

Forecasting Retirements: The Case of the New York City Government Workforce

Yevgeniy Davydov^{1,2}

¹Columbia University
yd2398@columbia.edu

²Department of Citywide Administrative Services
ydavydov@dcas.nyc.gov

ABSTRACT

By 2020, more than 25% of the United States workforce will be aged 55 or older [22]. With an aging workforce that is at or near retirement, there is a need for workforce planning strategies to ensure continued delivery of critical services. This study uses retirement data from the New York City (NYC) government workforce to develop a retirement forecasting application. The application utilizes automatic time-series forecasts to give retirement projections across several retirement categories such as employee type, agency, and job category. A complimentary model validation tool is also developed for gauging out-of-sample model accuracy.

Keywords: workforce planning, retirement, forecasting, ARIMA, curse of dimensionality, data visualization

1. INTRODUCTION

At the close of Fiscal Year 2016, the City of New York employed 383,704 people in professions including teachers, police officers and firefighters; analysts, engineers and inspectors of every description; caseworkers and nurses; administrative and clerical support staff; park workers and road repairers; as well as many other professions [15]. As of Fiscal Year 2016, the median age of a New York City (NYC) government employee was 44, 17% of the NYC workforce was eligible to retire, and 30% of the NYC workforce would be eligible to retire within 5 years [15]. In the face of an aging workforce, NYC government needs workforce planning strategies to ensure delivery of critical services by City agencies to the people of the City of New York.

The goal of this paper is to use data science approaches to provide NYC government agencies with a workforce planning tool. The paper utilizes machine-learning to select the best forecasting model and to estimate retirements for several years ahead of the current year. In particular, forecasts are generated via time-series models that are selected automatically based on statistical criteria. The R Shiny application [5] is used to build a retirement forecasting tool. The tool allows non-technical users such as hiring managers and human resources professionals to visualize retirement forecasts as interactive charts.

2. DATA

The data was taken from the Payroll Management System (PMS), which is the City of New York’s automated payroll and timekeeping processing and record keeping system. The data spans thirty years (FY 1988 - FY 2017). The data was cleaned to make sure that there were no duplicates by employee number. In addition, only the sample of full-time employees was selected to reduce noise in the retirements data.

The main variable of interest that is used in this study is the retirement to headcount ratio. The retirement to headcount ratio (r) is constructed by dividing the number of employees that retired in a specific category by the total number of employees in that category for each fiscal year. There are 5 variables in the employee type category (e.g., Manager, Civilian), 84 City agencies, and 29 job categories (e.g., Administrators, Social Scientists). It is also useful for agencies to look at retirements at their agency for a specific job category at their agency. Therefore, agency-job category

combinations are also constructed (e.g., DCAS – Clerical Supervisors). Since every agency has unique needs, not every job category matches jobs at a specific agency. This yields a total of 1,035 agency-job category combinations.

Considering all the categories we get a data set, D , with 1,153 unique univariate retirement time-series for each category i , $i = (1, \dots, 1153)$ and each year, t , $t = (1988, \dots, 2017)$:

$$D = \begin{pmatrix} r_{1,1988} & r_{2,1988} & \dots & r_{1135,1988} \\ r_{1,1999} & r_{2,1999} & \dots & r_{1135,1999} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1,2017} & r_{2,2017} & \dots & r_{1135,2017} \end{pmatrix}, \quad (1)$$

where $r_{i,t}$ is the retirement to headcount ratio for category i at time t . Structuring the data in this way provides a natural setting to analyze the data using time-series methods.

3. METHODOLOGY

3.1 Model Selection

The first step to building a retirement forecasting tool is to select an appropriate time-series model for each retirement time-series. The Auto-regressive Integrated Moving Average (ARIMA) class of models developed by Box and Jenkins [3] is chosen for this tool for several reasons. This class of time-series models is standard and has been applied in diverse fields to forecast unemployment rates, stock prices, energy demand, and server workload [14, 1, 6, 4]. These models are also intuitive and the meaning of the parameter estimates can easily be explained.

Formally, the standard ARIMA(p, d, q) model can be written for each retirement time-series as follows:

$$\phi_i(B)(1 - B)^d r_{i,t} = c_i + \theta_i(B)\varepsilon_{i,t} \quad (2)$$

where $\phi_i(B)$ and $\theta_i(B)$ are polynomials of auto-regressive and moving average components respectively:

$$\phi_i(B) = (1 - \phi_{i,1}B - \phi_{i,2}B^2 - \dots - \phi_{i,p}B^p), \quad (3)$$

$$\theta_i(B) = (1 + \theta_{i,1}B + \theta_{i,2}B^2 + \dots + \theta_{i,q}B^q), \quad (4)$$

where p is the order of the auto-regressive lag, q is the order of moving average lag, d is the differencing parameter (if $d \geq 1$ the series is not stationary), B is the backshift operator where $B^k r_{i,t} = r_{i,t-k}$, $\varepsilon_{i,t} \stackrel{i.i.d.}{\sim} N(0, \sigma_i^2)$ are identically independently distributed innovations following a Gaussian White Noise process with mean 0 and variance σ_i^2 , and c_i is the constant parameter for each time series category i . In addition, the models must satisfy the stationarity and invertibility conditions such that the roots of the polynomials $\phi_i(B)$ and $\theta_i(B)$ are outside of the unit circle.

One of the main challenges of ARIMA modeling is that the selection of the orders (p, d, q) is subjective and difficult to apply [9]. Various measures such as the Akaike Information Criteria (AIC), the corrected AIC (AICc) and the Schwarz Bayesian Information Criteria (known as SBIC, SIC, or BIC) have been developed over the years to help with model selection [2, 21, 18]. These information criteria are all penalized likelihood approaches that can be used to compare models within the same class. Model selection based on AIC, AICc, or BIC would result in a parsimonious model as these criteria penalize a model for having too many parameters. In

the case of ARIMA(p, d, q) models the AIC, AICc, and BIC can be written as functions of their parameters as in [8]:

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1) \quad (5)$$

$$\text{AICc} = \text{AIC} + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2} \quad (6)$$

$$\text{BIC} = \text{AIC} + [\log(T) - 2](p + q + k + 1), \quad (7)$$

where L is the maximized likelihood of the model fitted to the data, $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$, 1 is added to account for the variance of the residuals, σ^2 , and T denotes the length of the time-series. The objective is to choose the model with the minimum information criteria (i.e., minimize AIC).

3.2 The Curse of Dimensionality

While the information criteria allow us to compare models that were already estimated, it is often infeasible or impractical to estimate all possible ARIMA models in the model space. In the case of retirement prediction, this problem is due to the curse of dimensionality coming from two sources: 1) the model space, and 2) the retirement category space. A non-seasonal ARIMA(p, d, q) model can theoretically have state space of $p \times d \times q$ possible models. Even though in most practical applications $\{p, d, q\} \in [0, 2]$, there is still a total of 27 possible models with this restriction. Combining the restriction on the model state space with 1,153 retirement time-series we get 31,131 possible models. If seasonal components are considered as well, and if the model state space increases then dimensionality problem gets exacerbated very quickly.

3.2.1 Automatic ARIMA Forecasting

As we have seen due to the curse of dimensionality, the model state space can be difficult to traverse via manual selection. This creates a need for automatic model selection. Hyndman and Khandakar [9] proposed a step-wise algorithm to traverse the model space. Essentially the algorithm places some initial restrictions on the model space, then traverses the model space by increasing or decreasing the orders $\{p, d, q\}$ ¹ until the model with the minimal information criteria is found. The full description of the algorithm can be found in [9].

The `forecast` package [7] in R with the `auto.arima(.)` function is used for practical implementation of the Hyndman and Khandakar algorithm [9]. The `auto.arima(.)` function begins by utilizing the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests [11] for unit roots to select the order of the differencing component d . This ensures that the final model will be stationary. Then, the orders of p and q are chosen by minimizing the AICc and traversing the model space by a step-wise procedure as in [9, 8]. There are also options to select the Augmented Dickey-Fuller (ADF) [17] and the Phillips-Perron (PP) [16] unit root tests instead of the KPSS test. Alternative information criteria which include the AIC and BIC can be selected as well instead of the AICc. In most cases choosing a different unit root test or different information criteria should not alter the final model, but in some cases due to the strengths or weaknesses of each method the final model may be different.

¹The algorithm also considers orders $\{P, D, Q\}$ to check for seasonality.

3.2.2 The Retirement Category Space

While automatic ARIMA forecasting takes care of the dimensionality problem in the model space, we are still left with the dimensionality problem in the retirement category space. With 1,135 retirement categories it would be difficult and time-consuming to evaluate and report the results of each retirement time-series model with traditional approaches. The solution to this dimensionality problem in the retirement category space is to use data visualization techniques. The idea is for the user select a retirement category they are interested in, and the plot of the retirement forecasts and the original data for a particular retirement category will be shown to them.

The R Shiny application is chosen for data visualization. This tool has been used in various applications such as a web tool for the generation of box-plots [20] as well as for visualizing anomaly detection in bioinformatics data [13]. Using R Shiny, a dashboard is created where users are able to view retirement forecasts as interactive charts. Users are given options to select retirement forecasts by employee type, City agency, job category, and agency-job category combination. When the user selects a retirement category, the `auto.arima(.)` function runs “under the hood” and selects the best possible ARIMA model for that retirement time-series. Then, the application utilizes the `ggplot2` package [23] and the `plotly` package [19] to output the forecast results as interactive charts.

4. RETIREMENT FORECASTING APPLICATION

The retirement forecasting tool is shown in Figure 1. The tool gives retirement projections for five fiscal years ahead of the current year. Since the data ends in FY 2017, forecasts go up to FY 2022. The tool allows the user to view retirement forecasts as four separate charts for employee type, City agency, job category, and agency-job category combination.

This tool can assist government agencies in implementing successful workforce planning strategies ahead of projected retirements. Agencies can also compare and benchmark their retirements to citywide retirements by looking at forecasts for their agency compared to all city employees.

4.1 Case Study: Agency Level Retirements

A representative agency is selected to show a more in-depth analysis of the forecasting results. To illustrate the use of the retirement tool at the agency level, an agency was chosen as a representative retirement time-series if it had no missing values and it has been in existence throughout the entire sample period.

Forecasts for the representative agency are shown in Figure 2. The retirements for this agency are on average expected to hover around 4% over the next five fiscal years. For FY 2021, the projected retirements at this agency would be 4.09% with a 90% prediction interval of [1.10%, 7.09%]. As expected of forecasting models, the forecast variance gets wider as the forecast horizon increases. This means that forecasts are less accurate as the forecast horizon increases.

Retirement Forecasting Tool



Overview

Reliable retirement projections are an essential component for crafting and implementing successful workforce planning strategies. Utilizing thirty years of retirement data (1988-2017), this tool provides City agencies with five-year retirement projections (i.e., fiscal years 2018-2022) as a percent of total headcount by *Employee Type, Agency, and Job Category*.

Projections are based on an algorithm that learns from the historical data and chooses the best statistical model to provide the most reliable forecast for each selection. To account for variability in retirements over time, the range of possible values are highlighted by a yellow band. The charts below are interactive and include features such as hovering to see the data points.

Projected Retirements as Percentage of Headcount

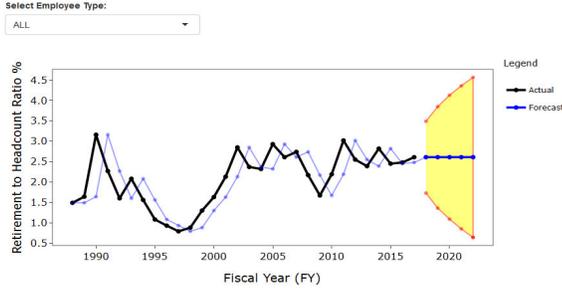


Figure 1: General view of the retirement forecasting tool. Actual retirements are shown on the black line, the model fit is shown on the light blue line, and the forecast is shown on the dark blue line. The 90% prediction interval is given in the yellow region with the red boundary being the lower and upper 90% prediction values.

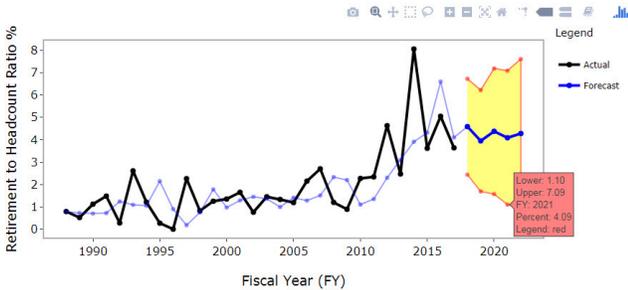


Figure 2: Retirement projections for a representative agency. Actual retirements are shown on the black line, the model fit is shown on the light blue line, and the forecast is shown on the dark blue line. The 90% prediction interval is given in the yellow line. The hover text for the red boundary shows the lower 90% predicted value (Lower), the upper 90% predicted value (Upper), the fiscal year (FY), the retirement to headcount ratio (Percent), and which line is selected (Legend).

The forecast in Figure 2 was generated by the `auto.arima()` function choosing the ARIMA(1,1,0) model as the best ARIMA model for the Department for the Aging retirement time-series using the AICc information criteria. The model parameters are shown in Table 1. The estimated AR(1) coefficient ϕ_1 is -0.67 and statistically significant. The stationarity condition of the roots of the AR polynomial $\phi(B)$ being outside the unit circle lead to a condition of $|\phi_1| < 1$ for an AR(1) model. Since the absolute value of the AR coefficient is $|\phi_1| = 0.67 < 1$, the inverse root of the AR polynomial is inside the unit circle, which means that the model satisfies the stationarity condition. The Ljung-Box test [12] yielded a p -value of 0.53 which means that we cannot reject the null that the residuals follow a Gaussian White Noise process. The innovations ε_t are normally distributed as $\varepsilon_t \sim N(0, 1.70)$, which means that random shocks on the order of $\sigma = \sqrt{\sigma^2} = \sqrt{1.70} = 1.35\%$ standard deviations would be affecting changes in retirements.

Table 1: Representative Agency Model

Model: ARIMA (1,1,0)	
	Coefficients
ϕ_1	-0.67^{***} (0.13)
σ^2	1.70
AIC	101.31
AICc	101.77
BIC	104.04
Ljung-Box Test p -value	0.53

The standard errors are given in parenthesis.

*** denotes significance at the 1% level.

4.2 Model Validation

Now that a forecasting tool was constructed it is also useful to have a model validation tool to evaluate the accuracy of the forecasts. The first step is to split the time-series data into a training set and a test set. Taking the retirement data, D , described in equation (1) we get the training and test data sets by splitting D into D_{train} and D_{test} as follows:

$$D_{train} = \begin{pmatrix} r_{1,1988} & r_{2,1988} & \dots & r_{1135,1988} \\ r_{1,1999} & r_{2,1999} & \dots & r_{1135,1999} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1,2012} & r_{2,2012} & \dots & r_{1135,2012} \end{pmatrix} \quad (8)$$

and

$$D_{test} = \begin{pmatrix} r_{1,2013} & r_{2,2013} & \dots & r_{1135,2013} \\ r_{1,2014} & r_{2,2014} & \dots & r_{1135,2014} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1,2017} & r_{2,2017} & \dots & r_{1135,2017} \end{pmatrix}, \quad (9)$$

where D_{train} contains all retirement time-series from FY 1988 to FY 2012, and D_{test} contains all retirement time-series from FY 2013 to FY 2017. The ARIMA model in equation (2) would then be estimated on D_{train} . This allows us to test out-of-sample model accuracy for 5 fiscal years.

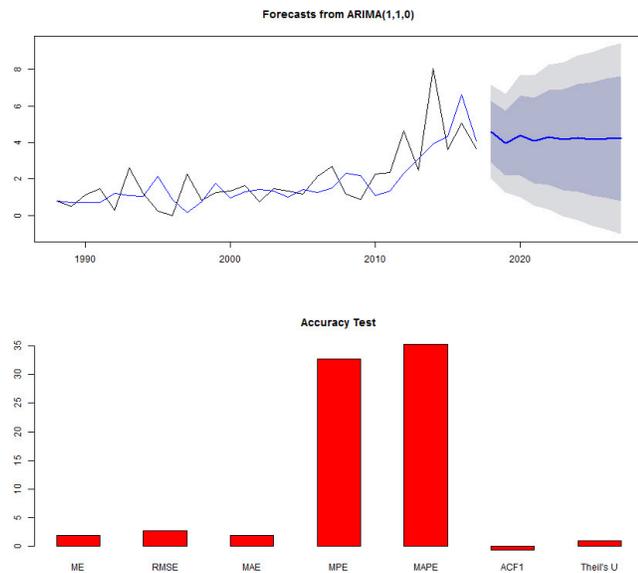


Figure 3: These charts provide a model validation dashboard. Retirement projections for the representative agency in Figure 2 are shown. Actual retirements are shown on the black line, the model fit is shown on the light blue line, and the forecast is shown on the dark blue line. The 80% and 95% prediction intervals are given in the shaded regions. The accuracy test shows the forecast accuracy based on the ME (Mean Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MPE (Mean Percentage Error), MAPE (Mean Absolute Percentage Error), ACF1 is first order auto-correlation of the residuals, and Theil's U is Theil's U statistic where lower U implies better forecast accuracy.

The model validation tool is built the same way as the retirement forecasting tool. This tool is intended for technical users who are familiar with forecast accuracy measures and it is not included as part of the main retirement forecasting tool. The model validation tool is shown in Figure 3. The

retirement forecasts are still shown as a chart, however this time the model specification is shown.

In general, the retirement forecast error can be defined for each retirement category as:

$$e_{i,t} = r_{i,t} - \widehat{r}_{i,t}, \quad (10)$$

where $e_{i,t}$ is the difference between actual retirements $r_{i,t}$ at time t , and forecasted retirements $\widehat{r}_{i,t}$. The accuracy test chart shows the forecast accuracy based on the ME (Mean Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), MPE (Mean Percentage Error), MAPE (Mean Absolute Percentage Error), ACF1 is first order auto-correlation of the residuals, and Theil's U. Multiple accuracy measures are provided since standard measures such as the mean error and root mean squared error are scale based measures and are sensitive to outliers [10]. This allows the analysts and technical users to evaluate the out-of-sample forecasting power of the models for a specified forecast horizon h , where $h = 1, \dots, n$.

5. CONCLUSION

In the face of an aging workforce, it is a challenge to create successful workforce planning strategies ahead of retirements. Knowing how many employees will retire ahead of time would help organizations hire and train enough new employees to replenish the outgoing workforce. This would ensure the continuation of critical services in the organization and in broader society. This study provides a retirement forecasting tool which can help HR professionals, policy makers, and other key stakeholders in tackling this longstanding workforce planning problem.

The retirement forecasting tool provides retirement forecasts for five fiscal years ahead of the current year by using the ARIMA class of time-series models. The model selection is automated using information criteria such as the AICc and the tool allows users to see forecasts for various retirement categories as interactive charts. A model validation tool is also provided to help practitioners and technical users to evaluate the out-of-sample forecasting power of the selected model for each retirement category. Future work that could improve retirement forecast accuracy would involve testing the ARIMA class of models against other time-series models such as Exponential Smoothing (ETS) and structural time-series models such as the Basic Structural Model (BSM).

6. ACKNOWLEDGEMENTS

I would like to thank Benjamin Packer for assistance in retrieving and cleaning the data. The views reflected here do not reflect the views of the Department of Citywide Administrative Services (DCAS). All errors are my own.

References

- [1] A. A. Adebisi, A. O. Adewumi, and C. K. Ayo. Comparison of ARIMA and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics*, 2014, 2014.
- [2] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [3] G. E. Box and G. M. Jenkins. *Time series Analysis: Forecasting and Control*. Holden-Day, 1970.
- [4] R. N. Calheiros, E. Masoumi, R. Ranjan, and R. Buyya. Workload Prediction Using ARIMA Model and Its Impact on Cloud Applications' QoS. *IEEE Transactions on Cloud Computing*, 3(4):449–458, 2015.
- [5] W. Chang, J. Cheng, J. Allaire, Y. Xie, and J. McPherson. Shiny: web application framework for R. R package version 0.11. <https://cran.r-project.org/web/packages/shiny/shiny.pdf>, 2015.
- [6] V. Ş. Ediger and S. Akar. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, 35(3):1701–1708, 2007.
- [7] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen. *forecast: Forecasting functions for time series and linear models*, 2018. R package version 8.4.
- [8] R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2 edition, 2018.
- [9] R. J. Hyndman and Y. Khandakar. Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, 26(3), 2008.
- [10] R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.
- [11] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159–178, 1992.
- [12] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.
- [13] G. Monaco, H. Chen, M. Poidinger, J. Chen, J. P. de Magalhães, and A. Larbi. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics*, 32(16):2473–2480, 2016.
- [14] A. L. Montgomery, V. Zarnowitz, R. S. Tsay, and G. C. Tiao. Forecasting the U.S. Unemployment Rate. *Journal of the American Statistical Association*, 93(442):478–493, 1998.
- [15] New York City Department of Citywide Administrative Services (DCAS). *FY 2016 New York City Government Workforce Profile Report*. 2017.
- [16] P. C. Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.
- [17] S. E. Said and D. A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.
- [18] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.
- [19] C. Sievert, C. Parmer, T. Hocking, S. Chamberlain, K. Ram, M. Corvellec, and P. Despouy. *plotly: Create Interactive Web Graphics via 'plotly.js'*, 2017. R package version 4.7.1.
- [20] M. Spitzer, J. Wildenhain, J. Rappsilber, and M. Tyers. Boxplotr: a web tool for generation of box plots. *Nature Methods*, 11:121–122, 2014.
- [21] N. Sugiura. Further Analysis of the Data by Akaike's Information Criterion and the Finite Corrections. *Communications in Statistics-Theory and Methods*, 7(1):13–26, 1978.
- [22] M. Toossi. Labor force projections to 2020: a more slowly growing workforce. *Monthly Labor Review*, 135(1):43–65, 2012.
- [23] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.