An Approach to Corroborating the Impact of Recruit Quality and Recruiting Mission on Resource Requirements

SCEA - June 2009

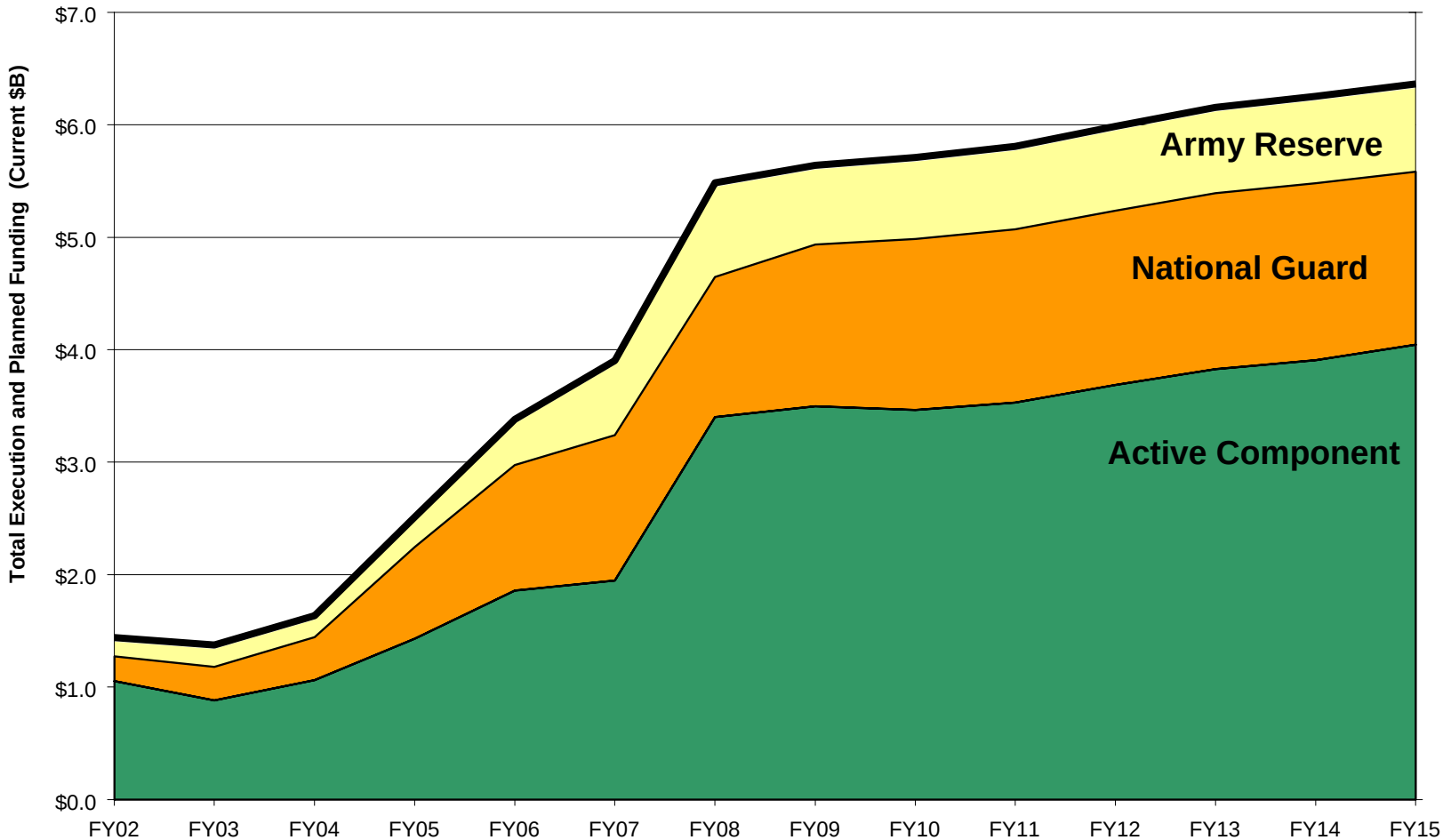
Bob Clemence, Jeremy Heusner, Robert Love, Raissa Nourieva, Meredith Sachs

Agenda

- Project Background
- Data
- Methodology
- Regression Analysis
- Optimization Modeling
- Neural Network Analysis
- Summary/Questions

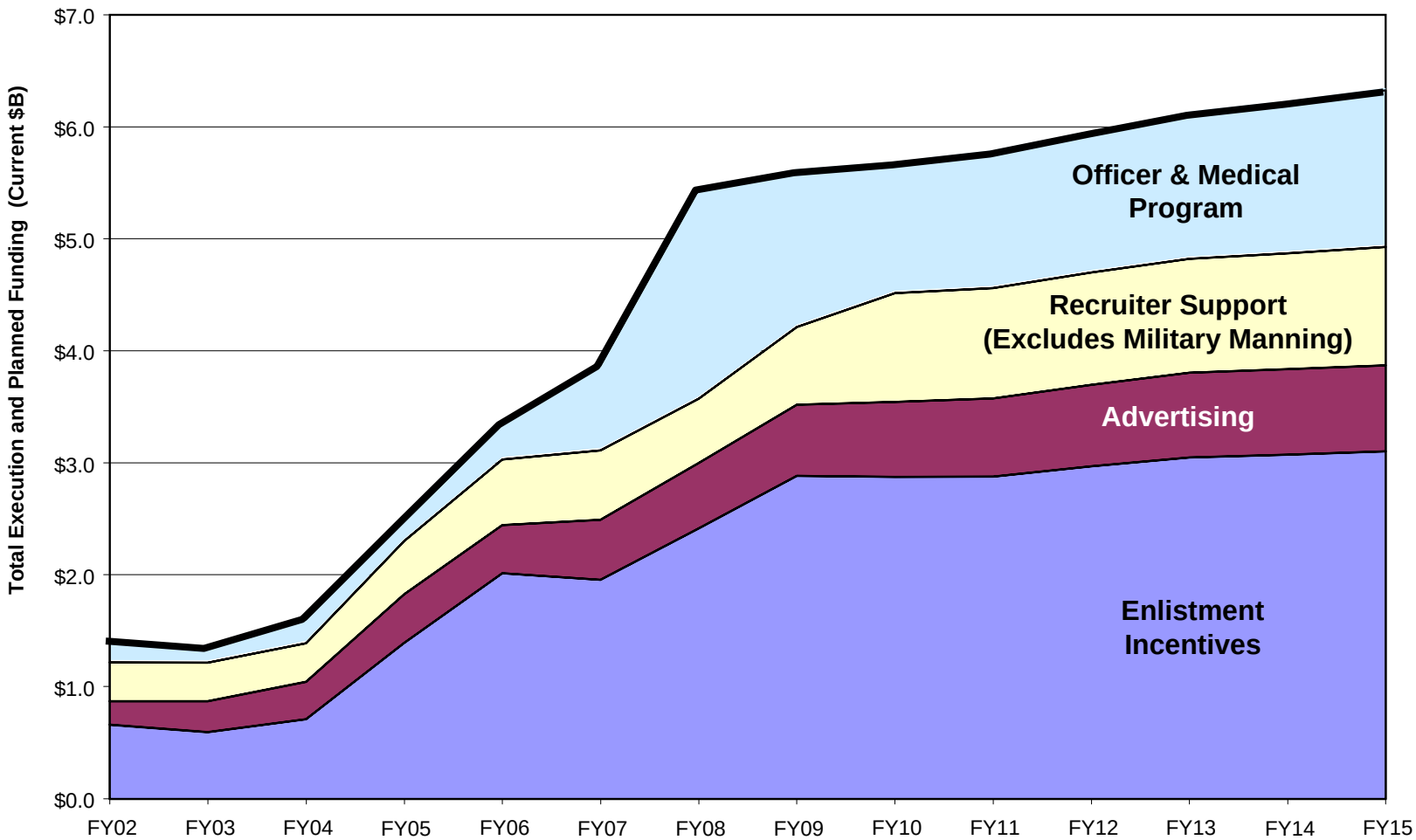
Project Background

Army Recruiting & Retention Expenditures



Project Background

Recruiting & Retention Expenditures by Program



Project Background
Objectives

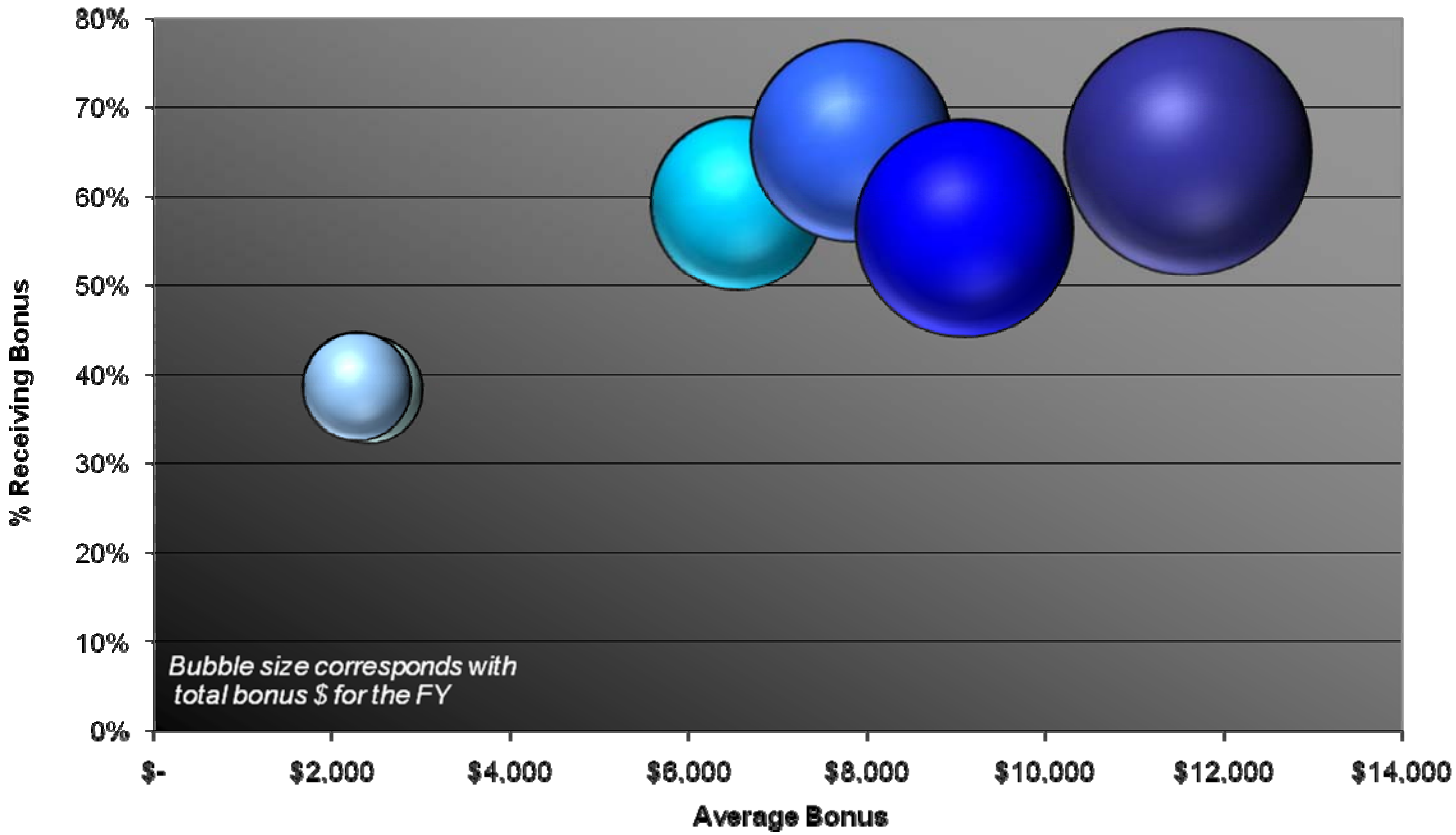
- Provide G-1 with the capability to more accurately forecast the resources required to recruit a high quality force.
  - Assess the impact of changes in funding on recruiting.
  - Assess the impact of changes in force requirements (strength, career field needs, and soldier quality) on programmed recruiting resources.
  - Assess the impact of changes in factors external to the Army, such as unemployment, public opinion concerning the Global War on Terror (GWOT) and perceived levels of danger, might affect recruiting.

Data
Characteristics

- Source: Headquarters Department of the Army G-1 (November 2008)
- 467,000 records of Active Component enlistments between October 2003 and September 2008.
- Information provided in each record included:
  - Armed Forces Qualification Test (AFQT) Percentile Score
  - Enlistment Effective Date
  - Contract Sign Date
  - Military Occupational Specialty (MOS)
  - Career Management Field (CMF)
  - Enlistment Contract Length
  - Amount of Incentive Taken
  - Date of Birth
  - Prior Service Status

Data
Variation by FYs

	Total Bonus $	% Receiving Bonus	Avg Bonus $	Max Bonus Offered	% Receiving Max Bonus Offered
FY03	$ 179,813,000	38.2%	$ 2,426	$ 20,000	3.3%
FY04	$ 183,896,000	38.7%	$ 2,286	$ 20,000	2.6%
FY05	$ 460,501,000	59.1%	$ 6,545	$ 20,000	34.0%
FY06	$ 622,615,086	66.2%	$ 7,830	$ 40,000	0.4%
FY07	$ 731,750,000	56.4%	$ 9,100	$ 40,000	0.9%
FY08	$ 933,670,000	65.0%	$ 11,597	$ 40,000	11.5%



Methodology

Desired Attributes for a Bonus Estimator

- **SIMPLICITY**- Use the fewest independent variables necessary to estimate the dependent variable.
- **EXPLAINABILITY**- Have a plausible causal relationship between the independent variables and the dependent variable, i.e., the form of the model precedes the method.
- **REPEATABILITY**- Good statistical fit when applied to a new set of data.
- **STABILITY**- Impact on output can be anticipated, both in direction and severity, when independent variables change value.

Methodology

Original Modeling Methodology

	Regressions	Forecasts	Supply Curve Development	Optimization
ACTION	Linear regressions conducted by career field	Values for independent variables forecasted and used to predict future bonuses	Cost relationships developed by career field Aggregated results estimate total Army cost of enlistment bonuses	Linear optimization model determines optimal mix of soldiers to recruit for different funding scenarios and quality constraints.
OUTPUT	Regression Equation	Predicted Bonus Values	Supply Curves	Recruit Quality Distribution and Annual Funding Needs

Methodology
Regression Validation and Deficiencies

- Variable Testing
  - Compiled independent variables that were suspected to influence bonus amounts based on recruit attributes, economic conditions, and sentiments towards the Army and GWOT.
  - Variables measuring similar conditions/attributes tested for multicollinearity; those exhibiting high correlation with other independent variables were removed from analysis. Final variables selected based on causality and statistical significance measured by t-tests.
- Equation Verification
  - Split sample testing was performed to ensure that the estimators provided a good fit when applied to a new data sample.
- Equation Validation
  - A 2008 Estimate was computed for total bonuses paid and compared to historical data, the aggregate error of the estimate was +3%.
- Drawbacks
  - Individual forecasts display positive and negative errors that offset each other in aggregate.
  - Data is not inherently linear, preference for non-linear model.
  - Changes in independent variables do not always provide an intuitive impact on overall cost to meet mission.

Methodology

Revised Modeling Methodology

Artificial Neural Networks

ANNs replace the current regression analysis, and produce an average forecasted bonus value

	Regressions	Forecasts	Supply Curve Development	Optimization
ACTION	Linear regressions conducted by career field	Values for independent variables forecasted and used to predict future bonuses	Cost relationships developed by career field Aggregated results estimate total Army cost of enlistment bonuses	Linear optimization model determines optimal mix of soldiers to recruit for different funding scenarios and quality constraints.
OUTPUT	Regression Equation	Predicted Bonus Values	Supply Curves	Recruit Quality Distribution and Annual Funding Needs

Methodology
Neural Network Functionality

- Artificial Neural Networks (ANNs) help to establish interrelationships for large and complex datasets that evade traditional data mining methodologies.
- Employs a self-training technique that allows model to “learn” how changing input values will impact output.
- Performs rigorous search for all possible dependencies in the data.
- Analysis can be performed in a range of complexity from basic to very sophisticated.
- Calculations take place in a hidden layer not visible to the user.

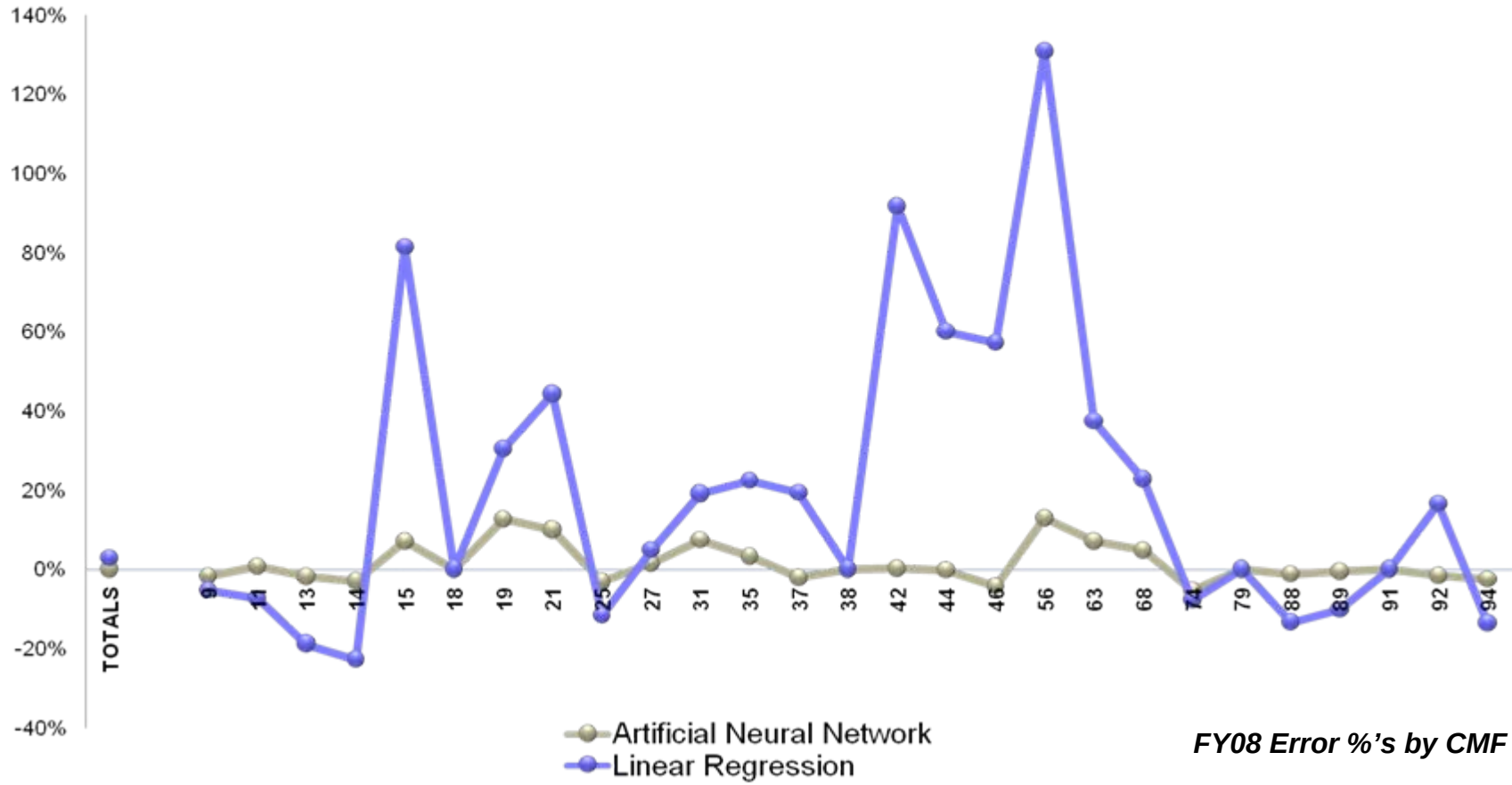
Methodology Comparison

Characteristics of Neural Networks vs. Linear Regression

	Linear Regressions	Artificial Neural Networks
Assumptions	Dependent upon assumptions of linearity between dependent and independent variables	Purely data driven models that do not depend on assumptions about functional form.
Cost Estimating Relationship	Cost estimating equation which defines the impact of changes in independent variables to the dependent variable	"Black box" cost estimating relationship which makes it difficult to predict the impact of changes in independent variables
Multicollinearity	Stability of coefficients is affected by multicollinearity	Capable of accepting a larger number of potential cost drivers and will accommodate multicollinearity
Resources Availability	Methodology guidance is readily available from texts and software	Fewer resources available and learning curve in building and interpreting neural networks is more imposing
Versatility	Produces a cost estimating relationship that may be easily embedded into cost models	There is no cost estimating equation to embed into separate software platform

Methodology Comparison

Accuracy of Neural Networks vs. Linear Regression



FY08 Error %'s by CMF

Methodology

Neural Network Impact on Desired Model Attributes

Desired Attribute	Condition	ANN Impact
SIMPLICITY	Use the fewest independent variables necessary to estimate the dependent variable	Potential to reduce the number of independent variables
EXPLAINABILITY	Have a plausible causal relationship between the independent variables and the dependent variable	Preliminary statistical analysis consistent with regression
REPEATABILITY	Good statistical fit when applied to a new set of data	Enhanced ability to respond to changing conditions
STABILITY	Impact on output can be anticipated, both in direction and severity, when independent variables change value	No traceability for deltas produced in “what if” analysis

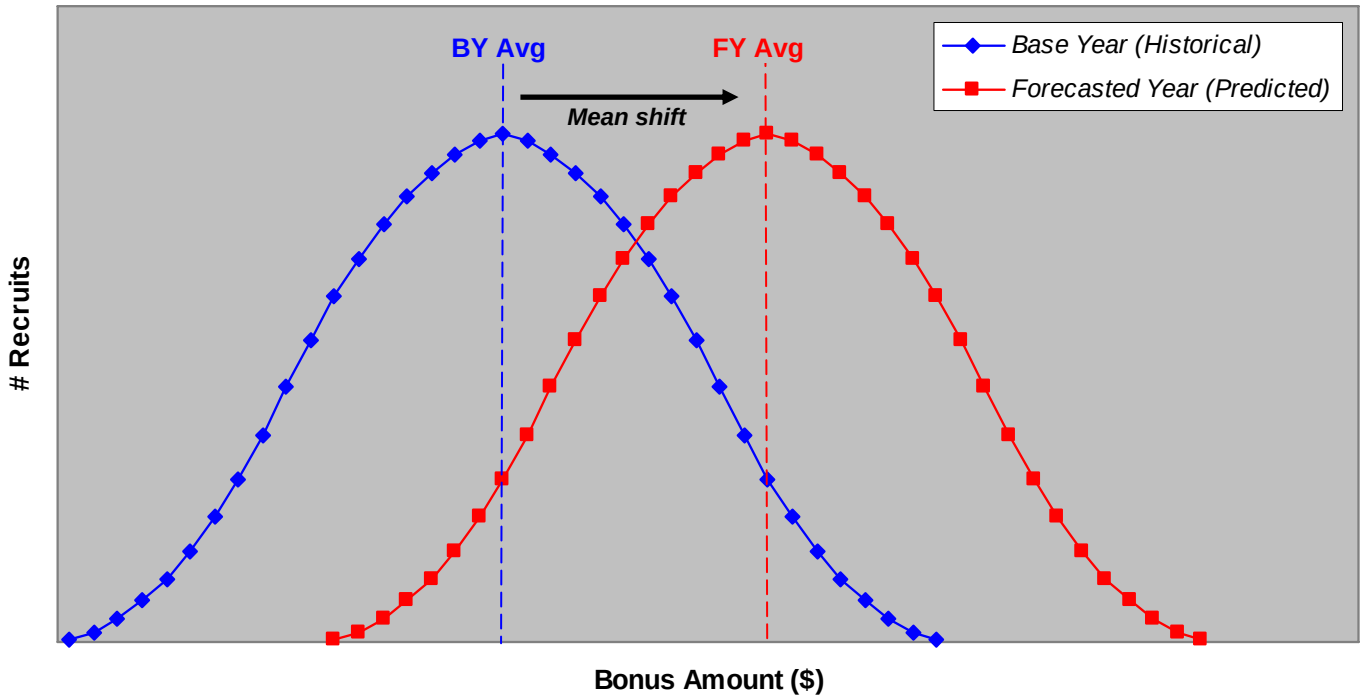
Methodology
Forecasting Values of Independent Variables

Variable	Type	Forecasting Methodology
Enlistment Length	Recruit Specific	Average length of enlistments per month for each CMF and CAT
ASVAB	Recruit Specific	Average ASVAB score per month for each CMF and CAT
Prior Service/ Non-Prior Service	Recruit Specific	Percentage of Non-Prior Service recruits per month for each CMF and CAT
Age	Recruit Specific	Average age of recruits at time of contract signing per month for each CMF and CAT
Contract Sign Date Lag	Recruit Specific	Number of months between the month of contract signing and the end of the fiscal year in which the recruit ships (indicator of urgency on behalf of the Army to meet mission lift)
Job Openings (3 month lag)	Economic	Average Job Openings Rate for the last 1 year of actual data, lagged three months prior to contract signing
Unemployment 16+	Economic	Cyclical average of Unemployment Rate for the last 5 years of actual data for persons over 16 years of age
War Disapproval	War/Sentiment	Average War Disapproval Rate beginning with 6 months after troop surge to last month Gallup poll was conducted
Military Casualties (1 month lag)	War/Sentiment	Average number of US Army deaths as reported by DoD Personnel and Procurement for the last 1 year
Propensity to Enlist	War/Sentiment	Average propensity to enlist value reported by JAMRS DoD Youth Poll for the last 1 year

Optimization

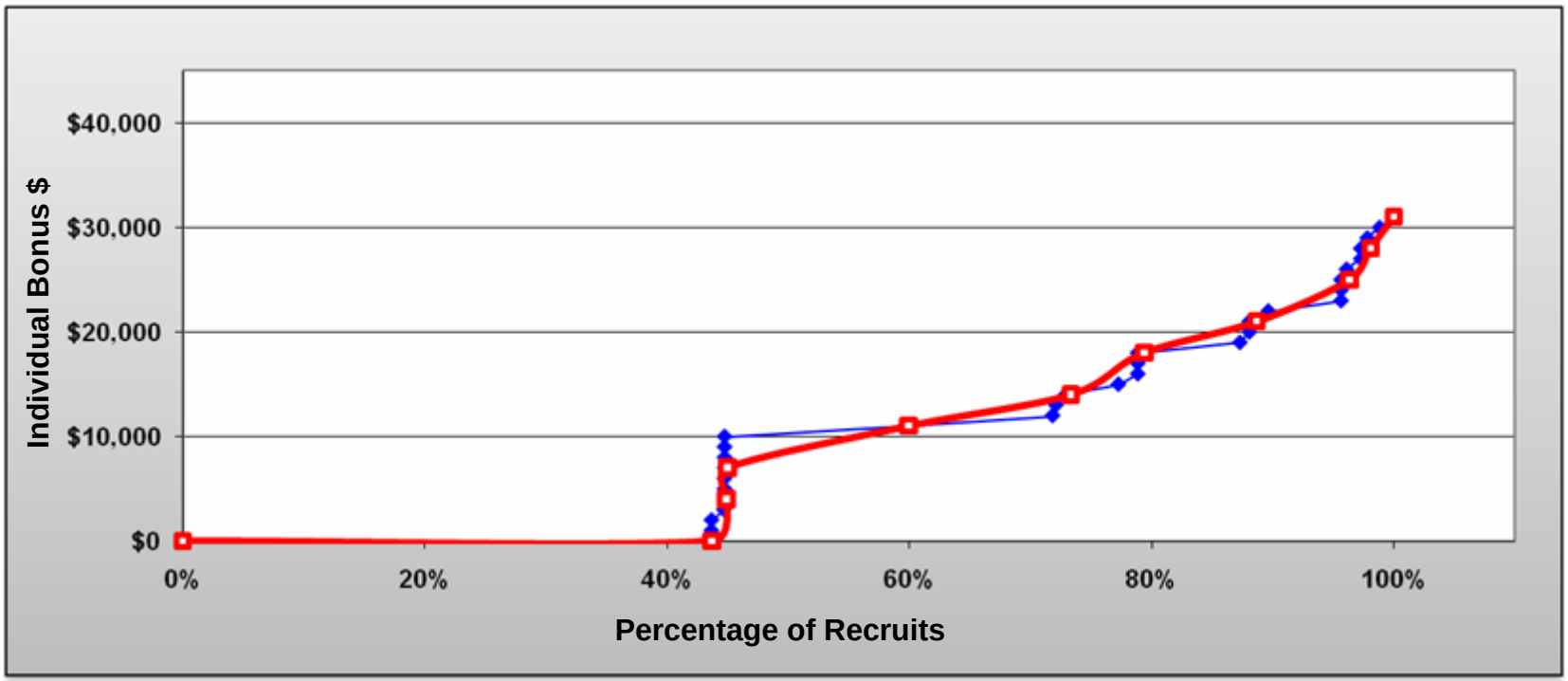
Bonus Distribution for Supply Curve Development

- Forecasted distribution of bonus values assumed to resemble distribution of the base year.
- The difference between the forecasted annual average bonus and the base year annual average bonus is used to shift the mean of the base year distribution.



Optimization
Supply Curve Development

- Forecasted bonus values and distribution are graphed to construct cost vs. quantity plots for each CMF by Soldier CAT.
- Supply curves developed from plots and approximated with piecewise linear functions.



Optimization
Problem Modeled as a Multi-Time Period Linear Program

- The **Objective Function** is to minimize the sum of total costs and budget over-runs in all years
  - Minimizing cost allows implementation of the piecewise linear supply curve without use of integer variables.
  - Including an over-run (shortage) variable identifies the amount and year of budget gaps, when they exist.
- The **Decision Variables** are
  - The number of people to recruit in each segment of the supply curves in each year.
  - The amount of additional funding needed each year to meet Mission and Quality constraints.
- The **Constraints** are
  - Meet mission for each career field (historical percentage of total mission) for each year.
  - Spend no more than is funded each year.
  - Do not violate historical career field quality profiles in any year.
  - Meet or exceed Army High Quality goals in each year.
  - No more than 4% of all recruits can be in TSC category IV in any year.

Summary

- ANN methodology enhances the functionality of forecasting model
  - Improved overall accuracy and reduced CMF error percentages
  - Does not depend on assumptions about functional form
  - Model's learning mechanism allows for continual improvement and incorporation of new information
- Drawbacks of using ANN overshadowed by more valued results
  - Hidden algorithm does not provide insight to the impact of changes in independent variables on the dependant variable
  - Non-standard software poses cost considerations, limits the number of licensed users that can directly interface with the model, and increases learning curve times

Questions?

- Points of Contact
  - Dr. Robert Clemence
    clemence_robert@bah.com
  - Jeremy Heusner
    heusner_jeremy@bah.com
  - Robert Love
    love_robert@bah.com
  - Meredith Sachs
    sachs_meredith@bah.com
  - Raissa Nourieva
    nourieva_raissa@bah.com