

Large Scale Prediction Models and Algorithms

by

Matthieu Monsch

B.Sc., Ecole Polytechnique (2009)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

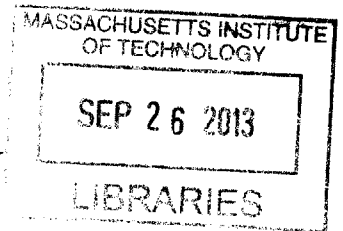
Doctor of Philosophy in Operations Research

at the

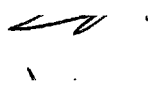
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

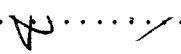
September 2013

ARCHIVES




© Massachusetts Institute of Technology 2013. All rights reserved.

Author

Sloan School of Management
August 9, 2013

Certified by

Vivek Farias
Robert N. Noyce Professor of Management
Thesis Supervisor

Certified by
.....
Georgia Perakis
William F. Pounds Professor of Management
Thesis Supervisor

Accepted by

Dimitris Bertsimas
Co-Director, Operations Research Center

Large Scale Prediction Models and Algorithms

by

Matthieu Monsch

Submitted to the Sloan School of Management
on August 9, 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

Over 90% of the data available across the world has been produced over the last two years, and the trend is increasing. It has therefore become paramount to develop algorithms which are able to scale to very high dimensions. In this thesis we are interested in showing how we can use structural properties of a given problem to come up with models applicable in practice, while keeping most of the value of a large data set. Our first application provides a provably near-optimal pricing strategy under large-scale competition, and our second focuses on capturing the interactions between extreme weather and damage to the power grid from large historical logs.

The first part of this thesis is focused on modeling competition in Revenue Management (RM) problems. RM is used extensively across a swathe of industries, ranging from airlines to the hospitality industry to retail, and the internet has, by reducing search costs for customers, potentially added a new challenge to the design and practice of RM strategies: accounting for competition. This work considers a novel approach to dynamic pricing in the face of competition that is intuitive, tractable and leads to asymptotically optimal equilibria. We also provide empirical support for the notion of equilibrium we posit.

The second part of this thesis was done in collaboration with a utility company in the North East of the United States. In recent years, there has been a number of powerful storms that led to extensive power outages. We provide a unified framework to help power companies reduce the duration of such outages. We first train a data-driven model to predict the extent and location of damage from weather forecasts. This information is then used in a robust optimization model to optimally dispatch repair crews ahead of time. Finally, we build an algorithm that uses incoming customer calls to compute the likelihood of damage at any point in the electrical network.

Thesis Supervisor: Vivek Farias
Title: Robert N. Noyce Professor of Management

Thesis Supervisor: Georgia Perakis
Title: William F. Pounds Professor of Management

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisors, Prof. Vivek Farias and Prof. Georgia Perakis for everything they have done for me. Their support, their guidance and their ever present enthusiasm have been instrumental not only in completing this thesis but also in making my stay at MIT a wonderfully enriching experience. It has truly been a privilege to work with both of them over the past five years.

I would also like to thank Prof. Itai Ashlagi who served on my thesis committee. I thoroughly enjoyed our discussions and am thankful for the opportunity I had to present this work to his students.

I am very grateful to Laura Rose, Andrew Carvalho, and Paulette Mosley for their administrative assistance. A special thank you to Laura for her remarkably flexible schedule and for always finding ways to make things work.

My time in the Operations Research Center wouldn't have been nearly as enjoyable and memorable without Viroth 'Ta' Chiraphadhanakul, Joline Uichanco, Gareth Williams, Yuan Zhong, and all my fellow 'ORCers'. I feel lucky to have made such great friends.

Thank you to Hussam Zebian, Meelap Shah, Ahmed Helal, Burcu Erkmen, Youssef Mroueh, Steve Voinea, Jit Hin Tan, and my wonderful friends at Sidney Pacific. You made Boston my home and I have grown so much from getting to know all of you. Thank you also to Janet Komatsu for being there when it mattered the most.

MIT was a unique opportunity to meet incredible people. I am forever grateful to Claude Grunitzky and David Ly for being such model teammates, and to Hoda Eydgahi for inspiring me more than words can express.

Finally, my deepest gratitude goes to my family for their love and support. There is no better person to dedicate this work to than my mother, to whom I owe so much.

Contents

Introduction	13
Notation	14
I Dynamic Pricing in Revenue Management:	
A Tractable Notion of Equilibrium	15
1 Introduction	16
1.1 Motivation	16
1.2 Potential Equilibrium Concepts	18
1.3 Contributions	19
1.4 Outline	20
2 Model	21
2.1 The Game	21
2.2 Assumptions	22
2.3 Strategies	23
2.4 Value Functions	26
2.5 Best Response Dynamics	29
3 Stationary Equilibrium	31
3.1 Simplicity	31
3.2 Learnability	33

3.3	Asymptotic Optimality	34
3.4	Empirical Support	36
4	Simulations	39
5	Conclusions and Future Work	41
6	Proofs	42
6.1	Proof of Theorem 3.2.1	43
6.2	Proof of Theorem 3.3.1	48
II	Predicting Power Outages in Electricity Networks:	
A	Framework to Reduce Restoration Time After Large Storms	60
1	Introduction	61
1.1	Motivation	62
1.2	Literature Review	62
1.3	Contributions and Outline	65
2	Predicting Outages	67
2.1	Data	68
2.1.1	Electrical Network Data	68
2.1.2	Historical Weather Data	70
2.1.3	Historical Outage Data	74
2.2	Model	74
2.2.1	Formulation	74
2.2.2	Without Censoring	77
2.2.3	With Censoring	78
2.3	Application	78
2.3.1	Segment Clustering	79
2.3.2	Using Actual Weather Logs	79
2.3.3	Using Weather Forecasts	80
2.3.4	Implementation	84

3	Optimizing Crew Dispatch	88
3.1	Deterministic Crew Optimization	88
3.1.1	Master Formulation	88
3.1.2	Relaxed Formulation	91
3.1.3	Comparison	92
3.2	Robust Optimization	93
3.2.1	Using Box Constraints	94
3.2.2	Using Bertsimas-Sim Uncertainty Sets	95
3.3	Calibrating the Model for the Power Utility	97
4	Estimating Damage From Calls	103
4.1	Model	103
4.1.1	Network	103
4.1.2	States	104
4.1.3	Evidence	106
4.2	Algorithm	107
4.2.1	Results	109
4.2.2	Proofs	110
4.3	Application to the Power Utility	115
4.3.1	The Graph	116
4.3.2	States and Transition Probabilities	118
4.3.3	Implementation	122
5	Conclusions and Future Work	126
	Concluding Remarks	128
	Bibliography	129

List of Figures

3.1	Price (in dollars) for four different substitutable flights from New York to San Francisco as a function of the number of days until departure	38
3.2	Average price (in dollars) over the same four flights from New York to San Francisco as a function of the number of days until departure	38
2.1	High level representation of the electricity network	69
2.2	Density of the electrical network in Massachusetts	71
2.3	Landcover distribution among segments	72
2.4	Map of weather stations	73
2.5	Outage prediction comparison sample map	81
2.6	Wind speed forecasting error statistics	85
2.7	Wind speed forecasting error histogram	85
2.8	Online prediction tool, test interface	86
2.9	Online prediction tool, sample run results	86
3.1	Job to platform assignment representation	100
3.2	Crew workloads for three randomly sampled sets	101
3.3	Historical Non-Storm Outage Histogram	102
3.4	Simulated historical Non-Storm Outage Histogram	102
4.1	Sample directed tree	105
4.2	Illustration of partial updates from new evidence ('data')	111
4.3	Histogram of the number of devices per circuit.	119
4.4	Histogram of the number of customers per circuit.	119

4.5	In-browser real-time damage estimation tool. View when no customer has lost power on the circuit.	123
4.6	In-browser real-time damage estimation tool. Updating with a new customer call.	123
4.7	In-browser real-time damage estimation tool. View after having notified by several customers of power loss. Note that there are several damage points noted as likely, such as the one selected.	124
4.8	In-browser real-time damage estimation tool. Same situation as above, with graph vertices sized by the probability of the corresponding segment being damaged.	124

List of Tables

4.1	Optimality gap of Nash equilibria and stationary equilibria for different capacity and time-horizon values	39
2.1	Storm outage data	74
2.2	Out of sample correlation for outage prediction	81
2.3	Out of sample correlation for outage prediction using ‘day of’ forecast data. The smaller numbers indicate the loss compared to the correlations using actual weather data from Table 2.2	83
2.4	Out of sample correlation for outage prediction at the platform level using forecast data from different days ahead	83
3.1	Master Formulation Variable Notation	89
3.2	Relaxed Formulation Variable Notation	91
3.3	Master Formulation vs Relaxed Formulation Results	93

Introduction

The world's technological per-capita capacity to store data has roughly doubled every three years for the past thirty years¹. Processing power has increased comparatively slowly, and consequently there has been a strong push towards scalable algorithms which are able to make the most of the vast amount of data now available.

However, many problems, both theoretical and practical, are not currently amenable to be solved by such algorithms. Moreover, in spite of unprecedented data collection, key data often remains missing. It is therefore paramount to find concise ways to combine the data available with additional information about the problem at hand. By choosing an adequate model we can significantly reduce the amount of data necessary and sometimes even make up for missing data.

In this thesis we demonstrate this approach through two separate examples. Our first application focuses on finding a tractable way to model competition in dynamic pricing problems. We then consider the problem of reducing power restoration times after major storms by predicting where power outages are most likely to occur.

¹Hilbert, Martin; Lpez, Priscila (2011). "The World's Technological Capacity to Store, Communicate, and Compute Information". *Science* 332 (6025): 6065

Notations

In the rest of this thesis, we will make use of the following notations:

- $[v]_i$ is the i -th component of vector v .
- $(v_i)_i$ is the vector such that $\forall i, [(v_i)_i]_i = v_i$.
- $v_{-i} = (v_j)_{j \neq i}$ so that $v = (v_i, v_{-i})$.
- $(x)_+$ is the positive part of x .
- $\times_i \mathcal{S}_i$ is the Cartesian product of sets \mathcal{S}_i .
- $\mathbb{N}_N = \{0, \dots, N\}$ and $\mathbb{N}_N^* = \{1, \dots, N\}$.

Part I

Dynamic Pricing in Revenue

Management:

A Tractable Notion of Equilibrium

1 Introduction

In this first part of the thesis, we study finite horizon stochastic dynamic games with a large number of players. Such games can be used to model a wide array of applications such as dynamic pricing decisions under competition. Stochastic games with many players are usually difficult to study in the traditional game theoretical framework (e.g. because of the 'curse of dimensionality'). We focus our analysis on a novel approach to model competition using stationary policies.

Under these dynamics, each firm (player) is only influenced by the initial average market price. We find necessary conditions for existence and convergence to equilibria. Moreover, we show bounds on optimal profit gains from unilateral deviations under such equilibria. We finally present a numerical study to test the robustness of our assumptions to real problems.

1.1 Motivation

The modern practice of Revenue Management (RM) extends across a swathe of industries, ranging from airlines to the hospitality industry to retail. 'Solving' a revenue management problem frequently entails (among many other things) the solution of some form of dynamic optimization problem. Such problems capture, in essence, the tension of allocating a limited supply of some good to uncertain demand, frequently using price as a lever. Often, these optimization problems are high dimensional. There exists a vast body of research devoted to the solution of these problems (for example in [26, 21, 20]). Now consider for a moment a customer looking to purchase an air ticket from New York to San Francisco departing

at some specific time. The customer will typically face a plethora of alternatives and with web tools available today be well aware of these alternatives. The internet has, by reducing search costs for customers, potentially added a new challenge to the design and practice of RM strategies – accounting for competition. This chapter considers a novel approach to modeling dynamic pricing in the face of competition. Our approach makes precise the notion that firms practicing RM can often compete ‘implicitly’ [41]. The approach is pragmatic, and easily compared to other well studied, but potentially less pragmatic notions of equilibria in dynamic RM games.

Modeling and accounting for competition in RM is challenging. Whereas in a monopolistic setting, a firm pricing a limited quantity of some perishable product might employ a pricing strategy contingent on remaining inventory and time to go, in a competitive setting the pricing decisions of competing firms immediately become relevant to any pricing strategy, and therefore indirectly, so do the inventory levels of these competing firms. To begin with, firms are unlikely to reveal information relevant to remnant inventory over time. Second, even assuming competitors’ inventory levels are visible, it is easy to see that the pricing policy called for will be high dimensional and thus hard to compute or execute. With this in mind, a viable proposal must satisfy several criteria:

1. The firm’s dynamic pricing strategy continues to remain contingent on the firm’s inventory levels; in particular, the firm responds to large demand shocks it may face.
2. The pricing strategy is learnable and emerges ‘naturally’.
3. The pricing strategy prescribed is at least an approximate notion of equilibrium in the sense that if all competing firms adopted the policy, there is little to gain from a unilateral deviation.

We consider an oligopoly where each individual firm is endowed with some initial inventory that it must sell over a finite horizon. The demand faced by a given firm is a function of the price it posts as well as those posted by its competitors; we assume that this demand system satisfies the law of demand. Firms can adjust their prices over the course of the selling horizon; they may do this to effectively ration their own inventory (the traditional role of RM) and also to respond to their competitors. This describes a dynamic stochastic

game for which one might consider several equilibrium concepts.

1.2 Potential Equilibrium Concepts

One notion of equilibrium in the game above is that of a Markov Perfect Equilibrium (MPE). Assuming inventory levels are common knowledge, an MPE must specify a pricing policy for each firm contingent on that firm's inventory level as well as those of its competitors. Due to natural complementarities in the game above, it is not difficult to show that an MPE exists and that a number of natural learning dynamics converge to this equilibrium; see for instance [16, 3]. While attractive, this notion is a poor fit to our requirements:

1. Arriving at such an equilibrium is hard computationally.
2. The resulting policies are high dimensional and it is unreasonable to expect firms to employ a dynamic pricing strategy with the information requirements of an MPE.

At the other end of the spectrum, one may consider employing the notion of a Nash equilibrium where firms commit to a *fixed* pricing strategy and do not deviate from this strategy over the course of the selling season; see [40, 27, 20]. This 'open loop' behavior is typically justified by making assumptions on model primitives that guarantee that a fixed pricing strategy is approximately optimal. While attractive in its simplicity, the notion that a firm does not respond to shocks it may face is perhaps unreasonable. Put another way, the absence of sub-game perfection is questionable.

Another notion of equilibrium which is closely related to this work is that of 'oblivious equilibrium' [43, 2, 1], with a subtle but important distinction as we will shortly note. This notion of equilibrium assumes that a firm commits to a dynamic pricing policy contingent on its own inventory levels but oblivious to the inventory levels of its competitors; instead the firm assumes a demand system that incorporates a 'fixed' competitor effect. This notion of equilibrium bears some resemblance to the notion that firms practicing RM effectively incorporate competition via the demand models they assume. At first it is unclear how such a prescription could preclude the possibility that a firm could gain by deviating from this strategy and tracking, say, the actual inventory levels of its competitors. What we exploit

is the fact that in demand systems of interest to us, at any given point in time, a firm is impacted by its competitors only through the ‘average’ of their prices at that point in time. While the price path of a single one of these competitors is by itself stochastic, the average of these price paths is less so and for a large number of competitors will look approximately deterministic.

1.3 Contributions

We make several contributions relative to the notion of equilibrium proposed:

1. **Simplicity:** the notion we posit effectively asks for firms to employ policies that are structurally similar to those a monopolist practicing dynamic pricing might use. Importantly, these policies allow firms to adjust prices in response to demand shocks they face.
2. **Learnability:** We show that best response dynamics converge to this equilibrium. It is interesting to note here, that if one employed a notion of equilibrium wherein the assumed average competitor price were allowed to vary with time it is no longer clear that best response dynamics are isotone so that the learnability of such an equilibrium in the RM setting is unclear. This latter notion would be in analogy with that studied in [43].
3. **An ‘Approximate’ Equilibrium:** We prove that gains from a unilateral deviation even with *perfect* information on competitor inventory levels are small when a firm faces a large number of competitors. This rests on establishing two phenomena: (a) the averaging of prices *across* competing firms at a given point in time (this is non-obvious since pricing policies across firms get correlated over time) and (b) a structural observation specific to dynamic pricing showing that the expected price of a given firm cannot vary much over the course of the selling season. This latter observation permits many of the attractive features offered by our notion of equilibrium.
4. **Empirical Support:** Using pricing data collected over several months for a set of substitutable transcontinental flights we observe that while the price trajectories for

an individual flight are potentially volatile, the average trajectory is far less so, and essentially constant. These are precisely the requirements for the notion of equilibrium we posit to be a reasonable one.

In summary, we posit a notion of equilibrium that we demonstrate is a natural candidate for dynamic RM applications.

1.4 Outline

This part of the thesis is structured as follows. In Chapter 2 we introduce the general framework for our dynamic competition model. In Chapter 3 we then define the notion of *stationary equilibrium* and show that it satisfies the four axioms posited earlier: simplicity, empirical support, learnability and near-optimal profits. Finally, in Chapter 4, we evaluate the performance of the stationary equilibrium compared to other common equilibrium concepts (Nash equilibrium, Markov perfect equilibrium). Finally, we conclude by summarizing the key features of the stationary equilibrium along with stating a few directions for future research. For clarity, proofs can be found in Chapter 6.

2 Model

In this chapter we present our dynamic competition model. We start by introducing the setting in which firms act (Subsection 2.1); in the following three sections, we then describe *how* firms act. We start by presenting a framework that captures the variability in information available for firms to make their pricing decisions (Subsection 2.3), then we present the corresponding notion of profit which each firm will try to maximize (Subsection 2.4), and finally we propose a simple and natural dynamic to model how firms respond to each other's pricing choices (Subsection 2.5). We conclude this chapter by presenting the assumptions under which we will study the model.

2.1 The Game

We consider a dynamic pricing problem for a single perishable product in a competitive market, modeled as a stochastic game, i.e. a tuple $\Gamma = (\mathcal{I}, \mathcal{T}, \mathcal{C}, \mathcal{A}, \lambda)$:

- **Firms.** There are finitely many firms indexed by $i \in \mathcal{I} = \mathbb{N}_I^*$.
- **Time.** The game is played in discrete time. Periods are indexed by $t \in \mathcal{T} = \mathbb{N}_T$. Time is indexed starting from the final stage up to the initial stage: $t = T$ corresponds to the initial time-step and $t = 0$ to the final time-step.
- **Capacities.** Initially, each firm i is endowed with some capacity $c_i \in \mathbb{N}$. At any time t , the current capacities of all the firms is represented by a vector $\mathbf{c}_t \in \mathbb{N}^n$ (e.g. $c_{i,t}$ denotes the capacity of firm i at period t). As will be made precise below, capacities are non-increasing over time and we can therefore define $\mathcal{C} = \times_i \mathcal{C}_i = \times_i \mathbb{N}_{c_{i,T}}$ as the set of all possible capacity vectors in the game.

- **Prices.** At each time-step t , each firm i sets a price $p_{i,t}$ from a set $\mathcal{A}_i = [0, p_i^\infty]$ with $p_i^\infty > 0$. We denote by \mathbf{p}_t the vector of all prices chosen by the firms at time t and by $\mathcal{A} = \times_i \mathcal{A}_i$ the set of all possible such vectors. Finally we assume that when firm i has no residual capacity, its price is always equal to p_i^∞ .
- **Sales.** Given a vector of prices \mathbf{p}_t and a vector of capacities \mathbf{c}_t , firm i 's probability $\lambda_i(\mathbf{p}_t, \mathbf{c}_t)$ of making a sale at time step $t > 0$ is given by the following equation:

$$\lambda_i(\mathbf{p}_t, \mathbf{c}_t) = \mathbb{1}_{\{c_{i,t} > 0\}} \min \left(1, \left(\alpha_i - \beta_i p_{i,t} + \sum_{j \neq i} \gamma_{ij} p_{j,t} \right)_+ \right) \quad (2.1)$$

where $\forall i, j, \alpha_i > 0, \beta_i > 0, \gamma_{ij} \geq 0$. In the final time-step $t = 0$, demand is assumed to be null: no firm can make a sale. If firm i makes a sale, it earns a payoff p_i and its capacity is decreased by 1 in the following time step: $c_{i,t-1} = c_{i,t} - 1$, otherwise its payoff is zero and capacity remains the same: $c_{i,t-1} = c_{i,t}$. To simplify notation, from now on, we will drop \mathbf{c}_t from λ_i when $c_{i,t} > 0$.

The objective of each firm during a game is to maximize its total expected profit over the course of the time horizon, that is the expected total profit generated from selling its initial inventory. Any remaining inventory at the end of the game ($t = 0$) is discarded at no value.

2.2 Assumptions

In order to guarantee that equation (2.1) is a valid probability and ensure the stability of the dynamics introduced previously, we require the following three assumptions:

Assumption 2.2.1. Diagonal dominance:

$$\forall i, \beta_i > \sum_{j \neq i} \gamma_{ij} \quad (2.2)$$

This assumption is standard in the dynamic pricing literature (cf. for example in [27, 14]) and is verified in most real world applications. A simple interpretation of this assumption

is a requirement that overall sales decrease when each firm increases its price: it is not possible for the total market demand to grow as prices go up.

Assumption 2.2.2. Upper bounds on prices:

$$\forall i, p_i^\infty = \frac{\alpha_i + \sum_{j \neq i} \gamma_{i,j} p_j^\infty}{\beta_i} \quad (2.3)$$

This assumption is without loss of generality as demand for each firm will also be equal to 0 for any price above the upper bound. It is easy to show that such a set of prices exists and is unique. Indeed, we can rewrite equation (2.3) as $M(p_i^\infty)_i = (\alpha_i)_i$ with M being the square matrix defined by $[M]_{ii} = \beta_i$ and $[M]_{ij} = -\gamma_{ij}$ for $i \neq j$. From Assumption 2.2.1, M is a diagonally dominant matrix and therefore invertible. Hence existence and uniqueness of the $(p_i^\infty)_i$.

Assumption 2.2.3. Valid demand:

$$\forall i, \beta_i p_i^\infty \leq 1 \quad (2.4)$$

Note that this can be rewritten in terms of the $\alpha_i, \beta_i, \gamma_{ij}$ using the matrix M introduced above: $\forall i, M^{-1}(\alpha_i)_i \leq (1/\beta_i)_i$. This assumption guarantees that the expression inside the minimum in the demand function λ in equation (2.1) is always smaller than 1. From equation (2.4) and (2.2), $\forall(i, \mathbf{p}), \alpha_i - \beta_i p_i + \sum_{j \neq i} \gamma_{ij} p_j \leq \alpha_i + \sum_{j \neq i} \gamma_{ij} p_j \leq \alpha_i + \sum_{j \neq i} \gamma_{ij} p_j^\infty = \beta_i p_i^\infty \leq 1$. Therefore $\forall(i, \mathbf{p}, \mathbf{c}), \lambda_i(\mathbf{p}, \mathbf{c}) \in [0, 1]$. This allow us the rewrite the demand more simply as follows:

$$\lambda_i(\mathbf{p}_t, \mathbf{c}_t) = \mathbb{1}_{\{c_{i,t} > 0\}} \left(\alpha_i - \beta_i p_{i,t} + \sum_{j \neq i} \gamma_{ij} p_{j,t} \right)_+ \quad (2.5)$$

2.3 Strategies

In this section, we define the notion of *strategy* which captures all the information necessary to describe a firm's behavior. In general, a firm's strategy could depend on many things (ex-

ternal market indicators, past events, predicted supply outages, etc.), however models very quickly become intractable as the number of such factors grows. This is why, traditionally in stochastic games (for example in [19, 35]), firm strategies are assumed to be Markovian: a firm’s actions only depend on the current state of the game. In this paper, we follow this line of thought and focus our analysis on Markovian strategies: the pricing decisions made at a given time-step by a firm depend solely on its and its competitors’ capacities at that time-step.

Definition 2.3.1. A *strategy* $\mu_i : \mathcal{C} \times \mathcal{T} \rightarrow \mathcal{A}_i$ is a mapping from capacities and time to a set of prices which entirely determines a firm’s pricing decisions. Given capacities \mathbf{c}_t and time t , firm i will choose price $\mu_i(\mathbf{c}_t, t)$. We denote by $\boldsymbol{\mu} = (\mu_i)_i$ the vector of strategies across firms. Finally, the combination of a capacity vector \mathbf{c} and a time-step t will often be referred to as a *state* of the game.

Given a strategy vector $\boldsymbol{\mu}$, we are therefore able to determine each firm’s prices in any possible state of the game. In particular, since demand at any given time-step is only a function of the firms’ current capacities and prices, given a strategy vector $\boldsymbol{\mu}$, we can write sale probabilities as a function of the state:

$$\begin{aligned} \lambda_i^\mu : \mathcal{C} \times \mathcal{T} &\rightarrow \mathbb{R} \\ (\mathbf{c}_t, t) &\mapsto \lambda_i(\boldsymbol{\mu}(\mathbf{c}_t, t), \mathbf{c}_t) \end{aligned}$$

It is now straightforward to see that the random vector of capacities at each time-step can be modeled as a Markov chain.

Definition 2.3.2. Define the *capacity vector* $\mathbf{C}_t = (C_{i,t})_i$ as the random vector corresponding to all the firms’ capacities at time t . Since strategies are Markovian, the capacity vector at time-step t only depends on the time-step t and its value at the previous time-step $t + 1$, therefore is a (non-homogeneous) Markov Chain on \mathcal{C} with the following transition rates

and boundary condition:

$$\begin{aligned} \mathbf{C}_T &= \mathbf{c}_T \\ \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, C_{i,t-1} &= \begin{cases} C_{i,t} - 1, & \text{w.p. } \lambda_i^\mu(\mathbf{C}_t, t) \\ C_{i,t}, & \text{otherwise} \end{cases} \end{aligned}$$

Notice that, up until now, we have assumed that firm strategies depend on the *entire* vector of current capacities. This means that in order to set its price, each firm has to have access to all of its competitors' capacity levels. In many cases in practice, as was mentioned above in Section 1.1, this is an unreasonable assumption. Furthermore, even if firms were able to observe those capacity levels, there might be too many to store effectively and we might want to restrict the number of competitors' capacities which firms need to keep track of. For this reason we introduce the concept of *observed capacities*:

Definition 2.3.3. We denote by $\sigma_i(\mathbf{c}_t)$ the vector of *observed capacities* by firm i at time t where $\sigma_i : \mathcal{C} \rightarrow \mathcal{C}$ is a measurable function such that $\sigma_i \circ \sigma_i = \text{id}$. We also define $\forall i \in \mathcal{I}, \mathcal{S}_i = \sigma_i(\mathcal{C}_i)$ as the set of capacities observable by firm i .

Given an observation vector σ , we say that a strategy vector μ is σ -observable if it is uniquely determined by each firm's observations. More formally, μ is σ -observable if and only if $\exists \mu' : \sigma_i(\mathcal{C}) \times \mathcal{T} \rightarrow \mathbb{R}$ such that $\forall(\mathbf{c}, t), \mu_i(\mathbf{c}, t) = \mu'(\sigma_i(\mathbf{c}), t)$ or equivalently, if and only if $\forall(i, t), \mu_i(\cdot, t) : (\mathcal{C}, \Sigma_i) \rightarrow (\mathbb{R}, \mathcal{B})$ is measurable, where Σ_i is the σ -algebra on \mathcal{C} generated by σ_i and \mathcal{B} the Borel algebra on \mathbb{R} . For convenience, we define $\mu_i^{\sigma_j}$ such that $\mu_i^{\sigma_j}(\mathbf{c}_t, t) = \mu_i(\sigma_j(\mathbf{c}_t), t)$ and the associated $\mu^\sigma = (\mu_i^{\sigma_i})_i$, $\mu^{\sigma_j} = (\mu_i^{\sigma_j})_i$.

The concept of observed capacities allows our model to capture a wide array of strategies. One could for example consider fixed-price strategies where firms post a fixed price until they run out of inventory by choosing σ such that $[\sigma_i(\mathbf{c}_t)]_i = \mathbb{1}_{\{c_{i,t} > 0\}}$ and $[\sigma_i(\mathbf{c}_t)]_j = c_{j,T}, \forall j \neq i$. Another less trivial example would be strategies contingent uniquely on a firm's own inventory, in which case we would pick σ such that $[\sigma_i(\mathbf{c}_t)]_j = c_{i,t}$ and $[\sigma_i(\mathbf{c}_t)]_j = c_{j,T}, \forall j \neq i$. Finally, this modeling choice is without loss of generality since by picking σ equal to the identity, we retrieve the full information model.

Note finally that even though firms are no longer in general able to observe the capacity levels of their competitors at any time step t , the initial capacity vector \mathbf{c}_T is still known by all firms.

2.4 Value Functions

Given the information that is has available, the goal of each firm is to maximize its expected profit i.e. maximize its total revenue from sold inventory over the course of a game.

The standard form to express this expected profit is through the use of a *value function* which represents the profit to go from any state of the game. In particular, in our setting, the value function maps a capacity vector \mathbf{c} and a time-step t to the expected profit from that state and takes the following form:

Definition 2.4.1.

$$V_{\mu_i, \mu_{-i}} : \mathcal{C} \times \mathcal{T} \rightarrow \mathbb{R}$$

$$(\mathbf{c}, t) \mapsto \mathbb{E} \left[\sum_{t'=t}^T \mu_i(\mathbf{C}_{t'}, t') \lambda_i^\mu(\mathbf{C}_{t'}, t') \middle| \mathbf{C}_t = \mathbf{c} \right]$$

However, recall that only some capacity vectors can be observed by each firm, firms are therefore in general unable to know the true vector of capacities at a given time step. We must amend the previous equation to take this fact into account. Given observations σ and strategies μ , we define $V_{\mu_i, \mu_{-i}}^\sigma : \mathcal{S}_i \times \mathcal{T} \rightarrow \mathbb{R}$ as the value function of firm i from any given observable state (recall that by definition, \mathcal{S}_i is the set of all capacities observable by firm i):

Definition 2.4.2.

$$V_{\mu_i, \mu_{-i}}^\sigma : \mathcal{S}_i \times \mathcal{T} \rightarrow \mathbb{R}$$

$$(\mathbf{s}_i, t) \mapsto \mathbb{E} \left[\sum_{t'=t}^T \mu_i(\mathbf{C}_{t'}, t') \lambda_i^\mu(\mathbf{C}_{t'}, t') \middle| \sigma_i(\mathbf{C}_t) = \mathbf{s}_i \right]$$

Given a strategy vector μ and an observed state s_i , firm i is able to infer the underlying probability distribution of the true capacity vector c (since all firms know the initial vector of capacities c_T) and from there compute its expected profit $V_{\mu_i, \mu_{-i}}^\sigma(s_i, t)$. Note that, in general, $\sigma_i(C_t)$ is not a Markov chain.

However, computing this new value function still requires that each firm know each of its competitors' entire strategy function. Indeed, in order to compute its expected demand at any time-step, a firm must know what each of its competitors' price will be. In most cases this is an unreasonable assumption and firms aren't able to infer how their competitors make their pricing decision which makes it impossible for them to compute their expected profit using the previous equation. We alleviate this difficulty by allowing firms to calculate their profit against an estimated but simpler demand function which only depends on variables it can observe:

Definition 2.4.3. Given observations σ and a strategy vector μ , each firm i will assume that demand is given by the function below, where η is the *assumed demand*:

$$\begin{aligned} \eta_i^\mu : \mathcal{S}_i \times \mathcal{T} &\rightarrow \mathbb{R} \\ (s_i, t) &\mapsto [\eta(\mu)]_i(s_i, t) \end{aligned}$$

Similarly to Definition 2.3.2, we can define a random vector corresponding to how capacities would evolve under assumed demand η . Note that Definition 2.4.3 allows each firm to assume a different demand function, therefore the random vector of *assumed capacities* is firm dependent: each firm will have its own Markov chain $\{S_t\}_t$ of capacities associated with its assumed demand η_i^μ .

Definition 2.4.4. Define the *assumed capacity vector* $S_t = (S_{i,t})_i$ as the random vector corresponding to all the firms' assumed capacities at time t . Similarly to C_t , we know that S_t is a (non-homogeneous) Markov Chain on \mathcal{C} with the following transition rates and

boundary condition:

$$[\mathbf{S}_T]_i = \sigma_i(c_T)$$

$$\forall i \in \mathcal{I}, \forall t \in \mathcal{T}, S_{i,t-1} = \begin{cases} S_{i,t} - 1, & \text{w.p. } \eta_i^\mu(\mathbf{S}_t, t) \\ S_{i,t}, & \text{otherwise} \end{cases}$$

It is important to observe that in general \mathbf{S}_t does not correspond to the true vector of observed capacities, i.e. $\mathbf{S}_t \neq \sigma_i(C_t)$ for some firm i and time-step t . Indeed, by construction, the demand assumed by a firm only depends on what it can observe, whereas in truth the demand is still ruled by equation (2.1), which depends on the capacities and strategies of all firms (and \mathbf{S}_t is always a Markov chain, $\sigma_i(C_t)$ in general is not). Moreover, the assumed capacities across firms do not necessarily match: there can be two firms i and j such that at time-step t , $S_{j,t} \neq S_{i,t}$.

For the aforementioned reasons, it is simpler to think of each assumed capacity vector only in the context of its corresponding firm. From the point of view of a firm, it provides an observable and simpler (yet generally inexact) model for how its competitors' capacities and prices evolve. More importantly, it enables each firm to compute its *assumed value function*:

Definition 2.4.5.

$$V_{\mu_i, \mu_{-i}}^{\sigma, \eta} : \mathcal{S}_i \times \mathcal{T} \rightarrow \mathbb{R}$$

$$(\mathbf{s}_i, t) \mapsto \mathbb{E} \left[\sum_{t'=t}^T \mu_i(\mathbf{S}_{t'}, t') \eta_{i,i}^\mu(\mathbf{S}_{t'}, t') \mid \mathbf{S}_t = \mathbf{s}_i \right]$$

The expected profit computed using this assumed value function is in general, of course, also inexact. That is, it differs from the true expected profit as computed using Definition 2.4.1. This is to be expected since its underlying demand doesn't correspond to the true demand. However, by controlling the shape of the assumed demand, we are able to strike a balance between tractability and precision.

2.5 Best Response Dynamics

In this section we are interested in how firms learn and react to each other's strategies. Indeed, through the demand model (expressed in equation (2.1)), each firm's pricing decisions can potentially impact all of the other firms' profits. For example if firm i globally decreases its prices, the demand rate λ_j observed by any firm $j \neq i$ will increase (it will remain constant if the corresponding γ_{ij} is equal to 0). It therefore stands to reason that firms take into account their competitors' strategies in their own pricing decisions: if their competitors' strategies change, the strategy which maximizes their own profit will also change.

We model this dynamic process iteratively. Firms play the game described above (in Section 2.1) repeatedly and each time choose a new pricing strategy which maximizes their expected profit given the current strategies used by their competitor, i.e. the *best response* to the strategies currently used. We use the notion of best response as a natural proxy for how rational firms behave in practice:

Definition 2.5.1. Given observations σ and assumed demand η , we define the (σ, η) -best response operator $\Phi^{\sigma, \eta}$ as follows: $\forall i \in \mathcal{I}, \forall \mathbf{s}_i \in \mathcal{S}, \forall t \in \mathcal{T}$,

$$[\Phi^{\sigma, \eta}(\boldsymbol{\mu})]_i(\mathbf{s}_i, t) = \arg \sup_{\mu^i} V_{(\mu^i, \boldsymbol{\mu}_{-i})}^{\sigma, \eta}(\mathbf{s}_i, t)$$

Using standard compactness arguments, it can be shown that $\Phi^{\sigma, \eta}$ is non-empty and we denote by $\phi^{\sigma, \eta}(\boldsymbol{\mu})$ an element of $\Phi^{\sigma, \eta}(\boldsymbol{\mu})$ (referred to as a best response to strategy $\boldsymbol{\mu}$).

From any initial vector of strategies $\boldsymbol{\mu}^0$, we can apply the best response operator iteratively and define iterated (σ, η) -best response as follows: $\forall n \geq 1$,

$$\begin{aligned} \boldsymbol{\mu}^n &= \phi^{\sigma, \eta}(\boldsymbol{\mu}^{n-1}) \\ &= (\phi^{\sigma, \eta})^n(\boldsymbol{\mu}^0) \text{ in the composition sense} \end{aligned}$$

Note that all firms update their strategy at each iteration (by computing their own best response). This implies that in general a firm's strategy choice won't actually be optimal in

the next realization of the game since its competitors' pricing policies have changed since the last iteration. In this chapter, we are particularly interested in the reverse case: where a firm's best response remains optimal in the next iteration of the game. We then say that we have reached an *equilibrium*:

Definition 2.5.2. A strategy vector μ is a (σ, η) -equilibrium if and only if it is a (σ, η) -best response to itself, i.e. a fixed point of the (σ, η) -best response operator:

$$\mu \in \Phi^{\sigma, \eta}(\mu)$$

Intuitively, an equilibrium corresponds to a situation where no single firm can increase its expected profit by changing its strategy. In other words, once such an equilibrium has been reached, a firm's strategy is already a best response to its competitors' strategies. In other words, best response dynamics have converged.

Finally, we note that, in general, the convergence of best response dynamics, and the existence of a corresponding equilibrium are hard problems.

3 Stationary Equilibrium

By appropriately choosing the observations vector σ and the assumed demand η , we are able to cover a wide spectrum of equilibria. In particular, as mentioned in the previous chapter, we are able to recover two extreme cases: a fixed price Nash equilibrium, eminently tractable but arguably too simple, and a Markov perfect equilibrium over all the firms' capacities, optimal among Markovian strategies but unpractical in most settings. Here we aim to find the right balance between computability and performance.

3.1 Simplicity

We start with the following two requirements: a firm needs to be able to respond to shocks in its own demand (as opposed for instance to the fixed price Nash equilibrium), and a firm has to be able to optimize its pricing decisions while taking competition into account but with minimal information from its competitors (as opposed to the full Markov perfect equilibrium setting).

The first condition requires that a firm be able to track its own inventory, which is in practice a reasonable assumption (most industries have implemented some kind of automated inventory tracking, be it in the airline industry, in the hotel business or even more recently in large scale grocery chains, etc.). This motivates a choice of observations vector of the form: $[\sigma_i(\mathbf{c}_t)]_j = c_{i,t}$ and $[\sigma_j(\mathbf{c}_t)]_j = c_{j,T}, \forall j \neq i$. Intuitively, each firm is able to observe its current inventory level but assumes nothing about its competitors' inventory (for example by letting it remain at their initial levels).

In order to fulfill the second condition, one needs to find a way to model the effect com-

petition has on a firm's demand in a concise way. A step in that direction is to replace the 'competition factor' $\sum_{j \neq i} \gamma_{ij} \mu_j(c_{j,t}, t)$ in the demand function by its expected value $\mathbb{E} \left[\sum_{j \neq i} \gamma_{ij} \mu_j(c_{j,t}, t) \right]$. This is the approach taken in the oblivious equilibrium literature ([43, 2, 1]). This however still has the drawback of requiring each firm to know its competitors' pricing strategy across all time-steps. We therefore go one step further and assume that this competition factor remains constant over time, equal to its initial value $\sum_{j \neq i} \gamma_{ij} \mu_j(c_j, T)$. Even though the resulting assumed demand will diverge likely more from the true demand than a time dependent one, we argue in the rest of this chapter that there is in fact sufficient information in the initial time-step to find near-optimal pricing strategies.

This finally leads us to define the following equilibrium concept, which satisfies both our previous criteria:

Definition 3.1.1. Let $\Gamma = (\mathcal{I}, \mathcal{T}, \mathcal{C}, \mathcal{A}, \lambda)$ be a game as described in Section 2.1. Let now σ^s and η^s be the observations vector and assumed demand respectively such that: $\forall (i, j) \in \mathcal{I}^2, \forall t \in \mathcal{T}$,

$$[\sigma_i^s(c_t)]_j = \begin{cases} c_{i,t}, & \text{if } j = i \\ c_{j,T}, & \text{otherwise} \end{cases} \quad (3.1)$$

$$[\eta^s(\mu)]_i(s_i, t) = \alpha_i - \beta_i \mu_i(s_i, t) + \sum_{j \neq i} \gamma_{ij} \mu_j^{\sigma^s}(c_T, T) \quad (3.2)$$

We define a *stationary equilibrium* as a (σ^s, η^s) -equilibrium.

In summary, firm i 's pricing decision relies on the following three factors:

- the number of time-steps to go t ,
- its own inventory level $c_{i,t}$,
- its 'competition factor' $\sum_{j \neq i} \gamma_{ij} \mu_j(c_j, T)$.

Note that since competition between firms is captured through a single term, the 'competition factor', the dimension of the state space of the dynamic problem a firm would need to solve to optimize its revenue is greatly reduced (it is actually of dimension two). In

particular, notice that its size is now independent of the total number of firms which allows our notion of equilibrium to remain practical for very large numbers of firms. In fact, we will show in Section 3.3 that as the number of firms grows, the ‘quality’ of the equilibrium increases (in a sense that is formally defined in Theorem 3.3.1).

To simplify notation, we will from now on omit the superscript s from the observations vector σ^s and assumed demand η^s . Unless otherwise stated, any mention of σ and η will refer to their corresponding stationary expressions defined in equations (3.1) and (3.2) respectively.

3.2 Learnability

Now that we have posited a notion of equilibrium, one of the first steps is to check under which conditions it exists and more importantly, whether firms’ pricing strategies will naturally converge to it. This is the focus of our first theorem:

Theorem 3.2.1. *Under Assumptions 2.2.1, 2.2.2 and 2.2.3, there exists at least one stationary equilibrium. Moreover, iterated stationary best response starting from an initial low-pricing strategy converges to such an equilibrium.*

The key result in the proof of this theorem is the following: each firm’s best response strategy increases with its competition factor. Intuitively, when firms are pricing low, the competition factor is also low and firms do not benefit much from competition in their sales probability. Therefore they cannot afford to price very high either when computing their best response. As firms increase their prices, so does their demand, which in turn leads firms to further increase their prices. However, as prices are bounded (by p^∞), we know that firms’ strategies converge. From there it is straightforward to show that this limit is in fact a stationary equilibrium. The step by step proof of this result can be found in Chapter 6.

An interesting fact is that this is no longer the case when if we pick a time-dependent assumed demand such as the expected competition factor at each time-step. Indeed, under some parameter choices, we observe a cyclic behavior in best response strategies. By fixing

the competition factor to be equal to its initial value, we avoid this issue and are able to guarantee monotonicity of best response strategies.

3.3 Asymptotic Optimality

In the previous section, we have shown that when firms myopically optimize their profit, i.e. ‘best respond’ to their competitors, they will eventually reach a stable pricing strategy. We are now interested in estimating how well these policies perform.

In particular, we would like to bound the profit gains a firm can make by unilaterally deviating from a stationary equilibrium. This is a non-obvious question as the demand against which firms optimize their profit (the assumed demand) is not equal to the true demand which governs when sales are realized. If there is a lot to gain for firms by deviating from a stationary equilibrium, the equilibrium becomes artificial: it is a compelling concept only if profits under the assumed demand and the true demand match closely.

Our second theorem demonstrates that as the number of firm grows, under some technical conditions, we can bound unilateral profit gains at a stationary equilibrium, thereby justify our notion of stationary equilibrium as an approximate-equilibrium:

Theorem 3.3.1. *Let $\{\Gamma^m\}_{\{m \geq 1\}}$ be a sequence of games such that:*

- *There are N types of firms. We denote by u_i the type of firm i , and by k_u^m the total number of firms of type u in game m . Firms of a similar type share a common demand function and initial capacity: $\forall u, u'$, there exist $\alpha_u, \beta_u, \gamma_{uu'}$, and c_u such that $\forall m, \forall i, \alpha_i^m = \alpha_{u_i}, \beta_i^m = \beta_{u_i}, c_i = c_{u_i}$ and*

$$\forall j, \gamma_{ij}^m = \begin{cases} \frac{\gamma_{u_i u_j}}{k_{u_j}^m}, & \text{if } u_i = u_j \\ \frac{\gamma_{u_i u_j}}{k_{u_i}^m - 1}, & \text{otherwise} \end{cases}$$

- $\forall m, \forall u, k_u^m \geq 2$ and $\min_u k_u^m \xrightarrow{m \rightarrow \infty} \infty$
- *Assumptions 2.2.1, 2.2.2 and 2.2.3 are satisfied*

$\forall m$, let μ^m be a stationary equilibrium policy of the game, σ^m the corresponding observation vector, and define $\delta_i^m = \max_{t \in \mathcal{T}} \mathbb{E} [\mu_i^m(\sigma_i^m(\mathbf{C}_t), t) - \mu_i^m(\sigma_i^m(\mathbf{C}_T), T)]$ the maximum expected difference between a firm's current price and its initial price. We can then bound the optimal profit gain a firm can make by unilaterally deviating from a stationary equilibrium as the number of firm grows large:

$$\lim_{m \rightarrow \infty} \frac{V_{(\mu_i^m, \mu_{-i}^m)}(\sigma_i^m(\mathbf{c}_T), T)}{\max_{\mu'} V_{(\mu', \mu_{-i}^m)}(\mathbf{c}_T, T)} \geq 1 - \frac{16 \sum_{j \neq i} \gamma_{ij} \delta_j^\mu}{(\alpha_i + \beta_i p_i^\infty)^2} \quad (3.3)$$

$$\geq 1 - \frac{8(\beta_i p_i^\infty - \alpha_i)}{(\alpha_i + \beta_i p_i^\infty)^2} \quad (3.4)$$

where the right hand terms are taken for $m = 1$.

The proof of this result is two-fold: we first introduce a new assumed demand which incorporates a time-dependent competition factor. In fact, it is self-concordant (the competition factor it assumes is the one that it generates). We then bound how much profit a firm can gain by unilaterally deviating from a stationary equilibrium under this new assumed demand. This is done by comparing the competition factors under both assumed demands and showing that they do not differ too much. Secondly we show that as the number of firms grows, the true demand function approaches the new time-dependent assumed demand. This is intuitively the case because demand shocks average out across firms. Tying both these statements together yields our result. The proof of which can once again be found in its entirety in Chapter 6.

The second bound (equation (3.4)) is derived from the first (equation (3.3)) using a relatively loose bound on the δ_i (namely that $\forall i, \delta_i < p_i^\infty/2$). In practice, from having run extensive simulations for a wide variety of parameters, we actually observe that $\delta_i < 0.2p_i^\infty$. This once again indicates that stationary equilibrium policies perform significantly better than the second bound would suggest.

Note that both these upper bounds are independent of initial capacities and of the time-horizon. Moreover, they hold for any strategy. In particular, they include strategies which are not σ -observable, i.e. which can not readily be applied (e.g. because they use information about competitors' current inventory levels and pricing strategies). If we were to restrict ourselves to strategies with requirements similar to our equilibrium strategies, we

would likely also have much tighter bounds.

Finally, we consider three limit cases where the bound yields intuitive results:

- As $\forall i, \delta_i \rightarrow 0$, this implies that price shocks become very small: the competition factor becomes essentially constant, i.e. equal to its initial value. This implies that the assumed demand η becomes exact. It therefore makes sense that in this case a stationary equilibrium becomes an exact equilibrium.
- As $\forall i, j \gamma_{ij} \rightarrow 0$, the competition factor vanishes: each firm's demand is affected by its competitors' prices very little. The game degenerates to a collection of monopolistic and separate dynamic pricing problems. In this case also the assumed demand and the stationary equilibrium becomes exact.
- When there is only one type of firm, Theorem 3.3.1's statement simplifies to the following:

Corollary 3.3.2. *Let $\{\Gamma^m\}_{\{m \geq 2\}}$ be a sequence of games such that each game has m firms and there exist $\alpha_0, \beta_0, \gamma_0$, and c_0 such that $\forall m, \forall i, \alpha_i^m = \alpha_0, \beta_i^m = \beta_0, c_i^m = c_0$ and $\forall j, \gamma_{ij}^m = \gamma_0/(m-1)$. Assume also that assumptions 2.2.1, 2.2.2 and 2.2.3 are satisfied.*

$\forall m$, let μ^m be a stationary equilibrium policy of the game, σ^m the corresponding observation vector. We can then bound the optimal profit gain a firm can make by unilaterally deviating from a stationary equilibrium as the number of firm grows large:

$$\lim_{m \rightarrow \infty} \frac{V_{(\mu_i^m, \mu_{-i}^m)}(\sigma_i^m(c_T), T)}{\max_{\mu'} V_{(\mu', \mu_{-i}^m)}(c_T, T)} \geq 1 - 8 \frac{\gamma_0(\beta_0 - \gamma_0)}{\alpha_0(2\beta_0 - \gamma_0)^2} \quad (3.5)$$

$$\geq 1 - 4 \frac{\gamma_0}{\alpha_0 \beta_0} \quad (3.6)$$

3.4 Empirical Support

So far we have shown that our notion of stationary equilibrium performs well when the competition factor each firm is faced with remains relatively constant. In this section we analyze a sample of airline pricing data in order to try and find empirical evidence that this is indeed the case.

Our data set consists of airline ticketing data for different flights from New York to San Francisco operated by seven distinct airlines over the course of a year. In particular we focus on the posted ticket price as a function of the number of days until departure. The flights were grouped into separate categories by departure time and day of the week so that inside each category each flight would be substitutable for another (e.g. all weekday morning flights).

A plot of the prices as a function of the number of days until departure for one such category (containing 4 flights) can be found in Figure 3.1. As we can see, individual flight prices are very volatile (with variations over \$100 from one day to the next). This is in line with what customers typically experience when purchasing tickets.

Figure 3.2 shows the evolution of the average ticket price across these four flights. Even though the sample size is very small (we are only considering four flights), the curve is noticeable smoother and flatter than individual price trajectories. One could imagine that, as we increase the number of flights further, the curve will continue to appear smoother and flatter.

This particular example therefore validates our assumption that demand shocks are independent across different firms which is precisely the condition for stationary strategies to perform well in practice.

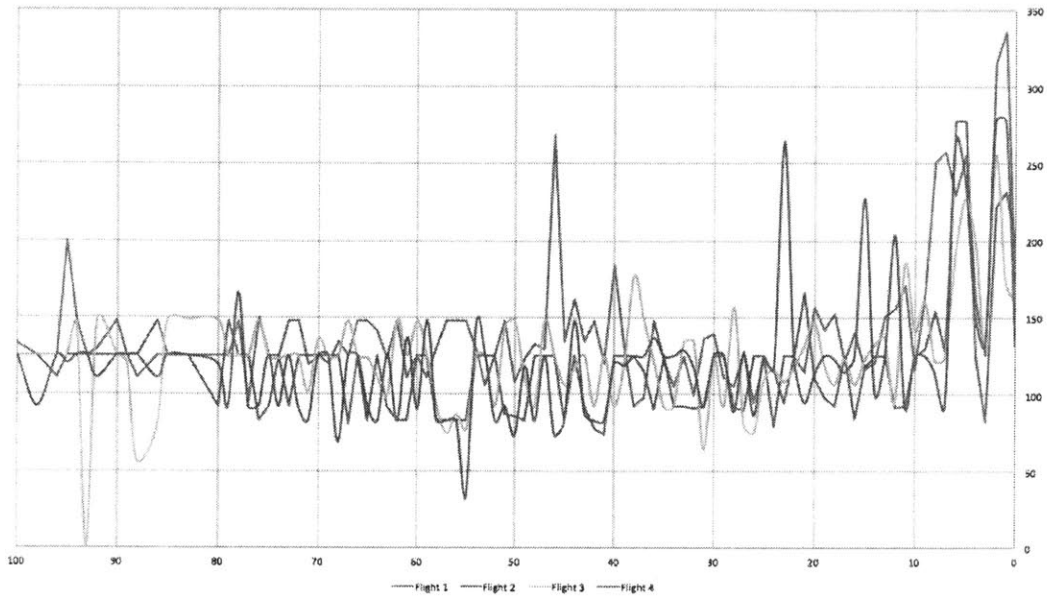


Figure 3.1: Price (in dollars) for four different substitutable flights from New York to San Francisco as a function of the number of days until departure

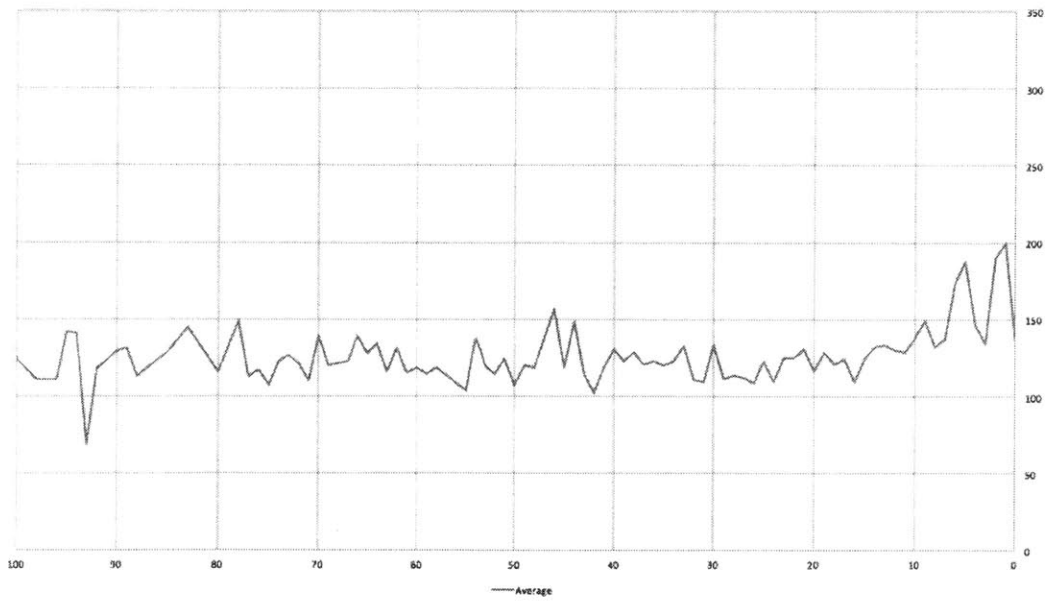


Figure 3.2: Average price (in dollars) over the same four flights from New York to San Francisco as a function of the number of days until departure

4 Simulations

Theorem 3.3.1 applies in the limit of the number of firms going to infinity. We are however also interested in evaluating how our bounds perform in the practical case of a game with finitely many firms. In order to do this we have simulated how different equilibrium concepts compare over a similar game. Note that since this requires computing strategies contingent on the entire state space of the game (the dimensionality of which is exponential in the number of firms), we can only do this for very small instances of our game. The setting we chose is the following: 4 firms with symmetrical demand with parameters $\alpha = 0.2, \beta = 0.2, \gamma = 0.01$ (expectations are approximated by averaging over 200 runs).

Table 4.1 shows the optimality gap (i.e. the maximum profit a firm can make by unilaterally deviating from equilibrium) for a stationary equilibrium Δ_{SE} and for a Nash equilibrium Δ_{Nash} . Under these conditions equation (3.4) predicts a gap of 60% which is more conservative than the numbers found below. This further validates the concept our stationary equilibrium concept while motivating future work to tighten our current bounds.

C	T	Δ_{Nash}	Δ_{SE}
10	50	20%	5.5%
10	100	16%	10%
10	150	13%	10%
10	200	13%	8.1%
15	50	18%	4.2%
15	100	15%	9.1%
15	150	12%	8.5%
15	200	12%	9.3%

Table 4.1: Optimality gap of Nash equilibria and stationary equilibria for different capacity and time-horizon values

In this table we have also included the largest gains from unilateral deviations under a

Nash equilibrium, Δ_{Nash} . These gains seems to always be larger than the corresponding stationary equilibrium setting which indicates that firms would be less likely to deviate from a stationary equilibrium than a Nash equilibrium as they have less to gain from doing so.

5 Conclusions and Future Work

In this part we have posited a notion of equilibrium which is both practical and provably asymptotically optimal, that of *stationary equilibrium*. In particular, we note that it arises from a behavior that is naturally exhibited by rational profit maximizing firms, best response, and can thus be learned efficiently. Moreover, we show that its assumptions seems to be verified in practice through the analysis of typical airline pricing data. Finally we favorably compared numerically the performance of the stationary equilibrium strategies to other commonly used notions of equilibria such as Nash equilibrium and Markov perfect equilibrium.

There are several directions for future work related to the results in this chapter. Proving a tighter bound on the δ_i (as defined in Section 3.3) would better the performance guarantee of stationary equilibriums strategies. A result on the difference in profit for different equilibrium concepts (such as was done numerically in Chapter 4) would also be very valuable. Finally, we believe it would be worth investigating whether similar results hold under different demand models (e.g. exponential or multinomial logit demand).

6 Proofs

For convenience in this chapter, we will rewrite the assumed demand $[\eta(\mu)]_{i,i}(c_i, t)$ as $\alpha_{i,t}^\mu - \beta_i \mu_i(c_i, t)$ where $\forall t, \alpha_{i,t}^\mu = \alpha_i^\mu = \alpha_i + \sum_{j \neq i} \gamma_{ij} \mu_j^\sigma(c_1, 1)$. From this expression it is easy to see that each firm's strategy is a monopolistic strategy: it doesn't depend on the other firms' capacities. From now on we will therefore write $V_{(\mu_i, \mu_{-i})}$, μ_i , and $[\eta^\mu]_{i,j}$ as functions of a firm's own capacity c_i and t . Finally, we introduce the notion of *marginal value of capacity* $\Delta_c V$:

Definition 6.0.1. Given observations σ , assumption function η and strategies μ , we denote by $\Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}$ the marginal value of one unit of capacity for firm i :

$$\begin{aligned} \Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta} : \mathcal{C}_i^* \times \mathcal{T} &\rightarrow \mathbb{R} \\ (c_i, t) &\mapsto V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t) - V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i - 1, t) \end{aligned}$$

Using the marginal value of capacity, we can write the standard Bellman equations for our problem as follows:

Definition 6.0.2. Bellman equation:

$$\begin{aligned} \forall t \in \mathcal{T}, V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(0, t) &= 0 \\ \forall c_i \in \mathcal{C}_i, V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, 0) &= 0 \\ \forall (c_i, t) \in \mathcal{C}_i^* \times \mathcal{T}, V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t) &= V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t - 1) + \max_{p_i \in \mathcal{A}_i} (\alpha_{i,t}^\mu - \beta_i p_i)_+ (p_i - \Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t - 1)) \end{aligned}$$

The second boundary condition comes from the fact that there is no salvage value at the end of the horizon. In the next section we will show that we are able to find a recursive closed form solution to this Bellman equation.

6.1 Proof of Theorem 3.2.1

The proof of this theorem relies on four technical lemmas which we now prove.

Lemma 6.1.1. *Let $\phi_i = [\phi^{\sigma,\eta}(\boldsymbol{\mu})]_i$ be the best response by firm i to strategy vector $\boldsymbol{\mu}$, we are able to bound the marginal value of capacity of firm i using strategy ϕ_i . More precisely: $\forall (c_i, t) \in \mathcal{C}_i \times \mathcal{T}$,*

$$0 \leq \Delta_c V_{(\phi_i, \boldsymbol{\mu}_{-i})}^{\sigma,\eta}(c_i + 1, t) \leq \Delta_c V_{(\phi_i, \boldsymbol{\mu}_{-i})}^{\sigma,\eta}(c_i, t) \leq p_i^\infty$$

Proof. The proof is done by induction on t . To simplify notation we temporarily drop the subscripts and superscripts from $V_{(\phi_i, \boldsymbol{\mu}_{-i})}^{\sigma,\eta}$ as they won't change in this proof. We start by proving the lower bound and monotonicity:

- $t = 0$: $\forall c_i \in \mathcal{C}_i$,

$$\begin{aligned} V(c_i, 0) = 0 &\Rightarrow \forall c_i \in \mathcal{C}_i, \Delta_c V(c_i, 0) = 0 \\ &\Rightarrow \forall c_i \in \mathcal{C}_i, 0 \leq \Delta_c V(c_i + 1, 0) \leq \Delta_c V(c_i, 0) \end{aligned}$$

- $t - 1 \rightarrow t$: From the Bellman equation and by definition of ϕ_i : $\forall c_i \in \mathcal{C}_i$,

$$V(c_i, t) = V(c_i, t - 1) + (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i, t))(\phi_i(c_i, t) - \Delta_c V(c_i, t - 1)) \quad (6.1)$$

Taking differences twice on c , we get:

$$\begin{aligned}
\Delta_c V(c_i, t) - \Delta_c V(c_i - 1, t) &= \Delta_c V(c_i, t - 1) - \Delta_c V(c_i - 1, t - 1) \\
&\quad + (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i, t))(\phi_i(c_i, t) - \Delta_c V(c_i, t - 1)) \\
&\quad - 2(\alpha_{i,t}^\mu - \beta_i \phi_i(c_i - 1, t))(\phi_i(c_i - 1, t) - \Delta_c V(c_i - 1, t - 1)) \\
&\quad + (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i - 2, t))(\phi_i(c_i - 2, t) - \Delta_c V(c_i - 2, t - 1))
\end{aligned}$$

Moreover, by definition of $\phi_i(c_i - 1, t)$:

$$\phi_i(c_i - 1, t) \in \arg \max_p \left[(\alpha_{i,t}^\mu - \beta_i p)(p - \Delta_c V(c_i - 1, t - 1)) + V(c_i - 1, t - 1) \right]$$

Therefore, we have:

$$\begin{aligned}
&(\alpha_{i,t}^\mu - \beta_i \phi_i(c_i - 1, t))(\phi_i(c_i - 1, t) - \Delta_c V(c_i - 1, t - 1)) \\
&\quad \geq (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i, t))(\phi_i(c_i, t) - \Delta_c V(c_i - 1, t - 1)) \\
&\quad \geq (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i - 2, t))(\phi_i(c_i - 2, t) - \Delta_c V(c_i - 1, t - 1))
\end{aligned}$$

Plugging this into the previous equality:

$$\begin{aligned}
&\Delta_c V(c_i, t) - \Delta_c V(c_i - 1, t) \\
&\quad \leq \Delta_c V(c_i, t - 1) - \Delta_c V(c_i - 1, t - 1) \\
&\quad + (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i, t))(\Delta_c V(c_i - 1, t - 1) - \Delta_c V(c_i, t - 1)) \\
&\quad + (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i - 2, t))(\Delta_c V(c_i - 1, t - 1) - \Delta_c V(c_i - 2, t - 1))
\end{aligned}$$

Which in turn yields:

$$\begin{aligned}
&\Delta_c V(c_i, t) - \Delta_c V(c_i - 1, t) \\
&\quad \leq \overbrace{(1 - \alpha_{i,t}^\mu + \beta_i \phi_i(c_i, t))}^{\geq 0} \overbrace{(\Delta_c V(c_i, t - 1) - \Delta_c V(c_i - 1, t - 1))}^{\leq 0 \text{ from the induction hypothesis}} \\
&\quad + \overbrace{(\alpha_{i,t}^\mu - \beta_i \phi_i(c_i - 2, t))}^{\geq 0} \overbrace{(\Delta_c V(c_i - 1, t - 1) - \Delta_c V(c_i - 2, t - 1))}^{\leq 0 \text{ from the induction hypothesis}}
\end{aligned}$$

From where we get the monotonicity result:

$$\Delta_c V(c_i, t) - \Delta_c V(c_i - 1, t) \leq 0 \quad (6.2)$$

Note now that by taking differences of equation (6.1) once on c , we get:

$$\begin{aligned} \Delta_c V(c_i, t) &= \Delta_c V(c_i, t - 1) \\ &+ (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i, t))(\phi_i(c_i, t) - \Delta_c V(c_i, t - 1)) \\ &- (\alpha_{i,t}^\mu - \beta_i \phi_i(c_i - 1, t))(\phi_i(c_i - 1, t) - \Delta_c V(c_i - 1, t - 1)) \\ &\geq \Delta_c V(c_i, t - 1) + \overbrace{(\alpha_{i,t}^\mu - \beta_i \phi_i(c_i, t))}^{\geq 0} \overbrace{(\Delta_c V(c_i - 1, t - 1) - \Delta_c V(c_i, t - 1))}^{\geq 0 \text{ from equation (6.2)}} \\ &\geq \Delta_c V(c_i, t - 1) \\ &\geq 0 \text{ from the induction hypothesis} \end{aligned}$$

Finally, remains to prove that $\forall(c_i, t), \Delta_c V(c_i, t) \leq p_i^\infty$. Notice that by definition of ϕ_i , we have that $\forall(c_i, t), \phi_i(c_i, t) \geq \Delta_c V(c_i, t - 1)$, otherwise the achieved profit at that stage would be lower than that of setting the price to p_i^∞ . This directly implies our desired result: $\forall(c_i, t), \Delta_c V(c_i, t) \leq p_i^\infty$ (we can artificially increase T to cover the case where $t = T$). \square

Lemma 6.1.2. *Let $\phi_i = [\phi^{\sigma, \eta}(\mu)]_i$ be the best response by firm i to strategy vector μ , we have the following equations:*

$$\phi_i(c_i, t) = \frac{1}{2} \left(\frac{\alpha_{i,t}^\mu}{\beta_i} + \Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t + 1) \right) \quad (6.3)$$

$$V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i, t) = V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i, t + 1) + \frac{\beta}{4} \left(\frac{\alpha_{i,t}^\mu}{\beta_i} + \Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t + 1) \right)^2 \quad (6.4)$$

$$\begin{aligned} \Delta_c V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i, t) &= \Delta_c V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i, t + 1) \\ &+ \beta \left(\frac{\alpha_{i,t}^\mu}{\beta_i} - \frac{\Delta_c V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i - 1, t + 1) + \Delta_c V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i, t + 1)}{2} \right) \\ &\times \left(\frac{\Delta_c V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i - 1, t + 1) - \Delta_c V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i, t + 1)}{2} \right) \end{aligned} \quad (6.5)$$

Proof. Solving the Bellman equation maximization problem using first order conditions, we

get that:

$$\phi_i(c_i, t) = \min \left\{ p_i^\infty, \frac{1}{2} \left(\frac{\alpha_{i,t}^\mu}{\beta_i} + \Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t+1) \right) \right\}$$

It is easy to see that $\frac{\alpha_{i,t}^\mu}{\beta_i} \leq p_i^\infty$ and from Lemma 6.1.2, we know that $\Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t+1) \leq p_i^\infty$, therefore we can simply write $\phi_i(c_i, t)$ as in equation (6.3). Replacing this value in the Bellman equation, we directly get equation (6.4). Finally, taking differences and simplifying the rightmost terms we get equation (6.5). \square

Lemma 6.1.3. *Let $\phi_i = [\phi^{\sigma, \eta}(\boldsymbol{\mu})]_i$ be the best response by firm i to strategy vector $\boldsymbol{\mu}$, the marginal value of capacity for firm i using strategy ϕ_i increases with $\boldsymbol{\mu}$. More formally, $\forall (c_i, t, t') \in \mathcal{C} \times \mathcal{T}^2, \Delta_c V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i, t)$ is continuously differentiable w.r.t. $\alpha_{i,t'}^\mu$ and:*

$$\frac{\partial \Delta_c V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}(c_i, t)}{\partial \alpha_{i,t'}^\mu} \geq 0$$

Proof. Let us prove this result by backwards induction on t . Once again, we temporarily drop the subscripts and superscripts from $V_{(\phi_i, \mu_{-i})}^{\sigma, \eta}$ as they won't change in this proof.

- $t = 0$: We have $\Delta_c V(c_i, t) = 0$, therefore its derivative satisfies the induction hypothesis.
- $t + 1 \rightarrow t$: Differentiating equation (6.5):

$$\begin{aligned} \frac{\partial \Delta_c V(c_i, t)}{\partial \alpha_{i,t'}^\mu} &= \frac{\partial \Delta_c V(c_i, t-1)}{\partial \alpha_{i,t'}^\mu} \\ &+ \beta_i \left(\frac{\mathbb{1}_{\{t'=t\}}}{\beta_i} - \frac{\partial}{\partial \alpha_{i,t'}^\mu} \left(\frac{\Delta_c V(c_i-1, t+1) + \Delta_c V(c_i, t-1)}{2} \right) \right) \\ &\times \left(\frac{\Delta_c V(c_i-1, t-1) - \Delta_c V(c_i, t-1)}{2} \right) \\ &+ \beta_i \left(\frac{\alpha_{i,t'}^\mu}{\beta_i} - \frac{\Delta_c V(c_i-1, t-1) + \Delta_c V(c_i, t-1)}{2} \right) \\ &\times \frac{\partial}{\partial \alpha_{i,t'}^\mu} \left(\frac{\Delta_c V(c_i-1, t-1) - \Delta_c V(c_i, t-1)}{2} \right) \end{aligned}$$

Aggregating terms together:

$$\begin{aligned} \frac{\partial \Delta_c V(c_i, t)}{\partial \alpha_{i,t'}^\mu} &= \overbrace{\frac{\partial \Delta_c V(c_i, t-1)}{\partial \alpha_{i,t'}^\mu}}^{\geq 0} \left(1 - \frac{\alpha_{i,t'}^\mu}{2} + \overbrace{\frac{\beta_i}{2} \Delta_c V(c_i, t-1)}^{\geq 0} \right) \\ &+ \overbrace{\frac{\partial \Delta_c V(c_i-1, t-1)}{\partial \alpha_{i,t'}^\mu}}^{\geq 0} \frac{\beta_i}{2} \overbrace{\left(\frac{\alpha_{i,t'}^\mu}{\beta_i} - \Delta_c V(c_i-1, t-1) \right)}^{\geq 0} \\ &+ \mathbb{1}_{\{t'=t\}} \overbrace{\frac{\Delta_c V_f^{\pi f}(c-1, t+1) - \Delta_c V_f^{\pi f}(c, t+1)}{2}}^{\geq 0} \end{aligned}$$

To get the desired result, suffices therefore that: $1 - \frac{\alpha_{i,t'}^\mu}{2} \geq 0$ i.e. $\alpha_i + \sum_{j \neq i} \gamma_{ij} \mu_j(c_T, T) \leq 2$.

2. This is true from Assumption 2.2.3.

Therefore the induction hypothesis holds for all t , and in particular $\Delta_c V(c_i, t)$ is also continuously differentiable w.r.t. $\alpha_{i,t'}^\mu, \forall t'$. \square

Lemma 6.1.4. *Let $\phi_i = [\phi^{\sigma, \eta}(\mu)]_i$ be the best response by firm i to strategy vector μ , then ϕ_i is increasing in μ . More precisely, $\forall (c_i, t, t') \in C_i \times \mathcal{T}^2$, $\phi_i(c_i, t)$ is continuously differentiable w.r.t. $\alpha_{i,t'}^\mu$ and:*

$$\frac{\partial \phi_i(c_i, t)}{\partial \alpha_{i,t'}^\mu} \geq 0$$

Proof. Recall equation (6.3):

$$\phi_i(c_i, t) = \frac{1}{2} \left(\frac{\alpha_{i,t}^\mu}{\beta_i} + \Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t+1) \right)$$

From Lemma 6.1.3, ϕ_i is differentiable w.r.t. $\alpha_{i,t'}^\mu$ and:

$$\frac{\partial \phi_i(c_i, t)}{\partial \alpha_{i,t'}^\mu} = \frac{1}{2} \left(\frac{\mathbb{1}_{\{t'=t\}}}{\beta_i} + \frac{\partial \Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, t+1)}{\partial \alpha_{i,t'}^\mu} \right) \geq 0$$

\square

From this last lemma we can prove Theorem 3.2.1 as follows:

Proof. Once again let $\phi_i = [\phi^{\sigma, \eta}(\boldsymbol{\mu})]_i$. Recall that $\forall t, \alpha_{i,t}^\mu = \alpha_i + \sum_{j \neq i} \gamma_{ij} \mu_j^\sigma(\mathbf{c}_1, 1)$. Combining this with Lemma 6.1.4, since $\forall i, j, \gamma_{ij} \geq 0$, we have: $\forall (c_i, t) \in \mathcal{C}_i \times \mathcal{T}$,

$$\forall j \neq i, \frac{\partial \phi_i(c_i, t)}{\partial \mu_j^\sigma(\mathbf{c}_1, 1)} \geq 0 \quad (6.6)$$

Now let $\boldsymbol{\mu}_0$ be the null strategy vector, i.e. $\forall (i, c, t) \in \mathcal{I} \times \mathcal{C}^* \times \mathcal{T}, \mu_i(c, t) = 0$. From equation (6.6) the sequence $\{(\phi^{\sigma, \eta})^n(\boldsymbol{\mu}_0)\}_{n \geq 0}$ (in the composition sense) is increasing. Since we also know that it is bounded from above (by $(p_i^\infty)_i$), it converges to a strategy $\boldsymbol{\mu}^*$. Finally from Lemma 6.1.4, we have that ϕ_i is continuously differentiable therefore $\boldsymbol{\mu}^*$ must satisfy $\phi(\boldsymbol{\mu}^*) = \boldsymbol{\mu}^*$ i.e. $\boldsymbol{\mu}^*$ is a stationary equilibrium. \square

6.2 Proof of Theorem 3.3.1

Theorem 3.3.1 bounds the profit sub-optimality gap from using a stationary equilibrium policy. The sub-optimality comes from the mismatch between assumed demand and actual demand. Indeed, recall that a stationary equilibrium policy considers the ‘competition term’ in the demand function to be constant over the entire game horizon, which in reality depends on the prices posted by other firms and therefore fluctuates over time.

In order to prove this result, we proceed in two steps. We first compare profits obtained under the stationary assumed demand with profits obtained under the following assumed demand:

Definition 6.2.1. We define iteratively the *expected assumed demand* $\tilde{\eta}$ as follows:

$$[\tilde{\eta}(\boldsymbol{\mu})]_{i,j}(c, t) = \begin{cases} \alpha_i - \beta_i \mu_i(c, t) + \mathbb{E} \left[\sum_{j \neq i} \gamma_{ij} \mu_j(\tilde{S}_{j,t}, t) \right], & \text{if } j = i \\ 0, & \text{otherwise} \end{cases}$$

where \tilde{S} is the associated capacity random variable.

This assumed demand is by definition monopolistic but not constant over time as opposed

to the stationary assumed demand η . It captures the interaction between firms' strategies and the demand function they face. The second step of the proof relies on showing that this expected assumed demand asymptotically approaches the true demand as the number of firms grows.

Once again, this proof is structured into four lemmas which we will first state and prove before demonstrating the proof of our main result.

Lemma 6.2.2. *Let μ be a stationary equilibrium. We first show that the expected price for each firm under assumed demand η increases as the game goes on. More formally, since time is indexed from T to 0: $\mathbb{E}[\mu_i(S_{i,t}, t)]$ is decreasing in t .*

Proof. In this proof, we will drop the subscript and superscript i from $S_{i,t}$ and the subscripts and superscripts from $\Delta_c V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}$ as they won't change. Let $t \in \mathcal{T}^*$:

$$\begin{aligned}
& \mathbb{E}[\mu_i(S_{t-1}, t-1)] - \mathbb{E}[\mu_i(S_t, t)] \geq 0 \\
\Leftrightarrow & \mathbb{E}[\mu_i(S_{t-1}, t-1) - \mu_i(S_t, t)] \geq 0 \\
\Leftrightarrow & \mathbb{E}[\mathbb{E}[\mu_i(S_{t-1}, t-1) - \mu_i(S_t, t) | S_t]] \geq 0 \\
\Leftrightarrow & \mathbb{E}[(1 - \alpha_i^\mu + \beta_i \mu_i(S_t, t))\mu_i(S_t, t-1) + (\alpha_i^\mu - \beta_i \mu_i(S_t, t))\mu_i(S_t - 1, t-1) - \mu_i(S_t, t)] \geq 0 \\
\Leftrightarrow & \mathbb{E}[\mu_i(S_t, t-1) - \mu_i(S_t, t) + (\alpha_i^\mu - \beta_i \mu_i(S_t, t))(\mu_i(S_t - 1, t-1) - \mu_i(S_t, t-1))] \geq 0
\end{aligned}$$

Substituting the value for μ_i from equation (6.3) into the previous equation:

$$\begin{aligned}
& \mathbb{E}[\mu_i(S_{t-1}, t-1)] - \mathbb{E}[\mu_i(S_t, t)] \geq 0 \\
& \Leftrightarrow \mathbb{E}[\Delta_c V(S_t, t-1) - \Delta_c V(S_t, t)] \\
& \quad + \mathbb{E}\left[(\alpha_i^\mu - \beta_i \Delta_c V(S_t, t)) \left(\frac{\Delta_c V(S_t-1, t-1) - \Delta_c V(S_t, t-1)}{2}\right)\right] \geq 0 \\
& \Leftrightarrow \mathbb{E}[\Delta_c V(S_t, t) - \Delta_c V(S_t, t-1)] \\
& \quad \leq \mathbb{E}\left[(\alpha_i^\mu - \beta_i \Delta_c V(S_t, t)) \left(\frac{\Delta_c V(S_t-1, t-1) - \Delta_c V(S_t, t-1)}{2}\right)\right] \\
& \Leftrightarrow \mathbb{E}\left[\left(\alpha_i^\mu - \beta_i \frac{\Delta_c V(S_t-1, t-1) + \Delta_c V(S_t, t-1)}{2}\right) \left(\frac{\Delta_c V(S_t-1, t-1) - \Delta_c V(S_t, t-1)}{2}\right)\right] \\
& \quad \leq \mathbb{E}\left[(\alpha_i^\mu - \beta_i \Delta_c V(S_t, t)) \left(\frac{\Delta_c V(S_t-1, t-1) - \Delta_c V(S_t, t-1)}{2}\right)\right]
\end{aligned}$$

Where the final equivalence comes from equation (6.5). Note also that from equation (6.4):

$$\begin{aligned}
\Delta_c V(c, t) - \Delta_c V(c, t-1) &= \frac{1}{4\beta_i} \left[(\alpha_i^\mu - \beta_i \Delta_c V(c, t-1))^2 - (\alpha_i^\mu - \beta_i \Delta_c V(c-1, t-1))^2 \right] \\
&= \overbrace{\alpha_i^\mu}^{\leq 1} \left[\frac{\Delta_c V(c-1, t-1) - \Delta_c V(c, t-1)}{2} \right] \\
& \quad + \frac{\beta_i}{4} \overbrace{\left[\Delta_c V(c, t-1)^2 - \Delta_c V(c-1, t-1)^2 \right]}^{\leq 0} \\
&\leq \frac{\Delta_c V(c-1, t-1) - \Delta_c V(c, t-1)}{2}
\end{aligned}$$

Finally:

$$\begin{aligned}
& \Delta_c V(S_t, t) - \Delta_c V(S_{t-1}, t-1) \leq \frac{\Delta_c V(S_t - 1, t-1) - \Delta_c V(S_t, t-1)}{2} \\
\Leftrightarrow & \Delta_c V(S_t, t) - \Delta_c V(S_t, t-1) \leq \Delta_c V(S_t - 1, t-1) - \Delta_c V(S_t, t) \\
\Leftrightarrow & \Delta_c V(S_t, t) \leq \frac{\Delta_c V(S_t - 1, t-1) + \Delta_c V(S_t, t-1)}{2} \\
\Leftrightarrow & \alpha_i^\mu - \beta_i \frac{\Delta_c V(S_t - 1, t-1) + \Delta_c V(S_t, t-1)}{2} \leq \alpha_i^\mu - \beta_i \Delta_c V(S_t, t) \\
\Leftrightarrow & \left(\alpha_i^\mu - \beta_i \frac{\Delta_c V(S_t - 1, t-1) + \Delta_c V(S_t, t-1)}{2} \right) \overbrace{\left(\frac{\Delta_c V(S_t - 1, t-1) - \Delta_c V(S_t, t-1)}{2} \right)}{\geq 0 \text{ from lemma 4.6}} \\
& \leq (\alpha_i^\mu - \beta_i \Delta_c V(S_t, t)) \overbrace{\left(\frac{\Delta_c V(S_t - 1, t-1) - \Delta_c V(S_t, t-1)}{2} \right)}{\geq 0 \text{ from lemma 4.6}} \\
\Rightarrow & \mathbb{E} \left[\left(\alpha_i^\mu - \beta_i \frac{\Delta_c V(S_t - 1, t-1) + \Delta_c V(S_t, t-1)}{2} \right) \left(\frac{\Delta_c V(S_t - 1, t-1) - \Delta_c V(S_t, t-1)}{2} \right) \right] \\
& \leq \mathbb{E} \left[(\alpha_i^\mu - \beta_i \Delta_c V(S_t, t)) \left(\frac{\Delta_c V(S_t - 1, t-1) - \Delta_c V(S_t, t-1)}{2} \right) \right] \\
\Leftrightarrow & \mathbb{E} [\mu_i(S_{t-1}, t-1)] - \mathbb{E} [\mu_i(S_t, t)] \geq 0
\end{aligned}$$

□

Lemma 6.2.3. *Let μ be a stationary equilibrium. We show that the expected price for each firm under assumed demand $\tilde{\eta}$ is always greater than its initial value:*

$$\mathbb{E} [\mu_i(\tilde{S}_{i,t}, t)] \geq \mu_i(c_i, T)$$

Proof. In order to prove this result, we prove by induction on $t \in \mathcal{T}$ the following:

$$\begin{cases} \mathbb{P}(\tilde{S}_{i,t} \leq c) \geq \mathbb{P}(S_{i,t} \leq c) \\ \mathbb{E}[\mu_i(\tilde{S}_{i,t}, t)] \geq \mathbb{E}[\mu_i(S_{i,t}, t)] \end{cases} \quad (6.7)$$

- $t = T$: Equation (6.7) is true since $\forall i \in \mathcal{I}, \tilde{S}_{i,T} = S_{i,T} = c_i$.

- $t \rightarrow t - 1$: Notice that by combining equation (6.7) with Lemma 6.2.2 we get that $\mathbb{E} \left[\mu_i(\tilde{S}_{i,t}, t) \right] \geq \mathbb{E} [\mu_i(S_{i,T}, T)] = \mu_i(c_i, T)$. This implies in particular that $\forall i \in \mathcal{I}, \forall c \leq c_i, \tilde{\eta}_i(c, t) \geq \eta_i(c, t)$. Therefore:

$$\begin{aligned}
\mathbb{P} \left(\tilde{S}_{i,t-1} \leq c \right) &= \mathbb{P} \left(\tilde{S}_{i,t} \leq c \right) + \tilde{\eta}_i(c+1, t) \mathbb{P} \left(\tilde{S}_{i,t} = c+1 \right) \\
&= \mathbb{P} \left(S_{i,t} \leq c \right) + \underbrace{\mathbb{P} \left(\tilde{S}_{i,t} \leq c \right) - \mathbb{P} \left(S_{i,t} \leq c \right)}_{\geq 0 \text{ from equation (6.7)}} + \underbrace{\tilde{\eta}_i(c+1, t)}_{\leq 1} \mathbb{P} \left(\tilde{S}_{i,t} = c+1 \right) \\
&\geq \mathbb{P} \left(S_{i,t} \leq c \right) + \underbrace{\tilde{\eta}_i(c+1, t)}_{\geq \eta_i(c+1, t) \text{ from above}} \left(\underbrace{\mathbb{P} \left(\tilde{S}_{i,t} \leq c+1 \right)}_{\geq \mathbb{P} \left(S_{i,t} \leq c+1 \right) \text{ from equation (6.7)}} - \mathbb{P} \left(S_{i,t} \leq c \right) \right) \\
&\geq \mathbb{P} \left(S_{i,t} \leq c \right) + \eta_i(c+1, t) \left(\mathbb{P} \left(S_{i,t} \leq c+1 \right) - \mathbb{P} \left(S_{i,t} \leq c \right) \right) \\
&= \mathbb{P} \left(S_{i,t} \leq c \right) + \eta_i(c+1, t) \mathbb{P} \left(S_{i,t} = c+1 \right) \\
&= \mathbb{P} \left(S_{i,t-1} \leq c \right)
\end{aligned}$$

Remains to prove the second part of the induction hypothesis:

$$\begin{aligned}
\mathbb{E} \left[\mu_i(\tilde{S}_{i,t-1}, t-1) \right] &= \sum_{c=0}^{c_i} \mathbb{P} \left(\tilde{S}_{i,t-1} = c \right) \mu_i(c, t-1) \\
&= \mathbb{P} \left(\tilde{S}_{i,t-1} \leq 0 \right) \mu_i(0, t-1) \\
&\quad + \sum_{c=1}^{c_i} \left(\mathbb{P} \left(\tilde{S}_{i,t-1} \leq c \right) - \mathbb{P} \left(\tilde{S}_{i,t-1} \leq c-1 \right) \right) \mu_i(c, t-1) \\
&= \mathbb{P} \left(\tilde{S}_{i,t-1} \leq 0 \right) \mu_i(0, t-1) + \mu_i(c_i, t-1) \\
&\quad + \sum_{c=1}^{c_i-1} \mathbb{P} \left(\tilde{S}_{i,t-1} \leq c \right) \underbrace{\left(\mu_i(c, t-1) - \mu_i(c+1, t-1) \right)}_{\geq 0 \text{ from Lemma 6.1.1}} \\
&\geq \mathbb{P} \left(S_{i,t-1} \leq 0 \right) \mu_i(0, t-1) + \mu_i(c_i, t-1) \\
&\quad + \sum_{c=1}^{c_i-1} \mathbb{P} \left(S_{i,t-1} \leq c \right) \left(\mu_i(c, t-1) - \mu_i(c+1, t-1) \right) \\
&= \mathbb{E} \left[\mu_i(S_{i,t-1}, t-1) \right]
\end{aligned}$$

In particular we therefore have: $\forall t \in \mathcal{T}$,

$$\begin{aligned} \mathbb{E} \left[\mu_i(\tilde{S}_{i,t}, t) \right] &\geq \mathbb{E} [\mu_i(S_{i,t}, t)] \\ &\geq \mathbb{E} [\mu_i(S_{i,T}, T)] \text{ from Lemma 6.2.2} \\ &= \mu_i(c_i, T) \end{aligned}$$

□

Lemma 6.2.4. *Let μ be a stationary equilibrium. We can lower bound the optimal profit obtained by unilaterally deviating from this equilibrium under assumed demand $\tilde{\eta}$:*

$$\max_{\mu'_i} V_{(\mu'_i, \mu_{-i})}^{\sigma, \tilde{\eta}}(c_i, T) \geq \frac{c_i}{16\beta_i} (\alpha_i + \beta_i p_i^\infty)^2$$

Proof.

$$\begin{aligned} \max_{\mu'_i} V_{(\mu'_i, \mu_{-i})}^{\sigma, \tilde{\eta}}(c_i, T) &= \max_{\mu'_i} \mathbb{E} \left[\sum_{t=1}^T \mu'_i(\tilde{S}_{i,t}, t) \left(\alpha_i - \beta_i \mu'_i(\tilde{S}_{i,t}, t) + \mathbb{E} \left[\sum_{j \neq i} \gamma_{ij} \overbrace{\mu_j(\tilde{S}_{j,t}, t)}^{\geq p_j^\infty / 2} \right] \right) \right]_+ \\ &\geq \max_{\mu'_i} \mathbb{E} \left[\sum_{t=1}^T \mu'_i(\tilde{S}_{i,t}, t) \left(\alpha_i - \beta_i \mu'_i(\tilde{S}_{i,t}, t) + \frac{1}{2} \sum_{j \neq i} \gamma_{ij} \overbrace{p_j^\infty}^{= \beta_i p_i^\infty - \alpha_i} \right) \right]_+ \\ &= \max_{\mu'_i} \mathbb{E} \left[\sum_{t=1}^T \mu'_i(\tilde{S}_{i,t}, t) \left(\frac{1}{2} (\alpha_i + \beta_i p_i^\infty) - \beta_i \mu'_i(\tilde{S}_{i,t}, t) \right) \right]_+ \\ &\geq \mathbb{E} \left[\sum_{t=1}^T \hat{\mu}_i(\tilde{S}_{i,t}, t) \left(\frac{1}{2} (\alpha_i + \beta_i p_i^\infty) - \beta_i \hat{\mu}_i(\tilde{S}_{i,t}, t) \right) \right]_+ \end{aligned}$$

Where $\hat{\mu}_i$ is a fixed price policy pricing at $\frac{\rho_i}{\beta_i} = \frac{\alpha_i + \beta_i p_i^\infty}{4\beta_i}$ (except in the case where capacity is null, in which case price is set to p_i^∞ as usual). Under policy $\hat{\mu}_i$, at each time step the

probability of making a sale is equal simply to ρ_i . We can therefore write:

$$\begin{aligned}
\max_{\mu'_i} V_{(\mu'_i, \mu_{-i})}^{\sigma, \bar{\eta}}(c_i, T) &\geq \frac{\rho_i}{\beta_i} \sum_{t=0}^T \binom{T}{t} \rho_i^t (1 - \rho_i)^{T-t} \overbrace{\min(t, c_i)}^{\geq tc_i/T} \\
&\geq \frac{\rho_i}{\beta_i} \sum_{t=0}^T \binom{T}{t} \rho_i^t (1 - \rho_i)^{T-t} \frac{tc_i}{T} \\
&= \frac{\rho_i c_i}{\beta_i T} \overbrace{\sum_{t=0}^T \binom{T}{t} \rho_i^t (1 - \rho_i)^{T-t}}^{= \rho_i T} \\
&= \frac{\rho_i^2 c_i}{\beta_i} \\
&= \frac{c_i}{16\beta_i} (\alpha_i + \beta_i p_i^\infty)^2
\end{aligned}$$

□

Lemma 6.2.5. *Let μ be a stationary equilibrium. We demonstrate an upper bound on the maximum relative profit gains from unilaterally deviating from this equilibrium under assumed demand $\bar{\eta}$:*

$$\frac{V_{(\mu_i, \mu_{-i})}^{\sigma, \bar{\eta}}(c_i, T)}{\max_{\mu'_i} V_{(\mu'_i, \mu_{-i})}^{\sigma, \bar{\eta}}(c_i, T)} \geq 1 - \frac{16 \sum_{j \neq i} \gamma_{ij} \tilde{\delta}_j^\mu}{(\alpha_i + \beta_i p_i^\infty)^2} \quad (6.8)$$

$$\geq 1 - \frac{8(\beta_i p_i^\infty - \alpha_i)}{(\alpha_i + \beta_i p_i^\infty)^2} \quad (6.9)$$

where, similarly to δ_i^μ , $\tilde{\delta}_i^\mu = \max_{t \in \mathcal{T}} \left(\mathbb{E} \left[\mu_i(\bar{S}_{i,t}, t) \right] - \mu_i(c_i, t) \right)$.

Proof. Let $\tilde{\mu}_i = \arg \max_{\mu'_i} V_{(\mu'_i, \mu_{-i})}^{\sigma, \eta}(c_i, T)$, we then have:

$$\begin{aligned}
& V_{(\mu'_i, \mu_{-i})}^{\sigma, \eta}(c_i, T) \\
&= \mathbb{E} \left[\sum_{t=1}^T \tilde{\mu}_i(\tilde{S}_{i,t}, t) \left(\alpha_i - \beta_i \tilde{\mu}_i(\tilde{S}_{i,t}, t) + \mathbb{E} \left[\sum_{j \neq i} \gamma_{ij} \mu_j(\tilde{S}_{j,t}, t) \right] \right) \right] \\
&= \mathbb{E} \left[\sum_{t=1}^T \tilde{\mu}_i(\tilde{S}_{i,t}, t) \left(\alpha_i - \beta_i \left(\overbrace{\tilde{\mu}_i(\tilde{S}_{i,t}, t) - \frac{\mathbb{E} \left[\sum_{j \neq i} \gamma_{ij} (\mu_j(\tilde{S}_{j,t}, t) - \mu_j(c_j, t)) \right]}{\beta_i}}^{:= \hat{\mu}_i(\tilde{S}_{i,t}, t)}} \right) + \sum_{j \neq i} \gamma_{ij} \mu_j(c_j, t) \right) \right]
\end{aligned}$$

We can rewrite the previous expression in terms of $\hat{\mu}_i$:

$$\begin{aligned}
& V_{(\mu'_i, \mu_{-i})}^{\sigma, \eta}(c_i, T) \\
&= \mathbb{E} \left[\sum_{t=1}^T \hat{\mu}_i(\tilde{S}_{i,t}, t) \left(\alpha_i - \beta_i \hat{\mu}_i(\tilde{S}_{i,t}, t) + \sum_{j \neq i} \gamma_{ij} \mu_j(c_j, t) \right) \right] \\
&+ \frac{1}{\beta_i} \mathbb{E} \left[\sum_{t=1}^T \mathbb{E} \left[\sum_{j \neq i} \gamma_{ij} (\mu_j(\tilde{S}_{j,t}, t) - \mu_j(c_j, t)) \right] \left(\alpha_i - \beta_i \hat{\mu}_i(\tilde{S}_{i,t}, t) + \sum_{j \neq i} \gamma_{ij} \mu_j(c_j, t) \right) \right] \\
&\leq \mathbb{E} \left[\sum_{t=1}^T \mu_i(\tilde{S}_{i,t}, t) \left(\alpha_i - \beta_i \mu_i(\tilde{S}_{i,t}, t) + \sum_{j \neq i} \gamma_{ij} \overbrace{\mu_j(c_j, t)}^{\leq \mathbb{E}[\mu_j(\tilde{S}_{i,t}, t)]} \right) \right] \\
&+ \frac{\sum_{j \neq i} \gamma_{ij} \bar{\delta}_j^\mu}{\beta_i} \mathbb{E} \left[\sum_{t=1}^T \left(\alpha_i - \beta_i \hat{\mu}_i(\tilde{S}_{i,t}, t) + \sum_{j \neq i} \gamma_{ij} \mu_j(c_j, t) \right) \right] \\
&\leq \mathbb{E} \left[\sum_{t=1}^T \mu_i(\tilde{S}_{i,t}, t) \left(\alpha_i - \beta_i \mu_i(\tilde{S}_{i,t}, t) + \mathbb{E} \left[\sum_{j \neq i} \gamma_{ij} \mu_j(\tilde{S}_{j,t}, t) \right] \right) \right] + \frac{c_i \sum_{j \neq i} \gamma_{ij} \bar{\delta}_j^\mu}{\beta_i} \\
&= V_{(\mu_i, \mu_{-i})}^{\sigma, \eta}(c_i, T) + \frac{c_i \sum_{j \neq i} \gamma_{ij} \bar{\delta}_j^\mu}{\beta_i}
\end{aligned}$$

Reorganizing terms, we get:

$$\begin{aligned} \frac{V_{(\mu_i, \mu_{-i})}^{\sigma, \tilde{\eta}}(c_i, T)}{\max_{\mu'_i} V_{(\mu'_i, \mu_{-i})}^{\sigma, \tilde{\eta}}(c_i, T)} &\geq 1 - \frac{c_i \sum_{j \neq i} \gamma_{ij} \tilde{\delta}_j^\mu}{\beta_i \max_{\mu'_i} V_{(\mu'_i, \mu_{-i})}^{\sigma, \tilde{\eta}}(c_i, T)} \\ &\geq 1 - \frac{16 \sum_{j \neq i} \gamma_{ij} \tilde{\delta}_j^\mu}{(\alpha_i + \beta_i p_i^\infty)^2} \text{ from Lemma 6.2.4} \end{aligned}$$

This proves equation (6.8). In order to prove equation (6.9), first recall that $\forall c_i, \forall t, \mu_i(c_i, t) \in [p_i^\infty/2, p_i^\infty]$, therefore $\tilde{\delta}_i^\mu \leq p_i^\infty/2$ and:

$$\begin{aligned} \frac{V_{(\mu_i, \mu_{-i})}^{\sigma, \tilde{\eta}}(c_i, T)}{\max_{\mu'_i} V_{(\mu'_i, \mu_{-i})}^{\sigma, \tilde{\eta}}(c_i, T)} &\geq 1 - \frac{8c_i \sum_{j \neq i} \gamma_{ij} p_j^\infty}{(\alpha_i + \beta_i p_i^\infty)^2} \\ &= 1 - \frac{8c_i (\beta_i p_i^\infty - \alpha_i)}{(\alpha_i + \beta_i p_i^\infty)^2} \text{ from Assumption 2.2.2} \end{aligned}$$

□

From Lemma 6.2.5, we are able to prove Theorem 3.3.1 as follows:

Proof of Theorem 3.3.1. We first define the random variable H_u^m as follows:

$$H_u^m(c, t) = \frac{\{\text{number of firms of type } u \text{ with inventory level } c \text{ at time } t\}}{k_u^m} \quad (6.10)$$

In this proof as in the statement of this theorem, we will omit the superscript m when it is equal to 1. We start by showing by induction on t that when at most one firm per type deviates from the stationary equilibrium policy μ^m , then:

$$\forall i \in \mathcal{I}, \forall c \in \mathcal{C}_i, \forall t \in \mathcal{T}, H_{u_i}^m(c, t) \xrightarrow[m \rightarrow \infty]{a.s.} \mathbb{P}(\tilde{S}_{i,t} = c) \quad (6.11)$$

For $t = T$, the result is true since all firms start with a deterministic inventory level: $\forall i, \forall c, H_{u_i}^m(c, T) = \mathbb{1}_{\{c=c_i\}} = \mathbb{P}(\tilde{S}_{i,t} = c)$. Let us now assume that the result is true for t and show that equation (6.11) hold for $t - 1$. For each firm type u , we denote by i_u the index of the firm deviating from the stationary equilibrium policy μ . We denote by μ'_{i_u} the policy used by this firm. If no firm of type u deviates from μ , we pick an arbitrary firm of type u and set $\mu'_{i_u} = \mu_u$. We start by noticing that $H_{u_i}^m$ verifies the following equation:

$\forall c < c_i,$

$$k_u^m H_u^m(c, t-1) = \sum_{i=1}^{k_u^m H_u^m(c+1, t)-1} A_i + \sum_{i=1}^{k_u^m H_u^m(c, t)-1} B_i + A'_i + B'_i \quad (6.12)$$

Where the A_i and B_i are independent identically distributed Bernoulli random variables with respective rates $\lambda_i^m(c+1, t)$ and $\lambda_i^m(c, t)$. Intuitively, equation (6.12) expresses the fact that firms with inventory level c at time $t-1$ are either firms that had inventory level $c+1$ at time t and realized a sale at time t or firms that had inventory level c at time t and didn't make a sale at time t . The A'_i and B'_i random variables correspond to the firms who potentially deviate from the stationary equilibrium policy (and therefore have a different rate). We can rewrite equation (6.12) as follows:

$$\begin{aligned} H_u^m(c, t-1) &= \frac{k_u^m H_u^m(c+1, t) - 1}{k_u^m} \left(\frac{1}{k_u^m H_u^m(c+1, t) - 1} \sum_{i=1}^{k_u^m H_u^m(c+1, t)-1} A_i \right) \\ &+ \frac{k_u^m H_u^m(c, t) - 1}{k_u^m} \left(\frac{1}{k_u^m H_u^m(c, t) - 1} \sum_{i=1}^{k_u^m H_u^m(c, t)-1} B_i \right) \\ &+ \frac{A'_i + B'_i}{k_u^m} \end{aligned}$$

In the case where $c = c_i$, the equation is the same but without the A_j and A'_j . Let now i such that firm i is using the stationary policy μ , we then have that:

$$\begin{aligned} \lambda_i^m(c_t, t) &= \alpha_{u_i} - \beta_{u_i} \mu_{u_i}^m(c_i, t) \\ &+ \sum_{u' \neq u_i} \frac{\gamma_{u_i u'}}{k_{u_i}^m} \left(\sum_{c'=0}^{c_{u'}} k_{u'}^m H_{u'}^m(c', t) \mu_{u'}(c', t) + \mu'_{i_{u'}}(c_t, t) - \mu_{i_{u'}}(c_{i_{u'}, t}, t) \right) \\ &+ \frac{\gamma_{u_i u_i}}{k_{u_i}^m - 1} \left(\sum_{c'=0}^{c_{u_i}} k_{u_i}^m H_{u_i}^m(c', t) \mu_{u_i}(c', t) + \mu'_{i_{u_i}}(c_t, t) - \mu_{i_{u_i}}(c_{i_{u_i}, t}, t) \right) \\ &= \alpha_{u_i} - \beta_{u_i} \mu_{u_i}^m(c_i, t) \\ &+ \sum_{u' \neq u_i} \gamma_{u_i u'} \left(\sum_{c'=0}^{c_{u'}} H_{u'}^m(c', t) \mu_{u'}(c', t) + \frac{\mu'_{i_{u'}}(c_t, t) - \mu_{i_{u'}}(c_{i_{u'}, t}, t)}{k_{u'}^m} \right) \\ &+ \frac{k_{u_i}^m}{k_{u_i}^m - 1} \gamma_{u_i u_i} \left(\sum_{c'=0}^{c_{u_i}} H_{u_i}^m(c', t) \mu_{u_i}(c', t) + \frac{\mu'_{i_{u_i}}(c_t, t) - \mu_{i_{u_i}}(c_{i_{u_i}, t}, t)}{k_{u_i}^m} \right) \end{aligned}$$

From the induction hypothesis and since the rightmost terms are bounded, we get:

$$\begin{aligned}
\lambda_i^m(\mathbf{C}_t, t) &\xrightarrow[m \rightarrow \infty]{a.s.} \alpha_{u_i} - \beta_{u_i} \mu_{u_i}^m(\mathbf{C}_{i,t}, t) + \sum_{u' \neq u_i} \gamma_{u_i u'} \sum_{c'=0}^{c_{u'}} \mathbb{P}(\tilde{S}_{u',t} = c') \mu_{u'}(c', t) \\
&\quad + \gamma_{u_i u_i} \sum_{c'=0}^{c_{u_i}} \mathbb{P}(\tilde{S}_{u_i,t} = c') \mu_{u_i}(c', t) \\
&= \alpha_{u_i} - \beta_{u_i} \mu_{u_i}^m(\mathbf{C}_{i,t}, t) + \sum_{u' \neq u_i} \gamma_{u_i u'} \mathbb{E}[\mu_{u'}(\tilde{S}_{u',t}, t)] + \gamma_{u_i u_i} \mathbb{E}[\mu_{u_i}(\tilde{S}_{u_i,t}, t)] \\
&= [\tilde{\eta}(\mu)]_{ii}(\mathbf{C}_{i,t}, t)
\end{aligned}$$

Proceeding as previously, it is easy to show that the same result holds for firms not using the stationary policy μ . Therefore:

$$\forall i \in \mathcal{I}, \lambda_i^m(\mathbf{C}_t, t) \xrightarrow[m \rightarrow \infty]{a.s.} [\tilde{\eta}(\mu)]_{ii}(\mathbf{C}_{i,t}, t) \quad (6.13)$$

Combining this result with equation (6.12), using the law of large numbers we get that:

$$\begin{aligned}
H_{u_i}^m(c, t-1) &\xrightarrow[m \rightarrow \infty]{a.s.} \mathbb{P}(\tilde{S}_{u_i,t} = c+1) [\tilde{\eta}(\mu)]_{ii}(c+1, t) + \mathbb{P}(\tilde{S}_{u_i,t} = c) [\tilde{\eta}(\mu)]_{ii}(c, t) \\
&= \mathbb{P}(\tilde{S}_{u_i,t-1} = c)
\end{aligned}$$

This concludes our induction proof. We now show that this result translates to a result on value functions:

$$\begin{aligned}
\max_{\mu'} V_{(\mu', \mu_{-i}^m)}(\mathbf{c}_T, T) &= \max_{\mu'} \mathbb{E} \left[\sum_{t'=t}^T \mu'(\mathbf{C}_{t'}, t') \lambda_i^{\mu'}(\mathbf{C}_{t'}, t') \right] \\
&\xrightarrow[m \rightarrow \infty]{a.s.} \max_{\mu'} \mathbb{E} \left[\sum_{t'=t}^T \mu'(\mathbf{C}_{t'}, t') [\tilde{\eta}(\mu')]_{ii}(\mathbf{C}_{i,t'}, t') \right] \\
&= \max_{\mu'} V_{(\mu', \mu_{-i}^m)}^{\sigma_i^m, \tilde{\eta}^m}(\sigma_i^m(\mathbf{c}_T), T)
\end{aligned}$$

Finally, combining this last result with Lemma 6.2.5, we get:

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{V_{(\mu_i^m, \mu_{-i}^m)}^{\sigma^m, \eta^m}(\sigma_i^m(c_T), T)}{\max_{\mu'} V_{(\mu', \mu_{-i}^m)}(c_T, T)} &\geq 1 - \frac{16 \sum_{j \neq i} \gamma_{ij} \delta_j^\mu}{(\alpha_i + \beta_i p_i^\infty)^2} \\ &\geq 1 - \frac{8(\beta_i p_i^\infty - \alpha_i)}{(\alpha_i + \beta_i p_i^\infty)^2} \end{aligned}$$

□

Part II

Predicting Power Outages in Electricity Networks: A Framework to Reduce Restoration Time After Large Storms

1 Introduction

Severe weather incidents frequently cause large-scale power outages, creating significant problems in the electricity supply industry. In the past five years, almost one million customers have lost power in Massachusetts alone. Sometimes for up to eleven days. Such events cause massive amounts of damage: Hurricane Sandy alone cost over 50 billion dollars in repairs, a significant portion of which went into restoring the electrical network.

In order to address this growing concern, we use machine learning and robust optimization to build a two-part model that makes preventative emergency planning more efficient. Using physical properties of the electrical network and historical data, we construct a model that predicts outages based on the weather forecast. We then use this prediction to optimally allocate repair crews in advance of a storm in order to minimize network-wide restoration time. At the conclusion of this work we find that our model's solution lies within 5% of the optimal scheduling for crews, while accounting for 90% of the worst-case repair time scenarios. Our predictive model is based on machine learning and will continuously improve in granularity and accuracy through the incorporation of additional information. As a data-driven model it provides an invaluable tool for decision making before a storm, which is currently motivated primarily by intuition from industry experience.

The work presented in this part of the thesis was done in collaboration with Anna Papush (graduate student at the Operations Research Center at MIT) Sean Whipple (LGO student at MIT) and a power utility company in the North East of the United States.

1.1 Motivation

Due to the unpredictable nature of severe weather events (cf. for example [8]), emergency storm planning is a particularly challenging problem. The resulting damage to electrical networks may prompt over a week's worth of restoration work and even longer customer interruption times. In just the past two years New England was hit by both Hurricanes Sandy and Irene, coupled with ice storms and massive blizzards such as Nemo. The outages caused by these events were of such a large scale that in some regions of Massachusetts hundreds of thousands of customers were out of power, some for up to ten days. Though this is detrimental to both electricity providers and consumers, few analytical studies have been conducted with the intent of improving emergency planning. Much of the previous work has been done in either predicting damage or restoration times, but not both in conjunction.

During, and in the days following a major event such as a large storm or a hurricane, mobility is often greatly reduced. There might be dozens of inches of snow covering the road, or tree limbs blocking access, especially in remote and non-urban areas such as those covered by the power company. There is therefore also an important need to carefully plan crew repair routes, to minimize transportation delays and reduce power downtime.

1.2 Literature Review

Society's increasing dependence on technology, media and communication has created a growing reliance on the electrical supply industry. Stemming from this dependence, weather-based power outages have recently become a very significant concern for both distributors and consumers. In the past decade there has been a good deal of research done with respect to this particular problem, from multiple angles. Most of the literature related to this field can largely be categorized into three distinct lines of work.

The first of these is research related to climate change and weather-based forecasting. Climate variation over the past several decades has sparked a great deal of academic interest both in terms of data collection and modeling. Synoptic weather typing is

a scheme for the classification of weather conditions into distinct types. It has frequently come into use as a valuable tool for work in climate impact applications such as pollution, disaster planning, agriculture and human health. In [15] and [29], we see an automated and manual approach to this kind of weather analysis. The work in [15] predicts occurrences of freezing rain by using automated synoptic typing on differentiations in air mass. By studying hourly meteorological readings, they identify weather types correlated with freezing rain and apply stepwise logistic regression to predict its likelihood. Similarly, [29] also employs airborne particle concentrations and daily weather data to build a manual synoptic typing that categorizes storm types in advance. This branch of research also encompasses the effects of climate change on a socio-economic level. Through time series modeling, [22] predicts daily variability in ski resort attendance based on a combination of surrounding urban and mountain weather. In [38] and [6], they consider the potentially harmful impacts of climate variability on temperature-related mortality and air pollution-related health effects by analyzing correlations with weather parameters.

A second branch of the literature considers electrical system reliability with respect to weather and the environment. Foundational work in this direction is done by Billinton et al. in [8], [9] and [10]. These papers propose single and two-state weather models, then expand these to a three state weather model that captures normal, adverse and extreme weather circumstances. Through the resulting calculations of reliability indices, they demonstrate the need for weather to be considered in practical system assessments. Sensitivity studies show that disregarding weather effects produces overly optimistic system appraisals, and that inclement weather conditions must be divided into a minimum of two types. Prior to this work, multiple investigative studies considered specific types of weather events and their impacts on system reliability. These works, such as [28], [4], and [12], ultimately aim to improve reliability through system redesign. The work in [28] analyzes drought conditions and their resulting effects on tree faults; by using the Palmer Drought Index, they present the influence of drought on tree-caused power outages. The latter two works consider lightning storms and ice storms, respectively. By modeling system response and storm characteristics,[4] presents a Monte Carlo simulation that evaluates system reliability and helps identify weaker areas for system redesign. Using a similar approach, [12] models weather, vulnerability and restoration times in order to estimate

system component reliability during severe ice storms. More recently, there have been general weather reliability studies following [8]. In order to present a cost-benefit analysis for overhead-to-underground line conversions, the work in [45] estimates damage rates based on hurricane wind speeds and simulates the resulting restoration process. The paper by Caswell et al. [13] considers correlations between reliability indices and various weather parameters to account for system variability.

The third line of work entails the prediction of weather-related electrical power outages. Some of the earlier considerations of this problem are demonstrated in [36], [17] and [18]. The approach in [36] utilizes artificial neural networks (ANNs) in order to predict the number of power interruptions based on inputted weather parameters. This approach combines time series and regression to develop a learning algorithm. The follow-up works [17] and [18] consider the effects of normal daily weather conditions on distribution system interruptions; by using Poisson regression models they determine the significant weather parameters that contribute most to daily outages. Later work such as [44] and [47] show the incorporation of other statistical techniques. In [44], contingency probability estimators are computed through the use of maximum likelihood (ML), to predict a transmission failure rate, and multiple linear regression on transformed weather data. Using both a Poisson regression model and a Bayesian network model, [47] proposes a method for predicting the number of annual overhead distribution line failures caused by weather.

The seminal series of papers by Liu et al., including [30] and [31], address a statistical approach to predicting spatial distribution of power outages and restoration times resulting from hurricane and ice storm damage. They employ a generalized linear mixed regression model (GLMM), however instead of using quantitative characteristics of each storm, they created indicator variables that map each outage to its respective storm. Furthermore, their model predicts damage on an outage level, meaning that it indicates whether a given device will open. This lacks granularity in that an outage may be caused by 5 trees falling across the lines or only by one broken pole. Their spatial prediction is executed on a 3 km x 3 km grid cell in a given area serviced by a utility company. Building on this approach, [24] uses generalized linear models (GLM) as well as generalized additive models (GAM), in addition to measurable storm data that replaced indicator variables. In order to avoid

variable collinearity, the data was transformed using principal component analysis (PCA), which insures that the data is not correlated. This work also predicts on a grid level, now 3.66 km x 2.44 km, in order to estimate numbers of outages, customers without power and damaged poles and transformers. Although this approach increases the prediction granularity, it still makes several assumptions on conditions that cause outages, such as the wind speed necessary to down a pole or uproot a tree. In the more recent work by Hongfei et al. [25], a Bayesian hierarchical statistical modeling approach is used to predict the number of outages and capture uncertainty in the outage data. Although the prediction is not categorized by type of outage, the model also geographically displays the uncertainty of the damage forecasts.

Finally, optimizing operations in industry is not a novel concept and has been utilized across many fields and industries. Recently, a similar optimization model was produced with the purpose of reducing overtime of gas utility repair crews [5]. However to the best of our knowledge there has not been a utility firm that has combined optimization with this level of uncertainty to their repair crews (particularly in a storm response scenario).

1.3 Contributions and Outline

Our work addresses the aforementioned issues by introducing a model, and subsequently a tool, that aid the decision making process of an electricity distributor ahead of a storm. Our model's unique approach combines information about the weather and geographical data with properties of the electrical network to predict which parts of the network are most vulnerable to damage. This allows both for higher accuracy and higher granularity than currently existing models. The resulting tool stems from a two-part algorithm that first forecasts damage based on weather reports, then assigns crews across staging platforms to expedite the power restoration process.

The quality and quantity of the data available, along with the intrinsic complexity of weather to damage interactions make this problem particularly challenging. Our purely data-driven approach is well suited to continuously improve as more data gets available and provide an increasingly reliable counterpoint to the purely intuition-based decision making currently

utilized in practice. Our optimization model relies heavily on Bertsimas-Sim uncertainty sets to introduce robustness in the problem formulation. This is critical to getting reliable results as the inputs for our model are typically hard to estimate.

Furthermore, we provide a secondary framework to guide repairs during and right after a storm. By combining call information (e.g. the address and the time of loss of power of each customer) with the failure probabilities of the components of the electrical network (obtained from the first model), we are able to efficiently estimate the probability that a particular part of the network has been damaged.

We show that these improvements could lead to a significant reduction in power restoration delays, by both helping the utility company to better stage resources ahead of a storm and significantly cutting down crew transportation times.

2 Predicting Outages

Storm warnings are usually issued by various weather forecasting agencies several days to a few weeks before a storm hits. Depending on the expected severity of the storm, the power utility company can bring in extra repair crews from out of state to help recovery. Once these crews have arrived (which can take multiple days), they are assigned along local repair crews the day before the storm to areas serviced by the company where damage is predicted to be most significant. This step is crucial as mobility during and right after a storm is greatly reduced (trees blocking access ways, heavy snowfall or rain, etc.) and bad crew placement can lead to extensive repair delays.

Dispatching crews, especially from out of state, is expensive so it is in the company's best interest to adequately size the extent of the repairs needed. It has happened that the company has come under fire for spending too much money preparing for storms that eventually caused little to no damage.

In this chapter, we develop a data-driven model that addresses this problem. We combine physical properties of the power network (e.g. number of electrical poles, wire framing), landcover data (e.g. tree cover, soil composition), and weather data (e.g. wind speed, pressure) to assess the risk of failure in each serviced area. We implement an algorithm, train it on six important past storms and results show that our model is able to capture the relationship between weather and damage significantly better than results found in the literature (e.g. [23]).

2.1 Data

In order to construct the predictive damage model, we first built a database from several distinct sources. Since our goal is to forecast damage to the network based on the anticipated weather, we need to understand the network structure and its vulnerabilities during various types of storms. To accomplish this we collected data on the following three categories: physical network properties, historical weather logs and historical outages during different severe weather events.

2.1.1 Electrical Network Data

To predict outages across an electrical network, we first needed to acquire data describing its structure. By using data from an electrical distributor in New England, we were able to construct an interactive mapping of their network across the state of Massachusetts. This included information on 300 circuits and 60,000 devices, such as reclosers and fuses, which serve approximately 1.2M customers across the state.

We began this mapping by attaining a physical description of the 280,000 segments that make up the state-wide network. As shown in Figure 2.1, a segment is a grouping of consecutive poles and wires that all share the same physical properties. It may contain a device as described above, but does not necessarily need to. The given list of physical properties included 35 parameters such as: wire insulation, above or below ground wiring, pole age, upstream and downstream segment coordinates, framing and length. By utilizing all of this information we were able to create a connected graph of all the segments for each given circuit.

Within this network we now define a new term, known as an *asset*. An asset is a collection of segments originating at a given device and includes all of the segments directly downstream of it. In other words, should that device open due to a short circuit, all customers serviced by the network beyond this point would be out of power. As explained in the next subsection, this new definition is necessary to allow modeling between historical outage logs and the physical electrical network. We then aggregated assets for a prediction on a 2 by 2 square

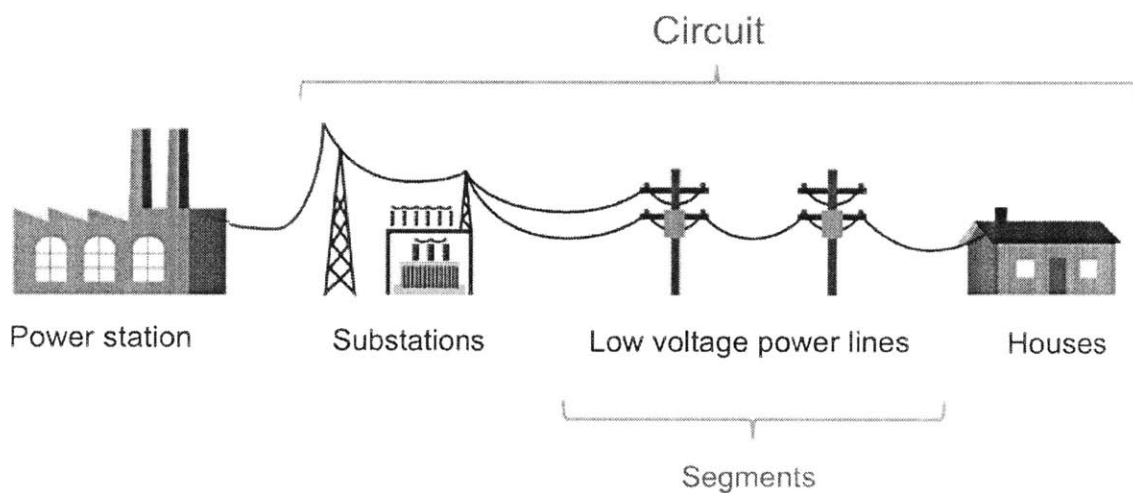


Figure 2.1: High level representation of the electricity network. In this work, we are focusing on the final two elements of the distribution network: circuits and segments. Recall that each segment corresponds to a small portion of the electricity network from substations (where power arrives to each circuit) to individual houses. We do not consider the network upstream of the power generation substations. Devices and segments are then grouped into 'assets'. Each asset comprises of a single device and all the segments that would directly trigger it to open.

mile area, as represented by Figure 2.2.

Furthermore, to encompass external properties surrounding the network we overlaid a geographical land cover mapping that added tree coverage, elevation and population density information to each asset. These factors were crucial to capture as they are very likely to influence outages, given severe weather factors such as high wind speeds or heavy precipitation. Figure 2.3 shows the distribution of segments across landcover types.

2.1.2 Historical Weather Data

Our primary objective is to identify future damage based on weather reports before a storm. However, obtaining historical weather forecast data proved to be a challenge, so we initially used historical weather data at the actual time of the storm. This allowed us to identify significant factors and test the models predictive accuracy when it was given ideal retrospective information. The historical weather data contained approximately 5.2M hourly logs over the course of severe weather events between 2008 to 2012. These logs came from 234 stations across Massachusetts and contained records of 20 weather parameters including: time, wind speed, temperature, pressure and humidity. We used a triangulation algorithm to assign the weather to a given asset. This algorithm created a weather vector by averaging the numerical hourly factors at the weather stations closest to an asset.

However, this data was not sufficient for our desired prediction, as weather typically deviates a great deal from forecasts. We therefore acquired historical weather forecast data that was comprehensive, but of less granular quality (as illustrated by Figure 2.4). Instead of hourly logs, the model incorporating forecasts used daily logs. There are only 20 stations from which we obtained forecast information. Furthermore, there was a great deal of forecasting error depending on how many days before the event the data was taken. The greatest discrepancy between the two data sources is the lack of key weather features in the forecast logs. Of the five most significant factors (as identified by the model during the training on the historical weather logs), four were not available in the forecasting logs.



Figure 2.2: This figure illustrates the density of the local distribution electrical network in Massachusetts. The state is partitioned into square areas of 2 miles per side which are colored by the number of assets they contain. In particular, notice that this electricity provider covers most of the rural areas but not the Boston agglomeration. Finally, the orange circles represent the platforms where the company stages repair crews before and during large storms.

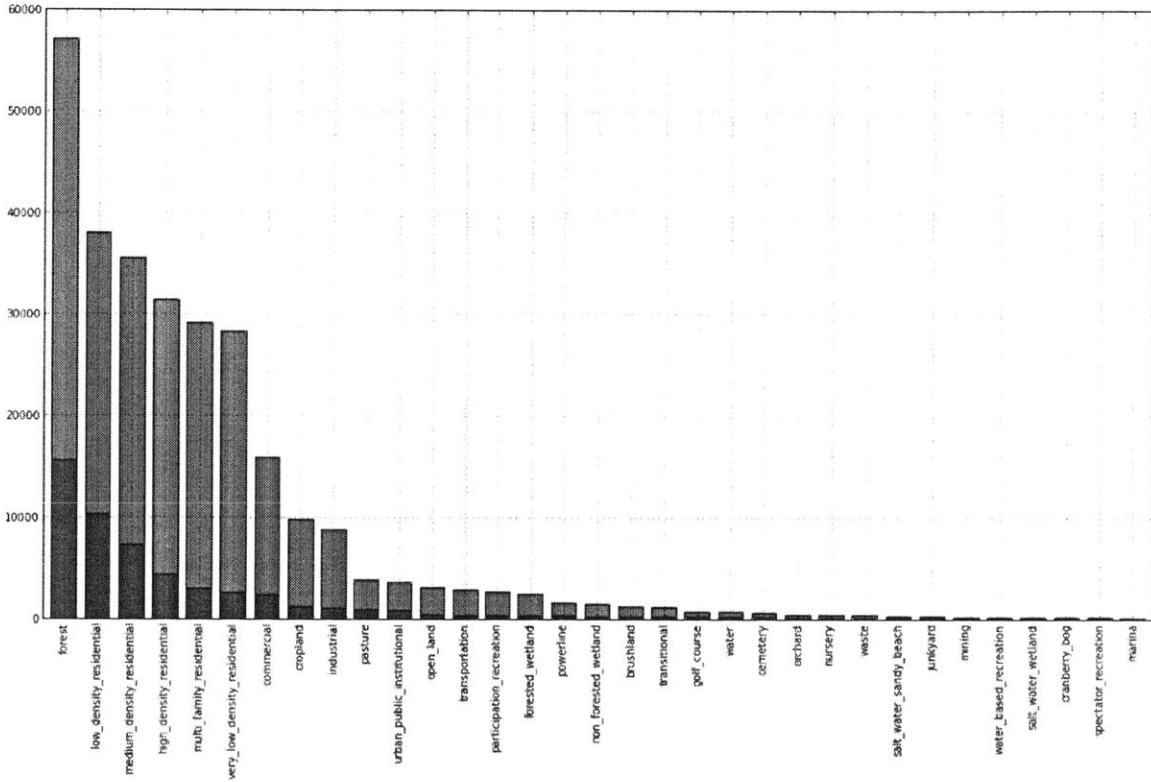


Figure 2.3: This histogram depicts the variety of landcover types where segments lie. The green bars represent the total number of segments corresponding to each category, while the blue bars represent the total number of segments corresponding to each category that are a part of an asset which has been damaged in one of the six storms considered in this work.

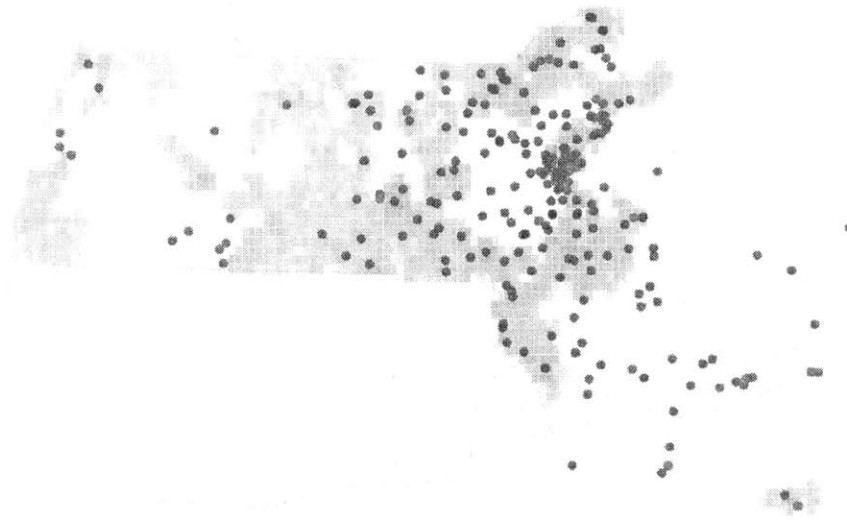


Figure 2.4: This map shows the locations of the weather stations used in this work. The blue circles represent Weatherbug (<http://weather.weatherbug.com/>) stations which provide hourly weather readings (but not forecasts), the orange circles represent weather stations from a variety of companies and aggregated by Intellovations (<http://www.intellovations.com/>) which provide daily weather forecasts. The grey areas in the background are colored by density of assets (darker means more assets in that area).

2.1.3 Historical Outage Data

The final piece in building our database consisted of outage logs from the weather events of interest. These were records from 6 major storms and two years worth of minor weather incidents. The large storms resulted in about 6,000 outages and 1.1M customers out of power (see Table 2.1 below), and this information was supplemented by approximately another 25,000 outages from the minor events. This indicates that only approximately 2% of assets failed during severe weather, emphasizing the challenge of predicting due to such a small scale.

Storm name	First outage	Days	Outages	Customers out
Winter Storm December 2008	2008-12-12	10	1784	185931
Wind Storm February 2010	2010-02-24	6	615	151350
Winter Storm December 2010	2010-12-26	4	444	106347
Tropical Storm Irene 2011	2011-08-28	8	1715	225567
Winter Storm October 2011	2011-10-29	11	2746	291672
Hurricane Sandy 2012	2012-10-29	7	1466	180416

Table 2.1: Storm outage data

2.2 Model

Our framework aims to capture the complex interactions between the electrical network and the weather which cause power outages. We start by presenting the general optimization model, then consider two different cases: a more exact one which requires more information from the power utility to run, and a coarser one which can readily be applied by the power utility with the data it has currently available.

2.2.1 Formulation

The interactions between weather conditions and the properties of the power network which generate failures are difficult to model. In particular, they are highly non-linear and aren't amenable to most function families found in the literature. For this reason, many traditional models (e.g. logistic regression) perform relatively poorly when used directly. However, we can use properties of the problem to make the following assumption:

Assumption 2.2.1. Damaging events happen independently on the power network as a Poisson process. The corresponding rate depends on both the weather features and the physical properties of the network at the time and place considered.

Moreover, since there is relatively little data available to us, we make use of the structure of the network to capture as much information as possible. The electrical network can be thought of abstractly as a tree where branches can represent different types of power lines (e.g. overhead lines with no framing, underground lines, etc.). Some types may be more vulnerable to failures than others. For example we can expect that a bare wire supported by poles in a wooded area will be more likely to fail than an underground line. The network is then partitioned into *segments*. Segments correspond to a small portion of power lines (of length roughly a few hundred feet), which allows us assume the following:

Assumption 2.2.2. The damage rate along each segment can be assumed constant.

Combining Assumption 2.2.1 and Assumption 2.2.2, we can then write that failure events happen on these segments independently across the network and at a rate proportional to their length, and linearly dependent in the surrounding weather features w_t :

$$\lambda_{s,t} = l_s g_{c_s}^* w_t \tag{2.1}$$

l_s is the length of segment s and g_{c_s} is a vector of vulnerabilities for segment type c_s (c_s being the type of segment s). We do not assume a priori any form for the weather features, which therefore permits us to capture arbitrary functions of the weather. Note also that our framework allows for segment types to be more prone to failures under certain weather conditions but less under other weather conditions. As we will show later in this chapter, this turns out to be a key driver in the final performance of our predictions.

Ideally, we would be able to study failures at the segment level. However, the granularity of the data available doesn't currently support it. We therefore batch failure events by *asset* (as described in Section 2.1). Since events are assumed to happen independently across the

network, we get that events happen on each asset a with rate:

$$\begin{aligned}\lambda_{a,t} &= \sum_{s \in a} l_s g_{c_s}^* w_t \\ &= \sum_c l_{a,c} g_c^* w_t \\ &= l_a^* g w_t\end{aligned}$$

where l_a is the vector of lengths of each segment type in asset a , and g is the matrix consisting of the vectors g_c assembled column-wise. The previous equation is linear in terms of the coefficients of the matrix g . By rearranging the terms, we can therefore rewrite it in a more common way as a scalar product between vectors:

$$\lambda_{a,t} = \gamma^* x_{a,t} \tag{2.2}$$

where $x_{a,t}$ is the vector representation of the matrix $l_a w_t^*$ which contains the features for an asset under specific weather conditions. From now on, we will use the subscript i to represent the combination (a, t) of a given asset and a particular time. Therefore, by definition of a Poisson process, the total number Y_i of events on each asset is a Poisson random variable with parameter:

$$Y_i \sim \mathcal{P}(\gamma^* x_i) \tag{2.3}$$

From here we also get the probability distribution of a failure happening on a given asset. Let Z_i be the indicator variable corresponding to at least an event having happened on an asset: $Z_i = 1$ if and only if at least one event occurred on one of its segments, i.e. $Z_i = \min(Y_i, 1)$. Therefore Z_i is distributed according to the following exponential distribution:

$$Z_i \sim \mathcal{E}(1 - e^{\gamma^* x_i}) \tag{2.4}$$

We are now interested in estimating γ . We consider two cases. First the case where Y_i is observable (i.e. we have access to the underlying event data), then the case where only Z_i

is observed. In the next two subsections, we show that the maximum likelihood estimator for γ can be efficiently estimated in both cases.

2.2.2 Without Censoring

In this case, we are able to recover the exact count of events that occurred on an asset. The corresponding likelihood function is therefore given by:

$$\hat{\mathcal{L}}(\gamma) = \mathbb{P}(\forall i, Y_i = y_i | \gamma) \quad (2.5)$$

As is commonly done in the literature, we will focus on the log-likelihood loss function (where K is a constant that doesn't affect the optimal choice of parameters):

$$\begin{aligned} \hat{L}(\gamma) &= -\ln \hat{\mathcal{L}}(\gamma) + K \\ &= \sum_i [\gamma^* x_i - y_i \ln(\gamma^* x_i)] \end{aligned}$$

We then have the following result:

Theorem 2.2.3. *The following maximum likelihood problem is convex:*

$$\max_{\gamma} \hat{L}(\gamma) \quad (2.6)$$

Proof. Convexity is immediate when observing that the Hessian matrix of the likelihood function can be written as follows:

$$\begin{aligned} H &= \left(\sum_i y_i \frac{x_{i,k} x_{i,l}}{(\gamma^* x_i)^2} \right)_{(k,l)} \\ &= (Dx)^*(Dx) \end{aligned}$$

D is a diagonal matrix with entries $D_{i,i} = \frac{\sqrt{y_i}}{\gamma^* x_i}$. H is therefore a positive semi-definite matrix. \square

2.2.3 With Censoring

Unfortunately, we are unable to optimize the log-likelihood function described in the previous subsection with the data currently available. We therefore, at least temporarily, focus ourselves on the following likelihood function instead:

$$\tilde{\mathcal{L}}(\gamma) = \mathbb{P}(\forall i, Z_i = z_i | \gamma) \tag{2.7}$$

We use the corresponding log-likelihood function:

$$\begin{aligned} \tilde{L}(\gamma) &= -\ln \tilde{\mathcal{L}}(\gamma) + K \\ &= \sum_{z_i=0} \gamma^* x_i - \sum_{z_i=1} \ln(1 - e^{-\gamma^* x_i}) \end{aligned}$$

We get a similar result in this setting:

Theorem 2.2.4. *The following maximum likelihood problem is convex:*

$$\max_{\gamma} \tilde{L}(\gamma) \tag{2.8}$$

Proof. Once again, the Hessian matrix is positive semi-definite as can be seen in the following expression:

$$\begin{aligned} H &= \left(\sum_i z_i \frac{x_{i,k} x_{i,l}}{(e^{\gamma^* x_i} - 1)^2} \right)_{(k,l)} \\ &= (Dx)^*(Dx) \end{aligned}$$

where D is a diagonal matrix with entries $D_{i,i} = \frac{z_i}{e^{\gamma^* x_i} - 1}$. □

2.3 Application

The data described in Section 2.1 was consolidated into a single database in order to run the model described in Section 2.2 efficiently. We now got through the steps involved to

generate outage predictions from the raw data.

2.3.1 Segment Clustering

The first step is to group the different segments into types. We do this using a clustering algorithm: k-means (also referred to as Lloyd’s algorithm, cf. [32]). This algorithm aims to partition the segments into k clusters in which each segment is associated with the center of its corresponding cluster. The number k of centers was initially set to 20, and eventually brought down to 10 (which yielded better overall prediction performance). A higher number of centers leads to better representation of the segments but also more parameters to optimize in the final model. In order for the clustering algorithm to work, some of the raw features were normalized by the length of the segment: the total number of poles (respectively total number of customers) was converted to pole density (respectively customer density). Without this preprocessing step, the clustering algorithm gives very poor results as there is a large variance in segment lengths (from 1 meter to 86 miles).

Each segment type is then further categorized by landcover (the 33 landcover categories were brought down to 10 by grouping similar ones together): forested areas, highly residential areas, open rural areas, etc.

We therefore now have 100 different segments types corresponding to different physical properties of the segment (e.g. framing, pole density, wire width, etc.) and the surrounding landcover. By simply aggregating the total length of each segment type for each asset, we have generated features compatible with the model described in Section 2.2.

2.3.2 Using Actual Weather Logs

We first ran our algorithm using historical weather logs, i.e. hourly weather readings from the 234 Weatherbug stations of previous storm days across Massachusetts. We first analyzed the raw data to discover correlations between the different variables (e.g. ‘wind speed’ and ‘average wind speed’, which both correspond to the same quantity averaged differently) and removed the highly-correlated and non-relevant columns (for example indoor temperature). We then normalized the data and pruned outliers before transforming the data by

performing Principle Component Analysis (PCA).

Combining these with the physical segment features from Subsection 2.3.1, we are ready to run the model. We maximize equation (2.6) using a bounded gradient ascent algorithm, adding an $L1$ regularization factor (this doesn't change the convexity results from the previous section since we are restricting ourselves to non-negative parameter values), and measure the performance of the resulting parameters by doing cross-validation on the six storms: we train the model on five storms and test on the left-out storm. We use the Pearson correlation between predicted outages and true outages as performance indicator. This can be done at several levels of aggregation. Indeed, although the prediction granularity is at the asset (i.e. device) level, predictions by area (i.e. on a 2 by 2 square mile basis) and by platform are also extremely relevant. The last one is sufficient for the power utility to plan how it will dispatch repair crews. Recall that a platform is a staging area where crews are positioned before a storm, from where they work on a daily basis during restoration times. In practice, approximately 6 platforms are opened during severe weather events, and there are about 30 total across Massachusetts. Each platform accounts for a subset of specific towns and cities, so this aggregation was convenient since each outage log entry contained a device location.

The results for different levels of aggregation are show in Table 2.2.

Figure 2.5 shows a visual representation of the distribution of outages across platforms. Both the predicted and actual counts are displayed and easily compared. The example shown is for Hurricane Sandy, which hit Massachusetts relatively hard, mostly along the coast. For this particular run, the algorithm performed very well (with a correlation factor of 85%). Crew placements using these expected outages would have been very close to optimal.

2.3.3 Using Weather Forecasts

As described under the data subsection in Section 2.1, we were limited in the number of available weather stations that could provide historical forecast logs. Specifically, we had less than 10% of the previously available number of stations, as depicted by the map in

Storm name	Aggregation level		
	Device	Area	Platform
Winter Storm December 2008	0.17	0.37	0.55
Wind Storm February 2010	0.10	0.23	0.50
Winter Storm December 2010	0.04	0.37	0.48
Tropical Storm Irene 2011	0.17	0.56	0.79
Winter Storm October 2011	0.22	0.50	0.67
Hurricane Sandy 2012	0.16	0.52	0.85
Average	0.14	0.43	0.64

Table 2.2: Out of sample correlation for outage prediction

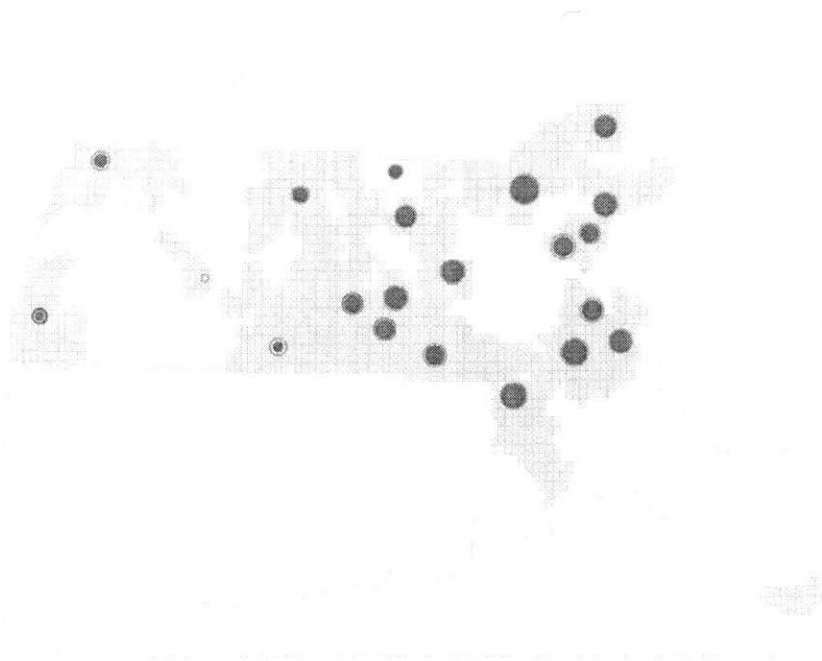


Figure 2.5: Comparison between the true and expected number of outages aggregated by platform for Hurricane Sandy. The blue circles, centered on staging platforms, are proportional to the predicted number of outages at that platform. The hollow orange circles are proportional to the true number of outages. This map corresponds to a correlation of 85% (cf. Table 2.2).

Figure 2.4.

Consequently, to cover the entirety of the assets across the state when using forecast weather data, we had to expand the minimum distance between devices and weather stations. If we use only a 5 mile radius around a given asset, we capture information for only 17.5% of the devices (10,814 out of 61,704). We enlarged this to a 20 mile radius and were able to capture 97.7% of all devices (60,289), and about 98.6% of customers (953,656 out of 966,750). Note that this entails a significant decrease in prediction accuracy and granularity since the weather features for a given asset may be taken from a location up to 20 miles away from it.

Moreover, the weather forecast logs severely limit how much information we can input in terms of factors that were available in the historical weather logs. The list of the available factors includes: temperature high and low, probability of precipitation, average wind speed and direction, presence of haze or fog, extreme heat or cold indicators, wind category, chances of rain, snow or thunderstorms and the sky condition. However, we are missing precipitation rates, hourly gusts and several other factors available in the historical logs and which were identified as significant when training the model in Subsection 2.3.2.

We then found numerical results using the same methodology and out of sample testing, but now using forecast data from the day of the weather event as input. These results are shown in Table 2.3.

Addressing the practical purposes of our tool, the day-of weather forecast is not sufficient for a prediction because a distributor contracts crews up to five days ahead and places them at platforms as least a day in advance. Therefore, we obtained numerical results using the forecast data each day or up to a week before the storm. The results are presented in Table 2.4.

Additionally, we found that as the number of days before the storm increases, the forecast data demonstrates increasing discrepancies with actual data. As an example, consider the figures below. Fig. 5 demonstrates the error in the forecast of wind speed first 3 days, then 7 days and finally 14 days before the actually measured amount, which is indicated by the grey bar in the center. If we then consider Figure 2.6, we see the mean of the forecast error

Storm name	Aggregation level					
	Device		Area		Platform	
Winter Storm December 2008	0.07	-0.10	0.25	-0.12	0.35	-0.20
Wind Storm February 2010	0.05	-0.05	0.20	-0.03	0.25	-0.25
Winter Storm December 2010	0.03	-0.01	0.22	-0.15	0.26	-0.22
Tropical Storm Irene 2011	0.08	-0.09	0.37	-0.19	0.63	-0.16
Winter Storm October 2011	0.11	-0.11	0.34	-0.16	0.47	-0.20
Hurricane Sandy 2012	0.07	-0.09	0.32	-0.20	0.68	-0.17
Average	0.07	-0.07	0.28	-0.15	0.44	-0.20

Table 2.3: Out of sample correlation for outage prediction using ‘day of’ forecast data. The smaller numbers indicate the loss compared to the correlations using actual weather date from Table 2.2

Storm name	Days ahead								
	0	1	2	3	4	5	6	7	
Winter Storm December 2008	0.35	0.34	0.34	0.30	0.30	0.32	0.29	0.26	
Wind Storm February 2010	0.25	0.24	0.24	0.23	0.20	0.20	0.18	0.20	
Winter Storm December 2010	0.26	0.24	0.24	0.25	0.25	0.24	0.23	0.22	
Tropical Storm Irene 2011	0.63	0.56	0.58	0.51	0.51	0.53	0.52	0.51	
Winter Storm October 2011	0.47	0.47	0.40	0.34	0.32	0.31	0.31	0.29	
Hurricane Sandy 2012	0.68	0.67	0.66	0.68	0.66	0.67	0.63	0.64	
Average	0.44	0.42	0.41	0.39	0.37	0.38	0.36	0.35	

Table 2.4: Out of sample correlation for outage prediction at the platform level using forecast data from different days ahead. Note that the quality of the prediction doesn’t always decrease as the forecasting horizon increase. This is because of the error fluctuations in weather forecasting: forecasts from a week in advance might be closer to reality than forecasts from three days in advance.

and the standard deviation of the error, which increases drastically with time.

2.3.4 Implementation

The algorithm previously described was implemented as a Python application, which directly interfaces with the MySQL database where the data is stored. Most of the data processing and analysis was done using Python libraries (pandas¹, scikit-learn², numpy³, scipy⁴) which are fast enough to allow for the relatively large size of our datasets. Our application also included a front end, built with Flask⁵, which serves an API exposing the database and a two-part web tool.

The first part of the tool is an interface to run tests from uploaded weather forecast data (which is then saved and can be reused for later runs). The utility can select which platforms are open, which will affect the final display (as expected total outages will be aggregated only toward open platforms). The training set of storms can also be defined at this stage by choosing storms from the available historical data. This is helpful for example in order to train the algorithm on certain types of storms exclusively (for instance wind storms, which cause very different damage from ice storms). Finally, some of the model parameters may also be tuned from this page (e.g. the number of segment clusters, the horizon of training forecasts, etc.). Figure 2.8 below shows a screen capture from this interface.

Once the algorithm has converged, the tool displays a second screen with a visualization of outage localizations similar to that of Figure 2.5, built using the JavaScript library d3.js⁶. A sample screen capture can be seen in Figure 2.9. Note that only open platforms are present in the table on the right.

The second part of the tool allows the company to add training data to the model (namely historical weather forecasts and power outages), which will then be added the available sets in the testing interface.

¹<http://pandas.pydata.org/>

²<http://scikit-learn.org/>

³<http://www.numpy.org/>

⁴<http://www.scipy.org/>

⁵<http://flask.pocoo.org/>

⁶<http://d3js.org/>

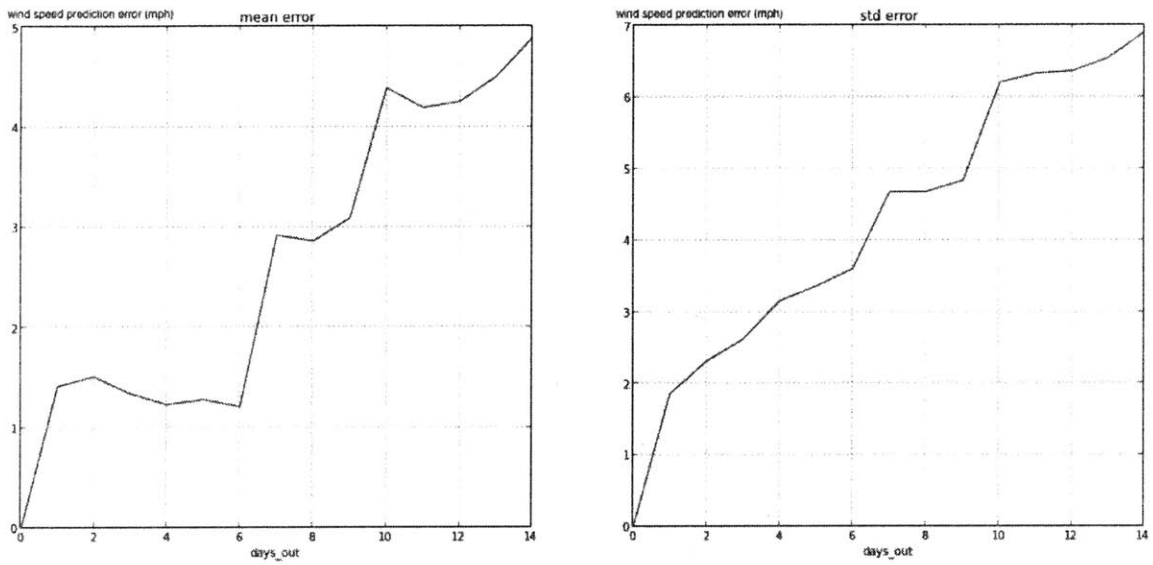


Figure 2.6: Mean (left) and standard error (right) of wind speed forecasting error (in miles per hour) as a function of the number of days ahead the prediction was made.

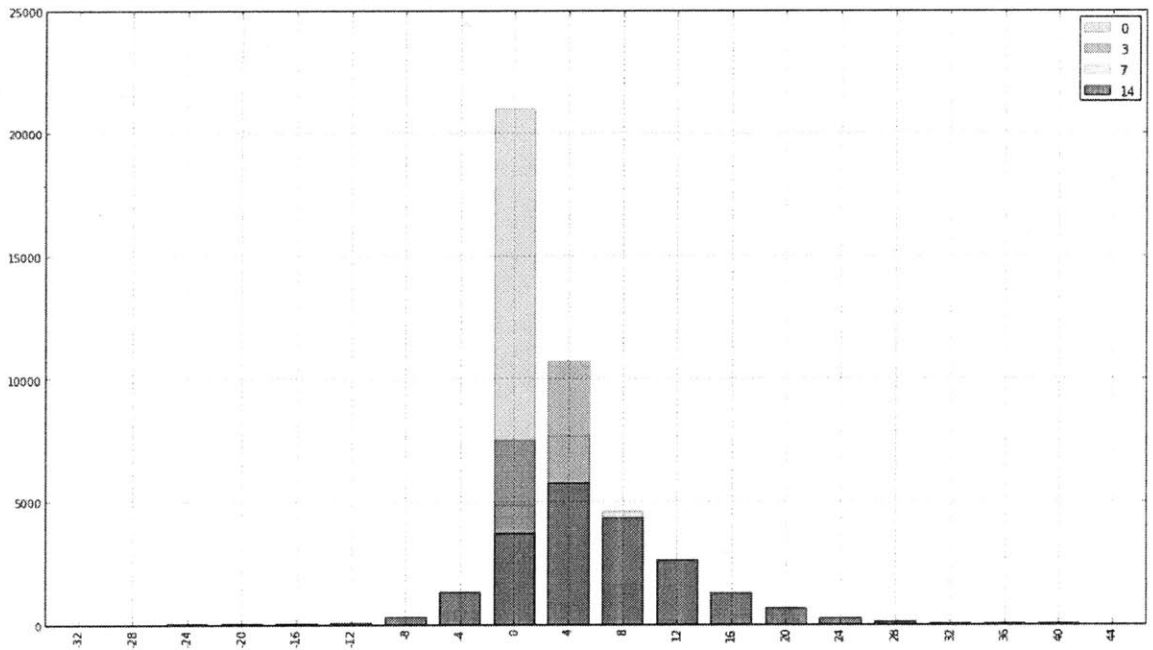


Figure 2.7: Histogram of wind speed forecasting error (in miles per hour) for predictions from different days ahead.

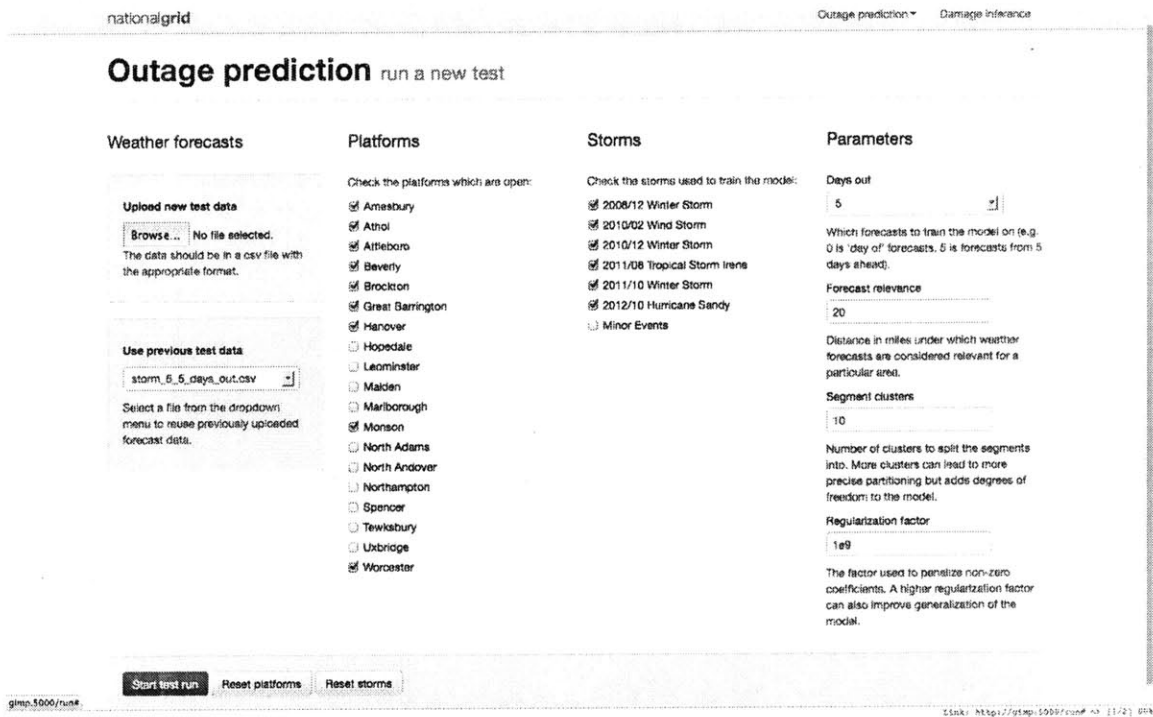


Figure 2.8: Online prediction tool, test interface

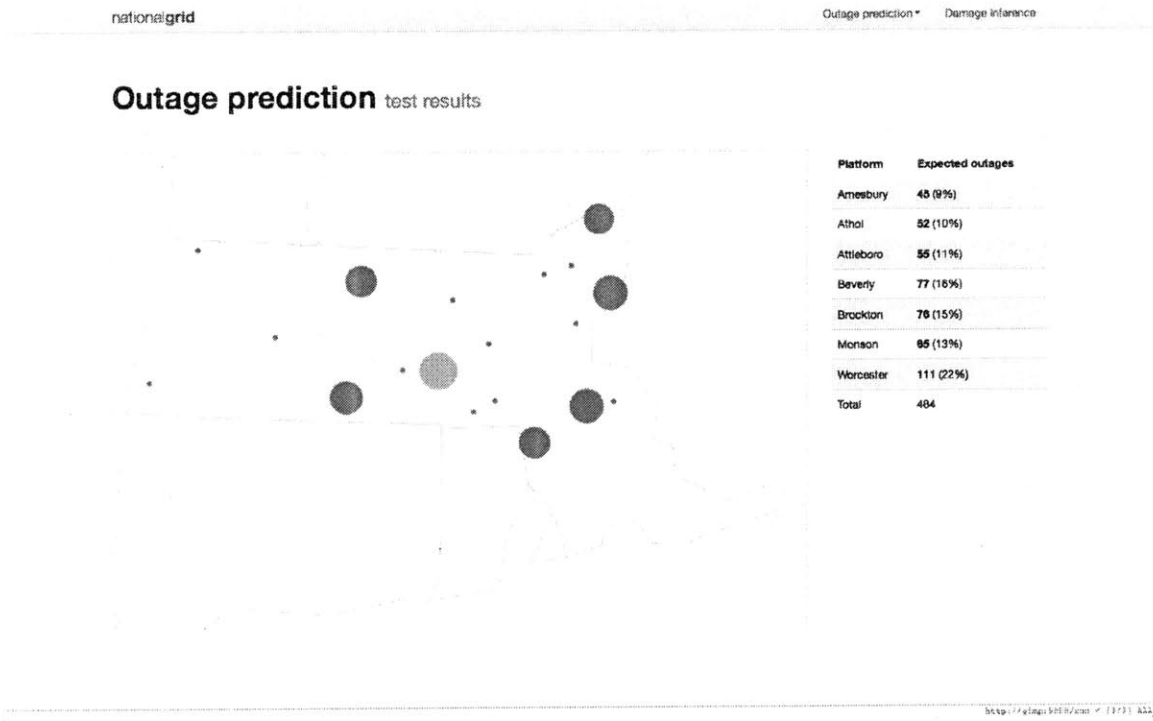


Figure 2.9: Online prediction tool, sample run results

As we will describe in the next chapter, the output of the outage prediction model is used to find the optimal crew assignments across platforms. However, this web tool also allows the utility to run the prediction model for different combinations of parameters, and potentially use the results directly to help in emergency response decision taking. Having some insight on which areas are likely to be hit hardest could for example be useful to find out which platforms to open.

3 Optimizing Crew Dispatch

During storm events the electricity provider stations its repair crews at sites called platforms. These platforms are the home base for these crews through the duration of the storm (or until they re-allocate crews). Once stationed crews repair damage to the company's electrical system when outages occur in locations that are also associated with their platform. Storm response is not only a large driver of public opinion, it also has a significant impact on costs. Large storms incur total repairs that number in the millions of dollars, much of it which the company may not be reimbursed by the Department of Public Utilities (DPU).

The previous prediction model produced predicted outages at the platform level based upon weather forecast input. The next step is to utilize those predictions to aid in storm response planning. An understanding of where they anticipate damage will allow the power utility to station crews so that they are best suited to repair damage in the fastest time possible. This will ultimately return service to customers quicker, reduce costs, and aid in justifying their actions to the DPU in their regulatory filings.

3.1 Deterministic Crew Optimization

3.1.1 Master Formulation

To solve the problem of storm response planning we created a formulation that ultimately decides where to station crews and what jobs each crew will complete. While this does not necessarily need to be a 'day to day' scheduler it ultimately gives the company the ability to better understand the anticipated work of each repair crew as a result of expected outages

from the damage prediction model. The complete formulation and explanation of variables is given below

	Notation	Description
Decisions	X_{ijk}	Crew i assigned to job j at platform k
	X_{ik}	Crew i assigned to platform k
	Z	Objective value equal to the worst case repair time
Data	γ_{jk}	Time required to do job j from platform k (including travel)
	M_k	Crew capacity for platform k

Table 3.1: Master Formulation Variable Notation

It should be noted that we are currently examining γ_{jk} values that take on a known value. In reality these repair times are unknown, as repairs can be caused by any combination of problems with varying repair time, and these unknowns will have serious effects on the optimization formulation. Moreover their variability can be large (especially since they include travel time from the platform to the damage location). The uncertainty of γ_{jk} and its effects on the optimization problem will be discussed later in this chapter.

Our objective is to minimize the time until all customers have their power restored. Other possible objectives could include time until a given fraction (e.g. 90%) of these customers have power back or minimizing cost, however the total time until all repairs are complete is often the metric used by regulators (which is closer to our chosen objective) and we currently lack any historical cost data to perform any such analysis.

In order to achieve this objective, the company currently relies on the intuition and experience of employees who have witnessed several past storms. Our approach aims to provide a quantitative alternative to this process. The model can be represented mathematically as follows:

Objective: Minimize Z subject to the following constraints

$$\sum_{j,k} \gamma_{jk} X_{ijk} \leq Z \quad \forall i \quad (3.1)$$

$$\sum_i X_{i,k} \leq M_k \quad \forall k \quad (3.2)$$

$$\sum_{i,k} X_{ijk} \geq 1 \quad \forall j \quad (3.3)$$

$$\sum_k X_{ik} \leq 1 \quad \forall i \quad (3.4)$$

$$X_{ijk} \leq X_{ik} \quad \forall i, j, k \quad (3.5)$$

$$X_{ijk}, X_{ik} \in \{0, 1\} \quad \forall i, j, k \quad (3.6)$$

Platform capacities are ensured by constraint (3.2) while constraint (3.3) ensures that all jobs are completed. Constraints (3.4) and (3.5) ensure that crews are assigned to only one station and they can only repair jobs that are assigned to that station. Constraint (3.1) stipulates that no repair crew's total repair time can exceed that of the "worst repair time" given by the objective value Z .

This solution does obtain optimal solutions however at very long solve times. Given the nature of variables and constraints, serious storm events will have a significant increase in the number of decision variables and constraints. Consider a case of 300 crews, 600 outages, and 6 platforms to station those crews. The resulting formulation has approximately 2 million decision variables and constraints. Solve time for this particular case was on the order of days (relaxing X_{ijk} to a continuous variable on the range $[0, 1]$ still produced integral solutions but the solve time was still insufficiently long). A typical storm, such as hurricane Sandy, is roughly ten times as big, which renders this formulation impractical for real-time use cases.

It is possible that the software was exploring an extraneous number of solutions given the potential symmetry of those solutions. However even achieving gap values that were close to 3% would take hours. Given the importance of timing in storm response planning an alternate simplified formulation was explored.

3.1.2 Relaxed Formulation

Deeper analysis of the previous formulation showed that while we were obtaining optimal solutions, crews were essentially splitting the work evenly at each platform. In order to achieve the best state wide completion time, the optimization would drive solutions such that all crews were completing their work at approximately the same time. By making the assumption that crews evenly split the work at each platform we can drastically simplify the problem. We no longer need to assign crews to jobs and platforms. Now our decisions are reduced to assigning jobs to platforms and the number of crews to each platform.

	Notation	Description
Decisions	X_{jk}	Job j assigned to platform k
	C_k	Platform k workload
	C_k^*	Number of crews assigned to platform k
	C	Objective value indicating worst platform workload
Data	γ_{jk}	Time required to do job j from platform k
	M_k	Crew capacity for platform k
	C^*	Total number of crews available

Table 3.2: Relaxed Formulation Variable Notation

The new mathematical model can be represented as the following:

Objective: Minimize C subject to the following constraints

$$\sum_k C_k \leq C \tag{3.7}$$

$$\sum_j \gamma_{jk} X_{jk} \leq C_k \quad \forall k \tag{3.8}$$

$$\sum_k X_{jk} \geq 1 \quad \forall j \tag{3.9}$$

We interpret the number of crews at each platform as the following:

$$C_k^* = \frac{C_k}{C} C^* \tag{3.10}$$

$$C_k^* \leq M_k \quad \forall k \tag{3.11}$$

The constraints are analogous to those in the master formulation. The major difference being that we ensure a statewide completion time with a combination of constraints (3.7)

and (3.8). Constraints for ensuring all jobs are completed and all platform capacities are met are similarly modified from the original formulation. Figure 3.1 demonstrates how the optimization assigns jobs to each platform.

This new formulation is a significant decrease in problem complexity. In our master solution we examined one potential scenario of 300 crews, 600 jobs, and 6 platforms which contained approximately 2 million decision variables and constraints. The relaxed formulation reduces the problem to approximately 7 thousand decision variables and constraints and solve time is now on the order of seconds. Typical storms will result in a relaxed formulation with less than 20 thousand variables and solve in under a minute.

3.1.3 Comparison

While the improvement in computation time for the relaxed solution will allow the power utility to adequately utilize the model within the time constraints imposed by storm operations, it is important to verify that the new formulation produces adequately optimal results given the assumptions made. In order to validate the relaxed model we will first look at how accurate our assumption that the model allocates work among crews evenly in the master solution. Next we will compare workload of crews in the master formulation and the relaxed formulation at a platform level (it should be noted that here we are still using nominal γ_{jk} values as input data into our models).

In order to test the validity of our assumption we ran the master formulation on several notional data set scenarios. The data sets were created by randomly sampling outages from previous storms and only opening two platforms for stationing. These are representative of actual storm scenarios that the company faces but are on a smaller scale to manage solve time. The histograms below the crew workloads for each platform for all three scenarios.

Ultimately the crew workload differ by only minutes (compared to hours of total work). The difference is due to the fact that we don't allow crews to go "help" another crew that is still finishing a job if they have finished early. Overall these differences suggest that our assumption of equal distribution of workload is a reasonable one.

Now we will examine the same three scenarios and compare the optimization results from the original and relaxed formulations. Ultimately we want to ensure that both optimization results are producing similar numbers of crews at the platform level. Returning to the same three scenarios we see the following results (the tuples in each entry represent the results for simulations one, two, and three respectively).

	Platform 1	Platform 2
Master Formulation	(6,11,10)	(33,29,28)
Relaxed Formulation	(7,9,11)	(32,31,27)

Table 3.3: Master Formulation vs Relaxed Formulation Results

Our relaxed formulation never deviates from the master solution by more than a few crews. Given the level of granularity of data and our current accuracy for outage prediction, we believe that this is sufficiently close to the optimum solution. Individual crew workloads may differ by values more than 10% on occasion but the ultimate operational question we are trying to answer is quite close to the master counterpart.

3.2 Robust Optimization

As stated earlier we have been dealing with known values of γ_{jk} in our optimization formulations. However, calibrating these values accurately is very hard. We cite below a few reasons behind this:

- Predictions currently give estimates of \mathbb{P} (outage). Inferring damage from these outage probabilities decreases the prediction precision. Therefore the total number of jobs expected might be relatively further from the actual number of jobs than the correlations from the previous chapter might suggest.
- Any given outage can be caused by a combination of issues (i.e. an outage can be caused by a broken pole and three down trees or it can be caused icing on a transformer causing open breakers and downed wires) which lead to variable repair times

It is impossible to escape these uncertainties so we must program them into our optimization to ensure that our solutions are always valid despite the variation in potential workload

values. Recall from our relaxed formulation that the only constraint using our unknown γ_{jk} values was the following:

$$\forall k, \sum_j \gamma_{jk} X_{jk} \leq C_k$$

We can amend this constraint to include all possible values of γ_{jk} and ensure that any solution given by our optimization will be a valid one with the following:

$$\sum_j \gamma_{jk} X_{jk} \leq C_k, \gamma \in \mathcal{U}, \forall k$$

Here \mathcal{U} is the set of all values that γ can potentially take on. Given this new constraint all solutions from the model are guaranteed to be valid. However in this current form we no longer have a mixed integer problem (MIP). Understanding the nature of \mathcal{U} is important to re-modeling the formulation back to a MIP [7].

3.2.1 Using Box Constraints

The simplest method is to assume all γ_{jk} values reside in a box $[l_{jk}, u_{jk}]$. Now our uncertainty set \mathcal{U} takes the following form:

$$\mathcal{U}_k = \{\gamma | \forall j, l_{jk} \leq \gamma_{jk} \leq u_{jk}\} \forall k$$

Recall our original constraint which contained uncertainty:

$$\sum_j \gamma_{jk} X_{jk} \leq C_k, \gamma_{jk} \in \mathcal{U}_{jk}, \forall k$$

The above constraint can now be re-written as:

$$\sum_j u_{jk} X_{jk} \leq C_k, \forall k$$

While the implementation of this solution is easy note that we now assume that all jobs take on the worst case scenario value. Given that we assumed all predictions of outages happened

independently we can assume that their respective repair times are also independent random variables. Assuming that all of the restoration times will take on their upper bound value simultaneously is highly unlikely and yields significantly lower objective value.

3.2.2 Using Bertsimas-Sim Uncertainty Sets

Instead of assuming all restoration times assume a worst case value we can now consider a case where the uncertainty set \mathcal{U} is comprised of values where a fraction assume a worst case scenario and the remainder are forced to their nominal values. Consider the following representation of the uncertainty set:

$$\mathcal{U}_k = \left\{ \gamma | \forall k, \gamma_{jk} \in [\gamma_{jk}^- - \hat{\gamma}_{jk}, \gamma_{jk}^- + \hat{\gamma}_{jk}], \sum_j \frac{|\gamma_{jk} - \gamma_{jk}^-|}{\hat{\gamma}_{jk}} \leq \Gamma \right\}$$

In the above specification of \mathcal{U}_k , γ_{jk}^- represents the nominal value of γ_{jk} and $\hat{\gamma}_{jk}$ represents the half width on the interval of which γ_{jk} can reside. Γ then bounds the total deviation from the nominal γ_{jk} values in the uncertainty set. Ultimately Γ is a parameter that specifies the number of values that assume the extreme values ($\gamma_{jk}^- - \hat{\gamma}_{jk}, \gamma_{jk}^- + \hat{\gamma}_{jk}$) [7]. Selection of the Γ parameter will be discussed later in this chapter.

To make the constraint robust we can dictate that:

$$\gamma_{jk} = \gamma_{jk}^- + \hat{\gamma}_{jk} u_{jk}$$

where all u_{jk} reside in the following uncertainty set:

$$\mathcal{U}_{k,u} = \left\{ u | \forall j, u_{jk} \in [-1, 1]; \sum_j |u_{jk}| \leq \Gamma \right\}$$

Taking this representation of our uncertainty set $\mathcal{U}_{k,u}$ the original constraint can now be represented as the following:

$$\sum_j \gamma_{jk}^- X_{jk} + \max_{u \in \mathcal{U}_{k,u}} \sum_j u_{jk} \hat{\gamma}_{jk} X_{jk} \leq C_k$$

The max problem on the right hand side is a linear optimization of the following problem:

$$\begin{aligned}
& \text{maximize} && \sum_j (u_{jk}^+ - u_{jk}^-) \hat{\gamma}_{jk} X_{jk} \\
& \text{subject to} && \sum_k u_{jk}^+ + u_{jk}^- \leq \Gamma, && \forall k \\
& && 0 \leq u_{jk}^+, u_{jk}^- \leq 1, && \forall j, k
\end{aligned}$$

The above optimization has a bounded finite region, thus attaining a finite optimum value. By strong duality the following the optimization is also feasible and will obtain the same optimum value [7].

$$\begin{aligned}
& \text{minimize} && \Gamma R_k + \sum_j r_{jk}^+ + r_{jk}^- \\
& \text{subject to} && R_k + r_{jk}^+ \geq \hat{\gamma}_{jk} X_{jk}, && \forall j, k \\
& && R_k + r_{jk}^- \geq -\hat{\gamma}_{jk} X_{jk}, && \forall j, k \\
& && R_k, r_{jk}^+, r_{jk}^- \geq 0 && \forall j, k
\end{aligned}$$

Using the above transformations, our original robust constraint:

$$\sum_j \bar{\gamma}_{jk} X_{jk} + \max_{u \in \mathcal{U}_{k,u}} \sum_j u_{jk} \hat{\gamma}_{jk} X_{jk} \leq C_k$$

Is equivalent to the following set of constraints:

$$\begin{aligned}
& \sum_j [\bar{\gamma}_{jk} X_{jk} + r_{jk}^+ + r_{jk}^-] + \Gamma R_k \leq C_k && \forall k \\
& R_k + r_{jk}^+ \geq \hat{\gamma}_{jk} X_{jk}, && \forall j \\
& R_k + r_{jk}^- \geq -\hat{\gamma}_{jk} X_{jk}, && \forall j \\
& R_k, r_{jk}^+, r_{jk}^- \geq 0 && \forall j, k
\end{aligned}$$

Combining this back with our original relaxed formulation we now have a robust solution that does not unnecessarily limit the objective value of the solution. Note however that the

above solution does include the possibility that some constraints will be violated (i.e. it is possible for the solution to be infeasible) [7]. The probability of this occurrence is dictated by Γ . Larger Γ values will produce more robust formulations but hinder the resulting objective.

Our formulation can now be written completely as the following:

Objective: Minimize C subject to the following constraints

$$\sum_j [\hat{\gamma}_{jk}^- X_{jk} + r_{jk}^+ + r_{jk}^-] + \Gamma R_k \leq C_k \quad \forall k \quad (3.12)$$

$$R_k + r_{jk}^+ \geq \hat{\gamma}_{jk} X_{jk}, \quad \forall j \quad (3.13)$$

$$R_k + r_{jk}^- \geq -\hat{\gamma}_{jk} X_{jk}, \quad \forall j \quad (3.14)$$

$$R_k, r_{jk}^+, r_{jk}^- \geq 0 \quad \forall j, k \quad (3.15)$$

$$\sum_k X_{jk} \geq 1 \quad \forall j \quad (3.16)$$

$$\sum_k C_k \leq C \quad (3.17)$$

We interpret the number of crews at each platform as the following:

$$C_k^* = \frac{C_k}{C} C^* \quad (3.18)$$

$$C_k^* \leq M_k \quad \forall k \quad (3.19)$$

In the next section we will discuss the trade offs between Γ by examining the historical repair information and how that informs our decisions when choosing our half width values ($\hat{\gamma}_{jk}$) and Γ

3.3 Calibrating the Model for the Power Utility

In order to choose the parameters of our robust optimization sufficiently we must first obtain an understanding of previous damaging events. Because our prediction model only predicts outages and not damage (which is a limitation of the data available and not the model), examining historical information will be critical.

Information on damage is rather limited but we have a much richer data set on previous outages. The data set contains data on what device opened, when it opened, how long until it was restored, and how many customers were affected by this event. In order to build a proxy for repair time we decided to use the difference between open and close times and exclude outages during storm events. During a storm a particular outage may be repaired by the circuit might not be energized for a number of reasons

- A circuit will not be re-energized if other crews are repairing damage that a given device feeds to
- Only certain qualified employees can re-energize a circuit, a contract crew may make the repairs but the system will not turn on until someone re-energizes the circuit thus affecting the turn on time
- Storms have "emergency" mode where crews are only repairing damage that causes a risk to the public. Other non-threatening outages therefore have a longer downtime not because repairs take longer but simply because crews are not authorized to work

Because of these operational differences a restoration time is not a good proxy for repair time during a storm. Excluding storm outages from the historical data set we get the following restoration profile (see Figure 3.3).

While the data clearly does not show a normal (or even centered) distribution choosing a nominal γ_{jk} such that it equals the average and a half width value that covers a large portion of the histogram is sufficient. The above histogram then yields that $\bar{\gamma}_{jk} = 166.67$ and a half width value $\hat{\gamma}_{jk} = 544.09$ (equal to 2 times the standard deviation of the data). It should be noted that the above values were all non-storm values restoration times so an average value that is higher than a median is still reasonable. Similarly our half width value, $\hat{\gamma}_{jk}$ we have selected to encompass all of the distribution (with a select few major outliers excluded). These values are a bit conservative but they are still a valid representation of the data and the implementation of the robust solution is significantly easier.

We can generate an imputed Γ histogram using simulation by random draws and calculating

Γ for each simulation using the following.

$$\Gamma = \sum_j \frac{|\gamma_{jk} - \bar{\gamma}_{jk}|}{\hat{\gamma}_{jk}}$$

Running 50 simulations with a scenario that contains 600 outages, using random draws from our historical data with replacement produces Figure 3.4.

Selecting a final Γ parameter such that it encompasses a significant portion of this histogram will ensure that our solution will remain valid with very high probability. A 90th percentile $\Gamma = 15.81$ and is sufficiently robust to meet the operational situation that the utility faces. While even large deviations (both in number of outages and simulations) yield relatively similar Γ values it is recommended that the simulation be re-run with each storm.

It should be noted that while the nominal γ_{jk} value is given by 166.67 in our scenario, we still apply a weighted distance to that value when applying it to each of the platforms. Even though the repair time will be the same the added time helps to account for variables such as travel. This weighted value is tailor-able by the user who can make estimates based upon weather conditions, logistical operations at the time, and other factors that will impact how much time crews are traveling.



Figure 3.1: This figure demonstrates the method used by the optimization to assign each job to a platform. The amount of time to complete a job from a given platform is represented by the length of the corresponding dashed line. Intuitively, a given repair will take longer if the crew assigned to it is dispatched from a further platform.

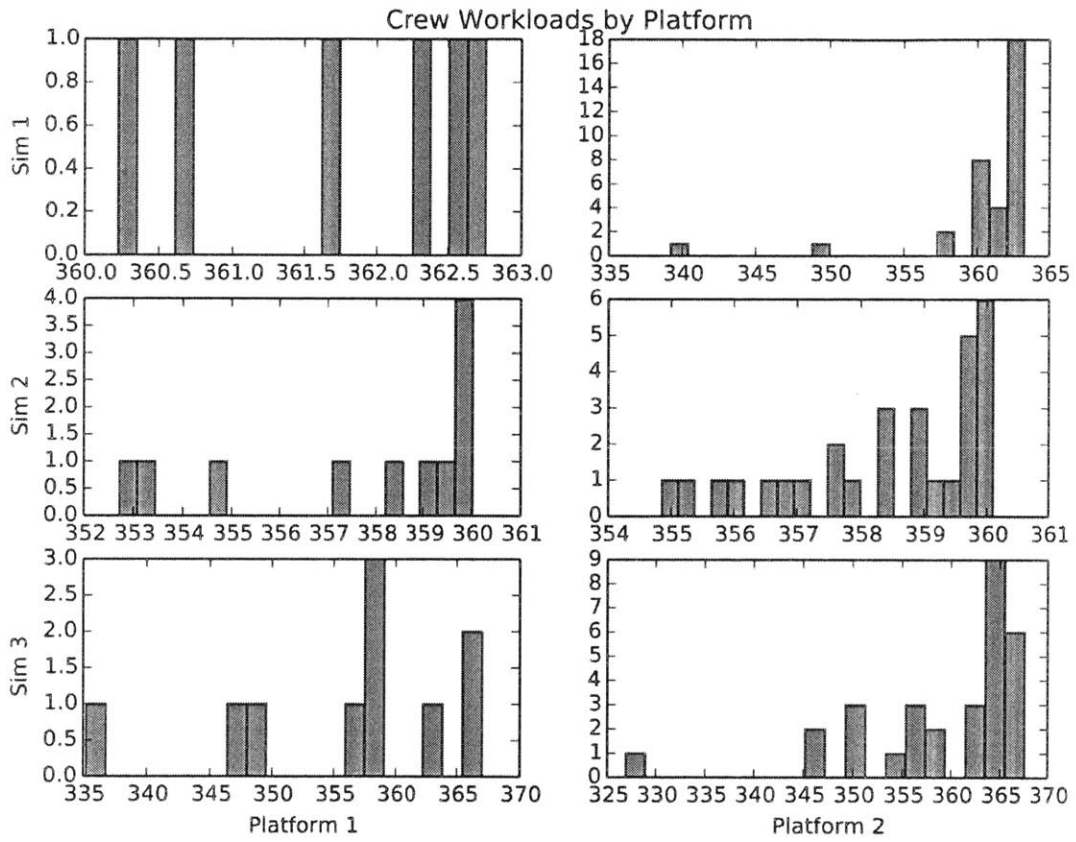


Figure 3.2: Crew workloads for three randomly sampled sets

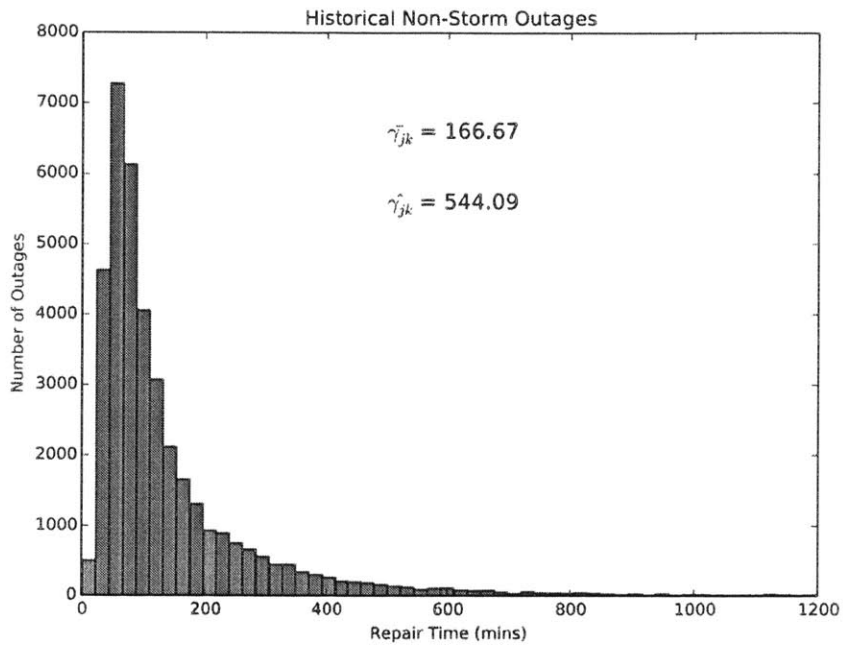


Figure 3.3: Historical Non-Storm Outage Histogram

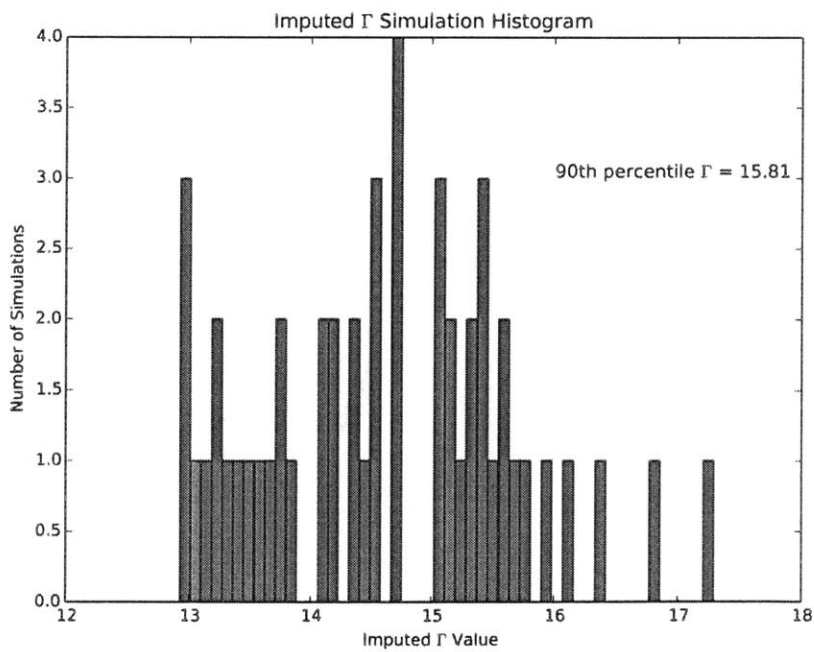


Figure 3.4: Simulated historical Non-Storm Outage Histogram

4 Estimating Damage From Calls

As a storm progresses, more people lose power in their homes and call the utility company to notify them. Typically, at least during the daytime, nearly every loss of power is reported, so the volume of calls is high. It is straightforward to see how this data can provide the company with very valuable live information on network damage. In this chapter we build a model and develop an algorithm which uses the localization of these calls to track storm damage on a much finer scale than what is possible using only weather forecasts. The goal is to guide emergency repair crews towards the areas where damage is the most likely to have occurred.

We first describe the framework and algorithm considered (in Sections 4.1 and 4.2 respectively). Then in Section 4.3, we show that the problem described above can indeed be cast as a special instance of this model and present a prototype implementation.

4.1 Model

In this section we describe the general framework that we will use to infer posterior probabilities from partial information on a graph.

4.1.1 Network

We consider a specific kind of graphs: *directed trees*. A directed tree is a directed graph that would be a tree if the edge directions were ignored, i.e. a graph in which any two vertices are connected by a unique undirected path. We denote by $\mathcal{T} = (\mathcal{V}, \mathcal{B})$ such a tree where \mathcal{V} is the set of vertices and \mathcal{B} is the set of branches (directed edges). Such networks

are known to be very amenable to message passing algorithms of which belief propagation is an instance (cf. for example [34, 33]).

Furthermore, we will assume that each vertex has at most one incoming branch (and therefore at most one parent). It is simple to show that this also implies that the tree \mathcal{T} has exactly one root. This assumption will allow us to further simplify belief propagation algorithms commonly found in the signal processing literature (e.g. the standard belief propagation algorithm introduced by Pearl and studied in [37, 42]).

For any vertex i , we denote by \mathcal{C}_i the set of its (direct) children, \mathcal{D}_i the set of its descendants (including i), \mathcal{A}_i the set of its ancestors, and if it exists, we denote the parent by p_i . Note that for any vertex i , $\mathcal{D}_i \cup \mathcal{A}_i = \mathcal{V}$. To illustrate these definitions, consider Figure 4.1. We have $\mathcal{C}_6 = \{9, 10\}$, $\mathcal{D}_6 = \{9, 10, 14, 15, 16\}$, $\mathcal{A}_6 = \{0, 2, 4, 6\}$, and $p_6 = 4$.

Finally, without loss of generality, we will assume that the root is indexed by $i = 0$ and we denote by \mathcal{V}^* the set of vertices that have a parent.

4.1.2 States

Each vertex $i \in \mathcal{V}$ of the graph is assigned a tuple $(x_i, y_i) \in \mathcal{X}_i \times \mathcal{Y}_i$ where:

- x_i is the state of the vertex and lies in a finite set \mathcal{X}_i
- y_i represents vertex specific characteristics and belongs to a (non-necessarily finite) set \mathcal{Y}_i

Note that the sets \mathcal{X}_i and \mathcal{Y}_i can be different across vertices: the network can consist of different ‘types’ of vertices which each have their own states and characteristics. For example in an electricity network, some vertices could correspond to power lines and others to residential customers. In practice, characteristics of a power line could then be the distance between supporting poles, the width of the cable, the wire framing whereas characteristics of a customer could be the number of residents, any past call history, etc.

In general the state of each vertex is unknown and is therefore represented by a random variable X_i . We assume that the underlying distribution of these states $\mathbb{P}(\{X_i\}_{i \in \mathcal{V}})$ can be factored according to the directed tree \mathcal{T} . In particular, this implies that $\mathbb{P}(\{X_i\}_{i \in \mathcal{V}})$

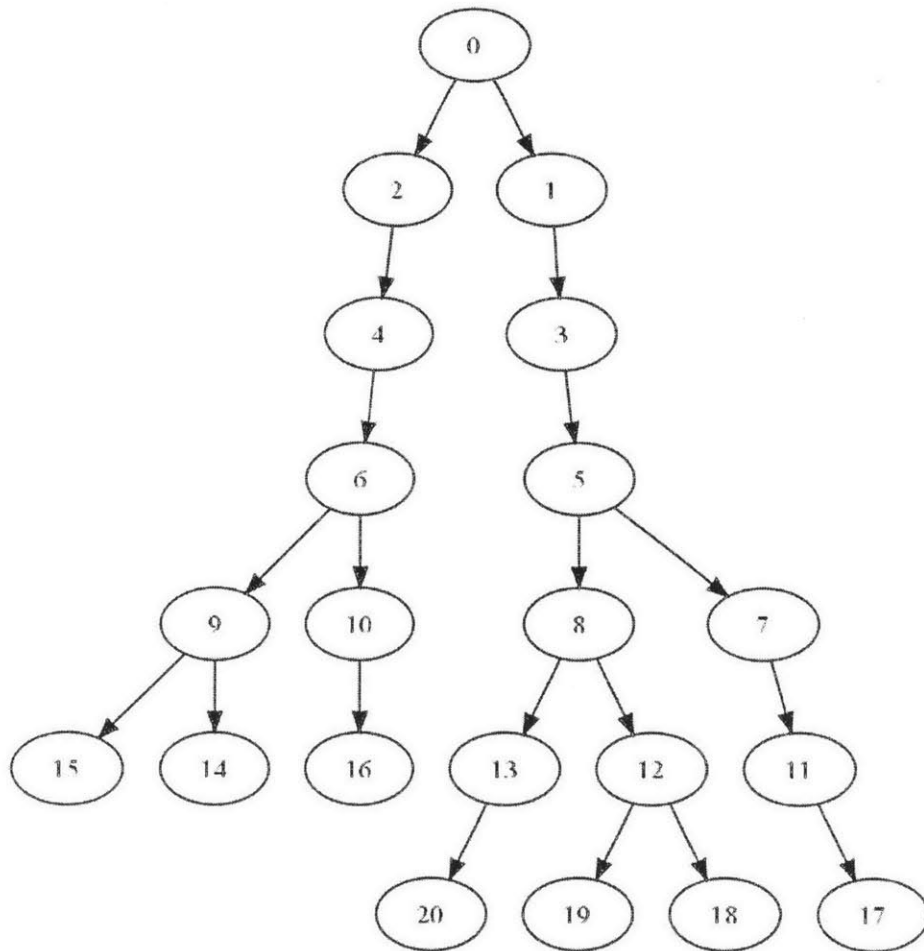


Figure 4.1: Sample directed tree

satisfies the Markov property:

$$\forall i \in \mathcal{V}^*, \forall x \in \mathcal{X}_i, \mathbb{P}(X_i = x) = \sum_{x' \in \mathcal{X}_{p_i}} \mathbb{P}(X_i = x | X_{p_i} = x') \mathbb{P}(X_{p_i} = x') \quad (4.1)$$

This equation implies that the state of any vertex only depends on the states of its ancestors. Moreover, it implies the important property that given the state of a vertex i , the state of any of its descendants is conditionally independent from the state of its ancestors. We can formulate this as follows:

$$\forall i \in \mathcal{V}, \forall j \in \mathcal{D}_i, \forall k \in \mathcal{A}_i, \forall x \in \mathcal{X}_k, \mathbb{P}(X_k = x | X_i, X_j) = \mathbb{P}(X_k = x | X_i) \quad (4.2)$$

4.1.3 Evidence

Finally, the state of some vertices might be known. We denote this evidence by a tuple $(\mathcal{E}, \mathbf{e})$ where \mathcal{E} is the set of vertices whose state is known and \mathbf{e} the vector of their corresponding states. For example, in the sample tree represented in Figure 4.1, we could have $\mathcal{E} = \{5, 10\}$ and $\mathbf{e} = (A, E)$ where A and E correspond to arbitrary states vertices 5 and 10 can respectively take.

Not all possible combinations of evidence have to be feasible. Given the form of conditional probabilities described in equations (4.2), it might be that the probability of witnessing a set of states indicated by a given evidence is zero. These correspond to combinations of state that cannot be realized by the network.

Conversely, we say that the evidence is *consistent* if it is a feasible state of the network, i.e.

$$\mathbb{P}(\forall i \in \mathcal{E}, X_i = e_i) > 0 \quad (4.3)$$

From now on we will assume that the evidence $(\mathcal{E}, \mathbf{e})$ is consistent. Moreover, in order to simplify notation, we will write $\mathbb{P}(\mathcal{E})$ as shorthand for $\mathbb{P}(\forall i \in \mathcal{E}, X_i = e_i)$.

4.2 Algorithm

Using the framework detailed in Section 4.1, we can rephrase the goal of this section as follows:

Assume we are given:

- a directed tree $\mathcal{T} = (\mathcal{V}, \mathcal{B})$
- transition probabilities
 - $\forall i \in \mathcal{V}^*, \forall x \in \mathcal{X}_i, \forall x' \in \mathcal{X}_{p_i}, \mathbb{P}(X_i = x | X_{p_i} = x')$
 - $\forall x \in \mathcal{X}_0, \mathbb{P}(X_0 = x)$
- consistent evidence $(\mathcal{E}, \mathbf{e})$ (possibly empty)

We would like to calculate:

$$\forall i \in \mathcal{V}, \forall x \in \mathcal{X}_i, \mathbb{P}(X_i = x | \mathcal{E}) \tag{4.4}$$

A naive combinatorial approach to calculate these conditional posterior probabilities would be to simply enumerate all state combinations and sum over all possible combinations. However the number of possibilities to consider grows exponentially with the number of vertices and quickly becomes impractical. For example, for a network consisting of 100 vertices with only 2 possible states per vertex, the number of possibilities one must consider is greater than 10^{30} . This would render such an approach inapplicable for our end use case: a typical circuit consists of several hundred segments and thousands of customers (each of which corresponds to a vertex).

This type of problem, inference on graphical models, has been studied extensively in the signal processing literature and has been commonly solved using ‘belief propagation’ algorithms (e.g. [11, 34, 33, 46, 37]). The first such algorithm was proposed by Judea Pearl in [39]. These are known to exact on trees and very efficient: their complexity is linear in the size of the graph (as opposed to the naive exponential complexity mentioned above). At a high level, belief propagation uses the graphical structure of the tree to compute the

desired marginals.

We therefore present a variant of a belief propagation algorithm specialized to the specific directed trees described in the previous section and show that it can indeed be used to efficiently solve (4.4):

Definition 4.2.1. *State inference algorithm*

Consider the following set of equations (where products over an empty set are considered equal to 1):

$$\forall i \in \mathcal{V}, \forall j \in \mathcal{C}_i, \forall x \in \mathcal{X}_i, m_{i \leftarrow j}(x) = \sum_{x' \in \mathcal{X}_j} d_j(x') \mathbb{P}(X_j = x' | X_i = x) \quad (4.5a)$$

$$m_{i \rightarrow j}(x) = \begin{cases} \delta_{\{x=e_i\}} a_i(x) \prod_{j' \in \mathcal{C}_i, j' \neq j} m_{i \leftarrow j'}(x), & \text{if } i \in \mathcal{E} \\ a_i(x) \prod_{j' \in \mathcal{C}_i, j' \neq j} m_{i \leftarrow j'}(x), & \text{ow.} \end{cases} \quad (4.5b)$$

$$a_i(x) = \begin{cases} \mathbb{P}(X_i = x), & \text{if } i = 0 \\ \sum_{x' \in \mathcal{X}_{p_i}} \mathbb{P}(X_i = x | X_{p_i} = x') m_{p_i \rightarrow i}(x'), & \text{ow.} \end{cases} \quad (4.5c)$$

$$d_i(x) = \begin{cases} \delta_{\{x=e_i\}}, & \text{if } i \in \mathcal{E} \\ \prod_{j \in \mathcal{C}_i} m_{i \leftarrow j}(x), & \text{ow.} \end{cases} \quad (4.5d)$$

The algorithm proceeds in two phases, repeated sequentially:

Bubbling phase

Starting from the leaves of the tree, the algorithm travels towards the root, computing d_i and $m_{p_i \leftarrow i}$ for each node $i \in \mathcal{V}^*$ visited

Capturing phase

Once the root is reached, it then travels back towards the leaves and computes $m_{i \rightarrow j}$ and a_i for each node $i \in \mathcal{V}$

In the next section we show how this algorithm relates to solving equation (4.4).

4.2.1 Results

Given the formulation of Definition 4.2.1, it is non-obvious how we initialize the algorithm. Indeed, not all vertices initially have values assigned to the quantities defined in equations (4.5). An important other consideration is the convergence of these quantities. A priori, the algorithm could keep updating values in the tree without ever terminating. The first step is therefore to show that the algorithm can in fact be applied, and moreover, that it converges in linear time:

Theorem 4.2.2. *The state inference algorithm (Definition 4.2.1) is well defined (i.e. the quantities defined in equations (4.5) can be computed by induction), and converges after the first capturing phase.*

Intuitively, equations (4.5) provide just enough initial information for the bubbling phase to run (a capturing phase wouldn't be able to update its values at this time). Once this first phase is done, the 'downstream' messages have all been assigned and now allow for the capturing phase to happen. The termination of the algorithm after two phases comes from the fact that neither of these messages change after being assigned once.

This algorithm is very fast: each vertex is visited at most twice (the root is visited only once), and only requires computing two messages per (undirected) edge along with the quantities a and d (each of which must be computed for each possible state of the vertex). From our assumptions on the shape of the graph, there are one less total edges than vertices, therefore the total number of computations required is of the order $2 \sum_{i \in \mathcal{V}} |\mathcal{X}_i|$, where $|\mathcal{X}_i|$ stands for the cardinality of set \mathcal{X}_i . In particular, if all vertices' states take values in a same set \mathcal{X} then the previous quantity simply becomes $|\mathcal{V}||\mathcal{X}|$.

We now show how this algorithm can be used to compute the marginals we are interested in:

Theorem 4.2.3. *Once the state inference algorithm (Definition 4.2.1) has converged, we are able to compute the desired posterior state probabilities using the following equation:*

$\forall i \in \mathcal{V}, \forall x \in \mathcal{X}_i,$

$$\mathbb{P}(X_i = x | \mathcal{E}) = \frac{d_i(x)a_i(x)}{\sum_{x' \in \mathcal{X}_i} d_i(x')a_i(x')} \quad (4.6)$$

It is interesting to note that it is possible to optimize the previous algorithm to allow for incremental arrival of evidence. Indeed, one can avoid visiting some vertices twice by observing that upstream messages from descendant of vertices where evidence has just arrived will not change. This allows for even faster partial updating of the marginals in practical ‘real-time’ applications such as the one we are interested in. This is illustrated in Figure 4.2, which illustrates how information propagates from a vertex which has just received new evidence.

4.2.2 Proofs

We now give the proofs omitted in the previous subsection.

Proof of Theorem 4.2.2. Initially, we are given the values of d_i for all nodes with evidence ($d_i(x) = \delta_{\{x=e_i\}}$) and all leaf nodes ($d_i(x) = 1$). We are also able to compute their outgoing upstream messages $m_{p_i \leftarrow i}$, and recursively repeat this procedure until we reach the root of the tree (the bubbling phase). Once, we have reached the root, all the upstream messages have been computed and we have all the information required to move back towards the leaves computing the downstream messages $m_{p_i \rightarrow i}$ and a_i as we go (the capturing phase).

To prove the second part of the theorem, notice that all the messages computed during the initial bubbling phase only depend on the evidence and transition probabilities, therefore they never change from their initial value. Once the root is reached, we have enough information to compute all other quantities by doing a breadth first exploration of the tree. Since the upstream messages never change, once a node is visited in this capturing phase, none of its incoming messages will change, and the algorithm has converged. \square

The proof of Theorem 4.2.2 relies on 2 lemmas (Lemma 4.2.4 and Lemma 4.2.5), which we first state and prove. These also yield some insight on what information the messages and

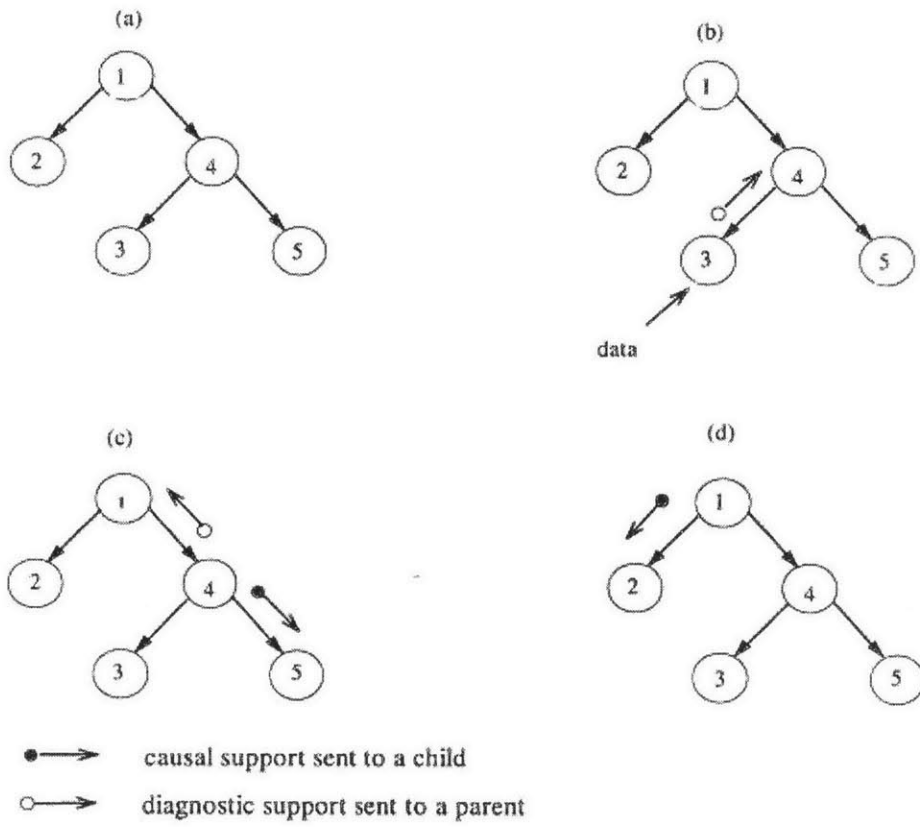


Figure 4.2: Illustration of partial updates from new evidence ('data')

quantities from equations (4.5) contain.

Lemma 4.2.4.

$$\forall i \in \mathcal{V}, \forall j \in \mathcal{C}_i, \forall x \in \mathcal{X}_i, m_{i \leftarrow j}(x) = \mathbb{P}(\mathcal{E} \cap \mathcal{D}_j | X_i = x) \quad (4.7a)$$

$$\forall i, \forall x \in \mathcal{X}_i, d_i(x) = \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x) \quad (4.7b)$$

Proof. By induction, starting from the leaves.

Initialization i leaf

- (4.7a) is true because i has no children.
- By construction: if $i \in \mathcal{E}$, $d_i(x) = \delta\{x = e_i\} = \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x)$, else $d_i(x) = 1 = \mathbb{P}(\emptyset | X_i = x) = \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x)$. Therefore in all cases, (4.7b) is true.

Recursion $i \Rightarrow p_i$

- (4.7a)

$$\begin{aligned} \forall x \in \mathcal{X}_i, m_{i \leftarrow j}(x) &= \sum_{x' \in \mathcal{X}_j} d_j(x') \mathbb{P}(X_j = x' | X_i = x) \\ &= \sum_{x' \in \mathcal{X}_j} \mathbb{P}(\mathcal{E} \cap \mathcal{D}_j | X_j = x') \mathbb{P}(X_j = x' | X_i = x) \quad \text{HR} \\ &= \sum_{x' \in \mathcal{X}_j} \mathbb{P}(\mathcal{E} \cap \mathcal{D}_j, X_j = x' | X_i = x) \\ &= \mathbb{P}(\mathcal{E} \cap \mathcal{D}_j | X_i = x) \end{aligned}$$

- If $i \in \mathcal{E}$, $d_i(x) = \delta_{\{x=e_i\}} = \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x)$ and the (4.7b) is true. Otherwise,

$$\begin{aligned}
\forall x \in \mathcal{X}_i, d_i(x) &= \prod_{j \in \mathcal{C}_i} m_{i \leftarrow j}(x) \\
&= \prod_{j \in \mathcal{C}_i} \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x) \\
&= \mathbb{P}\left(\mathcal{E} \cap \bigcup_{j \in \mathcal{C}_i} \mathcal{D}_j | X_i = x\right) \\
&= \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x) \text{ since } i \notin \mathcal{E}
\end{aligned}$$

□

Lemma 4.2.5.

$$\forall i \in \mathcal{V}^*, \forall x \in \mathcal{X}_{p_i}, m_{p_i \rightarrow i}(x) = g_i(\mathcal{E}) \mathbb{P}(X_{p_i} = x | \mathcal{E} \cap \mathcal{D}_i^c) \quad (4.8a)$$

$$\forall i \in \mathcal{V}, \forall x \in \mathcal{X}_i, a_i(x) = g_i(\mathcal{E}) \mathbb{P}(X_i = x | \mathcal{E} \cap \mathcal{D}_i^c) \quad (4.8b)$$

where $g_i(\mathcal{E}) = \prod_{j \in \mathcal{A}_i} \frac{1}{f_j(\mathcal{E})}$ is independent of x and positive.

Proof. By induction, starting from the root.

Initialization $i = 0$

- (4.8a) is true since i has no parent
- $a_i(x) = \mathbb{P}(X_i = x) = \mathbb{P}(X_i = x | \emptyset) = \mathbb{P}(X_i = x | \mathcal{E} \cap \mathcal{D}_i^c)$ which proves (4.8b)

Recursion $p_i \Rightarrow i$

- Let's start with (4.8a). If $i \notin \mathcal{D}$:

$$\begin{aligned}
\forall x \in \mathcal{X}_{p_i}, m_{p_i \rightarrow i}(x) &= a_{p_i}(x) \prod_{j \in \mathcal{C}_{p_i}, j \neq i} m_{p_i \leftarrow j}(x) \\
&= g_{p_i}(x) \mathbb{P}(X_{p_i} = x | \mathcal{E} \cap \mathcal{D}_{p_i}^c) \prod_{j \in \mathcal{C}_{p_i}, j \neq i} m_{p_i \leftarrow j}(x) \\
&= g_{p_i}(x) \mathbb{P}(X_{p_i} = x | \mathcal{E} \cap \mathcal{D}_{p_i}^c) \prod_{j \in \mathcal{C}_{p_i}, j \neq i} \mathbb{P}(\mathcal{E} \cap \mathcal{D}_j | X_{p_i} = x) \\
&= g_{p_i}(x) \mathbb{P}(X_{p_i} = x | \mathcal{E} \cap \mathcal{D}_{p_i}^c) \mathbb{P}\left(\mathcal{E} \cap \bigcup_{j \in \mathcal{C}_{p_i}, j \neq i} \mathcal{D}_j | X_{p_i} = x\right) \\
&= g_{p_i}(x) \mathbb{P}(X_{p_i} = x | (\mathcal{E} \cap \mathcal{D}_i^c) \cap \mathcal{D}_{p_i}^c) \mathbb{P}((\mathcal{E} \cap \mathcal{D}_i^c) \cap \mathcal{D}_{p_i} | X_{p_i} = x) \\
&= \frac{g_{p_i}(x)}{f_i(\mathcal{E})} \mathbb{P}(X_{p_i} = x | \mathcal{E} \cap \mathcal{D}_i^c) \\
&= g_i(\mathcal{E}) \mathbb{P}(X_{p_i} = x | \mathcal{E} \cap \mathcal{D}_i^c)
\end{aligned}$$

Otherwise, suffices to notice that: $\delta_{\{x=e_i\}} = \mathbb{P}(\mathcal{E} \cap \{p_i\} | X_{p_i} = x)$ and the previous proof holds.

- We are now ready to prove (4.8b)

$$\begin{aligned}
\forall x \in \mathcal{X}_i, a_i(x) &= \sum_{x' \in \mathcal{X}_{p_i}} \mathbb{P}(X_i = x | X_{p_i} = x') m_{p_i \rightarrow i}(x') \\
&= \sum_{x' \in \mathcal{X}_{p_i}} \mathbb{P}(X_i = x | X_{p_i} = x') g_i(\mathcal{E}) \mathbb{P}(X_{p_i} = x' | \mathcal{E} \cap \mathcal{D}_i^c) \\
&= g_i(\mathcal{E}) \sum_{x' \in \mathcal{X}_{p_i}} \mathbb{P}(X_i = x | X_{p_i} = x') \mathbb{P}(X_{p_i} = x' | \mathcal{E} \cap \mathcal{D}_i^c) \\
&= g_i(\mathcal{E}) \mathbb{P}(X_i = x | \mathcal{E} \cap \mathcal{D}_i^c)
\end{aligned}$$

□

We now tie the two previous lemmas together and prove Theorem 4.2.3:

Proof of Theorem 4.2.3. We start by showing that we can decompose the posterior proba-

bilities as follows:

$$\forall i \in \mathcal{V}, \forall x \in \mathcal{X}_i, \mathbb{P}(X_i = x | \mathcal{E}) = f_i(\mathcal{E}) \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x) \mathbb{P}(X_i = x | \mathcal{E} \cap \mathcal{D}_i^c) \quad (4.9)$$

where $f_i(\mathcal{E}) = \frac{\mathbb{P}(\mathcal{E} \cap \mathcal{D}_i^c)}{\mathbb{P}(\mathcal{E})}$ is independent of x and positive.

Indeed, notice that: $\forall i \in \mathcal{V}, \forall x \in \mathcal{X}_i,$

$$\begin{aligned} \mathbb{P}(X_i = x | \mathcal{E}) &= \mathbb{P}(X_i = x | \mathcal{E} \cap \mathcal{D}_i, \mathcal{E} \cap \mathcal{D}_i^c) \\ &= \frac{\mathbb{P}(X_i = x, \mathcal{E} \cap \mathcal{D}_i | \mathcal{E} \cap \mathcal{D}_i^c)}{\mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | \mathcal{E} \cap \mathcal{D}_i^c)} \\ &= \frac{\mathbb{P}(\mathcal{E} \cap \mathcal{D}_i^c)}{\mathbb{P}(\mathcal{E})} \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x, \mathcal{E} \cap \mathcal{D}_i^c) \mathbb{P}(X_i = x | \mathcal{E} \cap \mathcal{D}_i^c) \end{aligned}$$

Let $f_i(\mathcal{E}) = \frac{\mathbb{P}(\mathcal{E} \cap \mathcal{D}_i^c)}{\mathbb{P}(\mathcal{E})}$, we can now write: $\forall i \in \mathcal{V}, \forall x \in \mathcal{X}_i,$

$$\begin{aligned} \mathbb{P}(X_i = x | \mathcal{E}) &= f_i(\mathcal{E}) \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x, \mathcal{E} \cap \mathcal{D}_i^c) \mathbb{P}(X_i = x | \mathcal{E} \cap \mathcal{D}_i^c) \\ &= f_i(\mathcal{E}) \mathbb{P}(\mathcal{E} \cap \mathcal{D}_i | X_i = x) \mathbb{P}(X_i = x | \mathcal{E} \cap \mathcal{D}_i^c) \end{aligned}$$

The last part of the previous equation is a consequence of equation (4.2). Using the results from lemmas 4.2.4 and 4.2.5, we now get: $\forall i \in \mathcal{V}, \forall x \in \mathcal{X}_i,$

$$\begin{aligned} \mathbb{P}(X_i = x | \mathcal{E}) &= \frac{f_i(\mathcal{E})}{g_i(\mathcal{E})} d_i(x) a_i(x) \\ &= h_i(\mathcal{E}) d_i(x) a_i(x) \end{aligned}$$

Since $f_i(\mathcal{E})$ and $g_i(\mathcal{E})$ are positive, $h_i(\mathcal{E})$ is well defined. We get the desired result by noticing that $h_i(\mathcal{E})$ doesn't depend on x . □

4.3 Application to the Power Utility

The power utility company we worked with has a tool which tries to infer which device has opened on a circuit, given customer calls reporting loss of power the company has received. It is however relatively basic.

The device is chosen in the following way: it is the single most downstream device which can explain all the power losses on this circuit. During ‘blue sky’ days, this prediction is relatively accurate as most often only one device actually opens and crews can rapidly check the area to find the damaged segments. It can also be noted that even on non-storm days calls often come in batches: a single tree branch falling on an electrical wire can cause hundreds of people to lose power simultaneously. This doesn’t alter the quality of the prediction, as long as there is a single point of failure.

When damage occurs on the circuit at multiple points, which is almost certainly the case during every major event, this tool yields very imprecise results. Indeed, when several devices open on distant parts of a circuit (caused by separate damage events), it will try to explain all the resulting power losses with a single open device, which will typically be much more upstream than the actual open devices. Our framework takes this fact into account and allows for multiple open devices.

Moreover, the current tool doesn’t provide any information on where the damaging event is likely to be. At best it gives the location of the next upstream open device. In some cases, the electrical company might have received some information from customers about such damages (e.g. a customer saw a pole break next to his house), but in general, this can lead to repair crews wasting a lot of time trying to find the source of the open device. This is especially important when they are faced with difficult transport conditions.

In this section, we show how we can apply the framework from the previous sections to the problem of inferring damage locations on the electrical network from customer calls, allowing for multiple points of failure and probability estimates at each damage point.

4.3.1 The Graph

Recall that in this work, we are interested in the electricity network at the circuit level. Each circuit can be viewed effectively as a tree with a generating substation as root. There might be several circuits starting from the same substation but there are typically few possibilities of interconnections between each other. For the purpose of this work, we will assume that no such interconnections are realized and each circuit can be modelled independently as a

tree.

Since our algorithm also requires information gained from customers, it is straightforward to see that these must also be represented in the graph, in addition to the ‘electrical’ portion of the network. More formally, the set of all vertices can be divided into two subsets:

$$\mathcal{V} = \underbrace{\{assets\}}_{\mathcal{V}_a} \cup \underbrace{\{customers\}}_{\mathcal{V}_c} \quad (4.10)$$

The first subgraph, consisting of the vertices in \mathcal{V}_a , is built similarly to Chapter 2 on the ‘segment model’ used by the electricity company. Each circuit is represented by a graph: each vertex in the graph representing an asset, and each branch an existing path between two assets (an asset consists in a single device and all the downstream segments that directly trigger it). Note another difference with the first chapter of this part where we were ignoring any connectivity between assets and focused on each asset individually. By construction, this subgraph satisfies the assumptions of the directed tree stated in Section 4.1. Namely, each vertex has a unique parent and there exists a single unique undirected path between any two nodes in the graph.

The segments belonging to each asset provide power to multiple customers, all of whom will lose power if the asset’s device opens (along with all the customers of any assets further downstream). When such an event happens, customers will call the electrical company with some probability to inform them of the state of their neighboring device. This is represented in the graph by a customer vertex with an incoming branch from its corresponding asset node. An important consequence of this structure is that customers will only be found on leaves of the tree, thus there are exactly two types of branches:

$$\mathcal{B} = \underbrace{\{asset \rightarrow asset\}}_{\mathcal{B}_a} \cup \underbrace{\{asset \rightarrow customer\}}_{\mathcal{B}_c} \quad (4.11)$$

From equation (4.11), it is easy to see that the combined graph of assets and customers also satisfies the assumptions from Section 4.1.

To summarize, we have shown that we can model each circuit as a directed tree $\mathcal{T} = (\mathcal{V}, \mathcal{B})$

where vertices represent either assets or customers. To give an idea of the size of these graphs in practice: the number of assets encountered in a circuit is typically on the order of 50, but can go up to several thousands. The distribution of the total number of assets and customers per circuit are represented in figures 4.3 and 4.4 respectively.

4.3.2 States and Transition Probabilities

Now that we have constructed the graph representing devices and customers across each circuit, the next step is to determine the set of possible states for each vertex, i.e. $X_i, \forall i \in \mathcal{V}$. We consider the two subgraphs \mathcal{V}_a and \mathcal{V}_c separately and define the following states:

- Asset states. $\forall i \in \mathcal{V}_a$, X_i can take the following values:

$$X_i = 0 \text{ OK}$$

$$X_i = 1 \text{ No power}$$

$$X_i = 2 \text{ Damaged}$$

- Customer states. $\forall i \in \mathcal{V}_c$, X_i can take the following values:

$$X_i = 0 \text{ No call}$$

$$X_i = 1 \text{ Call}$$

Note that even though we are not trying to predict the probability that an asset vertex has lost power (we are interested in prediction the likelihood of damage occurring at that asset, that is of the device being open), we incorporate a *no power* state. Indeed, without this third state, we would be unable to factor the distribution according to equation (4.1). An asset downstream from a damaged asset might not be damaged but still have its customers call because the power is out.

Furthermore, this formulation allows the electrical company to take into account information gained from dispatching crews. When a crew drives to a particular street from the depot, they are able to observe the state of nearby devices and power lines (be it intact, damaged, or out of power), thereby giving information which can readily be fed back into the model.

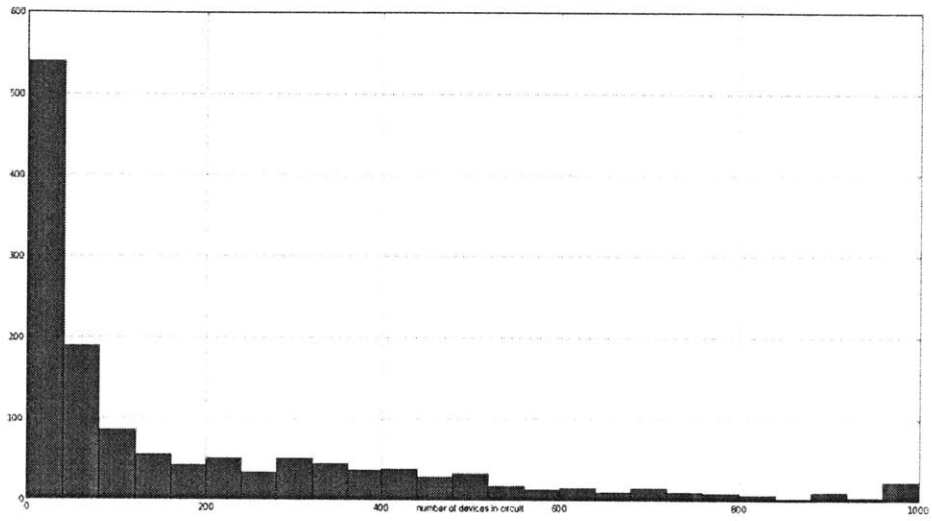


Figure 4.3: Histogram of the number of devices per circuit.

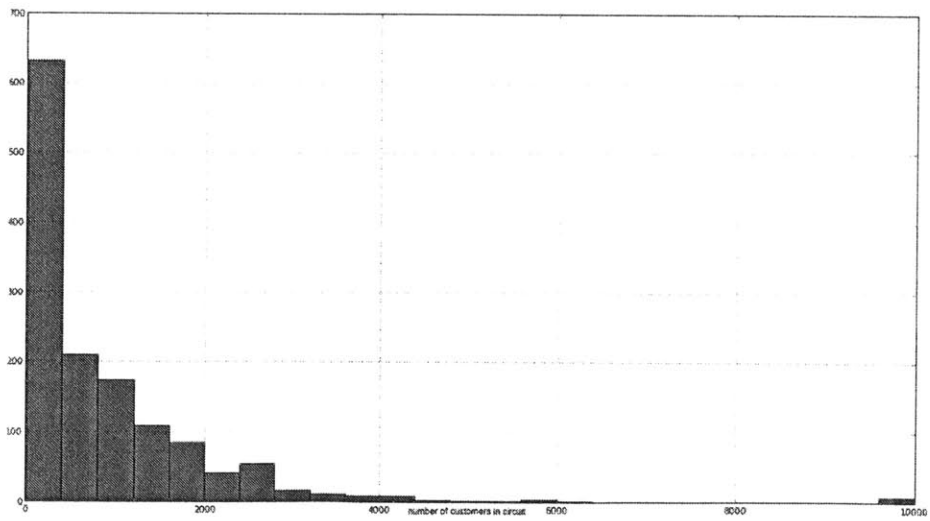


Figure 4.4: Histogram of the number of customers per circuit.

Given our earlier assumption that no reconnections happen on the tree, we can use an important property of power distribution circuits: if power is lost at a given asset, power is also lost at all the assets downstream from this asset. This implies in particular that we can model the physical reality of the electrical network while also satisfying equation (4.1) by considering the following transition matrices $\{\mathbf{P}_i\}_i$ among the vertices of subgraph \mathcal{V}_a :

$$\forall i \in \mathcal{V}_a, \mathbf{P}_i = \begin{pmatrix} 1 - \pi_i & \pi_i & 0 \\ 0 & \pi_i & 1 - \pi_i \\ 0 & \pi_i & 1 - \pi_i \end{pmatrix} \quad (4.12)$$

π_i is a number between 0 and 1 that represents the probability that the asset corresponding to vertex i is damaged and the transition matrix \mathbf{P}_i is defined as the matrix such that:

$$\forall x_i, x_{p_i} \in \mathcal{X}_a, \mathbf{P}_{x_{p_i}, x_i} = \mathbb{P}(X_i = x_i | X_{p_i} = x_{p_i}) \quad (4.13)$$

Let us describe case by case the transition matrix described in equation (4.12):

- If the parent p_i of asset i has power, by definition of π_i , with probability π_i asset i will incur some damage and the corresponding device will open. Otherwise, asset i behaves as normal.
- If the parent p_i of asset i has no power, asset i will never have any power. However, with probability π_i it might also be damaged.
- Likewise, if the parent p_i of asset i has been damaged, asset i will never have any power, and with probability π_i might also be damaged.

π_i is computed using weather data and the features computed in Chapter 2. The computation is essentially the same as in the first chapter, but the weather features will now be more precise as this model will be run while the storm is happening. Therefore the weather data won't suffer for forecasting errors. Moreover, the fact that this model is run in real time makes it possible to use weather logs instead of forecasts, which as was discussed in Chapter 2 contain much more information (e.g. rainfall rate, wind gust speeds, pressure, etc.) to allow for significantly higher prediction accuracy.

Similarly to the asset vertices, we can also define a transition matrix for the vertices corresponding to customers:

$$\forall i \in \mathcal{V}_c, \mathbf{P}_i = \begin{pmatrix} 1 & 0 \\ 1 - \pi'_i & \pi'_i \\ 1 - \pi'_i & \pi'_i \end{pmatrix} \quad (4.14)$$

π'_i now represents the probability that customer i calls given that there is no power in the corresponding asset. This probability can be estimated from previous storm call statistics. In this work, we did not pursue this direction and used a reasonable number for our tests, and leave precise estimation for future research.

Note that equation (4.14) implies in particular that there are no ‘false alerts’, i.e. there are no customers who wrongly notify the company of a loss of power. This enables us, without loss of generality, to group all direct children customer vertices from an asset i into one ‘super vertex’ with call probability $\pi'' = 1 - \prod_{j \in \mathcal{C}_i} (1 - \pi'_j)$. This transformation allows us to significantly reduce the number of vertices in the graph (as can be seen from Figure 4.4) and speed up the algorithm. However, this change also makes the implicit assumption that there are no internal failures inside an asset that leave the device closed. For example, we do not allow a house to lose power on its own, whereas in reality this is sometimes the case (the end line connecting a house to the electrical network could be severed without affecting other houses).

Finally, recall that only assets are represented in the graph. In order to recover granularity at the segment level we once again use results from Chapter 2. This time we take the individual segment features and combine them with the weather data to generate individual segment failure probabilities. Since we assume that failures happen independently across the graph, we are able to recover segment failure likelihoods. More simply, we are able to rank, inside each asset, the segments which are the most vulnerable and, from there, guide repair crews to a target damage location.

4.3.3 Implementation

In order to simulate how the company might run this model, the algorithm was implemented in JavaScript (compiled from CoffeeScript) and can be run directly from any HTML5 compatible browser (Google Chrome, Mozilla Firefox, and recent versions of Internet Explorer). The tool includes visualizations of the graph using d3.js, supporting both geographical coordinates of each asset and customer or a generated force-repulsion layout. The figures below (4.5, 4.6, 4.7, and 4.8) present screenshots of this interface in different states. This interface also allows information to be directly inputted into the model by toggling the state of each vertex in real time.

Figure 4.5 presents the way a particular circuit is displayed when no customer has notified the company of a power loss. The vertices are placed according to the geographical location of their corresponding entity (each asset vertex is centered on its device). Blue vertices represent assets while green vertices represent customers, and the generation substation is connected to the circuit via the vertex closest to the top right corner of the graph.

The next figure, Figure 4.6, shows the updated state of the circuit after a single called has arrived: the corresponding customer vertex has turned from green to orange. At the same time, the two closest asset vertices have turned red and black respectively. Red indicates that damage is very likely at this vertex, whereas black indicates that with high likelihood the asset is out of power but otherwise intact.

In Figure 4.7, many customers have called reporting power outages and as a result there are many more assets likely out of power (the black vertices). Notice that the algorithm flags three vertices as probable points of failure (in red), such as the selected one where the computed a posteriori failure probability is 59.4%. Figure 4.8 shows the graph in the same state but with vertices sized by failure probability: the three previous vertices are now easily spotted.

The algorithm is very fast. On a circuit such as this one (with 145 asset and 65 customer vertices) updating the marginals is virtually instant. Even for larger circuits, such as those mentioned above, computations take less than a second, which makes this algorithm very practical for real time updates. Note that if inconsistent evidence is entered (such as setting

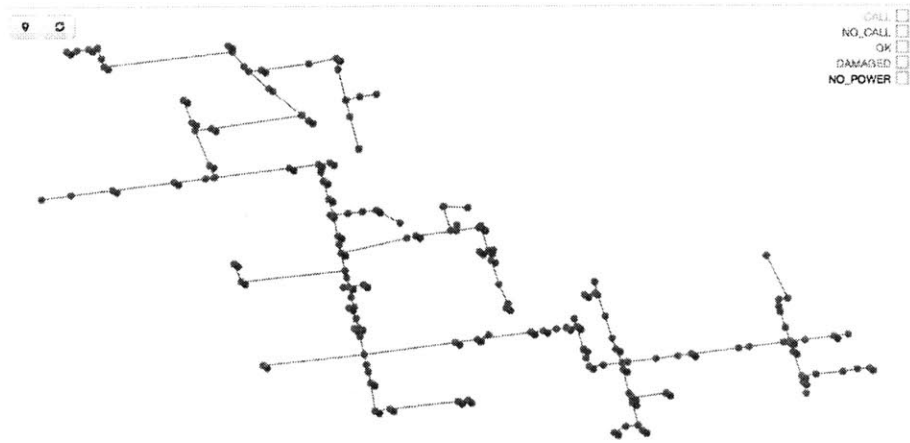


Figure 4.5: In-browser real-time damage estimation tool. View when no customer has lost power on the circuit.

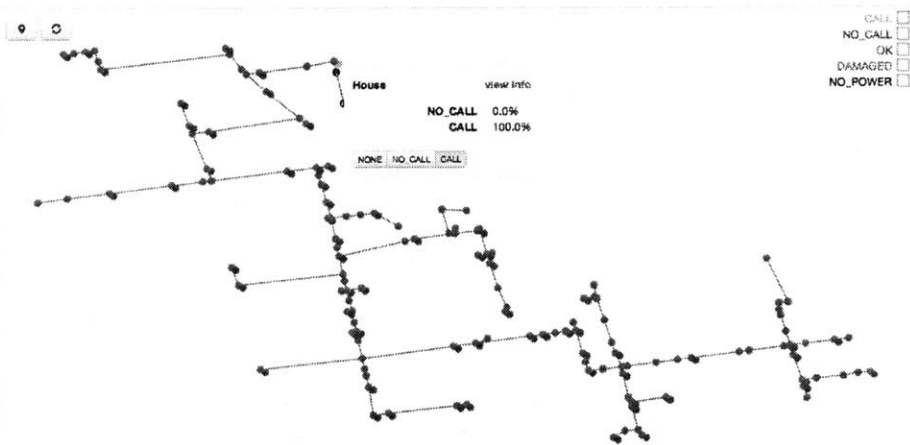


Figure 4.6: In-browser real-time damage estimation tool. Updating with a new customer call.

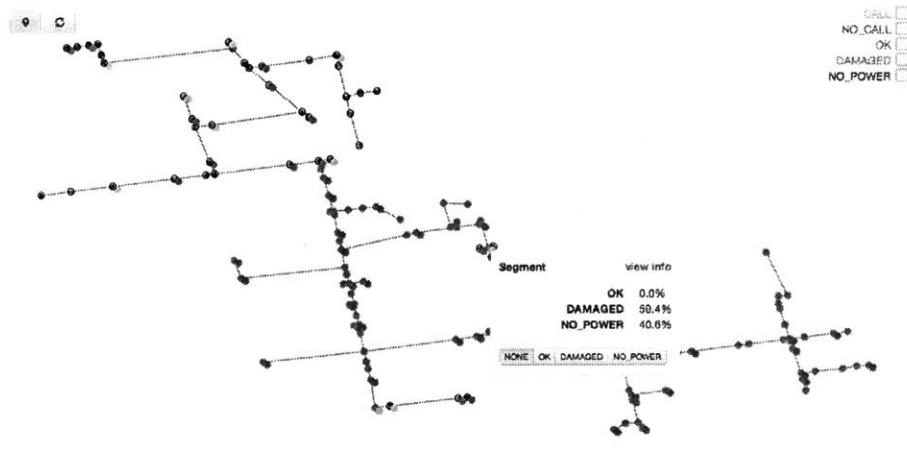


Figure 4.7: In-browser real-time damage estimation tool. View after having notified by several customers of power loss. Note that there are several damage points noted as likely, such as the one selected.

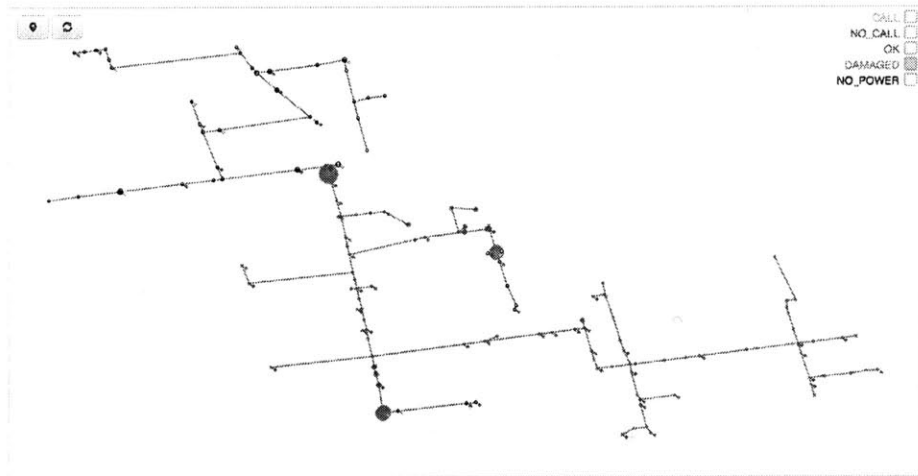


Figure 4.8: In-browser real-time damage estimation tool. Same situation as above, with graph vertices sized by the probability of the corresponding segment being damaged.

the state of an asset vertex to normal downstream of a damaged asset), the update will fail (the algorithm yields a division by zero error), and the current implementation will return to its state right before this update.

5 Conclusions and Future Work

The way power utility companies handle responses to storm events is directly linked with their ability to provide quality energy services to customers and to minimize costs during these storm responses. There is a unique opportunity here to significantly improve their process of getting ready for such major events, for example by learning ahead of time where the electricity network is most likely to see damage and what the best method of preparing repairs for that damage is.

The frameworks described in this part of the thesis could therefore add a lot value to such a company's emergency response system, especially during major storms. The tool shown in Chapter 2 can help the utility scale their response (e.g. how many out of state crews to bring in), while the second tool presented in Chapter 3 will guide initial crew placement among areas likely to be hit. During major events, the damage estimation algorithm implemented in Chapter 4 would yield much more precise estimated damage locations than the regular tool currently used.

It is important to note that as the quality of the data available grows, the precision of our models will continuously improve. Indeed, the models presented here are driven by weather forecasts and historical outage data. As a result, both the outage prediction and crew optimization models are limited by the errors inherent in these data sets. The better a company is able to quantify how it handles power restoration (e.g. repair times, restoration times, crew assignments, etc.), the better the proposed framework will work. A richer data set would also allow for more complex constraints in the optimization model and is therefore critical to a more realistic representation of actual crew dispatching (for example, one which would incorporate travel times between platforms).

Finally, there are several directions of interest for the future in order to improve the algorithms presented. With more granular outage data, the damage prediction model could function at the segment level instead of at the asset level. The interactions between land-cover and physical properties of the network could also be explored further. Even though the formulation is built to be robust, the crew optimization model relies on having reasonable estimates of repair times for which some deeper statistical analysis would be beneficial. The circuit damage estimation algorithm could also be extended to allow for circuit reconnections: in some cases the electricity company is able to restore power to certain assets that previously had lost power by isolating parts of the circuit and closing emergency switches. This could be done by allowing the model to capture dynamic changes to the structure of the tree.

Ultimately we believe that the insights gained from such tools will allow power utilities to provide better service to their customers (particularly in an environment where regulator demands are increasing) at a lower cost.

Concluding Remarks

In recent years, a lot of research has been focused on processing large quantities of data. However, in practice not all the data required to solve most problems is readily available. In this thesis, through two very different applications, we have presented an approach which utilizes structural properties of the problem considered to reduce data requirements and make up for missing data.

Our first example studied a dynamic pricing problem. We demonstrated policies that allows firms to compete asymptotically optimally with very low information requirements. The insight here is that by appropriately choosing summary statistics, we are able to capture most of the impact of competition. Policies similar to the ones we presented could be applied in practice by firms practising Revenue Management in highly competitive markets (e.g. in the airline or hotel industries).

In the second part of this thesis we focused on reducing power outage durations caused by major storms. In spite of missing historical data and the complexity of weather related damage, we are able to accurately pinpoint areas where major repairs are likely to be required. We built a practical tool around our model that a power utility can run to prepare itself ahead of a storm and reduce the total time until power is restored to all customers. Indeed, our predictions can be used not only to station crews preventively ahead of a storm but also to guide them during the event.

Bibliography

- [1] S Adlakha and R Johari. Mean field equilibrium in dynamic games with complementarities. 2010.
- [2] S Adlakha, R Johari, and G Weintraub. Oblivious equilibrium for large-scale stochastic games with unbounded costs. *Proceedings of the IEEE Conference on Decision and ...*, Jan 2008.
- [3] R Amir. Complementarity and diagonal dominance in discounted stochastic games. *Annals of Operations Research*, Jan 2002.
- [4] Nagaraj Balijepalli, Subrahmanyam S Venkata, Charles W Richter Jr, Richard D Christie, and Vito J Longo. Distribution system reliability assessment due to lightning storms. *Power Delivery, IEEE Transactions on*, 20(3):2153–2159, 2005.
- [5] Siddharth Balwani. Operational Efficiency through Resource Planning Optimization and Work Process Improvement. Master’s thesis, Massachusetts Institute of Technology, 2012.
- [6] Susan M Bernard, Jonathan M Samet, Anne Grambsch, Kristie L Ebi, and Isabelle Romieu. The potential impacts of climate variability and change on air pollution-related health effects in the united states. *Environmental Health Perspectives*, 109(Suppl 2):199, 2001.
- [7] Dimitris Bertsimas and Melvyn Sim. The Price of Robustness. *Operations Research*, 52(1):35–53, 2004.
- [8] R Billinton and JR Acharya. Weather-based distribution system reliability evaluation. *IEE Proceedings-Generation, Transmission and Distribution*, 153(5):499–506, 2006.
- [9] R Billinton and G Singh. Application of adverse and extreme adverse weather: modelling in transmission and distribution system reliability evaluation. In *Generation, Transmission and Distribution, IEE Proceedings-*, volume 153, pages 115–120. IET, 2006.
- [10] Roy Billinton and Janak Acharya. Distribution system reliability assessment incorporating weather effects. In *Universities Power Engineering Conference, 2006. UPEC’06. Proceedings of the 41st International*, volume 1, pages 282–286. IEEE, 2006.
- [11] A. Braunstein, M. Mezard, and R. Zecchina. Survey propagation: An algorithm for satisfiability. *Random Structures and Algorithms*, 2005.

- [12] Elin Broström and Lennart Söder. Modelling of ice storms for power transmission reliability calculations. In *in Proc. 15th Power Systems Computation Conference PSCC2005, Liege*, 2005.
- [13] Heidemarie C Caswell, Vincent J Forte, John C Fraser, Anil Pahwa, Tom Short, Mark Thatcher, and Val G Werner. Weather normalization of reliability indices. *Power Delivery, IEEE Transactions on*, 26(2):1273–1279, 2011.
- [14] Oben Ceryan, Ozge Sahin, and Izak Duenyas. Dynamic pricing of substitutable products in the presence of capacity flexibility. *MSOM*, 2013.
- [15] Chad Shouquan Cheng, Heather Auld, Guilong Li, Joan Klaassen, Bryan Tugwood, and Qian Li. An automated synoptic typing procedure to predict freezing rain: An application to ottawa, ontario, canada. *American Meteorological Society*, 19(4):751–768, April 2004.
- [16] L Curtat. Markov equilibria of stochastic games with complementarities. *Games and Economic Behavior*, Jan 1996.
- [17] A Domijan Jr, A Islam, WS Wilcox, RK Matavalam, JR Diaz, L Davis, and J D’Agostini. Modeling the effect of weather parameters on power distribution interruptions. 7th IASTED Int. Conf. Power and Energy Systems, Clearwater Beach, Fl, USA, 2004.
- [18] A Domijan Jr, RK Matavalam, A Montenegro, WS Wilcox, YS Joo, L Delforn, JR Diaz, L Davis, and JD Agostini. Effects of norman weather conditions on interruptions in distribution systems. *International journal of power & energy systems*, 25(1):54–61, 2005.
- [19] Drew Fudenberg and Jean Tirole. *Game Theory*. The MIT Press, 1991.
- [20] G Gallego, W Huh, W Kang, and R Phillips. Price competition with the attraction demand model: Existence of unique equilibrium and its stability. *Manufacturing & Service Operations Management*, 8(4):359, 2006.
- [21] Guillermo Gallego and Ming Hu. Dynamic pricing of perishable assets under competition. *Agifors Reservations and Yield Management Study Group Annual Meeting Proceedings*, 2007.
- [22] Lawrence C Hamilton, Cliff Brown, and Barry D Keim. Ski areas, weather and climate: time series models for new england case studies. *International Journal of Climatology*, 27(15):2113–2124, 2007.
- [23] Seung Ryong Han. Estimating hurricane outage and damage risk in power distribution systems. *Doctoral Thesis*, 2008.
- [24] Seung-Ryong Han, Seth D Guikema, Steven M Quiring, Kyung-Ho Lee, David Rosowsky, and Rachel A Davidson. Estimating the spatial distribution of power outages during hurricanes in the gulf coast region. *Reliability Engineering & System Safety*, 94(2):199–210, 2009.

- [25] Li. Hongfei, L.A. Treinish, and J.R.M. Hosking. A statistical model for risk management of electric outage forecasts. *IBM Journal of Research and Development*, 54(3):8:1–8:11, May-June 2010.
- [26] M Hu. Strategic substitutes in competitive revenue management with multiple predetermined options. pages 1–41, Jan 2009.
- [27] M Hu and G Gallego. Dynamic pricing of perishable assets under competition. pages 1–43, Jan 2010.
- [28] Vincent J. Forte Jr. Faults caused by drought stressed trees. Distribution system engineering, National Grid, November 2004.
- [29] Barry D Keim, Loren David Meeker, and John F Slater. Manual synoptic climate classification for the east coast of new england (usa) with an application to pm2.5 concentration. *CLIMATE RESEARCH*, 28:143–154, 2005.
- [30] Haibin Liu, Rachel A Davidson, and T Apanasovich. Statistical forecasting of electric power restoration times in hurricanes and ice storms. *Power Systems, IEEE Transactions on*, 22(4):2270–2279, 2007.
- [31] Haibin Liu, Rachel A Davidson, and Tatiyana V Apanasovich. Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms. *Reliability Engineering & System Safety*, 93(6):897–912, 2008.
- [32] Stuart Lloyd. Least squares quantization in pcm. *IEEE*, 1982.
- [33] Hans-Andrea Loeliger. An introduction to factor graphs. *Signal Processing Magazine, IEEE*, 2004.
- [34] David MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press, 2003.
- [35] Eric Maskin and Jean Tirole. A theory of dynamic oligopoly, ii: Price competition, kinked demand curves, and edgeworth cycles. *Econometrica*, 1988.
- [36] Roop Kishore R Matavalam. *Power Distribution Reliability as a Function of Weather*. PhD thesis, University of Florida, 2004.
- [37] Robert J. McEliece, David J. C. MacKay, and Jung-Fu Cheng. Turbo decoding as an instance of pearl’s belief propagation algorithm. *IEEE Journal*, 1998.
- [38] Michael A McGeehin and Maria Mirabelli. The potential impacts of climate variability and change on temperature-related morbidity and mortality in the united states. *Environmental Health Perspectives*, 109(Suppl 2):185, 2001.
- [39] Judea Pearl. Reverend bayes on inference engines: A distributed hierarchical approach. *AAAI*, 1982.
- [40] G Perakis and A Sood. Competitive multi-period pricing for perishable products: A robust optimization approach. pages 1–41, Feb 2006.
- [41] R Philips. *Pricing and Revenue Optimization*. Stanford University Press, 2005.
- [42] Linda van der Gaag. Pearl’s belief propagation: the proofs. *RUU-CS*, 1992.

- [43] G Weintraub, C Benkard, P Jeziorski, and B Van Roy. Nonstationary oblivious equilibrium. 2008.
- [44] Fei Xiao, James D McCalley, Yan Ou, John Adams, and Steven Myers. Contingency probability estimation using weather and geographical data for on-line security assessment. In *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, pages 1–7. IEEE, 2006.
- [45] L. Xu and R.E. Brown. A framework of cost-benefit analysis for overhead-to-underground conversions in florida. In *Power & Energy Society General Meeting, 2009. PES '09. IEEE*, pages 1–7, Calgary, AB, July 2009.
- [46] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. *MEL*, 2001.
- [47] Yujia Zhou, Anil Pahwa, and Shie-Shien Yang. Modeling weather-related failures of overhead distribution lines. *Power Systems, IEEE Transactions on*, 21(4):1683–1690, 2006.