Analysis and Modeling of U.S. Army Recruiting Markets

Sponsor: Mr. Joe Baird
HQ USAREC/G-2
Ft. Knox, KY

MAJ Joshua McDonald
Advisor: Dr. Edward D. White
Department of Mathematics and Statistics (ENC)
Air Force Institute of Technology

PURPOSE
Provide US Army Recruiting Command (USAREC) leadership with focused, relevant, and quantitative insight into its missioning process. Missioning is the process whereby missions (recruiting equivalents of a sales goals) are assigned to subordinate recruiting units in order to maximize the number of enlistments produced.

RESEARCH QUESTION
To what extent can we accurately express the relationship between enlistment supply and demand factors, and enlistment contract production?

BACKGROUND
Previous efforts to model behaviors of recruiting markets leave two primary knowledge gaps:
1. Market boundaries which are not directly applicable to the 38 current USAREC unit boundaries (show n):
   - USAREC unit boundaries are defined by ZIP Code Tabulation Areas (ZCTAs) whereas reliable data exists only for the county and higher level.
   - Previous studies have not adequately accounted for this incongruity.
2. No use of validation datasets to thoroughly test the stability of enlistment production models, or provide estimates of response uncertainty in the future.

PURPOSE

METHODOLOGY

1. Data Collection
   - Operational Variables (PMESII-PT) and corresponding subvariables describe aspects of recruiting supply factors
   - Mission Variables (METT-TC) describe salient aspects of recruiting demand factors
   - 26 total metrics represent both types of variables from county-level open sources and USAREC-internal databases
     - Countylevel data weighted to ZCTAs then aggregated to recruiting battalion levels with

\[
Z'_j = \sum_{m=1}^{M} \sum_{n=1}^{N} w_{m,n} Z_n \quad \text{where} \quad Z_n = \frac{\text{the proportional population of county } n \text{ residing in ZCTA } m}{z_i} \quad \text{as of 2010 Census} \quad \text{and } \quad z_i = \text{the available datapoint for county } i.
\]

   - Stochastic Mean Value Imputation applied to create monthly data points from gaps in available annual data

2. Variance Reduction
   - Principal Components Analysis used to reduce multicollinearity between the inter-related aspects of the collected data
     - Horn's criteria helps determine how many principal components of variance are useful to keep

3. Estimation: Stepwise Ordinary Least Squares

\[
y = Xb + e \quad \text{with} \quad b = (X'X)^{-1} X'y
\]

   - Stepwise methods estimate the OLS model to optimize fit (i.e., values of \(R^2(\alpha=0.05)\)) parsimony is obtained by iteratively entering statistically significant terms (\(\alpha_{IN} = 0.05\)) and removing non-significant terms (\(\alpha_{OUT} = 0.1\))
     - Categorical variables model recruiting units and quarters
     - Studentized residuals, \(r_i\), account for x-space distance
     - Lag-1 auto correlations evaluated with Durbin-Watson test
     - Multicollinearity assessed for Variance Inflation Factors (VIFs) of \(> 10\)
     - Approx. Box-Cox transformations correct variance heteroscedasticity

3. Validation
   - 25% of data set aside to test OLS models' predictive accuracy
   - Prediction Intervals account for forecast inputs

\[
100(1-\alpha)\% \text{PI} = \hat{y}_{t+h} \pm z_{\alpha/2} \sqrt{\hat{\sigma}^2 + \sum_{i=1}^{q} \hat{\sigma}_{t+i}^2} \quad t = T, T+1, \ldots, T+\tau-1
\]

   - Validation metrics:

\[
\text{MAPE} = 100 \% \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad \text{MAD} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|, \quad \text{RMSE} = \left[ \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \right]^{1/2}
\]

RESULTS & ANALYSIS

Key Takeaway#1
- A set of 5 continuous variables + 40 categorical variables for battalions and quarters explains 70%, 74%, and 81% of the estimation data for SA, GA, and OTH enlistment contracts.

Key Takeaway#2
- Validation shows models are stable even with forecast inputs

Key Takeaway#3
- Markets can be individually characterized according to predictive accuracy; Example: GA

CONCLUSIONS
- We achieve 530, 170, and 170% relative increases over previous efforts to model SA, GA, and OTH enlistment contract production.
- Market-specific models will likely allow USAREC to complete the optimization portion of its missioning process more effectively.
- Markets at the extremes of production and having poor accuracy in validation warrant further investigation.