

# A New Look at Adjusted Plus/Minus for Basketball Analysis

Dapo Omidiran  
University of California, Berkeley,  
Berkeley, California, USA, 94707  
Email: [dapo@eecs.berkeley.edu](mailto:dapo@eecs.berkeley.edu)

## Abstract

We interpret the Adjusted Plus/Minus (APM) model as a special case of a general penalized regression problem indexed by the parameter  $\vec{\lambda}$ . We provide a fast technique for solving this problem for general values of  $\vec{\lambda}$ . We then use cross-validation to select the parameter  $\vec{\lambda}$  and demonstrate that this choice yields substantially better prediction performance than APM.

## 1 Introduction

We model basketball as a sequence of  $N$  independent events between two teams. During each event, the home team scores  $Y_i$  net points ( $Y_i$  is actually defined slightly differently, but this is close enough for intuition.) These events are called “possessions” in basketball. For each event, we have a list of players on the floor, five for the home team and five for the away team. We also have  $p$  total players in the league. We can then represent the current players on the floor for possession  $i$  with a vector  $X_i$  of length  $p$  which is defined as follows:

$$X_{ij} = \begin{cases} 1 & \text{Player } j \text{ is on the floor for the home team} \\ -1 & \text{Player } j \text{ is on the floor for the away team} \\ 0 & \text{otherwise} \end{cases}$$

We want to define a functional relationship between  $X_i$  and  $Y_i$ , i.e., find a function  $f$  such that  $Y \approx f(X)$ . One natural way to do this is through a linear regression model, which defines the relationship

$$Y_i = \beta_{hca} + X_i^T \beta + \epsilon_i$$

Associated with each “possession” is an integer weight for each term,  $w_i$ . Thus each observation  $Y_i$ ,  $X_i$  actually represents  $w_i$  possessions.

For notational convenience, we stack the variables  $Y_i$ ,  $W_i$ , and  $\epsilon_i$  into the  $n$  vectors  $Y$ ,  $W$ , and  $\epsilon$  and the variables  $X_i$  into the  $n \times p$  matrix  $X$ .

This yields the matrix expression

$$Y = 1_n \beta_{hca} + X \beta + \epsilon \quad (1)$$

The scalar variable  $\beta_{hca}$  in Equation (1) represents a home court advantage term, while the  $p$  vector  $\beta$  is then interpreted as the net number of points each of the  $p$  players in the league “produce” per minute. Note that this is different from the number of points they directly score. The model described by Equation (1) recognizes players who by their presence on the floor, their team is more effective (e.g. players who set hard screens, play good defense, etc.)

Given observations  $(Y, X)$ , a natural technique for learning the parameters  $\beta$  is the least squares regression estimate:

$$(\hat{\beta}_{LS}, \hat{\beta}_{hca}) = \arg \min_{\beta, \beta_{hca}} \|Y - 1_n \beta_{hca} - X \beta\|_2^2 \quad (2)$$

The values  $\hat{\beta}_{LS}$  are called the adjusted plus/minus (APM) values for the players. See the website [Basketballvalue.com](http://Basketballvalue.com) [4] for computed  $\hat{\beta}_{LS}$  values for different players in the league for several recent seasons.

## 2 Improvements to APM

While algorithm (2) is a reasonable approach for estimating the parameters  $\hat{\beta}_{LS}$  and  $\hat{\beta}_{hca}$ , it doesn't take into account the following three pieces of information:

1. Sparsity. By and large, the NBA is a game dominated by elite players. Lesser players have far less impact on wins and losses. While this is not some sort of fact, it is certainly commonly accepted wisdom and thus informs player acquisitions, salaries, etc. For example, with say a \$60 million budget on players, one would much rather hire three elite \$15 million players and fill out the rest of the roster with cheap roleplayers, then spend tons of money on roleplayers and skimp on elite players (this is the road the Boston Celtics did in the summer of 2007 when trading

their roleplayers and other assets for Kevin Garnett and Ray Allen [8], and the Miami Heat did in the summer of 2010 by building a team around three star players [9].) We shall incorporate this prior information through  $\ell_1$  regularization. This effectively penalizes models that are not sparse, and should cause only elite players to stand out in the regression.

2. Box score information. Another valuable piece of information useful in inferring player worth is to take into account box score statistics. You expect good players to not only have good APM numbers, but also to generally produce assists, blocks, steals, etc. Thus, we prefer ratings vectors  $\hat{\beta}$  which are consistent with box score statistics.
3. We expect that some sort of weighted sum of player ratings should be close to zero, simply because the net number of points scored as a whole in the NBA is zero. Thus, the inner product of the vector  $v = y^T |X|$  and  $\beta$  should be close to zero.

We can encode the above prior information through the following function  $g$  defined as

$$g(\beta_{hca}, \beta, z_0, z, \beta_{median}; \vec{\lambda}) = \begin{aligned} & \|y - 1_n \beta_{hca} - x \beta\|_2^2 \\ & + \lambda_1 \|\beta - \beta_{median}\|_1 \\ & + \lambda_2 \|\beta - z_0 - Rz\|_2^2 + \lambda_3 \|z\|_1 + \lambda_4 \beta^T v v^T \beta \end{aligned}$$

where  $\vec{\lambda}$  is shorthand notation for  $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ . In the above expression

1.  $R$  is a  $p \times d$  matrix containing the box-score statistics of the  $p$  different players. There are  $d$  different box score statistics.
2. The variable  $z$  gives us weights for each of the box score statistics.
3. The variable  $z_0$  can be interpreted as an average player rating.
4. The variable  $\beta_{median}$  is used to center the betas around whatever the correct offset is.
5. The parameter  $\vec{\lambda}$  are different regularization weights.

We can then find a model consistent with both the data and the prior information by solving the following convex optimization problem:

$$\beta_{hca}, \beta, z_0, z, \beta_{median} = \arg \min g(\beta_{hca}, \beta, z_0, z, \beta_{median}; \vec{\lambda}) \quad (3)$$

We call the procedure described by Equation (3) the  $\vec{\lambda} PM$  algorithm.

One important difference between  $\vec{\lambda} PM$  and certain other approaches is that it yields both a player rating vector  $\beta$  and a box score weights vector  $z$ . Thus for player  $i$ , we can compare the variable  $\beta_i$  to the variable  $R_i z + \beta_{mean}$  to understand how “overrated” or underrated player  $i$  is relative to his box score production (note that  $R_i$  is the  $i^{th}$  row of the matrix  $R$ .) This is pretty useful, since many players produce great box score stats, but are less effective than their stats indicate due to poor defense, ballhogging, etc.

Furthermore, it is worth noting that  $\vec{\lambda} PM$  encompasses both the APM algorithm [1] and the SPM algorithm [3]. APM corresponds to the choice of  $\vec{\lambda} = 0$ , while SPM can be viewed as the limit of  $\vec{\lambda} PM$  as  $\lambda_1 \rightarrow 0$ , with  $\lambda_2, \lambda_3, \lambda_4$  fixed at zero.

For  $\vec{\lambda} PM$  to be useful, we need to be able to select the parameters  $\vec{\lambda}$  very quickly. Our general methodology is to first derive a fast algorithm for solving  $\vec{\lambda} PM$  for a fixed value of  $\vec{\lambda}$ , then use cross-validation to select the correct value of  $\vec{\lambda}$ . We can then compare the predictive power of this method to other approaches by evaluating predictive power on a test set.

## 3 Experimental Results

### 3.1 Cross Validation

We implemented the cyclical coordinate descent solver for  $\vec{\lambda} PM$  as well as a cross-validation procedure and ran it on the 2007-2008 NBA dataset. Experimentally we determined that the best values of  $\vec{\lambda}$  are

$$\vec{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (2^{15}, 2^{11}, 2^5, 1) \quad (4)$$

Our procedure for selecting this value was to start from the vector  $\vec{\lambda}_0 = (1, 1, 1, 1)$ , and then for each of the four components  $\lambda_1, \lambda_2, \lambda_3$  and  $\lambda_4$ , iterate over the following sequence of values

$$\lambda_{first} = \{(2^k, 1, 1, 1)\}_{k=-4}^{kmax}$$

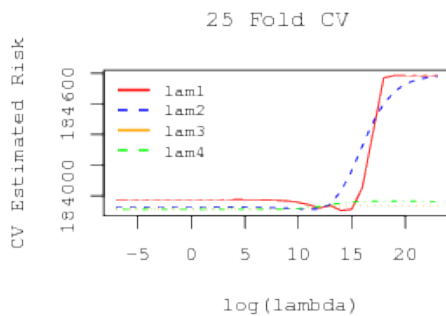
$$\lambda_{second} = \{(1, 2^k, 1, 1)\}_{k=-4}^{kmax}$$

$$\lambda_{third} = \{(1, 1, 2^k, 1)\}_{k=-4}^{kmax}$$

$$\lambda_{fourth} = \{(1, 1, 1, 2^k)\}_{k=-4}^{kmax}$$

where  $2^{kmax}$  is a value that causes the optimization variables to all be zero.

For each value  $\lambda_i$ , we get the following plot:



**Figure 1:** 25 fold CV, exponential grid in  $\lambda_i$ , all other components fixed. Best value roughly  $(2^{15}, 2^{11}, 2^5, 1)$

Again, the above plots were generated with  $\vec{\lambda}_0 = (1, 1, 1, 1)$ . We can take the best point on each plot to obtain the parameter vector  $\vec{\lambda}_1 = (2^{15}, 2^{11}, 2^5, 1)$ , and re-run to get similar plots. While this doesn't necessarily mean that  $\vec{\lambda}_1$  is the best possible point, it is suggestive. It seems plausible that a rigorous justification for this heuristic could be made using convex analysis.

### 3.2 Predictive Power

$\vec{\lambda} PM$  with the choice  $\vec{\lambda} = (2^{15}, 2^{11}, 2^5, 1)$  fits a particular model. But how do we evaluate this model? In classical linear regression, there are various goodness-of-fit tests [5] that we can do, which are valid under certain assumptions on the underlying model (e.g., gaussianity.)

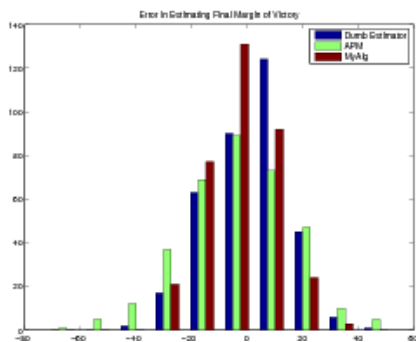
Another choice is to test predictive power on a sample of games. To be explicit, we can estimate the 1230 NBA regular season games, build a model on say the first 800, then evaluate the model's ability to predict the final margin of victory on the remaining 430 games.

We compare the following three estimates:

1. A dummy estimator that predicts that the home team wins each possession by 3.5 points.
2. The APM estimate, which computes player ratings using least squares and uses those in predicting the margin of victory for the home team.
3. The  $\vec{\lambda} PM$  estimate with the choice of parameters  $\vec{\lambda} = (2^{15}, 2^{11}, 2^5, 1)$ . Recall that this choice of parameters was obtained via cross-validation in the previous subsection.

Once we train the above models, we get an estimate of the net number of points scored by the home team per possession. Since we don't know a priori the number of possessions that will be involved in the game, we need an estimate of this. A very crude estimate is just the average number of possessions used per game in the 2007-2008 season, which is 92.4 [2].

Figure 3.2 is a plot of the resulting errors over the course of the last 348 games of the 2007-2008 NBA season, using the 882 games that came before it for cross-validation.



**Figure 2:** Comparison of Home Court Advantage estimate, APM estimate, and  $\vec{\lambda} PM$  estimate.

It is pretty clear from the above that  $\vec{\lambda} PM$  produces better estimates than APM. We will also look at some statistical properties of the error variable for each game:

	HCA	APM	$\vec{\lambda} PM$
Fraction of Games Guessed Wrong	0.3793	0.3333	0.3017
Mean of Absolute Error	11.54	15.83	10.06
Variance of AE	70.38	149.02	58.61
Median of AE	9.50	13.88	9.10
Min of AE	0.3	0.00	0.00
Max of AE	48.30	72.64	33.50
Empirical $P(AE > 1)$	0.95	0.95	0.92
Empirical $P(AE > 3)$	0.89	0.87	0.81
Empirical $P(AE > 5)$	0.76	0.81	0.66
Empirical $P(AE > 10)$	0.49	0.61	0.45

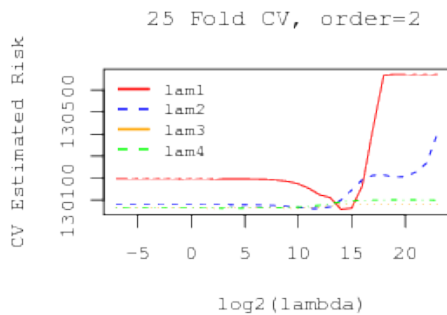
As the above table indicates, this method has substantially better predictive power than either the dummy model or the APM approach. As an aside, while the above two plots might reflect poorly on APM, it doesn't necessarily mean that APM is useless. Part of the problem with using pure APM is that it is simply too sensitive to outliers. A player who plays only one minute of the first 882 games, but very good things happen in that minute might have an outrageously high  $\beta_i$  associated with him. If he then starts playing major minutes later in the season, then his high rating will cause lots of errors on games. It is likely that a Huber loss function might lead to better results for APM than the quadratic loss.

### 3.3 Polynomials in $R$

We also took a look at different polynomial transformations of the box-score matrix  $R$ . It appears that different transformations of these variables seems to improve performance. For example, we can expand the matrix  $R$  to

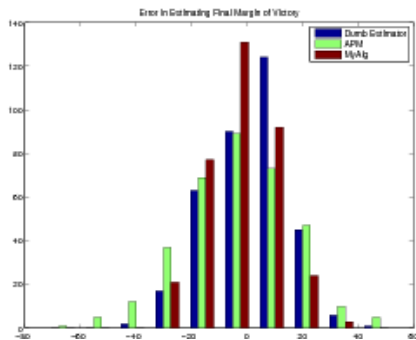
include all pairwise product of the basic variables. If  $R$  is say a  $p$  by 14 matrix, this leads to a  $p$  by  $14 + \binom{14}{2}$  matrix  $\tilde{R}^2$ .

### 3.3.1 $\tilde{R}^2$



**Figure 3:** 25 fold CV using  $\tilde{R}^2$  exponential grid in  $\lambda_i$ , all other components fixed. Best value roughly  $(2^{14}, 2^{12}, 2^2, 2^2)$ .

This yields the following plot



**Figure 4:** Comparison of Home Court Advantage estimate, APM estimate, and  $\vec{\lambda} PM$  estimate with matrix  $\tilde{R}^2$ .



It isn't completely clear from the above plot that  $\vec{\lambda}PM$  is the best. We examine the statistical properties of the absolute error variable over the block of games we want to predict:

	HCA	APM	$\vec{\lambda}PM$
Fraction of Games Guessed Wrong	0.3793	0.3333	0.2874
Mean of Absolute Error	11.54	15.83	10.58
Variance of AE	70.38	149.02	65.07
Median of AE	9.50	13.88	9.10
Min of AE	0.3	0.00	0.00
Max of AE	48.30	72.64	41.02
Empirical $P(AE > 1)$	0.95	0.95	0.94
Empirical $P(AE > 3)$	0.89	0.87	0.80
Empirical $P(AE > 5)$	0.76	0.81	0.70
Empirical $P(AE > 10)$	0.49	0.61	0.46

This more convincingly demonstrates the superiority of  $\vec{\lambda}PM$  with the expanded box score matrix  $\tilde{R}^2$ .

## 4 Concluding Remarks

We have introduced  $\vec{\lambda}PM$ , a powerful new statistical inference procedure for the NBA, which contains APM as a special case. We found a fast, iterative numerical algorithm for solving the convex optimization problem defining  $\vec{\lambda}PM$ . We then utilize cross-validation to perform parameter selection. Finally, we compare the statistical performance of our approach to APM and find that the performance is dramatically improved.



## References

- [1] 82games.com. Adjusted plus/minus. <http://www.82games.com/ilardi1.htm>.
- [2] basketball reference.com. Pace estimate. [http://www.basketball-reference.com/leagues/NBA\\_2008.html](http://www.basketball-reference.com/leagues/NBA_2008.html).
- [3] basketball reference.com. Statistical plus/minus. <http://www.basketball-reference.com/blog/?p=1351>.
- [4] basketballvalue.com. Player ratings. <http://www.basketballvalue.com>.
- [5] Julian J. Faraway. *Linear Models with R*. Chapman & Hall/CRC, Boca Raton, FL, 2004. ISBN 1-584-88425-8.
- [6] Jerome Friedman, Trevor Hastie, Holger Höfling, and Robert Tibshirani. Pathwise coordinate optimization. *Annals of Applied Statistics*, 2:302–332, 2007.
- [7] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.
- [8] NBA.com. Garnett Trade. <http://www.nba.com/celtics/news/press073107-garnett.html>.
- [9] sun sentinel.com. LeBron James, Chris Bosh join the Heat. [http://articles.sun-sentinel.com/2010-07-08/sports/sfl-lebron-james-decision-070810\\_1\\_james-and-bosh-dwyane-wade-lebron-james](http://articles.sun-sentinel.com/2010-07-08/sports/sfl-lebron-james-decision-070810_1_james-and-bosh-dwyane-wade-lebron-james).
- [10] Robert Tibshirani, Iain Johnstone, Trevor Hastie, and Bradley Efron. Least angle regression. *The Annals of Statistics*, 32(2):407–499, April 2004.
- [11] Tong Tong Wu and Kenneth Lange. Coordinate descent algorithms for lasso penalized regression. *Annals of Applied Statistics*, 2:224–244, 2008.