Forecasting Demand for Traveling Nurses

American Traveler PIC Math Project

Navika Timsina and Daniel Walford

April 26, 2024

Problem

The nursing industry is one of the most important industries in healthcare in the United States of America. Nursing is the nation's largest healthcare profession, with nearly 4.7 million registered nurses (RNs) nationwide. One of the growing field within the nursing industry is travel nurses. In 2019, only about 2% of all registered nurses were travel nurses. By the fall of 2021, travel nurse openings increased by nearly 500% compared to January 2020. There are over 1,733,502 traveling nurses currently employed in the United States. For our project, we are designing a statistical model to predict how the demand for nurses will change over time in the next five years. Our main mathematical and computational objectives we need to tackle are the supply and demand for nurses and the factors and data influencing this supply and demand.

Approach

We collected our data from government/reliable databases: The Bureau of Labor Statistics (BLS), US Census, and The State Health Access Data Assistance Center (SHADAC). We input this data into R, which is our preferred coding language. We used packages like `dplyr()` and `tidyr()` to filter and format the data. It was a challenge normalizing the data from three different sources into a single dataset, and this was a significant part of the beginning of our project.

We collected data at both state and national levels. We gathered the population and the number of employed registered nurses per state. We also collected data on the population over 65, as older populations typically require more medical care. We included median income because, unfortunately, lower incomes tend to correlate with worse health outcomes. The last piece of data we used at the state level is the median out-of-pocket cost. Other studies (Scheffler and Arnold, 2019) have explained that more spending results in higher utilization of medical services. We collected NCLEX pass rates and nurse retirement data on a national level only. We were only able to find the NCLEX data at a national level, as where one graduates from nursing school doesn't directly determine where a nurse could end up working. For this reason, we conducted our supply and demand forecasts at a national level.

We curated the demand equation and forecasted supply and demand equations. Due to limited data for most of our factors, such as retirement data, and time constraints for our overall supply and demand trend, we have decided to go national rather than regional.

Mathematical Concepts

Equations

We created equations for supply and demand and learned more about ARIMA math. Finding these equations included reading some similar research in the past ([**2**, **4**]) and generating an equation using similar factors for both supply and demand. For ARIMA, we used the two articles mentioned above to see how to set up an ARIMA equation and how the factors we have will affect ARIMA forecasting.

Our demand model expands on previous methods developed by Scheffler et al. (2008). In our model addition to income (i.e. median income per state), we include the size of the population aged 65 or over as an indicator of aging, as it will increase the demand for health services in that state. We also include private per capita household out-of-pocket (OOP) spending on medical care as an indicator for healthcare spending since the portion spent OOP is largely determined by the level of health coverage by health insurance. Less generous health coverage leaves individuals to pay more OOP, which is expected to lower the demand for and use of health services. Thus, higher OOP health spending correlates with lower demand for health workers.

In terms of variables, μ_i represents a vector of country-fixed effects, ξ_{it} is the disturbance term, and β_i coefficients are unknown parameters to be estimated from the model.

Demand Equation:

$$\begin{aligned}\ln(\text{nurses per 1000 population}_{it}) = \beta_0 &+ \beta_1 \cdot \ln(\text{median income per state}_{it-1}) \\ &+ \beta_2 \cdot \ln(\text{median income per state}_{it-4}) \\ &+ \beta_3 \cdot \ln(\text{median income per state}_{it-5}) \\ &+ \beta_4 \cdot \ln(\text{OOPPC}_{it-2}) \\ &+ \beta_5 \cdot \ln(\%\text{Pop65}_{it-3}) \\ &+ \mu_i + \xi_{it}\end{aligned}$$

Supply Equation:

Nurses per 1000 population = $(\beta_0 + \beta_1 \cdot \text{year}_t + \xi_t)$

To predict the supply of health workers in 2030, we use historical data to predict the changes in health worker densities nationwide. We assume that each state's historical growth rate of health worker density will continue into 2030 at the same rate.

Our variables include t, time(in years), ε_t, the random disturbance term, and β_0, and β_1 are unknown parameters (found through ARIMA), with the last two parameters representing the linear growth rates to be estimated from the model.

ARIMA:

Our supply projections use historical data on nurse supply by states. We assume that the factors that determine supply are static and project the number of nurses in 2030 using trends that result from fitting an autoregressive integrated moving average (ARIMA) model for each state for nurses. ARIMA models are applied when the data shows evidence of non-stationarity and an initial differencing of the data (the 'integrated' part) is applied to reduce the non-stationarity. ARIMA models are generally denoted ARIMA(p, d, q) with parameters p, d and q being non-negative integers which denote the number of time lags of the autoregressive model (p), the degree of differencing (d) and the order of the moving average model (q).

1. ARIMA(p,d,q) non-negative integers parameters
   (a) p = the number of time lags of the autoregressive model
   (b) d = the degree of differencing

(c) q = order of moving average model

2. Let Y denote the original series and y the stationarised series.

3. $y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{t-p}$ - $\theta_1 e_{t-1} - ..... - \theta_q e_{t-q}$

    (a) $\phi_1 y_{t-1} + ... + \phi_p y_{t-p}$ = AR Term (lagged values of y)

    (b) $\theta_1 e_{t-1} - ..... - \theta_q e_{t-q}$ = MA Terms ((lagged errors)

    (c) c = constant

4. On the supply side, the number of nurses in every state is a function of

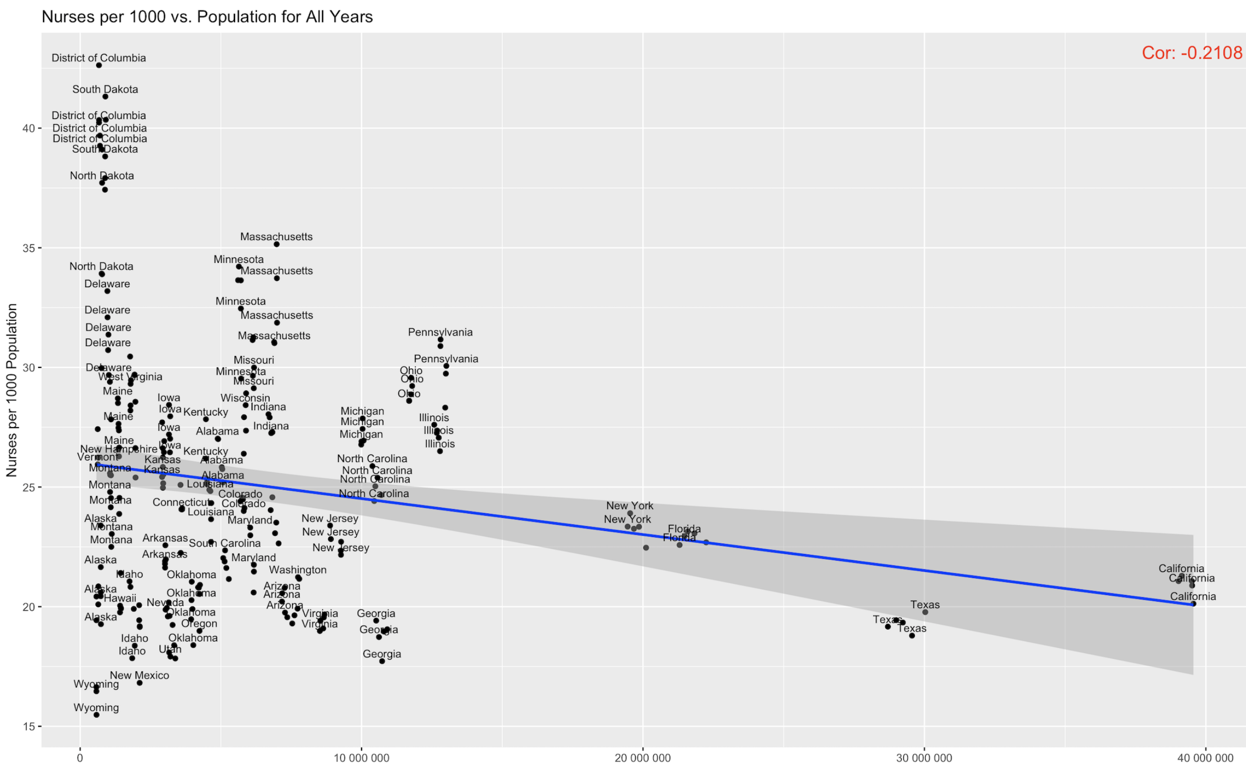
    (a) the number of graduates

    (b) the import or export of health workers and

    (c) deaths and retirements of health workers.

Correlations

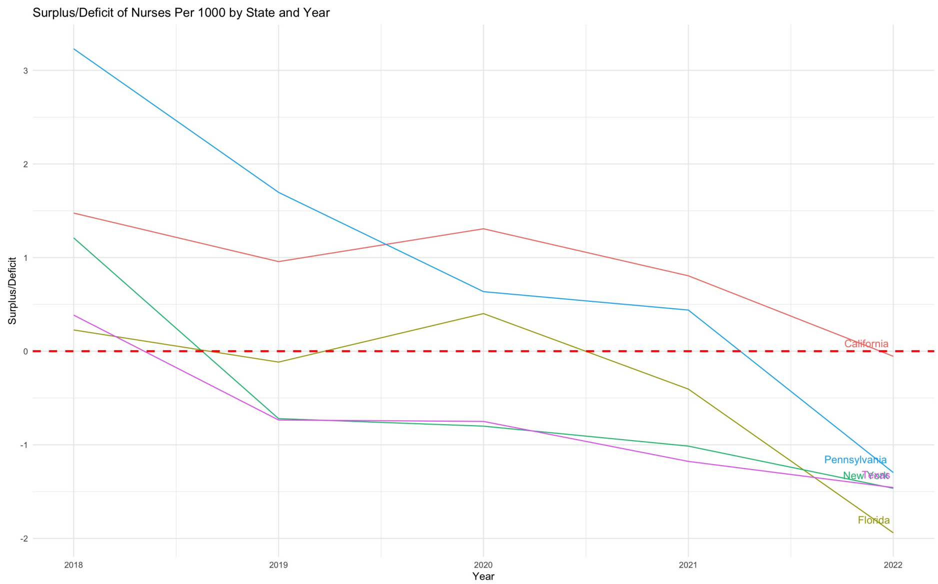
While reviewing the data looking at all the states, correlations between the data and contributing factors appeared weak. Also, the smaller states have a greater variance in data, a phenomenon attributed to the law of large numbers and sampling variability. This principle indicates that as the sample size increases, the sample mean approaches the population mean, resulting in reduced variability. A graph that helps illustrate this in our data is shown below.



This graph highlights all of our data points. There are 5 points for each state, one for each year in out dataset. The values are plotted and a linear regression was computed. The shaded region with the line is a confidence interval of the slope and intercept estimates. The wider it is the less confident the model. It gets wider towards the right of the graph as there is less data on that side.

Current Trends

As discussed earlier, looking at the data for all 50 states can make it hard to notice any trends. Because of this, we focused on the five states with the largest populations to view current trends. We also thought this might be advantageous to our industry partner, as these larger states will need more coverage. In the image below, we visualize the supply and demand of these states from the years 2018 to 2022. All the states are displaying a negative trend, and for the first time in the data, all states experienced a deficit in 2022.



Limitations

Part of our limitations was dealing with real-world data. Real-world data can be messy and may not always behave as you would expect. All of the predictors we used have been utilized in similar previous studies but did not appear strongly correlated with the data across the states. This presented issues when refining our model. Another limitation with our data was obtaining all of the necessary information. For example, we are not able to effectively compare future supply at the state level because one of our key predictors, the NCLEX Pass Rate, is only measured at the national level. Furthermore, even if this data were accessible by state, graduating in a state doesn't mean that a person will work in that state. They may return home, choose to migrate, or, like in our project, they may opt to be travel nurses. Because of this significant limitation, we conducted our prediction at a national level.

Prediction

Going back to our approach, we have decided to go on National level when it comes to predicting overall Supply and Demand, and to do that we will use the R package. The R package `forecast` helps businesses build predictive models, visualize data, and make accurate forecasts by providing methods and tools for displaying and analyzing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modeling. The `forecast` package allows the user to explicitly specify the order of the model using the `arima` function, or automatically generate a set of optimal (p, d, q) using auto.

If our current equation does not cooperate, we will look into the Markov Model. Markov Chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. In R, there is a package called `markovchain`, which will be used to create a Transition Matrix and use functions from the package to forecast future distribution and long term stability analysis.

Conclusion

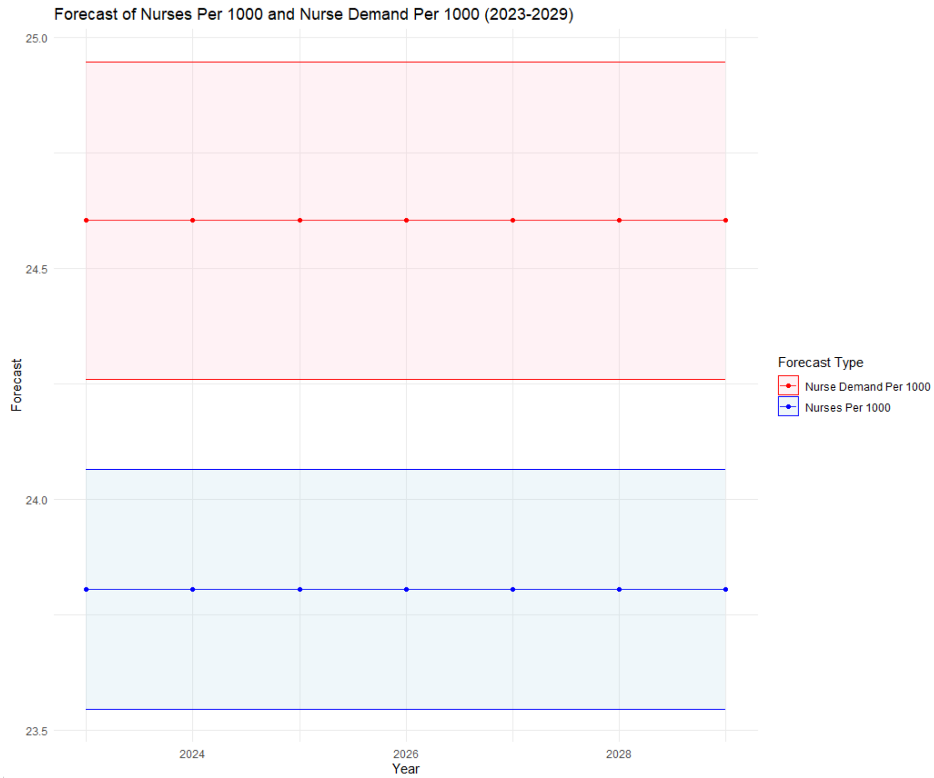
In conclusion...

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2023	24.6047	24.38002	24.82938	24.26108	24.94832
2024	24.6047	24.38002	24.82938	24.26108	24.94832
2025	24.6047	24.38002	24.82938	24.26108	24.94832
2026	24.6047	24.38002	24.82938	24.26108	24.94832
2027	24.6047	24.38002	24.82938	24.26108	24.94832
2028	24.6047	24.38002	24.82938	24.26108	24.94832
2029	24.6047	24.38002	24.82938	24.26108	24.94832

```
> print(forecast_nursesd)
     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2023        24.6047 24.38002 24.82938 24.26108 24.94832
2024        24.6047 24.38002 24.82938 24.26108 24.94832
2025        24.6047 24.38002 24.82938 24.26108 24.94832
2026        24.6047 24.38002 24.82938 24.26108 24.94832
2027        24.6047 24.38002 24.82938 24.26108 24.94832
2028        24.6047 24.38002 24.82938 24.26108 24.94832
2029        24.6047 24.38002 24.82938 24.26108 24.94832
```

Above is the results of our ARIMA models. While the model was successfully computed there are some problems. The main problem is that the prediction is constant year over year. The constant value of the forecast across years indicates that the model does not anticipate any significant changes in the ratio of nurses per 1000 people within the forecasted period, given the historical data and trends it analyzed. This might suggest that the factors driving the number of nurses per 1000 people are expected to remain relatively stable, or it could be a limitation of the model in capturing potential future changes based on the available data. Below is the visualization of the results. While the prediction is constant, the model does predict the demand to outweigh the supply in the next five years

```
# Plotting with ggplot2
ggplot(combined_forecast, aes(x = Year, y = Forecast, color = Type)) +
  geom_line() +
  geom_ribbon(aes(ymin = Lower, ymax = Upper, fill = Type), alpha = 0.2) +
  geom_point() +
  scale_color_manual(values = c("Nurses Per 1000" = "blue", "Nurse Demand Per 1000" = "red")) +
  scale_fill_manual(values = c("Nurses Per 1000" = "lightblue", "Nurse Demand Per 1000" = "pink")) +
  labs(title = "Forecast of Nurses Per 1000 and Nurse Demand Per 1000 (2023-2029)",
       y = "Forecast", x = "Year") +
  theme_minimal() +
  guides(color = guide_legend(title = "Forecast Type"), fill = guide_legend(title = "Forecast Type"))
```


Forecast of Nurses Per 1000 and Nurse Demand Per 1000 (2023-2029)
25.0
24.5
24.0
23.5
Forecast
2024
2026
2028
Year
Forecast Type
Nurse Demand Per 1000
Nurses Per 1000

‘

Acknowledgment

1. PIC Math is a program of the Mathematical Association of America (MAA) and the Society for Industrial and Applied Mathematics (SIAM). Support is provided by the National Science Foundation (NSF grant DMS-1722275).

2. Dr. Zvi Rosen (Professor at FAU)

3. Robbie Bok (VP at American Traveler)

4. American Traveler (Industry Partner)

References

1. Juraschek, S. P., Zhang, X., Ranganathan, V., & Lin, V. W. (2019). Republished: United States Registered Nurse Workforce Report Card and Shortage Forecast. *American Journal of Medical Quality, 34*(5), 473-481.

2. Liu, J. X., Goryakin, Y., Maeda, A., Bruckner, T., & Scheffler, R. (2016). Global Health Workforce Labor Market Projections for 2030. *Health Nutrition and Population Global Practice Group.*

3. National Council of State Boards of Nursing. (2022). NCLEX statistical brief: Q4 pass rates.

4. Roussel, A., Biard, R., Deschamps, M., Diss, M. (2023, September). Modeling medical material shortage using Markov processes [Review of Modeling medical material shortage using Markov processes].

5. Scheffler, R. M., & Arnold, D. R. (2019). Projecting shortages and surpluses of doctors and nurses in the OECD: what looms ahead. *Health Economics, Policy and Law, 14*(2), 274–290.

6. Smiley, R. A., et al. (2022). The 2022 National Nursing Workforce Survey. *Journal of Nursing Regulation, 14*.

7. State Health Access Data Assistance Center (SHADAC). (2022). Median Out-of-Pocket Medical Spending, by Insurance Coverage: 2018-2022.

8. U.S. Bureau of Labor Statistics. (2022). Occupational Employment and Wages, Registered Nurses: 2018-2022.

9. U.S. Census Bureau. (2022). State Population Totals: 2018-2022.