INSTITUTE FOR DEFENSE ANALYSES

Forecasting Demand for Air National Guard Enlisted Training

WEAI Presentation

Julie Lockwood
Joe King
Jay Dennis
Akshay Jain
Minerva Song

June 2021
Approved for public release; distribution is unlimited.
IDA Paper NS P-22653
Log: H 21-000160

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

IDA

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication

This work was conducted by the Institute for Defense Analyses under contract HQ0034-14-D-0001, project DZ-6-4720, "Forecasting Demand for Air National Guard (ANG) BMT and Tech School Slots to Improve Readiness" for the Chief of National Guard Bureau (CNGB) and the Director of the Air National Guard (DANG). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

For More Information:

Dr. Julie A. Lockwood, Project Leader
jlockwood@ida.org, 703-578-2858

ADM John C. Harvey, Jr., USN (ret) Director, SFRD
jharvey@ida.org, 703-575-4530

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper NS P-22653

Forecasting Demand for Air National Guard Enlisted Training

WEAI Presentation

Julie Lockwood
Joe King
Jay Dennis
Akshay Jain
Minerva Song

This page is intentionally blank.

Executive Summary

The Air National Guard (ANG) relies on the U.S. Air Force to provide enlisted entrants with basic and technical skills training. This enlisted training pipeline begins with Basic Military Training (BMT), which lasts approximately 8 weeks and provides individuals with basic skills to become airmen. Afterwards, individuals complete career-specific technical training at one of 210 different Tech Schools, which are tightly coupled with airmen's eventual occupation: their Air Force Specialty Code (AFSC).

Because BMT and Tech Schools are operated by the Air Force, the ANG must reserve training slots for these schools in advance, typically forecasting school slot demand up to 36 months in the future and budgeting for training costs 5 years in advance. Currently, 150 individual ANG Career Field Managers (CFMs) collect data through various methods to forecast school slot demand to replace airmen that exit the ANG, to fill vacancies created by airmen switching to another AFSC, and to upside/downsize particular AFSCs. While reserving BMT slots is typically unproblematic, ANG experiences long delays (up to a year) for critical technical schools due to inaccurate or incomplete school slot forecasts. These training delays harm unit readiness and contribute to inefficient resource expenditure while airmen await training, to the overall detriment of the ANG.

Using the Retention Prediction Model (RPM), a Machine Learning (ML) package developed by the Institute for Defense Analyses (IDA), this analysis provides the ANG with enhanced training demand forecasts. Given that training demand is filtered through AFSCs, we provide the ANG with the expected number of exits at the detailed (3-digit) AFSC-level, 5 years into the future. Our forecast model uses Defense Manpower Data Center (DMDC) personnel data, combined with ANG training data, spanning 2005–2020. These data comprise about 1.4 million person-years, and about 215,000 unique enlisted airmen.

Using a 5-year test set, interim results suggest that the model correctly classifies about 71% of service exits and about 79% of service retentions across the period. On balance, our model achieves about 77% accuracy and performs about 19 percentage points better than a naïve model that uses historical average AFSC exit rates.

This page is intentionally blank.

Forecasting Demand for Air National Guard Enlisted Training

WEAI Presentation

Julie Lockwood
Joe King
Jay Dennis
Akshay Jain
Minerva Song

30 June 2021

The Air National Guard (ANG) faces inefficiencies in forecasting training school slot demand

- ANG asked IDA to develop and apply a person-level Retention Prediction Model (RPM)
  - Forecasting enlisted school demand is just one application
- **Approach:** We use a gradient-boosted tree model to forecast individual retention/exit probabilities
- **Performance:** Using a 2005–2015 training set to forecast exits/stays in 2016–2020, we
  - Correctly predict **76.7%** of all exit-survival observations,
  - Achieve **51.8%** precision and **70.6%** recall, and
  - Produce predictions with **~20%** less error than average historical rate

Environment

The Air National Guard (ANG) is one of seven reserve components with both federal and state missions

On average, ANG annually has

20,000 officers

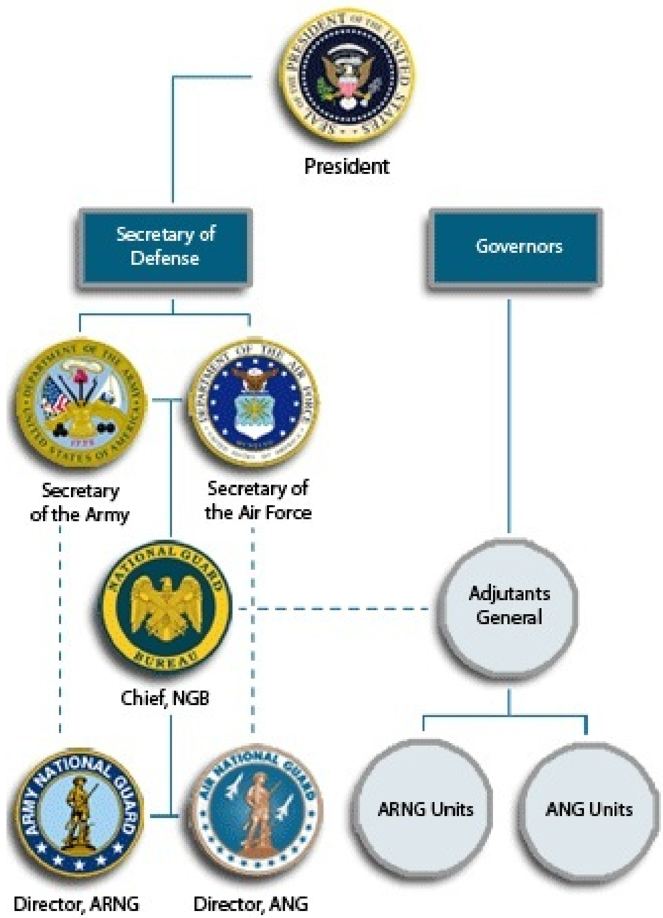
90,000 enlisted airmen

ANG offers careers to

Non-prior service and prior service

Full-time and part-time personnel

Most airmen are part time and maintain a regular civilian job



National Guard Bureau - New Jersey Army National Guard, http://www.njarmyguard.com/organization/

Technical school training programs for ANG airmen are run by the active Air Force

Technical school training corresponds to an Air Force Specialty Code (AFSC), a person's occupation

The most common enlisted AFSCs include aircraft maintenance, security, and cyber

ANG requests its expected number of school slots from the Air Force up to **36 months** in advance

ANG budgets for training costs over a **5-year** period

Existing ANG forecast model is inaccurate, leading to significant delays for technical schools

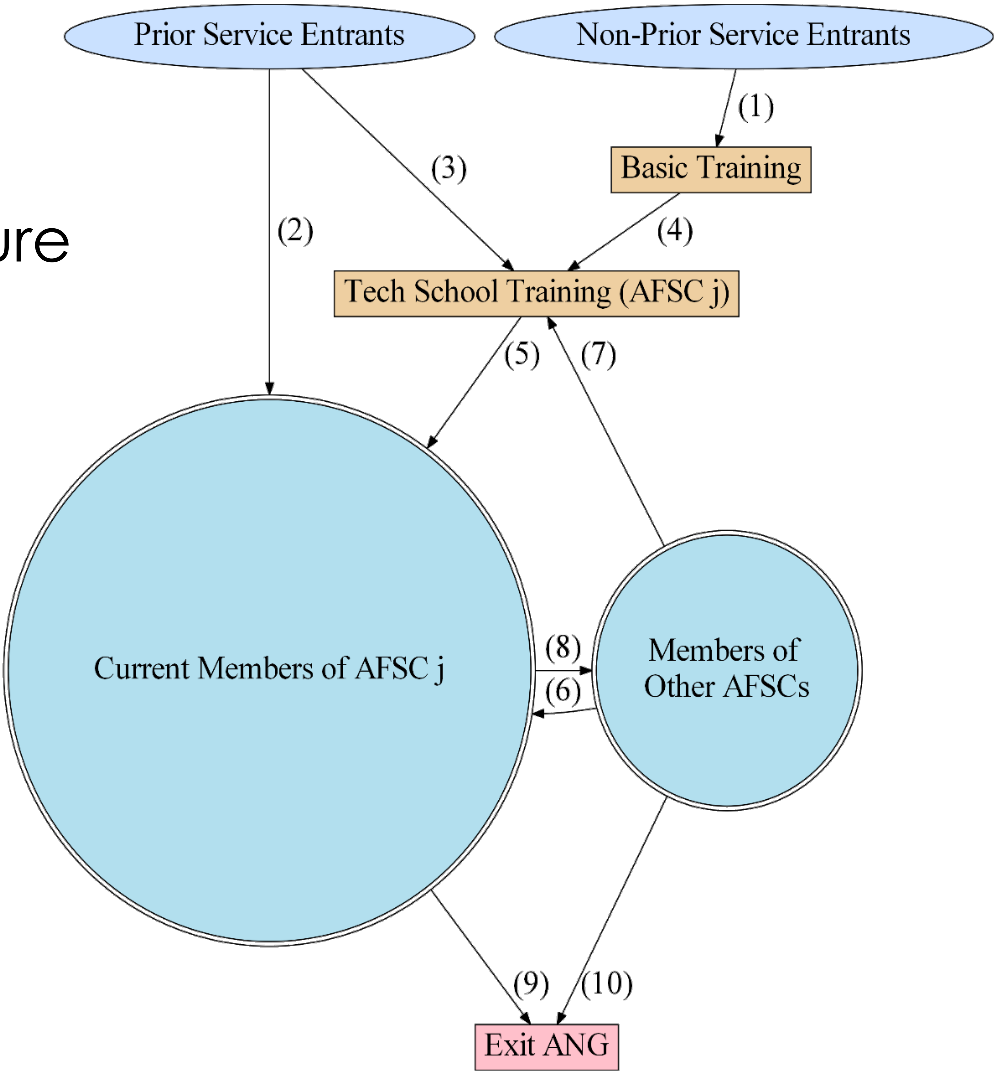
Research Question:

How can we better forecast demand for
technical school slots
using individual ANG retention prediction?

What drives AFSC demand?

1. Backfilling those exiting
2. Changes in force structure

This chart depicts the stock and flow of airmen within AFSC *j*

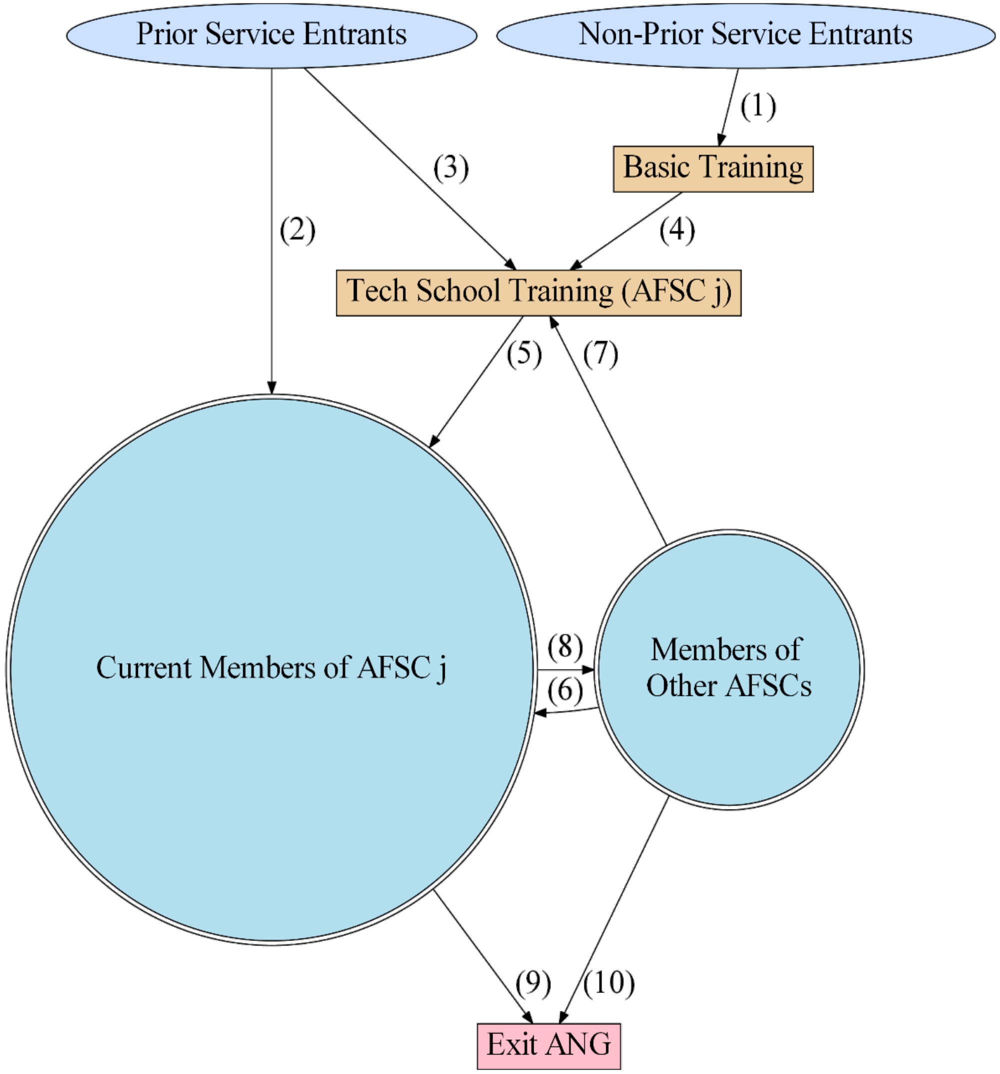


AFSC tech school demand equals the number of billets to be filled by untrained individuals

Training demand depends on how the AFSC demand is filled

Source	Requires Basic Training?	Requires Tech School?
Non-prior military service entrants	✓	✓
ANG AFSC transfers without required occupation training	✗	✓
ANG AFSC transfers with required occupation training	✗	✗
Prior military service entrants without required occupation training	✗	✓
Prior military service entrants with required occupation training	✗	✗

AFSC tech school demand equals the number of billets to be filled by untrained individuals



ANG seeks to forecast net demand for school slots ($D_{j,t}$) for AFSC j in year t

$$\Delta_{j,t} + \varphi_{j,t} + \lambda_{j,t} = \gamma_{j,t} + \theta_{j,t} \quad (1)$$

$$\gamma_{j,t} = \Delta_{j,t} + \varphi_{j,t} + \lambda_{j,t} - \theta_{j,t} \quad (2)$$

$$D_{j,t} = \max(0, \Delta_{j,t} + \varphi_{j,t} + \lambda_{j,t} - \theta_{j,t}) \quad (3)$$

Where

- $\Delta_{j,t}$ = **ANG structural adjustment** in AFSC *j* desired headcount
- $\varphi_{j,t}$ = **Exits from service** from AFSC *j* between *t-1* and *t*
- $\lambda_{j,t}$ = **Switch-outs** to other AFSCs between *t-1* and *t*
- $\gamma_{j,t}$ = **Non-trained new entrants** to AFSC *j* between *t-1* and *t*
- $\theta_{j,t}$ = **Pre-trained new entrants** to AFSC *j* between *t-1* and *t*

"New entrants" are non-prior service entrants, prior service entrants, and transfers

This research phase focuses on exits from service

$$\Delta_{j,t} + \boldsymbol{\varphi_{j,t}} + \lambda_{j,t} = \gamma_{j,t} + \theta_{j,t} \qquad (1)$$

$$\gamma_{j,t} = \Delta_{j,t} + \boldsymbol{\varphi_{j,t}} + \lambda_{j,t} - \theta_{j,t} \qquad (2)$$

$$D_{j,t} = \max(0, \Delta_{j,t} + \boldsymbol{\varphi_{j,t}} + \lambda_{j,t} - \theta_{j,t}) \qquad (3)$$

Where

$\Delta_{j,t}$ = **ANG structural adjustment** in AFSC *j* desired headcount

$\boldsymbol{\varphi_{j,t}}$ **= Exits from service** from AFSC *j* between *t-1* and *t*

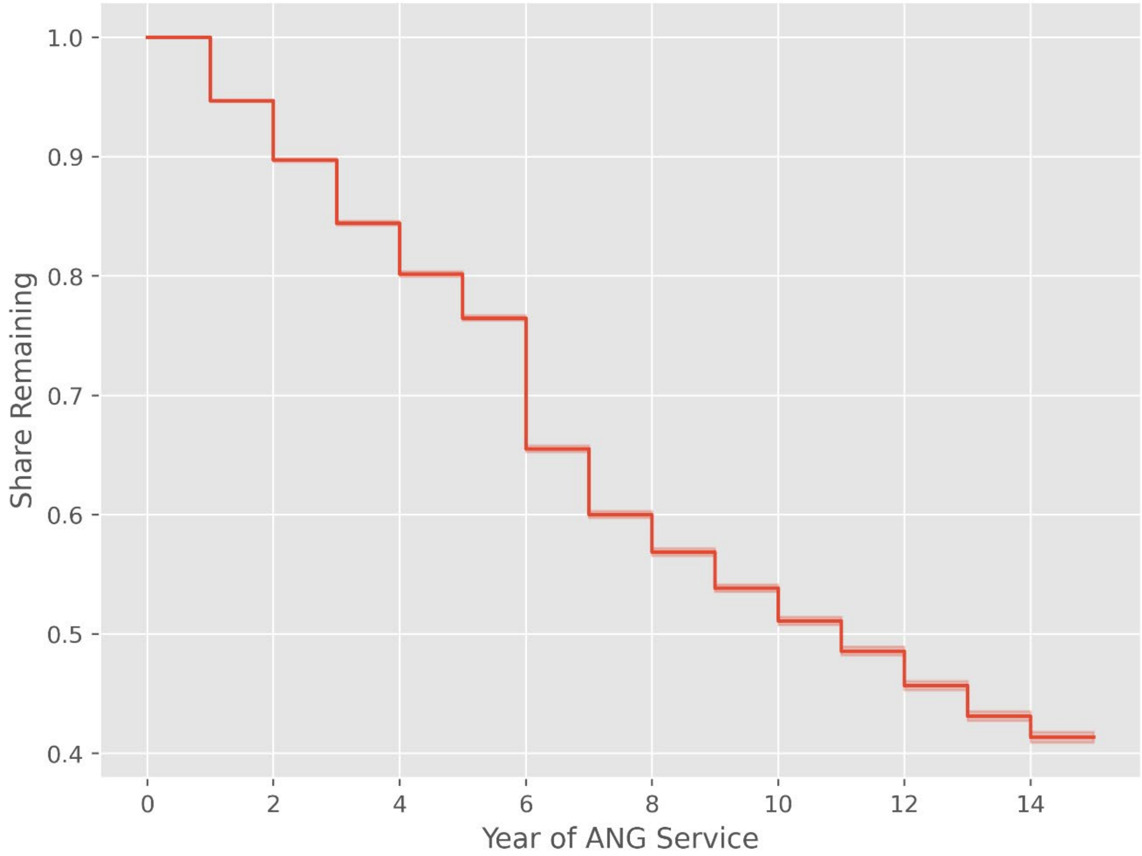
$\lambda_{j,t}$ = **Switch-outs** to other AFSCs between *t-1* and *t*

$\gamma_{j,t}$ = **Non-trained new entrants** to AFSC *j* between *t-1* and *t*

$\theta_{j,t}$ = **Pre-trained new entrants** to AFSC *j* between *t-1* and *t*

"New entrants" are non-prior service entrants, prior service entrants, and transfers

Share of enlisted ANG personnel remaining by year of service



Note: 2005–2020 data. Kaplan-Meier estimates with 95% confidence intervals shown

Methodology

The Finite Interval Forecasting Engine (FIFE) is a survival analysis tool

Free, open-source Python package offering an array of Machine Learning (ML) and other models for discrete-time survival analysis

Designed by the Institute for Defense Analyses (IDA) and sponsored by the Office of the Undersecretary of Defense for Personnel and Readiness (OUSD(P&R))

We use a Light Gradient Boosting Machine (LightGBM) survival model

FIFE estimates individual survival (or retention) probabilities for each person and period

Estimates probability of individuals in period *t* remaining in service τ periods into future, conditional on prior survival

Survival probabilities τ periods into future defined as

$$\hat{y}_{x,\tau} = Pr\left(T_{x,t} \geq \tau \middle| F_{x,t}\right) \qquad (4)$$

Where

x = Person x

$F_{x,t}$ = vector of feature values for individual x at time t

$T_{x,t}$ = number of consecutive future periods x remains after time t

We sum individual exit/retention probabilities by AFSC to obtain the expected number of exits per AFSC

Recall

$\varphi_{j,t+\tau}$ = Exits from service ***from AFSC j***,

between t + $\tau - 1$ and t + τ

We estimate

$\tilde{\varphi}_{j,t+\tau}$ = Exits from service ***of those in AFSC j in t***,

between t + $\tau - 1$ and t+τ

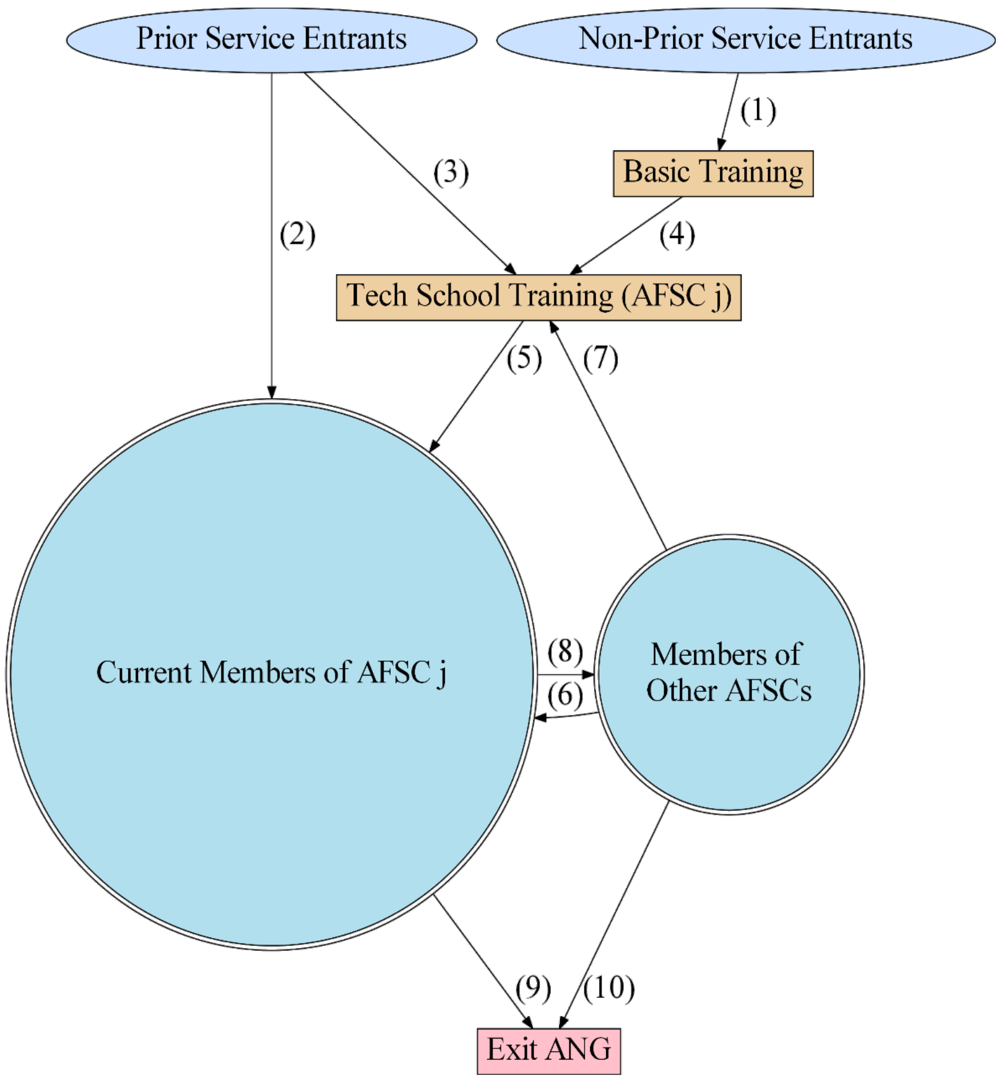
$$= \sum_{x \in X_{j,t}} \hat{y}_{x,t+\tau} \tag{5}$$

Where

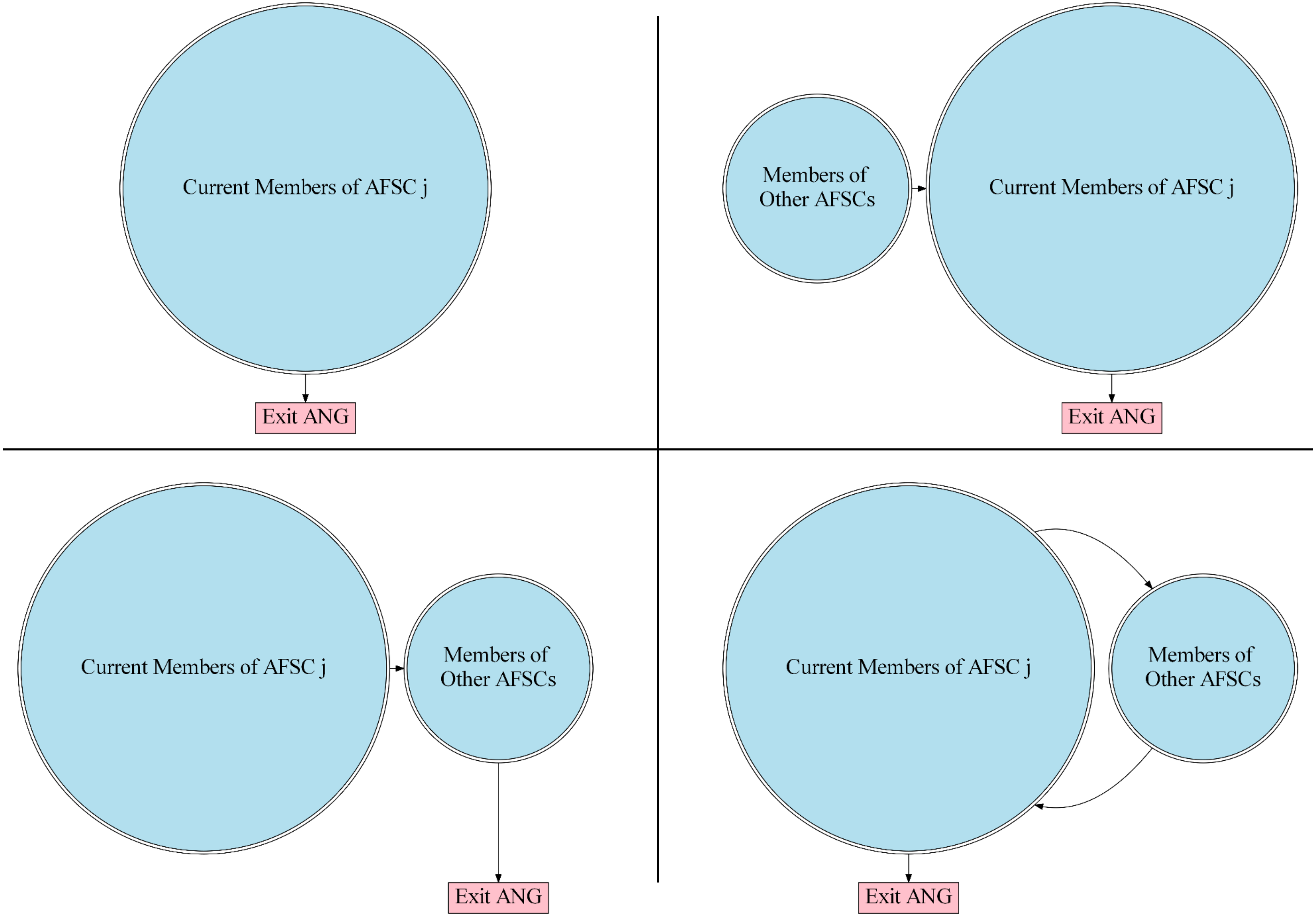
$X_{j,t}$ = Individuals in AFSC j in period t

$\hat{y}_{x,t+\tau}$ = Exit probability for individual x in period $t + \tau$

ANG exit forecasting for school slot forecasting requires adjusting for AFSC transfers and new entrants



There are several ways an airman can leave AFSC j



FIFE survival forecasts are one part of the exit component in the demand equation

$$\varphi_{j,t+\tau} = \sum_{\forall x \in X_{j,t}} \boldsymbol{E}_{\,t+\tau}(\boldsymbol{x}) - \sum_{\forall x \in X_{j,t}} E_{\,t+\tau}(x) * \left(1 - I_{j,t+\tau-1}(x)\right) + \sum_{\forall y \in Y_{j,t}} E_{\,t+\tau}(y) * I_{\,j,t+\tau-1}(y) + \sum_{\forall n \in N_{j,t+\tau}} E_{\,t+\tau}(n) * I_{\,j,t+\tau-1}(n) \quad (6)$$

Where

$X_{j,t+\tau}$ = individuals **in AFSC j** in t+τ

$Y_{j,t+\tau}$ = individuals in service but **not in AFSC j** in t+τ

$N_{j,t+\tau}$ = **New entrants** to AFSC j after t but before t+τ

$I_{j,t+\tau}$*(person p)* = 1 if p **is in AFSC j** in t+τ

$E_{t+\tau}$*(person p)* = 1 if p **exited** service between t+τ-1 and t+τ

Note: We estimate the part in **red**

We train using all years of data <u>except</u> the last few and then use these as a performance test

We construct a panel dataset (individual, year)

Using last few years as test set most closely emulates real-world performance in period of interest (2022–2026)

<u>Training set</u>: 2005–2015

<u>Test set</u>: 2016–2020

Hyper-optimized model (200 trials, 25% validation share)

Scope and data

Analytic set for modeling: all enlisted ANG personnel serving between June 2005 and June 2020

Before modeling, we exclude

- Direct commissioned officers and any post-commission observations
- Post-ANG observations (e.g., move other service/component)
- Observations before year 2005 due to poor data quality

Observations occur each June (annual frequency)

Results presented today use full population of ANG personnel who meet these criteria

Our dataset comprises DMDC, CPS, and ANG data

Dataset	Source	Description
Master	Defense Manpower Data Center (DMDC)	Demographics, career history
Pay	DMDC	Various pay fields
Deployments	DMDC	Deployment history
Family/Defense Enrollment Eligibility Reporting System (DEERS)	DMDC	Family and dependents
Activations file	DMDC	Title 10 federal activations
Casualties file	DMDC	Personal injuries, fatalities in unit
Civilian labor market data	Current Population Survey (June)	External labor market conditions
Supplementary ANG training data	ANG	ANG training

Total columns: 638

Results:

How would we have done if you asked us 5 years ago?

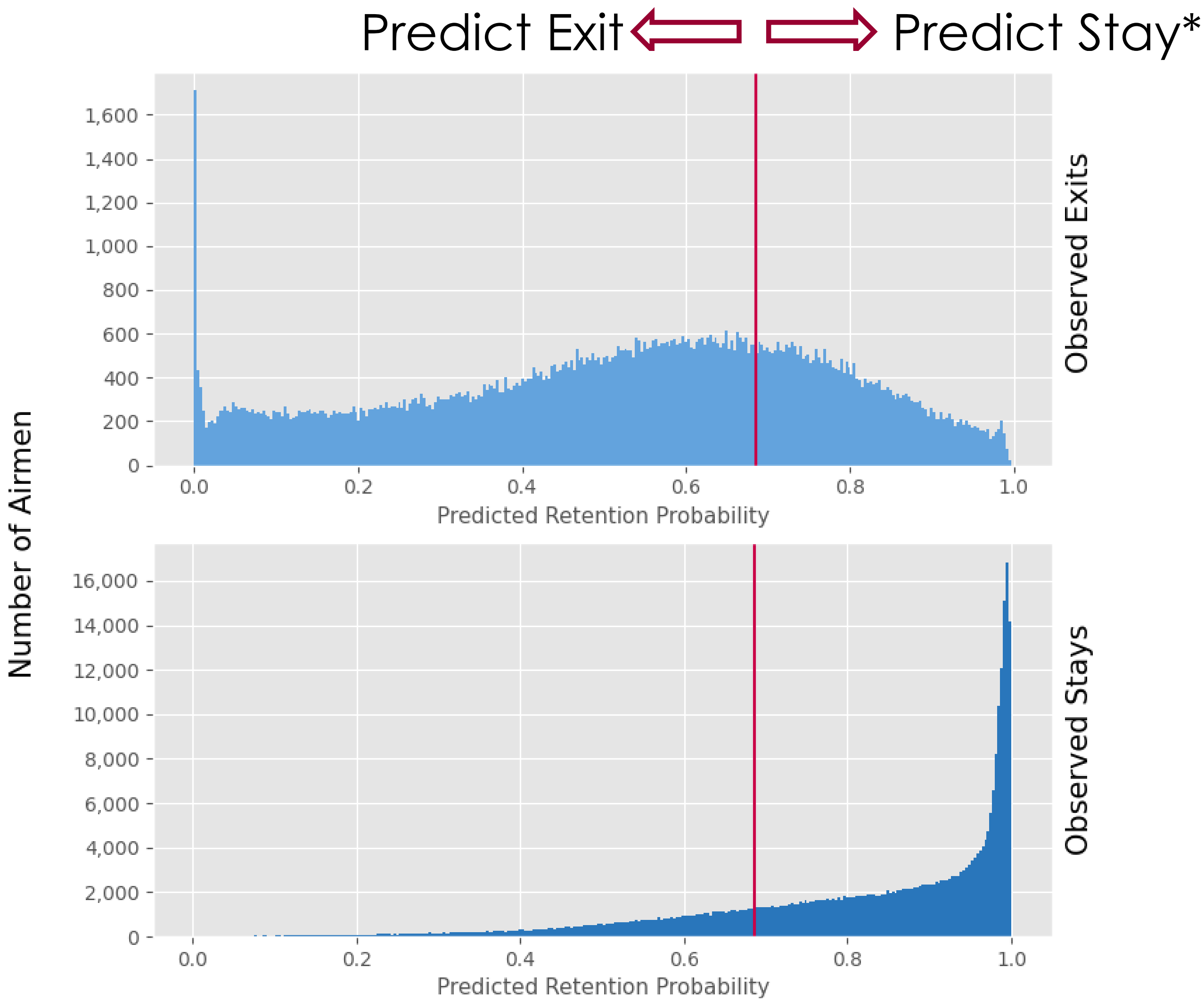
We accurately predict 70.6% of exits and 78.7% of retentions between 2016 and 2020

Forecasted Outcome	Actual Outcome: Exit	Actual Outcome: Remain
Exit	70.6%	21.3%
Remain	29.4%	78.7%

Overall accuracy: 76.7%

Note: Full ANG population meeting analytic restrictions (test: N=90,805). DMDC annual (June) data, 2005–2015. Estimates based on FIFE LightGBM model. The confusion matrix was produced with a threshold of **0.685** (i.e., PR(Y) > 0.685 classified as remain, else exit)

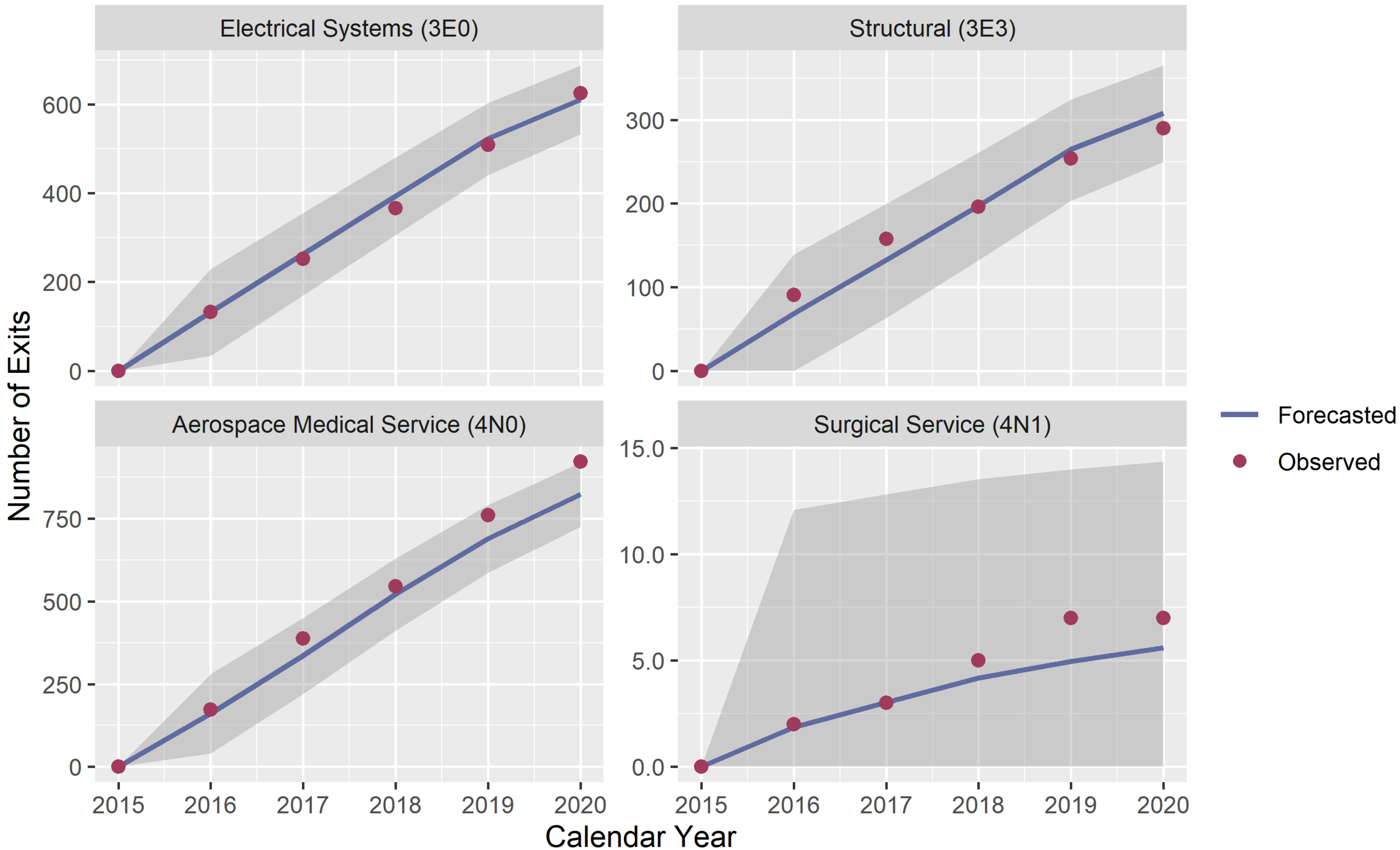
Predicting stays is easier than predicting exits



* Decision threshold: 0.685

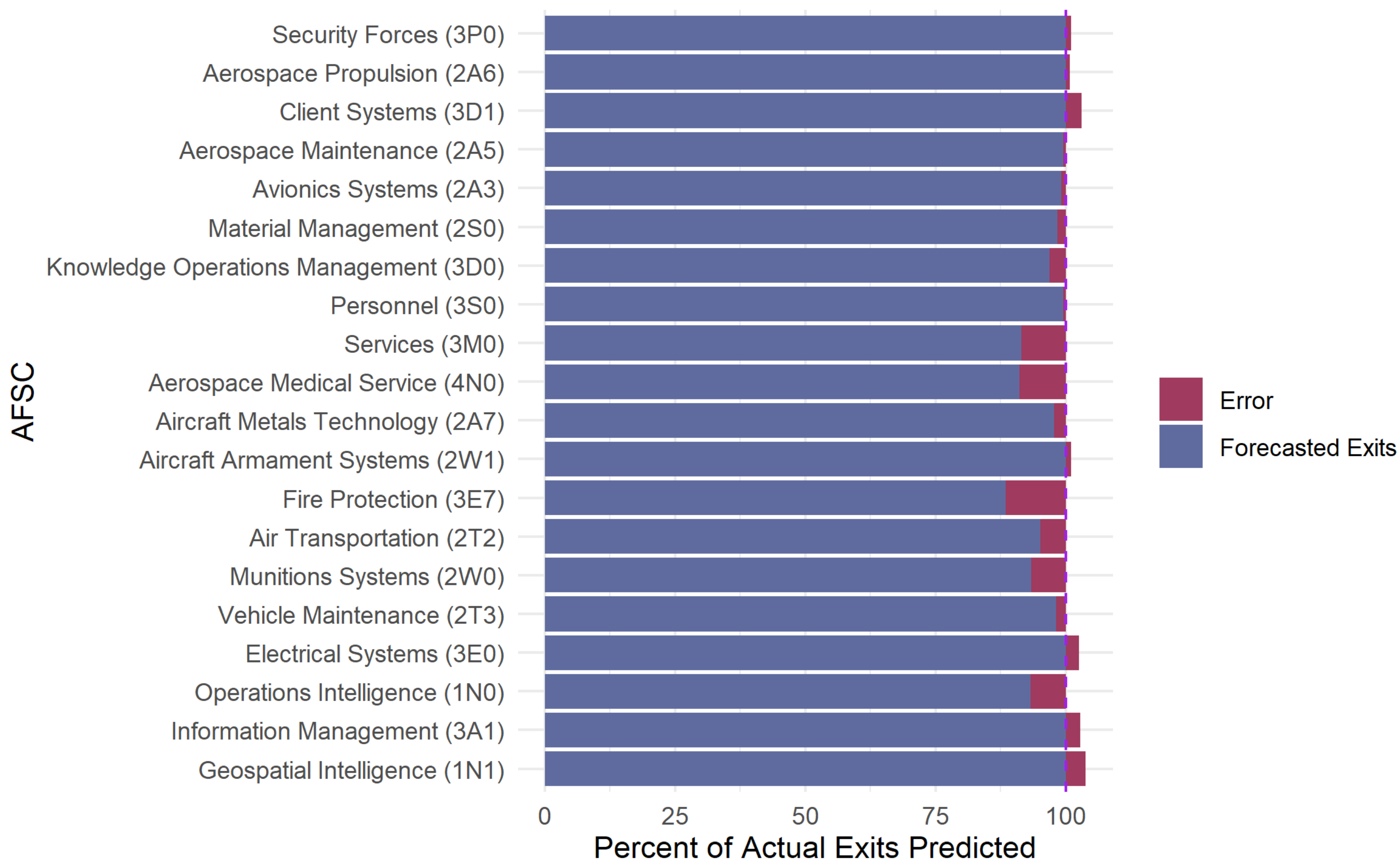
How well can our prototype forecast exits within AFSCs?

Forecasted vs. observed ANG exits for select AFSCs, 2016–2020



Forecast errors among 20 largest AFSCs

Covering 69% of enlisted personnel, 2016–2020



Note: Twenty largest AFSCs according to total personnel in (June) 2015

RPM-ANG performs 19% better than a model using average AFSC exit rate in 2005–2015

Mean Absolute Error				
Year	Forecast Horizon	FIFE	AFSC Average	% Improvement Using FIFE
2016	$t+1$	0.042	0.031	-35.5%
2017	$t+2$	0.027	0.031	12.9%
2018	$t+3$	0.051	0.033	-54.5%
2019	$t+4$	0.041	0.068	39.7%
2020	$t+5$	0.031	0.066	53.0%
Overall		**0.029**	**0.046**	**19.3%**

Note: Outcome is the share of exits per AFSC-year. Both models trained using 2005–2015, using 2016-2020 as test set. Baseline model uses average AFSC exit rate in 2005–2015. FIFE model sums individual exit forecasts by AFSC-year and error per AFSC-year

Discussion and conclusion

Accomplished

- Modeled one component of demand equation (exits from service)
- Successfully created model that outperforms existing methods used by the ANG
- Developed aggregate prediction method using expected values

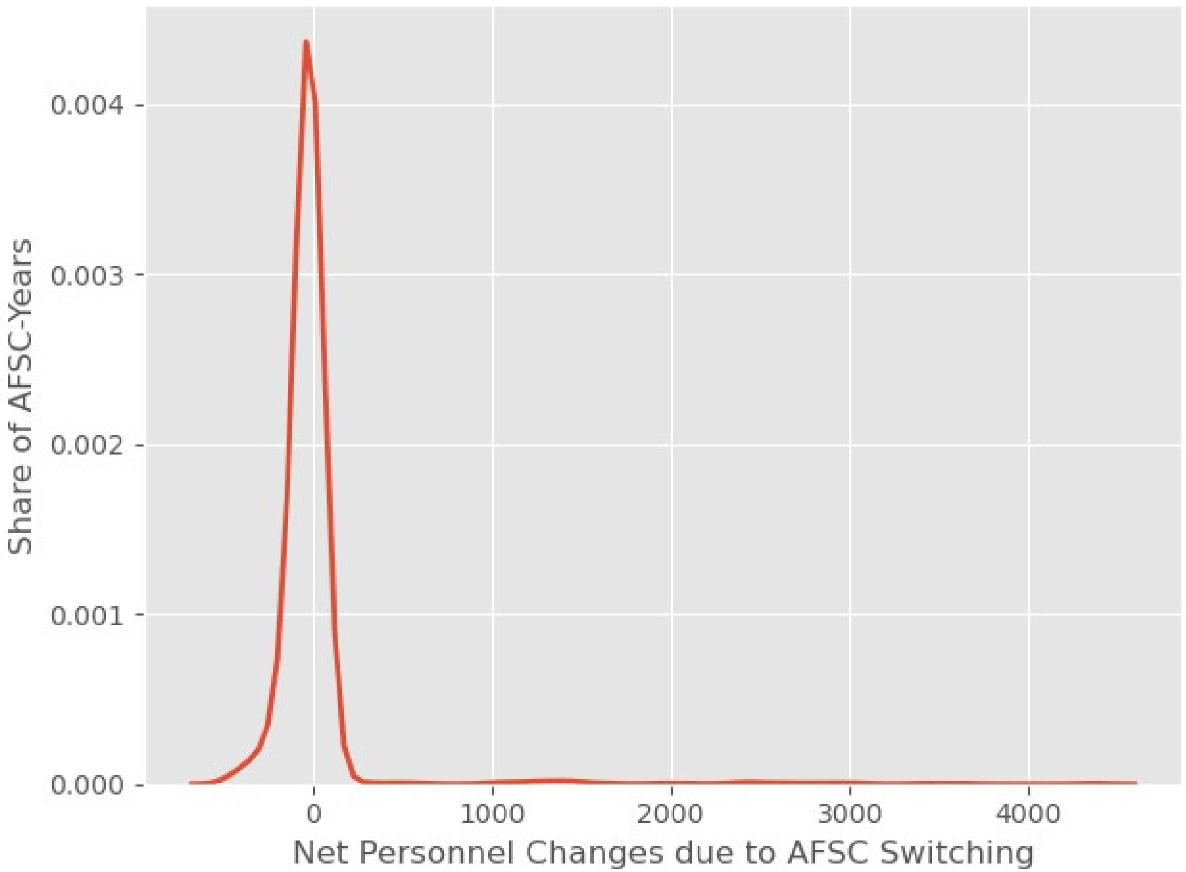
Remaining

- Model other components of demand equation (switching, new entrants)
- Make post-modeling forecast adjustments to further improve proposed method
- Compare other regression and time-series models on an AFSC level

IDA

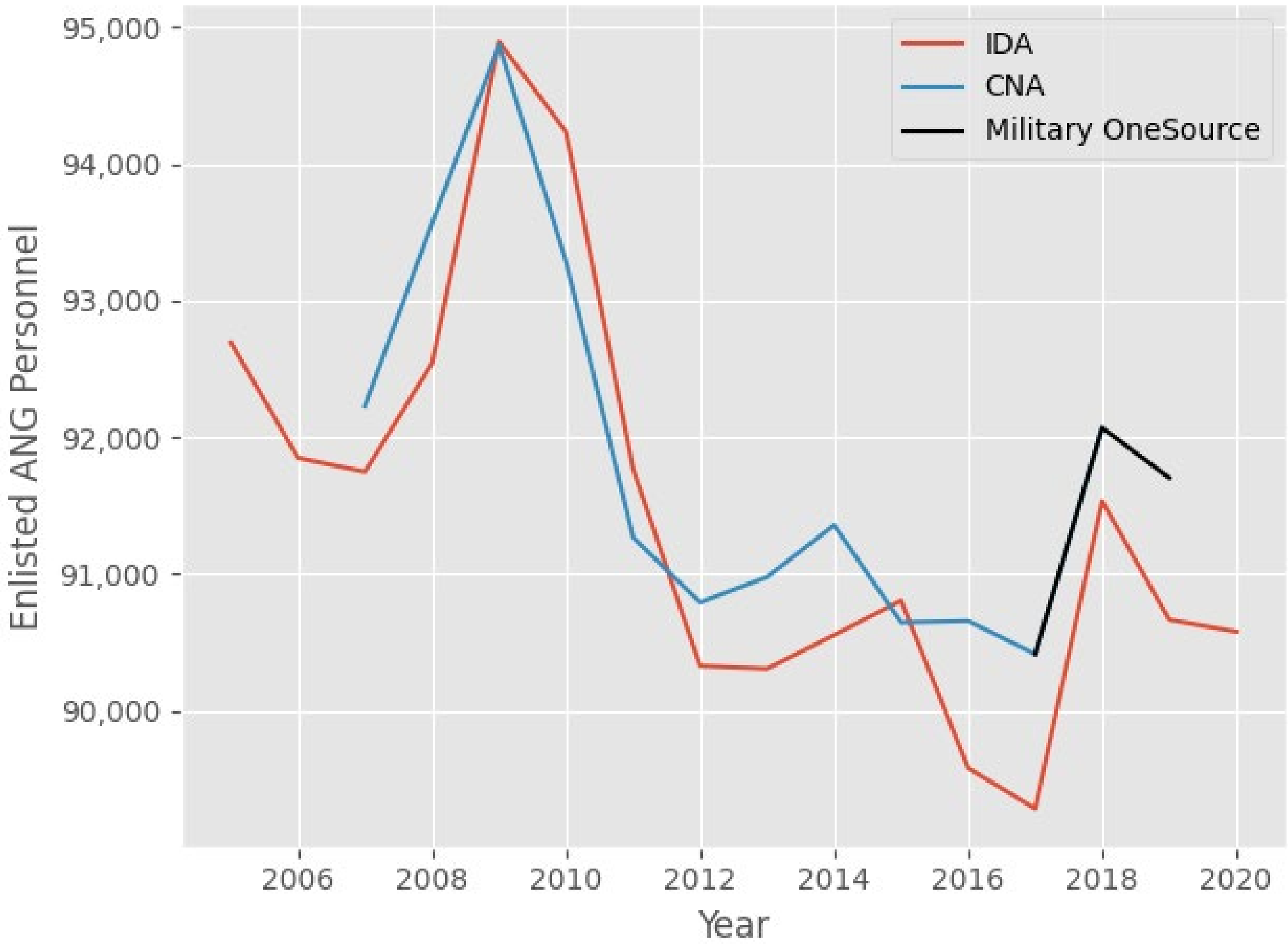
Backup

Net personnel changes across AFSC-years



Note: Changes calculated as AFSC switch-outs minus switch-ins. We exclude recoded or disbanded AFSCs by restricted to AFSC-years in which the number of remaining individuals in an AFSC >1 and the number of stayers is >1 (1535/1668). Mean=1.66, std=324

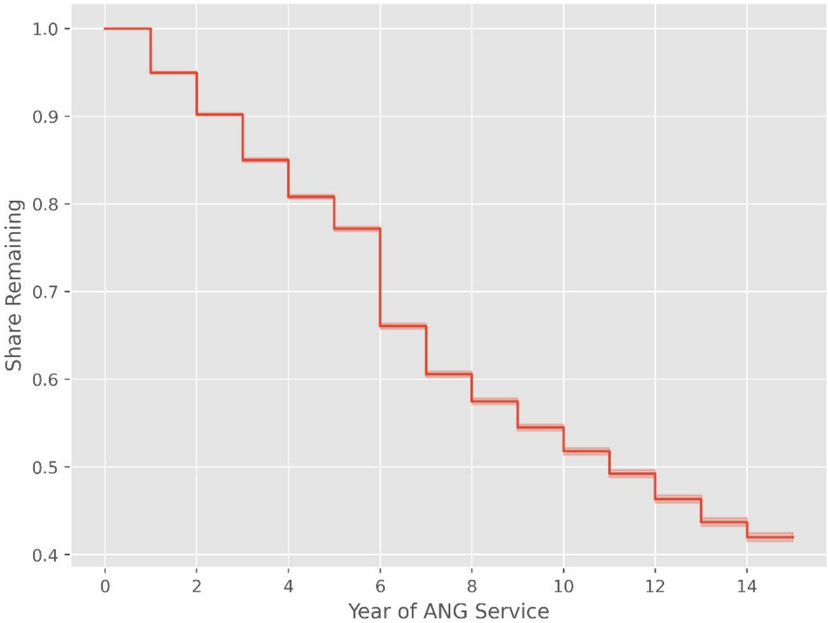
Enlisted ANG personnel in analytic set in 2005–2020



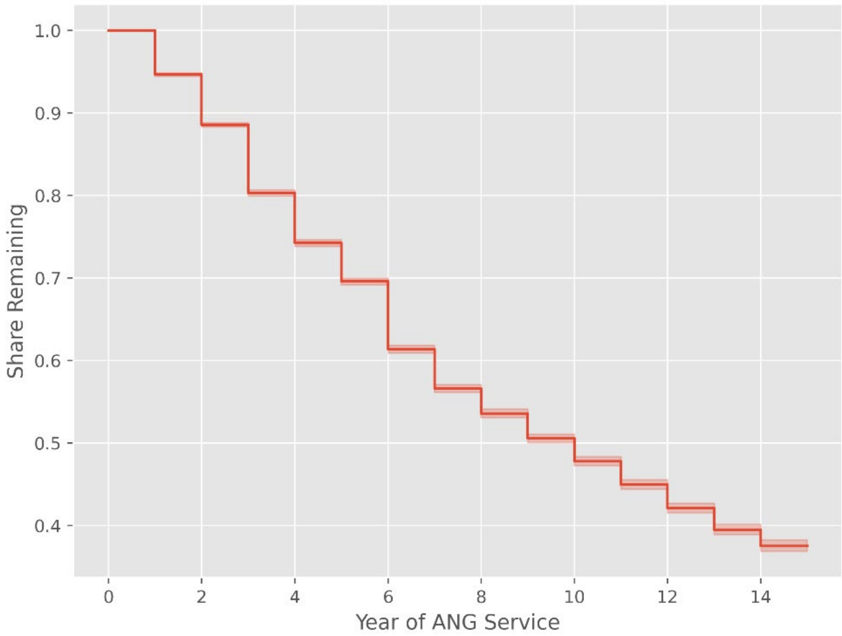
215,879 unique persons observed over 1,463,343 person-years

Retention among enlisted direct entrants vs. enlisted prior service individuals

Direct
entrants

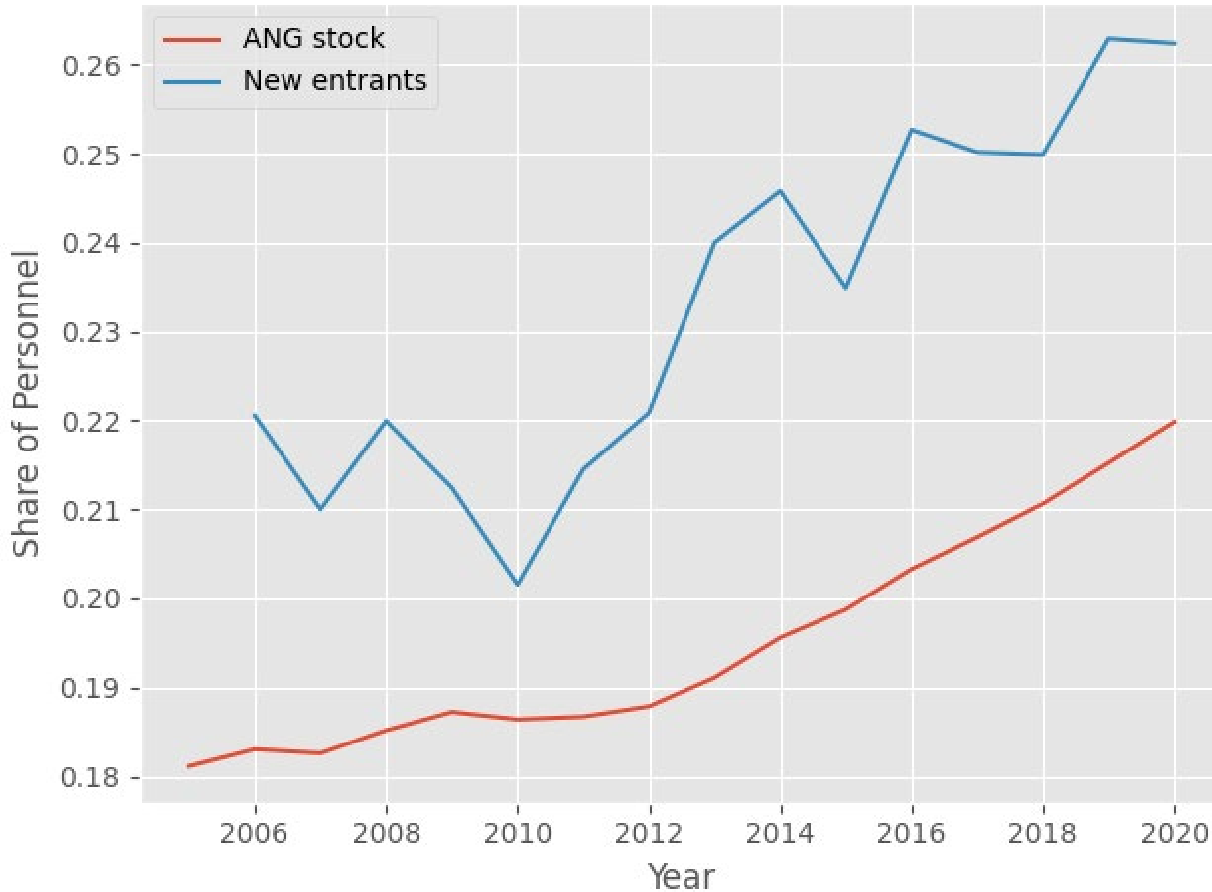


Prior
service



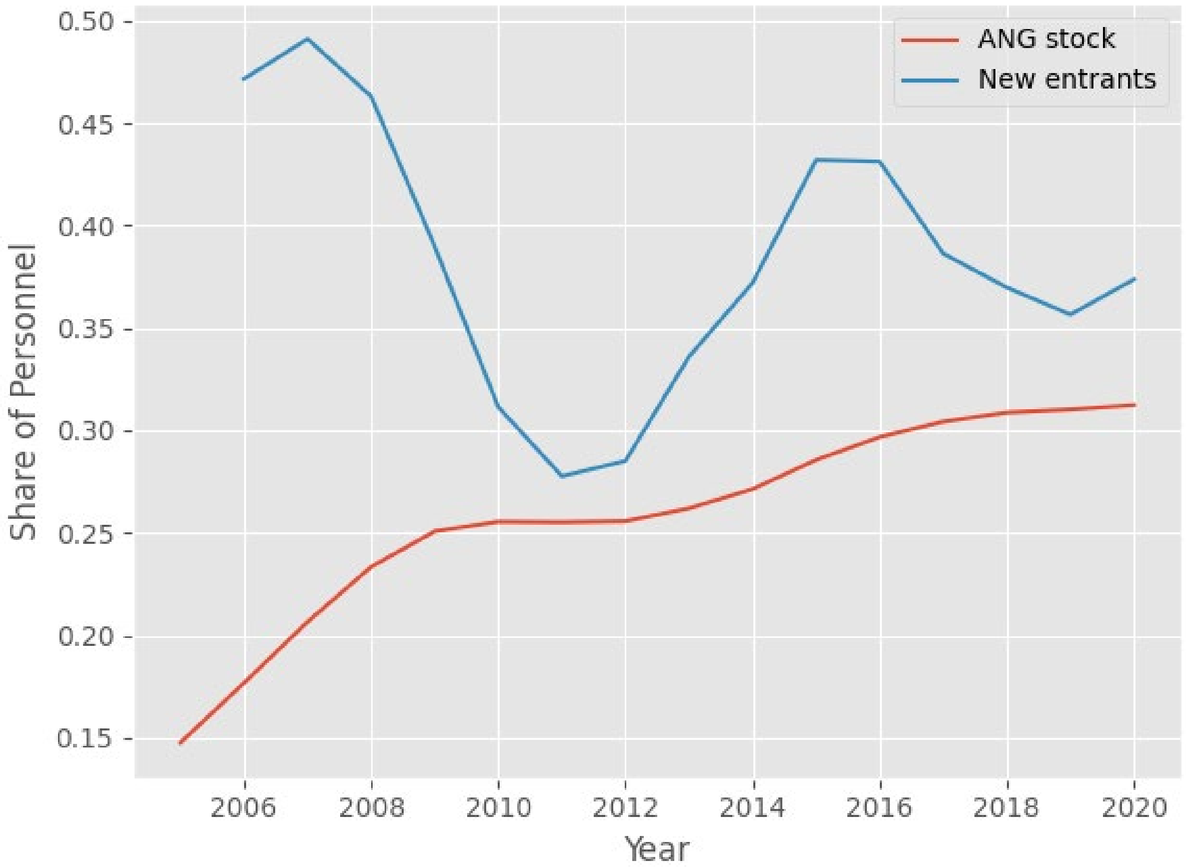
Note: 2005–2020 data. Kaplan-Meier estimates with 95% confidence intervals shown

Share of females among ANG enlisted stock and new ANG entrants



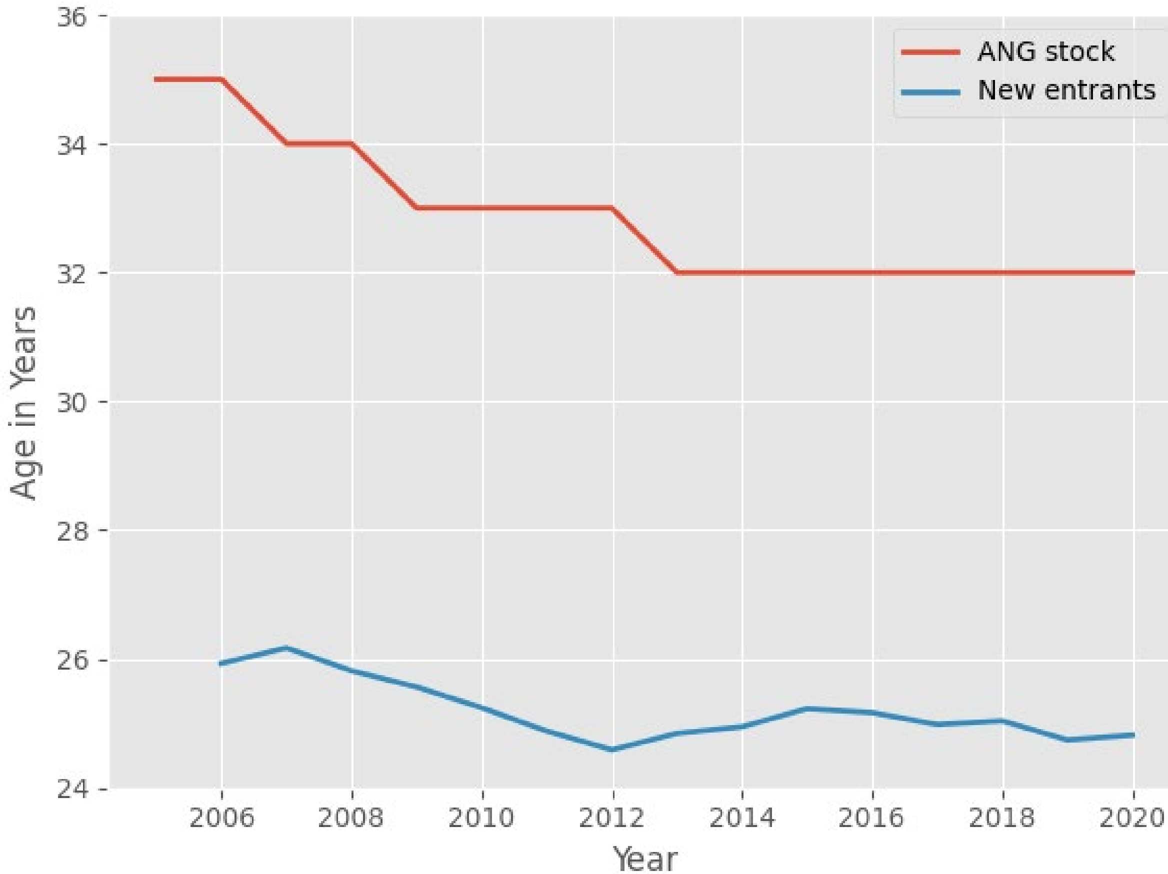
Note: 2005–2020 data, own analyses. New entrants defined as persons not present in previous period(s)

Share of prior service (any service/component) among ANG stock and new ANG entrants



Note: 2005–2020 data, own analyses. New entrants defined as persons not present in previous period(s)

Median age among ANG stock and new ANG entrants



Note: 2005–2020 data, own analyses. New entrants defined as persons not present in previous period(s)

Precision-recall curve

Depicts tradeoff between precision and recall for various probabilistic thresholds

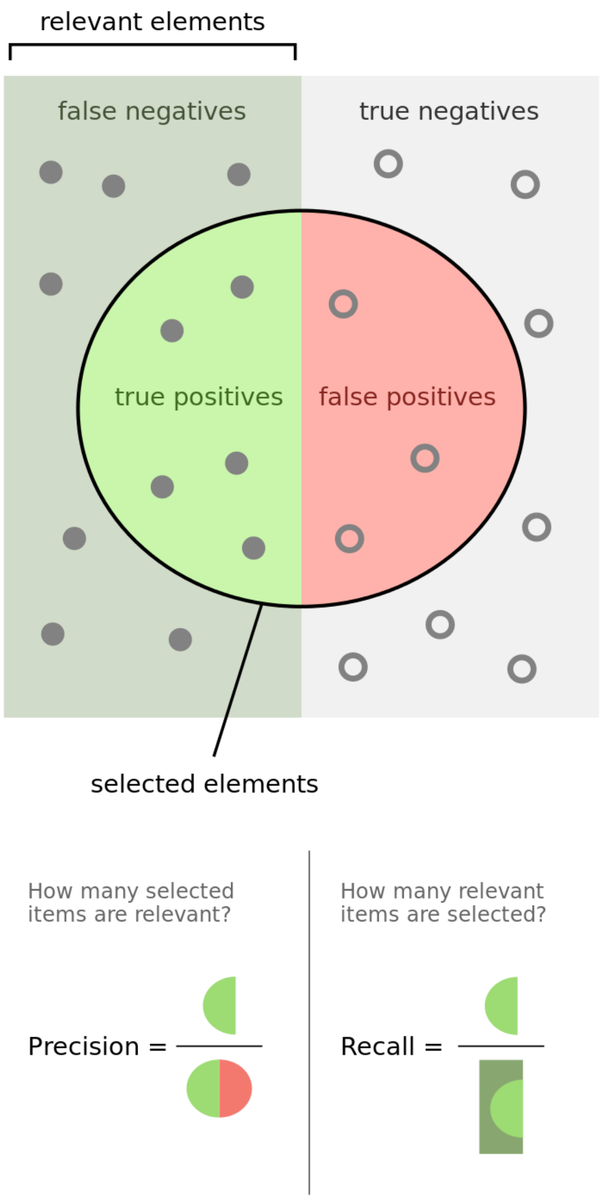
Positive class = exit

Precision: share of true exiters correctly identified

$$\frac{True\ Positives}{True\ Positives + False\ Positives}$$

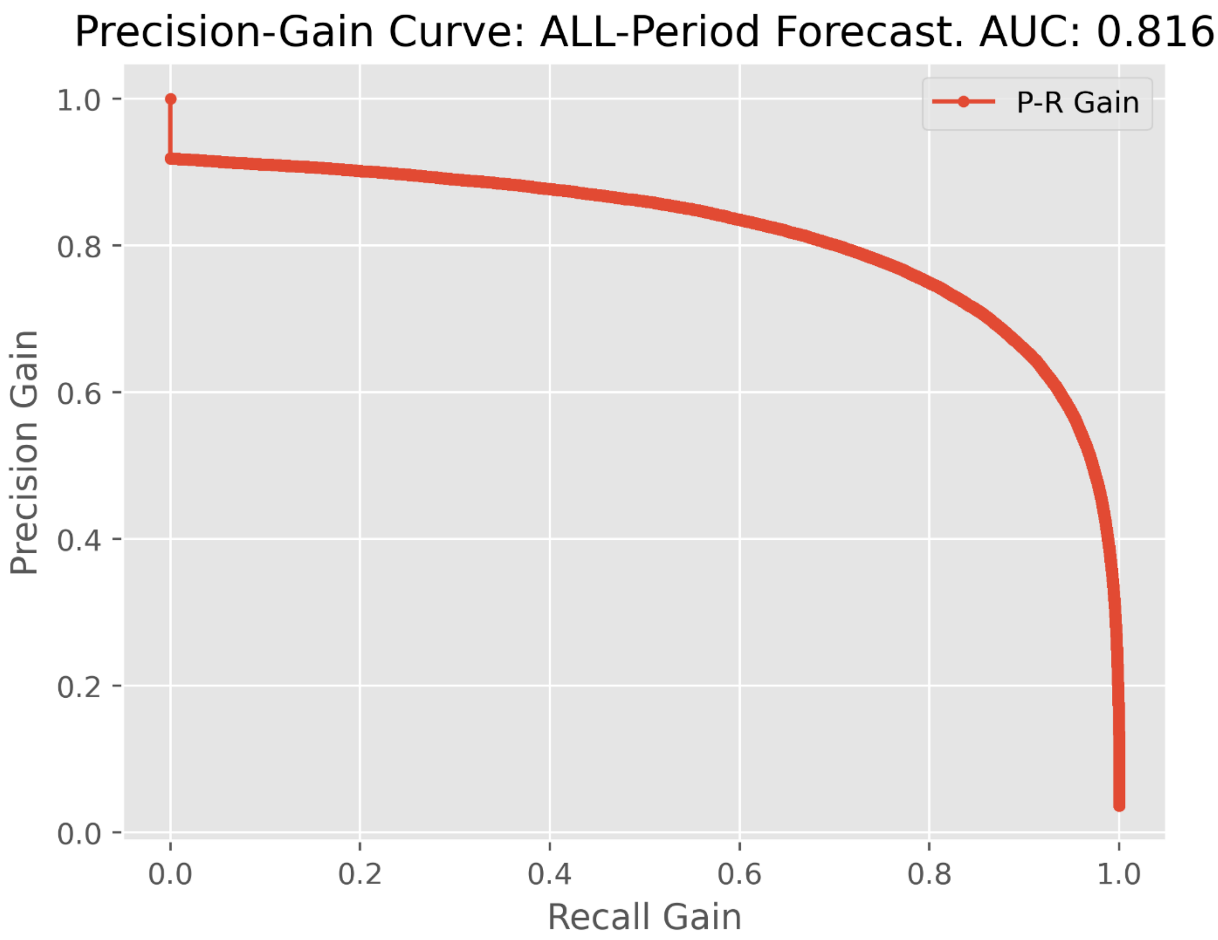
Recall: Share of true exiters identified of all exiters

$$\frac{True\ Positives}{True\ Positives + False\ Negatives}$$



Credit: https://en.wikipedia.org/wiki/Precision_and_recall

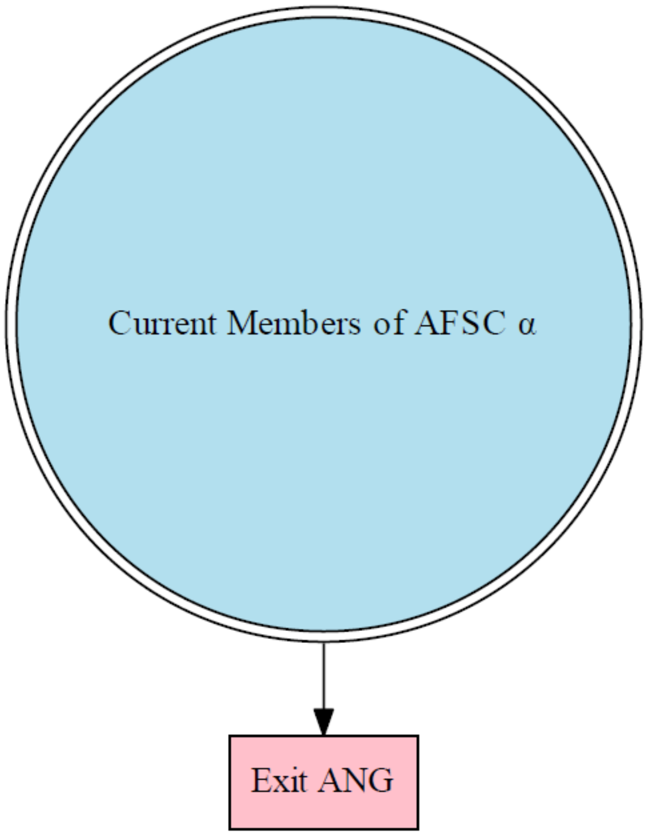
Precision-recall gain curve for 2016–2020 test



Note: Full ANG population meeting analytic restrictions (test: N=90,805). DMDC annual (June) data, 2005–2015. Baseline rate is observed attrition rate (i.e., when precision-recall curve threshold is zero)

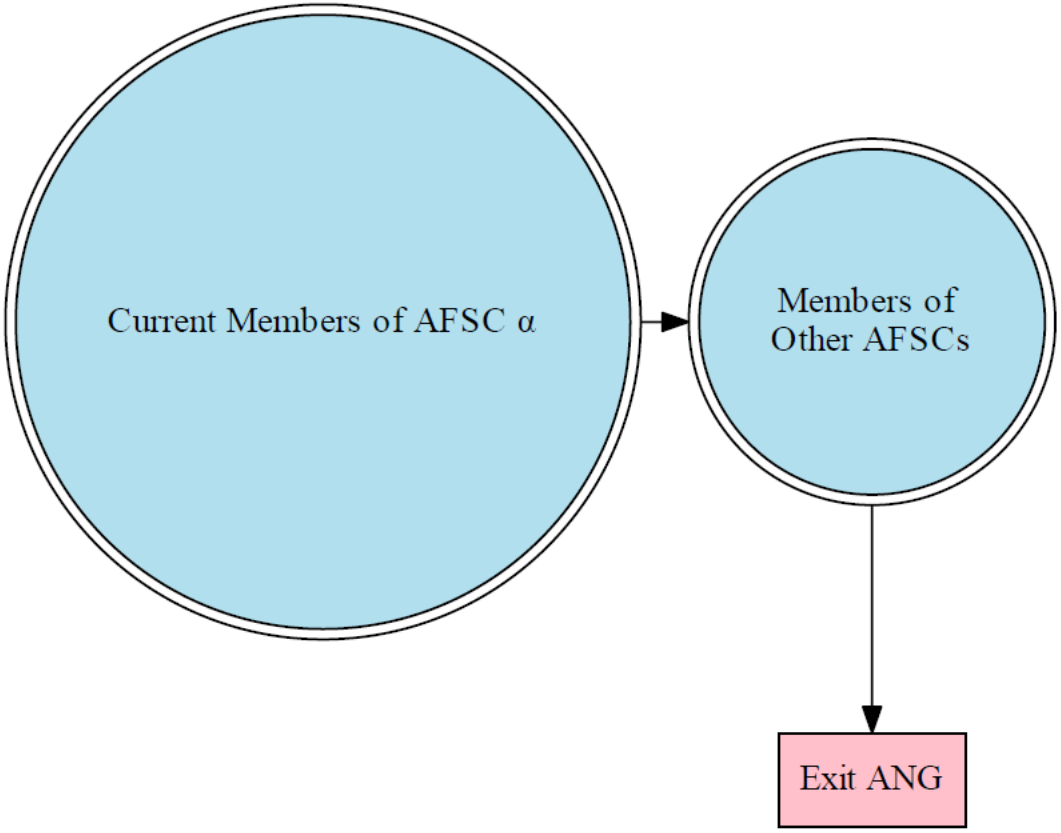
There are several ways for an airman to leave AFSC α

Option 1: Directly exit service from AFSC α



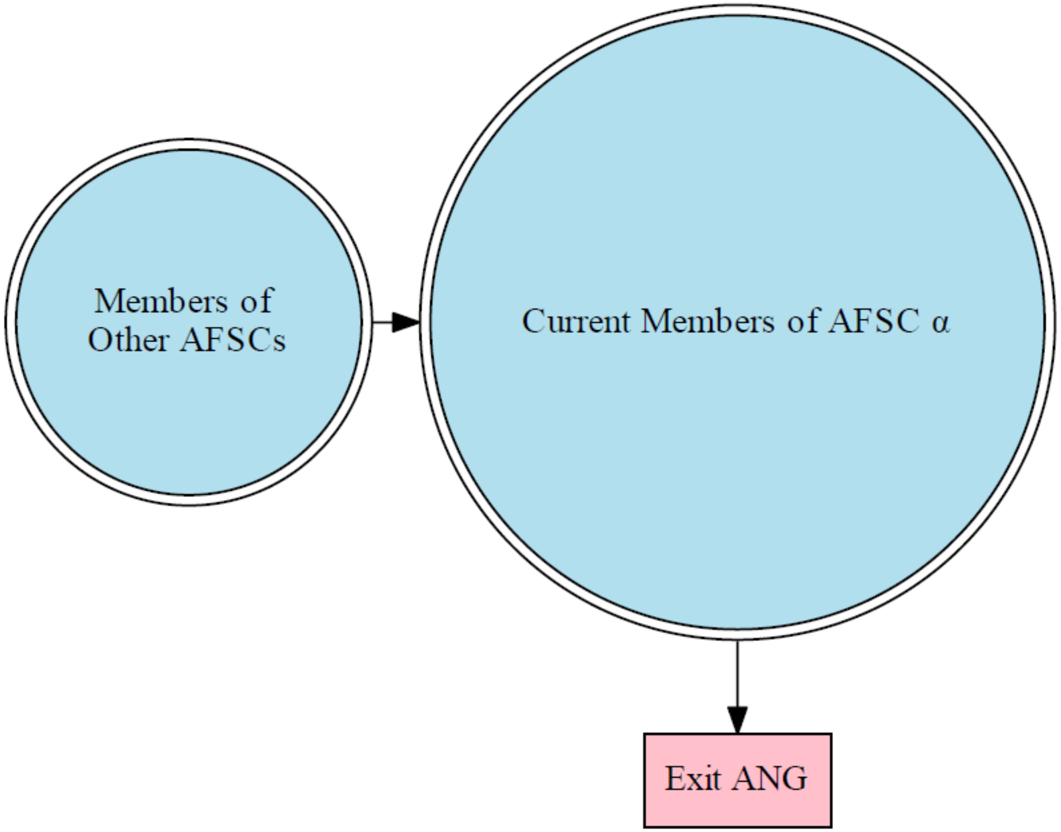
There are several ways for an airman to leave AFSC α

Option 2: Start in AFSC α, switch to another AFSC, then exit service



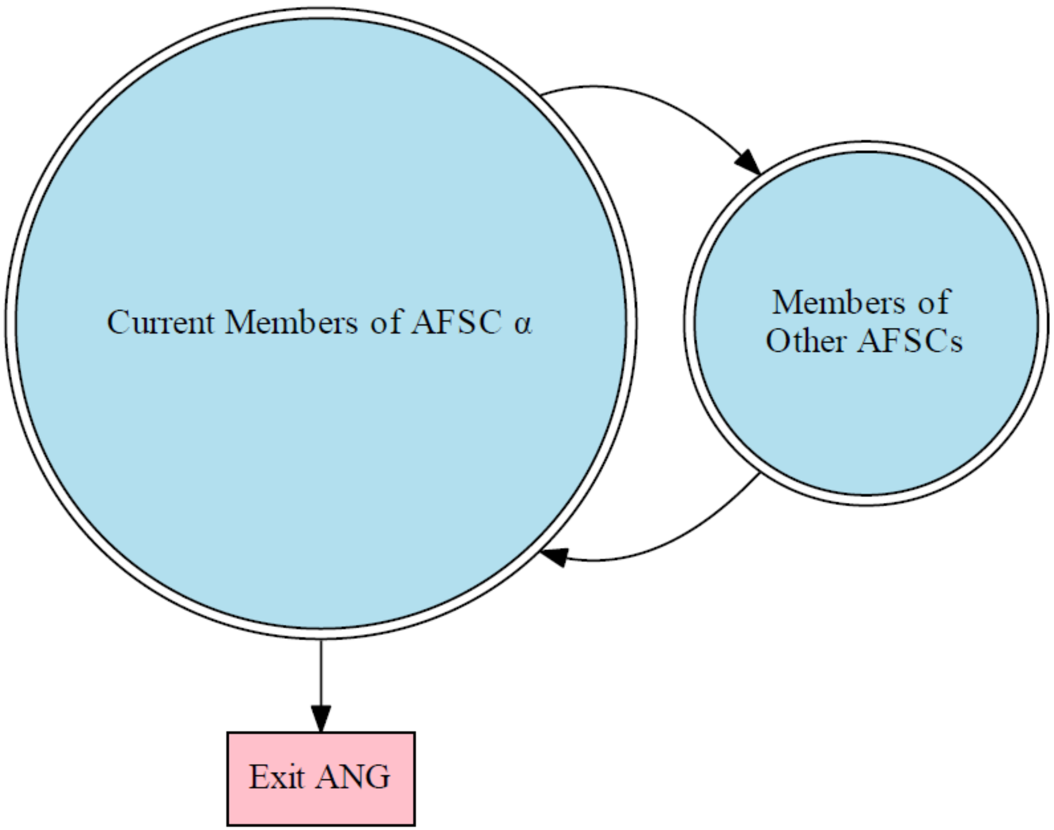
There are several ways for an airman to leave AFSC α

Option 3: Start in another AFSC, switch to AFSC α, then exit service



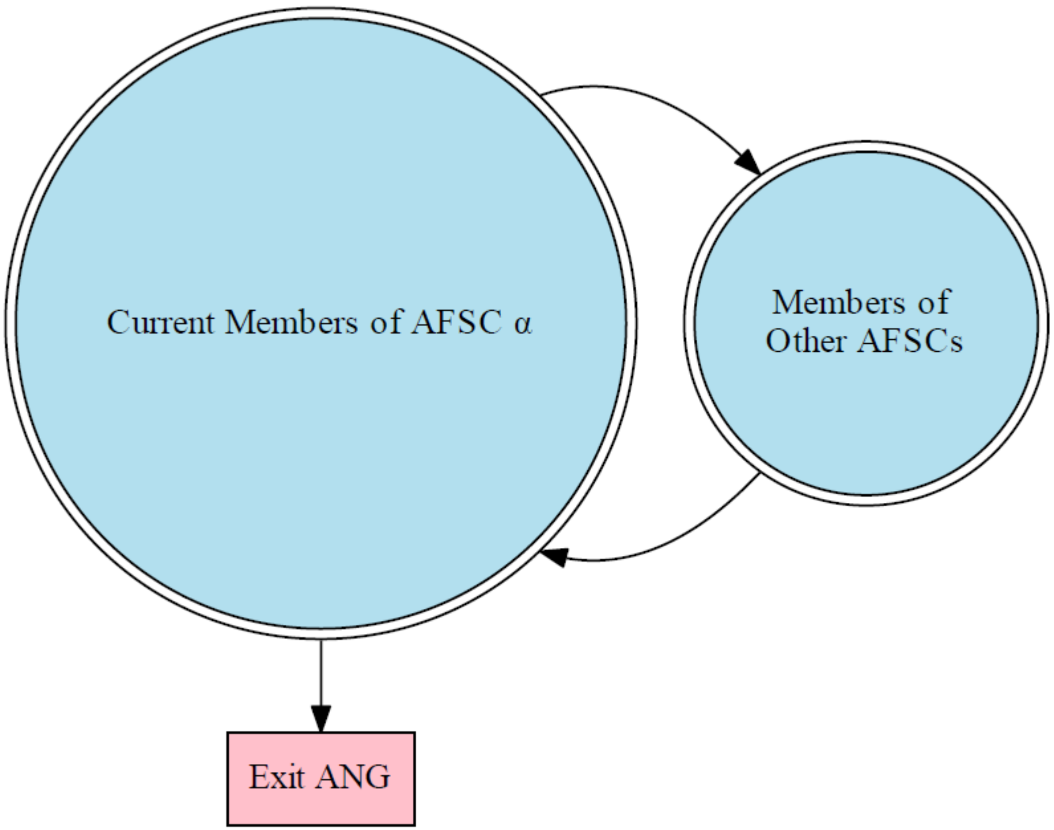
There are several ways for an airman to leave AFSC α

Option 4: Start in another AFSC, switch to AFSC α, switch to another AFSC, then exit service



There are several ways for an airman to leave AFSC α

Option 5: Start in AFSC α, switch to another AFSC, return to AFSC α, then exit service



REPORT DOCUMENTATION PAGE		*Form Approved OMB No. 0704-0188*

1. REPORT DATE (DD-MM-YY)	2. REPORT TYPE	3. DATES COVERED (From – To)
xx-06-2021	Final	

4. TITLE AND SUBTITLE	5a. CONTRACT NO.
Forecasting Demand for Air National Guard Enlisted Training *WEAI Presentation*	HQ0034-14-D-0001
	5b. GRANT NO.
	5c. PROGRAM ELEMENT NO(S).
6. AUTHOR(S) Julie Lockwood Joe King Jay Dennis Akshay Jain Minerva Song	5d. PROJECT NO.
	5e. TASK NO. DZ-6-4720
	5f. WORK UNIT NO.
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311-1882	8. PERFORMING ORGANIZATION REPORT NO. IDA Paper NS P-22653 Log: H 21-000160
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air National Guard Readiness Center Joint Base Andrews NAF, MD 20762	10. SPONSOR'S / MONITOR'S ACRONYM(S) ANG
	11. SPONSOR'S / MONITOR'S REPORT NO(S).

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The Air National Guard (ANG) relies on the U.S. Air Force to provide enlisted entrants with basic military training (BMT) and technical skills (Tech Schools). Because of this, the ANG must reserve training slots for these schools far in advance, requiring them to forecast school demand up to 36 months in the future. While reserving BMT slots is unproblematic, ANG experiences long delays (up to a year) for critical Tech Schools, which harms unit readiness and contributes to inefficient resource expenditure while airmen await training. Using the Retention Prediction Model (RPM), a machine learning tool developed by the Institute for Defense Analyses (IDA), we provide the ANG with enhanced training demand forecasts. Given that demand is filtered through Air Force Specialty Codes (AFSCs), we provide the ANG with the expected number of future exits at the detailed AFSC-level. Our model utilize Defense Manpower Data Center (DMDC) data and ANG training data, spanning 2005-2020 and comprising about 1.4 million person-years for about 215,000 enlisted airmen. Results show the model correctly classifies 71% of service exits and 79% of service retentions across the period, and performs about 19 percentage points better than a naïve model using historical average AFSC exit rates.

15. SUBJECT TERMS

ANG, Retention Prediction Model (RPM), Defense Manpower

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NO. OF PAGES	19a. NAME OF RESPONSIBLE PERSON MSgt Peter Whalen
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include Area Code)
U	U	U	U	54	240-612-7923

This page is intentionally blank.