



**AIR FORCE OFFICER ATTRITION: AN
ECONOMETRIC ANALYSIS**

THESIS

Jacob T Elliott, 1st Lt

AFIT-ENS-MS-18-M-118

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

**DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED..**

The views expressed in this document are those of the author and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States

AFIT-ENS-MS-18-M-118

AIR FORCE OFFICER ATTRITION: AN ECONOMETRIC ANALYSIS

THESIS

Presented to the Faculty
Department of Operational Sciences
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Operations Research

Jacob T Elliott, BS

1st Lt, USAF

22 March 2018

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED..

AFIT-ENS-MS-18-M-118

AIR FORCE OFFICER ATTRITION: AN ECONOMETRIC ANALYSIS

THESIS

Jacob T Elliott, BS
1st Lt, USAF

Committee Membership:

Civilian Raymond R. Hill, PhD
Chairman

Major Thomas P. Talafuse, PhD
Member

Committee Member 2 Name
Member

Abstract

The abstract of the thesis.

AFIT-ENS-MS-18-M-118

The dedication.

Acknowledgements

The acknowledgement.

Table of Contents

	Page
Abstract	iv
List of Tables	viii
List of Figures	ix
I Introduction	1
1.1 Background	1
1.2 Scope	1
1.3 Assumptions and Limitations	2
1.4 Outline.	2
II Literature Review	3
2.1 Chapter Overview.	3
2.2 The Retention Problem	3
2.3 Previous Research.	4
2.4 Insights.	8
III Analysis and Results	10
3.1 Data Composition.	10
Introduction.	10
HAF/A1XDX	10
Federal Reserve Bank of St. Louis	11
Cleaning and Preparation	13
3.2 Model Selection.	14
Initial Exploration	14
Dynamic Regression	19
Lagged Indicators	27
Officer Subsetting	27
IV Conclusions	28
V TO-DO	29
Bibliography	30
Vita	32

List of Tables

Table		Page
1	Naive Results	16
2	Seasonal Naive Results	17
3	Seasonal Naive RMSE Comparison	18
4	Estimated Coefficients - Initial Model	24
5	Initial Model - Autocorrelation Test	25
6	Model RMSE Comparison	26

List of Figures

Figure		Page
1	Participation and Unemployment	12
2	Monthly Officer Separations	14
3	Seasonal Plot: Total Separations	15
4	Simple and Seasonal Naive Forecasts	16
5	Separations - Outliers Removed	17
6	Seasonal Plot: Outliers Removed	18
7	Seasonal Naive Forecast After Imputation	19
8	Correlation Matrix - Economic Indicators	21
9	Economic Indicators - Raw	22
10	Economic Indicators - Differenced	23
11	Initial Attrition Model - Residual Analysis	25
12	Initial Model - Forecasts Against Validation Data	26

AIR FORCE OFFICER ATTRITION: AN ECONOMETRIC ANALYSIS

I. Introduction

1.1 Background

1.2 Scope

For both releasability and compatibility reasons, the personnel data has been aggregated to the national level, limiting the detail to which relationships can be explored. This was done to match the national economic data available, and to protect personal information of the individuals included in the analysis.

The military personnel data concerns those serving during the 2004-2017 time frame, and the economic data matches. Some extraordinary events occurred during that period, notably the Great Recession beginning in 2008, which may have altered normal behavior and affected relationships. In the vein, the U.S. military is transitioning to a new retirement system in 2018 which may likewise alter the relationship (if any) between the economic environment and military members' separation decisions.

1.3 Assumptions and Limitations

As with any analytic endeavor, several assumptions are made in order to facilitate the modeling of real world phenomena. Perhaps most central to this thesis is the assumption that there exists at least one economic indicator (but ideally many) that helps inform an individual military member's decision to stay or leave active duty service. It is also assumed that if these variables do not directly inform retention decisions, they serve as adequate proxies for unobservable or abstract factors that do influence the individual's decision. For instance, members may not follow the movements of the Consumer Price Index (CPI), but that movement should provide information on the cost of living which may affect the decision to stay in the military. We also assume that the skills held by the Air Force officer corps are largely transferrable to civilian labor markets (citation/justification??). This relationship may not hold equally across all Air Force career fields, however, and is investigated later in Chapter III. Standard assumptions for regression modeling and forecasting are made (independent, normal, and homoscedastic errors) and are tested, as well.

1.4 Outline

This chapter introduced the retention problem investigated and discussed the foundational motivations and thoughts underpinning the thesis. The next chapter reviews the related literature in detail - the efforts used to better frame the problem and previous attempts to model it. The third chapter focuses on the methodology. It documents how and why the data was attained (i.e. sources and selection criteria), as well as any transformations necessary to conduct analysis. The chapter continues by discussing the modeling procedure in detail, including general steps and specific mathematical formulations. Lastly, the results are examined and insights or conclusions are highlighted.

II. Literature Review

2.1 Chapter Overview

Managing personnel and modeling retention behaviors have, appropriately, long been a concern of the Department of Defense. This chapter summarizes the retention problem, examines previous research endeavors, and finally discusses the impetus for the econometric approach used in this research.

2.2 The Retention Problem

All organizations have some problem associated with retaining their people. This is especially true of the the military, wherein members are routinely confronted with deployments, long duty hours, and frequent relocations - factors generally not found in non-military organizations. These factors produce high stress on the military members and their families, who play a significant role in a member's retention decision [8]. Evidence suggest that individuals serving in the military are generally more tolerant of these conflicts [5], but the causes of attrition involve more than just familial concerns. Kane [13] argues the military suffers from a chronic personnel mismanagement problem: Members' merit is not rewarded nearly as well as it is in the private sector, in terms of personal recognition and upward movement, partly due to heavy bureaucratic restrictions. This disparity can lead to frustration and job dissatisfaction, damaging the member's commitment to the organization and incentivizing attrition [5].

Compounding the internal frustrations, civilian labor markets can offer intense incentives for leaving. Barrows [4] details the mechanisms underpinning U.S. Air Force pilot attrition

to civilian airlines, framing the problem with human capital theory. The military offers a unique opportunity for developing highly desired skill sets, placing member's in positions of high stress and responsibility at early stages of professional development [13]. Furthermore, evidence suggests that military as an institution is quite adept at attracting intelligent and capable individuals [2]. Providing innately talented individuals with a high degree of general and specific training fosters the development of high-performers with desirable and broadly applicable skill sets. Therein lies the problem, civilian firms are typically more flexible in their ability to compensate individuals through organizational advancement and wage, often outcompeting the military [13]. These phenomena are in direct contradiction to the principles for successful retention laid out by Asch [3]. She explains that in order for military compensation to be attractive, it needs to be at least as great as the member's the expected wages and benefits offered by civilian labor markets. Compensation should also be contingent upon performance, reflecting the individual's value to the organization, to maintain motivation and disincentivize attrition [3]. In order to best determine compensation, then, it behooves the military to develop methods for anticipating the effects of labor market conditions on military members' retention decisions.

2.3 Previous Research

There have been many forays into personnel retention modeling and forecasting. Saving et al. [15] find a significant interaction between labor markets and military retention by analyzing individual career fields within the U.S. Air Force. Their results indicate that demographic factors such race and education level are influential to retention at early stages, but exhibit diminished effects as careers progress. Additionally, their work supports the conjecture that civilian wages, unemployment rates, and other economic variables affect retention.

In 1987, Grimes [11] investigated the retention problem by applying a variety of regression

methods (ordinary multiple linear regression, with logarithmic transformations on response and/or explanatory variables) to predict officer loss estimates 6-12 months in the future. He was unable to provide adequate effects estimates or reliable predictions, concluding that the chronological nature of the data led to serial correlation errors.

Fugita and Lakhani [8] use survey and demographic data compiled by the Defense Manpower Data Center to estimate hierarchical regression equations to describe retention behaviors in Reservists and Guard members. Hierarchical regression models are useful when there exists some causal ordering among predictors, as is often the case with demographic and economic data. This causal relationship can lead to high multicollinearity, increasing the estimated standard error of coefficient estimates and resulting in non-significant predictors. They find that, for both officers and enlisted, retention probabilities tend to rise with increased earnings, years of service, and spousal attitude towards retention. Their work reinforces the importance of including demographic variables in retention modeling, and that wages are in the forefront of a member's mind when deciding to stay.

Gass [9] takes a more general view by modeling the manpower problem in three different ways: as a Markov chain with fixed transition rates between nodes, as a minimum-cost network flow problem, and as a goal-programming problem. While potentially easier to interpret, these models can present a too-sanitized picture of an enormously complex system, particularly the current military personnel system.

Barrows [4] analyzes retention, specifically for Air Force pilots, through the lens of human capital and internal labor market theories. He argues two points important to this thesis: the degree of specific training is inversely correlated with attrition, and that the Air Force personnel system suffers from the inefficiencies typical of an internal labor market.

To the first point, the military offers a high degree of general and specific training. General training is conducive to attrition, as it allows the individual to more easily transfer between jobs. Specific training, on the other hand, decreases worker transferability and helps enforce retention. This effect is seen in differing retention rates between general pilots (cargo,

heavies) and those with more specific skill sets (helicopters, fighters). One can imagine this would also reveal itself in the non-rated officer population; that is, career fields with transferable skill sets suffer more from attrition than those with specific skill sets, e.g. think logistics (general) versus aircraft maintenance (specific).

Regarding the second point, workers are somewhat insulated from the competition posed by outside labor markets (e.g. a Major does not have to worry about a civilian being hired to replace her), and are paid according to position as opposed to productivity. Shielding employees from outside competition can possibly remove incentive for performance; individuals who feel more secure in their jobs may not try as hard. Not paying according to performance can be damaging in two ways: high-performers can feel undervalued and motivated to leave, and under-performers could be receiving more than they produce.

Looking to the Navy, specifically Junior Surface Warfare Officers (SWOs), Gjurich [10] found that one of the most important factors affecting retention was marriage. Single officers are more likely to leave than those with families. This actually may be a proxy for risk aversion, however. Those officers with dependents may be less likely to risk unemployment by leaving the military, choosing instead to keep a relatively secure job. Again, the importance of demographic factors was reinforced, but little is said of the economic considerations.

In 2002, Demirel [6] used logit regression to analyze retention behaviors for officers at the end of their initial service obligation and at ten years of service. While the focus of this endeavor was to identify any changes in retention related to commissioning source, several other demographic factors - such as marital status, education level, and sex - were found to be statistically significant. This reinforces conclusions about demographic factors drawn by previous research efforts, and shows evidence that these trends generally apply to the military population, instead of particular service branches.

Ramlall [14] takes a less technical approach and surveys the existing employee motivation theories to offer an explanation of how employee motivations affect retention, and how the disregard for the principles contained therein motivate attrition. Many causes are discussed,

and a few are consistent (or at least common) amongst the spectrum of motivation theories. When wages and promotions are not seen to be tied to performance, individuals are disincorporated and do not feel as loyal to the institution. Also, a lack of flexibility within job scheduling and structure is seen as disloyal or disrespectful to the individual. Lastly, when managers fail to act as coaches or aren't seen as facilitators to employees' career's, turnover rates tend to be greater. Given that civilian labor markets are generally more flexible in both pay structure and work scheduling, this research underpins the importance of incorporating civilian labor market conditions.

More recently, Schofield [16] employs a logisitic regression model to identify key demographic influencing the retention decisions of non-rated Air Force Officers. She finds that career field grouping, distinguished graduate status at commissioning source, years of prior enlistment, and several other structural variables were significant. She then utilizes these factors to generate a series of survival functions describing retention patterns and behavior. Again, the importance of demographic factors is reinforced. However, any possible effects of economic factors were unexplored.

Looking at the rated officer corps, Franzen [7] takes a similar approach - again using logisitic regression to identify significant factors used to generate survival functions. However, this attempt differs from Schofield by choosing to also assess the influence of economic, demographic, and other variables exogenous to the military. She finds that marital status, number of dependents, gender, source of commissioning, prior enlisted service, and the New Orders value from the Advance Durable Goods Report were all significant. The first couple support the aforementioned notion that familial strain caused by military service affects retention, the next few (gender, source of commissioning, and prior service) reaffirm the work conducted by Schofield. The last variable, New Orders, suggests that indicators of economic health play some role in retention decisions. This last observation is a motivation for this thesis.

In that vein, we look to the work conducted by Jantscher [12] where she conducts cor-

relation analysis to determine the relationship between a host of economic indicators and retention rates for each Air Force Specialty Code (AFSC). The results of the preliminary correlation analysis provide a subset of economic indicators shown to be significantly correlated with retention, such as unemployment rates, gross national savings, real GDP growth, etc. She then attempts to form a regression model to forecast retention, but was unable due to high multicollinearity between many indicators. Nonetheless, her correlation analysis provides a starting point from which additional modeling techniques may be applied.

2.4 Insights

Based on the review of the literature, several key themes arise:

- Demographic and economic factors play a significant role in a member's attitude towards retention
- Military members are aware of and incorporate opportunities in the civilian labor market when deciding to remain in or leave military service
- Logistic regression on demographic data yields promising results when predicting whether an individual will remain in service, but may be inappropriate for modeling aggregate trends
- Effects estimation of economic factors through regression can be difficult, as many indicators are highly correlated

What is also apparent is that there are several topics yet unexplored:

- Modeling the military population with performance-based pay structures and advancement schemes to examine effects on retention
- Exactly how comparable the military population is to the civilian, how easily the professional skills sets exhibited by the former transfer to the latter
- Applying other forecasting techniques (ARIMA, Exponential Smoothing, Dynamic Regression) to the retention problem

This thesis focuses on the last point. We attempt to forecast Air Force Non-rated officer retention with a dynamic regression model in order to estimate the effects of different economic indicators, documenting the process in the next chapter.

III. Analysis and Results

3.1 Data Composition

3.1.1 Introduction

Before any predictive or descriptive analysis can begin, the data must be understood. Every data set has its idiosyncracies, its own unique challenges. Understanding these characteristics and the meaning of the data - what the variables represent and how they might interact with each other - is key to any successful analytic endeavor. Here, the data are described in detail - sources, meaning, and peculiarities.

3.1.2 HAF/A1XDX

The Strategic Analysis branch of the Force Management Division (AF/A1XDX) provided the data on Air Force personnel used in this thesis. The data are extracted from the Military Personnel Data System (MilPDS), a database containing Air Force personnel data for every airman over his or her career. The data are input by trained personnelists or are automatically updated within the system (e.g. age will automatically increase annually) The data were originally split into two separate `.sas7bdat` files, one containing monthly attrition numbers for each Air Force Specialty Code (AFSC) and the other detailing monthly assigned levels for each AFSC. Each file contains information starting in October of 2004 through September of 2017, for a total of 156 observations across 67 AFSCs.

3.1.3 Federal Reserve Bank of St. Louis

The Federal Reserve Bank of St. Louis is one of 13 entities which comprise the United States' central bank (the others being 11 regional reserve banks and the Board of Governors). As a whole, this group is responsible for deciding on and enacting monetary policy for the U.S. They maintain expansive databases containing information about the economic environment - financial data, national employment statistics, private sector business data, etc. Fortunately, the Federal Reserve Bank of St. Louis offers public access to the Federal Reserve Economic Data (FRED) database via online interface. From here, historical data on several economic indicators were retrieved: the nation unemployment rate (both seasonally adjusted and non-adjusted), the labor force participation rate (LFPR), job openings (adjusted and not), total nonfarm job quits, the FED's labor market momentum index, real GDP per capita, and the consumer price index (CPI). Each indicator consists of monthly recordings across varying time spans (e.g. 1990-2016 vs 2001-2017).

The LFPR is the percentage of the population actively employed or looking for employment. Its movement can give insight into the strength of the economy - e.g. rising participation is usually associated with growth. When paired with unemployment rates, the LFPR can also reveal people's attitude about the economy. For example, the steady decline of participation from 2010 onward (seen in Figure 1) might indicate that the decrease in unemployment over the same period is somewhat exaggerated; people seeking, but unable to find, work may become discouraged and exit the labor force, artificially decreasing the unemployment rate. It is possible that this perception of economic health affects retention decisions. In this thesis, LFPR is restricted to members of the civilian labor force with at least a baccalaureate and no younger than 25 years of age, the subset most closely related to officers.

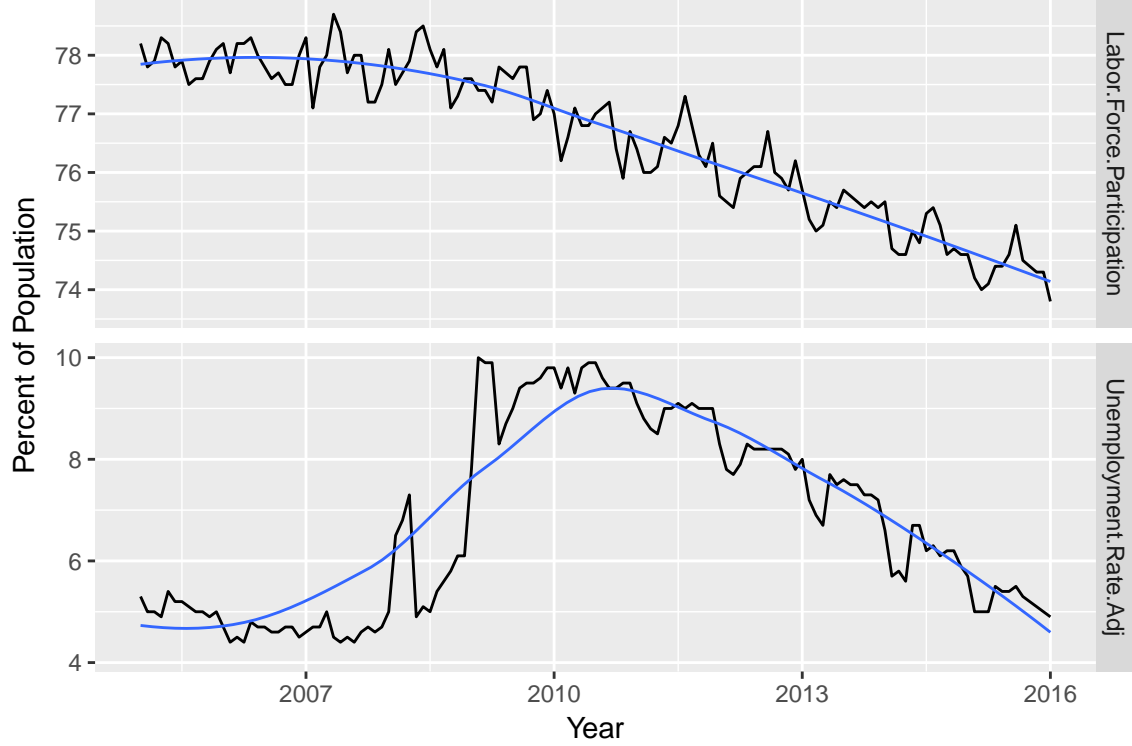


Figure 1. Participation and Unemployment

It is assumed that the skillsets of our target population (non-rated line officers) are most transferrable to those jobs covered by nonfarm payrolls. Nonfarm is a category that excludes proprietors, private household employees, unincorporated self-employment, unpaid volunteers, and farm employees [1]. Job quits are generally voluntary separations and may reflect workers' willingness to leave; it may be that the a higher propensity to volutarily leave a job translates to a positive outlook on obtaining another.

The labor market momentum index compares current labor market conditions to historical averages. A negative value indicates conditions below the long-term average, and a positive value indicates favorable conditions. The CPI examines the weighted average price of a basket of consumer goods and services; it is used to estimate the cost of living. There is some uncertainty involving employment in separation from the military, so cost of living information may be especially important to that decision as the military is excluded from CPI statistics.

By including these variables in a regression model and estimating their effects, this thesis

is attempting to capture military members' perceptions of economic health and job prospects, and use that information as a means to forecast attrition.

3.1.4 Cleaning and Preparation

Perfect data is rarely found or received outside of the classroom, and such is the case here. Before exploration and modeling, several steps are taken to produce a useable data set.

The personnel data is first converted from long to wide format. This procedure generates missing values, and they need to be dealt with appropriately. Missing values can result from several underlying issues: data storage corruption, entry errors, miscommunication between software, etc. Luckily, none of those apply here. Since the attrition data is a monthly count of people exiting USAF service, the intuition is that these missing values simply represent a lack of an observation (i.e. zero separations). This is confirmed by the data's provider, and so all missing values are replaced with zero. Initially, observation dates are stored as the number of days since 1 Jan 1960 (the standard for SAS). This is transformed into YYYY-MM-DD, in order to merge with the economic data. Several columns are added, each the total separations for a specific subset of the officer corps: rated, non-rated line, non-line, etc. These sums will be the responses used for modeling.

The economic data does not require much treatment, as it comes from a professionally managed database. One of the indicators, real GDP per capita, occurs in quarterly intervals - the rest are monthly. In order to combine, the observation for a quarter is carried through the quarter (e.g. the observation for Q1 2006 is applied to January, February, and March 2006). Then, variables are renamed for clarity. Finally, economic data are merged with the personnel through an inner join, preserving only those observations in common.

3.2 Model Selection

3.2.1 Initial Exploration

Before throwing around modeling techniques, the data is visually inspected. Plotting one response, total separations over all career fields, in Figure 2 shows significant spikes during 2005, '06, '07, and '14. We know that during these periods, special separation incentive programs were introduced in order to artificially downsize the air force. The effects of these periods merit investigation later on, as they could negatively affect model performance.

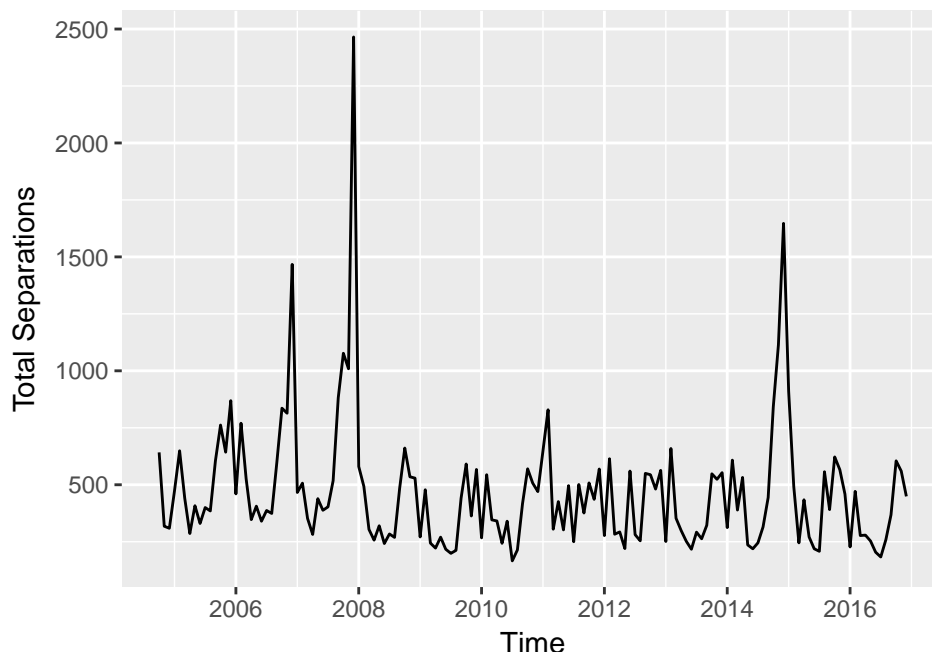


Figure 2. Monthly Officer Separations

No seasonality is immediately obvious in Figure 2. However, if each year is plotted separately, a clearer picture emerges. First, Figure 3 shows that the extreme points noticed above seem to be relegated to the November-December time frame. Second, it is easier to witness the seasonality: bowing across the year, with higher counts at the beginning and end.

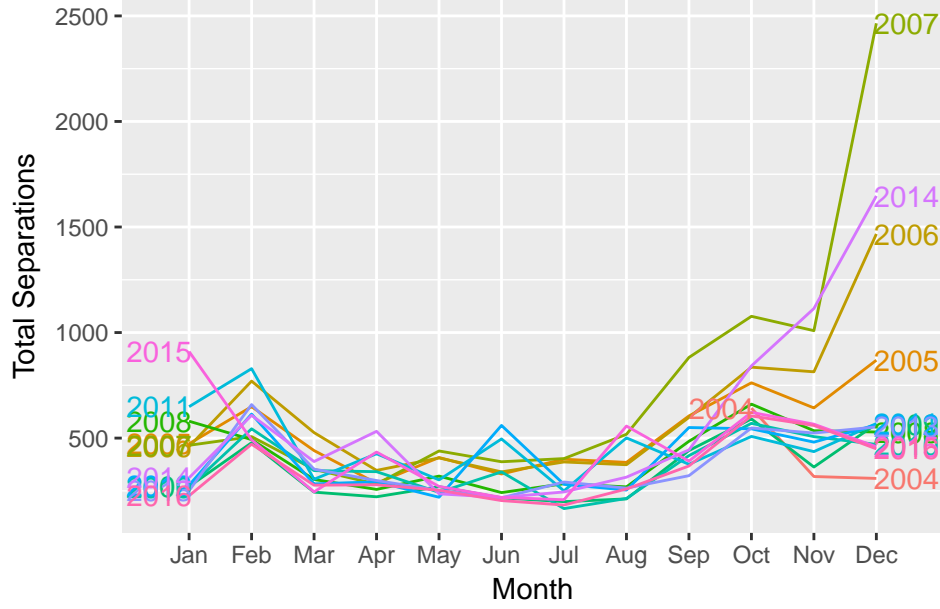


Figure 3. Seasonal Plot: Total Separations

Considering these plots, we expect that a seasonal model performs best and that we will have to perform some alteration to accomodate the outliers. To confirm, we fit naive models to the data and examine to results. Beyond revealing seasonlity and outlier effects, fitting naive models serves to establish a baseline to compare later models against. Naive models are as simplistic as possible, so if later models perform worse or only marginally better, it implies they are not capturing much information.

Figure 4 gives evidence to the negative effects of outliers, notice the large confidence intervals surrounding the naive forecast and the 2014 spike carried through in the seasonal forecast.

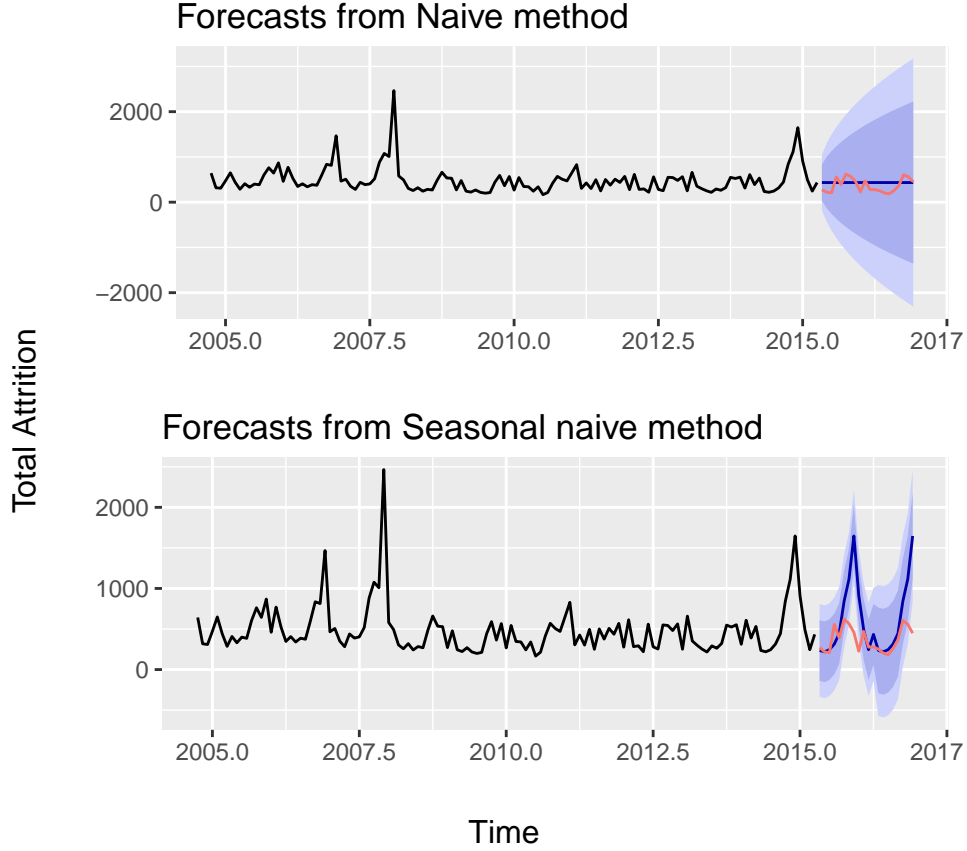


Figure 4. Simple and Seasonal Naive Forecasts

Tables 1 and 2 show different error metrics for each of the two models. By mean square error (MSE), the seasonal model generally fits the training data better, possibly indicating presence of seasonality effects. However, there is a large disparity between validation MSEs, possibly caused by the major spike in 2014 - reaffirming the earlier intuition about outlier effects.

It is known that during years 2005, '06, '07, and '14 special separation programs were implemented. Given the effect those years appear to have on modeling, they ought to be dealt with before continuing. Before deciding how, the explicit points in question need to be

Table 1. Naive Results

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	-1.650794	312.5230	195.9841	-12.53034	42.09254	1.2607358
Test set	-62.600000	160.6421	144.3000	-37.26534	51.56901	0.9282598

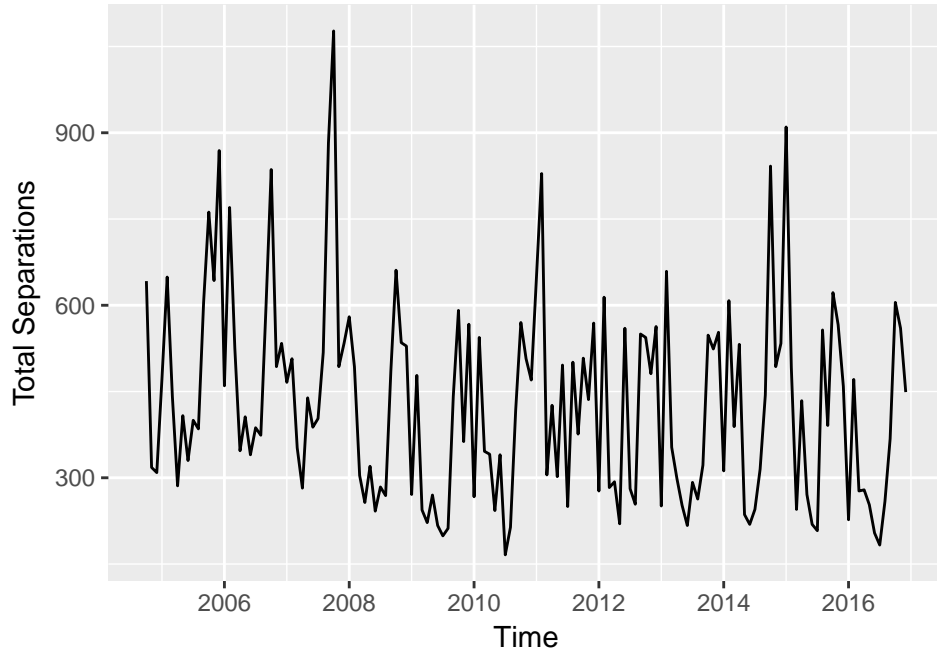
Table 2. Seasonal Naive Results

	ME	RMSE	MAE	MPE	MAPE	MASE
Training set	16.49565	291.7907	155.4522	-7.373817	30.45221	1.000000
Test set	-238.85000	454.2881	271.4500	-59.852083	67.31547	1.746196

identified. To help, refer back to Figure 3. We note above that the spikes generally occur in November and December. However, the observations from 2005 are close enough to those from other years that they may have resulted naturally. We do not want to remove too much information from the data set, only that which is misleading. So, November and December observations from 2006, '07, and '14 are selected for replacement.

Given the seasonality in the data set, the replaced values should stem from matching observations in previous years, as opposed to previous observations within the same year. The outliers are replaced with the arithmetic mean of all years not being replaced (e.g. November 2006, '07, '14 is replaced with the mean separations in November for all other years).

Replotting the response in Figure 5 shows a much better behaved data set. The data look fairly stationary, setting the stage for more forecasting complex models.

**Figure 5.** Separations - Outliers Removed

With the outliers replaced, seasonal effects are much more apparent as well (see Figure 6 below), further enforcing the need for a seasonal model.

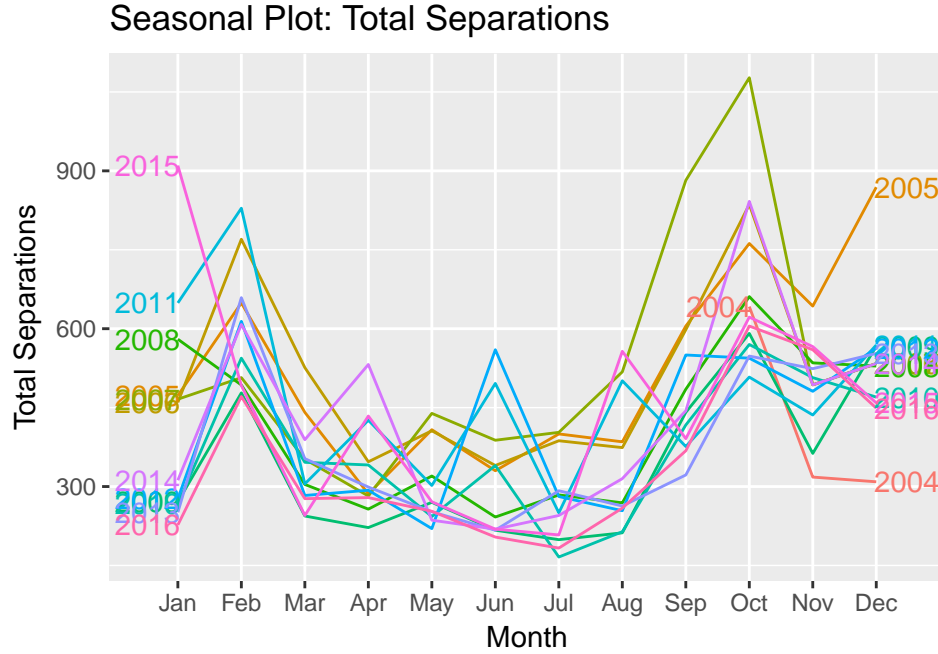


Figure 6. Seasonal Plot: Outliers Removed

The purpose of imputing those values was to help produce better models. That can be assessed by retraining the naive models and comparing against previous results.

Table 3 below compares the seasonal naive RMSEs before and after imputating the identified outliers. The results indicate that removing and replacing the extreme values for November and December improved the model’s ability to capture information. This is further reflected by the forecast (shown in blue) in Figure 7, which follows the validation data (shown in orange) more closely than those in Figure 4. Overall, these results imply that imputation of the selected observations is a worthwhile endeavor.

Table 3. Seasonal Naive RMSE Comparison

	Raw Data	Imputed Data
Training	291.7907	161.2624
Validation	454.2881	186.5842

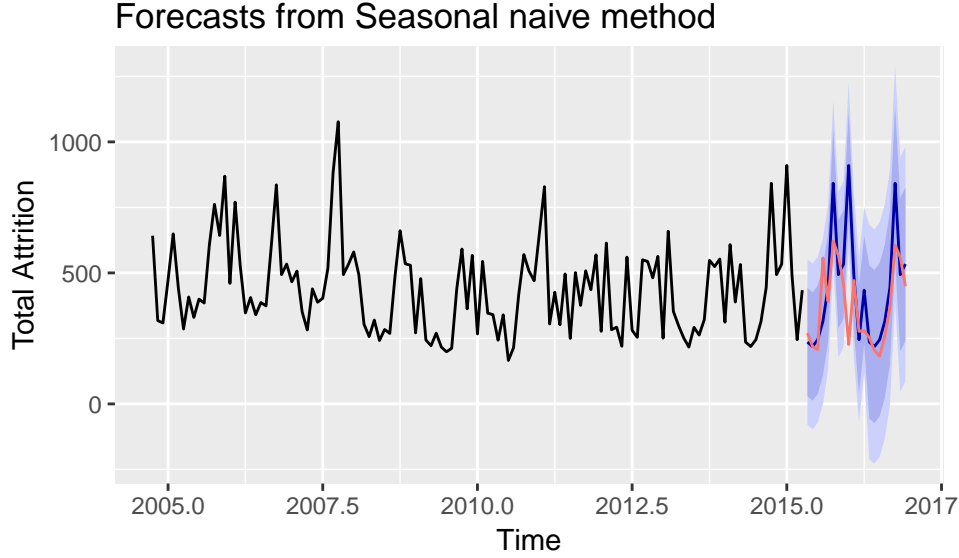


Figure 7. Seasonal Naive Forecast After Imputation

Though an important step in the modeling process, these naive models are not the end goal. Rather, they help identify initial road bumps in the modeling process and create a baseline against which to compare other, more sophisticated models.

3.2.2 *Dynamic Regression*

One major shortfall of the naive model is the inability to include exogenous information, and this is the case for many forecasting methods (ARIMA, Decomposition, Exponential Smoothing, etc). However, these methods do allow the inclusion of autoregressive information, which normal regression techniques do not.

The solution, then, is to combine efforts with dynamic regression. Dynamic regression is just a multivariate regression model with an ARIMA model on the errors. The regression piece allows independent variables to be used in predicting a response, and the ARIMA portion helps capture and account for autoregressive information which often exists in time-series data (thereby helping preserve the assumption of independent errors). The general formulation of a dynamic regression model with ARIMA(1,1,1) errors, for example, errors is:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + n_t$$

where,

$$(1 - \phi_1 B)(1 - B)n_t = (1 + \theta_1 B)e_t$$

and e_t is white noise.

Before fitting the dynamic regression model, several steps must be taken to ensure other key assumptions are not violated. First is the issue of collinearity. Dependence between any subset of the regressors will primarily lead to innaccurate coefficient estimates, a result contrary to the goal of this endeavor. To help avoid this pitfall, a correlation heat map of possible regressors is built. Figure 8 shows the correlation between all pair-wise combinations of variables considered in the economic data set. There are many instances of collinearity, which is expected - many economic indicators are variations or flavors of the same information. There are some independent subsets, though, and these will be the candidates for the dynamic regression model.

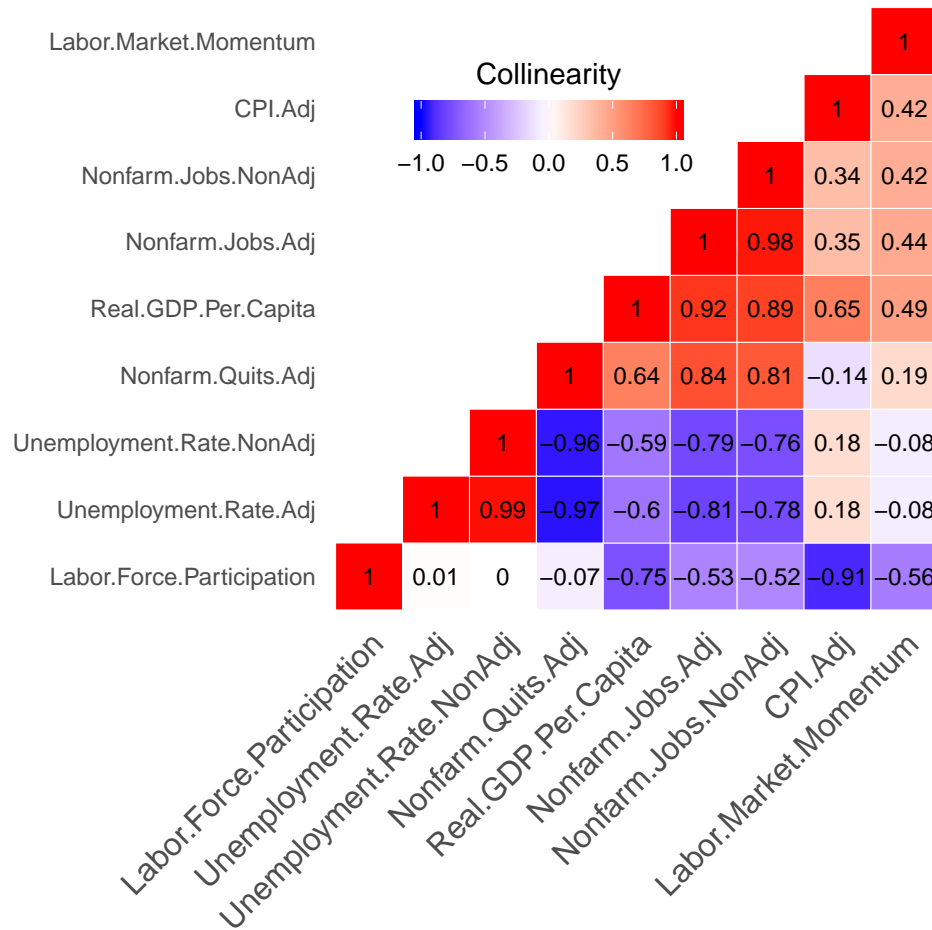


Figure 8. Correlation Matrix - Economic Indicators

For the first stab at modeling, the Unemployment Rate (Adj.), Labor Force Participation Rate, and Market Momentum Index are selected. At this point, the stationarity of the regressors must be ensured. Non-stationary regressors also produce inconsistent coefficient estimates - even if they are independent. To assess, we plot the three indicators below in Figure ??.

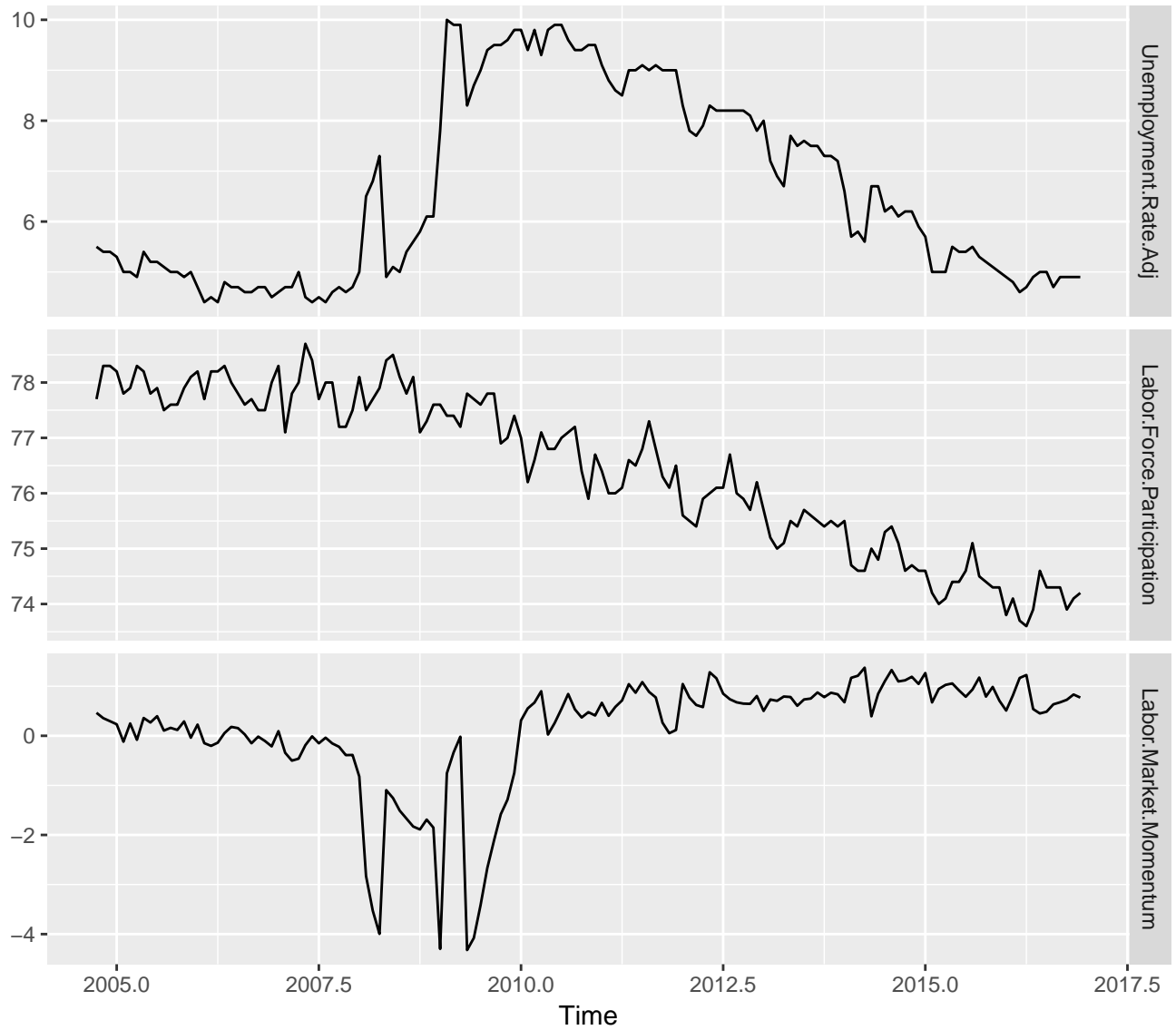


Figure 9. Economic Indicators - Raw

Each of the three indicators show evidence of a trend (i.e. are non-stationary), and to handle this the data are differenced. The resulting data are shown below in Figure 10.

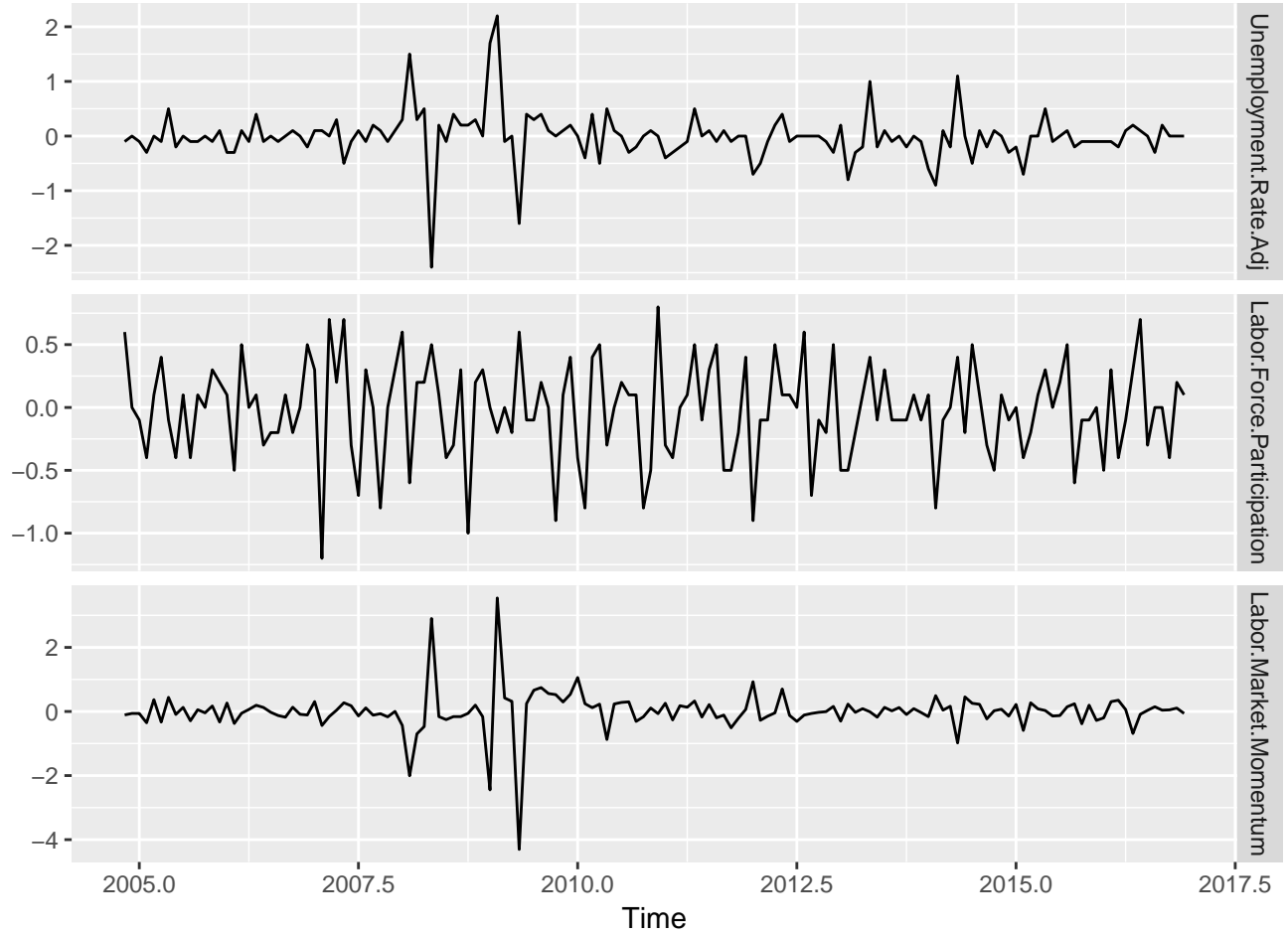


Figure 10. Economic Indicators - Differenced

Simple differencing appears to produce the desired effects; there is no evident trend or seasonality in any of the indicators. It is important to note the regressors now show the month-to-month change, which will be important when interpreting results.

With stationary and independent economic indicators, we can move on from regression to the ARIMA piece of the model. Up to six parameters must be specified: the order of autoregression, degree of differencing, and order of the moving average (p , d , and q , respectively) and their seasonal counterparts (P , D , and Q). For the sake of time, these are not manually found; instead, a brute-force algorithm is used. A range is specified for each parameter, and a model is fit for every combination within the specified ranges; the model with the lowest corrected Akaike Information Criteria (AICc) is selected. For the first, and all subsequent, dynamic regression models in the thesis, the following ranges/values were used: $p, q \in [0, 5]$,

$d, D = 0$, and $P, Q \in [0, 2]$.

For this first pass, the model selected was a regression model with a fourth-order moving average and first-order seasonal autoregression on the errors. More explicitly:

$$y = \beta_0 + \beta_1 x'_{1,t} + \beta_2 x'_{2,t} + \beta_3 x'_{3,t} + n_t$$

where,

$$(1 - \Phi_1 B^{12})n_t = (1 + \theta_4 B)^4 e_t$$

and,

$$x'_{i,t} = x_{i,t} - x_{i,t-1}$$

The coefficient estimates are below in Table 4:

One of the primary sources of model assessment is analysis of the residuals. When inspecting the residuals, we are looking for evidence of leftover autocorrelation, satisfaction of normality assumptions ($e_t \sim N(0, \sigma^2)$), and outlier effects. Figure 11 below combines a few plots to help answer those questions. The top subfigure plots the raw model residuals, and is used to identify possible trends, seasonality, or heteroscedasticity. Fortunately, none of those features are apparent. The bottom-left searches for significant autocorrelation in the residuals. Raw residuals are correlated with themselves in increased lags, and any significant correlations would indicate a possible violation of the independence of the residuals. The current model's results only show one lag-period with significant autocorrelation, which may mean that there is information unaccounted for by the model. Overall autocorrelation, however, appears to be insignificant, as further evidenced by the results of a Ljung-Box test for autocorrelation (Table 5). The bottom-right plot shows a histogram of the residuals,

Table 4. Estimated Coefficients - Initial Model

	θ_1	θ_2	θ_3	θ_4	Φ_1	β_0	β_1	β_2	β_3
Coeff	0.218	0.145	0.336	0.260	0.576	429.875	-13.390	-15.949	-2.819
StdErr	0.087	0.088	0.092	0.092	0.082	47.602	22.286	35.016	11.494

comparing the raw distribution against the ideal normal. The plot shows slight skewness, but overall the data appear normal. In short, there is little - if any - misbehavior in the model's residuals.

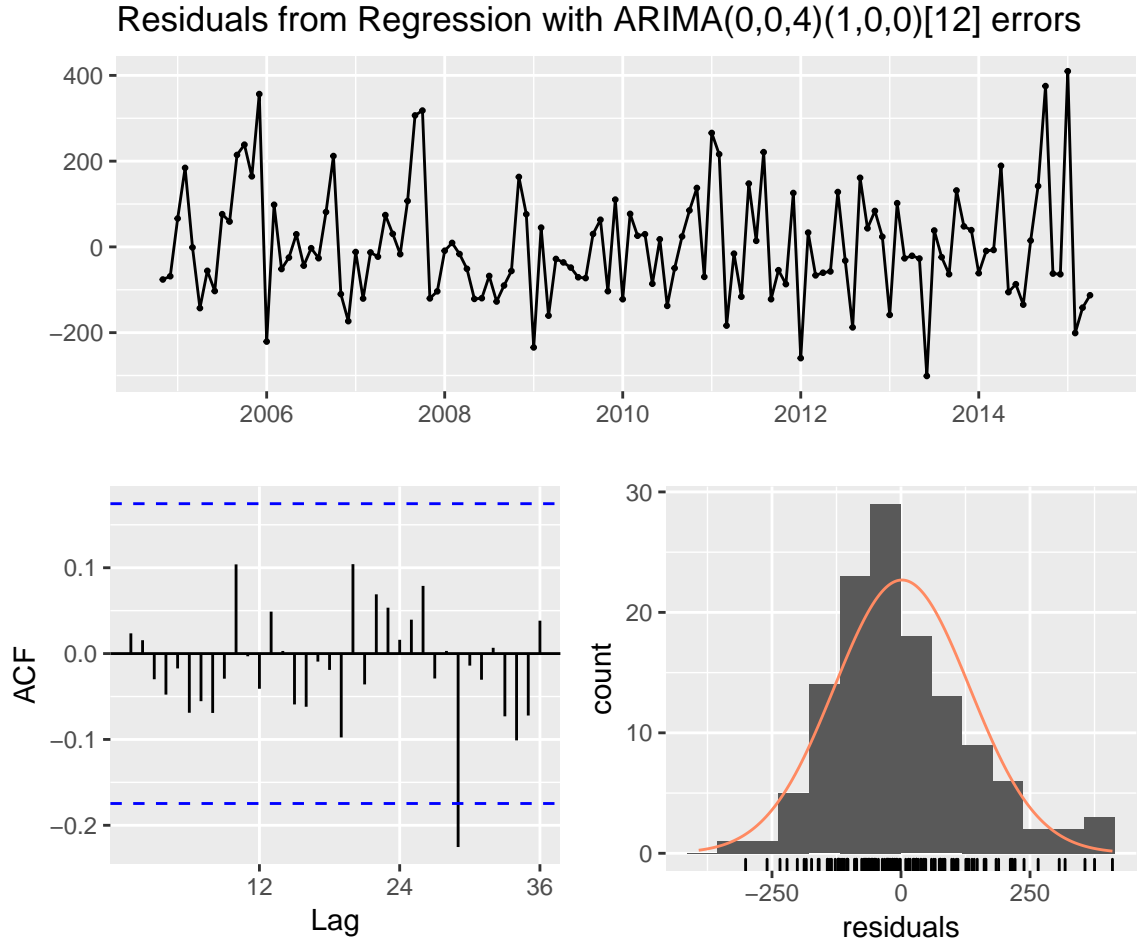


Figure 11. Initial Attrition Model - Residual Analysis

Next, forecasts are generated from the training data and compared against the validation data. Figure 12 shows a model that looks to track attrition in much greater detail than the naive models, which is expected. Large movements are generally captured, even if not perfectly forecast. To get a sense of modeling progress, the RMSEs are from our models are

Table 5. Initial Model - Autocorrelation Test

Test type	Test statistic	p-value
Box-Ljung test	10.121	0.812

compared in Table 6.

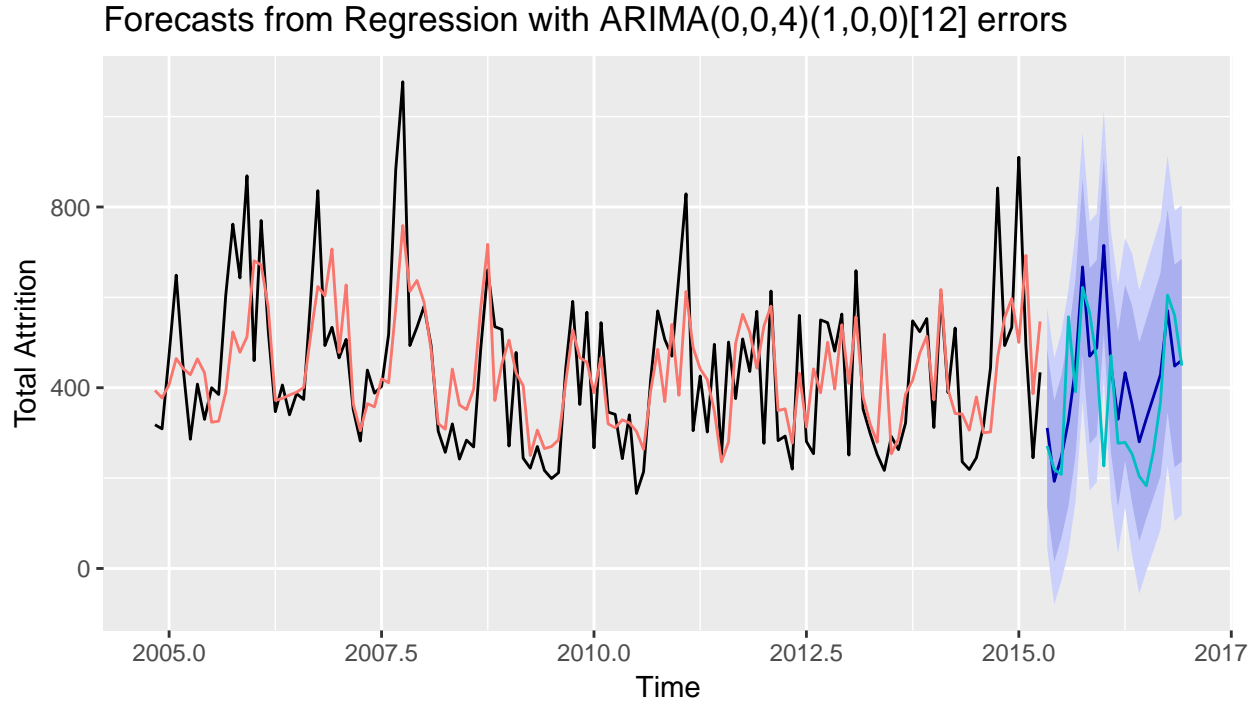


Figure 12. Initial Model - Forecasts Against Validation Data

Dynamic regression, thankfully, demonstrates greater ability to forecast the attrition data than the naive models. However, all is not well. The high standard errors of our regression coefficients (Table 4) indicate that none of the economic indicators are statistically significant, meaning the ARIMA model is handling all the forecasting and the regression is providing little insight. Essentially, no information is gained from the economic indicators as they have been specified. This could be for several reasons:

- Having been differenced, the indicators represent month-to-month changes. For most observations in the data set those changes are marginal, resulting in an insignificant effect on attrition - numerically.

Table 6. Model RMSE Comparison

	Simple Naive	Seasonal Naive	Dynamic Regression
Training	199.8321	161.2624	130.7455
Validation	160.6421	186.5842	142.9881

- As they are, the indicators show only the previous month's change. That is, the regression coefficients represent the effects last month's changes have on this month's attrition. Intuitively, this does not seem correct. Voluntary separation from the military is a long, bureaucratic process; as such, one would think the time-line is well beyond a month.
- The economic and personnel data are both aggregated to the national level. It is possible that such a degree of aggregation includes so much noise that any economic effects are being masked.

Unfortunately, there is not much that can be done about the first point. As mentioned earlier, the indicators must be stationary in order to ensure the reliability of any potential effects, and the data must be differenced to be stationary. The rest of the thesis, then, focuses on the last two points. It is simple enough to introduce lagged regressors, and that process is introduced next. As for aggregation, the economic data cannot be broken down, but the personnel data can (to an extent). Currently, the personnel data encompasses all AFSCs. It may be that different subsets of the Air Force officer corps are more sensitive to economic conditions than others, a possibility investigated in subsequent analysis.

3.2.3 Lagged Indicators

3.2.4 Officer Subsetting

IV. Conclusions

V. TO-DO

- Fix some of the Table references
- Add complete set of lagged variable analysis
- Introduce subsetting of officer population (total vs rated-only vs non-rated line vs etc)
- Adjust legends and coloring in some figures
- Conclusion (<- wait until lagged effects are fully explored before drawing insights)
- Introduction (<- wait until end? not sure how to introduce)

Bibliography

- [1] All Employees: Total Nonfarm Payrolls, Dec 2017. URL <https://fred.stlouisfed.org/series/PAYEMS>.
- [2] B. Asch and J. R. Hosek. Looking to the Future: What Does Transformation Mean for Military Manpower and Personnel Policy. Technical report, DTIC Document, 2004.
- [3] B. J. Asch. Designing Military Pay: Contributions and Implications of the Economics Literature. Technical report, RAND Corporation, 1993.
- [4] S. P. Barrows. Air Force Pilot Retention: An Economic Analysis. Technical report, DTIC Document, 1993.
- [5] J. Capon, O. S. Chernyshenko, and S. Stark. Applicability of Civilian Retention Theory in the New Zealand Military. *New Zealand Journal of Psychology*, 36(1):50, 2007.
- [6] T. Demirel. A Statistical Analysis of Officer Retention in the U. S. Military. Master’s thesis, Naval Postgraduate School, 2002. URL <https://pdfs.semanticscholar.org/d117/d7abcf4ea6ac90ba8a206e80cb9515ce235.pdf>.
- [7] C. N. Franzen. Survival Analysis of US Air Force Rated Officer Retention. Master’s thesis, Air Force Institute of Technology, 2017.
- [8] S. S. Fugita and H. A. Lakhani. The Economic and Noneconomic Determinants of Retention in the Reserve/Guard Units. Technical report, DTIC Document, 1991.
- [9] S. I. Gass. Military Manpower Planning Models. *Computers & Operations Research*, 18(1):65 – 73, 1991. ISSN 0305-0548. doi: [https://doi.org/10.1016/0305-0548\(91\)90043-Q](https://doi.org/10.1016/0305-0548(91)90043-Q). URL <http://www.sciencedirect.com/science/article/pii/030505489190043Q>.
- [10] G. D. Gjurich. *A Predictive Model of Surface Warfare Officer Retention: Factors Affecting Turnover*. PhD thesis, Monterey, California. Naval Postgraduate School, 1999.
- [11] G. R. Grimes. The Effects of Economic Conditions on Overall Air Force Officer Attrition. Technical report, DTIC Document, 1987.
- [12] H. L. Jantscher. An Examination of Economic Metrics as Indicators of Air Force Retention. Master’s thesis, Air Force Institute of Technology, 2016.
- [13] T. Kane. *Bleeding Talent: How the US Military Mismanages Great Leaders and Why It’s Time for a Revolution*. Palgrave Macmillan, 2012.

- [14] S. Ramlall. A Review of Employee Motivation Theories and their Implications for Employee Retention within Organizations. *Journal of American Academy of Business*, 5(1/2):52–63, 2004.
- [15] T. R. Saving, B. M. Stone, L. T. Looper, and J. N. Taylor. Retention of Air Force Enlisted Personnel: An Empirical Examination. Technical report, DTIC Document, 1985.
- [16] J. A. Schofield. Non-Rated Air Force Line Officer Attrition Rates Using Survival Analysis. Master’s thesis, Air Force Institute of Technology, 2015.

Vita

The vita.

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>						
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE			3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE				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 NAMES(S) AND ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR ACRONYM(S)		
				11. SPONSOR/MONITOR REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT						
15. SUBJECT TERMS						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. PHONE NUMBER (Include area code)	