Estimating the Effects of COVID-19 on the Economy and PM2.5 in California

Knowledge Discovery and Data Mining

Srikar Katta
srikar.katta@temple.edu

September 14, 2020

Abstract

Studies of the Coronavirus' impact on the economy and environment have compared observations recorded immediately before the virus to observations after the virus; these comparisons do not take into account the general trend and/or seasonality that may exist. So, in order to derive a better causal estimate, we develop counterfactual estimates of unemployment rate from unemployment insurance filers as a proxy for economic activity and PM2.5, a component of air quality, as a proxy for environmental changes using time series techniques. We evaluated Least Squares Regressions, the Auto-regressive Integrated Moving Average model, Holt Winter's forecasting technique, the Facebook Prophet model, and random forest regression trees to find the best predictor. We then integrated our predicted estimates together to understand the effect of the Coronavirus, considering both the economy and environment side-by-side rather than separately. While we hypothesized that including PM2.5 in our accounting of the virus' impact on the economy would reduce the magnitude of its effect, we discovered that the accounting actually *increased* the costs of the virus.

Acknowledgements: I would like to thank Dr. Kevin Kuruc from the University of Oklahoma for offering initial directions and experimental design construction, Dr. Douglas Webber from Temple University for offering advice on method evaluation, Dr. Viviane Sanfelice for her discussions on techniques in causal inference, Vivek Trivedy for discussing time series prediction techniques, Lee Mai for proofreading offering input on the paper, Omkar Katta for offering insight on the research process, and Dr. Zoran Obradovic for his guidance over the course of the semester and this project.

1 Introduction

The Coronavirus (Covid-19) has shaken America's bedrock. The Covid-19 death toll is greater than the number of deaths during the Vietnam war, education systems are entirely online, and socialization is now a forbidden fruit [8]. It seems as though the virus has impacted everything around us, and the economy and environment have certainly not escaped that stronghold.

After the Coronavirus, the Federal Reserve announced that they will be taking quantitative easing measures to help deal with the potential economic losses we may incur, supply chains have been damaged, and US unemployment insurance claims have seen historic highs [15] [10]. In fact, the International Monetary Fund announced that the global economic impacts of the virus are worse than that of the 2008 Great Recession, indicating a massive economic downturn [11].

While it seems as though everything has worsened, many places around the world have demonstrated improvements in environmental conditions after the Coronavirus. For instance, China's particulate matter concentration has decreased by 20 - 40% since the start of the pandemic [14], carbon monoxide levels in New York City have decreased up to 50% since the virus began, and its methane levels have decreased by 5-10% [26]. Due to the environmental improvements, some have mentioned discussions of curtailing environmental policies in order to boost the economy [35]. This response may be partly due to our treatment of the environment and the economy as disparate concepts. By integrating them into a single account, we might be able to understand the impact of the virus on the economic costs, while ensuring that we keep environmental factors in consideration as we make our decisions. I hypothesized that including PM2.5 in our accounting would decrease the economic impact of the virus.

In order to account for both environmental and economic conditions, we first need to understand the effect of Covid-19 on each, considering a series of historical observations rather than just those recorded the week before the virus. This way, we can differentiate Coronavirus related setbacks and the pre-existing, general trend. In order to understand the effects of COVID-19 related actions on the environment and the economy, we need a "control," an account of what would have happened if the Coronavirus had not impacted communities. In other situations, we might have been able to use techniques in causal inference in which we utilize areas not impacted by this shock to develop a quasi-control (for examples, see [2] and [7]); however, given the wide reach of the pandemic, those natural experimental techniques carry little utility. So, in order to develop this counterfactual, we use time series techniques on our proxy for environmental conditions, particulate matter smaller than 2.5 microns (PM2.5), and the economy, insured unemployment rate, specifically for the state of California.

PM2.5 data was webscraped at the daily frequency for each county in California, except for Yuba, from the Environmental Protection Agency's AirNow page [5]. Insured unemployment rate data was gathered from the Federal Reserve Bank of St. Louis' economic data dashboard, FRED, at the weekly, state level by the Department of Labor for their unemployment insurance claims report [32]. We evaluated least squares regressions, the Facebook Prophet model [36], random forest regressions, autoregressive integrated moving average models, and Holt Winter's exponential smoothing method to find the best predictor.

Based on our analyses, we discovered that a least squares regression predict unemployment rate best, and that the Facebook Prophet predicts PM2.5 better than other approaches using root mean squared error as our evaluation criterion. We also discovered that the percent difference between actual and counterfactual average PM2.5 concentrations after the Coronavirus ranged from about -0.3% to about 0.1%. With respect to economic losses, we discovered that there exists a maximum weekly unemployment rate difference of about 6%, which translates to a maximum weekly GDP change of about \$700,000 after the Coronavirus. Including PM2.5 concentrations in our accounting indicated that there is actually a decrease in state GDP, the maximum difference now being \$1.3 million.

2 Related Work

2.1 The Economy After the Coronavirus

The Coronavirus has deeply hurt the economy, impacting small businesses, international supply chains, and financial markets. At the region level, manufacturing indexes from the Federal Reserve Banks of Philadelphia, Dallas, and New York all demonstrated the lowest readings they have seen since the Great Recession, suggesting a steep decline in current economic activity, which could have major ramifications for the economy in the future [28] [31] [30]. In fact, economists claim that that US GDP is poised to decline by up to 5% for each month we remain in lockdown and by 25% over the next two quarters [25] [27]. While I did not aim to estimate future GDP changes - only nowcasts - my calculations suggest that California state GDP decreased by about 21%, which lay within other's economic projections for the entire country.

2.2 The Environment After the Coronavirus

Even though the economy is drowning and a significant portion of Americans are left unemployed, possible improvements to air quality seem as though there may be a silver lining. According to the European Space Agency, pollution levels across Europe and Nitrogen Dioxide levels have dropped drastically after the institution of lockdowns [3] [4]. However, these estimates do not capture the possible pre-existing trends associated with PM2.5 concentrations - only a change after the shock, which makes it difficult to disassociate whether the changes are a result of air quality's seasonal patterns or due to the Coronavirus; while inverse modeling and several different data types are necessary to estimate changes in chemical composition, a computationally expensive and long process [3], we can hope to estimate the average treatment effect of the institution of Coronavirus-related policies by comparing the observed concentration levels to a counterfactual measure of the concentrations had the virus not occurred.

2.3 Forecasting Methods

In order to develop our counterfactual, we looked to time series methods. With respect to forecasting PM2.5 concentrations, the EPA suggests several techniques, including tree based methods building off of the Classification and Regression Tree algorithm (CART), regression based approaches, and Auto-regressive Integrated Moving Average model (ARIMA) [5]. Additionally, Mahajan et al. suggest using exponential smoothing methods, such as Holt Winter's algorithm [22] [21]. Furthermore, others have demonstrated that the Facebook Prophet model is also an effective estimator of PM2.5 predictions [34] [38]. While the EPA suggests that the ARIMA model forecasts better than regression and CART based approaches [5], I found that the Facebook Prophet model predicts PM2.5 concentrations better than other methods that we tested, using out of sample root mean squared error as my evaluation criterion, validating Samal et al's findings [34].

Past work pertaining to predicting unemployment rates generally uses non-auto-regressive features. For instance, some have used Google trends and search reports to nowcast unemployment statistics [33] while others have used surveys [9]. However, due to the wide impact of the Coronavirus, using non-unemployment rate variables might introduce bias in creating our counterfactual, so we must rely on auto-regressive techniques. A common, simple, and strong method, the least squares regression is the foundation of many time series methods. One such extension is the the auto-regressive integrated moving average (ARIMA) technique to predict unemployment rates [23] [33]. Others have demonstrated that exponential smoothing approaches, another model that builds of the least squares regression, are strong methods for forecasting unemployment rates [13].

2.4 Using the Labor Market as an Indicator for the Economy

Because unemployment moves very closely with GDP, labor market data is consistently used as as an indicator for the economy. The Federal Reserve Bank of Philadelphia even announced that "The monthly unemployment rate is typically the most influential factor underlying a state's coincident index for most states," suggesting a strong association between unemployment rates and state GDP levels [1]. Furthermore, building off of Okun's law, which postulates that Gross National Product (GNP) and unemployment are strongly and inversely correlated, economists from the Federal Reserve System demonstrate this relationship continues exists at the state level, providing strong evidence for using unemployment rates as a method of tracking the economy [16].

2.5 Integrating Economic and Environmental Variables

Environmental economists have formulated methods of integrating PM2.5 concentrations into accounts of economic production using the public health costs that may come as a result of PM2.5, such as medical costs and the value of a statistical life [37] [24]. A novel technique for estimating the social costs of PM2.5 specifically for America, subset at the county level, offers a unique method of translating emissions into dollar amounts in this study [17]. While this has not been used to study the economic and environmental changes as a result of Covid-19, which is relatively limited, past work does provide some grounding for using this method. In the context of COVID-19, the Energy Policy Institute at the University of Chicago has weighed PM2.5 emissions and energy consumption, a historically strong predictor of economic activity, and PM2.5 levels nationally, internationally, and for select US states, including San Francisco, for which they discovered an increase in particulate pollution as compared to December 2019 [6]. In comparison with my results, in San Fransisco county, I discovered that there existed a -13.8% percent difference between the actual concentration and my predicted values.

3 Methods

3.1 Data

In order to capture changes in air quality, I web scraped daily concentration levels for particulate matter smaller than 2.5 microns per meters cubed (PM2.5) dating back to 2002 for each California county except for Yuba from the Environmental Protection Agency's AirNow page [5]. Missing data was imputed with a k-nearest neighbor regression using date, latitude, and longitude as my features with a k of 5. In order to capture economic trends, I used the Department of Labor's California-level, non-seasonally adjusted, weekly unemployment insurance, unemployment rate (UR) data and the Bureau of Economic Analysis' quarterly, seasonally adjusted state GDP data from the Federal Reserve Bank of St. Louis' FRED database [32] [29]. The data had no missing values to impute.

3.2 Data Preparation

For my PM2.5 data, I added 90, 180, 270, and 360-day lagged terms as possible features and removed missing values, shifting my date interval from January 1st, 2002 to January 1st, 2003. For my UR data, I added similar lagged terms, but at the weekly level - 13, 26, 39, and 52 week lags - shifting the start date of my data from August 1st, 1987 to July 30th, 1988. For model evaluation, both sets of data were separated by a 90/10, training-validation split for dates until March 14, 2020 (the date I used as the start of the virus). Additionally, I used out-of-sample root mean squared error (rmse) to compare model fits:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}},$$

where y_i, $\hat{y}_i$, and n represent the true values, the predicted values, and the number of observations respectively.

3.3 Forecasting Techniques

In order to establish a baseline for model evaluation, we used a naïve forecast, imputing the last recorded value of our training set as the prediction for all of our test cases. From there, we applied Holt Winter's forecasting technique for both UR and PM2.5, simple least squares regression with lagged terms for UR prediction, a regression model fixed for county and week of year for PM2.5 predictions, an Auto-regressive Integrated Moving Average model (ARIMA), classification and regression tree (CART) based approaches, and the Facebook Prophet model.

3.3.1 Holt Winter's Forecasting Technique

Holt Winter's forecasting technique is an exponential smoothing based approach to forecasting that considers three components: trend, seasonality, and an average, each of which are expressed as portrayed as three types of exponential smoothing. Using historical data, the algorithm predicts future values. In the following equation, let $x_t, y_t,$ and s_t denote overall smoothing, trend, and seasonality respectively. We can then represent Holt Winter's as follows:

$$\begin{aligned}
x_t &= \alpha(n_t - x_{t-T} + (1-\alpha)(x_{t-1} + y_{t-1}) \\
y_t &= \gamma(x_t - x_{t-1}) + (1-\gamma)y_{t-1} \\
s_t &= \beta(n_t - x_t) + (1-\beta)s_{t-T} \\
n_{t+h} &= x_t + hy_t + x_{t-T+h},
\end{aligned}$$

where n refers the number of historical observations, α, β, γ are parameter values for smoother forecasting and the subscripts denote time values. The three components, x_t, y_t, s_t are used to then predict h steps forward, denoted as n_{t+h} [21].

Given the seasonal hiring and firing patterns, I believed that using Holt Winter's method would be a method giving strong considering for our counterfactual prediction of UR. Several papers have used exponential smoothing based forecasting approaches to predict PM2.5, which have found decent prediction results [22] [21]; given the precedence in the field, and the seasonality associated with PM2.5 concentrations, I believed it would be a strong forecasting technique.

3.3.2 Least Squares Regression

Regression as a forecasting tool is one of the simplest, yet most effective time series methods that exist. By minimizing the sum of squares, regressions reduce bias and offer a strong prediction. It is also the framework for more integrated time series approaches, such as the ARIMA model.

In our prediction of UR, we used an auto-regressive model that considered lagged values of 13, 26, 39, and 52 weeks and a dummy variable indicating week of year to capture seasonal changes, as follows:

$$UR_t = \beta_0 + \beta_1 UR_{t-13} + \beta_2 UR_{t-26} + \beta_3 UR_{t-39} + \beta_4 UR_{t-52} + \beta_5 \mathbb{1}week_t + \varepsilon.$$

In our prediction of PM2.5, we used another auto-regressive model consisting of 90, 180, 270, and 360-day lagged terms. This regression forecast also used week-of-year dummy variables in order to capture any weekly seasonality. Furthermore, because each county is different from one another, I used a fixed effects model that controlled for any county level differences that may occur while still considering the trend for all counties:

$$PM2.5_t = \beta_0 + \beta_1 PM2.5_{i,t-90} + \beta_2 PM2.5_{i,t-180} + \beta_3 PM2.5_{i,t-270} + \beta_4 PM2.5_{i,t-360} + \beta_5 \mathbb{1}week_t + \varepsilon_t,$$

where the subscript i denotes each county and t relates to time index. This integration of entity level and group level patterns allows regressions to be a particularly powerful tool for longitudinal forecasting. While the EPA proposes regressions as a useful technique for forecasting PM2.5 concentrations, they also recommend incorporating other variables, such as precipitation and temperature to build a stronger prediction. Here, we can not include outside features because, if they were impacted by the Coronavirus, which I am assuming they are, they would bias our counterfactual and suggest that the control we are developing is more similar to the actual values than they truly are. Furthermore, the EPA's regression recommendation outlines the use of $t-1$ through $t-k$ lagged terms because of higher auto-correlation, rather than $t-90$ and beyond; when we move this far back, we lose that predictive reliability. But predictions including $t-1$ terms would involve using values that happened after the Coronavirus, we decided to use terms that were about one season away.

3.3.3 Auto-regressive Integrated Moving Average (ARIMA)

The ARIMA model, a widely used time series technique, considers three components: an auto-regressive structure, an integrated structure, and a moving average structure. The auto-regressive piece of ARIMA allows the forecaster to use historical data to predict future values, while the moving average element smooths the data to allow for better forecasting. In order to stationarize our data (i.e., have a constant mean and variance over time), we can difference our data until we remove any trend, which builds the integrated element of ARIMA. The order of these three pieces are denoted by p, q, and d, respectively. Let X_t be a time series such that t denotes the time index. Then, we can write an ARIMA(p,d,q) model as follows:

$$(1-B)^d(1-\sum_{i=1}^{p}\phi_l B^l)X_t = (1+\sum_{i=1}^{p}\theta_l B^l)\varepsilon_t,$$

where B refers to backward shift operator, ϕ is the auto-regressive parameter and θ_l is the moving average parameter, and ε_t is the error [21]. In order to find p, d and q, I used the forecast package in R, which grid searches to find the best p and q parameters that minimizes the Akaike Information Criterion (AIC) after differencing the series until it is stationary, tested by the Augmented Dickey Fuller test [18].

The ARIMA is a strong linear predictor method whose results are easily interpreted. The EPA also suggests the ARIMA method and considers it to be a better predictor of PM2.5 than regression based approaches [5].

The seasonal ARIMA (SARIMA), an extension of ARIMA, allows us to consider seasonal auto-regressive and moving average parameters (whose orders are denoted by P and Q respectively) that might violate the stationarity assumption we make when using the regular ARIMA model. The seasonal auto-regressive (P) and moving average (Q) components of SARIMA can be expressed as

$$\Phi_P(B^s)Z_t = \Theta(B^s)\varepsilon_t, \tag{1}$$

where s represents the number of seasonal periods. Additionally, the seasonal difference of order D, which we use to remove seasonality, can be written as

$$\forall_s^D Z_t = (1-B^s)^D Z_t.$$

Integrating the three components of SARIMA, the final structure, $SARIMA(p,d,q)(P,D,Q)_s$, can be formed as

$$\Phi_P(B^S)\phi(B)\forall_s^D\forall^d Z_t = \Theta_Q(B^s)\theta(B)\varepsilon_t,$$

where $\forall_s^D\forall^d Z_t$ is an ARIMA(p,0,q) model with many coefficients statistically not different than 0 [12]. Because of the seasonality associated with unemployment rates, SARIMA seemed to be a strong model to evaluate; however, the one downside is that SARIMA is much more computationally time consuming than the ARIMA model. While forecasting for one state is relatively quick, developing separate seasonal models for PM2.5 would be incredibly time consuming, especially since other researchers have used the non-seasonal ARIMA to forecast PM2.5 [22] [20].

3.4 Random Forest and Tree Based Methods

Classification and regression tree (CART) algorithms function by splitting "nodes" - the features of our dataset - based on a splitting criteria, for which we used mean squared error, in order to reduce variation in our data and predict our outcome variable based on patterns in the training data [5]. Because CART methods can capture periodicity and non-linearities in data, which are both very evident in PM2.5 data, I believed that they would be a strong method to consider. However, CART has a tendency to overfit the training data [5], so I used the random forest ensemble approach to minimize predictive error. Because we are working with time series and there exists temporal continuity in our data, we can not bootstrap random samples from our dataset in order to develop separate trees. So, we created a random forest ensemble using the entire dataset with mean squared error as our evaluation criteria.

3.5 Facebook Prophet Model

The Facebook Prophet model is a newer time series technique that utilizes a Bayesian state space in order to predict future values. Similar to Holt Winter's, the Prophet model inherently accounts for trend and seasonality, but it also considers holidays and other events that may be associated with changes in time series. The model can be represented mathematically as

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t,$$

where $g(t)$ models the non-periodic trends in the time series, $h(t)$ models irregularly scheduled holidays, and $s(t)$ models periodic/seasonal trends using a Fourier series transform, which can be written as follows:

$$s(t) = \sum_{n=1}^{N} a_n cos(\frac{2\pi nt}{p}) + b_n sin(\frac{2\pi nt}{p}),$$

where p represents periodicity [36] [34]. Because of the apparent seasonality in our data and Prophet's easy implementation, Prophet seemed like a wise choice in considering the creation of our counterfactual.

3.6 Translating Unemployment Rate and PM2.5 to Dollar Amounts

Because UR and PM2.5 are measured on different scales, integrating the two into a single model while keeping both in the same units could be difficult. So, I translated UR to US dollars using seasonally adjusted, California quarterly GDP by performing the following regression:

$$ln(GDP_t) = \beta_0 + \beta_1 t + \beta_2 MedianUR_t$$

where $MedianUR$ represents the median unemployment rate for that quarter. I then used the true unemployment rate to predict GDP changes for that month, which I return to a dollar scale using an exponential transformation of base e. While this is an experimental translation technique that I have not seen used in the literature, with a strong model fit (R^2 = 0.982), there is statistical evidence for this translation. In essence, this translation of UR to GDP is similar to a regression-based imputation method for missing-data.

In order to translate PM2.5 into dollar values, I applied a novel and easy tool to estimate the social/public health costs associated with emissions - The Estimating Air pollution Social Impact Using Regression model (EASIUR) [17]. After forecasting the counterfactual concentration level, I took the difference between the predicted and true values, converted the difference from $\mu g/m^3$ for each county to metric tons and then converted it to a dollar amount. In order to account for PM2.5 benefits with UR data, I took a sum of the EASIUR estimates for each week.

4 Results

4.1 Evaluation of Time Series Techniques for PM2.5

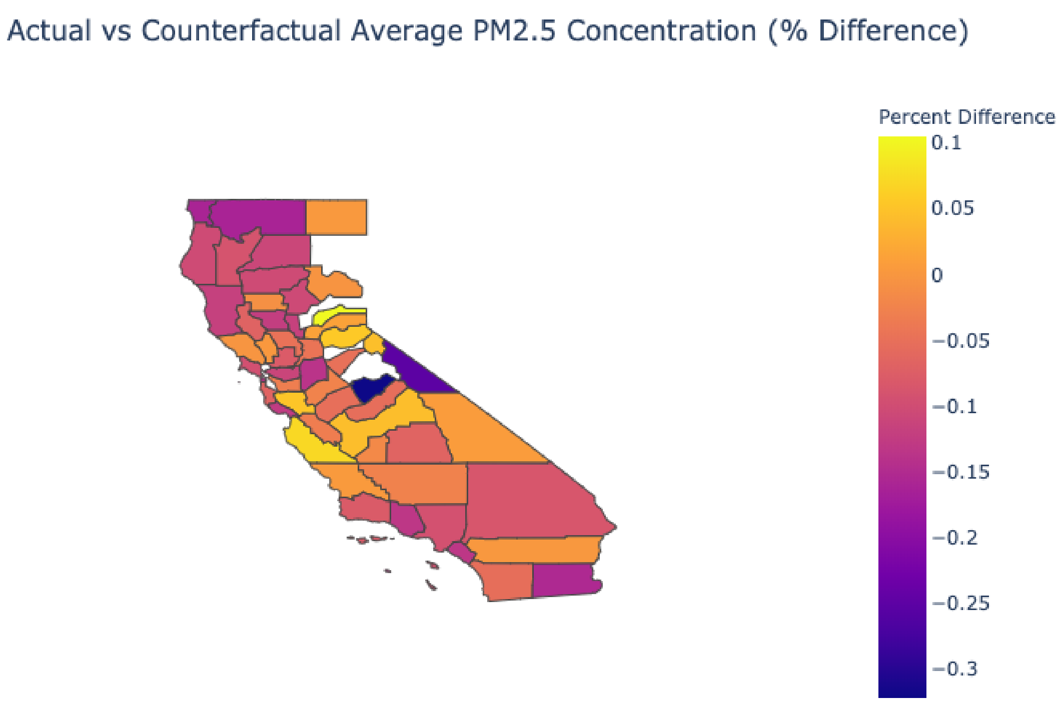
From our naïve forecast, we establish the baseline for evaluating our forecasting techniques: an out of sample rmse of 11.87. A good time series model should perform much better than a simple imputation, considering the high seasonality in PM2.5 concentrations. From our evaluations, we notice that the Fixed Effect and Random Forest models have a lacking performance with rmse's of 12.35 and 14.28. However, most of our other techniques perform better than our baseline, by a slight margin. Our ARIMA models had an average rmse of 11.67, Holt Winter's exponential smoothing method had an average rmse of 11.4, and the Prophet model had an average rmse of 11.26, indicating that the Prophet model would be the best method to develop our counterfactual.

Average RMSE over 634 Test Cases

Method	**RMSE**
Naive	11.87
Random Forest	14.28
Prophet	11.26
Holt Winter's	11.40
Fixed Effect	12.35
ARIMA	11.67

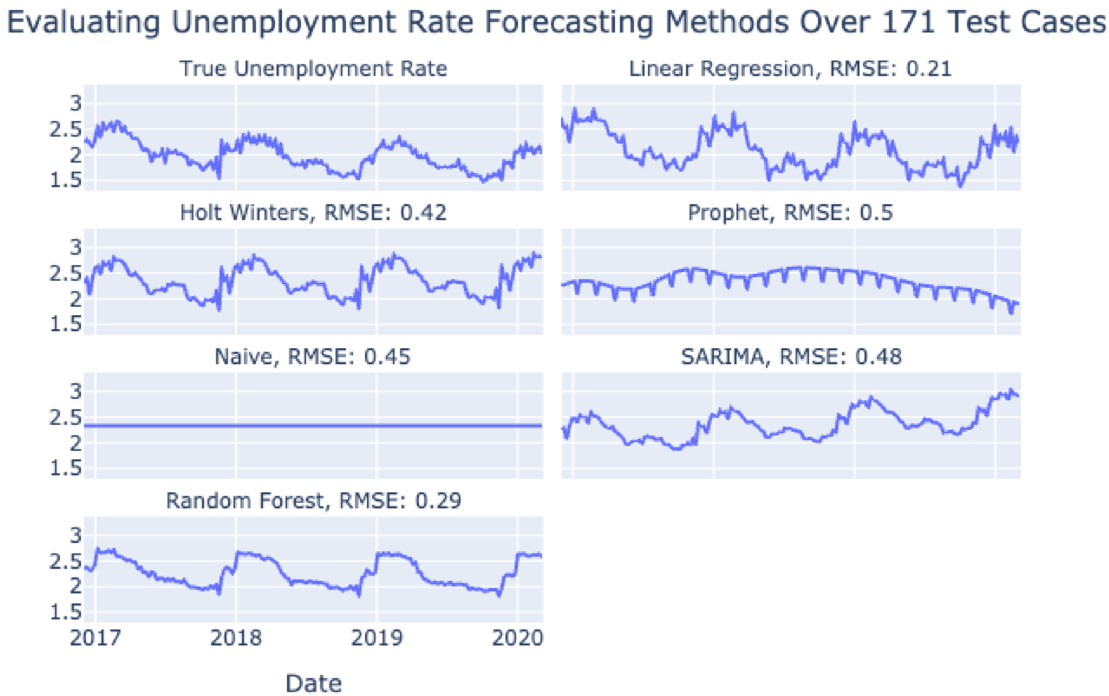
4.2 Impact of Coronavirus on Air Quality

Assuming the predicted PM2.5 values are the true counterfactual, the percent difference between the average PM2.5 concentrations range from -0.3 to 0.1 *percent*, suggesting that concentrations changed only slightly after the Coronavirus. Additionally, in several counties, concentrations were actually noted to have increased. PM2.5 generally improved in the outer regions of the state, while they maintained/increased concentration closer to the center.



4.3 Evaluation of Time Series Techniques for Unemployment Rate

From our naïve forecast, we established that a "poor" forecasting technique would have a root mean squared error (rmse) less than or equal to 0.45 over our 171 test cases. Surprisingly, the Facebook Prophet model had a rmse of 0.5, suggesting that it was a very poor choice to use. Similarly, the $SARIMA(21,0,4)(014)_52$ model had a rmse of 0.48, which was worse than the naïve prediction; even though it was able to capture the length of the seasonal trend, it was still unable to identify the proper direction of the trend. However, Holt Winter's method did do slightly better than the naïve regression with a rmse of 0.42. By visual inspection, we can see that it captured the seasonal movement of unemployment rate relatively well, but it was unable to determine the proper magnitude of the troughs and peaks. Additionally, its trend was not as downward sloping as the true data's. Next, the random forest regression was a relatively good predictor of unemployment rate with a rmse of 0.29, but it too was unable to capture the troughs and peaks properly; it overshot both. Lastly, I found that the least squares method was the best predictor of UR with a rmse of 0.21. By visual inspection, even though it slightly overshoots the true claims, it still generally trends with the data and seasonal pattern, which none of the other methods did very well.



4.4 Validity of Translating Unemployment Rate to GDP

From the regression table, we notice that unemployment rate is statistically significantly related to GDP, suggesting that unemployment might be a viable metric to use as a proxy for weekly GDP. Additionally, with an R^2 of 0.982 and F-statistic of 1596, much of the variance in $log(GDP)$ is captured, demonstrating the statistical validity of using unemployment rate to capture weekly changes in GDP. From the regression, we notice that a unit change in the unemployment rate relates to an approximately four percent decrease in GDP.

Dep. Variable:	log(GDP)	**R-squared:**	0.982
Model:	OLS	**Adj. R-squared:**	0.982
Method:	Least Squares	**F-statistic:**	1596
Log-Likelihood:	139.56	**Prob (F-statistic):**	9.11e-51
No. Observations:	60	**AIC:**	-273.1
Df Residuals:	57	**BIC:**	-266.8
Df Model:	2		

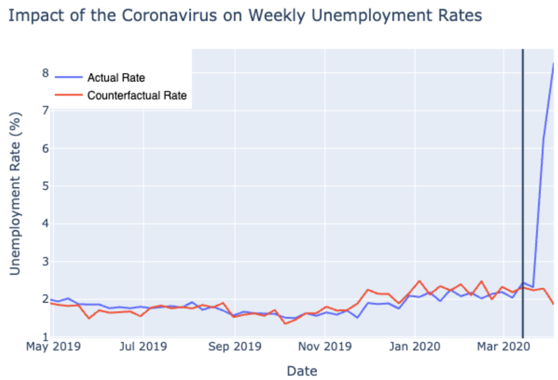
	coef	**std err**	**t**	**P>** $\|$**t**$\|$
Intercept	13.8643	0.024	571.043	0.000
Unemployment Rate	-0.0376	0.003	-11.044	0.000
Time	0.0007	1.49e-05	47.434	0.000

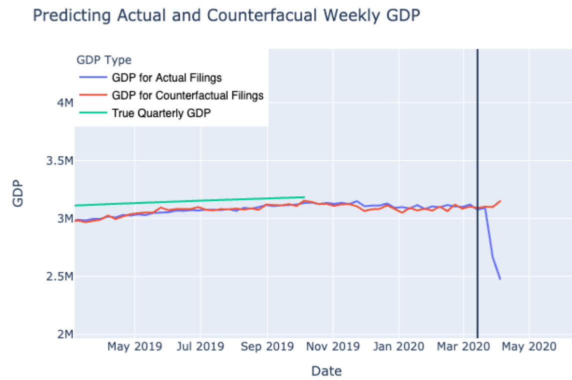
4.5 Economic Impact of the Coronavirus

Using a linear regression, we predicted the counterfactual unemployment rate. Based on the counterfactual forecast, the unemployment rate was set to trend downwards already. However, the counterfactual-factual comparison demonstrates that there was an approximately 8.5% increase in unemployment rates. From the regression of $log(GDP)$ on unemployment rate, that suggests an approximately 24% decrease in GDP.

When inspecting the impact of the Coronavirus on GDP, we notice that there is an approximately 21% difference between the counterfactual and actual GDP rates, a drop from the estimated \$3.15 million to the estimated \$2.48 million. This significant decrease pushes California state GDP back to about 2015 levels, a five year setback. To place into perspective, Dr. Kwasnicki from the University of Wroclaw details that the Great Recession regressed America's 2008 GDP back to its earlier 2004/2005 levels, and America took the better part of a decade to return to its normal economic conditions [19].

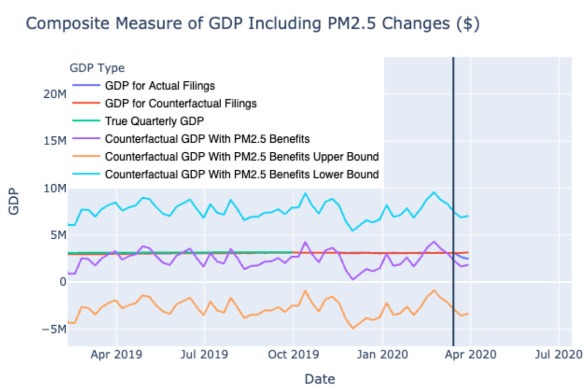
If America's GDP resembles California, America may have a difficult time saving itself from the looming recession.

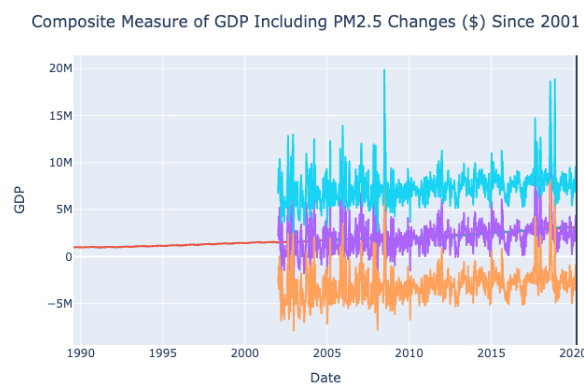




4.6 Integrating PM2.5 Into Our Economic Accounting

After accounting for changes in PM2.5 concentrations in our understanding of the Coronavirus' impact, we notice that there is in fact a significant decrease in GDP, refuting our hypothesis that including PM2.5 changes into our calculations of economic changes might mitigate the economic impact of the Coronavirus. However, this may be attributable to the Prophet's inability to capture minute PM2.5 differences. In the "Composite Measure of GDP Including PM2.5 Changes ($) From 2001" plot, we can see that including PM2.5 into our economic accounting calculations is in fact a very noisy process, and the decreased GDP we saw from incorporating PM2.5 after the Coronavirus might in fact be due to the volatility in estimated PM2.5 changes rather than true PM2.5 changes. Even when we include the upper and lower bounds for our predictions, we notice that the wide range of possibilities, which indicates that this technique may not be a strong estimator.

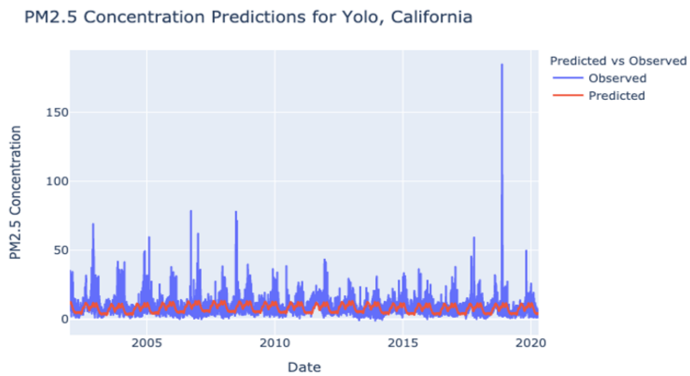




5 Discussion

Based on our results, integrating PM2.5 into the accounting of economic conditions suggests that the economic impact of the Coronavirus is even greater than we initially suspected. Additionally, the change in PM2.5 concentrations in California, using this methodology, suggests that the Coronavirus' did not impact air quality in California counties much, evidenced by a range of concentrations changes of -0.3% to 0.1%. Additionally, this change was not as drastic as those found in traditionally, incredibly polluted cities like New York and countries like Beijing, whose environmental improvements have been dominating the media.

Our methodology in predicting a counterfactual measure of both the economy and air quality has several flaws. For instance, even though the Prophet model had the lowest rmse of the models tested, in order to capture the true counterfactual, we would really need to consider prediction models with almost no error in order to make strong claims. While we may have calculated that incorporating PM2.5 would actually increase the costs as a result of the virus, these changes may be due to the Prophet model's inability to capture the intricacies of PM2.5 concentrations.



While we witnessed the Prophet model's relatively strong ability to predict the general trend and seasonal components, it still could not capture PM2.5's slight variations. Due to the highly non-linear structure of PM2.5 time series data, other methods that can capture this variation may be better suited for the task of creating a counterfactual. However, considering the wide use of these methods in other studies, these approaches seemed appropriate. We even observed the impact of poor model fit in our evaluation of unemployment rates when comparing models, highlighted by the regression's incapacity to fit the data perfectly. However, the tremendous upward shock in the unemployment rate after the Coronavirus is quite obvious and outweighs the model's inability to capture detailed changes for our discussions. An issue with the economic aspect of this study, however, is that unemployment insurance rates do not take into account unemployed gig economy workers, whose contibutions are far from insignificant. So, we may actually be under-predicting the true economic impact of the Coronavirus.

Another shortcoming of this methodology is that we were unable to control for exogenous shocks to both the economy and PM2.5 that may have occurred around the time of the Coronavirus since disassociating the source of these effects would be incredibly difficult. For instance, suppose Norhtern California, the area in which we witnessed improvements in air quality, consistently has wildfires every March. Now, also assume that, this year, there were no wildfires that may have caused a worsening in air quality. Perhaps the wildfire did not happen because people stayed at home and did not go camping, reducing the risk of accidental, human-induced fires. Then, this reduction of camping would be a byproduct of the Coronavirus and would be fair to include in our causal estimates. However, if there was no wildfire because of a new policy instituted at the beginning of March 2020, this method would have no way of disassociating the source of the improvement in air quality. That being said, the possible endogeneity in our model is understandable considering the wide reach of the Coronavirus limiting the use of other control variables. Discovering geographies unimpacted by the Coronavirus, directly or indirectly, is almost inconceivable, so it would be difficult to use a synthetic control or differences in differences approach. Furthermore, finding strong instruments not impacted by the virus would also be almost infeasible considering that every facet of the world itself is different.

The limitations in this study's design offer opportunities to further our understanding of the relationship between air quality, economic accounting, and the Coronavirus. First, by studying if accounting for other components of air quality, such as Nitrogen Dioxide, also lead to a similar directional shift in our economic impact evaluations, offering us a greater comprehension of the interaction between economic and air quality accounting after the Coronavirus. Furthermore, by using other methods that can better capture non-linearity in PM2.5 time series, such as convolutional neural networks, we can hope to develop stronger counterfactuals. Understanding better methods of building auto-regressive counterfactuals would extend the field of causal inference, allowing us to study situations ripe for quasi-experiments where the impacts of the shock are so wide reaching that synthetic controls are difficult to discover.

We were motivated to understand whether incorporating environmental changes as a result of the Coronavirus can help us understand the true impact of the virus. While we were unable to validate our hypothesis that including PM2.5 would decrease the perceived impact of the Coronavirus on the economy, it is apparent that we make improvements to our understanding of the virus' total impact and acknowledge that our economic and environmental needs must be considered together. Since air quality in California did not improve as greatly as others have suggested, there is little justification for the claim that, because of the perceived benefits in the environment, we can set environmental policies to the side in the hopes of boosting the economy. If anything, the lack of a significant change in air quality should open our eyes to the state of environmental conditions. We must make improvements to our environment, and we can not continue to place our economic needs in front of our environmental needs: they must go hand in hand.

References

[1] *Federal reserve bank of philadelphia*, Federal Reserve Bank of Philadelphia, (2020).

[2] A. Abadie and J. Gardeazabal, *The economic costs of conflict: A case study of the basque country*, American Economic Review, 93 (2003), pp. 113–132.

[3] E. S. Agency, *Coronavirus lockdown leading to drop in pollution across europe*, Mar 2020.

[4] ——, *Covid-19: nitrogen dioxide over china*, Mar 2020.

[5] U. S. E. P. Agency, *Outdoor air quality data*, Nov 2018.

[6] F. Burlig, S. Cicala, and A. Sudarshan, *Covid-19*, Apr 2020.

[7] D. Card and A. B. Krueger, *Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania*, The American Economic Review, 84 (1994), pp. 772–793.

[8] N. Chavez, *Us coronavirus cases surpass 1 million as projections show deaths could rise in coming weeks*, Apr 2020.

[9] O. Claveria, *Forecasting the unemployment rate using the degree of agreement in consumer unemployment expectations*, Journal for Labour Market Research, 53 (2019).

[10] M. Corkery and D. Yaffe-bellany, *U.s. food supply chain is strained as virus spreads*, Apr 2020.

[11] I. M. F. R. Dept., *World Economic Outlook, April 2020 : The Great Lockdown*, INTERNATIONAL MONETARY FUND, USA, 2020.

[12] C. Dritsaki, *Forecast of sarima models: Š¡n application to unemployment rates of greece*, American Journal of Applied Mathematics and Statistics, 4 (2016), pp. 136–148.

[13] K. Dumicic, A. . asni, and B. Žmuk, *Forecasting unemployment rate in selected european countries using smoothing methods*, 2015.

[14] E. C. for Medium-Range Weather Forecasts, *Amid coronavirus outbreak: Copernicus monitors reduction of particulate matter (pm2.5) over china*, Apr 2020.

[15] D. Goldman, *Federal reserve cuts rates to zero to support the economy during the coronavirus pandemic*, Mar 2020.

[16] A. Guisinger, R. Hernandez-Murillo, M. Owyang, and T. Sinclair, *A state-level analysis of okun's law*, Working paper (Federal Reserve Bank of St. Louis), (2015).

[17] J. Heo, P. J. Adams, and H. O. Gao, *Reduced-form modeling of public health impacts of inorganic pm 2.5 and precursor emissions*, Atmospheric Environment, 137 (2016), p. 80–89.

[18] 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*, 2020. R package version 8.12.

[19] W. Kwasnicki, *Was the us recovery from the great recession a success?: Witold kwasnicki*, Sep 2019.

[20] D. Liu and L. Li, *International journal of environmental research and public health application study of comprehensive forecasting model based on entropy weighting method on trend of pm 2.5 concentration in guangzhou, china*.

[21] S. Mahajan, L. Chen, and T. Tsai, *An empirical study of pm2.5 forecasting using neural network*, in 2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2017, pp. 1–7.

[22] S. Mahajan, L.-J. Chen, and T.-C. Tsai, *Short-term pm2.5 forecasting using exponential smoothing method: A comparative analysis*, Sensors, 18 (2018), p. 3223.

[23] U. Mahmudah, *Predicting unemployment rates in indonesia*, Economic Journal of Emerging Markets, 9 (2017), pp. 20–28.

[24] K. J. Maji, W.-F. Ye, M. Arora, and S. S. Nagendra, *Pm2.5-related health and economic loss assessment for 338 chinese cities*, Environment International, 121 (2018), p. 392–403.

[25] C. Makridis and J. Hartley, *The cost of covid-19: A rough estimate of the 2020 us gdp impact*, SSRN Electronic Journal, (2020).

[26] M. McGrath, *Coronavirus: Air pollution and co2 fall rapidly as virus spreads*, Mar 2020.

[27] J. MORGAN, *Fallout from covid-19: Global recession, zero interest rates and emergency policy actions*, Mar 2020.

[28] F. R. B. OF DALLAS, *Texas manufacturing activity contracts suddenly, outlook worsens*, Mar 2020.

[29] U. B. OF ECONOMIC ANALYSIS, *Total gross domestic product by industry for california [canqgsp].*

[30] F. R. B. OF NEW YORK, *Empire state manufacturing survey*, Apr 2020.

[31] F. R. B. OF PHILADELPHIA, *April 2020 manufacturing business outlook survey*, Apr 2020.

[32] F. R. B. OF ST. LOUIS, *U.s. employment and training administration, insured unemployment rate in california [cainsuredur], retrieved from fred.*

[33] J. PAVLICEK AND L. KRISTOUFEK, *Nowcasting unemployment rates with google searches: Evidence from the visegrad group countries*, PLOS ONE, 10 (2015), pp. 1–11.

[34] K. K. R. SAMAL, K. S. BABU, S. K. DAS, AND A. ACHARAYA, *Time series based air pollution forecasting using sarima and prophet model*, in Proceedings of the 2019 International Conference on Information Technology and Computer Communications, ITCC 2019, New York, NY, USA, 2019, Association for Computing Machinery, p. 80–85.

[35] B. STORROW, *5 ways the economic upheaval of coronavirus may impact co 2 emissions*, Apr 2020.

[36] S. J. TAYLOR AND B. LETHAM, *Forecasting at scale*, PeerJ, (2017).

[37] Y. XIE, H. DAI, Y. ZHANG, T. HANAOKA, T. MASUI, AND Y. WU, *Health and economic impacts of ozone pollution in china: a provincial level analysis*, Environment International, 13 0 (2019), p. 1–63.

[38] N. ZHAO, Y. LIU, J. VANOS, AND G. CAO, *Day-of-week and seasonal patterns of pm2.5 concentrations over the united states: Time-series analyses using the prophet procedure*, Atmospheric Environment, 192 (2018), pp. 116–127.