Variable Selection and Employee Performance Evaluation

Yi Tan

University of Kansas

Wei Xiong

University of Cincinnati

oal of our research

e propose a decision support framework that combines
riable selection techniques with stochastic frontier models
r evaluating employees. Differing to conventional methods
performance evaluation of employees, we evaluate them
ased on some organization-specific performance evaluation
etrics

oal of our research

e apply our idea into National Basketball Association (NBA) ams' players recruitment. We will act ourselves as the role someone who provide a service of evaluating or comparing ayer's performance for a team's coach.

Outlines

- Introduction
- Sports strategies
- Player performance measurement
- Stochastic frontier analysis (SFA)
- Empirical result
- conclusion

ntroduction

ur work is motivated by the well-known discovered field, rategic human resource management.

- As the business environment becomes more competitive, firms' human resources become more important to firm success (Wright, McMahan 2011).
- strategic management research has been extended through discussions of the resource-based approach (Barney, 1991; Mahoney, Pandian, 1992)
- Based on the assumption that firms competing in the same industries are homogeneous, individual firms are unique and composed of distinct bundles of resource (Wright, Smart, McMahan, 1995)

ntroduction

am managers perform trading in order to improve their am performance.

Steve Nash

Personal: 17.7 points, 7.2 assists/game--->18.6 points, 11.6 assists/game

Team: 29 wins/53 loses--->62 wins/20 loses

Steve Francis

Personal: 21 points, 6.2 rebounds 6.2 assists/game--->11.2 points/game

Team: 39 wins/43 loses--->33 wins/49 loses

ports strategies

ow should a team coach or a team manager evaluate ayers' performance?

- A coach may be more interested in players who can efficiently understand and execute his preferred team's game strategy

ports strategies

ight, P. M., et al. (1995). "Matches between human resources and ategy among NCAA basketball teams." Academy of Management urnal **38**(4): 1052-1074.

Summary: it examined the relationships among strategy, human resources, and performance among National Collegiate Athletic Association (NCAA) basketball teams. Based on their survey data, they indicated that coaches' preferred strategies influence the characteristics that they look for in recruits. Also, teams implementing a strategy different from a coach's preferred strategy performed less well than those implementing the preferred strategy

ome more recent articles

- Berger and Pope (2011) showed large and significant effects of being slightly behind an opponent increased success.
- Dobson and Goddard announced strategic choices, such as defensive, attacking, non-violent, and violent, which influence the probabilities of scoring and conceding goals at the current stage of the match and the probabilities that players are dismissed.
- Goldman and Rao (2011) found that players overall adhere quite closely to the theoretical predictions; overall they are suburb optimizers.
- Annis (2005) analyzed optimal end game strategy and found that intentionally fouling the opponent increases the chances of eventually wining the game.

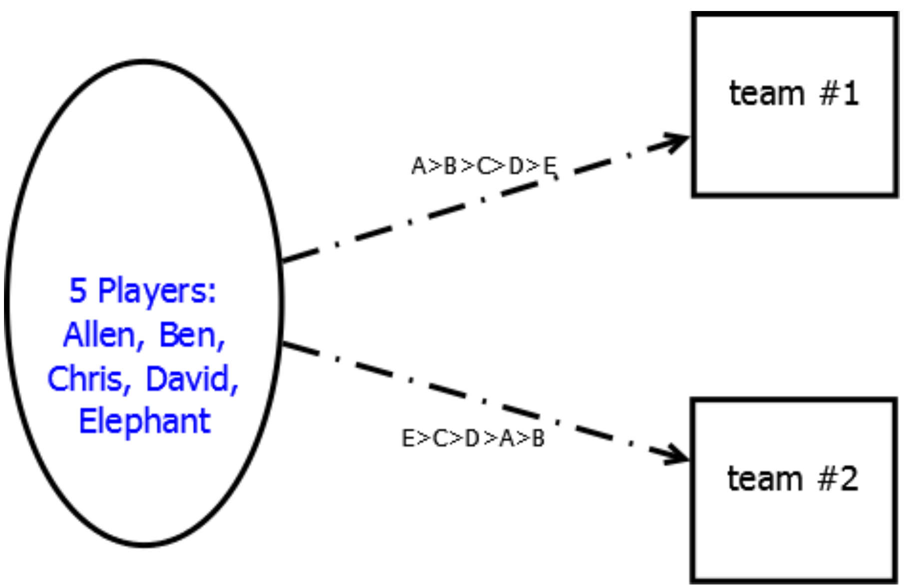
ports strategies examples

- ***full-court press:*** a full-court press is an attacking full-court defense with the purpose of trying to force a turnover or accelerate the pace of the game.
- ***Run-and-Gun*****:** Some teams like to push the ball up the floor and take the first possible shot.
- ***pick-and-roll:*** an offensive play where a player first sets a pick for his teammate who has the ball, then moves towards the basket (or "rolls" to the basket) to receive a pass

ck to 2004 – 2008 NBA seasons, Phoenix Suns played a fast
eak strategy (Pick n’ Roll), and highly focused in offense.
owever, San Antonio Spurs put more weight in defense,
ayed a relative slow offensive strategy, such as Post-Up.

Why game strategy matters?

If different teams use different game strategies, they would not have the same measurement of performance for targeted players. They need to recruit players who are most suitable/fitting for their game strategies.



How to learn game strategy?

ports strategies

- We use Generalized linear model (logistic regression) to analyze the game strategy for teams.
- Lasso variable selection method is applied to identify the significant features.

 Recent development in variable selection literature suggests a promising role
 penalized shrinkage approaches (Tibshirani, 1996, 2011; Zou, 2006; Meier et
, 2008), which select predictive variables through shrunken coefficients under
re-specified roughness penalty.

- We want to seek players who can avoid the negative effect and improve the positive effect for team's wins

ata structure:

Team datasets are come from *basketball-reference*

Player datasets are come from *NBAstuffer*

target	Number of variables	variables
player	26	Date, age, Opp, home/away, win/loss, GS, MP, FG, FGA, FG%, 3P, 3PA, 3P%, FT, FTA, FT%, ORB, DRB, TRB, AST, STL, BLK, TOV, PF, PTS, GmSc, +/-
team	37	Date, home/away, Opp, win/loss, FG, FGA, FG%, 2P, 2PA, 2P%, 3P, 3PA, 3P%, FT, FTA, ORB, DRB, TRB, AST, STL, BLK, TOV, PF, PTS, FG_opp, FGA_opp, 2P_opp, ……, PTS_opp
Player& Team	20	Date, home/away, Opp, win/loss, FG, FGA, FG% 2P, 2PA, 2P%, 3P, 3PA, 3P%, FT, FTA, ORB, DRB, TRB, AST, STL, BLK, TOV, PF, PTS

emove features

Non-strategic and redundant features: date, home/away, Opp”

Features that can be deduced: 2P, 2PA, 2P%, FG%,3P%, TRB

Points

13 independent strategic features remained: FG, FGA, 3P, 3PA, FT, FTA, ORB, DRB, AST, STL, BLK, TOV, PF

FG: field goal	FGA: field goal attempt
3P: three point	3PA: three point attempt
FT: free throw	FTA: free throw attempt
ORB: offensive rebound	DRB: defensive rebound
AST: assist	STL: steal
BLK: block	TOV: turnover
PF: personal foul	

ports strategies

ven 13 explanatory features and one response (win/lose), e use logistic regression with LASSO selection to learn the rategy of a specific team:

$$Y = \text{logit}((p(y=1)|\boldsymbol{\beta})) = \text{logit}(\frac{p}{1-p}) = \beta_0 + \boldsymbol{\beta}\mathbf{X} + \varepsilon$$

y=1: win
y=0: lose

ports strategies

	intercept	FG	FGA	3P	3PA	FT	FTA	OR	DR	A	PF	ST	TO	BL
Atlanta	-11.5272	0.28062	-0.13303	0.067178	-0.0078	0.069908	0	0.233469	0.237571	0.052672	0	0.183754	-0.14472	0.149318
Boston	-7.96456	0.207938	-0.08547	0.039006	0	0.048848	0	0.041621	0.139262	0.122289	-0.03975	0.074993	-0.05404	0.008436
Brooklyn	-10.0596	0.253388	-0.14802	0.203545	-0.03317	0.011562	0.075101	0.177306	0.279456	0.055407	-0.05987	0.18864	-0.13552	0.035018
Charlotte	-13.5027	0.25996	-0.0512	0	0	0.107051	0	0	0.158155	0.003694	0	0.083455	0	0.008384
Chicago	-5.2558	0.172905	-0.12319	0.071103	0	0.075773	0	0.100131	0.237236	0.066674	-0.06434	0.131353	-0.12217	0.019388
Cleveland	-5.9173	0.184822	-0.10045	0.107805	0	0.069039	0	0.100394	0.171258	0.042859	-0.08359	0.125282	-0.1499	0.072376
Dallas	-6.35954	0.232802	-0.10806	0.013432	0	0.100878	0.004599	0.094688	0.131196	0.093117	-0.0456	0.047207	-0.12312	0
Denver	-13.4294	0.118747	-0.01716	0.13372	-0.02465	0	0.088606	0.007142	0.145257	0.158	-0.05317	0.097473	-0.04331	0.127891
Detroit	-3.25235	0.305642	-0.18779	0.083436	-0.02311	0.056858	0.014762	0.099309	0.303919	0	-0.1624	0.115391	-0.17392	0
Golden State	-7.35825	0.179434	-0.11537	0.064275	-0.02403	0.099365	0	0.138387	0.192556	0.112485	-0.08048	0.150951	-0.11604	0.122296
Houston	-0.40923	0.210291	-0.149	0.113809	-0.06142	0.014627	0	0.091563	0.215247	0.013658	-0.14511	0.115949	-0.05707	0.024198
Indiana	-4.54646	0.225483	-0.18096	0.047632	-0.03108	0.052561	0	0.157902	0.241345	0.140157	-0.03441	0.251242	-0.2137	0.092342
LA Clippers	-5.20824	0.215793	-0.10522	0.019981	0	0.034722	0	0.026566	0.164516	0.038476	-0.02616	0.179945	-0.15647	0.091125
LA Lakers	-7.23694	0.164114	-0.06315	0.035021	0	0.015407	0	0	0.180361	0.097355	-0.09718	0.136381	-0.07169	0.032006
Memphis	-2.24869	0.182938	-0.11515	0.116814	0	0.023771	0	0.074038	0.17295	0.001423	-0.08857	0.134067	-0.09688	0.102286
Miami	-8.50905	0.420593	-0.21825	0.155666	-0.00517	0.226836	-0.1113	0.160231	0.252541	0.086514	-0.09696	0.29578	-0.20046	0.072157
Milwaukee	-7.62113	0.275252	-0.13263	0.115412	-0.00689	0.152497	0	0.058613	0.111091	0.030298	-0.02296	0.162948	-0.06532	0.097449
Minnesota	3.227333	0.226966	-0.22101	0.29281	-0.0723	0.062749	0	0.171417	0.194454	0.020913	-0.13406	0.157034	-0.20998	0.04756
New Orleans	-4.27599	0.234623	-0.1575	0	0	0.03816	0	0.1045	0.21674	0.048704	-0.02954	0.237711	-0.21351	0.095883
New York	-0.31087	0.231109	-0.17921	0.156925	0	0.003212	0	0.072183	0.17972	0.015178	-0.02675	0.201699	-0.15978	0
Oklahoma City	-5.55657	0.250723	-0.12173	0.238825	-0.06554	0.115032	0	0.091548	0.175492	0	-0.08914	0.190635	-0.14505	0
Orlando	-6.82862	0.14358	-0.12361	0.229993	0	0.100778	0	0.063308	0.223585	0.077535	-0.05434	0.10996	-0.09336	0.031917
Philadelphia	-6.1391	0.245496	-0.1296	0.091646	0	0.050185	0	0	0.220589	0.050042	-0.07464	0.118591	-0.16458	0.182379
Phoenix	-4.87813	0.166147	-0.1241	0.241902	-0.06023	0.071777	0	0.11055	0.242506	0.013362	-0.07437	0.208725	-0.14013	0
Portland	-5.01611	0.310424	-0.1774	0.169199	-0.04726	0.085166	0	0.105897	0.238784	0.002529	-0.07286	0.267327	-0.21078	0.10563
Sacramento	-2.33116	0.252249	-0.20192	0.140691	0	0.05282	0	0.185145	0.223156	0	-0.0462	0.234856	-0.18667	0
San Antonio	-8.81981	0.228536	-0.06991	0	-0.02089	0.07591	0	0	0.143644	0.069376	-0.05019	0.153555	-0.04374	0.053212
Toronto	-1.80832	0.209404	-0.17198	0.179546	0	0.077386	0	0.03775	0.218078	0	-0.04955	0.153384	-0.17287	0.095087
Utah	-5.05707	0.180476	-0.08425	0.028463	0	0.041468	0	0.031032	0.146453	0.101964	-0.10338	0.137085	-0.16126	0.122866
Washington	-8.13042	0.200358	-0.10453	0.125816	-0.06664	0.079448	0	0.050195	0.239584	0.038689	-0.04371	0.110569	-0.12424	0.120385

omparison

	Miami	San Antonio
FG	0.420593	0.2285359
FGA	-0.21825	-0.06990819
3P	0.155666	0
3PA	-0.00517	-0.020893468
FT	0.226836	0.075909545
FTA	-0.1113	0
OR	0.160231	0
DR	0.252541	0.1436443
A	0.086514	0.069375634
PF	-0.09696	-0.05018765
ST	0.29578	0.15355501
TO	-0.20046	-0.04374036
BL	0.072157	0.053212055

...easurement of player performance

- Players' performance are measured based on the features we selected and their coefficient we get from previous logistic regression
- We use linear weight method to measure players' performance for each game.
  - Harville (1977) used linear model methodology to simply rate college football teams and with expected accuracy.
  - Lackritz (1990) analyzed the impact of performance statistics from players to the current teams' winning percentages
  - Berri (1993) used an econometric model that links the players' statistics to teams' wins for determining the value of production from players.

ome notations

- i: subscript i indicates the player i and i=1 to N where N is the number of players in the dataset.
- j: subscript j indicates the feature j and j=1 to p where p is the number of features we selected using LASSO.
- $y_{i,j,t}$: denotes the output of j^{th} feature for player i in his t^{th} game and we define $Y_i^t = (y_{i,1,t}, \ldots, y_{i,p,t})^{\prime}$ to be a p-vector of outputs.
- α_j: denotes the weight(coefficient) for j^{th} feature

easurement of player performance

evaluate players' performance, an output aggregator is quired to deal with multiple outputs(features).

e define θ(Y↓j↑t) as a scalar function that aggregates these tputs:

$(Y↓j↑t)=∑j=1↑p▒α↓j y↓i,j,t$

ow could we transform our game by game aggregators into ayers' efficiency?

layer evaluation

A set of players with the same level of ability may have different performance for several reason.

Big Decline

21.4 points, 8.9 rebounds/game, 5 ALL STARS

18.2 points, 6.7 rebounds/game, 1 ALL STARS

layer evaluation

ome external we should eliminate:

- Teammates/team strategy effect
- Fixture effect
- Season/year effect
- Other team related effect (opponent, stadium and etc.)

layer evaluation

e include following explanatory variables to control these
kternal effect:

- $x_{1} \sim x_{29}$ = dummies for 29 of 30 teams that exist in 2010-2013 period (Thunder is omitted)
- $x_{30} \sim x_{31}$ = dummies for 2 of 3 seasons (2010-2011 is omitted)

tochastic Frontier Analysis Model

SFA: a method of economic modeling. It measures efficiency that explicitly account for random variation in inputs and outputs.

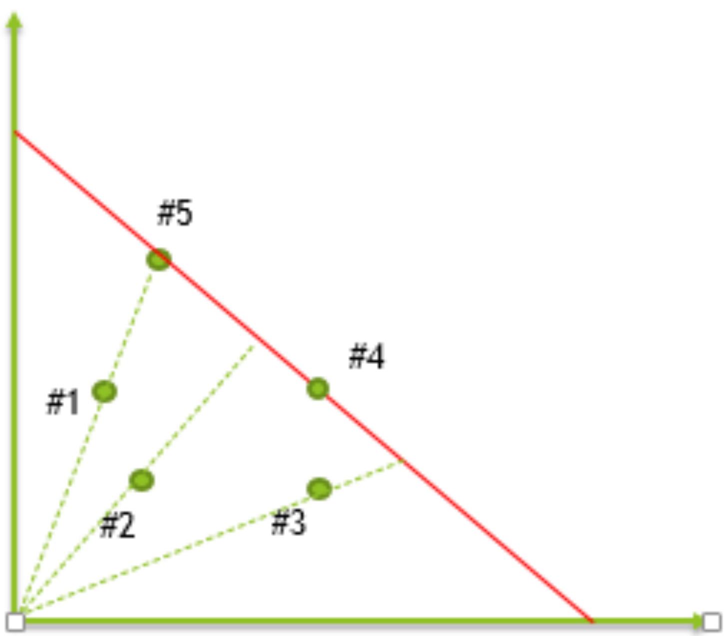
But why do we use SFA?

tochastic Frontier Analysis Model

- The great advantage of SFA is the possibility that it offers of decomposing productivity change into parts that have straightforward interpretation.

- SFA gets rid of external effect by comparing each individual player to his team frontier which is the best player in the team

tochastic Frontier Analysis Model



We define frontier as the best performance in the team, all players lie below the frontier curve

We use the ratio of distances as the measure of efficiency of each player: O#1/O#5. players are measured relative to the frontier curve define as the “best” performance

tochastic Frontier Analysis Model

- TE_i: denotes the efficiency of player i. TE_i always <=1
- $v_{i,t}$: denotes the random shock for player i in game t.

e introduce the SFA function form as:

$$Y_j^t = f(x_i^t, \beta) TE_i \exp(v_{i,t})$$

ere $f(x_i^t, \beta)$ is the frontier indicating the maximum amount
 aggregate output can be produced with given output.

tochastic Frontier Analysis Model

sume $f(x_j^t, \beta)$ takes the log-linear Cobb-Douglas form and rite $TE_i = \exp(-u_i)$, we take log-transformation:

$$(\theta(Y_j^t)) = X_i^t \beta - u_i + v_{i,t}$$

ome more notations

define a T-dimensional vector as:

$$Y)=(\theta(Y_1^1),\dots,\theta(Y_1^{T_1}),\dots,\theta(Y_n^{T_n}))'$$

ere T_i is the number of games player i played and $T=\sum_{i=1}^{n} T_i$. So:

$$g\theta(Y)=(log\theta(Y_1^1),\dots,log\theta(Y_1^{T_1}),\dots,log\theta(Y_n^{T_n}))'$$

also have:

$$=\begin{bmatrix} x_{1,1}^1 & \cdots & x_{1,K}^1 \\ \vdots & \ddots & \vdots \\ x_{n,1}^{T_n} & \cdots & x_{n,K}^{T_n} \end{bmatrix}$$

$$=\begin{bmatrix} 1_{T1} & \cdots & 0_{T1} \\ \vdots & \ddots & \vdots \\ 0_{Tn} & \cdots & 1_{Tn} \end{bmatrix}$$

d:

$$=(v_1,\cdots v_{t1},\cdots v_{tn})'$$

tochastic Frontier Analysis Model

hus the final form of our SFA model becomes:

$$g\theta(Y)=X\beta-DU+V$$

layer efficiency

ayers efficiency are not always consistent, it's better for to treat TE_i to be probabilistic in this research. Thus a ood distribution assumption for u is important to make our timation accurate.

on-negative, bell-shaped, more flexible form

amma distribution is accepted.

olve the model

e use Markov Chain Monte Carol(MCMC) to solve the model:

ome assumptions:

$u) \sim Gamma(\lambda \downarrow 1 , \lambda \downarrow 2)$

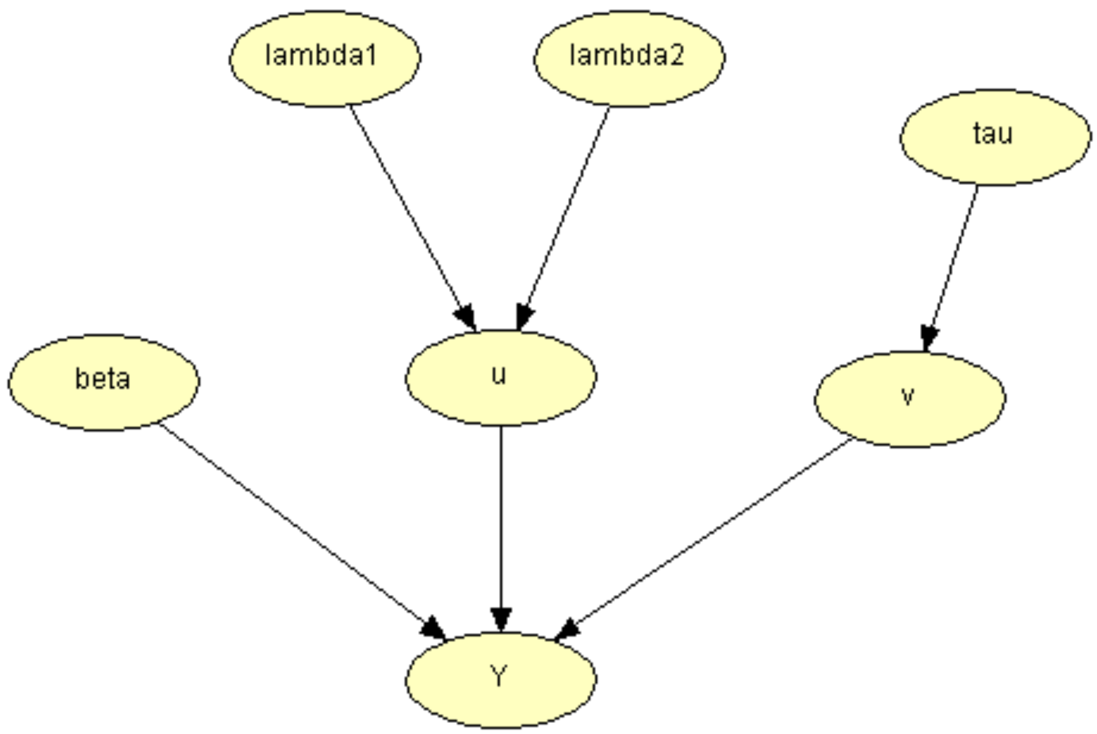
$\lambda \downarrow 1) \sim lambda(9,3)$ $p(\lambda \downarrow 2) \sim lambda(9,3)$

$p(\beta) \sim normal(3, -\ln(3))$

$normal(0,1/tau)$

$u \sim gamma(1, 10\hat{}6)$

MCMC



mpirical

e used NBA 2010-2013 regular season game by game data om *NBAstuffer.*

?9 games played during 2010-2013 regular seasons for each am

here are more than 600 players in the dataset

ata manipulation

- We only keep observations with min>10
- Players being considered should played at least 80% games in that regular season
- Only non-essential players will be available in the player trading market

mpirical

nally, we get

9 players in 29 teams, with totally 36237 observations
ame*player)

low does the model work

Step1: learn the team strategy for given team from teams game-by-game data using logistic LASSO.

Step2: calculate the output aggregators using players game-by-game data and game strategy information.

Step3: build the SFA model with aggregators.

Step4: solve the model with MCMC.

mpirical result

e choose New York as an example to help its coach to find e “best” players for it after 2012-2013 season.

ecord: 54-28

TS/G:100 Opp.PTS/G: 95.7

mpirical result

Here is the table of top 20 players for New York

PLAYER	efficiency	position
Dwight Howard	0.99005	C
Kendrick Perk	0.99005	C
Kevin Durant	0.99005	SF
Nick Collison	0.99005	PF/C
Reggie Evans	0.99005	PF
Serge Ibaka	0.99005	PF/C
Thabo Sefolos	0.99005	SG/SF
James Harden	0.980199	SG
Kris Humphrie	0.980199	PF/C
Omer Asik	0.980199	C
Russell Westb	0.980199	PG
Andray Blatch	0.970446	PF/C
C.J. Watson	0.970446	PG
Eric Maynor	0.970446	PG
Tyson Chandle	0.970446	C
Kevin Love	0.951229	PF/C
Marcus Camby	0.951229	C
Joakim Noah	0.941765	C
Al Horford	0.932394	PF/C

mpirical result

And here is the information about the trading after 2012-2013 season

In	efficiency	Out	efficiency
Beno Udrih	0.212248	Marcus Camby	0.951229
World Peace	0.160414	Jason Kidd	0.323033
Shannon Brown	0.115325	Steve Novak	0.145148

cord: 37-45

S/G:98.6 Opp.PTS/G: 99.4

onclusion

The NBA teams would not have exactly same important characteristics relate to wins. Teams need to find players who can fit their game strategies. In the same sense, firms will also have different characteristics relate to their benefit. They should develop and exploit distinctive competencies based on their own situation.

Thanks and questions?