

Variability Scaling and Capacity Planning in Covid-19 Pandemic

L. Jeff Hong

School of Management and School of Data Science, Fudan University, Shanghai, China
hong_liu@fudan.edu.cn

Guangwu Liu

College of Business, City University of Hong Kong, Kowloon Tong, Hong Kong, China
msgw.liu@cityu.edu.hk

Jun Luo*

Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China
jluo_ms@sjtu.edu.cn

Jingui Xie

School of Management, Technical University of Munich, Heilbronn, Germany
Munich Data Science Institute, Technical University of Munich, Munich, Germany
jingui.xie@tum.de

Abstract

Capacity planning is a very important global challenge in the face of Covid-19 pandemic. In order to hedge against the fluctuations in the random demand and to take advantage of risk pooling effect, one needs to have a good understanding of the variabilities in the demand of resources. However, Covid-19 predictive models that are widely used in capacity planning typically often predict the mean values of the demands (often through the predictions of the mean values of the confirmed cases and deaths) in both the temporal and spatial dimensions. They seldom provide trustworthy prediction or estimation of demand variabilities, and therefore, are insufficient for proper capacity planning. Motivated by the literature on variability scaling in the areas of physics and

*Corresponding author.

biology, we discovered that in the Covid-19 pandemic, both the confirmed cases and deaths exhibit a common variability scaling law between the average of the demand μ and its standard deviation σ , that is, $\sigma \propto \mu^\beta$, where the scaling parameter β is typically in the range of 0.65 to 1, and the scaling law exists in both the temporal and spatial dimensions. Based on the mechanism of contagious diseases, we further build a stylized network model to explain the variability scaling phenomena. We finally provide simple models that may be used for capacity planning in both temporal and spatial dimensions, with only the predicted mean demand values from typical Covid-19 predictive models and the standard deviations of the demands derived from the variability scaling law.

Key words: Covid-19; capacity planning; variability scaling; demand aggregation; network model; risk pooling effect

1 Introduction

The Covid-19 pandemic has brought a global public health crisis. On one hand, some critical medical resources are facing shortage and therefore require a fair and rational planning and allocation of these resources (Emanuel et al., 2020), including the medical personnel of nurses and physicians (Bersano and Pantoni, 2020; Vergano et al., 2020), the critical equipment of ventilator and ECMO service as well as the ICU management (Kotfis et al., 2020; Mehrotra et al., 2020; White and Lo, 2020), the personal protective equipment of masks and alcohol (Livingston et al., 2020), and so on. On the other hand, capacity planning and maintenance of these resources are often challenging and costly in practice. For instance, the New York City has issued an ambitious pandemic preparedness plan of stockpiling thousands of ventilators and millions of masks right after the 2006 SARS outbreak. Due to the budget constraints, the city only acquired 500 additional ventilators at the beginning, but all were auctioned off a few years later because the department could not afford the expensive maintenance cost (Elliott et al., 2020).¹

¹Interested readers may refer to the PROPUBLICA for a detailed report via <https://www.propublica.org/article/how-new-york-city-emergency-ventilator-stockpile-ended-up-on-the-auction-block>.

In order to provide professional advice and guidelines for decision makers to combat Covid-19, it is important to understand the evolution of Covid-19 infected cases and then establish capacity planning models and resource allocation rules over different time periods and across different regions. There emerge many works on data analysis and predictions of Covid-19 in the past several months. To name a few, Remuzzi and Remuzzi (2020) propose an exponential model to fit the data in Italy, and Crokidakis (2020) builds a susceptible-infectious-quarantined-recovered (SIQR) model for Covid-19 in Brazil, while Fanelli and Piazza (2020) use a susceptible-infected-recovered-deaths (SIRD) model to forecast the Covid-19 spreading trends in China, Italy and France. The team from the Institute for Health Metrics and Evaluation (IHME) at the University of Washington also develops modified susceptible-exposed-infectious-recovered (SEIR) models to predict the evolution trends in different scenarios in the US and other countries worldwide.² These models are in general developed to predict the mean of Covid-19 cases, e.g., infected and death cases, over different time periods and regions. There are also some recent works that combine Covid-19 data with other data to deliver a better prediction, e.g., Jia et al. (2020) use the mobile phone data to develop a spatio-temporal risk source model that can forecast the distribution of confirmed cases.

However, predicting only the mean demand over different time periods (i.e., in the temporal dimension) and over different geographical locations (i.e., in the spatial dimension) is not enough. Understanding the corresponding variability of demand is also very critical for determining medical resource planning and allocation. Some models also provide a 95% confidence interval of their estimates. However, it is not clear what the confidence intervals mean. For example, some critics have pointed out that the actual death rate was outside the 95% confidence interval of IHME's prediction in nearly 70% of the US states (Marchant et al., 2020), which highlights the need for better characterization and more accurate estimates of the variabilities.³ Based on classical theories on capacity planning in the areas of production and operations management, to hedge against the fluctuations in the random demand and

²More detailed information can be found via <http://www.healthdata.org/covid>.

³See article for more details <https://www.sydney.edu.au/news-opinion/news/2020/04/14/us-covid-19-deaths-poorly-predicted-by-ihme-model.html>.

to take advantage of the risk pooling effect, one needs to have good understanding of the variabilities in the demand of resources.

Traditionally, the variability may have already been specified by the stochastic processes as modeling the random demand arrivals. For instance, the Poisson and inhomogeneous Poisson processes are often used to model the patient arrivals to a healthcare system (Green, 2005), which implies that the standard deviation (i.e., a common approach to quantifying variability) of the demand is the square root of the mean demand. Then, the famous square-root staffing rule can be obtained (Yom-Tov and Mandelbaum, 2014). The square-root staffing rule indicates a strong pooling effect in the temporal dimension, which is due to the *independent increments* property of Poisson models. Similarly, the pooling effect has been obtained in the spatial dimension, e.g., by specifying the demand correlation structures in the inventory problem with multiple locations in Eppen (1979), Corbett and Rajaram (2006) and Mak and Shen (2014).

The pooling effect, which means fulfilling the demands for multiple time periods or different locations from a centralized system, is essential for planners to make proper inventory or resource allocation decisions to meet certain quality of service (QoS) levels (Bimpikis and Markakis, 2016). The magnitude of the pooling effect depends not only on the average demand level, but also on the scale of variability. Then, appropriately modeling the variability becomes very critical.

Motivated by the literature on variability scaling in the areas of physics and biology, in this study, we discovered that in the Covid-19 pandemic, both the confirmed cases and deaths exhibit a common *variability scaling law* between the average of the demand μ and its standard deviation σ , that is,

$$\sigma \propto \mu^\beta, \tag{1.1}$$

where β is the scaling parameter that determines the scaling effect, which is typically in the range of $[0.5, 1]$, and the scaling law exists in both the temporal and spatial dimensions. Notice that the Poisson and inhomogeneous Poisson process models also obey the variability scaling law, with $\beta = 0.5$. Indeed, one can prove that commonly used time series models, such as auto-regressive models and moving-average models, also obey the variability scaling

law with $\beta = 0.5$. Notice that the square-root staffing rule may be extended to these time series models and the square root corresponds to $\beta = 0.5$.

The relationship in Equation (1.1) captures the scaling phenomena in many areas, in order to understand the dynamics and the structure of a complex system. In ecology, Taylor was among the first who showed empirically that the mean and standard deviation of the number of individuals of a species in a region follows this scaling relation (Taylor, 1961), which is known as *Taylor's law* today. In physics, this law is known as *fluctuation scaling*, and it was found in heavy-ion collision experiments (Botet et al., 2001) and in fluxes of cosmic radiations (Uttley and McHardy, 2001). This law has also been discovered in some other systems, including stock markets (Eisler and Kertész, 2006), emails and printing jobs etc. (Barabási, 2005; Oliveira and Barabási, 2005)⁴. However, it appears that this law has not been discovered in Covid-19 or other healthcare systems.

By analyzing the Covid-19 data, we find that the values of β are typically greater than 0.5. In fact, they are between 0.7 and 1 in most situations. Note that $\beta < 1$ implies that there is a pooling effect and, therefore, there exist economies of scale in the medical resource planning. Meanwhile, $\beta > 0.7$ suggests that the pooling effect is not as strong as what is commonly expected under the Poisson models, which tends to underestimate the resource requirement and lead to a lower QoS if not taking the true variability scaling into consideration. We show that the variability scaling phenomena exist not only in Covid-19, but also in various healthcare demands, in both temporal and spatial dimensions, based on a number of datasets from different healthcare systems in different countries.

Based on the mechanism of contagious diseases, we further build a stylized network model to explain the variability scaling phenomena. In the model, the critical reason that the variability scaling phenomenon emerges is that the network is a small-world network, where any two persons in the network can be linked with a small of steps (West et al., 1999; Salathé et al., 2010). Therefore, the diseases can be transmitted quickly between any two random persons, which matches what we have seen in the Covid-19 pandemic. Furthermore,

⁴As an interesting note, Oliveira and Barabási (2005) showed in their *Nature* paper that the mail correspondence patterns between Darwin and Einstein follow the variability scaling law.

to illustrate how the variability scaling affects the capacity planning, we also provide a simple model that may be used for capacity planning and resource allocation in both temporal and spatial dimensions, with only the predicted mean demand values from typical Covid-19 predictive models. In this model, the standard deviations of the demands are derived from the variability scaling law.

The contributions of this study can be summarized in the following three folds. We first show that variability scaling exists in the Covid-19 pandemic in US, Italy, Brazil, China and worldwide, and also in many other healthcare systems. Then, we propose a network model to provide a reasonable explanation that can theoretically yield the scaling phenomena. Finally, we establish explicit formulas between variability scaling and capacity planning, and economies of scale. We show that high variability requires extra resource planning, and the high variability reduces the benefit of economies of scale.

We close this section by briefly reviewing some existing works in the operation management literature, which is closely related to our study. Eppen (1979) is a seminal work on the pooling effect of a multi-location newsvendor problem where the demand correlations are captured by the covariance matrix of a multivariate normal distribution. This work has been generalized by Corbett and Rajaram (2006) for non-normal dependent demand distributions and extended by Mak and Shen (2014) to a general supply chain network model with various dependence structures. These studies can be viewed as modeling works on the demand variability in the spatial dimension. In the temporal dimension, the doubly stochastic Poisson process proposed by Whitt (1999) is one of the pioneering works that model the demand correlation of the arrival process for call centers. Since then, in a stream of research, e.g., Chen and Henderson (2001); Avramidis et al. (2004), stochastic arrival rates of arrival processes have been studied in various queueing models for call center operations management. Empirical studies of Kim and Whitt (2014); Kim et al. (2015) investigate the Poisson and non-Poisson properties for call centers and healthcare systems to address data fitting issues, e.g., over-dispersion of Poisson models. Recently, Whitt and You (2019) use the index of dispersion for counts (IDC) introduced in Fendick and Whitt (1989) to emphasize the importance of dependence in single-server queueing systems. However, rather than providing the IDC to indicate the dispersion between the mean and standard deviation, we establish

the relation between the mean and standard deviation via the proposed variability scaling law. The variability scaling reveals the long-range dependence over time.

The rest of this paper is organized as follows. In Section 2, we summarize our empirical findings which show that variability scaling exists in the Covid-19 pandemic and all other investigated healthcare systems. We then propose a network model to explain the mechanism of variability scaling in Section 3. In Section 4, we show how the variability scaling affects the capacity planning using a stylized demand model. Section 5 concludes the paper.

2 The Evidence: What is Variability Scaling?

By taking the logarithmic transformations, the law of variability scaling in Equation (1.1) becomes

$$\log \sigma = \alpha + \beta \log \mu,$$

where α and β are the regression coefficients to be estimated, and the standard deviation σ and the average μ will be computed from the data. In other words, the statistical model can be expressed as follows,

$$\log \hat{\sigma}_i = \alpha + \beta \log \hat{\mu}_i + \epsilon_i, \quad i = 1, 2, \dots, N, \quad (2.1)$$

where $\log \hat{\mu}_i$ and $\log \hat{\sigma}_i$ are the “observable” variables, ϵ_i are the unobserved error terms, and N is the total number of “observations” for $\log \hat{\mu}_i$ and $\log \hat{\sigma}_i$.⁵ Both $\hat{\mu}_i$ and $\hat{\sigma}_i$ are computed from the corresponding sample average and sample standard deviation,

$$\hat{\mu}_i = \frac{1}{b_i} \sum_{j=1}^{b_i} Y_{ij}, \quad \text{and} \quad \hat{\sigma}_i = \left[\frac{1}{b_i - 1} \sum_{j=1}^{b_i} (Y_{ij} - \hat{\mu}_i)^2 \right]^{1/2},$$

where Y_{ij} , $j = 1, 2, \dots, b_i$ are random samples of size b_i from observation i .

The index i plays an essential role in our study, which could be used to represent different subsystems in different dimensions. For instance, in the temporal dimension, the subsystems

⁵In fact, the “observations” for $\log \hat{\mu}_i$ and $\log \hat{\sigma}_i$ are not directly observable but computed based on the sample average and sample standard deviation, which then bring the error terms into the model. We adopt to use the word “observation” simply following the traditional terminology in the statistic/econometric literature.

could be daily new cases of Covid-19 in one country in different consecutive daily intervals; in the spatial dimension, the subsystems could represent new cases occurred in different geographical state locations. To better calibrate the power relationship, subsystems must have different sizes, so that the scaling phenomena may be observed over different magnitude scales.

Note that the statistical model in Equation (2.1) is often called a *loglinear regression model* in econometric literature, which studies the elasticities rather than the marginal effects between σ_i and μ_i (see, for example, (Verbeek, 2008)). However, we interpret the model from a statistical point of view rather than an econometric perspective since the endogeneity between σ and μ is intrinsic and difficult to be removed. Nevertheless, the model is simple and has been widely studied in various research areas in the literature and it can provide us some managerial insights in Covid-19 and other healthcare management problems.

2.1 Empirical Study in Covid-19 Pandemic

To establish the variability scaling law in Covid-19 pandemic, we collect the data of Covid-19 daily new cases and daily new deaths worldwide, as well as the data from four different countries on four different continents, i.e., US on North America, Italy on Europe, Brazil on South America and China on Asia. Since the majority of the new cases and deaths occurred in China are from Hubei province, we investigate on Hubei province separately.⁶ We briefly introduce the datasets as follows, and refer interested readers to Appendix B.1 for more details about the original raw datasets.

- (1) Worldwide: The dataset is obtained from the World Health Organization (WHO), which consists of daily new cases and daily deaths in 237 countries or regions, over the time period from March 8, 2020 to September 30, 2021.
- (2) US: The dataset is obtained from the US Centers for Disease Control (CDC), which

⁶Among the 34 provincial-level regions in China, Hubei was first attacked by Covid-19 pandemic and suffered most. Up to September 30, 2021, there are about 71% new cases and 97.3% deaths in mainland China reported from Hubei.

consists of daily new cases and daily deaths in 51 states (including Washington D.C.), over the time period from March 8, 2020 to September 30, 2021.

- (3) Italy: The dataset is obtained from the Department of Civil Protection of Italy, which consists of the daily new cases and daily deaths in 20 regions, over the time period from March 8, 2020 to September 30, 2021.
- (4) Brazil: The dataset is obtained from the Ministry of Health of Brazil, which consists of daily new cases and daily deaths in 27 states, over the time period from March 22, 2020 to September 30, 2021.
- (5) China: The dataset is obtained from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, which consists of daily new cases and daily deaths in 33 provincial-level regions (including Hong Kong, Macao and Taiwan, but excluding Hubei province), over the time period from January 23, 2020 to September 30, 2021.
- (6) Hubei in China: The dataset is obtained from a medical technology company DXY.cn in China that collected the raw data reported by government sectors in China, which consists of daily new cases and daily deaths in 17 cities in Hubei province, over the time period from January 25, 2020 to April 26, 2020.

We would like to use all data from the beginning of the breakout until the end of September in 2021 because we notice there might be some data corrections in recent months. Then, we select March 8, 2020 as the starting time for US and Italy as well as worldwide because the US CDC decided to relax the requirements for the nucleic acid testing on March 4 and data from all states became available on March 8 (Team, 2020). For Brazil, We start from March 22, 2020 rather than March 8, 2020 since the data of all states in Brazil became available on that day. For China, we choose January 23, 2020 as the starting time because the first record of new cases in CSSE was reported on that day. For Hubei in China, we choose January 25, 2020 as the starting time because the first record of new cases in DXY.cn was available on that day, and choose April 26, 2020 as the ending time because the last hospitalized Covid-19 patient in Wuhan city was discharged on that day.

Table 1: Variability scaling of Covid-19 in different countries or regions

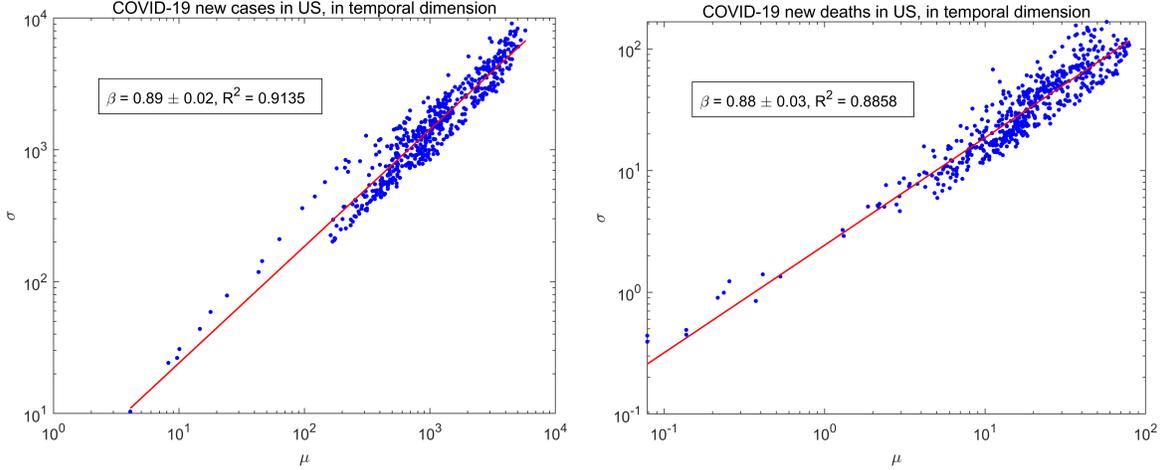
Country (Date)	Dimension	Aggregation level	β	R^2
Worldwide	Temporal	Daily new cases within 572 days	0.94 ± 0.02	0.9416
		Daily deaths within 572 days	0.84 ± 0.02	0.8853
(03/08/2020 – 09/30/2021)	Spatial	Daily new cases across 237 countries/regions	0.82 ± 0.02	0.9719
		Daily deaths across 237 countries/regions	0.76 ± 0.02	0.9496
US	Temporal	Daily new cases within 572 days	0.89 ± 0.02	0.9135
		Daily deaths within 572 days	0.88 ± 0.03	0.8858
(03/08/2020 – 09/30/2021)	Spatial	Daily new cases across 51 states	0.95 ± 0.04	0.9780
		Daily deaths across 51 states	0.85 ± 0.06	0.9490
Italy	Temporal	Daily new cases within 572 days	0.87 ± 0.02	0.9507
		Daily deaths within 572 days	0.87 ± 0.02	0.9436
(03/08/2020 – 09/30/2021)	Spatial	Daily new cases across 20 regions	0.96 ± 0.04	0.9913
		Daily deaths across 20 regions	0.91 ± 0.06	0.9841
Brazil	Temporal	Daily new cases within 558 days	0.95 ± 0.02	0.9463
		Daily deaths within 558 days	0.94 ± 0.02	0.9098
(03/22/2020 – 09/30/2021)	Spatial	Daily new cases across 27 states	1.04 ± 0.15	0.8926
		Daily deaths across 27 states	1.00 ± 0.08	0.9607
China	Temporal	Daily new cases within 617 days	0.98 ± 0.03	0.9057
		Daily deaths within 617 days	0.96 ± 0.02	0.9590
(01/23/2020 – 09/30/2021)	Spatial	Daily new cases across 33 provinces	0.77 ± 0.08	0.9260
		Daily deaths across 33 provinces	0.65 ± 0.04	0.9742
Hubei in China	Temporal	Daily new cases within 93 days	0.93 ± 0.04	0.9783
		Daily deaths within 93 days	0.96 ± 0.03	0.9842
(01/25/2020 – 04/26/2020)	Spatial	Daily new cases across 17 cities	0.94 ± 0.07	0.9800
		Daily deaths across 17 cities	0.91 ± 0.13	0.9397

On each dataset, we analyze the variability scaling law in both the temporal and spatial dimensions by fitting the loglinear regression model in Equation (2.1). Taking the Covid-19 data in US for example, in the temporal dimension analysis, we use the daily new cases and new deaths data of 51 states from March 8, 2020 to September 30, 2021 (about 19-month period with 572 days in total), and calculate the mean and standard deviation of new cases and deaths over the 51 states for each day. Then, we obtain $N = 572$ data points and use them to run the regression model. In the spatial dimension analysis, we calculate the mean and standard deviation of daily new cases and deaths over the whole time period for each state, and run the loglinear regression model using $N = 51$ data points, as shown in Figure 1. Similarly, we apply the same procedure to analyze all other datasets. All results are summarized in Table 1 and detailed plots of all other datasets are presented in Figures 4–8 in Appendix A.

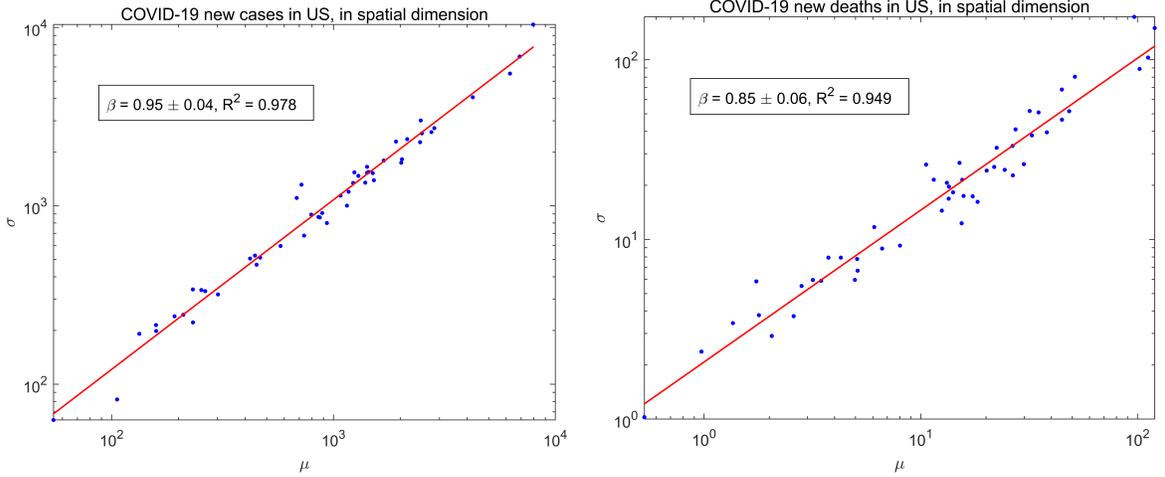
From the results shown in Table 1, we have five important findings. First, we find that there exists a strong linear relationship between $\log \sigma$ and $\log \mu$ in either the temporal or spatial dimensions, as all the values of R^2 are above 0.89. The results indicate that there is a common variability scaling law that governs the variability of Covid-19 new cases or new deaths in the temporal and spatial dimensions.

Second, unlike the commonly assumed Poisson models in the healthcare management literature, the exponents of the scaling relation, i.e., the values of β , are not always equal to 0.5. Notice that we may rewrite the variability scaling law as $\sigma/\mu \propto \mu^{\beta-1}$, where σ/μ is the *coefficient of variation* that measures the demand variability per unit of average demand, implying that there is a pooling effect when $\beta < 1$, but the pooling effect decreases as β becomes larger. In particular, the results of $\beta > 0.9$ in some scenarios demonstrate a large variability scaling effect but only mild pooling effect, which may lead to great challenges in medical resource planning in the Covid-19 pandemic.

Third, for new cases, the values of β in spatial dimension in US and Italy are significantly larger than those in temporal dimension, while the opposite results can be observed worldwide and in China. One plausible reason is that the traveling restrictions among different countries and strict traveling examinations among different cities in China may effectively



(a) new cases computed over all states for $N = 572$ days. (b) new deaths computed over all states for $N = 572$ days.



(c) new cases computed over all days for $N = 51$ states. (d) new deaths computed over all days for $N = 51$ states.

Figure 1: Variability scaling of Covid-19 in US.

lower the travel frequency, and thus lower the Covid-19 propagation geographically, compared with the situations in US and Italy.

Fourth, the values of β in Brazil, for both new cases and deaths, no matter in temporal or in spatial dimensions, are very close to 1,⁷ which is very similar to the results obtained in the

⁷We note that $\beta = 1.04$ for the new cases in the spatial dimension in Brazil, however, its 95% confidence interval is $[0.89, 1.19]$, which is relatively large compared with others but still covers values less than 1.

early time at Hubei province in China. As for Hubei province, which was first attacked by Covid-19 pandemic, there was no enough medical resource or effective actions to combat the Covid-19 at the beginning, could cause high variability scaling phenomena. We conjecture that the lack of sufficient medical resource or effective prevention and control measures could also be the problems in some developing countries, such as Brazil, then leading to a higher variability scaling, especially in new deaths, compared with developed countries, such as US and Italy. Note that the small value of $\beta = 0.65$ for new deaths in the spatial dimension in China also indicates the significant effectiveness of the Covid-19 prevention and treatment policies taken in China.

Last but not least, it is also worthwhile pointing out that, except the result in the temporal dimension in Italy and Hubei province, the values of β for new deaths tend to be smaller than those for new cases in the same dimension, indicating a smaller variability for new deaths, which may possibly be explained by the vaccine or treatment effect after being confirmed.

2.2 Empirical Study in Other Healthcare Systems

We also examine the variability scaling in other healthcare demands, from 10 datasets in 4 different countries with respect to different healthcare systems. These datasets include four detailed patient arrival admissions from a large hospital in China, Israel, Singapore and the US, respectively; one dataset on the daily appointments to a group of clinics in Singapore; one dataset on the admissions of all patients with the government insurance plan to a cluster of hospitals from a provincial capital city in China; two datasets on the all confirmed cases of two infectious diseases (i.e., hand-foot-mouth (HFM) disease and dysentery disease, respectively) with their location information from a prefecture-level city in China; one dataset on all emergency calls to an ambulance service center from one of the largest cities in China; and one dataset on the demands of orthopedic surgery materials from a major supplier at the provincial level in China (see Appendix B.2 for more details about the 10 datasets).

We test the scaling relation on all 10 datasets in the temporal and spatial dimensions

Table 2: Variability scaling in various healthcare systems.

Healthcare demand	Dimension	Aggregation level	β	R^2	Location (Year)
Inpatient admissions in a China hospital	Temporal	Hourly admission requests	0.89 ± 0.05	0.9833	China (2013)
	Spatial	Monthly requests from 262 cities	0.63 ± 0.01	0.9668	
Inpatient admissions in a Singapore hospital	Temporal	Hourly admissions	0.55 ± 0.05	0.9759	Singapore (2010)
All admissions in an Israel hospital	Temporal	Hourly admissions during weekdays	0.52 ± 0.02	0.9549	Israel (2004-2005)
Emergency admissions in a US hospital	Temporal	Hourly admissions into Emergency Department (ED)	0.50 ± 0.03	0.9848	US (2006-2007)
Clinic appointments in Singapore	Temporal	Hourly appointment from 8am to 8pm	0.81 ± 0.02	0.9927	Singapore (2012-2013)
	Spatial	Weekly requests from 26 postal districts	0.82 ± 0.03	0.9905	
Insured inpatient admissions	Spatial	Weekly admissions into 60 public hospitals in different locations	0.55 ± 0.04	0.9396	China (2012-2013)
Hand-Foot-Mouth (HFM) disease	Temporal	Daily incidents in each week	0.71 ± 0.05	0.8484	China (2010-2012)
	Spatial	Weekly incidents in 184 divisions	0.71 ± 0.03	0.9210	
Dysentery disease	Spatial	Weekly incidents in 170 locations	0.59 ± 0.02	0.9645	China (2013)
Ambulance service	Spatial	Monthly ambulance calls from 83 subdistricts	0.88 ± 0.08	0.8468	China (2011)
Orthopedics surgery materials	Spatial	Monthly sales to 150 different hospitals	0.72 ± 0.04	0.8742	China (2012-2013)

whenever the data is supported. The results are summarized in Table 2. We can draw similar conclusions as obtained from the Covid-19 data. Moreover, we have two additional interesting findings. First, compared with the HFM and dysentery infectious diseases, we notice that the values of β in Covid-19 tend to be larger than those in HFM and dysentery diseases, which implies larger fluctuations in Covid-19 pandemic. Second, the values of β for the patient admissions into a hospital in Singapore, Israel and the US in the temporal dimension are smaller than 0.55, which indicates that Poisson arrival processes are suitable to model these demands as widely used in the queueing literature. In the rest of the datasets, we may conclude that many of the currently used models of resource planning with Poisson assumptions tend to underestimate the resource requirement of the Covid-19 pandemic and various healthcare systems and lead to a lower service quality.

3 The Mechanism: What Causes Variability Scaling?

Different models have been proposed to explain variability scaling (Eisler et al., 2008). Some of these models are more conceptual. For instance, it was discovered that random walks on a complex network can generate the variability scaling phenomenon in the temporal dimension (De Menezes and Barabási, 2004). Some of these models are specifically designed to explain a certain type of variability scaling phenomenon in physics or biology. For instance, in mean-field thermodynamics, models have been proposed based on ferromagnetic Ising models (Eisler et al., 2008), and in biology, models have been developed based on plant reproduction processes (Koenig and Knops, 2000). However, these models may not be directly applicable to the explanation of the variability scaling phenomena that we observe in Covid-19 and other healthcare demands. In this paper, motivated by the mechanism of contagious diseases, we establish a simplified network model, which is essentially a space-filling fractal network model that has been used to explain allometric scaling by West et al. (1997, 1999), in order to capture the correlation among demands.

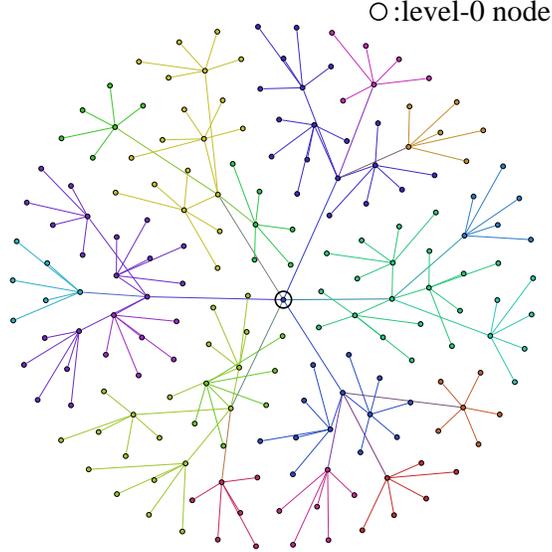


Figure 2: Graphical illustration of the network model with $m = 5$ and $k = 3$.

3.1 The Network Model

We consider the case where the infections are correlated through some internal mechanism, in which each demand node except those in the bottom level is directly connected to $m + 1$ demand nodes. To simplify the analysis, we assume that the network has no cycles, thus, is a tree structure.⁸ Denote the node at the center of the network as the level-0 node (i.e., the root or the central node in Figure 2). Note that the level-0 node has $m + 1$ descendant demand units while the nodes in all other intermediate levels, i.e., level $1, 2, \dots, k - 1$, have m descendants and one parent node. The leaf nodes in the bottom level, i.e., level k , have no descendant, where k is the number of levels of the network. For the purpose of mathematical rigorousness, we assume that $m \geq 2$ and $k \geq 3$.

Let n_i denote the number of nodes in level i , $0 \leq i \leq k$. It can be shown that

$$n_0 = 1, \quad n_i = (m + 1)m^{i-1}, \quad 1 \leq i \leq k.$$

Then the total number of nodes in the network is

$$n = n_0 + \dots + n_k = \frac{(m + 1)m^k - 2}{m - 1}. \quad (3.1)$$

⁸In medical literature, shortest path trees of human contact networks are often used to study the spread of infectious disease (Salathé et al., 2010).

For ease of presentation, we let the aggregated demand of all levels in the network be

$$D = X_1 + \cdots + X_n, \quad (3.2)$$

where $E(X_i) = \mu_0$ and $\text{Var}(X_i) = \sigma_0^2$ for all i , and assume the correlation of any pair of demands (X_i, X_j) is $\rho_d = \rho_0 \gamma^d$, where d is the distance between the two nodes. After some derivation, we can obtain the results for $\text{Var}(D)$ as follows.

Proposition 1. *Suppose that the aggregated demand D is defined in Equation (3.2) with all parameters specified above in the network model. We assume that $m \geq 2$ and $k \geq 3$. Then, as $n \rightarrow \infty$ (i.e., $k \rightarrow \infty$), if $m\gamma^2 < 1$,*

$$\text{Var}(D) = \sigma_0^2 \left(1 + \frac{(m-1)\rho_0\gamma^2}{1-m\gamma^2} \right) n[1 + o(1)];$$

and if $m\gamma^2 > 1$,

$$\text{Var}(D) = \sigma_0^2 \frac{(m-1)\rho_0\gamma^2}{m\gamma^2 - 1} n^{2+2\log\gamma/\log m} [1 + o(1)].$$

Proof. We let C_I and C_O denote the summations of correlations for all pairs of nodes within the same level and cross different levels, respectively. Then,

$$\text{Var}(D) = n\sigma_0^2 + \sum_{i \neq j} \text{Cov}(X_i, X_j) = n\sigma_0^2 + \sigma_0^2(C_I + C_O). \quad (3.3)$$

We first analyze C_I . Let W_i denote the sum of correlations for level i , $1 \leq i \leq k$. Notice that the distance between any two nodes at level 1 is 2, and there are $n_1 m$ such pairs of nodes. Then,

$$W_1 = \rho_2 n_1 m = \rho_0 \gamma^2 m(m+1).$$

Similarly, distance between any pair of nodes in level i ($i = 2, \dots, k$) may take values in $\{2, 4, 6, \dots, 2i\}$. In particular, the number of pairs that have distance $2s$ is $n_i(m-1)m^{s-1}$ for $1 \leq s \leq i-1$, while the number of pairs that have distance $2i$ is $n_i m^i$. Therefore, if $m\gamma^2 \neq 1$,

$$\begin{aligned} W_i &= \sum_{s=1}^{i-1} \rho_{2s} n_i (m-1) m^{s-1} + \rho_{2i} n_i m^i \\ &= \rho_0 \gamma^2 (m+1) (m-1) m^{i-1} \frac{1 - (m\gamma^2)^{i-1}}{1 - m\gamma^2} + \rho_0 \gamma^{2i} (m+1) m^{2i-1}, \end{aligned}$$

and

$$\begin{aligned} C_I &\equiv W_1 + W_2 + \dots + W_k \\ &= \rho_0(m+1)\gamma^2 m^k \frac{1 - (m\gamma^2)^k}{1 - m\gamma^2}. \end{aligned} \quad (3.4)$$

Next we analyze C_O . Let T_{ij} denote the sum of correlation terms between level- i and level- j nodes. Notice that distance between the node at level 0 and a node at level j is j . Then, for $j \geq 1$,

$$T_{0j} = \rho_j n_j = \rho_0 \gamma^j (m+1) m^{j-1}.$$

For $1 \leq i \leq k-1$, and $j > i$, the distance between a node at level i and a node at level j may take value in $\{j-i, j-i+2, \dots, j-i+2i\}$. In particular, there are $n_i m^{j-i}$ pairs with distance $j-i$, $n_i(m-1)m^{j-i}m^{s-1}$ pairs with distance $j-i+2s$ for $1 \leq s \leq i-1$, and $n_i m^{j-i} m^i$ pairs with distance $j-i+2i$. If $m\gamma^2 \neq 1$, for $1 \leq i \leq k-1$ and $j > i$, we have

$$\begin{aligned} T_{ij} &= n_i \left[\rho_{j-i} m^{j-i} + \sum_{s=1}^{i-1} \rho_{j-i+2s} (m-1) m^{j-i} m^{s-1} + \rho_{j+i} m^j \right] \\ &= \rho_0(m+1) \gamma^{j-i} m^{j-1} \frac{1 - (m\gamma^2)^{i+1}}{1 - m\gamma^2} - \rho_0(m+1) \gamma^{j-i} m^{j-2} \frac{m\gamma^2 - (m\gamma^2)^i}{1 - m\gamma^2}. \end{aligned}$$

Then by elementary algebra, if $m\gamma^2 \neq 1$ and $m\gamma \neq 1$,

$$C_O = 2 \sum_{i=0}^{k-1} \sum_{j=i+1}^k T_{ij} = \frac{2\rho_0(m+1)\gamma}{1 - m\gamma^2} \left[\frac{(1+\gamma)m^k}{m-1} + \frac{\gamma^{2k+1}m^{2k}}{1 - m\gamma} - \frac{1 - m\gamma^2}{(m-1)(1 - m\gamma)} \right],$$

Recall that $\text{Var}(D) = n\sigma_0^2 + \sigma_0^2(C_I + C_O)$, and

$$m^k = n(m-1)/(m+1) + 2/(m+1).$$

If $m\gamma^2 < 1$ and $m\gamma = 1$, it can be verified that, as $n \rightarrow \infty$,

$$C_I = \rho_0(m+1)\gamma^2 \frac{1 - \gamma^k}{1 - \gamma} m^k = \frac{\rho_0 \gamma^2 (m-1)}{1 - \gamma} n[1 + o(1)] = \frac{\rho_0}{m} n[1 + o(1)]$$

and

$$C_O = \frac{2\rho_0(m+1)}{1 - \gamma} \left[\frac{m+1}{m-1} m^{k-2} - \frac{2k(m-1) + m+1}{(m-1)m^2} \right] = \frac{2\rho_0(m+1)}{m(m-1)} n[1 + o(1)],$$

and thus

$$\text{Var}(D) = \sigma_0^2 \left(1 + \frac{\rho_0(3m+1)}{m(m-1)} \right) n[1 + o(1)].$$

If $m\gamma^2 < 1$ and $m\gamma \neq 1$, it can be seen that

$$C_I = \frac{\rho_0\gamma^2(m-1)}{1-m\gamma^2}n[1+o(1)],$$

and

$$C_O = \frac{2\rho_0(m+1)\gamma}{1-m\gamma^2}m^k \left[\frac{1+\gamma}{m-1} + \frac{\gamma(m\gamma^2)^k}{1-m\gamma} + o(1) \right] = \frac{2\rho_0\gamma(1+\gamma)}{1-m\gamma^2}n[1+o(1)],$$

and thus

$$\text{Var}(D) = \sigma_0^2 \left(1 + \frac{\rho_0\gamma}{1-m\gamma^2}(m\gamma + \gamma + 2) \right) n[1+o(1)].$$

If $m\gamma^2 > 1$, we have,

$$\begin{aligned} C_I &= \rho_0(m+1)\gamma^2 \left(\frac{m-1}{m+1}n + \frac{2}{m+1} \right) \frac{(m\gamma^2)^k - 1}{m\gamma^2 - 1} \\ &= \frac{\rho_0(m+1)\gamma^2}{m\gamma^2 - 1} \left(\frac{m-1}{m+1}n + \frac{2}{m+1} \right) \left((m\gamma^2)^{(\log n + \log(\frac{m-1}{m+1} + \frac{2}{(m+1)n})) / \log m} - 1 \right). \end{aligned}$$

Notice that $\log(m\gamma^2)/\log m > 0$, and

$$\begin{aligned} & (m\gamma^2)^{(\log n + \log(\frac{m-1}{m+1} + \frac{2}{(m+1)n})) / \log m} \\ &= \left(\frac{m-1}{m+1} + \frac{2}{(m+1)n} \right)^{\frac{\log(m\gamma^2)}{\log m}} n^{\frac{\log(m\gamma^2)}{\log m}} \\ &= \left(\frac{m-1}{m+1} \right)^{\frac{\log(m\gamma^2)}{\log m}} n^{\frac{\log(m\gamma^2)}{\log m}} [1+o(1)] \\ &= (m\gamma^2)^{\log \frac{m-1}{m+1} / \log m} n^{1+2\log \gamma / \log m} [1+o(1)]. \end{aligned}$$

Then,

$$C_I = \frac{\rho_0(m+1)\gamma^2}{m\gamma^2 - 1} \frac{m-1}{m+1} (m\gamma^2)^{\log \frac{m-1}{m+1} / \log m} n^{1+2\log \gamma / \log m} [1+o(1)].$$

Similarly, it can be verified that

$$\begin{aligned} C_O &= \frac{2\rho_0(m+1)\gamma}{m\gamma^2 - 1} \left[\frac{\gamma(m\gamma^2)^k m^k}{m\gamma - 1} + \frac{m\gamma^2 - 1}{(m\gamma - 1)(m-1)} - \frac{(1+\gamma)m^k}{m-1} \right] \\ &= \frac{2\rho_0(m+1)\gamma}{(m\gamma^2 - 1)(m\gamma - 1)} \left[\gamma \frac{m-1}{m+1} (m\gamma^2)^{\log \frac{m-1}{m+1} / \log m} n^{1+2\log \gamma / \log m} [1+o(1)] - \frac{1+\gamma}{m+1} n [1+o(1)] \right] \\ &= \frac{2\rho_0(m+1)\gamma}{(m\gamma^2 - 1)(m\gamma - 1)} \gamma \frac{m-1}{m+1} (m\gamma^2)^{\log \frac{m-1}{m+1} / \log m} n^{1+2\log \gamma / \log m} [1+o(1)], \end{aligned}$$

where the last equality follows from the fact that $1+2\log \gamma / \log m > 1$. Therefore, if $m\gamma^2 > 1$,

$$\text{Var}(D) = \sigma_0^2(n + C_I + C_O)$$

$$= \sigma_0^2 \frac{\rho_0(m-1)(m\gamma+1)\gamma^2}{(m\gamma^2-1)(m\gamma-1)} (m\gamma^2)^{\log \frac{m-1}{m+1} / \log m} n^{2+2 \log \gamma / \log m} [1 + o(1)].$$

□

Based on the result in Proposition 1, we are ready to state the main results to exhibit the variability scaling in this network model.

Theorem 1. *Suppose that the aggregated demand $D = X_1 + X_2 + \dots + X_n$, where $\{X_i, i = 1, 2, \dots, n\}$ is all atomic demand units positioned at all levels in the network model with k levels, in which level 0 has $m+1$ descendant nodes and level $1, 2, \dots, k-1$ have m descendant nodes and one parent node, and $n = [(m+1)m^k - 2]/(m-1)$ as shown in Equation (3.1). Let $E(X_i) = \mu_0$ and $\text{Var}(X_i) = \sigma_0^2$ denote the mean and variance for all i , and let $\rho_d = \rho_0\gamma^d$ denote the correlation between any pair of X_i and X_j with distance d defined in the network model, where $\rho_0 \in (0, 1]$ and $\gamma \in (0, 1)$. Then, as $n \rightarrow \infty$ (i.e., $k \rightarrow \infty$), the scaling parameter β defined in Equation (1.1) can be expressed as follows,*

$$\beta = 0.5 \text{ if } m\gamma^2 < 1, \text{ and } \beta = 1 + \frac{\log \gamma}{\log m} \text{ if } m\gamma^2 > 1. \quad (3.5)$$

Notice that, β increases with these two parameters m and γ . In addition, when $m\gamma^2 > 1$, Equation (3.5) implies that $0.5 < \beta \leq 1$. The results of both $\beta > 0.5$ and $\beta = 0.5$ match what we have observed empirically in Covid-19 and other healthcare systems.

Remark 1. (a). *The proposed network is a stylized model. One may consider each node as one person, then m is the average number of people contacted with others and γ is the individual infection rate. One may consider each node as one family, the m is the average number of contacted families and γ is the family infection rate. In fact, each node could be used to represent one region or one time-interval. Intuitively, the policies like travel restrictions and hygiene regulations can reduce the parameters m and γ , thus reduce the coefficient β .*

(b). *Even though we make strong assumptions of the proposed network model, such as each node (except the leaf nodes) connects to exactly the other $m+1$ nodes, the model can be easily extended to the case with a random number of connected nodes, which will be demonstrated from the simulation results in the following subsection.*

3.2 Simulation Results

We have conducted simulation experiments to test the robustness of the results for the network model. For all simulations, we set $\mu_0 = \sigma_0^2 = 1$ and $\rho_0 = 1$ and assume that the atomic demand X_i has the mean μ_0 and variance σ_0^2 and the correlation between any pair of X_i and X_j , $i \neq j$, is specified according to the network structure of the proposed model.

We take 12 values of γ , i.e., $\gamma = \{0.001, 0.01, 0.1, 0.2, 0.3, \dots, 1\}$. For each γ , we generate a fixed number of random networks with different size n . The size of network n varies from a few nodes to thousands of nodes. The distance of any two nodes d is the length of the shortest path between these two nodes. The expected total demand $E(D)$ is actually n , and the variance of demand $\text{Var}(D)$ can be calculated by Equation (3.3). Using linear regression (as discussed in the empirical analysis), we can estimate the coefficient β for each γ . In order to reflect the random nature of actual networks, we set the number of nodes connected any non-leaf node as a Poisson random variable with mean $m + 1 = 6$. In particular, we consider 5 random networks with 6 replications for each level $k = \{1, 2, \dots, 5\}$, resulting 30 random networks in total, i.e., 30 observations for the regression model.

The simulation results show that variability scaling also emerges in these random networks (as in Figure 3) and the exponents are close to the ones derived for the deterministic networks, which are derived in Equation (3.5) when the number of descendants or the number of connected nodes is fixed as $m + 1$ (also see Table 3). From Table 3, we notice that the value of β increases as γ increases, which is consistent with the theoretical results in Equation (3.5). While β in the random case may be slightly different from that in the deterministic case, which could be due to the estimation error of the mean and standard

Table 3: Comparison of variability scaling exponent β between the deterministic and random networks with $m + 1 = 6$.

γ	0.001	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$m\gamma^2$	5×10^{-6}	5×10^{-4}	0.05	0.20	0.45	0.80	1.25	1.80	2.45	3.20	4.05	5.00
deterministic	0.50	0.50	0.50	0.50	0.50	0.50	0.57	0.68	0.78	0.86	0.93	1.00
random	0.50	0.50	0.50	0.51	0.53	0.57	0.63	0.72	0.79	0.87	0.93	0.99

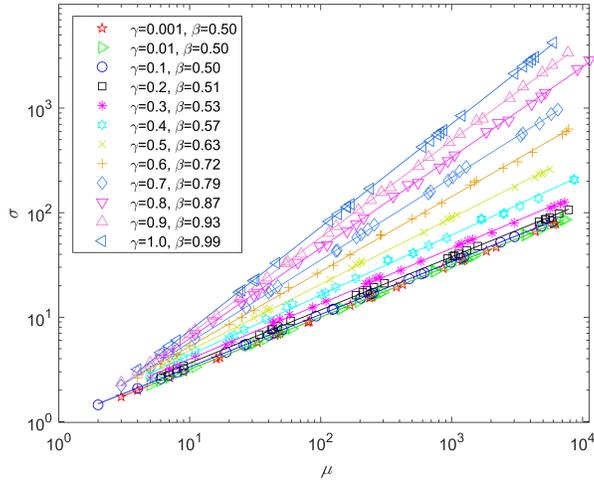


Figure 3: Simulation results when the number of connected neighbors follows a Poisson distribution with mean $m + 1 = 6$.

deviation. From Figure 3, we notice that the linear regression model fits the simulation data quite well for our proposed network model.

4 The Application: How to Apply Variability Scaling?

In this section, we illustrate how variability scaling affects the capacity planning via two simple inventory models. The first is the classical newsvendor problem with the relationship between the mean and standard deviation specified by the variability scaling in the demand function, and the second is a multi-period/multi-location inventory model with demand correlation structure specified by the proposed network model. We demonstrate that the safety stock level will increase while the pooling effect will decrease when the variability scaling effect, i.e., the value of β , becomes larger.

4.1 Capacity Planning for A Single-node Inventory Model

We first consider the classical newsvendor problem (Zipkin, 2000) for a certain type of medical resources for Covid-19. Suppose that the demand for this resource can be fully determined by the number of susceptible, new infected or death cases, which are random and the relationship

between the mean and standard deviation follow the variability scaling law. For example, if we know there are 100 new confirmed cases, then it requires 100 additional beds to admit these patients. For notational simplicity, we directly model the demand of the medical resource as a random variable D , with cumulative distribution function F . In practice, F is unknown and hard to estimate due to the high fluctuation in Covid-19 pandemic. Fortunately, different methods have been proposed in the literature to provide an accurate estimation of the mean $E(D)$. For illustrative purpose, we take a simple way by assuming that D follows a normal distribution $N(\mu, \sigma^2)$, where μ is known and $\sigma = \alpha\mu^\beta$ with α and β estimated from the loglinear regression model.

Let h and p denote the unit holding cost and unit penalty cost of unsatisfied orders, respectively. Let q denote the order quantity and the expected total cost is

$$H(q) = pE[\max(D - q, 0)] + hE[\max(q - D, 0)].$$

By minimizing the expected cost function $H(q)$, we can determine the optimal inventory level q^* as follows,

$$q^* = \mu + \bar{z}\alpha\mu^\beta, \tag{4.1}$$

where \bar{z} is the $p/(p + h)$ th quantile of the standard normal distribution. Then, the corresponding optimal cost is

$$H(q^*) = [h\bar{z} + (h + p)R(\bar{z})] \cdot \alpha\mu^\beta, \tag{4.2}$$

where $R(u) = \int_u^\infty (w - u) \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw$ is the right-hand unit normal linear-loss integral that is widely used in inventory literature.

From Equations (4.1) and (4.2), we can see that the optimal safety stock level and the total cost increase in an exponential rate as the variability scaling coefficient β increases. When β is close to 0.5, it implies there is a significant risk pooling effect since the safety stock level is in a near-square-root order of the mean level. When β is close to 1, it indicates there is almost no pooling effect as the safety stock increases nearly in the same order of the mean demand to satisfy a certain QoS level.

4.2 Capacity Planning for A Multi-period/location Inventory Model

In the following, we establish a simple multi-period/multi-location inventory model to illustrate how the variability scaling affects the optimal capacity planning decision for a central planner. We consider that there are n nodes in the proposed network model, which could represent n time periods or locations, and each has the demand of one particular medical resource X_i , where $i = 1, 2, \dots, n$. For simplicity, we assume that $X_i \sim N(\mu_0, \sigma_0^2)$, and the correlation of any pair of demands (X_i, X_j) is $\rho_d = \rho_0 \gamma^d$, with all parameters defined in Theorem 1.

Without pooling, we assume that each individual node solves the classic newsvendor problem without considering the correlations. Then, the optimal expected cost for each node is $C = c_0 \sigma_0$, where $c_0 = h\bar{z} + (h + p)R(\bar{z})$ with unit holding cost h and unit shortage cost p and \bar{z} is the $p/(p + h)$ th quantile of the standard normal distribution and $R(\cdot)$ is the right-hand unit normal linear-loss integral function as defined before. The optimal total expected cost for the system with n nodes is simply $TC_1 = nC = nc_0\sigma_0$.

With pooling, we assume that the central planner solves the classic newsvendor problem for the entire system with n nodes. Then, the optimal total expected cost for the system is $TC_2 = c_0\sigma_n$, where

$$\sigma_n = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}}$$

with σ_{ij} being the covariance of X_i and X_j . By the variability scaling result in Theorem 1, we know that $\sigma_n \propto n^\beta$, where the scaling coefficient β is typically between 0.5 and 1. Hence, the cost saving of pooling is

$$\frac{TC_2}{TC_1} = \frac{\sigma_n}{n\sigma_0} \propto n^{\beta-1}.$$

We have established a bridge between two important results, i.e., the variability scaling effect and the economies of scale, in which the former one is widely observed in complex systems and the latter one is well known in economics. Then, the cost savings by pooling n subsystems into a centralized system can be approximately estimated as

$$\text{Cost Savings} \approx (1 - n^{\beta-1}) \times 100\%.$$

If β is close to 0.5, we achieve strong cost savings by pooling demands together. If β is close to 1, as observed from the Covid-19 and some other healthcare datasets in Section 2, then the benefit of pooling becomes negligible. In practice, the system manager or government decision maker must be very careful when making pooling decisions when the demand variability of the system is large, e.g., the scaling effect β is significantly greater than 0.8. For instance, considering $\beta = 0.98$ for the new cases in the temporal dimension in China, there is only about 12.1% cost savings over the long time period of 617 days. While there is about 55.3% cost savings over the 33 provinces if using $\beta = 0.77$ for the new cases in the spatial dimension in China. To some extent, it means there is not so much pooling benefit if requiring to prepare some medical resources for a long time period that is more than one year.⁹ However, allowing different regions to share their resources could bring significant pooling benefit in the spatial dimension, which is definitely worthy to design some mechanism to guarantee the sharing system function in the future study.

5 Concluding Remarks

As the Covid-19 pandemic keeps going, it is very critical to design a reasonable and effective capacity planning of medical resources. Accurately calibrating the evolution of Covid-19 is the precondition to achieve this goal. However, most of the current works focus just on predicting mean demands of new cases and deaths and so on, but seldom provide trustworthy prediction or estimation of the demand variabilities, and therefore are insufficient for proper capacity planning.

In this paper, we adopt the commonly used approach to quantifying the variability by the standard deviation and establish a variability scaling law between the mean and standard deviation. Our empirical findings show that variability scaling is common in the Covid-19 pandemic and various other healthcare demands, most of which yield the variability scaling coefficient β larger than 0.5 and even greater than 0.7 for most Covid-19 datasets. This indicates that some existing demand models with β equal to 0.5 may underestimate the

⁹Recall the ventilator and mask stockpiling example in New York City mentioned in Section 1, it could be even worse if taking the maintenance and other costs into consideration in practice.

variation of the medical demands and overestimate the risk pooling effects. We then build a stylized network model to explain the internal demand/infections correlation, from which we can see how correlations among the network affect the intensity of variability scaling. At last, we use two simple inventory models to illustrate that larger variability scaling effect leads to higher inventory stock level, and lower pooling effect.

Our proposed models serve as a first step towards the understanding of the mechanisms behind variability scaling. In the future, as more information about Covid-19 becomes available, e.g., the imposed social distancing mandates, factors that affect the variability scaling effect deserve further investigation. The change of capacity of the nucleic acid testing and other cure resources could also affect the variability of new cases and new deaths for Covid-19, which is also interesting to study. While we assume a linear relationship between the mean and standard deviation in our study, it may be possible to establish a nonlinear relationship as more data are available. It is also worthwhile to model the variability scaling effect in different dimensions by developing advanced stochastic or complex network models. In terms of examining the impact on the capacity planning and resource allocation problems in the Covid-19 pandemic, one may propose more sophisticated problem formulations rather than the stylized newsvendor model.

References

- Avramidis, A. N., A. Deslauriers, and P. L'Ecuyer (2004). Modeling daily arrivals to a telephone call center. *Management Science* 50(7), 896–908.
- Barabási, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature* 435(7039), 207–211.
- Bersano, A. and L. Pantoni (2020). On being a neurologist in Italy at the time of the COVID-19 outbreak. *Neurology* 94(21), 905–906.
- Bimpikis, K. and M. G. Markakis (2016). Inventory pooling under heavy-tailed demand. *Management Science* 62(6), 1800–1813.

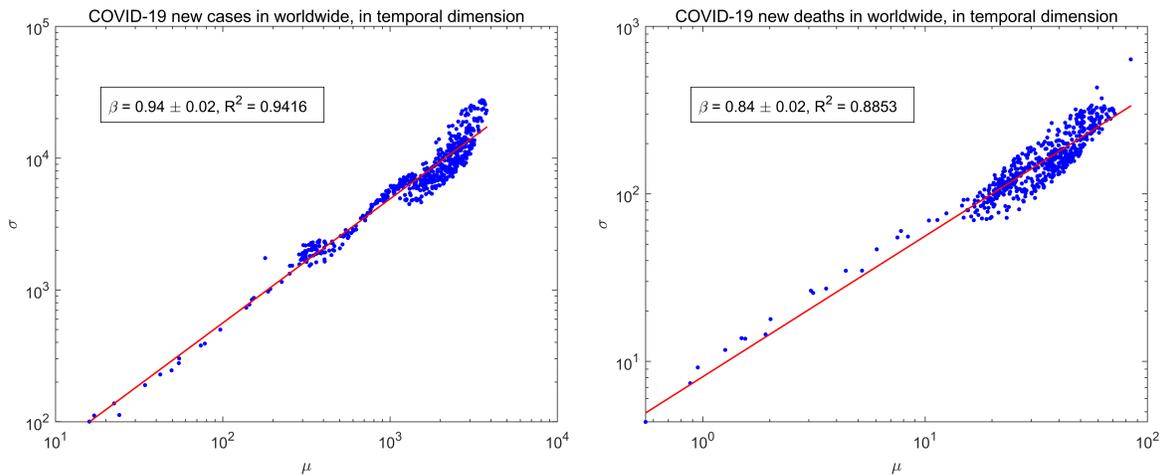
- Botet, R., M. Płoszajczak, A. Chbihi, B. Borderie, D. Durand, and J. Frankland (2001). Universal fluctuations in heavy-ion collisions in the Fermi energy domain. *Physical Review Letters* 86(16), 3514.
- Chen, B. P. K. and S. G. Henderson (2001). Two issues in setting call centre staffing levels. *Annals of Operations Research* 108(1), 175–192.
- Corbett, C. J. and K. Rajaram (2006). A generalization of the inventory pooling effect to nonnormal dependent demand. *Manufacturing & Service Operations Management* 8(4), 351–358.
- Cota, W. (2020). Monitoring the number of COVID-19 cases and deaths in Brazil at municipal and federative units level. *SciELO Preprints:362*.
- Crokidakis, N. (2020). Data analysis and modeling of the evolution of COVID-19 in Brazil. *arXiv preprint arXiv:2003.12150*.
- De Menezes, M. A. and A.-L. Barabási (2004). Fluctuations in network dynamics. *Physical Review Letters* 92(2), 028701.
- Eisler, Z., I. Bartos, and J. Kertész (2008). Fluctuation scaling in complex systems: Taylor’s law and beyond. *Advances in Physics* 57(1), 89–142.
- Eisler, Z. and J. Kertész (2006). Scaling theory of temporal correlations and size-dependent fluctuations in the traded value of stocks. *Physical Review E* 73(4), 046109.
- Elliott, J., A. Waldman, and J. Kaplan (2020, April). How New York City’s emergency ventilator stockpile ended up on the auction block. <https://www.propublica.org/article/how-new-york-city-emergency-ventilator-stockpile-ended-up-on-the-auction-block>.
- Emanuel, E. J., G. Persad, R. Upshur, B. Thome, M. Parker, A. Glickman, C. Zhang, C. Boyle, M. Smith, and J. P. Phillips (2020). Fair allocation of scarce medical resources in the time of Covid-19.
- Eppen, G. D. (1979). Note—Effects of centralization on expected costs in a multi-location newsboy problem. *Management Science* 25(5), 498–501.

- Fanelli, D. and F. Piazza (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals* 134, 109761.
- Fendick, K. W. and W. Whitt (1989). Measurements and approximations to describe the offered traffic and predict the average workload in a single-server queue. *Proceedings of the IEEE* 77(1), 171–194.
- Green, L. V. (2005). Capacity planning and management in hospitals. In *Operations Research and Health Care*, pp. 15–41. Springer.
- Jia, J. S., X. Lu, Y. Yuan, G. Xu, J. Jia, and N. A. Christakis (2020). Population flow drives spatio-temporal distribution of COVID-19 in china. *Nature* 582(7812), 389–394.
- Kim, S.-H., P. Vel, W. Whitt, and W. C. Cha (2015). Poisson and non-Poisson properties in appointment-generated arrival processes: The case of an endocrinology clinic. *Operations Research Letters* 43(3), 247–253.
- Kim, S.-H. and W. Whitt (2014). Are call center and hospital arrivals well modeled by nonhomogeneous Poisson processes? *Manufacturing & Service Operations Management* 16(3), 464–480.
- Koenig, W. D. and J. M. H. Knops (2000). Patterns of annual seed production by northern hemisphere trees: a global perspective. *The American Naturalist* 155(1), 59–69.
- Kotfis, K., S. Williams Roberson, J. E. Wilson, W. Dabrowski, B. T. Pun, and E. W. Ely (2020). COVID-19: ICU delirium management during SARS-CoV-2 pandemic. *Critical Care* 24(1), 1–9.
- Livingston, E., A. Desai, and M. Berkwits (2020). Sourcing personal protective equipment during the COVID-19 pandemic. *Jama* 323(19), 1912–1914.
- Mak, H. Y. and Z. J. M. Shen (2014). Pooling and dependence of demand and yield in multiple-location inventory systems. *Manufacturing & Service Operations Management* 16(2), 263–269.

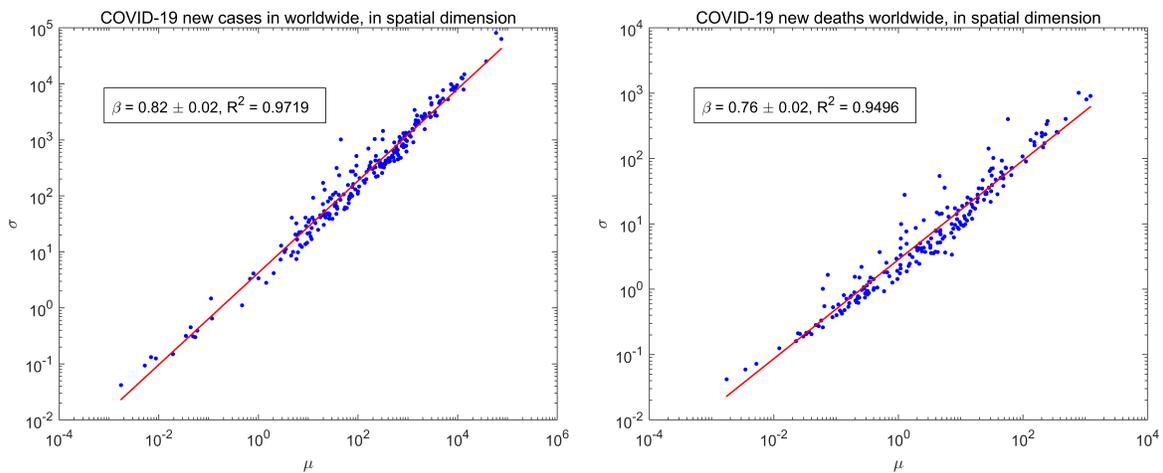
- Marchant, R., N. I. Samia, O. Rosen, M. A. Tanner, and S. Cripps (2020). Learning as we go: An examination of the statistical accuracy of COVID19 daily death count predictions. *arXiv:2004.04734*.
- Mehrotra, S., H. Rahimian, M. Barah, F. Luo, and K. Schantz (2020). A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat COVID-19. *Naval Research Logistics (NRL)* 67(5), 303–320.
- Oliveira, J. G. and A.-L. Barabási (2005). Darwin and Einstein correspondence patterns. *Nature* 437(7063), 1251–1251.
- Remuzzi, A. and G. Remuzzi (2020). COVID-19 and Italy: what next? *The Lancet* 395(10231), 1225–1228.
- Salathé, M., M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones (2010). A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences* 107(51), 22020–22025.
- Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature* 189(4766), 732–735.
- Team, C. C.-. R. (2020). Geographic differences in COVID-19 cases, deaths, and incidence - United States, February 12-April 7, 2020. *Morbidity and Mortality Weekly Report (MMWR)* 69(15), 465–471.
- Uttley, P. and I. M. McHardy (2001). The flux-dependent amplitude of broadband noise variability in X-ray binaries and active galaxies. *Monthly Notices of the Royal Astronomical Society* 323(2), L26–L30.
- Verbeek, M. (2008). *A Guide to Modern Econometrics* (Third edition ed.). John Wiley & Sons Ltd.
- Vergano, M., G. Bertolini, A. Giannini, G. R. Gristina, S. Livigni, G. Mistraletti, L. Riccioni, and F. Petrini (2020). Clinical ethics recommendations for the allocation of intensive care treatments in exceptional, resource-limited circumstances: the Italian perspective during the COVID-19 epidemic.

- West, G. B., J. H. Brown, and B. J. Enquist (1997). A general model for the origin of allometric scaling laws in biology. *Science* 276(5309), 122–126.
- West, G. B., J. H. Brown, and B. J. Enquist (1999). A general model for the structure and allometry of plant vascular systems. *Nature* 400(6745), 664–667.
- White, D. B. and B. Lo (2020). A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *Jama* 323(18), 1773–1774.
- Whitt, W. (1999). Dynamic staffing in a telephone call center aiming to immediately answer all calls. *Operations Research Letters* 24(5), 205–212.
- Whitt, W. and W. You (2019). The advantage of indices of dispersion in queueing approximations. *Operations Research Letters* 47(2), 99–104.
- Yom-Tov, G. and A. Mandelbaum (2014). Erlang-r: A time-varying queue with reentrant customers, in support of healthcare staffing. *M&SOM* 16, 283–299.
- Zipkin, P. (2000). *Foundations of Inventory Management*. McGraw-Hill Companies.

A More Results in Covid-19 Pandemic

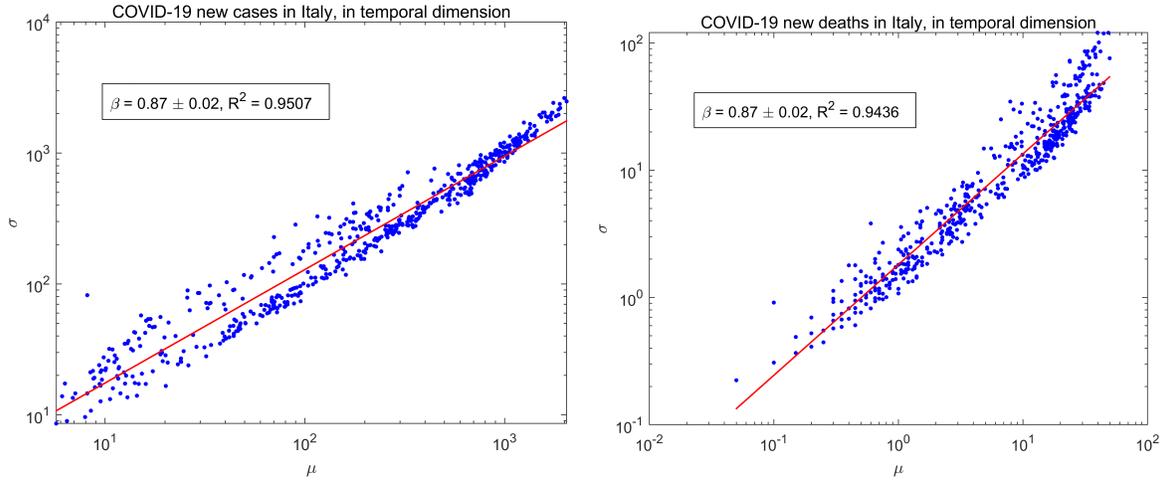


(a) new cases computed over all countries/regions for $N = 572$ days. (b) new deaths computed over all countries/regions for $N = 572$ days.

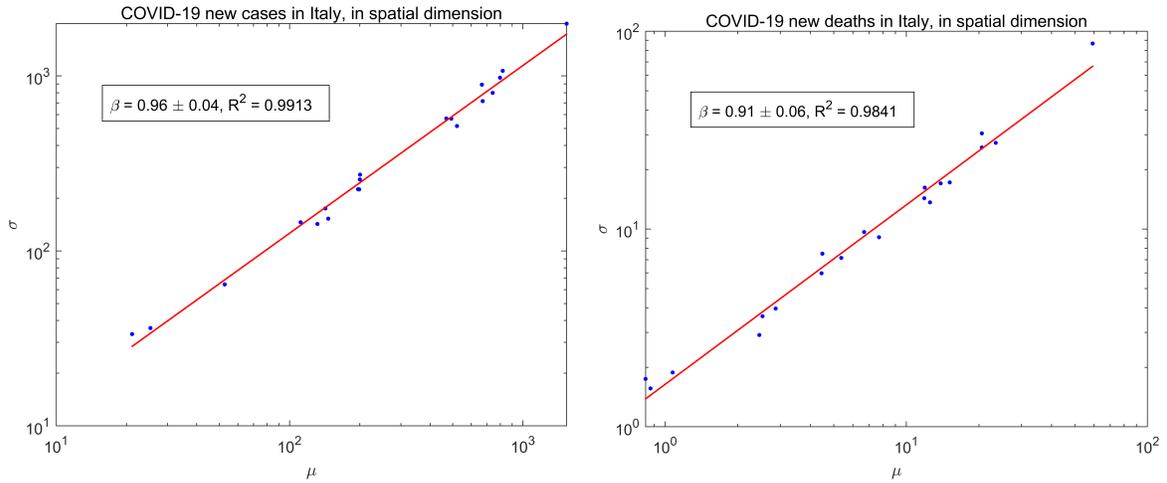


(c) new cases computed over all days for $N = 237$ countries/regions. (d) new deaths computed over all days for $N = 237$ countries/regions.

Figure 4: Variability scaling of Covid-19 worldwide.

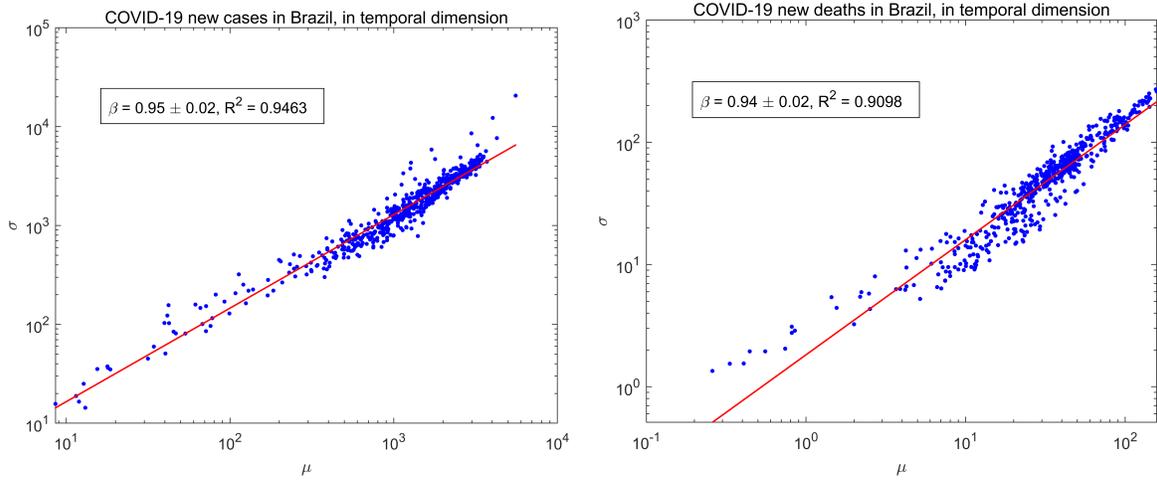


(a) new cases computed over all regions for $N = 572$ days. (b) new deaths computed over all regions for $N = 572$ days.

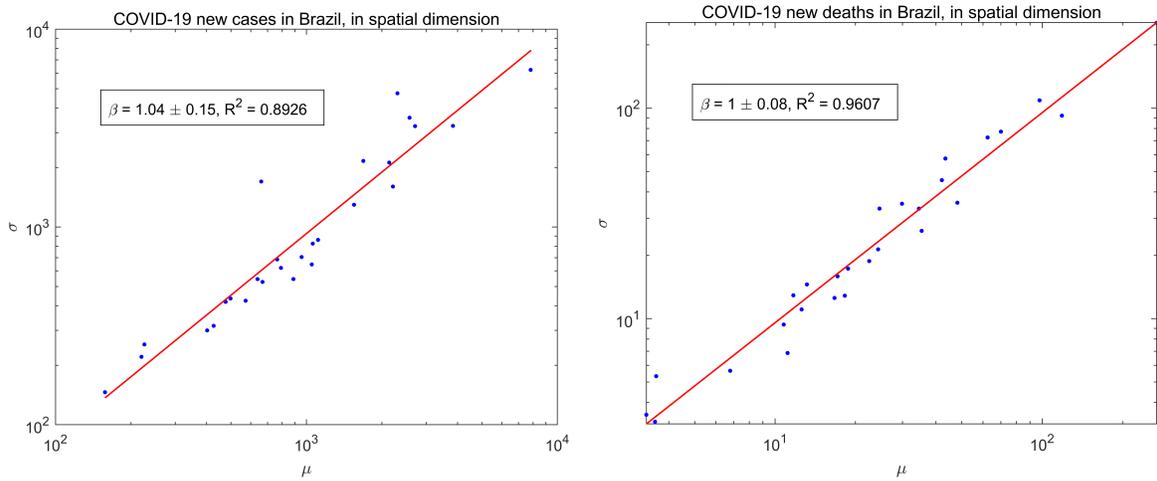


(c) new cases computed over all days for $N = 20$ regions. (d) new deaths computed over all days for $N = 20$ regions.

Figure 5: Variability scaling of Covid-19 in Italy.

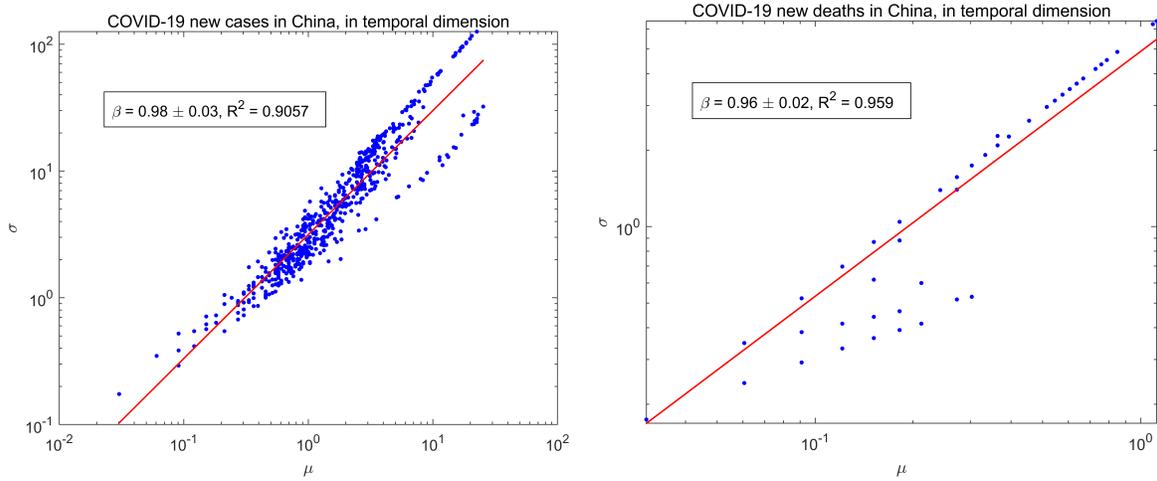


(a) new cases computed over all states for $N = 558$ days. (b) new deaths computed over all states for $N = 558$ days.

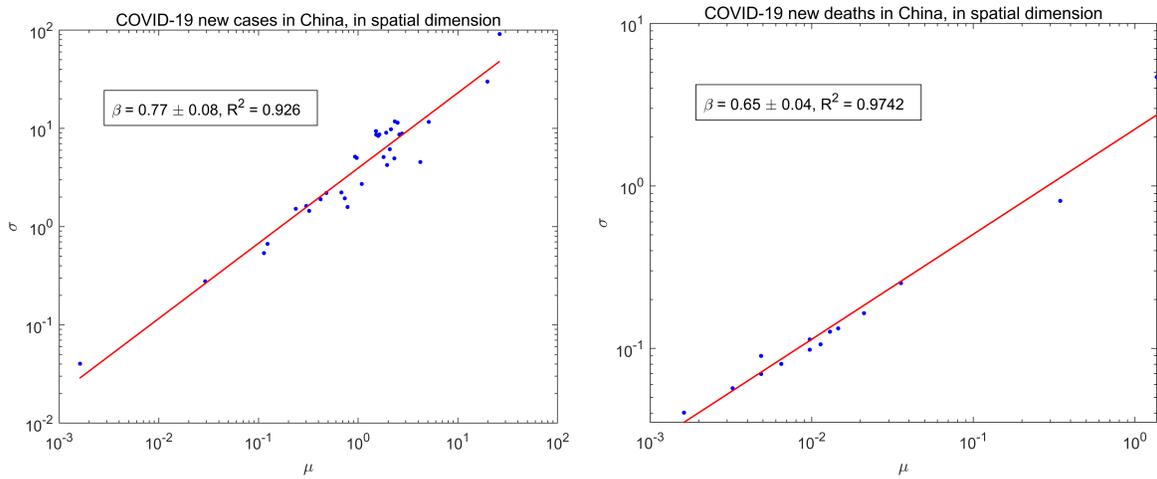


(c) new cases computed over all days for $N = 27$ states. (d) new deaths computed over all days for $N = 27$ states.

Figure 6: Variability scaling of Covid-19 in Brazil.

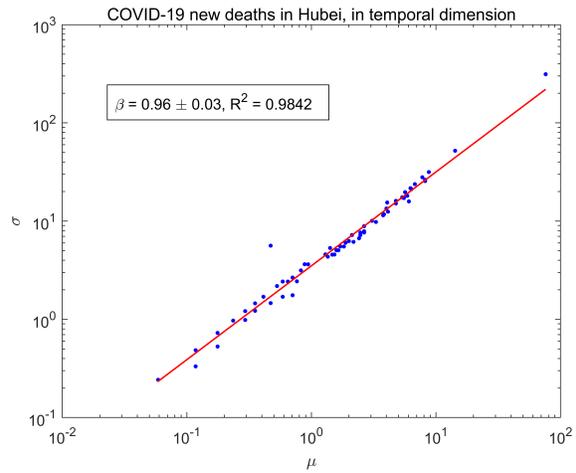
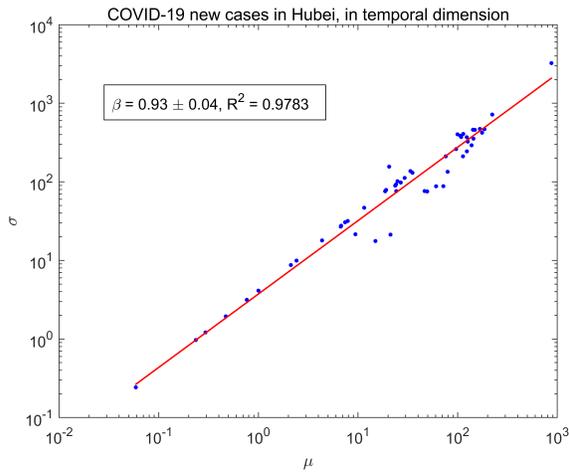


(a) new cases computed over all provinces for $N = 617$ days. (b) new deaths computed over all provinces for $N = 617$ days.

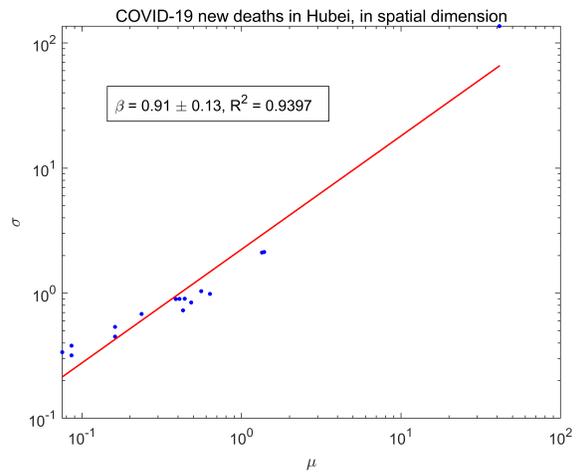
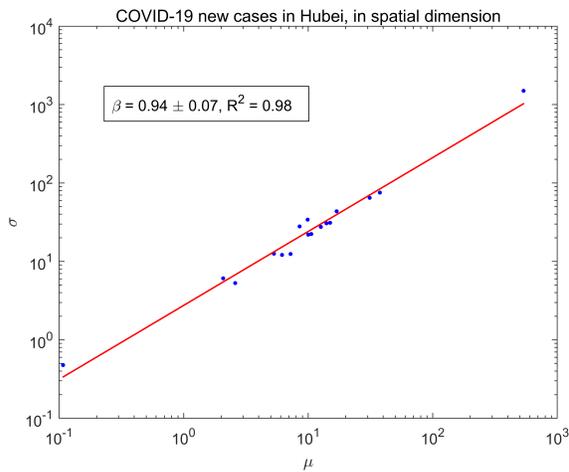


(c) new cases computed over all days for $N = 33$ provinces. (d) new deaths computed over all days for $N = 33$ provinces.

Figure 7: Variability scaling of Covid-19 in China.



(a) new cases computed over all cities for $N = 93$ days. (b) new deaths computed over all cities for $N = 93$ days.



(c) new cases computed over all days for $N = 17$ cities. (d) new deaths computed over all days for $N = 17$ cities.

Figure 8: Variability scaling of Covid-19 in Hubei in China.

B Data Sources and Descriptions

B.1 Data Sources and Descriptions for Covid-19

- (1) Worldwide: The dataset is provided by the WHO. It includes daily new cases and daily deaths of 237 countries or regions in the world, which correspond to the columns named “New_cases” and “New_deaths” in the original file.
- (2) US: The dataset is provided by the US CDC. It contains daily new cases and daily deaths (including both confirmed and probable cases) in 60 regions of US (including 50 states, New York City, Washington D.C. and 8 overseas territories). We exclude the data of all territories and combine the data of New York City with New York state, and use only the daily new confirmed cases (i.e., column “new_case”) and daily deaths (i.e., column “new_death”) from the 50 states and Washington D.C. Note that the CDC may correct the historical data, so the dataset may be slightly different if downloading in different time. The date we downloaded the dataset is December 8, 2021.
- (3) Italy: The dataset is provided by the Civil Protection Department of Italy. The raw dataset records the cases, deaths and the usage of medical resources in 21 regions every-day. Note that Trentino-Alto Adige region consists of Bolzano and Trento provinces, which have been recorded separately. In our analysis, we combine Bolzano and Trento together as one region, and use the daily new cases (i.e., column “nuovi_positivi”) and calculate daily new deaths from the accumulative deaths (i.e., column “deceduti”) by subtracting the accumulative deaths at previous date from the accumulative deaths at current date.
- (4) Brazil: The dataset is obtained from the Federal University of Viçosa (Cota, 2020), which is originally provided by the Ministry of Health of Brazil. The Federal University of Viçosa research team has collected the data of new cases, deaths, tests and vaccinations in 27 states. We use the accumulative cases and deaths data (i.e., columns “totalCasesMS” and “deathsMS”) to calculate the daily new cases or daily deaths by

a similar subtraction as done for Italy. Note that we select March 22, 2020 as the starting time since the first record is on March 21, 2020 in this dataset.

- (5) China: The CSSE at Johns Hopkins University provides the dataset in two separate files as accumulative cases and accumulative deaths. The two files record the accumulative cases and deaths in 195 countries or regions. We include Taiwan in our analysis and deleted the records with “unknown” in column “Province/State” in raw datasets. Then, we use the accumulative cases and deaths data in China at province level, and calculate the daily new cases and daily deaths by subtraction.
- (6) Hubei in China: We use the file called “DXYArea.csv” in the dataset, which is provided by a medical technology company DXY.cn in China. The dataset includes new cases, deaths and cure data in the city level in China. We exclude the data of “Area not defined” and those left blank in column “cityEnglishName”, and use the accumulative cases and deaths (i.e., columns “city_confirmedCount” and “city_deadCount”) of 17 cities in Hubei province to calculate the daily new cases and daily deaths by subtraction. Note that there may be more than one record in some dates, we just use the last record in these dates.

B.2 Data Sources and Descriptions for Other Healthcare Systems

We follow the order of the datasets reported in Table 2 in the paper.

- (1) Inpatient admission requests include 243,686 inpatient admission requests in one of the largest hospital in China from 262 cities in 2013.
- (2) Inpatient admissions refer to inpatient admissions of 11 major clusters (i.e., Cardiology, Dental, EDTU, Ear-Nose-Throat, Eye, Medicine, Gynaecology-and-Obstetrics, Oncology, Orthopedics, Paediatrics and Surgery) into a general hospital in Singapore in 2010.

- (3) Inpatient admissions into an Israel hospital were from a large hospital in Israel, over a period of 2 years (January 2004 – December 2005). This hospital consists of about 1000 beds.
- (4) ED arrivals refer to the patient arrivals of an emergency department (ED) of a university hospital in the US, starting from July 1st, 2010 to June 30th, 2011, which contains 44,217 arrival records in total.
- (5) Clinic appointments refer to appointment requests in a group of clinics from 28 postal districts in Singapore from 2012 to 2013.
- (6) Insured inpatient admissions include 125,585 inpatient admission records in 60 public hospitals in a provincial capital city of China from 2012 to 2013. Those inpatients are subsidized by the medical insurance for urban residents.
- (7) Hand-foot-mouth disease refers to all the cases of hand-foot-mouth disease in a prefecture-level city of China from 2010 to 2012.
- (8) Dysentery disease refers to all the cases of dysentery in a prefecture-level city of China in 2013.
- (9) Ambulance service refers to monthly ambulance calls from 83 postal districts in one of the largest cities of China in 2011.
- (10) Orthopedics surgery materials refer to the sales data from a company that sells orthopedics surgery products. There are 5,909 products in the inventory, which can be further aggregated into 1,035 types of products. The dataset includes all the sales data from June 2012 to April 2013.