The Dissertation Committee for Vishwakant Malladi
certifies that this is the approved version of the following dissertation:

Dependence in Operations: Modeling and Applications

Committee:

Kumar Muthuraman, Co-Supervisor

Efstathios Tompaidis, Co-Supervisor

Guoming Lai

John Hasenbein

Dependence in Operations: Modeling and Applications

by

Vishwakant Malladi

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2018

Dedicated to my parents

Acknowledgments

I would firstly like to thank my family for supporting me throughout my academic career, especially my mother, Girija for showing me the joy in learning and my father, Lakshmana for showing me the joy in exploring. My professional life is largely built on the foundation laid by them. I would also like to thank my sisters Sampreeti and Sarvani, for tolerating my idiosyncrasies through the duration of my doctoral studies.

I am extremely thankful for the guidance my advisors, Dr. Stathis Tompaidis and Dr. Kumar Muthuraman, have given me along my path through graduate school. They have helped me become a better and a more rounded researcher. I hope to continue to collaborate with them as I begin my career as an academic. I extend a special thanks to Dr. Rafael Mendoza-Arriaga who has been a co-mentor during my time at the Red McCombs School of Business. I would also like to thank Dr. Guoming Lai and Dr. John Hasenbein who have graciously agreed to serve on my dissertation committee and have given me great feedback on my work.

Certainly, I cannot but thank my friends at the McCombs School of Business. Changseung Yoo and Abhishek Roy, your company through the years was a major

reason for my duration here being enjoyable and pleasant. I would also be remiss if I omit the names of my fellow graduate students at the McCombs school of business, Chinmoy and all the other students who have shared their experiences and their knowledge. Our regular interactions were great and served as short respites from the stressful life of doctoral candidacy,.

Dependence in Operations: Modeling and Applications

Vishwakant Malladi, Ph.D.
The University of Texas at Austin, 2018

Co-Supervisors: Kumar Muthuraman
Efstathios Tompaidis

Operations Management is replete with examples where dependence in risk has large financial consequences. Modeling dependence in multi-dimensional systems is a hard problem. Traditional models o dependence suffer from the curse of dimensionality. This dissertation introduces a framework, called subordinated Markov chains, to model dependence in risk and demonstrates the efficacy of the framework for two applications in operations management. First, the mathematical validity of the framework is established along with simulation algorithms for the framework. Next, the framework is used to model dependent outages of electricity power plants and to model the joint disruptions in the facility location problem.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Risk Modeling in Operations Management

Modern businesses are susceptible to a variety of risks that adversely influence their operations. These risks manifest themselves as, for example, the interruption of production schedules, inefficiencies in the distribution network, decreased reliability of the production units, etc. They can vary from weather related occurrences, like hurricanes, geo-political factors like social upheaval, economic downturns, or other sources of disruptions. The risks can have a large financial impact on a firm's operations. For example, the cost to businesses due to damage from hurricane Harvey in 2017 is estimated at $125 billion. Thus, it is understandable that firms spend a considerable portion of their resources planning, preparing, and protecting against risks. In this context, the modeling of risk is essential to modern business decision making.

In this dissertation we address the modeling of dependence in risk in an Operations Management context. We interpret modeling risk as the building of a mathematical framework for the joint probabilities of the different states and for the

dynamics of the operational state of the components in the system. More specifically, we introduce a framework to model the dependence in the state of the system due to common risk factors.

Certain risks that affect a firm's operations impact multiple units or components within the firm. For example, a single hurricane might disrupt multiple warehouses of a firm that operate within the particular geography affected by the hurricane. Since multiple warehouses are disrupted, the firm faces exorbitant costs towards servicing and maintenance. Thus, such events, while being rare, have a disproportionately large financial impact on the firm due to their tendency to affect multiple components of the firm simultaneously. Because of their high impact, a characteristic that needs to be addressed while modeling risk is this intrinsic dependence between the risk affecting multiple components. Accurate modeling of the risk to these aspects of a firms business can help firms better protect themselves against events with high impact.

While modeling dependence is an important aspect of risk modeling, it is nonetheless a difficult problem, especially when dealing with systems with a high number of components. Consider a simple example with 10 components with each component in the "ON" or the "OFF" state. This system can exist in 2^{10} or 1024 different possible states. Under the assumption of independent risk factors, the probability distribution can be reproduced with the 10 marginal disruption probabilities of the 10 components. While this is computationally easy, it explicitly assumes independence. A possible model for the state of this system that considers dependence would be specifying the 1024 steady-state probabilities explicitly. As the number of

components increases, the number of probabilities that need to be specified increases exponentially.

A comprehensive approach would model not only the probability of the various disruption scenarios but also the dynamics of the system. For example, consider a system with two components, each with two possible states ("ON" and "OFF"). We want to build a framework that can model, along with steady state probabilities, aspects like; How often do both the components go down together? How long do both the components stay in the "OFF" state? etc. Modeling the dynamics of the system would be even more difficult than modeling just the steady state probabilities since need to consider the entire state space and then model the transition dynamics from each state to every other state.

The estimation of parameters is also a critical aspect in risk modeling. Since large scale disruptions are relatively rare, limited data is available to estimate the parameters. Thus, models that use a large number of parameters are infeasible. Models of risk that are easy to estimate and work with few parameters are needed to use in real life applications.

We focus on building a parsimonious model for the dynamics of the state of a system with a large number of components that allows for dependence. To do this, we use subordinated Markov chains. Subordinated Markov chains build on traditional Markov Chains by introducing a Lèvy process to the system which induces dependence between the previously independent Markov chains. A single subordinator can induce dependence in multiple Markov Chains. Thus, dependence can be introduced in the multiple components in a parsimonious fashion. In addition to

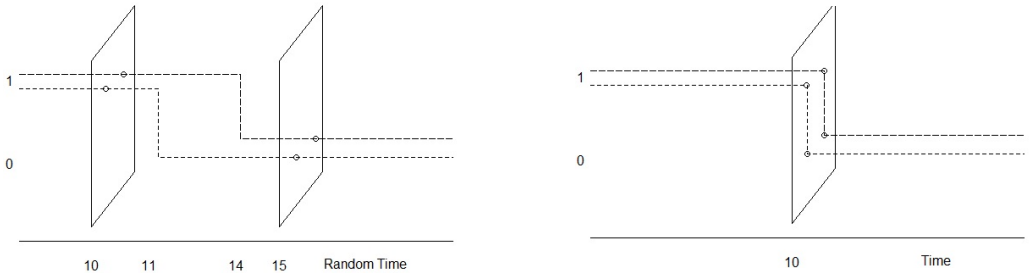
being parsimonious, subordinated Markov chains are easy to simulate and estimate, making them appropriate in a multitude of applications in Operations Management

1.2 Dependent Risk Modeling using Subordinated Markov Chains

We provide a brief intuition into Subordinated Markov Chains below. A rigorous construction of Subordinated Markov Chains is available in Chapter 2

A subordinated Markov chain is a Markov chain imposed on a Lèvy subordinator. Subordinated Markov Chains are constructed by first starting with independent Markov Chains in virtual time, and then translating the virtual time to actual time using a subordinator. A subordinator can be thought of as a virtual time clock that jumps at random times by a random positive amount. That is, the state evolution of the Markov chain in actual time looks like the state evolution in virtual time with random sections corresponding to the jumps wiped out and pasted together. The wiped out sections are dictated by the jumps in the virtual clock. The pasted together processes can transition from one state to another at the same time, which is not possible if the two Markov Chains were independent.

To better illustrate the mechanism, we present a simple graphical example for subordinated Markov chains.



(a) Independence in T_t (b) Dependence in t

Figure 1.1: Dependence through Subordination

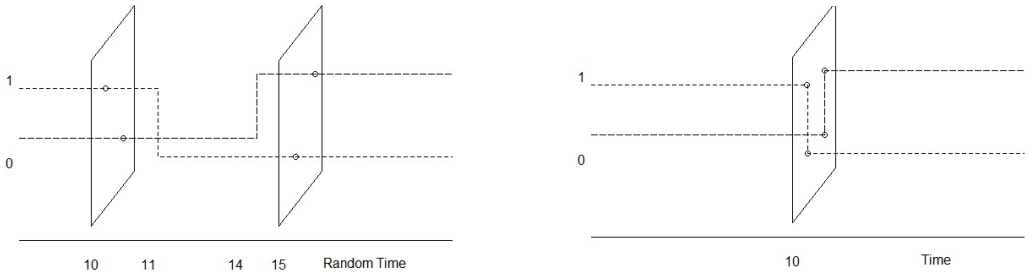
Figure (1.1a) represents the evolution of two independent Markov chains with two states 1 and 0. "1" represents the "ON" state and "0" the "OFF" state. The first Markov Chain, represented by the short dashed line, transitions from "1" to "0" at time 11. The second Markov Chain, represented by the long dashed line, transitions at time 14. We also consider a virtual clock (or subordinator) that jumps at time 10 by 5 units. The vertical planes represent the interval associated with the jump of the subordinator. In Figure (1.1b), we erase the interval where subordinator jumps which in this case (10, 15) – to construct the new processes. The two new processes transition at exactly time 10, thereby inducing dependence. Or, put another way, the two Markov chains are disrupted at the exact same time. To induce this dependence, we only need to define the interval (10,15) that is to be erased in the virtual time.

This construction can be extended to more than two independent continuous-time Markov chains (CTMCs) and a single subordinator. A single subordinator can be used to induce dependence in multiple CTMCs. Multiple subordinators can be used on the same set of CTMCs to model complex dependence among a group of CTMCs.

Discontinuities in the subordinator can be thought of as shocks to the system. They are analogous to external factors like weather-related events, which results in disruptions of the individual components. Multiple subordinators can be associated with individual risk factors like weather, geography etc. Thus, subordinators can not only model dependence, but also help us identify the causes of dependence among the multiple components.

A simple subordinator can be described by two parameters; the arrival rate of these random jumps and the size of the jumps. Hence it is possible to construct a joint distribution that allows for dependence, with only a few additional parameters – those of the subordinator.

While this example illustrates the parsimony of the subordinated Markov chain framework in introducing dependence in Markov chains, it is not applicable as-is in an Operations Management context because it treats disruptions and recoveries symmetrically. Consider the graphical example provided below:



(a) Independence in T_t (b) Contrary dependence in t

Figure 1.2: Dependence through Subordination

Continuing with the same setup as above, we present two independent Markov

chains which transition in the opposite directions at time 14 and 11. The short dashed Markov chain goes from state "ON" to "OFF" while the long dashed Markov chain goes from "OFF" to "ON". Time-erasure in this case results in simultaneous transitions as shown in Figure (1.2b). Associating a transition from "0" $\rightarrow$ "1" with a recovery and a transition from "1" $\rightarrow$ "0" with a disruption, we see that, at time 10 in Figure (1.2b), one component is disrupted while the other recovers at the exact same time. This is counter-intuitive to real-life applications where shocks tend to cause disruptions and not recoveries. For instance, it is not possible for a hurricane to cause the disruption of a warehouse and the recovery of another disrupted warehouse. Hence, a simple subordinated Markov chain model is not applicable in most operations management problems. To ensure the direction of the dependence is also accounted for, we introduce partially subordinated Markov chains where dependence in disruptions and recoveries can be independently modeled.

In partially subordinated Markov chains, a subordinator is uniquely associated with transitions of a particular direction. For example, a subordinator is only associated with the disruptions of the components. The recoveries might be independent, or might have a separate subordinator. This lets one control the direction of dependence and ensure that a risk factor has a uniform effect across the multiple components. Partially subordinated Markov chains provide a richer class of CTMCs with dependence and can be used in real life OM applications.

The second chapter of this dissertation focuses on the mathematical description of dependence in Operations Management. It provides a brief introduction to continuous-time Markov chains before developing the theory behind subordinated

Markov chains and partially subordinated Markov chains. In it, we demonstrate the mathematical validity of the construction of a partially subordinated Markov Chain. We also present various algorithms to simulate subordinated Markov chains.

1.3 Applications of Dependent Risk Modeling

In addition to providing a mathematical framework for dependent risk modeling, this dissertation also covers two applications of subordinated Markov chains in Operations Management. The first application studies the impact of dependence of outages in electricity power plants. The second is related to the reliable facility location problem. Both these applications have been studied before in the literature with rather simplistic assumptions like independence in disruptions. We see that in both these problems, dependence is a major aspect that has not been addressed and one that has operational implications. We use our framework, the partially subordinated Markov chains, to estimate and analyze the impact of dependence in both these applications.

In the first application, covered in chapter 3, we examine the impact of dependence in outages in electricity power generating plants. We calibrate our model for power plants in the Electric Reliability Council of Texas and the Western Electricity Coordinating Council regions using a unique dataset that we obtained from the North American Electric Reliability Corporation. The subordinated Markov chain framework allows for commonality in outages to be driven by common risk factors. We find strong evidence of dependence in power plant outages both system-wide, and based on the input fuel of the plants. We illustrate the consequences of out-

age dependence on the reliability of electricity supply for both regions and evaluate policy implications; e.g., the impact of building additional capacity.

For the second application, covered in chapter 4, we study the facility location problem with joint disruptions. We use subordinated Markov chains to model the probability of joint disruptions. This approach offers a more realistic and easy-to-simulate disruption distribution using fewer parameters. We also propose algorithms to calibrate the subordinated Markov chain model and optimize for the facility location choices under dependent disruptions. Calibration of the disruption data from the National Oceanic and Atmospheric Administration shows a dependence structure that is defined by geographical proximity. We also show the efficacy of the optimization algorithm for the standard benchmark case with 49 locations, corresponding to the states in continental United States.

We conclude by discussing a few other directions where subordinated Markov chains can be used to model and understand the impact of dependence in Operations Management applications. We also discuss some areas where the framework itself can be extended to model more complex versions of dependence structures.

Chapter 2
Subordinated Markov Chains

2.1 Overview

We induce dependence in continuous time Markov chains using a procedure called subordination. The resulting processes, called subordinated Markov chains, exhibit, in addition to dependence, certain properties that make them an attractive option to model dependence. In this chapter, we present the methodology to model dependent continuous-time Markov chains.

We first provide a brief overview of simple continuous-time Markov chains and systems with multiple independent continuous-time Markov chains and then show how dependence can be induced using subordinated Markov chains. We then extend the theory of subordinated Markov chains to construct partially subordinated continuous-time Markov chains. Partially subordinated continuous-time Markov chains allow a greater degree of control on dependence and thus are more suitable for Operations Management applications. Finally we present algorithms to simulate

subordinated and partially subordinated continuous-time Markov chains. [1]

2.2 Introduction

Subordination of continuous stochastic processes is an efficient way to model stochastic processes with jumps and discontinuities and induce dependence. It is a common technique in finance to model processes with shocks. Some examples of the usage of subordination in literature include Madan et al. (1998) to model the variance gamma process, Barndorff-Nielsen (1997) to model the normal inverse Gaussian process, Lim et al. (2012) and Mendoza-Arriaga et al. (2014) for interest rate modeling, Carr et al. (2002) for the CGMY process, Mendoza-Arriaga et al. (2010) for credit-equity modeling, Eberlein and Prause (2002) for the generalized hyperbolic process. Sun et al. (2017) is an example where dependence in default times is modeled using subordinators for credit risk applications.

Classical subordination, as used in finance, is not applicable as-is to Operations Management. Consider for example the variance-gamma process used to model stock prices in Madan et al. (1998), where subordination models the jumps in stock prices. As with stock prices, jumps induced via subordination can be positive or negative, making subordination suitable for such applications.

However, the impact of shocks is different in Operations Management. For example, shocks like hurricanes only have a negative impact in the form of shutdowns, production delays, etc. Classical subordination induces both positive and negative

[1] This chapter is based on the material covered in the paper *Modeling Dependent Outages in Electricity Generators* written with Dr. Stathis Tompaidis and Dr. Rafael Mendoza-Arriaga.

behavior in the stochastic processes and hence is not applicable in Operations Management applications. We introduce partially subordinated Markov chains to adapt the classical subordination framework to applications in Operations Management.

2.3 Continuous Time Markov Chains

In this section, we briefly introduce continuous-time Markov Chains. For a more thorough exposition, see Wolff (1989).

A continuous-time Markov Chain (CTMC), $X(t)$, is a stochastic process with a finite or countable state space $\mathcal{S}$. $X(t)$ is a CTMC if and only if, for any $s, t \geq 0$ and any $i, j \in \mathcal{S}$, the transition probabilities satisfy:

$$P(X(s+t) = i | X(u); u \leq s) = P(X(s+t) = i | X(s)). \tag{2.1}$$

This is called the Markov property and the resulting process a continuous-time Markov chain (CTMC). Intuitively, a CTMC is a process whose evolution can be described with only information of the current state of the process. CTMCs are a popular mechanism to model the state evolution of various real life processes in Operations Management. For example, CTMCs are widely used in queuing, population analysis, disruption analysis, etc.

A subclass of CTMCs is CTMCs with stationary probabilities, also called time-homogeneous CTMCs. Time-homogeneous CTMCs have the additional characteristic that the transition probabilities over a given duration of time are stationary over time. i.e., $\forall\ s, t \geq 0$ and $\forall\ i, j \in \mathcal{S}$

$$P(X(s+t) = i | X(s) = j) = P(X(t) = i | X(0) = j). \tag{2.2}$$

In this thesis, we only consider CTMCs with stationary probabilities. The transition probabilities for time-homogeneous CTMCs are functions of the length of the time interval alone. Denote transition probabilities from state i to state j in time t by, $P_{ij}(t) := P(X(t) = i|X(0) = j)$,$\forall\ i, j \in \mathcal{S}$. We use matrix $\mathbf{P}(t)$ to denote all possible transition probabilities, where $P_{ij}(t)$ corresponds to the element in row i and column j in matrix $\mathbf{P}(t)$.

The evolution of a time-homogeneous CTMC can be described by a single generator matrix $\mathbf{Q}$ with the following properties:

(i) all row sums of the matrix $\mathbf{Q}$ equal zero,

(ii) all off-diagonal elements of $\mathbf{Q}$ are non-negative, and

(iii) element $q_{ij}, i \neq j,\ i, j \in \mathcal{S}$ in the matrix $\mathbf{Q}$ corresponds to the transition intensity of the CTMC from state i to state j.

Each row and column of $\mathbf{Q}$ correspond to a state in the state space of the CTMC. $\mathbf{Q}$ is the instantaneous rate of change of the transition probabilities of the CTMC and defines the transition probabilities of the CTMC. Given P_0, a vector of the probabilities of the state of the CTMC at time 0, the infinitesimal generator can be used to evaluate the transition probabilities of the CTMC as:

$$\mathbf{P}(t) = P_0^T e^{\mathbf{Q}t}. \tag{2.3}$$

A time-homogeneous CTMC, has a steady-state probability vector which is the limiting distribution of the state of the CTMC; i.e., $\mathbf{P}(t)$ as $t \to \infty$. The steady-

state distribution vector, π of a CTMC with generator $\mathbf{Q}$ is given by:

$$\begin{aligned} \sum_i \pi_i &= 1, \\ \pi^T \mathbf{Q} &= 0. \end{aligned} \tag{2.4}$$

Example

Consider a CTMC with two states; "OFF" and "ON". If the intensity of transition from "OFF"$\rightarrow$"ON" is α and the intensity of transition from "ON"$\rightarrow$"OFF" is β, then:

- The infinitesimal generator of the CTMC is:

$$\mathbf{Q} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}, \tag{2.5}$$

- The transition probability at time t is:

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = \begin{pmatrix} 1 - \frac{\alpha}{\alpha+\beta} f(t) & \frac{\alpha}{\alpha+\beta} f(t) \\ \frac{\beta}{\alpha+\beta} f(t) & 1 - \frac{\beta}{\alpha+\beta} f(t) \end{pmatrix}, \tag{2.6}$$

where $f(t) = 1 - e^{-(\alpha+\beta)t}$. The first row and the first column of $\mathbf{Q}$ and $\mathbf{P}(t)$ correspond to state "OFF" and the second row and second column correspond to state "ON". Using Equation (2.4), the steady state probability for the above CTMC is

$$\pi = \begin{pmatrix} \frac{\beta}{\alpha+\beta} \\ \frac{\alpha}{\alpha+\beta} \end{pmatrix}. \tag{2.7}$$

It can be easily seen that each row of $\mathbf{P}(t) \rightarrow \pi$, or the steady state probability of the CTMC as $t \rightarrow \infty$.

2.4 Independent CTMCs

Extending the above set up, we model a system of multiple independent CTMCs as another CTMC. Each individual CTMC can correspond to a component within a firm or a business.

Consider a system of N individual CTMCs; $X = \{X^1, \ldots, X^N\}$, each with an associated infinitesimal generator $\mathbf{Q}^i$ and a transition probability matrix $\mathbf{P}^i(t)$, $i \in \{1, \ldots, N\}$. When the individual X^i's, $i \in \{1, \ldots, N\}$ are independent, the joint system X is also a CTMC with a state space of all possible combinations of the states of the individual CTMCs. The infinitesimal generator $\mathbf{Q}$ and the transition probabilities at time t, $\mathbf{P}(t)$ of the joint system X are given by:

$$\begin{aligned} \mathbf{Q} &= \oplus_{i=1}^{N} \mathbf{Q}^i, \\ \mathbf{P}(t) &= \otimes_{i=1}^{N} \mathbf{P}^i(t), \end{aligned} \tag{2.8}$$

where $\otimes$ is the Kronecker product of matrices and $\oplus$ is the Kronecker sum of the matrices.[2]

[2]If $\mathbf{A}$ and $\mathbf{B}$ are two matrices, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \ldots & a_{1n}\mathbf{B} \\ \vdots & \ldots & \vdots \\ a_{m1}\mathbf{B} & \ldots & a_{mn}\mathbf{B} \end{pmatrix}. \tag{2.9}$$

If $\mathbf{A}$ is an $n \times n$ matrix and $\mathbf{B}$ is an $m \times m$ matrix, the Kronecker sum is:

$$\mathbf{A} \oplus \mathbf{B} = A \otimes \mathbf{I}_m + \mathbf{I}_n \otimes \mathbf{B}, \tag{2.10}$$

where $\mathbf{I}_k$ is an identity matrix of dimension k.

Example

Consider two independent CTMCs , X^1 and X^2 each with two states "OFF" and "ON". If, for both the CTMCs, the intensity of transition from "OFF"$\rightarrow$"ON" is α and the intensity of transition from "ON"$\rightarrow$"OFF" is β, then, the infinitesimal generator of the joint system of the 2 CTMCs is:

$$\begin{aligned}\mathbf{Q} &= \mathbf{Q}^1 \otimes \mathbf{I}_2 + \mathbf{I}_2 \otimes \mathbf{Q}^2 \\ &= \begin{pmatrix} -2\alpha & \alpha & \alpha & 0 \\ \beta & -(\alpha+\beta) & 0 & \alpha \\ \beta & 0 & -(\alpha+\beta) & \alpha \\ 0 & \beta & \beta & -2\beta \end{pmatrix}\end{aligned} \tag{2.11}$$

The rows and columns of $\mathbf{Q}$ correspond to the states ("OFF","OFF"), ("OFF","ON"), ("ON","OFF") and ("ON","ON") in that order. Here, we would like to make two observations about systems of multiple CTMCs. First, there is an explosion in the size of the joint state space of the system. The state space of the new system has all possible combinations of the states of the individual CTMCs. Since each CTMC in the above example has two state, the joint system has 2×2 or four states. If each individual CTMC X^i, $i \in \{1, 2, \ldots, N\}$ has S^i states, the state space of the joint CTMC X has $\prod_{i=1}^{N} S^i$ states.

Second, a fundamental characteristic of the system constructed above is that components of the system never transition simultaneously. For example, in the above system, the transition intensity from ("ON","ON") to ("OFF","OFF") is 0 i.e., two CTMCs never transition simultaneously. Thus, the individual components do not exhibit dependence which is often found in real life. A hurricane is generally followed by immediate disruptions in all the affected areas.

A possible way to introduce dependence in a system of CTMCs is to specify the entire generator matrix; i.e., to explicitly specify all the transition intensities in the generator matrix of the joint system. In the above example, we could specify all the 16 parameters of the generator matrix $\mathbf{Q}$ to ensure that simultaneous transitions occur with non-zero intensities.

This approach has many difficulties. First, the number of parameters increases exponentially. A small system, with 10 component CTMCs, each with two states, needs $2^{10} \times 2^{10}$, or roughly a million parameters. Second, the large number of parameters makes the system unwieldy and hence makes it difficult to study the impact of dependence. Finally, by defining the entire matrix, we lose out on properties like ease-of-simulation and closed form estimates of probabilities, which make CTMCs an attractive option.

Hence, defining the entire generator matrix is not a feasible option when dealing with systems with multiple dependent CTMCs . Subordination can overcome these difficulties while inducing dependence and simultaneous transitions in systems of multiple CTMCs.

2.5 Subordinated CTMC

Stochastic time-changing, known as subordination in the sense of Bochner – see Bochner (1949)– can be applied to any stationary Markov process. We, however, only consider subordination of CTMCs.

We explain the construction of subordinated Markov Chains by building on

a system of independent CTMCs introduced in section 2.4. Subordinated CTMCs are independent CTMCs in a different randomized time scale. The properties of the time scale are what ensure dependence in the components of the system.

Consider a CTMC $X = \{X^1, \ldots, X^N\}$ with the independent components X^i, $i = 1, \ldots, N$. As before, each component CTMC, X^i, $i = 1, \ldots, N$ has an associated infinitesimal generator $\mathbf{Q}^i$ and a transition probability matrix $\mathbf{P}^i(t)$. We construct a new process X^ϕ, with the same state space, where each component, $X^{\phi,i}$, is defined by time-changing each X^i with the random clock $T = \{T_t, t \geq 0\}$; i.e., $X^\phi = \{X^{\phi,1}, \ldots, X^{\phi,N}\}$ with $X^{\phi,i}(t) := \{X^i(T_t), t \geq 0\}$, $i = 1, \ldots, N$.

Since the clock T is random, the probability transition matrix $\mathbf{P}^\phi(t)$ of X^ϕ is given by

$$\mathbf{P}^\phi(t) = \mathbb{E}[\mathbf{P}(T_t)] = \mathbb{E}\left[\bigotimes_{i=1}^{N} \mathbf{P}^i(T_t)\right], \tag{2.12}$$

where the last equality of Equation (2.12) follows from the fact that, conditional on the realization of the random clock T, the CTMCs X^i are independent and the transition probability of a system of independent CTMCs is given by Equation (2.8). In the case when the random clock T_t is a Lévy subordinator; i.e., a non-negative, right continuous with left limits (RCLL) process with independent and stationary increments that starts at zero, X^ϕ is also a CTMC with dependent components. Intuitively, if T jumps, then all component CTMCs X^i, $i = 1, \ldots, N$, simultaneously experience the jump, resulting in a non-zero probability of simultaneous transitions or dependence in the multiple components. Thus, jumps of T can be interpreted as external common shocks that may cause disruptions in the operational state of the multiple components.

The Laplace transform of T is given by

$$\mathbb{E}[e^{-\lambda T_t}] = \int_{[0,\infty)} e^{-\lambda s}\pi_t(ds) = \exp\{-\phi(\lambda)t\}, \quad \lambda > 0,\ t \geq 0, \tag{2.13}$$

where π_t is the transition density of the subordinator T_t, and $\phi : [0,\infty) \mapsto [0,\infty)$ is the Lévy-Khintchine exponent given by

$$\phi(\lambda) = a\lambda + \int_{(0,\infty)} (1 - e^{-\lambda s})\nu(ds). \tag{2.14}$$

Here, $a \geq 0$ and $\nu(.)$ are the drift and Lévy measure of the subordinator, respectively. In addition, the Lévy measure satisfies the integrability condition $\int_{(0,\infty)}(1 \wedge s)\nu(ds) < \infty$. For $\nu((0,\infty)) = C < \infty$, the subordinator T is a compound Poisson process with intensity C. If $\nu((0,\infty)) = \infty$ then T is an infinite activity subordinator; i.e., T experiences an infinite number of jumps in each infinitesimal time interval dt. Since, in Operations Management applications, we expect the number of shocks to be finite for finite time intervals, we model T as a compound Poisson process with drift and exponential jump sizes. Such processes are defined by three parameters: the drift $a \geq 0$, the jump arrival rate $C > 0$, the rate parameter of the exponential distribution of the jump sizes $\eta > 0$.

In this case,

$$\begin{aligned} \nu(ds) &= C\eta e^{-\eta s} ds, \text{ and} \\ \phi(\lambda) &= a\lambda + \frac{C\lambda}{\lambda + \eta}. \end{aligned} \tag{2.15}$$

The subordinator can be represented as

$$T_t = at + \sum_{k=1}^{K(t)} \mathbf{1}_{\{t_k < t\}} J_k, \tag{2.16}$$

where $K(t) \sim \text{Poisson}(C)$ is the number of jumps, $\mathbf{t} = (t_1, \ldots, t_{K(t)})$ are the jump times, and $\mathbf{J} = (J_1, \ldots, J_{K(t)})$ are the jump sizes, $J_k \sim \text{Exp}(\eta)$.

Phillips' theorem – see, e.g., Theorem 32.1 in Sato (1999) – shows that the subordinated process X^ϕ is also a CTMC. Hence, we can characterize its transition matrix $\mathbf{P}^\phi(t)$ and infinitesimal generator $\mathbf{Q}^\phi$, in terms of the transition probability matrix $\mathbf{P}(s)$, the generator matrix $\mathbf{Q}$ of the underlying system of independent CTMCs and the subordinator T. Using Phillips' theorem, $\mathbf{P}^\phi(t)$ and $\mathbf{Q}^\phi$ are given by:

$$\begin{aligned} \mathbf{P}^\phi(t) &= \int_{[0,\infty)} \mathbf{P}(s)\pi_t(ds) \text{ and} \\ \mathbf{Q}^\phi &= a\mathbf{Q} + \int_{(0,\infty)} (\mathbf{P}(s) - \mathbf{I})\nu(ds), \end{aligned} \tag{2.17}$$

where $\mathbf{I}$ is an identity matrix of the same dimension as $\mathbf{P}(s)$.

2.5.1 Dependence via Subordination

To formally show that the components of X^ϕ are dependent, we use the eigenvalue decomposition of $\mathbf{Q}$. Let $\mathbf{V}^i$ and $\mathbf{\Lambda}^i$, $i = 1, \ldots, N$, be the matrix of eigenvectors and diagonal matrix of eigenvalues of the infinitesimal generator $\mathbf{Q}^i$, such that $\mathbf{Q}^i = \mathbf{V}^i \, \mathbf{\Lambda} \, (\mathbf{V}^i)^{-1}$, $i = 1, \ldots, N$. Then, $\mathbf{V} = \bigotimes_{i=1}^N \mathbf{V}^i$ is the matrix of eigenvectors and $\mathbf{\Lambda} = \bigoplus_{i=1}^N \mathbf{\Lambda}^i$ is the diagonal matrix of eigenvalues of the infinitesimal generator $\mathbf{Q}$ – see, e.g., Chapter 7 in Meyer (2001). Using the classical Dunford-Taylor integral – see Dunford and Schwartz (1958), Section VII.3 – the transition matrix $\mathbf{P}(t)$ is given by $\mathbf{P}(t) = \mathbf{V} e^{\mathbf{\Lambda} t} \mathbf{V}^{-1}$. Using once more the Dunford-Taylor integral

$$\begin{aligned} \mathbf{P}^{\phi}(t) &= \mathbf{V} e^{\phi(\mathbf{\Lambda})t} \mathbf{V}^{-1}, \text{ and} \\ \mathbf{Q}^{\phi} &= \mathbf{V} \phi(\mathbf{\Lambda}) \mathbf{V}^{-1}, \end{aligned} \tag{2.18}$$

where ϕ is the Lévy-Kintchine exponent (2.14) of the subordinator T. Observe that if $\lambda^{(i)}_{n^{(i)}}$ corresponds to the $n^{(i)}$th eigenvalue of $\mathbf{\Lambda}^i$, and N^i corresponds to the dimension of $\mathbf{\Lambda}^i$, then $\lambda_r = \sum_{i=1}^{N} \lambda^{(i)}_{n^{(i)}}$ is the rth eigenvalue of $\mathbf{Q}$ where $r = \big(\sum_{i=1}^{N-1}(n^{(i)}-1)\big(\prod_{j=i+1}^{N} N^j\big) + n^{(N)}\big)$. Hence, $\phi(\lambda_r) = \phi\big(\sum_{i=1}^{N} \lambda^{(i)}_{n^{(i)}}\big)$ corresponds to the rth eigenvalue of $\mathbf{Q}^{\phi}$, which is different from $\sum_{i=1}^{N} \phi(\lambda^{(i)}_{n^{(i)}})$. This difference implies dependence among the components of X^{ϕ}. The following proposition illustrates this point.

Proposition 1. *Let $X^{\phi} = \{X^{\phi,1}, \ldots, X^{\phi,N}\}$, be the subordinated CTMC resulting from time-changing X with a Lévy subordinator T, different from the trivial subordinator, whose Lévy-Kintchine exponent is denoted by ϕ. Then, the marginal CTMCs $X^{\phi,i}$, $i = 1, \ldots, N$, defined as $X^{\phi,i}_t = X^i(T_t)$, are dependent.*

Proof. By contradiction. Assume that $X^{\phi,i}_t = X^i(T_t)$ are independent. Then the infinitesimal generator of X^{ϕ} is given by $\mathbf{V} \bigoplus_{i=1}^{N} \phi(\mathbf{\Lambda}^i) \mathbf{V}^{-1}$. On the other hand, from (2.18) we have that $\mathbf{Q}^{\phi} = \mathbf{V}\phi(\mathbf{\Lambda})\mathbf{V}^{-1}$, which is a contradiction since $\bigoplus_{i=1}^{N} \phi(\mathbf{\Lambda}^i) \neq \phi(\bigoplus_{i=1}^{N} \mathbf{\Lambda}^i)$ implying $\mathbf{P}^{\phi}(t) \neq \bigoplus_{i=1}^{N} \mathbf{P}^{\phi,i}(t)$ unless $\phi(\lambda) = \lambda$; i.e., the trivial subordinator $T_t = t$. □

Example

Continuing from the previous example from section 2.4, consider two independent CTMCs, X^1 and X^2 each with two states, "OFF" and "ON", and a sub-

ordinator with a drift a, intensity of jumps C and expected jump size $1/\eta > 0$. If the intensity of transition from "OFF"$\rightarrow$"ON" is α and the intensity of transition from "ON"$\rightarrow$"OFF" is β, then by Equation (2.18), the infinitesimal generator of the subordinated joint system of the 2 CTMCs is:

$$\mathbf{Q}^{\phi} = \phi(\mathbf{I}_2 \otimes \mathbf{Q}^1 + \mathbf{Q}^2 \otimes \mathbf{I}_2) =$$
$$f(\alpha,\beta)\begin{pmatrix} 2\alpha\beta\Phi_1 + \alpha^2\Phi_2 & \alpha(\alpha+\beta)\Phi_1 - \alpha^2\Phi_2 & \alpha(\alpha+\beta)\Phi_1 - \alpha^2\Phi_2 & \alpha^2\Phi_2 - 2\alpha^2\Phi_1 \\ -\beta^2\Phi_1 + \alpha\beta\Phi_1 - \alpha\beta\Phi_2 & (\beta^2+\alpha^2)\Phi_1 + \alpha\beta\Phi_2 & \alpha\beta(\Phi_2 - 2\Phi_1) & \alpha^2\Phi_1 + \alpha\beta\Phi_2 - \alpha\beta\Phi_1 \\ -\beta^2\Phi_1 + \alpha\beta\Phi_1 - \alpha\beta\Phi_2 & \alpha\beta(\Phi_2 - 2\Phi_1) & (\beta^2+\alpha^2)\Phi_1 + \alpha\beta\Phi_2 & \alpha^2\Phi_1 + \alpha\beta\Phi_2 - \alpha\beta\Phi_1 \\ \beta^2\Phi_2 - 2\beta^2\Phi_1 & \beta(\alpha+\beta)\Phi_1 - \beta^2\Phi_2 & \beta(\alpha+\beta)\Phi_1 - \beta^2\Phi_2 & 2\alpha\beta\Phi_1 + \beta^2\Phi_2 \end{pmatrix}, \tag{2.19}$$

where

$$\begin{aligned} f(\alpha,\beta) &= \frac{1}{(\alpha+\beta)^2}, \\ \phi(\gamma) &= a\gamma + C\frac{\gamma}{\gamma+\eta}, \\ \Phi_1 &= \phi(-(\alpha+\beta)), \text{ and} \\ \Phi_2 &= \phi(-2(\alpha+\beta)). \end{aligned} \tag{2.20}$$

As before, the rows and columns of $\mathbf{Q}^{\phi}$ correspond to the states ("OFF","OFF"), ("OFF","ON"), ("ON","OFF") and ("ON","ON") in that order. Subordination leads to non-zero intensity rates of simultaneous transitions of the two component CTMCs. For example, the transition intensity from ("ON","ON") to ("OFF","OFF"); i.e., from both CTMCs being in the "ON" state to both CTMCs in "OFF" state, is non-zero. This was not the case in the system of independent CTMCs, as seen in Equation (2.11).

The same pattern is observed when dealing with more than two CTMCs. Subordinating multiple CTMCs with the same subordinator results in dependent

transitions of all component CTMCs. A single subordinator is needed to induce simultaneous transitions, irrespective of the number of CTMCs. The size of the parameter space thus does not grow exponentially with the number of CTMCs. Rather, it is a linear function of the number of CTMCs.

Since jumps in the subordinator are analogous to external shocks to the components, the degree of dependence can be controlled by the expected size of the jumps $1/\eta$. A lower η implies higher jumps sizes or bigger common shocks and thus greater degree of dependence. In fact, the intensity of independent transitions; i.e., the intensity that exactly one component CTMC transitions, tends to 0 as $\eta \to 0$. In the extreme case, when $\eta = 0$, the two CTMCs are completely dependent and only simultaneous transitions occur.

Thus, subordination allows not only dependence in a parsimonious fashion, but also is a framework where the degree of dependence can be controlled by a single parameter.

However, as mentioned, classical subordination is not applicable to Operations Management problems. In the setup described above, we see that the transition intensity in opposite directions is non-zero, i.e., it is possible for one CTMC to move from "OFF" to "ON" while another switches from "ON" to "OFF". This does not square up with transitions in Operations Management. Factors causing disruptions, for example, fuel shortages for power plants, can cause outages of power plants but not cause power plants to restart operations.

Hence, the formulation of a subordinated Markov chain is not applicable as-is

to Operations Management problems. A degree of customization is needed to make subordination suitable in Operations Management. We do that by constructing partially subordinated Markov chains. To mathematically establish partially subordinated Markov chains, we first introduce subordinated Markov chains with respect to their counting measures.

2.5.2 Subordinated CTMC and its corresponding counting measure

We formalize the subordinated CTMC model construction in terms of the counting measure of the transitions from state i to state j, for all $i, j \in \mathcal{S}$.

We assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given, and that the underlying CTMC process $X = \{X(t), t \geq 0\}$ with values in $\mathcal{S}$, and infinitesimal generator $\mathbf{Q}$, is well defined on this probability space. Each element $q_{i,j} = [\mathbf{Q}]_{i,j}$, $i, j \in \mathcal{S}$, of $\mathbf{Q}$ represents the intensity of jumps of X from i to j. These intensities satisfy the standard conditions, i.e., $q_{i,j} \geq 0$ for all $i \neq j$, and $q_{i,i} = -\sum_{j \neq i} q_{i,j}$. Let $\mathbb{F}^X = \{\mathcal{F}_t^X, t \geq 0\}$ be the completed natural filtration generated by X.

For any two states $i, j \in \mathcal{S}$, such that $i \neq j$, we represent the number of jumps from state i to state j over the time interval $(0, t]$ by the $\mathbb{F}^X$-optional random measure $N_{i,j}(t) = N_{i,j}((0, t])$, which is defined on $[0, \infty)$ by

$$N_{i,j}((0, t]) = \sum_{0 < s \leq t} \mathbb{1}_{\{X_{s^-} = i, X_s = j\}}. \tag{2.21}$$

From Lemma 5.1 in Bielecki et al. (2008), it follows that the $\mathbb{F}^X$-compensator; i.e., the dual predictable projection, $\nu_{i,j}(dt)$ of the counting measures $N_{i,j}(dt)$, $i, j \in \mathcal{S}$ is

of the form

$$\nu_{i,j}(dt) = \mathbb{1}_{\{X_{t-}=i\}} q_{i,j} dt, \tag{2.22}$$

such that

$$M_{i,j}(t) = N_{i,j}(t) - \int_0^t \nu_{i,j}(ds) = N_{i,j}(t) - q_{i,j} \int_0^t \mathbb{1}_{\{X_{s-}=i\}} ds \tag{2.23}$$

is an $\mathbb{F}^X$-martingale. Moreover, any set of counting measures $N_{i,j}(t)$, $i, j \in \mathcal{S}$, such that their compensators $\nu_{i,j}(dt)$ are defined as in Equation (2.22), defines a CTMC M on $\mathbb{F}^X$ with generator $\mathbf{Q}$.

Assume that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supports a subordinator $T = \{T_t, t \geq 0\}$ that is independent of M. Let $L = \{L_t, t \geq 0\}$ be the right-inverse of the subordinator process T; i.e., $L_t = \inf\{s \geq 0 : T_s > t\}$, for all $t \geq 0$. Note that L is RCLL since T is RCLL. Let $\mathbb{L} = \{\mathcal{L}_t, t \geq 0\}$ denote the completed natural filtration of L, and also let $\mathbb{H} = \{\mathcal{H}_t, t \geq 0\}$ denote the enlarged filtration with $\mathcal{H}_t = \mathcal{F}_t^X \vee \mathcal{L}_t$. Then T is an increasing family of $\mathbb{H}$-stopping times, and we can define the time-changed filtration $\mathbb{H}^\phi = \{\mathcal{H}_t^\phi, t \geq 0\}$ by $\mathcal{H}_t^\phi = \mathcal{H}_{T_t}$ for all $t \geq 0$. The subordinated CTMC process $X^\phi(t) = X(T_t)$ is clearly $\mathbb{H}^\phi$-adapted and RCLL. Moreover, the process X^ϕ is an $\mathbb{H}^\phi$-semimartingale – see Mendoza-Arriaga and Linetsky (2014). From Theorem 3.2(iv) in Mendoza-Arriaga and Linetsky (2014), it follows that for all $f \in \mathrm{Dom}(\mathbf{Q})$, the process

$$\mathcal{M}^f(t) = f\big(X^\phi(t)\big) - f\big(X^\phi(0)\big) - \int_0^t \phi(\mathbf{Q}) f\big(X^\phi(s)\big) ds \tag{2.24}$$

is an $\mathbb{H}^\phi$-martingale. Here, the matrix $\phi(\mathbf{Q})$ obtained from Equation (2.17). This implies that $\phi(\mathbf{Q})$ is indeed the infinitesimal the generator of the subordinated process X^ϕ. Hence, from Lemma 5.1 in Bielecki et al. (2008), it follows that

$$\nu_{i,j}^{\phi}(t) = \mathbb{1}_{\{X^{\phi}(t-)=i\}} \left[\phi(\mathbf{Q})\right]_{i,j} dt, \quad i,j \in \mathcal{S} \tag{2.25}$$

is the $\mathbb{H}^{\phi}$-compensator of the counting measure $N_{i,j}^{\phi}(t) := N_{i,j}(T_t)$. From Lemma 5.1 in Bielecki et al. (2008), we can also conclude that any set of counting measures $N_{i,j}^{\phi}(t)$, $i,j \in \mathcal{S}$, such that their compensators $\nu_{i,j}^{\phi}(dt)$ are defined as in (2.25), defines the CTMC $X^{\phi}(t) = X(T_t)$ on $\mathbb{H}^{\phi}$ with generator $\phi(\mathbf{Q})$.

2.6 Partially Subordinated CTMC: a factor-like model

Factor models are conceptually appealing for systems of multiple components since they allow us to distinguish between idiosyncratic factors; i.e., proper to a particular component, and systematic factors; i.e., common to some, if not to all, components in the system. We would like to model disruptions of individual component CTMCs – idiosyncratic behavior – as well as system-wide events that simultaneously affect several component CTMCs – systematic behavior – using subordination. We extend the theory of subordinated CTMCs to allow for multiple factors in a model that we call partially Subordinated CTMC

Consider the CTMC $X = \{X^1, \ldots, X^N\}$ with an infinitesimal generator $\mathbf{Q}$ given by the linear combination $\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2$, where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are also infinitesimal generators. We interpret $\mathbf{Q}$ as the sum of an idiosyncratic factor $\mathbf{Q}_1$ and a systematic factor $\mathbf{Q}_2$. That is, we consider $\mathbf{Q}_1$ as the collection of idiosyncratic factors affecting each component of X; i.e., $\mathbf{Q}_1 = \bigoplus_{i=1}^{N} \mathbf{Q}_1^i$, while we consider $\mathbf{Q}_2$ as the portion of $\mathbf{Q}$ which drives the dependent behavior of X. In particular, $\mathbf{Q}_2 = \phi(\widetilde{\mathbf{Q}}_2)$ for some

infinitesimal generator $\widetilde{\mathbf{Q}}_2 := \bigoplus_{i=1}^{N} \widetilde{\mathbf{Q}}_2^i$. Note that in this case, the infinitesimal generator of the individual component CTMC i is $\mathbf{Q}^i = \mathbf{Q}_1^i + \mathbf{Q}_2^i = \mathbf{Q}_1^i + \phi(\widetilde{\mathbf{Q}}_2^i)$

Continuing the discussion of section 2.5.2, we consider a counting measure $N_{ij}(t) := \tilde{N}_{i,j}(t) + \hat{N}_{i,j}(t)$ with the individual counting processes $\tilde{N}_{i,j}(t)$ and $\hat{N}_{i,j}$ having compensators $\tilde{\nu}_{i,j}$ and $\hat{\nu}_{i,j}$ where:

$$\tilde{\nu}_{i,j}(t) = \mathbb{1}_{\{X(t-)=i\}} \left[(\mathbf{Q}_1)\right]_{i,j} dt,$$
$$\hat{\nu}_{i,j}(t) = \mathbb{1}_{\{X(t-)=i\}} \left[\phi(\widetilde{\mathbf{Q}}_2)\right]_{i,j} dt.$$

It then follows from Lemma 5.1 in Bielecki et al. (2008) that the set of counting measures $N_{i,j}, i, j \in \mathcal{S}$ uniquely defines a CTMC X with the generator $\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2$ with $\mathbf{Q}_2 = \phi(\widetilde{\mathbf{Q}}_2)$. Consequently, this model can be interpreted as a CTMC with idiosyncratic factor $\mathbf{Q}_1$ and systematic factor $\mathbf{Q}_2$ acting together. As explained in section 2.5, the infinitesimal generator $\mathbf{Q}_2 = \phi(\widetilde{\mathbf{Q}}_2)$ of X^2 results from Bochner subordination, and hence, it allows us to model dependence.

Example

Consider two CTMCs , X^A and X^B each with two states "ON" and "OFF" . For each CTMC, the transitions from "OFF"$\rightarrow$"ON" are independent and occur with a transition intensity α. Thus, the infinitesimal generator associated with"OFF"$\rightarrow$ "ON" transitions is $\mathbf{Q}_1 = \begin{pmatrix} -\alpha & \alpha \\ 0 & 0 \end{pmatrix}$. To model dependent "ON"$\rightarrow$"OFF" transitions, we associate an infinitesimal generator $\mathbf{Q}_2 = \begin{pmatrix} 0 & 0 \\ \gamma & -\gamma \end{pmatrix}$ for "ON"$\rightarrow$"OFF" transitions and a subordinator with Laplace transform $\phi(.)$.

Then, the infinitesimal generator of each individual CTMC is:

$$\begin{aligned}\mathbf{Q}^A = \mathbf{Q}^B &= \mathbf{Q}_1 + \phi(\mathbf{Q}_2) \\ &= \begin{pmatrix} -\alpha & \alpha \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \phi(\gamma) & \phi(\gamma) \end{pmatrix} \\ &= \begin{pmatrix} -\alpha & \alpha \\ \phi(\gamma) & -\phi(\gamma) \end{pmatrix}\end{aligned} \tag{2.26}$$

The joint system is a CTMC with an infinitesimal generator given by:

$$\begin{aligned}\mathbf{Q} &= \mathbf{Q}_1 \oplus \mathbf{Q}_1 + \phi(\mathbf{Q}_2 \oplus \mathbf{Q}_2) \\ &= \begin{pmatrix} -2\alpha & \alpha & \alpha & 0 \\ 0 & -\alpha & 0 & \alpha \\ 0 & 0 & -\alpha & \alpha \\ 0 & 0 & 0 & 0 \end{pmatrix} + \phi\left(\begin{pmatrix} 0 & 0 & 0 & 0 \\ \gamma & -\gamma & 0 & 0 \\ \gamma & 0 & -\gamma & 0 \\ 0 & \gamma & \gamma & -2\gamma \end{pmatrix} \right) \\ &= \begin{pmatrix} -2\alpha & \alpha & \alpha & 0 \\ 0 & -\alpha & 0 & \alpha \\ 0 & 0 & -\alpha & \alpha \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ \phi(\gamma) & -\phi(\gamma) & 0 & 0 \\ \phi(\gamma) & 0 & -\phi(\gamma) & 0 \\ 2\phi(\gamma) - \phi(2\gamma) & \phi(2\gamma) - \phi(\gamma) & \phi(2\gamma) - \phi(\gamma) & -\phi(2\gamma) \end{pmatrix} \\ &= \begin{pmatrix} -2\alpha & \alpha & \alpha & 0 \\ \phi(\gamma) & -(\alpha + \phi(\gamma)) & 0 & \alpha \\ 0 & 0 & -(\alpha + \phi(\gamma)) & \alpha \\ 2\phi(\gamma) - \phi(2\gamma) & \phi(2\gamma) - \phi(\gamma) & \phi(2\gamma) - \phi(\gamma) & -\phi(2\gamma) \end{pmatrix}\end{aligned} \tag{2.27}$$

Note that, in this example, the two component CTMCs have different generators that have the same values for transition intensities. As before, the rows and columns of $\mathbf{Q}^\phi$ correspond to the states ("OFF","OFF"), ("OFF","ON"), ("ON", "OFF") and ("ON","ON") . From the final transition intensities it is clear that while transitions from "OFF" to "ON" are idiosyncratic – the system cannot transition from ("OFF","OFF") to ("ON","ON") simultaneously – the transitions from "ON" to "OFF" are dependent and can hence occur simultaneously for both CTMCs. The two component CTMCs cannot have simultaneous transitions in opposing directions

(the transition intensity from ("OFF","ON") to ("ON","OFF") is 0). Extending a similar set up to multiple component CTMCs, with each CTMC having two associated generators, results in a system with dependence only in particular directions.

In addition to parsimony and ease of simulation, partially subordinated CTMCs have three advantages.

First, the flexibility of partially subordinated CTMCs in controlling the direction of dependence, particularly for problems in Operations Management. In the above example, the disruptions ("ON"$\rightarrow$"OFF" transitions) are dependent while the recoveries are idiosyncratic. This behavior is in line with disruptions caused by risk factors like hurricanes where shocks affect the multiple components simultaneously. The recoveries after disruptions like hurricanes are independent. This independence is also captured in the above example. Other options include having an idiosyncratic disruption intensity in the infinitesimal generator $\mathbf{Q}^1$ or having multiple causes or factors of dependent disruptions by including multiple subordinators and each subordinator having an infinitesimal generator for each component CTMC. The recoveries can also be made dependent just by the inclusion of a subordinator exclusively for the recoveries.

Second, partially subordinated CTMCs are factor models and hence a sense of attribution is possible for the dependence. Each subordinator can be related to an observable cause of dependent behavior. For example, shared demographics, common geography etc. can be reasons for dependent behavior of the components.

Third, partially subordinated Markov chains are suitable for "lumpability".

Lumpability was first introduced in Kemeny and Snell (1960) and is a method for reducing the state space of large CTMCs. Under certain conditions on the infinitesimal generator, elements of the state space of the CTMC can be lumped into disjoint subsets. The lumped subsets can be considered the state space of another CTMC whose infinitesimal generator can be evaluated from the infinitesimal generator of the original CTMC. The transition probabilities of the lumped CTMC are the probabilities that the process transitions from a lumped subset to another lumped subset of the original state space. Since the lumped state space is smaller, it is easier to handle mathematically. The consequence of this property is that, when dealing with a system of partially subordinated Markov chains, we do not have to consider the entire state space to analyze the dynamics of a subset of CTMCs. For example, in a system with 10 component CTMCs, we can analyze any three CTMCs by considering only their generators and the associated subordinator. Hence, partially subordinated CTMCs are easier to slice and dice and analyze only the components of interest in a system with a large number of components.

2.6.1 Setting the Subordinator parameters

The partially subordinated CTMC system is defined by the parameters of the infinitesimal generators and the subordinators. In this section, we further simplify the parameter space of the subordinator in the context of Operations problems while using partially subordinated Markov chains.

The main purpose of using subordinators is to induce dependence among multiple components. Dependence is induced by the jumps and jump sizes in the

subordinator. The drift parameter a is immaterial in the context of modeling dependence and only results in idiosyncratic behavior in the component CTMCs. In the partially subordinated Markov chain framework, we model idiosyncratic behavior using a distinct infinitesimal generator for each component that is not associated with any subordinator. Thus we set the drift component a of the subordinators to zero and model the subordinators only as a summation of jumps. Since we only consider subordinators with no drift ($a = 0$), the Laplace transform of a subordinator is:

$$\phi(\gamma) = C\frac{\gamma}{\gamma + \eta}. \tag{2.28}$$

From Equation (2.28),it is clear that there is degeneracy in the γ and the η parameters. Consider two subordinators, T and T^k with the same arrival rate of jumps C, but different jump size parameters, η and $k\eta$ respectively. If the Laplace transform of the two subordinators are $\phi(.)$ and $\phi^k(.)$, it is easy to see that for any value of γ,

$$\phi(\gamma) = \phi^k(\gamma k). \tag{2.29}$$

Thus, it is possible to normalize the jump size parameter η to 1 and induce dependence only through the γ parameter. This allows a further reduction of the parameter space of the partially subordinated CTMC by considering only subordinators with expected size of jumps of one.

2.7 Simulation of the subordinated CTMC

An important advantage of subordinated CTMCs is that they allow efficient numerical simulation for large systems. We provide a Monte Carlo algorithm that allows the generation of sample paths of the partially subordinated CTMC introduced in section 2.6 in several steps. Algorithm 1 simulates a CTMC. Algorithm 2 simulates the sample path of a single subordinated CTMC X^{ϕ} with infinitesimal generator $\phi(\mathbf{Q})$. Algorithm 3 simulates the random path of a subordinator. Algorithm 4 simulates a multivariate case of the subordinated CTMC $X^{\phi} = \{X^{\phi,1}, \ldots, X^{\phi,N}\}$. Algorithm 5 simulates a partially subordinated CTMC. Algorithm 6 simulates multiple partially subordinated CTMCs with multiple subordinators.

First, we present a standard algorithm for simulating a continuous-time Markov chain up to time $t = S$.

Algorithm 1. *Simulating a continuous-time Markov chain X_t up to time $t = S$.* Assume the infinitesimal generator $\mathbf{Q}$ is available; i.e., numerically tractable.

1. Choose an initial state for the process $X = \omega$. Set $k = 0$ and $t = t_0 = 0$.

2. Generate the random variable $\tau_0 \sim \exp(-q_\omega)$, where $q_\omega = [\mathbf{Q}]_{\omega,\omega} < 0$ corresponds to the ω-th diagonal element of $\mathbf{Q}$. Set $t = t_1 = \tau_0$.

3. If $t < S$, generate the landing state Y_k as follows:

   (a) compute the vector of transition probabilities $\mathbf{p}_\omega(\tau_k)$ using $p_{\omega\tilde{\omega}}(\tau_k) = q_{\omega\tilde{\omega}}/(-q_\omega), \forall i \neq j$.

(b) use a uniformly distributed random variable U to determine the landing state Y_k from $\mathbf{p}_\omega(\tau_k)$ using the usual inverse transform sampling method.

(c) set $\omega = Y_k$. Then set $k = k + 1$ and $X_{t_k} = \omega$. Generate $\tau_k \sim \exp(-q_\omega)$, set $t = t + \tau_k$ and $t_{k+1} = t$.

4. If $t \geq S$ Stop. Otherwise, return to Step 3.

This algorithm provides the number of transitions $K(S) = \max\{k : t_k \leq S\}$ occurring during the time interval $(0, S]$, the sample path X_t, $0 \leq t \leq S$, and the corresponding jump times $t_1, \ldots, t_{K(S)}$.

We use Algorithm 1 to simulate sample paths of a CTMC X^ϕ. Let T be a Lévy subordinator with Lévy-Kintchine exponent ϕ defined as in (2.14). We choose the parameters $a = 0$; i.e., no drift, and $\eta = 1$ such that T is a compound Poisson process. Next, let X be a CTMC independent of T. Then, the construction $X^\phi(t) = X(T_t)$ defines a subordinated CTMC with infinitesimal generator $\phi(\mathbf{Q})$. Since X^ϕ is also a CTMC then the simulation algorithm for X^ϕ is the same as Algorithm 1, as long as $\phi(\mathbf{Q})$ is numerically tractable. That is, the CTMC X^ϕ can be simulated using Algorithm 2.

Algorithm 2. *Simulating a subordinated CTMC X_t^ϕ up to time $t = S$.* Assume the infinitesimal generator $\phi(\mathbf{Q})$ is available.

1. Set $\mathbf{Q} = \phi(\mathbf{Q})$ in Algorithm 1.

Algorithm 2 is simply a variation of Algorithm 1 when the CTMC is subordinated. Unfortunately, Algorithm 2 is not efficient when the number of components,

N is large since, it makes simulation computationally intractable due to the large size of $\mathbf{Q}$ and consequently $\phi(\mathbf{Q})$. This is the advantage of introducing dependence via subordination.

Since we only consider subordinators with no drift, the subordinated CTMC X^ϕ transitions from one state to another only at the jump times of T.

In this case the subordinated CTMC is defined by $X^\phi = (X^{\phi,1}, \ldots, X^{\phi,N})$, with $X^{\phi,i} = X^i(T)$ for all $i = 1, \ldots, N$. Conditional on the subordinator T, each CTMC X^i is independent of the others and the transition probability of X_t^ϕ is given by

$$\hat{\mathbf{P}}(t) = e^{\mathbf{Q}T_t} = \prod_{h=1}^{N} e^{\mathbf{Q}^h T_t} = \bigotimes_{h=1}^{N} e^{\widetilde{\mathbf{Q}}^h T_t} = \bigotimes_{h=1}^{N} \widetilde{\mathbf{P}}^h(T_t), \tag{2.30}$$

where $\tilde{\mathbf{Q}}^h$ is the infinitesimal generator of the underlying CTMC X^h, $h = 1, \ldots, N$; and $\widetilde{\mathbf{P}}^h(t) = e^{\widetilde{\mathbf{Q}}^h t}$ the corresponding transition probability matrix. Hence, the transition probability $\mathbf{P}_t^\phi$ of $X^\phi(t)$ is given by $\mathbf{P}_t^\phi = \mathbb{E}[\hat{\mathbf{P}}(t)]$. Taking advantage of the independence of all X^h and T, we provide a simulation algorithm for a subordinated CTMC. We first provide the algorithm to simulate the subordinator T up to time $t = S$ and then the algorithm to simulate a subordinated CTMC with subordinator T.

Algorithm 3. *Simulating subordinator T up to time $t = S$.* Assume arrival rate C given and the rate parameter $\eta = 1$.

1. Set $t = 0$ and the set of jump times $\mathbf{t}$ and the corresponding jump sizes $\mathbf{J}$ to the null.

2. Generate the random variable $\tau \sim exp(C)$. Update $t = t + \tau$

3. If $t < S$, Go to Step 4 else Stop.

4. Append the jump times set $\mathbf{t}$ with t and the jump sizes set $\mathbf{J}$ with a random variable generated as $exp(1)$. Go to Step 2.

Algorithm 3 simulates the subordinator up to a given time S. The Algorithm returns a sequence of jump times given by the set $\mathbf{t}$ and jump sizes given by the set $\mathbf{J}$. The jumps times and the jumps sizes are inputs to Algorithm 4 which simulates the subordinated Markov Chain up to time $t = S$.

Algorithm 4. *Simulating* $X_t^{\phi} = (X_t^{\phi,1}, \ldots, X_t^{\phi,N}) = (X^1(T_t), \ldots, X^N(T_t))$ *up to time* $t = S$. Assume that T is a finite activity Lévy subordinator as in Equation (2.16) with no drift $a = 0$.

1. Set $k = 0$ and $t_k = 0$, and choose an initial state $X^{\phi}(0) = (X^1(0), \ldots, X^N(0)) = (\omega_1, \ldots, \omega_N)$.

2. Simulate the Lévy subordinator T up to time $t = S$ using, e.g., Algorithm 3. That is, obtain the number of jumps $K(t)$, the jump times $\mathbf{t} = (t_1, \ldots, t_{K(t)})$, and the jump sizes $\mathbf{J} = (J_1, \ldots, J_{K(t)})$.

3. Since the subordinator has zero drift; i.e., $a = 0$, the components of X^{ϕ} can only transition to another state at the jump times $\mathbf{t}$ of T. Then, for $k = 1, \ldots, K(S)$:

   (a) Set $(\omega_1, \ldots, \omega_N) = (X^1(t_{k-1}), \ldots, X^N(t_{k-1}))$,

(b) Compute the landing states Y^h for $h = 1, \ldots, n$ by simulating each CTMC X^h for time J_k and generator $\widetilde{\mathbf{Q}}^h$ using Algorithm 1 to set $i_h = Y^h$ and $X^{\phi,h}(t_k) = i_h$.

Algorithm 4 allows to obtain sample paths of the subordinated CTMC X^ϕ with dependent components. In particular, for each sample path of the subordinator T, the components $X^{\phi,h}$, $h = 1, \ldots, N$, may shift to another state only at the jump times $\mathbf{t}$ of T. It follows that the transition probability matrix of the joint CTMC X^ϕ, $\mathbf{P}^\phi(t) = \mathbb{E}[\hat{\mathbf{P}}(t)] = \mathbb{E}[\bigotimes_{h=1}^{N} \widetilde{\mathbf{P}}^h(t)]$, can be simply estimated by counting and averaging the transitions from each simulation. We now present the algorithm for the simulation of a subordinated mixed CTMC.

Algorithm 5. *Simulating a Partially Subordinated CTMC up to time $t = S$.* Assume that T is a finite activity Lévy subordinator as in Equation (2.16) with no drift; i.e., $a = 0$, and the infinitesimal generators $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are available either in closed form or numerically.

1. Simulate the Lévy subordinator T up to time $t = S$ using, e.g., Algorithm 3. That is, obtain the number of jumps $K(t)$, the jump times $\mathbf{t} = (t_1, \ldots, t_{K(t)})$, and the jump sizes $\mathbf{J} = (J_1, \ldots, J_{K(t)})$.

2. Start at $l = 0$ and set $t_0 = 0$, $X_{t_l} = X_0$.

   (a) With starting state X_{t_l}, simulate the CTMC with generator $\mathbf{Q_1}$ for time $t_{l+1} - t_l$. Record the end state in $X_{t_{l+1}}$.

(b) With starting state $X_{t_{l+1}}$, simulate the CTMC with generator $\mathbf{Q_2}$ for time J_{l+1}. Record the end state in $X_{t_{l+1}}$.

(c) Set $l = l + 1$. If $l > K(t)$ proceed to Step 3. Else go to Step 2(a).

3. Simulate the CTMC with generator $\mathbf{Q_1}$ for time $S - t_{K(t)}$ and record the landing state at X_S.

Algorithm 5 is used to simulate the state of a subordinated mixed CTMC at time S given the starting state of the subordinated mixed CTMC and the associated generators and subordinator. It can also be used to simulate the state of the subordinated mixed CTMC at different equispaced points in time.

Multiple Subordinators

We have provided a simulation algorithm for the case of a partially subordinated Markov chain with a single subordinator. We can easily extend Algorithm 5 to the case of multiple subordinators. Below we describe the Algorithm to simulate a partially subordinated Markov chain with multiple subordinators.

Algorithm 6. *Simulating a partially subordinated CTMC with multiple subordinators up to time $t = S$* Assume that T^i, $i = 1, \ldots p$ are finite activity Lévy subordinators as in Equation (2.16) with no drift; i.e., $a = 0$. Associated with each subordinator T^i, $i = 1, \ldots p$, infinitesimal generator $\mathbf{Q_i}$ is available either in closed form or numerically. Also an idiosyncratic infinitesimal generator $\mathbf{Q}$ is available in closed form or numerically.

1. Simulate each Lévy subordinator T^i, $i = 1, \ldots p$ up to time $t = S$ using, e.g., Algorithm 3. That is, for each subordinator T^i, obtain the set of jump times $\mathbf{t^i}$, and the set of jump sizes $\mathbf{J^i}$.

2. Concatenate the set of all jump times and jump sizes $\mathbf{J} = \cup_{\mathbf{i}} \mathbf{J^i}$ and $\mathbf{t} = \cup_{\mathbf{i}} \mathbf{t^i}$.

3. Sort the set of jump times $\mathbf{t}$ and the set of corresponding jump sizes $\mathbf{J}$ in the increasing order or jump times. Start at $l = 0$ and set $t_0 = 0$, $X_{t_l} = X_0$.

   (a) With starting state X_{t_l}, simulate the CTMC with generator $\mathbf{Q}$ for time $t_{l+1} - t_l$. Record the end state in $X_{t_{l+1}}$.

   (b) With starting state $X_{t_{l+1}}$, simulate the CTMC with generator $\mathbf{Q_i}$ for time J_{l+1}, where $\mathbf{Q_i}$ corresponds to the generator associated with the subordinator that jumps at time t_{l+1}. Record the end state in $X_{t_{l+1}}$.

   (c) Set $l = l + 1$. If $l > K(t)$ proceed to Step 3. Else go to Step 3(a).

4. Simulate the CTMC with generator $\mathbf{Q}$ for time $S - t_{K(t)}$ and record the landing state at X_S.

Algorithms 5 and 6 can be extended to systems where multiple components have a common subordinator. Then, as in Algorithm 4, we first simulate a common subordinator path for all component CTMCs and then simulate the transitions of the component CTMCs using the common subordinator paths.

Chapter 3

Dependent Outages in Electricity Power Plants

3.1 Introduction

Reliable supply of electricity is critical to the modern economy. As additional capacity comes online, it is necessary to establish public policies that incentivize the production of reliable energy. One problem in providing electricity reliably is the occurrence of joint outages of power plants. Although rare, joint outages have a disproportionate impact; e.g., the tsunami of 2011 in Japan, caused the joint outages of multiple nuclear plants, and eventually the shutdown of all nuclear plants. This event caused severe power shortages in Japan and, as a consequence, major power users were asked to cut peak power usage by 15% – McCurry (2018).

In this research, we provide a framework to estimate the reliability of power generation capacity that accounts for the joint outages and recoveries of electricity power plants. The framework is a step towards establishing better policies to achieve higher reliability. We estimate the reliability of two systems by using our framework on real outage data.

Modeling joint outage and recovery dynamics of electricity power plants is challenging. In a region with n electricity power plants, and allowing for only two operating states for each plant (up/down), a full model of the dynamics of the system has 2^n different states and $2^n(2^n - 1)$ possible transition probabilities. Rather than modeling the transitions between the 2^n states directly, we induce dependence by introducing the concept of dependence sets. A dependence set identifies which power plants may be susceptible to common outages. Dependence sets may overlap, allowing for individual power plants to exhibit multiple types of dependence. For example, power plants can exhibit dependence due to common input fuel or geographic proximity.

The advantage of using dependence sets to induce correlation in outages is that they significantly reduce the number of parameters that need to be calibrated. To keep the model tractable, we induce correlation within each dependence set using ideas from time subordination. In a time-subordinated process, joint outages and recoveries are more (or less) likely depending on the jumps of a random clock that is common to power plants in the same dependence set. Due to the jumps of the common random clock, power plants exhibit dependence, while conditional on the value of the random clock, transitions are independent. Rather than modeling all possible transitions, we only need to choose parameters related to the process followed by the random clock for each dependence set.

We calibrate our model with a unique dataset from the National Electricity Reliability Corporation (NERC) which includes all outages of power plants with ca-

pacity greater that 50 MWs from 2010-2012.[1] We focus on two regions: the Electricity Reliability Council of Texas (ERCOT), and the Western Electricity Coordinating Council (WECC). The ERCOT region covers most of the state of Texas, while the WECC region covers the western United States and Canada. We choose the ERCOT and the WECC because they are regions that are also interconnections; i.e., areas across which power plants share resources, generation capacities and loads.

We find evidence of dependence in both regions. Plants that use natural gas exhibit dependence in outages in both the ERCOT and the WECC regions. In addition, in the WECC region, plants that use hydroelectric energy, distillate oil, geothermal energy and other gas exhibit dependence. After accounting for input fuel dependence, we also show that system-wide dependence in outages exists in WECC.[2] These results imply that accounting for dependence is necessary to accurately estimate reliability. For example, in the ERCOT region, when we take dependence into account, we find that the minimum capacity that is available at least 99.5% of the time, decreases by 0.4% or by approximately 200 MWs. It decreases by 0.2%, or approximately 350 MWs in the WECC.

We quantify the policy implications of our research in two ways. First we estimate the outage probability of a new plant of a particular input fuel for different operational levels of the existing generation capacity. For example, while a natural gas plant has an unconditional probability of outage of 0.7%, we find that in the

[1]NERC is a not-for-profit international regulatory authority to assure the reliability of power transmission in North America.

[2]We do not identify the cause of system-wide dependence. Possible examples include weather, the transmission network etc.

ERCOT region, when 5% of the existing capacity is down, the outage probability increases to 4.4%. In the WECC, when 5% if the existing generation capacity is down, the probability of outage of a new natural gas plant increases to 10%. The result is even more pronounced for a new plant that uses geothermal energy in the WECC region. For such a plant, the probability of outage increases from 0.3% to 21.5%.

The second way we quantify the policy implications of our research is to estimate the additional capacity needed to achieve a given reliability level. For example, allowing only additional natural gas capacity, we find that to achieve an availability of 48 GWs of capacity at least 98.5% of the time in the ERCOT, an additional 2370 MWs of natural gas generation are needed. Perhaps surprisingly, this amount is lower than the amount of generation capacity that would be needed if we assume that outages of natural gas plants were independent. The reason for this difference is that increased correlation in outages does not necessarily result in lower reliability. Rather, it results in larger, rare, outages which can increase or decrease the generation capacity available at different levels of reliability. This example illustrates that, to evaluate reliability, we need to consider different measures. Our research allows for the estimation of such measures while accounting for dependence in outages, and can be a building block for policymakers that design incentives for building new generation capacity.

3.2 Literature Review

The review papers by Allan et al. (1999) and Billinton et al. (2001) provide a broad overview of research in capacity reliability using probabilistic models. Almutairi et al. (2015) provides a detailed road map of adequacy analysis of generation capacity. The prevailing framework for modeling reliability in the literature, described in Billinton and Allan (1992), models the aggregate capacity available, commonly called the supply stack. Each capacity level of the supply stack has an associated probability used to calculate the reliability of the region. While this approach has the benefit of being applicable to systems with large numbers of power plants, it does not model individual plants and does not capture common sources of outages among plants.

Much of the literature that models individual plants assumes that outages are independent. Nasiri and Huang (2007) do study the capacity planning problem with multiple power plants to meet environmental targets, but do not consider outage correlation between different power plants. Mavrotas et al. (1999) address the same problem assuming an availability factor for each power plant. These papers allow variations in availability but, unlike our research, do not estimate how rare events can affect supply for a large number of power plants with multiple input fuels. Harvey et al. (2005) model the outages in thermal plants in California by considering the hazard rates as functions of the characteristics of different plants. While the approach includes dependence in hazard rates, it does not imply that outages happen simultaneously.

A few papers in the literature capture the dependence in outages between

power plants with the same input fuel by modeling the availability of the fuel. For example, the availability of wind for wind turbines or the presence of sunlight for solar panels. Borges and Pinto (2008) capture the impact of dependence of outages of hydroelectric power plants by modeling the inflows using a Markov model. Lopes and Borges (2015) capture outage dependence in a system with both wind and hydroelectric power plants. Rocha and Borges (2010) consider the case of wind and biomass generation facilities. These papers capture dependence in outages between power plants with specific input fuels but cannot capture dependence across the entire system.

A separate stream of literature considers correlation in disruptions of a transmission grid due to weather events – see the review paper Wang et al. (2016). Bernstein et al. (2014) specifically look at the vulnerability of power grids to natural disasters using a graph-theoretic framework for the grid. In contrast, our paper looks at the generation aspect of the problem where, unlike transmission lines, there may not be a direct physical connection between different power plants.

3.3 Dependence Framework for Outages and Recoveries

We adapt the subordination framework to the specific problem of modeling dependent outages for electricity power plants. We first define the stochastic behavior of each power plant and then introduce dependent outages using dependence sets.

3.3.1 Continuous Time Markov Chain model

We model the operational state of each electricity power plant with a Continuous Time Markov Chain (CTMC), which we denote by $X := \{X(t), t \geq 0\}$. For simplicity, we consider only two states for each power plant: we say that $X(t) = 1$ ("ON"), if the power plant can operate in full capacity at time $t \geq 0$; and $X(t) = 0$ ("OFF"), when the power plant does not operate, or operates sub-optimally; i.e., at less than 100% available capacity.[3]

Dependent Transitions

To induce dependence of the transitions, we introduce the concept of *dependence sets*, $\mathbf{D}_j := \{D_j^1, D_j^2, \ldots, D_j^N\}$, $j = 1, \ldots, l$. The number of dependence sets, l, corresponds to the number of factors of common outages. The indicator, D_j^i, is a binary variable that indicates whether plant $i = 1, \ldots, N$ is affected by common factor j. For example, let Λ_j be the set of power plants with the same primary fuel. We associate plants having the same primary input fuel with a common cause of outages, the indicator for the dependence set for primary fuel j is given by:

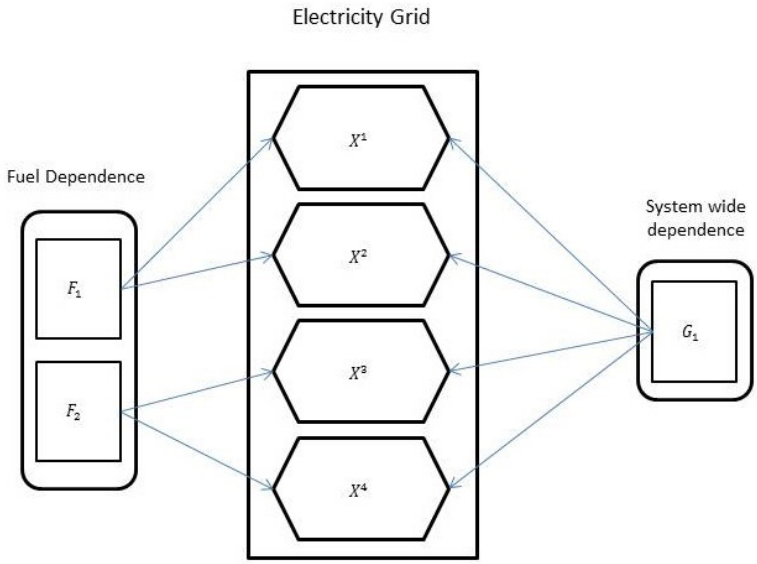
$$D_j^i = \begin{cases} 1 & \text{if } i \in \Lambda_j, \\ 0 & \text{if } i \notin \Lambda_j, \ i = 1, \ldots, N. \end{cases} \tag{3.1}$$

In addition to defining a dependence set for each input fuel, we capture dependence among all plants in a region by defining a dependence set for the entire region with all elements equal to 1. Our model can simultaneously handle multiple common

[3]The framework can be extended to accommodate additional states of operation without drastically affecting numerical tractability.

factors of outages based on fuel, geography, etc., through multiple dependence sets. Figure (3.1) provides an example for a system with 4 plants and 3 dependence sets. Each plant is affected by 2 common factors: a system-wide factor and a fuel specific factor.

Figure 3.1: Dependence Sets



The state, X^i, of each power plant, i, may exhibit fuel-specific dependence and system-wide dependence. For instance, power plant X^1 belongs to the dependence sets for the first input fuel factor and the system-wide factor.

With each dependence set, we associate a subordinator. The theory of subordinated Markov chains, introduced in chapter 2 shows that subordination is an effective mechanism to induce dependence. Let the ϕ_j be the Laplace transform of the subordinator associated with dependence set j. Then, the infinitesimal generator

for outages caused by common factor j is given by the $2^N \times 2^N$ matrix:

$$\mathbf{Q}_j = \phi_j \left(\oplus_{i=1}^{N} \mathbf{Q}_j^i D_j^i \right).$$

It is possible to allow for both independent, idiosyncratic, transitions and dependence through common factors. The infinitesimal generator for a system with N power plants and l common factors, with both independent and dependent transitions is given by

$$\mathbf{Q} = \oplus_{i=1}^{N} \mathbf{Q}^i + \sum_{j=1}^{k} \phi_j \left(\oplus_{i=1}^{N} \mathbf{Q}_j^i D_j^i \right),$$

where the matrices $\mathbf{Q}^i$ capture idiosyncratic transitions, while the functions ϕ_j induce dependence among power plants with common factor j. Since the transformation ϕ preserves the infinitesimal generator property, $\mathbf{Q}$ is also an infinitesimal generator.

We associate each dependence set with a specific subordinator. Hence plants belonging to each dependence set exhibit dependence in outages.

Simulation

Although subordination is a parsimonious way to induce dependence between CTMCs, it is difficult to compute the transition probabilities in closed form. Nonetheless, subordinated CTMCs are computationally tractable because they are easy to simulate.

The simulation of a subordinated CTMC involves three stages. We describe each of these stages here using the Algorithms described in Chapter 2.

First, we generate sample paths of the subordinators; i.e., a set of jump times and jump sizes for each subordinator. Second, given a particular subordinator path, during the interval between jump times each component, i, of the CTMC evolves independently according to the idiosyncratic infinitesimal generators $\mathbf{Q}^i$. Third, at times when the subordinator associated with the dependence set j jumps, each component, i, of the CTMC evolves simultaneously for a period equal to the jump size, according to the infinitesimal generator $\mathbf{Q}^i_j D^i_j$.

This procedure allows us to simulate the state of every power plant consistent with the dependence structure captured by the various dependence sets.

3.4 Adapting the framework to a system of power plants

We make two assumptions, which we subsequently verify empirically. First, outages can exhibit three behaviors: (a) they can be independent; (b) they may depend on an input fuel factor; or, (c) they may exhibit system-wide dependence. Second, we assume that recoveries occur independently.

Independent Transitions

We allow independent outages as well as recoveries by associating an infinitesimal generator, $\mathbf{Q}^i$, to each power plant i. For tractability, we assume that the parameters of the infinitesimal generator $\mathbf{Q}^i$ are the same for all power plants with the same fuel. The infinitesimal generator associated with independent transitions for

power plant i, with fuel j, is given by

$$\mathbf{Q}^i = \begin{pmatrix} -\alpha_j & \alpha_j \\ \beta_j & -\beta_j \end{pmatrix}, i = 1, \ldots, N, \tag{3.2}$$

where j is the primary input fuel for plant i and N is the number of power plants.[4] Note that since α_j and β_j correspond to transition intensities, independent transitions can occur both for outages as well as recoveries.

Dependent Outages

We assume that dependence in outages is based on two factors: input fuel, and a system-wide factor. Input fuel dependence is captured by a dependence set for each fuel. The infinitesimal generators associated with each fuel dependence set are given by

$$\mathbf{Q}^i_j = \begin{pmatrix} 0 & 0 \\ \gamma_j & -\gamma_j \end{pmatrix}, \quad j = 1, \ldots, k, \quad i = 1, \ldots, N, \tag{3.3}$$

where j is the primary input fuel for plant i, k is the number of input fuels, and N is the number of power plants. The form of the infinitesimal generators — in particular setting the parameters for the transition intensity from the "OFF" to the "ON" state to zero — implies that dependence among plants of the same fuel is assumed only in outages. The parameter γ_j captures the sensitivity of the probability of outages for plants with input fuel j to the shocks associated with the common factor.

We allow for system-wide dependence in outages by using a dependence set that includes all the power plants. Similar to the case of fuel dependence, we assume

[4] To fix notation, we set the first row and column to correspond to the "OFF" state.

that system-wide dependence exists only in outages. The infinitesimal generator associated with system-wide dependence is given by

$$\mathbf{Q}_{k+1}^{i} = \begin{pmatrix} 0 & 0 \\ \gamma_j & -\gamma_j \end{pmatrix}, i = 1, \ldots, N, \tag{3.4}$$

where j is the primary input fuel for plant i. The functional form of the infinitesimal generators for system-wide dependence assumes that the sensitivity to system-wide shocks for each power plant is the same as the sensitivity to the input fuel factor. This assumption is made for reasons of tractability — to reduce the overall number of parameters. This choice implies that, for any plant, the infinitesimal generators for the fuel-based and the system-wide factors can be combined.

We assume that the jumps of the subordinator for factor j, are discrete. The times between the jumps are given by a Poisson process with arrival rate c_j. We also assume that the jump sizes are exponentially distributed. The functions that capture dependence among power plants in the same dependence set, ϕ_j, are the Lévy-Khintchine exponentials of the subordinator

$$\phi_j(\lambda) = \frac{c_j \lambda}{\lambda + 1}, j = 1, \ldots, k+1. \tag{3.5}$$

The coefficients c_j capture the degree of dependence within each dependence set. We note that fuel-based dependence sets and the system-wide dependence set have the same infinitesimal generators for each plant of the same fuel. However, the intensity of outages due to a fuel-specific factor, $c_j, j = 1, \ldots, k$, can be different from the intensity of outages due to the system-wide factor, c_{k+1}.

3.5 Calibration Methodology

For any region with N power plants, we calibrate the model parameters iteratively. We start by calibrating model parameters for plants with the same input fuel, and then calibrate the residual system-wide dependence.

To define the optimization objective, consider a set of N power plants and n time intervals. We denote by μ_i the number of intervals during which i plants switch from state "ON" to state "OFF" (outages), and by ν_j the number of intervals during which j plants switch from state "OFF" to state "ON" (recoveries). Given a set of model parameters Θ, we define the model-expected number of intervals with i outages as μ_i^Θ, and the model-expected number of intervals with j recoveries as ν_j^Θ.

Rather than base our calibration on maximum likelihood estimation using the history of the state of every plant, we instead minimize a measure based on the system-wide Pearson χ^2 metric.[5] The measure is given by

$$\min_\Theta H(\Theta) := \sum_{i=0}^{N} \frac{(\mu_i - \mu_i^\Theta)^2}{\mu_i^\Theta} + \sum_{i=0}^{N} \frac{(\nu_i - \nu_i^\Theta)^2}{\nu_i^\Theta} \tag{3.6}$$

Given our assumption of independent recoveries, we can match the empirically observed average time between an outage and a recovery, by setting the transition rate α_j appropriately. In particular, α_j can be set to equal the inverse of the empirically observed average time between an outage and a recovery for plants with

[5] If the residuals are normally distributed χ^2 and maximum likelihood are equivalent. Berkson (1980) argues that minimizing χ^2 is preferable in many cases. Using a χ^2 metric increases the relative weight associated with multiple outages. Since the model parameters determine the frequency of multiple outages, this choice allows for a stable and efficient calibration.

input fuel j. We optimize the objective for the remaining parameters; i.e., choose the fuel-specific parameters $\beta_j, \gamma_j, c_j,\ j = 1, \ldots, k$, and the system-wide subordinator parameter that captures the intensity of system-wide outages, c_{k+1}, using a simulation optimization method developed in Yang and Tompaidis (2013). For each set of potential values of the parameters we use simulation to estimate the model-expected number of transitions and then calculate the corresponding value of the objective. As with any simulation optimization algorithm, the objective is calculated with sampling error: larger sample sizes reduce variability but are more costly computationally. The method developed in Yang and Tompaidis (2013) balances computational cost and accuracy. Briefly, the method starts with a search area in the parameter space and progressively decreases its size based on the sampled values of the objective within the area, occasionally backtracking to the previous search area and increasing the sample size if the objective deteriorates. Assuming that a non-degenerate solution that optimizes the objective exists within the search area, Yang and Tompaidis (2013) show that this procedure will ensure convergence.

We carry out the optimization in two steps. First for each input fuel we find the parameters that optimize the objective function for transitions of plants using that fuel. If the χ^2 statistic associated with the outages of plants of a particular input fuel is not statistically significantly different from zero at a confidence level of 95%, we eliminate the dependence set for that particular input fuel. Second, we estimate the parameters for the system-wide dependence factor by optimizing the objective for transitions of all the power plants in the region.

3.6 Data

We use a unique dataset we obtained from NERC. The dataset includes an exhaustive history of events at power plants. Owners are required to report incidents for every plant in Canada and the United States (excluding Alaska and Hawaii) with capacity above 50 MWs. Several plants from other regions of the world, as well as plants with smaller capacities, voluntarily report to NERC.

We illustrate our framework for the ERCOT and the WECC regions. These regions are interconnections, i.e., they are separate areas across which power plants share resources, generation capacities and loads.

The dataset covers the period 2010-2012. In that period there are 1815 power plants in the data in the ERCOT and the WECC. The dataset includes a unique identifier for each plant, the region and the state the plant is located, the input fuel for the plant, the generating capacity for the plant, the start time and the end time of each event, a cause code and a description for each event, and a measure of the operational capacity of the plant during the event. The causes listed range from reserve shutdowns, when a plant does not operate due to lack of demand, to shortages of fuel, to plant failures due to mechanical, operational, and/or weather related factors. The total number of cause codes is 892. Additional information is provided in Curley (2006) .

We discretize the 2010-2012 period into 4384 six-hour windows and consider only events that cause a forced shutdown of the plant.[6] A plant is considered to be

[6]A forced shutdown is defined by NERC as an event where the plant has to cease operations

Table 3.1: Outages by Region

Region	# Plants	Avg. # Outages	Avg. Capacity (MW)
ERCOT	204	18.0	235.5
FRCC	248	12.4	219.5
MRO	502	13.2	107.2
NPCC	839	27.2	103.4
RF	1397	23.1	159.4
SERC	1350	13.5	180.0
SPP	318	15.4	171.5
WECC	1611	16.4	108.1

This table lists the summary statistics for outages of electricity power plants by region. # Plants is the number of plants in the data for each region, Avg. # Outages is the average number of six-hour periods that plants in each region are in the "OFF" state out of the 4384 possible six-hour intervals, and Avg. Capacity (MW) is the average capacity of plants in each region measured in MegaWatts. The regions are: Electricity Reliability Council of Texas (ERCOT); Florida Reliability Coordinating Council (FRCC); Midwest Reliability Organization (MRO); Northeast Power Coordinating Council (NPCC); Reliability First (RF); SERC Reliability Corporation (SERC); Southwest Power Pool, Inc. (SPP); Western Electricity Coordinating Council (WECC).

in the "OFF" –outage– state if for any period of time during the six hour window, it is under a forced shutdown. It is considered to be in the "ON" state otherwise.

Tables (3.1) and (3.2) present summary statistics for the power plants and outages. The tables include the average number of six-hour intervals that the plants were "OFF" as well as the number of plants and average plant capacity.

Table (3.1) shows that the ERCOT region has fewer plants with larger capacity, while the WECC region has a large number of smaller plants. The average number of forced shutdowns over the three year period per plant, is 18.0 for the

before the end of the upcoming weekend.

Table 3.2: Outages by Fuel

Fuel	# Plants	Avg. # Outages	Avg. Capacity (MW)
CC	1009	27.8	318.3
DI	536	9.9	40.0
GG	2488	15.5	142.9
GE	37	12.7	37.4
NU	90	6.1	1002.1
OO	125	33.9	270.5
WA	1973	15.9	50.5
WH	186	19.1	187.6
KE	56	14.4	40.8
OG	49	143.6	52.2
Other	90	19.4	47.4

This table lists the summary statistics for outages of electricity power plants by input fuel. # Plants is the number of plants in the data for each primary fuel, Avg. # Outages is the average number of six-hour periods that plants of each primary fuel are in the "OFF" state out of the 4384 possible six-hour intervals, and Avg. Capacity(MW) is the average capacity of plants of each input fuel measured in MegaWatts. Fuel Codes are: Coal (CC); Distillate Oil (DI); Natural Gas (GG); Nuclear (NU); Other Oil (OO); Water (WA); Waste Heat (WH); Kerosene (KE) and Other which includes fuels such as Jet Fuel, Solar, Sludge Gas and Wind.

ERCOT and 16.4 for the WECC.

Table (3.2) shows the same statistics by input fuel. Natural gas, hydroelectric energy, and coal are the most common input fuels, while nuclear and coal plants are the largest. Nuclear plants have the fewest number of outages, while other gas, oil, and coal plants have the largest number of outages.

Table (3.3) presents the observed average outage and recovery times for plants of different input fuels. The table shows that the recovery times for all plants are, on average, short, while the times between outages are relatively long. Both the average time between outages and the average time between recoveries across regions

Table 3.3: Average Transition times

	ERCOT		WECC	
Fuel	Recovery	Outage	Recovery	Outage
CC	1.23	172.2	1.22	120.9
GG	1.24	301.1	1.27	291.6
NU	NA		1.10	750.3
WA	NA		1.21	369.0
WH	1.27	406.9	1.30	282.5
DI	NA		1.30	847.1
GE	NA		1.36	423.2
OG	NA		1.36	1010.1

This table lists the average transition times for plants of the same input fuel for the ERCOT and the WECC regions. In cases where there are fewer that five plants in a region, we do not report any results. Fuel is the input fuel of the plant. Fuel codes are described in the caption to table (3.2). Outage is the average time between a recovery and an outage measured in number of six-hour intervals. Recovery is the average time between an outage and a recovery measured in number of six-hour intervals.

are comparable for plants with the same input fuel.

3.7 Dependence in the ERCOT and the WECC regions

We present the results of the calibration in table (3.4) for the ERCOT region and in table (3.5) for the WECC region. The tables show, side-by-side, the parameter estimates, first assuming independent transitions, and then allowing for dependence through dependence sets. The p-values, reported for each primary fuel, correspond to the χ^2 test for the empirical vs. model outages and recoveries for each case.

Table 3.4: Parameter Estimates for the ERCOT region

		Independent				Subordinated						
Fuel	# Gen	α^f	β^f	p-value (Outages)	p-value (Recoveries)	α^f	β^f	γ^f	C^f	C^g	p-value (Outages)	p-value (Recoveries)
GG	163	1.74 (0.0037)	0.01 (0.0002)	0	0	1.74 (0.0215)	0.004 (0.0002)	0.01 (0.0019)	0.28 (0.038)	0	1	1
CC	23	1.70 (0.012)	0.01 (0.001)	1	1				NA			
WH	15	1.65 (0.015)	0.01 (0.005)	1	1				NA			

This table lists the results of calibration for the ERCOT region. The Fuel and # Gen columns are the fuel code and number of electricity plants in the region respectively. The α^f and β^f columns in the Independent section of the table are model parameters assuming independence in outages. The columns p-value (Outages) and p-value (Recoveries) in the Independent section correspond to the p-values of the null hypotheses that the outages and recoveries are independent. The Subordinated section reports parameter estimates when outages are modeled with a subordinated process. The columns α^f, β^f and γ^f are the infinitesimal generator parameters for the fuel factor in the subordinated case and C^f and C^g are the subordinator parameters for the fuel and system-wide factors. The columns p-value (Outages) and p-value (Recoveries) in the Subordinated section correspond to the p-values of the null hypotheses that the outages and recoveries are explained by the subordinated process. The values in parentheses, wherever applicable, are the bootstrapped standard deviations of the estimates. For Fuel Codes, refer to the caption of table (3.2).

In the ERCOT region, table (3.4) shows that natural gas (GG) is the only fuel where the evidence is consistent with dependence in outages. There is no evidence that a system-wide factor exists. Our calibration results are consistent with the outage patterns in the data. During the period we study, in the ERCOT region, outages involving 7 or more plants occurred 20 times. In one case, in January 2010, 10 plants went down due to shortage of natural gas; in another case, where 7 plants went down, the cause listed in the data was a cold wave resulting in frozen pipelines, which also affected the natural gas plants.

Table 3.5: Parameter Estimates for the WECC region

		Independent				Subordinated						
Fuel	# Gen	α^f	β^f	p-value (Outages)	p-value (Recoveries)	α^f	β^f	γ^f	C^f	C^g	p-value (Outages)	p-value (Recoveries)
CC	94	1.78 (0.036)	0.02 (0.0003)	1	1				NA			
DI	18	1.75 (0.18)	0 (0.0004)	0	0	1.75 (0.11)	0 (0.0003)	0.06 (0.024)	0 (0.018)	0.01 (0.006)	1	1
GE	37	1.39 (0.085)	0 (0.0003)	0	0	1.39 (0.07)	0 (0.0006)	0.02 (0.008)	0.15 (0.076)	0.01 (0.006)	1	0.89
GG	509	1.62 (0.022)	0.01 (0.0001)	0	0	1.62 (0.020)	0 (0.0001)	0.01 (0.0006)	0.47 (0.038)	0.01 (0.006)	1	1
NU	6	2.53 (0.492)	0.0 (0.0007)	0.4	0.58				NA			
OG	6	1.43 (0.248)	0.0 (0.0005)	0	0	1.43 (0.23)	0 (0.0002)	0.12 (0.06)	0 (0.0145)	0.01 (0.006)	0.97	1
WA	879	1.76 (0.0182)	0.01 (0.0000)	0	0	1.76 (0.021)	0 (0.0)	0.01 (0.0005)	0.47 (0.035)	0.01 (0.006)	1	1
WH	56	1.54 (0.042)	0.01 (0.003)	1	1				NA			

This table lists the results of calibration for the WECC region. The Fuel and # Gen columns are fuel codes and number of electricity plants in the region respectively. The α^f and β^f columns in the Independent section of the table are model parameters assuming independent outages. The columns p-value (Outages) and p-value (Recoveries) in the Independent section correspond to the p-values of the null hypotheses that the outages and recoveries are independent. The Subordinated section reports estimates when outages are modeled with a subordinated process. The columns α^f, β^f and γ^f are the infinitesimal generator parameters for the fuel factor in the subordinated case and C^f and C^g are the subordinator parameters for the fuel and system-wide factors. The columns p-value (Outages) and p-value (Recoveries) in the Subordinated section correspond to the p-values of the null hypotheses that the outages and recoveries are dependent. The values in parentheses, wherever applicable, are the bootstrapped standard deviations of the estimates. For Fuel Codes, refer to the caption of table (3.2).

Table (3.5) shows the results of the calibration for the WECC region. In this case the evidence is consistent with dependence in outages for five different fuel types. Dependence exists for power plants using natural gas, distillate oil, other gas, geothermal energy and hydroelectric energy. The size of the coefficients for plants using natural gas and hydroelectric energy shows that dependence in those cases is stronger. Specific instances of outages confirm the calibration results: for hydroelectric plants there are multiple instances of outages due to "dirty water," a common cause for many hydroelectric plants. There are also outages of multiple

plants due to freezing temperatures which causes dependence across the entire region. This is confirmed in our calibration, where we find evidence consistent with system-wide dependence.

We validate our calibration results in three ways. First, we compare the model estimates for average time to recovery and outage for each primary fuel and each region. We find that our estimates largely match the observed values in table (3.3). The biggest discrepancy is for the case of plants that use distillate oil in the WECC region with differences of up to 6%.

Second, we estimate the sampling error via bootstrapping. Using the calibrated values of the parameters, we simulate scenarios of outages and recoveries for the same number and type of power plants, and the same number of periods as in the data. For each simulated scenario, we re-estimate the values of the parameters. The bootstrapped standard deviation of the parameter estimates are reported in tables (3.4) and (3.5). We find that, in most cases, the variance of the bootstrapped, estimated, parameters is small, indicating that the sampling error is small and the calibration robust. We do note that the variance of the system-wide subordinator is high in the WECC region indicating that the evidence for system-wide dependence is weak.

Finally, in figure (3.2), we compare the empirical frequency of multiple shutdowns to the frequency calculated using the calibrated values from our model. While a model with independent outages predicts close to a zero probability of these rare events, we see that a subordinated model that accounts for dependence of outages, predicts multiple outages much more accurately for both regions.

3.8 Policy Implications

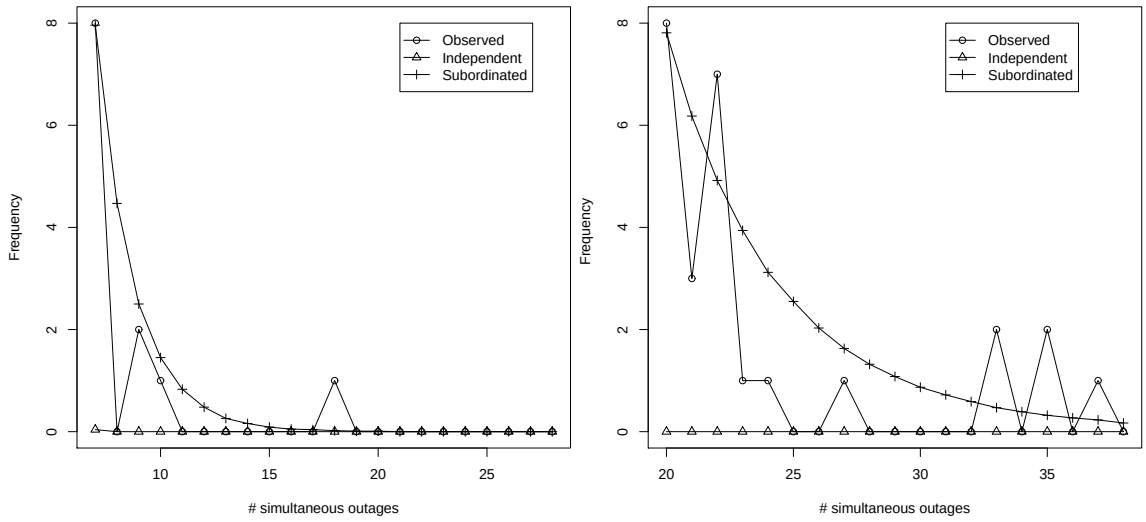
Our framework allows the calculation of several measures relating to system reliability, accounting for dependence between plant outages. We illustrate the policy implications of accurately accounting for dependence in three ways: (a) by determining the difference in available capacity at a given reliability level with and without outage dependence; (b) by estimating how the outage probability of marginal addition to capacity depends on the state of the system; and, (c) by estimating the amount of additional generation required to ensure the availability of a certain level of capacity at a given level of reliability.

Impact of dependence on reliability

To evaluate the impact of dependence on the available generation capacity, we construct reliability curves under both dependent and independent outages. The reliability curve represents the minimum amount of generation available a given percent of the time.

Figure (3.3) shows the reliability curves for the ERCOT and the WECC regions. In both regions, the available capacity is lower when outages are dependent. For example, in the ERCOT region, 97.1% of total capacity would be available 99% of the time, if outages were independent. Accounting for dependence reduces the available capacity by 0.2%; i.e., to 96.9%, of total capacity. This difference corresponds to a gap of 100 MWs. The gap grows when we consider the amount of capacity available 99.9% of the time to 300 MWs, or 0.6% of total capacity.

Figure 3.2: Outage Frequencies for the ERCOT and WECC regions

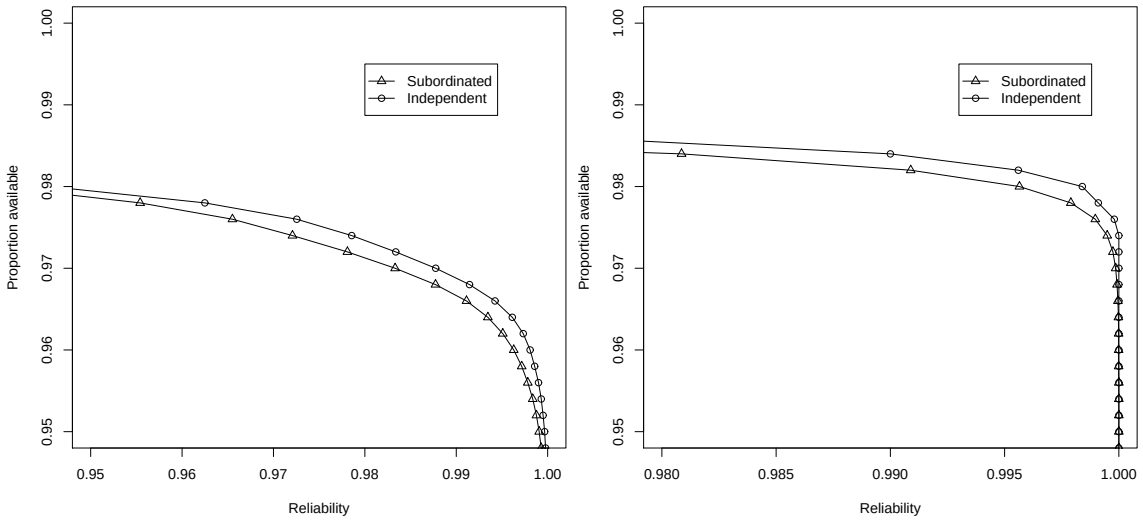


(a) # simultaneous outages - ERCOT (b) # simultaneous outages - WECC

The above figure shows the frequency of a given number of outages for the ERCOT (left panel) and WECC (right panel) regions. The horizontal axis corresponds to the number of simultaneous outages, while the vertical axis corresponds to their frequency; e.g., over the period in the data there is one six-hour window with 10 simultaneous outages in the ERCOT region.

The difference is more pronounced in the WECC region, where dependence in outages is stronger. If outages were independent, at least 98.4% of total capacity would be available 99% of the time, and at least 97.8% of total capacity would be available 99.9% of the time. Accounting for dependence, the available capacity is reduced to 98.2%, or by 600 MWs, and 97.6%, or by 350 MWs, respectively.

Figure 3.3: Reliability Curves



(a) ERCOT (b) WECC

The figure shows the supply available at different reliability levels. The horizontal axis represents the reliability level, while the vertical axis represents the proportion of generation available. The left panel corresponds to the ERCOT region and the right panel to the WECC region.

Marginal additions to capacity

When outages are dependent, the probability that a power plant will be available depends on the state of the system. Accounting for this dependence is important when we consider adding a new plant to the generation mix of a region.

In table (3.6) we show how the probability of outage of a new plant depends on the proportion of available, existing, capacity for the ERCOT and the WECC regions. The figure shows that the probability of outage increases when a large proportion of the existing capacity is unavailable. The difference can be large. For the ERCOT region, for natural gas plants, the probability of outage of a new natural

Table 3.6: Marginal additions to capacity

Panel A: ERCOT

Fuel	Independent	Subordinated					
		≤95%	96%	97%	98%	99%	≥ 99%
GG	0.4%	4.5%	3.3%	2.5%	1.7%	0.7%	0.3%

Panel B: WECC

Fuel	Independent	Subordinated					
		≤95%	96%	97%	98%	99%	≥ 99%
DI	0.1%	18.2%	18.2%	18.2%	18.2%	18.2%	0.2%
GE	0.3%	21.5%	21.2%	20.9%	20.2%	19.5%	0.3%
GG	0.4%	9.3%	8.8%	8.4%	7.9%	7.3%	0.4%
OG	0.1%	25.3%	25.5%	25.5%	25.7%	25.9%	0.2%
WA	0.3%	8.9%	8.5%	8.2%	7.7%	7.1%	0.3%

This table lists the outage probability of a marginal addition to generation capacity of different fuels as the available, existing, generation capacity varies for the ERCOT (Panel A) and the WECC (Panel B) regions. The Fuel column is the fuel code. The Independent column represents the outage probabilities assuming the outages are independent. In the Subordinated section, each column represents the proportion of available generation capacity. For the ERCOT region, we only present results for natural gas (GG), while for the WECC region we present results for natural gas (GG), geothermal energy (GE), distillate oil (DI), hydroelectric energy (WA) and Other gas (OG).

gas plant is 0.3% when more than 99% of existing capacity is available, but increases to 4.5% when only 95% of existing capacity is available. In the ERCOT region, only natural gas plants are affected because these plants exhibit dependence in outages. In the WECC region, where there is evidence of dependence for multiple input fuels, as well as for a system-wide factor, the effect is more pronounced. For example, the probability of outage of a natural gas plant is 0.4% when 99% of the existing capacity is available but increases to 9.3% when 95% of the existing capacity is available. The case of geothermal energy is even more extreme, where the outage probability increases from 0.3% to 21.5% respectively.

Capacity Expansion

Another criterion in the planning of capacity expansion is the amount of additional generation required to achieve a certain amount of capacity a given percent of the time.

We use our model to estimate this amount by adding power plants of a single input fuel. Table (3.7) shows the results for the ERCOT region for natural gas plants. Setting a target of having at least 47 GWs across the region available 98.5% of the time requires an additional 1406 MWs of natural gas generation accounting for dependence in outages.[7] Perhaps surprisingly, the additional amount of generation needed when outages are assumed independent is higher, at 1435 MWs. This result illustrates that dependence in outages does not imply that more capacity is needed to achieve a certain level of reliability. The results in table (3.8), for the WECC region, reinforce this point. While, for example, to have at least 172.5 GWs available across the region 98.5% of the time, 1373 MWs of new natural gas capacity are needed when accounting for dependence in outages vs. 1496 MWs when assuming independence, the situation is reversed when considering hydroelectric plants. In that case 1544 MWs are needed accounting for dependence in outages while only 1433 MWs are needed when we assume independence.

The results in tables (3.7) and (3.8) illustrate that accounting for dependence can have unpredictable results. The additional capacity needed to achieve a target

[7]Recall that our data only includes plants that have capacity of at least 50 MWs. Total generation capacity in both the ERCOT and the WECC regions is higher than the total capacity in our data.

Table 3.7: Capacity Expansion - ERCOT

Fuel	Model	Capacity Target: 47 GW		Capacity Target: 48 GW	
		99.5%	98.5%	99.5%	98.5%
GG	Independent	1779	1435	2781	2437
	Dependent	1702	1406	2726	2370

This table lists the additional capacity needed in natural gas plants in the ERCOT region to achieve a target of available capacity of at least 47 GW and 48 GW across the region, 98.5% and 99.5% of the time. The additional capacity needed is given in MWs under models of independent and dependent outages.

level is a complex function that depends on the average capacity of plants, the target chosen, and the degree of outage dependence. The results suggest that accurately accounting for dependence and being able to evaluate different reliability measures for proposed expansion plans is critical.

Table 3.8: Capacity Expansion - WECC

Fuel	Model	Capacity Target: 172.5 GW		Capacity Target: 175 GW	
		99.5%	98.5%	99.5%	98.5%
DI	Independent	1916	1412	4453	3911
	Dependent	1543	1101	4068	3613
GE	Independent	1950	1409	4485	3948
	Dependent	1540	1101	4073	3613
GG	Independent	2003	1496	4632	3998
	Dependent	1926	1373	4474	3908
OG	Independent	2099	1374	4607	4092
	Dependent	1549	1100	4095	3624
WA	Independent	1943	1433	4509	4000
	Dependent	2224	1544	4786	4097

This table lists the additional capacity needed in natural gas plants in the WECC region to achieve a target of available capacity of at least 172.5 GW and 175 GW across the region, 98.5% and 99.5% of the time. The additional capacity needed is given in MWs under models of independent and dependent outages. or Fuel Codes refer to the caption of t able (3.2).

Chapter 4

Facility Location Problem with Joint Disruptions

4.1 Introduction

The classical facility location problem is a widely studied problem in the supply chain literature. It seeks to choose a set of locations to commission facilities, say warehouses, that can serve a fixed set of client locations. By increasing the number of warehouses, one can decrease the cost of servicing customers (like transportation) at the expense of increasing setup costs (like real estate, building etc.). The objective is to find the choice of warehouse locations that strikes the best balance between these two costs. Apart from its obvious relevance in helping locate facilities like warehouses, distribution centers, airports and transit terminals, the facility location problem also serves important pedagogical purposes in Operations Management. It is also becoming more relevant in other areas outside operations, like self-configuration in wireless sensor networks to determine the functions of the nodes (Frank and Römer (2007)), locating and storing data in different servers on the internet (Sen et al. (2016)), and in computational biology in the selection of a

subset of treatments to provide maximum utility for a population of targets (Dueck et al. (2008)).

The classical facility location problem does not incorporate the possibility of these facilities or warehouses getting disrupted. Modern supply chains are however susceptible to disruptions caused by factors like inclement weather or other social, political and economic events. Understandably, the resulting choice of facilities from the classical formulation can be far from acceptable. In most cases, the costs of disruptions are too significant to be ignored. For example, the impact of disruptions caused by hurricane Irma on lost sales is estimated at $1.45 billion – see Sharon Edelson (2017). Hurricane Harvey is estimated to impact 10% of all US trucking activity – Bonney and Cassidy (2017). With increasing effects of global warming, such events are predicted to be more frequent and have a greater economic impact in the future – Gledhill et al. (2013).

When disrupted, the facilities become inoperable and cannot service customers. Client locations have to then be serviced by other non-disrupted and more expensive facilities. When disruption possibilities are accommodated in the model, we seek to minimize the expected servicing costs plus the set up cost.

While it is easy to argue against the omission of disruptions in the facility location problem, it is much harder to choose the right probabilistic structure one should use to capture disruptions. Factors like weather or political events that cause disruptions mostly affect more than one location. Hence, disruptions in facilities, especially those that share common factors like geography, ownership, and suppliers exhibit dependence. The reliability of the resulting location choice, understandably,

ends up depending on our ability to specify or model the joint disruptions accurately. When joint disruptions occur, two locations that are affected by very similar and common factors end up becoming less likely to both be a part of our location choice. Incorporating dependent disruption possibilities in our analysis will favor the selection of locations that not only have lower disruption probabilities but also those with disruptions that are more independent. It has been pointed out in literature that accurate models for supplier disruptions should play a central role in designing supply chains – Saghafian and Van Oyen (2012).

There are two major challenges in choosing the joint probability distribution for disruptions. First, given that disruptions are rare events, data available for calibration are scarce. For example, take the standard benchmark case, used in Cui et al. (2010), Snyder and Daskin (2005), with 49 possible locations (the 48 contiguous states in mainland US and Washington D.C) and hurricanes as surrogates for disruptions. If we use data from the National Oceanic and Atmospheric Administration to model disruptions, we observe around 24,000 disruptive events over a span of 60 years. The number of joint probabilities that need to be calibrated with this data is around $2^{49} \approx 10^{15}$. Given the very limited data, estimation of all the joint probabilities becomes impossible without any structural assumption. Following this calibration challenge, comes the challenge of solving the formulated optimization problem. Facility location problems with disruptions are much harder to solve primarily because of the number of disruption scenarios to plan for. The choice of the joint distribution and its structure directly dictates the number and structure of these scenarios.

The calibration challenge and the optimization challenge have both been dealt with in prior literature by using parsimonious representations of disruptions that lend themselves well towards simplifying the optimization problem. They make up for their inability to accurately capture the dependence structure by offering significant computational benefits.

The two most popular assumptions for disruptions are the total independence assumption and the so called worst-case disruption model. The independent model needs estimates of only the marginal probabilities rather than the entire joint distribution. Further, to help solve the resulting optimization problem, an assumption about the structure of the optimal solution is made whereby each client location can only be serviced by a fixed number of facilities. This simplifies the optimization problem considerably. The worst-case disruption model, which is also referred to as the robust disruption model is also built using only the marginal disruption rates. Dependence is built by assuming that a riskier facility always fails when a less riskier facility fails. The number of possible scenarios reduce drastically under the robust distribution making the optimization easier. Lu et al. (2015) show that the solution yielded using this distribution model is the same as the solution to a robust (given marginals minimize worst case cost) version of the original problem. Hence it is understandable that the solution is extremely conservative and in test cases has been shown to be far from optimal. We discuss these assumptions in detail while reviewing the literature in the following section.

Contributions

In this research we propose the use of a new joint probability structure for modeling disruptions. The structure we propose is that of having the joint probabilities be dictated by subordinated Markov chains. Subordination is simply a way of taking independent stochastic processes and building a dependence structure in a parsimonious way. While subordination of Brownian motions are prevalent in finance (Clark (1973)), subordinated Markov chains have only recently been introduced to Operations Management literature (Malladi et al. (2017)).

We adapt the subordinated Markov chain framework developed in Chapter 2, to the context of the facility location problem. We also develop and present a calibration algorithm that can be used to deduce the dependence amongst locations and to estimate the respective parameters. The algorithm can calibrate the model using sparse data and also adapt itself to a more complex dependence structure as more data becomes available. Our calibration framework provides consistent estimates of the parameters. The framework is also computationally inexpensive even for problems with a large number of facilities.

Lastly, we develop a methodology to solve for the optimal choice of facility locations under the dependent disruptions modeled by the subordinated Markov chains. Our solution algorithm solves the benchmark case, in a reasonable amount of time and can be extended to problems with a larger number of locations.

In terms of insights and results, the research contributes to the literature by demonstrating the efficacy of this modeling approach and methodology using the

benchmark facility location problem. Using the data for hurricanes as surrogates for disruptions, we first calibrate our framework to the 49 locations (the 48 contiguous states in mainland US and Washington D.C.) of the United States to model the disruption probabilities. Each facility is associated with a single factor causing disruptions. Facilities that share a common factor, exhibit dependent disruptions. Our calibration of real life disruption data results in six unique factors that are the causes of disruptions amongst 49 facility locations. We find that, states in geographical proximity are associated with a common factor. This is reassuring for our model and the calibration algorithm, since hurricanes affecting a particular state, have a high probability of affecting neighboring states.

We also solve for the optimal choice of the facility locations under dependent disruptions dictated by the subordinated Markov chain framework. We find that our optimization algorithm results in a choice of locations that is different from the choice when the worst-case distribution is assumed for the disruptions. The overall cost of the optimal solution under the dependent distribution is around 9% lower than the overall cost of the optimal solution under the worst-case distribution.

Literature Review

Previous literature pertinent to our research can be classified into three categories; one, research into the reliable facility location problem; two, research in the modeling of dependent Bernoulli variables and; three, research into subordinated stochastic processes.

The review papers, Snyder (2006) and Melo et al. (2009), provide a detailed

overview of research into the facility location problem. Here we discuss the literature most relevant to our research. A possible model for the disruption probabilities is to assume independence among disruptions. This approach requires only the estimation of the marginal disruption rates of individual locations. Snyder and Daskin (2005) and Cui et al. (2010) use this approach, with the former assuming equal marginal probabilities and the latter assuming unequal marginal disruption probabilities. To solve for the optimal choice of facilities, both these papers impose a structure on the facility location choice, whereby each customer has a preference order of a fixed number of locations. The independence assumption and the special structure on the facility location choice allows them to restructure the model as a linear program. Using Lagrangian relaxation, they solve for the optimal facility location choice for the benchmark case and computationally show the tightness of the bounds for the optimal solution. Shen et al. (2011) also model disruptions as independent. However, they do no impose a structure on the optimal solution. They propose heuristic algorithms for the solutions that can produce near-optimal solutions for the problem. For the special case, when the probability that a facility fails is constant, they provide an approximation algorithm and a bound for the solution.

Lu et al. (2015) propose using a robust joint distribution to account for dependence in disruptions. The robust model requires only the marginal disruption rates and hence is parsimonious in its specification. The model assumes the highest possible degree of dependence between disruptions of locations. It ties in dependence by dictating that all riskier facilities fail when a safer (lower marginal probability) facility fails. This essentially equates to assuming there is only a single source and

location of disruptions with the only uncertainty being its magnitude. The robust distribution provides conservative solutions in the sense that it tends to place facilities at locations with lower marginal disruption rates. It also yields solutions with the highest possible cost for given marginal disruption probabilities because of the robust nature of the underlying probability distribution. This distribution has a few possible scenarios, and hence, optimization under this distribution is computationally easy.

Lim et al. (2010) formulate the facility location problem where each location can have a reliable or an unreliable facility. They assume the disruptions as independent and use the Lagrangian relaxation to solve for the optimal choice of facilities. They derive structural results for the problem and show that a threshold exists for the failure rates beyond which it is only optimal to commission reliable facilities and show that the solution algorithm proposed is capable of solving large-scale problems. Santoso et al. (2005) propose solving the facility location problem under disruptions using the sample average approximation method using Bender's decomposition algorithm and suggest various accelerating techniques for optimization. They computationally show the efficiency of the acceleration schemes over traditional solution methods. However, they do not specify a particular probability distribution to model the disruptions.

The problem of modeling joint Bernoulli variables has been widely studied in the literature. Cox (1972) provides a brief description of the various methods to model multivariate Bernoulli distributions. The problem in modeling joint Bernoulli variables is the large number of parameters required to completely describe the joint

distribution. The complete joint distribution of n Bernoulli variables requires 2^n parameters. Using parametric models for joint disruptions also requires a high number of variables. For example, Banerjee et al. (2008) propose a method where the joint distribution of n Bernoulli random variables is defined by the n marginals and the moments of the variables, taken two at a time. They propose algorithms to estimate the parameters and computationally show the efficiency of these algorithms. To our knowledge, joint Bernoulli variables have not been used to model dependent disruptions in the facility location problem.

Subordination of Lévy processes is a prevalent technique in finance, introduced in Clark (1973) to model asset prices – see, e.g., Mendoza-Arriaga et al. (2010) and Sun et al. (2013)). Subordinated Markov chains, are a sub-class of subordinated Lévy processes which can be used to model multiple Markov chains with dependent transitions with fewer parameters.

Subordinated Markov chains were used in an Operations Management context in Malladi et al. (2017) . They use them to model joint outages of electricity power plants using dependence sets to group the electricity plants into subsets with a common cause of outages. They assume the characteristics of the dependence sets are given and only estimate the marginals and the strength of the dependence from the data. This research does not involve any optimization of a system that uses such a joint probability model. In contrast to this, the calibration algorithm we propose here learns the dependence structure and all the relevant parameters together. Moreover, we provide and solve, to our knowledge, the first optimization problem built on such subordinated Markov chain dynamics, thereby making the case for the use

of such models and techniques in Operations Management.

The rest of the chapter is organized as follows. In section 4.2, we describe both the classical facility location problem and the modified facility location problem that allows for disruptions. In section 4.3, we adapt the model of partially subordinated Markov chains, developed in chapter 2 for the specific problem of modeling joint disruptions in the facility location problem. In section 4.4 we describe the algorithms to estimate the parameters of the subordinated Markov chains used to describe the disruption probabilities. We specify the metric we use to calibrate the subordinated Markov chains and the algorithms to estimate the strength of dependence, the number of subordinators and the association of individual subordinators to different locations. In section 4.5, we cover the algorithms to optimize the facility location choices under our disruption model. We describe the iterative approach we take to solve for the optimal choice of facilities using the solution to the classical problem. We discuss the results in section 4.6 for the benchmark case. We show that states that exhibit dependent disruptions are in geographical proximity which is consistent with our intuition. We also present the optimal choice of facilities for the benchmark case.

4.2 Model

In this section we present the formulation and the notation we use for the facility location problem. We first present the classical facility location problem that does not consider disruptions, also called the Uncapacitated Facility Location Problem (UFLP) – Erlenkotter (1978). The UFLP has been widely studied in the

literature and multiple solution techniques have been developed for the UFLP including, among others, Erlenkotter (1978), Khumawala (1973), Cornuéjols et al. (1983). We then introduce the facility location problem that considers the possibility of disruptions, called the reliable uncapacitated fixed-charge location problem (RUFLP) – Lu et al. (2015).

4.2.1 Uncapacitated Facility Location problem

Let $\mathcal{I} = \{1, 2, \ldots, I\}$ be the set of possible facility locations indexed by i and $\mathcal{J} = \{1, 2, \ldots, J\}$, the set of customers indexed by j. The cost of setting up a facility at location i is f_i and the cost of servicing customer j from the facility at location i is c_{ij}. The optimization problem is to minimize the total cost of set up and servicing while satisfying the demand at each customer location. Let the x_i, $i \in \mathcal{I}$ be the binary variables which take the value 1 if a facility is set up in location i and 0 otherwise and y_{ij}, $\forall\ i, j, i \in \mathcal{I}, j \in \mathcal{J}$ be binary variables which take the value 1 if the facility at location i serves customer j and 0 otherwise. Let x be the vector containing all x_i variables. The UFLP is a two stage problem:

$$\begin{aligned} \min\ & \sum_i f_i x_i + h(x) \\ \text{s.t } & x_i \in \{0,1\}\ \forall\ i \in \mathcal{I} \end{aligned} \qquad \begin{aligned} \text{where } h(x) = \min\ & \sum_{ij} c_{ij} y_{ij} \\ \text{s.t } & y_{ij} \leq x_i, \quad \forall\ i \in \mathcal{I}, j \in \mathcal{J} \\ & \sum_i y_{ij} = 1, \forall\ j \in \mathcal{J} \\ & y_{ij} \in \{0,1\}\ \forall\ i \in \mathcal{I}, j \in \mathcal{J} \end{aligned} \tag{4.1}$$

In the above formulation, $\sum_i f_i x_i$ is the total cost of setting up facilities, while $h(x)$ is the total cost to service all customers given a particular choice of facilities x.

$h(x)$, the optimal value of the second stage problem which solves for the minimum cost of servicing all customers given a choice of facilities x. The first constraint of the second stage problem ensures that a client is serviced only by a location where a facility is placed. The second constraint of the second stage problem ensures that every customer is serviced by some facility. Since the facilities are uncapacitated, in the second stage problem, every client is serviced by the closest available facility location.

4.2.2 Reliable Uncapacitated fixed-charge location problem

The Reliable Uncapacitated fixed-charge Location problem (RUFLP) incorporates potential facility disruptions and seeks to minimize the sum of set up costs and expected servicing costs. Disruptions are modeled by the random vector $\xi = \{\xi_1, \xi_2, .., \xi_I\}$ with each ξ_i being a binary variable taking the value 0 if location i is disrupted and 1 otherwise. The random vector ξ follows the probability distribution $\mathbb{P}$.

In the classical case described in section 4.2.1, each y_{ij} defines the allocation of locations to customers. However, in the RUFLP, assignment of facilities to customer locations depends on the choice of facilities x and the realization of random variable for disruptions ξ. A facility installed in location i cannot service customers if the facility is disrupted, i.e., $\xi_i = 0$. Hence, for each scenario ξ, we will use y_{ij}^{ξ} to represent the allocation variables under scenario ξ. y_{ij}^{ξ} is a binary variable which indicates whether facility i services customer j under scenario ξ. Moreover, in this case we add a penalty c_{0j} for not servicing client j, which is possible if all facilities

are disrupted and customer j cannot be serviced. This non-servicing penalty also allows clients to not be serviced when, under a disruption scenario, the servicing cost from the operational facilities are higher than the penalty cost. Just as the allocation variable for each scenario, y_{ij}^{ξ}, we introduce binary variables y_{0j}^{ξ}, which take the value 1 if client j is not serviced by any facility under scenario ξ and 0 otherwise. We will use $h(x, \xi)$ to represent the total servicing costs (including penalties, if applicable) under scenario ξ.

The RUFLP is set up as a two stage problem as follows:

$$\begin{aligned}
&\min \sum_i f_i x_i + \mathbb{E}_{\mathbb{P}}[h(x,\xi)] \\
&\text{s.t } x_i \in \{0,1\} \ \forall \ i \in \mathcal{I}
\end{aligned}
\qquad
\begin{aligned}
&\text{where } h(x,\xi) = \min \sum_{ij} c_{ij} y_{ij} + \sum_j c_{0j} y_{0j} \\
&\text{s.t } y_{ij}^{\xi} \leq \xi_i \ \forall \ i,j,\xi \\
&\quad y_{ij}^{\xi} \leq x_i \ \forall \ j,i,\xi \\
&\quad \sum_i y_{ij}^{\xi} + y_{0j}^{\xi} = 1, \forall \ j \in \mathcal{J}, \forall \xi \\
&\quad y_{ij}^{\xi} \in \{0,1\} \ \forall \ i,j,\xi, \ y_{0j}^{\xi} \in \{0,1\}, \ \forall \ \xi, j
\end{aligned} \tag{4.2}$$

The objective function in the first stage problem of the RUFLP has the sum of set up costs $\sum_i f_i x_i$ and the expected value of the servicing costs $\mathbb{E}_{\mathbb{P}}[h(x, \xi)]$ over all the possible scenarios. The second stage problem solves the minimum servicing costs for a given choice of facilities x and a given disruption scenario ξ. The first and the second constraint of the second stage problem ensure that an allocation from location i to customer j is only made when a facility is placed in location i and it is not disrupted under scenario ξ respectively. The third constraint ensures that under

every scenario, each customer is assigned a facility or a penalty is accrued.

It is clear that the optimal solution of the RUFLP, described in (4.2), is dependent on $\mathbb{P}$, the joint probability distribution of the disruptions. Hence, it is imperative to use an accurate distribution of the probabilities $\mathbb{P}$ before optimizing for the choice of the facilities. However, estimation of $\mathbb{P}$ using only the disruption data has two distinct problems.

First, the state space of the random variable ξ has 2^I different elements. For high values of I, handling a probability distribution with an exponentially high number of elements becomes mathematically inconvenient. Second, disruptions are inherently rare. Hence, limited data is available towards the calibration of the probability of different scenarios or the distribution $\mathbb{P}$. Due to the limited data, very few or no instances are observed of multiple facilities disrupting. Thus the probability of joint disruptions, as estimated from the data, will be zero or close to zero. However, as argued before, due to common factors of disruptions, joint disruptions occur and the non-zero probability of joint disruptions impacts the choice of optimal facility locations. Thus, naive models for $\mathbb{P}$ might result in sub-optimal choice of facilities.

We propose using subordinated Markov chains to model the disruption distribution $\mathbb{P}$. Subordinated Markov chains address both the above issues. Since subordinated Markov chains use limited parameters, they can be calibrated with limited data. They also result in non-zero probabilities for the joint disruptions. This ensures that the choice of facilities selected while using probabilities from the subordinated Markov chain framework account for the non-zero joint disruption probabilities.

4.3 Subordinated Markov Chain

In this section we adapt the subordinated Markov chain framework, developed in chapter 2, to model the probability of joint disruptions in the facility location problem.

4.3.1 Modeling disruption probabilities using subordinated Markov chains

The state of each location in the RUFLP is modeled as a partially subordinated Markov chain. As explained in chapter 2, this ensures that the disruptions among various locations are dependent. Each location, or Markov chain, is associated with a single subordinator. Due to the nature of weather patterns, disruptions of locations separated by geography do not show dependence while those in close proximity tend to exhibit dependence. For example, the disruptions in Florida might be dependent on the disruptions in South Carolina, and be independent of the disruptions in California. Hence, we use multiple subordinators to model the dependence in the Markov chains. Markov chains associated with the same subordinator exhibit dependence while Markov chains associated with different subordinators are independent. The probability distribution $\mathbb{P}$ used in the RUFLP is the steady state probability of this system of Markov chains.

We model each facility location as having two possible states "ON" and "OFF". Thus the related Markov chain has two states – "ON" and "OFF". A partially subordinated Markov chain is defined by two infinitesimal generators $\hat{Q}^i$ and $\tilde{Q}^i$ and a subordinator. For each location, we define $\hat{Q}^i$ as the infinitesimal generator for the recoveries and $\tilde{Q}^i$ as the infinitesimal generator for disruptions.

Disruptions are dependent and hence, for each location, only $\tilde{Q}^i$ is subordinated, and $\hat{Q}^i$ models the independent recovery of locations. Each location i has a recovery rate of α^i and a disruption sensitivity to subordinator of γ, i.e.,

$$\begin{aligned} \hat{Q}^i &= \begin{pmatrix} -\alpha^i & \alpha^i \\ 0 & 0 \end{pmatrix} \\ \tilde{Q}^i &= \begin{pmatrix} 0 & 0 \\ \gamma & -\gamma \end{pmatrix} \end{aligned} \tag{4.3}$$

In the above example, the first row and first column of the two infinitesimal generators $\tilde{Q}^i$ and $\hat{Q}^i$ correspond to the "OFF" state and the second row and second column correspond to the "ON" state. Notice that $\hat{Q}^i$ has a zero intensity of "OFF" to "ON" transitions while $\tilde{Q}^i$ has a zero intensity of "ON" to "OFF" transitions. Thus, $\hat{Q}^i$ only governs the recoveries and $\tilde{Q}^i$ only the disruptions. Subordinating only the $\tilde{Q}^i$ generators for different Markov chains results in dependent disruptions and independent recoveries.

We assume that each facility has the same level of reinforcements against the common shocks implying the sensitivity to subordinator shocks γ is equal for all facilities. We also model the recoveries as independent. Both these assumptions can be relaxed to include, different disruption intensities or dependent recoveries. The calibration algorithm in that case will be more involved and there is the risk of over-fitting data to the model.

We use multiple subordinators to model the dependence among the Markov chains. The set $\mathcal{K} = \{1, 2, \ldots, K\}$ denotes the K subordinators defining the system. We use the set of binary variables $\mathcal{R} = \{r_{ik}\}, i \in \mathcal{I}, k \in \mathcal{K}$ to define the subordinator

associations. The variables in the $\mathcal{R}$ are described by:

$$\begin{aligned} r_{ik} = &\begin{cases} 1, \text{if Markov chain } i \text{ is associated with subordinator } k, \\ 0 \text{ otherwise.} \end{cases} \\ &\sum_k r_{ik} = 1, \ \forall \ i, \text{and} \\ &r_{ik} \in \{0, 1\}, \ \forall \ i, k. \end{aligned} \tag{4.4}$$

The summation in (4.4) ensures that each Markov chain is associated with exactly one subordinator.

We now present the infinitesimal generator governing the dynamics of the individual Markov chains and the system of Markov chains. Let $\phi^k(.)$, $k \in \mathcal{K}$ be the Lèvy-Khintchine exponential of subordinator k. As shown in chapter 2, the infinitesimal generator, Q^i for Markov chain i is:

$$Q^i = \hat{Q}^i + \sum_k r_{ik} \phi^k(\tilde{Q}^i)$$

A subordinator is defined by the arrival rate of jumps C and the size of jumps η. The η parameter can be normalized to 1 due to degeneracy with the γ parameter of the infinitesimal generator of the Markov chain. Additionally, since we are working with only the steady-state probabilities, for this particular application, we can normalize the arrival rate of jumps parameter C to 1. Since all subordinators used in the system have the same jump arrival rate C and jump intensity parameter η, the Lèvy-Khintchine exponential of the subordinator $\phi^k(.), \forall \ k$ is the same. Consequently, we drop the index k for the Lèvy-Khintchine exponential and the infinitesimal generator Q^i of any Markov chain i is:

$$Q^i = \begin{pmatrix} -\alpha^i & \alpha^i \\ \phi(\gamma) & -\phi(\gamma) \end{pmatrix} \tag{4.5}$$

The steady-state probability that a Markov chain with an infinitesimal generator given by Equation (4.5) will be in the "OFF" state is:

$$P = \frac{\phi(\gamma)}{\phi(\gamma) + \alpha^i} \tag{4.6}$$

We will use the steady-state probability given by Equation (4.6) in the calibration of the α^i parameter.

All Markov chains linked to subordinator k exhibit dependence. Hence, as shown in chapter 2, the infinitesimal generator for the system of Markov chains associated with subordinator k is:

$$\bigoplus_{i|r_{ik}=1} \hat{Q}^i + \phi\Big(\bigoplus_{i|r_{ik}=1} \tilde{Q}^i \Big) \tag{4.7}$$

Finally, since Markov chains that are associated with different subordinators are independent, the infinitesimal generator for the system of Markov Chains, is simply the kronecker sum of the generators for the Markov chains with a common subordinator. The infinitesimal generator for the system is given by:

$$Q = \bigoplus_{k \in \mathcal{K}} \Big(\bigoplus_{i|r_{ik}=1} \hat{Q}^i + \phi\Big(\bigoplus_{i|r_{ik}=1} \tilde{Q}^i \Big) \Big) \tag{4.8}$$

In the following section, we discuss the calibration of the α,γ and $\mathcal{R}$ parameters given disruption data.

4.4 Calibration

Estimating the optimal facility location involves the calibration of $\mathbb{P}$ under the subordinated Markovian model and subsequently, solving the RUFLP using the estimated $\mathbb{P}$, which is modeled using subordinated Markov chains. To estimate the parameters of the subordinated Markov chains, we use disruption data, which includes the state of each Markov chain or location at different points in time.

The calibration of the subordinated Markov chain system involves the estimation of three sets of parameters; (i) The set $\mathcal{A} = \{\alpha^i\}$, $\forall i \in \mathcal{I}$ corresponding to the recovery rates of individual Markov chains, (ii) γ, the parameter corresponding to the sensitivity of the disruptions to the subordinator, and (iii) $\mathcal{R}$, the set of variables defining the allocation of each Markov chain to a subordinator. Estimation of $\mathcal{R}$ includes the estimation of K, the number of subordinators defining the system, which is not known a priori, and the allocation of individual Markov chains to the subordinators.

The calibration of the parameters is carried out in two phases. In the first phase, described in section 4.4.1, we fix the marginal disruption rates for individual Markov chains to equal the observed marginal rates. This fixes the α^i parameters of individual Markov chains and enables us to focus on calibrating the parameters $\{\gamma, \mathcal{R}\}$. Before proceeding to the next phase of the calibration, we define the metric we use for estimating γ and $\mathcal{R}$, which we do in section 4.4.2. In section 4.4.3, we describe the second phase of the calibration, the estimation of γ and $\mathcal{R}$.

4.4.1 Calibrating the Marginal probabilities

In this step, we calibrate the set of the parameters $\mathcal{A}$ as functions of the γ parameter. Let $P(\{\mathcal{A}, \gamma, \mathcal{R}\}, \{i\}), \forall\ i \in \mathcal{I}$ be the model marginal disruption rate for facility i. Let N_i be the number of instances in the data with facility i in the "OFF" state and N be the total number of observations in the data, or the instances for which we have the recorded state of the Markov chains. We fix the steady state marginal disruption probability for each facility i to N_i/N, i.e., using Equation (4.6), we parametrize α^i in terms of γ as:

$$\begin{aligned} P(\{\mathcal{A}, \gamma, \mathcal{R}\}, \{i\}) &= \frac{\phi(\gamma)}{\alpha_i + \phi(\gamma)} = \frac{N_i}{N} \\ \implies \alpha_i &= \frac{N_i}{N - N_i} \frac{\gamma}{\gamma + 1},\ \forall i \in \mathcal{I}, \end{aligned} \tag{4.9}$$

where the second equation follows from the arrival rate of jumps, C of all subordinators being equal to 1 and the closed form expression of $\phi(\gamma)$[1].

In this step we are only fixing the ratio of α_i to γ for each facility. Setting the values of $\mathcal{A}$ using Equation (4.9) ensures that the marginal disruption probabilities equal the observed marginal disruption probabilities for any choice of γ. Thus, given a value of γ, since the set $\mathcal{A}$ is fixed, we now need to calibrate $\{\gamma, \mathcal{R}\}$.

Although the marginal disruption rates have been fixed, the strength of dependence, measured by γ and the structure of dependence, manifested by the parameters

[1]For the class of subordinators the Lévy-Khintchine exponential for a subordinator with jump intensity 1 is:

$$\phi(\gamma) = \frac{\gamma}{\gamma + 1}$$

in the set $\mathcal{R}$ are yet to be calibrated. For the same marginal disruption rates, the dependence structure can be different based on the γ and $\mathcal{R}$ variables.

The next phase of the calibration is to estimate the γ and the $\mathcal{R}$ that best explain observed disruption data. After the α^i's are fixed in terms of the γ parameter, the disruption probabilities are functions of γ and $\mathcal{R}$. Before discussing the calibration of γ and $\mathcal{R}$, we present the metric using which we calibrate γ and $\mathcal{R}$.

4.4.2 Calibration metric

Given disruption data, a possible option for parameter estimation is maximizing the likelihood across all the 2^I possible scenarios. This option is computationally cumbersome due to the exponentially high number of scenarios and the lack of closed form expressions for the steady-state probabilities of different scenarios for a given choice of parameters. Thus, we propose an alternate metric that we minimize in order to calibrate the γ and $\mathcal{R}$ parameters.

Denote by $P(\{\gamma, \mathcal{R}\}, S), S \subset \mathcal{I}$, the steady-state probability of all Markov chains in set S are in the disrupted state given the subordinated Markov chain system parameters $\{\gamma, \mathcal{R}\}$. Note that the marginal rates of disruption are fixed and hence the α parameters are not included in the probability function. Let $N_{ij},\ i, j \in \mathcal{I}$ be the number of observations with both facilities i and j in the "OFF" state.

To calibrate the γ and $\mathcal{R}$, we solve the problem:

$$\{\gamma^{opt}, \mathcal{R}^{opt}\} = \text{argmin}_{\{\gamma,\mathcal{R}\}} T(\{\gamma, \mathcal{R}\}), \tag{4.10}$$

where

$$T(\{\gamma, \mathcal{R}\}) := \sum_{i,j} \left(\frac{N_{ij}}{N} - P(\{\gamma, \mathcal{R}\}, \{i, j\}) \right)^2 \; i, j \in \mathcal{I} \tag{4.11}$$

$$P(\{\gamma, \mathcal{R}\}, \{i, j\}) = \begin{cases} P_i P_j + (1 - P_i)(1 - P_j)[\frac{2\phi(\gamma) - \phi(2\gamma)}{\alpha_j + \alpha_i + \phi(2\gamma)}], \text{ if } r_{ik} = r_{jk} \; \forall k. \\ P_i P_j, \text{ otherwise} \end{cases}, \tag{4.12}$$

where

$$P_i = P(\{\gamma, \mathcal{R}\}, \{i\})$$

$$P_j = P(\{\gamma, \mathcal{R}\}, \{j\})$$

If the Markov chains i and j share a common subordinator, i.e., $r_{i,k} = r_{j,k}, \; \forall k$, the joint disruption probability of Markov chains i and j is given by first part of Equation (4.12) [2]. If the two Markov chains do not share a common subordinator, they are independent and thus the probability of a joint disruption is simply the product of the two marginal disruption probabilities. It is easy to see that, as expected, the

[2] If two Markov chains i and j share a subordinator, the infinitesimal generator of the joint Markov Chain is:

$$Q = \begin{pmatrix} -(\alpha_i + \alpha_j) & \alpha_i & \alpha_j & 0 \\ \phi(\gamma) & -(\alpha_j + \phi(\gamma)) & 0 & \alpha_j \\ \phi(\gamma) & 0 & -(\alpha_i + \phi(\gamma)) & \alpha_i \\ 2\phi(\gamma) - \phi(2\gamma) & \phi(2\gamma) - \phi(\gamma) & \phi(2\gamma) - \phi(\gamma) & -\phi(2\gamma) \end{pmatrix}$$

The joint disruption probability of Markov chains i and j is calculated by solving the equations:

$$\pi Q = 0$$

$$\sum \pi = 1$$

and substituting $\frac{\phi(\gamma)}{\phi(\gamma) + \alpha_i} = P_i$ and $\frac{\phi(\gamma)}{\phi(\gamma) + \alpha_j} = P_j$.

steady-state disruption probability of two independent Markov chains is lower than the disruption probability of two dependent Markov chains.

The metric $T(\{\gamma, \mathcal{R}\})$ is the sum of the square of the difference of the model predicted and observed disruption probabilities of two Markov chains across all possible combinations of Markov chains taken two at a time. Thus, the summation is over the $\binom{I}{2}$ terms corresponding to the number of ways two Markov chains can be selected from the I possible locations. Minimizing the metric matches the steady state joint disruption probability of two Markov chains to be as close as possible to that observed in the data. Considering the disruption probability of two locations at a time ensures that we capture the effect of dependence in the disruptions. The largest impact of joint disruptions comes from the probability of two Markov chains, or two facilities being disrupted. We capture this effectively using the subordinated Markov chain model. Furthermore, the joint disruption probability of more than two locations is estimated as non-zero according to the probability model we use, and thus, the impact of the rarer events is also captured.

We now prove that minimizing the metric $T(\gamma, \mathcal{R})$ gives a consistent estimator of the parameters γ and $\mathcal{R}$.

Lemma 1. *Let $P^{'}(S)$ be the disruption probability of all locations in set $\mathcal{S}$ with fixed marginals when the disruptions are independent. Then, $P(\{\gamma, \mathcal{R}\}, S) \rightarrow P^{'}(S)$ as $\gamma \rightarrow 0$.*

Proof. The proof is by contradiction. Let $Q(\mathcal{R}, \gamma)$ be the infinitesimal generator associated with the subordinated Markov chain system defined with subordinator

association set $\mathcal{R}$ and sensitivity γ. Also, let $Q^{'}$ be the infinitesimal generator for independent Markov chains with steady state probabilities defined by $P^{'}(\mathcal{S})$. Then $Q(\mathcal{R}, \gamma) \rightarrow Q^{'}$ as $\gamma \rightarrow 0$ irrespective of $\mathcal{R}$. Additionally, we know that $P^{'}(\mathcal{S})$ is the steady state probability for the generator matrix $Q^{'}$. Since the steady state probability for any Markov chain is unique, we must have:

$$P(\{\gamma, \mathcal{R}\}, \mathcal{S}) \rightarrow P^{'}(\mathcal{S}) \text{ as } \gamma \rightarrow 0. \tag{4.13}$$

□

The above result allows us to expand the state space of the γ parameter to $[0, \infty)$ by defining the probability distribution at $\gamma = 0$ as the probabilities under indepenent disruptions for a given set of marginal disruption probabilities.

Proposition 2. *The optimal* $(\gamma^{opt}, \mathcal{R}^{opt})$ *obtained from the minimization of* (4.10) *is a consistent estimator of the true parameters* $(\gamma^*, \mathcal{R}^*)$

Proof. For this particular proof, we follow the conditions for consistency in Hayashi (2000). The objective function satisfies the following conditions:

- As the number of observations N increases, $T(\gamma, \mathcal{R}) \rightarrow T^*(\gamma, \mathcal{R}) := \sum_{i,j}(P(\{\gamma^*, \mathcal{R}^*\}, \{i, j\}) - P(\{\gamma, \mathcal{R}\}, \{i, j\})^2$
- $T^*(\gamma, \mathcal{R})$ is uniquely minimized at $(\gamma, \mathcal{R}) = (\gamma^*, \mathcal{R}^*)$

Additionally, our construction of the probability measure ensures the continuity and convergence in probability. Using Proposition (7.1) in Hayashi (2000), this implies

that the minimization of $T(.)$ gives a consistent estimator of the true values of $(\gamma, \mathcal{R})$

□

While the minimizing the metric $T(\gamma, \mathcal{R})$ over the possible values of γ and $\mathcal{R}$ results in consistent estimates of the true parameters of the subordinated Markov chains, the minimization is not an easy problem for three reasons. First, the association set $\mathcal{R}$ is not clearly defined because, a priori, the number of subordinators, K is unknown. Second, given a certain number of subordinators, calibration of associations is combinatorial in nature. Any combination of allotments of subordinators to Markov chains is possible. Lastly, both the number of subordinators K and the association set $\mathcal{R}$ are functions of the γ parameter. Thus, we have to consider the optimization of K and $\mathcal{R}$ for different values of γ individually.

The following section explains the calibration of the γ and $\mathcal{R}$.

4.4.3 Calibration of $(\gamma, \mathcal{R})$

Estimating $(\gamma, \mathcal{R})$ is more involved than the estimation of the marginal disruption rates. It involves three nested sub-problems;

1. Calibration of the γ parameter, which we assume to be the same for all Markov chains.

2. Given value of γ, an estimate of the number of subordinators K that define the dependence within the system.

3. Given a value of γ and K, the partition of the Markov chains into K sets where Markov chains in each set are associated with a single subordinator.

The calibration of these three problems is carried out in reverse. We first fix the value of γ. For different values of the number of subordinators K, we calibrate $\mathcal{R}$ that minimizes the error function. We pick the number of subordinators K and $\mathcal{R}$ that best describes the observed data. We do this for different values of γ to select the optimal value of γ based on the minimum error function.

Below, we describe first, calibration of the $\mathcal{R}$ given K and γ. Then we present the methodology to select the optimal K for a given value of γ. Finally, we present the mechanism to select the optimal γ.

Calibration of the Partition

The input for this step of the problem are the parameters K, the number of subordinators and γ the strength of dependence between two Markov chains associated with the same subordinator. The objective is to find the subordinator association set $\mathcal{R}$ with the minimum error function. The subordinators are indexed by k and the set $\mathcal{R} = \{r_{i,k}\},\ \forall\ i \in \mathcal{I}, k \in \mathcal{K}$ where $r_{i,k} = 1$ implies Markov chain i is associated with subordinator k and $r_{i,k} = 0$ otherwise.

The optimization problem for the calibration of the partition is:

$$\begin{aligned} &\min_{\mathcal{R}} \sum_{i,j} \left(\frac{N_{i,j}}{N} - P(\{\gamma, \mathcal{R}\}, \{i, j\}) \right)^2 \\ &\sum_{k} r_{i,k} = 1,\ \forall\ i \in I \\ &r_{i,k} \in \{0, 1\},\ \forall\ i \in \mathcal{I}, k \in \mathcal{K} \end{aligned} \tag{4.14}$$

The first constraint in (4.14) ensures that each Markov chain is associated with exactly one subordinator. The above problem is a binary integer problem with a non

linear objective function and can be solved by introducing a penalty function of the form $M \sum_{i,k} r_{i,k}(1 - r_{i,k})$ for some large M. It can be shown that for a sufficiently large M, the solution to (4.14) yields an optimal solution – see Murray and Ng (2010). However, for a large M, the problem will have a locally optimal solution at every feasible solution due to the convexity of the penalty function. We propose an alternate, SIMPLEX like solution methodology to solve the above problem.

Algorithm 1 Optimizing $\mathcal{R}$, given γ and K

1. Choose a starting allocation set $\mathcal{R}$, i.e, choose a random allocation of the subordinators to the Markov chains.

2. For each $i \in \mathcal{I}$, solve for $H(i) := \min_k T(\{\gamma, \mathcal{R}_{i,k}\}) - T(\{\gamma, \mathcal{R}\})$ where $\mathcal{R}_{i,k}$ is the allocation set where Markov chain i is associated with subordinator k. The subordinator allocations of all the other Markov chains in $\mathcal{R}_{i,k}$ are the same as the allocations in $\mathcal{R}$.

3. Solve for $E^{'} := \min_i H(i)$ and $i^{'} = \text{argmin}_i H(i)$.

4. If $E^{'} = 0$, Stop. Else set $\mathcal{R} = \mathcal{R}_{i^{'},k^{'}}$, where $k^{'} = \text{argmin}_k T(\{\gamma, \mathcal{R}_{i,k}\}) - T(\{\gamma, \mathcal{R}\})$.

5. Go to Step 2.

Starting with a random allocation of subordinators, in every iteration of algorithm 1, the subordinator associated with one Markov chain is changed. The choice of the Markov chain whose subordinator is changed, and the choice of the subordinator to which it is changed is based on the new allocation that will most decrease

the error function. This is reflected in step 2 of the Algorithm where, first for each Markov chain and each possible change in the subordinator association, the difference in the error function due to the change is calculated. Then, in step 3, we pick the Markov chain and the subordinator for which the change in allocation causes the most decrease in the error function. In step 4, we make the change of allocation and start the new iteration. The algorithm terminates when no possible change in the subordinator allocations causes a decrease in the error function.

Proposition 3. *Algorithm 1 converges to a locally optimal solution.*

Proof. The error function is $T(\gamma, \mathcal{R})$, by definition positive. Since the algorithm terminates if the decrease in the error function is 0, at every iteration, the error function decreases. Hence the algorithm must converge to a locally optimal solution. □

We try different starting points of the initial solution $\mathcal{R}$ to ensure we arrive at a globally optimal solution.

Calibration of the Number of Subordinators

We now describe the calibration of the number of subordinators. The input for this step is the γ parameter. For a given value of γ and each possible value of the number of subordinators K, we calculate the best subordinator allocation and subsequently, the error function for the best allocation, using Algorithm 1. Based on the error function for optimal allocation for each value of K, we select K that minimizes an Adjusted R^2-like metric.

Since our metric for calibration is a least squares metric, we define:

$$G = 1 - \frac{\sum_{i,j} P(\{\gamma, \mathcal{R}\}, \{i,j\}) - \frac{N_{i,j}}{N})^2}{\sum_{i,j} (P(\{\gamma, \mathcal{R}\}, \{i\}) P(\{\gamma, \mathcal{R}\}, \{j\}) - \frac{N_{i,j}}{N})^2} \tag{4.15}$$

The above metric is R^2-like in that it is the ratio of the explained variation in probability to the total variation in probability due to the dependence. Similar to standard regression metrics, we also define *Adj G* metric as:

$$Adj\ G = 1 - (1 - G)\frac{\binom{I}{2} - 1}{\binom{I}{2} - K - 1}, \tag{4.16}$$

where I is the number of locations and K is the number of subordinators. $\binom{I}{2}$ is the number of observations we use in the calibration of the subordinated Markov chain and hence can be thought of as the degrees of freedom in the data. Our choice for the number of subordinators K^* is the one with the highest *Adj G* metric. As with regression, the choice of K that maximizes the *Adj G* metric cushions for over fitting while ensuring the model explains dependence sufficiently.

Calibration of γ

The last part of the methodology to calibrate the model parameters is the evaluation of γ. We propose to optimize for γ by searching over a discrete grid of values for the choice of γ that minimizes the error function. Since the domain of γ is $[0, \infty)$, we instead construct a grid on the variable $\phi(\gamma)$ which has a domain of $[0, 1]$.

The procedure to calibrate γ and $\mathcal{R}$ can be described as follows:

1. Create a grid of possible values for $\phi(\gamma)$ between 0 and 1.
2. For each value of $\phi(\gamma)$:
   (a) For each value of $K = 1, 2, \ldots$:
   (b) Using algorithm 1, to calibrate the optimal allocation and calculate the *Adj G* metric for the optimal allocation.
   (c) Select the K with the highest *Adj G*.
3. Select the γ, and the corresponding optimal K and $\mathcal{R}$, that minimizes the error function.

The only observed probabilities we use in the calibration, the marginal disruption probabilities and the joint disruption probabilities of Markov chains taken two at a time, have low standard errors. Thus, the calibration of the parameters is robust.

4.5 Optimization

As mentioned before, getting to a solution to the facility location problem is dependent on the number of random disruptions scenarios. Our probability model considers all possible disruption scenarios, i.e, given I facilities, there are 2^I possible disruption scenarios with non-zero probabilities. Solving the facility location problem under an exponentially large number of disruption scenarios is difficult using

traditional methods. In this section we propose algorithms to solve for the optimal facility location.

4.5.1 Optimization through bounding

The algorithm for optimization presented here assumes equal marginal disruption rates and a single subordinator for the subordinated Markov chains and ensures a globally optimal solution. It can, however, be extended to a problem with multiple subordinators and unequal marginal disruption rates.

We begin with the UFLP, defined in (4.1). Given a cost structure, which includes the set up costs $f_i, \forall i$, and transport costs $c_{ij}, \forall\ i, j$, solving the UFLP is easier due to the fewer number of variables and constraints in the problem relative to the RUFLP. Hence, we prefer solving the UFLP multiple times than attempt to solve the reliable facility location problem.

We first note that the feasibility region for the choice of facility locations, x, for both the RUFLP and UFLP are the same. We construct linear lower and upper bounds for the cost of any selection of facilities x in the RUFLP, in terms of the total cost for the same choice of facilities in the UFLP. Using these bounds, we restrict the variable space over which we search for the optimal x and solve the UFLP multiple times to arrive at an optimal solution for the RULP.

We first present the expressions for the lower and upper bounds that we use in our algorithm. The expressions for the lower and upper bound for the total cost for a choice of facilities under the RUFLP, in terms of costs under the UFLP, are dependent on the number of facilities in any choice of x. Thus we fix the number of

facilities to be set up, say n, and solve for the optimal choice of facilities given that exactly n facilities are to be set up. The optimal solution for the RUFLP can be obtained by comparing the cost of the optimal facility location choice for different values of n.

We define the following parameters which would help in establishing the lower bounds. For each facility $i \in \mathcal{I}$, we define q_i, the servicing costs if facility i was exclusively used to serve all customers, i.e., $q_i = \sum_j c_{ij}$. Let n be the number of facilities that are to be set up. Since the marginal rates of disruption are equal and there is exactly one subordinator in the system, we define $P_z,\ z = 0, \dots n$ as the probability that in a setup with n facilities, exactly z are in the "ON" state. Note that for any choice of facilities x, $h(x)$ is the total servicing costs in the UFLP and $\mathbb{E}[h(x,\xi)]$ is the expected client servicing cost in the RUFLP.

Proposition 4. *Let T be the penalty cost when none of the customers are serviced. ($T = \sum_j c_{0j}$). Then:*

$$\mathbb{E}[h(x,\xi)] \geq (\sum_{z>2} P_z)h(x) + (P_1 + (n-1)P_2)\sum_i q_i x_i + (P_0 - n(n-1)P_2)T$$

$$\implies \sum_i f_i x_i + \mathbb{E}[h(x,\xi)] \geq \sum_i f_i x_i + (\sum_{z>2} P_z)h(x) + (P_1 + (n-1)P_2)\sum_i q_i x_i + (P_0 - n(n-1)P_2)T \tag{4.17}$$

Proof. The proof follows from the fact that servicing costs are super modular in nature and hence the servicing costs of a set of facilities is less than the servicing costs of any subset of facilities. □

The above proposition gives a lower bound to the cost under the RUFLP for a given choice x of facilities. Observe that the terms in the right hand side

of the inequality, the lower bound, are independent of various disruption scenarios. Therefore, the lower bound of the costs in the RUFLP can be expressed as a linear combination of the cost under the UFLP. Thus, the lower bound constructed in Proposition (4) can be added to the UFLP to reduce the state space of possible facility location choices. For example, given a choice of facility locations, x, we can calculate the costs for x under the RUFLP. We can then add a constraint to the UFLP using the above proposition, to restrict the feasible region, and only look for solutions whose lower bound is less than the calculated costs for the RUFLP.

Similarly, an upper bound to the RUFLP is also constructed using the super modular nature of the servicing costs for any set of facilities. We define the parameters:

$$\begin{aligned} C_{Full} &= \sum_{i>1} \frac{n}{i} P_i \binom{n-2}{i-2} \\ C_q &= \sum_{i<n} P_i \frac{\binom{n-1}{i}}{n-1} \end{aligned} \tag{4.18}$$

Proposition 5. *Let T be the penalty cost when none of the customers are serviced. ($T = \sum_i c_{0i}$). Then:*

$$\begin{aligned} &\mathbb{E}[h(x,\xi)] \leq \mathbb{P}_0 T + \sum_i C_q q_i x_i + C_{Full} h(x) \\ \implies &\sum_i f_i x_i + \mathbb{E}[h(x,\xi)] \leq \sum_i f_i x_i + P_0 T + \sum_i C_q q_i x_i + C_{Full} h(x) \end{aligned} \tag{4.19}$$

Proof. Using super modularity on individual terms. □

Just as with the lower bound, the upper bound can also be expressed as a linear combination of the cost under the UFLP. Thus, we can use the upper bound

expression to search for the optimal solution of the RUFLP through the UFLP. We now present the algorithm to solve for the optimal choice of facilities under the RUFLP, using the bounds created in the two propositions above. As mentioned, we present the algorithm for a fixed number of facilities. The overall solution can be evaluated by varying the number of selected facilities.

Algorithm 2 Solving the RUFLP under dependent disruptions with equal marginal disruption rates and n, the number of selected facilities.

1. Evaluate the probability of different disruption scenario $P_z,\ z = 0, \ldots, n$.

2. Evaluate C_{Full} and C_q using Equation (4.18).

3. Set LB = Inf. Define UBL as:

$$\begin{aligned} &\min \sum_i f_i x_i + P_0 T + \sum_i C_q q_i x_i + C_{Full} h(x) \\ &\text{where } h(x) = \min \sum_{ij} c_{ij} y_{ij} \\ &\text{s.t } \sum_i y_{ij} = 1,\ \forall\ j \\ &\quad y_{ij} \leq x_i,\ \forall\ i, j \\ &\quad \sum_i x_i = n \\ &\quad x_i \in \{0,1\}\ \forall\ i, y_{ij} \in \{0,1\}\ \forall\ i, j \end{aligned}$$

$$\sum_i f_i x_i + (\sum_{z>2} P_z) h(x) + (P_1 + (n-1)P_2) \sum_i q_i x_i + (P_0 - n(n-1)P_2) T \leq LB$$

4. Solve UBL for new candidate facility location choice $x^{'}$. If UBL is infeasible, stop.

5. Evaluate $LB^{'} := \sum_i f_i x_i^{'} + \mathbb{E}[h(x_i^{'}, \xi)]$.

6. If $LB^{'} < LB$, set $x^* = x$ and $LB = LB^{'}$ [3].

7. Introduce a cut in the UBL that excludes solution $x^{'}$. Go to Step 4.

The underlying idea behind the algorithm is to iteratively search for possible facility location choices based on the lower bound and upper bound of the true cost under disruptions. Given any starting choice of the facilities, the lower bound constraint first reduces the possible facility location choices based on the true cost of the starting choice of the facilities which can be evaluated. The algorithm then searches for a new candidate solution based on the facility choice that has the lowest possible upper bound for the RUFLP. This search is conducted on the much smaller UFLP and hence is computationally easy. If the new candidate solution has a true cost (cost when considering disruptions) less than the current optimal solution, first the current optimal is updated to the new candidate solution (step 6) and the lower bound constraint in the UBL is changed to reflect the new solution. Else, a constraint is added to the UBL to reject the new candidate solution.

Proposition 6. *Algorithm 2 terminates at the optimal choice of facilities.*

Proof. At each iteration, we exclude a solution that is sub-optimal. Additionally, the number of possible facility location choices are finite. By construction, the algorithm is continued till the UBL is infeasible, i.e., there are no more possible facility location

[3]This step affects the lower bound constraint in UBL

choices with costs potentially lower than the current solution. Thus the algorithm will converge at the optimal choice of facilities. □

The algorithm can be modified to solve for a near-optimal solution with bounds by stopping the algorithm not when the UBL is infeasible but when the current optimal solution falls within an interval of the best lower bound.

The above algorithm can be adapted to the case when the system of Markov chains that describes disruption probabilities has multiple subordinators and unequal marginal disruption rates. We do so through the following proposition.

Proposition 7. *For any choice of facilities x:*

- *The expected servicing cost $\mathbb{E}[h(x,\xi)]$ is increasing in the individual marginal rates for any configuration of the subordinators.*
- *The expected servicing cost $\mathbb{E}[h(x,\xi)]$ is greater than or equal to the expected servicing cost in the case when the disruptions are independent.*

Proposition (7) allows us to rewrite the lower and upper bounds in terms of the lowest and highest marginal disruption rates. For example, if the marginal disruption rates are unequal, the upper bound is constructed by using the highest marginal disruption probability for all Markov chains and assuming a single subordinator defines the disruptions to evaluate C_{Full} and C_q. Similarly, while constructing the lower bound, probabilities P_0 and P_1 are evaluated by assuming the lowest marginal disruption rates for each plant and assuming independent disruptions amongst the facilities.

Algorithm 2 works best when the marginal disruption rates are low and there is a single subordinator. When the marginal disruption rates are low, the bounds specified in Propositions (4) and (5) are tight and hence the set of facility locations is constrained to a large extent. The algorithm is not preferable when the marginal disruption rates are high or the dependence structure is more complex. For problems with more complex dependence structures or higher marginal disruption rates, we solve for the optimal facility location via simulation. This approach is described in the next section.

4.5.2 Optimization through Simulation

One of the advantages of using subordinated Markov chains to model the disruption probabilities is that they can be simulated easily. This allows us to generate random samples of a given dependence structure efficiently. Thus, an alternate method to optimize for the facility locations is via simulation. This option of optimization via simulation is preferable when the underlying dependence structure is complex due to multiple subordinators defining the probability distribution or when the marginal disruption rates are high and the bounds developed the previous section are not tight.

Proposition 8. *For any choice of the facility locations, if all the transportation costs (c_{ij}) are unique, then the LP relaxation of the second stage problem of the RUFLP has an unique integer solution.*

Proof. The proof follows from the fact that since the facilities are uncapacitated,

there can only be one facility that can be assigned to a customer location to minimize costs. □

The above theorem allows us to restate the RUFLP as a two stage stochastic optimization problem with linear variables in the second stage. Thus we use techniques like Bender's decomposition to solve the RUFLP – see Santoso et al. (2005). We now present the algorithm to solve for the optimal facility location choice for the RUFLP using simulation.

Algorithm 3 Bender's decomposition Algorithm for the RUFLP. Set UB = +Inf. LB = -Inf. Set $Iter = 0$. $x^{'}$ is the current solution. Let $\epsilon > 0$ be the acceptable tolerance level.

1. Solve the Master Problem:

$$\begin{aligned} \text{LB} \ &= \min_x \sum_i f_i x_i + \theta \\ & x_i \in \{0.1\}, \ \forall \ i = 1, \ldots, \mathbb{I} \\ & a^k + x b^k \geq \theta, \forall \ k = 1, \ldots, K \end{aligned}$$

2. Simulate n different scenarios, $(\omega^1, \ldots, \omega^n)$ of the state of the facilities using algorithms described in chapter 2.

3. Evaluate UB using the simulated scenarios or the true disruption probabilities. For each simulated scenario, evaluate the total cost under the current optimal solution. The sample average of the evaluated total costs is the UB.

4. If $(UB - LB) \leq \epsilon$ stop. Else Go to step 5.

5. Given a solution x^k to the master problem and for each scenario $l = 1, 2, \ldots n$, solve the second stage problem:

$$\begin{aligned} &\min_y \sum_{ij} c_{ij} y_{ij} + \sum_j c_{0j} y_{0j} \\ &s.t. \sum_i y_{ij} + y_{0j} = 1 \ (\mu_j^l) \\ &y_{ij} \leq x_i^k \omega_i^l, \ \forall i, j \ (\nu_{ij}^l) \\ &y_{0j} \leq 1 \ , \forall j (\chi_j^l) \\ &y_{ij} \geq 0 \ \forall i, j. \end{aligned}$$

where the parameters in the parentheses are the duals of the constraints.

6. Set $k = k + 1$ and evaluate the new cut coefficients:

$$b_i^k = \frac{1}{n} \sum_{lj} \nu_{ij}^l$$

$$a^k = \frac{1}{n} \sum_{lj} \chi_j^l$$

7. Go to Step 1.

At each iteration of algorithm 3, a new cut is added to the master problem which bounds the optimal solution of the second stage problem. Progressively, the cuts become tighter and thus the solution to the master problem yields the optimal choice of facility locations.

4.6 Results

In this section we discuss the results of both our calibration model and our optimization algorithm. We use the data set described in Cui et al. (2010) for our

analysis. We consider a facility location problem with 49 possible locations. The locations are the capitals of the 48 contiguous states of US and Washington D.C. They serve as both the facility locations and the client locations. The set up costs are proportional to the median cost of a house in the capital of each state. The servicing costs from any location to any location are proportional to the product of the distance between the two locations and the population of the client location state. We assume a fixed penalty costs for not servicing a client which is higher than all the servicing costs in the data set.

For the purposes of the disruption probabilities, we use data from the National Oceanographic and Atmospheric Authority (NOAA) for the tornadoes affecting the 48 contiguous states of the United States and DC. For every tornado from 1950-2015, NOAA records the date of a tornado, the location of a tornado including the state, the intensity of the tornado. NOAA provides this data for all the states of the United States. We aggregate the data on a daily basis and hence we have 24,105 days in out data set for the calibration of the 2^{49} distinct disruptions probabilities. We consider any state affected by a hurricane to be disrupted on that day. We observed 11,225 days with hurricanes affecting various locations of our dataset. Each hurricane, on average affects 2.3 locations.

4.6.1 Calibration of the Disruption Model

As explained in section 4.4, we first estimate the marginal disruption probabilities for each of the 49 locations. We find that the state of Texas has the highest disruption probability of 0.12 while states in the mid-west have low disruption prob-

abilities. Washington D.C. and Rhode Island have the lowest marginal disruption probabilities.

For the second phase of our calibration, described in section 4.4.3, we create a grid of possible values for $\phi(\gamma)$ which can take values from $[0, 1]$. We solve for the optimal partition using algorithm 1 for different values of $\phi(\gamma)$ and different values of K, the number of subordinators. The results are presented in table (4.1)

Table 4.1: Error function $T(\gamma, \mathcal{R})$ $(*10^{-3})$ by K and γ.

$K \setminus \phi(\gamma)$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	3.56	2.55	6.54	13.40	21.86	31.14	40.75	50.41	59.92	69.17	78.11
2	3.56	1.84	2.76	5.34	8.80	12.74	16.90	21.11	25.30	29.40	34.14
3	3.56	1.67	2.12	3.54	5.79	7.78	10.20	12.90	15.47	17.94	20.36
4	3.56	1.65	1.96	2.76	4.17	5.67	7.22	9.08	10.87	12.63	14.32
5	3.56	1.64	1.83	2.41	3.24	4.71	5.50	6.75	8.03	9.24	11.10
6	3.56	1.64	1.77	2.18	2.78	4.19	4.54	5.55	6.59	7.83	8.76
7	3.56	1.64	1.75	2.03	2.53	3.75	4.22	4.88	5.57	6.81	7.49
8	3.56	1.64	1.74	1.97	2.38	3.10	3.88	4.38	5.03	5.64	6.34

The above table presents the value of the Error Function for the optimal partition of the facilities obtained using algorithm 1. The error function is presented for each combination of values of γ, given in the columns and the values of K given in the rows.

Table (4.1) shows that the error function, $T(\gamma, \mathcal{R})$, first decreases and then increases with the value of γ. We note here that lower values of γ imply low levels of dependence and higher values of γ imply high levels of dependence. Thus, while there is evidence of dependence amongst disruptions, the dependence is not complete in the sense that a disruption of any facility does not always occur simultaneously with a disruption of every other facility with a common subordinator. The error function decreases with K, the number of subordinators initially, but then plateaus.

This is because having additional subordinators need not necessarily explain joint disruptions better. However, a model with a greater number of subordinators should be just as good, if not better, as one with fewer subordinators. Table (4.1) shows that the optimal choice of $\phi(\gamma)$ is 0.1, across all values of the number of subordinators. This translates to a γ value of 0.11. Intuitively, this implies that every common shock or discontinuity of the subordinator has a one in nine chance of resulting in a disruption of the Markov chain.

The next phase of the calibration involves the selection of K. Using the *Adj G* metric defined in Equation (4.16), we choose the optimal K from the options available.

Table 4.2: *Adj G* by K for $\phi(\gamma) = 0.1$

K	1	2	3	4	5	6	7	8
Adj G	0.284	0.484	0.531	0.536	0.538	0.538	0.538	0.538

The above table presents the value of the *Adj G* metric for different values of K calculated according to Equation (4.16)

Based on the values in the table (4.1) and table (4.2), we choose $K = 6$ and $\phi(\gamma) = 0.1$ as the optimal number of subordinators and the optimal strength of dependence required to describe the disruptions probabilities for the 49 facility locations.

Each subordinator corresponds to a source of dependence in the outages of different facility locations. Locations that are associated with the same subordinator exhibit dependent disruptions while those associated with different subordinators

exhibit independence. For the six subordinators and $\phi(\gamma) = 0.1$, table (4.3) presents the states associated with each subordinator.

Table 4.3: Optimal facility locations

Subordinator	States
1	Arkansas, Colorado, Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, Oklahoma, South Dakota, Texas, Wisconsin, Wyoming
2	Alabama, Georgia, DC, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia
3	California, Idaho, Oregon, Washington
4	Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Rhode Island, Vermont, West Virginia
5	Arizona, Nevada, Utah
6	Montana, New Mexico

The above table shows the states associated with a single source of disruptions or a single subordinator.

In table (4.3), we see that South-Eastern states like Georgia, Louisiana, Alabama, Mississippi, Tennessee and Arkansas have a common subordinator and hence dependent disruptions. This is in line with our expectations since there is a high probability that hurricanes that affect one of these states will also affect the others. There is also such an expected grouping in the Western part of the country where Washington and Oregon and California share a common subordinator. Similarly, the North-Eastern states have a common subordinator, as do the central regions. It must be noted that New Mexico and Montana have a common a subordinator which is contrary to expectations. This could be a byproduct of limiting the number of

subordinators using the *Adj G* metric.

The above dependence structure is validated by the average correlation between any two facilities within the same subordinator grouping. The average correlation of disruptions between any two facilities is 0.0273. The average correlation between disruptions of any pair of states with a common subordinator, for each subordinator is presented in table (4.4):

Table 4.4: Average Correlation by Subordinator

Subordinator	1	2	3	4	5	6
# States	17	12	4	11	3	2
Avg. Correlation	0.087	0.065	0.027	0.090	0.023	0.024

The above table presents the average correlation between a pair of facilities associated with the same subordinator and the number of locations associated with the subordinator according to our calibration.

In table (4.4), disruptions of locations associated with the three subordinators show a higher average correlation between pairs of states than the global average. These three subordinators account for most of the 49 states in our study. We do acknowledge that the average correlation for the other subordinators is lower than the population average. This is because we limit the number of subordinators using the *Adj G* statistic. Additionally, the number of states associated with these subordinators is low and the states associated with these subordinators are in geographical proximity leading us to believe that they do tend to share common weather patterns and thus dependent disruptions.

4.6.2 Optimization

We now present the results for the RUFLP for the dependence structure defined by the subordinated Markov chain parameters and associations described in section 4.6.1. We optimize for the best choice of facility locations using algorithm 2. Below, we present the results of the algorithm for different number of facilities, the number of iterations and the time taken to arrive at the optimal solution.

Table 4.5: Optimal facility locations

N	Optimal Solution	# iterations	Time (in s)	Total Cost (*10^4)
2	WV, NV	51	41.1	1.46
3	PA, IL, NV	38	30.6	1.24
4	PA, IL, MS, NV	29	33.4	1.18
5	PA, IL, MS, NV, VT	5	42.27	1.15
6	PA, MI, AL, OK, NV, VT	27	21.45	1.18
7	PA, MI, AL, OK, IA, NV, VT	20	19.63	1.23

The above table presents the results of the optimization algorithm for different number of facilities. The columns N corresponds to the number of facilities to be chosen. The column Optimal Solution is the optimal selection of the facilities. The column # iterations is the number of cuts in algorithm 1 required to solve the problem. The column Time is the time taken by the algorithm to solve for the optimal solution and the column Total Cost is the sum of the Set up and the Expected Servicing Costs of the facility location choice under the disruption model defined by the subordinated Markov chain system.

From table (4.5) we see that the optimal facility location choice involves placement of facilities at Pennsylvania, Illinois, Mississippi, Nevada and Vermont. The average solution times are around 30-40 seconds for each choice of the number of plants and we need around 30 iterations for each choice. It is important to note that, in this particular scenario, the optimal solution when the disruptions are assumed to be independent is the same as the case when disruptions are dependent.

This is because the optimal solution under dependent disruptions chooses states with different subordinators. Since individual states in our optimal solution have different subordinators, their disruptions are independent. However, the true costs of the optimal solution are higher for the dependent disruption case. Additionally, we find that the optimal solution for the dependent and independent disruptions is not the same when the set up costs are lower. In that case, a higher number of facilities are needed, amplifying the effect of dependence in the choice of the optimal facility locations.

Table 4.6: Optimal facility Locations - Robust disruption probabilities

States	Cost under Robust Prob. ($*10^4$)	Cost Under Sub. Prob. ($*10^4$)
KY, NV, VT	1.32	1.25

The above table presents the optimal choice of facilities under the robust distribution and the Total Expected Cost under the robust disruption probabilities and Subordinated Disruption probabilities.

We compare the optimal facility location under the two disruption probabilities distributions (subordinated and robust). We find that using a robust distribution for the disruptions leads to a choice with fewer facilities as shown in table (4.6). As expected, the robust solution forces facilities in locations with very low marginal disruptions rates. The three locations in table (4.6) have marginal disruption rates of 0.0163, 0.0031, 0.0015 respectively, far below the marginal disruption rates of the facilities under the dependent disruption solution. The robust probability distribution assumes a high degree of dependence between disruptions. Hence, only facilities with a lower marginal disruption rate are chosen ensuring the impact of the high risk multiple disruption events is kept to a minimum.

Chapter 5
Concluding Remarks

This dissertation develops a new framework for modeling dependence in risk and demonstrates its applicability and flexibility in modeling dependence in Operations Management problems. The impact of dependence in risk due to common disruptive factors underscores the importance of such a framework. The framework, called subordinated Markov chains, uses time-changing to model dependence in continuous time Markov chains in a parsimonious fashion. We also show the adaptability of the framework to two applications. While this dissertation shows the advantage of subordinated Markov chains to the two applications mentioned, it can be used for further analysis in not only the two applications covered, but also in other areas in Operations Management. Additionally, the framework itself can be modified to include effects like seasonality in the underlying stochastic processes.

In chapter 2, we constructed subordinated Markov chains and partially subordinated Markov chains. Subordinated Markov chains are classical continuous time Markov chains constructed in a random time frame. The random time-frame is defined by a Lèvy process called a subordinator. The subordinator has jumps or dis-

continuities which result in dependent behavior in the individual Markov chains. The jumps or discontinuities can be interpreted as shocks or the common risk factors that cause disruptions in real life. Since the impact of risk factors is only negative, classical subordinated Markov chains cannot be implemented as-is in an Operations Management context. Thus, we construct partially subordinated Markov chains which allow us to control the direction of dependence of the disruptions. Partially subordinated Markov chains incorporate uniformity in the effect of factors that cause dependence in Markov chains and can be applied to Operations Management problems. An advantage of using subordination to model dependence is the ease of simulation that follows. Thus, we also present algorithms to simulate subordinators, subordinated Markov chains and partially subordinated Markov chains.

We can expand the scope of the framework by using other models for the subordinators. The subordinators that we use are Poisson jump processes which are time-homogeneous. The dynamics of the resulting Markov chains are also time-homogeneous. However, certain shocks can impact components over a period of time or have cyclical patterns. There is a wide class of subordinators that could be used to model more complex dependence in the multiple Markov chains. For example, a sub-class of subordinators is the shot noise process – see e.g. Gaspar and Schmidt (2010). Subordination of Markov chains using shot noise subordinators leads to a dependence structure where the effect of a shock decays with time. Thus, the same shock can affect different Markov chains at different points in time. Another class of subordinators, the additive subordinators can be used to model temporal effects like cyclicity in effect of common shocks to the system of Markov chains. These

extensions to the subordinated Markov chain framework can help us model more complex dependence structures.

In chapter 3, we model the dependence in outages of electricity power plants. We calibrate the partially subordinated Markov chain model for electricity power plants for two interconnections in the North American Electric Reliability Corporation (NERC); the Electricity Reliability Council of Texas (ERCOT) and the Western Electricity Coordinating Council (WECC). We find evidence of dependence in outages of power plants based on common fuel, like natural gas, in both the ERCOT and the WECC. In the case of the WECC, we also find geography based dependence which affects all plants. We show the impact of this dependence on reliability of the existing generation capacity and on the reliability of new capacity addition.

Our framework can be used as a building block towards a model for electricity prices that accounts for dependent outages in the supply curve. In a deregulated market, the price of electricity is determined by the marginal cost of production of the last power plant needed to satisfy demand. Simultaneous power plant failures lead to large, short lived, spikes in electricity prices. A stochastic model of the supply curve based on our model, would allow for a parsimonious model of spikes in prices in the spot electricity market. Examples of contracts that are particularly sensitive to such spikes include interruptible and direct load control contracts – see Baldick et al. (2006).

While we have only considered reliability when evaluating capacity additions, further analysis, from a public policy perspective, would include the cost of the primary fuel used when comparing the impact of capacity addition. This cost-benefit

analysis of the capacity addition based on primary fuel prices would help policy planners compare the various options available for capacity addition from an economic perspective.

Another possible extension to this research could be to learn the clustering of the dependence set associations from the data. In our research, we have assumed the dependence sets are defined based on commonalities in fuel or geography. That is, we have defined the common source of shocks to be defined by fuel or geography. However, it is possible that some plants might share a common fuel and yet do not have a common sources of outages. Alternatively, it is also possible that there are other charactertistics of power plants that define the commonality of shocks. This dependence structure could be learnt from the data. For example, algorithms could be constructed to learn the dependence set associations in an iterative fashion from the commonality in outages. This approach might be difficult considering the number of electricity plants involved and consequently the exponentially high number of dependence set combinations possible. However, this could lead to a truer representation of the dependence in outages between the power plants.

In chapter 4, we consider the uncapacitated facility location problem with joint disruptions. The underlying probability distribution for disruption affects the choice of the facility locations. Since disruptions are caused by factors common to multiple facilities, we expect dependence in the disruptions. Hence, we use partially subordinated Markov chains to model the joint disruption probabilities. We propose algorithms to calibrate the subordinated Markov chain system to model the probability of joint disruptions, and for finding the optimal choice of facilities under the

proposed probability structure. Calibration of the framework on real hurricane data results in a dependence structure where locations that are geographically closer to each other exhibit dependent disruptions and disruptions of locations farther away from each other are independent. This is in line with our expectations given the nature of hurricanes which tend to affect regions in geographical proximity.

Extensions to this research include considering joint disruptions in the capacitated facility location problem with disruptions – see e.g. Santiváñez and Carlo (Santiváñez and Carlo) and to the dynamic uncapacitated facility location problem, mentioned in Van Roy and Erlenkotter (1982), with disruptions. The transition matrix in our disruption model is available in closed form making such an adaptation convenient. Multiple sources of disruption for each location can also be considered. For example, in our model for disruptions, each Markov chain is associated with a single subordinator. We could associated each Markov chain with multiple sources of risk or multiple subordinators. This represents a more accurate model for disruptions. However, calibration of the subordinators and the intensities might be more involved.

The strength of the framework is in its adaptability to multiple problems in Operations Management where dependence is a significant factor. Here, we would like to mention a few other application areas where the partially subordinated Markov chain framework can be useful to model dependence and from which we can derive managerial insights.

Inventory policy in the presence of a single supplier with disruptions has been widely studied. However, an under addressed area in Operations literature is to

study the impact of dependence in disruptions of multiple suppliers on the optimal inventory policy. The lack of a specific framework that models dependent disruptions of suppliers is an impediment towards such an analysis. Modeling the individual supplier behavior using subordinated Markov chains is a way to understand the impact of dependent supplier disruptions on optimal inventory policy. Preliminary results in this direction show that higher dependence among suppliers producing complementary products results in firms having to hold lower inventories. The impact of dependence is opposite when dealing with substitutable suppliers where increased dependence causes firms to hold higher inventory levels.

Another application area where partially subordinated Markov chains can be used is the modeling of correlated credit rating transitions of firms. For example, the credit rating of two auto-makers could be dependent on common factors like economic cycles, shocks to common raw material manufacturers, etc. Firm credit ratings are generally modeled as Markov chains. Thus, subordinated Markov chains can be used to model the dependence in the credit rating transitions. Calibration of the dependence structure in credit ratings using subordinators could help us identify large scale risk factors and their impact on the firms in the economy. This could also help in pricing of Credit Default Swaps and other default related debt instruments.

Subordinated Markov chains, by their very nature, could be useful tools in better understanding business practices and theories where dependence is an integral aspect. This dissertation has a grander goal of introducing and advocating the potential benefits of using the proposed subordinated Markov chains in Operations Management models.

Bibliography

Allan, R., R. Billinton, A. Breipohl, and C. Grigg (1999). Bibliography on the Application of Probability Methods in Power System Reliability Evaluation. *IEEE Transactions on Power Systems 14*(1), 51–57.

Almutairi, A., M. H. Ahmed, and M. Salama (2015). Probabilistic Generating Capacity Adequacy Evaluation: Research Roadmap. *Electric Power Systems Research 129*, 83–93.

Baldick, R., S. Kolos, and S. Tompaidis (2006). Interruptible Electricity Contracts from an Electricity Retailer's Point of View: Valuation and Optimal Interruption. *Operations Research 54*(4), 627–642.

Banerjee, O., L. E. Ghaoui, and A. dAspremont (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine learning research 9*(Mar), 485–516.

Barndorff-Nielsen, O. E. (1997). Processes of normal inverse gaussian type. *Finance and stochastics 2*(1), 41–68.

Berkson, J. (1980). Minimum chi-square, not Maximum Likelihood! *The Annals of Statistics*, 457–487.

Bernstein, A., D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman (2014). Power Grid Vulnerability to Geographically Correlated FailuresAnalysis and Control Implications. In *INFOCOM, 2014 Proceedings IEEE*, pp. 2634–2642. IEEE.

Bielecki, T. R., J. Jakubowski, A. Vidozzi, and L. Vidozzi (2008). Study of Dependence for Some Stochastic Processes. *Stochastic Analysis and Applications 26*(4), 903 – 924.

Billinton, R. and R. N. Allan (1992). *Reliability evaluation of engineering systems*. Springer.

Billinton, R., M. Fotuhi-Firuzabad, and L. Bertling (2001). Bibliography on the Application of Probability Methods in Power System Reliability Evaluation 1996-1999. *IEEE Transactions on Power Systems 16*(4), 595–602.

Bochner, S. (1949). Diffusion Equation and Stochastic Processes. In *Proceedings of the National Academy of Sciences*, Volume 35 (7), pp. 368–370.

Bonney, J. and W. Cassidy (2017). Impact of harvey flooding ripples through logistics networks. (Accessed on 09/29/2017).

Borges, C. L. and R. J. Pinto (2008). Small Hydro Power Plants Energy Availability Modeling for Generation Reliability Evaluation. *IEEE Transactions on Power Systems 23*(3), 1125–1135.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2002). The fine structure of asset returns: An empirical investigation. *The Journal of Business 75*(2), 305–332.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, 135–155.

Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes*. Boca Raton: CRC Press.

Cornuéjols, G., G. L. Nemhauser, and L. A. Wolsey (1983). The uncapacitated facility location problem. Technical report, Carnegie-mellon univ pittsburgh pa management sciences research group.

Cox, D. R. (1972). The analysis of multivariate binary data. *Applied statistics*, 113–120.

Cui, T., Y. Ouyang, and Z.-J. M. Shen (2010). Reliable facility location design under the risk of disruptions. *Operations Research 58*(4-part-1), 998–1011.

Curley, G. (2006). Power plant performance indices in new market environment: IEEE standard 762 working group activities and gads database. In *Power Engineering Society General Meeting, 2006. IEEE*, pp. 5–pp. IEEE.

Dueck, D., B. Frey, N. Jojic, V. Jojic, G. Giaever, A. Emili, G. Musso, and R. Hegele (2008). Constructing treatment portfolios using affinity propagation. In *Research in Computational Molecular Biology*, pp. 360–371. Springer.

Dunford, N. and J. T. Schwartz (1958). *Linear Operators*, Volume I of *Pure and Applied Mathematics Volume VII*. Wiley-Interscience Publishers.

Eberlein, E. and K. Prause (2002). The generalized hyperbolic model: financial derivatives and risk measures. In *Mathematical FinanceBachelier Congress 2000*, pp. 245–267. Springer.

Erlenkotter, D. (1978). A dual-based procedure for uncapacitated facility location. *Operations Research 26*(6), 992–1009.

Frank, C. and K. Römer (2007). Distributed facility location algorithms for flexible configuration of wireless sensor networks. *Distributed computing in sensor systems*, 124–141.

Gaspar, R. M. and T. Schmidt (2010). Credit risk modelling with shot-noise processes.

Gledhill, R., D. Hamza-Goodacre, and L. P. Low (2013). Business-not-as-usual: Tackling the impact of climate change on supply chain risk. *Resil. Win. Risk PWC 2*, 15–20.

Harvey, S. M., W. W. Hogan, and T. Schatzki (2005). A Hazard Rate Analysis of Mirants Generating Plant Outages in California (Jan-04). Technical report, Harvard University, John F. Kennedy School of Government. Avaliable at `https://sites.hks.harvard.edu/fs/whogan/Harvey_Hogan_Schatzki_Toulouse_010204.pdf`. (Accessed July 2017).

Hayashi, F. (2000). Econometrics princeton university press.

Kemeny, J. G. and J. L. Snell (1960). Finite markov chains.

Khumawala, B. M. (1973). An efficient heuristic procedure for the uncapacitated warehouse location problem. *Naval Research Logistics (NRL) 20*(1), 109–121.

Lim, D., L. Li, and V. Linetsky (2012). Evaluating callable and putable bonds: an eigenfunction expansion approach. *Journal of Economic Dynamics and Control 36*(12), 1888–1908.

Lim, M., M. S. Daskin, A. Bassamboo, and S. Chopra (2010). A facility reliability problem: Formulation, properties, and algorithm. *Naval Research Logistics (NRL) 57*(1), 58–70.

Lopes, V. S. and C. L. Borges (2015). Impact of the Combined Integration of Wind Generation and Small Hydropower Plants on the System Reliability. *IEEE Transactions on Sustainable Energy 6*(3), 1169–1177.

Lu, M., L. Ran, and Z.-J. M. Shen (2015). Reliable facility location design under uncertain correlated disruptions. *Manufacturing & Service Operations Management 17*(4), 445–455.

Madan, D. B., P. P. Carr, and E. C. Chang (1998). The variance gamma process and option pricing. *Review of Finance 2*(1), 79–105.

Malladi, V., R. Mendoza-Arriaga, and S. Tompaidis (2017). Modeling dependent outages of electricity power plants.

Mavrotas, G., D. Diakoulaki, and L. Papayannakis (1999). An Energy Planning Approach Based on Mixed 0–1 Multiple Objective Linear Programming. *International Transactions in Operational Research 6*(2), 231–244.

McCurry, J. (2018, May). Japan faces power shortages due to nuclear shutdowns — environment — the guardian. `https://www.theguardian.com/world/2011/jul/06/japan-power-shortages-nuclear-shutdowns`. (Accessed on 05/07/2018).

Melo, M. T., S. Nickel, and F. Saldanha-Da-Gama (2009). Facility location and supply chain management–a review. *European journal of operational research 196*(2), 401–412.

Mendoza-Arriaga, R., P. Carr, and V. Linetsky (2010). Time-changed markov processes in unified credit-equity modeling. *Mathematical Finance 20*(4), 527–569.

Mendoza-Arriaga, R. and V. Linetsky (2014). Time-Changed CIR Default Intensities with Two-Sided Mean-Reverting Jumps. *The Annals of Applied Probability 24*(2), 811–856.

Mendoza-Arriaga, R. and V. Linetsky (2016). Multivariate Subordination of Markov Processes with Financial Applications. *Mathematical Finance 26*(4), 699–747.

Mendoza-Arriaga, R., V. Linetsky, et al. (2014). Time-changed cir default intensities with two-sided mean-reverting jumps. *The Annals of Applied Probability 24*(2), 811–856.

Meyer, C. D. (2001). *Matrix Analysis and Applied Linear Algebra Book and Solutions Manual*. Philadelphia: Society for Industrial and Applied Mathematics.

Murray, W. and K.-M. Ng (2010). An algorithm for nonlinear optimization problems with binary variables. *Computational Optimization and Applications 47*(2), 257–288.

Nasiri, F. and G. Huang (2007). Capacity Planning for Electricity Generation with Energy-Environmental Targets.

Phillips, R. S. (1952). On the generation of semigroups of linear operators. *Pacific Journal of Mathematics 2*(3), 343–369.

Rocha, L. F. and C. L. Borges (2010). Probabilistic Generation and Interruption Costs and Other Economic Aspects Related to Distributed Generation integration. In *Power and Energy Society General Meeting, 2010 IEEE*, pp. 1–6.

Saghafian, S. and M. P. Van Oyen (2012). The value of flexible backup suppliers and disruption risk information: newsvendor analysis with recourse. *IIE Transactions 44*(10), 834–867.

Santiváñez, J. A. and H. J. Carlo. Reliable capacitated facility location problem with service levels. *EURO Journal on Transportation and Logistics*, 1–27.

Santoso, T., S. Ahmed, M. Goetschalckx, and A. Shapiro (2005). A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research 167*(1), 96–115.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

Sen, G., M. Krishnamoorthy, N. Rangaraj, and V. Narayanan (2016). Facility location models to locate data in information networks: a literature review. *Annals of Operations Research 246*(1-2), 313–348.

Sharon Edelson (2017). Hurricane irma could pack a $1.45 billion retail punch - planalytics, inc. `http://www.planalytics.com/hurricane-irma-pack-1-45-billion-retail-punch/`. (Accessed on 09/29/2017).

Shen, Z.-J. M., R. L. Zhan, and J. Zhang (2011). The reliable facility location problem: Formulations, heuristics, and approximation algorithms. *INFORMS Journal on Computing 23*(3), 470–482.

Snyder, L. V. (2006). Facility location under uncertainty: a review. *IIE Transactions 38*(7), 547–564.

Snyder, L. V. and M. S. Daskin (2005). Reliability models for facility location: the expected failure cost case. *Transportation Science 39*(3), 400–416.

Sun, Y., R. Mendoza-Arriaga, and V. Linetsky (2013). Marshall-Olkin Distributions, Subordinators, Efficient Simulation, and Applications to Credit Risk. *Preprint*.

Sun, Y., R. Mendoza-Arriaga, and V. Linetsky (2017). Marshallolkin distributions, subordinators, efficient simulation, and applications to credit risk. *Advances in Applied Probability 49*(2), 481514.

Van Roy, T. J. and D. Erlenkotter (1982). A dual-based procedure for dynamic facility location. *Management Science 28*(10), 1091–1105.

Wang, Y., C. Chen, J. Wang, and R. Baldick (2016). Research on Resilience of Power Systems Under Natural Disasters—a Review. *IEEE Transactions on Power Systems 31*(2), 1604–1613.

Wolff, R. W. (1989). *Stochastic modeling and the theory of queues*. Pearson College Division.

Yang, C. and S. Tompaidis (2013). An Iterative Simulation Approach for Solving Stochastic Control Problems in Finance. *Preprint. Available at SSRN 2295591*.