Yale SCHOOL *of* MANAGEMENT

YALE UNIVERSITY
School of Public Health

Brock University

YALE UNIVERSITY
School of Engineering and
Applied Science

Yale College

A Puck In The Net Beats Four Men In The Box

Edward H. Kaplan[1], Kevin Mongeon[2], and John Ryan[3]

[1]Schools of Management, Public Health, and Engineering, Yale University, New Haven, CT USA

[2]Department of Sport Management, Brock University, St. Catherines, ON Canada

[3]Yale College, Yale University, New Haven, CT USA

In These Slides We…

- Review the basic Poisson model of hockey
- Expand to account for penalty-induced manpower differential
- Calibrate with second-by-second data from the 2008-2012 NHL regular seasons (4920 games = 17,712,000 seconds; big data?)
- Report the state-dependent probability of the home team winning at any second during a game
- Show that *a puck in the net beats four men in the box!*
- Show that this is a property of the data as opposed to the model
- Develop new *Win Probability Added* (*WPA*) metric for players
- Examine the extent of model/data concordance
- Determine degree of stochastic variation in observed win probabilities one should expect based on the model

Poisson Model of Hockey

Siméon Poisson and The National Hockey League

GARY M. MULLET*

Using only goals scored and goals given up, home and away, for each team in the National Hockey League for the 1973–74 season, hypotheses tests indicated that mean goals for and against both home and away are all distributed according to a member of the Poisson family. Further analysis indicated that goals for and goals against at home and away are independent random variables. This latter conclusion came by assuming independence, explaining won-loss records and using several hypotheses tests to look for contradictions. No serious ones were found.

The American Statistician, 31:1, 8-12, 1977

- Also Morrison (1976), Morrison and Wheat (1986), Erkut (1987), Nydick and Weiss (1989), …
- Most relevant for us are Washburn (1991; state space), Beaudoin and Swartz (2010) and Buttrey, Washburn and Price (2011) (manpower dependent goal scoring rates)

Poisson Model of Hockey

- Let λ (μ) denote the expected number of goals scored by the home (away) team in a game, and let X (Y) denote the (random) number of home (away) goals scored; assume X (Y) independent and Poisson
- Home team beats away team in regular time if $X > Y$ (home scores more goals)
- But, in hockey can tie in regular time; we'll assume win in overtime/shootout with conditional probability 0.5

Poisson Model of Hockey

- Assuming home, away scoring is independent, we have

$$\Pr\{X > Y\} = \sum_{x=1}^{\infty} \Pr\{X = x\} \Pr\{Y < x\}$$

- Also, for the probability of a tie we have

$$\Pr\{X = Y\} = \sum_{x=0}^{\infty} \Pr\{X = x\} \Pr\{Y = x\}$$

- Pr{Home Win} = $\Pr\{X > Y\}$ + 0.5*$\Pr\{X = Y\}$

Poisson Model of Hockey

- In 2008-09/2011-12 regular seasons (1230 games/year, 4920 games overall)…
- Average goals/game = 2.75 for the home team and 2.47 for the away team
- Poisson model prediction: probability home team wins equals 0.547
- Over 4920 games, home team won 2702 or 54.9%

Poisson Model of Hockey

- But with almost no additional effort, we can learn much more from this model
- Let $w(x, t)$ denote the probability that the home team wins the game, given that with t time units remaining in the game, the home team leads by x goals ($x < 0$ means home team trails)
- Note that Pr{Home Win} = $w(0, 60 \text{ mins})$

Poisson Model of Hockey

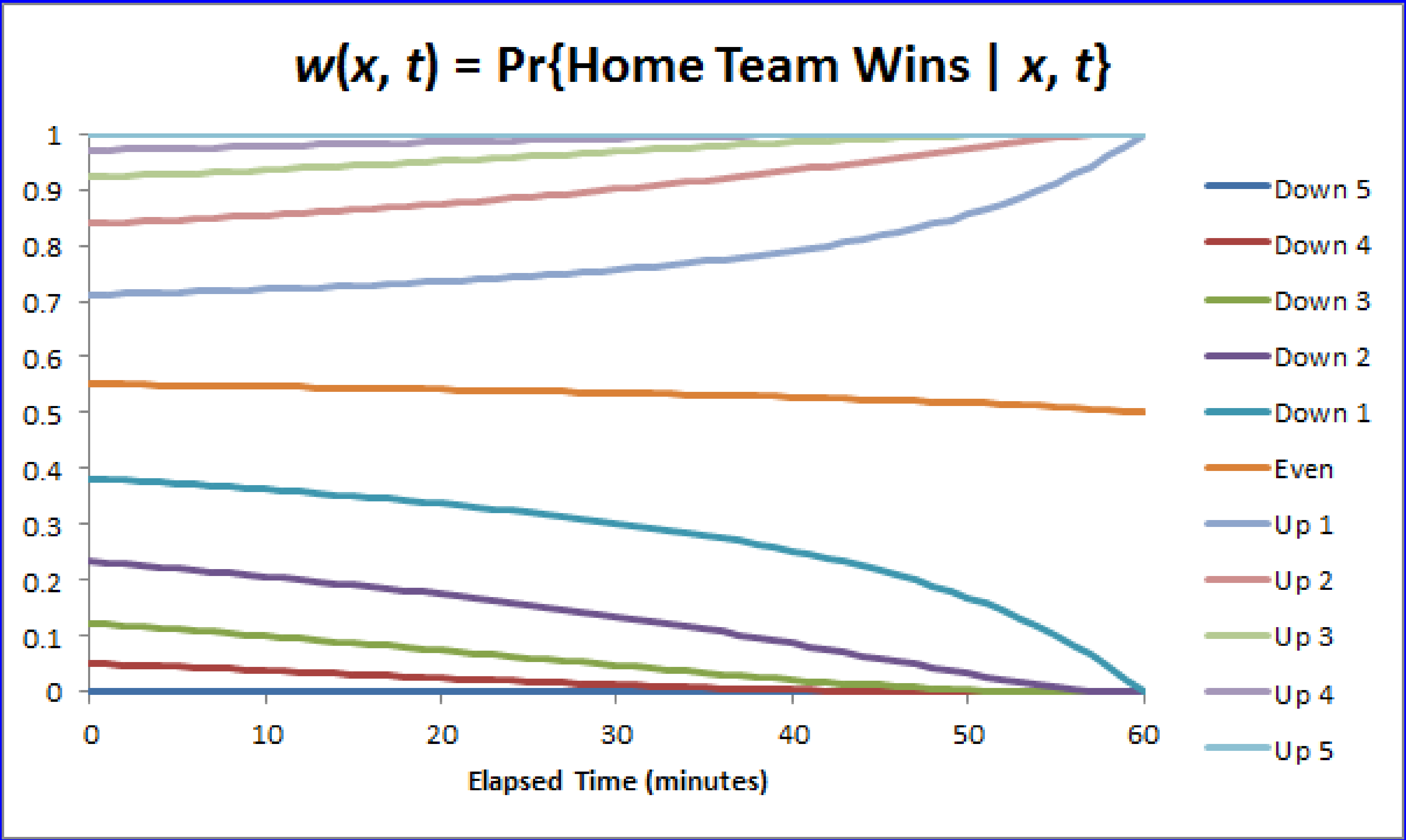
	goes to		with probability	what happened
		$w(x+1, t-\Delta t)$	$\lambda\Delta t$	home team scored!
$w(x,t)$		$w(x, t-\Delta t)$	$1 - \lambda\Delta t - \mu\Delta t$	nothing
		$w(x-1, t-\Delta t)$	$\mu\Delta t$	away team scored!

$$\frac{dw(x,t)}{dt} = \lambda w(x+1,t) + \mu w(x-1,t) - (\lambda+\mu)w(x,t)$$

$$w(x,0) = \begin{cases} 1 & x > 0 \\ \frac{1}{2} \quad \text{if} & x = 0 \\ 0 & x < 0 \end{cases}$$

$$w(5,t) = 1; w(-5,t) = 0 \quad \text{(mercy rule!)}$$

Poisson Model of Hockey



Incorporating Manpower Differential

- Hockey has penalties! Penalized teams lose players for 2 minutes (minor, most common), 5 minutes (major, less common), 10 minutes (misconduct), or ejection from game (game misconduct)
- Minor penalties can also end with a power play goal, major penalties last the full 5 minutes, while misconducts are usually accompanied by other penalties (minor or major) that create a manpower advantage after which the player with the misconduct must remain in the penalty box, but otherwise the team plays even strength

Incorporating Manpower Differential

- We seek to answer the following question: with time t remaining in regulation, what is the probability that the home team wins, given that the home team leads by x goals while enjoying a manpower differential of y players
- Denote this by $w(x, y, t)$
- Following NHL rules, we only consider manpower differentials y of –2, –1, 0, 1, 2
- For convenience we apply a mercy rule at $x = +/- 5$

Markov Model With Manpower Differential

- General model: let $\lambda_{xy}^{x'y'}$ denote the conditional transition rate to state (x', y') from state (x, y)
- Note that transitions can involve scoring, start and end of penalties, or both (power play goal ends minor penalty)
- Note that real penalties last for a fixed duration unless ended by a power play goal (minors)
- In the model, do not keep track of elapsed penalty time

Markov Model With Manpower Differential

- Equations for model with manpower differential:

$$\frac{dw(x,y,t)}{dt} = \sum_{(x',y')\neq(x,y)} \lambda_{xy}^{x'y'} w(x',y',t) - \left(\sum_{(x',y')\neq(x,y)} \lambda_{xy}^{x'y'}\right) w(x,y,t)$$

$$w(x,y,0) = \begin{cases} 1 & x > 0 \\ w(0,y,0) & x = 0 \\ 0 & x < 0 \end{cases} \quad w(0,y,0) = \Pr\{\text{OT Win}\}$$

$$w(5,y,t) = 1; w(-5,y,t) = 0 \quad \text{(mercy rule!)}$$

$$\frac{dw(0,y,t)}{dt} = \sum_{(x',y')\neq(0,y)} \lambda_{0y}^{x'y'} w(x',y',t) - \left(\sum_{(x',y')\neq(0,y)} \lambda_{0y}^{x'y'}\right) w(0,y,t)$$

$$w(1,y,t) = 1; w(-1,y,t) = 0 \quad \text{(sudden death!)}$$

$$w(0,y,-5) = \tfrac{1}{2} \quad \text{(shoot out!)}$$

Data Description

- The data consist of the regulation clock times that goals are scored (home or away) and penalties are issued (home or away) for the 4920 NHL regular season hockey games in the 2008-09 to 2011-12 seasons (17,712,000 seconds), along with who won the game

Sample Data

VISITOR		HOME
5	**Game Summary**	4
DALLAS STARS Game 2 Away Game 2	Saturday, October 9, 2010 Attendance 13,351 at Nassau Coliseum Start 7:16 EDT; End 9:55 EDT Game 0015 Final	NEW YORK ISLANDERS Game 1 Home Game 1

SCORING SUMMARY

G	Per	Time	Str	Team	Goal Scorer	Assist	Assist	DAL on Ice	NYI on Ice
1	1	9:52	PP	DAL	14 J.BENN(1)	91 B.RICHARDS(2)	29 S.OTT(1)	6, 10, 14, 29, 32, 91	4, 24, 26, 39, 51
2	1	11:59	EV	DAL	3 S.ROBIDAS(1)	91 B.RICHARDS(3)	6 T.DALEY(1)	3, 6, 18, 21, 32, 91	4, 24, 26, 39, 91
3	2	6:21	PP	NYI	20 J.WISNIEWSKI(1)	93 D.WEIGHT(1)	15 P.PARENTEAU(1)	3, 14, 29, 32, 37	12, 15, 20, 26, 39, 93
4	2	13:45	EV	DAL	10 B.MORROW(2)	63 M.RIBEIRO(2)	2 N.GROSSMAN(1)	2, 3, 10, 16, 32, 63	10, 15, 20, 26, 28, 39
5	2	15:24	PP	NYI	93 D.WEIGHT(1)	15 P.PARENTEAU(2)	20 J.WISNIEWSKI(1)	3, 32, 37, 63	12, 15, 20, 26, 39, 93
6	3	7:12	EV	DAL	10 B.MORROW(3)	63 M.RIBEIRO(3)		5, 10, 29, 32, 37, 63	16, 27, 39, 47, 57, 93
7	3	8:45	EV	NYI	57 B.COMEAU(1)	24 R.MARTINEK(1)	12 J.BAILEY(1)	6, 13, 28, 29, 32, 91	4, 12, 24, 28, 39, 57
8	3	17:21	PP	NYI	26 M.MOULSON(1)	51 F.NIELSEN(1)	93 D.WEIGHT(2)	2, 10, 28, 32, 63	7, 20, 26, 39, 51, 93
9	SO			DAL	63 M.RIBEIRO				

PENALTY SUMMARY

DALLAS STARS

#	Per	Time	Player	PIM	Penalty
1	1	1:53	29 S.OTT	2	Hooking
2	1	3:44	18 J.NEAL	2	Tripping
3	2	6:15	17 T.PETERSEN	2	Hooking
4	2	14:06	2 N.GROSSMAN	2	Interference
5	2	14:31	29 S.OTT	2	Tripping
6	3	10:56	91 B.RICHARDS	2	Holding
7	3	16:17	3 S.ROBIDAS	2	Delaying Game-Puck over glass
8	3	17:53	3 S.ROBIDAS	2	Delaying Game-Puck over glass
TOT (PN-PIM) 8-16					
Power Plays (Goals-Opp./PPTime) 1-7/11:12					

NEW YORK ISLANDERS

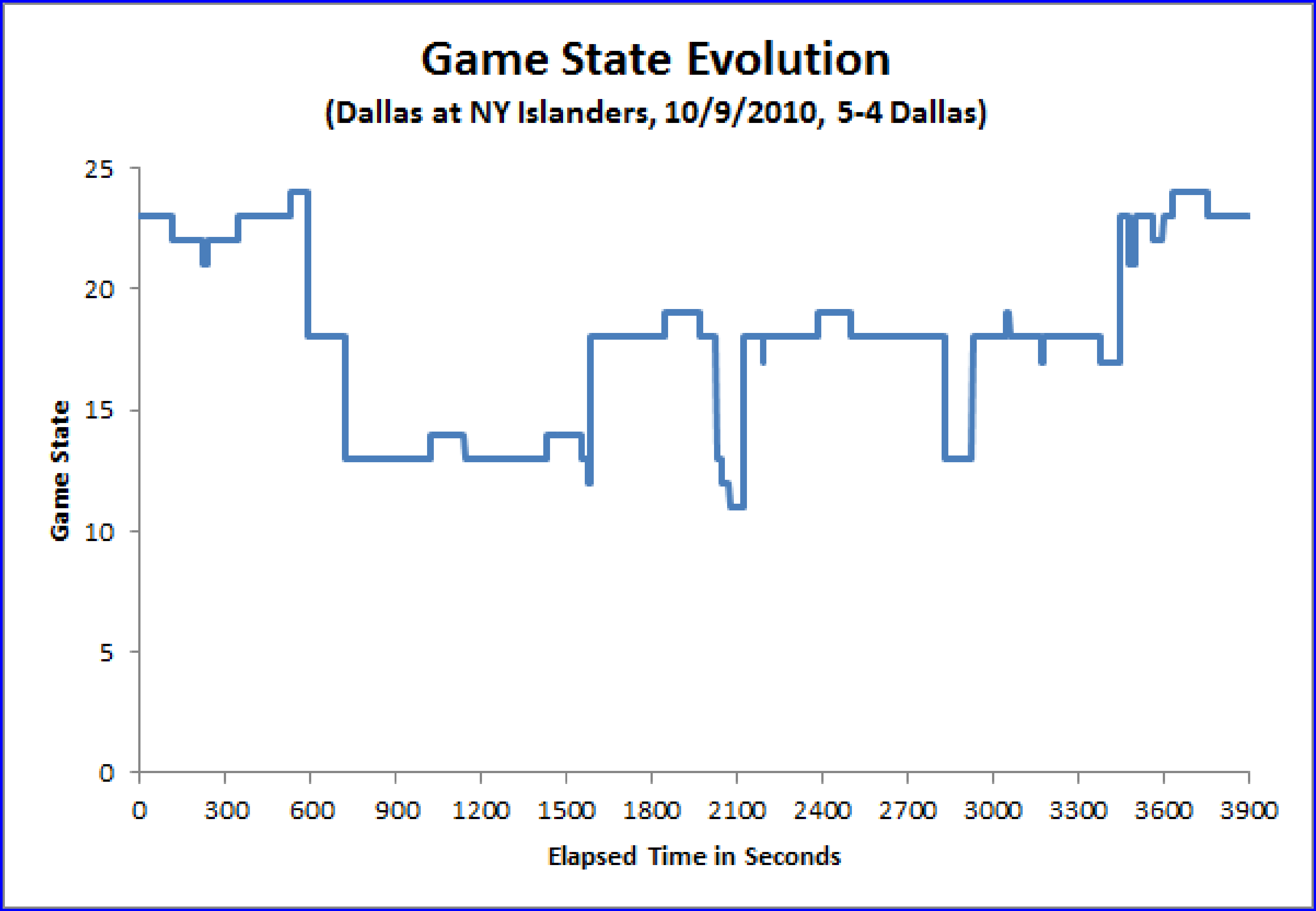
#	Per	Time	Player	PIM	Penalty
1	1	8:49	57 B.COMEAU	2	Boarding
2	1	[illegible]	4 M.EATON	2	Interference
3	2	3:51	TEAM	2	Too many men/ice - bench
4	2	10:46	93 D.WEIGHT	2	Hi-sticking
5	2	19:40	20 J.WISNIEWSKI	2	Slashing
6	3	10:47	16 J.SIM	2	Roughing
7	OT	0:29	57 B.COMEAU	2	Cross checking
TOT (PN-PIM) 7-14					
Power Plays (Goals-Opp./PPTime) 3-8/09:35					

- Note that need to *infer* end of penalty
- e.g. Dallas goal at 9:52…
- ends NYI boarding penalty

Data Description

- Convenient to convert two dimensional state (x, y) to one dimensional $s(x, y)$
- Define $s(x, y) = 23 + 5x - y$
  - Tie game/even strength: $s(0, 0) = 23$
  - 2 goal lead/2 men down: $s(2, -2) = 35$
  - 4 goals down/2 men up: $s(-4, 2) = 1$
  - 4 goals up/2 men down: $s(4, -2) = 45$
  - Also set $s(5, y) = 46$; $s(-5, y) = 0$
- At end of regulation, home team wins if $s > 25$, loses if $s < 21$, and OT if $21 \leq s \leq 25$

Convert Text to Sample Path



Observation: Home Edge

- "…'officials' bias' is the most significant contributor to home field advantage." (Moskowitz and Werthem, *Scorecasting*, 2011)
- Home team averages 2.75 goals/game; away team averages 2.47 goals per game; home wins 55% of games
- Home team penalized on average 4.64 times per game; away team penalized on average 4.99 times per game

Manpower State	Home Goals/60 mins	Away Goals/60 mins
Home Up 2	11.73	0.49
Home Up 1	6.11	0.89
Even Strength	2.49	2.25
Home Down 1	1.00	5.65
Home Down 2	0.58	11.66

Parameter Estimation

- Over all 4920 games, define:
  - $\tau(s)$ = total time spent in state s
  - $n(s,s')$ = number transitions from s to $s' \neq s$
- Then $\lambda_s^{s'} = \dfrac{n(s,s')}{\tau(s)}$
- But, allowing arbitrary data-defined transitions would require 45 x 45 = 2025 transition rates; we know most are impossible (e.g. 2 goals down man up to 2 goals up man down in one transition)

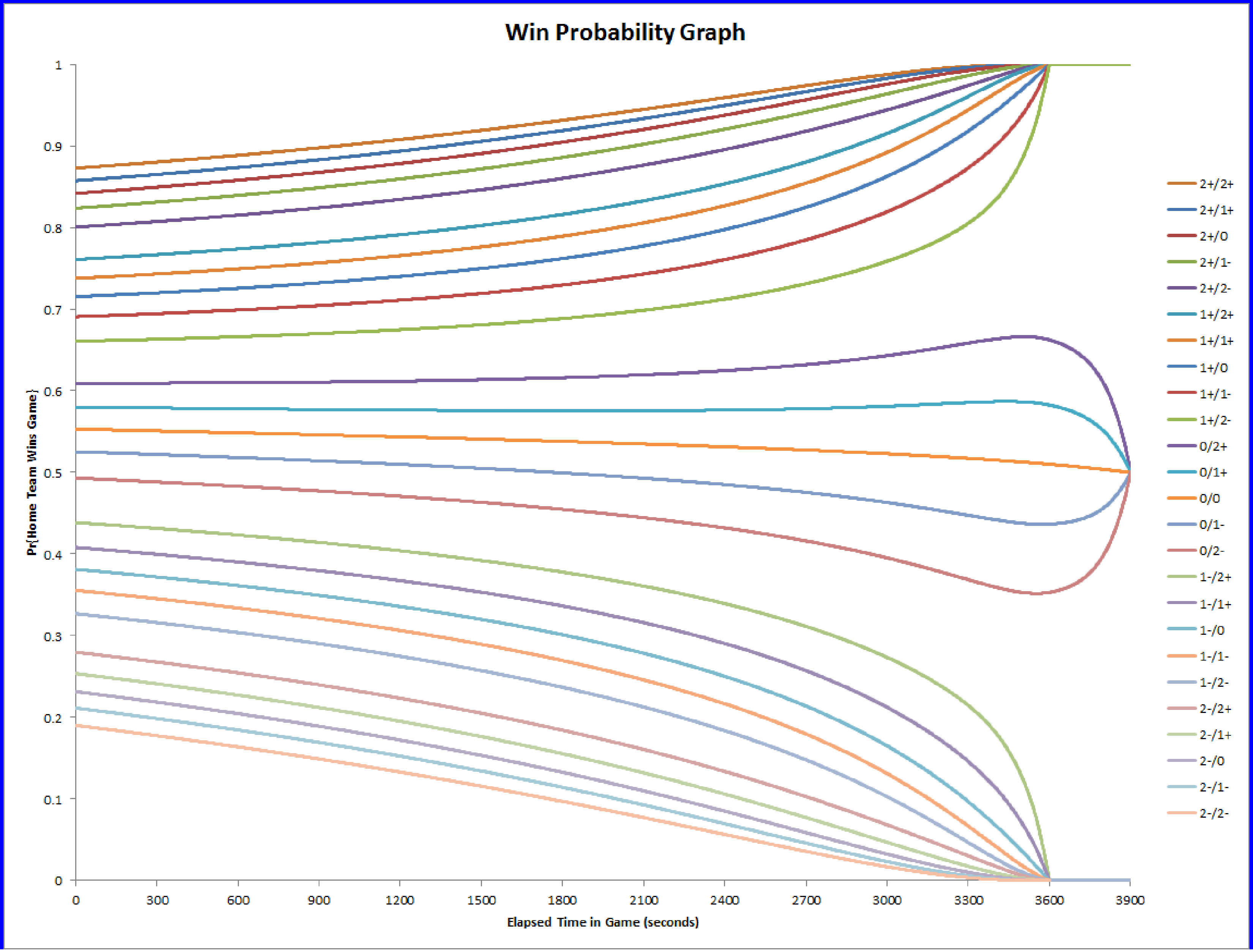
Simplify Transition Rate Structure

- Remove dependence on goal differential of starting state; remove impossible transitions

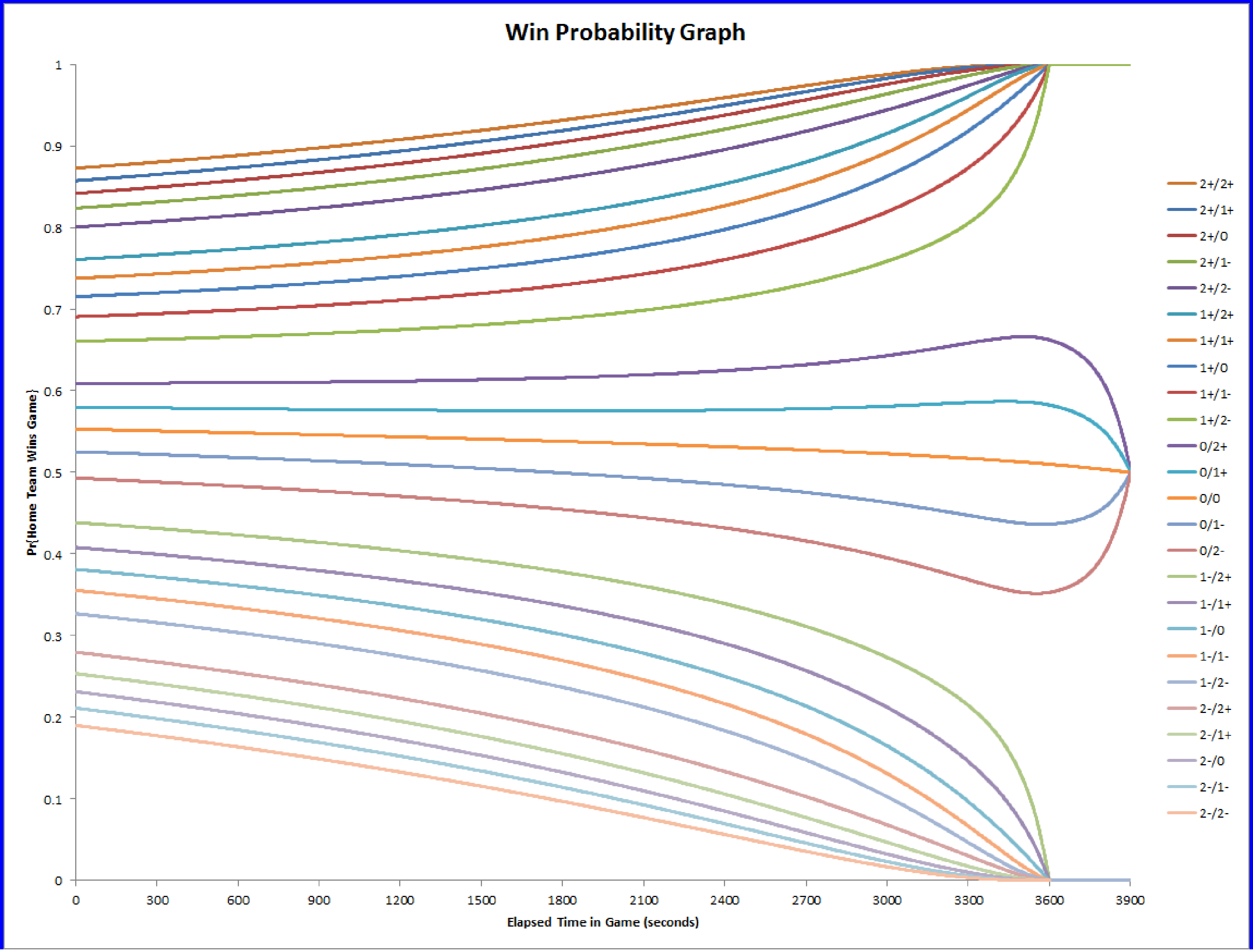
$x' - x$	1	1	1	1	1	0	0	0	0	0	-1	-1	-1	-1	-1
y / y'	2	1	0	-1	-2	2	1	0	-1	-2	2	1	0	-1	-2
2	λ	λ	λ	$-$	$-$	$-$	λ	λ	$-$	$-$	λ	$-$	$-$	$-$	$-$
1	λ	λ	λ	$-$	$-$	λ	$-$	λ	λ	$-$	$-$	λ	λ	$-$	$-$
0	$-$	λ	λ	λ	$-$	λ	λ	$-$	λ	λ	$-$	λ	λ	λ	$-$
-1	$-$	$-$	λ	λ	$-$	$-$	λ	λ	$-$	λ	$-$	$-$	λ	λ	λ
-2	$-$	$-$	$-$	$-$	λ	$-$	$-$	λ	λ	$-$	$-$	$-$	λ	λ	λ

- Leaves 38 λ's to estimate from 17.7 M seconds

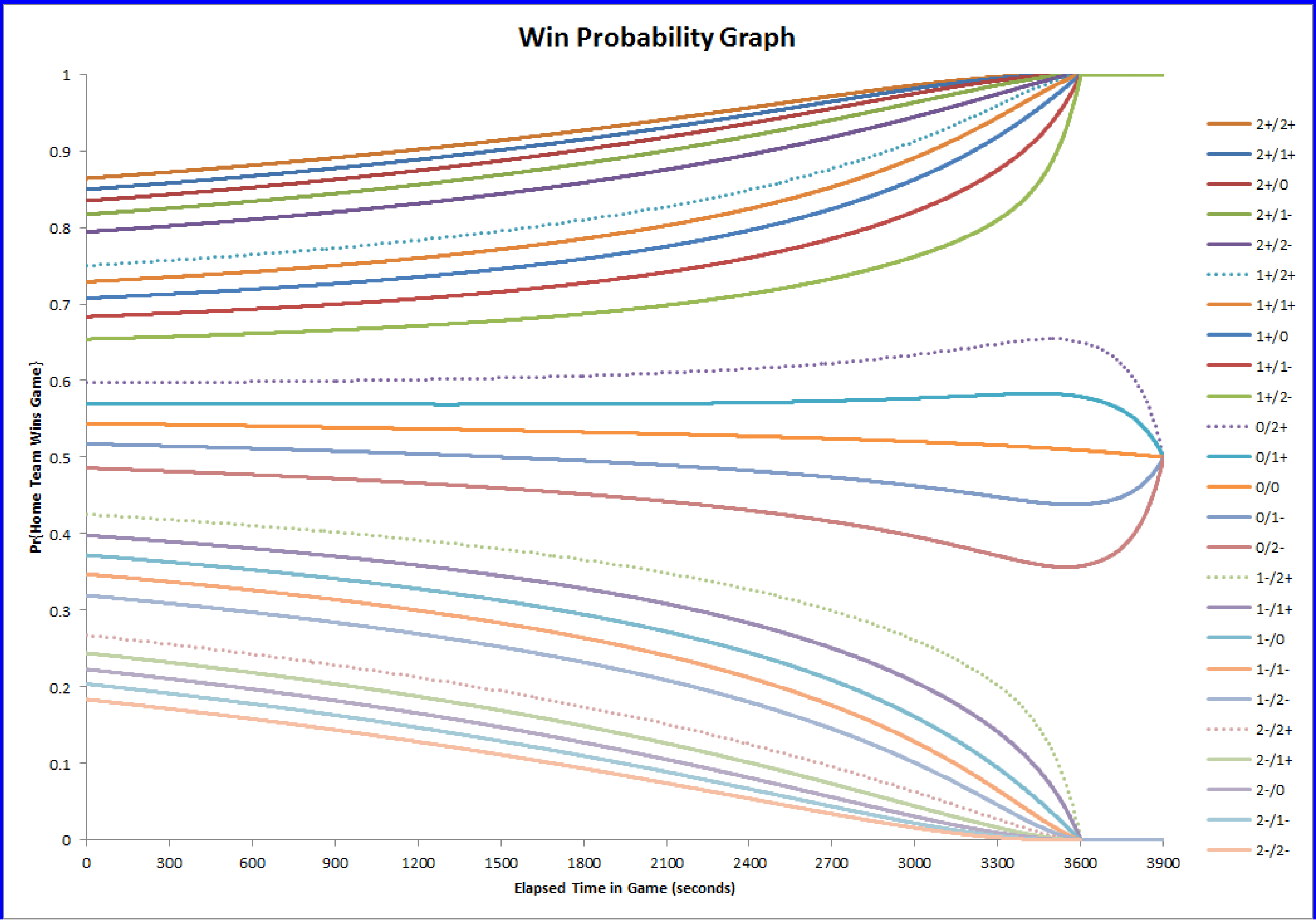
Resulting Win Probability Model



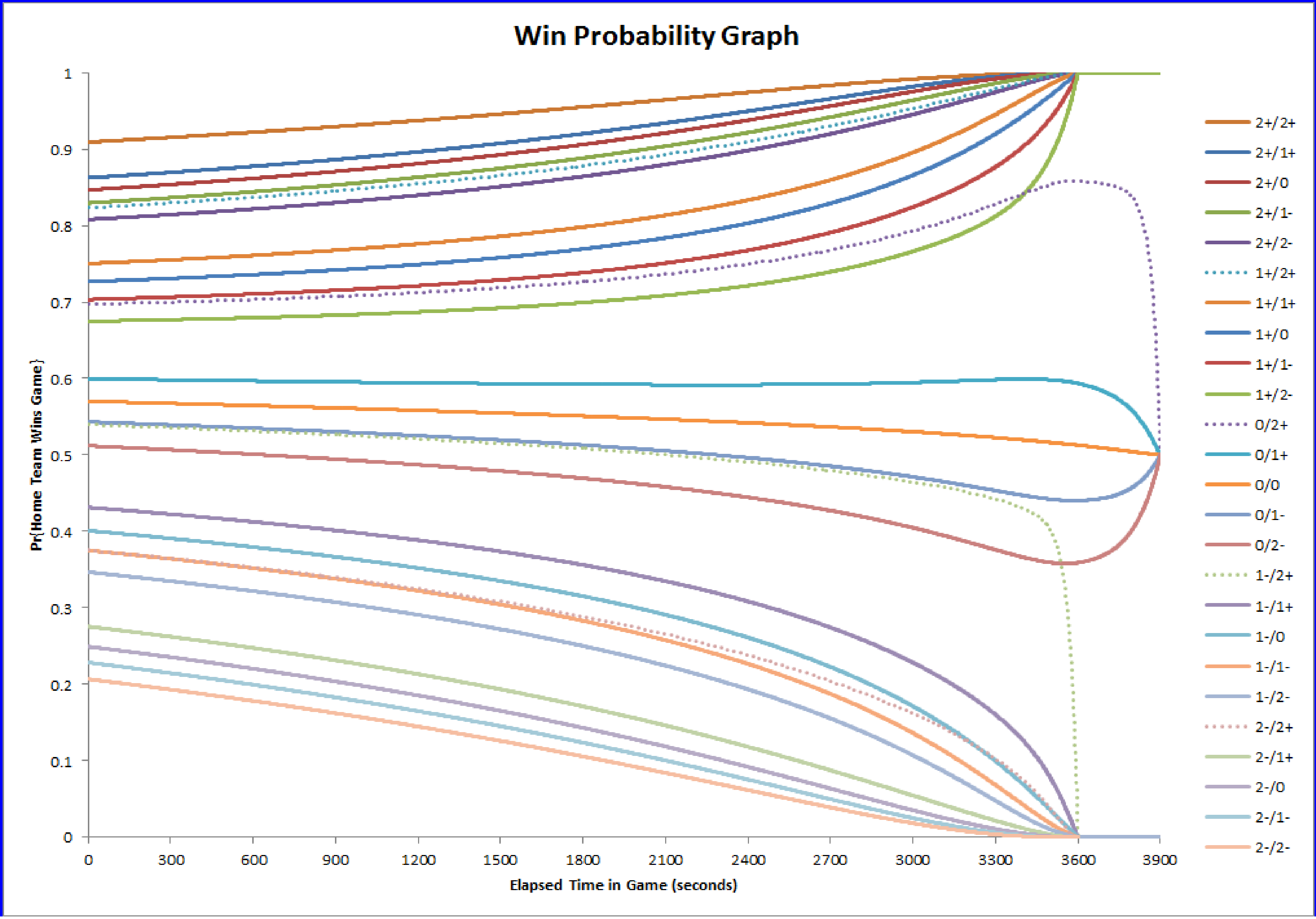
A Puck In The Net Beats Four Men In The Box



Focus On 2 Man Advantage States



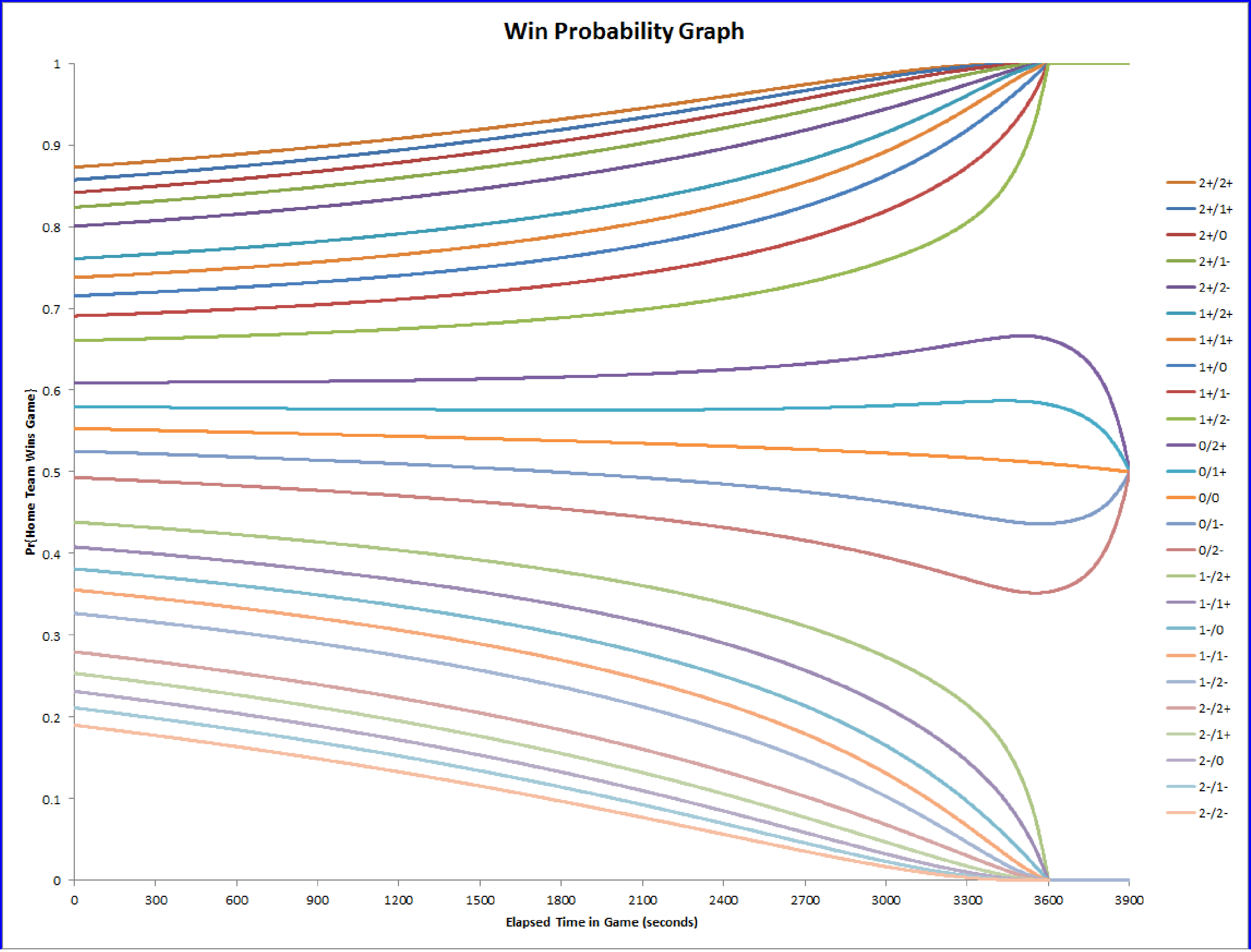
Must A Puck In The Net Beat Four Men In The Box?



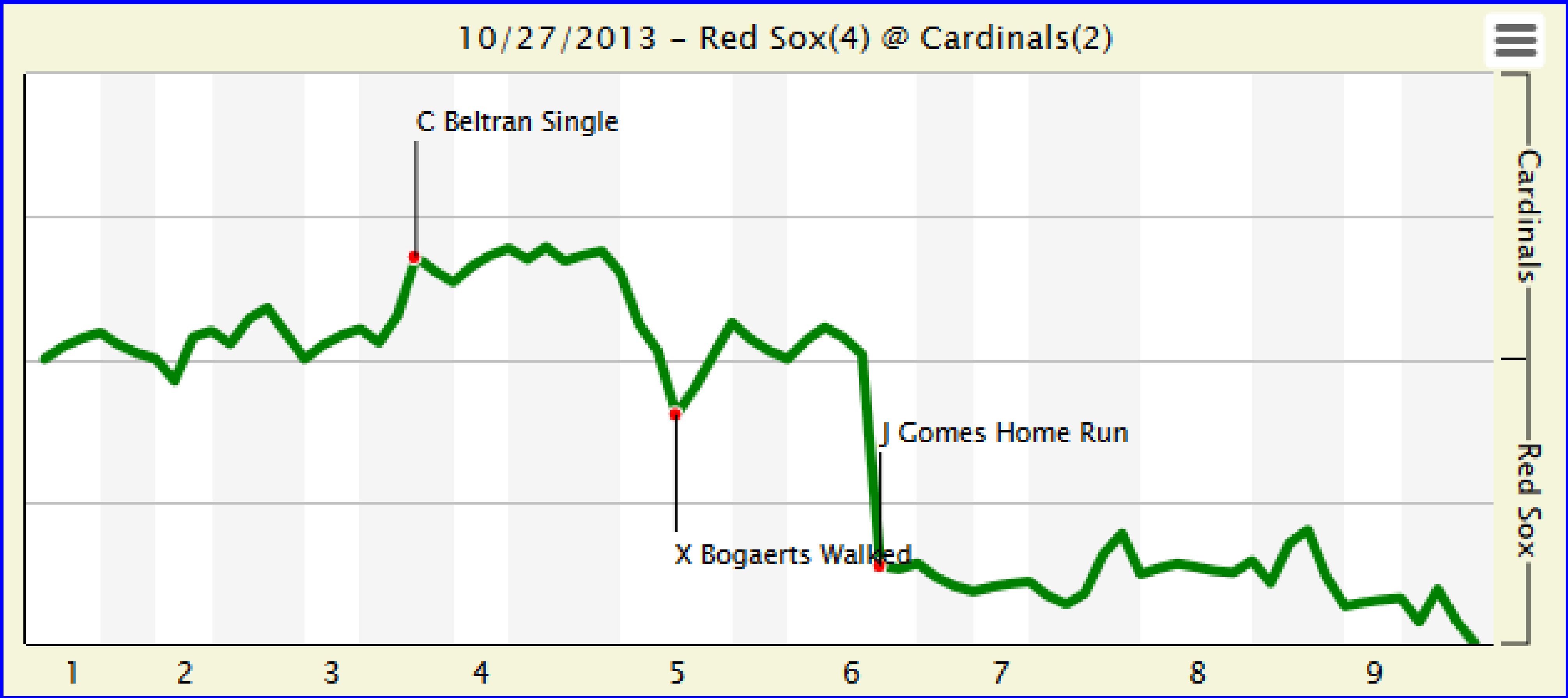
A Puck In The Net Beats Four Men In The Box…

- …is a property of NHL hockey as opposed to a property of the hockey model
- Counterexamples show that were the data different (e.g. much higher home team goal scoring rate with 2 man advantage), a puck in the net would *not* beat four men in the box
- With actual observed manpower-specific goal scoring rates, generally a good idea to take a penalty to save a goal!

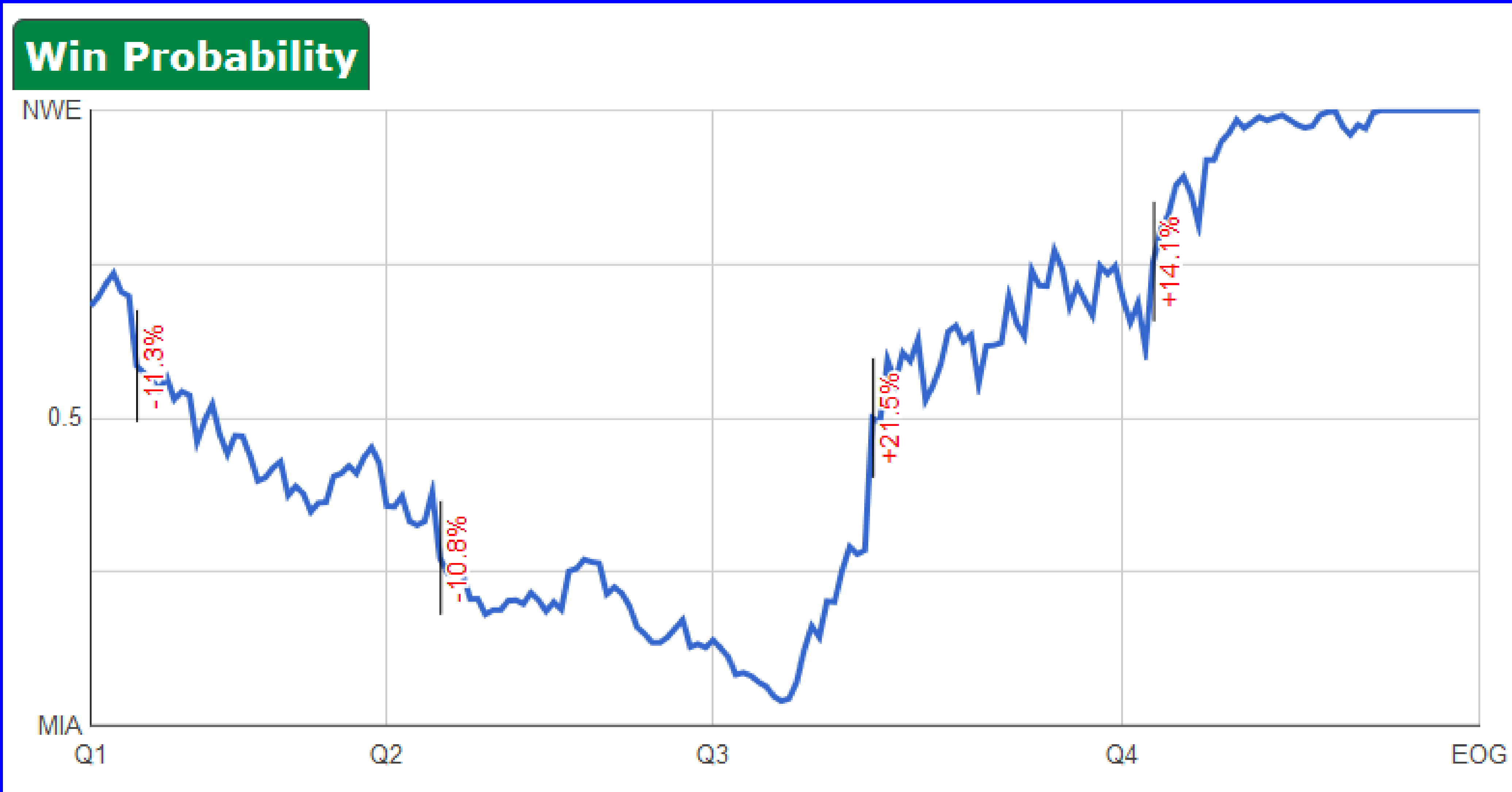
How Might You Use This?



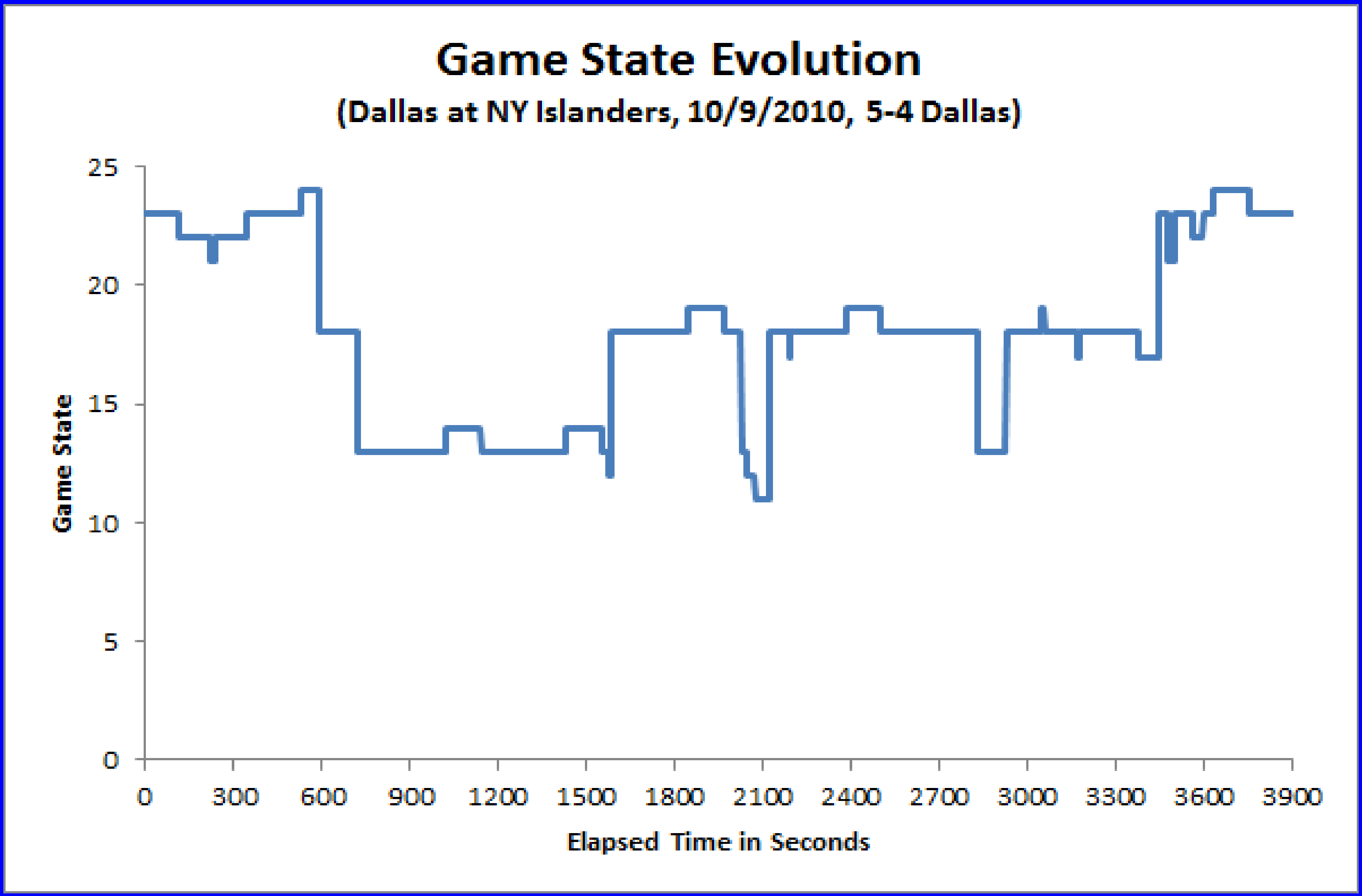
In Baseball… (fangraphs.com)



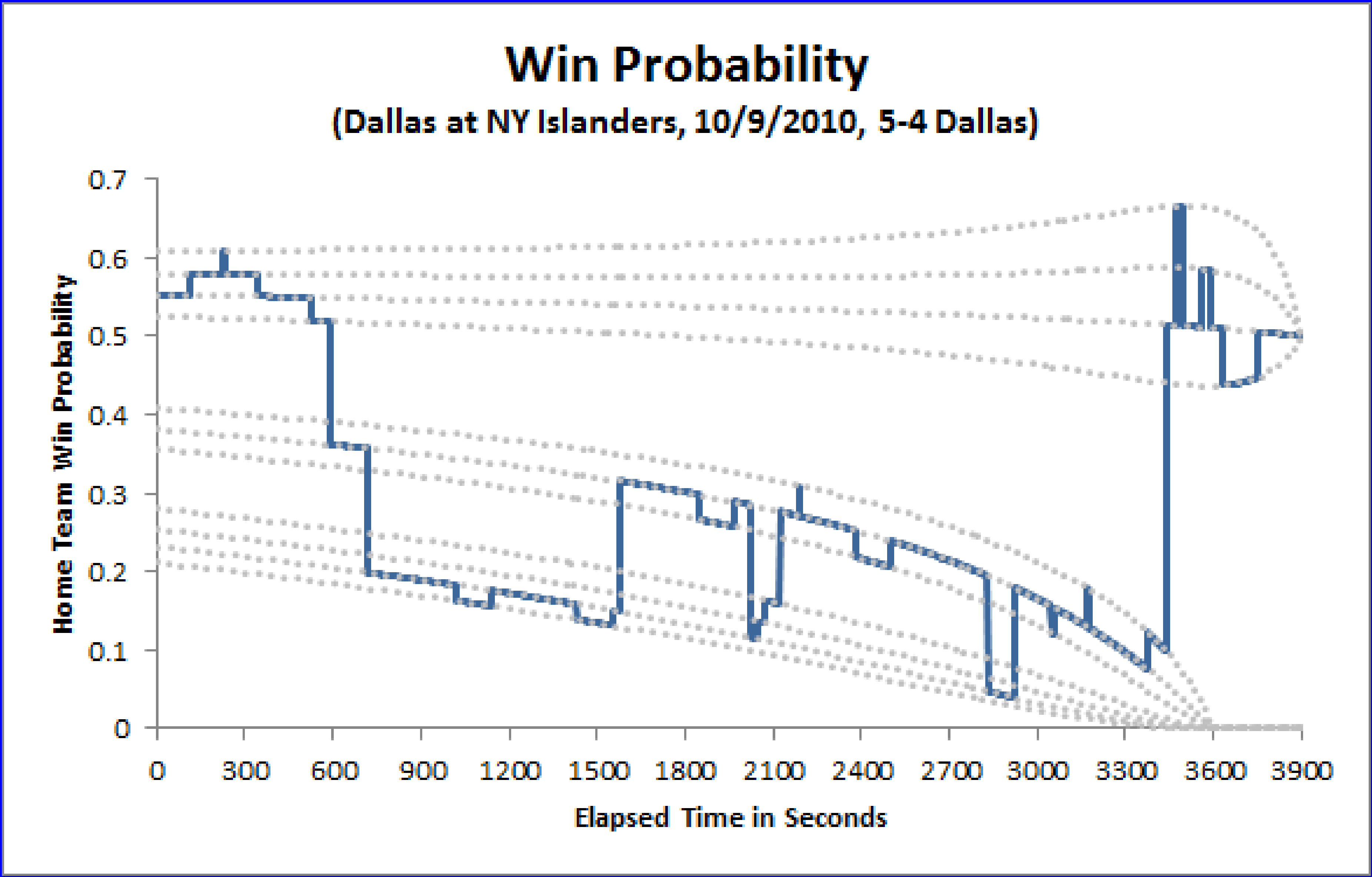
In Football… (pro-football-reference.com)



Use Hockey Model To Translate This…

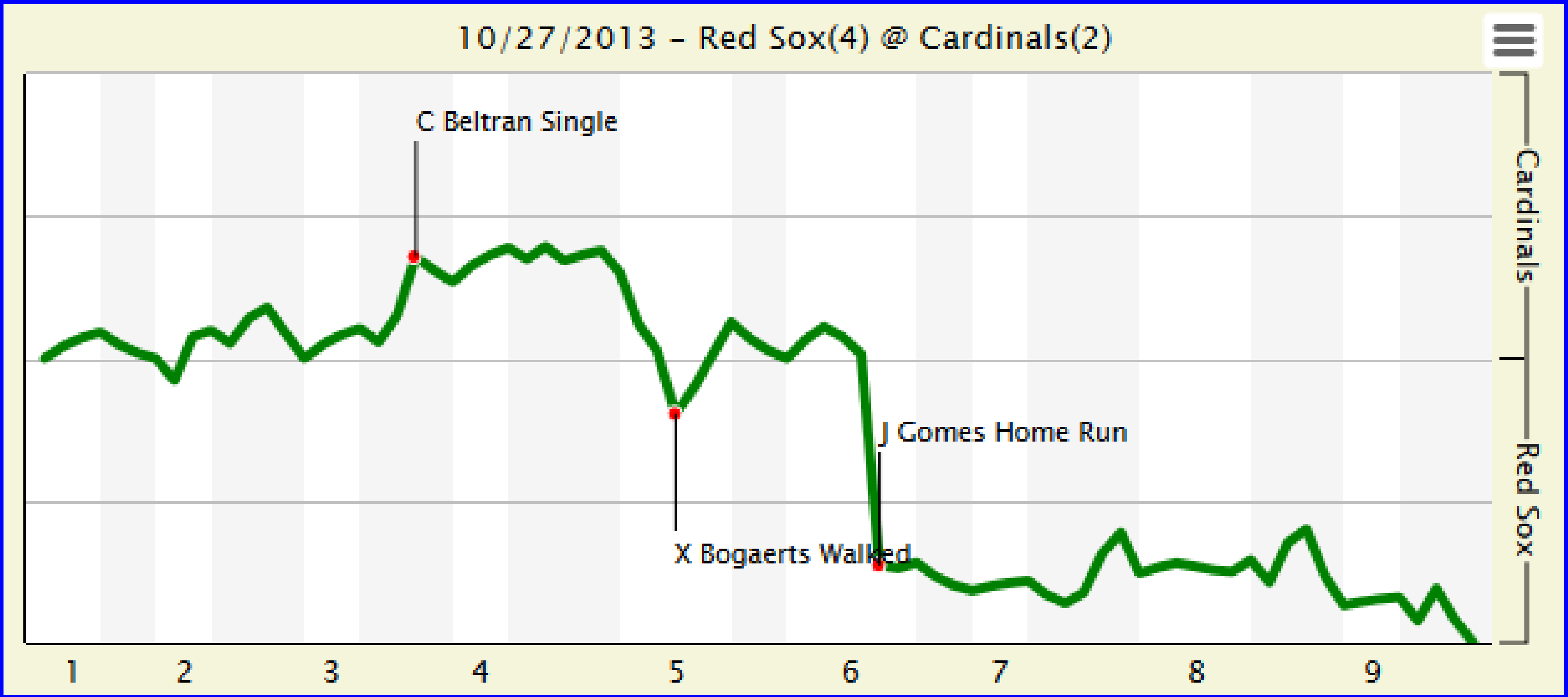


To This!!



Win Probability
(Dallas at NY Islanders, 10/9/2010, 5-4 Dallas)
Home Team Win Probability
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
Elapsed Time in Seconds

Break Down Win Probability By Individual Player Contributions (*WPA*)



Break Down Win Probability By Individual Player Contributions (*WPA*)

Pitcher	IP	H	HR	ER	BB	SO	pLI	WPA
F Doubront	2.2	1	0	1	0	3	0.73	.133
K Uehara	1.0	1	0	0	0	0	1.66	.081
J Lackey	1.0	0	0	0	0	0	1.85	.080
J Tazawa	0.1	0	0	0	0	0	2.86	.071
C Buchholz	4.0	3	0	0	3	2	1.02	.057
C Breslow	0.0	1	0	0	1	0	1.15	-.105
Total	9.0	6	0	1	4	5	1.18	.317

Pitcher	IP	H	HR	ER	BB	SO	pLI	WPA
L Lynn	5.2	3	0	3	3	5	1.27	.129
J Axford	1.1	0	0	0	1	2	0.42	.026
K Siegrist	0.2	1	0	0	0	0	0.33	.012
R Choate	0.1	0	0	0	0	0	0.15	.004
S Maness	1.0	2	1	1	0	1	0.73	-.354
Total	9.0	6	1	4	4	8	0.95	-.183

Batter	AB	R	H	HR	RBI	BB	SO	pLI	WPA
J Gomes	2	1	1	1	3	2	0	1.67	.344
D Ortiz	3	2	3	0	0	1	0	0.92	.164
X Bogaerts	3	0	1	0	0	1	1	1.09	.085
Q Berry	0	0	0	0	0	0	0	0.42	.006
M Napoli	1	0	0	0	0	0	1	0.13	-.003
F Doubront	1	0	0	0	0	0	1	0.23	-.005
C Buchholz	1	0	0	0	0	0	1	0.45	-.011
D Pedroia	4	1	1	0	0	0	2	0.47	-.012
M Carp	1	0	0	0	0	0	0	2.28	-.057
D Nava	4	0	0	0	0	0	0	0.70	-.068
J Ellsbury	4	0	0	0	0	0	0	0.88	-.086
S Drew	3	0	0	0	1	0	0	1.24	-.087
D Ross	4	0	0	0	0	0	2	0.96	-.087
Total	31	4	6	1	4	4	8	0.95	.183

Batter	AB	R	H	HR	RBI	BB	SO	pLI	WPA
C Beltran	3	0	1	0	1	1	0	1.05	.103
Y Molina	4	0	1	0	0	0	0	1.10	.102
A Craig	1	0	1	0	0	0	0	1.07	.057
S Robinson	1	1	1	0	0	0	0	0.34	.020
M Carpenter	5	1	2	0	1	0	1	1.23	.011
K Wong	0	0	0	0	0	0	0	1.45	-.040
J Jay	2	0	0	0	0	2	0	1.15	-.050
L Lynn	2	0	0	0	0	0	1	1.27	-.063
D Descalso	3	0	0	0	0	1	0	1.31	-.094
M Adams	4	0	0	0	0	0	1	1.06	-.107
M Holliday	4	0	0	0	0	0	1	1.22	-.120
D Freese	4	0	0	0	0	0	1	1.37	-.136
Total	33	2	6	0	2	4	5	1.18	-.317

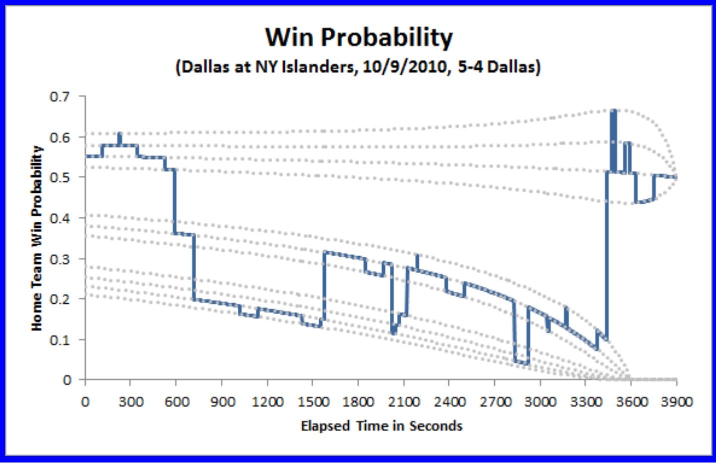
Hockey Model Enables Computation of Hockey *WPA*

- Excluding goalie, let

$$\xi_i^H(t) = \begin{cases} 1 & \text{home team player } i \text{ on ice at time } t \\ 0 & \text{otherwise} \end{cases}$$

$$n_H(t) = \sum_i \xi_i^H(t) = \text{number home players on ice at time } t$$

$$\omega(t) = \text{home team win probability at time } t$$



Computing Player *WPA*

- *WPA* for i^{th} player on home team equals

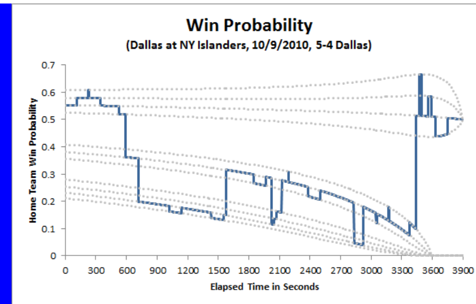
$$WPA_i^H = \int_0^\tau \frac{\xi_i^H(t)}{n_H(t)}\, d\omega(t)$$

- Analogous measure for i^{th} away player is

$$WPA_i^A = -\int_0^\tau \frac{\xi_i^A(t)}{n_A(t)}\, d\omega(t)$$

- For all players, $E(WPA) = 0$ since $E(d\omega(t)) = 0$

Example:



Islanders	WPA	Mins	Stars	WPA	Mins
Josh Bailey	0.1097	22.15	Trevor Daley	0.0700	25.78
Trent Hunter	0.0660	17.35	Adam Burish	0.0530	12.35
Doug Weight	0.0549	20.33	Steve Ott	0.0381	16.47
James Wisniewski	0.0353	24.17	Brad Richards	0.0348	24.58
Nino Niederreiter	0.0305	11.45	Brenden Morrow	0.0292	19.97
Frans Nielsen	0.0303	18.25	Toby Petersen	0.0272	11.38
Andrew MacDonald	-0.0007	23.58	James Neal	0.0271	20.42
Zenon Konopka	-0.0031	11.38	Matt Niskanen	0.0229	17.75
Trevor Gillies	-0.0049	1.95	Karlis Skrastins	0.0168	19.60
Matt Moulson	-0.0100	21.52	Loui Eriksson	0.0091	23.58
Jon Sim	-0.0113	12.52	Jamie Benn	-0.0097	13.23
Milan Jurcina	-0.0119	18.18	Brian Sutherby	-0.0129	6.55
Pa Parenteau	-0.0373	17.08	Stephane Robidas	-0.0141	26.05
John Tavares	-0.0467	5.30	Tom Wandell	-0.0147	10.25
Radek Martinek	-0.0470	20.60	Krystofer Barch	-0.0236	6.00
Mark Eaton	-0.0575	18.82	Mike Ribeiro	-0.0256	22.20
Blake Comeau	-0.0609	20.90	Nicklas Grossman	-0.0696	19.33
Mike Mottau	-0.0884	21.10	Mark Fistric	-0.1051	12.13

Team *WPA* Properties

- Total home team *WPA* satisfies

$$WPA^H = \sum_i WPA_i^H = \begin{cases} 0.45 & \text{home team wins} \\ -0.05 & \text{OT tie} \\ -0.55 & \text{home team loses} \end{cases}$$

- Analogous away team *WPA* satisfies

$$WPA^A = \sum_i WPA_i^A = \begin{cases} 0.55 & \text{away team wins} \\ 0.05 & \text{OT Tie} \\ -0.45 & \text{away team loses} \end{cases}$$

Team *WPA* Properties

- Obviously $E(WPA^H) = E(WPA^A) = 0$ but to see this directly:
  - Let p = Pr{Home Win in Reg or OT}
  - Let q = Pr{Game goes to a shootout}
- Then

$$E(WPA^H) = p^*(.45) + q^*(-.05) + (1-p-q)^*(-.55)$$

$$= p + 0.5^*q - 0.55$$

$$= 0 \text{ (since Pr\{Win\}} = p + 0.5^*q = 0.55)$$

Season Team *WPA*

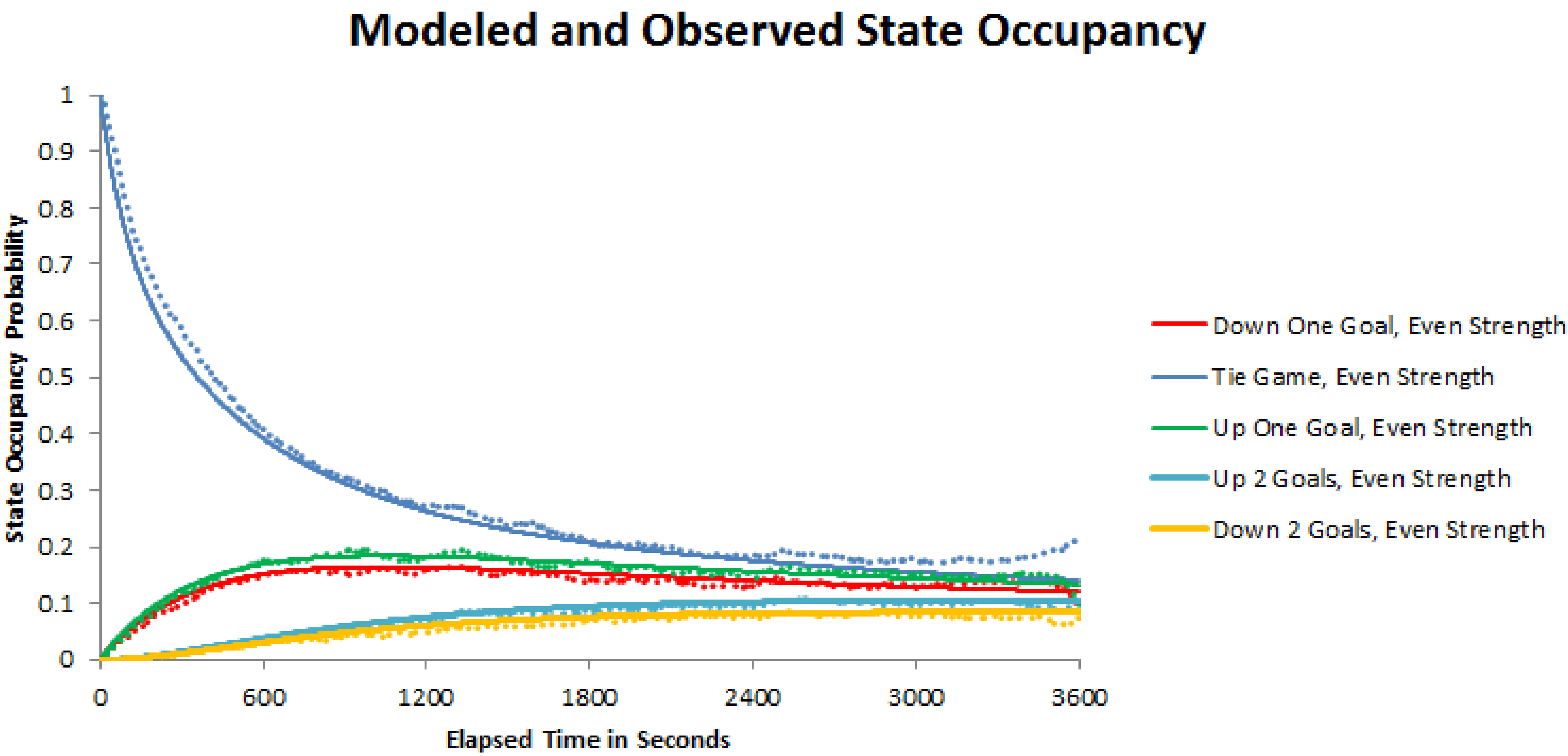
- Total team WPA over course of season

$$\text{Team } WPA = 0.45 \times \#HW - 0.05 \times \#HT - 0.55 \times \#HL$$
$$+ 0.55 \times \#AW + 0.05 \times \#AT - 0.45 \times \#AL$$

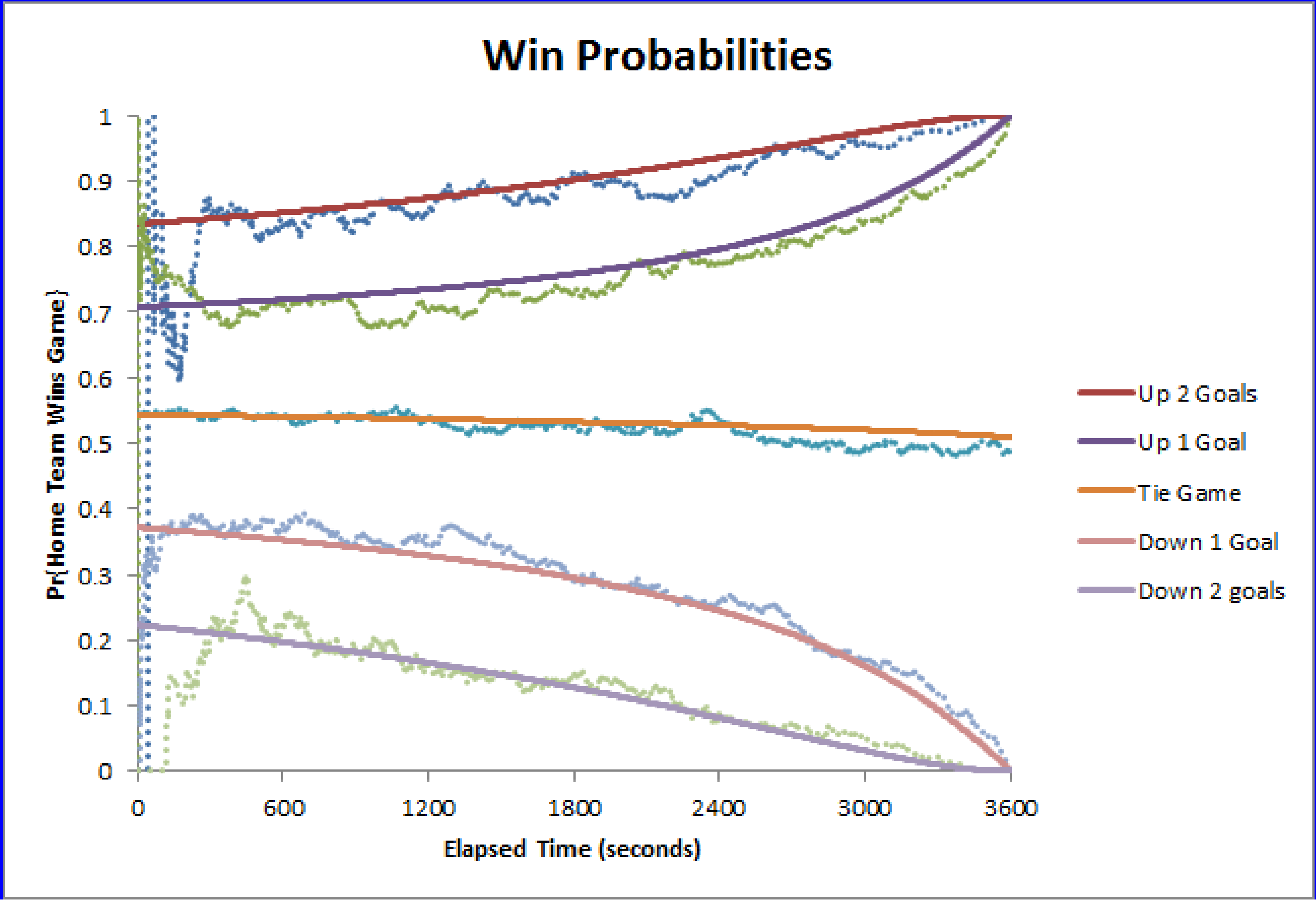
- But $\#HT = 41 - \#HW - \#HL; \#AT = 41 - \#AW - \#AL$, so

$$\begin{aligned}\text{Team } WPA &= 0.45 \times \#HW - 0.05 \times \#HT - 0.55 \times \#HL \\ &\quad + 0.55 \times \#AW + 0.05 \times \#AT - 0.45 \times \#AL \\ &= 0.5 \times \#HW - 0.5 \times \#HL - 0.5 \times 41 \\ &\quad + 0.5 \times \#AW - 0.5 \times \#AL + 0.5 \times 41 \\ &= 0.5(\text{Wins} - \text{Losses}) \\ &= \text{Wins} - 41 + \frac{\text{OT Ties}}{2}\end{aligned}$$

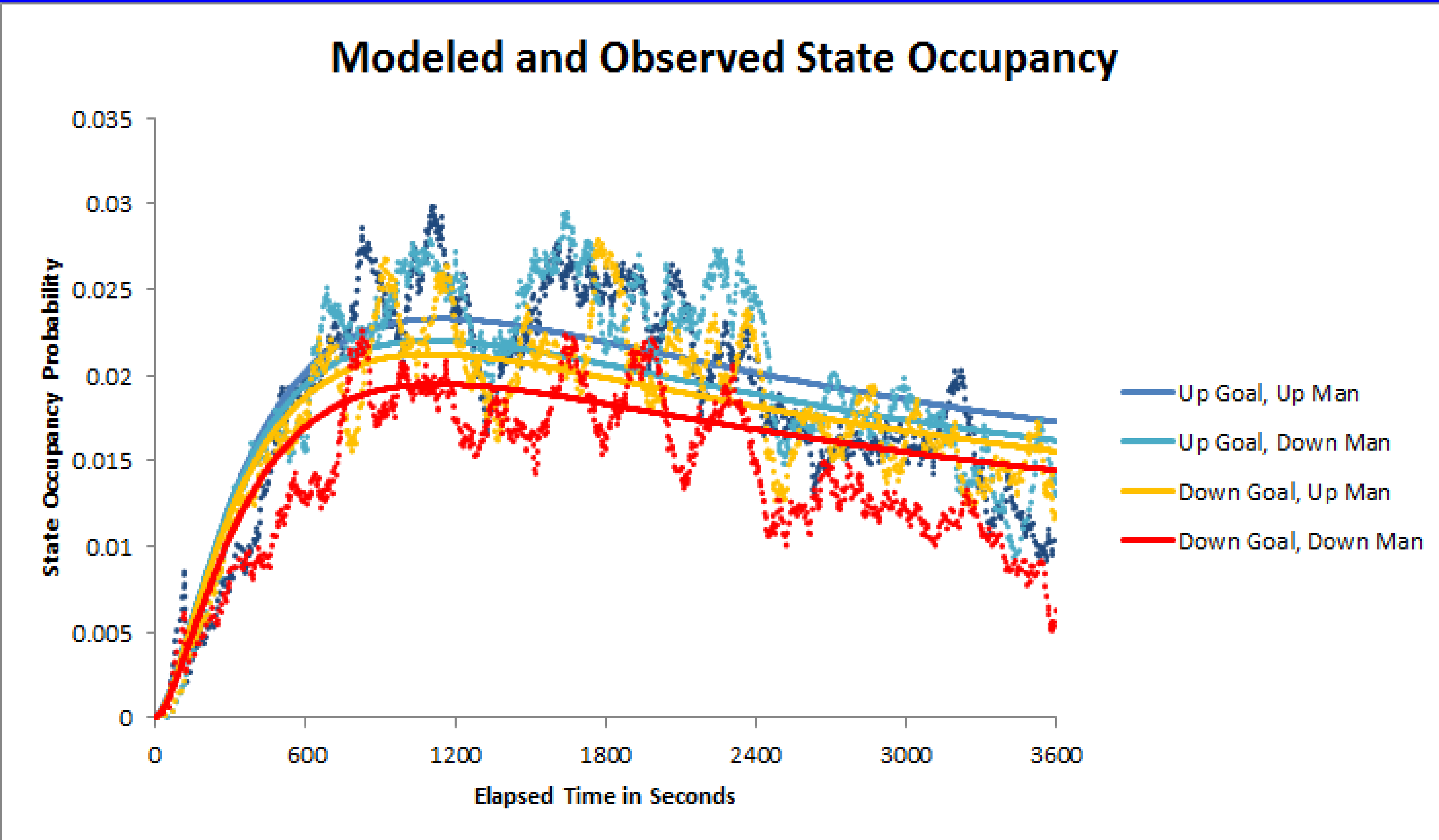
Observations Versus The Model



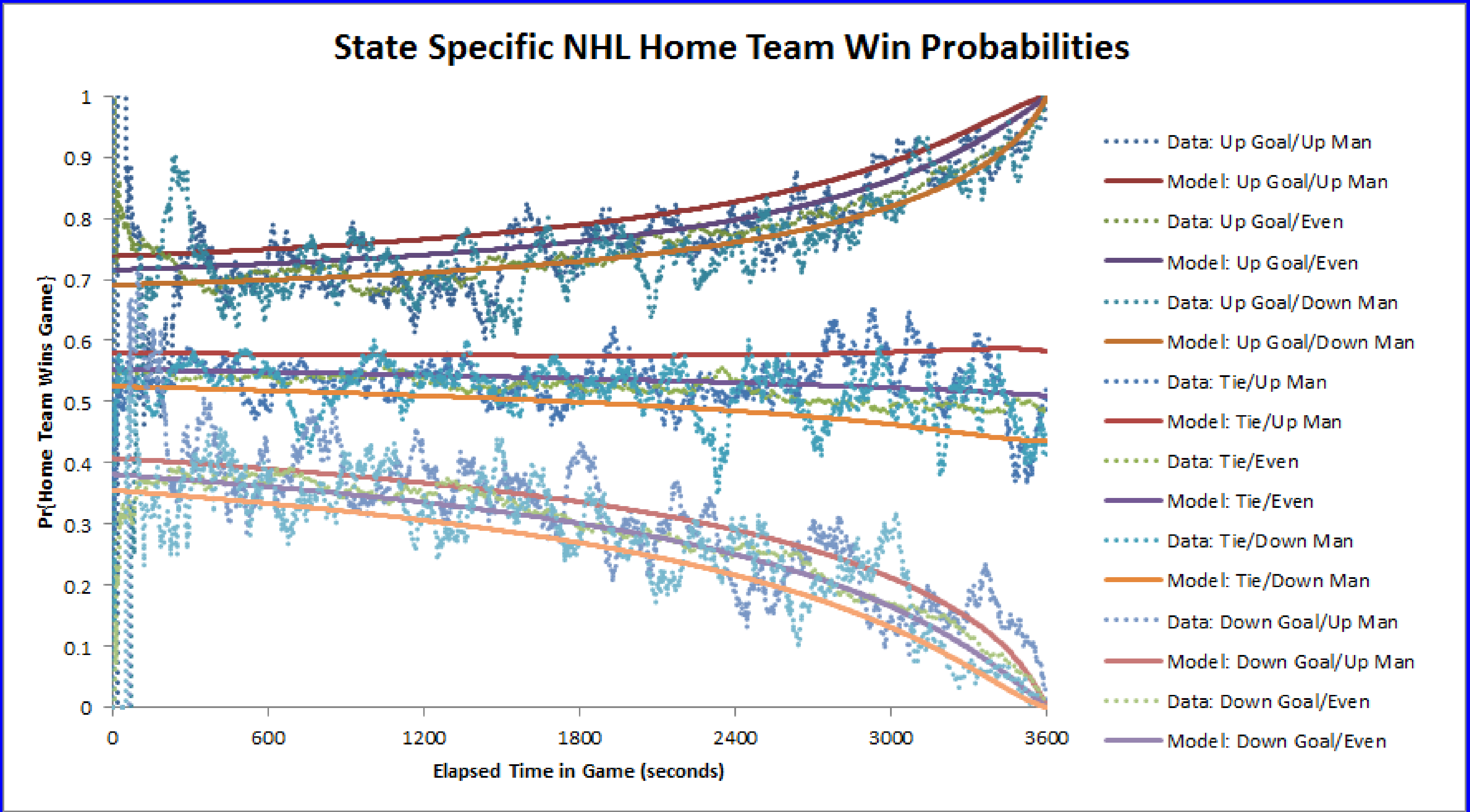
Observations Versus The Model



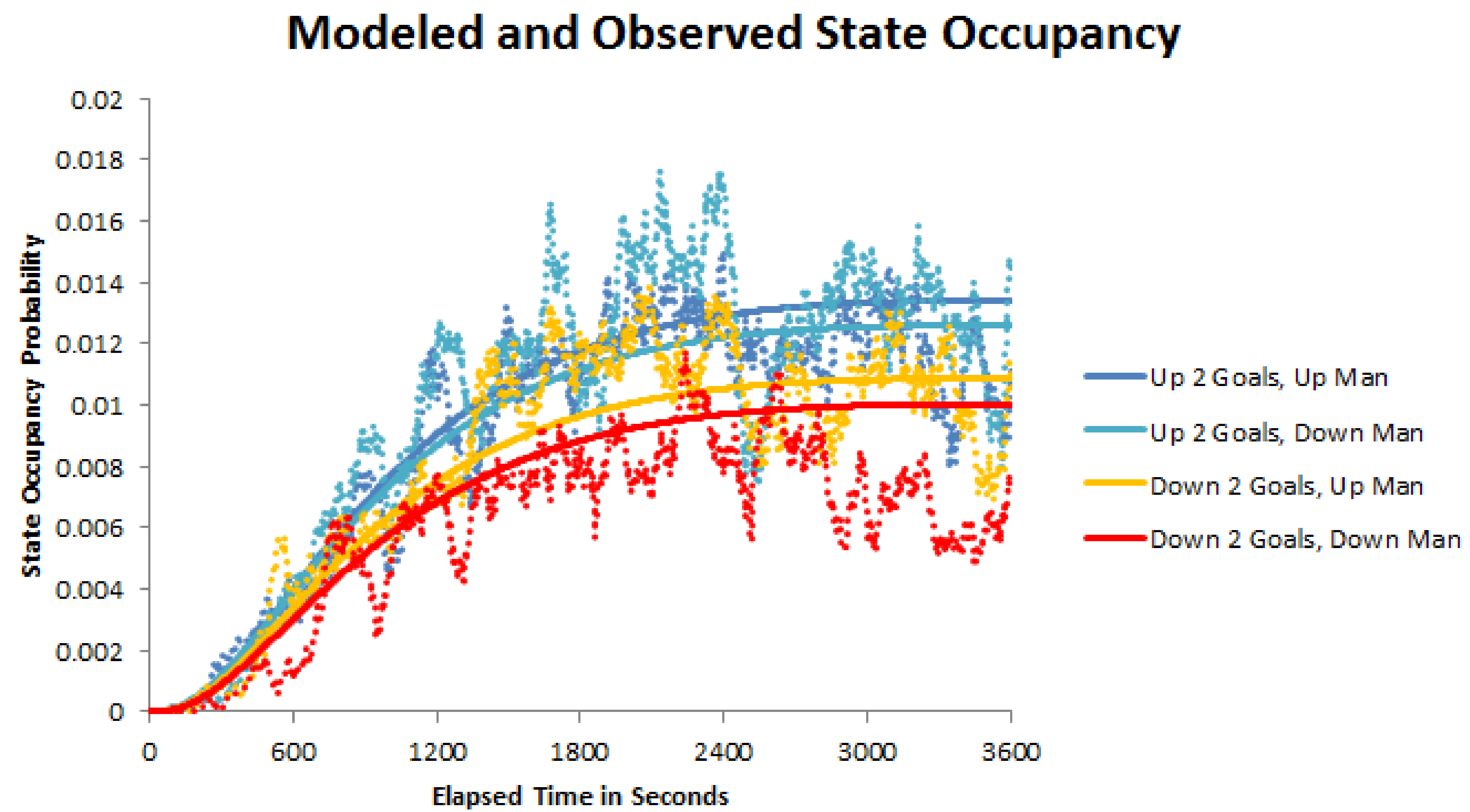
State Occupancy At Uneven Strength



Win Probability By Manpower State



Even Rarer State Occupancy



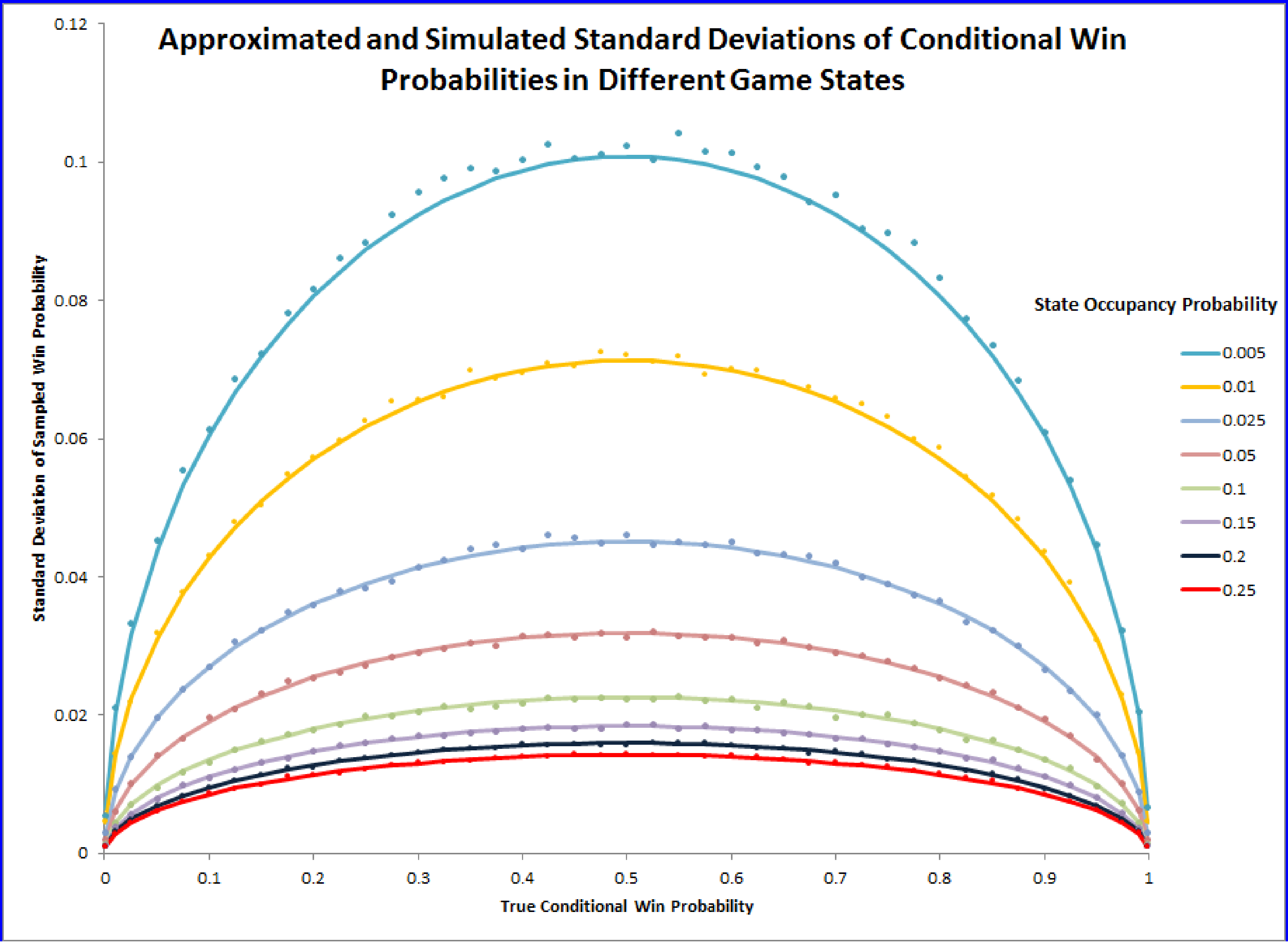
What Variability Should We Expect In Win Probability?

- Focus on some goal/manpower differential state s at an arbitrary time in game
- Define T_s as total games in that state at that time
- Define W_s as number times home team wins given that state at that time
- p_s (w_s) are (model derived) probability of being in state s (conditional probability of winning given s)
- T_s is binomial (4920, p_s), while W_s is binomial (T_s, w_s), and $Cov(T_s, W_s) = w_s Var(T_s)$

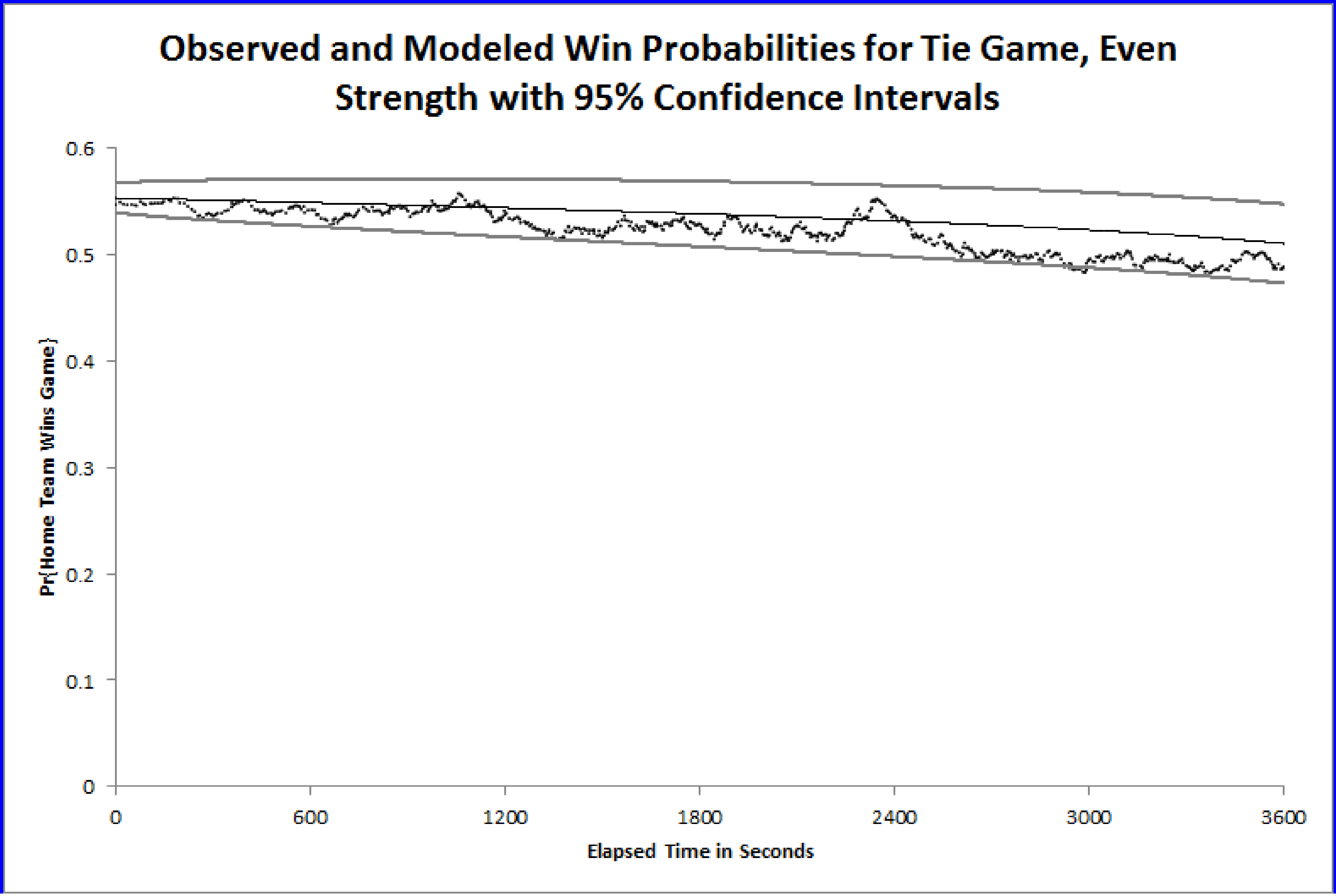
What Variability Should We Expect In Win Probability?

- Now define $R_s = W_s / T_s$ as empirical win probability (adjustment if $T_s = 0$)
- Then via delta method, R_s is approximately normal with mean w_s and variance equal to $w_s(1-w_s) / (4920p_s)$
- Delta method requires assuming that T_s, W_s are approximately bivariate normal, which works providing p_s is not too small
- Can always simulate if don’t trust normal approximation

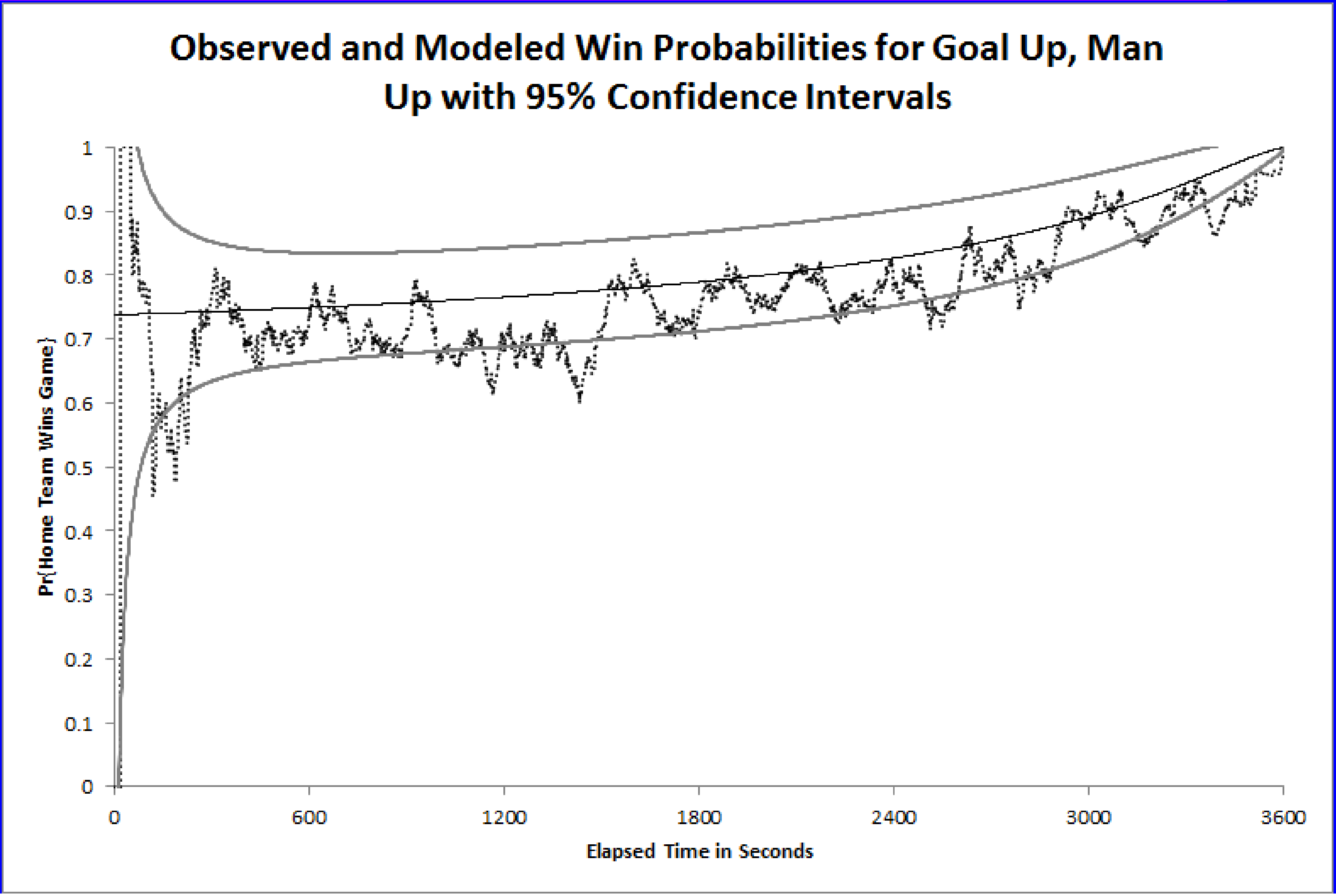
Model-Implied Variation in Empirical Win Probabilities



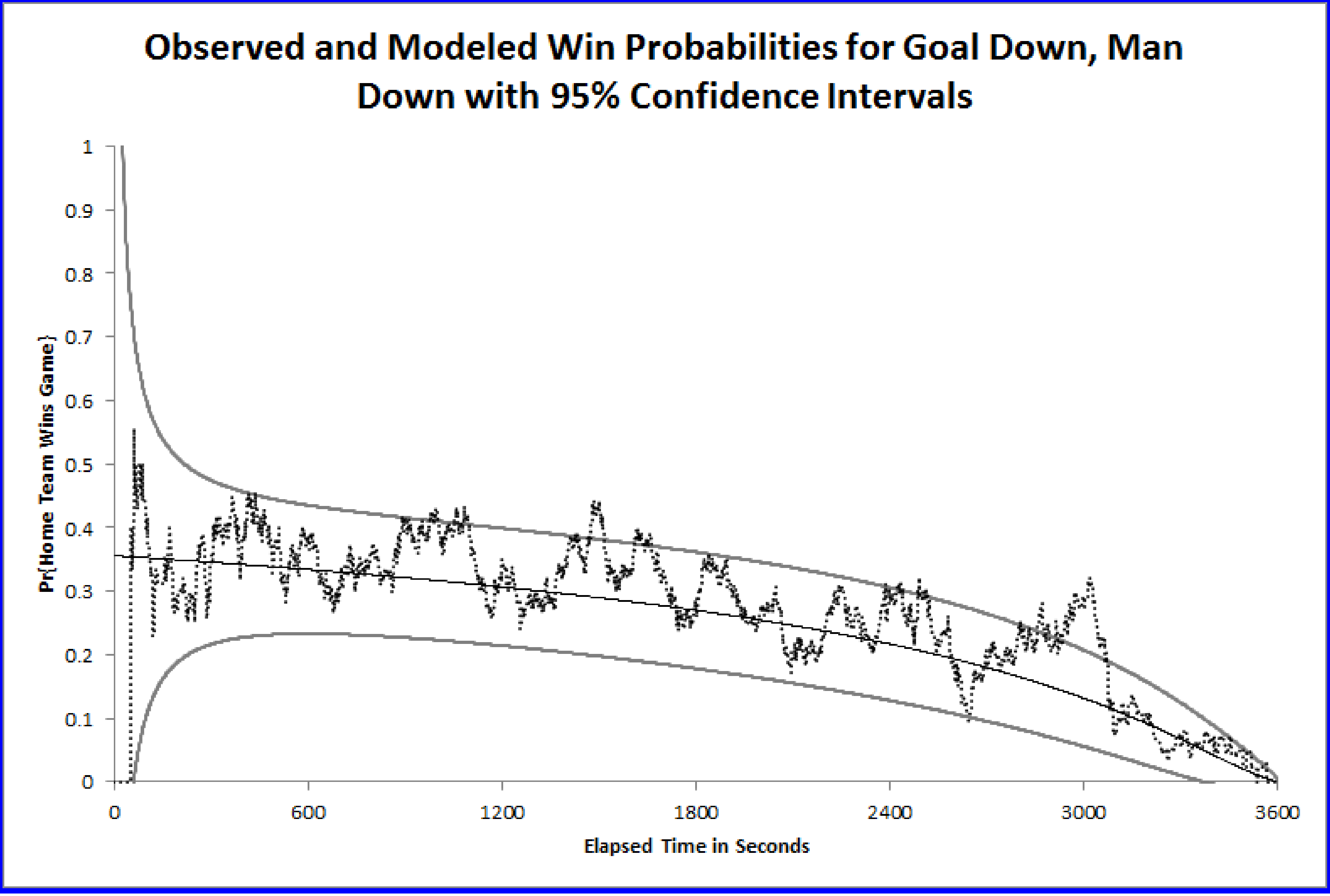
Accounting for Variability



Accounting for Variability



Accounting for Variability



Summing Up…

- New Markov model for hockey win probability that incorporates penalties/manpower differential
- Calibrated for four NHL seasons (4920 games)
- Showed that for NHL, a puck in the net beats four men in the box!
- Model can provide in-game win expectancies
- Model leads to new system for estimating player *WPA*