MathWorks Math Modeling Challenge
A program of siam

PREVIEW PAPER: EXCELLENT

The team had a good summary and was somewhat consistent with the requirement that it be made for a high level political figure. Many teams struggled with this aspect of the problem statement. Overall, the team did a good job with respect to the writing and presentation, and their tables and graphs were clearly labeled, described well, and annotated nicely. The team provided a strengths and weakness subsection for each question, but they did not provide an adequate sensitivity analysis.

The team did a good job of stating their model, and they clearly indicated the way they calculated their estimates, but the estimates for some parameters were missing. The results were presented in a clear, succinct manner, and the results seemed reasonable. The team examined the data within the data sets with a nice comparison across the data sets. Additionally, the team provided a good reference for Mercer's constant, and even though their assumptions were problematic in some respects, their efforts were consistent with the assumptions. Another positive is that the team identified surprising results and explicitly noted which results were not expected.

The team's approach to question two was consistent with their work for question one. For question two, though, they based their projections on all jobs and did not restrict their calculations to remote-ready jobs. On the plus side, they remained consistent in how they interpreted their results. The team made use of two factors, age and occupation, which was good, but other teams were able to incorporate more factors in their models.

**from among the screened sample of papers examined during pre-triage work.*

Remote Work: *Fad or Future?*

1. Executive Summary

Dear Mr. President,

As the COVID-19 pandemic continues to impact our world, from education to employment, we become increasingly reliant on remote work. However, there are some jobs that require a physical presence that are simply unable to be completed to a satisfactory degree remotely, given our current technology. Our team aims to create models that predict the percentage of workers who can and will work remotely in 2024 and 2027, as well as evaluate the magnitude of the impact that remote work will have on any given city.

We first created a model to predict the percentage of workers with “remote-ready” jobs in Seattle, Omaha, Scranton, Liverpool, and Barry in 2024 and 2027. First, we calculated the number of workers in remote-ready jobs, and consequently the percent of jobs in a city that are remote-ready. By applying this equation to all the cities over past years, we calculated the percentage of remote-ready workers in each city over time. Finally, using the *scipy.optimize* module in Python, we fit a linear regression model to this data to predict the percentage of remote-ready workers in 2024 and 2027 for all five cities. Our model predicts that Seattle and Omaha are both growing in remote-readiness: +0.146% and +0.0960% per year, respectively. Barry is growing at a slighter rate of +0.0313% per year. Scranton’s change of –0.003% per year is relatively static, while Liverpool is decreasing in remote-readiness, at –0.142% per year.

Next, we were tasked with finding the percentage of the employed workforce of a specific occupation who, given the option, would elect to work remotely. Taking into account age, job type, and education level, we produced an equation that can be applied to any job category to determine the desired population. To test our model, we applied it to a niche population (works in “Life, Physical, and Social Science”, lives in Seattle, aged between 25-54, and holds a Bachelor’s degree) and found the percentage of this population who will choose to work from home if offered. With data from D1 (City Employment Data) and D3 (Remote Work Data), we concluded that if given the option, 32.35% of Life, Physical, and Social Science workers who live in Seattle, hold a Bachelor’s degree, and are aged between 25-54 would elect to work remotely.

For Part III, we applied our previous models to determine an individual's likelihood of working from home to each member of a city’s entire working population to get an aggregate model for the whole city that is extremely accurate. We summed up all the combinations to produce a precise and accurate estimate of the remote work participation rate in various cities across the western world over multiple years. Recognizing a linear relationship within this time-series, we determined the remote work participation rate for the rest of the decade and ranked the most impacted cities accordingly. In 2024 and 2027, from most to least impacted, we found the ranking to be Seattle, Omaha, Barry, Scranton, and Liverpool.

Table of Contents

2. Part I: Ready or Not

2.1 Problem Restatement

In this problem, we created a model that calculates the percentage of workers in a city whose jobs are "remote-ready", which we define in the assumptions below. We then applied this model to Seattle, Omaha, Scranton, Liverpool, and Barry to predict the percentage of workers with remote-ready jobs in these cities in 2024 and 2027, in which we also accounted for how inputs to our model changed over time.

2.2 Assumptions

1. *"Remote-ready" jobs are defined as jobs that can be performed at an adequate standard primarily online.* These are jobs that have the potential to be properly done from home.

2. *We consider the following job categories from D1 City Employment Data and D3 Remote Work Data [4] to be equivalent to each other*. This enables us to better compare Remote and City data.

<table>
<tr><th><u>D1 City Employment Data (Job Categories)</u></th><th colspan="4"><u>D3 Remote Work Data (Sectors)</u></th></tr>
<tr><td>Mining, Logging, Construction</td><td colspan="2">Installation, Maintenance, & Repair</td><td colspan="2">Construction & Extraction</td></tr>
<tr><td>Manufacturing</td><td colspan="2">Farming, Fishing, & Forestry</td><td colspan="2">Production</td></tr>
<tr><td>Trade, Transportation, Utilities</td><td colspan="2">Building and grounds cleaning & Maintenance</td><td colspan="2">Transportation & Material Moving</td></tr>
<tr><td>Information</td><td colspan="2">Computer & Mathematical</td><td colspan="2">Architecture & Engineering</td></tr>
<tr><td>Financial Activities</td><td colspan="4">Business & Financial Operations</td></tr>
<tr><td>Professional & Business Services</td><td>Legal</td><td>Management</td><td>Office & Administrative</td><td>Sales & Related</td></tr>
<tr><td>Education & Health Services</td><td>Education, Training & library</td><td>Community & Social Service</td><td>Healthcare Practitioners & Technical</td><td>Healthcare Support</td></tr>
<tr><td>Leisure & Hospitality</td><td colspan="2">Arts, Design, Entertainment, Sports & Media</td><td>Personal Care & Service</td><td>Food Preparation & Service Related</td></tr>
<tr><td>Government</td><td colspan="4">Protective Service</td></tr>
<tr><td>Other Services</td><td colspan="4">Life, Physical, & Social Science</td></tr>
</table>

Table 1. Classifying professions from D3 (Remote Work Data) into comparable categories from D1 (City Employment Data) [4]

3. *In the next decade, the change in percentage of jobs that can be done at home is negligible.* Despite the impacts of the pandemic in recent years, we are assuming that remote-work statistics will remain somewhat consistent in the next few years, until 2027.

4. *Jobs such as construction, farming, fishing, and other similar manual and labor intensive professions will remain primarily non-remote through 2027.* We are assuming that these jobs won't be mechanized within the next 10 years, thus there will not be an exponential increase or any other sort of dramatic shift in the percentage of remote-workers in these professions.

5. *The unemployed population is not taken into account in our model.* Since we are only given data for the currently employed population, we are unable to predict the career trends of the unemployed, thus we are excluding them from our calculations and only considering those who are currently employed.

2.3 Designing the Model

Analyzing the data we are given in the D1 (City Employment Data) and D3 (Remote Work Data) sheets [4], we realized we have data for the number of people in each job category in the given cities over the past years. We also have the percentage of those jobs that have the possibility of being done at home, which is what we consider remote-ready. Each profession also has its own standard of qualification for what is defined as remote-ready.

When examining a specific job category, we multiplied **the number of workers per job (pop_{job})** by **the percentage of jobs that are remote-ready (rr_{job})** to calculate **the percentage of workers with remote-ready jobs (rr_{pop})**. We repeated this process for every job category and summed the results, which produced **the total number of workers with remote-ready jobs**. Then, we divided this number by **the total number of workers in the workforce (pop_{tot})** in order to obtain rr_{pop} in the entire city. We repeated these calculations for each year of the data provided to us. This produced a dataset of rr_{pop} over time over which we fit a linear regression model to predict rr_{pop} in 2024 and 2027. The entire process thus far was then repeated with data from each of the five cities.

2.3.1 Remote-Ready Jobs by City

The variables of this equation are defined as followed:

rr_{pop} = *the percentage of the population that are workers with remote-ready jobs**
rr_{job} = *the percentage of total jobs that are remote-ready*
pop_{job} = *the number of employed workers in each categorical profession*
pop_{tot} = *the sum of pop_{job} for all categorical professions; the total employed workforce*

**rr_{pop} *describes the "remote-readiness" of a city, and is used interchangeably with this term.*

To calculate **the percentage of workers with remote-ready jobs**, we divided **the total number of workers with remote-ready jobs** by **the total employed workforce**. This produced the following equation:

$$rr_{pop} = \frac{\sum \left(pop_{job} \cdot rr_{job} \right)}{pop_{tot}}$$

Eq 1. Equation to Calculate Percentage of Workers with Remote-Ready Jobs

Applying this equation to all the cities produced a set of rr_{pop} values over past years for each city, which describes numerically the progression of remote-readiness in each city. This produces the following data:

City	2000	2005	2010	2015	2019	2020	2021
Seattle	28.46%	29.20%	29.76%	29.80%	30.69%	30.89%	32.56%
Omaha	29.77%	30.40%	31.38%	31.74%	31.86%	31.88%	31.65%
Scranton	25.81%	25.61%	26.05%	26.47%	26.13%	25.57%	25.35%
Liverpool	N/A	27.3%	27.72%	26.87%	25.65%	24.79%	26.06%
Barry	N/A	33.45%	35.57%	35.44%	34.77%	35.78%	33.45%

Table 2. Results for Percentage of Remote-Ready Workers in Each City

The data from Table 2 will be the inputs for our linear regression model in the next section.

2.3.2 Linear Regression Model

In order to predict the percentage of workers with remote-ready jobs for a city's future, we fit a linear regression model to the results of Eq. 1 (Table 2) for past years for each city, where the independent variable is *years* and dependent variable is rr_{job}. We did this with the *scipy.optimize* module in Python.

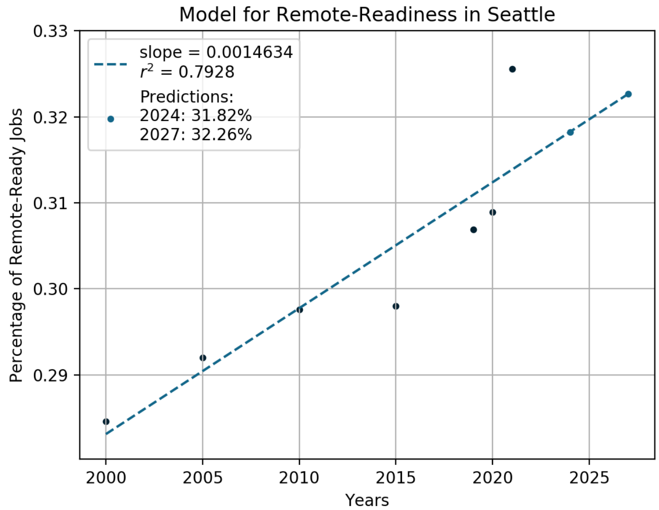

Fig 1. Linear Regression Model for Seattle's Remote-Readiness

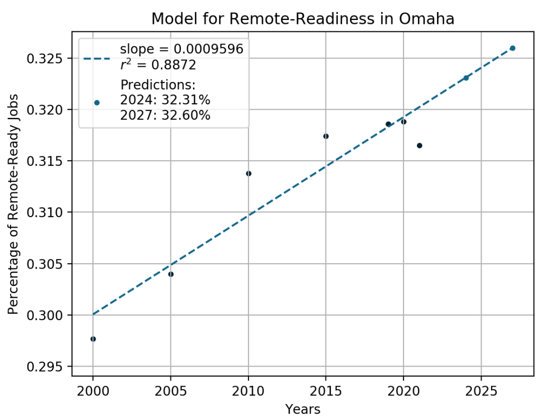

Fig 2. Linear Regression Model for Omaha's Remote-Readiness

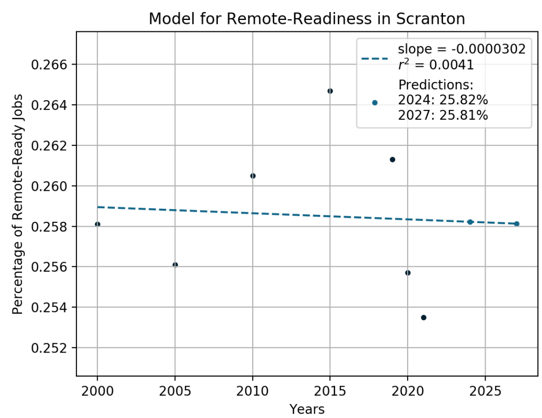

Fig 3. Linear Regression Model for Scranton's Remote-Readiness

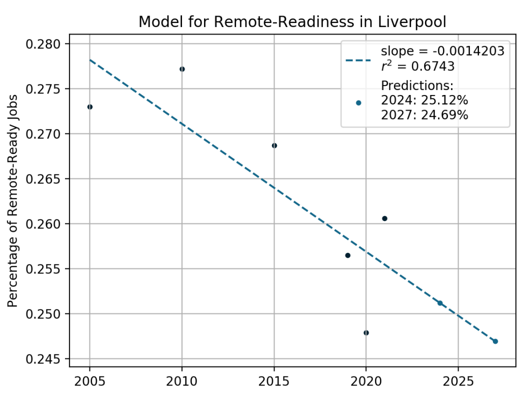

Fig 4. Linear Regression Model for Liverpool's Remote-Readiness

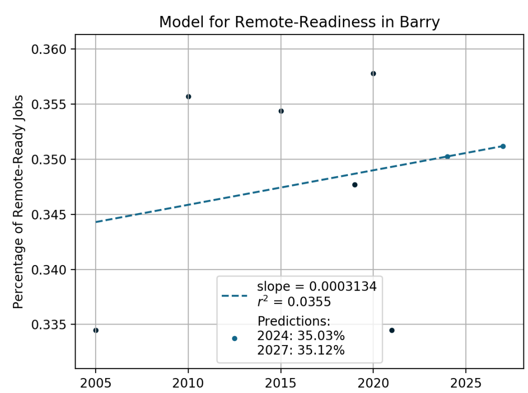

Fig 5. Linear Regression Model for Barry's Remote-Readiness

2.4 Results

Using our linear regression model, we were able to predict the remote-readiness of each city in 2024 and 2027, as outlined in the table below:

City	Predicted Remote-Readiness in 2024	Predicted Remote-Readiness in 2027
Seattle	31.82%	32.26%
Omaha	32.31%	32.60%
Scranton	25.82%	25.81%
Liverpool	25.12%	24.69%
Barry	35.02%	35.12%

Table 3. Predicted Remote-Readiness of Each City in 2024 and 2027

Seattle (+0.146% per year) and Omaha (+0.0960% per year) are both increasing in remote-readiness. Barry (+0.0313% per year) is also growing, yet at a more slight rate. Scranton's change (–0.003% per year) is negligible enough to be considered static. Liverpool (–0.142% per year) is decreasing in remote-readiness.

2.5 Strengths and Weaknesses

Our prediction of remote-readiness in five cities has the following strengths and weaknesses:

2.5.1 Strengths

1. Our model will continue to be reliable as long as a city's workforce policies and conditions remain relatively constant, as our model is based upon statistics of past employment trends.
2. The linear regression model is performed by a robust and reputable Python module, which adds confidence to our results.

2.5.2 Weaknesses

1. Our model excludes the currently-unemployed workers of the total workforce population, thus we are unable to account for their growth and contributions to the workforce.
2. Since our data was extrapolated from historical statistical trends, it does not account for future policy changes in each city.
3. Our assumption that the working population in each sector is evenly distributed among the job categories creates a source of inaccuracy, which causes error proportional to the number of sectors combined in the category in the calculation of rr_{job}.

3. Part II: Remote Control

3.1 Problem Restatement

In Part II, we created a model that predicted whether an employed worker whose job is “remote-ready” will be permitted to work from home by their employer, and if so, based on their age and education level, whether the worker will choose to work from home.

3.2 Assumptions

1. *Firms will operate in the way that most optimally maximizes their profits.* We assume that the only factor influencing whether a firm will decide to allow its workers to work from home is whether it is the most profitable move for the firm.

2. *An increase in worker productivity, whether in-person or remote, leads to an increase in company profit*. We are disregarding other factors that could play into the growth and decline of company profit, and only focusing on the impact of worker productivity.

3. *Remote productivity will remain constant across multiple years.* As we are using past data in our model, we assume that the rate of job productivity will not change over time.

4. *We used Mercer’s constant (β =0.94), which assumes that 94% of firms will allow remote work if it is possible to be done remotely.* Mercer, an HR consulting firm, found through a survey of 800 employers conducted in 2020 during the COVID-19 pandemic, at the height of remote work, that 94% of employers stated that their employees’ productivity had improved or remained constant during remote work. [3][6]

5. *The only factors that influence someone’s decision to work remotely are their age, education level, and job type.* While there are many other factors that influence whether someone decides to work at home, because of the limitations of our data we will only be examining the influence of age, education level, and job type.

6. *Age and education level (National Qualifications Framework level) distributions are equal across all job sectors*. This assumption was made by observing that the distributions of education level did not vary substantially among different age groups. [5]

7. *The same assumptions persist from Part I.*

3.3 Designing the Model

In the next question, we created a model that **uses a person's age, job type, and level of education** quantified by their National Qualifications Framework level and returns **the percentage of people that would choose to work from home if given the option**. We assumed that only about 94% of businesses would implement remote-work based on data from Mercer, an HR consulting firm [3][6]. We used this constant to first calculate **the percentage of the total employed workforce of a job type allowed to work from home**. We then specified an expression to represent the total number of employed workforce allowed to work from home of a **certain age and education level who would choose to work from home if able**. We then divided this expression by the **total employed workforce** of that age and education level. This gives us **the percentage of people that would choose to work from home if given the option**.

3.3.1 Percentage of the Employed Workforce in a Particular Occupation Allowed to Work Remotely

The variables for this equation are defined as follows:

$ß$ = 0.94 (Mercer's Constant)
rr_{job} = the percentage of total jobs that are remote-ready
pop_{job} = the number of employed workers in each categorical profession
Ar_{job} = the percentage of the employed workforce of a particular job (pop_{job}) ***allowed*** *to work remotely*

To calculate **the percentage of the employed workforce of a particular job allowed to work remotely (*Ar_{job}*)**, we first multiplied **the number of employed workers in each profession** by **the percentage of total jobs that are remote-ready** and **Mercer's Constant**. We then took that product and divided it by the **the total number of employed workers in each profession**, which produced the following equation:

$$Ar_{job} = \frac{pop_{job} \cdot rr_{job} \cdot ß}{pop_{job}}$$

Eq. 2 Percentage of Employed Workforce of a Particular Job Allowed to Work Remotely

This equation then simplifies to eliminate pop_{job}; this makes sense because the percentage of jobs that will offer remote work is simply **the product** of **the percent of jobs that are remote-ready (*rr_{job}*)** and **the probability that a firm will offer remote work (*Mercer's Constant ß*)**. The simplified equation is:

$$Ar_{job} = rr_{job} \cdot ß$$

Eq. 3 Simplified Version of Eq. 2

The resulting percentage of the employed workforce of a particular job allowed to work remotely (Ar_{job}) will be used in the next part of our model.

3.3.2 Percentage of the Employed Workforce in a Particular Occupation that will Elect to Work Remotely if Offered

The variables for this equation are defined as follows:

rw_{age} = the percentage of the total employed workforce of a particular age that works remotely
rw_{NQF} = the percentage of the total employed workforce of a particular NQF that works remotely
w_{age} = the number of workers of a particular age in a particular occupation (proportional to the total amount of workers of that age in the US out of the total number of workers in the US)
w_{NQF} = the number of workers of a particular NQF in a particular occupation (proportional to the total amount of workers of that NQF in the US out of the total number of workers in the US)
ER_{job} = the percentage of workers in a particular occupation who will choose to work remotely

To calculate **the percentage of workers in a particular occupation who will choose to work remotely (*ER_{job}*)**, we took **the number of people in a particular age and education group who both agreed to and was offered to work remotely**, and divided it by **the number of people in a particular age and education group in the workforce for that job type**. The numerator of this equation is as follows:

$$\left(Ar_{job}\right)\left[\left(rw_{age}\cdot rw_{NQF}\right)\left(w_{age}+w_{NQF}\right)\right]$$

Eq. 4 Expression for the Number of People in a Particular Age and Education Group who Agreed to Work Remotely when Offered

This expression was calculated by taking **the percentage of the employed workforce of a particular job allowed to work remotely (*Ar_{job}*)** and multiplying it by **the percentage of the people with particular education level NQF and particular age range,** and **the number of people of that age and NQF in that particular job**. This produces **the number of people in a particular age and education group who agreed to work remotely when offered**. The denominator of this equation is as follows:

$$pop_{job}\left(\frac{w_{age}}{pop_{job}}\right)\left(\frac{w_{NQF}}{pop_{job}}\right), \text{ which simplifies to } \frac{w_{age}\cdot w_{NQF}}{pop_{job}}$$

Eq. 5 & Eq. 6 Number of People of a Particular Age and Education Group in the Workforce for a Job Type

This expression was calculated by multiplying **the number of employed workers in a particular job (*pop_{job}*)** by **the proportion of people of a particular age in the workforce** and **the proportion of people of a particular education level in the workforce**. This produces **the number of people in a particular age and education group in the workforce for that job type**. To then calculate **the percentage of workers in a particular occupation who will choose to work remotely (*ER_{job}*)**, we take Eq. 4 and divide it by Eq. 6, as previously explained. This results in the following equation model:

$$ER_{job} = \frac{\left(Ar_{job}\right)\left[\left(rw_{age} \cdot rw_{NQF}\right)\left(w_{age} + w_{NQF}\right)\right]}{\left(\frac{w_{age} \cdot w_{NQF}}{pop_{job}}\right)}$$

Eq. 7 Percentage of Workers in a Particular Occupation who Chose to Work Remotely When Given the Option

3.4 Results

To test our model, we determined the percentage of people who will choose to work from home if offered in a **specific scenario**, defined as follows:

Profession in the category "Life, Physical, and Social Science"
Lives in Seattle, WA
Aged between 25-54
Holds a Bachelor's degree (NQF Level 6)

With these specifications, using data from D1 (City Employment Data) and D3 (Remote Work Data) [4], we were able to make the following determinations:

$rr_{job} = 0.54$	$rw_{age=25\text{-}54} = 0.29$	$rw_{NFQ=6} = 0.41$	$w_{age=25\text{-}54} = 38201$	$w_{NFQ=6} = 15222$	$pop_{job} = 59300$

With this data, we were able to apply our two equations and make the following two conclusions:

$$Ar_{job} = 0.51$$
$$ER_{job=\text{Life, Physical, \& Social Science}} = 0.3235$$

Using our model, we conclude that if given the option, **32.35%** of Life, Physical, and Social Science workers who live in Seattle, hold a Bachelor's degree, and are aged between 25-54 **would elect to work remotely**.

3.5 Strengths and Weaknesses

Our model predicting a worker's ability and desire to work remotely has the following strengths and weaknesses:

3.5.1 Strengths

1. Our model employs multiple variables to calculate the percentage of workers who are able to, and who will elect to, work remotely. We consider age as well as education level in tandem with one another, increasing the reliability and accuracy of our model.

3.5.2 Weaknesses

1. Our model does not take into account the situation of if a worker wants to work from home but their employer doesn't allow them, which could make up a sizable portion of the population.
2. Moreover, our model does not factor in data regarding hybrid models which are a combination of both work from home and in person work.
3. In our calculations, we assumed that age, education level, and job type were the only factors influencing one's decision to work remotely. We do not take into account gender, family size, or other factors that may impact a person's life if they choose to work remotely.

4. Part III: Just a Little Home-work

4.1 Problem Restatement

In the final section, we integrated our models from the previous two questions to create a model that estimates the percentage of remote workers in any given city. Using the five cities from the previous sections, we made predictions for the remote-working populations of the cities in 2024 and 2027, then ranked these cities in terms of how much remote work will impact them.

4.2 Assumptions

1. *The same assumptions persist from Part I and Part II.*

4.3 Designing the Model

To create our model for Part III, we first found the **total population of specific city** and found the **percentage distribution of workers in certain categories**. For instance, to find the amount of people in a city with no high school diploma in the age range of 16 to 24, we'd multiply **the distribution percentages of each category** and **the population** together. We assume that these age and education distributions for the total US workforce population will remain the same for the individual cities. Because there are 5 different education categories (NQF 1, NQF 2-3, NQF 4-5, NQF 6, and NQF 7-8) and 4 different age-ranges (16-24, 25-54, 55-64, 65+), there are **20 possible age-education combinations for each job category**. Each of these numbers of people will be multiplied by a distinct ER_{job} (percentage of job population that will work from home) from the previous model. Recall that ER_{job} is dependent on an individual's age, education, in addition to the profession in which they work. The sum of the amounts of remote workers for each distinct combination is then divided by the total workforce population for the city in a given year. We repeat this **10 times** for each aggregate category of job.

Mining, logging, construction
Manufacturing
Trade, transportation, and utilities
Information
Financial activities
Professional and business services
Education and health services
Leisure and hospitality
Other services
Government

After getting all of those values we repeat for each of our cities in each year for which we were provided data (2005-2021). From this time-series we are able to construct a linear regression model to predict the future remote work participation rate for each city.

4.3.1 Calculation for one cell

The variables for this equation are defined as follows:

$a_{age,\ NQF}$ *= percentage of total working population that belongs in the specified age and NQF range*
P_{job} *= population in the specified job category*
$ER_{job}(age_i\ ,\ NQF_i\ )$ *= the percentage of workers in a particular occupation and specified age and NQF range who will choose to work remotely*

To calculate the number of remote workers for one specified group of age and NQF, we created the following equation:

$$W_{remote}(age, NQF) = a_{age,\ NQF} \cdot \sum_{jobs} \left[P_{job} \cdot ER_{job}(age, NQF) \right]$$

Eq. 8 Number of remote workers as a function of age group and NQF group.

This shows **the number of remote workers** for each group of age and NQF is **the percentage of all workers** who belong in the age and NQF group (a_{age}) multiplied by **the number of workers in each job category** who will work remotely ($P_{job} \cdot ER_{job}(age, NQF)$) summed over all job categories.

4.3.2 Example Table for Seattle in 2021

By applying the previous equation to all possible age and NQF groups in a city, we can calculate the number of remote workers for each group, as shown in this example with data for **Seattle** in **2021**:

NQF\Age	16-24	25-54	55-64	65+
1	9202.30472458	65535.17099143	22977.00424284	14005.83813591
2-3	9445.94445016	56761.24413608	20448.72462206	18001.26812816
4-5	4068.69877591	23288.13292458	9643.03049663	6585.25553758
6	2024.57729865	11879.60962295	4823.0067812	3248.79223267
7-8	392.6568276	3583.21720479	1050.68988921	529.90496318

Table 4. Number of Remote Workers for Each Group of Age and NQF in Seattle in 2021

We then summed the value of all cells in the table to arrive at the total number of remote workers in a city, which was then divided by the number of total workers to calculate the percentage of remote workers in a given city in a given year.

$$Percent_{remote} = \frac{W_{remote}}{W_{total}}$$

Eq. 9 Percentage of remote workers in a given city for a given year.

We then repeated this above process for each year to get a time series of remote worker percentage, over which we performed linear regression, similar to Part 2.3.

4.3.3 Linear Regression Model

To predict the percentage of remote workers for each city, we fit a linear regression model using data generated from iterating our previous model from Part II (*ER*) with hundreds of different combinations. We then produced a time series where the independent variable is *years* and the dependent variable is the *percent of remote workers*. We did this with the *scipy.optimize* module in Python.

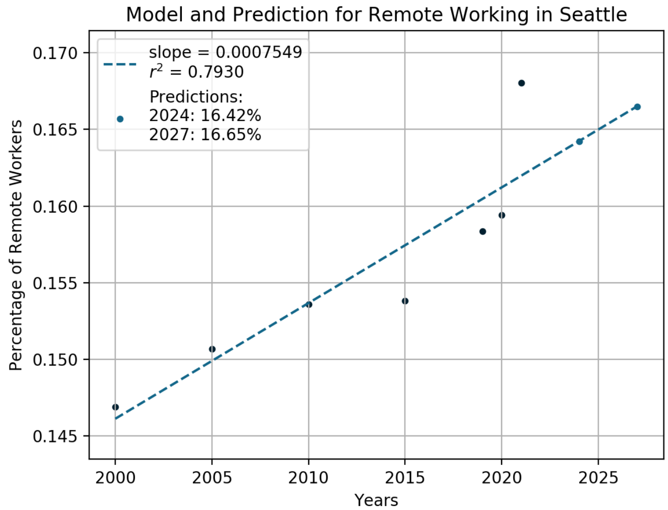

Fig 6. Linear Regression Model for Seattle's Remote Workers

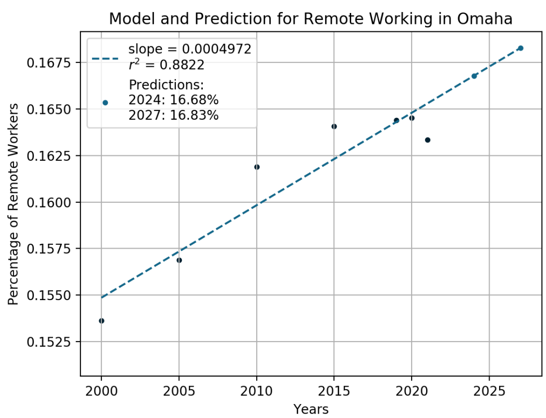

Fig 7. Linear Regression Model for Omaha's Remote Workers

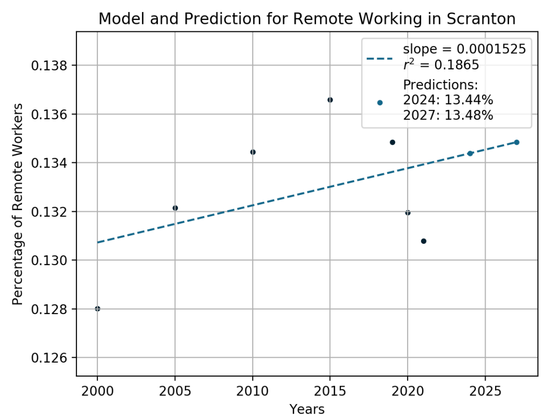

Fig 8. Linear Regression Model for Scranton's Remote Workers

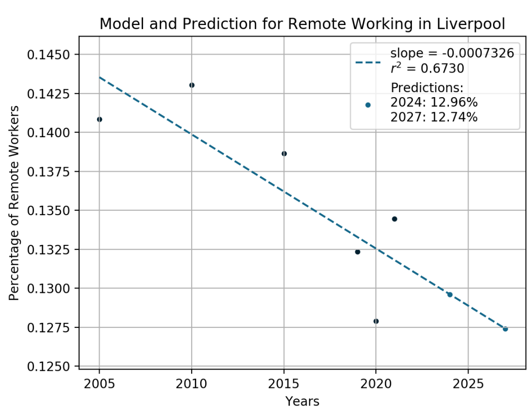

Fig 9. Linear Regression Model for Liverpool's Remote Workers

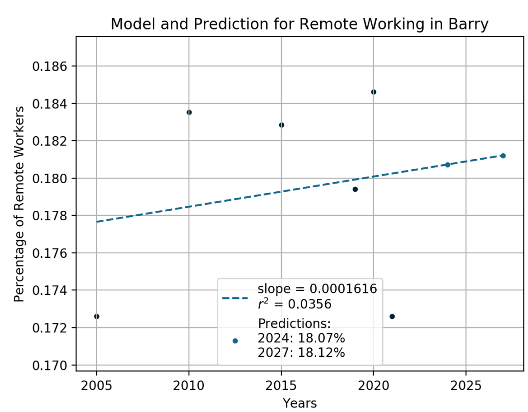

Fig 10. Linear Regression Model for Barry's Remote Workers

4.4 Results

Using our linear regression model, we were able to predict the percentage of remote workers in each city in 2024 and 2027, as outlined in the table below:

City	Predicted Remote Workers in 2024	Predicted Remote Workers in 2027	Projected increase in Remote Workers per year (slope)
Seattle	16.42%	16.65%	0.0007549
Omaha	16.68%	16.83%	0.0004972
Scranton	13.44%	13.48%	0.0001525
Barry	18.07%	18.12%	0.0001616
Liverpool	12.96%	12.74%	-0.0007326

Table 3. Predicted Remote Workers in 2024 and 2027, and Projected Increase per Year

The linear regression model for Liverpool's remote work participation rate may be declining for a plethora of reasons. For one, there may be educated individuals in Liverpool's workforce who would rather relocate to other areas of the world, especially considering the fact that they are now more likely to be allowed to work from home. alternatively, the proportion of jobs that are remote-ready may be lower than other cities given Liverpool's unique economic conditions and primary industries.

4.4.1 Most-Impacted Cities Ranking

Using the predictions from our regression model, we are able to rank the cities by "most-impacted by remote work". We measured impact by the percentage of remote workers in 2024 and 2027. Therefore, the rankings are as follows, with 1 being "most-impacted" to 5 being "least-impacted":

In 2024:

1. Barry, Wales (18.07%)
2. Omaha, NE (16.68%)
3. Seattle, WA (16.42%)
4. Scranton, PA (13.44%)
5. Liverpool, England (12.96%)

In 2027:

1. Barry, Wales (18.12%)
2. Omaha, NE (16.83%)
3. Seattle, WA (16.65%)
4. Scranton, PA (13.48%)
5. Liverpool, England (12.74%)

4.5 Strengths and Weaknesses

Our model for estimating the percentage of remote workers in any given city has the following strengths and weaknesses:

4.5.1 Strengths

1. Our model is robust in that it factors in age and education level into an individual's likelihood of electing to work from home.
2. Our model can easily be applied to different cities by simply changing the city population variable (our model is transferable).
3. Our model is highly unbiased. We analyzed all of our data objectively without trying to selectively choose statistics which matched our predictions.

4.5.2 Weaknesses

1. All the weaknesses from the models defined in Part I and Part II persist.
2. Our model contains discrepancies within its calculations because we included a population of 16-24 year-olds holding postgraduate degrees, despite the fact that this phenomenon is unlikely to occur due to educational convention of the United States.

5. Conclusion

5.1 Further Studies

For clarity of calculation, our models excluded populations such as the currently-unemployed portion of the workforce, or those who want to work remotely but are restricted from doing so by their superiors. In our first model, we did not account for the potential of policy changes in the future, nor did we consider the unemployed workforce. Our second model depends heavily upon the assumption that age, job type, and education level are the only variables that play into one's decision to work remotely. Whereas in real life, there are a multitude of factors pertaining to family, transportation, or personal life that can affect one's decision. Our third model has similar weaknesses as the previous two, as its data is contingent upon variables from the first two models. Given more time, we could refine our models to accommodate for such discrepancies, which would greatly improve its accuracy and reliability.

5.2 Summary

We first predicted the percentage of workers with "remote-ready" jobs in five given cities in 2024 and 2027. We did this by applying an equation involving the number of workers in remote-ready jobs and percentage of remote-ready jobs in a city, over time. Finally, we fit a linear regression model to this data with Python. Our model predicts that Seattle and Omaha are both growing in remote-readiness at +0.146% and +0.0960% per year, respectively; Barry grows slightly at +0.0313% per year; Scranton's changes negligibly at –0.003% per year; Liverpool decreases at –0.142% per year.

Next, to find the percentage of the employed workforce of a specific occupation who, given the option, would elect to work remotely, we considered age, job type, and education level. This led to an equation that can be applied to any job category to determine the desired population. We used a population that works in “Life, Physical, and Social Science”, lives in Seattle, aged between 25-54, and holds a Bachelor’s degree to test this model, finding the percentage of this population choosing to work from home if offered to be 32.35%.

For our 3rd model, we applied our individual model to each member of a city's population, keeping in mind their unique categorical identifiers. By summing up all of these hundreds of combinations we were able to gain an accurate and precise representation of the remote work participation rate in various cities across both the U.S. and U.K. over 15+ years. Recognizing a linear pattern in this time-series data, we were able to deduce the remote work participation rate for future years and rank the cities accordingly. In 2024, from most to least impacted, we determined the ranking to be Seattle, Omaha, Barry, Scranton, and Liverpool.

Ultimately, remote-work is on the rise, and thanks to the pandemic, it is finally being highlighted as an area with potential for growth in many cities. A comprehensive overview and prediction of future trends in remote-working could be greatly beneficial to the analysis of a company’s productivity rates, profitability, and growth in the coming years. With a continued focus and a growing understanding of what the future holds for remote-workers, we can reduce the ecological impact of commutes as well as their cost to employees, reduce the operation costs of firms providing work spaces for employees, and maximize the efficiency and productivity of a firm’s workforce thus increasing a firm's profit margin and stimulating the broader economy. All in all, these predictions have the potential to greatly improve quality of life for both employees and employers.

We thank you for your efforts in maintaining the prosperity of the United States of America and hope that you are able to use this research when you are working towards passing government policies concerning American labor markets.

Sincerely,
M3 Challenge Team #15646

6. References

[1] - Bloom, Nicholas, et al. "Does Working from Home Work? Evidence from a Chinese Experiment." *National Bureau of Economic Research*, Mar. 2013, www.nber.org/papers/w18871.

[2] - Bloom, Nicholas A., et al. "Does Working from Home Work? Evidence from a Chinese Experiment." *Stanford Graduate School of Business*, Mar. 2013, www.gsb.stanford.edu/faculty-research/working-papers/does-working-home-work-evidence-chinese-experiment.

[3] - Maurer, Roy. "Study Finds Productivity Not Deterred by Shift to Remote Work." *Society for Human Resource Management*, 16 Sept. 2020, www.shrm.org/hr-today/news/hr-news/pages/study-productivity-shift-remote-work-covid-coronavirus.aspx.

[4] - Remote Work: Fad or Future? Data, *MathWorks Math Modeling Challenge 2021*, https://m3challenge.siam.org/node/559.

[5] - Ryan, Camille L., and Kurt Bauman. "Educational Attainment in the United States: 2015." *U.S. Department of Commerce*, Mar. 2016, www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf.

[6] - "Statistics on Remote Workers That Will Surprise You (2022)." *Apollo Technical Engineered Talent Solutions*, 16 Jan. 2021, www.apollotechnical.com/statistics-on-remote-workers/.

7. Appendix

Code used for analysis in Part I and Part III.

```
# Created by Jack Zhang of Team-15646 for the 2022 M3 Challenge

# Instructions for using the file:
# In a python environment, compile the file, then run commands "problem1()" and
"problem3()"

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy.optimize

def predict_p1(city='',years=[],percentage=[]):
    # constants for plotting
    c_data = '#012030'
    c_fit = '#13678A'
    c_prediction = c_fit
    size = 10

    # load data
    x = years; y = percentage

    # plot data
    plt.figure()
    plt.clf()
    plt.xlabel('Years')
    plt.ylabel('Percentage of Remote-Ready Jobs')
    plt.title('Model for Remote-Readiness in '+city)
    plt.scatter(x,y,s=size,c=c_data)

    # fit the data
    def model(t,a,b):
        return b+a*t

    popt, pcov = scipy.optimize.curve_fit(model,x,y)
    perr = np.sqrt(np.diag(pcov)) # 1-sigma errors on parameters
```

```
    # find r squared
    residuals = y - model(np.array(x),popt[0],popt[1])
    squared = y - np.mean(y)
    rss = np.sum(residuals**2)
    tss = np.sum(squared**2)
    r = 1 - rss/tss

    # make prediction
    x_predict = np.array([2024,2027])

    prediction = model(x_predict, popt[0], popt[1])
    prediction_text = "\n2024: %.2f%%\n2027: %.2f%%" %
(prediction[0]*100,prediction[1]*100)

    plt.scatter(x_predict, prediction, s=size,
c=c_prediction,label='Predictions:'+prediction_text)

    # plot fit
    x_fit = np.linspace(min(x),max(x_predict),100)
    y_fit = model(x_fit, popt[0], popt[1])
    label = 'slope = %.7f\n' % popt[0]+r'$r^2$'+' = %.4f' % r
    plt.plot(x_fit,y_fit,'--',color=c_fit,label=label)
    plt.legend()
    plt.grid()

    plt.savefig('p1_'+city+'.png')
    return prediction

def problem1():
    years = [2000,2005,2010,2015,2019,2020,2021]

    city = 'Seattle'
    percentage = np.array([28.46,29.20,29.76,29.80,30.69,30.89,32.56])/100
    prediction = predict_p1(city=city,years=years,percentage=percentage)

    city = 'Omaha'
    percentage = np.array([29.77,30.40,31.38,31.74,31.86,31.88,31.65])/100
    prediction = predict_p1(city=city,years=years,percentage=percentage)
```

```
    city = 'Scranton'
    percentage = np.array([25.81,25.61,26.05,26.47,26.13,25.57,25.35])/100
    prediction = predict_p1(city=city,years=years,percentage=percentage)

    years = [2005,2010,2015,2019,2020,2021]
    city = 'Liverpool'
    percentage = [0.273,0.2772,0.2687,0.2565,0.2479,0.2606]
    prediction = predict_p1(city=city,years=years,percentage=percentage)

    city = 'Barry'
    percentage = np.array([33.45,35.57,35.44,34.77,35.78,33.45])/100
    prediction = predict_p1(city=city,years=years,percentage=percentage)

    return

# end problem 1
# ---------------
# begin problem 3

# load data
b=0.94
A = np.array([
    [0.0084,0.0459,0.0123,0.0047],
    [0.0295,0.1618,0.0432,0.0166],
    [0.0314,0.1723,0.046,0.0177],
    [0.03,0.17,0.04,0.02],
    [0.018,0.0989,0.0264,0.0102]
])
age_remote = np.array([
    [17176,2051],
    [94221,27178],
    [25184,6245],
    [9673,2155]
])
nqf_remote = np.array([
    [10414,345],
    [36722,3224],
    [39112,6612],
    [37539,15226],
```

```
    [22468,12223]
])

# xaxis: years; yaxis: population per job category
workforce_data = {'Seattle': np.array([
    [101700,104700,83600,107100,127600,129900,109600],
    [212800,171300,167000,188200,184300,168400,142200],
    [325600,313200,301600,354400,398000,390300,332600],
    [79500,77700,87700,97500,128400,133700,139000],
    [101800,106700,92100,95900,101400,100400,87600],
    [220500,214400,220700,268600,302100,295700,277500],
    [183700,198400,231500,251300,283000,272100,223500],
    [145800,152500,155700,185200,207800,150600,133000],
    [57800,61800,63200,70200,78700,71100,59300],
    [236000,252100,264200,270300,275500,266000,206700]
]),
'Omaha': np.array([
    [23500,25700,20900,25800,30500,30400,30700],
    [35700,32900,31200,32700,33600,33000,33500],
    [108100,99700,94100,98200,96100,91800,94100],
    [15300,13300,11200,11600,10500,9900,9800],
    [35800,37600,40500,42200,46000,45500,44100],
    [60400,61700,63500,73600,73100,70900,71900],
    [55200,61200,71500,76100,79700,78000,79600],
    [41100,42200,43800,48400,52000,43300,47500],
    [14400,16400,17800,18300,18600,17700,18300],
    [55300,59900,65300,65900,66900,65200,65200]
]),
'Scranton': np.array([
    [10700,10600,9400,10200,10500,9800,10300],
    [45600,34900,27800,27000,28600,26900,27200],
    [55600,58500,58900,62600,63500,61900,63900],
    [7000,6300,5000,3500,2900,2600,2500],
    [13700,13400,12400,12600,13100,13000,13000],
    [23000,23400,25000,29800,28300,25500,26100],
    [45300,49100,52200,51900,55200,51500,50500],
    [19000,22000,21800,23300,23500,17800,18200],
    [10000,10000,8300,8500,8800,7500,7700],
    [31200,31700,31700,29400,29000,27900,28300]
```

```
]),
'Liverpool': np.array([
   [141000,138700,138000,150300,153500,146240],
   [80200,73900,80500,100200,107500,103120],
   [92000,109100,128300,146900,145800,146100],
   [59900,68400,66800,72100,73300,73120],
   [22400,21105,22890,22820,20160,25592],
   [41600,39195,42510,42380,37440,47528],
   [29000,24450,22900,23850,21450,23900],
   [69800,59800,66400,64000,69600,69700],
   [70400,78700,73800,80200,75000,73120],
   [29000,24450,22900,23850,21450,26560]
]),
'Barry': np.array([
   [4100,3500,4500,4600,3300,4100],
   [5700,4300,3500,4900,4800,5700],
   [1400,900,1400,800,1200,1400],
   [4000,4400,3800,3900,3600,4000],
   [3045,2940,3710,3535,4095,3045],
   [5655,5460,6890,6565,7605,5655],
   [9700,9800,10550,10850,11550,9700],
   [9500,10800,11200,13000,8000,9500],
   [2400,4000,2800,3700,3100,2400],
   [9700,9800,10550,10850,11550,9700],
])}
RRjob = np.array([0.005,0.010,0.015,0.805,0.880,0.693,0.355,0.340,0.540,0.060])


def ER(x,y,job,year,city):
   """
   helper function for percentage_remote()
   calculates ER values which is used in percentage_remote()
   """

   global b, age_remote, nqf_remote, RRjob, workforce_data

   POPjob = workforce_data[city][job,year]
   ARjob = RRjob[job] * b
```

```
    RWage = age_remote[:,1]/age_remote[:,0]
    per_dist = np.array([i/np.sum(age_remote[:,0]) for i in age_remote[:,0]])
    Wage = per_dist[x]*workforce_data[city][job,year]

    RWnqf = nqf_remote[:,1]/nqf_remote[:,0]
    per_dist = np.array([i/np.sum(nqf_remote[:,0]) for i in nqf_remote[:,0]])
    Wnqf = per_dist[y]*workforce_data[city][job,year]

    out = ARjob*POPjob * (RWage[x]*RWnqf[y])*(Wage+Wnqf)/(Wage*Wnqf)
    return out

def percentage_remote(year,city):
    """
    calculates the percentage of workforce that will go remote per year in the given
city
    """
    global A, workforce_data
    # x is index of age ranges
    # y is index of nqf ranges
    # job is index of job categories
    xs = range(4)
    ys = range(5)
    jobs = range(10)

    chart = np.array([])
    for y in ys:
        row = []
        for x in xs:
            cell = 0
            for job in jobs:
                cell += ER(x,y,job,year,city)*workforce_data[city][job,year]
            cell *= A[y,x]
            # row.append(cell)
            row = np.append(row, cell)
        # chart.append(row)
        chart = np.append(row, chart)
    print('%d:' % year)
    print(chart)
    rr_pop = np.sum(chart) # population who will go remote per year per city
```

```
    work_pop = np.sum(workforce_data[city][:,year]) # total workforce per year per city
    remote_percentage = rr_pop/work_pop # percentage of workforce that will go remote
per year per city
    return remote_percentage

def predict_p3(city):
    """
    predicts the percentage of workforce that will go remote in 2024 and 2027 per city
    """
    # constants for plotting
    c_data = '#012030'
    c_fit = '#13678A'
    c_prediction = c_fit
    size = 10

    if city == 'Seattle' or city == 'Omaha' or city == 'Scranton':
        years = [2000,2005,2010,2015,2019,2020,2021]
    else:
        years = [2005,2010,2015,2019,2020,2021]

    percentage = np.array([])
    for year in range(len(years)):
        percentage = np.append(percentage,percentage_remote(year,city))

    # load data
    x = years; y = percentage

    # plot data
    plt.figure()
    plt.clf()
    plt.xlabel('Years')
    plt.ylabel('Percentage of Remote Workers')
    plt.title('Model and Prediction for Remote Working in '+city)
    plt.scatter(x,y,s=size,c=c_data)

    # fit the data
    def model(t,a,b):
        return b+a*t
```

```
   popt, pcov = scipy.optimize.curve_fit(model,x,y)
   perr = np.sqrt(np.diag(pcov)) # 1-sigma errors on parameters
   # find r squared
   residuals = y - model(np.array(x),popt[0],popt[1])
   squared = y - np.mean(y)
   rss = np.sum(residuals**2)
   tss = np.sum(squared**2)
   r = 1 - rss/tss

   # make prediction
   x_predict = np.array([2024,2027])

   prediction = model(x_predict, popt[0], popt[1])
   prediction_text = "\n2024: %.2f%%\n2027: %.2f%%" %
(prediction[0]*100,prediction[1]*100)

   plt.scatter(x_predict, prediction, s=size,
c=c_prediction,label='Predictions:'+prediction_text)

   # plot fit
   x_fit = np.linspace(min(x),max(x_predict),100)
   y_fit = model(x_fit, popt[0], popt[1])
   label = 'slope = %.7f\n' % popt[0]+r'$r^2$'+' = %.4f' % r
   plt.plot(x_fit,y_fit,'--',color=c_fit,label=label)
   plt.legend()
   plt.grid()

   plt.savefig('p3_'+city+'.png')
   return

def problem3():
   cities = ['Seattle','Omaha','Scranton','Liverpool','Barry']
   for city in cities:
       print('For %s:'%city)
       predict_p3(city)
       print('\n\n')
   return
```