A Multi-factor Model of COVID-19 Epidemic in California

Ted G. Lewis
Naval Postgraduate School (ret).
tedglewis@icloud.com

Abstract

We describe a multi-factor model of the spread of COVID-19 across the 58 counties of California from March 2020 to June 2023. The model provides estimates of cumulative cases and duration of the epidemic versus 5 independent variables. The independent variables are the following correlated factors: population, population density, family income, Gini coefficient, and land area (size) of each county. The correlation coefficients of these factors are used to reduce the error in models of cumulative cases and duration.

The model produces linear equations – one for cumulative cases and the other for duration of infection. The cumulative case estimate is highly correlated with population, but the estimate is improved by considering all 5 factors. The duration of infection[1] estimate is improved by considering population and income level. We also find that infection rate (per capita cases) varies highly and roughly obeys a normal distribution, suggesting randomness, rather than correlation with one or more of the 5 factors.

We also observe the vast differences between high and low cumulative cases across the 58 counties. Using the same model but with different values of correlation of the same factors, we obtain very accurate models of both bottom-half and top-half counties. i.e., counties with 49,000 or fewer cumulative cases versus counties with 50,000 – 3.5 million cases. This suggests that multi-factor models are a suitable alternative to traditional diffusion models especially if an analysis of causal factors is preferred versus estimates of cumulative cases versus time.

An attempt to predict the cumulative cases at the end of the pandemic (1191 days) at one- and two-year intervals reveals the challenge of prediction. The model improves as more data is accumulated but lacks predictive power. We speculate that the model might be extended by noting the convergence rate of successive approximations and then extrapolating. This was left as future work.

Keywords

Covid-19, diffusion of contagions, estimating spread of disease, multi-factor model, public health policy, causal factors in epidemiology.

---

[1] Duration is number of days from first infection to last infection reported.

https://doi.org/10.32388/I4DVAG.2

What Is a Multi-factor Model?

Many models have been developed for estimating and predicting the spread of contagions via contact. They are essentially diffusion models with compartments devised to capture variable diffusion rates. For example, the famous Kermack-McKendrick differential equation model defined three compartments – susceptible, infectious, and recovered (SIR) [7]. Subsequent compartments have been added to the original model; for example, the SEIR (S-Exposed-IR) model. SEIR has been extended in many different ways to capture other properties of the population under study. Typically, diffusion rates for each component are estimated by optimal least squares (OLS) curve fitting to empirical data.

The Kermack-McKendrick model and its many extended derivatives are used to estimate the size and duration of epidemics [1, 2, 5, 6, 8-14]. The spread of disease via contact is assumed to be principally dependent on the infection rate (probability that contact will transfer the disease), while the death rate depends on a different diffusion rate known as death rate (conditional probability that the patient will die assuming they have the disease). More elaborate models incorporated additional factors such as population size, region, public health policy and political leanings. In most cases, the diffusion within a compartment is used to estimate some parameter of interest using OLS curve fitting to arrive at an estimate or prediction.

Diffusion models most often result in logistic equations for cumulative cases of the form:

$$I(t) = \frac{1-\exp(-rt)}{1+a*exp(-rt)}$$

Where parameters *a* and *r* are obtained by optimal least squares (OLS) curve fitting. In [6] the author applies the Bass product adoption model to easily obtain these parameters without requiring the solution to a differential equation [8].

Another alternative approach is to use network models of contagion that yields insights into the effects of topology on contagion spreading [5, 6]. Susceptible populations are assumed to be non-uniform, forming clusters and sparse groupings which either accelerate or slow spreading. These network models reveal properties of scale-free structure containing hubs which are designated as super-spreaders. In [5] the author shows how scale-free structure is a natural consequence of contact over throughout the epidemic.

In this work, we depart from compartmentalization, diffusion/differential equation modeling, and making assumptions about how COVID-19 spread, and instead apply a *data science approach to modeling* the spread of COVID-19 throughout California during the pandemic from March 2020 to June 2023 using data collected by public health for each of the 58 counties of California. [2] Eleven hundred and ninety-one days of cumulative and daily cases were recorded for each county and downloaded into Excel. This data was augmented by family income, Gini coefficients, and country size from the California database.[3] That is, our model is based on 58 data points provided by the state of California and correlations calculated by the author. This

---

[2] https://data.chhs.ca.gov/dataset/covid-19-time-series-metrics-by-county-and-state.

[3] https://data.ca.gov/dataset

multi-factor model is purely statistical in the sense that it applies correlation coefficients and OLS linear equations. It does not attempt to relate the spread of COVID-19 to biological or medical processes or public policy. Nor does in use diffusion as a mechanism for spreading.

The 58 counties vary widely in terms of population size and density, income and income inequality (Gini coefficient) and cumulative infections. Specifically, population and population densities range from 1,200 to 9.9 million and 1.6 to 17,688 people per square mile, respectively. Incomes range from \$45,000 to \$120,000 and Gini coefficients[4] range from 0.389 (below the US average of .414) to 0.507 (above the US average), respectively. Cumulative cases range from 139 to 3.5 million – several orders of magnitude differences. The variations are very large resulting in large variations of cumulative cases, see Figure 1. Figure 1 illustrates the vast differences in cumulative cases across the state for these outliers.

Using time series data from California, recorded over the period of the pandemic (2020 – 2023), we show that estimating the *number of cases and duration* of the epidemic is tractable in terms of several independent variables such as population, population density, family income, Gini coefficient, and size (land area) of the county. We build linear models to estimate the size and duration of California's COVID-19 epidemic but find that estimating the infection rate is impossible because the infection rate behaves like a random variate.

In addition, we test the flexibility of the multi-factor modeling approach by constructing two sub-models – one based on the counties with 49,000 or fewer cases, and another based on the top counties with 50,000 – 3.5 million cases. This shows the flexibility of the multi-factor model across a wide range of factors and cumulative cases. The results support our belief that multi-factor models can be used effectively for large and small outbreaks to identify potential causal factors.

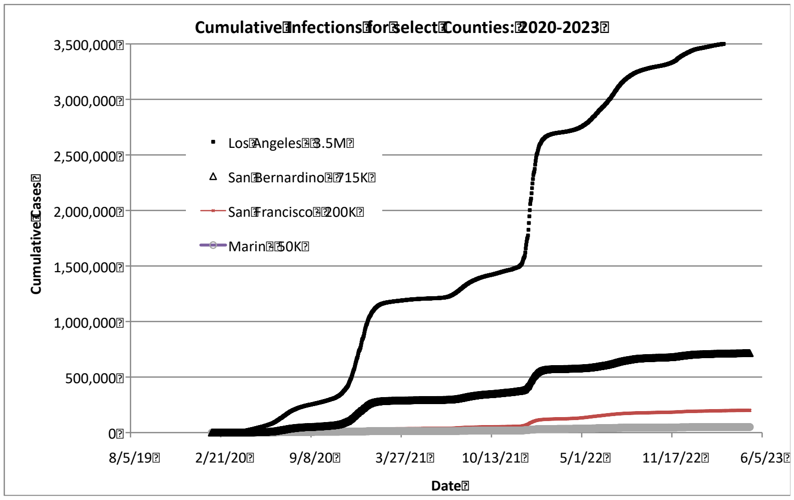


Figure 1. An illustration of the vast differences in cumulative cases: These counties were chosen based on the fact they are the counties with maximum values of factors per Table I. Los Angeles

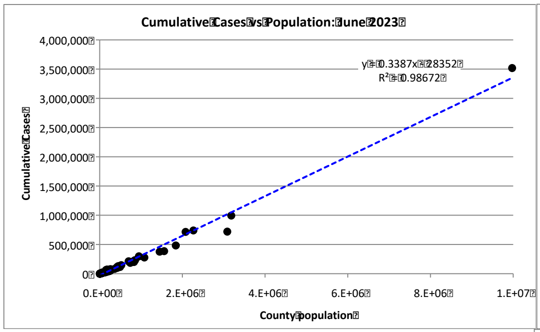
---

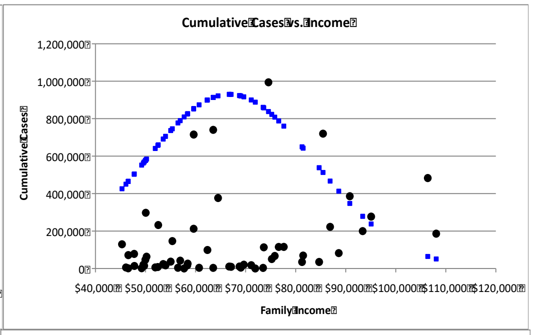
[4] Gini is an international measure of wealth inequality based on Pareto distribution.

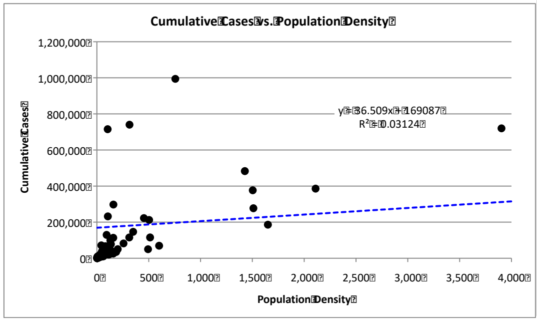
has the largest population; San Bernardino the largest land area (size); San Francisco the highest density and Gini coefficient; Marin the highest family income. Clearly, there is a large difference among counties.

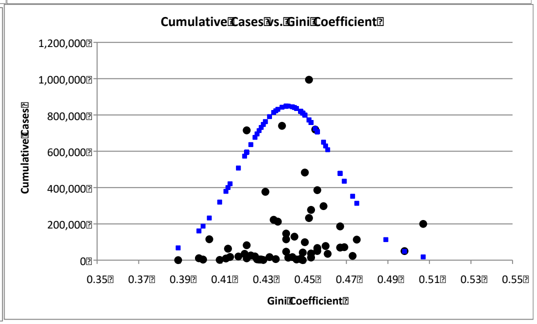
Table I. Independent Variables as factors in the formulation of a predictor vector, x. Los Angeles has the highest population with nearly 10 million inhabitants. San Francisco has the highest Gini coefficient and population density. Marin is the wealthiest county as measured by family income. San Bernardino County has the largest land area. This variation leads to highly variable cumulative cases and duration.

#i	Factor x_i	County with maximum value
1	Population	Los Angeles
2	Income	Marin
3	Population Density	San Francisco
4	Gini Coefficient	San Francisco
5	Land Area of County(Size)	San Bernardino









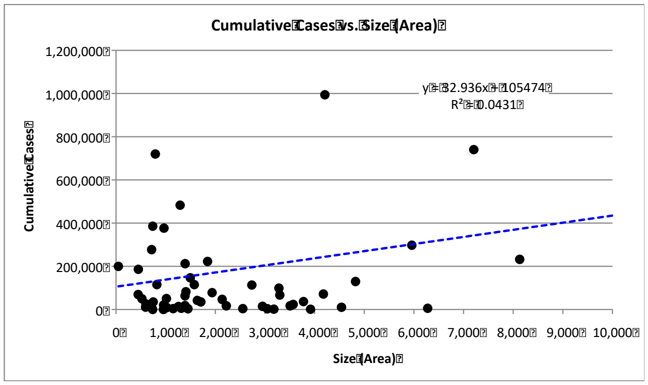


Figure 2. California COVID-19 data shows very little correlation between cumulative cases and independent factors of population density, income, Gini coefficient, and size. Only population is strongly correlated. Income and Gini coefficient even appear to obey a normal distribution and lack any indication of a trend.

Methods and Models

Generally, one purpose of modeling is to reveal cause-and-effect in a phenomenon. In the case of COVID-19, it is suspected that rates of spreading, duration, and ultimate number of infected people are the result of factors such as crowding, economic diversity, and public health policy. Is it possible to construct a model that is based on such factors? In the following, we isolate 5 factors that intuitively impact the number of cases infected and the duration of the epidemic in California. All 58 counties are included in the model. See Table I.

The approach taken here is not unique, but rather it has a precedent. The Fama-French equation uses a similar approach to estimate the value of stocks in a portfolio [3, 4]. However, the Fama-French equation does not use correlations as predictors. As far as the author knows, this is a novel approach.

It is reasonable to use correlation coefficients because they are a measure of “similarity” or “similitude” for modeling. Still, there is no rule for what factors to use or what level of correlation makes sense. In practice, it is a trial-and-error process to obtain close OLS fits to the actual data. We used available factors and their correlations to construct “predictor” equations that form a bases for estimation, similar to principle component analysis, only much simpler.

First, we show that single-factor estimation is insufficient to obtain accurate estimates of duration. Duration of epidemic is improved greatly by using more than one factor and even cumulative case estimation is improved by using multiple factors. Regardless of the number of factors used, the estimation of the infection rate remains beyond the reach of our method, as the infection rate appears to be random across the 58 counties.

Then we show that multiple factors improve estimates of cumulative cases and duration. But this approach has limitations – it cannot predict total cumulative cases or duration early in the epidemic because multi-factor analysis lacks predictive power. We show that estimates improve with more data, but the model is unable to predict the ultimate size or duration of an epidemic.

Single Factors

Figure 2 shows the results of the OLS curve fitting of cumulative cases versus each of the 5 independent variables across all 58 counties. This approach failed. There is little correlation between cumulative cases measured at the end of the epidemic and four of the five independent variables considered one at a time. For income and Gini coefficients, the dependence appears to be random, roughly obeying a Normal distribution, although the r-squared error R^2 for fitting a Normal distribution indicates a poor fit.

Table II. Correlation Coefficients for each x-factor across all 58 counties.

x-Factor	Cumulative Cases Correlation	Duration Correlation
x_1 : Population	0.993	0.323
x_2 : Income	0.082	0.277
x_3 : Density	0.177	0.220
x_4 : Gini	0.177	0.340
x_5 : Size	0.208	0.140

Combining Factors

An alternative to single-factor modeling is to use many factors in combination to gain more accuracy. The multi-factor model developed here is based on 5 independent variables – x-factors – collected per county across all 58 counties. Table I lists the factors – independent variables – used to estimate cumulative cases and duration. These were selected because they are available, the author hypothesized that they are relevant to estimation of infections, and they are for the most part, positively correlated. Other factors may be just as valid – there is nothing magical about the chosen factors, however, highly correlated factors yield greater accuracy.

In each case, we fit a linear equation to a “predictor” function based on the correlation coefficients obtained by correlating cumulative cases or duration with each of the 5 x-factors. The cumulative number of cases and duration at the end of the epidemic are accurately modeled in terms of independent factor vectors I(x) and D(x) as follows:

$$Cases = 22{,}802I(x) - 20{,}361;\ R^2 = 0.99$$[5]

$$Duration = 2{,}765D(x) + 302.5;\ R^2 = 0.67$$

The infection rate is shown to be uncorrelated and appears as a random variable, roughly obeying a normal distribution across counties, see Figure 5.

Linear predictor equations use correlated factors to predict key parameters of the epidemic. They are similar to the Fama-French equations used in finance to value a stock. Let P(x) be a linear predictor equation consisting of a weight sum divided by a normalization function, as follow:

---

[5] R-squared error from Excel.

$$P(x) = \frac{Weighted\ sum}{normalization} = \sum_{i=1}^{5} c_i\ x_i / \sum_{i \in s} x_i$$

where,

x_i value of factor, c_i correlation coefficient.
s is the inclusion set, e.g., factors that are included in the summation over x.

Note that only factors that improve the OLS curve fit are included in the normalization summation. These form the inclusion set s. In the following analysis we find that income and density provide the best OLS fit for cumulative cases and population and income provide the best fit for duration. They form the inclusion set s.

The correlation coefficients for cumulative cases and duration are listed in Table II. The correlation coefficients c_i were tested to determine the best fit for the data. Population and income were the major drivers of cumulative infections and duration, with density and size in marginally improving the prediction of cumulative cases.

For cumulative cases the predictor inclusion set is s = {2, 3}:

$$I(x) = \frac{.993x_1 + .082x_2 + .177x_3 + .177x_4 + x.208x_5}{x_2 + x_3}$$

For duration the predictor inclusion set is s = {1, 2}:

$$D(x) = \frac{.323x_1 + .277x_2 + .220x_3 + .140x_5}{x_1 + x_2}$$

Where, of course, the x's are actual data points taken from the California database. To get the best OLS fit, D(x) uses four of the factors in Table 1 to obtain the numerator and only two factors to obtain the denominator. This is due to the fact that Gini coefficients vary only slightly across the 58 counties. [This is problematic because Gini is orders of magnitude smaller than the other factors].

Figures 3, 4, and 5 show the results. Figure 3 OLS curve fit is nearly idea using all factors in the numerator, N(c, x), but only two normalization factors, income and density. Figure 4 OLS curve fit is modest using only population and income in the normalization term. Figure 5 shows that infection rate in terms of per capita cumulative cases approximates a random Normal distribution unsuitable for curve fitting.

Gini coefficient and size do not exert a major influence on cumulative cases or duration. This is partly due to the relative magnitude of Gini coefficients which vary between zero and one, compared with population and income which are many orders of magnitude larger. However, when attempting to fit cumulative cases and duration to Gini coefficients, alone, no reasonable OLS curve fits was possible. The correlations are simply not significant, because Gini coefficients vary only slightly from their average value – Gini = $0.448 \pm .059$.

Size also has little effect on cumulative cases and duration due to the fact that it is incorporated in density, and density has a physical meaning when considering contagions. It is presumed that more dense populations are prone to greater infections, although that cannot be claimed here. Density includes population, which exerts the greatest influence overall.

Similarly, duration varies only slightly across the 58 counties with the shortest and longest span of 975 to 1191 days. It appears to be unusual that income is influential in duration with correlation of 0.277. Both population and income were used to normalize the predictor. One can speculate that greater prosperity leads to greater duration! However, the greatest correlation coefficient is with Gini coefficient (0.340) suggesting wealth inequality as a larger influence. See Table II.

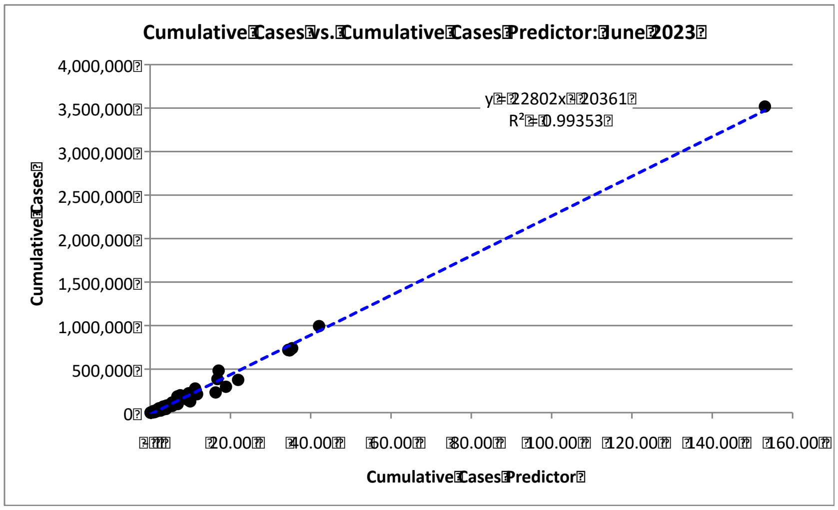


Figure 3. Cumulative cases are accurately estimated by population, income, density, Gini coefficient and size, with income and density, only, as normalization. Data collected by the 58 counties in California is plotted versus cumulative cases predictor I(x).

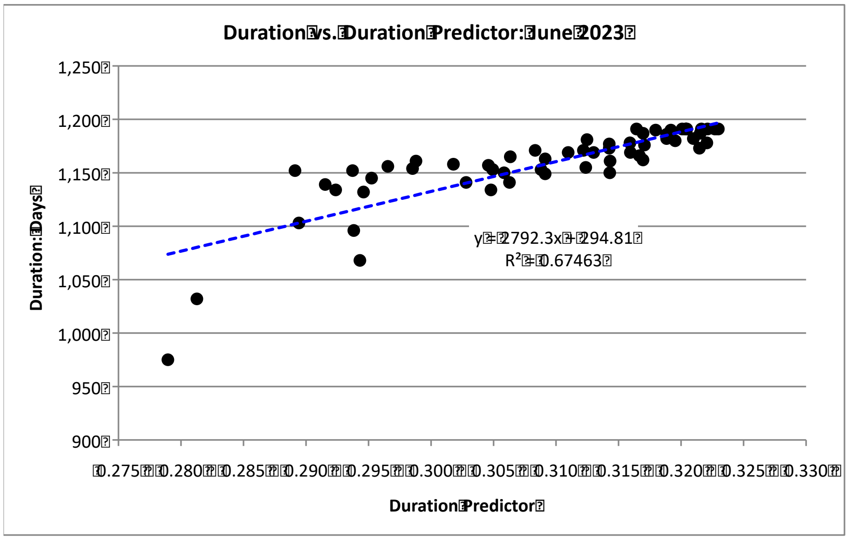


Figure 4: Duration (days) is somewhat accurately estimated by population, income, density, and size with population and income, only, as normalization. Data collected by the 58 counties in California is plotted versus cumulative cases predictor D(x).

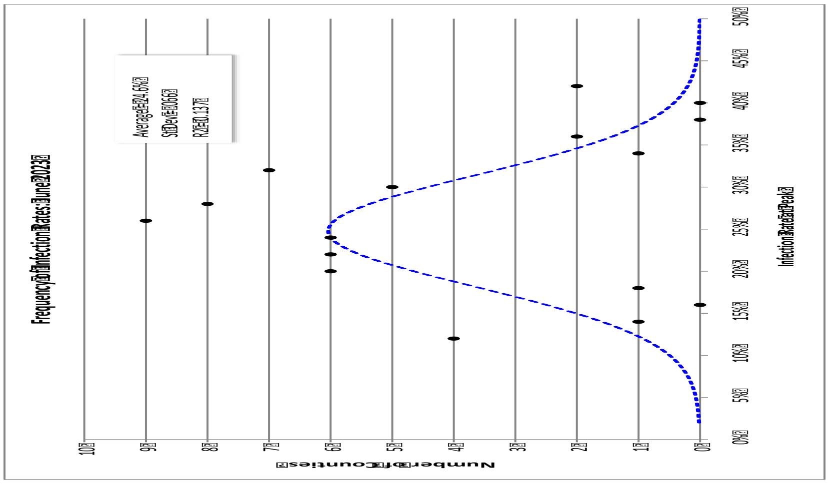


Figure 5. The infection rate varies like a random variable that most likely obeys a Normal distribution with an average value of 24%. This graph was constructed by counting counties falling into bins of size 2%.

Factors in High/Low Cumulative Cases

The counties were divided into a bottom half and a top half by cumulative cases to determine if counties with a high number of infections were different from counties with low numbers. The bottom half incurred 139 to 49,000 infections while to top half incurred 50,000 to 3.5 million infections. The correlation results are shown in Table III.

Correlation with population remains nearly the same as the state-wide result, but the other factors vary greatly. For the bottom half (fewer cases of infected people), population density (size) and income (Gini) are more strongly correlated. Specifically, lower population density and lower inequality (lower Gini) resulted in a lower number of infections. Cumulative cases are negatively correlated with size.

The top half (greater number of cases) correlates strongly with the population as before, but much less than with the other factors. In fact, income is negatively correlated with a high number of cases – greater wealth implies fewer cases. Generally, the Gini coefficient shows a difference when dividing the counties along cumulative infection lines. Compared with the state-wide correlations, this suggests that density plays a bigger role in infections than income or the Gini coefficient. Perhaps weak correlation with Gini coefficient is due to the nearly flat values of Gini centered on $0.448 \pm .059$.

Estimating infection rate, an essential element in most diffusion models, is made more difficult because there is no known correlation with basic factors like population, population density, income, Gini, or size. Infection rates for COVID-19 in California ranged from 10% to 42% – a very large spread.

Improved models for low- and high-infection counties are easily obtained using the correlations in Table III. We re-compute correlation coefficients for each half and then plug the coefficients into the predictor function and identify the best OLS fit.

Table III. Correlation Coefficients: Cumulative cases bottom half versus top half. Inclusion value of 1 means the factor is included in set s.

Cumulative Cases	*Population*	*Income*	*Density*	*Gini*	*Size*
Inclusion (Bottom)	0	1	1	0	0
Inclusion (Top)	0	1	1	0	0
Bottom Half	0.970	0.227	0.799	0.374	(0.183)
Top Half	0.994	(0.061)	0.093	0.241	0.214

Figure 6 shows the results of recomputing models for the two populations – one with fewer than 49,000 cases and the other with 50,000 to 3.5 million cases. The models are very different but still dominated by population. In both cases, (a) and (b) predictors use the correlation coefficients given in Table III but normalized by income and density. That is, the denominator consists of income and density and specifically ignores population.

For bottom half cases set s = {2, 3}:

$$B(x) = \frac{.970x_1 + .227x_2 + .799x_3 + .394x_4 + x - .183x_5}{x_2 + x_3}$$

For top half cases set s = {2, 3}:

$$T(x) = \frac{.994x_1 - 0.061x_2 + .093x_3 + .241x_4 + .214x_5}{x_2 + x_3}$$

where x_2 and x_3 are income and density. A better fit is obtained by eliminating population in the normalization function. The best OLS fit is obtained without population, Gini coefficient, and size in the denominators.

$$\text{Bottom Half } Cases = 11{,}945B(x) - 994;\ R^2 = 0.86$$
$$\text{Top Half } Cases = 22{,}970T(x) - 27{,}758;\ R^2 = 0.99$$

Compared with the combined model:

$$Cases = 22{,}802I(x) - 20{,}361;\ R^2 = 0.99$$

Table IV. Correlations for 1-year and 2-years into the pandemic. Inclusion of 1 means the factor is included in the normalization set s.

Cumulative Cases	Population	Income	Density	Gini	Size
Inclusion (365)	0	1	1	0	0

Inclusion (730)	0	1	1	0	0
Correlations (365)	0.981	0.033	0.143	0.286	0.261
Correlations (730)	0.989	0.061	0.162	0.301	0.224
Correlations (All)	0.993	0.082	0.177	0.177	0.208

Predictive Power

Do these multi-factor models have predictive power? That is, can they reveal cumulative cases and duration midway through an epidemic? To find out, multi-factor analysis was applied at days 365 (1/30/21) and 730 (1/30/22) of the covide-19 pandemic. The multi-factor predictors and OLS curve fits were repeated to determine if it is possible to make reasonable predictions of cumulative cases based on 365 and 730 days of data. We obtained OLS curve fits with R^2 in excess of 0.98 for both terms. The results are summarized in Table VI.

OLS curve fitting results in the following predictors and linear models:

$$I_{365}(x) = \frac{.981x_1+.033x_2+.143x_3+.286x_4+x.261x_5}{x_2+x_3}$$

$$I_{730}(x) = \frac{.989x_1+.061x_2+.162x_3+.301x_4+x.224x_5}{x_2+x_3}$$

Resulting in linear models:

$$\text{Cumulative } Cases\ 365 = 7{,}110 I_{365}(x) - 11{,}392;\ R^2 = 0.98$$

$$\text{Cumulative } Cases\ 730 = 16{,}335 I_{730}(x) - 20{,}983;\ R^2 = 0.99$$

$$\text{Cumulative } Cases\ All = 22{,}802 I(x) - 20{,}361;\ R^2 = 0.99$$

Once again, the best OLS fit is obtained by limiting the normalization factors, in this case, income and density were the only factors in the denominator of the predictor. Note that population, income, and density correlations monotonically increase over time. That is, as more data are collected, the correlations grow stronger. Is this a trend or coincidence?

Using these predictors to estimate the number of cases in Los Angeles county, for examples, yields 1.1 million in 1/30/21 and 2.6 million in 1/30/22. The number of cases by the end of the pandemic was 3.5 million. As the amount of data increases, so does the accuracy of the predictions. More data means greater accuracy. Of course, this is of little consequence at the early stages of an epidemic. Accurate prediction of the ultimate number of casualties due to an epidemic such as covid-19 remains a challenging problem.

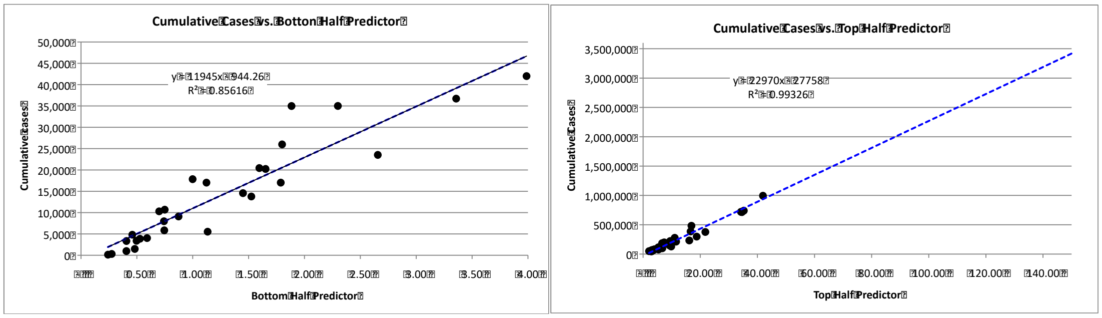
Conclusions

It is possible to improve the accuracy of epidemic models by using multiple factors rather than a single factor or classical SEIR equations. Using population, population density, income, income equality measure Gini, and land area size, we obtained very high accuracy for cumulative cases and modest accuracy for estimated duration. Population and density are strongly correlated with cases reported in low-infection counties, while other measures such as the Gini coefficient and size play a secondary role.

We provide a novel prediction model based on correlation coefficients but note that the infection rate behaves as a random variable, meaning we are not able to estimate it. The prediction model is a weighted sum of factors where correlation coefficients are the weights. This approach is scalable – it can be applied on a country, county, or local level whenever data are available.

Population, population density, and income play dominant roles in number of cases (per correlations). In fact, the bottom-half versus top-half cumulative cases corresponds with population and income per Table V. We see that large population counties incur large cumulative cases and high-income counties incur even higher cumulative cases, in general. Thus, population is a dominant factor.

Multi-factor models cannot predict ultimate outcomes of cumulative cases or duration any better than diffusion models. They can only show correlations between number of infections and possible causal factors. Estimating duration remains an extremely difficult challenge. However, it may be possible to apply an uber model to determine the rate of convergence of successive estimates in order to forecast the ultimate cumulative cases and duration. This was not done here.



(a). Less than 49,000 cumulative infections. (b). Greater than 50,000 cumulative infections.

Figure 6. Cumulative cases versus predictor models for (a) bottom half of infections, and (b) top half infections showing close OLS fits with very different parameters.

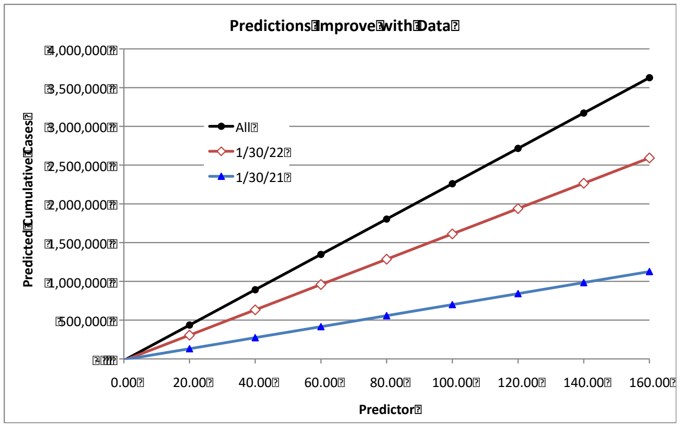


Figure 7. As empirical data is gathered correlations increase, in general, and predictions improve.

Table V. Population dominates infections as compared with income.

	Population Range	*Income Range*	*Per Capita Cases*
Bottom-half	1.2k-222k	$46k-$85k	0.10-0.31
Top-half	1.5m-9.9m	45k-120k	0.19-0.42

Discussion

The multi-factor approach may seem somewhat ad hoc because factor selection and use appears arbitrary and perhaps unconvincing. The author’s use of intuition can be criticized. However, multi-factor modeling provides a test of dependence on factors such as Gini coefficient, which, as it turns out, exhibits weak correlation with cumulative cases. On the other hand, population is clearly a major factor even as population density is less meaningful than intuition might suggest. Multi-factor models represent a simple form of principle factor analysis but still lack predictive power.

Can multi-factor models be used to predict eventual outcomes? It is doubtful that multi-factor modeling can render better predictions than diffusion models, which are reliable up to one week in the future. Figure 7 shows how estimation of the ultimate cumulative cases “converges” as more data are considered. Obviously, this information arrives late in a pandemic which limits its usefulness.

Multi-factor models do indicate broad “cause-and-effect” processes. It is highly likely, although speculative, that the next pandemic will ravish high population/dense areas with high incomes regardless of Gini coefficient or size of territory.

An improvement to this model whereby prediction of ultimate cumulative cases by trend analysis of subsequent estimates (made with additional data) promises to yield better estimates of ultimate cases. This is left as future work.

Citations

[1] Carcione, J. M., Santos, J. E., Bagaini, C., & Ba, J. (2020). A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model. *Frontiers in Public Health, 8.* https://doi.org/10.3389/fpubh.2020.00230

[2] Pei, Y., Li, J., Xu, S., & Xu, Y. (2022). Adaptive Multi-Factor Quantitative Analysis and Prediction Models: Vaccination, Virus Mutation and Social Isolation on COVID-19. *Frontiers in Medicine, 9.* https://doi.org/10.3389/fmed.2022.828691

[3] Fama, E. F., & French, K. R. (1998). Value versus Growth: The International Evidence. *The Journal of Finance, 53*(6), 1975-1999.

[4] Fama, E. F., & French, K. R. (1996). Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance, 51*(1), 55-84.

[5] Lewis T. G. (2021). Scale-Free Emergence of Contagion Networks from Random Populations. *Epidemol Int J*, 5(3): 000202.

[6] Lewis T. G. and Al Mannai W. I. (2021). Predicting the Size and Duration of the COVID-19 Pandemic. *Front. Appl. Math. Stat.* 6:611854.doi: 10.3389/fams.2020.611854

[7] Kermack, W. O. and McKendrick, A.G. (1927). A contribution to the Mathematical Theory of Epidemics, *Proceedings of the Royal Society A*,115, 1927, pp.700-721.

Bibliography

[8] Bass, Frank M., (1969). A new product growth for model consumer durables. Management Science, Vol. 15, No.5, pp.215-227.

[9] Hossain, T., Miah, M., and Hossain, B. (2017). Numerical Study of Kermack-McKendrick SIR Model to Predict the Outbreak of Ebola Virus Diseases Using Euler and RK4 Methods, American Scientific Research Journal for Engineering, Tech, and Sciences (ASRJETS), 2017, Volume 37, No.1, pp 1-21.

[10] Raissi,M., Ramezani, N., and Seshaiyer, P. On Parameter Estimation Approaches for Predicting Disease Transmission through Optimization, Deep Learning and Statistical Inference Methods, Letters in Biomathematics, Published by Informa U.K., Oct 2019.

[11] Yi-Cheng Chen, Ping-En Lu, Cheng-Shang Chang, and Tzu-Hsuan Liu, (28 April 2020), A Time-dependent SIR model for COVID-19 with Undetectable Infected Persons. http://gibbs1.ee.nthu.edu.tw/A_TIME_DEPENDENT_SIR_MODEL_FOR_COVID_19.PDF

[12] Murray, Christopher, Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. doi:https://doi.org/10.1101/2020.04.21.20074732

[13] Nave, Op., Shemesh, U., & HarTuv, I. (2021). Applying Laplace Adomian decomposition method (LADM) for solving a model of Covid-19. In Computer Methods in Biomechanics and Biomedical Engineering (Vol. 24, Issue 14, pp. 1618–1628). Informa UK Limited. https://doi.org/10.1080/10255842.2021.1904399

[14] Harris, J. E. (2021). Los Angeles County SARS-CoV-2 Epidemic: Critical Role of Multi-generational Intra-household Transmission. In Journal of Bioeconomics (Vol. 23, Issue 1, pp. 55–83). Springer Science and Business Media LLC. https://doi.org/10.1007/s10818-021-09310-2