Compensation and Promotion Models: A New Econometrics Approach

Amos Golan
American University and Info-Metrics

Personnel and National security:
A Quantitative Approach

Johns Hopkins University, January 25-29, 2010

I thank the office of the Navy Personnel Research, Studies, and Technology and Dr.Tanja Blackstone for help and support.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JAN 2010**	2. REPORT TYPE	3. DATES COVERED **00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE **Compensation and Promotion Models: A New Econometrics Approach**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **American University,4400 Massachusetts Avenue, NW,Washington,DC,20016**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Personnel and National Security: A Quantitative Approach (Unclass), 25-28 January 2010, Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland. U.S. Government or Federal Rights License

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**65**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Objective:

Discuss econometric modeling of Personnel data.

Concentrate on “Empirical Modeling” of compensation, retention and promotion models.

Specifically:

- Retention – Bonuses - Compensation
- Promotion
- Promotion and Job Maching
- (Attrition)

Outline of Talk:

- Objective – Background
- Brief on Theory
  - Compensation – Bonus - ACOL
  - Promotion
- Data
- Problems with current methods
- Information-Theoretic Methods of Estimation
  - Formulation
- Empirical Examples
  - Promotion (CCMT)
  - MODCOMP
  - Others
- Summary and Points for Discussion

Background

Model Compensation – Retention models

- A Brief Theory
- ACOL- Annualized Cost of Living (Warner and Goldberger, 1984):

  The present value of expected military compensation for staying a certain period (or periods) in the military is compared to the present value of expected civilian earnings for that individual.

- ACOL – 2

  Improves on the ACOL model by "remembering" each individual‟s history (Panel)

- Dynamic programming approach generalizes the ACOL – 2 model

  At any decision point, weights (probabilities) are used to calculate the discounted pay differences over all possible future departure dates. Thus, the "single – dominant – horizon" assumption is relaxed. Even though this approach may provide superior and more accurate estimates, it is computationally complex and inefficient. Even worse, it (usually) builds on very strong assumptions that may contradict economic intuition/theory.

- From a practical point of view, a number of researchers show that the multivariate (or bivariate) probit model (a panel probit) may outperform the dynamic programming type methods and is easier to handle.
- Empirical Problems with the ACOL Model:
- ACOL variable‟s Assumptions (very strong – mean characteristics of individuals: discount factor, promotion probs., life expectancy, maximal length of service, civilian wages, etc.)
- Other issues:

  Endogeneity

Promotion models

Depends on the objective and on the available data

Examples:

1. Conditional Markov Model (CCMT)
2. Conditional Markov Model (CCMT) with Environmental Effects
3. Two-stage model (Promotion, Reenlistment)

Personnel Data

Problems with data:

Highly collinear

A large number of dummy/binary variables

Could be very small

- Econometrics Framework:

Class of Discrete Choice Models

But

- ML may not work in some cases
- Strong assumptions
- ML (when works) may have poor predictive power

Other

- Endogeneity in both compensation and promotion models

What do we want from a good model?

Flexibility
Efficiency
Easy to apply
Can handle small/large data
Minimal assumptions on the DGP
Can handle highly collinear data
Can handle events with (very) small probability
Can incorporate priors and soft data (when/if available)

So we propose:

Information-Theoretic Methods of Estimation

(A promotion model is presented here – the compensation model is a special case of that one.)

Promotion Model (CCMT)

Objective of the model;

- Using Navy enlisted personnel data, we develop an empirical model for estimating the joint probabilities that an individual will achieve a variety of career goals within a particular time conditional on the individual‟s record, the standing of the individual within her cohort and other factors outside of the individual‟s control.
- These career goals include promotion, geographic stability, easier tasks, next assignment, as well as education and training opportunities.

Motivation

- Policy maker can use the analysis to study (and forecast) the force and the impact of Navy policies, wars and the state of the economy on that force.
- Identify the personal attributes of fast promoters and early exiters (at all levels and skills)
- Identify discrimination (if exists)
- Providing each sailor with the possibility of studying her/his potential career trajectory periodically (and consider potential scenarios).
- Simulate policies and scenarios

The Basic Promotion Path

Chart 1. The Time Dimension

```
|------|-----------|-------|--------|------------|-------------|---------------------------|------ ----|---→ time,  t
    E4        Min     E5   Min          E6          Min                          E7       Min
                     Max                Max                                      Max
                     Quit               Quit                                     Quit
```

E4 – E7 – Different pay grades for enlisted personnel (lowest is E1)

The Basic Promotion Path - The Discrete, Simplified Version

The Hierarchy of States

From E3 to E4
- Fast
- Slow
- Exit

From E4 to E5
- Fast
  - Fast before (t-1)[1]
  - Slow before (t-1)
- Slow
  - Fast before (t-1)
  - Slow before (t-1)
- Exit
  - Fast before (t-1)
  - Slow before (t-1)

From E5 to E6
- Fast
  - Fast before (t-1)
    - Fast before (t-2)
    - Slow before (t-2)
  - Slow before (t-1)
    - Fast before (t-2)
    - Slow before (t-2)
- Slow
  - Fast before (t-1)
    - Fast before (t-2)
    - Slow before (t-2)
  - Slow before (t-1)
    - Fast before (t-2)
    - Slow before (t-2)
- Exit
  - Fast before (t-1)
    - Fast before (t-2)
    - Slow before (t-2)
  - Slow before (t-1)
    - Fast before (t-2)
    - Slow before (t-2)

And so on

[1] The term “t-1” means the previous promotion for that individual. The term “t-2” means the promotion before the last for that individual.

The Empirical Model

The Basic Promotion Model – First Order Markov

Let y_{itj} be state *j* (grade level j=E3, …, E9) of individual *i* in period *t*. For each individual *i* (*i=1, 2, …,n*) let $y_{itj}=1$ if state *j* (*j=1, 2, …, K*) is observed at period *t* (*t=1, 2, …, T*), and $y_{itj}=0$ for all other *K-1* states.

Next, define the K by K matrix of transition (promotion) probabilities P (p_{kj}) representing the probability of promotion from state (Pay grade) *k* to state (Pay grade) *k+1=j*.

<u>Note:</u> One of the *K* states represents (voluntary or non-voluntary) exit.

The basic relationship between period (*t-1)* and period *t* is captured via the (K by K) matrix of transition (promotion) probabilities

$$y_{itj} = \sum_{\{y_{i,t-1,k}=0\}} p_{kj}\, y_{i,t-1,k} + \sum_{\{y_{i,t-1,k}\neq 0\}} p_{kj}\, y_{i,t-1,k}$$

$$\equiv \sum_{k=1}^{K} p_{kj}\, y_{i,t-1,k}$$

Taking into account the noise in the observed data, the correct noisy observed model is

$$y_{itj} = \sum_{\{y_{i,t-1,k}=0\}} p_{kj} y_{i,t-1,k} + \sum_{\{y_{i,t-1,k}\neq 0\}} p_{kj} y_{i,t-1,k} + \varepsilon_{itj}$$

$$\equiv \sum_{k=1}^{K} p_{kj} y_{i,t-1,k} + \varepsilon_{itj}$$

where $\varepsilon \in [-1, 1]$

Introducing the Individual Level Information

- There are two types of individual level covariates in this model: Time Dependent and Time Independent.
- To capture the relationship between the observed data, ***y***, the unknown probabilities *P*, and the covariates *X*, we introduce the following (cross moments) relationship:

$$\sum_{t=2}^{T}\sum_{i=1}^{N} y_{itj} x_{its} = \sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{\{y_{i,t-1,k}=0\}}^{K} p_{kj} x_{i,t-1,s} + \sum_{t=2}^{T}\sum_{i=1}^{N}\sum_{\{y_{i,t-1,k}\neq 0\}}^{K} p_{kj} y_{i,t-1,k} x_{i,t-1,s} + \sum_{t=2}^{T}\sum_{i=1}^{N} \varepsilon_{i,t-1,j} x_{i,t-1,s}$$

$$\equiv \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{\{y_{itk}=0\}}^{K} p_{kj} x_{its} + \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{\{y_{itk}\neq 0\}}^{K} p_{kj} y_{itk} x_{its} + \sum_{t=1}^{T-1}\sum_{i=1}^{N} \varepsilon_{itj} x_{its}$$

$$= \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{k=1}^{K} p_{kj} y_{itk} x_{its} + \sum_{t=1}^{T-1}\sum_{i=1}^{N} \varepsilon_{itj} x_{its}$$

Introducing the Global Variables

- The next step is to extend this framework to include economic and political (environmental) variables.
- Let Z be a *G*-dimensional vector of global variables for each period *t*, composed of *G* macro and policy variables

- <u>The General Formulation - Case A</u>

$$\sum_{t=2}^{T}\sum_{i=1}^{N} y_{itj} x_{its} z_{tg} = \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{k=1}^{K} p_{kj} y_{itk} x_{its} z_{tg} + \sum_{t=1}^{T-1}\sum_{i=1}^{N} \varepsilon_{itj} x_{its} z_{tg}$$

The Information-Theoretic - GME Estimation Model

The Basic Model

Rewriting the basic Eq. such that all unknowns are proper probabilities:

$$\sum_{t=2}^{T}\sum_{i=1}^{N} y_{itj}x_{its} = \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{k=1}^{K} p_{kj}y_{itk}x_{its} + \sum_{t=1}^{T-1}\sum_{i=1}^{N} x_{its}\varepsilon_{itj}$$
$$= \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{k=1}^{K} p_{kj}y_{itk}x_{its} + \sum_{t=1}^{T-1}\sum_{i=1}^{N}\sum_{m=1}^{M} x_{its}w_{itjm}v_{m}$$

with $\varepsilon_{itj} \equiv \sum_{m=1}^{M} w_{itjm}v_m$ and where $\sum_{m=1}^{M} w_{itjm} = 1$ for M>=2. Since $\varepsilon_{itj} \in [-1,1]$

for all *i, t, j*, then $v_m \in [-1,1]$ and ***v*** is a symmetric around zero

support space for each random error defined above.

(**v** is Poisson.)

The Information Theoretic, General Maximum Entropy (IT-GME) Model:

(Maximizing the joint **entropies** of the signal and noise subject to the linear cross moments, First Order Markov conditions, and the requirements that **w** are proper distributions)

$$\underset{\{p,w\}}{Max}\left\{-\sum_{k,j} p_{kj}\log p_{kj} - \sum_{itjm} w_{itjm}\log w_{itjm}\right\}$$

$s.t.$

$$\sum_{t=2}^{T}\sum_{i} y_{itj}x_{its} = \sum_{t=1}^{T-1}\sum_{k=1}^{K} p_{kj}y_{itk}x_{its} + \sum_{t=1}^{T-1}\sum_{i}\sum_{m} w_{itjm}x_{its}v_m$$

$$\sum_{j} p_{kj} = 1; \sum_{m} w_{itjm} = 1$$

Solution:

$$\hat{p}_{kj} = \frac{\exp\left(-\sum_{t=1}^{T-1}\sum_{i,s} y_{itk} x_{its} \hat{\lambda}_{js}\right)}{\sum_{j}\exp\left(-\sum_{t=1}^{T-1}\sum_{i,s} y_{itk} x_{its} \hat{\lambda}_{js}\right)} \equiv \frac{\exp\left(-\sum_{t=1}^{T-1}\sum_{i,s} y_{itk} x_{its} \hat{\lambda}_{js}\right)}{\Omega_k}$$

$$\hat{w}_{itjm} = \frac{\exp\left(-\sum_{s} x_{its} v_m \hat{\lambda}_{js}\right)}{\sum_{m}\exp\left(-\sum_{s} x_{its} v_m \hat{\lambda}_{js}\right)} \equiv \frac{\exp\left(-\sum_{s} x_{its} v_m \hat{\lambda}_{js}\right)}{\Psi_{itj}}$$

The Concentrated Model:

$$\ell(\lambda) = \sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{i,s} y_{itj} x_{its} \lambda_{sj} + \sum_{k} \log\left[\sum_{j} \exp\left(-\sum_{t=1}\sum_{i,s} y_{itk} x_{its} \lambda_{sj}\right)\right]$$

$$+\sum_{i,t,j} \log\left[\sum_{m} \exp\left(-\sum_{s} x_{its} v_{m} \lambda_{sj}\right)\right]$$

$$= \sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{i,s} y_{itj} x_{its} \lambda_{sj} + \sum_{k} \log \Omega_{k}(\lambda) + \sum_{i,t,j} \log \Psi_{itj}(\lambda),$$

Solution of IT-GME/GCE Model with Priors:

$$\tilde{p}_{kj} = \frac{p_{kj}^{0} \exp\left(\sum_{t=1}^{T-1} \sum_{i,s,g} y_{itk} x_{its} z_{tg} \tilde{\lambda}_{jsg} \right)}{\sum_{j} p_{kj}^{0} \exp\left(\sum_{t=1}^{T-1} \sum_{i,s,g} y_{itk} x_{its} z_{tg} \tilde{\lambda}_{jsg} \right)}$$

$$\equiv \frac{p_{kj}^{0} \exp\left(\sum_{t=1}^{T-1} \sum_{i,s,g} y_{itk} x_{its} z_{tg} \tilde{\lambda}_{jsg} \right)}{\Omega_{k}}$$

The Concentrated IT-GME/GCE Model with Priors (A Generalized ML-Logit for the Markov Model):

$$\ell(\lambda) = \sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{i,s,g} y_{itj} x_{its} z_{tg} \lambda_{sjg} - \sum_{k} \log\left[\sum_{j} p_{kj}^{0} \exp\left(\sum_{t=1}\sum_{i,s,g} y_{itk} x_{its} z_{tg} \lambda_{sjg}\right)\right]$$

$$-\sum_{i,t,j} \log\left[\sum_{m} w_{itjm}^{0} \exp\left(\sum_{s,g} x_{its} z_{tg} v_{m} \lambda_{sjg}\right)\right]$$

$$= \sum_{t=2}^{T}\sum_{j=1}^{K}\sum_{i,s,g} y_{itj} x_{its} z_{tg} \lambda_{sjg} - \sum_{k} \log \Omega_{k}(\lambda) - \sum_{i,t,j} \log \Psi_{itj}(\lambda)$$

(Note: same level of complexity as ML)

Notes on Empirical Model

- Censoring Issues (Left and Right)
- Estimation is done separately for each Pay Grade
- Errors' Support
- Other Possible Models
- "Actual Promotions" vs. "Selected for Promotion"

(Note: We Skip the Data Description)

The Skill Groups Analyzed (Covers all Navy Enlisted Personnel)

- Administration
- Submariner Electronic
- Surface Combat Electronics
- Surface Electrical
- Surface Engineering
- Surface Operations
- Crypto Intel
- Diver Spec War
- Nuclear
- Aviation Mechanical (3 subgroups)
- Aviation Air Crew
- Aviation Boatswain
- Aviation ATC
- Aviation Meteorologist
- Seabee
- Submariner Other
- Surface Combat Weapons
- Surface Repair
- Surface Deck
- Medical (2 subgroups)
- Supply (2 subgroups)

*Mapped based on Job families, functional work, SME input, instructions
Statistically validated.
**187 Skill groups aggregated to 22 groups
***Personnel and promotion files 1996-2005

Empirical Analysis and Results
(A partial example)

Administration (with transfers)

(Transfers are defined within or out of skill group)

Predicted Transition Probabilities

	E-3	E-4	E-5	E-6	E-7	LOSS
E-3	29.6	42.7	0.0	0.0	0.0	27.7
E-4	0.0	63.1	28.1	0.0	0.0	8.7
E-5	0.0	0.0	71.0	18.6	0.0	10.3
E-6	0.0	0.0	0.0	74.9	11.2	13.9

Within 1 year, 0.296 probability (29.6%) E-3 is not advanced

Within 1 year, 0.427 probability (42.7%) E-3 is advanced to E-4

Within 1 year, 0.277 probability (27.7%) E-3 will leave the force

Within 1 year, 0.186 probability (18.6%) E-5 is advanced to E-6

Administration Skill Group. Transition Matrix from E3 through E7 for T=3 and 5 periods into the future.

T= 3

	E3	E4	E5	E6	E7	Loss
E3	0.0260	0.2875	0.1968	0.0224	0.0000	0.4674
E4	0.0000	0.2515	0.3802	0.1096	0.0059	0.2529
E5	0.0000	0.0000	0.3583	0.2977	0.0512	0.2927
E6	0.0000	0.0000	0.0000	0.4202	0.2582	0.3216
E7	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000

T= 5

	E3	E4	E5	E6	E7	Loss
E3	0.0023	0.1248	0.2109	0.0812	0.0085	0.5724
E4	0.0000	0.1002	0.2867	0.1781	0.0352	0.3998
E5	0.0000	0.0000	0.1808	0.2645	0.1169	0.4378
E6	0.0000	0.0000	0.0000	0.2357	0.3403	0.4239
E7	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000

Nuclear Skill Group - Transition Matrix from E4 through E7

	E4	E5	E6	E7	Loss
E4	0.369	0.591	0.000	0.000	0.041
E5	0.000	0.497	0.176	0.000	0.327
E6	0.000	0.000	0.571	0.148	0.282

Nuclear Skill Group - Priors used.

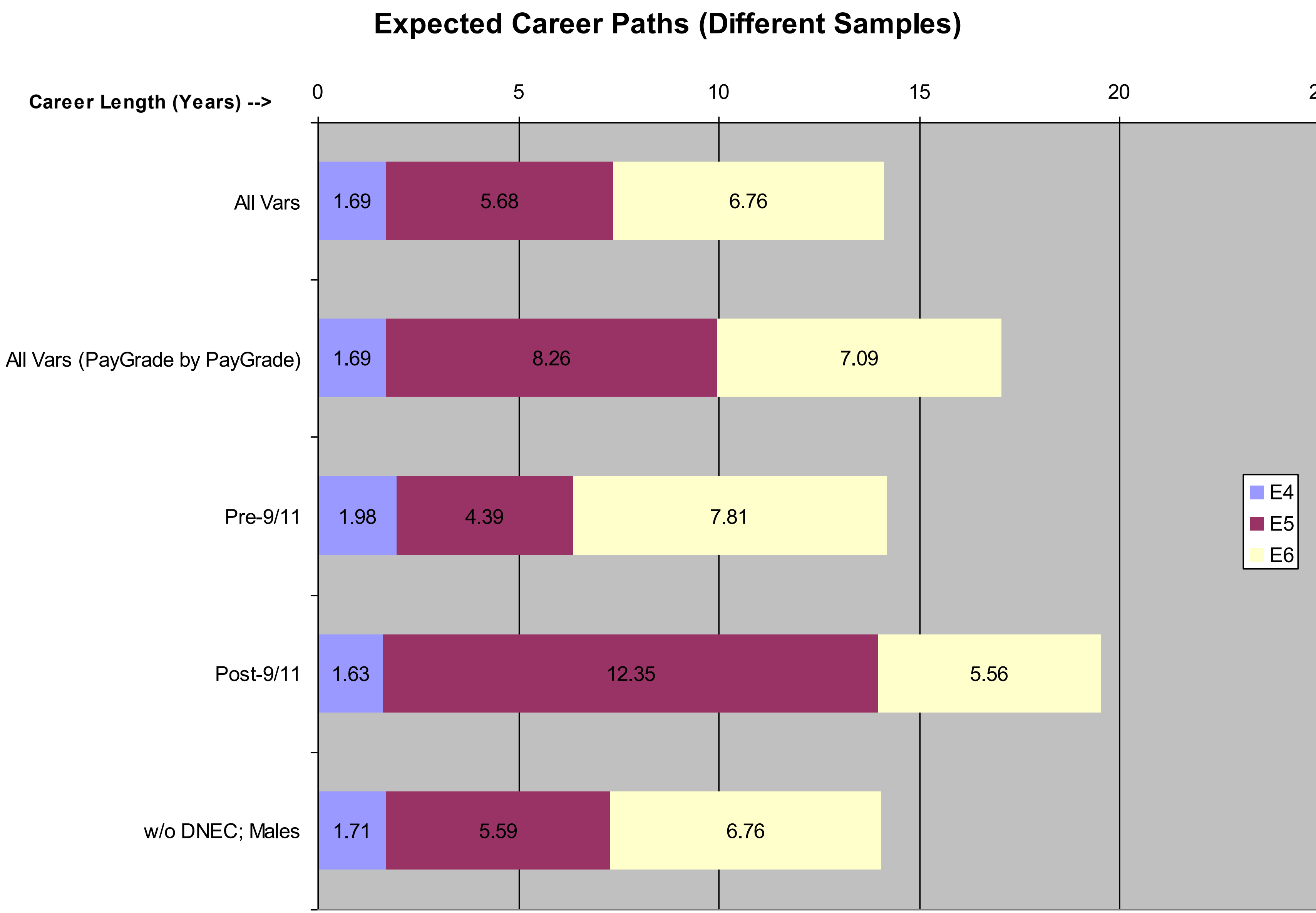
	E4	E5	E6	E7	Loss
E4	0.429	0.470	0.000	0.000	0.101
E5	0.011	0.589	0.121	0.000	0.279
E6	0.000	0.003	0.753	0.082	0.162

Transition Probs. – Pre War (Sep. 11)

	E4	E5	E6	E7	Loss
E4	0.382	0.506	0.000	0.000	0.112
E5	0.000	0.460	0.228	0.000	0.312
E6	0.000	0.000	0.487	0.128	0.385

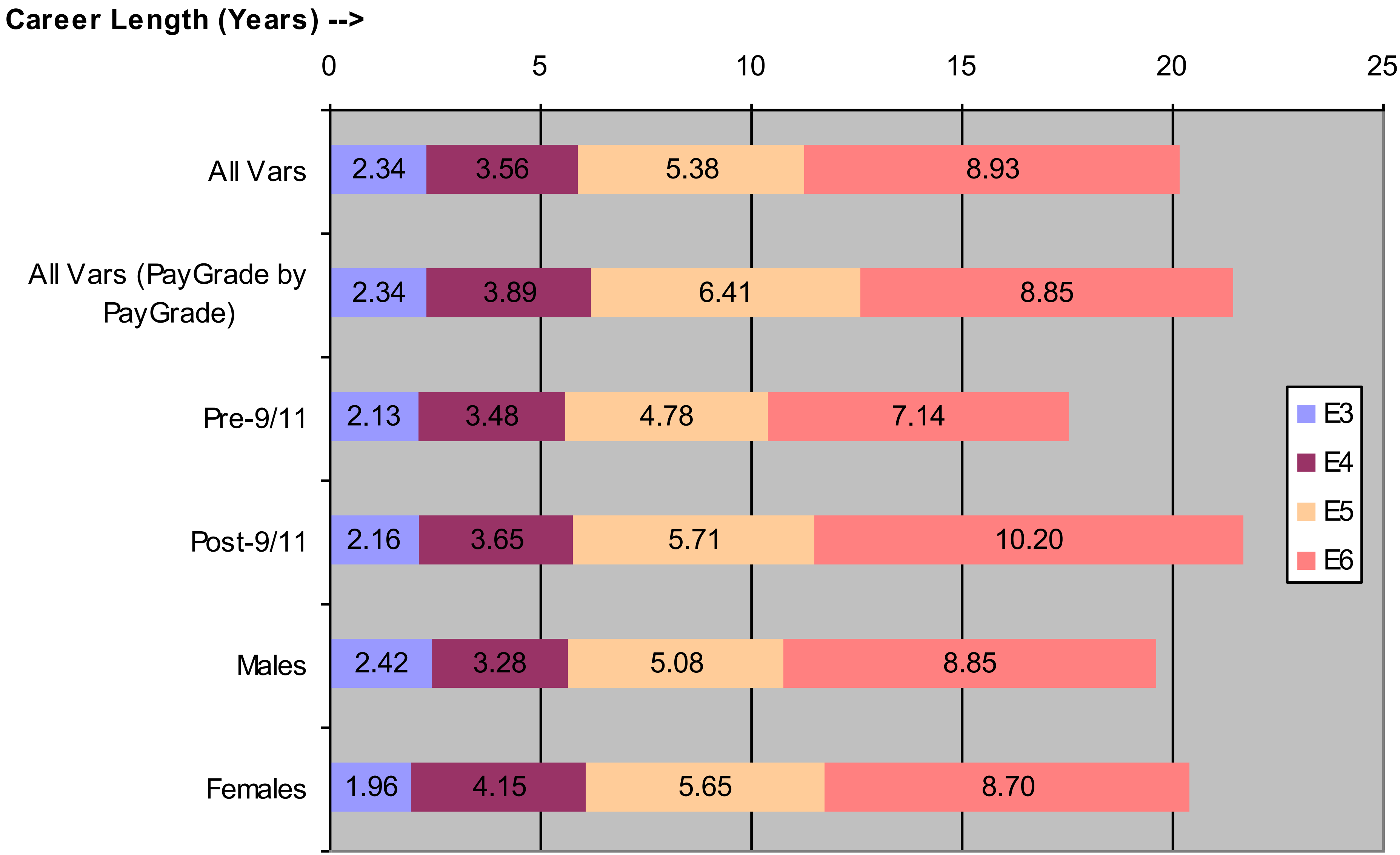
Transition Probs. – Post Sep. 11

	E4	E5	E6	E7	Loss
E4	0.386	0.614	0.000	0.000	0.000
E5	0.000	0.590	0.081	0.000	0.329
E6	0.000	0.000	0.648	0.180	0.173



Expected Career Paths (Different Samples)
Career Length (Years) -->
0
5
10
15
20
25
All Vars
1.69
5.68
6.76
All Vars (PayGrade by PayGrade)
1.69
8.26
7.09
Pre-9/11
1.98
4.39
7.81
Post-9/11
1.63
12.35
5.56
w/o DNEC; Males
1.71
5.59
6.76
E4
E5
E6

Figure 4. Expected Career Paths - Administration



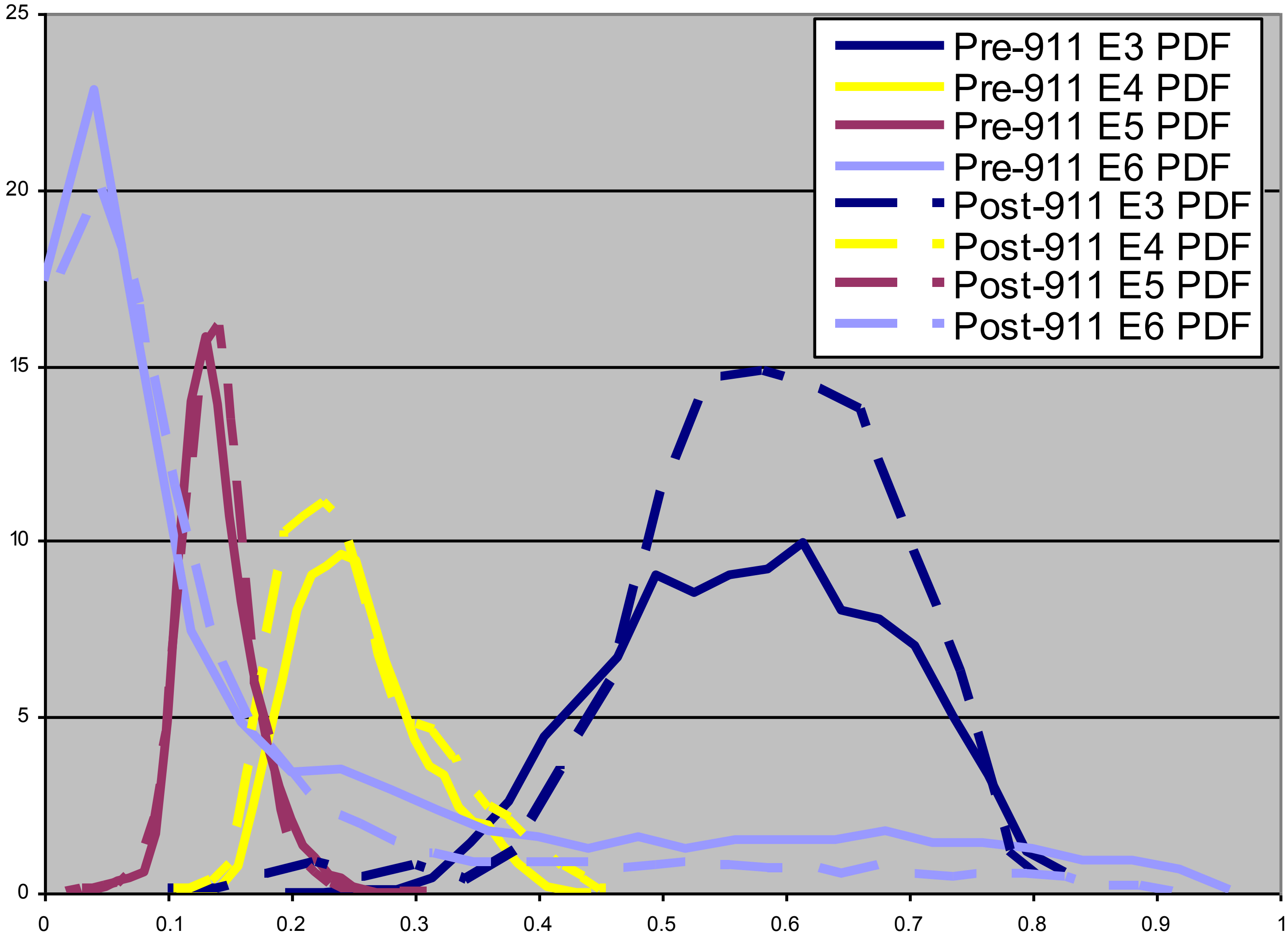
Brief analysis of the relative explanatory power (contribution to the explanation), and reduction of uncertainty, of each group of variables used.

- Individual attributes include: education vars, marital status, IQ, „patience" (if individual moved much), and performance.
- Environmental information includes: macro level indicators.
- Policy information includes: Navy demand and supply at each promotion cycle.

Results: individual attributes contribute approx. one half; environmental and policy variables contribute about one fourth each.

	Individual	**Environmental (econ)**	**Navy/Policy**
Admin	53.6%	21.4%	25.0%
Nuclear	60.2%	18.4%	21.4%

Pre and post war estimated probabilities (for all individuals) of promotion within a year - Administration



Probability of Loss within a year: Simulated individual, environmental and policy experiments – Nuclear.

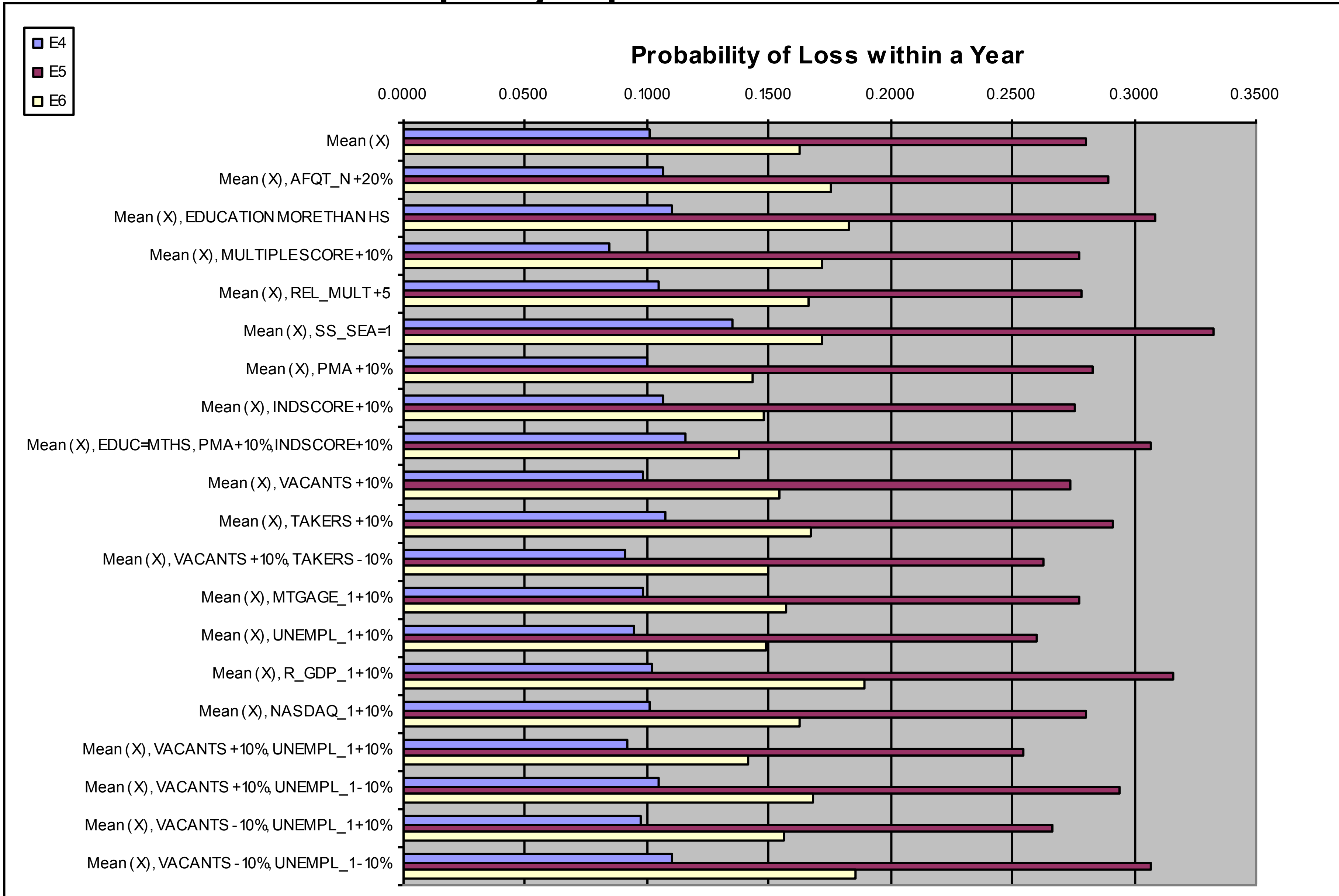


Figure 9A. Average Transition Probability Trajectories (7 years out), Skill Group = Medical, Starting State = E3

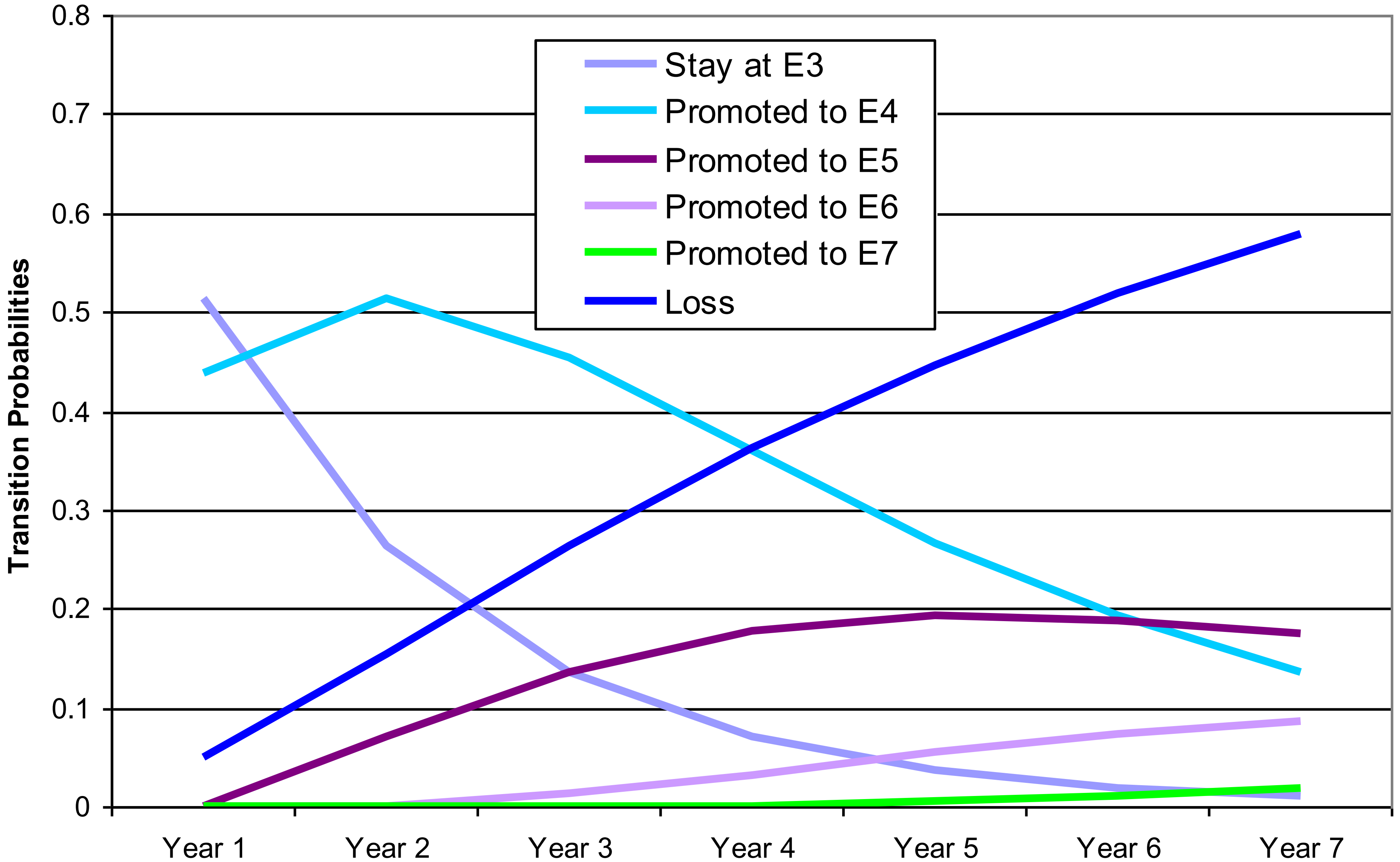


Figure 9B. Average Transition Probability Trajectories (7 years out), Skill Group = AIRCREW, Starting State = E3

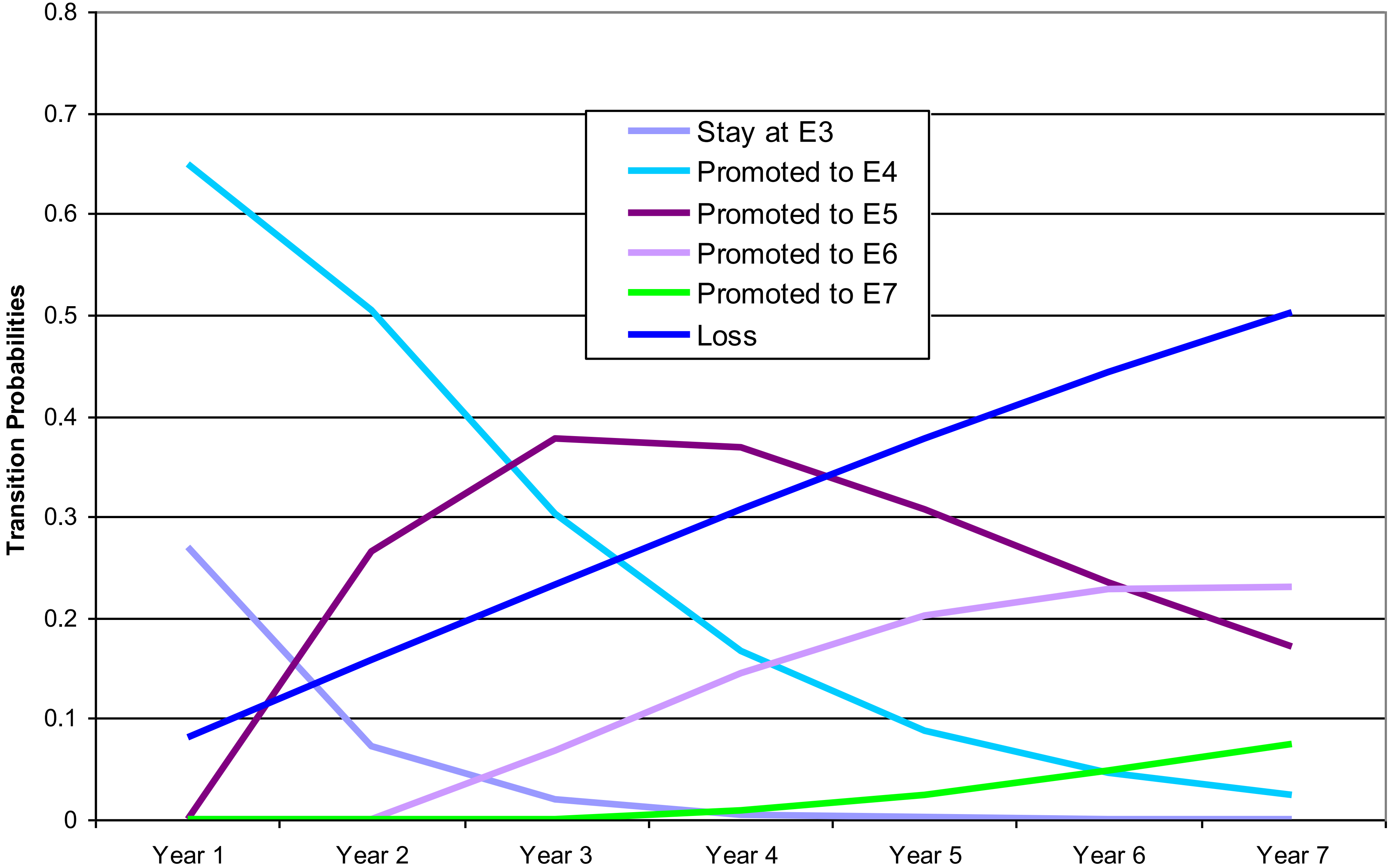
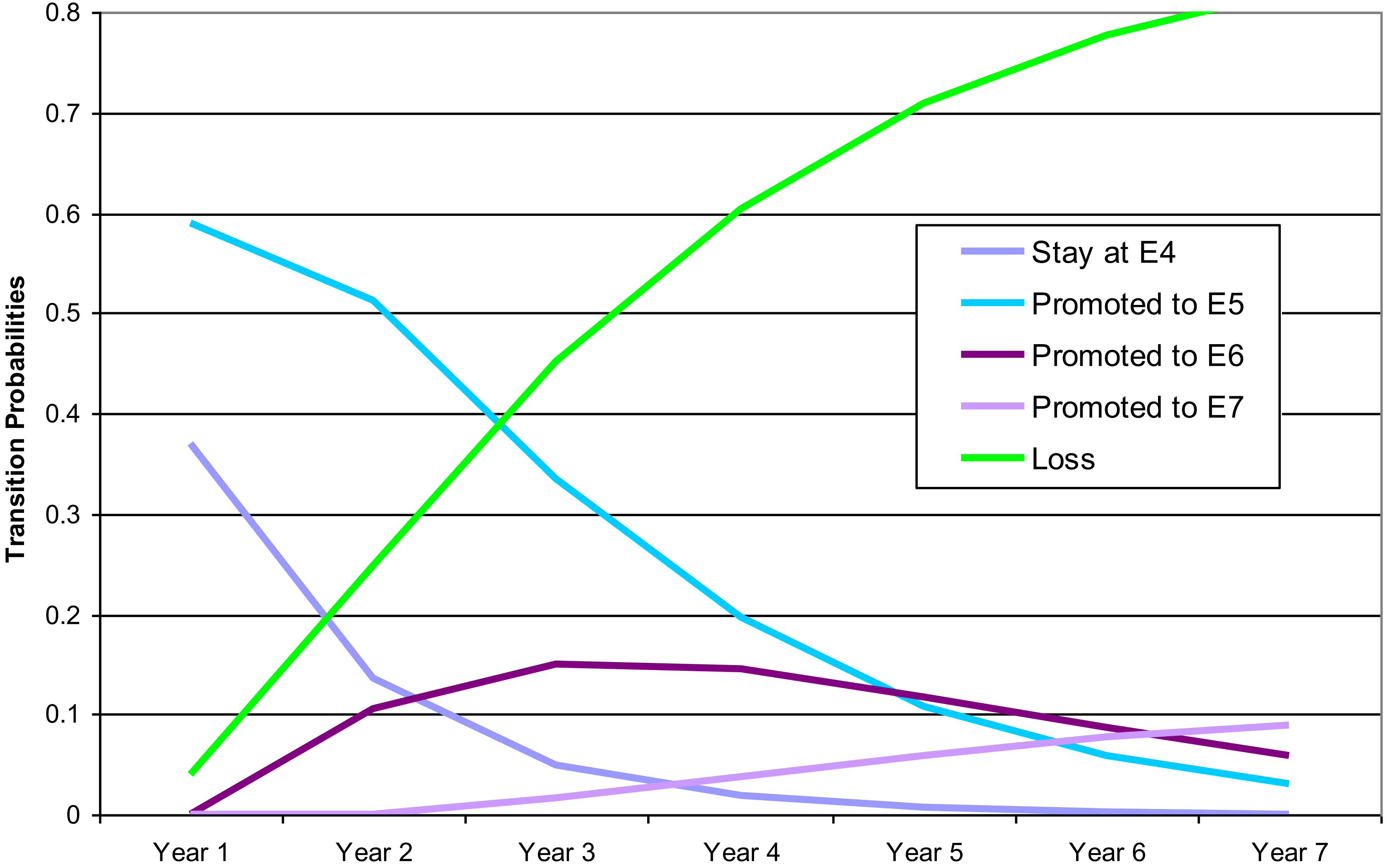
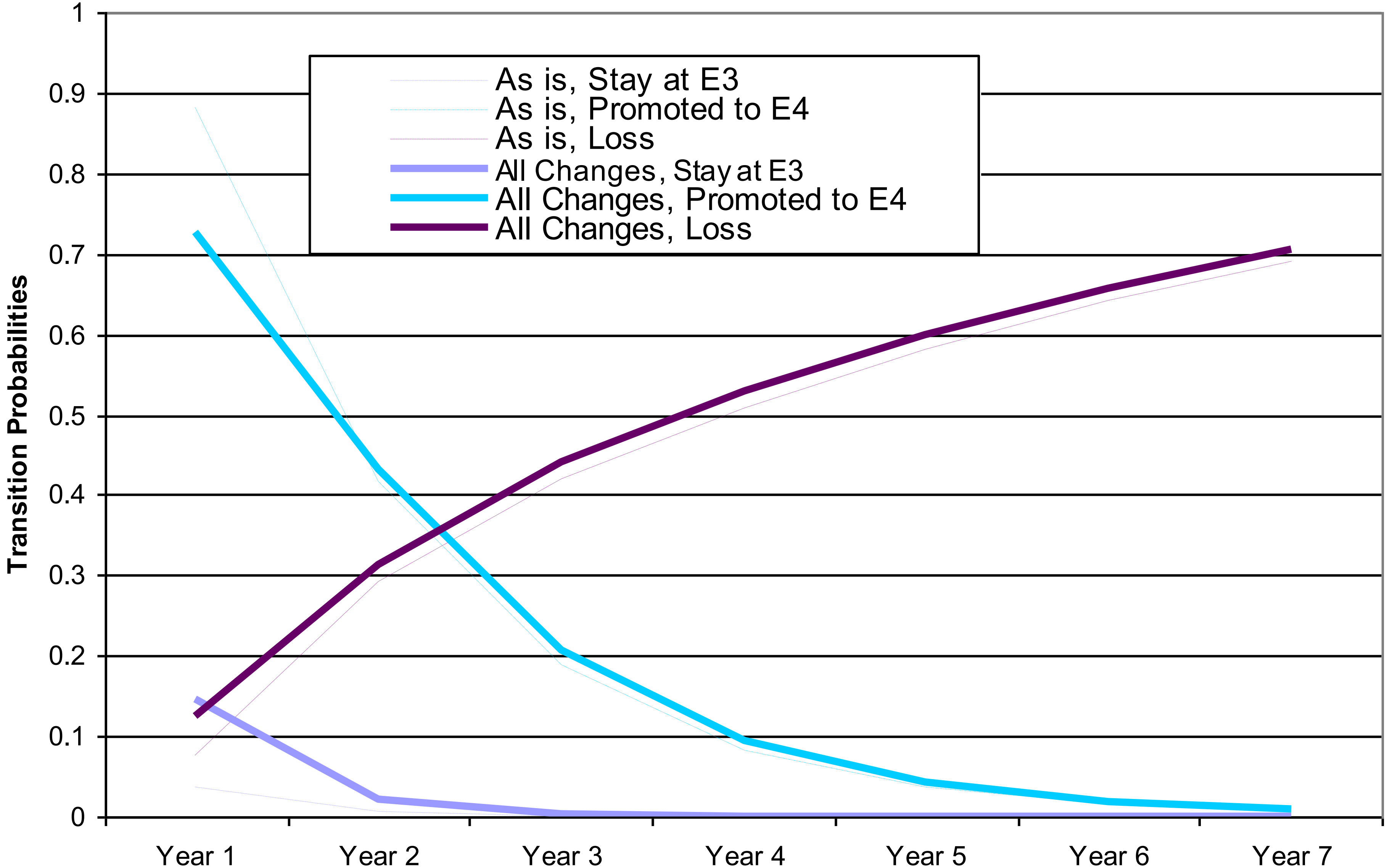


Figure 9C. Average Transition Probability Trajectories (7 years out), Skill Group = Nuclear, Starting State = E4



A particular individual - original trajectories vs. simulated effects of decrease in unemployment and mortgage rates, and a change of specialization



Marginal Effects - Theory

- Technically, The Marginal Effects of individual *i*, at period *t*, are:

$$\frac{\partial p_{kj}}{\partial x_{its}} = p_{kj} y_{itk} \left(\lambda_{sj} - \sum_j p_{kj} \lambda_{sj} \right)$$

Evaluated at sample means, sub-groups of interest or at individual levels

Marginal Effects - Empirics

Marginal effects:

- By Pay-Grade/Skill results show significant effects for:
  - Education
  - Marital status
  - Sea duty
  - Sea Months
  - Fleet Concentration Area (FCA)
  - FCA changes prior to and after advancement
  - Number of times a (pre-promotion) exam was taken
  - Performance evaluation
  - Stability

Marginal Effects - ADMIN

Variable	E-3 to E-4	E-4 to E-5	E-5 to E-6	E-6 to E-7	Loss
Sea Duty				12.77%	
FC Norfolk		-5.17%		12.65%	
H.S +		9.9%	7.7%		
AFQT	-	-	-	-	
PMA		10.7%			
Sea Months					< 1%

Surface Combat Weapon – An (Artificial) Individual Sailor

The sailor's characteristics:

- High School +
- AFQT = 87
- MOS = 25
- TIR = 8 months
- PMA = 3.6
- Sea Months =1
- FCA= Corpus Christi
- Cycle Cut = 179.75
- Final Multiple = 124.89
- DNEC 0981 (Vertical Launching Systems Maintenance)

	E-4	E-5	E-6	E-7	Loss
Year 1	1				
Year 2	.3635	.2578			.3786
Year 3	.1321	.2801	.0487		.539
Year 4	.048	.2365	.0946	.0005	.6203
Year 5	.0175	.1833	.1257	.0014	.6721

Currently in E-4: 5 Year Forecast

Some Notes:

- Endogeneity of the Model
- Model Refinements

Compensation (MODCOMP) Empirical Example and Model Comparisons

(Results from an IT-GME discrete choice model – a "Generalized Logit")

A note on Endogeneity

The Multinomial GME model used here was modified in such a way that it takes care of the endogenous relationship between retention and bonuses. To do so, we use a basic two-stage GME model. In the first stage we estimate the predicted value of the bonuses (or ACOL for the ACOL model) and then we use the predicted values as instrument in the GME second stage. We then follow Murphy and Topel (19--) to correct the covariance matrix.

- Examples: **Analyzing Special Reenlistment Bonus (SRB) and Other Pay Incentives**
- We demonstrate our approach with the main results obtained for the Weapon Control group (1995-2002) consisting of 66,509 individuals. The independent variables used (in addition to the intercept) are Gender, Race, Number of Children, AFQT Score, Base Pay, Total Allowance, Education Dummies (No High School and Above High School), See Duty, Dollar Amount SRB, Zone Dummies, Expected Civilian Wage (see MODCOMP Report for details), Lag Real Interest Rate, Lag Value of NASDAQ index, and the Unemployment Rate.

Table 3.1.Weapon Control – SRB Experiments for GME 3 Categories Model (Full Sample 66,509 observations)

	Implied Elasticities (0)	Prob Leave	Prob Extend	Prob Reenlist (infinity)	Prob Reenlist (0)	% Change Reenlist (infinity)	% Change Reenlist (0)	Mean $SRB	**Normalized % Change Reenlist (infinity)**	Additional Reenlisted Personnel	**Normalized % Change Reenlist (0)**
Base Case		*0.4615*	*0.2032*		*0.3353*			*1,244*			
SRB+0.5	0.45	0.4008	0.2118	0.3580	0.3875	6.8%	15.6%	1,677	**6.3%**	3,458	**12.3%**
SRB+1	0.40	0.3554	0.2135	0.3842	0.4311	14.6%	28.6%	2,129	**12.2%**	7,715	**23.0%**
SRB+1.5	0.36	0.3224	0.2115	0.4078	0.4661	21.6%	39.0%	2,582	**17.7%**	11,506	**31.7%**
SRB+2	0.33	0.2982	0.2075	0.4289	0.4943	27.9%	47.4%	3,034	**22.7%**	14,964	**38.8%**
SRB+2.5	0.30	0.2802	0.2025	0.4478	0.5173	33.6%	54.3%	3,486	**27.3%**	18,090	**44.7%**
SRB+3	0.28	0.2664	0.1971	0.4647	0.5364	38.6%	60.0%	3,938	**31.5%**	20,817	**49.5%**

Weapon Control – SRB Experiments for GME binary and 3 Categories and by zones

	Normalized % Change Reenlist (infinity) 3-Choice	% Change Reenlist (infinity) Binary	% Change Reenlist (0) 3-Choice	% Change Reenlist (0) Binary	**Normalized % Change Reenlist (0) 3-Choice**	**Normalized % Change Reenlist (0) Binary**	**Normalized % Change (infinity) Zone A 3-Choice**	**Normalized % Change (infinity) Zone B 3-Choice**	**Normalized % Change (infinity) Zone C 3-Choice**
SRB+0.5	**6.3%**	1.8%	15.6%	8.0%	**12.3%**	**6.7%**	**4.7%**	**9.8%**	**24.6%**
SRB+1	**12.2%**	6.4%	28.6%	15.9%	**23.0%**	**13.4%**	**9.3%**	**18.7%**	**44.8%**
SRB+1.5	**17.7%**	10.8%	39.0%	23.3%	**31.7%**	**19.8%**	**13.8%**	**27.5%**	**60.6%**
SRB+2	**22.7%**	15.0%	47.4%	30.2%	**38.8%**	**25.6%**	**18.2%**	**35.8%**	**72.8%**
SRB+2.5	**27.3%**	19.0%	54.3%	36.5%	**44.7%**	**NA**	**22.3%**	**43.5%**	**82.3%**
SRB+3	**31.5%**	22.8%	60.0%	42.2%	**49.5%**	**NA**	**26.3%**	**50.7%**	**89.7%**

Model Comparison

Out of Sample Prediction/Forecast

- Based on real data, in this example we study the validity of the SRB experiments for different models. Investigating the data we learn that between 1998 and 1999 there is a significant increase in SRB for both the Sensor Operations and Weapon Control groups. Specifically, from an SRB range of 0-4.5 and mean of 0.35 in 1998, the SRB range increased to 6.5 with a mean of 0.67 in 1999 for the Sensor Operations group. Similarly, for the Weapon Control, SRB increased from a range of 0-5.5 in 1998 to 0-7.0 in 1999. In both cases there is a significant increase in the number of reenlisted personnel in 1999. We used this observation and data to perform the following experiment.
- Estimate the pre 1999 data (two cases: just the 1998 – Case A; 1995-1998 – Case B).
- For each estimation model, perform SRB experiments with the pre-1999 data.
- Compare the impact of the SRB increases with the observed values of 1999.

Range of Estimated Percentage Point Change as a Result of Change in SRB – Pre 1999 Data.

	Sensor Operations		Weapon Control	
	Case A	Case B	Case A	Case B
Full Sample	**16.9**	**5.2**	**23.0**	**18.1**
Entitled Only	**19.3**	**10.0**	**27.4**	**16.5**
ACOL10-Full	0.4-1.8	0.3-0.6	0.3-0.5	0.3-0.5
ACOL10-Ent	4.7-24.6	8.9-17.8	6.8-10.5	5.8-8.9
ACOL20-Full	0.8-5.4	1.0-2.0	1.1-1.65	1.1-1.7
ACOL20-Ent	10.8-51.8	18.5-37.0	7.5-18.5	9.6-14.7
ACOL40-Full	NA	NA	3.6-5.4	3.6-5.4
ACOL40-Ent	NA	MA	15.0-26.7	11.6-20.4
GME-Full	8.6-15	12.6-22.5	8.4-12.6	12.9-18.5
GME-Ent	13.7-24.0	18.7-33.6	13.7-20.5	20.5-30.1

Notes:

1. Bold numbers (rows 3-4) reflect the correct observed changes
2. Case A. Using only the 1998 sample.
3. Case B. Using all data prior to 1999 (1995-1998).

All GME results are based on the higher discounted factor case

The most important results here are:

- The GME non-ACOL yielded better estimates based on the experiment (estimates that are closer to the observed data in 1999 – rows 3-5 of Table 4.1). For example, consider the Sensor Operation group. Based on the full data (Case A) the increase in reenlistment in 1999 was 16.9% and 19.3% for the entitled subgroup. With the 2 points increase in SRB, the GME predicted an increase of 15% for the full model and 24% for the entitled subgroup.
- The ACOL model for the entitled personnel provides much higher values than the traditional ACOL experiment done on the full model.
- The ACOL model is very sensitive to the discount factor and whether one analyzes the full data or just the entitled personnel.

Some thoughts and points for discussion

Model selection

Choice of estimator

Data

IT vs. ML

Unified data Sets and Models

Thank You

Data - Basics

- First observation in data is at E-3
- Analyzed Events (States of Nature)
  - Promotions from E-4 to E-7
  - Loss, Demotions
  - Eligible for Advancement
    - Either captured by minimum Time-in-Rate or sat for advancement exam
  - Non-advancement events
    - Change in duty station (geographic location, platform, UIC)
  - Community transfers or merges
    - Reenlistments
- Sea shore code = 1-4, 6
- Sea months – aggregate number of months at sea
- Individual characteristics – gender, marital status, children, age, education, AFQT, variable to capture relative placement within cohort
- Time periods – date of first observation and event date
- Event specific data – EMC, rating, pay grade, EAOS, UIC, Platform, CA (ATC), merge, transfer, LOS in months, total number of NECs held by individual at time of observation, etc.

Data – Detailed Description

VARIABLE	DESCRIPTION / SOURCE / INSTRUCTIONS / NOTES
Dummy Variables	
CS (3-8)	The sailor's current state (pay grade E3-E7, or Loss with Loss=8)
LS (3-6)	The sailor's lagged state (pay grade at the end of the previous year)
HS	*Label*: Whether or not the sailor has completed high school. *Source*: Raw
MTHS	*Label*: Whether or not the sailor has education higher than high school. *Source*: Raw
MARRIED	*Label*: Whether or not the sailor is currently married. *Source*: Raw
FLAG_TR	*Label*: Whether or not the sailor has ever been transferred. *Source*: Created *Instructions*: This is a flag that is set to 1 if the sailor has ever been transferred and 0 if never been transferred.
FCC_PROM	*Label*: Whether or not the sailor has changed Fleet Concentration since last promotion (or since start of current pay grade). *Source*: Created *Instructions*: Scan the Sailor's records since the last promotion (or since the beginning of the current pay grade) and set this flag to 1 if there is any change in Fleet Concentration found. If not, set this flag to 0.
SS_SEA	*Label*: Is the Sailor currently at sea. *Source*: Raw
SS_OTH	*Label*: Is the Sailor's current Seashore code something other than SEA or LAND. *Source*: Raw

Data (Cont.)

SSC_LO	*Label*: Whether the sailor never changed from Sea to Shore duty or if (s)he changed once, since the sailor was first observed in the data. *Source*: Created *Instructions*: Scan the sailor's complete history for changes in Sea Shore duty codes and set this flag to 1 if the sum of these changes is 0 or 1. Else, set the flag to 0.
SATCC_LO	*Label*: Whether the sailor's ATC code did not change since the last promotion or start of current pay grade. *Source*: Created *Instructions*: Scan the Sailor's record for change in ATC (since the last promotion or start f this pay grade) and set this flag to 1 if no ATC changes are found.
SATCC_ME	*Label*: Whether the sailor's ATC code has changed just once since the last promotion or start of current pay grade. *Source*: Created *Instructions*: Scan the Sailor's record for change in ATC (since the last promotion or start f this pay grade) and set this flag to 1 if only 1 ATC change is found.
EXAM_M	*Label:* If the sailor has been eligible for this promotion 2 or 3 times (the sailor took the exams more than once in the past). *Source:* Created *Instructions:* Scan the Sailor's record for the relevant event_id to see how many times has (s)he been eligible for promotion. The scans are pay grade specific. For example, if the sailor is currently at E3, and is eligible for E4, then scan the sailor's record to see how many times event_id 9 occurred in his/her record (since the last promotion). Set this flag = 1 if the current count is equal to 2 or 3 else set this flag to 0. The scanning is done through the last month of the current period.

Data (Cont.)

EXAM_S	*Label:* If the sailor has been eligible for this promotion 4 or more times. *Source:* Created *Instructions:* Scan the Sailor's record for the relevant event_id to see how many times has (s)he been eligible for promotion. The scans are pay grade specific. For example, if the sailor is currently at E3, and is eligible for E4, then scan the sailor's record to see how many times event_id 9 occurred in his/her record (since the last promotion). Set this flag = 1 if the current count is equal to 4 or more else set it to zero. The scanning is done through the last month of the current period.
NAR	*Label*: Whether or not the sailor is eligible for promotion (i.e., passed her/his minimum time in pay-grade). *Source*: Created *Instructions*: If a non-missing VACANTS variable is recorded to the current observation, then set this flag to 1, else set it to 0.
FLCONC00 FLCONC01 … FLCONC11	*Label*: Sailor's current Fleet Concentration. *Source*: Raw *Note*: In the actual models, when a particular FLCONC category is very rare for a given skill group (i.e., less than 2%) it is not included in the models.
DNECXXXX	*Label*: Sailor's DNEC code. *Source*: Raw *Note*: In the actual models, when a particular DNEC category is very rare for a given skill group (i.e., less than 2%) it is not included in the models. The DNEC categories are skill group specific.

Data (Cont.)

Continuous Sailor and Promotion Cycle Specific Variables	
AFQT_N	*Label*: Sailor's AFQT Score. *Source*: Raw *Note*: All sailors with AFQT Score less than 30 or higher than 99 are removed from the data.
SEAMONTH	*Label*: Number of months, to date, that the Sailor has spent continuously at Sea. *Source*: Raw
SEAMONTH2	*Label*: SEAMONTH Squared *Source*: Created *Instruction*: This is just the square of the raw variable SEAMONTH
LOS	*Label*: Sailor's Length of Service, to date. *Source*: Raw
LOS2	*Label*: Sailor's Length of Service Squared *Source*: Created *Instructions*: This is just the square of the raw variable (LOS)
TIR	*Label*: Sailor's Time in Current Rank, to date. *Source*: Raw
TIR2	*Label*: Sailor's time in current rank squared *Source*: Created *Instructions*: This is just the square of the raw variable (TIR)
VACANTS	*Label*: Approximate number of vacancies for the rating-pay grade combination to which the sailor is attempting to be promoted (exam cycle specific). *Source*: Raw
TAKERS	*Label*: Approximate number of sailors who had the opportunity to advance (exam cycle specific). *Source*: Raw

Data (Cont.)

PMA	*Label*: Most recent Performance Mark Average value. *Source*: Raw
CYCLECUT	*Label*: Minimum score used by promotion period to determine advancement. *Source*: Raw
INDSCORE	*Label*: Sailor's individual final multiple score for the latest promotion cycle. *Source*: Raw
MULTIPLE	*Label*: Sailor's pay grade specific Multiple Score *Source*: Created *Instructions*: This variable is created using the following pay grade specific calculations. **For E3 and E4**: MULTIPLE = 0.34*INDSCORE + 0.36*((PMA*60) - 156) + 0.13*((TIR*2) + 15) + 0.13*(2*PNA) **For E5**: MULTIPLE = 0.30*INDSCORE + 0.415*((PMA*60) - 130) + 0.13*((TIR*2) + 19) + 0.11*(2*PNA) **For E6**: MULTIPLE = 0.60*INDSCORE + 0.40*(PMA*13)
REL_MULT	*Label*: Sailors pay grade specific Multiple Score relative to appropriate cohort. *Source*: Created *Instructions*: This variable is created by subtracting, from the sailor's individual Multiple score (MULTIPLE), the average Multiple scores of all sailors in the same cohort (GROUP MEAN). A cohort is defined as the groups of sailors who, in any given year, have the same EMC code and are at the same pay grade. i.e., REL_MULT = MULTIPLE – MULTIPLE_GM.

Data (Cont.)

Continuous Macroeconomic Variables	
MTGAGE_1	*Label*: Mortgage rate from one month prior to observation. *Source*: Raw
UNEMPL_1	*Label*: Unemployment Rate from one month prior to observation. *Source*: Raw
R_GDP_1	*Label*: Real GDP from one month prior to observation. *Source*: Raw
NASDAQ_0	*Label*: NASDAQ closing index for the month of this observation. *Source*: Raw

Special Variables

DNECs

- To capture effect of specific assignment/job on advancement probability used DNEC
  - Looked at job characteristics
    - Location – Regional ATCs and Concentration areas
    - DNEC
    - Duty Type
  - NEC problematic – no/little variation in data
- Incorporated specific DNECs
  - Input from ECMs by skill group

How Do Advancements Work?

- Final Multiple (FM) Components
  - Exam Score
  - PMA – based on 1-4 scale
  - PNA – points awarded if individual passed exam but was not advanced
    - PNA points vary by Pay Grade (PG)
    - Higher exam score greater number of PNA pts awarded
    - PNA points can be lost if individual fails next exam
  - Minimum TIR (Time in Rank)
  - Awards – not included in (makes up < 4% of FM)
  - E-7 advancements only consider PMA, Exam, and board recommendation
  - Components of FM weighted by PG and TIPG
    - Greater weight placed on exam score for lower PGs w/greater weight on performance for E-6 and above – See BUPERSINST 1430 and 1610

How Do Advancements Work? (con't)

- Eligible for advancement – must meet Min TIPG
  - Varies by PG
- Exam score (standard score 20-80)
  - Performance Mark Average $\geq$ 3.6, in order to take the exam
  - E-4 to E-6 can take exam 2x/year
  - E-7 can take exam 1x/year
- Inverse relationship between vacancies and cut score
  - Cut Score – minimum exam score needed to be considered for advancement
    - Cut score is adjusted in every advancement cycle