Applied Regression Analysis

PREDICTING SUCCESS IN THE NBA

Final Project

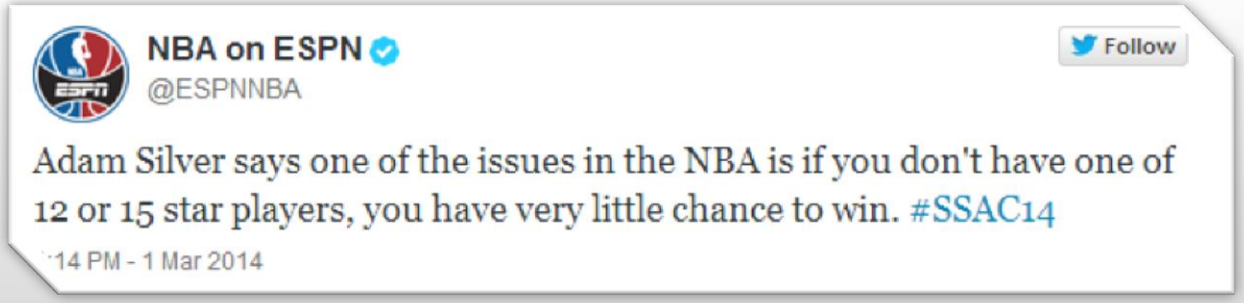
Damani Aaron, Sunanda Kameswaran, Matthew Leo, Debby Man,
Shantala Menon, Lucas Mohr, Gavin Mohrman

Date: March 4, 2014

PREDICTING SUCCESS IN THE NBA

Today, the 30 teams are worth $19 billion- yes that is 19 and 9 zeroes after that- There is no room for any mistakes in NBA drafting"
- Kurt Badenhausen, 2014

"Stern estimates NBA revenue up 20 percent to a record $5B"- *Forbes, 2013*



THE PROBLEM

Joe Alexander, No. 8 (2008)
10.8 Rookie PER
Played 2 seasons

Hasheem Thabet, No. 2 (2009)
12.9 Rookie PER
SI described “Benchwarmer”

Adam Morrison, No. 3 (2006)
7.9 Rookie PER
Left the league in 2010

Rafael Araujo, No. 8 (2004)
7.1 Rookie PER
Three year washout

WHAT IS THIS "PER"?

PER = Player Efficiency Rating

- Basketball stats are typically position-specific (assists, rebounds, shots scored, etc)
  - Difficult comparing box scores (20 points/10 rebounds vs. 12 points/10 assists/3 rebounds?)
- PER combines traditional stats into a single number that measures a player's per-minute productivity
  - Standardizes the comparison of player performance by adjusting for pace of play and setting league average to 15.0
- We used players' PER in third NBA season as a proxy for long-term production (some survivorship bias)

Full calculation at http://en.wikipedia.org/wiki/Player_efficiency_rating#Calculation

Project Background

Can we predict which college basketball players will have a high PER in the NBA?

Assumption & Hypothesis

- A players performance in the NBA can be summarized and measured by their Player Efficiency Rating
- An optimal combination of college stats and physical characteristics will have predictive power for college players who go on to the NBA

Applications

- Help team owners determine draft strategy based on expected contribution to team

Data Used

- We used key college stats for the 60 picks in each of the 2007, 2008, and 2009 drafts
- Data was gathered from a variety of sources, including ncaa.com, espn.com, and nba.com

What We Analyzed

We started (where else?) at the beginning with descriptive statistics.

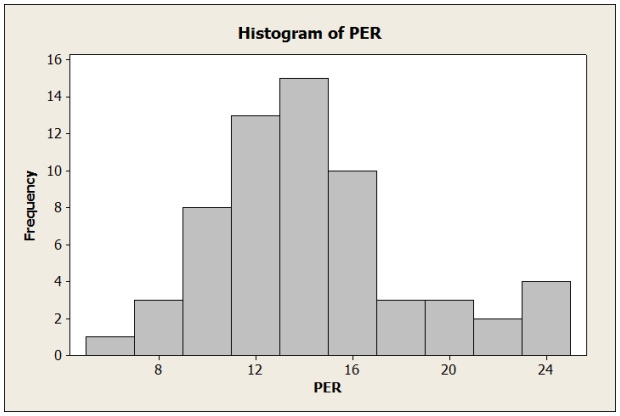
26 independent variables (13 continuous and 13 dummy variables):

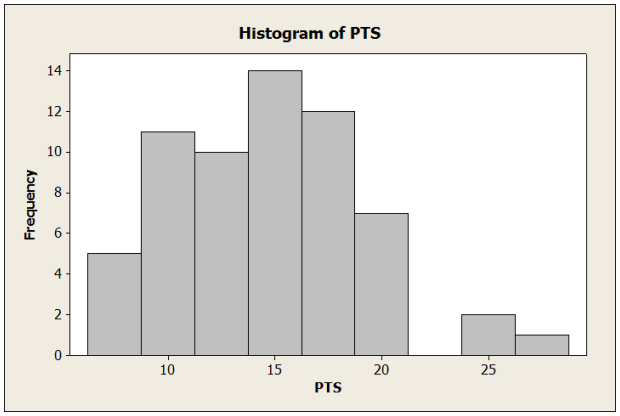
- Games played (G) – does experience count?
- Minutes (MP), points (PTS), rebounds (TRB), assists (AST), steals (STL), blocks (BLK), and turnovers (TOV) – per game production
- Field goal (FG%), three point (3P%) and free throw (FT%) percentages – efficiency
- We also looked at height and weight, position played, and conference (America's team plays in the Pac-12, fka Pac-10 as of our data set)

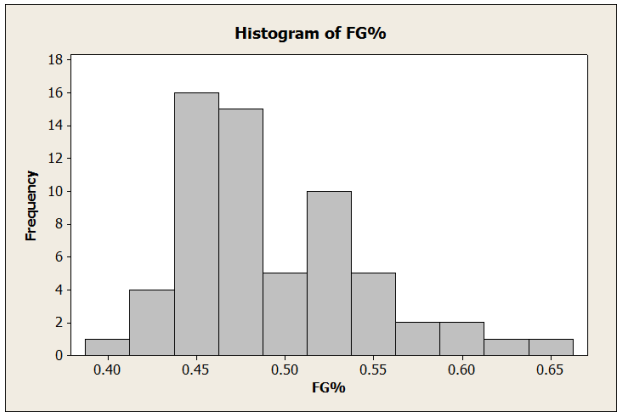
	PER	G	MP	FG%	3P%	FT%	TRB	AST	STL	BLK	TOV	PTS	Height	Weight
Mean	14.2	85	29.2	49.3%	33.8%	73.3%	5.7	2.3	1.2	0.8	2.4	14.7	78	217
StDev	4.1	38	4.7	5.2%	13.5%	8.1%	2.3	1.5	0.5	0.6	0.6	4.4	4	27
Min	6.7	30	16.6	40.0%	0.0%	51.4%	2.3	0.3	0.3	-	1.3	6.8	71	160
Max	24.4	145	36.8	63.8%	100.0%	89.0%	12.4	6.2	2.6	2.3	3.6	26.6	86	278

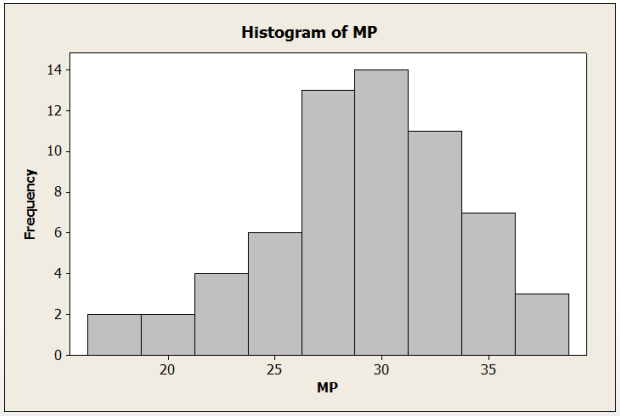
“We should always plot histograms of the *y* and *x* variables”

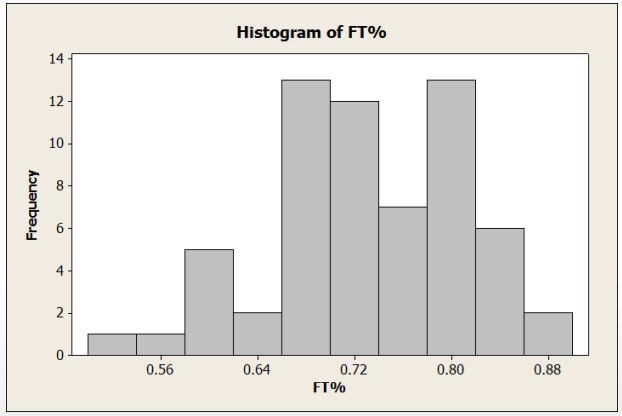
Most of our key variables showed some skewness, but PER is relatively normal.

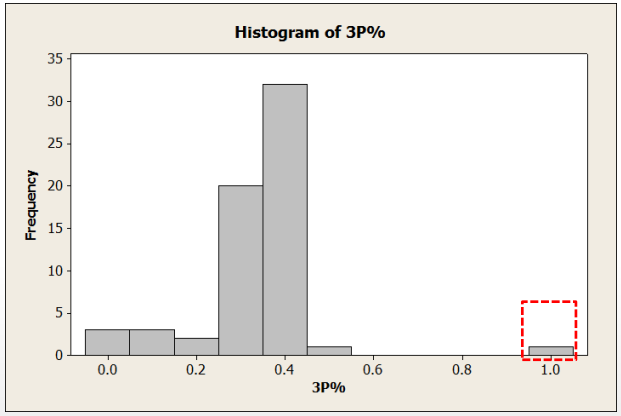






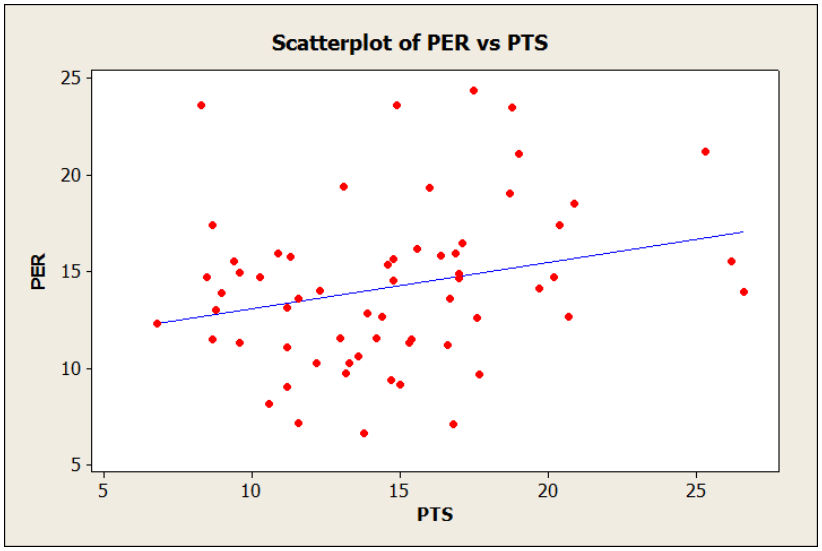


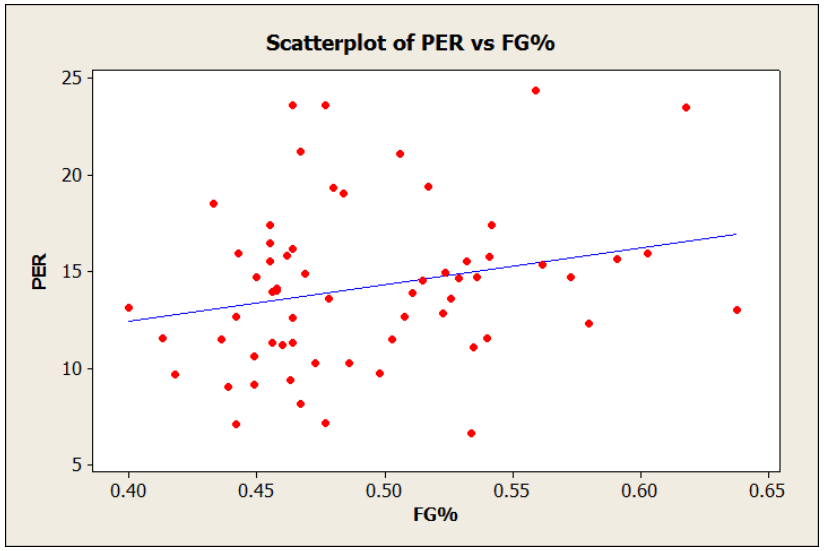


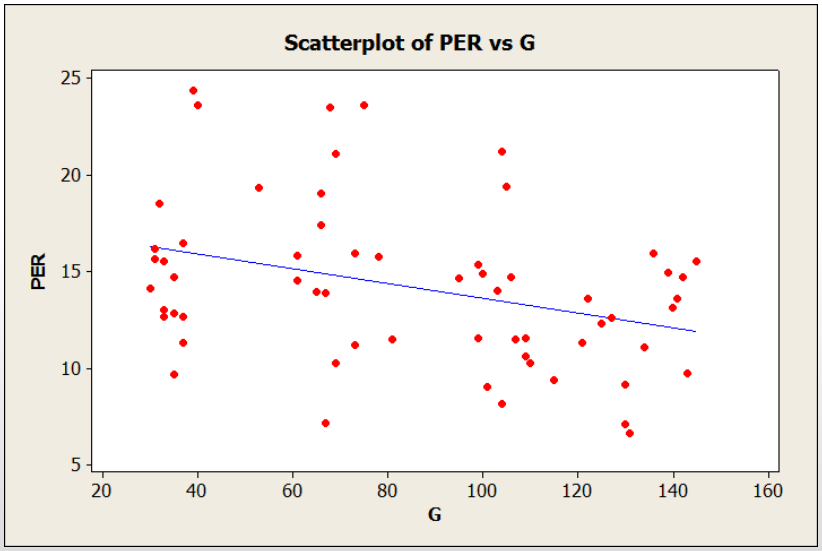


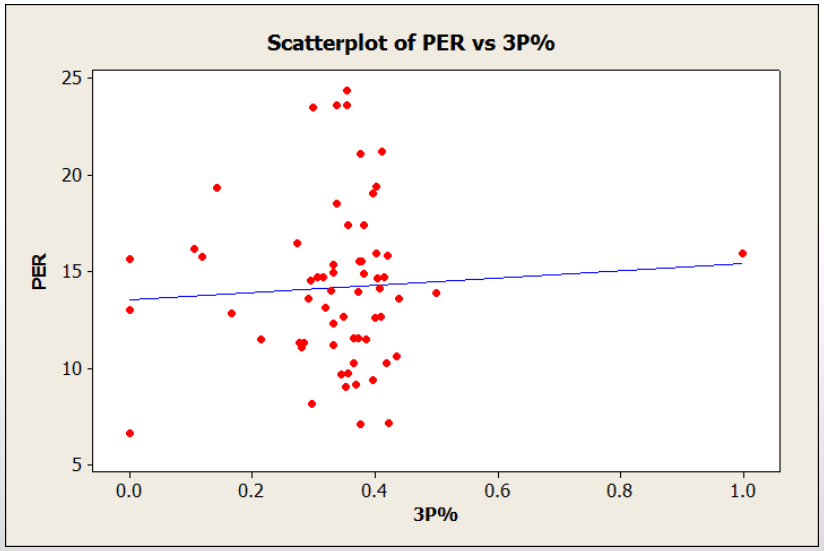
The Relationship Between PER and Variables

Some interesting relationships emerge – especially between PER and Games.









Issues With Collinearity

"We must be aware of any potential problems - and be vigilant!"

- "Position-based correlations" arise (e.g. TRB, BLK, AST, STL vs. HT, WT)
  - Shorter, lighter players with more AST and STL; taller heavier with more TRB, BLK
- VIFs will confirm if there are any issues

	PER	*G*	*MP*	*FG%*	*3P%*	*FT%*	*TRB*	*AST*	*STL*	*BLK*	*TOV*	*PTS*	*Height*	*Weight*
PER	1.00	(0.34)	0.04	0.24	0.06	0.04	0.17	0.11	0.03	0.09	0.16	0.25	0.02	0.06
G	(0.34)	1.00	(0.23)	0.02	0.27	0.01	(0.15)	0.02	0.08	(0.25)	(0.33)	(0.33)	(0.12)	(0.09)
MP	0.04	(0.23)	1.00	(0.44)	0.03	0.44	0.12	0.51	0.51	(0.27)	0.65	0.73	(0.40)	(0.33)
FG%	0.24	0.02	(0.44)	1.00	(0.13)	(0.49)	0.43	(0.49)	(0.43)	0.35	(0.43)	(0.19)	0.54	0.61
3P%	0.06	0.27	0.03	(0.13)	1.00	0.34	(0.27)	0.21	0.10	(0.04)	(0.03)	0.07	(0.12)	(0.15)
FT%	0.04	0.01	0.44	(0.49)	0.34	1.00	(0.32)	0.43	0.47	(0.41)	0.31	0.54	(0.56)	(0.53)
TRB	0.17	(0.15)	0.12	0.43	(0.27)	(0.32)	1.00	(0.46)	(0.25)	0.52	(0.01)	0.31	0.63	0.60
AST	0.11	0.02	0.51	(0.49)	0.21	0.43	(0.46)	1.00	0.67	(0.57)	0.53	0.26	(0.81)	(0.72)
STL	0.03	0.08	0.51	(0.43)	0.10	0.47	(0.25)	**0.67**	1.00	(0.47)	0.54	0.43	(0.68)	(0.64)
BLK	0.09	(0.25)	(0.27)	0.35	(0.04)	(0.41)	0.52	(0.57)	(0.47)	1.00	(0.17)	(0.11)	0.78	0.61
TOV	0.16	(0.33)	**0.65**	(0.43)	(0.03)	0.31	(0.01)	0.53	0.54	(0.17)	1.00	0.60	(0.40)	(0.41)
PTS	0.25	(0.33)	**0.73**	(0.19)	0.07	0.54	0.31	0.26	0.43	(0.11)	**0.60**	1.00	(0.22)	(0.15)
Height	0.02	(0.12)	(0.40)	0.54	(0.12)	(0.56)	**0.63**	**(0.81)**	**(0.68)**	**0.78**	(0.40)	(0.22)	1.00	0.84
Weight	0.06	(0.09)	(0.33)	**0.61**	(0.15)	(0.53)	**0.60**	**(0.72)**	**(0.64)**	**0.61**	(0.41)	(0.15)	**0.84**	1.00

Full Model Output (*first attempt*)

"Always start with the full model"

- Using 26 variables, including dummy variables for position and conferences (treating international experience as a conference), we still have 35 degrees of freedom with n=62
- **THE GOOD**: adjusted R^2 of 90.7% and regression p-value of 0.000
- **THE BAD**: the p-values for most variables are relatively high (>0.1), indicating that having 26 variables might give the highest adjusted R^2, but a simpler model may prove better
- **and THE UGLY**: Rampant multicollinearity with relatively few variables with VIFs < 5. AND Minitab kicked out Pac-10 dummy variable because it couldn't estimate a coefficient
- To confirm what the best variables are, we rerun the model with the "Best Subsets" function in Minitab

Full Model Output (*first attempt*)

"Always start with the full model"

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1.26401	94.49%	90.67%	*

Coefficients

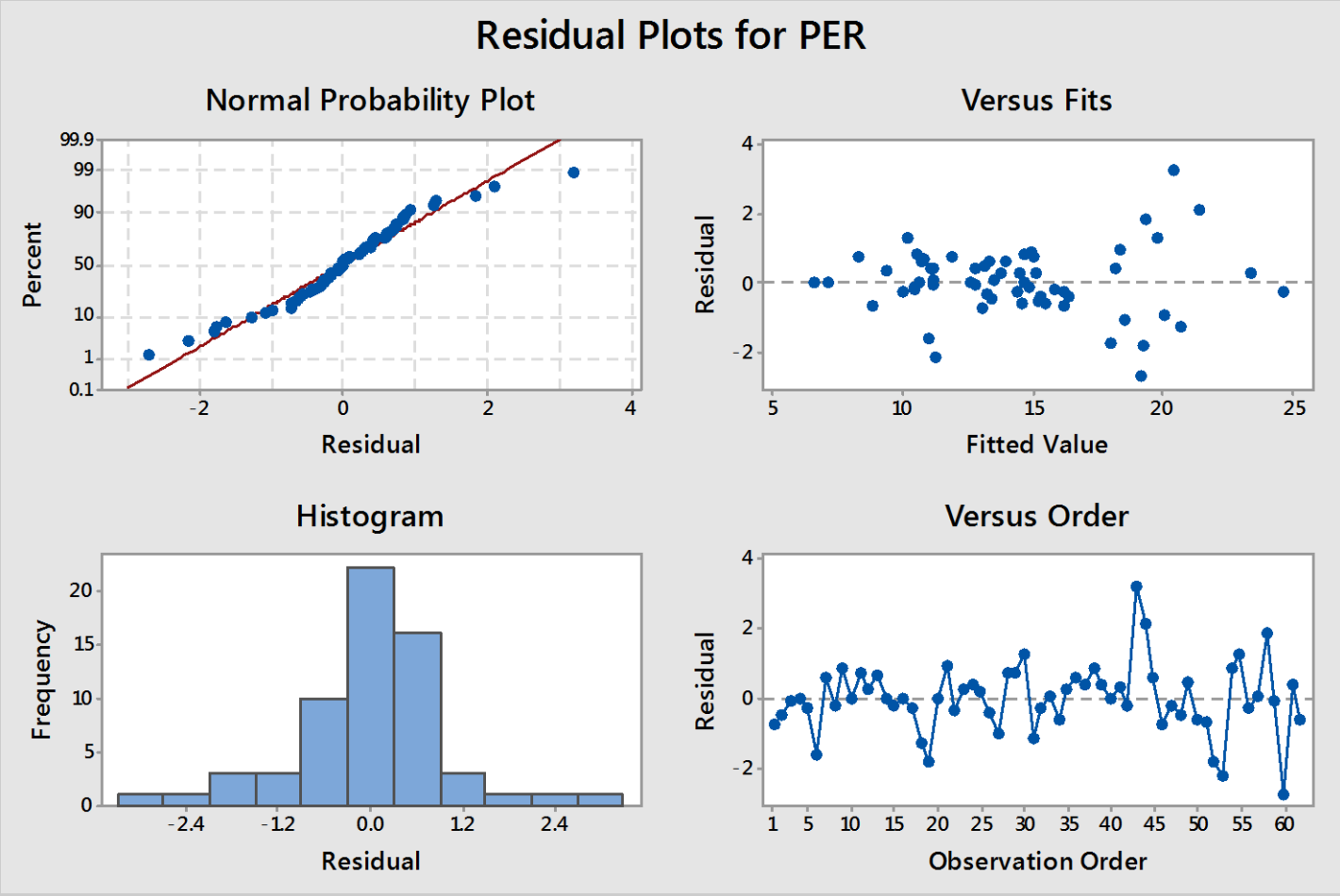
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.1	12.5	0.00	0.996	
G	-0.00196	0.00626	-0.31	0.756	2.12
MP	-0.0288	0.0719	-0.40	0.691	4.27
FG%	8.35	5.32	1.57	0.125	2.91
3P%	0.71	1.72	0.41	0.682	2.07
FT%	0.74	3.76	0.20	0.845	3.55
TRB	0.269	0.174	1.55	0.131	6.17
AST	0.369	0.257	1.44	0.159	5.55
STL	-0.433	0.532	-0.81	0.422	2.88
BLK	-0.809	0.559	-1.45	0.157	4.82
TOV	-0.149	0.532	-0.28	0.781	3.34
PTS	-0.0212	0.0960	-0.22	0.826	6.74
Height	0.030	0.153	0.20	0.845	12.02
Weight	-0.0048	0.0147	-0.32	0.749	6.08
Forward	-0.396	0.746	-0.53	0.599	5.37
Center	-0.23	1.44	-0.16	0.875	5.94
A-10	17.56	1.84	9.55	0.000	4.10
ACC	13.76	1.70	8.10	0.000	17.51
Big 12	9.95	1.75	5.69	0.000	16.05
Big East	8.57	1.69	5.07	0.000	13.75
Big Ten	7.45	1.66	4.49	0.000	9.34
CAA	7.73	2.32	3.33	0.002	3.33
CUSA	6.64	1.84	3.60	0.001	4.12
International	5.05	1.63	3.10	0.004	18.86
MAAC	3.70	1.82	2.03	0.049	4.00
MWC	2.30	1.98	1.16	0.253	4.75

Regression Equation

PER = 0.1 - 0.00196 G - 0.0288 MP + 8.35 FG% + 0.71 3P% + 0.74 FT% + 0.269 TRB + 0.369 AST - 0.433 STL - 0.809 BLK - 0.149 TOV - 0.0212 PTS + 0.030 Height - 0.0048 Weight - 0.396 Forward - 0.23 Center + 17.56 A-10 + 13.76 ACC + 9.95 Big 12 + 8.57 Big East + 7.45 Big Ten + 7.73 CAA + 6.64 CUSA + 5.05 International + 3.70 MAAC + 2.30 MWC

Full Model Output (*first attempt*)

"Always start with the full model"



Best Subsets (*first attempt*)

"Multicollinearity is a weapon against our enemies" and sometimes against ourselves

* ERROR * Predictor columns are highly correlated. Use REGR command to find correlated variables.
* ERROR * Completion of computation impossible.

- Minitab says *"woa! best subsets can't help you here…try again later after you've culled some of these multicollinear variables"*
- We revised the full model before trying again
  - Eliminated height and weight due to low correlation to PER and high multicollinearity
  - Reduced 10 conference dummy variables to one (NonMaj) because of high multicollinearity. Now equal to 1 if player comes from a non-major conference and 0 if player comes from ACC, Big 12, Big East, Big Ten, Pac 10, or SEC (kept International dummy variable)

Full Model Output (*second attempt*)

"Always start with the full model"

- Using 15 variables, including dummy variables for position and conference, we have 47 degrees of freedom with n=62
- **THE GOOD**: regression p-value of 0.002, limited multicollinearity with only PTS borderline with a VIF of 5.77
- **THE BAD**: adjusted R^2 came down to 33.8% and several variables still have p-values >0.1 suggesting a further slimmed down model may prevail
- **and THE UGLY**: Actually nothing too horrible!
- To confirm what the best variables are, we rerun the model with the "Best Subsets" function in Minitab

Full Model Output (*second attempt*)

"Always start with the full model"

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	15	522.55	34.836	3.07	0.002

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.36725	50.05%	33.76%	0.00%

Coefficients

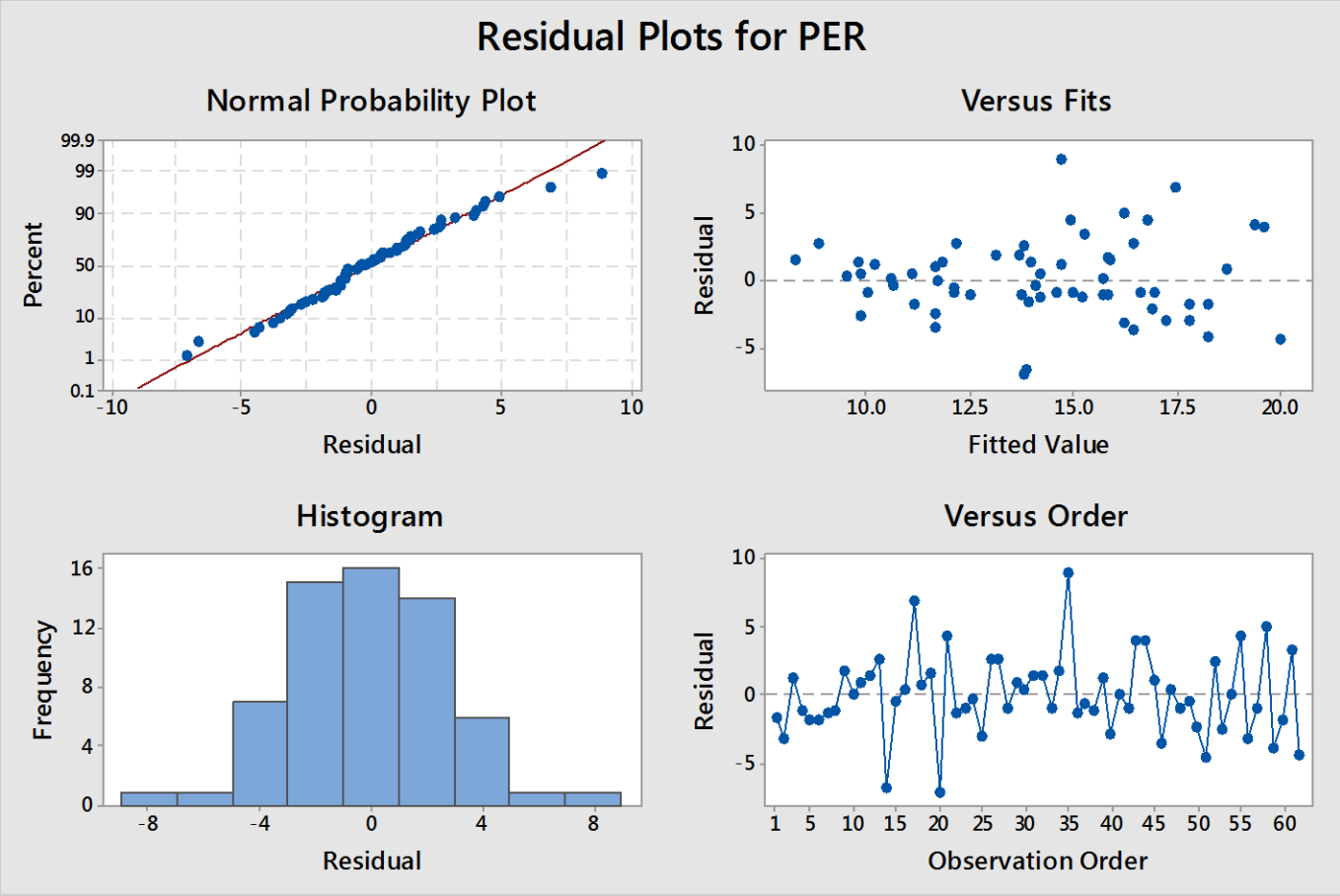
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	14.2	11.9	1.20	0.238	
G	-0.0351	0.0150	-2.33	0.024	1.72
MP	-0.326	0.178	-1.84	0.073	3.68
FG%	11.9	13.1	0.91	0.368	2.49
3P%	4.82	4.07	1.19	0.242	1.63
FT%	2.50	9.06	0.28	0.784	2.90
TRB	0.613	0.393	1.56	0.126	4.43
AST	0.951	0.556	1.71	0.094	3.66
STL	-0.76	1.35	-0.56	0.575	2.60
BLK	-0.61	1.11	-0.55	0.587	2.69
TOV	-0.03	1.29	-0.02	0.983	2.76
PTS	0.085	0.237	0.36	0.721	5.77
Forward	-1.76	1.71	-1.03	0.309	4.00
Center	-2.01	2.73	-0.74	0.464	3.01
NonMaj	-2.33	1.38	-1.69	0.099	1.30
International	-4.21	1.15	-3.65	0.001	1.33

Regression Equation

PER = 14.2 - 0.0351 G - 0.326 MP + 11.9 FG% + 4.82 3P% + 2.50 FT% + 0.613 TRB + 0.951 AST - 0.76 STL - 0.61 BLK - 0.03 TOV + 0.085 PTS - 1.76 Forward - 2.01 Center - 2.33 NonMaj - 4.21 International

Full Model Output (*second attempt*)

"Always start with the full model"



Best Subsets (*second attempt*)

"Run Best Subsets and call it a day"

- Three models maximize adjusted R^2 at 39.0% (one with 8 variables and two with 9)
- However, one of the 9 variable models makes the most sense in the context of the data
  - Values PTS and FG% over position designation and BLK

Best Subsets (*second attempt*)

"Run Best Subsets and call it a day"

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	G	MP	FG%	3P%	FT%	TRB	AST	STL	BLK	TOV	PTS	Forward	Center	NatMan	International
1	23.6	22.3	20.1	12.4	3.6461															X
1	11.9	10.4	6.4	23.1	3.9157	X														
2	29.9	27.5	24.5	8.6	3.5221	X														X
2	29.4	27.0	18.4	9.0	3.5345														X	X
3	34.6	31.2	22.0	6.3	3.4325	X													X	X
3	33.4	30.0	26.3	7.3	3.4623	X						X								X
4	40.0	35.8	30.0	3.2	3.3147	X		X				X								X
4	37.4	33.0	23.1	5.6	3.3862	X											X		X	X
5	41.7	36.5	25.4	3.7	3.2965	X		X				X							X	X
5	41.6	36.4	28.6	3.8	3.2999	X		X	X			X								X
6	43.9	37.7	28.1	3.7	3.2648	X	X	X				X				X				X
6	43.6	37.5	23.5	3.9	3.2719	X		X	X			X							X	X
7	45.9	38.9	24.1	3.8	3.2339	X	X	X				X				X			X	X
7	45.1	37.9	21.3	4.6	3.2592	X	X		X		X	X							X	X
8	47.0	39.0	21.9	4.8	3.2319	X	X	X	X			X				X			X	X
8	46.7	38.7	20.1	5.1	3.2402	X	X		X		X	X				X			X	X
9	48.0	39.0	18.3	5.9	3.2313	X	X	X	X		X	X				X			X	X
9	48.0	39.0	17.6	5.9	3.2315	X	X		X		X	X		X			X		X	X
10	48.8	38.7	13.1	7.2	3.2382	X	X	X	X		X	X					X	X	X	X
10	48.7	38.6	14.9	7.3	3.2416	X	X	X	X		X	X		X		X			X	X
11	49.3	38.2	8.9	8.7	3.2536	X	X	X	X		X	X		X			X	X	X	X
11	49.2	38.0	8.3	8.8	3.2572	X	X	X	X	X	X	X					X	X	X	X
12	49.6	37.3	3.4	10.4	3.2771	X	X	X	X	X	X	X		X			X	X	X	X
12	49.6	37.3	4.2	10.4	3.2773	X	X	X	X		X	X		X		X	X	X	X	X
13	50.0	36.4	2.5	12.1	3.2994	X	X	X	X		X	X	X	X		X	X	X	X	X
13	49.9	36.3	1.7	12.1	3.3014	X	X	X	X	X	X	X	X	X			X	X	X	X
14	50.0	35.2	0.0	14.0	3.3313	X	X	X	X	X	X	X	X	X		X	X	X	X	X
14	50.0	35.1	0.0	14.1	3.3340	X	X	X	X		X	X	X	X	X	X	X	X	X	X
15	50.0	33.8	0.0	16.0	3.3673	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Reduced Model

"Explain why your model is more complicated than the solar system's"

- Using 9 variables, including dummy variables for NonMaj and International, we have 53 degrees of freedom with n=62
- **THE GOOD**: regression p-value of 0.000, limited multicollinearity with no VIF > 5; maximized adjusted R^2 while using a simpler and more intuitive model
- **THE BAD**: not every p-value is significant at 0.05 alpha level
- **and THE UGLY**: nothing!

Reduced Model

"Explain why your model is more complicated than the solar system's"

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	501.18	55.69	5.33	0.000

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.23126	48.00%	39.00%	18.26%

Coefficients

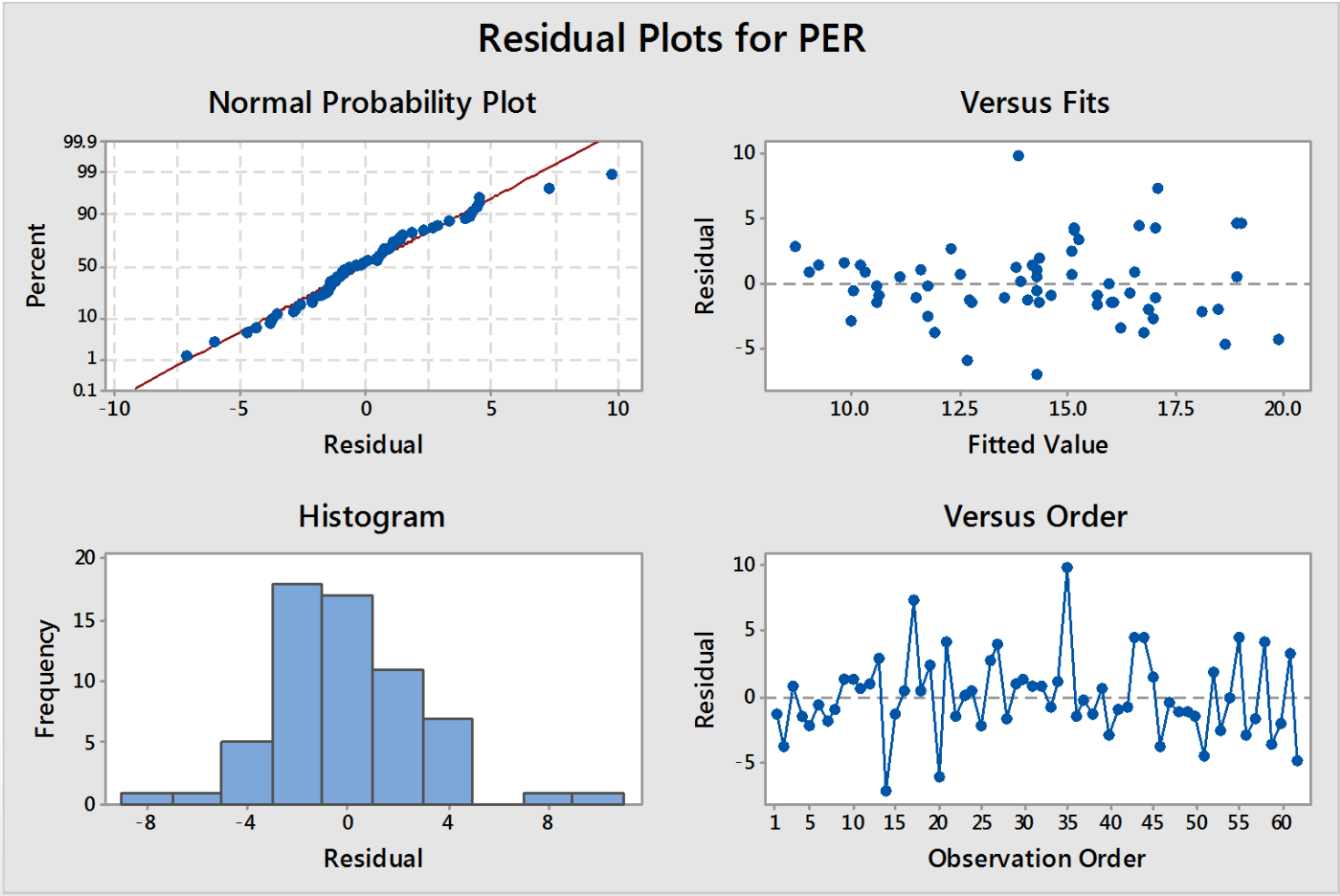
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	13.15	7.66	1.72	0.092	
G	-0.0323	0.0125	-2.59	0.013	1.29
MP	-0.333	0.167	-1.99	0.052	3.55
FG%	12.6	11.1	1.14	0.261	1.94
3P%	4.24	3.42	1.24	0.222	1.26
TRB	0.275	0.272	1.01	0.317	2.30
AST	1.118	0.419	2.67	0.010	2.26
PTS	0.201	0.156	1.29	0.202	2.71
NonMaj	-2.05	1.30	-1.58	0.121	1.25
International	-3.97	1.08	-3.68	0.001	1.27

Regression Equation

PER = 13.15 - 0.0323 G - 0.333 MP + 12.6 FG% + 4.24 3P% + 0.275 TRB + 1.118 AST
+ 0.201 PTS - 2.05 NonMaj - 3.97 International

Reduced Model

"Explain why your model is more complicated than the solar system's"



Key Findings

Results not perfect but still valuable

Can provide valuable data-driven insights to personnel officers evaluating prospects

- Be skeptical of international players and “Cinderellas” from non-major conferences
- Avoid upperclassmen – they stayed in college that long for a reason
- Efficiency is key – based on per minute production and shooting percentages
- Should value distributors of the ball regardless of position

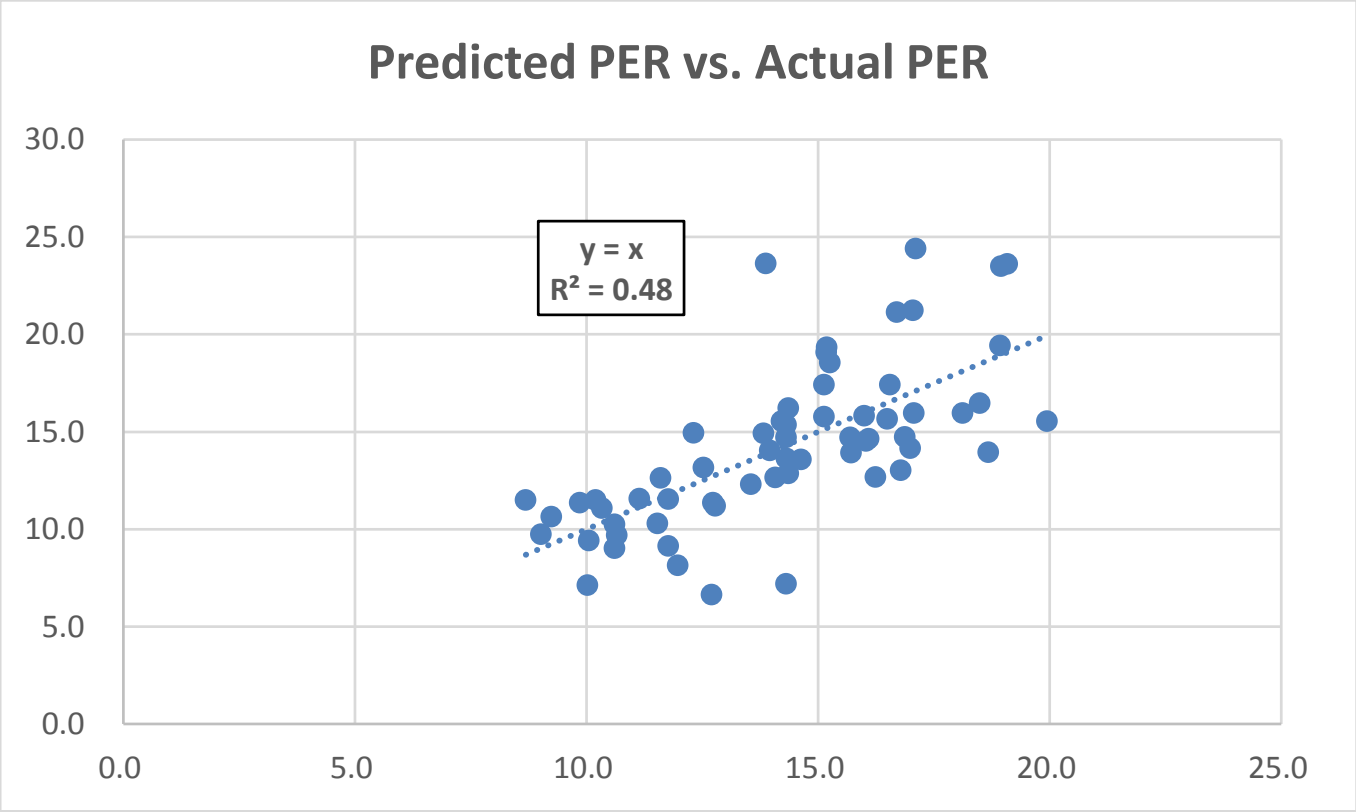
IDEAL PLAYER TRAITS

Passing	Court Vision	Shooting
Dishing the rock	*Orchestra Conductors*	*Overall shooting %*

Validating The Model

Comparing PER to PER-hat

Model's predicted PER with 48% correlation to actual PER



Validating The Model

2009 Draft

Our model seemed to work the best for the most recent draft we used in dataset

- Average absolute difference from actual PER ranking was nearly half a spot lower than actual draft position
- Over long term could save money and improve team performance

					Error vs. Actual PER Ranking			
		Rank by:			+/- spots:		absolute spots	
Pick #	Team	Draft Pick	Predicted PER	3rd Year PER	Draft	Model	Draft	Model
1	LAC	Blake Griffin	Blake Griffin	Blake Griffin	-	-	0	0
2	MEM	Hasheem Thabeet	Jonny Flynn	Stephen Curry	-5	-3	5	3
3	OKC	James Harden	James Harden	James Harden	-	-	0	0
4	SAC	Tyreke Evans	Tyreke Evans	Brandon Jennings	-6	-8	6	8
5	MIN	Ricky Rubio	Stephen Curry	Tyreke Evans	+1	+1	1	1
6	MIN	Jonny Flynn	DeMar DeRozan	Jordan Hill	-2	-3	2	3
7	GSW	Stephen Curry	Tyler Hansbrough	Tyler Hansbrough	-6	-	6	0
8	NYK	Jordan Hill	Gerald Henderson	Ricky Rubio	+3	-5	3	5
9	TOR	DeMar DeRozan	Jordan Hill	Gerald Henderson	-3	+1	3	1
10	MIL	Brandon Jennings	Hasheem Thabeet	Terrence Williams	-1	-1	1	1
11	NJN	Terrence Williams	Terrence Williams	DeMar DeRozan	+2	+5	2	5
12	CHA	Gerald Henderson	Brandon Jennings	Hasheem Thabeet	+10	+2	10	2
13	IND	Tyler Hansbrough	Ricky Rubio	Jonny Flynn	+7	+11	7	11
						Avg	**3.5**	**3.1**

Validating The Model

2008 Draft

About even with draft performance here

					Error vs. Actual PER Ranking			
		Rank by:			+/- spots:		absolute spots	
Pick #	Team	Draft Pick	Predicted PER	3rd Year PER	Draft	Model	Draft	Model
1	CHI	Derrick Rose	Michael Beasley	Kevin Love	-4	-1	4	1
2	MIA	Michael Beasley	Kevin Love	Russell Westbrook	-2	-5	2	5
3	MIN	O.J. Mayo	D.J. Augustin	Derrick Rose	+2	-1	2	1
4	SEA	Russell Westbrook	Derrick Rose	Brook Lopez	-6	-5	6	5
5	MEM	Kevin Love	Jerryd Bayless	Eric Gordon	-2	-3	2	3
6	NYK	Danilo Gallinari	O.J. Mayo	D.J. Augustin	-3	+3	3	3
7	LAC	Eric Gordon	Russell Westbrook	Danilo Gallinari	+1	-5	1	5
8	MIL	Joe Alexander	Eric Gordon	Michael Beasley	+6	+7	6	7
9	CHA	D.J. Augustin	Brook Lopez	Jerryd Bayless	-2	+4	2	4
10	NJN	Brook Lopez	Joe Alexander	Jason Thompson	-2	-3	2	3
11	IND	Jerryd Bayless	Brandon Rush	O.J. Mayo	+8	+5	8	5
12	SAC	Jason Thompson	Danilo Gallinari	Brandon Rush	-1	+1	1	1
13	POR	Brandon Rush	Jason Thompson	Joe Alexander	+5	+3	5	3
						Avg	**3.4**	**3.5**

Validating The Model

2007 Draft

Not as good here

					Error vs. Actual PER Ranking			
		Rank by:			+/- spots:		absolute spots	
Pick #	Team	Draft Pick	Predicted PER	3rd Year PER	Draft	Model	Draft	Model
1	POR	Greg Oden	Mike Conley	Kevin Durant	-1	-4	1	4
2	SEA	Kevin Durant	Joakim Noah	Greg Oden	+1	-4	1	4
3	ATL	Al Horford	Yi Jianlian	Al Horford	-	-9	0	9
4	MEM	Mike Conley	Spencer Hawes	Joakim Noah	-5	+2	5	2
5	BOS	Jeff Green	Kevin Durant	Acie Law	-6	-5	6	5
6	MIL	Yi Jianlian	Greg Oden	Thaddeus Young	-6	-3	6	3
7	MIN	Corey Brewer	Brandan Wright	Mike Conley	+3	+6	3	6
8	CHA	Brandan Wright	Julian Wright	Jeff Green	+3	-5	3	5
9	CHI	Joakim Noah	Thaddeus Young	Spencer Hawes	-1	+5	1	5
10	SAC	Spencer Hawes	Acie Law	Yi Jianlian	+4	+7	4	7
11	ATL	Acie Law	Corey Brewer	Corey Brewer	+4	-	4	0
12	PHI	Thaddeus Young	Al Horford	Julian Wright	-1	+4	1	4
13	NOK	Julian Wright	Jeff Green	Brandan Wright	+5	+6	5	6
						Avg	3.1	4.6

Validating The Model

Just for fun (because we couldn't get enough)!

Not the best predictive power when looking at some of the greatest players…

- However, players penalized for playing three or four seasons back in the day – adjusted to 29 games shows better accuracy
- We did predict Steve Nash would not be good early in his career – hit for being non-major
- Also, predicted that Iverson, Allen (dead-on!), and Durant would all be above-average
  - Predicted all would be above-average using adjusted PER-hat
- Completely off the mark on Jordan

	G	MP	FG%	3P%	TRB	AST	PTS	NonMaj	Int	PER-hat	PER - 3rd	Adjusted PER-hat
Michael Jordan	101	34	54.0%	44.7%	5.0	1.8	17.7	0	0	14.2	29.8	16.6
Ray Allen	101	28.6	48.7%	44.8%	6.0	2.4	19.0	0	0	16.6	18.9	18.9
Allen Iverson	67	32.5	44.0%	31.4%	3.6	4.6	23.0	0	0	17.8	22.2	19.0
Steve Nash	113	30.1	43.0%	40.0%	3.1	4.5	14.9	1	0	13.4	10.9	16.1
Tim Duncan	128	35.1	57.7%	32.1%	12.3	2.3	16.5	0	0	15.2	24.8	18.4
Kevin Durant	35	35.9	47.3%	40.4%	11.1	1.3	25.8	0	0	17.4	26.2	17.6

Implications for 2014

Using ESPN's Lottery Mock Draft top 14 players and stats as of 3/2/2014.

Joel Embiid
Kansas
Freshman

Andrew Wiggins
Kansas
Freshman

Dante Exum
Australia

Jabari Parker
Duke
Freshman

Julius Randle
Kentucky
Freshman

Marcus Smart
Oklahoma State
Sophomore

Aaron Gordon
Arizona
Freshman

Tyler Ennis
Syracuse
Freshman

Noah Vonleh
Indiana
Freshman

Dario Saric
Croatia

Rodney Hood
Duke
Sophomore

Jerami Grant
Syracuse
Sophomore

Gary Harris
Michigan State
Sophomore

Doug McDermott
Creighton
Senior

Implications for 2014

Large potential for busts in the "deepest draft in years."

- Potential for Andrew Wiggins and Dante Exum to be huge busts
  - Exum hurt by being International; Saric not as much
- Noah Vonleh and Doug McDermott being underrated – good potential upside!
- "Deepest draft in years" doesn't have any projected "dominant" players like Lebron

	Rank by:			
Pick #	**ESPN Picks**	**Model Picks**	**+/- Spots**	**PER-hat**
1	Joel Embiid	Joel Embiid	0	18.9
2	Andrew Wiggins	Noah Vonleh	-6	18.0
3	Dante Exum	Jabari Parker	-11	17.4
4	Jabari Parker	Julius Randle	1	17.1
5	Julius Randle	Tyler Ennis	1	16.7
6	Marcus Smart	Marcus Smart	0	16.5
7	Aaron Gordon	Aaron Gordon	0	15.8
8	Tyler Ennis	Andrew Wiggins	3	15.2
9	Noah Vonleh	Doug McDermott	7	14.9
10	Dario Saric	Jerami Grant	-3	14.7
11	Rodney Hood	Gary Harris	-1	14.1
12	Jerami Grant	Rodney Hood	2	14.0
13	Gary Harris	Dario Saric	2	11.6
14	Doug McDermott	Dante Exum	5	11.0

Discriminant Analysis

Can we predict All-NBA team members based on college statistics?

62cases were analyzed – 53 were correct = 85.5%

- Non All-NBA Team: 49 out of 57 were put into true group – 8 put into "All-NBA team" who didn't make it
- All-NBA Team: 4 out of 5 were put into true group – 1 put into "Non All-NBA Team" who did make it

Disappointing results…but maybe not:

- 4 of the 12 put into the "All-NBA Team" actually made the team = 33% accuracy
- However, only 15 players each year and dominated by small group – typically only 4-5 first-time selections per year

	True Group	
Put into Group	0	1
0	49	1
1	8	4
Total N	57	5
N correct	49	4
Proportion	0.860	0.800

N = 62 N Correct = 53 Proportion Correct = 0.855

Conclusions

Was model successful, accurate, useful- Yes, Yes and Yes

Limitations

- Our reduced model is only able to explain about 48% of the player's PER value. Hence there is still some room for other considerations and improvements.

Risks of using model

- As stated before, we used players' PER in third NBA season as a proxy for long-term production. This can lead to some survivorship bias.

Areas for further analysis

- Add additional athletic performance abilities – e.g. speed, vertical jump, bench press
- Check strength of schedule in college
- Isolate performance in March Madness tournament (higher pressure conditions against the best in the nation)

Since no model is 100% accurate, we believe that our reduced model is a great place to start and can be tweaked over the years to come, especially as some of the draft lottery rules are changed.