A simple explanation

A simple explanation

Let $\pi(\beta|\lambda) = \frac{\lambda}{4} e^{\frac{-\lambda}{2}|\beta|}$ and $f(y|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(y - \boldsymbol{x}'\boldsymbol{\beta})^2}$, then the *MAP* estimates are the maximum of the posterior distribution:

A simple explanation

Let $\pi(\beta|\lambda) = \frac{\lambda}{4} e^{\frac{-\lambda}{2}|\beta|}$ and $f(y|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(y - \boldsymbol{x}'\boldsymbol{\beta})^2}$, then the *MAP* estimates are the maximum of the posterior distribution:

$$f(\boldsymbol{\beta}|\boldsymbol{y}) \quad \propto \quad \prod_{j=1}^{p} \pi(\beta_j|\lambda) \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \sigma^2)$$

A simple explanation

Let $\pi(\beta|\lambda) = \frac{\lambda}{4} e^{\frac{-\lambda}{2}|\beta|}$ and $f(y|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(y - \boldsymbol{x}'\boldsymbol{\beta})^2}$, then the *MAP* estimates are the maximum of the posterior distribution:

$$
\begin{aligned}
f(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \prod_{j=1}^{p} \pi(\beta_j|\lambda) \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \sigma^2) \\
&\propto e^{\frac{-\lambda}{2} \sum_{j=1}^{p} |\beta_j|} e^{\frac{-1}{2} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2}
\end{aligned}
$$

A simple explanation

Let $\pi(\beta|\lambda) = \frac{\lambda}{4} e^{\frac{-\lambda}{2}|\beta|}$ and $f(y|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(y - \boldsymbol{x}'\boldsymbol{\beta})^2}$, then the *MAP* estimates are the maximum of the posterior distribution:

$$
\begin{aligned}
f(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \prod_{j=1}^{p} \pi(\beta_j|\lambda) \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \sigma^2) \\
&\propto e^{\frac{-\lambda}{2} \sum_{j=1}^{p} |\beta_j|} e^{\frac{-1}{2} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2} \\
\ell(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \frac{-\lambda}{2} \sum_{j=1}^{p} |\beta_j| - \frac{1}{2} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2
\end{aligned}
$$

A simple explanation

Let $\pi(\beta|\lambda) = \frac{\lambda}{4} e^{\frac{-\lambda}{2}|\beta|}$ and $f(y|\boldsymbol{\beta}, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}(y - \boldsymbol{x}'\boldsymbol{\beta})^2}$, then the *MAP* estimates are the maximum of the posterior distribution:

$$
\begin{aligned}
f(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \prod_{j=1}^{p} \pi(\beta_j|\lambda) \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \sigma^2) \\
&\propto e^{\frac{-\lambda}{2}\sum_{j=1}^{p}|\beta_j|} e^{\frac{-1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2} \\
\ell(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \frac{-\lambda}{2}\sum_{j=1}^{p}|\beta_j| - \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 \\
LASSO \rightarrow &\propto \sum_{i=1}^{n}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 + \lambda\sum_{j=1}^{p}|\beta_j|
\end{aligned}
$$

Bayesian approach and prior regularization

- σ_t and λ_j dictate the amount of penalization imposed on estimates

Bayesian approach and prior regularization

- σ_t and λ_j dictate the amount of penalization imposed on estimates
- Prior standard deviations for team-effects were set at $\sigma_t = 1$

Bayesian approach and prior regularization

- σ_t and λ_j dictate the amount of penalization imposed on estimates
- Prior standard deviations for team-effects were set at $\sigma_t = 1$
- Independent conjugate gamma hyperpriors were used for the scale parameters λ_j

Bayesian approach and prior regularization

- σ_t and λ_j dictate the amount of penalization imposed on estimates
- Prior standard deviations for team-effects were set at $\sigma_t = 1$
- Independent conjugate gamma hyperpriors were used for the scale parameters λ_j
  - $E[\lambda_j] = 15$ smallest penalty manageable while eliminating large non-zero β_j for players with little ice time

Point Estimation of Player Contribution

Point Estimation of Player Contribution

Two models considered using *MAP* estimation:

Point Estimation of Player Contribution

Two models considered using *MAP* estimation:

- Full team-player model as described in (2) and (3)

Point Estimation of Player Contribution

Two models considered using *MAP* estimation:

- Full team-player model as described in (2) and (3)
- Player-only model where ${\boldsymbol{x}'}_{T_i}\boldsymbol{\alpha}$ is replaced by a common α

Point Estimation of Player Contribution

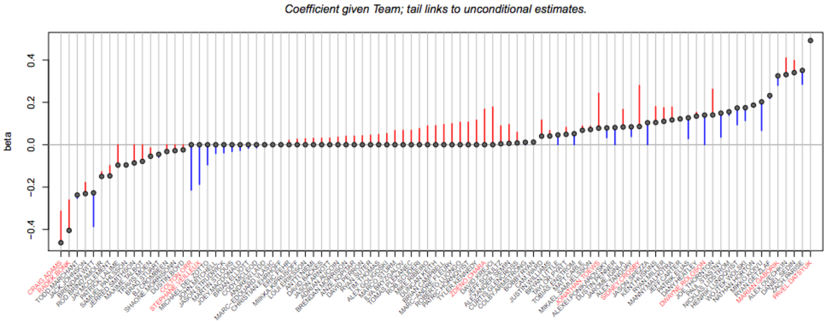


Figure 2: Comparing main effects for players in the team-augmented model (dots), to the player-only model. The lines point to the unconditional (player-only) estimates. The coefficients have been ordered by the dots. Players discussed in the text have their names colored in red. Players with coefficients estimated as zero under both models are not shown.

Point Estimation of Player Contribution

Findings:

Point Estimation of Player Contribution

Findings:

- Incorporating team effect causes more player effects to be zeroed out

Point Estimation of Player Contribution

Findings:

- Incorporating team effect causes more player effects to be zeroed out
- Sidney Crosby, Jonathan Toews', and Zdeno Chara's contributions drop after accounting for their teams (all captains)

Point Estimation of Player Contribution

Findings:

- Incorporating team effect causes more player effects to be zeroed out
- Sidney Crosby, Jonathan Toews', and Zdeno Chara's contributions drop after accounting for their teams (all captains)
- Dwayne Roloson (goaltender) is a star on bad teams

Point Estimation of Player Contribution

Findings:

- Incorporating team effect causes more player effects to be zeroed out
- Sidney Crosby, Jonathan Toews', and Zdeno Chara's contributions drop after accounting for their teams (all captains)
- Dwayne Roloson (goaltender) is a star on bad teams
- Craig Adams has the worst contribution, and is an even worse performer after accounting for his team

Point Estimation of Player Contribution

Findings:

- Incorporating team effect causes more player effects to be zeroed out
- Sidney Crosby, Jonathan Toews', and Zdeno Chara's contributions drop after accounting for their teams (all captains)
- Dwayne Roloson (goaltender) is a star on bad teams
- Craig Adams has the worst contribution, and is an even worse performer after accounting for his team
- Pavel Datsyuk is the best player by far according to this model

Point Estimation of Player Contribution

Findings:

- Incorporating team effect causes more player effects to be zeroed out
- Sidney Crosby, Jonathan Toews', and Zdeno Chara's contributions drop after accounting for their teams (all captains)
- Dwayne Roloson (goaltender) is a star on bad teams
- Craig Adams has the worst contribution, and is an even worse performer after accounting for his team
- Pavel Datsyuk is the best player by far according to this model
    - Posterior odds of contributing to a goal for his team are nearly 50% larger than the next best player

Comparison to traditional plus-minus

Comparison to traditional plus-minus

Comparison to traditional plus-minus

- Far fewer players are distinguishable from their team-average under the MAP estimation

Comparison to traditional plus-minus

- Far fewer players are distinguishable from their team-average under the MAP estimation
- Ability to check for statistical significance

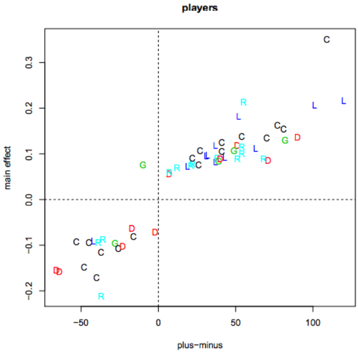
Comparison to traditional plus-minus

- Far fewer players are distinguishable from their team-average under the MAP estimation
- Ability to check for statistical significance
- Measures partial effect of a player on his respective team

Comparison to traditional plus-minus

- Far fewer players are distinguishable from their team-average under the MAP estimation
- Ability to check for statistical significance
- Measures partial effect of a player on his respective team
  - Players on good team need to be even better to get a positive β while average players on the same team may get good plus-minus

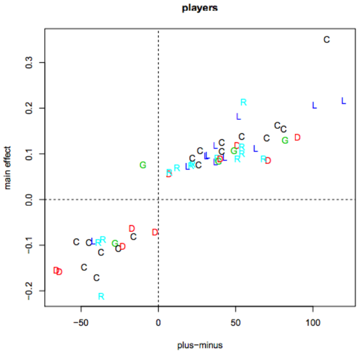
Comparison to traditional plus-minus



team	MAP	+/−
T.B	−0.165	−159
NYI	−0.126	−143
EDM	−0.077	−103
ATL	−0.045	−80
COL	−0.025	−60
OTT	−0.020	−56
MIN	−0.019	−52
TOR	−0.017	−70
NSH	0.019	9
S.J	0.045	93
BOS	0.103	109
CHI	0.106	99
PIT	0.112	114

Figure 3: *Left:* Comparing plus-minus, aggregated over the four seasons considered in our analysis, to the MAP partial effects $\hat{\beta}$. Plot symbols give positional information: C = center, L = left wing, R = right wing, D = defense, and G = goalie. *Right:* Comparing team partial effects $\hat{\alpha}$ to their plus-minus values.

Comparison to traditional plus-minus

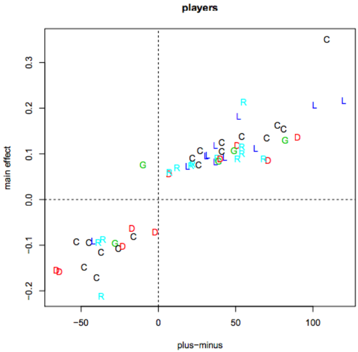


team	MAP	+/−
T.B	−0.165	−159
NYI	−0.126	−143
EDM	−0.077	−103
ATL	−0.045	−80
COL	−0.025	−60
OTT	−0.020	−56
MIN	−0.019	−52
TOR	−0.017	−70
NSH	0.019	9
S.J	0.045	93
BOS	0.103	109
CHI	0.106	99
PIT	0.112	114

Figure 3: *Left:* Comparing plus-minus, aggregated over the four seasons considered in our analysis, to the MAP partial effects $\hat{\beta}$. Plot symbols give positional information: C = center, L = left wing, R = right wing, D = defense, and G = goalie. *Right:* Comparing team partial effects $\hat{\alpha}$ to their plus-minus values.

- Dwayne Roloson was on T.B., NYI, EDM, and MIN

Comparison to traditional plus-minus



team	MAP	+/−
T.B	−0.165	−159
NYI	−0.126	−143
EDM	−0.077	−103
ATL	−0.045	−80
COL	−0.025	−60
OTT	−0.020	−56
MIN	−0.019	−52
TOR	−0.017	−70
NSH	0.019	9
S.J	0.045	93
BOS	0.103	109
CHI	0.106	99
PIT	0.112	114

Figure 3: *Left:* Comparing plus-minus, aggregated over the four seasons considered in our analysis, to the MAP partial effects $\hat{\beta}$. Plot symbols give positional information: C = center, L = left wing, R = right wing, D = defense, and G = goalie. *Right:* Comparing team partial effects $\hat{\alpha}$ to their plus-minus values.

- Dwayne Roloson was on T.B., NYI, EDM, and MIN
- This model attributes goals counting against him in his plus-minus to the teams he was on

Prior sensitivity analysis

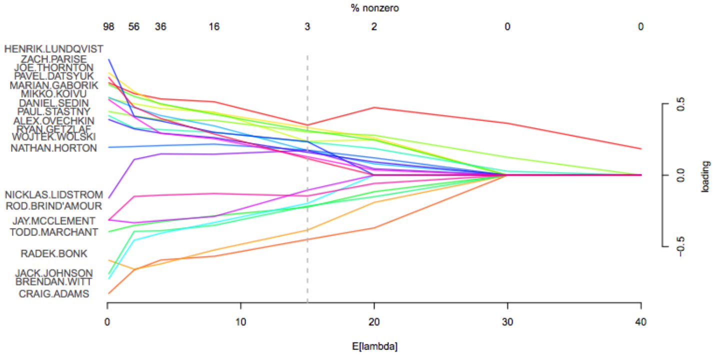


Figure 4: Coefficient estimates for a subset of players (chosen from all players with nonzero coefficients at $\mathrm{E}[\lambda_j] = 15$, our specification in Sections 3.1-2). The expected L1 penalty is shown along the bottom, with corresponding % of estimated $\beta_j \neq 0$ along the top and coefficient value on the right.

Prior sensitivity analysis

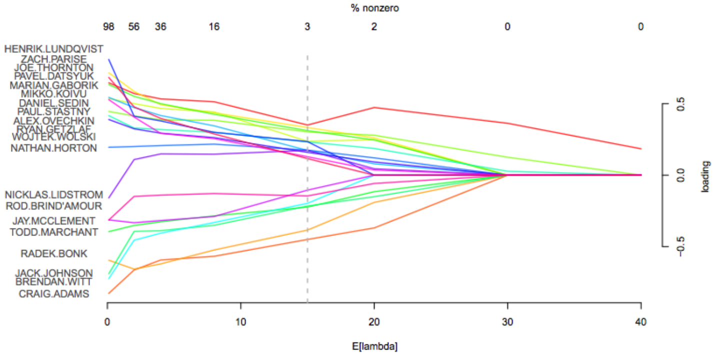


Figure 4: Coefficient estimates for a subset of players (chosen from all players with nonzero coefficients at $E[\lambda_j] = 15$, our specification in Sections 3.1-2). The expected L1 penalty is shown along the bottom, with corresponding % of estimated $\beta_j \neq 0$ along the top and coefficient value on the right.

- Averaging over penalty uncertainty will help eliminate sensitivity

Value for money

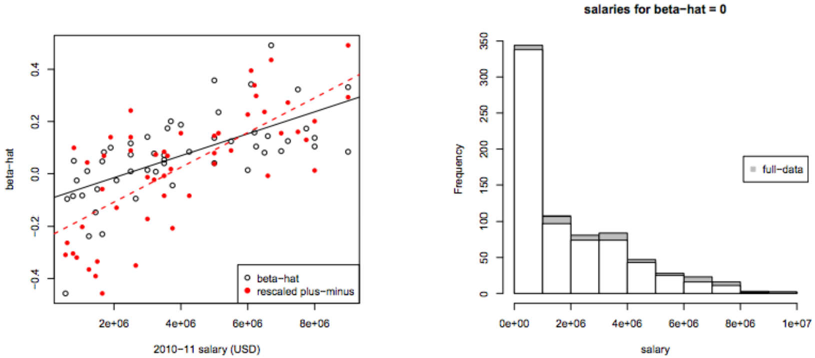


Figure 5: The left plot shows non-zero MAP $\hat{\beta}$ estimates versus 2010-11 salary, augmented with rescaled plus-minus points for comparison. Ordinary least squares fits are added to aid in visualization. The right plot shows the histogram of 2010-11 salaries for players with $\hat{\beta}_j = 0$, extending to the full set in gray.

Value for money

- Smaller standard error for slope of *MAP* estimates with salary indicates tighter relationship

Value for money

- Smaller standard error for slope of *MAP* estimates with salary indicates tighter relationship
- Slope is lesser with *MAP* estimates

Value for money

- Smaller standard error for slope of *MAP* estimates with salary indicates tighter relationship
- Slope is lesser with *MAP* estimates
- Pavel Datsyuk needed a raise whereas Sidney Crosby may be overpriced

Value for money

- Smaller standard error for slope of *MAP* estimates with salary indicates tighter relationship
- Slope is lesser with *MAP* estimates
- Pavel Datsyuk needed a raise whereas Sidney Crosby may be overpriced
- Relatively large proportion of high-paid players with $\beta_j = 0$

Value for money

- Smaller standard error for slope of *MAP* estimates with salary indicates tighter relationship
- Slope is lesser with *MAP* estimates
- Pavel Datsyuk needed a raise whereas Sidney Crosby may be overpriced
- Relatively large proportion of high-paid players with $\beta_j = 0$
    - Evgeni Malkin ($10M), Vincent Lecavalier ($10M), Duncan Keith ($9M)

Additional analyses

Additional analyses

- Posterior analyses of team-player Model

Additional analyses

- Posterior analyses of team-player Model
- Posterior player match-ups and line optimization

Additional analyses

- Posterior analyses of team-player Model
- Posterior player match-ups and line optimization
- Extension to player-player interactions

Additional analyses

- Posterior analyses of team-player Model
- Posterior player match-ups and line optimization
- Extension to player-player interactions
- Appendix with estimation details and software used to carry out this analysis

Additional analyses

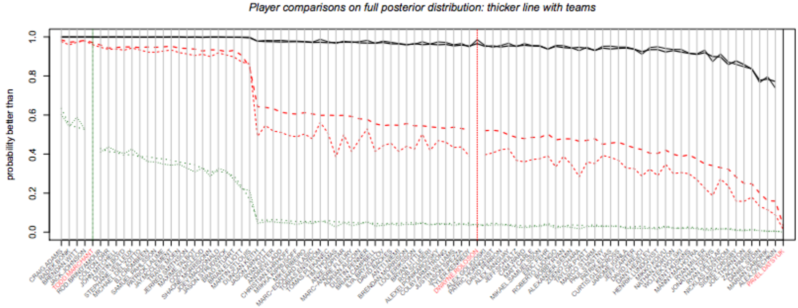


Figure 6: Comparing the ability of Datsyuk (black), Roloson (red), and Marchant (green) to the 90-odd other players with non-zero coefficients in either the team–player or player-only models. These three players are also indicated in red among the list of players on the X-axis. Thicker lines correspond to the team–player model.

Additional analyses

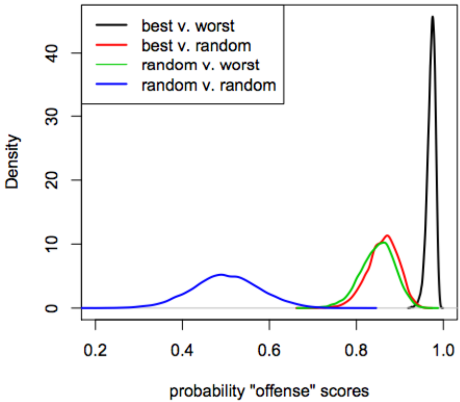


Figure 7: Posterior probability that "offense" scores in various line matchups (smoothed using a kernel density). Better team (listed first) is always considered to be the offense.

Additional analyses

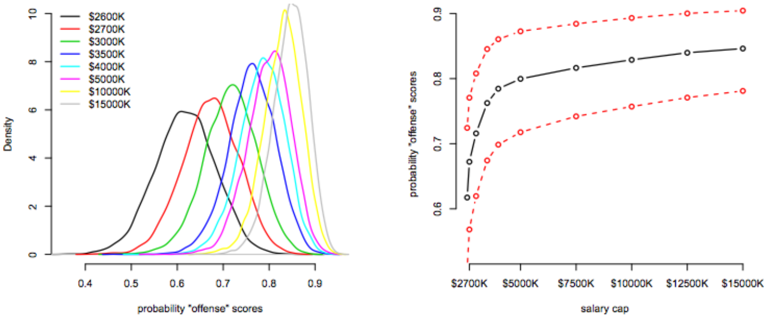


Figure 8: The *left* panel shows kernel density plots of the probability that an optimally chosen line scores against a random line according to the full posterior distribution of β and under several salary caps; the *right* panel shows the means and 90% predictive intervals of the same posterior as a function of those caps.

Reference

https://arxiv.org/pdf/1209.5026.pdf