

# Big Data: Homework 3

Will Clark & Matthew DeLio  
41201-01

April 21, 2015

## 1 Player Contribution Regression

The model for player contribution is:

$$\log \left[ \frac{Pr(y = 1)}{1 - Pr(y = 1)} \right] = \beta_0 + \alpha_{\text{team,season}} + \alpha_{\text{config}} + \sum_{\text{homeplyr}} \beta_j + \sum_{\text{awayplyr}} \beta_j$$

For a given player  $j$ , the model estimates the odds multiplier that a goal scored while player  $j$  is on the ice was scored by the his team. It includes the following two control factors:

- **team,season:** This should control for high- or low-offense years, certain arenas that provide a special home-ice advantage, or coaches that are better/worse than average; and
- **config:** This should control for disproportionate playing time in power plays or end-of-game situations where the goalie has been pulled.

To use one example, the coefficient on Alex Ovechkin is 0.30. This means that a goal scored while Ovechkin is on the ice is  $\exp(0.30) = 1.35$  times as likely to be scored by his team, the Washington Capitols, than by their opponents. Put differently, if a goal is scored while Ovi is on the ice, it is 35 percent more likely that it is scored by the Caps than by their opponents.

We can sort the array of player coefficients to determine the 10 most and least valuable players in the data set. We show the results in Table 1 and Table 2. This evaluation metric accords with our intuition about hockey. The list of 10 best players includes some of the conventionally-regarded best players of the last decade, which tells us the model is doing a reasonably good job of quantifying performance. It also includes one player (Tyler Toffoli) who has only played since 2013, although his brief career has been successful to date. Table 1 and Table 2 have a column labeled **G**, displaying the number of goals a player was on the ice for. We can see that Toffoli is indeed an inexperienced player compared to veterans like Joe Thornton and Pavel Datsyuk

By this performance metric, the best and worst players are both outliers and most players have a 0 rating. The sample includes 2439 players and only 646 have non-zero ratings (390 have net positive and 256 have net negative ratings). We can see in Figure ?? that only a small handful of players are significantly better or significantly worse than average.

	Player	Rank	$\beta_j$	$\exp(\beta_j)$	G
1	PETER FORSBERG	1	0.7548	2.1272	532
2	TYLER TOFFOLI	2	0.6293	1.8762	93
3	ONDREJ PALAT	3	0.6284	1.8746	140
4	ZIGMUND PALFFY	4	0.4427	1.5569	197
5	SIDNEY CROSBY	5	0.4131	1.5115	1568
6	JOE THORNTON	6	0.3838	1.4678	1740
7	PAVEL DATSYUK	7	0.3762	1.4567	1725
8	LOGAN COUTURE	8	0.3682	1.4451	513
9	ERIC FEHR	9	0.3677	1.4444	369
10	MARTIN GELINAS	10	0.3578	1.4301	460

Table 1: Top 10 NHL Players (2002-2014)

	Player	Rank	$\beta_j$	$\exp(\beta_j)$	G
1	RYAN HOLLWEG	2430	-0.2989	0.7417	78
2	RAITIS IVANANS	2431	-0.3129	0.7313	81
3	DARROLL POWE	2432	-0.3340	0.7161	337
4	CHRIS DINGMAN	2433	-0.3342	0.7159	30
5	MATHIEU BIRON	2434	-0.3512	0.7038	203
6	THOMAS POCK	2435	-0.3844	0.6809	131
7	NICLAS HAVELID	2436	-0.3855	0.6801	1041
8	P. J. AXELSSON	2437	-0.4284	0.6516	121
9	JOHN MCCARTHY	2438	-0.5652	0.5683	45
10	TIM TAYLOR	2439	-0.8643	0.4213	148

Table 2: Bottom 10 NHL Players (2002-2014)