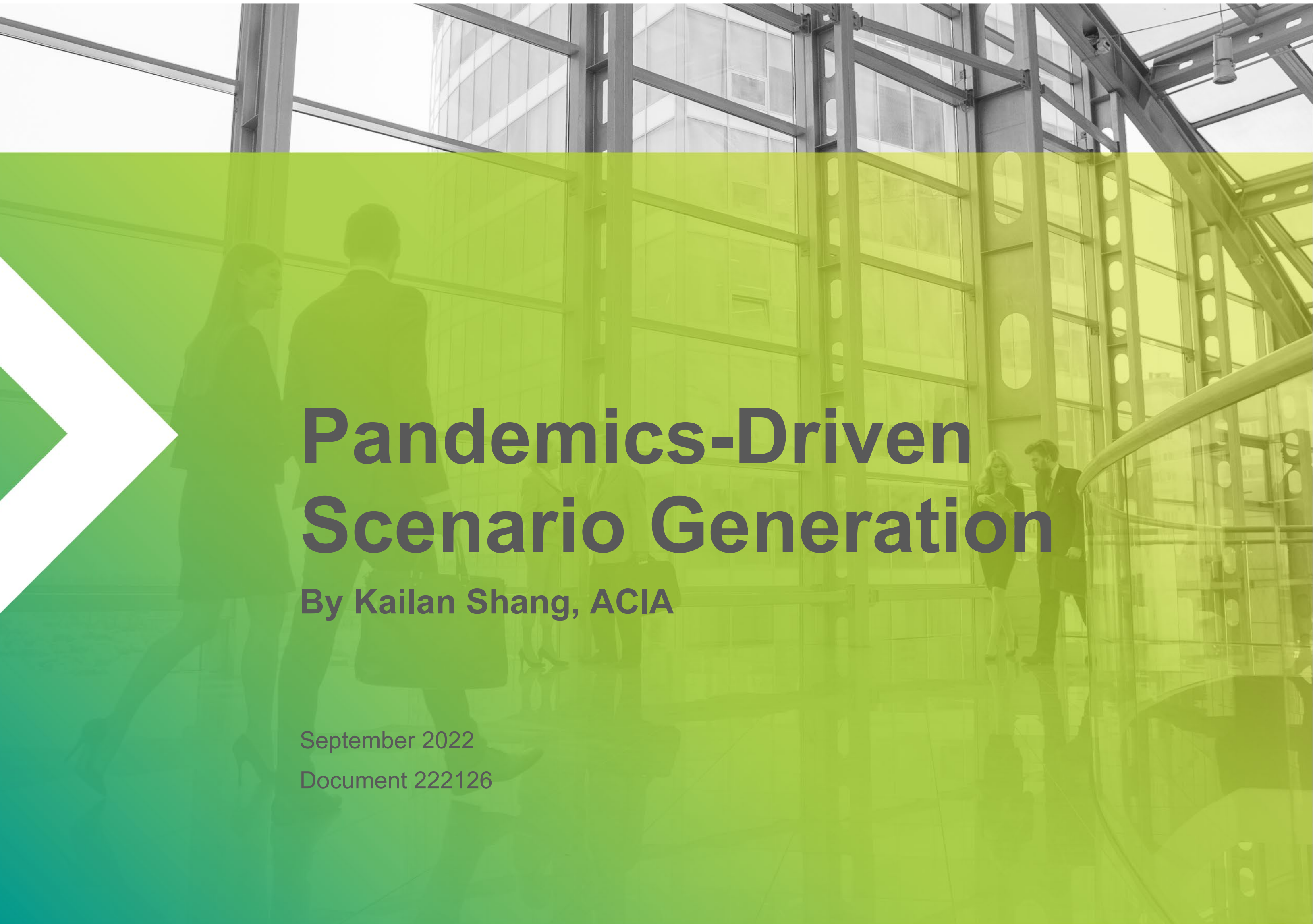
Canadian Institute of Actuaries
Institut canadien des actuaires



Pandemics-Driven Scenario Generation

By Kailan Shang, ACIA

September 2022
Document 222126

Table of contents

Executive Summary

Pandemic and epidemic events cause losses to the insurance industry not only through insurance claims but also disturbances to social and economic activities. Stress testing is typically used by insurance companies to assess the exposure to pandemic risk. The stress scenarios are usually based on single historical events such as the 1918 Spanish flu, or a pandemic model which uses epidemiological models such as the SIR (susceptible, infectious, and removed) model and its variants. However, the stress-testing approach may be insufficient as it does not reflect the distribution of pandemic events and their financial impacts. An integrated scenario generator that incorporates insurance risk factors, economic risk factors, and capital market factors can provide a more holistic view of pandemic events.

To supplement the existing stress-testing approach, an analysis of the pandemic and epidemic events in recorded human history has been undertaken. More than 800 historical events were used to study the frequency, severity, and duration of pandemics/epidemics. Pandemic and epidemic events can be modelled by using different statistical distributions, extreme value theory, and state-dependent models such as hidden Markov models.

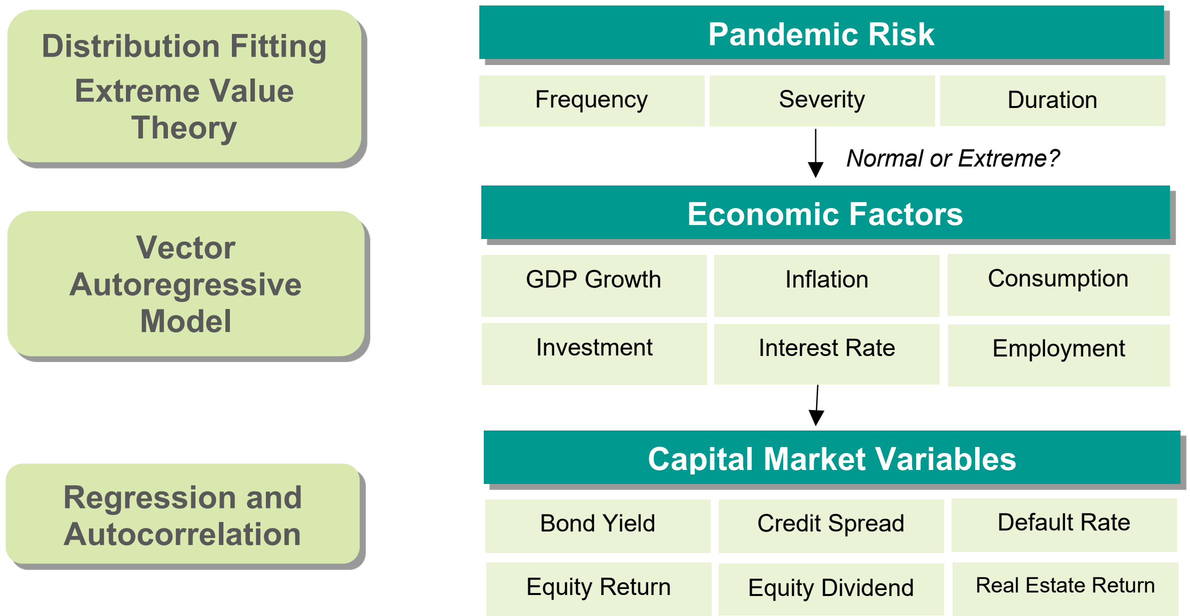
The impact of pandemic and epidemic events on the economic system can be material, as witnessed in the recent COVID-19 outbreak. Therefore, a study of the relationships between discrete events and their impact on economic system has been conducted:

1. The relationship between pandemics and gross domestic product/population growth was studied using more than 2,000 years of history, while acknowledging that economic and social structural changes are likely to change the future pattern significantly.
2. The quarterly macroeconomic data and capital market data of the past 70 years were used to quantify both the contemporary and temporal relationships, given the occurrence of a pandemic event.

With the analysis of both the pandemic risk and its relationship with the economy, a pandemics-driven scenario generator (PSG) has been designed based on the findings. It uses a basket of fitted distributions to generate pandemic events including frequency, severity, and duration. The generated pandemic events will then drive the generation of macroeconomic factors based on their severity. Macroeconomic factors are then reflected in the generation of capital market variables such as bond yield, credit spread, default rate, and equity return.

Figure E.1

Pandemics-Driven Scenario Generation Framework



The PSG covers pandemic risk factors, economic factors, and capital market variables that are needed to quantify the impact on claim experience, capital adequacy, new business, and so on. It also reflects the relationships among these risk factors observed in the past. Methods such as multiple correlation matrices, copulas, or even structured models are tested against historical data. Structured models represent the evolvement of variables and their relationships using formulas in an explicit way, and are recommended to reflect both contemporary and temporal nonlinear relationships.

This paper relies on historical data for PSG calibration, to present an objective view as much as possible. The generated scenarios show similar features and relationships to those observed in the historical data. However, the PSG is built to be flexible, to use different views of the pandemic risk and its impact on the economic system and insurance business. This can be achieved by either changing the data input for PSG calibration or setting the parameters directly based on alternative analysis.

In summary, this research contributes to the existing literature in three ways. First, it collects and analyzes past pandemic and epidemic events in recorded human history since 1050 BC. While acknowledging some missing information for many early events, the dataset is helpful for frequency and severity analysis.

Second, it provides a holistic PSG that explicitly models pandemic/epidemic events and, more importantly, their interactions with the economic system, for its users to evaluate the possible financial impact and probability of these events on the insurance business.

Third, sample implementation codes are made public for educational purposes and hosted at https://github.com/windwill/pandemic_esg.

Although this research uses as much data as are available globally, different views of data relevancy and geographical focuses can be reflected by adjusting data inputs or model parameters and applying the analysis leveraging on the R codes.

1 Introduction

A pandemic not only poses dangers to human lives but can also disrupt medical systems, economic activities, and human behaviours. In the recent COVID-19 outbreak around the globe, we observed events that were unique and often not covered by existing economic scenario generators (ESGs) and stress scenarios. This is understandable because the historical data most ESG calibration relies on do not include a severe pandemic. Neither the 2009 H1N1 pandemic nor the Ebola epidemics are comparable to COVID-19 in terms of its severity and economic impact, which are unlikely to be covered in the scenario generator before 2020. A revisit of pandemic risk modelling and its economic impact is needed.

A pandemic is when a new infection spreads to multiple countries and continents at the same time, infecting a lot of people. An epidemic is an outbreak that is larger than usual but stays confined to a single location or region. Epidemic events can be devastating even though they affect relatively fewer people. For example, in 2003 SARS (severe acute respiratory syndrome) was considered an epidemic because of its low infection rate of 8,000 people, but cost the world economy more than 50 billion dollars.

The recorded history of pandemic and epidemic events dates to 11th century BC when the plague of Ashdod struck the Philistines when they tried unsuccessfully to conquer the Hebrews, as explained in Kohn (2008). While many past diseases have been eradicated due to medical advancement, unknown or new viruses keep emerging, and it takes time to study them and find effective treatments or preventive measures. Historical experience can still provide some insights into the possible scenarios of future pandemic events. For example, the impact of influenza pandemic events in the 20th century has been observed in mortality data and, to some extent, has been reflected in actuarial assumptions, but it is beneficial to look at all of human history and consider other possible scenarios.

In addition to insurance risk, it is also important to assess how these scenarios will play out with current socio-economic conditions. The economic and social impact of coronavirus is unprecedented and provides ample data to study the resulting economic risk. Pandemics often involve highly correlated underlying risk drivers across space and time. For example, given the disruption to economic activities, an extreme financial event may well involve a bear equity market, a low-interest-rate environment, a poor credit environment, and expansionary fiscal and monetary policies. For insurers, this extends to insurance risk, including mortality, morbidity, unemployment insurance and lapses, and business risk, such as business disruption and product demand.

Leveraging on the historical data of both pandemic/epidemic events and their economic, financial, social, and business impact, this research builds a comprehensive PSG that explicitly reflects the impact of potential pandemic events. The components of building the PSG and its results are discussed below:

- Section 2 (The History of Pandemics and Epidemics) summarizes more than 800 pandemic and epidemic events recorded in human history, and provides an overview of the frequency, severity, and duration of each using descriptive statistics and graphs. This provides helpful information to determine how statistical models can be used to represent the patterns in the data, as presented in the following section.

- Section 3 (Pandemic Frequency and Severity) studies the frequency and severity of pandemic events. Distribution fitting is performed using historical data recognizing different measures of severity and different distributions for fat tails. Fitted distributions are used to simulate pandemic events as part of the PSG.
- Section 4 (Correlation) analyzes the relationships among different risk types during pandemics, including gross domestic product (GDP) growth rate, inflation rate, unemployment rate, consumption, investments, central bank rates, government bond yields, credit spreads, default rates, equity returns, and fiscal spending, wherever possible. Correlation matrices, copulas, and structured models are used to study both contemporary and temporal relationships.
- Section 5 (Scenario Generation) describes the PSG based on frequency, severity, and correlation modelling. Consistent pandemic scenarios, including insurance risk factors, economic factors, and capital market variables, are generated. The scenarios can be used to quantify both the short- and long-term financial impact on insurers. Reasonableness checks of simulated scenarios and methods of adjustment are also discussed.
- Section 6 (Conclusion) summarizes the key points of this research and concludes the main body of the report.
- The appendices contain the technical details of the PSG and its calibration. The information can be helpful for any adjustment or recalibration of the PSG.

Readers who are only interested in the PSG for simulation purposes can start reading from Section 5, and skip sections 2 to 4, which focus on data inputs and model calibration.

2 The History of Pandemics and Epidemics

The written history of pandemics and epidemics dates back to 1050 BC., as documented in Kohn (2008), when an outbreak of plague happened in the region that is now Israel. It became known as the Philistine Plague, or the Plague of Ashdod. Since then, hundreds of pandemic and epidemic events have been recorded, which provides good references to analyze patterns, assess impacts, and offer insights for the future.

2.1 Pandemic Diseases

The diseases that caused pandemics and epidemics are countable. One disease may cause several major pandemic events, with its cure or vaccine being gradually developed with medical advancement. To better utilize historical data for predicting the future, it is important to have a basic understanding of these diseases. Table 1 describes the main types of diseases that have caused pandemics and epidemics with no fewer than 10 recorded outbreaks where a sudden increase in the number of cases of a disease occurred in the affected region(s).

Table 1

Pandemic Diseases

Disease	No. of recorded outbreaks	Description	Earliest recorded	Latest outbreak	Deadliest outbreak
Plague	127	Plague is an infectious disease caused by Yersinia pestis bacteria, usually found in small mammals, and is transmitted when people are bitten by fleas. The most common form of plague is called the bubonic plague, which is a serious infection of the lymphatic system that causes fever, vomiting, bleeding, organ failure, and open sores. In a septicaemic plague, the disease reaches the bloodstream and creates clots, preventing blood from reaching the tissues. In a pneumonic plague, the disease reaches the lungs, which can lead to a lot of coughing and rapid death.	Philistine Plague (1050 BC)	Madagascar Plague (2008)	Black Death (mid-14th century), with 50 million deaths
Cholera	102	Cholera is a virus caused by a bacterium called Vibrio cholerae. Cholera causes vomiting and diarrhea. As early as 12 hours after the infection, the bacteria produce toxins that bind to small intestine enzymes that control water secretion from the rest of the body. The toxin makes the enzymes flood the intestines with water, and that leads to dehydration so severe that it can cause death unless treated properly.	Asiatic cholera pandemic (1817–23)	Yemen cholera epidemic (2016–21)	Japanese cholera epidemics (1858–59), with 3 million deaths
Smallpox	95	Smallpox was a disease caused by the variola virus. It started with headaches and small fevers. Then small bumps full of infectious fluid would appear all over the victim's body. In survivors, the pox would eventually scab and turn into scars.	Antonine Plague (Plague of Galen) (AD 165–80)	Indian smallpox epidemic (1973–74)	Mexican smallpox epidemic (1520–21), with 2.75 million deaths

Disease	No. of recorded outbreaks	Description	Earliest recorded	Latest outbreak	Deadliest outbreak
Influenza	66	Influenza has had many different outbreaks. Some symptoms of influenza are fever, cough, sore throat, a runny or stuffy nose, muscle or body aches, headaches, fatigue, and vomiting and diarrhea. There are more than 100 subtypes of influenza, and the ones in circulation change from season to season. The flu's genetic code also makes it mutate more quickly than any other virus. The flu spreads by turning a host's own cells into viral production factories. When the virus is engulfed by a host cell, it expels its genetic material, which makes its way to the nucleus. There, cells that normally copy the host's genes start copying viral genes instead, creating more and more copies of the virus. New viruses are repackaged and stuffed into the cell until it bursts, sending fresh new influenza viruses to infect other cells.	Roman Pestilence (212 BC)	U.S. flu season (2017–18)	Spanish influenza pandemic (1917–19), with 21 million deaths
Typhus	63	Typhus, also known as typhus fevers, "are a group of diseases caused by bacteria that are spread to humans by fleas, lice, and chiggers. Typhus fevers include scrub typhus, murine typhus, and epidemic typhus. Chiggers spread scrub typhus, fleas spread murine typhus, and body lice spread epidemic typhus. The most common symptoms are fever, headaches, and sometimes rash."	Cava typhus epidemic (1083)	Japanese–Korean typhus epidemic (1945–46)	German typhus epidemics (1813–14), with 469,000 deaths

Disease	No. of recorded outbreaks	Description	Earliest recorded	Latest outbreak	Deadliest outbreak
Yellow fever	44	Yellow fever is caused by flavivirus, a viral disease spread by mosquitoes. Most people exposed to flavivirus experience symptoms such as fever, chills, aches, fatigues, and vomiting. Around 15% of infected patients get much worse symptoms, such as jaundice, bleeding, and multi-organ failure that could lead to death. It gets transmitted when people are bitten by infected mosquitoes.	Barbadian yellow fever epidemic (1647)	Nigerian yellow fever epidemic (2020)	Rio de Janeiro yellow fever epidemics (1849–1902), with 70,000 deaths
Measles	40	”Measles is a highly contagious, serious disease.” It is “caused by a virus in the paramyxovirus family and it is normally passed through direct contact and through the air.” Symptoms include a high fever, a runny nose, a cough, red and watery eyes, and a rash	Plague of Cyprian (251 AD)	Congo measles epidemic (2019)	Fiji Islands measles epidemics (1875), with 40,000 deaths
Dengue fever	33	Dengue fever is a mosquito-borne tropical disease caused by the dengue virus. Those stricken are attacked by the virus within a short time (three to 12 days) after the crucial mosquito bite. Their symptoms include severe frontal headaches and excruciating pains in their joints (hence “breakbone fever”), as well as a rash, chills, nausea, and prostration.	U.S. and Caribbean dengue epidemic (1826–28)	Asia-Pacific and Latin America dengue epidemic (2019–20)	Vietnamese dengue epidemics (1970s), with 1,424 deaths
Poliomyelitis	30	“Poliomyelitis is a disabling and life-threatening disease caused by the poliovirus. The virus spreads from person to person and can infect a person’s spinal cord, causing paralysis.”	Stockholm poliomyelitis epidemic (1887)	Angolan poliomyelitis epidemic (1999)	U.S. poliomyelitis epidemic (1916), with 7,000 deaths

Disease	No. of recorded outbreaks	Description	Earliest recorded	Latest outbreak	Deadliest outbreak
Malaria	27	Malaria is a disease caused by a parasite. The parasite is spread to humans through the bites of infected mosquitoes. Symptoms include violent chills, nausea, high fever, and sweats, and those infected may die in a coma or delirium without prompt treatment.	Henry IV's army epidemics (1081–83)	South African malaria epidemic (1999–2000)	Punjab malaria epidemic (1908), with 307,316 deaths
Cerebrospinal meningitis (CSM)	15	"Meningitis is an inflammation (swelling) of the protective membranes covering the brain and spinal cord. A bacterial or viral infection of the fluid surrounding the brain and spinal cord usually causes the swelling. However, injuries, cancer, certain drugs, and other types of infections also can cause meningitis." Symptoms include fever, severe headaches, and a stiff neck. Children often suffer convulsions. People have died after serious symptoms of vomiting, bleeding, and collapse.	Albenga meningitis epidemic (1815)	West African meningitis epidemic (2009–2010)	Nigerian meningitis epidemics (1949–50), with 16,055 deaths
Ebola	14	The disease caused by the Ebola virus is a hemorrhagic fever. It starts by attacking the immune systems by neutralizing its cells, allowing the virus to spread. Starting from two to 20 days after contracting the disease, symptoms like a high temperature, aching, and sore throat appear, but quickly escalate to vomiting, rashes, and diarrhea. As the virus spreads, it infects the lymph nodes and vital organs, causing them to lose function. Ebola only exists in bodily fluids. In order to spread, these must be transmitted from an infected patient to another's body through passageways.	Sudanese Ebola outbreaks (1976)	Congolese Ebola outbreak (2020)	Western African Ebola epidemic (2013–16)

Disease	No. of recorded outbreaks	Description	Earliest recorded	Latest outbreak	Deadliest outbreak
HIV/AIDS	14	"HIV (human immunodeficiency virus) is a virus that attacks the body's immune system. If HIV is not treated, it can lead to AIDS (acquired immunodeficiency syndrome). There is currently no effective cure. Once people get HIV, they have it for life." However, treatments are available to keep the virus under control, and patients may still live a long and healthy life.	HIV/AIDS pandemic (1975–2010)		25 million deaths
Diphtheria	11	"Diphtheria is a serious infection caused by strains of bacteria called Corynebacterium diphtheriae that make toxin (poison). It can lead to difficulty breathing, heart failure, paralysis, and even death."	Paris diphtheria epidemic (1576)	Russian diphtheria epidemic (1990s)	Italian diphtheria epidemic (1618), with 8,000 deaths
Dysentery	10	Dysentery is an intestinal inflammation, primarily of the colon. It can lead to mild or severe stomach cramps and severe diarrhea with mucus or blood in the feces. Without adequate hydration, it can be fatal. Infection with the Shigella bacillus, or bacterium, is the most common cause.	Plague of Xerxes (480 BC)	Sri Lankan dysentery epidemic (1942)	French dysentery epidemic (1779), with 175,000 deaths

Although Table 1 only shows 691 outbreaks, 817 pandemic/epidemic events are used in this study to model the total frequency of all diseases.[1] It is notable that the ongoing coronavirus pandemic is not on the list of most frequent diseases. It is caused by 2019-nCoV, which is a type of SARS-related coronavirus. It is a relatively newly identified virus, and the number of outbreaks is fewer than 10. Although smallpox has been eradicated and there are effective treatments and vaccines for many of the other diseases in Table 1, they can be helpful for gauging the possible paths of new or future diseases that can lead to pandemic or epidemic events.

[1] The diseases include plague, smallpox, cholera, typhus, influenza, yellow fever, measles, poliomyelitis, malaria, dengue, cerebrospinal meningitis (CSM), HIV/AIDS, diphtheria, dysentery, Ebola, relapsing fever, scarlet fever, typhoid, hepatitis, beriberi, Japanese B encephalitis (JBE), sleeping sickness, whooping cough, encephalitis, mumps, bronchopneumonia, Legionnaires' disease (LD), Rift Valley fever (RVF), scurvy, tuberculosis, West Nile virus (WNV), CoV, anthrax infection, ergotism, hantavirus, kala-azar (visceral leishmaniasis), schistosomiasis, sweating sickness, syphilis, acute hemorrhagic conjunctivitis (AHC), African sleeping sickness (African trypanosomiasis), Bornholm disease (pleurodynia), Chikungunya virus fever, Creutzfeldt–Jakob disease (vCJD), dancing mania, English sweat, hemorrhagic fever with renal syndrome (HFRS), Kyasanur Forest disease, Lassa fever, leptospirosis, Machupo virus, Nipah virus, onchocerciasis, polyarthritis, profuse sweating, SARS, sexually transmitted diseases (STDs), sprue, venereal disease, visceral leishmaniasis, Q fever, Chikungunya, and Zika virus.

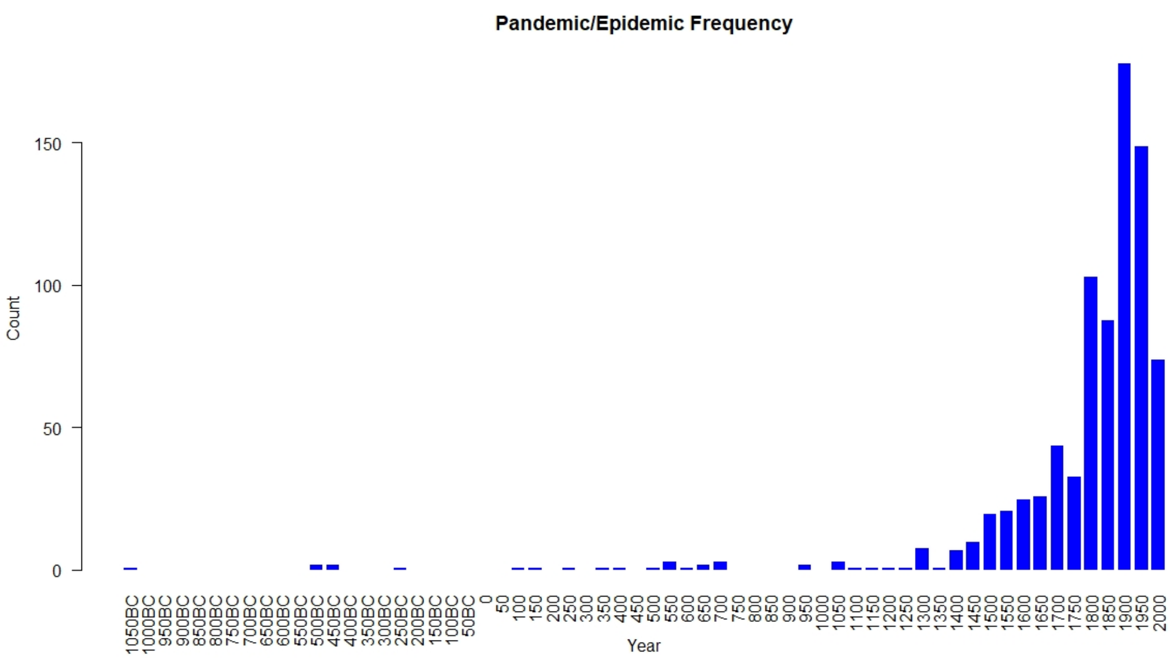
2.2 Historical Experience

Although more than 800 pandemic and epidemic events are recorded, the data quality varies by time and location. In the earlier stages of human history, fewer events may have been recorded. However, if one's interest is in pandemic events, this may not be a big concern given that the world was less connected at that time. The social, economic, and medical conditions of the human race also changed dramatically during the past 3,000 years. While studying the historical experience, we need to keep these structural changes in mind. This experience also indicates that recent history may be of more use for predictions of the future.

The 817 data records are composed of 30 pandemic events and 787 epidemic events. Figure 1 shows the frequency of pandemic/epidemic events every 50 years since 1050 BC.

Figure 1

Pandemic/Epidemic Frequency



In the medieval age, most of the pandemic/epidemic events happened during wars, which caused migration and unsanitary conditions. International trading activities contributed to the spread of diseases among continents to a large extent. Table 2 summarizes pandemic and epidemic events using descriptive statistics.

Table 2

Descriptive Statistics of Pandemic/Epidemic Data

Statistic	Duration (years)	Deaths	Confirmed cases	Case fatality rate	Mortality rate	Infection rate
No. of valid records	817	506	321	297	128	68
Min.	0	1	19	0.0%	0.0%	0.1%
Max.	200	50,000,000	400,000,000	100.0%	99.7%	100.0%
Median	2	3,572	12,767	16.7%	10.0%	21.1%
Mean	6	305,011	3,437,930	27.5%	17.3%	29.3%
Standard deviation	15	2,807,561	26,960,180	26.7%	21.7%	27.1%
Coefficient of variation	2.59	9.20	7.84	0.97	1.25	0.92

Notes:

1. Duration: Duration of the event measured in years; a few outliers include outbreaks consolidated as one incident covering two centuries
2. Case fatality rate: No. of deaths as a percentage of confirmed cases
3. Mortality rate: No. of deaths as a percentage of affected population
4. Infection rate: No. of confirmed cases as a percentage of affected population
5. Coefficient of variation: Standard deviation divided by mean which measures the dispersion of the distribution

Duration, case fatality rate, mortality rate, and infection rate are useful variables to measure the impact of a pandemic/epidemic event. Figure 2 shows the histogram of duration, and Figure 3 the distribution of three severity variables using boxplots.

Figure 2

Histogram of Pandemic/Epidemic Duration

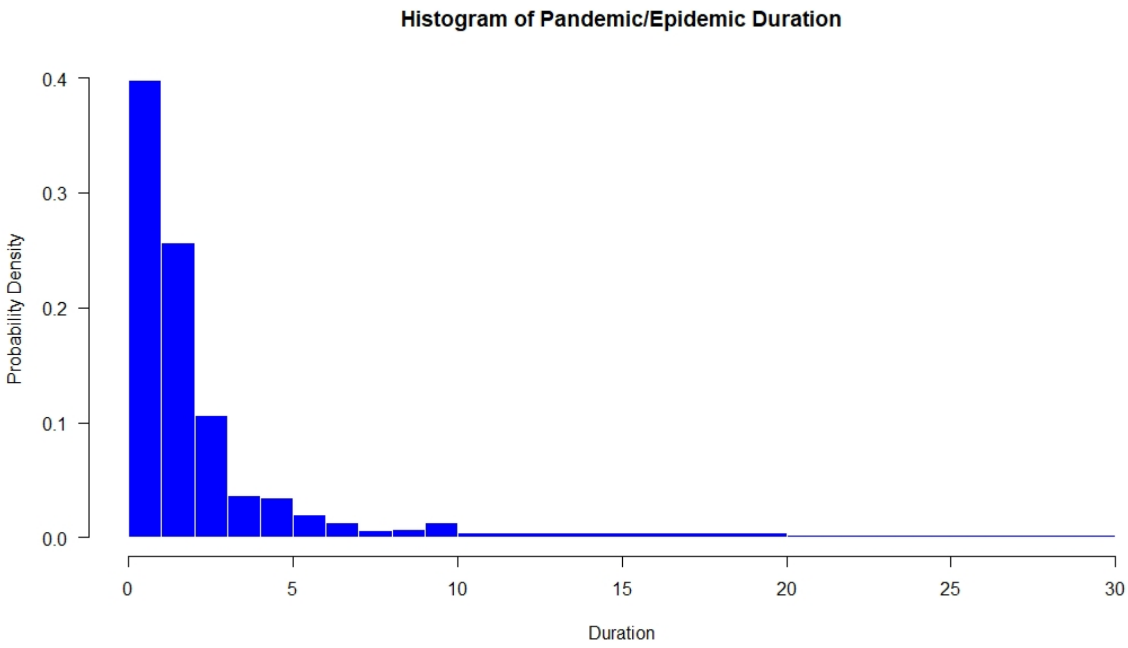
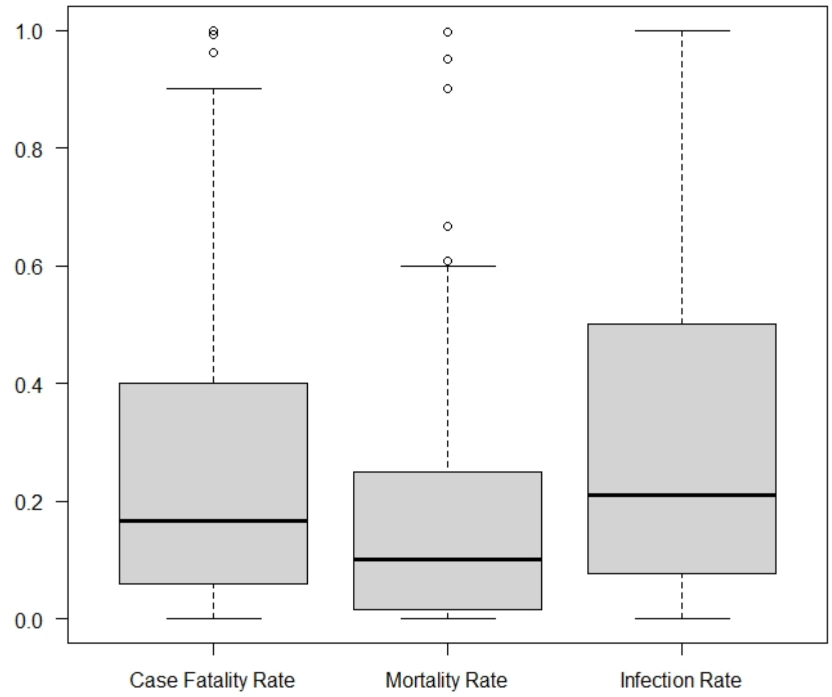


Figure 3

Boxplots of Pandemic/Epidemic Severity Measures

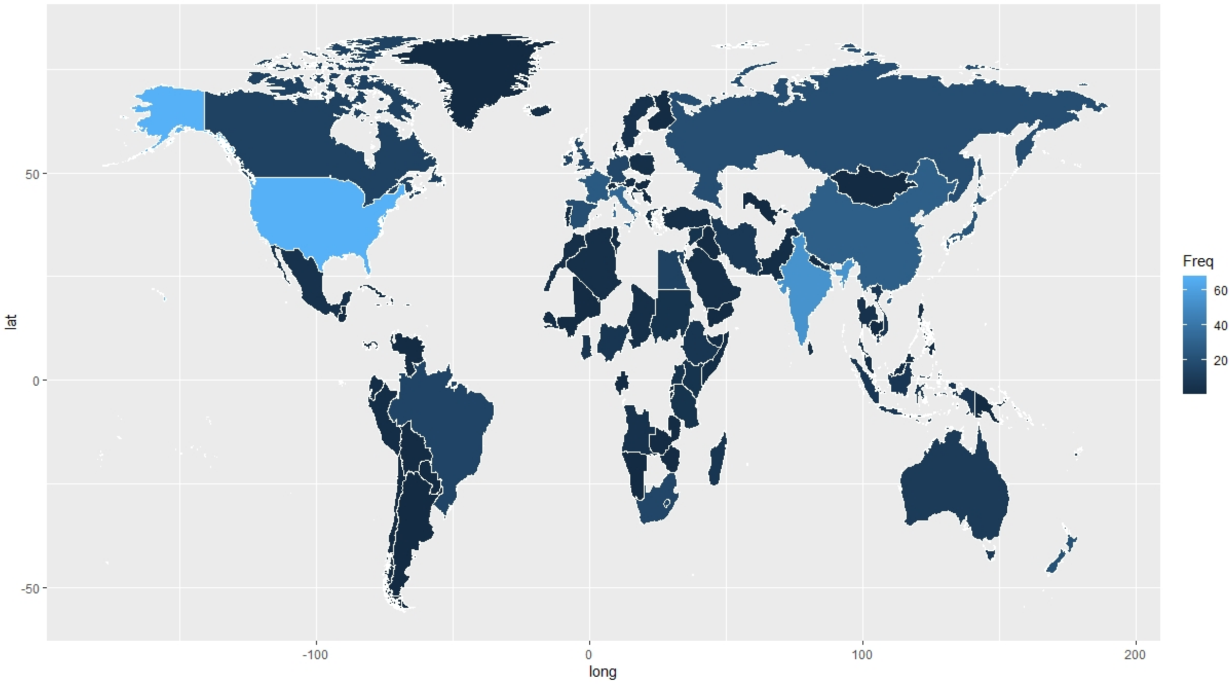


Each box represents the data between the first and third quantile (Q1 and Q3), with the line in the middle as the median. The box extends to a wider data range between [Q1 – 1.5(Q3 – Q1), Q3 + 1.5(Q3 – Q1)], where available. For outliers outside this range, they are plotted as individual dots. All rates are right-skewed with extremely severe incidents in the past.

Figure 4 shows the frequency map of pandemic/epidemic events based on the origin of the events. A specific region may be used for its own analysis, which can be more meaningful considering the different risk exposures for different regions.

Figure 4

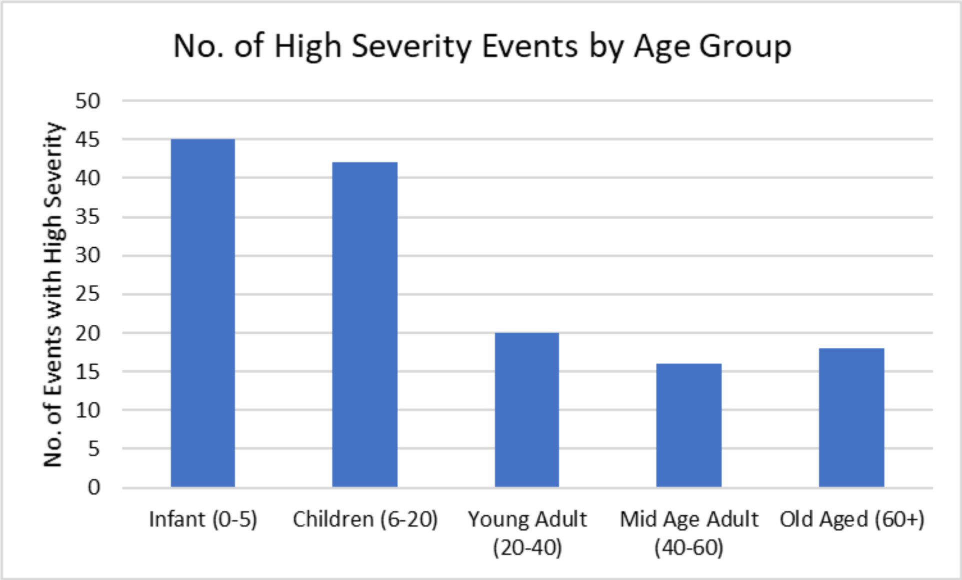
Map of Pandemic/Epidemic Events



The impact of pandemic/epidemic events varies by age group; for historical events where records of impacts on different age groups are available, Figure 5 summarizes the experience. Children are more vulnerable in general compared to adults. But for certain diseases, adults are the group that were more severely attacked.

Figure 5

Pandemic/Epidemic Impact by Age Group



Separate analyses of historical pandemic events that exclude epidemic events can be found in Appendix A.1.

3 Pandemic Frequency and Severity

With the historical data, a parametric approach is used to describe five key variables: frequency, duration, case fatality rate, mortality rate, and infection rate. This is the building block of the first part of the PSG: modelling the pandemic risk using distribution fitting and extreme value theory (EVT).

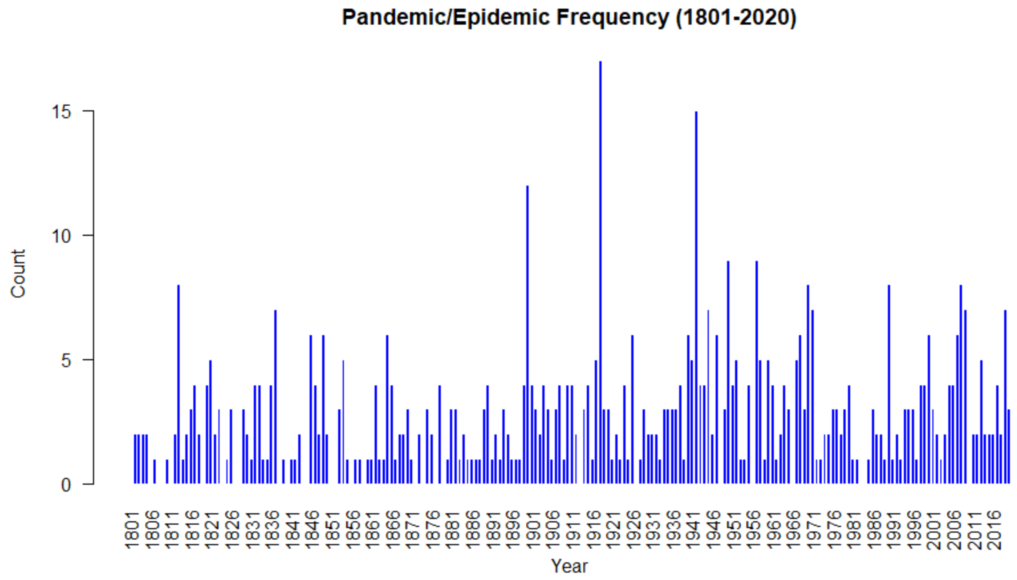
For each variable, a few distribution types are used to describe the historical data, with the results assessed using standard goodness-of-fit measures such as loglikelihood, Akaike information criteria (AIC), and Bayesian information criteria (BIC). In addition, empirical distribution and the chosen fitted distribution are compared to make sure the entire distribution, especially the tail behaviour, is captured.

3.1 Frequency

To model annual frequency, data between 1801 and 2020 are used, as shown in Figure 6. Recorded events are sparse before 1800 on an annual basis. Recent history may also be more relevant for predictions.

Figure 6

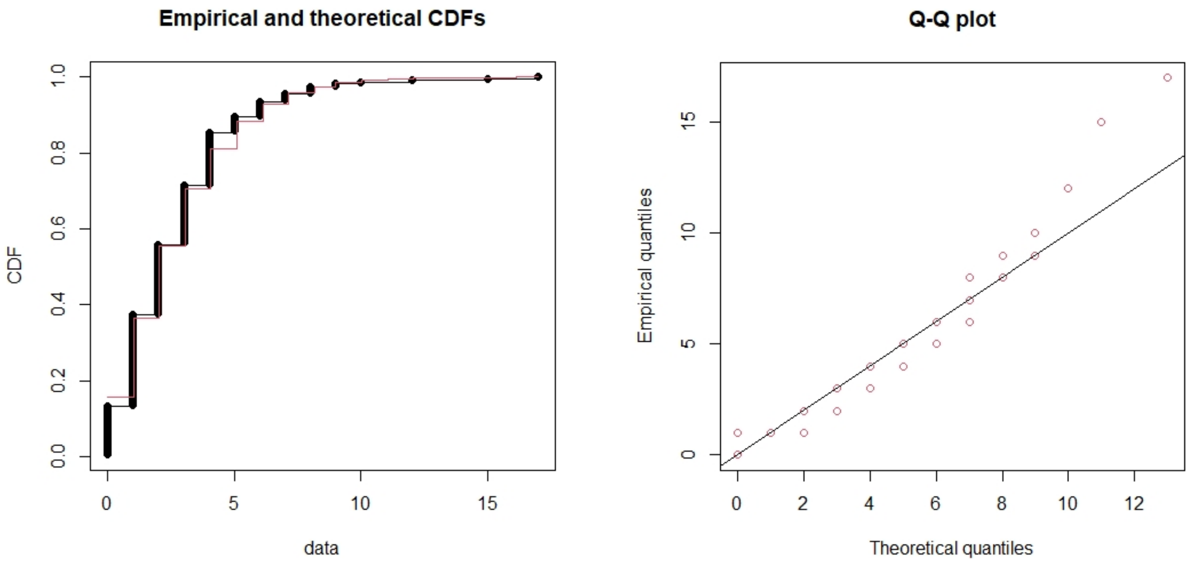
Pandemic/Epidemic Annual Frequency (1801–2020)



Compared to Poisson distribution and Geometric distribution, the negative binomial distribution has the highest loglikelihood, with a mean of 2.7 and a standard deviation of 2.3. Figure 7 compares empirical data with calibrated negative binomial distribution using a cumulative distribution function (CDF) and a quantile–quantile (Q–Q) plot.

Figure 7

Frequency Distribution Fitting Comparison: Negative Binomial Distribution

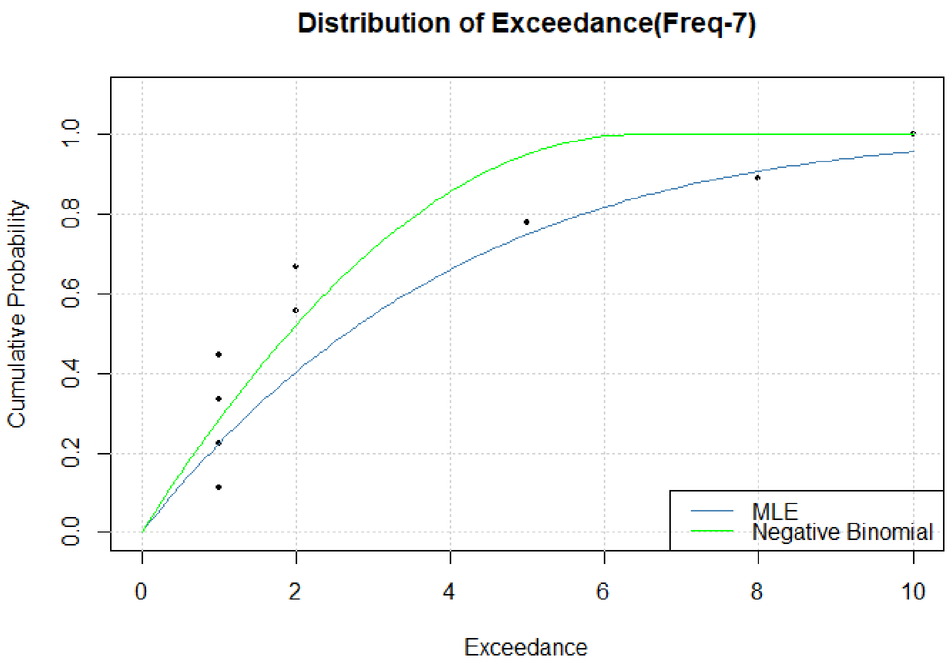


In the Q–Q plot, points above the line at the left end indicate a heavier left tail for the empirical data compared to the distribution. Points above the line at the right end indicate a heavier right tail for the empirical data. In order to capture the heavier tail, EVT can be used to model the extremes. EVT relies on the Fisher–Tippett–Gnedenko theorem and is formalized by Emil Julius Gumbel in his 1958 classic *Statistics of Extremes*. It can be used to model exceedance beyond a threshold and is called the “peak over threshold” (POT) method. In our case, a threshold of 7 is a possible choice beyond which a heavier tail is observed.

The distribution of exceedance (annual frequency subtracted by 7) based on the experience and fitted negative binomial distribution is illustrated in Figure 8. The Q–Q plot in Figure 7 showed a consistent underestimation beyond 7 by the negative binomial distribution. A higher threshold can lead to less satisfactory fit for both the negative binomial distribution due to the heavy tail and the generalized Pareto distribution (GPD) due to fewer data points. Note that the GPD fits the data much better than the negative binomial distribution.

Figure 8

Frequency Distribution Fitting: EVT



By combining the negative binomial distribution describing annual frequency no greater than 7 and the GPD modelling the exceedance beyond 7, the heavy tail can be fully captured while maintaining a goodness of fit for the entire distribution. More details of the frequency distribution fitting can be found in Appendix A.2.

An alternative approach is to use state-dependent models to describe the frequency pattern of the pandemic’s history. By modelling the state, an extra layer of complexity and flexibility is added to handle different frequency patterns. An example of using a hidden Markov model is included in Appendix A.5. Although it is not used in the example of the PSG in Section 5, it can be applied to represent more sophisticated frequency patterns, if desired.

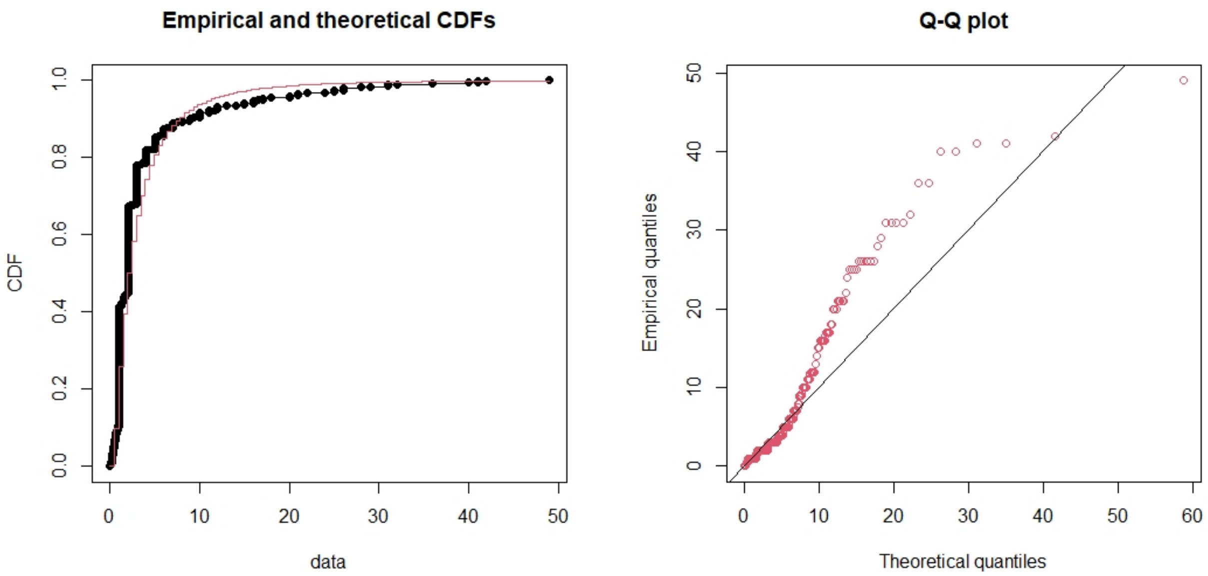
3.2 Duration

Some data points include several outbreaks in a long time period, and the detailed duration information for each outbreak is not available. These data points are excluded from the distribution fitting analysis to avoid the undesired impact of these inaccurate outliers. Only data points with a duration shorter than 50 years are included. Compared to exponential distribution, Gaussian distribution, gamma distribution, and Weibull distribution, the fitted lognormal distribution is a better choice in terms of loglikelihood, with a mean of two years and a standard deviation of 4.9.

Figure 9 compares the empirical data with the calibrated lognormal distribution using a Q–Q plot. The fitting is not perfect for durations larger than seven years. For pandemic events with longer durations, the lognormal distribution may underestimate the duration on average.

Figure 9

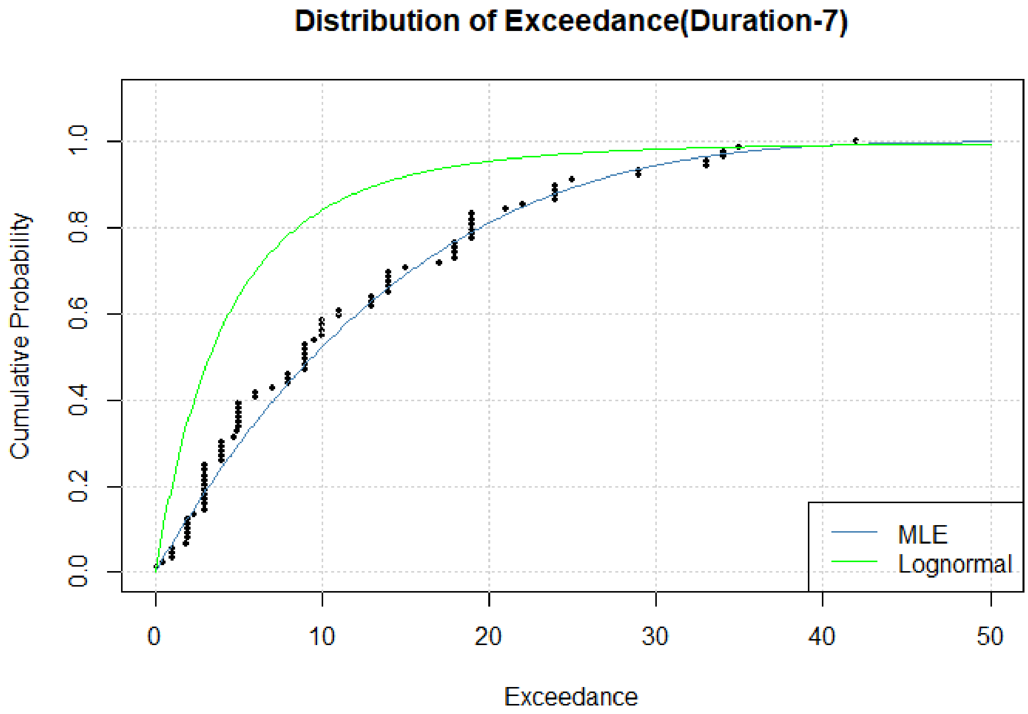
Duration Distribution Fitting Comparison: Lognormal Distribution



Similar to the frequency distribution, EVT can be applied to capture the heavy tail, as shown in Figure 10, using a threshold of seven years for the duration.

Figure 10

Duration Distribution Fitting: EVT



More details of the duration distribution fitting can be found in Appendix A.3.

3.3 Severity

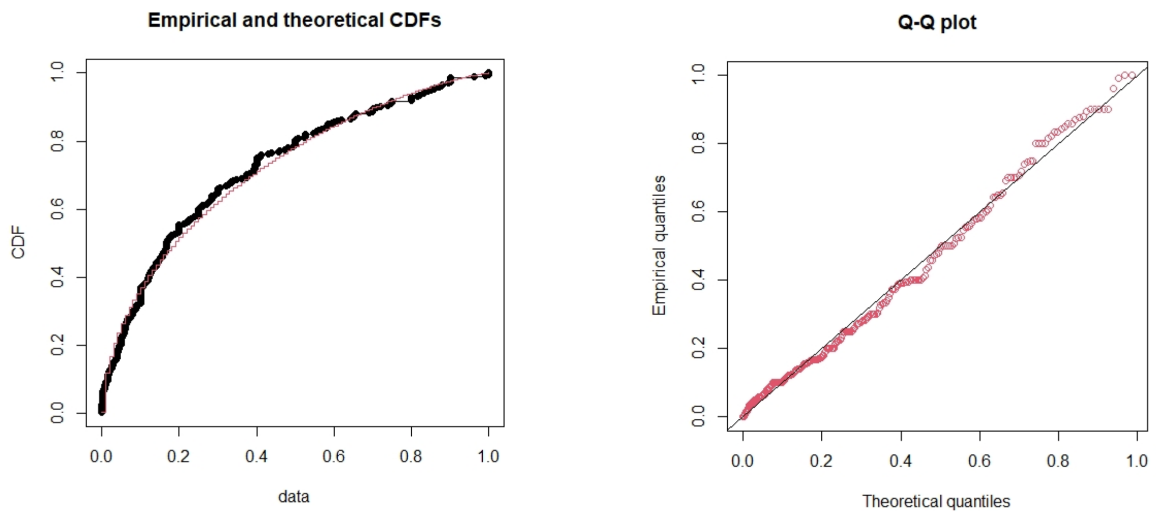
Three rate measures are used to analyze severity: case fatality rate, mortality rate, and infection rate. These rate measures are chosen to provide a consistent measurement among pandemic and epidemic events regardless of time and the geolocation of their occurrence. Other severity measures such as death count and number of people hospitalized depend on the populations of infected regions and are not directly usable for generating future events. Time-independent distributions are used because the historical data in general do not justify a time-varying assumption.

Case Fatality Rate

Compared to exponential distribution, Gaussian distribution, lognormal distribution, gamma distribution, and Weibull distribution, the fitted beta distribution is a better choice in terms of matching the entire distribution, with both the mean and the standard deviation as 27%. Figure 11 compares the empirical data with the calibrated beta distribution.

Figure 11

Case Fatality Rate Distribution Fitting Comparison: Beta Distribution

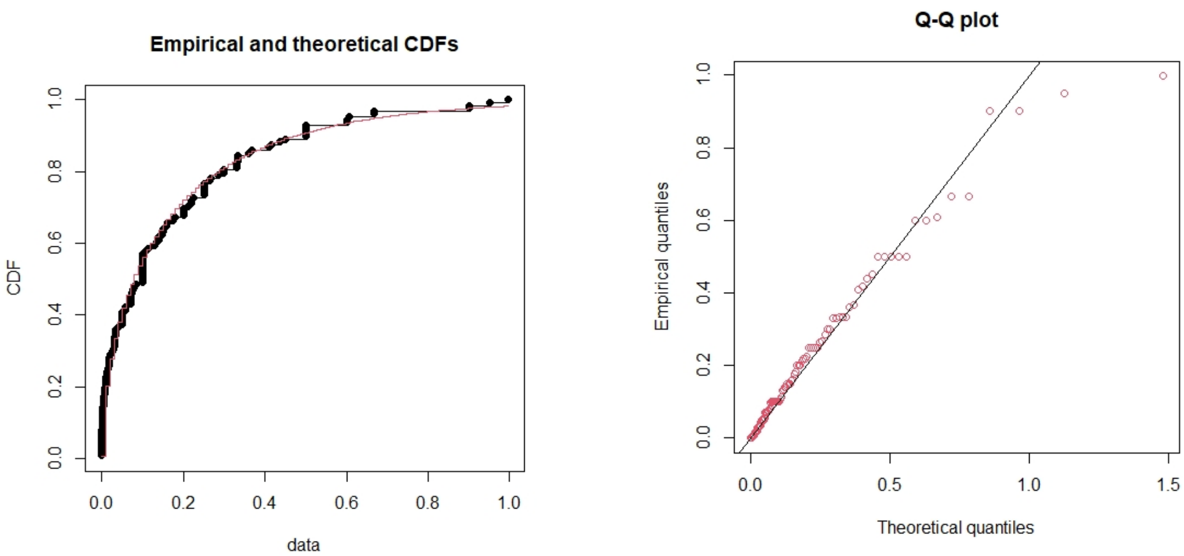


Mortality Rate

Compared to exponential distribution, Gaussian distribution, lognormal distribution, Weibull distribution, and beta distribution, the calibrated gamma distribution has the largest loglikelihood, with a mean of 17% and a standard deviation of 25%. Figure 12 compares the empirical data with the calibrated gamma distribution, without any underestimation of the tail risk.

Figure 12

Mortality Rate Distribution Fitting Comparison: Gamma Distribution

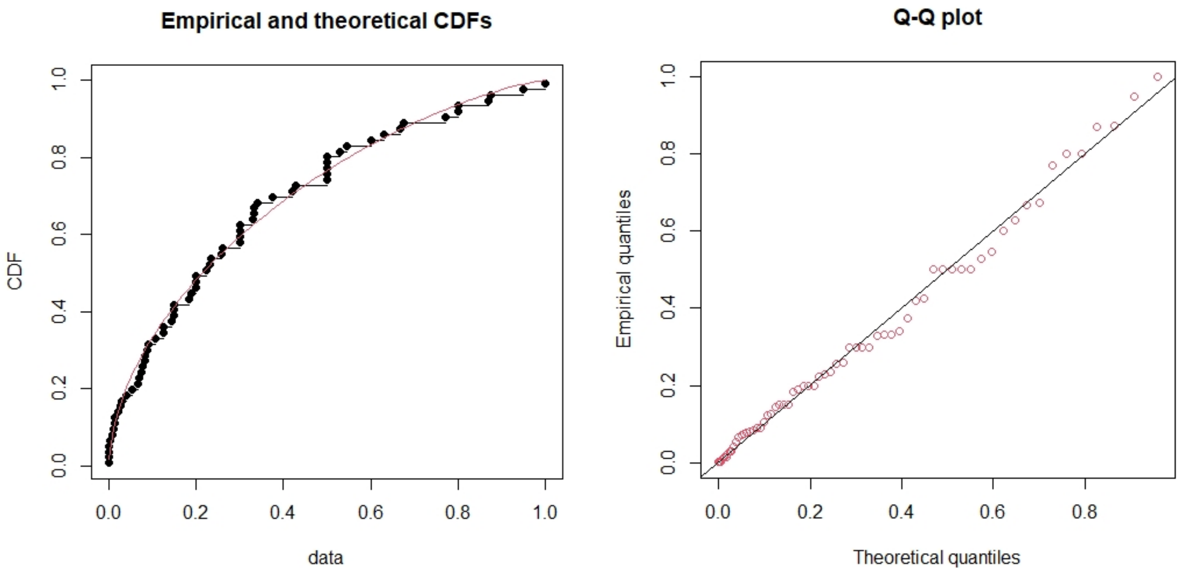


Infection Rate

Compared to exponential distribution, Gaussian distribution, lognormal distribution, gamma distribution, and Weibull distribution, the fitted beta distribution is a better choice in terms of matching the empirical distribution of the historical infection rate, with a mean of 29% and a standard deviation of 27%. Figure 13 compares empirical data with the calibrated beta distribution.

Figure 13

Infection Rate Distribution Fitting Comparison: Beta Distribution



More details of the severity distribution fitting can be found in Appendix A.4.

4 Correlation

The impacts of pandemic and epidemic events include not only the mortality experience, but also potentially significant economic consequences if the events cause social and economic disruptions. It is helpful to understand the possibility and magnitude of the economic impacts due to extreme pandemic/epidemic events. This is the second and third part of the PSG: modelling economic factors and capital market variables which depend on the occurrence of extreme pandemic events.

4.1 Extreme Pandemic Events and Economic Conditions

Severity measures such as death count and number of confirmed cases can be used to identify extreme events which had a profound impact on the historical economic and social system. Figures 14 and 15 show the total deaths and confirmed cases from the years 1500–2200, with 1500–1600, 1600–1700, and 1700–1820 condensed into the years 1600, 1700, and 1820. Data after 1820 are presented at an annual frequency, where available.

Figure 14

Total Deaths of Historical Pandemic/Epidemic Events (Years 1500–2020)

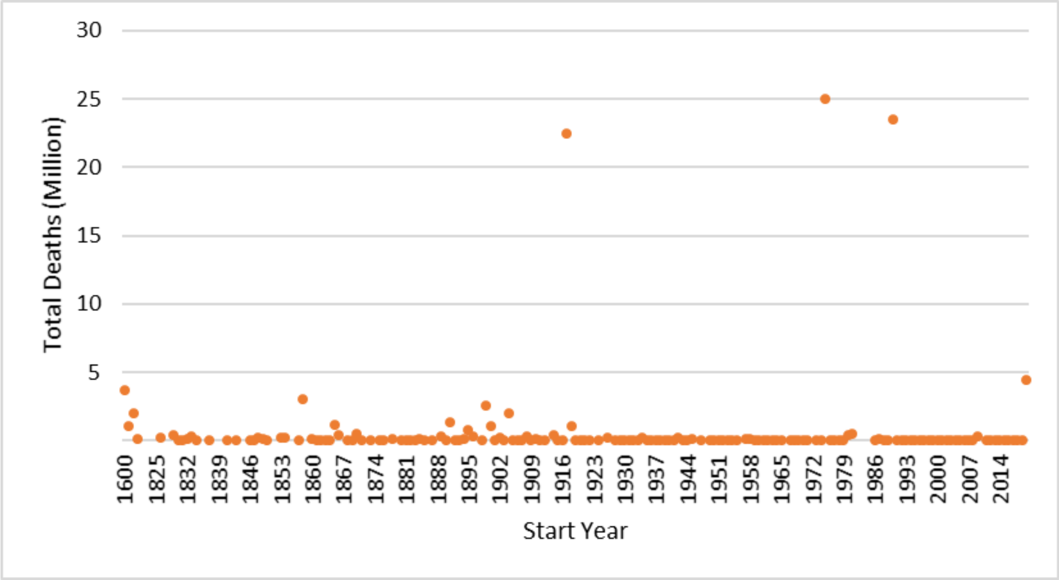
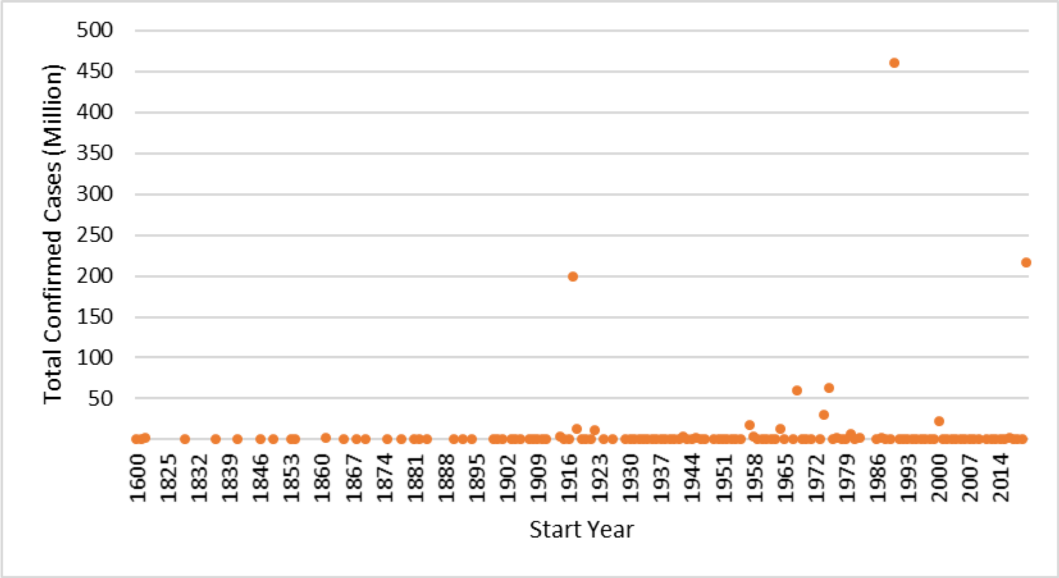


Figure 15

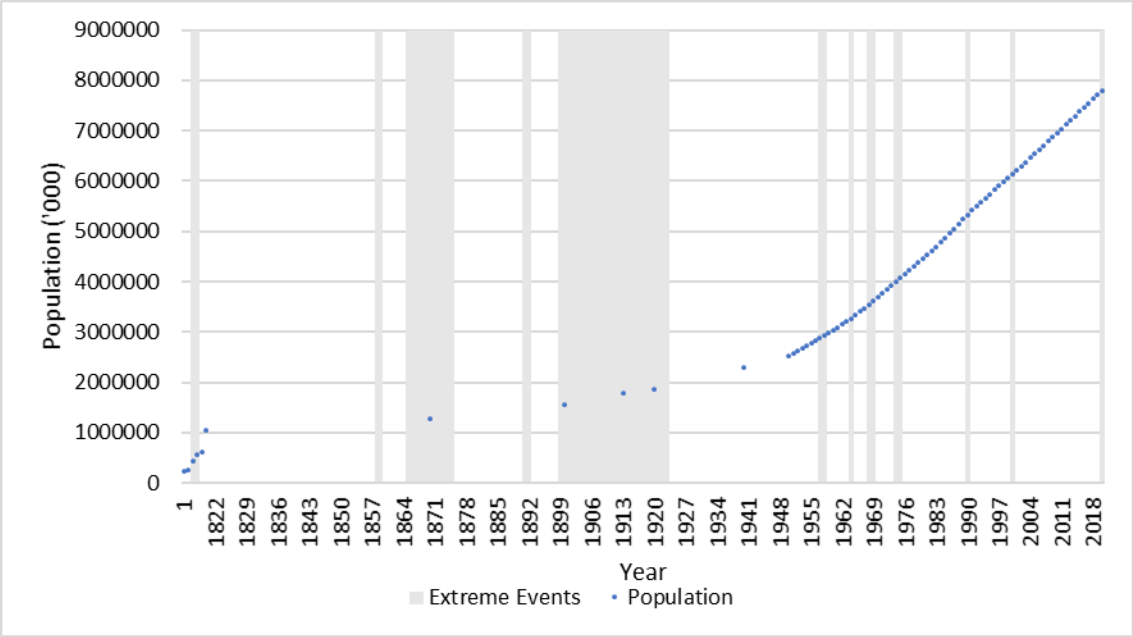
Total Confirmed Cases of Historical Pandemic/Epidemic Events (Years 1500–2020)



In this report, a threshold of 800,000 is used for death counts, and a threshold of 12,000,000 for case counts. When one or both conditions are met, the period is considered as an extreme period regarding pandemic/epidemic events. Figures 16 and 17 show the historical population and GDP per capita with the periods of extreme pandemic/epidemic events highlighted.

Figure 16

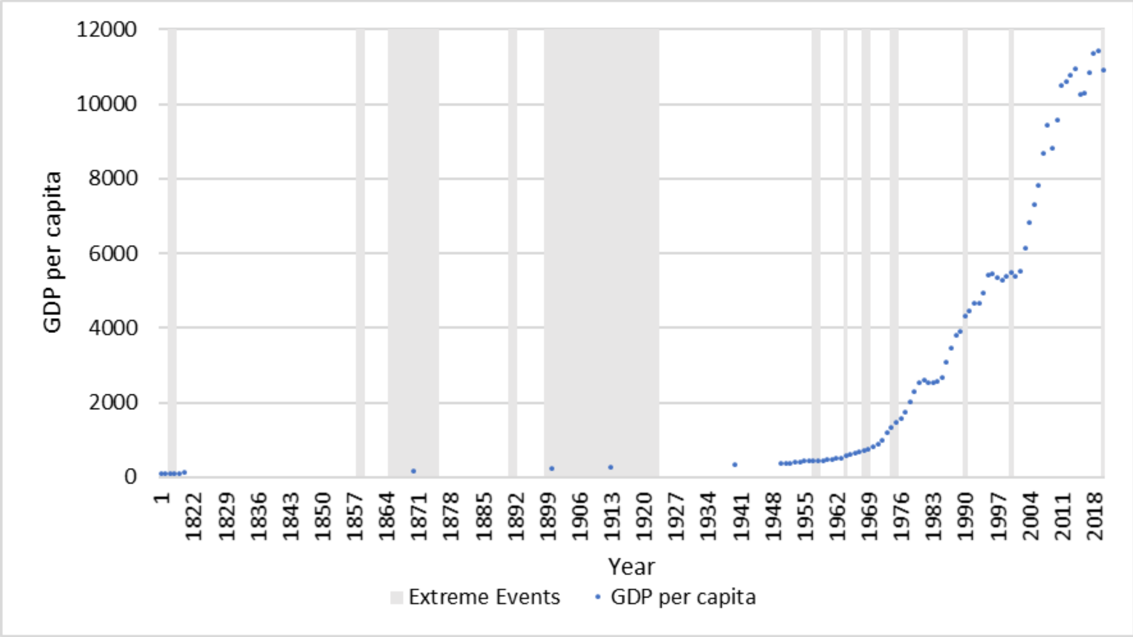
Historical World Population



Data sources: Maddison Project Database, version 2020; United Nations (UN) Database: https://population.un.org/wpp/Download/Standard/Population/

Figure 17

Historical World GDP Per Capita

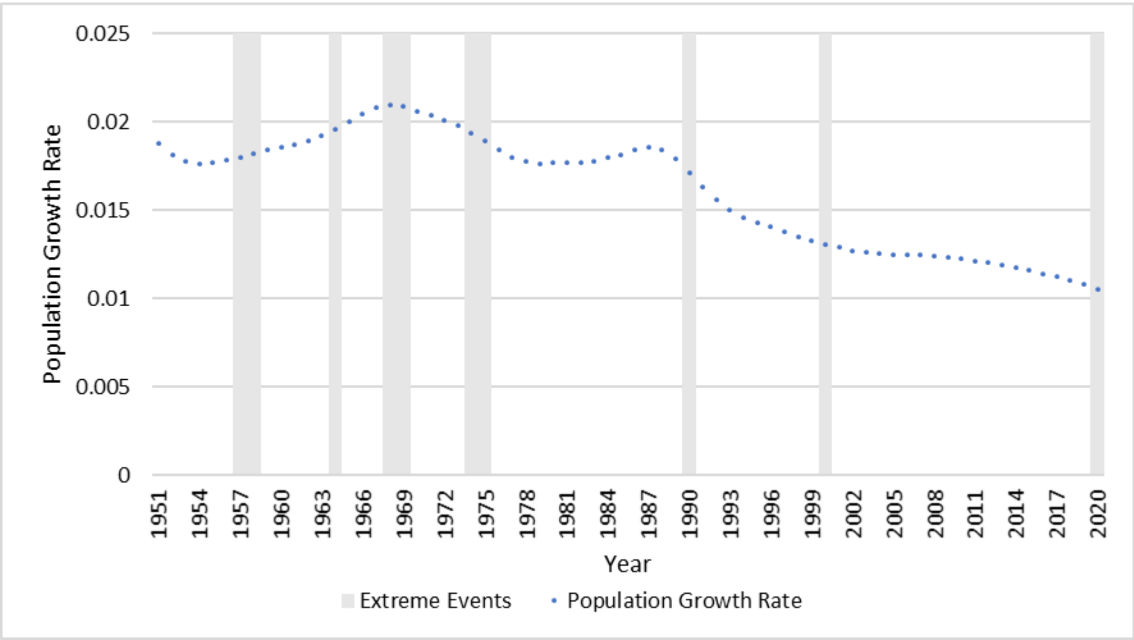


Data sources: Maddison Project Database, version 2020; World Bank Database:
https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

From 1 AD to 2020, it seems that pandemic and epidemic events had little impact on the long-term upward trend of world population, but they can lead to short-term disturbance to GDP growth. Figures 18 and 19 show the growth rates of population and GDP per capita in the past 70 years. While it is hard to tell any disruption of the world population growth by pandemic/epidemic events, it is observed that the growth of GDP per capita slowed down to different degrees during the influenza pandemic of 1957–58, the HIV/AIDs pandemic (1975–2010), and the recent COVID-19 pandemic (2020 to present), for certain periods.

Figure 18

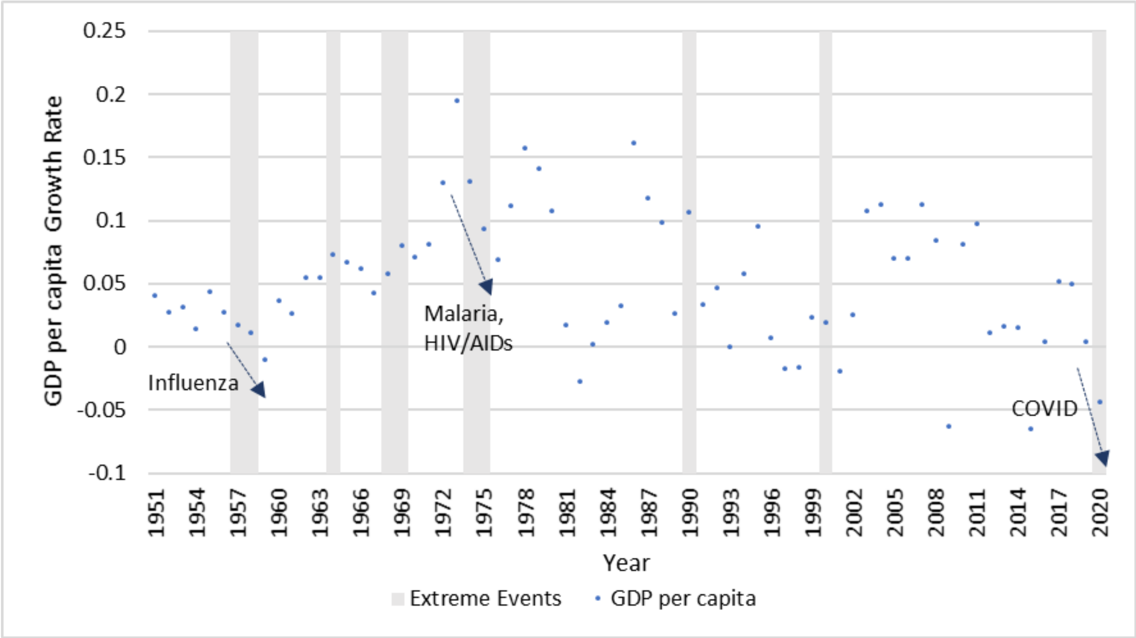
Historical Population Growth Rate (1951–2020)



Data sources: Maddison Project Database, version 2020; UN Database:
https://population.un.org/wpp/Download/Standard/Population/

Figure 19

Historical World GDP Per Capita Growth Rate (1951–2020)



Data sources: Maddison Project Database, version 2020; World Bank Database: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

On the other hand, many other risk factors play an important role in economic well-being and could be more influential than pandemics. Numerous studies have examined the economic impact of pandemic events. For example, *The Impact of AIDS* (United Nations Department of Economic and Social Affairs/Population Division, 2004) listed conclusions on the impact on GDP from a variety of research reports. The estimates varied from the statistically insignificant to a 2–4% annual reduction. These studies focused on attribution analysis. For our purpose, we still want to capture the economic patterns due to multiple risk factors while reflecting the impact of pandemic risk in an explicit way. Therefore, the focus is on different patterns with and without extreme pandemics. But these patterns are considered in a holistic way and attributed not only to pandemic events but also their domino effect.

When assessing the aggregated impact of a risk event, quantification of the diversification benefit is important. A small change in the correlation structure often leads to a significant change in the total required capital. The preferred way to model the relationships in an extreme event is to model them among their underlying risk drivers. For the purposes of this report, we define a risk driver to be a random variable that lends itself to univariate statistical modelling and simulation. Examples of risk drivers include the GDP growth rate, unemployment rate, equity return, and government bond yield, and the mortality rate for a particular risk class. To quantify the correlation between extreme event risk drivers, it is important that the practitioner first filter out non-extreme data because correlations associated with an extreme event can differ significantly from those observed during non-extreme periods.

Two approaches are available to model the correlation among risk factors: the statistical approach, including correlation matrices and copulas, and the structured approach.

4.2 Statistical Approach: Correlation Matrix

A correlation matrix contains the correlation coefficients among exposures to individual risk factors with the assumption of linear relationships.

$$\mathrm{RE}_{\mathrm{Total}} = \sqrt{(RE_1 \quad RE_2 \quad RE_3)\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}\begin{pmatrix} RE_1 \\ RE_2 \\ RE_3 \end{pmatrix}}$$

where

RE_{Total} is the aggregated risk exposure.
RE_i is the risk exposure for risk factor i.
ρ_{ij} is the correlation coefficient of risk factors i and j.

While one correlation matrix can define a set of linear relationships, ideally the correlation matrix at each confidence level is unique to reflect the nonlinear relationship in reality. However, because of the lack of data, it is difficult to construct them credibly. In addition, the correlation between risk drivers is not necessarily the same as the correlation between risk exposures. Therefore, it may need to be adjusted to reflect product features that can strengthen or weaken the relationship. Many relevant studies have been done by the insurance industry and regulators. In 2009 the CRO Forum issued an article, “Calibration recommendation for the correlations in the Solvency II standard formula,” that contains a suggested range of correlation among major risk types. The European Insurance and Occupational Pensions Authority (EIOPA) also provided information on the appropriate correlation matrix used in solvency requirement calculation at a confidence level of 99.5%. These can be used as a reference so that our customized correlation matrix does not deviate much from the industry standard. At the same time, a thorough analysis using pandemic data is needed to reflect any necessary differences. Table 3 shows the correlation matrix from 1871 to 2020 where data are available. Except for GDP and GDP per capita growth rates, U.S. data including macroeconomic variables and capital market information are used, given their availability.

Table 3

Economic Risk Correlation Matrix (1871–2020)

Variable	gdp_gr	gdppc_gr	inflation	tby_10yr	tby_1yr	sp500_rtn	sp500_divd	Baa_cs	Aaa_cs	BBB_default	ur	Fed_rate	pce	fpi	ge	mhp_gr	rent_gr
gdp_gr	1.00	0.99	0.33	0.32	0.39	-0.02	0.21	-0.10	-0.21	-0.12	-0.06	0.43	0.57	0.37	0.16	0.30	0.28
gdppc_gr	0.99	1.00	0.33	0.30	0.36	-0.03	0.17	-0.06	-0.16	-0.13	-0.05	0.41	0.54	0.35	0.15	0.30	0.29
inflation	0.33	0.33	1.00	0.27	0.23	0.13	0.00	0.26	-0.10	0.07	0.30	0.72	0.63	0.07	0.47	0.07	0.68
tby_10yr	0.32	0.30	0.27	1.00	0.84	0.02	-0.01	-0.01	-0.31	0.17	0.37	0.90	0.29	0.21	0.15	0.16	0.40
tby_1yr	0.39	0.36	0.23	0.84	1.00	0.00	0.17	-0.18	-0.48	0.10	0.15	0.98	0.27	0.30	0.22	0.18	0.42
sp500_rtn	-0.02	-0.03	0.13	0.02	0.00	1.00	-0.26	-0.22	-0.12	-0.35	-0.26	0.04	0.45	0.39	-0.02	0.20	0.20
sp500_divd	0.21	0.17	0.00	-0.01	0.17	-0.26	1.00	-0.28	-0.58	0.09	0.08	0.60	0.09	0.18	0.42	0.18	-0.03
Baa_cs	-0.10	-0.06	0.26	-0.01	-0.18	-0.22	-0.28	1.00	0.87	0.32	0.50	-0.13	-0.22	-0.43	0.03	-0.27	0.29
Aaa_cs	-0.21	-0.16	-0.10	-0.31	-0.48	-0.12	-0.58	0.87	1.00	0.28	0.31	-0.43	-0.44	-0.40	-0.27	-0.18	0.01
BBB_default	-0.12	-0.13	0.07	0.17	0.10	-0.35	0.09	0.32	0.28	1.00	0.17	0.12	-0.04	-0.42	0.38	-0.12	0.02
ur	-0.06	-0.05	0.30	0.37	0.15	-0.26	0.08	0.50	0.31	0.17	1.00	0.08	-0.08	-0.30	-0.19	-0.03	0.19
Fed_rate	0.43	0.41	0.72	0.90	0.98	0.04	0.60	-0.13	-0.43	0.12	0.08	1.00	0.76	0.34	0.63	0.17	0.74
pce	0.57	0.54	0.63	0.29	0.27	0.45	0.09	-0.22	-0.44	-0.04	-0.08	0.76	1.00	0.67	0.33	0.35	0.76
fpi	0.37	0.35	0.07	0.21	0.30	0.39	0.18	-0.43	-0.40	-0.42	-0.30	0.34	0.67	1.00	-0.08	0.46	0.25
ge	0.16	0.15	0.47	0.15	0.22	-0.02	0.42	0.03	-0.27	0.38	-0.19	0.63	0.33	-0.08	1.00	0.19	0.36
mhp_gr	0.30	0.30	0.07	0.16	0.18	0.20	0.18	-0.27	-0.18	-0.12	-0.03	0.17	0.35	0.46	0.19	1.00	0.17
rent_gr	0.28	0.29	0.68	0.40	0.42	0.20	-0.03	0.29	0.01	0.02	0.19	0.74	0.76	0.25	0.36	0.17	1.00

Notes:

- gdp_gr: World GDP growth rate (source: World Bank Database)
- gdppc_gr: World GDP per capital growth rate (source: World Bank Database)
- inflation: inflation rate based on the Consumer Price Index (CPI; source: Consumer Price Index for All Urban Consumers: All Items in U.S. City Average, Bureau of Labor Statistics)
- tby_10yr: 10-year treasury bond yield (source: 10-Year Treasury Constant Maturity Rate, FRED Economic Database; Shiller, 1992)
- tby_1yr: 1-year treasury bond yield (source: 1-Year Treasury Constant Maturity Rate, FRED Economic Database; Shiller, 1992)
- sp500_rtn: S&P 500 index return (source: Yahoo! Finance; Shiller, 1992)
- sp500_divd: S&P 500 dividend yield (source: Yahoo! Finance; Shiller, 1992)
- Baa_cs: Baa-rated corporate bond credit spread (source: Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity, FRED Economic Database)
- Aaa_cs: Aaa-rated corporate bond credit spread (source: Moody's Seasoned Aaa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity, FRED Economic Database)
- BBB_default: BBB-rated corporate bond default rate (source: S&P Global, 2021)
- ur: U-3 unemployment rate (source: U.S. Unemployment Rate, FRED Economic Database)
- Fed_rate: Effective Fed funds rate (source: FRED Economic Database)
- pce: Personal consumption expenditures (source: FRED Economic Database)
- fpi: Fixed private investment (source: FRED Economic Database)
- ge: Government consumption expenditures and gross investment (source: FRED Economic Database)

- mhp_gr: Medium house price growth rate (source: FRED Economic Database)
- rent_gr: Rent inflation rate (source: FRED Economic Database)

  Some variables have longer historical data available than others. Pairwise variables with the longest available data are used to construct each cell in the correlation matrix. These variables are used in many tables and figures in the rest of this paper.

Table 4 shows the correlation matrix with data during extreme pandemics as defined earlier in this section. It is clear that many correlations are higher, either positive or negative, than the normal correlations shown in Table 3.

Table 4

Economic Risk Correlation Matrix during Extreme Pandemics (1871–2020)

Variable	gdp_gr	gdppc_gr	inflation	tby_10yr	tby_1yr	sp500_rtn	sp500_divd	Baa_cs	Aaa_cs	BBB_default	ur	Fed_rate	pce	fpi	ge	mhp_gr	rent_gr
gdp_gr	1.00	1.00	0.68	0.89	0.86	-0.35	0.57	0.14	-0.15	0.99	0.15	0.81	0.81	0.31	0.63	-0.23	0.58
gdppc_gr	1.00	1.00	0.69	0.89	0.86	-0.34	0.54	0.19	-0.11	0.99	0.17	0.81	0.80	0.31	0.63	-0.24	0.61
inflation	0.68	0.69	1.00	0.18	0.24	-0.05	0.08	0.53	0.12	0.98	0.62	0.47	0.59	-0.16	0.85	0.24	0.52
tby_10yr	0.89	0.89	0.18	1.00	0.76	-0.22	0.11	0.24	-0.05	1.00	0.05	0.89	0.82	0.30	0.59	-0.33	0.18
tby_1yr	0.86	0.86	0.24	0.76	1.00	-0.26	0.29	0.16	-0.07	0.95	-0.17	0.96	0.86	0.48	0.52	-0.38	0.50
sp500_rtn	-0.35	-0.34	-0.05	-0.22	-0.26	1.00	-0.37	-0.34	0.01	0.62	-0.63	-0.13	-0.34	0.38	-0.75	-0.44	-0.03
sp500_divd	0.57	0.54	0.08	0.11	0.29	-0.37	1.00	-0.14	-0.55	0.82	0.44	0.15	0.47	-0.27	0.67	0.08	0.18
Baa_cs	0.14	0.19	0.53	0.24	0.16	-0.34	-0.14	1.00	0.87	0.23	0.51	0.23	0.08	-0.25	0.49	0.24	0.75
Aaa_cs	-0.15	-0.11	0.12	-0.05	-0.07	0.01	-0.55	0.87	1.00	-0.33	0.24	0.06	-0.12	-0.10	0.11	0.23	0.49
BBB_default	0.99	0.99	0.98	1.00	0.95	0.62	0.82	0.23	-0.33	1.00	-0.16	0.93	0.82	0.23	0.99	-0.99	0.90
ur	0.15	0.17	0.62	0.05	-0.17	-0.63	0.44	0.51	0.24	-0.16	1.00	-0.25	-0.17	-0.74	0.42	0.33	0.34
Fed_rate	0.81	0.81	0.47	0.89	0.96	-0.13	0.15	0.23	0.06	0.93	-0.25	1.00	0.77	0.54	0.43	-0.45	0.71
pce	0.81	0.80	0.59	0.82	0.86	-0.34	0.47	0.08	-0.12	0.82	-0.17	0.77	1.00	0.58	0.66	-0.01	0.51
fpi	0.31	0.31	-0.16	0.30	0.48	0.38	-0.27	-0.25	-0.10	0.23	-0.74	0.54	0.58	1.00	-0.05	-0.17	0.06
ge	0.63	0.63	0.85	0.59	0.52	-0.75	0.67	0.49	0.11	0.99	0.42	0.43	0.66	-0.05	1.00	0.40	0.60
mhp_gr	-0.23	-0.24	0.24	-0.33	-0.38	-0.44	0.08	0.24	0.23	-0.99	0.33	-0.45	-0.01	-0.17	0.40	1.00	-0.04
rent_gr	0.58	0.61	0.52	0.18	0.50	-0.03	0.18	0.75	0.49	0.90	0.34	0.71	0.51	0.06	0.60	-0.04	1.00

If the correlation matrix approach is used to model the nonlinear relationships, different correlation matrices are needed for simulation periods with and without extreme pandemic events.

4.3 Statistical Approach: Copulas

A copula is used as a general way of formulating a multivariate distribution in such a way that various general types of dependence can be represented. A copula is used to formulate a multivariate distribution via a simple transformation being made of each marginal variable in such a way that each transformed marginal variable has a uniform distribution. Its theoretical foundation is Sklar's theorem of 1959, which says that every multivariate CDF can be written as a function of the marginal distribution functions. For a bivariate CDF, $\mathrm{P}(X \leq x, Y \leq y) = C\big(\mathrm{P}(X \leq x), \mathrm{P}(Y \leq y)\big)$. The copula function *C* is a parameterized model that describes the relationship of multiple variables. Dependence modelling with copula functions is widely used in applications of financial risk assessment and actuarial analysis.

Table 5 illustrates several copula functions for bivariate analysis, all of which can be extended to multivariate analysis to accommodate three or more variables. With the same marginal distributions, different copulas exhibit different joint distributions. The correlation at the tail implied by the Gumbel copula is the highest in the example. The Clayton copula shows a negative correlation in the example. Although the example is for two variables, copulas can be easily applied to multiple variables as well.

Table 5

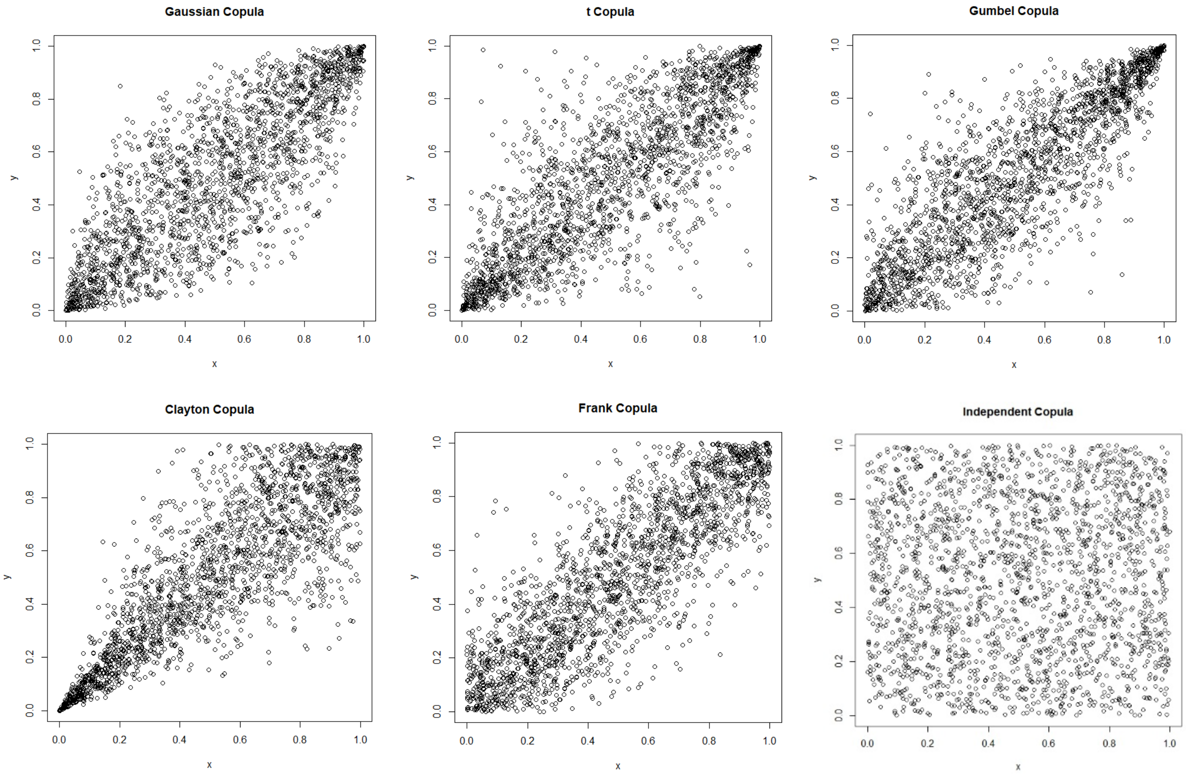
Copula Example

Marginal distribution		
u	$\mathrm{P}(X \leq x)$	0.95
v	$\mathrm{P}(Y \leq y)$	0.95
Joint distribution		
Gaussian copula	Bivariate normal distribution Φ with correlation coefficient ρ $C(u,v) = \Phi_\rho(x,y)$	$\mathrm{P}(X \leq x, Y \leq y) = 0.928$ when $\rho = 0.85$
***t* copula**	Bivariate t distribution with correlation coefficient ρ and the number of the degrees of freedom v	$\mathrm{P}(X \leq x, Y \leq y) = 0.932$ when $\rho = 0.85$ and $v = 5$
Gumbel copula	$C(u,v) = \exp\left(-\left((-\log(u))^\theta + (-\log(v))^\theta\right)^{1/\theta}\right)$	$\mathrm{P}(X \leq x, Y \leq y)$ = 0.937 when θ=3
Clayton copula	$C(u,v) = (u^{-\theta} + v^{-\theta})^{-1/\theta}$	$\mathrm{P}(X \leq x, Y \leq y) = 0.799$ when θ=4
Frank copula	$C(u,v) = -\frac{1}{\theta}\log\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right)$	$\mathrm{P}(X \leq x, Y \leq y) = 0.916$ when θ=9.5
Independent	$C(u,v) = u \times v$	$\mathrm{P}(X \leq x, Y \leq y) = 0.9025$

Unlike the correlation matrix approach in which a nonlinear relationship needs to use multiple matrices, a copula can describe a nonlinear relationship. Copulas allow us to parsimoniously reflect a nonlinear dependence in stochastic scenario generations. Figure 20 illustrates a few simulated copulas as used in Table 5. We show five sets of simulated data, each set with two variables that have a correlation coefficient of about 0.85, compared to the independent copula with a correlation coefficient of 0. The Gaussian copula models the linear relationship, which is exactly the same as the correlation matrix approach. The *t* copula has a higher correlation at both ends, the Gumbel copula a higher correlation at the right end, the Clayton copula a higher correlation at the left end, and the Frank copula a lower correlation at both ends. However, like the correlation matrix approach, it is difficult to consider the order and timing of extreme events. Copulas are a complicated statistical concept with many more types and possible applications than discussed above. Roger B. Nelsen's well-known book *An Introduction to Copulas* (1999) provides more theoretical background.

Figure 20

Copula Simulation



For a comprehensive PSG with many variables of interests, the relationships among those variables are rarely represented by copulas other than Gaussian and *t* copulas. The reason is that the other copulas mentioned above are parsimonious and it is difficult to capture all the variation in the relationships. Alternatively, a few key variables with strong nonlinear relationships may be selected and modelled by copulas.

As an example, we want to investigate five variables using copulas: gdp_gr (nominal GDP growth rate), inflation (inflation rate), ur (unemployment rate), tby_10y (10-year treasury bond yield), and sp500_rtn (S&P 500 index return). Some variables have longer historical data than others. Tables 6 and 7 use pairwise data points as much as possible to construct each cell in the correlation matrix. To model the relationships among the five variables using a single copula, complete data records are used. These five variables showed different correlation coefficients during periods with extreme pandemic events, as shown in Figure 21.

Figure 21

Correlation Matrices of Sample Variables Using Complete Data Records (1948–2020)

Correlation Matrix (Normal)

Variable	gdp_gr	inflation	ur	tby_10yr	sp500_rtn
gdp_gr	1.00	0.33	-0.06	0.32	-0.02
inflation	0.33	1.00	0.34	0.67	-0.20
ur	-0.06	0.34	1.00	0.36	-0.28
tby_10yr	0.32	0.67	0.36	1.00	-0.04
sp500_rtn	-0.02	-0.20	-0.28	-0.04	1.00

Correlation Matrix (Extreme Periods)

Variable	gdp_gr	inflation	ur	tby_10yr	sp500_rtn
gdp_gr	1.00	0.68	0.15	0.89	-0.35
inflation	0.68	1.00	0.62	0.72	-0.83
ur	0.15	0.62	1.00	0.05	-0.63
tby_10yr	0.89	0.72	0.05	1.00	-0.34
sp500_rtn	-0.35	-0.83	-0.63	-0.34	1.00

The goodness of fit in terms of copula fitting can be measured by comparing the empirical multivariate distribution and the fitted distribution using statistical tests such as the Cramér–von Mises test and Kolmogorov–Smirnov test. Genest et al. (2009) reviewed and compared a variety of goodness-of-fit tests for copulas. Table 6 lists the calibration results with goodness-of-fit tests.

Table 6

Copula Calibration Example

<table>
<tr><th>Copula type</th><th>Parameters</th><th>Sn</th><th>p-value</th></tr>
<tr><td>Gaussian copula</td><td>ρ:
<table>
<tr><td>Variable</td><td>gdp_gr</td><td>inflation</td><td>ur</td><td>tby_10yr</td><td>sp500_rtn</td></tr>
<tr><td>gdp_gr</td><td>1.00</td><td>0.39</td><td>-0.01</td><td>0.39</td><td>-0.05</td></tr>
<tr><td>inflation</td><td>0.39</td><td>1.00</td><td>0.27</td><td>0.65</td><td>-0.24</td></tr>
<tr><td>ur</td><td>-0.01</td><td>0.27</td><td>1.00</td><td>0.31</td><td>-0.27</td></tr>
<tr><td>tby_10yr</td><td>0.39</td><td>0.65</td><td>0.31</td><td>1.00</td><td>-0.11</td></tr>
<tr><td>sp500_rtn</td><td>-0.05</td><td>-0.24</td><td>-0.27</td><td>-0.11</td><td>1.00</td></tr>
</table></td><td>0.039</td><td>0.127</td></tr>
<tr><td>t copula</td><td>ρ:
<table>
<tr><td>Variable</td><td>gdp_gr</td><td>inflation</td><td>ur</td><td>tby_10yr</td><td>sp500_rtn</td></tr>
<tr><td>gdp_gr</td><td>1.00</td><td>0.37</td><td>0.01</td><td>0.38</td><td>-0.08</td></tr>
<tr><td>inflation</td><td>0.37</td><td>1.00</td><td>0.31</td><td>0.65</td><td>-0.26</td></tr>
<tr><td>ur</td><td>0.01</td><td>0.31</td><td>1.00</td><td>0.31</td><td>0.31</td></tr>
<tr><td>tby_10yr</td><td>0.38</td><td>0.65</td><td>0.31</td><td>1.00</td><td>-0.08</td></tr>
<tr><td>sp500_rtn</td><td>-0.08</td><td>-0.26</td><td>-0.21</td><td>-0.08</td><td>1.00</td></tr>
</table>
$v = 9$</td><td>0.030</td><td>0.321</td></tr>
</table>

Copula type	Parameters	Sn	p-value
Gumbel copula	$\theta = 1.081$	0.044	0.147
Clayton copula	$\theta = 0.143$	0.046	0.107
Frank copula	$\theta = 0.571$	0.049	0.082

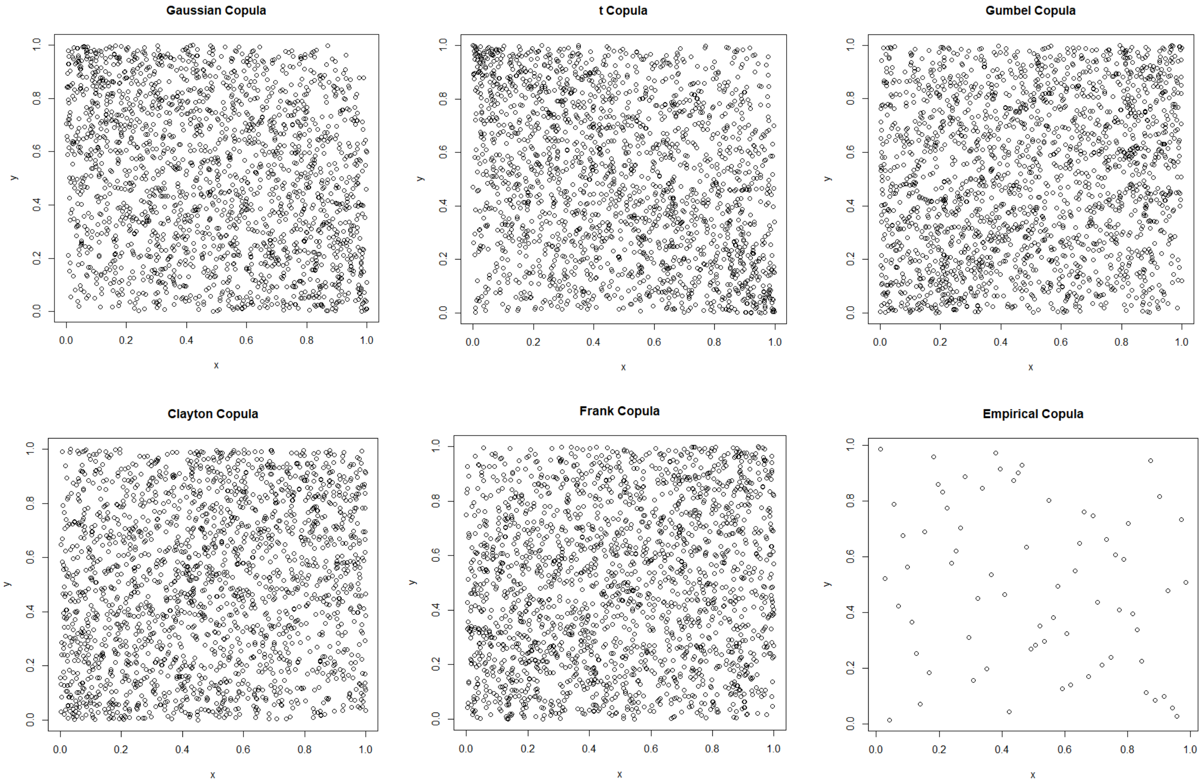
Notes:

1. Sn: Cramér–von Mises statistic introduced by Genest, Remillard, and Beaudoin (2009).
2. P-value: P-value of Sn test using the multiplier method introduced by Kojadinovic and Yan (2011).

The *p*-value of the Sn test shows that the *t* copula has the best fit, as a higher *p*-value indicates that the empirical data are more likely to follow the calibrated copula. The correlation coefficient matrix ρ of both the Gaussian and *t* copulas is close to the correlation matrix using all the complete data in Figure 21. However, although the statistical test indicates the goodness of fit in general, it is difficult to tell whether the higher correlation observed in extreme pandemic events is captured in any of the calibrated copulas. Inflation rates and S&P 500 index returns are used as an illustration of necessary visualization. The correlation coefficient of these two variables changes from -0.2 in all periods to -0.83 in extreme periods. Figure 22 compares the calibrated copulas and the empirical data. Of all the calibrated copulas, the *t* copula has a higher correlation at both ends, which is desired at least at the bottom right end. The empirical copula shows a very high correlation at the bottom right end, but not necessarily at the top left end.

Figure 22

Copula Simulation



Notes:

- x: Inflation
- y: sp500_rtn

If a copula is needed to model the nonlinear relationships, the t copula seems to be the best choice even though it is still not ideal to match the tail behaviours at both ends. Although the copula approach provides more flexibility and preserves the parsimony, it is not an easy task to find the most appropriate copula. The data used for copula calibration may be sparse, and the goodness of fit can be low.

4.4 Structured Model Approach

Compared to the statistical approaches where contemporary relationships are the main subject to study, structured models are more flexible to deal with both contemporary and temporal relationships at the same time. A monetary policy may be the result of a crisis but can also dampen the impact and duration of the crisis. As evidenced in the recent COVID-19 outbreaks, the proactive monetary and fiscal policies helped improve market liquidity and reversed the course of a bear market. Correlated simulation models can be used to reflect nonlinear correlation and the timing of events.

However, the data requirement is higher for the structured models. Annual data were used for statistical approaches to utilize as long a history as possible. But some temporal relationships usually exist at a higher frequency. For example, central banks may react quickly during extreme events. In the following analysis using structured models, quarterly data are used given data availability, policy-makers' reaction time, and typical actuarial valuation frequency. In addition, U.S. data are used for all economic variables given their long history, available from 1947, and the country's dominating position in the world economy post Second World War.

4.4.1 Economic Factors

A vector autoregressive (VAR) model is used to study nonlinear relationships among economic factors, including GDP growth rate, inflation rate, personal consumption growth rate, fixed private investment growth rate, government consumption and investment growth rate, Fed fund rate, and unemployment rate. Asset return scenarios are determined based on their relationships with these economic factors and their own recent experience. These economic factors are chosen to capture the economic situation in a succinct and manageable way.

1. **GDP growth rate** is a direct reflection of real economic activities. It is an aggregate indicator of consumer spending, private- or public-sector investment, and imports/exports.
2. **Inflation rate** affects economic activities in various ways. In general, the inflation rate is determined by the money supply growing faster than the economy in the long term. It can reduce the burden of public and private debt. However, hyperinflation and unexpected inflation can be harmful to the economy, discourage investment and exports, and even cause social unrest. A very low or negative inflation rate usually accompanies economic recession.
3. **Personal consumption growth rate** is affected by personal income, which grows faster during an economic expansion. Higher consumption also means higher spending and a higher GDP growth rate.
4. **Investment growth rate** is an indicator of the investment portion of the economy. A higher investment growth rate means a higher economic growth rate, *ceteris paribus*.
5. **Government consumption and investment growth rate** reflects the direction and magnitude of fiscal policies, which is an important force in modern economies.
6. **Fed fund rate** is determined by central-bank monetary policies, which are used to smooth the impact of economic volatilities. In an economic recession, whether caused by a pandemic or not, the short-term interest rate is usually low to spur economic growth. In an economic expansion, the short-term interest rate is usually high to cool the economy down.
7. **Unemployment rate** is another important indicator of the economy. It reflects the balance of the labour market. An economic recession usually comes with a high unemployment rate.

Although these factors are used in this report, other factors can be included as well, depending on the features of the economy and purpose of a model. For example, to model an export-driven economy, exports/imports and foreign investment positions can be included in the list of fundamental economic factors. Table 7 describes the historical data of economic factors used in this example.

Table 7

Economic Factor Historical Data

Economic factor	Indicator	Notation	Annualized*	Data source
GDP growth rate	U.S. quarterly notional GDP growth rate	gdp_gr	Y	FRED Economic Database
Inflation rate	Consumer Price Index for All Urban Consumers: All Items in U.S. City Average, Index 1982–1984=100, Seasonally Adjusted	cpi	Y	Bureau of Labor Statistics
Personal consumption growth rate	PCEA (Personal Consumption Expenditures, Seasonally Adjusted Annual Rate)	pce	N	FRED Economic Database
Investment growth rate	FPI (Fixed Private Investment, Seasonally Adjusted Annual Rate)	fpi	N	FRED Economic Database
Government consumption and investment growth rate	GCE (Government Consumption Expenditures and Gross Investment, Seasonally Adjusted Annual Rate)	ge	N	FRED Economic Database
Fed fund rate	Effective Federal Funds Rate, Not Seasonally Adjusted	Fed_rate	Y	FRED Economic Database
Unemployment rate	U-3 rate	ur	Y	Bureau of Labor Statistics

* If a variable is annualized, quarterly growth rate is converted to annual growth rate.

Quarterly historical data from 1947Q1 to 2021Q2 are used, except that unemployment rate data start from 1948Q1 and Fed fund rate data from 1954Q3. The data contain a few periods in which extreme pandemic events occurred, as listed below:

- 1957–58: British influenza epidemic, as part of the Asian flu pandemic
- 1964: U.S. rubella epidemic
- 1968–69: Russian influenza epidemic
- 1974–75: Indian malaria epidemic
- 1975: HIV/AIDS pandemic
- 1990: Sub-Saharan African HIV/AIDS pandemic; tuberculosis pandemic; U.S. venereal disease epidemics
- 2000: Indian HIV/AIDS epidemic
- 2020–21: COVID-19 pandemic

A VAR model is used to describe the relationship of the fundamental economic factors based on this historical data. By incorporating lagging variables into the analysis through VAR, relationships among leading, coincident, and lagging economic factors can be better reflected. For example, the short-term interest rate is largely controlled by the U.S. Federal Reserve Board (the Fed), after reviewing economic growth, unemployment, and other economic conditions. Time is needed before making rate decisions. For simplicity, VAR(1) is used so that the evolution of fundamental economic factors is affected by their values in the previous quarter. A quarter is likely to be enough for the interaction among fundamental economic factors. Having a higher order of VAR model can only improve the results marginally in this example:

$$\mathbf{E}_{\boldsymbol{t}} = \mathbf{c} + \boldsymbol{A_E}\mathbf{E}_{\boldsymbol{t-1}} + \mathbf{e}_{\boldsymbol{t}}$$

where

$\mathbf{E}_{\boldsymbol{t}} =$ $(\mathrm{gdp\_gr}_t, \mathrm{inflation}_t, \mathrm{pce}_t, \mathrm{fpi}_t, ge_t, \mathrm{Fed\_rate}_t, \mathrm{ur}_t)^T$, a column vector with seven elements as the value of economic factors at time t or during period t;

$\mathbf{c} =$ a column vector with seven elements to represent the constant terms of the seven economic factors;

$\boldsymbol{A_E} =$ a 7×7 matrix containing the model parameters describing the linear dependence of economic factors; and

$\mathbf{e}_{\boldsymbol{t}} =$ a column vector with seven elements to store the error terms that cannot be explained by linear models.

Table 8 shows the fitted model parameters ($\boldsymbol{A_E}$ and **c**) based on the historical data. It also shows σ, the standard deviation of error vector $\mathbf{e}_{\boldsymbol{t}}$.

Table 8

VAR(1) Model Parameters

Variable	A_E							c	s
	gdpgr	inflation	pce	fpi	ge	Fed_rate	ur		
gdpgr	-0.32	-0.34	0.07	0.98	0.59	0.29	0.99	-0.02	0.05
inflation	0.11	-0.01	-0.20	0.06	0.27	0.23	0.11	0.00	0.02
pce	-0.02	-0.09	-0.16	0.17	0.24	0.08	0.25	0.00	0.01
fpi	0.00	-0.04	-0.15	0.60	0.09	-0.02	0.35	-0.01	0.02
ge	-0.01	-0.05	0.14	-0.05	0.02	0.13	-0.10	0.01	0.01
Fed_rate	-0.08	-0.07	0.20	0.17	0.01	0.97	-0.02	0.00	0.01
ur	-0.04	0.03	0.05	-0.05	0.06	0.01	0.83	0.01	0.01

Based on the fitted VAR(1), the stable values of fundamental risk factors $\bar{\mathbf{E}}$ can be derived.

$$\bar{\mathbf{E}} = \mathbf{c} + \boldsymbol{A_E}\bar{\boldsymbol{E}}$$

Table 9 lists the stable values based on VAR(1), along with the historical mean and standard deviation. The VAR(1) suggests a lower future economic growth rate, inflation rate, interest rate, and consumption and investment growth rate than in the post–Second World War era, which is a phenomenal period of fast development. These model-implied expectations are good checkpoints to assess the model's reasonableness against a model user's view on future economic development.

Table 9

VAR(1) Stable Values

Economic factors	VAR(1)	Historical data	
	Stable value (%)	Mean (%)	Standard deviation (%)
gdp_gr	4.54	6.45	5.52
inflation	2.23	3.51	3.56
pce*	1.14	1.57	1.31
fpi*	1.20	1.64	2.54
ge*	0.93	1.58	1.82
Fed_rate	2.28	4.68	3.66
ur	5.98	5.78	1.74

* Quarterly growth rate

The VAR model reflects the relationships among economic factors through a system of linear equations. In addition, the error terms are not independent of each other. It is important for the economic scenario generator to capture the correlation when developing future scenarios. Table 10 shows the correlation matrix of the error vector et.

Table 10

Correlation Matrix of VAR(1) Error Terms

	gdp_gr	inflation	pce	fpi	ge	Fed_rate	ur
gdp_gr	1.00	0.30	0.90	0.73	0.24	0.34	-0.82
inflation	0.30	1.00	0.34	0.25	0.15	0.22	-0.23
pce	0.90	0.34	1.00	0.66	0.21	0.20	-0.84
fpi	0.73	0.25	0.66	1.00	0.10	0.18	-0.57
ge	0.24	0.15	0.21	0.10	1.00	-0.04	-0.05
Fed_rate	0.34	0.22	0.20	0.18	-0.04	1.00	-0.25
ur	-0.82	-0.23	-0.84	-0.57	-0.05	-0.25	1.00

Ideally, two calibrations of the VAR models are used to represent different patterns with and without extreme pandemic events. However, not all extreme pandemic events lead to heightened economic volatilities, and some financial crises led to the same, if not worse, economic conditions. In addition, the data volume is another obstacle to support two separate calibrations. Therefore, one set of VAR models is used to describe the economic variables.

4.4.2 Capital Market Variables

With the relationships among economic factors represented by the VAR(1), returns in each asset class can be constructed based on their relationships with economic factors and autocorrelation. In this report, several key capital market variables are used to represent market conditions, as shown in Table 11. Other variables of interests can be easily added using the same methodology described in this section.

Table 11

History Data of Capital Market Variables

Asset class	Return type	Notation	Annualized*	Time period	Data source
1-year Treasury bond yield	Yield	tby_1yr	Y	1953Q2–2021Q2	FRED Economic Database: 1-Year Treasury Constant Maturity Rate
10-year Treasury bond yield	Yield	tby_10yr	Y	1953Q2–2021Q2	FRED Economic Database: 10-Year Treasury Constant Maturity Rate
Aaa-rated corporate bond	Credit spread	Aaa_cs	Y	1953Q2–2021Q2	Moody's Seasoned Aaa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity
Baa-rated corporate bond	Credit spread	Baa_cs	Y	1953Q2–2021Q2	Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity
BBB-rated corporate bond	Default rate	BBB_default	Y	1981Q1–2021Q2	2020 S&P Annual Global Corporate Default Study and Rating Transitions Report; S&P Credit Markets Update North America Q2 2021
Public equity, large cap	Dividend yield	sp500_divd	Y	1947Q1–2021Q2	Yahoo Finance: S&P 500 Index (ticker name: ^GSPC); Shiller (2015)

Asset class	Return type	Notation	Annualized*	Time period	Data source
	Capital return	sp500_rtn	N		
Real estate	Cap rate	rent_gr	N	1947Q1–2021Q2	Consumer Price Index for All Urban Consumers: Rent of Primary Residence in U.S. City Average
	Capital return	mhp_gr	N	1963Q1–2021Q2	Median Sales Price of Houses Sold for the United States

* If a variable is annualized, quarterly growth rate is converted to annual growth rate.

Seven model types are used to build the relationships among capital market variables and economic factors: linear regression, lasso regression, ridge regression, k-nearest neighbours (KNN), classification and regression tree (CART), randomForest (RF), and gradient boosting machine (GBM).

Linear regression assumes a linear relationship between economic factors and capital market variables including asset returns, bond yields, credit spreads, and default rates. Lasso and ridge regression are variations of the linear regression model with different methods of regularization to prevent overfitting. In addition to minimizing the squared errors, lasso models add the sum of the absolute value of parameters into the error function. Ridge regression uses the sum of squared parameters as the regularization term. The weight of the regularization term is determined by cross-validation to get the highest model accuracy.

KNN is a type of nonparametric model that predicts the explained variable based on the values of the k closest neighbours. In this report, the closeness is measured by the Euclidean distance based on macroeconomic factors. Different numbers of neighbours are tested, and five nearest neighbours are used for all models for relatively better performance. The average return of five nearest neighbours in historical data is then used to predict the asset return. CART models build trees to split the data based on economic factors. At each split, a factor is used to separate the data into two subgroups. The factor and split rule are chosen to provide the best split that improves the purity of the data in the subgroups. In the calibration of CART models for capital market variables, two criteria are used to limit the size of the tree and avoid overfitting:

- The minimum number of data points in a node for splitting is 10.
- The minimum improvement of data purity is 0.001 for each split.

Each terminal node has an estimation based on the average value of the explained variable. For example, if the terminal node has 10 historical data records, the capital market variable is calculated as the average return of the 10 data records.

RF is a decision-tree-based ensemble method. Each tree is a weak estimator utilizing a random subset of the training data and a random subset of all available explanatory variables. The final estimation is based on the average of estimates of all trees. In this example, 70% of the training set were randomly selected each time to fit the next tree. The minimum observations in each node are set to be two, and 500 trees were used.

GBM is also a decision-tree-based ensemble method. However, unlike RF, it is a sequential model. Each tree is a weak estimator trying to estimate the residual error that the estimation of previous trees has caused. Gradually, with a sufficient number of decision trees, the estimation error will decline to a very low level. GBM is usually proven to have better accuracy than many other methods when presented with nonlinear relationships. In this example, 70% of the training set were randomly selected each time to fit the next tree. The depth of the trees is set to be six, and 500 trees were used. The minimum observations in each node are set to be two, given the limited amount of training data.

For all the model testing, the data are randomly split into a training dataset (80%) and a validation dataset (20%). The training dataset is used to find the parameters that minimize the sum of squared prediction errors for the training data. The validation dataset is used to measure the prediction accuracy based on out-of-sample data. Table 12 shows the aggregate performance of the seven model types.

Table 12

Model Aggregate Performance for Capital Market Variables

Model	Training data		Validation data	
	RMSE	R2	RMSE	R2
LM	0.01	0.71	0.01	0.54
Lasso	0.01	0.68	0.01	0.67
Ridge	0.01	0.67	0.01	0.64
KNN	0.01	0.62	0.02	0.37
CART	0.01	0.91	0.02	0.32
RF	0.01	0.94	0.01	0.62
GBM	0.00	0.99	0.01	0.56

Notes:

- RMSE: Root mean square error
- R^2: Coefficient of determination

Given the data volume, overfitting is a critical issue in this example. Tree-based models, including CART, RF, and GBM, showed very good estimation on the training data but worse estimation than some linear models on the validation data. Linear regression models, especially lasso models, did a good job in general. For better performance, linear models such as lasso models and ridge regression models are used for modelling capital market variables in this example:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + B_0 E_t + B_1 E_{t-1} + B_2 E_{t-2} + \sigma\varepsilon_t$$

Where

- $\mathbf{y}_t$ is a column vector containing capital market variables during period *t*;
- $\boldsymbol{\alpha}$ is a column vector containing the constant terms of all capital market variables;
- $\boldsymbol{\phi}_i$ is a column vector containing parameters to govern the relationship between current value and previous value for all capital market variables during period *t* − *i*;
- $\boldsymbol{B}_i$ is a matrix with seven columns that contains all capital market variables' model parameters for the seven economic factors during period *t* − *i*;
- $\boldsymbol{E}_t$ is a column vector with seven elements as the value of economic factors at time *t* or during period *t*;
- $\boldsymbol{\sigma}$ is a column vector containing the standard deviation of error terms of all capital market variable models; and
- $\boldsymbol{\varepsilon}_t$ is a column vector containing independent random variables following a standard normal distribution for all capital market variable models.

Table 13 lists the parameters of linear models, standard deviation of error terms, and the adjusted R^2 of the linear models. Adjusted R^2 indicates the portion of variable volatility that can be explained by the linear relationships. Certain variables, including S&P 500 index return, real estate cap rate, and capital return, have a low adjusted R^2. They are more driven by idiosyncratic factors than by fundamental economic factors. The second-to-last row of Table 13 contains the standard deviation of the idiosyncratic factors that cannot be explained by the linear models.

Table 13

Linear Model Parameters

	Lag (quarter)	Asset class	tby_1yr	tby_10yr	Aaa_cs	Baa_cs	BBB_default	sp500_rtn	sp500_divd	mhp_gr	rent_gr
Model*			Lasso	Lasso	Lasso	Lasso	Lasso	Lasso	Ridge	Lasso	Lasso
Intercept	0	α	0.000	0.001	0.002	0.004	0.000	0.007	0.002	-0.010	0.004
Auto_correlation	2	ϕ_2	0.104	0.000	0.138	0.000	0.000	-0.029	0.332	-0.056	0.275
	1	ϕ_1	0.512	0.834	0.757	0.843	0.754	0.000	0.506	-0.281	0.000
Economic risk factors	2	gdp_gr	0.007	0.000	-0.010	0.000	0.000	0.000	0.003	0.000	0.000
	1		0.016	0.009	-0.007	0.000	0.000	0.000	-0.007	0.132	0.000
	0		0.004	0.003	-0.004	0.000	0.000	0.000	-0.001	0.157	0.000
	2	inflation	0.009	0.000	0.000	0.005	0.000	0.000	0.010	-0.044	0.023
	1		-0.032	0.000	0.010	0.010	0.000	-0.034	0.022	0.000	0.000
	0		0.000	0.022	-0.004	0.000	0.000	-0.063	0.006	0.000	0.000
	2	ur	-0.022	0.000	0.004	0.000	0.000	0.000	0.010	0.000	0.000
	1		0.020	0.021	0.000	0.000	0.000	0.148	0.007	0.023	0.000
	0		0.000	0.000	0.000	0.009	0.000	0.000	0.004	0.172	0.000

	Lag (quarter)	Asset class	tby_ 1yr	tby_ 10yr	Aaa_ cs	Baa_ cs	BBB_ default	sp500_ rtn	sp500_ divd	mhp_ gr	rent_ gr
	2	Fed_ rate	0.000	0.000	0.004	0.000	0.002	0.000	0.011	0.000	0.034
	1		-0.185	0.000	0.039	0.000	0.000	0.000	0.009	0.000	0.000
	0		0.527	0.104	-0.042	0.000	0.000	0.000	-0.009	-0.113	0.000
	2	pce	-0.002	0.000	0.002	0.000	0.000	0.000	-0.047	0.000	0.000
	1		-0.006	0.005	0.000	0.000	0.000	0.000	-0.018	0.113	0.000
	0		-0.047	0.000	0.000	-0.042	0.000	0.538	0.024	0.000	0.000
	2	fpi	-0.005	0.000	0.018	0.000	-0.002	-0.368	0.029	0.190	0.000
	1		0.024	0.000	0.011	0.000	-0.004	-0.062	-0.020	-0.117	0.000
	0		0.060	0.021	-0.025	-0.052	0.000	1.022	-0.017	0.031	0.000
	2	ge	0.057	0.000	0.015	0.000	0.000	-0.178	0.016	-0.074	0.000
	1		0.000	0.000	-0.016	0.000	0.016	0.000	0.024	0.154	0.029
	0		0.003	0.000	0.007	0.000	0.000	-0.427	0.028	0.013	0.000
	0	σ	0.004	0.004	0.002	0.003	0.001	0.065	0.002	0.025	0.006
Adjusted R^2			0.982	0.976	0.882	0.861	0.673	0.129	0.978	0.161	0.243

*All capital asset variables are using a lasso model except for sp500_divd. Even though the lasso model has a slightly better performance than ridge regression for sp500_divd, the resulting autoregressive model is close to non-stationary. The stationary condition for an autoregressive model with a lag of 2 is that the roots of the following equation have an absolute value greater than 1.

$$1 - \boldsymbol{\phi_1 x} - \boldsymbol{\phi_2 x^2} = \mathbf{0}$$

One of the roots of the calibrated lasso model for sp500_divd is very close to the threshold of 1 and therefore ridge regression is used instead.

The relationship between asset returns and fundamental economic factors is not always linear. During stressful periods, higher correlations are often observed. Table 14 shows the volatility of idiosyncratic factors (error terms) and the correlation between systemic factors (prediction by lasso models) and idiosyncratic factors, using either all the data or the data in extreme pandemic events. Correlation behaved quite differently in extreme periods. These values capture second-order impact in addition to the linear relationships built in lasso models.

Table 14

Linear Model Idiosyncratic Factors: Volatility and Correlation

Capital market variable	Idiosyncratic factor volatility		Correlation with economic factors	
	Normal periods	Pandemic periods	Normal periods	Pandemic periods
tby_1yr	0.42%	0.41%	2.6%	-6.2%
tby_10yr	0.43%	0.34%	9.8%	20.9%
Aaa_cs	0.19%	0.20%	2.7%	6.0%
Baa_cs	0.30%	0.31%	10.2%	10.0%
BBB_default	0.13%	0.05%	6.6%	22.3%
sp500_rtn	6.52%	6.87%	9.7%	24.9%
sp500_divd	0.16%	0.15%	1.1%	13.8%
mhp_gr	2.52%	2.28%	8.8%	0.0%
rent_gr	0.57%	0.49%	11.2%	27.3%

In addition to the relationships between capital market and economic factors which represent systemic risk, idiosyncratic risk factors of the capital market variables also have nonlinear relationships that can be important to capture. Figure 23 shows the correlation matrix of residuals from the linear models during normal and extreme periods, separately. Heightened correlations are observed during extreme periods.

Figure 23

Correlation Matrices of Idiosyncratic Factors of Capital Market Variables

Correlation Matrix (Normal Periods)

Variable	tby_1yr	tby_10yr	Aaa_cs	Baa_cs	default	0_rtn	0_divd	mhp_gr	rent_gr
tby_1yr	1.00	0.77	-0.53	-0.46	0.06	0.00	0.11	-0.04	-0.06
tby_10yr	0.77	1.00	-0.59	-0.56	0.00	0.05	-0.05	0.03	-0.09
Aaa_cs	-0.53	-0.59	1.00	0.78	-0.02	-0.29	-0.01	-0.10	0.12
Baa_cs	-0.46	-0.56	0.78	1.00	0.06	-0.44	0.21	-0.07	0.16
BBB_default	0.06	0.00	-0.02	0.06	1.00	0.01	0.09	0.06	0.05
sp500_rtn	0.00	0.05	-0.29	-0.44	0.01	1.00	-0.38	0.14	0.00
sp500_divd	0.11	-0.05	-0.01	0.21	0.09	-0.38	1.00	-0.12	0.03
mhp_gr	-0.04	0.03	-0.10	-0.07	0.06	0.14	-0.12	1.00	0.14
rent_gr	-0.06	-0.09	0.12	0.16	0.05	0.00	0.03	0.14	1.00

Correlation Matrix (Extreme Periods)

Variable	tby_1yr	tby_10yr	Aaa_cs	Baa_cs	default	0_rtn	0_divd	mhp_gr	rent_gr
tby_1yr	1.00	0.72	-0.58	-0.47	0.41	0.38	-0.04	-0.07	0.40
tby_10yr	0.72	1.00	-0.67	-0.72	0.59	0.29	-0.19	-0.09	0.36
Aaa_cs	-0.58	-0.67	1.00	0.80	-0.25	-0.40	0.10	0.07	-0.28
Baa_cs	-0.47	-0.72	0.80	1.00	-0.49	-0.40	0.22	0.01	-0.34
BBB_default	0.41	0.59	-0.25	-0.49	1.00	0.27	0.18	-0.27	0.46
sp500_rtn	0.38	0.29	-0.40	-0.40	0.27	1.00	-0.31	-0.26	0.10
sp500_divd	-0.04	-0.19	0.10	0.22	0.18	-0.31	1.00	0.08	0.13
mhp_gr	-0.07	-0.09	0.07	0.01	-0.27	-0.26	0.08	1.00	0.13
rent_gr	0.40	0.36	-0.28	-0.34	0.46	0.10	0.13	0.13	1.00

Similar to economic factors based on VAR(1), Table 15 lists the stable value of capital market variables along with the historical mean and standard deviation. The linear models suggest a lower future interest rate level and dividend yield compared to the entire post–Second World War period. This is somehow consistent with the low-interest-rate environment we have been in for more than a decade. These model-implied expectations are good checkpoints to assess the model's reasonableness against a model user's view on future capital market behaviours.

Table 15

Capital Market Variable Stable Values based on Linear Models

Capital market variable	Linear	Historical data	
	Stable value (%)	Mean (%)	Standard deviation (%)
tby_1yr	2.45	4.69	3.31
tby_10yr	3.83	5.66	2.92
Aaa_cs	1.21	0.94	0.58
Baa_cs	2.10	1.92	0.83
BBB_default	0.16	0.19	0.25
sp500_rtn*	2.06	2.15	7.17
sp500_divd	2.39	3.29	1.42
mhp_gr*	1.11	1.36	2.89
rent_gr*	0.80	0.95	0.69

* Quarterly growth rate

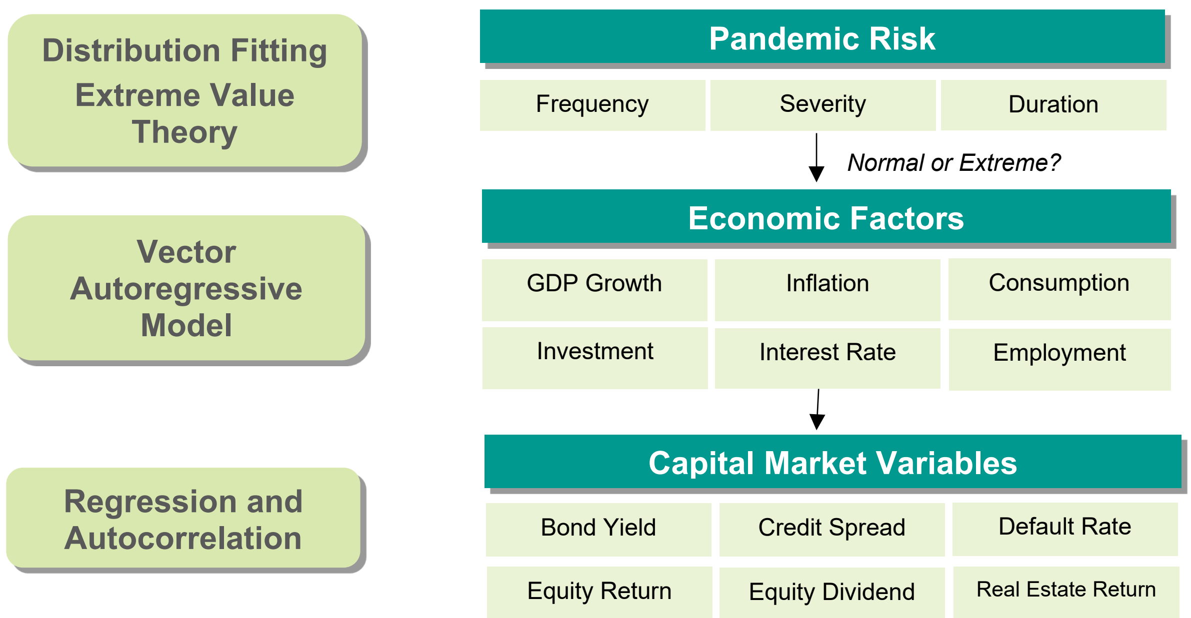
Compared to statistical approaches, structured models provide a much more sophisticated framework to capture more complicated relationships, either contemporary or temporal.

5 Scenario Generation

With the analysis of pandemic risks and their relationships to economic risks based on historical data in previous sections, a PSG is built in this section to reflect our observations of the history, while allowing forward-looking assumptions to be reflected. The structure of the PSG is shown in Figure 24.

Figure 24

Pandemics-Driven Scenario Generation Framework



It starts with the pandemic risk by simulating frequency, severity, and duration of pandemic events that are used for insurance risk factors. For each simulated event, the PSG assesses if it is an extreme event and justifies different volatility and correlation levels for financial risks. Based on the state of simulated pandemic events, economic factors that define the systemic risk level are simulated using the VAR model. Regression models that contain lagged variables are then used to simulate capital market variables based on simulated economic factors.

5.1 Scenario-Generation Process

Based on the analysis in the previous two sections, this section describes the general process of generating scenarios using the calibrated models with the best goodness of fit to historical data. Technical details can be found in Appendix B.

Step 1

Generate pandemic/epidemic event frequencies. A negative binomial distribution with the GPD describing the tail is chosen, given it a good fit to the historical experience. The annual frequency is expected to be 2.7 with a standard deviation of 2.3. The probability of having more than seven pandemic/epidemic events is 4.2%.

Step 2

For each event simulated from Step 1, generate duration and severity measures, including case fatality rate and infection rate. The duration follows a lognormal distribution with a mean of two years and a standard deviation of 4.9, with a chance of 11.4% for a simulated pandemic/epidemic event lasting more than seven years. The case fatality rate (CFR) follows a beta distribution with a mean of 27% and a standard deviation of 27%. The infection rate (IR) follows a beta distribution as well, with a mean of 29% and a standard deviation of 27%.

When generating the three variables in Step 2, the correlation matrix in Table 16 is used based on historical data. Correlated marginal cumulative probabilities are generated first, and values of variables are calculated based on the cumulative probability functions and generated cumulative probabilities.

Table 16

Duration and Severity Correlation Matrix (1050 BC to AD 2020)

	Duration	CFR	IR
Duration	1.00	0.06	-0.23
CFR	0.06	1.00	0.17
IR	-0.23	0.17	1.00

The severity measure generated above is at the total level. As different age groups may experience different severity levels, a multinomial distribution based on the findings in Figure 5 can be used to simulate the fatalities in different age groups: infant (age 0–5) with a share of 32%, children (6–20) with a share of 30%, young adult (21–40) with a share of 14%, mid-aged adult (41–60) with a share of 11%, and old-aged adult (60+) with a share of 13%.

Step 3

Based on the simulated severity results for each pandemic/epidemic event, determine whether there is any extreme event and how long will such an event last. The extreme pandemic events were defined in Subsection 4.1 as events with no fewer than 800,000 death counts and/or no fewer than 12,000,000 infected cases to study nonlinear relationships with the financial risk.

- For all the events in previous steps, a binomial distribution with a probability of 3.7% is used to determine whether a generated event is a pandemic event. The probability 3.7% is derived as the ratio of 30 historical pandemics and 817 pandemics and epidemics in total.
- If a generated event is tagged as a pandemic event, the number of infected cases will be calculated as the product of the current world population and the simulated IR for that event. Death counts will be calculated as the product of the number of infected cases and the simulated CFR for that event. If either one of them exceeds the threshold, the generated event will be determined as an extreme event for that specific

scenario. Different economic scenario assumptions, such as the volatility of idiosyncratic risks and correlation among idiosyncratic risks and the systemic risk, will be used depending on the occurrence of a simulated extreme pandemic event.

- For each simulated epidemic event, the death count and the number of infected cases are not explicitly calculated given the limitation of modelling the geographic reach of the event. Although the historical data used in this report do not support such details, a possible way to use these epidemic events is to apply them directly to the regions with written insurance businesses.

Step 4

Generate economic factors based on VaR models.

$$\mathbf{E}_t = \mathbf{c} + \boldsymbol{A}_E \mathbf{E}_{t-1} + \boldsymbol{\sigma\varepsilon}_t$$

Table 8 lists the calibration results for $\mathbf{c}$ and $\boldsymbol{A}_E$, and σ, the standard deviation of error vector $\mathbf{e}_t$. Table 10 lists the correlation matrix among the seven economic factors.

Step 5

Generate capital market variables based on linear models.

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \mathbf{y}_{t-2} + \boldsymbol{B}_0 \mathbf{E}_t + \boldsymbol{B}_1 \boldsymbol{E}_{t-1} + \boldsymbol{B}_2 \mathbf{E}_{t-2} + \boldsymbol{\sigma\varepsilon}_t$$

Table 13 lists the calibration results for constant term $\boldsymbol{\alpha}$, autocorrelation coefficient $\boldsymbol{\phi}$, relationships with economic factors $\boldsymbol{B}$, and volatility of the error term $\boldsymbol{\sigma}$. The idiosyncratic part of capital market variables $\boldsymbol{\sigma\varepsilon}_t$ is adjusted to reflect non-constant volatility and nonlinear relationships.

5.2 Sample Simulations

Although statistical tests and model validation were performed to ensure the calibrated distributions, correlations, and regression models are robust, as described in sections 3 and 4, it is necessary to check the reasonableness of the generated scenarios. For all simulated variables, including insurance risk factors, economic factors, and capital market variables, simulated data are compared with the historical data and the range of simulated values is displayed. One thousand scenarios of 100 quarters are simulated in this analysis, which contains around 70,000 pandemic/epidemic events, and 100,000 records of economic factors and capital market variables.

5.2.1 Insurance Risk Factor

Table 17 compares simulated insurance risk factors with historical data in terms of mean and standard deviation. Most statistical measures are close, apart from some noticeable differences in the standard deviation of IR. The historical data of IRs at event level contains only around 70 events, which may partly lead to the difference. Since the calibration was done using the maximum likelihood estimation (MLE) method, if standard deviation is a key concern, the moment matching (MM) method can be used instead and is likely to narrow the gap. On the other hand, using MM can cause other issues, such as poor performance in mimicking the tail risk given that the number of parameters may be insufficient to match high-degree moments such as skewness and kurtosis.

Table 17

Simulated Insurance Risk Factors vs. Historical Data

Variable	Simulated data		Historical data	
	Mean	Standard deviation	Mean	Standard deviation
Frequency	2.82	1.98	2.77	2.48
Duration	4.04	6.29	4.09	6.54
Case fatality rate	30.6%	29.0%	26.1%	21.8%
Infection rate	25.4%	24.4%	27.6%	12.3%
Chance of a pandemic	3.6%	18.6%	3.7%	18.8%
Infant death portion	32.0%		31.1%	
Children death portion	30.0%		29.5%	
Young adult death portion	14.0%		13.9%	
Mid-aged death portion	11.0%		13.1%	
Old-aged death portion	13.0%		12.3%	

5.2.2 Economic Factor

Table 18 compares simulated economic factors with historical data in terms of mean and standard deviation. Noticeable differences include a lower expected economic growth rate and Fed rate, a higher unemployment rate than historical data, and an inflation rate closer to the 2% target set by the Fed. This indicates the PSG captured recent trends and acknowledged that post–Second World War booming may not sustain at the same level.

Table 18

Simulated Economic Factors vs. Historical Data

Economic factors	Simulated data		Historical data	
	Mean (%)	Standard deviation (%)	Mean (%)	Standard deviation (%)
gdp_gr	4.55	6.18	6.45	5.52
inflation	2.18	2.87	3.51	3.56
pce*	1.16	4.23	1.57	1.31
fpi*	1.22	4.69	1.64	2.54
ge*	0.91	1.91	1.58	1.82
Fed_rate	1.92	6.62	4.68	3.66
ur	6.44	5.16	5.78	1.74

*Quarterly growth rate

As discussed earlier, the historical data do not support a separate calibration of the VAR models for extreme pandemic events; the generated economic variables are expected to follow the same distribution. As observed in historical data, many extreme pandemic events have had very short-lived impacts on the economy in modern history, which lead to extreme pandemic but non-extreme economic factors. On the other hand, financial crises can cause more extreme impacts on the economic system that do not necessarily happen during pandemics.

5.2.3 Capital Market Variable

Table 19 compares simulated capital market variables with historical data in terms of mean and standard deviation. Noticeable differences include lower expected bond yields, slightly heightened credit spreads, and higher market volatility. Consistency among simulated capital market variables exists, such as longer-term bond yield being higher than short-term bond yield in the long run, and higher-rated bond credit spread being smaller than that of lower-rated bonds.

Table 19

Simulated Capital Market Variables vs. Historical Data

Capital market variable	Simulated data		Historical data	
	Mean (%)	Standard deviation (%)	Mean (%)	Standard deviation (%)
tby_1yr	2.10	5.16	4.69	3.31
tby_10yr	3.39	3.98	5.66	2.92
Aaa_cs	1.25	3.67	0.94	0.58
Baa_cs	2.13	4.88	1.92	0.83
BBB_default	0.24	0.29	0.19	0.25
sp500_rtn*	2.12	9.98	2.15	7.17
sp500_divd	2.94	3.01	3.29	1.42
mhp_gr*	1.12	3.24	1.36	2.89
rent_gr*	0.75	0.82	0.95	0.69

*Quarterly growth rate

Appendix B.2 contains graphs of simulated risk factors which show the percentiles to help visualize the range of the simulated scenarios.

5.3 Adjustments with Forward-Looking Views

Although the historical data are an appropriate choice to calibrate the PSG, they may not generate desired forward-looking patterns. A pure data-driven calibration makes it difficult to make manual adjustments. In general, there are three ways to achieve this goal:

1. **Adjust data inputs.** Although this paper uses as much historical data as possible, the desired pattern may be present in more recent historical data. During the PSG calibration process, the pattern may be dominated by older historical data. Removing some older data may be helpful. However, there is a risk that the data are not enough to get statistically credible conclusions. In addition, users may have forward-looking views of possible events that never happened in history, such as a prolonged period of negative interest rates in the U.S. in future pandemics. In these difficult cases, predicted data inputs that reflect the views may be used together with historical data. For example, if a negative-interest-rate environment is expected in the next extreme pandemic event, a few years of future data may be created based on that assumption and used for PSG calibration. The logic is to add more weight to data that reflect the new views so that the model can learn and reflect it. Data inputs may be adjusted to reflect any difference by geographic region or country as well. For example, Canada-specific economic data may be used to reflect differences from the U.S. economic data, even though a high correlation between the two economies during a pandemic event is expected.

2. **Adjust model parameters.** Adjusting model parameters directly may be convenient in some cases to reflect forward-looking views. It is fairly easy to adjust mean and volatility in the PSG for either distribution-based pandemic risk factors, or model-based economic and financial risk factors. For example, if the model generates higher bond yields than expected on average, we can adjust the constant term a in the following PSG to lower the stable value. If higher-than-expected volatility is observed in the simulated scenarios, we can adjust parameter s to lower the volatility of the idiosyncratic risk.

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \boldsymbol{y}_{t-2} + \boldsymbol{B}_0 \mathbf{E}_t + \boldsymbol{B}_1 \boldsymbol{E}_{t-1} + \boldsymbol{B}_2 \mathbf{E}_{t-2} + \boldsymbol{\sigma}\boldsymbol{\varepsilon}_t$$

   Adjusting modelled relationships by adjusting parameters is also possible. Rather than having a single parameter to control all correlations, the PSG models autocorrelation, correlation with systemic risk, and correlation among idiosyncratic risks separately. In the example of the PSG used for capital market variables, $\boldsymbol{\phi}$ controls the autocorrelations, $\boldsymbol{B}$ controls the correlation with systemic risk, and $\boldsymbol{\varepsilon}_t$ is generated from correlated random variables that represent idiosyncratic risks of capital market variables. However, it is not a straightforward adjustment given a specific correlation coefficient of two capital market variables, as they are linked together through both the systemic risk and correlated idiosyncratic risks.

3. **Adjust the generated scenarios.** Generated scenarios may be adjusted directly to reflect different views and to remove unreasonable scenarios. Default rates need to be floored at zero. Credit spreads may also be floored if sovereign risk is believed to be smaller than corporate credit risk. Interest rates and bond yields may be floored at zero or a negative level if a negative interest rate is deemed possible. Modelled variables may also be capped at the historical maximum, and scenarios may also be adjusted to change the average value or add certain trends. This approach is not suitable for adjusting volatilities and relationships.

Although model adjustment is possible to embed a forward-looking view, it is important to understand the impact of the adjustment on all modelled variables and their relationships.

6 Conclusion

Pandemic and epidemic events can cause significant loss to human society, economies, and the insurance industry. To evaluate the impact of these events on the insurance business, a holistic scenario generator is needed to simulate these events directly and, based on these events, to generate appropriate economic scenarios.

The recorded history of pandemic and epidemic events provides a good source to calibrate the frequency, severity, and duration of pandemics. Many diseases have cures or effective ways of containment, and historical experience may not apply in the future. New infectious diseases may appear; in particular, many ancient diseases may reappear due to global warming, against which human beings may no longer have antibodies. Therefore, the historical data, at least to a certain extent, provides insights into possible futures. The frequency, severity, and duration of pandemic and epidemic events can be well represented with common distribution types, including negative binomial, lognormal, gamma, and beta distribution. When the goodness of fit is not satisfactory with a single distribution, EVT-based GPD and state-dependent models such as hidden Markov models can be used to reflect additional volatility.

Unlike the analysis of the frequency and severity of pandemic events, the economic impacts of these events are studied using only the past 180 years of history or more recent ones driven by data availability and change in economic structures. Multiple correlation matrices, copulas, and structured models can be used to describe the relationship between pandemic events and the economic system, although structured models are more suitable for more complex relationships.

A comprehensive PSG is built in this paper that can simulate nonlinearly correlated pandemic events and economic variables, and is calibrated based on available historical data. It provides simulated variables to help insurers measure the impacts of pandemics on claim experience, financial conditions, new business, and such in a holistic way. The PSG can be used to generate stochastic scenarios to facilitate the quantification and management of pandemic-related risks. Extreme pandemic events are embedded into individual scenarios, which also govern the behaviours of the economic system during those simulated events.

The calibration of the PSG depends heavily on the historical data in this study. Undesired calibration results may be caused by limitations on historical data such as incomplete information on ancient events, inconsistent survey methods in the past, and the difficulties of gauging the actual number of infected cases, some of which may never be recorded, as has occurred sometimes in the COVID-19 pandemic. This, however, can be adjusted with views and estimates that are different from available data.

7 Acknowledgments

The author would like to thank all members of the Project Oversight Group (POG) tasked with providing governance on this research project. This paper would not have attained its current level of completeness and relevance to practitioners without the POG’s guidance, feedback, insightful inputs, and reviews.

The “Scenario Generation for Pandemics” POG members are:

- Ibukun Adejuyigbe
- Stella Ma
- Jean-Philippe Morin
- Krizsia Praznik
- Melanie Skiarski
- Salina Young (Chair)

The author would also like to thank Shlomit Jacobson and Esteban Rivera for their effective management and coordination of this project, as well as the sponsorship and funding support of the Canadian Institute of Actuaries.

8 Bibliography

Centers for Disease Control and Prevention, "Diphtheria."
https://www.cdc.gov/diphtheria/index.html

Centers for Disease Control and Prevention, "HIV Basics."
https://www.cdc.gov/hiv/basics/whatishiv.html

Centers for Disease Control and Prevention, "Meningitis."
https://www.cdc.gov/meningitis/index.html

Centers for Disease Control and Prevention, "What is Polio?"
https://www.cdc.gov/polio/what-is-polio/index.htm

Centers for Disease Control and Prevention, "Typhus Fevers."
https://www.cdc.gov/typhus/index.html

CEIOPS, 2010. "CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: SCR STANDARD FORMULA Article 111(d) Correlations."

CRO Forum. 2009. "Calibration recommendation for the correlations in the Solvency II standard formula."
www.thecroforum.org/2009/12/04/solvency-ii-calibration/

Fenner, Frank, Donald A. Henderson, Isao Arita, Zdenek Jezek, Ivan Danilovich Ladnyi, et al. 1988. "Smallpox and its eradication." World Health Organization.
https://apps.who.int/iris/handle/10665/39485

Genest, Christian, Bruno Remillard, and David Beaudoin. 2009. "Goodness-of-fit tests for copulas: A review and a power study." Insurance: Mathematics and Economics 44(2): 199–213.

Kohn, George Childs. 2008. *Encyclopedia of Plague and Pestilence: From Ancient Times to the Present*. New York, NY: Facts on File.

Kojadinovic, Ivan, and Jun Yan. 2011. "A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems." *Statistics and Computing* 21: 17–30.

Maddison, Angus. 2010. "Statistics on world population, GDP and per capita GDP, 1-2008 AD."
www.ggdc.net/maddison/Historical_Statistics/vertical-file_02-2010.xls

Maddison Project Database, version 2020. Bolt, Jutta, and Jan Luiten van Zanden, 2020. *Maddison Style Estimates of the Evolution of the World Economy: A New 2020 Update*.
www.rug.nl/ggdc/historicaldevelopment/maddison/publications/wp15.pdf

Nelsen, Roger B. 1999. *An Introduction to Copulas*. New York, NY: Springer.

S&P Global. 2021. "Default, transition, and recovery: 2020 Annual Global Corporate Default and Rating Transition Study." www.spglobal.com/ratings/en/research/articles/210407-default-transition-and-recovery-2020-annual-global-corporate-default-and-rating-transition-study-11900573

S&P Global. 2021. *Credit Markets Update North America Q2 2021*. www.spglobal.com/_assets/documents/ratings/research/100049977.pdf

Shang, Kailan, and Zakir Hossen. 2019. *Liability-Driven Investment: Benchmark Model*. Society of Actuaries. www.soa.org/globalassets/assets/files/resources/research-report/2019/liability-driven-investment.pdf

Shang, Kailan, and Marc Vincelli. 2015. *Report on Extreme Events for Insurers: Correlation, Models and Mitigations*.

Society of Actuaries. www.soa.org/globalassets/assets/files/research/projects/2015-extreme-events-for-insurers.pdf

Shiller, Robert J. 1992. "Market volatility." Data link: www.econ.yale.edu/~shiller/data/chapt26.xlsx

Shiller, Robert J., 2015. "Irrational Exuberance." Data link: www.econ.yale.edu/~shiller/data/ie_data.xls

United Nations Department of Economic and Social Affairs/Population Division. 2004. *The Impact of AIDS*, Chapter 8: Impact on Economic Growth. www.un.org/en/development/desa/population/publications/pdf/hiv/impact/chap8.pdf

World Health Organization. 2019. "Measles." https://www.who.int/news-room/fact-sheets/detail/measles

Appendix A. PSG Calibration

A.1 Historical Pandemic Events

In Subsection 2.2, the historical data of both pandemic and epidemic events are studied. It is also helpful to analyze pandemic events on their own when designing more severe yet possible scenarios. Figure A.1 shows the frequency of pandemic events every 50 years since 500 AD.

Figure A.1

Pandemic Frequency

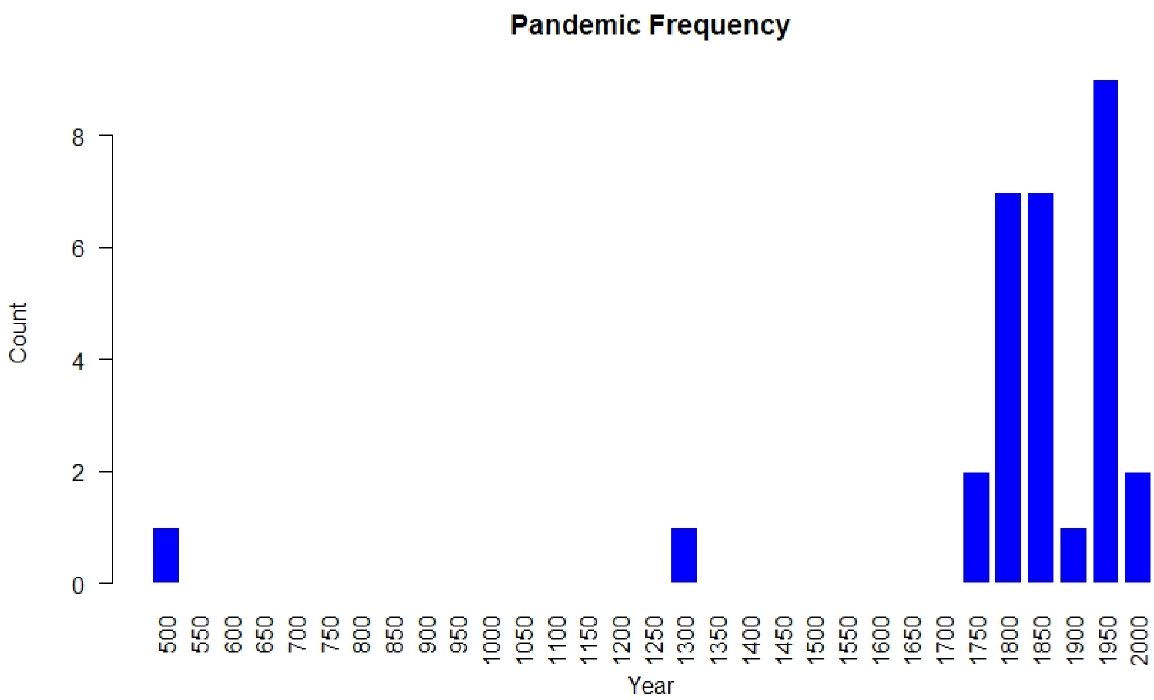


Table A.1 shows the descriptive statistics of the 30 pandemic events where available. As expected, average numbers of deaths and confirmed cases are higher compared to Table 2. Both case fatality rate (CFR) and infection rate (IR) are higher on average as well. Average mortality rate is lower, but the affected population is much larger. The average mortality rate for all pandemic and epidemic events is skewed upward by a few outliers, such as the Hispaniola smallpox epidemic of 1518, which wiped out almost all the affected population.

Table A.1

Descriptive Statistics of Pandemic Data

Statistic	Duration (years)	Deaths	Confirmed cases	Case fatality rate	Mortality rate	Infection rate
No. of valid records	30	16	7	12	3	8
Min.	1	14,756	210,000	0.5%	0.1%	15.0%
Max.	110	50,000,000	200,000,000	81.6%	10.0%	80.0%
Median	3	734,433	24,500,000	29.9%	0.1%	39.8%
Mean	13	7,891,782	65,851,873	33.3%	3.4%	41.4%
Standard deviation	22	14,124,921	82,656,882	28.0%	5.7%	19.8%
Coefficient of variation	1.71	1.79	1.26	0.84	1.68	0.48

Figure A.2 shows the histogram of duration, and Figure A.3 the distribution of the three severity measures using a boxplot.

Figure A.2

Histogram of Pandemic Duration

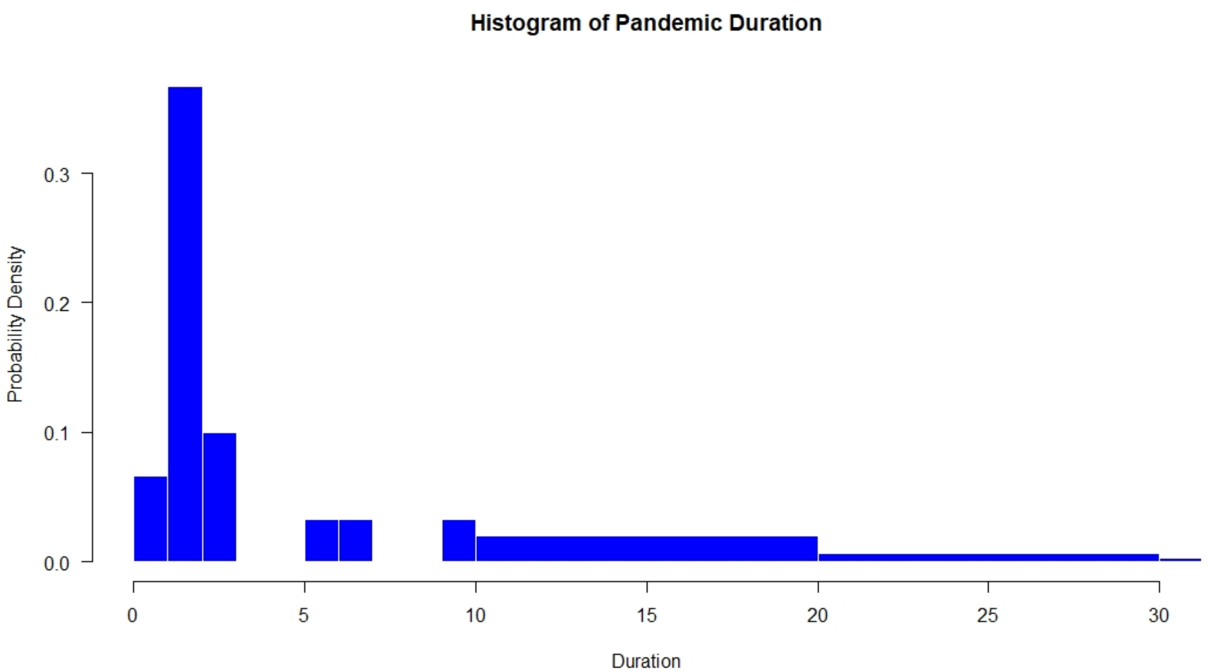
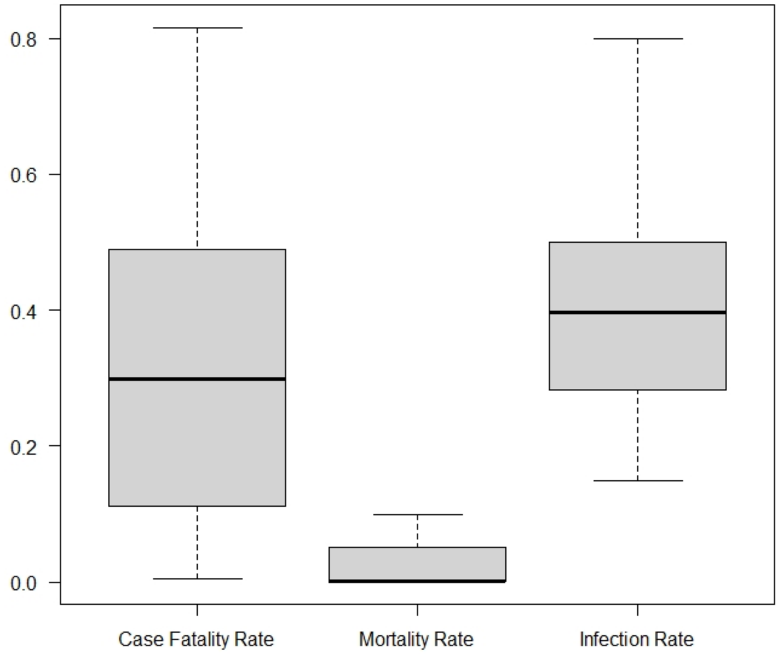


Figure A.3

Boxplots of Pandemic Severity Measures



A.2 Frequency Distribution Fitting

Table A.2 lists the fitting results for the annual frequency data, with negative binomial distribution having the best goodness of fit measured by loglikelihood, Akaike information criteria (AIC), and Bayesian information criteria (BIC).

Table A.2

Frequency Distribution Fitting Result

Distribution	Estimate	Std. error	Loglikelihood	AIC	BIC
Poisson	lambda = 2.7	lambda: 0.11	-495	992	995
Geometric	prob = 0.27	prob: 0.016	-474	951	954
Negative binomial	size = 2.7 mu = 2.7	size: 0.52 mu: 0.16	-458	920	927

The frequency is modelled by combining the negative binomial distribution describing annual frequency no greater than 7 and the extreme value theory (EVT)-based generalized Pareto distribution (GPD) modelling the exceedance beyond 7. According to the EVT, if the threshold u is large, the distribution of exceedances follows a GPD.

$$\Pr(X - u < y | X > u) \sim F(y) = \begin{cases} 1 - (1 + \xi \dfrac{y}{\sigma_u})^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma_u}}, & \xi = 0 \end{cases}$$

Where

u: threshold
ξ: tail index
σ_u: scaling factor

Exceedances without an upper bound will have $\xi \geq 0$. To model a bounded series of exceedances, $\xi < 0$ and the upper bound is $-\frac{\sigma_u}{\xi}$.

A.3 Duration Distribution Fitting

Table A.3 lists the fitting results with the lognormal distribution having the best goodness of fit.

Table A.3

Duration Distribution Fitting Result

Distribution	Estimate	Std. error	Loglikelihood	AIC	BIC
Exponential	rate = 0.26	rate: 0.0093	-1860	3722	3726
Gaussian	mean = 3.8 sd = 6.3	mean: 0.22 sd: 0.16	-2597	5199	5208
Lognormal	meanlog = 0.68 sdlog = 1.05	meanlog: 0.037 sdlog: 0.026	-1709	3422	3431
Gamma	shape = 0.89 rate = 0.23	shape: 0.039 rate: 0.013	-1856	3716	3725
Weibull	shape = 0.85 scale = 3.41	shape: 0.021 rate: 0.151	-1836	3675	3685

A.4 Severity Distribution Fitting

Case Fatality Rate

Table A.4 shows the fitting results of CFR, with the gamma distribution having the largest loglikelihood.

Table A.4

Case Fatality Rate Distribution Fitting Result

Distribution	Estimate	Std. error	Loglikelihood	AIC	BIC
Exponential	rate = 3.6	rate: 0.21	87	-171	-168
Gaussian	mean = 0.28 sd = 0.27	mean: 0.015 sd: 0.011	-29	62	70
Lognormal	meanlog = -2.1 sdlog = 1.8	meanlog: 0.104 sdlog: 0.074	33	-62	-55
Gamma	shape = 0.73 rate = 2.64	shape: 0.051 rate: 0.257	98	-192	-185
Weibull	shape = 0.84 scale = 0.25	shape: 0.04 rate: 0.018	94	-184	-176
Uniform	min = 0.000015 max = 1	NA	0.0045	4	11
Beta	shape1 = 0.49 shape2 = 1.30	NA*	NA	NA	NA

* Sometimes, the optimization routine finishes without achieving the default goodness-of-fit criterion, and it may produce certain outputs. However, it does not mean the fitted distribution is worse than those with all the outputs.

However, the Q–Q plot shows that the beta distribution is doing a better job matching the tails. Figures A.4 and A.5 compare the empirical data with the calibrated gamma and beta distribution, respectively.

Figure A.4

Case Fatality Rate Distribution Fitting Comparison: Gamma Distribution

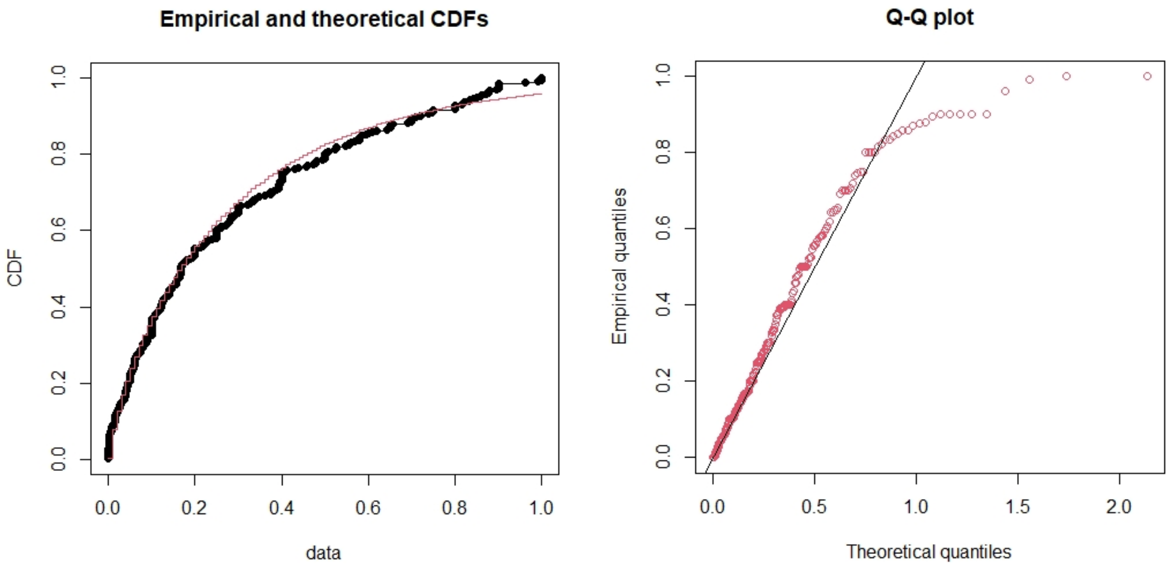
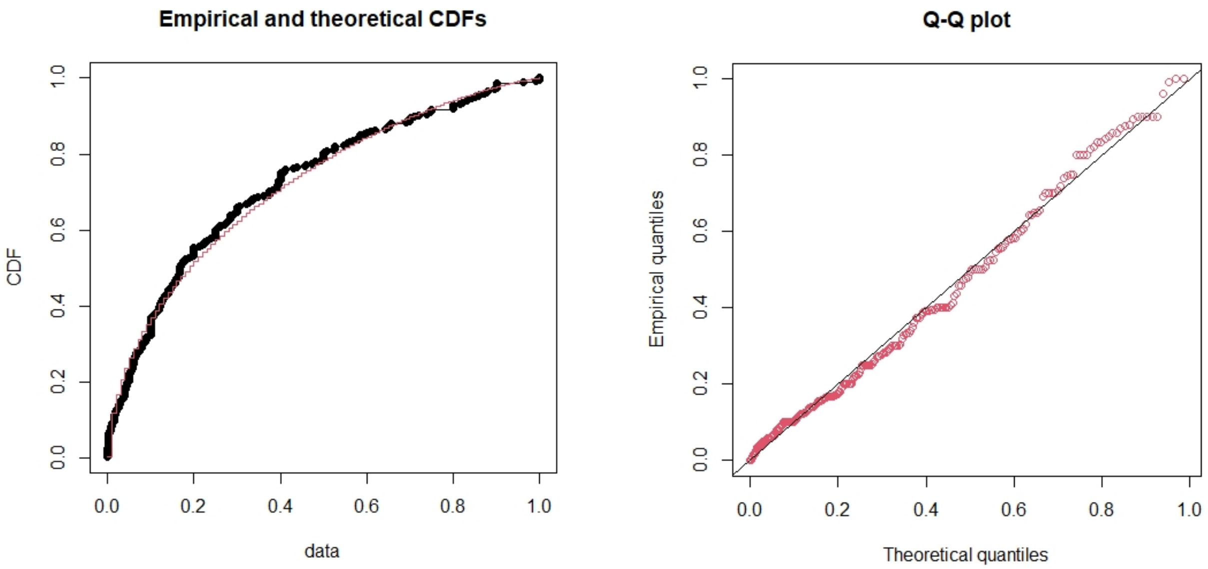


Figure A.5

Case Fatality Rate Distribution Fitting Comparison: Beta Distribution



Mortality Rate

Table A.5 shows the fitting results of mortality rate, with the gamma distribution having the largest loglikelihood.

Table A.5

Mortality Rate Distribution Fitting Result

Distribution	Estimate	Std. error	Loglikelihood	AIC	BIC
Exponential	rate = 5.8	rate: 0.51	97	-191	-189
Gaussian	mean = 0.17 sd = 0.22	mean: 0.019 sd: 0.013	15	-25	-20
Lognormal	meanlog = -3.1 sdlog = 2.4	meanlog: 0.21 sdlog: 0.15	103	-203	-197
Gamma	shape = 0.48 rate = 2.76	shape: 0.049 rate: 0.454	129	-254	-248
Weibull	shape = 0.61 scale = 0.13	shape: 0.045 rate: 0.019	125	-246	-240
Uniform	min = 0.000003 max = 0.996667	NA	0.43	3.1	8.8
Beta	shape1 = 0.39 shape2 = 1.61	shape1: 0.039 shape2: 0.232	128	-253	-247

Infection Rate

Table A.6 shows the fitting results of mortality rate, with the gamma distribution having the largest loglikelihood.

Table A.6

Infection Rate Distribution Fitting Result

Distribution	Estimate	Std. error	Loglikelihood	AIC	BIC
Exponential	rate = 3.4	rate: 0.41	15	-29	-27
Gaussian	mean = 0.29 sd = 0.27	mean: 0.033 sd: 0.023	-7.1	18	23

Distribution	Estimate	Std. error	Loglikelihood	AIC	BIC
Lognormal	meanlog = -2.1 sdlog = 1.8	meanlog: 0.22 sdlog: 0.16	2.4	-0.84	3.6
Gamma	shape = 0.73 rate = 2.47	shape: 0.11 rate: 0.51	18	-32	-28
Weibull	shape = 0.85 scale = 0.27	shape: 0.087 rate: 0.041	17	-30	-25
Uniform	min = 0.00052 max = 1	NA	0.035	3.9	8.4
Beta	shape1 = 0.55 shape2 = 1.32	NA	NA	NA	NA

Similar to the CFR, the Q–Q plot shows that the beta distribution is doing a better job matching the tails. Figures A.6 and A.7 compare the empirical data with the calibrated gamma and beta distribution, respectively.

Figure A.6

Infection Rate Distribution Fitting Comparison: Gamma Distribution

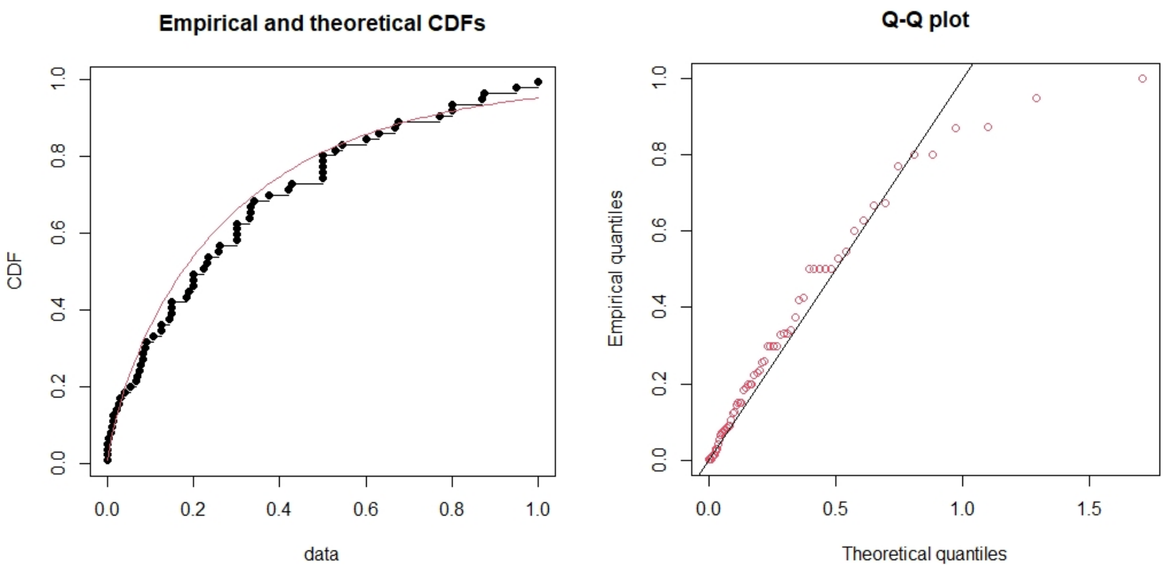
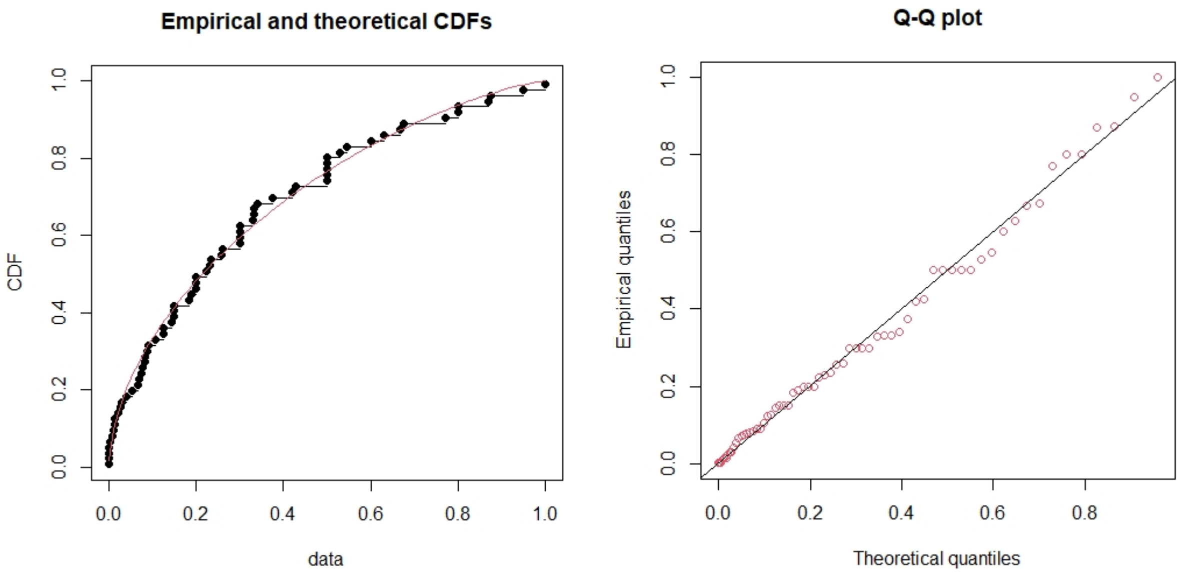


Figure A.7

Infection Rate Distribution Fitting Comparison: Beta Distribution



It is also helpful to identify any structural changes of these severity rate measures given the social, medical, and economic changes during human history. Figures A.8 to A.10 show the changes by the start year of the pandemic/epidemic events. The range of CFR seems to be quite stable over time, with old diseases eradicated but new diseases emerging. The mortality rate has a downward trend, largely due to better medical care and effective containment measures. The IR was also going down, but the data may not be sufficient to have a definite answer.

Figure A.8

Case Fatality Rate by Year

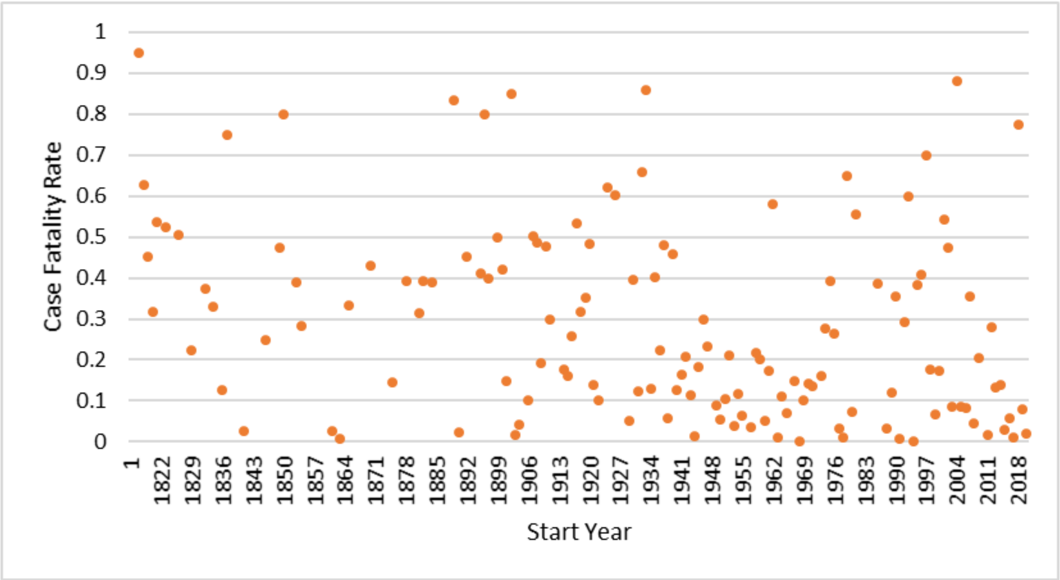


Figure A.9

Mortality Rate by Year

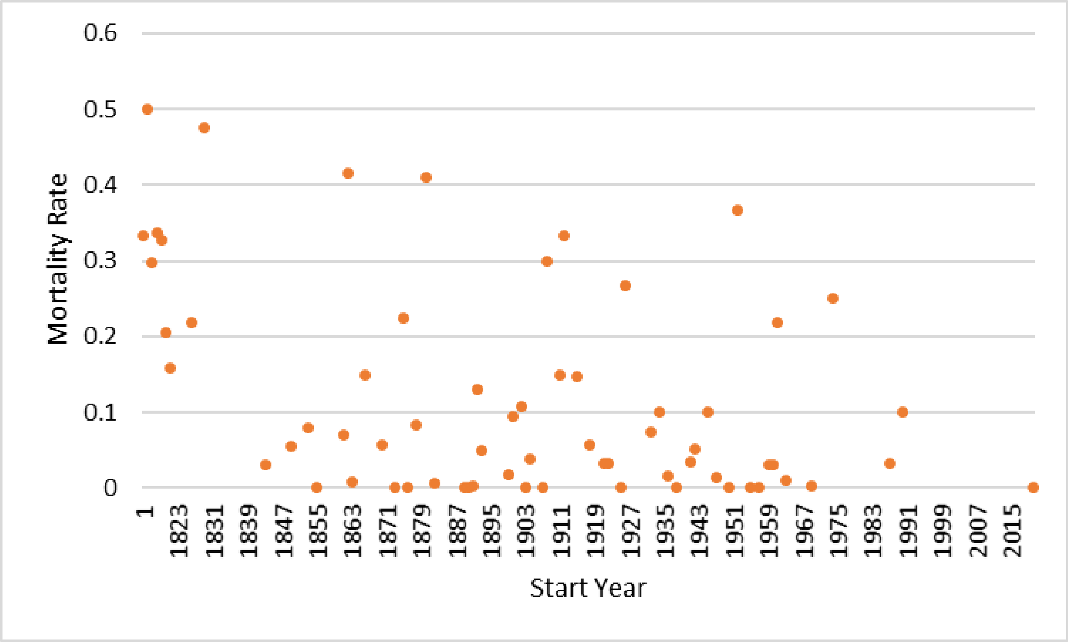
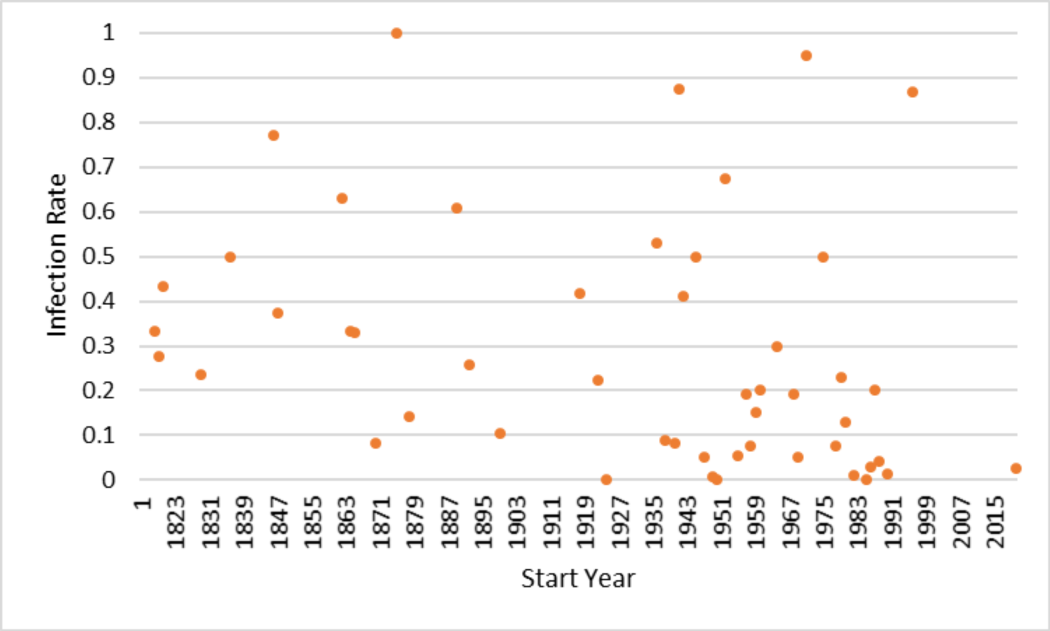


Figure A.10

Infection Rate by Year



A.5 Hidden Markov Model

A hidden Markov model (HMM) studies the Markov process of hidden states with observations that highly depend on the hidden states. The next hidden state depends on the current hidden state, but not the history of the hidden states. In most cases, a transition matrix is used to define the probability of the next state given the current one. The distribution of observations changes with the hidden state. Based on actual observation, an inference system built on Bayes's rule can be used to predict future hidden states. The states can be considered different phases of a cycle. For example, an HMM can be used to identify high frequency and low frequency. Distributions of frequency and severity can be different between the high-frequency state and low-frequency state. An HMM uses randomness to describe the transition between different states.

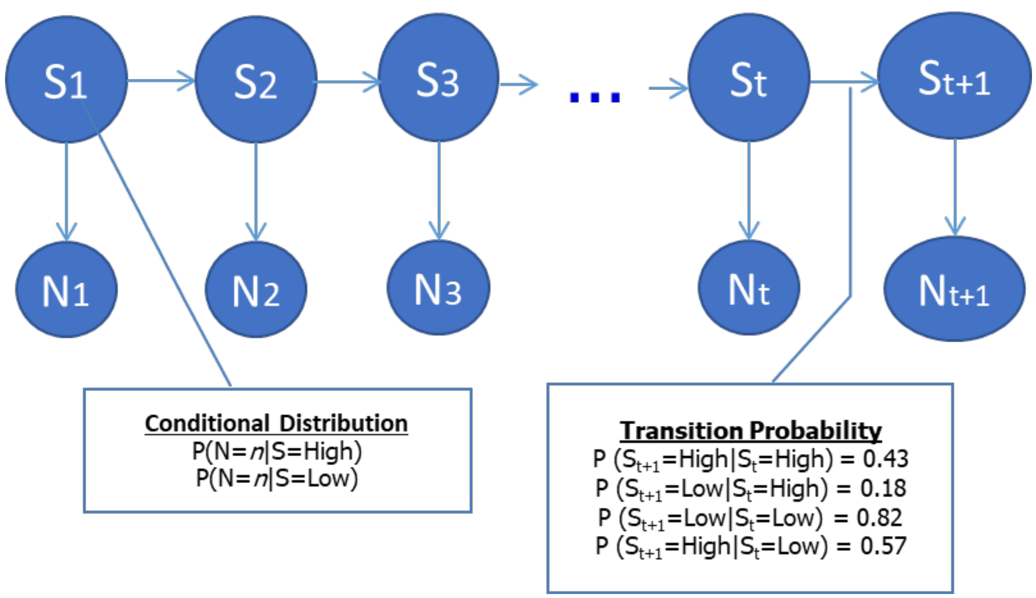
The basic setup of HMM is the same as the Markov Chain Monte Carlo (MCMC) model and the regime-switching model. However, in a HMM, states are not observable and can be inferred based only on observation of other related variables. As the related variables are affected by the state, their values can indicate the current state in a probabilistic way. The state inferred from the observations is important in projecting the future paths of states and their outcomes. Similar to a regime-switching model, a HMM captures additional variability introduced by state changes that could not otherwise be reflected by a single distribution.

In Figure 6, where the annual frequency of pandemic/epidemic events is shown from 1801 to 2020, there were multiple-year periods when more pandemic/epidemic events than average occurred. To capture this pattern, a HMM is calibrated to add another layer of uncertainty in frequency modelling on top of the distribution fitting approach.

In different phases of history, the distributions of pandemic/epidemic frequency have been different. There are two states assumed: "High" frequency and "Low" frequency. The annual frequency of pandemic/epidemic events can be any integer from 0 to 6, with any value above 6 truncated at 6 given a low probability of 6%. Figure A.11 illustrates the model structure, with the transition probabilities shown as an example.

Figure A.11

HMM Structure Example



Based on historical frequency data from 1800 to 2020, the Baum–Welch algorithm[2] is applied, and the calibrated HMM is given in Table A.7. The calibrated conditional distributions are reasonable when comparing between the High and Low states, which represent the high and low frequency.

Table A.7

HMM Calibration Result

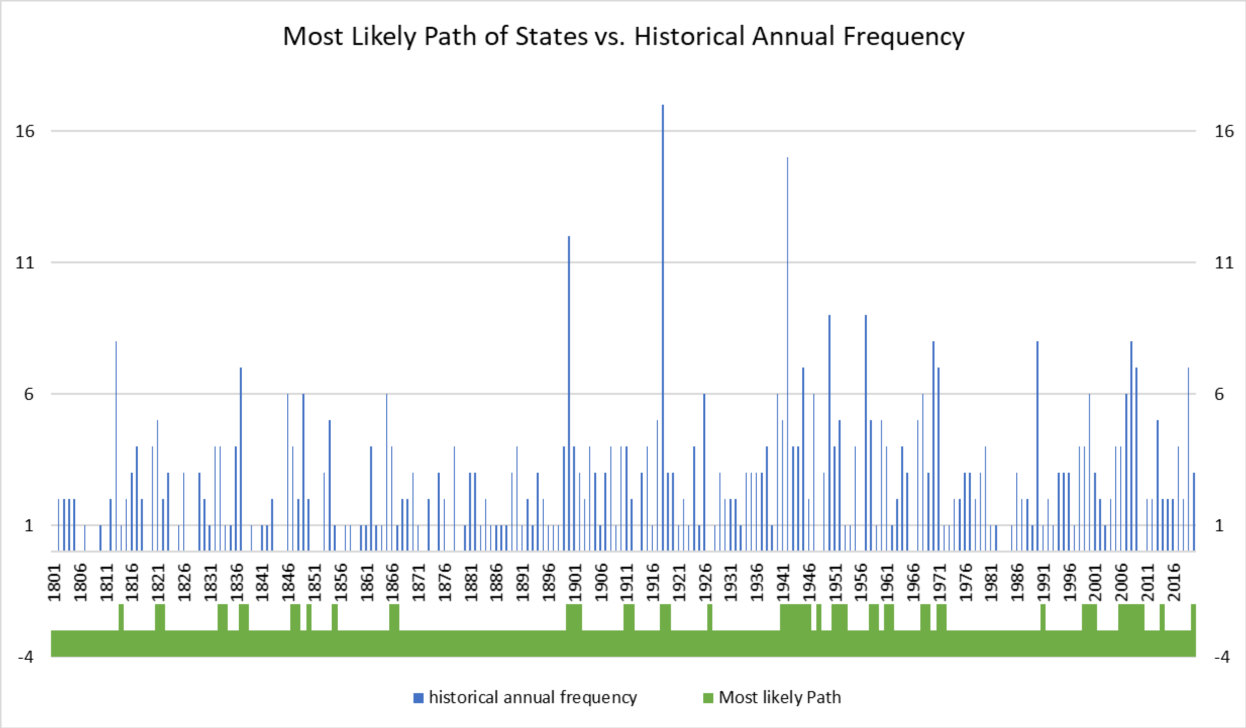
<table>
<tr><th>Two states</th><th>“H”: High, “L”: Low</th></tr>
<tr><td>Initial state probability</td><td>$\pi_{\mathrm{H}} = 20\%$, $\pi_{\mathrm{L}} = 80\%$</td></tr>
<tr><td>Transitional probability</td><td>
<table>
<tr><td>Probability</td><td>S_{t+1} = “H”</td><td>S_{t+1} = “L”</td></tr>
<tr><td>$S_t = \mathrm{H}$</td><td>43%</td><td>57%</td></tr>
<tr><td>$S_t = \mathrm{L}$</td><td>18%</td><td>82%</td></tr>
</table>
</td></tr>
<tr><td>Annual frequency</td><td>0, 1, 2, 3, 4, 5, and 6+, denoted as a, b, c, d, e, f, and g, respectively.</td></tr>
<tr><td>Conditional distribution of annual frequency</td><td>
<table>
<tr><td>Probability</td><td>S_i = “H”</td><td>S_i = “L”</td></tr>
<tr><td>$\mathrm{P}(a \mid S_i)$</td><td>0.0%</td><td>17%</td></tr>
<tr><td>$\mathrm{P}(b \mid S_i)$</td><td>0.0%</td><td>31%</td></tr>
<tr><td>$\mathrm{P}(c \mid S_i)$</td><td>0.0%</td><td>24%</td></tr>
<tr><td>$\mathrm{P}(d \mid S_i)$</td><td>0.027%</td><td>20%</td></tr>
<tr><td>$\mathrm{P}(e \mid S_i)$</td><td>38%</td><td>7%</td></tr>
<tr><td>$\mathrm{P}(f \mid S_i)$</td><td>18%</td><td>0%</td></tr>
<tr><td>$\mathrm{P}(g \mid S_i)$</td><td>45%</td><td>0%</td></tr>
</table>
</td></tr>
</table>

Using the Viterbi algorithm,[3] widely used in dynamic programming, the most likely path of states in the historical data can be found. Figure A.12 shows the most likely path against the historical annual frequency of pandemic/epidemic events. A high green bar stands for a state of High from the calibrated model, with the rest having a state of Low. The path fits the historical experience but with some uncertainty. For some years with a high number of pandemic/epidemic events, the state is Low. For example, the years 1878 and 1890 are considered to be in the Low state, while the year 1991 is considered to be in the High state. All of these years have four events recorded.

[2] Details can be found at http://en.wikipedia.org/wiki/Baum%E2%80%93Welch_algorithm.
[3] Details can be found at http://en.wikipedia.org/wiki/Viterbi_algorithm#History.

Figure A.12

Most Likely Path for Pandemic/Epidemic Historical Frequency (1800–2020)



With the calibrated HMM, the future distribution of annual frequency can be simulated. We will test two scenarios: (a) where the current state is High, and (b) where the current state is Low. Given the updated HMM, the state and frequency for the next period are simulated. For example, if the current state is High, the simulator will first simulate the state of the next period based on the transition matrix. After that, the model will simulate the number of pandemic/epidemic events based on the conditional distribution given the state. Simulation results are summarized in Table A.8.

Table A.8

HMM Simulation vs. Historical Data

Statistic	No. of pandemic/epidemic events in one year		
	Historical experience[1]	Simulation: Initial state Low	Simulation: Initial state High
Average	2.47	2.23	3.12
Standard deviation	1.82	1.70	2.04
VaR(70)[2]	3	3	4
VaR(80)	4	4	5
VaR(90)	6	5	6
VaR(95)	6	6	6

Notes:

1. Historical experience includes data from 1801–2020, as used in frequency distribution fitting.
2. VaR(X): Value at risk at the Xth percentile.

In general, we can see that the average and VaR values of the simulated paths starting in the High state are higher than those of the historical experience. For extreme risk modelling and management, the simulated paths starting in the High state are the most useful of the three datasets because they more fully capture the potential adverse outcomes.

If a state-based model is used to model the frequency, it is helpful to understand that in addition to the frequency itself, other variables, such as severity and duration, also vary by state. Table A.9 shows the descriptive statistics by state based on historical data. This can be used to define the attributes of states which can be used in state-based scenario generation.

Table A.9

State-Dependent Severity and Duration Statistics

Statistic	State	Case fatality rate	Mortality rate	Infection rate	Duration	Frequency
Average	L	0.29	0.10	0.29	5.39	2.68
	H	0.24	0.09	0.26	3.62	3.02
Standard deviation	L	0.23	0.12	0.29	7.84	9.46
	H	0.20	0.13	0.24	3.92	2.82

Appendix B. PSG Simulation

B.1 Simulation Process

The detailed process of generating scenarios using the calibrated models with the best goodness of fit to historical data is described below. This appendix is an expansion of Subsection 5.1.

Step 1

Generate pandemic/epidemic event frequencies. For annual frequency of pandemic/epidemic events, the negative binomial distribution with the GPD describing the tail is chosen, giving it a good fit to the historical experience. Its probability density function is given below:

$$f(freq = x) \sim \begin{cases} \binom{size + x}{x} p^{size}(1-p)^x & if\ x \le u \\ \frac{1}{\sigma_u}\left(1 + \xi \frac{x-u}{\sigma_u}\right)^{-\frac{1}{\xi}-1} (1 - F(freq \le u)) & if\ x > u \end{cases}$$

Calibrated parameters based on historical data:

- $size$ = 2.7
- p = 0.5
- u = 7
- σ_u = 4.03
- ξ = -0.16

HMM can be used as well, but the advantage is immaterial given the similar level of severity under high and low frequency. The actual occurrence date follows a uniform distribution within a year.

Step 2

For each event simulated from Step 1, generate duration and severity measures including CFR and IR.

Duration (D)

$$f(D = x) \sim \begin{cases} \frac{1}{\sqrt{2\pi}x\sigma} e^{-\frac{(lnx-\mu)^2}{2\sigma^2}} & if\ x \le u \\ \frac{1}{\sigma_u}\left(1 + \xi \frac{x-u}{\sigma_u}\right)^{-\frac{1}{\xi}-1} (1 - F(D \le u)) & if\ x > u \end{cases}$$

Calibrated parameters based on historical data:

- μ = 0.68
- σ = 1.05
- u = 7
- σ_u = 14.86
- ξ = -0.27

Case Fatality Rate (CFR)

$$f(CFR = x) \sim \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\left(1 + \xi \frac{x-u}{\sigma_u}\right)^{-\frac{1}{\xi}-1}$$

Calibrated parameters based on historical data:

- α = 0.49
- β = 1.30

Infection Rate (IR)

$$f(IR = x) \sim \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\left(1+\xi\frac{x-u}{\sigma_u}\right)^{-\frac{1}{\xi}-1}$$

Calibrated parameters based on historical data:

- α = 0.55
- β = 1.32

When generating the three variables in Step 2, the correlation matrix in Table B.1 is used based on historical data. Correlated marginal cumulative probabilities are generated first, and values of variables are calculated based on the cumulative probability functions and generated cumulative probabilities.

Table B.1

Duration and Severity Correlation Matrix (1050 BC to AD 2020)

	Duration	CFR	IR
Duration	1.00	0.06	-0.23
CFR	0.06	1.00	0.17
IR	-0.23	0.17	1.00

The severity measure generated above is at the total level. As different age groups may experience different severity levels, a multinomial distribution based on the findings in Figure 5 can be used to simulate the fatalities in different age groups: infant (age 0–5), with a share of 32%; children (6–20), with a share of 30%; young adult (21–40), with a share of 14%; mid-aged adult (41–60), with a share of 11%; and old-aged adult (60+), with a share of 13%.

Step 3

Based on the simulated severity results for each pandemic/epidemic event, determine whether there is any extreme event and how long will such an event last. The extreme pandemic events were defined as events with no fewer than 800,000 death counts and/or no fewer than 12,000,000 infected cases in Subsection 4.1 to study nonlinear relationships with the financial risk.

- For all the events in previous steps, a binomial distribution with a probability of 3.7% is used to determine whether a generated event is a pandemic event. The probability 3.7% is derived as the ratio of 30 historical pandemics and 817 pandemics and epidemics in total.
- If a generated event is tagged as a pandemic event, the number of infected cases will be calculated as the product of the current world population and the simulated IR for that event. Death counts will be calculated as the product of the number of infected cases and the simulated CFR for that event. If either one of them exceeds the threshold, the generated event will be determined as an extreme event for that specific scenario. Different economic scenario assumptions such as the volatility of idiosyncratic risks and correlation

among idiosyncratic risks and the systemic risk will be used depending on the occurrence of a simulated extreme pandemic event.

- For each simulated epidemic event, the death count and the number of infected cases are not explicitly calculated given the limitation of modelling the geographic reach of the event. Although the historical data used in this report do not support such details, a possible way to use these epidemic events is to apply them directly to the regions with written insurance businesses.

Step 4

Generate economic factors based on VAR models.

$$\mathbf{E}_t = \mathbf{c} + \boldsymbol{A}_E \mathbf{E}_{t-1} + \boldsymbol{\sigma\varepsilon}_t$$

Table 8 lists the calibration results for $\mathbf{c}$ and $\boldsymbol{A}_E$, and σ, the standard deviation of error vector $\mathbf{e}_t$. Table 10 lists the correlation matrix among the seven economic factors. By using Cholesky decomposition, a correlation matrix $\mathbf{CM}$, such as that in Table 19, can be decomposed as the product of a lower triangular matrix $\boldsymbol{L}$ and its transpose $\boldsymbol{L}^T$, given that the correlation matrix is positive definite. $\boldsymbol{L\varepsilon}_t$ has the same correlation matrix from which $\boldsymbol{L}$ is derived, as shown below:

$$\mathrm{Cov}(\boldsymbol{L\varepsilon}_t, \boldsymbol{L\varepsilon}_t) = \mathbb{E}\left(\boldsymbol{L\varepsilon}_t(\boldsymbol{L\varepsilon}_t)^T\right) = \mathbb{E}\left(\boldsymbol{L\varepsilon}_t(\boldsymbol{\varepsilon}_t)^T\boldsymbol{L}^T\right) = \boldsymbol{L}\mathbb{E}\left(\boldsymbol{\varepsilon}_t(\boldsymbol{\varepsilon}_t)^T\right)\boldsymbol{L}^T = \boldsymbol{L} \times \boldsymbol{I} \times \boldsymbol{L}^T = \boldsymbol{L} \times \boldsymbol{L}^T = \mathbf{CM}$$

The economic factors can then be generated as below:

$$\mathbf{E}_t = \mathbf{c} + \boldsymbol{A}_E \mathbf{E}_{t-1} + \boldsymbol{\sigma} \cdot \boldsymbol{L\varepsilon}_t$$

Step 5

Generate capital market variables based on linear models.

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \mathbf{y}_{t-2} + \boldsymbol{B}_0 \mathbf{E}_t + \boldsymbol{B}_1 \boldsymbol{E}_{t-1} + \boldsymbol{B}_2 \mathbf{E}_{t-2} + \boldsymbol{\sigma\varepsilon}_t$$

Table 13 lists the calibration results for constant term $\boldsymbol{\alpha}$, autocorrelation coefficient $\boldsymbol{\phi}$, relationships with economic factors $\boldsymbol{B}$, and volatility of the error term $\boldsymbol{\sigma}$. The idiosyncratic part of capital market variables $\boldsymbol{\sigma\varepsilon}_t$ needs to be adjusted to reflect non-constant volatility and nonlinear relationships as in the following:

- Construct two correlation matrices from the error terms of lasso models. The first correlation matrix, $\mathbf{CM}_{Normal}$, as shown on the left of Figure 23, describes the relationships among error terms of all modelled capital market variables, using historical data during normal periods. The second correlation matrix, $\mathbf{CM}_{\mathbf{Extreme}}$, as shown on the right of Figure 23, describes the relationships among error terms during periods with extreme pandemic events. Cholesky decomposition can then be performed to get the lower triangular matrices $\boldsymbol{L}_{\mathbf{Normal}}$ and $\boldsymbol{L}_{\mathbf{Extreme}}$ to generate correlated idiosyncratic factors.
- Generate correlated idiosyncratic factors $\mathbf{I}_t^i$ for all modelled capital variables during period *t* under scenario *i*:

$$\mathbf{I}_t^i = \begin{cases} \boldsymbol{\sigma}_{\mathbf{Normal}} \cdot \boldsymbol{L}_{\mathbf{Normal}} \boldsymbol{\varepsilon}_t^i & \text{if in normal period during period } t \text{ under scenario } i \\ \boldsymbol{\sigma}_{\mathbf{Extreme}} \cdot \boldsymbol{L}_{\mathbf{Extreme}} \boldsymbol{\varepsilon}_t^i & \text{if in extreme period during period } t \text{ under scenario } i \end{cases}$$

where

$\boldsymbol{\sigma}_{\mathbf{Normal}}$ is a column vector containing the standard deviation of error terms of all capital market variables in normal periods, as shown in column 2 of Table 23;

$\boldsymbol{\sigma}_{\mathbf{Extreme}}$ is a column vector containing the standard deviation of error terms of all asset return models in normal periods, as shown in column 3 of Table 23;

$\boldsymbol{\varepsilon}_t^i$ is a column vector containing independent random variables following a standard normal distribution for all capital market variables;

$\boldsymbol{L}_{\mathbf{Normal}}$ is a lower triangular matrix so that the error term correlation matrix $\mathbf{CM}_{\boldsymbol{Normal}}$ can be decomposed as $\boldsymbol{L}_{\mathbf{Normal}} \times \boldsymbol{L}_{\mathbf{Normal}}{}^{T}$; and

$\boldsymbol{L}_{\mathbf{Extreme}}$ is a lower triangular matrix so that the error term correlation matrix $\mathbf{CM}_{\mathbf{Extreme}}$ can be decomposed as $\boldsymbol{L}_{\mathbf{Extreme}} \times \boldsymbol{L}_{\mathbf{Extreme}}{}^{T}$.

Till this step, non-constant volatility and nonlinear relationships among idiosyncratic factors of capital market variables have been taken care of. The economic-scenario-generation formula becomes

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \mathbf{y}_{t-2} + \boldsymbol{B}_0 \mathbf{E}_t + \boldsymbol{B}_1 \boldsymbol{E}_{t-1} + \boldsymbol{B}_2 \mathbf{E}_{t-2} + \mathbf{I}_t^i$$

$$\text{Let } \mathbf{y}_t^p = \boldsymbol{\alpha} + \boldsymbol{\phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \mathbf{y}_{t-2} + \boldsymbol{B}_0 \mathbf{E}_t + \boldsymbol{B}_1 \boldsymbol{E}_{t-1} + \boldsymbol{B}_2 \mathbf{E}_{t-2}$$

$$\mathbf{y}_t = \mathbf{y}_t^p + \mathbf{I}_t^i$$

- Adjust $\mathbf{I}_t^i$ to reflect non-zero correlation between idiosyncratic factors and systemic factors during extreme pandemic events:

$$\mathbf{I}(y_t^i) = \begin{cases} \left(\boldsymbol{\rho}_r \cdot (\mathbf{y}_t^p - \bar{y}) + \sqrt{\mathbf{1} - \boldsymbol{\rho}_r{}^2} \cdot \mathbf{I}_t^i\right) \dfrac{\boldsymbol{\sigma}_{\mathbf{Extreme}}}{\sqrt{\boldsymbol{\rho}_r{}^2 \left(\boldsymbol{\sigma}_{\mathbf{Extreme}}^p\right)^2 + (\mathbf{1} - \boldsymbol{\rho}_r{}^2)(\boldsymbol{\sigma}_{\mathbf{Extreme}})^2}} & \text{if in extreme period during period } t \text{ under scenario } i \\ \mathbf{I}_t^i & \text{otherwise} \end{cases}$$

where $\boldsymbol{\rho}_r$ is a column vector containing the non-zero correlation between idiosyncratic factors and systemic factors, as shown in column 5 of Table 23. $\bar{y}$ is the expected/stable value of variable $\boldsymbol{y}$.

- The economic-scenario-generation formula of capital market variables is as follows:

$$\mathbf{y}_t = \boldsymbol{\alpha} + \boldsymbol{\phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\phi}_2 \mathbf{y}_{t-2} + \boldsymbol{B}_0 \mathbf{E}_t + \boldsymbol{B}_1 \boldsymbol{E}_{t-1} + \boldsymbol{B}_2 \mathbf{E}_{t-2} + \mathbf{I}(\mathbf{y}_t^i)$$

B.2 Sample Simulations

Percentiles of sample simulated risk factors are presented in the graphs below. This appendix is an expansion of Subsection 5.2.

Figures B.1 to B.4 show the percentiles of frequency, duration, CFR, and IR by time period. No time-based trend is spotted, which is consistent with our findings based on the historical data.

Figure B.1

Percentiles of Simulated Quarterly Frequency

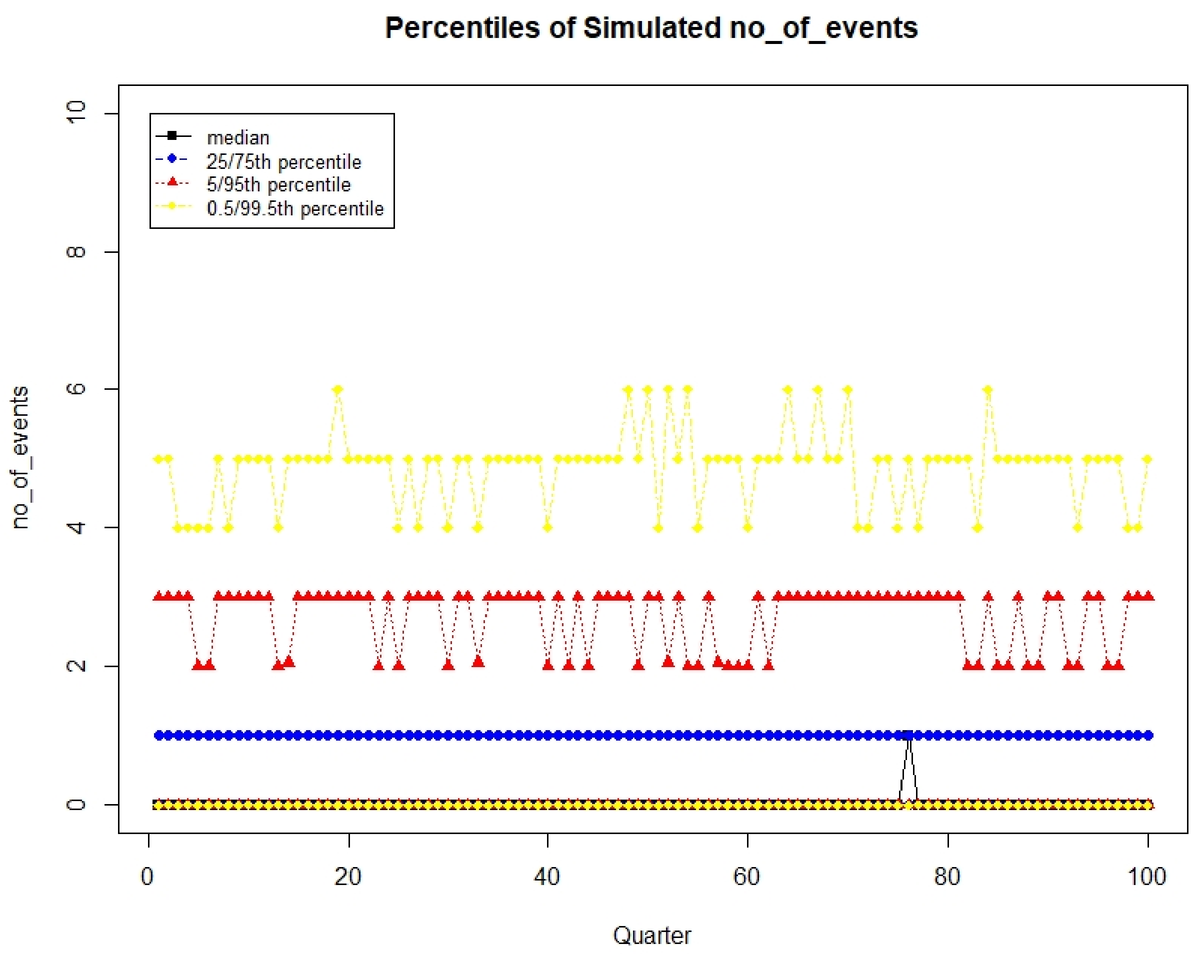
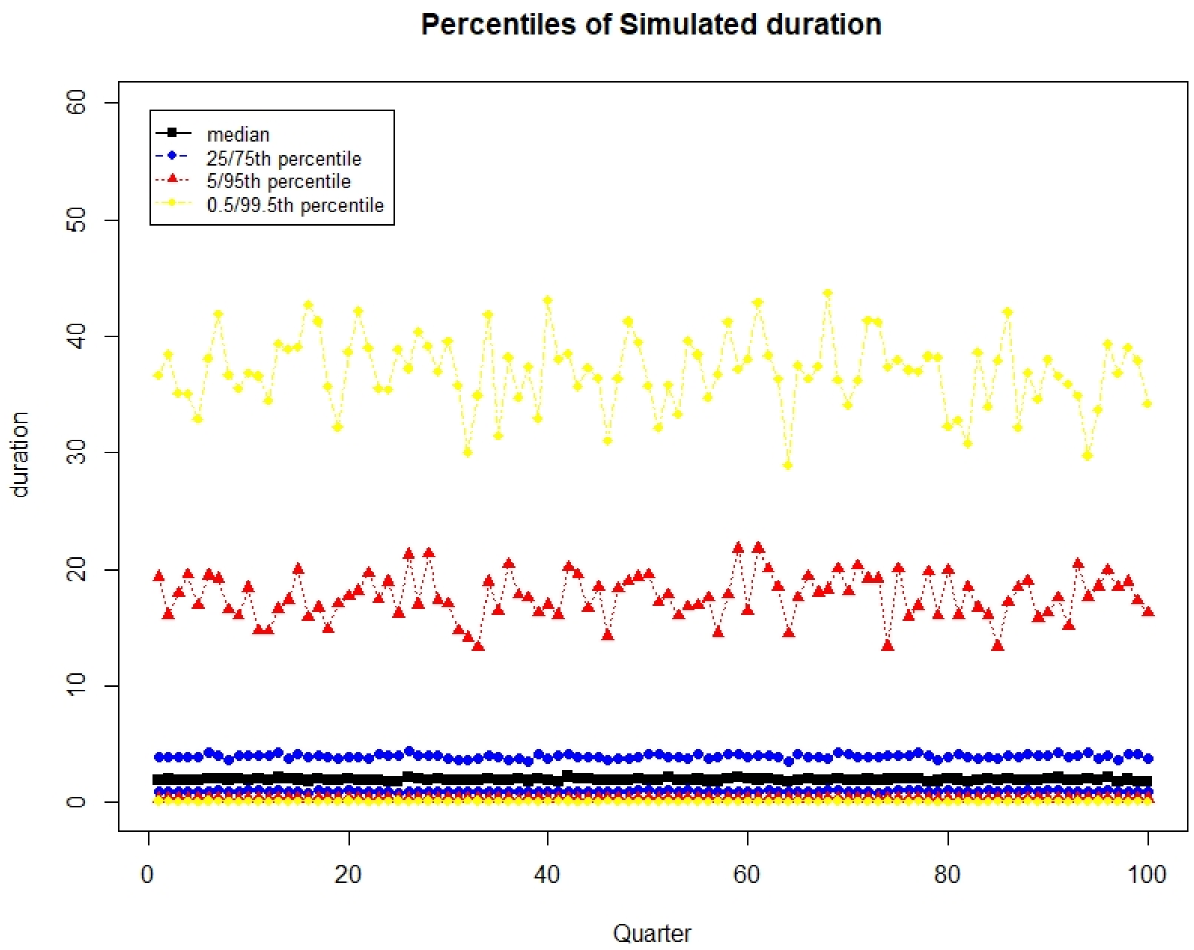


Figure B.2

Percentiles of Simulated Duration



Both short- and long-term pandemic/epidemic events are simulated by the PSG.

Figure B.3

Percentiles of Simulated Case Fatality Rate

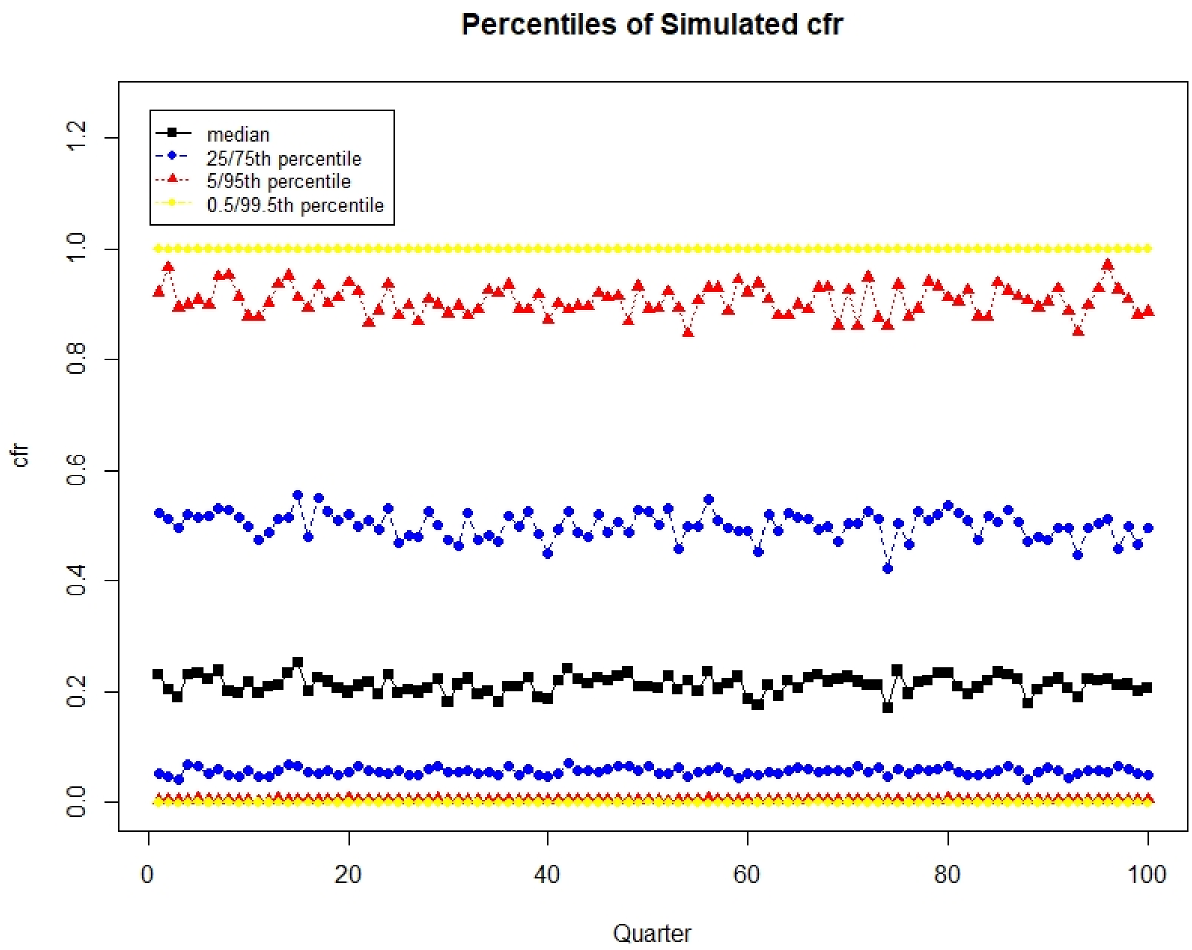
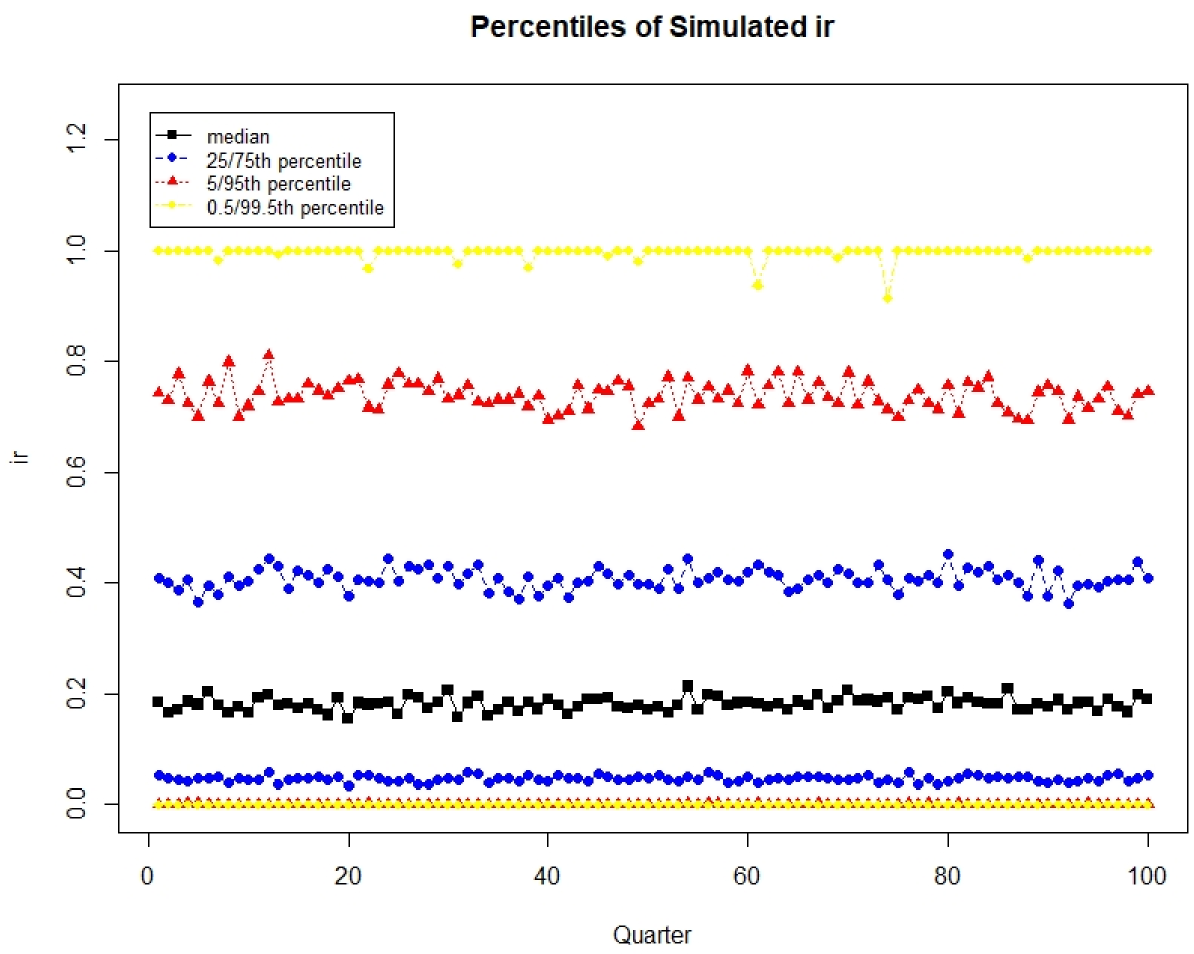


Figure B.4

Percentiles of Simulated Infection Rate



Figures B.5 to B.11 show the percentiles of the simulated economic factors by time period, without observing any abnormal patterns.

Figure B.5

Percentiles of Simulated GDP Growth Rate

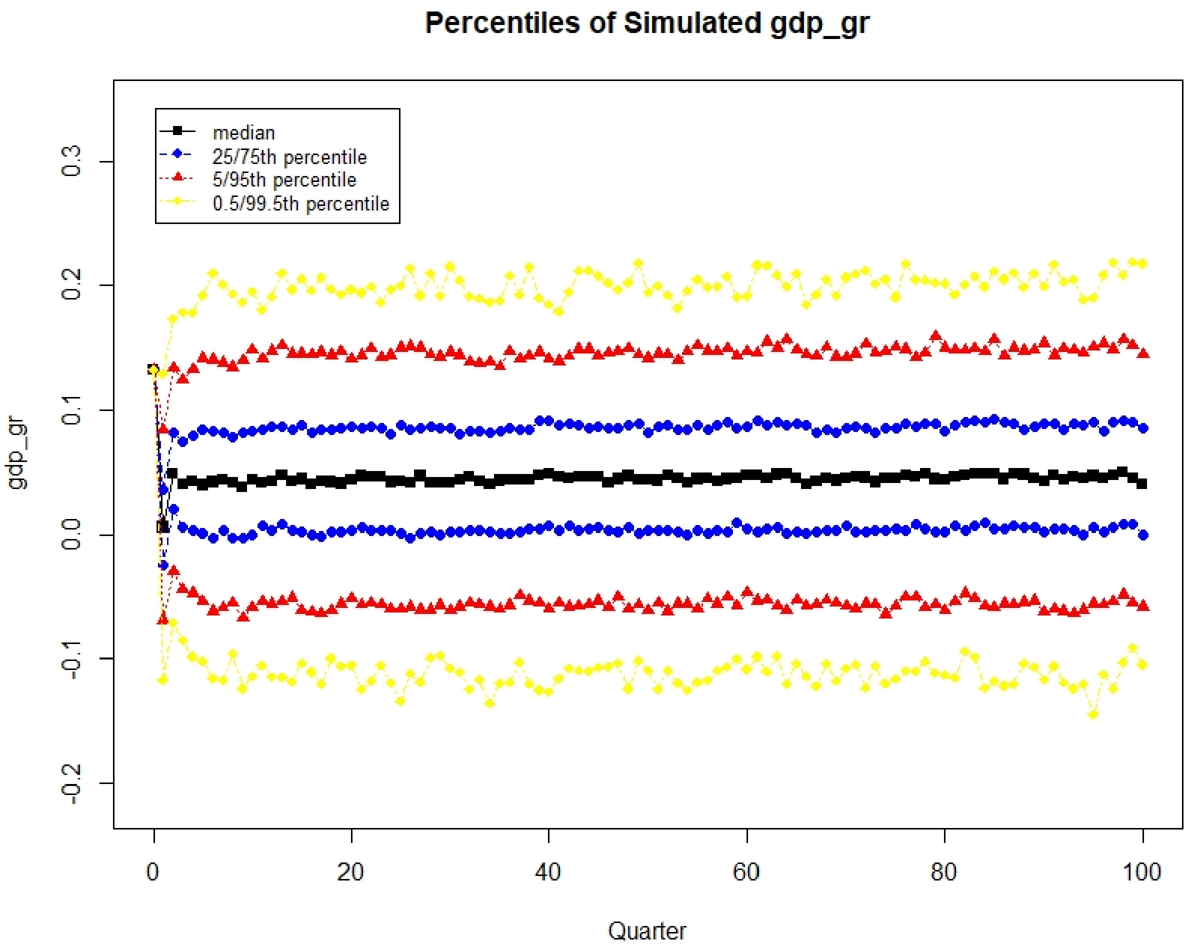


Figure B.6

Percentiles of Simulated Inflation Rate

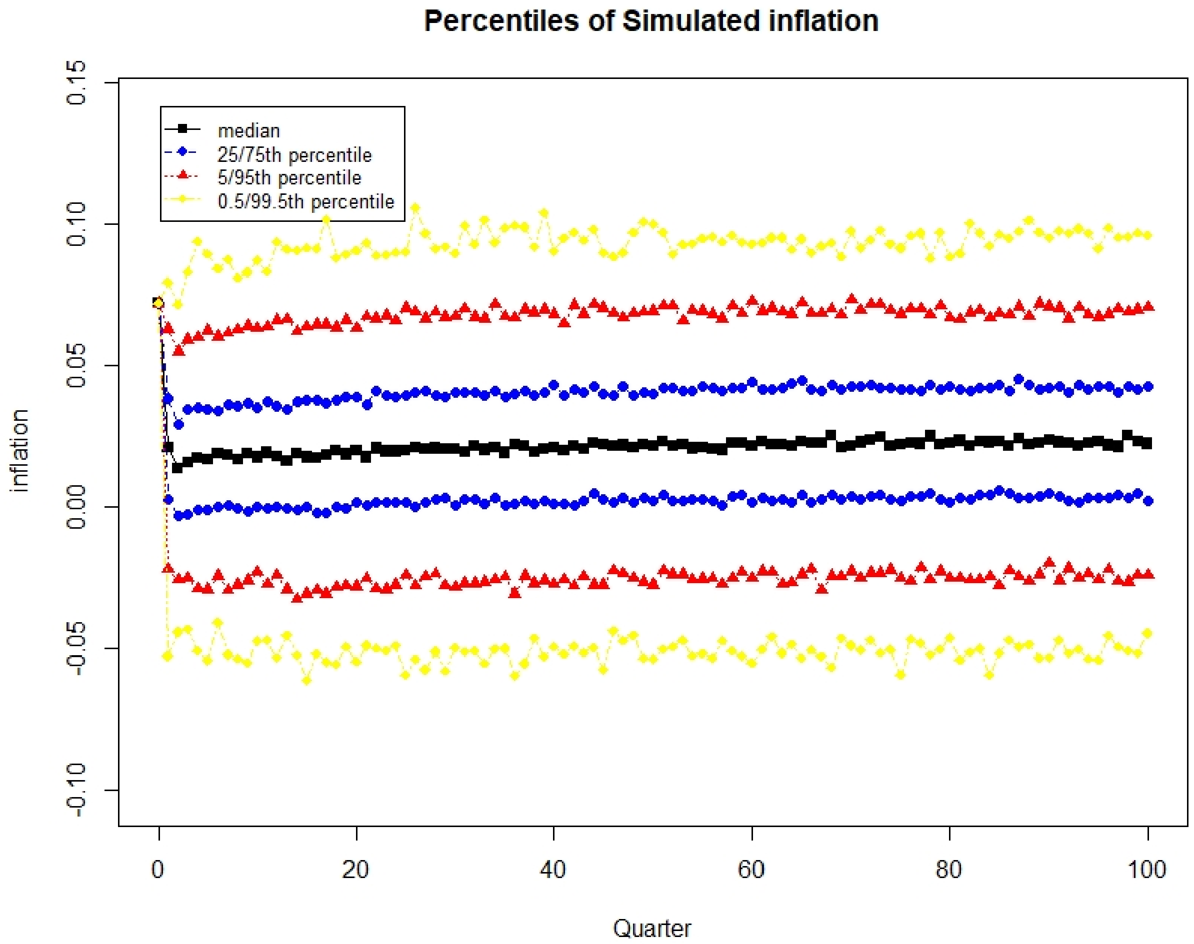


Figure B.7

Percentiles of Simulated Personal Consumption Growth Rate

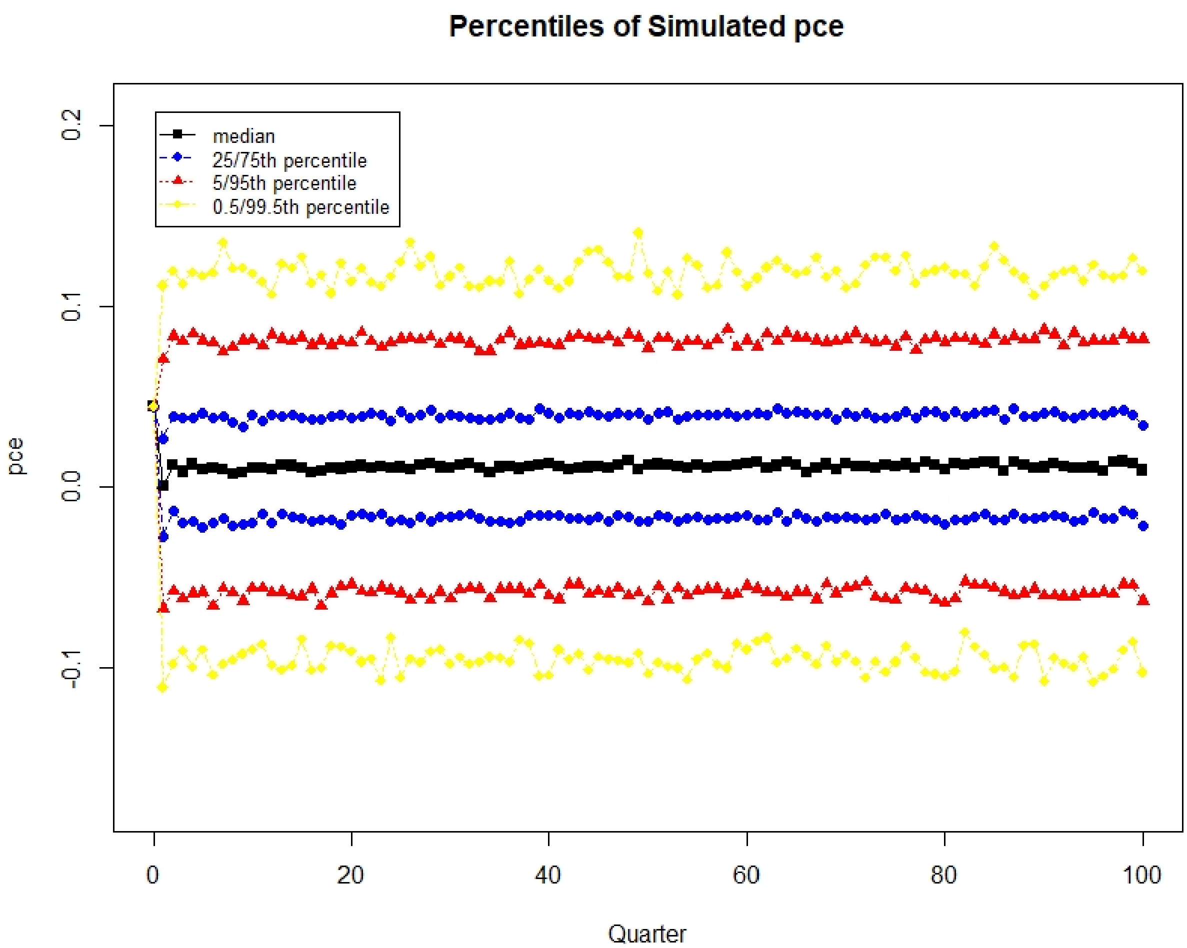


Figure B.8

Percentiles of Simulated Investment Growth Rate

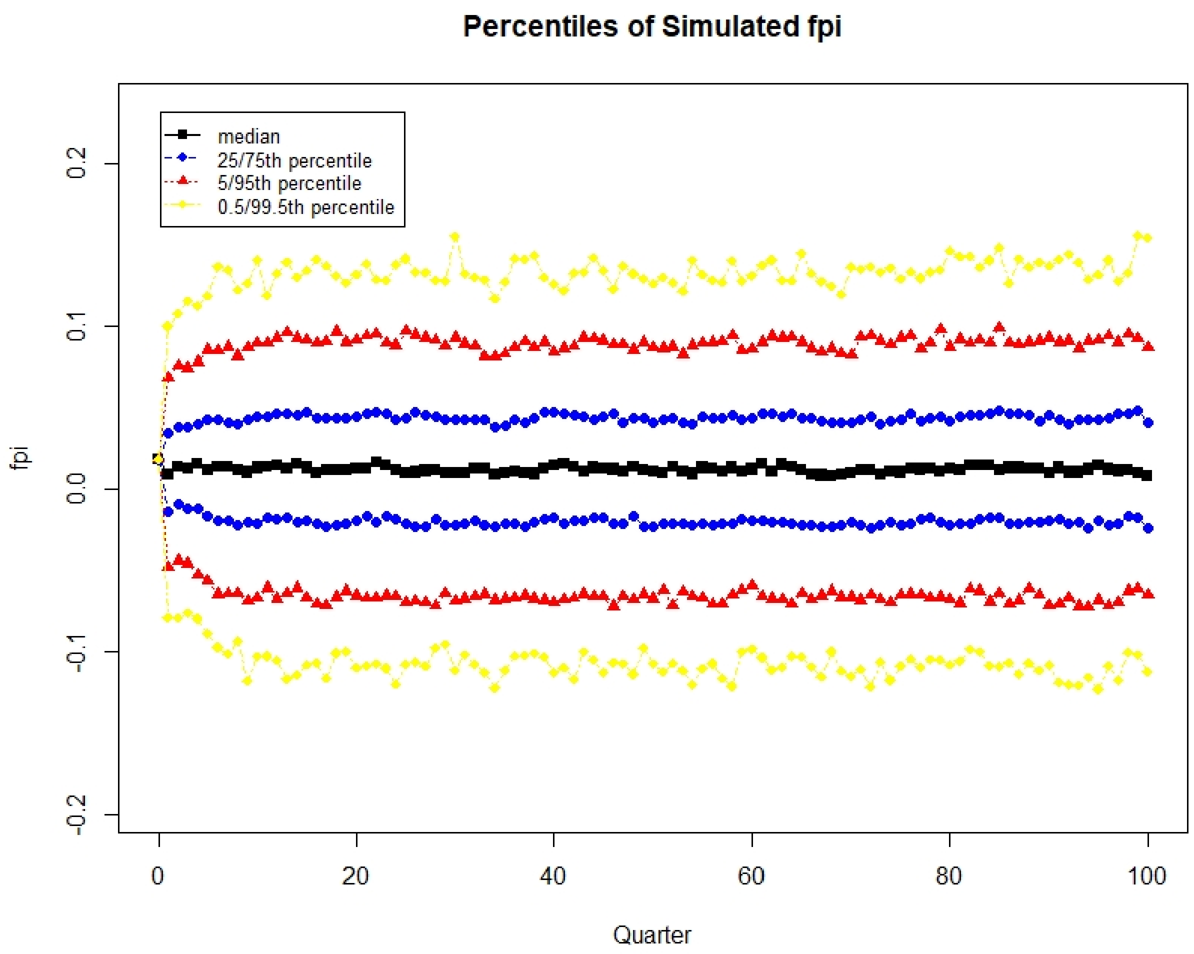


Figure B.9

Percentiles of Simulated Government Expense Growth Rate

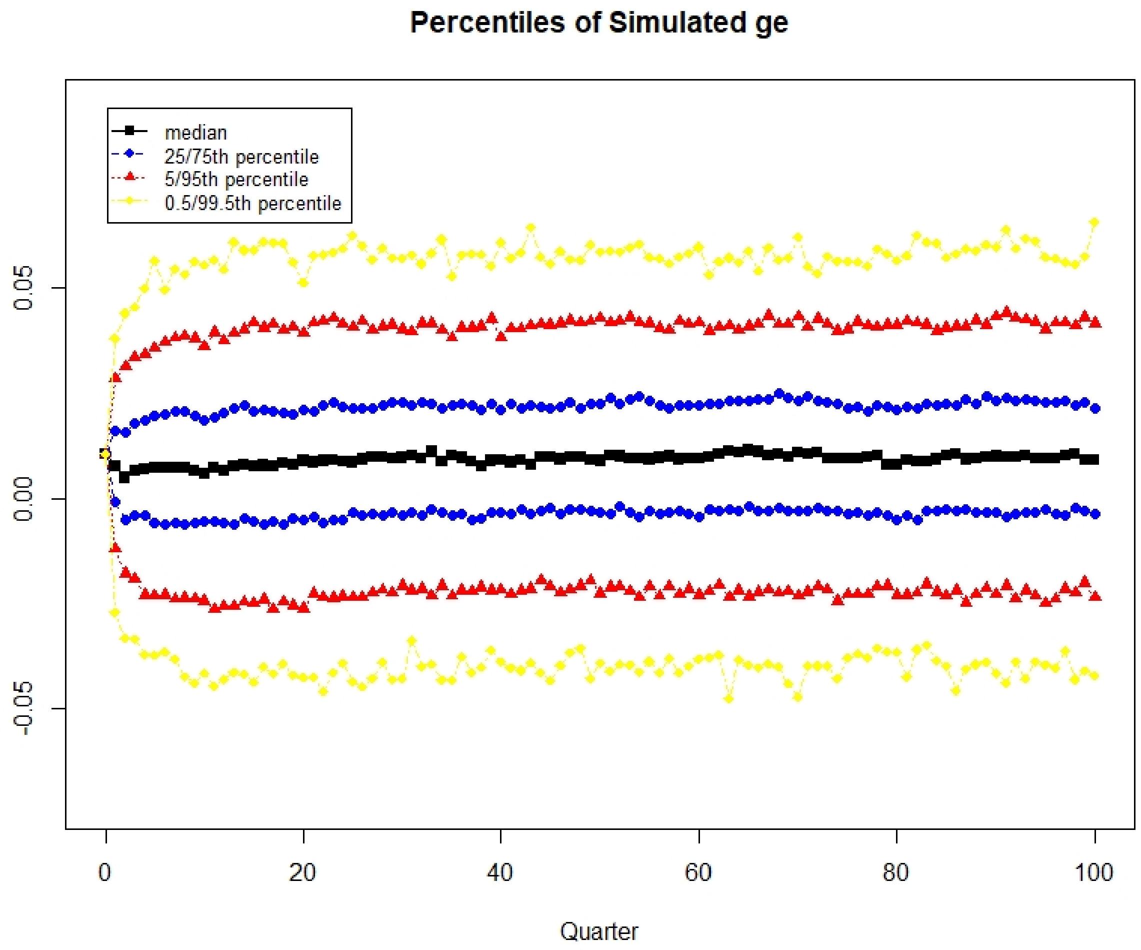


Figure B.10

Percentiles of Simulated Fed Fund Rate

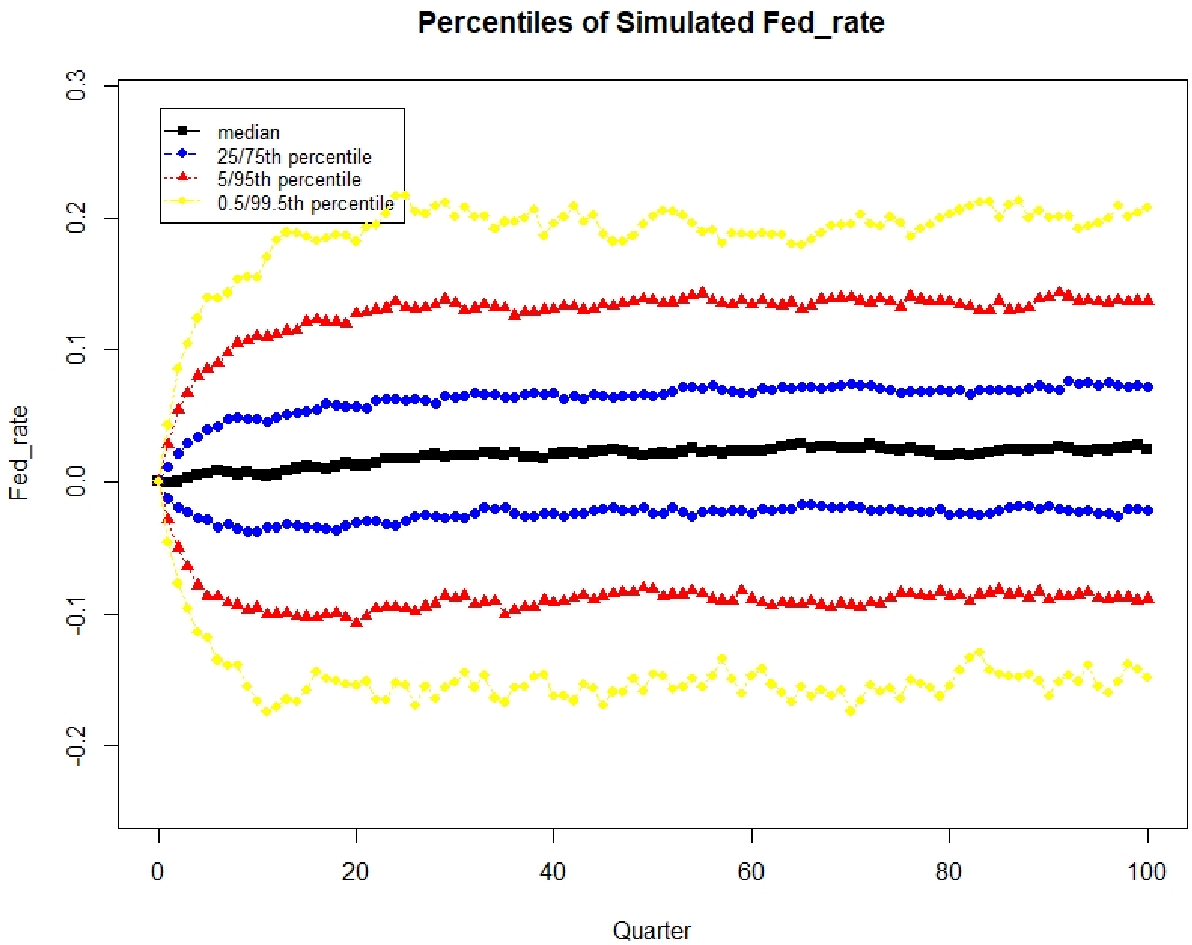
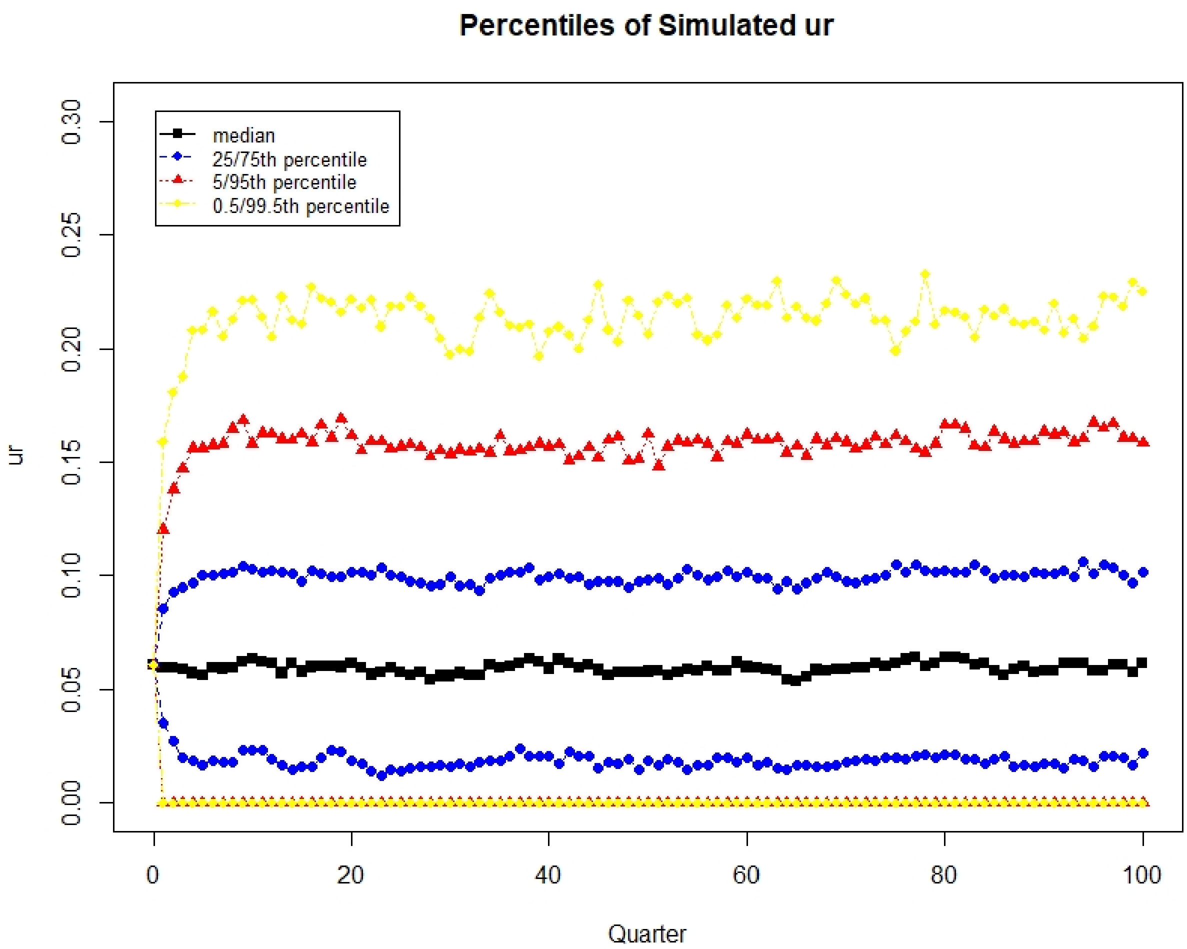


Figure B.11

Percentiles of Simulated Unemployment Rate



Figures B.12 to B.20 show the percentiles of the simulated capital market variables by time period, without observing any abnormal patterns.

Figure B.12

Percentiles of Simulated 1-Yr Treasury Bond Yield

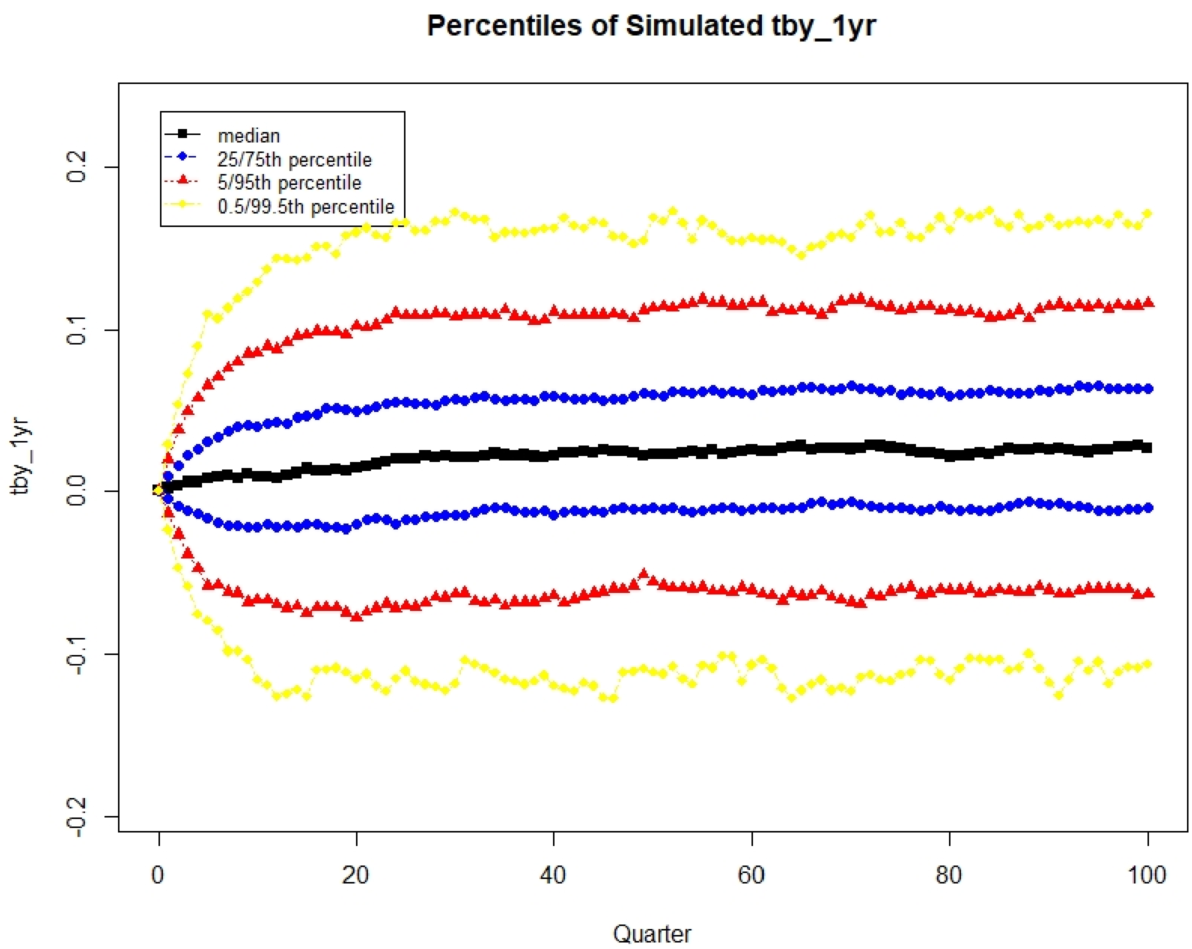


Figure B.13

Percentiles of Simulated 10-Yr Treasury Bond Yield

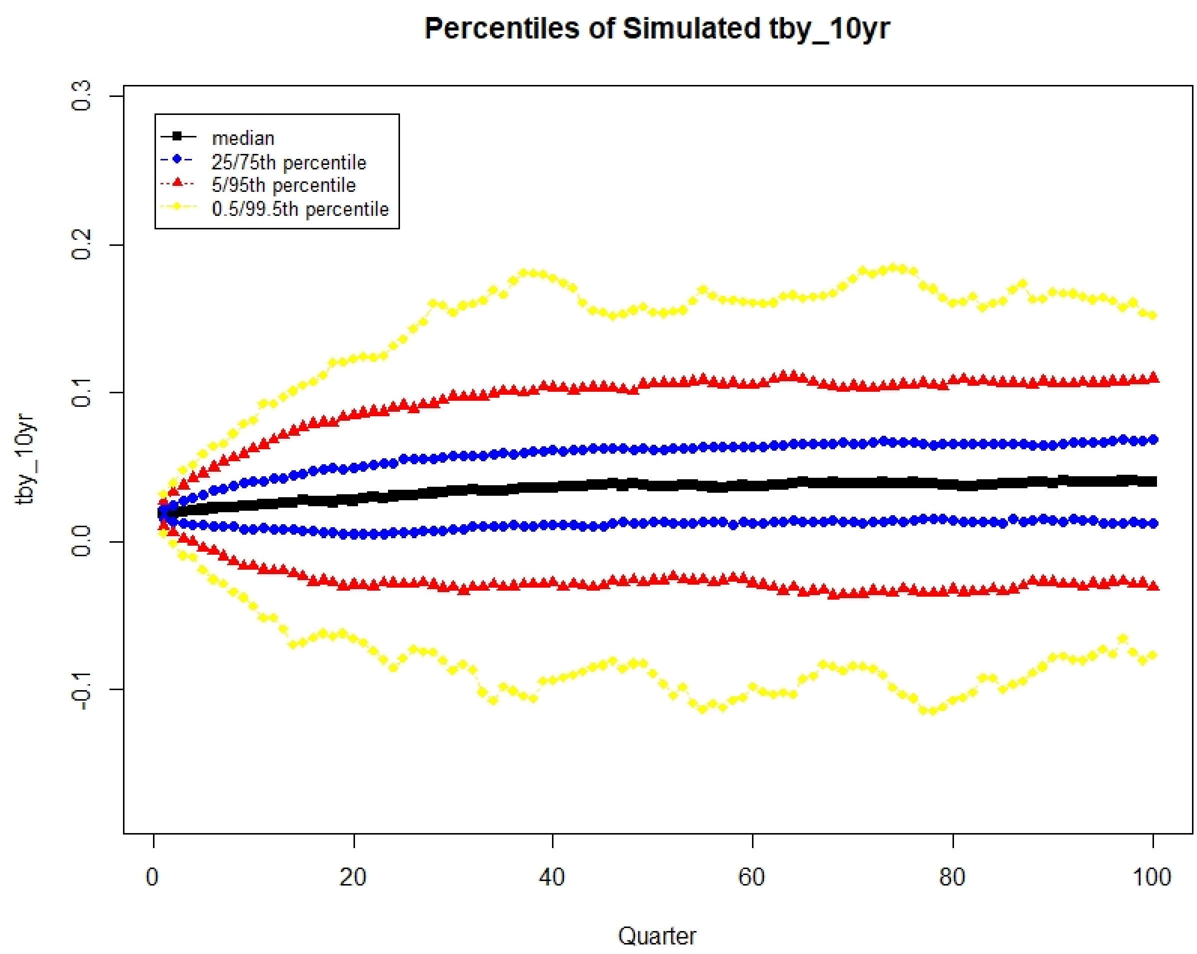


Figure B.14

Percentiles of Simulated Aaa-Rated Bond Credit Spread

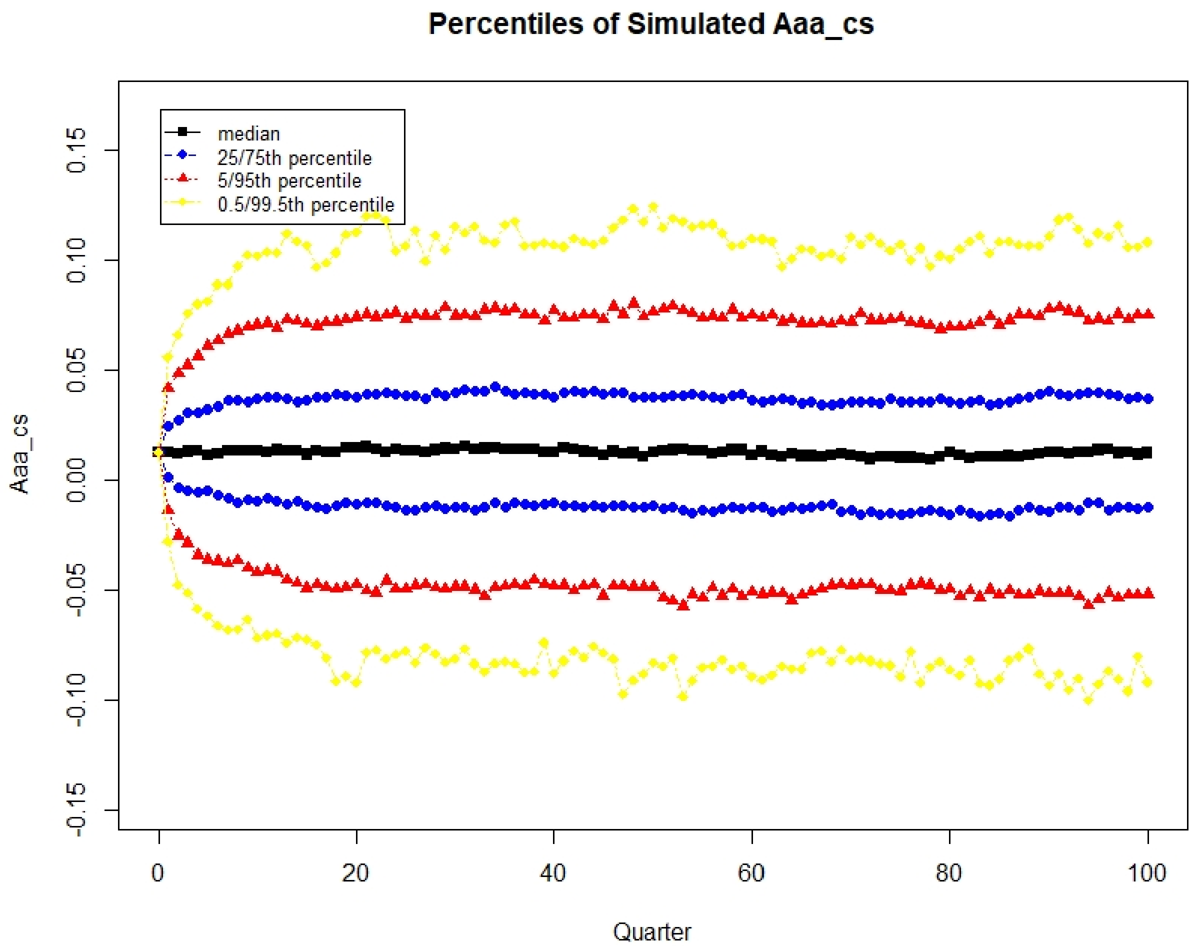


Figure B.15

Percentiles of Simulated Baa-Rated Bond Credit Spread

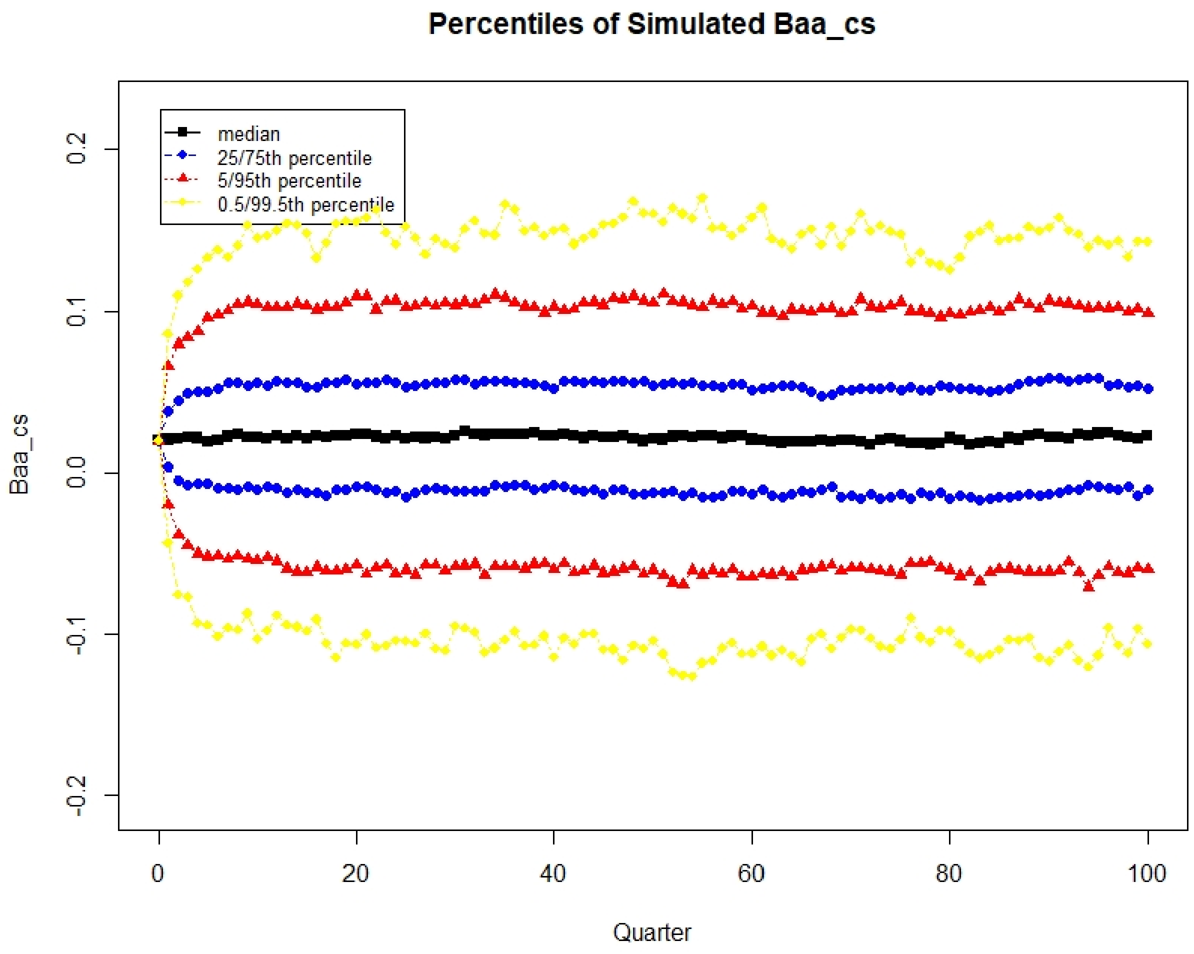


Figure B.16

Percentiles of Simulated BBB-Rated Bond Default Rate

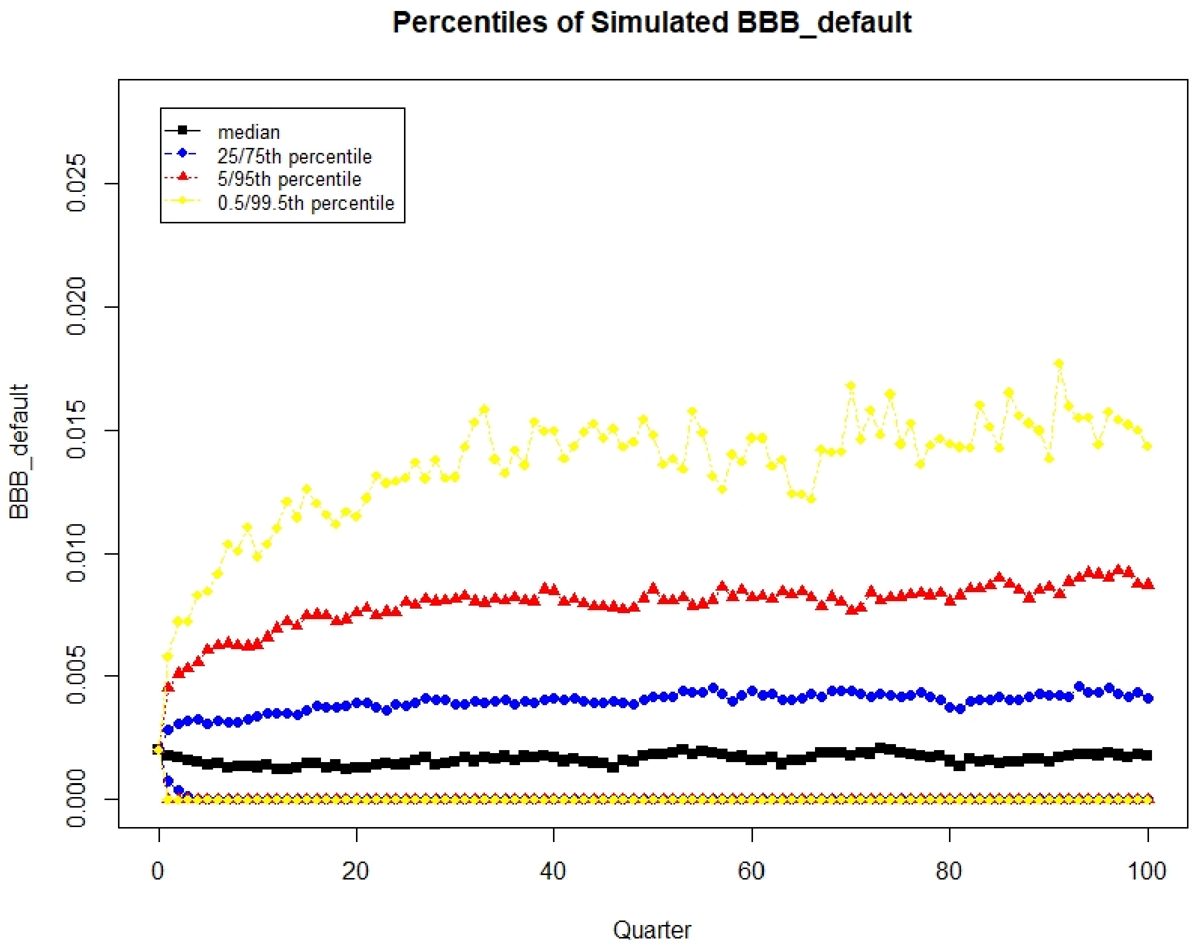


Figure B.17

Percentiles of Simulated S&P 500 Index Return

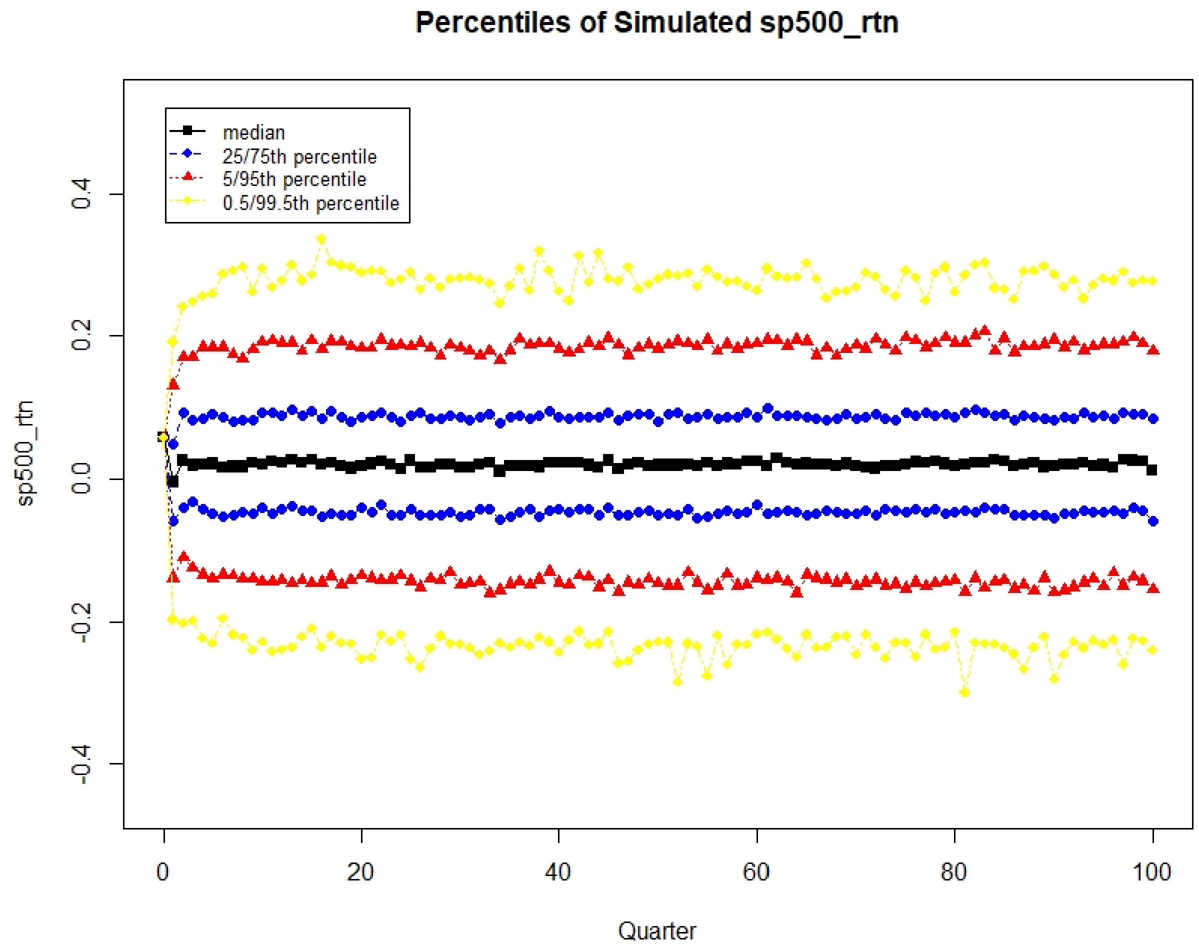

Percentiles of Simulated sp500_rtn
median
25/75th percentile
5/95th percentile
0.5/99.5th percentile
sp500_rtn
0.4
0.2
0.0
-0.2
-0.4
0
20
40
60
80
100
Quarter

Figure B.18

Percentiles of Simulated S&P 500 Dividend Yield

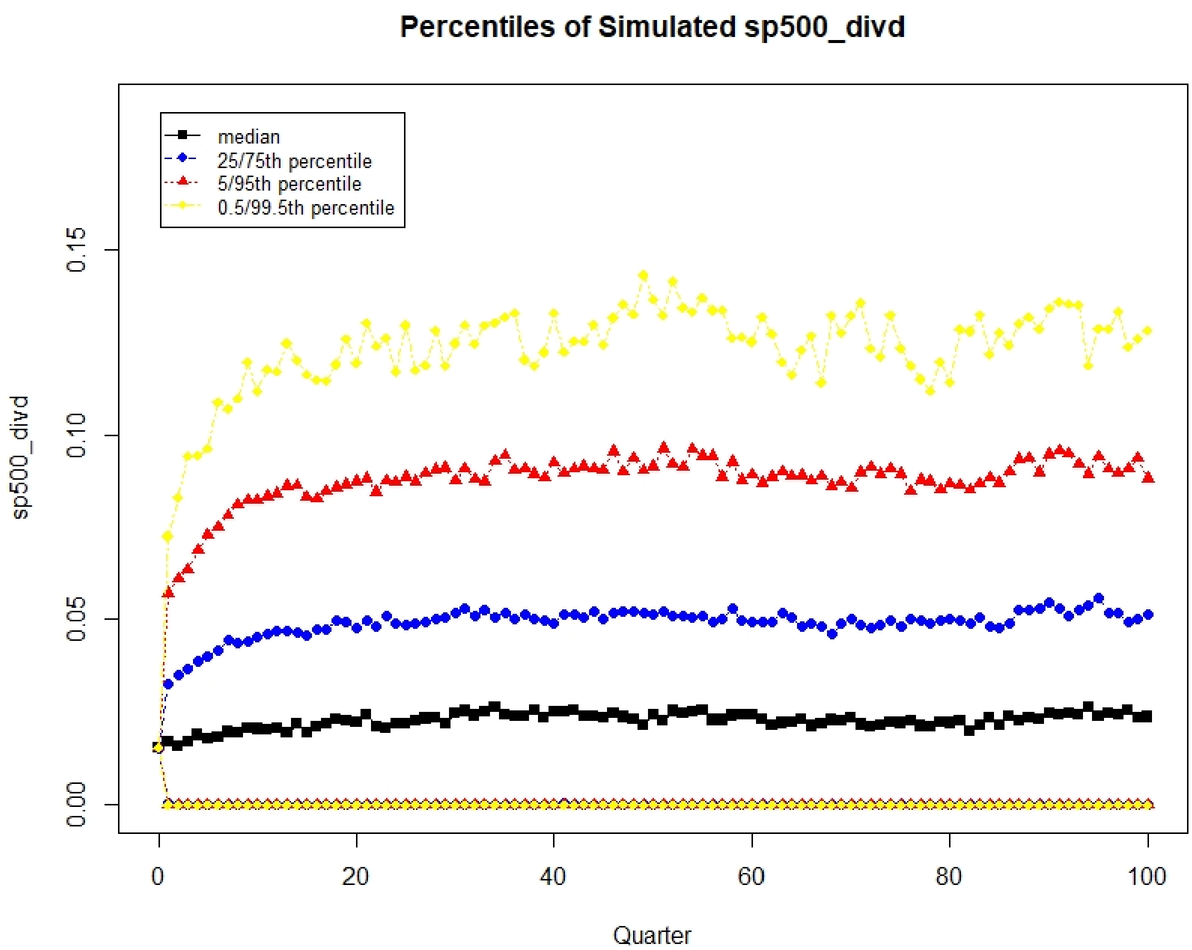


Figure B.19

Percentiles of Simulated Median House Price Growth Rate

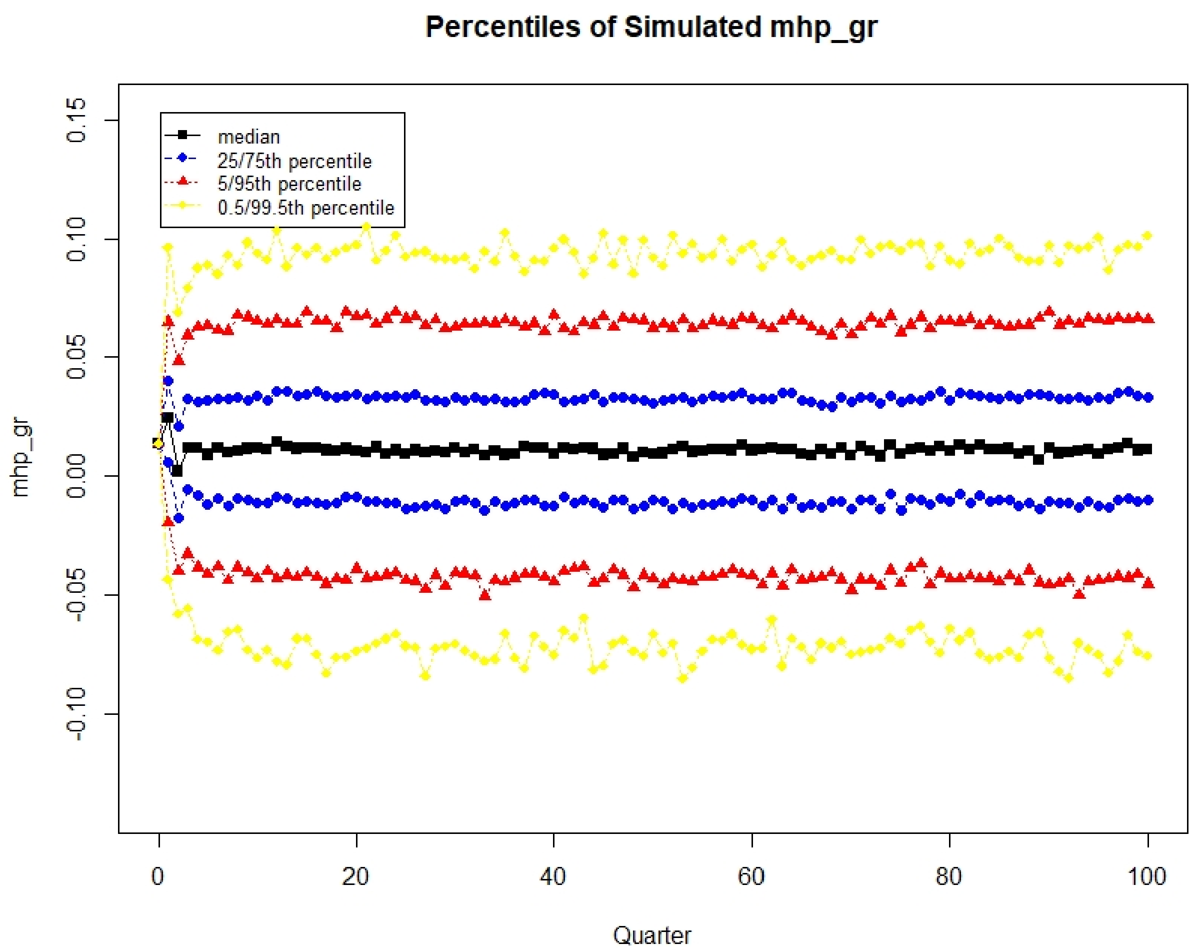
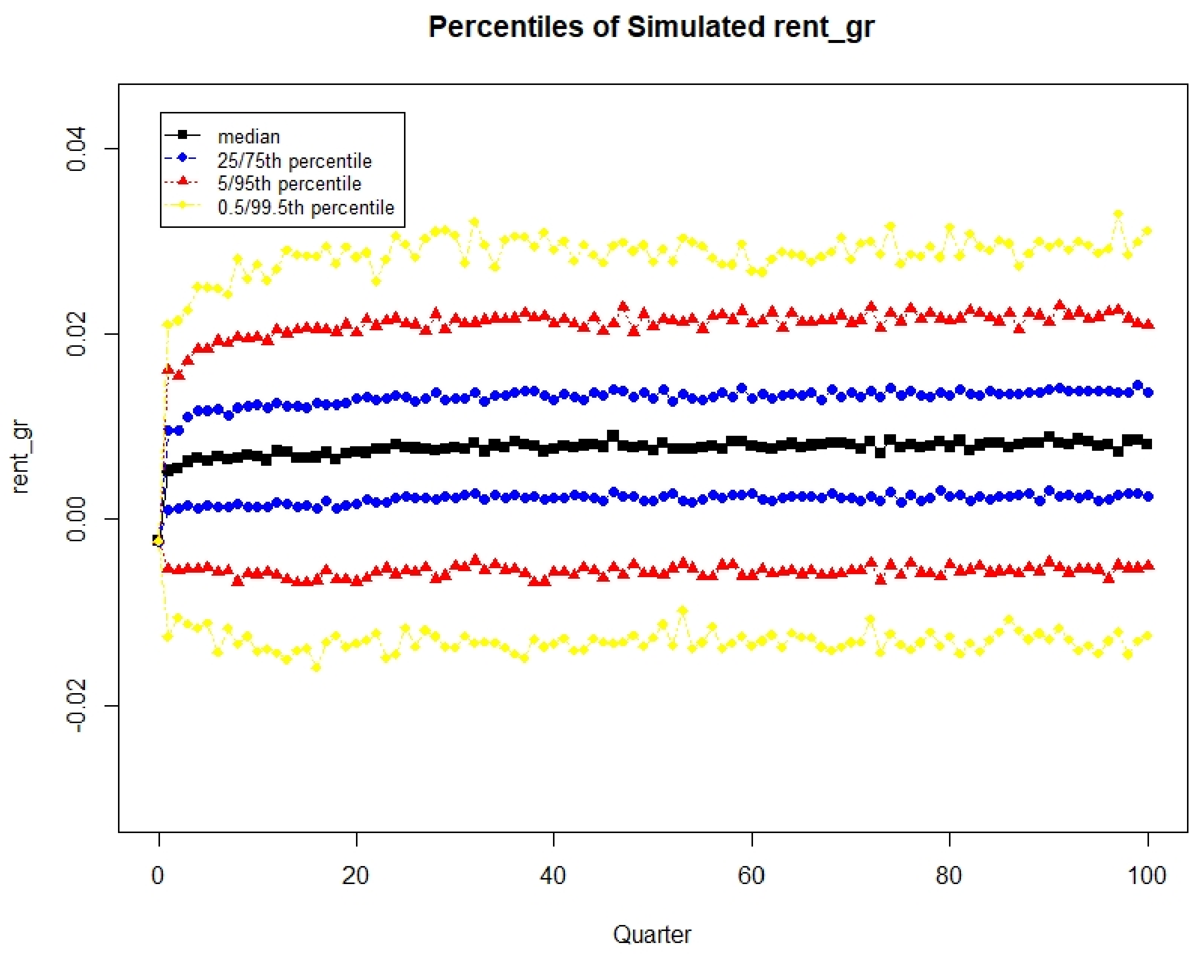


Figure B.20

Percentiles of Simulated Rent Growth Rate



Canadian Institute of Actuaries
360 Albert Street, Suite 1740
Ottawa, ON K1R 7X7
613-236-8196
head.office@cia-ica.ca
cia-ica.ca
seeingbeyondrisk.ca

The Canadian Institute of Actuaries (CIA) is the qualifying and governing body of the actuarial profession in Canada. We develop and uphold rigorous standards, share our risk management expertise, and advance actuarial science to improve lives in Canada and around the world. Our more than 6,000 members apply their knowledge of math, statistics, data analytics, and business in providing services and advice of the highest quality to help Canadian people and organizations face the future with confidence.