Stochastic Optimization in Disaster Operations Management: Medical Capacity Planning During Epidemic, Optimal Subsidy Policy For Renewal Technology Adoption and Data-driven Robust Supply Chain Optimization

Swapnil Rayal

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Apurva Jain, Chair

Matthew Lorig

Leela Nageswaran

Program Authorized to Offer Degree:
Foster School of Business

University of Washington

Abstract

Stochastic Optimization in
Disaster Operations Management: Medical Capacity Planning During Epidemic, Optimal Subsidy Policy For Renewal Technology Adoption and Data-driven Robust Supply Chain Optimization

Swapnil Rayal

Chair of the Supervisory Committee:
Apurva Jain
Information Systems and Operations Management

Each of the problems addressed in this thesis falls under the broader umbrella of disaster operations management, where the focus is on minimizing the impact of unforeseen disruptions on critical systems. Whether it's managing the surge in healthcare demand during a pandemic, promoting sustainable energy adoption to mitigate long-term environmental crises, or optimizing supply chains to withstand production disruptions, the core challenge remains the same: to design resilient systems capable of navigating uncertainty and resource constraints. These problems highlight the need for well-informed, strategic decision-making to ensure that essential services continue to function even in the face of crises.

The first problem deals with managing medical equipment capacity during the early spread of an infection in a region, a critical research problem during COVID pandemic. After a brief introduction to infectious disease modeling, we develop a model for a regional decision-maker to analyze the requirement of medical equipment capacity in the early stages of a spread of infections. We use the model to propose and evaluate ways to manage limited equipment capacity. Early stage infection growth is captured by a stochastic differential equation (SDE) and is part of a two-period community spread and shutdown model. We use the running-maximum process of a geometric Brownian motion to develop a performance metric, probability of breach, for a given capacity level. Decision-maker estimates costs of

economy versus health and the time till the availability of a cure; we develop a heuristic rule and an optimal formulation that use these estimates to determine the required medical equipment capacity. We connect the level of capacity to a menu of actions, including the level and timing of shutdown, shutdown effectiveness, and enforcement. Our results show how these actions can compensate for the limited medical equipment capacity in a region. We next address the sharing of medical equipment capacity across regions and its impact on the breach probability. In addition to traditional risk-pooling, we identify a peak-timing effect depending on when infections peak in different regions. We show that equipment sharing may not benefit the regions when capacity is tight. A coupled SDE model captures the messaging coordination and movement across regional borders. Numerical experiments on this model show that under certain conditions, such movement and coordination can synchronize the infection trajectories and bring the peaks closer, reducing the benefit of sharing capacity.

We then turn our attention to the problem of devising optimal subsidy to encourage adoption of solar technology in a region. Each household in a population characterized by income heterogeneity faces random demand for electricity and decides if and when it should adopt a solar product, rooftop solar or community solar. A central planner, aiming to meet an adoption level target within a set time, offers net metering and subsidy on solar products and minimizes its total cost. Our focus is on analyzing the interactions of three new features we add to the literature: income diversity, availability of community solar, and consideration of adoption timing. We develop a bilevel optimization formulation to derive the optimal subsidy policy. The upper level (planner's) problem is a constrained non-linear optimization model in which the planner aims to minimize the average subsidy cost. The lower level (household's) problem is an optimal stopping formulation, which captures the adoption decisions of the households. We derive a closed-form expression for the distribution of optimal adoption time of households for a given subsidy policy. We show that the planner's problem is convex in the case of homogeneous subsidy for the two products. Our results underscore the importance for planners to consider three factors - adoption level tar-

get, time target, and subsidy budget - simultaneously as they work in tandem to influence the adoption outcome. The planners must also consider the inclusion of community solar in their plans because, as we show, community and rooftop solar attract households from different sides of the income spectrum. In the presence of income inequality, the availability of community makes it easier to meet solar adoption targets.

The third problem we consider is a classic mixed integer programming formulation for supply chain optimization of a semi-conductor supplier in presence of supply disruptions. We provided a deterministic formulation for optimal outbound network routing. We integrate the past publicly available data in the optimization model and show the shortcomings of its optimal solution using a numerical experiment. We then designed a practical robust formulation to handle uncertain production capacity. We numerically study the optimal solutions for low and high tolerance to deviation from deterministic optimal solution and generate optimal production and transportation plan which is not only more resilient to random shocks but also requires marginal change from current plan.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My journey as a researcher has been profoundly shaped by my time at the University of Washington (UW). UW offered a unique academic environment where I had the freedom to explore courses across different departments, nurturing both my curiosity and interdisciplinary thinking. It was a place that encouraged exploration, welcomed mistakes, and embraced learning through experience. Throughout my PhD, I've been fortunate to meet exceptional mentors, scholars, and friends—relationships that I will carry with me for a lifetime.

I am especially grateful to my advisor, Prof. Apurva Jain. Over the past six years, he has been an unwavering source of guidance, encouragement, and wisdom. More than just a mentor, Apurva has been a true advocate for my growth, both professionally and personally. He consistently gave me the space to develop and pursue my ideas, while also challenging me to think deeper and push beyond my comfort zone. His thoughtful feedback, patient listening, and belief in my potential made a lasting impact on the way I approach research and problem-solving. Beyond academics, his integrity, kindness, and humility have served as an example of the kind of scholar and person I aspire to be. I feel incredibly fortunate to have had the opportunity to work with him, and I will always carry his mentorship with me.

I would also like to extend sincere thanks to Prof. Matthew Lorig, whose exceptional support and guidance have been instrumental in shaping my research. Matt's technical prowess is truly remarkable—his deep understanding of complex concepts, combined with his ability to break them down into clear, manageable components, has had a profound influence on my work. His systematic approach to problem-solving and research has provided me with invaluable insights into how to approach challenging questions with both rigor and creativity. His course on Brownian motion sparked my initial interest in the field, and

the advanced techniques I learned under his tutelage have become crucial tools in my own research.

I am thankful to my doctoral committee members, Prof. Leela Nageswaran and Prof. James Ritcey, for their valuable guidance and support during my time at UW. I also appreciate the Foster School of Business and staff members Shawna Reimers, Nuzulita Budhiari, and Beau Kirkeby for cultivating a supportive and welcoming environment for the PhD program.

Above all, I am eternally grateful to my parents, whose unwavering support and belief in me have been the bedrock of my academic journey. Their encouragement continues to inspire my pursuit of knowledge and meaningful contribution to society. Finally, I owe my heartfelt thanks to my wife, Ankita, whose patience, love, and unwavering support have been a constant source of strength—I could not have completed this journey without her by my side.

DEDICATION

I dedicate this thesis to my wife Ankita, my mother Suchita, my father Sanjay and my brother Snehil.

Chapter 1

INTRODUCTION

Disasters, whether unfolding slowly like climate change or striking suddenly like a pandemic, have a profound ability to disrupt lives, economies, and systems we rely on every day. In recent years, the world has been reminded of just how fragile our interconnected infrastructure can be. Hospitals pushed beyond their limits, energy systems under pressure to transition sustainably, and global supply chains brought to a standstill—all underscore the urgent need to think ahead. These moments of crisis are not only tests of endurance but also opportunities to design smarter, more resilient responses that can better withstand the next shock. This thesis explores three such moments of disruption through the lens of operations management.

Pandemics particularly represent an alarming threat: unpredictable in timing, rapid in spread, and deeply damaging in their impact. The COVID-19 pandemic starkly revealed how even the most advanced healthcare systems can be brought to their knees by a sudden surge in critically ill patients. Medical equipment such as ventilators became emblematic of both our technological capabilities and our logistical unpreparedness. Beyond the immediate health crisis, the ripple effects extended into economic slowdowns, mental health struggles, and long-term shifts in how societies function. These experiences have shown that the threat of pandemics is not hypothetical—it is a lived reality with the potential to upend nearly every facet of modern life. In the face of such risk, proactive planning is not just wise—it is essential. Chapter 2 considers the early days of a potential pandemic asking how we can plan ventilator capacity when the scale of need is uncertain but potentially overwhelming, when should we impose lock-downs and should we share resources with the neighboring states.

While pandemics strike with sudden urgency, climate change presents a slower but no less devastating threat—one that is already unfolding around us. Rising global temperatures,

intensifying storms, prolonged droughts, and devastating wildfires are no longer future possibilities; they are present-day realities affecting millions across the globe. The consequences of inaction are mounting, with vulnerable communities bearing the brunt of environmental degradation, economic instability, and resource scarcity. At the heart of the response to this crisis lies the need to reduce our dependence on fossil fuels and transition to cleaner, renewable sources of energy. Among these, solar technology stands out as a scalable and increasingly affordable solution. Yet, despite its promise, widespread adoption remains uneven, often hindered by high upfront costs and policy barriers. Encouraging faster uptake of solar power through well-designed subsidy policies is not just an economic decision—it is a critical step toward mitigating the long-term damage of climate change and building a more sustainable future. Chapter 3 investigates how subsidy policies can be structured to encourage faster adoption of solar technology.

Among the most vivid examples of how deeply disasters can disrupt global systems is the semiconductor industry. These highly sophisticated components are essential to everything from consumer electronics and electric vehicles to medical equipment and military technology. Yet despite their critical role, semiconductor supply chains are stretched thin across continents, relying on just-in-time production models and limited manufacturing nodes. This fragility has been repeatedly exposed by disasters of various kinds—natural events like earthquakes and floods, pandemics that shut down production facilities, and geopolitical tensions that threaten cross-border flows. The impact of these disruptions is far-reaching: global shortages, production delays, skyrocketing prices, and lost revenue across multiple sectors. The semiconductor shortage during the COVID-19 pandemic, for instance, brought entire auto factories to a standstill and highlighted how localized shocks can cascade into global crises. In such a landscape, resilience is no longer a luxury but a necessity. Chapter 4 takes us into the world of semiconductor manufacturing, where even a brief production halt can ripple across industries; here, the focus is on finding the right level of buffer inventory and routing strategies to keep operations running smoothly despite disruptions.

Each project stands on its own, but together, they offer a broader perspective on preparing for and managing crises in critical systems.

Chapter 2

MANAGING MEDICAL EQUIPMENT CAPACITY WITH EARLY SPREAD OF INFECTION IN A REGION

2.1 Introduction

During the early spread of COVID-19, there has been widespread reporting about shortages of medical equipment like ventilators, vital sign monitors, and intensive-care-unit beds (45). Institutions like the FDA and WHO have provided information about monitoring and tracking such shortages (33). The demand for such equipment is driven by the evolution of infections in different regions. The supply is severely constrained. The regional administrators and healthcare authorities must determine the level of capacity they need in their regions. Given the limited availability, they must also identify ways to use this capacity to improve health outcomes in their regions. Our goal in this paper is to develop an analytical model to guide such decisions.

The determination of the right capacity and its management to improve performance is a core concern of the production and operations management (POM) research. The literature, however, has paid limited attention to managing demands that are driven by the spread of a contagious infection. The growth pattern of infections, an early exponential rise followed by a decline or a plateau, determines how much capacity will be needed and when. The literature is also relatively silent about situations where decision-makers, facing considerable uncertainty about infection-growth patterns and other model inputs, exhibit a wide range in the decisions they make (38). We aim to develop a model that can capture the essential new features of this context: uncertainty in how infections grow and drive demand, a wide range of decision-making choices in different regions, and the need to compensate for the limited capacity.

We take the perspective of a regional decision-maker (she) to address the following questions. Based on her region's infection growth and her estimates of its impact on the region's

health and economy, how should she determine the medical equipment capacity? Given that such capacity may not be fully available, what is the impact of compensating actions—from early shutdowns to consistent enforcement messaging and regional capacity sharing—on her region's health outcomes? Our approach to answering these questions is analytical, but the model provides an understanding of trade-offs and insights into the value of collaboration that is practically useful. On the theoretical side, our contributions include developing a model that (i) derives medical equipment demand from the stochastic spread of infections, (ii) accounts for the variety in decision-makers' actions, and (iii) proposes and evaluates ideas for improving health outcomes with constrained supply. Our focus is on the demand side; we do not model the acquisition of medical equipment supply.

Specifically, we analyze a stochastic differential equation (SDE) of infection growth in the early stages. Based on this growth, we develop a measure of the chance that the need for medical equipment over a period will exceed a given capacity, the breach probability. We then describe a formulation of a regional decision maker's beliefs about costs related to economy and health and the expectation of a cure. We show how her preferred estimates interact with the spread of infections to determine the infection level at which a region should shut down to reduce its infection rate. We also analyze the capacity needed at the shutdown time. We present a heuristic rule for this decision and an optimal stopping formulation. We show that while a lack of capacity can drive up the breach probability, actions like reusability, early or high-intensity shutdowns, and consistent messaging can reduce it. We then develop a coupled SDE model to analyze the impact of capacity coordination across two regions. The model allows for travel between the regions and for the coordination of messaging and shutdown policy. We analyze the impact of such coordination on the regions' breach probability.

We develop the following results and insights:

- Modeling uncertainty in infection rates in an SDE formulation induces the needed realism in the model while still maintaining tractability for analyzing outcomes like the breach probability.

- A region can shut down to reduce infection rate; should it do so and when? The

model clarifies the trade-off between incurring a cost to shutting down the economy and reducing the probability of a breach. A heuristic decision rule determines the level of infections at which to shut down.

- At the shutdown level of infections, capacity determination is a function of the decision maker's two estimates: the relative cost impacts of an economic shutdown versus a breach of health capacity and the time until the availability of a cure or a vaccine. A simple visual guide lets her determine the capacity she needs corresponding to her estimates. The guide maps the variety of decision-makers' beliefs and estimates to the range of preparatory actions we observe across various US states.

- An optimal stopping formulation of the problem leads to observations similar to those based on the heuristic rule.

- If capacity is limited, what actions can a region take to reduce its breach probability? The toolbox includes reusability, early shut down or high-intensity shut down, and consistent enforcement messaging. The model can determine the reduction in breach probability that these actions can achieve. Consistency in messaging reduces noise by limiting the variation in people's behavior; its impact on the breach probability is not always beneficial.

- Looking beyond a region's boundaries, what is the impact of sharing capacity with a neighboring region? We model the dynamics as a coupled SDE and prove a technical result for the solution's existence and positivity. The model captures travel between regions and the correlation between the random processes underlying the two regions' infec- tions.

- Without any travel and correlation, capacity-sharing between the two regions will reduce the breach probability. Interestingly, this holds only if the capacity is not too tight. We identify two separate drivers for this effect: a pooling effect and a peak-timing effect.

- Movement across the border of the two regions can increase the breach probability with shared capacity.

- Consistency in pandemic related messaging and joint efforts to maintain data accuracy correlates with the random processes underlying the two regions' infections. An increase in correlation can increase the breach probability with shared capacity.

In addition to these practical insights, our theoretical contribution to the POM literature includes offering a tractable model to determine the equipment capacity as a function of the running maximum of infections evolving as a geometric Brownian motion. We believe that the paper contributes to the POM literature as an early effort to introduce uncertainty in the pandemic-driven demand model while keeping it tractable for capacity questions. The coupled SDE formulation for regional coordination is new to the literature as well; the proof of the condition for its solution's existence and positivity is a technical contribution to the literature. The insights into the drivers of capacity-sharing benefits introduce the idea of peak-timing, which is new to the literature.

Section 2 describes how our model connects to three different streams of literature. Section 3 presents the model, and Section 4 offers an analysis of the shutdown level and the capacity. Section 5 presents the impact of compensatory actions. Section 6 presents the regional collaboration coupled SDE model. Section 7 discusses possible extensions and concludes the paper.

2.2 Literature Survey

We draw on several different streams of literature: epidemiology literature, stochastic finance literature, and capacity models with diffusion-driven demand in operations management. This section highlights how our model connects to these streams and how we differentiate ourselves from the existing models. Some of the recent and rapidly growing literature on pandemic modeling is also discussed.

Our starting point is a model of how infections spread in a pandemic. There is a vast body of literature in epidemiology that develops and analyzes such models. The best known class of such models is compartment models, with a well-known example being the susceptible,

infectious, recovered (SIR) model. These types of models divide the population into separate compartments and estimate transition rates between these compartments. If we drop the recovered compartment, it is called an susceptible, infectious, susceptible (SIS) model. If we add an exposed compartment, it is a susceptible, exposed, infectious, recovered model. If post- infection immunity fades away, the model is called susceptible, infectious, recovered, susceptible. Demographic and geographic details can be added to population transition rate inputs. The resulting ordinary differential equations can be solved numerically or simulated to project the number of infections over time. For a good introduction, see (59).

Our goal is to analyze capacity questions in a setting that captures the uncertainty inherent in the spread of infections. Most compartment models, however, do not offer analytical solutions even without any uncertainty. Our approach is to present a simplified, early-stage version of the SIR model before we add uncertainty to it. The early-stage approach, discussed in (18) among others, relies on the observation that at the beginning of the pandemic, the number of susceptible (almost the whole population) is extremely large, compared to the number of infections. The next section discusses the early-stage model in detail and offers other supporting references for this approach.

We add uncertainty by explicitly modeling the probabilistic spread using an SDE. Some recent reporting (see, e.g., (68)) has argued that classic epidemiological models tend to use constant reproduction rate numbers even though, in reality, the spread depends on uncertain factors. There is, however, plenty of discussion of stochastic modeling in epidemiology. For basic computational methods for analyzing probabilistic spread, see (91). The direct approach to introduce stochasticity in a compartment model is to use continuous-time Markov chains (CTMC). For the CTMC approach, now known as the stochastic general epidemic model, see (11). In general, these models yield only to computational analysis. The use of SDEs to model the spread is relatively new. The literature justifies the use of SDE models for epidemic modeling in several ways. (6) presents discrete-state CTMC models and then shows how they can be approximated by SDEs. In some cases, SDEs are introduced as a reasonable way to model growth without any need for further justification; see, for example, (90). Another approach is to perturb the infection rate parameter by adding noise to it. This is the approach we take; the next section describes it in detail with supporting

references. Our contribution is to combine the SDE method with the early-stage modeling discussed above to present an approach that is tractable enough to analyze questions about capacity, the area of interest in the production and operations management literature.

In contrast to population-based compartment models, an alternative approach is to treat each individual as an agent and develop the epidemic dynamics based on interactions between agents. These individual-based models allow for the incorporation of agents' choices and reactions to testing and shutdown policies. See, for example, (1) and (29). Given our focus on the aggregated need for capacity and on how it evolves over time, we do not take this individual agent-based approach. Our paper fits a rapidly growing literature on modeling the impact of economic, social, and operational policies in the context of the recent pandemic. Some examples of papers that focus on operations issues include (60), (7), and (94).

The second stream of literature that we draw upon is stochastic finance. While we do not interpret our model using finance terminology, the analytical techniques we use are similar to those used in finance. Analysis and solution of SDEs is a foundation of stochastic finance. See, for example, (78). The pricing of exotic options requires an analysis of running maximum; we use a similar approach, see, for example, (20). Finally, we will model the decision to shut down as a stopping problem. Stopping problems are rigorously presented in (77) and (67)), but they are also used in finance to make investment decisions, see, for example, (26).

We extend this literature by presenting a new application of SDEs to model the spread of infections in two neighboring regions. The regions influence each other due to the possibility of people traveling across the border. They are also related through the correlation of Wiener processes that underlie the evolution of infections. This leads to the creation of a set of two coupled SDEs. Our contribution to this literature is a new result to show the existence and positivity of a solution to these equations.

Finally, the third stream of literature we relate to is the analysis of capacity or inventory decisions in the presence of demands that follow the increasing-then-decreasing pattern discussed above. While we are not aware of any work that directly addresses demands influenced by the spread of infections, the first-increasing-then-decreasing pattern is similar

to the new product diffusion. There has been some work in the operations management literature that relates to diffusion demand. (49) and (53) compare the option of delaying the launch with the option to reject customers when capacity is limited. (76) include pricing as a decision variable in addition to capacity and sales policy. These papers follow a new product demand model based on (14) diffusion, a deterministic approach. The use of SDEs and optimal stopping formulation for other POM problems is comparatively limited; see (54) and (9) for two recent examples. We contribute to this literature by offering a stochastic diffusion model to consider capacity questions in a new context, the spread of infections in a pandemic.

2.3 Stochastic Model

We model a region with a decision-maker. The region is observing the beginning of the spread of infections. A continuous variable I_t represents the number of infections in the region at time t. We will first present the dynamics of the spread and then discuss its underlying assumptions based on epidemiology literature.

2.3.1 Dynamics of the infection spread

We model the growth of infections by focusing on the new infections generated by an infected individual. We define the transmission rate b to be equal to the new infections as the instantaneous fraction of the infected population. To consider the recovery of infected individuals, we model the recovery rate d as the instantaneous fraction of the infected population that leaves the infected compartment due to recovery or death. For notational brevity, we define the effective infection spread rate as $r = b - d$, a constant estimated based on actual observations. The rate of change in the number of infections is as follows

$$\frac{dI_t}{dt} = (b - d)I_t = rI_t \tag{2.1}$$

2.3.2 Early-stage modeling

Our approach is a simplification of the SIR model as applied to the early spread stage. In the standard SIR model, the infection dynamics are presented as $\frac{-dI_t}{dt} = \frac{\beta S_t I_t}{N} - \gamma I_t$ where

S_t represents the number of susceptible at time t, N is the fixed total population and β, γ are rate parameters. For detailed explanations and parameter estimation, please see the references available in the previous section. The first term captures the increase in the number of infections as a fraction of interactions between the susceptible and infected populations. In the early stage spread of infections, the whole population N of the region, an extremely large number, is susceptible to infection. That is, there is an early stage in which $\frac{S}{N} \approx 1$ is a reasonable assumption. Substituting this in the above SIR equation leads to our model. In other words, when the number of susceptible (almost the whole population) is extremely large, compared to the number of infections, it is reasonable to assume that each infection generates a certain number of new infections irrespective of how many susceptible remain. Therefore, it is sufficient to track only the number of infections to model early-stage spread. This early stage approach is well supported in the literature. Employ this approach to model the spread of COVID-19 in stages. For a unit-size population, they assume $S \approx 1$ in the first stage of the epidemic. (18) suggests that "The key reason for having an approximation during the early stages of an outbreak when there are many initial susceptible is that it is then very unlikely that any of the first numbers of infectious contacts happen to be with the same susceptible individual. Conversely, it is very likely that all of the first set of infectious contacts happen with distinct individuals." This formulation is equivalent to our model. (73) model the spread in two stages where the first (early) stage is analyzed without the recovered compartment.

The modeling advantage of this early-stage approach is that it allows us to introduce uncertainty in the model while keeping it tractable for capacity considerations. The disadvantage is that in the absence of an intervention, the number of infections in the model will continue to grow. We address this later by adding a second stage where a shutdown decreases the infection rate.

2.3.3 Stochastic spread

The above dynamics are deterministic, but there is a great deal of variation in how infected individuals interact with others and spread the infection. At the beginning of the pandemic,

the drivers of this variation, such as people's behavior, the accuracy of information, penetration of testing, and so forth, are poorly understood. The decision-maker perceives them as unpredictable variability in the spread. We model the infection spread rate as a stochastic process $\tilde{r}$ where $d\tilde{r} = rdt + sdW_t$, s is the volatility, and dW_t is the increment of the standard Brownian motion (Wiener process) $\{W_t\}_{t\geq 0}$. We can now rewrite the dynamics of the infection spread as follows:

$$dI_t = (rdt + sdW_t)I_t = rI_t dt + sI_t dW_t. \tag{2.2}$$

The stochastic process I_t, which satisfies the above SDE is defined on an underlying probability space (Ω, F, P), which is equipped with a filtration $F = \{F_t\}_{t\geq 0}$and the stochastic process is adapted to F. Parameters r and s are constants.

We have briefly reviewed the literature on stochastic epidemic models in the previous section. The parameter perturbation approach (addition of a time-dependent noise in the infection rate) that we have used to create the SDE model is directly supported by (44), (21), (85), and references therein.

It is useful to consider the interpretation of the volatility parameter s and the underlying Wiener process W_t because these notations are not usually part of infection-spread models. The SDE is interpreted to mean that in successive small intervals $[t, t + dt), [t + dt, t + 2dt), ..., [t + (n - 1)dt, t + ndt)$, the numbers of new infections created by an infected individual are independent and identically distributed random variables that follow a normal distribution with mean rdt and variance s^2dt. Thus, the Wiener process captures the time dependent variability of the number of contacts an infected individual will make in a small period, and s can be adjusted to scale its impact on the spread. This variability is driven by the inconsistency in the behavior of the infected individual and the behavior of others, a natural part of human interaction. A lack of standard messaging will add to this variability. Incomplete and non-uniform testing leads to noise in the observed number of infections and that too will add to this variability. A possible way to estimate s is to collect estimates of infection rates from various sources like hospitals, neighborhood health centers and use this sample to estimate it. For the interpretation and estimation of the infection rate, we refer to (35). In general, for the entire timeline defined below, we assume $r - \frac{s^2}{2} \geq 0$ in order to

avoid scenarios in which $I_t \to 0$ a.s. when $t \to \infty$.

2.3.4 The timeline

We frame our model in terms of two periods, beginning at two discrete points in time. At time, t_0, when a very small number of initial infections I_{t_0} are observed, the regional decision maker starts tracking the number of infections. The infection rate is r_0 and the volatility is s_0, both given constants. At a time t_1 of her choosing, she observes the number of infections I_{t_1} and decides if she will take action (e.g., mask mandates, shutdowns) to decrease the infections' spread. If she chooses to shut down, we change the infection rate to $r_1 < r_0$ and the volatility to s_1, assumed to be equal to s_0, $s_1 = s_0 = s$, unless otherwise mentioned. If she chooses not to shut down, the rate and volatility parameters remain unchanged. Even though it is a shutdown-or-not decision, for convenience, we refer to it as a shutdown decision.

t_o :	A given constant, representing the start of the decision-maker's observations.
I_{t_0} :	A given constant, representing the very small number of infections observed at the time t_0.
t_1 :	Decision time at which the decision-maker observes the number of infections I_{t_1} and determines whether to shut down. For now, it is assumed to be a given constant $t_1 > t_0$. Later, we develop a decision rule to determine t_1.
I_t or $I_t(I_u, r, s)$:	Stochastic process I_t at time $t > u$, starting at time u with a known constant value I_u, and following the early-stage SDE described above with known constant parameters r and s. For $t_0 \leq t < t_1$, we have $I_u = I_{t_0}, r = r_0, s = s_0$. For $t > t_1$,the process evolves conditional on the filtration up to the time t_1 and the decision taken at that time: If the decision at t_1 is to shutdown,we have $I_u = I_{t_1}, r = r_1 < r_0, s = s_1$;and if the decision is to not shut down, we have $I_u = I_{t_1}$, $r = r_0, s = s_0$. Depending on the context, we will drop the arguments and simply refer to the process as I_t.

We refer to the interval $[t_0, t_1)$ as the early community spread period. During this period, some medical understanding of the pandemic is starting to build, but there is no clear indication about if and when a therapeutic cure or vaccine may be discovered. There is no common understanding of when a medical solution may be available to halt the spread of infections. Nevertheless, a regional decision-maker must look ahead toward such a possibility when making a shutdown decision. That is, at the decision time t_1, she needs to have an estimate of the horizon for her decision problem. It may be possible to argue that given the state of medical knowledge about the pandemic at the decision time, one can estimate a random variable to represent the horizon she should consider. A more likely scenario is that, given the extreme uncertainty, decision makers let their preferences guide the level of optimism they feel about the pandemics' end. See, for example, (17) and (43) for differences

in decision-makers' estimates of how long the pandemic will last. Therefore, we chose to model the decision horizon T as a preferred estimate of the decision-maker, which would vary from one decision-maker to the other.

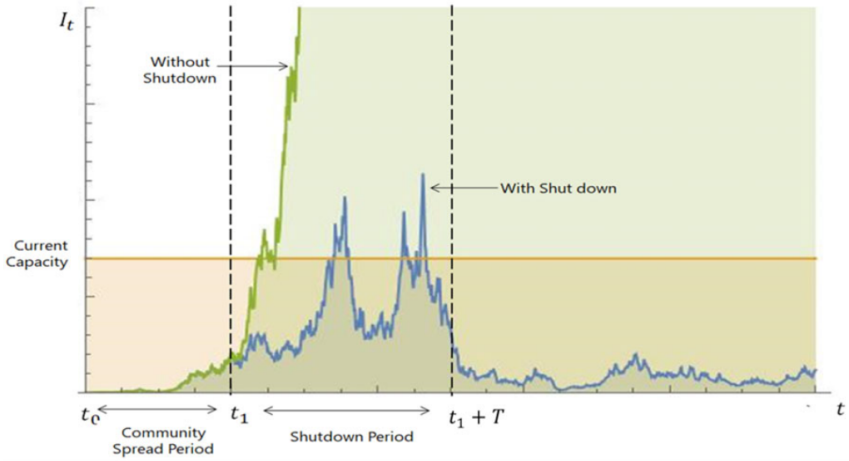


Figure 2.1: Spread of infection in a region with and without shutdown during different time periods

If a shutdown is implemented at a time t_1, we refer to $[t_1, t_1 + T]$ as the shutdown period, but we emphasize that $t_1 + T$ does not model the actual reopening time. It only rep- resents the decision maker's estimate of how far she prefers to look ahead. Given our focus on the capacity question during the early spread, we do not specifically model the reopening decision. In the extensions section, we provide a brief discussion of the reopening decision. Figure 2.1 shows two sample trajectories with and without shutdown.

We believe that this two-period framing reflects how the recent pandemic unfolded in various regions, U.S. states, and other countries. At the same time, it is flexible enough to capture the variety of choices made by the regional decision makers about the timing and intensity of shutdowns. Our focus in this paper is on the decision time t_1 and the shutdown period. This is the period in which the region experiences a peak in infections and in which the questions about medical equipment capacity have the most relevance.

2.3.5 Cost and performance considerations

As the decision-maker watches the infections increase in the early community spread period, it becomes clear that the disease will continue to grow and impose adverse health consequences. The degree of this adverse impact will depend on the region's level of preparation (medical equipment capacity). In the absence of therapeutics or vaccines, the decision-maker must decide if she should take any preventive measures and, if so, when. These measures can range from a ban on large gatherings to a wholesale shutdown of businesses. Different measures may bring different levels of improvements in the infection rates, but they will all impose an adverse economic consequence on the region. In our basic model, we capture these adverse health and economic consequences by the following parameters.

T :	The decision-horizon, a constant; the decision-maker who is considering a shutdown decision at a time $t_1 > t_0$ will consider the impact of her decisions in $[t_1, t_1 + T]$.
m :	Preparation level or medical equipment capacity, a known constant, expressed in terms of the maximum number of infections that the regional health system can simultaneously manage. Note that there are two types of equipment, those that cannot be reused and those that can be. The former's required capacity depends on the cumulative number of infections and the latter on the maximum number of infections. The required capacity for equipment like ventilators is of the latter type; this is how we define m here.
$PrB(I_{t_1}, m; r_i, s_i, T)$:	Breach probability; the probability that starting with I_{t_1} number of infections at the time t_1 and evolving with the rate r_i and volatility s_i, $i = 0, 1$, the maximum number of infections in $[t_1, t_1 + T]$ will exceed the capacity m.
C_h :	A fixed health cost incurred if the maximum number of infections exceeds the capacity m.
C_e :	A fixed economy cost that the decision-maker believes will be incurred if she chooses to implement a regional shutdown.

The primary performance measure in our model is the probability of a breach. If a breach occurs, the region incurs a fixed health cost. The practical relevance of these measures is evident from reporting such as (71) that discusses the risk of shortage and efforts to prepare for a breach. The region can influence this probability by incurring a fixed economy cost. Once again, the extreme uncertainty surrounding these events makes it unlikely that regional decision-makers have precise estimates for such costs. We believe that a model built on the decision-makers' preferred estimates or beliefs regarding such costs better suits the reality. See, for example, (58) to support our modeling of differences in decision-makers' estimates of these costs. Our performance measures capture this situation's essential trade-off while

directly connecting it to the paper's main new feature, infection rate uncertainty. In the extensions section, we briefly discuss other, more detailed metrics to capture the impact on health and economic performance.

Decision-makers in different regions have different preferences about how they view the health versus economy debate and how optimistic they are about the availability of a cure. These preferences guide their estimates of C_h, C_e and T. Given these preferences, a decision-maker can compare the costs of shutdown versus not shutting down as we analyze below.

2.4 Analysis

In this section, we analyze the expression for the number of infections, develop a formula for the breach probability, and then analyze the decision to shut down and how it relates to the capacity level.

2.4.1 Expression for I_t

The first step in the analysis is to obtain an expression for the number of infections at time t, I_t. We use standard methods (78) from the analysis of SDEs to analyze the infection dynamics to obtain the following solution.

Proposition 1.

$$I_t = \begin{cases} I_{t_0} e^{(r_0 - \frac{s_0^2}{2})(t-t_0) + s_0(W_t - W_{t_0})} & t_0 \leq t < t_1 \\ I_{t_1} e^{(r_1 - \frac{s_1^2}{2})(t-t_1) + s_1(W_t - W_{t_1})} & t_1 \geq t_1 \textit{ if a shutdown is implemented} \\ I_{t_1} e^{(r_0 - \frac{s_0^2}{2})(t-t_1) + s_0(W_t - W_{t_1})} & t_1 \geq t_1 \textit{ if a shutdown is not implemented} \end{cases}$$

It is straightforward to recognize the dynamics of infections as a geometric Brownian motion, leading to the above result. Proofs are available in the Appendix in Supporting Information. The equations describe how the number of infections evolves, first with community spread in the period $t_0 \leq t < t_1$ and then with shutdown interventions in the period $t \geq t_1$.

2.4.2 The breach probability

We next develop an expression for the probability that the number of infections exceeds the capacity, a measure we have defined as the breach probability. The explicit expression of It in the above result allows us to determine the probability that the maximum number of simultaneous infections will exceed a given level of capacity over a time horizon. We evaluate this probability at the decision time t_1, conditional on the filtration up to the time t_1. Let $\widehat{I_{t_1}} = \max_{t_1 \le t' \le t_1+T} I_{t'}$. For $I_{t_1} \ge m$, the breach probability is trivially 1. For $0 < I_{t_1} < m$, we have the following result:

Proposition 2.

$$PrB(I_{t_1}, m; r_i, s_i, T) = P(\widehat{I_T} > m) = 1 - \Phi\left(\frac{\frac{1}{s_i}\log\left(\frac{m}{I_{t_1}}\right) - \frac{1}{s_i}(r_i - \frac{s_i^2}{2})T}{\sqrt{T}}\right) \tag{2.3}$$
$$+ e^{2\frac{1}{s_i}\log\left(\frac{m}{I_{t_1}}\right)\frac{1}{s_i}(r_i - \frac{s_i^2}{2})}\Phi\left(\frac{-\frac{1}{s_i}\log\left(\frac{m}{I_{t_1}}\right) - \frac{1}{s_i}(r_i - \frac{s_i^2}{2})T}{\sqrt{T}}\right)$$

for i={0,1}, where Φ is the standard normal cumulative distribution function.

We briefly discuss the plan for the rest of this section. As the time progresses forward from t_0, we arrive at the given decision time $t_1 > t_0$ and observe I_{t_1} . We first propose a heuristic shutdown rule to determine if the decision-maker should shut down at time t_1. Next, rather than using a given time t_1, we use the rule to determine t_1 when the region should shut down and, at this time, what should be the region's shutdown level capacity. We then extend the basic model and present an optimal stopping formulation of the problem.

2.4.3 The shutdown rule

The decision-maker has her preferred estimates C_h, C_e, and T, knows the current rate and volatility parameters r_0, s_0, and estimates that a shutdown will change them to r_1 and s_1. At time t_1 with I_{t_1} infections, should she shut down or not? We note that she is making this decision in a context characterized by a great deal of uncertainty about the pandemic's evolution and difficulty estimating precise cost parameters. Therefore, our first approach is to present a simple heuristic rule that can capture the major drivers of this decision and

analyze their relationships and trade-offs. The rule is as follows.

At a given time t_1 with I_{t_1} infections, a decision-maker will choose to remain open if for all $i \leq I_{t_1}$

$$PrB(i, m; r_1, s_1, T) * C_h + C_e \geq PrB(i, m; r_0, s_0, T) * C_h, \tag{2.4}$$

that is if, $\Delta PrB(i, m; r_0, s_0, r_1, s_1, T) = PrB(i, m; r_0, s_0, T) - PrB(i, m; r_1, s_1, T) \leq \frac{C_e}{C_h}$ and will choose to shut down otherwise.

The heuristic is directly motivated by the underlying trade- off. If a regional decision-maker chooses not to shut down, the higher infection rate r_0 will continue, and the region will incur the health cost C_h with a probability that, given her decision horizon, she will estimate as $PrB(I_{t_1}, m; r_0, s_0, T)$. If she chooses to shut down, the infection rate will decrease to r_1. The region will incur the economy cost C_e and will also incur the health cost C_h with probability $PrB(I_{t_1}, m; r_1, s_1, T)$. The heuristic decision rule suggests that the decision-maker will remain open as long as the shutdown has a higher cost. The rule relies on the assumption that the region starts its observation at a very small number of infections, I_{t_0}. As it observes the number of infections grow from a small number, the rule recommends a shutdown at the first time when the estimated economic cost of shutdown exceeds the health cost of remaining open. We will highlight the limitations of this rule in the next section, but its advantage is clear. It is a simple connection between the decision-maker's economy and health beliefs, and it roots the shutdown decision in the dynamics of how infections evolve.

Next, to make sure that the rule is easily implementable, it will be useful to show that the decision can simply be based on the observed number of infections at the decision time. The following result shows that there exist thresholds on different input parameters (costs, decision horizon, capacity level) such that if they are crossed, the region will not shut down. Otherwise, the rule is well-formed in the sense that there is a unique threshold on the number of infections that would determine the decision.

Proposition 3. *If* $b = \frac{1}{s} \log\left(\frac{m}{I_t}\right)$, $\alpha = \frac{1}{s}(r_1 - \frac{s^2}{2})$ *and* $s_0 = s_1 = s$ *then for a given horizon,* T,

(i) There exists $\widehat{b}$, *such that if* $b > \widehat{b}$, *it is optimal not to shutdown irrespective of the* $\frac{C_e}{C_h}$

preference level.

(ii) There exists threshold $\widehat{\frac{C_e}{C_h}}$*, such that if* $\frac{C_e}{C_h} > \widehat{\frac{C_e}{C_h}}$*, it is optimal not to shut down*

(iii) Given the preference level, $\frac{C_e}{C_h}$*, such that* $0 < \frac{C_e}{C_h} < \widehat{\frac{C_e}{C_h}}$*, there exist* b^**, such that if* $b < b^*$*, it is optimal to shut down.*

(iv) There exists $\widehat{T}$*, such that if* $T < \widehat{T}$*, then it is optimal not to shut down.*

The benefit we gain by shutting down, represented by the ΔPrB function, is the difference in breach probabilities achievable due to a shutdown-induced reduction in the infection rate. As capacity increases, individual breach probabilities may decrease, but the difference first increases then decreases, achieving its maximum when the capacity is neither too low nor too high. Therefore, at some high level of cost ratio $\frac{C_e}{C_h}$, even the peak benefit may be less than the cost. That is why the result states that if the preferred estimate of $\frac{C_e}{C_h}$ is too high, the rule will not recommend a shutdown. An overly optimistic outlook for the availability of a cure or vaccine will cause the decision-maker to estimate a very short horizon T and that too will lead to the rule to not recommend a shutdown. Otherwise, at a given decision time t_1, if the number of infections is more than a threshold I^h (where h refers to the heuristic rule), then the rule recommends a shutdown.

The above result also shows the impact of some parameters on the threshold Ih. The more effective the shutdown is, the less the shutdown-induced infection rate r_1 is. This increases the benefit of the shutdown leading the rule to recommend shutdown at a lower threshold. The proof of Proposition 2 allows us to understand that it is actually the ratio of capacity to the number of infections m/I_{t_1} that determines the threshold behavior. A higher capacity allows for a higher number of infections before shutting down.

Figure 2.2 visualizes the decision rule and the relationship between m and I_{t_1} . The heuristic decision rule trades off the shutdown-induced reduction in the breach probability (due to a lower infection spread rate) against the fixed economy cost. As the number of infections increases, the breach probability with different infection rates increases at different rates. The difference, ΔPrB (the yellow surface in the figure), can be shown to increase first and then decrease. If the capacity or the economy-to-health cost ratio (C_e/C_h, the blue plane in the figure) is too high, ΔPrB never reaches the threshold, and no shutdown is

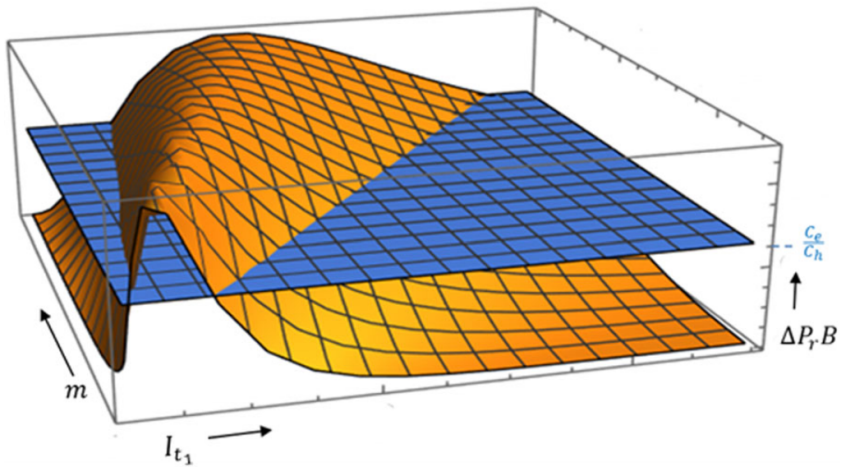


Figure 2.2: Plot by ΔPrB, in m and I_{t_1}, the intersecting plane of C_e/C_h $[T = 3, s_0 = s_1 = 0.4, r_0 = 0.8, r_1 = .6]$

recommended. Otherwise, the rule recommends a shutdown when the difference in breach probabilities hits the cost ratio. As can be seen in the figure, an increase in the number of infections decreases the benefit, which can be compensated by an increase in capacity.

2.4.4 The shutdown level capacity

At this stage, a natural question for the decision-maker is to ask how much capacity to acquire? There was a great shortage of medical equipment like ventilators at the beginning of the pandemic; prices were reported to be temporarily high and variable. In this context, we focus only on the demand side and ask the following question: what should be the regional decision-maker's target capacity at the time she chooses to shut down? In the severely capacity-constrained environment, this can guide the decision-maker's efforts to acquire capacity.

We note that the capacity question is intertwined with the shutdown decision. As the number of infections grows, the decision-maker is working to acquire capacity and also considering a shutdown. Rather than taking the shutdown time t as given, we consider a shutdown time when starting with a low I_{t_0} , the number of infections reaches the threshold

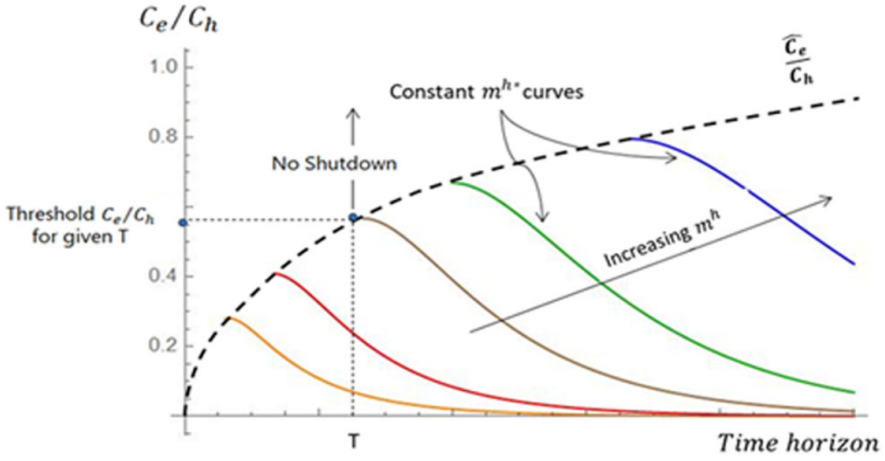


Figure 2.3: Shutdown-level capacity, m^h, as a function of the regional decision-makers preferences $(T, C_e/C_h)$

I from below, and the heuristic rule recommends a shut down for the first time. At this time, the equation $\Delta PrB = C_e/C_h$ is satisfied (we note that a rigorous formulation of stopping time is delayed till the next section.) We are interested in the capacity at the shutdown time, referred to as the shutdown level capacity, $m^h = m(C_e/C_h, T, I_t)$ satisfying $\Delta PrB = C_e/C_h$. This measure of capacity level has practical significance because it represents the capacity a decision-maker will aim to have at the time she decides to shut down. This level, of course, depends on the decision-maker's preferred estimates of C_e/C_h and T. Therefore, we can now directly analyze the influence of the decision-maker's preferences on the capacity level. Figure 2.3 displays this relationship.

Figure 2.3 is instructive as it allows us to map the variety of preferences and beliefs decision-makers have on their shut- down and capacity choices. As discussed earlier, different regional decision-makers bring different preferences to this decision. We organize these preferences along two dimensions: (i) estimate of C_e/C_h, and (ii) estimate of T. The former is an expression of the decision maker's desire not to mandate rules for closing business activity; the stronger is this desire, the higher is her estimate of C_e/C_h. The latter is an expression of

her hopefulness towards discovering a quick solution, a new drug, a vaccine, or just a gradual disappearance of the virus. It is not difficult to find quotes from various state governors expressing their differences of the more optimistic she is, the smaller is her estimate of T opinions regarding the importance they give to keeping the economy open. Nor is it difficult to find differences in governors' opinions regarding how soon the pandemic will go away. Various state officials are on record expressing their differences regarding the importance they give to keeping the economy open and their opinions about how soon the pandemic will end.

The figure shows the shutdown level capacity as a function of these two preferences. For a given T, there is a threshold $\widehat{C_e/C_h}$ value above which the decision-maker will choose not to shut down; this gives the black dashed line in the figure. Regions whose preferences place them at or above this line will not shut down. For example, North Dakota, where the decision-maker put a high value on keeping the economy open, is one of the few states where no stay-at-home order was ever issued. Regions whose preferences intersect below this line can use this figure as a decision aid to determine capacity by noting the solid curve the intersection lies on; each solid curve indicates a specific m^h value (curves for only some selected m^h values are shown in the figure). The purpose of drawing these constant m^h curves is to observe that different preferences can lead to the same capacity level. Following any one of the constant m^h curve shows that an increase in decision horizon requires higher capacity but can be compensated by a decrease in C_e/C_h which would reduce the required capacity.

2.4.5 Per-infection cost: The optimal stopping formulation

In the previous section, we proposed a decision rule heuristic for the regional decision-maker based on the trade-offs she observes. We defended the decision rule approach based on the simplicity it offers in an uncertain environment with hard- to-estimate parameters. A main limitation of the decision rule is that it is limited to considering only fixed costs related to health and economy; there is no per-infection cost charged. Additionally, it takes a myopic view of the shutdown decision in the sense that a decision is made the first time when the

cost of shutting down becomes less than the cost of keeping it open; the option value of waiting under uncertainty is not considered. It also does not offer a rigorous formulation and definition of the stopping (shutdown) time of the underlying stochastic process. To include the per-infection cost and address other issues, we present an optimal stopping formulation of the problem.

The optimal stopping formulation offers a more general form of the problem than the decision rule formulation. Another way to consider the progression from the decision rule to the optimal stopping formulation is to think of the decision rule as capturing early decision-making with less clarity about costs and more heterogeneity in inputs and out- comes and to think of the optimal stopping as capturing later- period, more mature decision-making with additional cost information and more homogeneity in outcomes. Given the complexity of the objective function in the optimal stopping formulation presented below, our goal in this section is relatively limited. We primarily focus on rigorously proving the existence of a stopping time. We believe that the analysis below can serve as a first step towards considering extensions mentioned in a later section.

For the optimal stopping formulation, we assume that $t_0 = 0$, with $I_0 = x$. The total cost function is given by $J^\tau(x)$ where the decision-maker implements a shutdown at an admissible stopping time, τ, which is adapted to the underlying filtration. The number of infections at stopping time τ, I_τ , is a random variable. In the period before the shutdown, a per-infection cost rate $0 < c < C_e$ is incurred. If a shutdown is imple- mented at time τ, we incur a cost equal to $C_e - C_h \Delta PrB$, which includes a fixed economy cost of the shutdown, less the saving in health cost that would be achieved due to shutdown induced lower probability of a breach. The objective is to determine the optimal stopping time (shutdown time) that will minimize the total cost function. The optimal stopping time, τ^*, is the time τ at which the expected total cost $J^\tau(x)$ achieves its minimum. The function $C_e - C_h \Delta PrB$ is referred to as terminal cost and $V(x)$ is the optimal cost function. The formulation is as follows:

$$J^{\tau}(x) = E_x\left[\int_0^{\tau} cI_t(x, r_0, s_0)dt + C_e - C_h\Delta PrB(I_\tau, m; r_0, s_0, r_1, s_1, T)\right]$$

$$\tau^0 := \arg\inf_{\tau\geq 0} J^{\tau}(x), \tag{2.5}$$

$$V(x) = J^{\tau^0}(x) = \inf_{\tau\geq 0} E_x\left[\int_0^{\tau} cI_t(x, r_0, s_0)dt + C_e - C_h\Delta PrB(I_\tau, m; r_0, s_0, r_1, s_1, T)\right] \tag{2.6}$$

With some modifications to address the integration term in the optimal cost function, the above formulation has the form of a classical optimal stopping problem (see, (67)). The solution approach involves using the characteristic function of the underlying stochastic process to set up a differential equation. This equation, along with the value matching and smooth-pasting conditions, leads to a candidate solution $V(x)$ for the optimal cost function and the definition of a continuation region D that specifies the range of x in which the solution proposes not shutting down. We then use the verification theorem (see, e.g., (87)) to show that this candidate function satisfies the requirements to be the optimal cost function. The proof of Proposition 4 shows the above steps and is supported by three lemmas to prove that the verification conditions are satisfied. Finally, this leads us to the following existence result.

Proposition 4. *If* $c > \frac{r_0}{m^{2r_0/{s_0}^2}}$ *then there exists a threshold* $I^0 < m$ *such that it is optimal to shut down if* $I_t > I^0$.

The result shows that under the optimal policy, the continuation region can be written as $D := \{x | x \leq I^o\}$ and the shutdown time is the optimal stopping time defined as $\tau^o := \inf\{t > 0 : I_t \notin D\}$. The proof of Proposition 4 holds for any initial infection level and defines the threshold I^o as the supremum of values where smooth pasting condition is valid. If the infections exceed I^o , the cost of infections will exceed the cost of imposing a shutdown.The role of the assumption $c > \frac{r_0}{m^{2r_0/{s_0}^2}}$ is to ensure that $I^o < m$. In the case of small c, $0 < c < \frac{r_0}{m^{2r_0/{s_0}^2}}$, we still have a finite stopping time and a threshold I^o such that if $I_t > I^o$, it would always be optimal to shut down. This is so because the continuing infection cost increases in the number of infections, but in contrast, the terminal economy

and health costs are bounded and also the underlying geometric Brownian motion has a positive drift. Though there would be no guarantee that $I^o < m$, making the result less insightful in a practical setting.

It is useful to consider the difference between the decision rule and the optimal stopping approaches. Without a per-infection cost, the decision rule looks for the first time that the trade-off between fixed economy and health costs justifies shutdown. With the per-infection cost, the optimal stopping formulation shifts focus to how that cost determines a level at which the influence of the per-infection cost justifies a shutdown. The qualitative behaviors of the optimal stopping problem and the decision rule strategy are fundamentally different. For the decision rule strategy, without the per- infection cost, it is possible that it is never optimal to shut down. But with the optimal stopping problem, with a non-zero per-infection cost, the shutdown should eventually occur with probability 1. This is because the total per-unit infection cost will continue to increase and exceed the combined impact of health and economy costs as long as the latter costs are bounded.

2.5 Compensating for limited capacity

Starting with a decision maker's preferred estimates of costs of health and economy (C_h, C_e), and the level of her optimism about the availability of a cure or vaccine (T), the previous section analyzed the shutdown level of capacity. In a capacity-constrained capacity environment, the desired level of capacity may not be available. The objective would then be to reduce the probability of breach. Therefore, in this section, we analyze other actions a decision-maker can take to reduce the breach probability if the medical equipment capacity is constrained to a given value.

2.5.1 Reusability

We start with a simple observation regarding the impact of reusability of the capacity on the breach probability. See, for example, (80) for ventilator sharing. Recall that the capacity m was introduced in terms of the maximum number of infections that can be simultaneously treated. For example, let V represent the ventilator capacity and let the fraction α (40)

represent the fraction of infected persons who will need the equipment. Then, $m = \frac{V}{\alpha}$. Any reuse decreases α and, for a fixed V, increases m. As a result, the breach probability decreases.

2.5.2 Shutdown intensity r_1

Next, we discuss the impact of r_1, the infection spread rate in the shutdown period. The decision-maker may have a range of mandated behaviors as possible options, from stay-at-home orders to mask orders, which translate into different degrees of reduction in the infection rate post shutdown. For example, (36) lists recommended strategies ranging from self-isolation, ban on public events, social distancing, school closures, and lockdown with the corresponding percentage decrease in transmissibility from 5% to 45%. Our model can predict the impact of such reduction on the breach probability. Part (a) of the following result formalizes this observation.

Proposition 5. *Assuming* $r_1 - \frac{(s_1)^2}{2} \geq 0$ *and* $I_t \leq m$

(a) $\frac{\partial PrB(I_t,T;r_1,s_1,m)}{\partial r_1} \geq 0$ *(b)* $\frac{\partial PrB(I_t,T;r_1,s_1,m)}{\partial I_t} \geq 0$

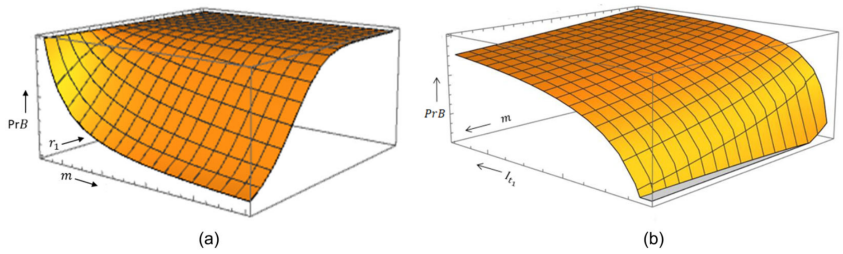


Figure 2.4: Variation in breach probability with variation in preparation level (m), rate of spread during shutdown (r_1) and shutdown level (I_{t_1})[(a) $T = 5, s_1 = 1, I_{t_1} = 10; (b)$ $T = 5, r_1 = 1.2, s_1 = 1$]

2.5.3 Shutdown level I_{t_1} or, equivalently timing t_1

The previous section proposed a heuristic rule and an optimal stopping formulation to determine the shutdown level, but if the region does not have the capacity it needs, it can

choose to deviate from the prescribed shutdown level. To formalize this argument, we first look at the impact of the shutdown level on the breach probability. Part (b) of the result shows that the breach probability increases as the shutdown level infections increase.

These observations allow us to point out substitutability between the capacity and the shutdown infection rate and shutdown level. As Figure 2.4 shows, a lack of capacity increases the breach probability, but it can be compensated by taking action to increase shutdown intensity. The figure also shows that any decrease in the breach probability due to lack of capacity can be compensated by starting the shut- down early at a smaller number of infections. These strategies can be observed in real contexts as well. For example, by some measures of available bed capacity (adjusted for size), states like Oregon and Washington are the ones with the tightest capacity, while Texas and Minnesota come out at the top (50). Both Washington and Oregon were the earliest states to issue stay-at-home orders, and Texas was one of the last ones (see (61)).

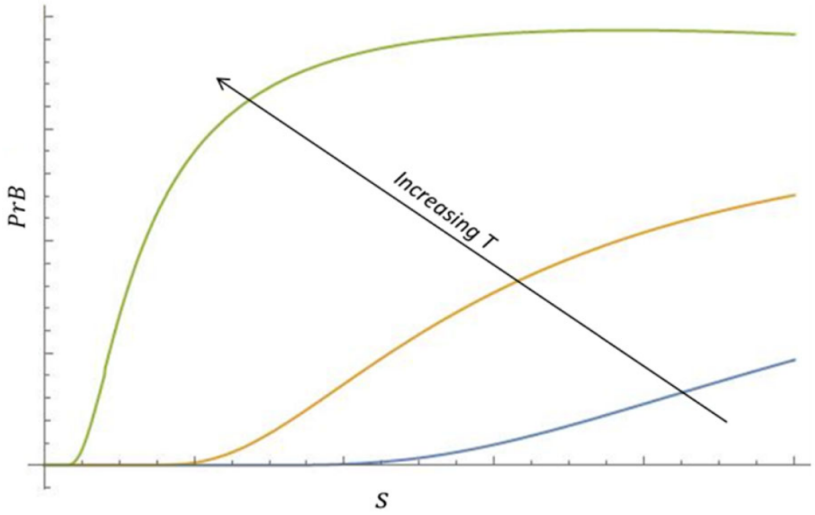


Figure 2.5: Variation in breach probability with volatility for short time horizon [$m = 50, I_{t_1} = 10, r_1 = 1.5, T = \{0.5, 1.0, 1.5\}$]

2.5.4 Shutdown enforcement s_1

Recalling our discussion for the introduction of volatility parameter in an earlier section, we note that volatility exists due to variations in how people behave in their interactions with each other. Before the shutdown, people follow their normal behavior patterns; some interact more and others less, with each interaction a possible source of infection. After the shutdown, the variation in people's behavior may be influenced by the strength of shutdown messaging and enforcement of its rules. Improvements in testing capacity and accuracy may also reduce noise in the system. We interpret these as reductions in volatility. Its impact on the breach probability can be complex. The intuition suggests that reduction in variability should benefit the performance. As Figure 2.5 shows, this is indeed the case. But, in cases with long time horizons, it can also increase the breach probability.

This section considered strategies that a region can implement on its own. In the next section, we consider strategies that require coordination across regions.

2.6 Coordination across regions

A major theme in reporting about the early spread of the disease in the United States is the lack of coordination between different states and regions. In the absence of coordination, many states and regions had closed down their borders, charted their own communication strategies, and relied on their own medical equipment capacity (34). This description fits closely with the model we have developed in the previous section. We will refer to the previous model as the individual-region model.

There are different ways we can conceive coordination between individual regions in order to improve outcomes. In the United States, a lack of coordination on acquiring and managing the ventilator capacity came in for early criticism. Later, several states changed their policies to make their beds and ventilators available to both neighboring and non-adjacent states ((63); (70)). A natural question to ask is about the improvement in outcomes that can be achieved by combining two individual regions' capacity. When two regions agree to make their total medical equipment capacity available to treat patients from either region, we refer to it as medical equipment capacity-sharing. It appears intuitively clear that such

a sharing of capacity should be better for the regions. Our model offers a direct way to check this intuition; does this always hold true? We address this question as the first step in this section.

Our interest, however, is not just in showing the existence of such capacity-sharing improvement. We are also interested in identifying and investigating other features of regional coordination that can influence the benefit of capacity sharing. As we better understand the drivers of this value, we aim to show that other features of regional coordination, such as information and movement coordination, can have either a positive or negative influence on the benefit of capacity sharing. Later sub-sections address such features.

2.6.1 Medical equipment capacity sharing in two independent regions

Let us identify two regions by assigning a superscript index k to each relevant parameter where $k = a, b$, with each index value representing one region. The regions are assumed to be independent in the sense that the infections evolve independently in each region. The next section considers the case of interdependent regions. We will focus on the dynamics of infections in the shutdown periods $t \geq t_1$ in each region, assuming a shutdown is implemented at t_1. Recall that in the individual region model discussed earlier, $\widehat{I^k_{t_1}}$ represents the t_1 maximum number of infections in $[t_1, t_1+T]$ and the capacity m^k represents the number of simultaneous infections that the existing medical equipment capacity can treat. The health cost is incurred if the maximum number of infections exceeds the capacity, an event we refer to as a breach. To keep the focus on the operational performance, we consider the impact of capacity sharing on the breach probability represented by $Pr[\widehat{I^a_{t_1} + I^b_{t_1}} > m^a + m^b]$in the capacity-sharing case and by $Pr[\widehat{I^k_{t_1}} > m^k]$, $k = a, b$ in the individual-region case.

Proposition 6. $Pr[\widehat{I^a_{t_1} + I^b_{t_1}} > m^a + m^b] \leq Pr[\widehat{I^a_{t_1}} > m^a] + Pr[\widehat{I^b_{t_1}} > m^b]$.

The result states that the maximum of the sum of the two regions' infections is less likely to exceed the combined capacity than the sum of individual regions' probabilities of exceeding their own capacities. The result confirms intuition and is also general in the sense that it does not depend on the relative sizes, infection rates, and capacities of the two regions. There is less risk of breaching capacity from the combined region's perspective

by combining medical equipment capacity. However, if we focus on the health cost as our metric, the benefit is less clear. It would depend on how we estimate each region's health cost in the capacity-sharing case. From an individual region's perspective, the benefit is less clear as it depends on the other region's size and capacity.

To investigate this further, we consider two identical and independent regions. That is, they shut down at the same time t_1 with an equal number of infections $I^k_{t_1}$ and equal preparation levels m^k, and that they have the same infection rate and volatility after shutdown. This allows us to focus only on the impact of sharing capacity, unmediated by any other differences. From each region's perspective, do we expect to see the breach probability always decrease if the two regions share a capacity? Interestingly, the next result suggests not always.

Proposition 7. *For two identical regions, there exists a preparation level threshold* $\overline{m_1}$*, such that* $P(\widehat{I^a_t + I^b_t} > m + m) < P(\widehat{I^b_t} > m) = P(\widehat{I^a_t} > m)$ *if* $m > \overline{m_1}$.

Note that even if the two regions have identical parameters, they have different Wiener processes in their evolution equations, and the number of infections evolves independently. The result is analytically limited to identical regions, but it is stronger than the previous result because each region will independently see benefits, and the total system health cost will be less. However, this is only true if the medical equipment capacity is high enough. If the capacity is tight, it is possible that by sharing capacity with another identical region, one would see an increase in the breach probability. Why is this so? An explanation will also help us understand the drivers of this benefit. We discuss this below.

The proof of the above result is based on the consideration of an intermediate system. This system focuses on the sum of maximum infections in the two regions, $\widehat{I^b_{t_1}} + \widehat{I^b_{t_1}}$, and corresponding breach probability $Pr[\widehat{I^b_{t_1}} + \widehat{I^b_{t_1}} > m + m]$. We note that this sum, $\widehat{I^a_{t_1}} + \widehat{I^b_{t_1}}$, does not necessarily represent the true requirement, $\widehat{I^a_{t_1} + I^b_{t_1}}$, of medical equipment at any time because the chance that two independent regions will experience the maximum need simultaneously is vanishingly small. As we will discuss in the next section, interdependence between regions will bring $\widehat{I^a_{t_1} + I^b_{t_1}}$ closer to $\widehat{I^a_{t_1}} + \widehat{I^b_{t_1}}$. The consideration of the sum of the maximums and the breach probability in the intermediate system provides a bridge to un-

derstand the breach probability we are really interested in, the maximum of the sum versus total capacity in the shared- capacity case. The proof first shows that the probability of $\widehat{I^a_{t_1}} + \widehat{I^b_{t_1}}$ breaching the total capacity is less than the probability of each region breaching its own capacity if and only if the capacity is above a threshold. The proof then shows that the probability of $\widehat{I^a_{t_1} + I^b_{t_1}}$ breaching the total capacity will be even less.

Our construction of this intermediate-system measure, $\widehat{I^a_{t_1}} + \widehat{I^b_{t_1}}$, not just helps with the above proof,it also helps build an intuitive explanation of the drivers of the benefits of capacity sharing, an explanation that will lay a foundation for the next section. We find it helpful to differentiate between two drivers of the capacity-sharing benefit. First, there is the traditional risk-pooling benefit. At any given instant, the sum of two random variables representing independent regional infections is less variable than individual regions' infections. One would expect it to drive some part of the sharing benefits. That is represented by the difference between $\widehat{I^a_{t_1}}$ or $\widehat{I^b_{t_1}}$ and $\widehat{I^a_{t_1} + I^b_{t_1}}$. Second, there is the peak-timing benefit. Given that each region's infections evolve independently, the time difference between the two regions' peaks should benefit the deployment and use of the limited total equipment capacity.That is represented by the difference between $\widehat{I^a_{t_1}} + \widehat{I^b_{t_1}}$ and $\widehat{I^a_{t_1} + I^b_{t_1}}$. We will use these two drivers, risk pooling and peak timing, to discuss our next section's observations. The discussion of peak-timing as a source of capacity-sharing benefits adds new insight to the literature.

However, why may capacity sharing not be beneficial if each region's capacity is below a threshold? If an individual region's capacity is so small that $\widehat{I^a_{t_1}}$(or $\widehat{I^b_{t_1}}$) is very likely to cross it, then $\widehat{I^a_{t_1}} + \widehat{I^b_{t_1}}$ is also very likely to cross the sum of the two capacities. If so, there should not be much value in sharing capacity. With a larger capacity, it becomes possible that one region crosses its capacity, but the other does not. Such cases open up the possibility that total infections may still be below the combined capacity. Both risk-pooling and peak-timing play a role in this phenomenon.

2.6.2 Evolution of infections with information and movement coordination

We are now ready to consider other aspects of coordination across regions that can promote interdependence between the two regions' infection evolutions. Specifically, we analyze two different features of regional coordination. First, inter-regional coordination may allow movement across the physical boundaries. We label such physical coordination leakage and seek to understand its impact on capacity sharing. Second, coordination may allow regions to exercise uniformity in communication strategies. We label this practice information coordination. We propose a way to incorporate it into our model and ask how this may influence capacity sharing benefits. Later, we briefly discuss another feature, shutdown coordination, where regions may synchronize their shutdown policies. This section models the evolution of infections with these new features, and the next section considers the impact of such interdependent evolution on the benefit of capacity sharing.

Unlike our discussion of capacity sharing in the previous section, where each region's infections evolved independently, the coordination features we describe in this section require us to model the interdependence between how regional infections evolve. To that end, we first propose a new dynamics of infection evolution that links the two regions. We introduce two new parameters: a correlation index ρ, and a leakage parameter l. They are explained in detail right after the dynamics presented below. Assuming a shutdown at t_1, for $t \geq t_1$:

$$dI_t^a = r_1^a((1-l)I_t^a + lI_t^b)dt + s_1^a I_t^a dW_t^a \tag{2.7}$$

$$dI_t^b = r_1^b((1-l)I_t^b + lI_t^a)dt + s_1^b I_t^b dW_t^b \tag{2.8}$$

where W_t^a and W_t^b are correlated Brownian motions. That is, $[W^a, W^b]_t = \rho t$, $\rho \in (-1, 1)$. To model physical coordination between different regions such that the travel between the regions is permitted but the travel from outside to either region is not, we use the leakage parameter l. Physical coordination across the two regions allows some movement across the borders, thus enlarging the area that can be insulated from outside influences. Such a policy has its advantages in providing the residents with a larger area for movement, and thus lessening the pressure brought on by the shutdown orders. For example, such physical coordination existed between states like New York and New Jersey in the early days of the pandemic. As seen in the equations above, a fraction l of the current infected population in

one region may leak to the other region. Once there, this infected fraction will contribute to the other regions spread.

To model information coordination, we focus on the sources of noise in the infection rate. Each region has its idiosyncratic character that drives the variation in the range of people's social behavior. The lack of testing capacity and inconsistent data reporting also introduce uncertainty. The Wiener process captures this variation. With information coordination, regions implement coordinated messaging and uniform standards of data capture and reporting. Initially, the lack of such coordination was heavily criticized. Later, multi-state collaborations were launched to coordinate data accuracy; see, for example, (15). The evidence of heterogeneity in health-related social behavior is available in (5). The connection between consistent messaging and uniformity in social behavior is discussed in (39). In this framing, coordination of information and messaging across two regions will induce consistency in people's social behavior. We model such coordination as an increase in the correlation between the two regions' Wiener processes.

Since we are proposing a new system of SDEs to model the spread of infections due to interactions between two regions, it is important to derive the conditions under which there exists a unique positive solution to the coupled equations. We proceed to do so in the following result.

Proposition 8. *Given that r^a, r^b, s^a, s^b and $l \in [0,1]$ are known positive constants, W_t^a and W_t^b are correlated Wiener processes such that $[W^a, W^b]_t = \rho t$, where $\rho \in (-1,1)$, and I_0^a and I_0^b are the initial points, the coupled equations*

$$\begin{aligned} dI_t^a &= r^a((1-l)I_t^a + lI_t^b)dt + s^a I_t^a dW_t^a \qquad I_0^a \\ dI_t^b &= r^b(lI_t^a + (1-l)I_t^b)dt + s^b I_t^b dW_t^b \qquad I_0^b \end{aligned}$$

has a unique positive solution if

$$max\left\{r^a(1-l) - r^b(1-l) + \frac{(s^b)^2}{2}, -r^a(1-l) + r^b(1-l) + \frac{(s^a)^2}{2}\right\} - \frac{\rho s^a s^b}{2} + l(r^a + r^b) > 0$$

We show the existence of a unique solution using the at- most linear growth and Lipschitz continuity condition. The proof of the positivity of the solution is more involved. We create

a sequence of stopping times, which are the first hit- ting times of a sequence of boundaries very close to 0. We then show that as the boundaries tend toward 0, the stopping time tends to infinity. The detailed proof is presented in the Appendix in Supporting Information. The literature on SDEs to model the spread of infections is limited. We believe that with good parameter estimation, the above equations can be used to model the early spread of infections.

Because the set of equations above does not yield an analytical solution, we present further insights based on the above system's simulation analysis. The analysis is carried out on a testbed of parameter ranges. The observations reported in the next section hold true for this testbed. For clarity, we present graphs for specific parameter combinations, which are mentioned with each graph.

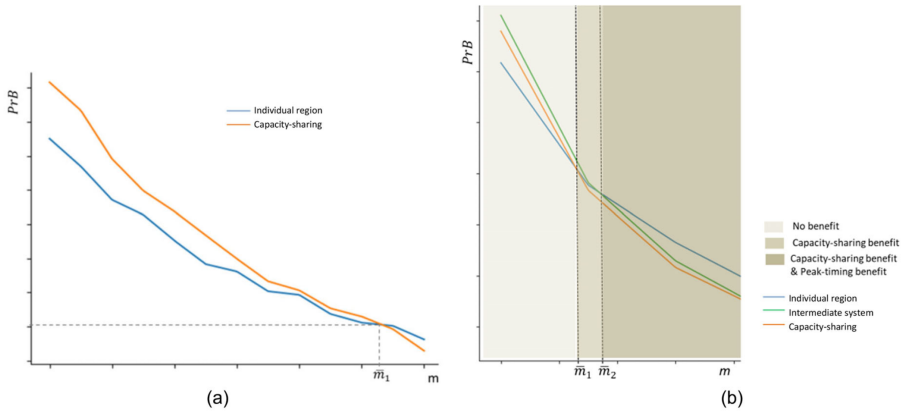


Figure 2.6: (a) Exitence of a threshold, $\overline{m_1}$, such that capacity coordination is beneficiary only if $m > \overline{m_1}$. (b) Three zones of no benefit, only risk-pooling benefit, and both risk-pooling and peak-timing benefit. $[r_1^a = r_1^b = 1.2, s_1^a = s_1^b = .5], I_{t_1}^a = I_{t_1}^b = 2, \max m = 2000, \min m = 800]$

2.6.3 Capacity sharing with information and movement coordination

Before analyzing the impact of regional interdependence, driven by information and movement coordination, on capacity sharing, we first use the simulation to set a baseline by

revisiting Section 6.1 where there was no movement coordination (zero leakage) and no information coordination (zero correlation). Figure 2.6(a) shows that as capacity increases, the breach probability for both the individual region case, $Pr[\widehat{I^a_{t_1}} > m]$, and the shared capacity case, $Pr[\widehat{I^a_{t_1} + I^b_{t_1}} > 2m]$, decreases as we expect them to. There is, however, a unique threshold capacity level below in which sharing capacity does not reduce breach probability. Benefits of capacity-sharing occur only if the capacity is above a threshold; this is in line with Proposition 7. Figure 2.6 (b) digs deeper into this by adding the breach probability for the intermediate system, $Pr[\widehat{I^a_{t_1}} + \widehat{I^b_{t_1}} > 2m]$. When capacity is above the threshold, the graph can differentiate between the risk-pooling effect and the peak timing effect as discussed below Proposition 7.

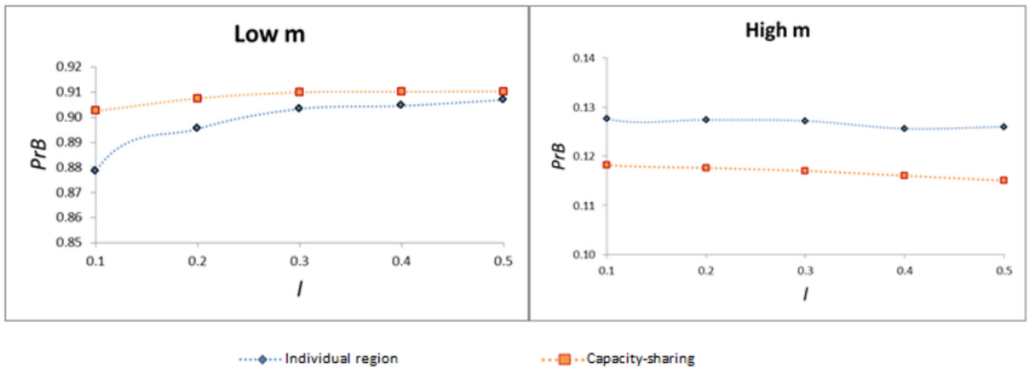


Figure 2.7: Variation in breach probability with leakage under capacity coordination and no capacity coordination. The graphs are generated under the scenarios of high and low individual capacity.$[r^a_1 = r^b_1 = 1.2, s^a_1 = s^b_1 = .5, I^a_{t_1} = I^b_{t_1} = 2, \rho = 0, high\ m = 1500, low\ m = 200]$

2.6.4 Impact of physical coordination

Physical coordination across the two regions allows move- ment across the borders increasing the leakage l and inducing interdependence between two regions' evolutions. Our focus is on its impact on the breach probability. We note that unlike Figure 2.6 and Section 6.1, the following discussion computes, $\widehat{I^a_{t_1}}$, $\widehat{I^b_{t_1}}$ and $\widehat{I^a_{t_1} + I^b_{t_1}}$, based on simulation of the coupled SDEs presented in Section 6.2.

While the existence of a capacity threshold mentioned in Proposition 7 was proved for the independent region case, it is notable that a similar notion holds in the case of interdependent regions. Figure 2.7 shows that capacity sharing improves breach probability for the high-capacity case but not for the low-capacity case. When capacity is high, the figure shows that the breach probability in both individual-region and capacity-sharing cases is relatively indifferent to increasing leakage, but when capacity is tight, increasing leakage increases breach probability in the capacity-sharing case. At any given time, the SDEs with leakage replace the individual regions' number of infections with a combination of both regions' infections. This synchronizes the two trajectories bringing their peaks closer. This peak-timing effect then leads to an increase in breach probability when capacity is shared. The sample trajectories in Figure 2.8 show a visual example of how an increase in leakage can bring the two regions' trajectories together. If we assume increased travel to be a proxy for increased leakage then this suggests inverse outcomes for an increase in travel, an observation similar to that made in some empirical work (see, e.g., (51)).

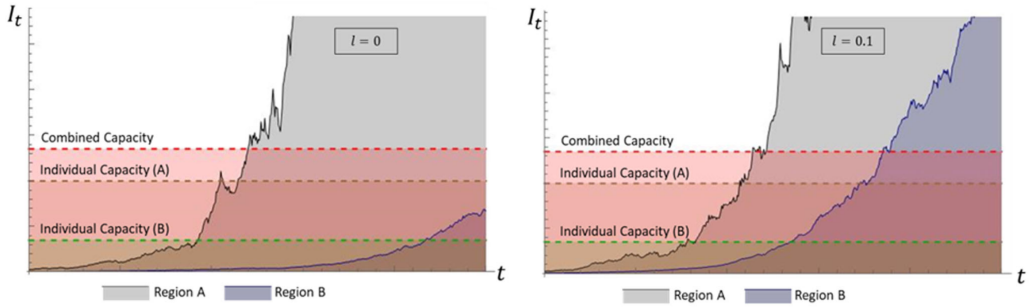


Figure 2.8: Difference in the spread of infection in region B due to highly infected region A. Left: Leakage $l = 0$ signifying closed borders; Right: Leakage $l = 0.1$ signifying some interactions

2.6.5 Impact of information coordination

Information coordination across the two regions means coordination on messaging and communications transmitted to populations in two regions. We interpret this as an increase in

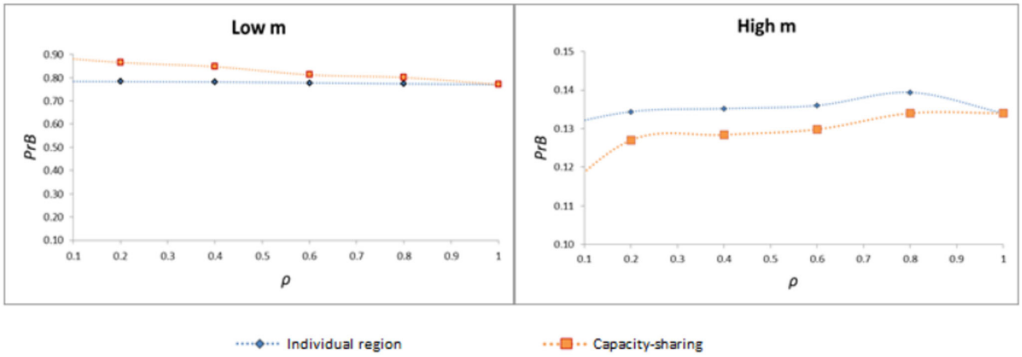


Figure 2.9: Variation in breach probability with correlation under capacity coordination and no capacity coordination. The graphs are generated under the scenarios of high and low individual capacity.[$r_1^a = r_1^b = 1.2, s_1^a = s_1^b = .5, I_{t_1}^a = I_{t_1}^b = 2, \rho = 0, high\ m = 1500, low\ m = 200$]

correlation between the two Wiener processes underlying the spread of infections in the regions. We consider the impact of the increasing correlation on the breach probability. Once again, Figure 2.9 supports the existence of a capacity threshold even with interdependent regions; capacity-sharing reduces breach probability only with high capacity. The figure shows that increasing correlation does not have a measurable influence on the individual region's breach probability. As the SDEs show, combining two Wiener processes still means the same degree of noise for the individual regions. The figure further shows that with high capacity, increasing correlation brings the individual-region and capacity-sharing breach probabilities closer, thus reducing the benefit derived by capacity-sharing. This too can be explained by the peak timing effect. The larger the correlation, the more likely it is that peaks occur close to each other, driving up the breach probability for capacity-sharing.

2.6.6 Shutdown coordination

We close this section with an observation related to the coordination of shutdown timing. In our analysis till now, we have assumed that the two regions' shutdown levels are the same and the shutdowns at the same time. However, given the random evolution of infection in

the early spread period, the two shutdown times will not be the same. This would be a further reason for the peaks to occur at different times. Therefore, the peak timing effect will drive the breach probability down even further. Conversely, we expect coordination on shutdown timing to decrease the breach probability under capacity sharing.

To summarize, this section shows that medical equipment capacity sharing across two regions can be beneficial but not when capacity is very tight. Allowing movement across the border, interestingly, can smooth the infection trajectories and offer further benefits. Any coordination that increases the correlation between the underlying random processes can bring peaks in sync and reduce the benefits. We note that these conclusions say nothing about other benefits regional coordination can bring. However, our results suggest that regions should assess the benefit of capacity sharing in light of other elements, such as leakage and correlation, included in the coordination agreement.

2.7 Discussion of extension and conclusion

We believe that a significant part of the paper's contribution stems from being the first to apply SDE-based modeling to pandemic-related capacity. The model can capture the uncertainty in infection spread and yet remain tractable enough to address capacity questions. Being early in this research stream, we have mainly focused on building a model that captures this situation's essential features and asks fundamental questions. However, other questions remain, raising the possibility of a variety of extensions. We discuss them in this section.

2.7.1 Additional details in the infection-spread dynamics

Our focus on the early-stage dynamics led to the SDE model we used here. The main modeling goal is to produce a trajectory that can capture an early exponential increase and then follow it up with a variety of possibilities leading to a plateau, long-term decrease, or resurgence. One extension would be to use time-dependent rate and volatility parameters in the geometric Brownian motion. Depending on how these parameters change, it is possible to produce a variety of trajectories. Such a model can better fit the wide range of long-term infection-spread patterns we have observed across US states and other regions.

Another extension would be to bring the infection-spread SDE closer to the epidemiological literature by modeling additional compartments. Specifically, we can include the susceptible compartment in the model where the sum of the numbers of susceptible people and infected people equals a fixed population size.

2.7.2 Additional details in performance measures

We have focused on the breach probability as our primary performance measure. There are, however, other performance measures that directly depend on medical equipment capacity. The total number of infections over the horizon or the cumulative number of infections for which capacity is unavailable are important performance measures to compute. On the economy side, the region's size and the shutdown length will be important determinants of the total cost. An extension can include these considerations in the optimal stopping problem.

Combining the previous two extensions, modeling long- term trajectories and additional performance metrics, would also allow us to include other decisions in the optimal stop- ping problem. The most important example is the reopening decision. The reopening decision will depend on how the infections are likely to spread after reopening and its implications for health and the economy.

2.7.3 Supply-side model

We limited the scope of our model to modeling the demand- side impact of capacity. In our judgment, that is the most relevant question in a constrained capacity environment. However, this leaves open the question of how this capacity is acquired. The most straightforward extension would be to include a per-unit capacity cost in the model. The reality, however, is more complex. Given the severe shortages, each region is competing against others to acquire limited capacity. Thus, an extension with a competitive supply model would be more appropriate.

2.7.4 Additional details in the coordination model

The previous point also offers a reason to extend our analysis of regional coordination. While we have focused on the coordination of existing capacity, the coordination on acquiring new capacity can be a worthy extension. This extension would be akin to the development of cooperative game models for distributing coordination benefits in newsvendor model setting; such models followed the earlier literature showing the benefits of pooling coordination. Other extensions for the coordination model would include consideration of non-identical regions and shutdown policies.

2.7.5 Testable propositions

Our theoretical model delivers conclusions that relate differences in decision-makers' preferences (about the relative costs of economy and health and expectation of a cure) to differences in their shutdown policies. Another conclusion we draw relates the value of capacity sharing to the tightness of capacity and the extent of movement and messaging coordination across neighboring regions. An extension that tests these theoretical conclusions against actual observations in the US states and other countries would be a useful exercise. We conclude by reciting some of the main themes in the paper. This paper is a first attempt to explicitly include infection rate uncertainty in the pandemic modeling while maintaining tractability to analyze the impact of limited capacity on the region's health performance. We develop an SDE model of the spread of infection in the early stages of a pandemic, analyze its running maximum process, and use it to compute the required medical equipment capacity. The performance measure of our choice is the breach probability.

Our results relate regional decision-makers' preferred estimates of economy and health costs and time-till-cure to the shutdown policy, the level of infections at which the region should shut down, and the corresponding medical equipment capacity. Given the shortage of capacity, we analyze a list of options to compensate for it. These options include reusability of the equipment, shutdown intensity, earlier shut down, and reduction in the messaging noise. The model analyzes the impact of these actions on the breach probability. We then consider sharing capacity across regions as a way to influence the breach probability. We show that

shared capacity may not benefit individual regions if the capacity is too tight. Movement across borders decreases the breach probability, and an increasing correlation between the random processes underlying two regions' infections increases the breach probability. As discussed above, we offer several ways to extend the model.

Chapter 3

ROOFTOP AND COMMUNITY SOLAR ADOPTION WITH INCOME HETEROGENEITY

3.1 Introduction.

Rooftop solar panels are a major component of governments' renewable energy strategies. State and local governments worldwide offer subsidy programs to promote the adoption of solar panels among households. The earliest such laws in the U.S. date to the 1980s (89). The popularity of tools like Google Project Sunroof shows widespread awareness and interest in household rooftop solar adoptions. As such programs proliferate and mature, new concerns about their impact and effectiveness have emerged. In addition to tracking how many households adopt rooftop solar, the metrics of interest have broadened to include how fast such adoptions occur and how well-off these adopters are. Such concerns about timing and equity raise interesting questions regarding the design of subsidy programs and how households of different income profiles respond to them (93).

With global warming's effects ever more visible, there is an increasing pressure on government planners to specify time targets for the adoption-quantity goals they set. According to the National Conference of State Legislatures (64), more than half of American states have set renewable energy targets, with ten aiming to achieve full conversion with deadlines ranging between 2030 and 2050. Some regions have set specific targets for solar energy; for instance, Washington D.C. aims at a 5% solar contribution in its complete transition to cleaner resources by 2032. The industry has the same focus; the Solar Energy Industries Association aims to reach 30% of U.S. electricity generation by 2030 (74).

The concerns about equity are also gaining prominence. As activists point out, one problem with the currently implemented solar subsidy policies is that they benefit the higher-income households more than the low-income households. For example, the federal tax credit for solar panel installation cannot be claimed by a household that rents an

apartment and cannot put a solar array on the roof. At the same time, it is easier for a homeowner to install a solar array and qualify for the tax credit (27). There is evidence of disparity in solar adoption based on income level, race, and ethnicity (37), (83). Generally, solar adopters tend to be from higher-income households living in neighborhoods with a relatively higher proportion of non-Hispanic white and Asian populations (13).

The recent emergence of community or subscription solar products may be able to address the concerns about speed and equity raised above. These products allow households to lease or subscribe to a portion of shared solar systems located away from homes, thus removing the disadvantages of high upfront cost and space constraints in rooftop solar installation. As of 2021, there are about 1600 community solar projects in 39 states, and 22 states have enacted community solar legislation (48). Many states, including Delaware, New Hampshire, Oregon, and Minnesota, are crediting surplus community solar generation at retail price (30), the same way they support rooftop solar products. The availability of rooftop solar and community solar products in a region gives the households a flexibility in payment schemes which can hasten their adoption decision. Thus, community and subscription solar products may provide planners with an additional resource to promote the adoption of solar technology.

This paper aims to develop a model that can incorporate important but previously unaddressed features discussed above - the timing of adoption, the impact of income inequality, and the availability of community solar - while maintaining tractability. Models that analyze the adoption timing of a new technology are an integral part of Operations Management(OM) research, but there has been limited work on their application to solar products. In addition, the interplay between household adoption timing decisions and the planners' time targets has not yet been explored in the literature. The modeling of income inequality also remains an open question in the OM literature; we take an early step in that direction. Finally, we believe that this is the first effort to model a household's choice between two different solar products, rooftop and community solar. In the following, we describe the model, the types of questions we ask, and the insights we draw from it.

We model a regional population where the households differ in income and electricity consumption. Drawing on energy economics literature, we argue that income diversity

correlates with differences in electricity demand and determines a household's propensity to invest in renewable energy. More specifically, a higher-income household consumes more electricity at a faster growth rate and, financially, finds it easier to make a fixed investment; we capture this by modeling a lower discount rate for a higher-income household. In other words, a lower-income household uses a higher discount rate to value the future benefits of renewable energy. See Section 3.3 for detailed support of these modeling choices.

A household's electricity demand is modeled as a continuous time stochastic process. Each household solves an optimal stopping problem to decide when to adopt a solar product, either rooftop or community solar. The household incurs an electricity consumption cost before adoption; must spend a fixed amount for rooftop solar, or pay a subscription fee for community solar adoption; and enjoys a net-metering billing system after adoption where it can sell any excess capacity back to the grid. Our modeling choices are informed by practice and earlier literature and are further explained in Section 3.3.

The central planner (representing the regional government) influences a household's decision by providing solar subsidies, modeled as percentage discounts applied to the adoption costs. The planner's problem is to minimize the total subsidy cost with a constraint to ensure that the adoption level target is achieved within the adoption time target. We model the central planner's problem as a bi-level optimization where the planner sets the subsidies with full knowledge of the household's problem, and then the households make their decisions about adoption timing and product choice.

In addition to the new modeling features mentioned above (timing of solar adoption, household income inequality, and solar product choice), we consider the following to be the paper's technical contributions in solving the household's problem and the planner's problem described above. (1) We offer an explicit characterization of a household's product choice decision by determining a threshold income level that divides the household population distribution into two parts: those who will choose rooftop solar and those who will opt for subscription solar. (2) We develop closed-form expressions for a household's adoption timing density function. (3) For the central planner's bilevel optimization formulation, we prove that it is a convex problem in the special case of equal subsidies, and for the general case, we prove the properties of the optimal solution and offer an algorithm to determine it. We

view the formulation and solution of a bilevel optimization problem where the lower-level problem is based on an optimal stopping problem for a continuous time stochastic process as a new contribution to the OM modeling literature.

The solutions we develop for the household's and planner's problems allow us to investigate the interaction between the planner's policy decisions and a household's responses to them. Based on numerical explorations, we contribute the following observations. (1) To set goals, a central planner must consider three factors simultaneously: adoption level target, adoption time target, and subsidy budget. A shorter target time will require a higher subsidy budget, as will a higher target adoption level. For a given subsidy budget, adoption level and adoption time can be substitutes for each other. (2) Community solar is a useful tool for planners to achieve targets at a lower cost. While it is intuitive to suggest that adding additional product choices to the rooftop solar will make the planner's problem easier, the impact of community solar goes deeper. This is so because community solar and rooftop solar attract households from the opposite ends of the income spectrum. The central planner can differentiate the subsidies it offers for both products to help it achieve its target in a timely manner. (3) Increasing income inequality in a region, as measured by its Gini coefficient, influences the adoption decision across the population. We show that the availability of community solar as a choice can help speed the adoption process; a low subsidy level with two products can deliver faster adoption than a high subsidy level with only one product.

The rest of this paper is organized as follows. In Section 3.2, we review the related literature. We present the notation and assumptions and then formulate the problem in Section 3.3. In Section 3.4, we first solve a special case of our model and then numerically solve the general model. We discuss the implications of the results and possible extensions of the model in Section 3.5 and Section 3.6 respectively. The proofs are deferred to the online appendix.

3.2 Literature Review

The emergence of distributed solar energy production and government incentives to promote it has raised interesting research questions. We start with summarizing this literature. Most

of these papers, however, deploy static models to address questions they are interested in. Our focus is on adding a time dimension, which requires us to develop a dynamic model of adoption decisions. There is limited work on developing dynamic models in the context of solar energy. Below, we describe how our model builds on the existing literature and then adds to it. Finally, our methodology has parallels in the real options literature. We close this section by highlighting these connections.

An early question in the literature was about the design of government subsidies to incentivize adoption by users. (19) analyzed the optimal rate of compensation for excess electricity generation in a distributed generation system when electricity generated is stochastic. (82) included the wholesale market dynamics in the analysis of the benefits of the net-metering policy to the utilities assuming random electricity demand. (79) studied the optimal tariff structure for the utilities to achieve certain welfare goals. (2) investigated the economic and environmental implications of non-ownership business models by endogenizing a solar power company's business model decisions. These papers use static models to develop their insights.

The literature on dynamic models in renewable energy operations is limited. (10) model electricity prices and investment cost as diffusion processes and compares the effectiveness of two different subsidy policies, feed-in tariff, and tax rebate policies. (9) study when and how much capacity a single customer should install with an uncertain demand process to minimize future costs. (4) offers a periodic model with information diffusion and learning over time with a goal to determine changes in the government's incentive over time. Our focus on the dynamic nature of the problem is for a different purpose. We analyze how individuals' adoption timing decisions interact with the planner's horizon for achieving the region's renewable energy goals. In addition, our model incorporates features like product choice and income inequality. Our focus on these new features has required us to simplify some of the other features available in some of the above papers, most notably the modeling of utilities and other intermediaries between the central planner and the electricity customer. We discuss such simplifications and assumptions after we present our model in the next section.

Our research also draws on real option valuation literature. A detailed exposition on

real options valuation is available in (26), (77), (78) and (67). In OM literature, techniques from real options valuation have been used to address investment timing, size, and choice decisions in various domains ((24), (54), (84), (55), (9)). In our work, we solve an optimal investment timing sub-problem with a choice of investment and use the results to solve a non-linear optimization problem in a bilevel formulation.

3.3 Problem and Model Formulation

We develop an optimal stopping formulation for a household's decision to adopt a solar product from a set of available choices. The households are heterogeneous in their discount rates and demand rates. They weigh the cost of adoption against future savings to determine which product to adopt and when to adopt it. Section 3.3.1 presents how we model household heterogeneity, and Section 3.3.2 lists the cost parameters. Section 3.3.3 puts these elements together in an optimization model for the household. Finally, a central planner, usually a state, sets targets related to adoption level and timing and chooses adoption subsidy levels to achieve them as it anticipates household responses. Section 3.3.3 models the central planner's problem.

3.3.1 Household Population

We consider a region with a population of households that are heterogeneous in their income levels. Let $r \in (0, \infty)$ denote the income level of a household. The function p denotes the probability density of income in the population, i.e., $p(r)\mathrm{d}r$ is the portion of the population with income in the range of $[r, r + \mathrm{d}r]$. As discussed earlier, we are interested in studying the effect of inequality on adoption decisions. We chose income, rather than wealth or consumption, as the modeling primitive to capture inequality in our model. This choice is primarily due to the availability of direct support in the literature for two links we want to capture in our model: the link between discount rates and income levels and the link between electricity consumption and income levels. We discuss these two features of our model below.

Demand Heterogeneity

We consider a complete filtered probability space $(\Omega, \mathcal{F}, F, P)$. The demand of electricity per unit time for a household belonging to income level r is represented by a stochastic process, $\{X_t^r\}_{t\geq 0}$. Let $\{\mathcal{F}_t^r\}_{t\geq 0}$ be the $\sigma-$algebra generated by the stochastic process $\{X_t^r\}_{t\geq 0}$, i.e., $\mathcal{F}_t^r = \sigma(X_s^r : 0 \leq s \leq t)$. The filtration $F = \sigma(\bigcup_r \bigcup_t \mathcal{F}_t^r)$. We assume that X_t^r follows the following dynamics

$$\mathrm{d}X_t^r = \mu(r)X_t^r \mathrm{d}t + \sigma X_t^r \mathrm{d}W_t^r. \tag{3.1}$$

Here, the functions $\mu : (0,\infty) \to (0,\infty)$ and parameter σ satisfy the at-most linear growth and Lipschitz continuity conditions.

The choice of representing electricity demand as a stochastic process allows us to capture the inherent variability of electricity consumption over time and provides us with a modeling platform to focus on the adoption timing decisions. This choice also follows recent literature; see (9) for a similar modeling approach.

A wide variety of sources provide empirical evidence for the observation that the electricity demand is increasing over time and that households with higher incomes consume more. For example, a report by NREL (81), states that the average electricity demand is expected to increase 2.6% by 2050 due to the widespread electrification of homes and modes of transport. The data (92) from 1971 to 2014 on per capita electricity consumption for three different major income levels shows that the electricity consumption for a higher income level is not just larger, it also displays a steeper gradient indicating faster growth for higher-income households. We capture these empirical observations in our model by assuming $\mu(r) > 0$ for all r and $\mu(r)$ is a non-decreasing function of r. We also assume that $X_0 = X_0^r = x$ for all $r \in (0,\infty)$ for the sake of tractability and expositional simplicity.

Discount Rate Heterogeneity

(12) observes that income level should not influence investment decisions if markets work perfectly, but that is not what we see in practice. The paper then shows how markets are inefficient and how such inefficiencies can explain the impact of inequality on investment

decisions. Moving from general arguments to the specific case of household investments in renewables, (8) provides evidence of such investments being less than expected and then lists a variety of factors that may be driving such underinvestment, including the fact that lower-income households may hold back from such investments. This underinvestment is sometimes called the "energy efficiency gap" (42). Investigations into why such a gap exists tend to focus on the role of the implicit discount rate; see, for example, (72), which describes it as the rate at which subjects discount the returns to energy-efficiency investments inferred ex-post from actual purchase decisions.

We seek to parsimoniously capture the effect of income differences on the adoption investment decision in our model. There may be many underlying drivers for how income levels influence household decisions – such as credit availability, liquidity constraints, and hidden costs for low-income households – but the above literature suggests that an effective way to capture their combined effect is to posit that the implicit discount rate which households use to make their decision is dependent on their income level.

In energy policy literature, the relationship between discount rates and income has long been a subject of interest. (47) is one of the earliest works to model a household purchase decision of an energy-using durable and estimate its parameters; the results show that the discount rate varies inversely with income. This early literature is summarized in (86), which reviewed results from more than 12 studies that show discount rates decreasing with income. (57) estimates that discount rates are three to five percent higher for households with low permanent incomes than for those with high permanent incomes. Newer studies like (32) continue to confirm the negative correlation between discount rates and income. A recent review, (52), graphically represents the results of 13 studies, further observing a correlation between high-income levels and low discount rates.

We consider this evidence strong enough to assume a similar relationship in our model. The income-dependent time preference discount rate is given by $\lambda(r)$. The function λ is assumed to be continuous, bounded, and decreasing in r. It satisfies $\lambda(r) > \mu(r)$ and $\lambda(r) \in [\underline{\lambda}, \overline{\lambda}]$ for all r.

3.3.2 Costs

From the perspective of a household considering the adoption of a solar product, we model the following costs: the price of electricity consumption, any adoption costs related to the acquisition of solar products, and net compensation after obtaining the solar product.

Let us first consider the consumption costs. Before adoption, a household pays for the electricity *consumption* at the retail price of p_b per unit. If it adopts a solar product, the cost is determined by what is generally known as the net-metering mechanism in the literature. Under net-metering, the central planner compensates the household for any electricity it delivers to the grid. The billing cycle of the compensation mechanism, usually a month or a year, is denoted by t_b. At the end of a cycle, if a household's electricity demand exceeds the total solar energy generated during a billing cycle, the household pays for the surplus consumption at retail price p_b per unit, and if a household generates more solar electricity than the total demand during a billing cycle, the household is compensated for excess production at the same rate p_b. We refer to a household's net cost as its *compensation*.

The net-metering solar compensation mechanism is widely used in the solar adoption models in the literature; see, for example, (82) and (2). It is also widely used in practice. While some variations exist, many states, including Delaware, New Hampshire, Oregon, and Minnesota (30), use compensation mechanisms similar to the one described above, crediting surplus solar generation at retail price.

The acquisition costs incurred by the household at the time of adoption depend on the choice of solar product. We assume that the household has a choice of two solar products: rooftop solar and subscription solar, with index $i = 1$ for the rooftop and $i = 2$ for the subscription solar. The former requires an upfront cost, and the latter charges a periodic fee. Including these two choices in our model reflects contemporary practice; these two products were simultaneously available in 39 states in the U.S. in 2021 (65). The capacity acquired by a household under either product choice is assumed to be the same and is denoted by c. The function $\eta(c, t_b)$ represents the electricity generated by solar capacity c during the billing cycle of duration t_b. The exogenous function η captures the innate inefficiency in solar products due to the semiconductor material used in solar cells and the

seasonality effects. The notion of a household purchasing similar capacity under both choices is supported by studies such as (2), which report similar average capacities of residential panels under sales and non-ownership business models.

The rooftop solar requires a fixed payment of K and a variable payment of k per unit capacity. The total cost of installing rooftop solar of capacity c is $K + kc$. The subscription solar requires a periodic payment of subscription cost per unit capacity, p_{sub}. The payment structure for the two products is different; rooftop solar requires a substantial upfront payment, while subscription solar requires periodic small payments. We refer to these as the *adoption* costs.

Finally, a household may receive a subsidy from the central planner in the form of a percentage reduction in a household's adoption costs. We model this subsidy as δ_i where the subscript $i \in \{1, 2\}$ refers to rooftop and subscription solar products, respectively. We will discuss the design of this subsidy further in Section 3.5.2. We summarize the notations described above in a table available in the online appendix.

3.3.3 Model Formulation

We present a bilevel formulation to derive the optimal subsidy policy for the central planner. The problem naturally fits in the bilevel optimization framework since the household's solar adoption decisions across the population impact the central planner's ability to achieve the targets, and the central planner can influence the household's decisions by setting the subsidy levels. Once the central planner sets the subsidy policy, the households react to these values and either choose one of the solar products or do not adopt solar technology. In our model, the central planner is the leader, minimizing the average discounted cost due to subsidies, and the households in the region are the followers, minimizing their expected total discounted consumption, adoption, and compensation costs. We will refer to the lower level (respectively upper level) problem as the household's (respectively central planner's) optimization problem. In the following, we first present a household's lower-level problem and then the central planner's upper-level problem.

Household's optimization problem

Consider a household at income level r whose electricity consumption per unit time follows the stochastic process $X^r = (X_t^r)_{t\geq 0}$. For now, we do not specify the dynamics for X^r; we just assume that X^r is a Markov process. For a given subsidy level δ, the household must decide (i) if and when it should adopt solar technology and (ii) which of the two products it should adopt: rooftop or subscription solar. Suppose the household chooses to adopt solar technology when its rate of electricity consumption is x. Then, its expected discounted future cost is given by

$$g(x;\delta,r) = E_x\left[\sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} p_b\left(\int_{(n-1)t_b}^{nt_b} X_s^r ds - \eta(c,t_b)\right)\right] \tag{3.2}$$

$$+ \min\big((1-\delta_1)(K+kc), (1-\delta_2)\sum_{n=0}^{\infty} e^{-n\lambda(r)t_b} p_{sub} c\big). \tag{3.3}$$

where E_x denotes the expectation under P given $X_0^r = x$.

The first term represents the expected discounted cost of electricity consumption. The integral, $\int_{(n-1)t_b}^{nt_b} X_s^r ds$, is the total electricity consumption in a billing cycle, and the function, $\eta(c,t_b)$ is the electricity generated by solar capacity c during the billing cycle. In any billing cycle of length t_b, if the total demand is more than the solar production with acquired capacity c, then the household pays the rate p_b for the excess electricity it consumes from the grid. Otherwise, the household receives the same rate for the excess electricity it produces and puts into the grid. This follows the net-metering policy described above. The index n counts the number of billing periods forever into the future, and $\lambda(r)$ represents the discount rate used to compute the present value at the time of adoption. Irrespective of which option is adopted, rooftop or subscription, the first term remains the same due to similar capacity offerings in both the products and compensation mechanism i.e., net-metering. If we have solar products with different capacities or different compensation mechanisms, the minimum function would be outside the expectation function. The assumptions of similar capacity and compensation mechanism among solar products are to keep the model analytically tractable. The second term represents the adoption cost and depends on the chosen option. The first term under the min function is the cost of acquiring rooftop solar, and the second term represents the present value of the stream of subscription fee payments incurred

at the end of each billing cycle. The household picks the minimum of the two costs.

We are now ready to formulate the household's optimization problem. We denote a household's adoption time given income level r by $\tau_r := (\tau|r)$. The random variable τ_r is adapted to the filtration $\{\mathcal{F}_t^r\}_{t\geq 0}$ generated by the demand process for income level r, $\{X_t^r\}_{t\geq 0}$. The optimal stopping time for a household, given the income level r, is denoted by τ_r^* , and it is a lower-level variable. The optimal cost function for a household of a given income level r and subsidy policy δ, $V(x;\delta,r)$, is given by the following optimal stopping problem.

$$J(x,\tau_r;\delta,r) := E_x\left[\int_0^{\tau_r} e^{-\lambda(r)s} p_b X_s^r ds + e^{-\lambda(r)\tau_r} g(X_{\tau_r}^r;\delta,r)\right], \tag{3.4}$$

$$V(x;\delta,r) := \inf_{\tau_r\geq 0} J(x,\tau_r;\delta,r), \tag{3.5}$$

$$\tau_r^* := \arg\min_{\tau_r\geq 0} J(x,\tau_r;\delta,r). \tag{3.6}$$

In the above optimal stopping formulation, the optimal cost function seeks to minimize the discounted sum of the running cost of electricity consumption till the stopping time and the post-adoption costs. In essence, a household incurs the running cost of electricity consumption before adoption. It has the option to adopt now and receive a terminal payoff equal to the sum of the adoption cost and the present value of future electricity consumption, adjusted for net metering. Or it has the option to wait and delay the adoption. The household determines an optimal stopping time to adopt one of the products. The formulation above will determine when it should adopt and, when it does, which of the two options it should choose.

Central planner's optimization problem

A central planner aims to meet the adoption level target Λ , expressed in terms of the fraction of households who adopt either the rooftop or the subscription solar product within an adoption time target horizon T. We consider the dual adoption level and time targets, (Λ, T), as an exogenous input to the central planner's problem. Later in the numerical section, we explore the impact of different targets.

Given a target, the central planner's objective is to achieve it at the minimum expected

average discounted cost of the subsidy the planner provides to the households that adopt. As formulated in the previous section, a household solves its own optimal stopping problem to determine the time to adopt the preferred solar product. Thus, household decisions across the population impact the central planner's ability to achieve the target, and the central planner can influence a household's decisions by the subsidy offered. Assuming that the central planner has complete knowledge of the household's problem, we formulate the central planner's problem as bilevel optimization with the level of subsidy δ as the upper-level variable. We present it below and then discuss it further.

We first define the adoption time of a randomly chosen household from the population, τ^*, and its density function $f_{\tau^*}(t;\delta,x)$. This metric is useful for the central planner as it represents a summary response of the households to a specific subsidy policy. Note that we can interpret the probability of τ^* not exceeding a given time-horizon as the average fraction of households that will adopt in that time horizon. We use this interpretation in the formulation below.

$$f_{\tau^*}(t;\delta,x) := \int_0^\infty f_{\tau_r^*}(t;\delta,x)p(r)dr. \tag{3.7}$$

$$P(\tau^* \leq t|X_0 = x) := \int_0^\infty P(\tau_r^* \leq t|X_0 = x)p(r)dr. \tag{3.8}$$

The bilevel formulation for the central planner is given by

$$\min_{\delta} E_x\left[\int_0^\infty e^{-\lambda(r)\tau_r^*}\left(1_{\{r\leq r^*(\delta)\}}\delta_2\sum_{n=0}^{\infty}e^{-n\lambda(r)t_b}p_{sub}c + 1_{\{r>r^*(\delta)\}}\delta_1(K+kc)\right)p(r)dr\right]$$

s.t.

$P(\tau^* \leq \mathcal{T}|r \in (r_{lb}^\Lambda, r_{ub}^\Lambda), X_0 = x) \geq \Lambda$

$\tau_r^* := \arg\min_{\tau_r\geq 0} J(x,\tau_r;\delta,r)$

$(1-\delta_1)(K+kc) - (1-\delta_2)p_{sub}c \quad \geq \epsilon$

$0 \leq \delta \leq 1.$

In the above formulation, the objective function minimizes the average discounted cost of providing a subsidy to a household. The choice of the household for solar product is embedded in the objective function using the indicator function. The first constraint ensures

that the probability of a randomly chosen household adopting a solar product within horizon T is at least Λ; that is, the central planner must meet or exceed his target (Λ, T). Note that we allow this randomly chosen household to belong to any specified income interval with a lower bound r_{lb}^{Λ} and an upper bound r_{ub}^{Λ} . This gives us the flexibility to impose an adoption target for any chosen set of income levels. We will use this flexibility later, but for now, we focus on meeting the target across the entire population: $r_{lb}^{\Lambda} = 0$ and $r_{ub}^{\Lambda} = \infty$. The second constraint corresponds to the lower-level problem and is valid for all r. The third constraint eliminates uninteresting subsidies like those that make the entire investment in rooftop solar less than subscription payment for a single period. Here, ϵ is a very small positive real number. The last constraint ensures that the subsidy percentage for each product is between 0 and 1.

3.4 Analysis of the Model

In Section 3.4.1, we solve the household's optimization problem and derive a closed-form expression for a threshold demand level and the density of optimal adoption time for each income level. In Section 3.4.2, we show that the central planner's optimization problem is convex in the case of homogeneous subsidies. In the case of non-homogeneous subsidies, the problem is non-convex; we offer a computationally tractable algorithm to solve the problem.

3.4.1 Analysis of the Household's Optimization Problem

To analyze this problem, we first show that it is possible to separate the two decisions: the timing of adoption and the type of product to adopt. In the following result, we show that the decision about the type of product depends only on cost parameters and can be characterized based on the income level of a customer. Specifically, we prove the existence of a subsidy-dependent threshold income level $r^*(\delta)$ that divides the range of incomes into two distinct subsets such that if a customer adopts a solar technology, she will choose roof-top solar if $r > r^*(\delta)$ and subscription solar if $r \leq r^*(\delta)$. In the following, we state the result in terms of a discount rate threshold λ_δ^* where $r^*(\delta)$=$\lambda^{-1}(\lambda_\delta^*)$ and $\lambda^{-1}(\cdot)$ is the inverse function of the income-dependent discount rate $\lambda(r)$, which is a decreasing function of r.

This allows us to state the result without assuming a specific form for $\lambda(r)$ and yet provide a closed-form expression for the threshold λ^*_δ.

Lemma 1. *Given* $\lambda^*_\delta = -\frac{1}{t_b}\ln\left(1 - \frac{(1-\delta_2)p_{sub}c}{(1-\delta_1)(K+kc)}\right)$,

1. *if* $\lambda^*_\delta \geq \overline{\lambda}$ *then* $r^*(\delta) = 0$.
2. *if* $\lambda^*_\delta \leq \underline{\lambda}$ *then* $r^*(\delta) = \infty$.
3. *if* $\underline{\lambda} < \lambda^*_\delta < \overline{\lambda}$ *then there exists* $r^*(\delta) \in (0, \infty)$.

Given the monotonous nature of $\lambda(r)$, the above result states that if λ^*_δ, exclusively determined by the cost and subsidy parameters, is too low, all customers, irrespective of their income, will adopt subscription solar. If λ^*_δ is too high, all customers will adopt rooftop solar. In between these two extremes is a region where customers below an income threshold will adopt subscription solar, and those above will adopt rooftop solar. A clear articulation of this threshold allows us to focus on the timing of adoption for the rest of this section. A direct advantage of the previous result is that it simplifies the terminal cost function g in (3.3). We first present the expressions for this terminal cost function.

Lemma 2. *Assuming* $\underline{\lambda} < \lambda^*_\delta < \overline{\lambda}$,

$$g(x; r, \delta) = \begin{cases} A(r)x - B(r) + \frac{(1-\delta_2)p_{sub}c}{1-e^{-\lambda(r)t_b}} & , r < r^*(\delta) \\ A(r)x - B(r) + (1-\delta_1)(K+kc) & , r \geq r^*(\delta). \end{cases}$$

where,

$$A(r) = \frac{p_b}{\mu(r)} \frac{(1 - e^{-\mu(r)t_b})}{e^{(\lambda(r)-\mu(r))t_b} - 1}. \tag{3.9}$$

$$B(r) = \frac{p_b}{e^{\lambda(r)t_b} - 1}\eta(c, t_b). \tag{3.10}$$

We are now ready to solve the optimal stopping problem. We first show that for a customer from a given income level, an adoption time exists. We prove it by defining a set of admissible discounting policies and a set of income levels such that the customers belonging to this set will not adopt a solar product. We then show that this set is empty. Then, we define and analyze the continuation region of the optimal stopping formulation and derive an expression for a threshold demand level.

Lemma 3. *For the given structure of function g and subsidy policy δ,*
(i) there exists an electricity demand threshold for a customer from an income level such that solar adoption is an optimal choice if current electricity consumption is greater than the demand threshold,
(ii) the threshold is given by

$$\overline{X}(r,\delta) = \left(\frac{\gamma_1(r)}{\gamma_1(r)-1}\right)\left(\frac{f(r,\delta)-B(r)}{\frac{p_b}{\lambda(r)-\mu(r)}-A(r)}\right) \tag{3.11}$$

where,

$$f(r,\delta) = 1_{\{r\leq r^*(\delta)\}}\left((1-\delta_2)\frac{p_{sub}c}{1-e^{-\lambda(r)t_b}}\right) + 1_{\{r>r^*(\delta)\}}(1-\delta_1)(K+kc). \tag{3.12}$$

The above result shows that customers will eventually adopt a solar product under an appropriate subsidy policy. But we are interested in when such adoptions occur. We address this question in two steps. The first result shows that a threshold electricity demand level exists, such that when a customer's demand exceeds this level, it is optimal for that customer to adopt a solar product. The following result provides a closed-form expression for this threshold and explicitly captures its dependence on the customer's income level and the subsidy policy. Later, in the second step, we use this threshold to determine the timing of adoption. The closed-form expression of the threshold is instructive, as it connects the adoption cost parameters, income, and subsidy to the adoption threshold. Note that the function $f(r,\delta)$ captures the upfront costs for either option, the adoption cost for the rooftop solar, and the discounted sum of regular payments for the subscription solar. As this function increases, the threshold increases too. Based on this threshold, we are now ready to state a result regarding a household's time of adoption. As a household observes the stochastic process that governs its demand, the demand level eventually meets the above threshold. At this time, the household adopts a solar product. We present a closed-form expression for the density function of this adoption time.

Lemma 4. *The probability density function of τ_r^* is given by*

$$f_{\tau_r^*}(t;\delta,r) = \frac{a(r,\delta)}{\sqrt{2\pi}t^{3/2}}e^{\frac{-(a(r,\delta)-b(r)t)^2}{2t}} \tag{3.13}$$

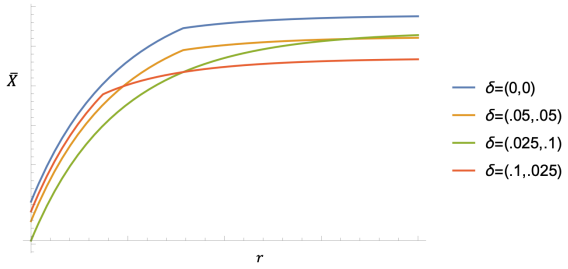


Figure 3.1: $\overline{X}(r)$ for different subsidy policies

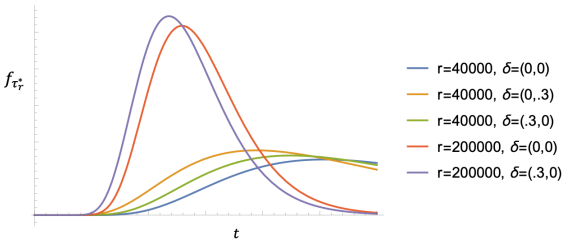


Figure 3.2: Probability density of τ_r^*

where,

$$a(r,\delta) := \frac{1}{\sigma}\log\left(\frac{\overline{X}(r,\delta)}{x}\right) \tag{3.14}$$

$$b(r) := \frac{1}{\sigma}\left(\mu(r) - \frac{\sigma^2}{2}\right) \tag{3.15}$$

We visualize the relationships between demand threshold and probability density of adoption time at different income levels and subsidies in Figure 3.1 and Figure 3.2 respectively. Figure 3.1 shows that the threshold decreases as the overall subsidy increases. However, the exact manner of the change in threshold depends on how that subsidy is divided between the two products. Figure 3.2 shows that the density of adoption time shifts to the right for households of different income levels as subsidy and income increases. We also observe that the peaks of the density curve for low and high income households can be brought closure with proper subsidy policy. To summarize this section, a household decides between the two products based on their adoption cost, income level, and the central planner's subsidy level. The product choice decision depends on comparing the upfront adoption cost of the rooftop solar product and the subscription cost stream of the subscription solar product. We show that the choice decision depends on an income-level threshold for which we provide a closed-form expression. Households from income levels higher than the threshold adopt rooftop solar, while those lower than the threshold adopt subscription solar. The adoption timing decision compares the value of waiting and the value of immediate adoption. We show that this decision, too, depends on a threshold for electricity demand, which is a function of income and subsidy levels. When a customer's demand exceeds this threshold,

she finds it optimal to adopt. We provide a closed-form expression for this threshold and for the resulting adoption time density.

3.4.2 Analysis of the Central Planner's Optimization Problem

In this section, we solve the bilevel formulation for the central planner presented in Section 3.3.3. We consider two cases: homogenous subsidy and non-homogenous subsidy. In the homogenous subsidy case, we assume that the central planner assigns similar subsidies to both the products, i.e., $\delta_1 = \delta_2 = \delta$. Under the non-homogenous subsidy case, δ_1 and δ_2 can be unequal.

We rewrite the objective function of the central planner's problem below and refer to it as $z(\delta)$:

$$z(\delta) := E_x\left[\int_0^\infty e^{-\lambda(r)\tau_r^*}\left(1_{\{r\leq r^*(\delta)\}}\delta_2\sum_{n=0}^{\infty}e^{-n\lambda(r)t_b}p_{sub}c + 1_{\{r>r^*(\delta)\}}\delta_1(K+kc)\right)p(r)dr\right]. \tag{3.16}$$

Homogenous subsidy

In this case, we can observe using Lemma 1 that $r^*(\delta)$ is independent of the discount policy δ. This allows us to prove the following result.

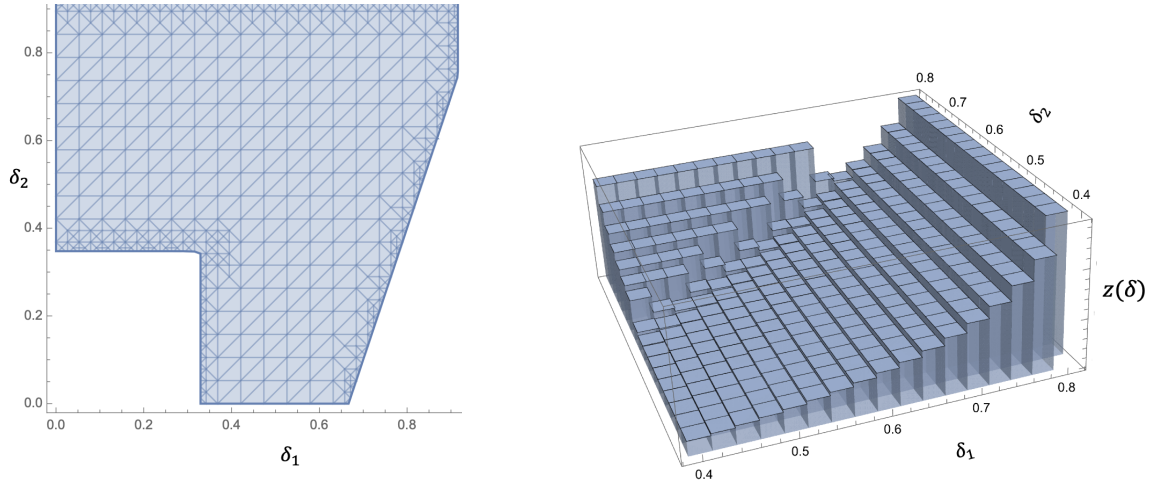
Lemma 5. *The feasible region of the central planner's problem is a convex set.*

Lemma 6. *$z(\delta)$ is a convex and increasing function of δ.*

The combination of Lemma 5 and Lemma 6 shows that the central planner's problem for the case $\delta_1 = \delta_2 = \delta$ is a convex optimization problem. Any traditional algorithm, like gradient descent, can be used to solve the problem optimally.

Heterogeneous subsidy

In the general case, where product subsidies can differ, the central planner's formulation is a non-convex optimization problem. In Figure 3.3, we show results for a specific case of parameters when the feasible region and objective function are non-convex.The finding



(a) (b)

Figure 3.3: In case of heterogeneous subsidy (a) Feasible region (b) Objective function

of non-convexity leads us to observe that a heuristic rule that simply combines two subsidy levels, each individually feasible, may not even provide a solution that satisfies all constraints.

We use a grid search algorithm to find the optimal subsidy, δ^*, which minimizes $z(\delta)$. We set a precision level, *prec*, and divide the $[0,1]^2$ space into disjoint grids of size *prec* x *prec*. We choose a sequence of equally spaced points (δ_1^i, δ_2^i) from the disjoint smaller grid, which satisfies the constraints of the central planner's optimization problem and evaluates $z(\delta)$ at each point. The pair (δ_1^i, δ_2^i), which gives the lowest value of $z(\delta)$ is the optimal subsidy level. We can increase the accuracy of the algorithm by choosing a lower value of *prec* and making the divisions finer, but it would increase the computational burden. We note that $z(\delta)$ is not Lipschitz continuous. Therefore, it is difficult to develop bounds for the algorithm.

3.5 Discussion of the Impact of Parameters

After deriving theoretical results in the previous section, we now turn to developing insights into the impact of input parameters and what such insights may mean for the central

planner. We first consider the role target level and target horizon play. Next, we analyze the impact of subsidy structure and what it means for the customers' decisions. We then develop insights into the impact of income distribution in the region.

In this section, we use numerical experiments to graph results and draw insights. The set of input parameter values for these numerical experiments is based on a variety of industry and research reports; we briefly mention them here. We use $p_b = 11.2$ cents/kWh, which was the average price of electricity to residential customers in Washington state in July 2023 (EIA) and $t_b = 1$ month. We use $c = 6.35$ kW, which is the mean of the average roof-top solar capacity of 6.21 kW and the average non-ownership solar capacity of 6.48 kW in the US (2). We assume $\eta(6.48, 1) = 500$ kWh per month for $c = 6.35$ kW based on the energy calculator (66). We set $p_{sub} = \$22$/kW per month, $K = \$10,000$ and $k = \$4,000$, which are in the range of prices without rebates by the government. We assume that $\underline{\mu} = .01$ and $\overline{\mu} = .04$ i.e., the annual rate of increase in electricity demand is bounded between 1% and 4%. We use annual discounting with rates $\underline{\lambda} = .045$ and $\overline{\lambda} = .06$. Based on (9), we have $x = 1.6$ kW per hour.

3.5.1 Adoption Level and Time Targets

As discussed earlier, a distinguishing feature of our model is the focus on the adoption level and adoption time targets for the central planner. This section addresses the impact of these targets on both the central planner and the customer. It is straightforward to see the impact of these targets on the central planner's problem. A longer adoption time target T leads to lower optimal subsidy costs. This is because a longer horizon results in a larger feasible region which will be a super-set of the original feasible region due to the monotonicity of the probability measure. Using similar arguments, we can say that the higher value of adoption level target Λ leads to higher optimal subsidy costs. In summary, more challenging targets, either high adoption levels or short planning horizons, increase the expected discounted subsidy cost.

Our numerical analysis underlines the impact on the optimal central planner cost as the adoption level and time targets change. It suggests that the two types of targets can be used

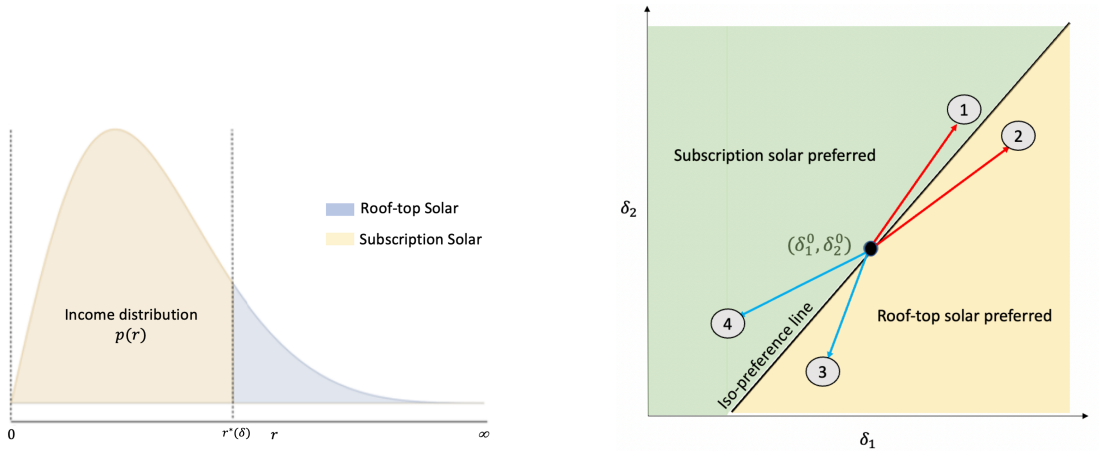
as substitutes for each other. If the central planner must set a tight adoption-level target, it can still control costs by relaxing the time targets. From another perspective, it is not enough to advocate for larger adoption-level targets. Such advocacy must also emphasize time horizons for achieving them; otherwise, central planners may choose the control costs by stretching the horizon over which they promise to meet the adoption level targets.

But as targets change, and as the central planner optimally adjusts subsidies in response, how do they influence a household's choice, rooftop or subscription solar? The length of the adoption time and level targets impact the central planner's subsidy decisions for different products. Though these targets are not part of a household's optimization problem, they impact the household's decision through subsidies assigned by the central planner to achieve the target adoption level in the target time. We explore the impact of subsidies in the next section.

3.5.2 Effect of Subsidy on Households' Product Choice

As we discussed in Section 3.4.1, the value of $r^*(\delta)$ indicates the threshold for a household's choice of different solar products in the region. A household belonging to income level r prefers subscription solar if $r \leq r^*(\delta)$ and roof-top solar if $r > r^*(\delta)$. We visualize this threshold in Figure 3.4a. A central planner can influence this choice by changing the subsidy policy. Though if the subsidy is homogeneous, i.e., $\delta_1 = \delta_2 = \delta$, using Lemma 1, we note that $r^*(\delta)$ is independent of the discount policy δ and, therefore, the central planner cannot control the product choice by altering the subsidy. Under a homogeneous subsidy policy, a change in targets does not affect the choice of a household and only influences the decision to adopt or not.

In the case of heterogeneous subsidy, the central planner has an additional lever at its disposal. It can differentiate between subsidies and thus also influence a household's product choice. To better understand this lever, we study the change in threshold income level, $r^*(\delta)$, when the subsidy for one of the products is changed. We show that the threshold moves to the left as δ_1 increases, i.e., there are households whose choice shifts from subscription solar to rooftop solar. The threshold moves to the right as δ_2 increases, i.e., there are some



(a) (b)

Figure 3.4: (a) Product preference in the region (b) Effect of subsidy change on customer product preference

households who chose rooftop solar earlier and now prefer subscription solar. Lemma 7, states this formally and confirms the intuition that a higher subsidy for a product choice will favor its adoption.

Lemma 7. *$r^*(\delta)$ is a non-increasing function of δ_1 and a non-decreasing function of δ_2.*

The impact on preference when subsidy for both products is changed simultaneously is not as intuitive. If both subsidies are increased (indicated by the red arrow in Figure 3.4b) or both subsidies are decreased (indicated by the blue arrow), the impact on preference is not immediate. We now present a method that allows us to understand the impact of such subsidy changes. Let $\delta^0 = (\delta_1^0, \delta_2^0)$ be the initial subsidy and $\delta^1 = (\delta_1^1, \delta_2^1)$ be the changed subsidy such that $\delta_1^0 \neq \delta_1^1$ and $\delta_2^0 \neq \delta_2^1$.

We find that there exists a line passing through (δ_1^0, δ_2^0) such that if (δ_1^1, δ_2^1) lies on the line, then the household preferences in the region do not change. We call such a line *iso-preference* line. If (δ_1^1, δ_2^1) lies above the iso-preference line, then preference is shifting towards subscription solar, and if (δ_1^1, δ_2^1) lies below the iso-preference line, then preference is shifting towards roof-top solar, as shown in the Figure 3.4b. From Lemma 1, we know

that

$$r^*(\delta) = \lambda^{-1}\left(\frac{-1}{t_b}\ln\left[1 - q(\delta)\frac{p_{sub}c}{K+kc}\right]\right), \tag{3.17}$$

where $q(\delta) := (1-\delta_2)/(1-\delta_1)$. In the equation above, for any δ^1 such that $q(\delta^1) = q(\delta^0)$, $r^*(\delta^1) = r^*(\delta^0)$. This gives us the equation of the iso-preference line, $\delta_2^1 = q(\delta^0)\delta_1^1 + 1 - q(\delta^0)$. If $\delta_2^1 > q(\delta^0)\delta_1^1 + 1 - q(\delta^0)$ then using the decreasing property of $\lambda(\cdot)$ function we can show that $r^*(\delta^1) > r^*(\delta^0)$. Similarly, if $\delta_2^1 < q(\delta^0)\delta_1^1 + 1 - q(\delta^0)$ then $r^*(\delta^1) < r^*(\delta^0)$. In Figure 3.4b, we see that since points 1 and 4 lie above the iso-preference line through δ^0, households' choice would move towards subscription solar. A similar analogy can be made for points 2 and 3.

How should the central planner use the difference in subsidies to influence a household's choice, and which product should it favor: rooftop or subscription? The central planner's choice of optimal subsidies for different products depends not only on the planning horizon and adoption level but also on the income distribution in the region. We know from Lemma 1 that the high (low) income customers prefer rooftop (subscription) solar. Depending on the targets and the income distribution, the central planner may allocate a higher subsidy to the product preferred by the higher number of households, thus promoting early adoptions. In our numerical experiments, when faced with a population distribution skewed toward lower incomes, the central planner finds it optimal to offer higher subsidies for community solar in order to meet tight adoption levels and time targets. We note that this is counter to what we find in practice where planners first offered incentives for rooftop solar products.

3.5.3 Income Inequality

Another distinctive feature of this work is the explicit modeling of the income spectrum. This allows us to address questions about the impact of income distribution on solar adoption: How does income inequality impact the adoption, and who gets the lion's share of subsidy offered by the central planner? To address these questions, we assume an explicit form for the income distribution in the region. The individual income in the region follows a log-logistic distribution, i.e., $p(r) = \frac{(\beta/\alpha)(r/\alpha)^{\beta-1}}{(1+(r/\alpha)^{\beta})^2}$. Here, $\alpha > 0$ is a scale parameter, and $\beta > 0$ is a shape parameter. According to the literature (23) ,(28), this is a well-defended

choice for capturing actual income dispersion (other possible choices of income distributions are log-normal, exponential, and Pareto). In a log-logistic income distribution, the median income of the region is given by the scale parameter α, and the Gini coefficient of the region is given by $1/\beta$. These two properties enable us to capture differences in income characteristics of different regional populations. This also provides a direct way for the central planner to incorporate income inequality in the region as it seeks an effective and fair optimal subsidy policy. We assume that $\beta > 1$; this limits us to $G < 1$ and makes the income distribution unimodal. The choice of log-logistic distribution is also useful because we can still derive the density function of the household adoption time, as we show in the next result.

Lemma 8. *The probability density function of τ^* is given by*

$$f_{\tau^*}(t; x, \delta, G, \alpha) = \frac{1}{\sqrt{2\pi}} \frac{1}{(G\alpha)t^{3/2}} \int_0^\infty a(r, \delta) \frac{(r/\alpha)^{\frac{1-G}{G}}}{(1 + (r/\alpha)^{1/G})^2} e^{\frac{-(a(r,\delta)-b(r)t)^2}{2t}} dr. \tag{3.18}$$

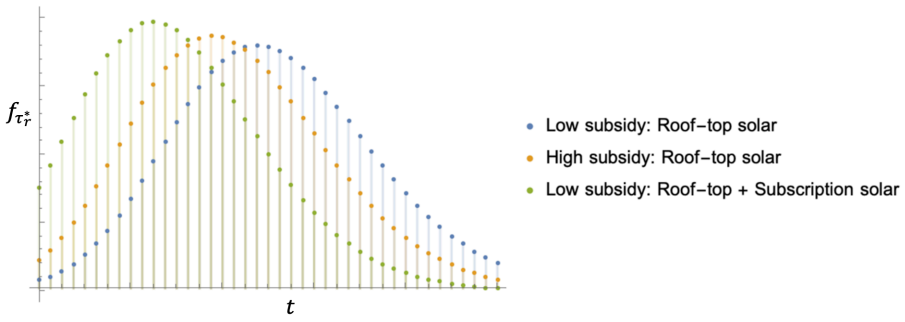


Figure 3.5: Probability density function of τ^* for different product-subsidy offerings

In the above Lemma 8, we develop a relationship between the density function of the optimal adoption time in the region and the Gini coefficient. We can interpret the probability density function of τ^* as a weighted average of the probability density of adoption time at different income levels where the weights depend on the region's income distribution. This Lemma allows us to study the interplay between income inequality, subsidy policies, and adoption time. In Figure 3.5 we compare the effectiveness of a policy of providing

households a choice of two solar products in a region of high-income inequality. The green curve represents a policy where the planner offers a low subsidy level but offers it for both products. In contrast, the orange curve represents a policy where the planner offers a higher subsidy level but only for rooftop solar products. As the Figure shows, the mode of the adoption time distribution is smaller for the former policy than for the latter. In other words, a low subsidy level with two products can deliver faster adoption than a high subsidy level with only one product. Further reduction in subsidy for the one-product policy can delay the adoption, as is apparent from the blue curve. This observation offers useful policy advice to the central planner: offering multiple options may reduce the total subsidy cost.

3.6 Extensions

In this section, we develop several extensions of our model. We start by generalizing the net-metering credit mechanism and then add other features.

3.6.1 Solar Energy Crediting Mechanisms

Our analysis till now assumed net-metering as the crediting mechanism for excess solar energy generation. Though a large number of states use net-metering, some states like Arizona, Kentucky, and South Carolina use net-billing. Under net-billing, the households are credited at a rate less than the retail electricity price. In Arizona, a household is compensated at wholesale electricity price, while in Kentucky and South Carolina, the crediting price is set between retail and wholesale electricity rates. As of 2024, some states like Alabama, South Dakota, and Tennessee do not offer any form of compensation for excess solar energy generated (31). In this section, we generalize the compensation function to incorporate net-billing and no-crediting options for excess generation along with net-metering.

We denote the price for crediting excess generation by p_s such that $0 \leq p_s \leq p_b$. If $p_s = p_b$, the household is compensated under the net-metering compensation mechanism. If $0 < p_s < p_b$, it is compensated under the net-billing compensation mechanism. If $p_s = 0$, it can offset the electricity bill but is not compensated for the excess solar generation.

We define the solar compensation function h_i as

$$h_i(x,c;r,p_s,p_b) := E_x\left[\sum_{n=1}^{\infty} e^{-n\lambda(r)t_b}\left(p_b\left[\int_{(n-1)t_b}^{nt_b} X_s^r ds - ct_b\right]^+ - p_s\left[ct_b - \int_{(n-1)t_b}^{nt_b} X_s^r ds\right]^+\right)\right]. \tag{3.19}$$

The function h_i represents the expected payments made by a household after adopting solar product i. It is the sum of expected discounted payments in a billing cycle. The payments in a billing cycle depend on the demand for electricity during a billing cycle. The term $p_b[\int_{(n-1)t_b}^{nt_b} X_s^r ds - ct_b]^+$ is the payment made if the total demand exceeds the solar electricity generation during a billing cycle. The term $p_s[ct_b - \int_{(n-1)t_b}^{nt_b} X_s^r ds]^+$ is the payment received if the solar electricity generation exceeds the total demand during a billing cycle. We can re-write the above equation as

$$h_i(x,c;r,p_s,p_b) := \sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} t_b\left(p_b E_x\left[\frac{\int_{(n-1)t_b}^{nt_b} X_s^r ds}{t_b} - c\right]^+ - p_s E_x\left[c - \frac{\int_{(n-1)t_b}^{nt_b} X_s^r ds}{t_b}\right]^+\right). \tag{3.20}$$

To analyze if the household's optimal stopping problem with the above modification can be solved, we refer to the stochastic finance literature. The term $E_x\left[\frac{\int_{(n-1)t_b}^{nt_b} X_s^r ds}{t_b} - c\right]^+$ resembles the payoff of an Asian call option and the term $E_x\left[c - \frac{\int_{(n-1)t_b}^{nt_b} X_s^r ds}{t_b}\right]^+$ resembles the payoff of an Asian put option (78). The household is a seller of p_b units of the Asian call option, and a buyer of p_s units of the Asian put option. Unlike American/European call-and-put options, there is no closed-form expression for the expected payoff of Asian options. This is primarily because the payoffs of Asian options are path-dependent due to the integral function, which makes payoffs non-Markovian. We conclude that an analytical solution for our problem may not be achievable. However, numerous numerical methods (41), (69), (88) exist for approximate valuation of the Asian options, which could provide better insights about different crediting mechanisms. Even though a complete analysis of this model is not possible, it turns out that we can derive limited results for the existence of solutions for the household's optimal stopping problems in even more general settings. In the next section, we further generalize our model before presenting these results.

3.6.2 Other Features

In our primary model, the demand dynamics follow geometric Brownian motion, the *compensation* to the household and various costs have fixed functional forms, and there is a choice of only two products with similar capacity. In a more general context, a household may have the option of considering different solar products with different capacities with compensation mechanisms other than net-metering, net-billing, and no solar credits. In addition, there may be other payment mechanisms available for solar adoption that require lower *upfront* costs than rooftop solar but higher periodic payment than subscription solar. We extend our formulation by adding these generalized features. Specifically, we now let the demand dynamics follow a general stochastic diffusion equation. The *compensation, running, and adoption costs* can assume a general form, and the household has a choice of n products with different capacities. These generalizations make detailed analytical exposition hard, but we show that exploring the edge cases is still possible. We derive a set of conditions to help the central planner devise better compensation schemes, adoption costs, and subsidy policies. We first present the model and then state the conditions.

Consider a complete filtered probability space $(\Omega, \mathcal{F}, F, P)$. The demand of electricity per unit time for a household belonging to income level r is represented by a stochastic process, $\{X_t^r\}_{t\geq 0}$. Let $\{\mathcal{F}_t^r\}_{t\geq 0}$ be the $\sigma - algebra$ generated by the stochastic process $\{X_t^r\}_{t\geq 0}$, i.e., $\mathcal{F}_t^r = \sigma(X_s^r : 0 \leq s \leq t)$. The filtration $F = \sigma(\bigcup_r \bigcup_t \mathcal{F}_t^r)$. We assume that X_t^r follows the following dynamics.

$$\mathrm{d}X_t^r = \mu(r, X_t^r)\mathrm{d}t + \sigma(r, X_t^r)\mathrm{d}W_t^r. \tag{3.21}$$

Here the functions $\mu : (0,\infty) \times (0,\infty) \to (0,\infty)$ and $\sigma : (0,\infty) \times (0,\infty) \to (0,\infty)$ satisfy the at-most linear growth and Lipschitz continuity conditions.

The central planner offers n solar products in the region. We use the notation i to denote a particular product such that $i \in \{1, .., n\}$. The subsidy and capacity for products are represented by n-dimensional vectors δ and c, respectively. The components of these vectors, δ_i, and c_i, are the subsidy provided by the central planner and the capacity offered for the product i, respectively. We assume $\delta \in [0,1]^n$ and $c \in R_+^n$. The function $f_i(c_i; r, \delta_i)$ is the *adoption* cost of the solar product i with capacity c_i given the income level r of the

household and subsidy δ_i offered by the central planner. The function $h_i(x; r, p_s, p_r, c_i)$ is the solar *compensation* offered by product i when demand for electricity is x, installed capacity is c_i, the household's income level is r, the retail price is p_r, and the re-selling price is p_s. We define the *net cost of solar adoption* for product i, $g_i(x; r, \delta_i)$, as

$$g_i(x; r, \delta_i) := E_x\big[h_i(x; r, p_s, p_r, c_i) + f_i(c_i; r, \delta_i)\big]. \tag{3.22}$$

The function, $g(x; r, \delta)$, is the net cost of solar adoption at demand level x given the income level r of the household and δ subsidy policy. We define $g(x; r, \delta)$ as

$$g(x; r, \delta) := \min_{i\in\{1,\dots n\}} g_i(x; r, \delta_i). \tag{3.23}$$

The function $w(x; p_b)$ is the cost of consuming x unit of electricity per unit of time at the retail rate p_b. The optimal cost functional for a household, given income level r and subsidy policy δ, is

$$V(x; r, \delta) := \inf_{\tau_r\geq 0}\left\{E_x\left[\int_0^{\tau_r} e^{-\lambda(r)s} w(X_s^r; p_b)ds + e^{-\lambda(r)\tau_r(\delta)} g(X^r_{\tau_r(\delta)}; r, \delta)\right]\right\}. \tag{3.24}$$

Lemma 9. *Assuming $g \in \mathcal{C}^2(R)$, $w \in \mathcal{C}^2(R)$ and a subsidy policy δ, if a household belongs to income level r such that*
(i) $\big(-\lambda(r)g(x; r, \delta) + \mathcal{L}_{X^r} g(x; r, \delta) + w(x; p_b)\big) \geq 0$ for all x then $\tau_r^* = 0$ a.s.
(ii)$\big(-\lambda(r)g(x; r, \delta) + \mathcal{L}_{X^r} g(x; r, \delta) + w(x; p_b)\big) < 0$ for all x then $\tau_r^* = \infty$ a.s.
Here $\mathcal{L}_{X^r}$ is the infinitesimal generator for the demand process $\{X_t^r\}_{t\geq 0}$.

As discussed earlier, the formulation above does not lend itself to a complete determination of adoption timing, but we are still able to identify situations in which the adoption time takes extreme values. In Lemma 9, we derive the conditions that lead to either the immediate adoption or rejection of solar technologies by a household. It connects the net adoption cost function g, running cost function w, and the dynamics of the demand process in (3.21) for a household with income level r. The result is useful for the central planner in designing net adoption cost function g and the discount policy δ, which prevents rejection of solar technology for any income level r. Using Lemma 9, we can construct a set of admissible discounting policies and a set of income levels that will not adopt solar technology at the given discount levels.

3.7 Conclusion

Governments and planners all around the world engage in setting targets for achieving alternative energy production. Solar energy production is a major source of alternative energy. The planners devise subsidy schemes to promote the adoption of solar products across populations. There is increasing urgency to set ambitious targets and achieve them quickly. We are motivated by a need to introduce a time dimension in the models that analyze this process: the setting of subsidies and how they influence solar adoption. We are also motivated by innovations in business models that may assist with faster adoption, specifically the emergence of community solar. Both the inclusion of the time dimension and modeling of the community solar option are unexplored topics in the solar adoption literature. We model a household's electricity demand as a continuous-time stochastic process. This choice allows us to focus on a household's adoption time decisions as well as its choice between community solar and rooftop solar. Another new feature in our model is discount rate heterogeneity, which allows us to focus on the adoption decisions across a range of income levels. Our technical contributions include a complete analysis of the household's decisions. We derive an income threshold that governs the choice of solar products and a closed-form expression for the probability density of the adoption time. We also derive technical results for the planner's bilevel optimization problem, where the planner determines subsidies to minimize its cost while meeting targets and assuming that the households will make their own adoption timing decision.

Numerical experiments lead to insights about target setting for the planner. Adoption level and adoption time targets behave as substitutes, allowing the planner flexibility in setting subsidies. The structure of this subsidy determines the range of household incomes that will prefer community solar over rooftop solar. Our examples show that depending on the income distribution, the planner may change its subsidy levels to achieve its adoption quantity and time targets. Finally, we show how inequality in income distributions affects the achievement of these targets. We extend our model to address general cost and compensation functions with multiple products, but several avenues for further work remain. In future work, we will continue to explore the effect of specific population densities and

inequality measures on the design of subsidies and adoption of competing technologies.

Chapter 4

DESIGNING A ROBUST OUT-BOUND NETWORK

4.1 Introduction

In today's highly interconnected and volatile global economy, designing a robust supply chain is critical for businesses seeking resilience, efficiency, and long-term competitiveness. A robust supply chain can mitigate risks associated with disruptions such as natural disasters, pandemics, geopolitical tensions, and cyberattacks. For example, during the COVID-19 pandemic, companies with flexible sourcing strategies and diversified suppliers—such as Unilever and Procter & Gamble—were able to sustain operations more effectively than those reliant on a single source. Robust supply chains also ensure consistent service levels, as seen with Amazon, which leverages a vast logistics network to meet customer expectations even under pressure. Furthermore, robust supply chains offer agility in responding to market volatility; Zara's ability to quickly adjust its production in response to changing consumer trends helps reduce inventory risks and lost sales. They also support sustainability and regulatory compliance, with companies like Apple ensuring ethical sourcing to avoid reputational and legal setbacks. Overall, a robust supply chain enables firms not only to survive disruptions but also to capitalize on opportunities, exemplified by Toyota's post-tsunami strategies to diversify suppliers and reduce lead times, which enhanced its resilience in later crises. Thus, robust supply chain design is not just a defensive strategy—it is a source of strategic advantage.

The importance of robust supply chain design is especially pronounced in the semiconductor industry, where long lead times, high capital intensity, and global interdependencies make the supply chain particularly vulnerable. The global chip shortage that began in 2020 exposed the fragility of semiconductor supply chains, severely impacting industries such as automotive, consumer electronics, and industrial manufacturing. Many companies had

relied on just-in-time inventory models and a small number of specialized foundries, such as TSMC in Taiwan, creating significant bottlenecks when demand surged and geopolitical risks escalated. In response, industry leaders and governments have begun to emphasize supply chain robustness through diversification of manufacturing locations, increased onshore production (e.g., Intel's investments in U.S. fabs), and strategic stockpiling of critical components. Robust design in this context also includes better demand forecasting, cross-industry collaboration, and digital visibility across the supply chain to anticipate and respond to disruptions proactively. As semiconductors are foundational to innovation in AI, electric vehicles, and advanced computing, a resilient supply chain is now viewed as not only a business imperative but also a matter of national and economic security.

A key aspect of designing robust supply chains involves the deliberate use of buffer inventory, data-driven decision-making, and the integration of uncertainty into supply chain modeling. Buffer inventory—such as safety stock or excess capacity—serves as a shock absorber during unexpected disruptions, whether from supplier failures, logistics bottlenecks, or demand spikes. For example, companies that had maintained buffer stocks of critical semiconductors were able to navigate the global chip shortage more effectively, avoiding costly delays and maintaining production continuity. While additional inventory entails higher costs, it significantly reduces the risk of stockouts and revenue loss during unforeseen events. Beyond inventory, incorporating uncertainty into supply chain models through stochastic optimization, scenario planning, and simulations allows businesses to evaluate strategies across a range of possible futures, rather than relying on average-case assumptions. This strengthens the system's ability to perform under volatility. Complementing these strategies, data-driven decision-making—enabled by predictive analytics, real-time tracking, and AI—allows for more accurate forecasting, early disruption detection, and agile response. By combining these tools, organizations can design supply chains that are not only optimized for efficiency under normal conditions but are also resilient, adaptive, and better equipped to navigate the uncertainties of a complex and rapidly changing world.

In this paper, we develop a data-driven stochastic optimization model to develop opti-

mal routing strategies and optimal buffer inventory of products. The aim of the research is to design a resilient supply chain optimization strategies for a semiconductor supplier. First, we analyze a publicly available dataset of the semiconductor supplier. We, then, integrate the information from the past data in a mixed integer programming model using GUROBI. We analyze the solution of the model and empirically show the inefficiencies of such a solution in case of a supply chain disruption. We use chance constraint formulation ((22) and (62)) to incorporate randomness in the production capacity of the semi conductor. In deterministic optimization, constraints must be satisfied 100% of the time. Due to random disruption in supply, instead of hard constraints, we use probabilistic constraints to reflect this supply uncertainty. Addition of such constraints provide much more robust production and transportation plans but such plans might not be feasible due to huge deviations from the current plans. We add two constraints to give decision makers control on the deviation from the current plans, thus making the solutions not only robust but also practically feasible. In the end, we generate these solutions for the given dataset using GUROBI.

The rest of the paper is organized as follows. In section 2 we study the supplier's outbound network data. The model assumptions, parameters, variables and both, deterministic and stochastic, formulations are in section 3. In section 4 we present the numerical findings and key sights from the models. We conclude the paper with a commentary on future work.

4.2 Dataset

We test the optimization model on a publicly available dataset. The data represents the order log for a single day, costs and capacity of carriers associated with fulfilling orders. Table 4.1 is the description of the data in different tabs of the excel file. We use columns Order ID, Customer,Product ID, Destination Port, Unit quantity and Weight from *OrderList*. We assume that plant capacity in *WhCapacities* is tenfold to reduce the demand surplus in the data. Since Plant 19 is not assigned any product in *ProductsPerPlant*, we have assigned product (Id-1698759) of CND9, an invalid entry, to Plant19. We assume that the production cost of all products at a given plant is equal and is expressed in per unit quantity in *WhCosts*. The minimum cost and rate in *FreightRates* can be different for similar carriers

OrderList	9215 demand orders characterized by order ID, customer, product ID, weight and units ordered.
FreightRates	1540 feight options characterized by carrier, origin port, destination port, minimum weight capacity, maximum weight capacity, fixed rate, variable unit weight rate, estimated shipping time, available service level and mode of transport.
WhCosts	Unit production costs in 19 different plants
WhCapacities	Maximum capacity of the plants
ProductsPerPlant	Specific products which a plant can manufacture
VmiCustomers	Specific customers whose order a plant can service. If a plant is not on the list, it can service any customers request.
PlantPorts	Specific ports a plant can access

Table 4.1: Data description

and rate is linearly dependent on quantity instead weight.

4.3 Problem Formulation

In this section with introduce the assumptions, parameters and variables of the model. We also present a deterministic and a stochastic MIP formulation for the outbound network.

4.3.1 Assumptions

An order can be split and transported using different freight options. If an order quantity is divided and assigned multiple freights, the weights of the order in multiple freights is proportional to split quantities. We consider service type(svc_cd) as DTD and mode as AIR in the optimization model. We assume that production time is zero, hence time taken to fulfill an order just includes the time to transport the order. We notice that in the data provided the total quantity of orders far exceeds the total daily production capacity of the plants. Also, there are restrictive permissible combinations of plant, product and customer

(See *ProductsPerPlant*, *VmiCustomers*, *PlantPorts*). These two factors imply that all orders cannot be fulfilled. We introduce a dummy uncapacitated warehouse to store unassigned demand of different products. A dummy port and dummy freight is also introduced to ensure feasibility in the formulation.

4.3.2 Parameters

Sets

I: Set of plants including a dummy warehouse

J: Set of ports including a dummy port

C: Set of customers

D: Set of Product id

K: Set of Orders

M: Set of Freight options including a dummy uncapacitated option from dummy warehouse to dummy port

S: Set of permissible plant-product combinations

T: Set of permissible plant-port combinations

L: Set of permissible plant-customer combinations

Indices

i: Individual plant s.t $i \in I$

j: Indivdual post s.t $j \in J$

k: Individual order s.t $k \in K$

m: Single freight option s.t $m \in M$

i_{dummy}: Dummy warehouse

j_{dummy}: Dummy port

m_{dummy}: Dummy freight option

Order characteristic

c_k: Customer of order k

d_k: Product id of order k

q_k: Order quantity of order k

w_k: Weight of order k

Carrier characteristic

s_m: Carrier of freight option m

o_m: Origin port of freight option m

t_m: day count of freight option m

g_m: mode of freight option m

svc_m: service type of freight option m

Costs

p_i: Production cost at plant i

f_m: Fixed cost of freight option m

v_m: Variable cost of freight option m

Capacity

cap_i: Production capacity of plant i

w_m^{min}: Minimum weight of freight option m

w_m^{max}: Maximum weight of freight option m

4.3.3 Variables

z_{di}: Quantity of product d manufactured at plant i

x_{ijm}^k: Quantity of order k delivered from plant i through port j using freight option m

y_m: 1 if freight option m is chosen otherwise 0

4.3.4 Deterministic model formulation

$$\min \sum_i p_i(\sum_d z_{di}) + \sum_m f_m y_m + \sum_m \sum_k \sum_i \sum_j v_m x_{ijm}^k \tag{4.1}$$

$s.t$

$$\sum_d z_{di} \leq cap_i \quad \forall i \tag{4.2}$$

$$w_m^{min} y_m \leq \sum_k \sum_i \sum_j \frac{x_{ijm}^k}{q_k} w_k \leq w_m^{max} y_m \quad \forall m \tag{4.3}$$

$$\sum_k \sum_j \sum_m x_{ijm}^k = z_{di} \quad \forall i \tag{4.4}$$

$$\sum_i \sum_j \sum_m x_{ijm}^k \geq q_k \quad \forall k \tag{4.5}$$

$$z_{di} = 0 \quad \forall (i,d) \notin S \tag{4.6}$$

$$x_{ijm}^k = 0 \quad \forall k, m \text{ and } (i,j) \notin T \tag{4.7}$$

$$x_{ijm}^k = 0 \quad \forall j, m \text{ and } (i,k) \notin L \tag{4.8}$$

$$x_{i_{dummy}jm}^k = 0 \quad \forall k, j \in J/j_{dummy}, m \tag{4.9}$$

$$x_{ij_{dummy}m}^k = 0 \quad \forall k, i \in I/i_{dummy}, m \tag{4.10}$$

$$x_{i_{dummy}j_{dummy}m}^k = 0 \quad \forall k, m \tag{4.11}$$

$$x_{ijm_{dummy}}^k = 0 \quad \forall k, i \in I/i_{dummy}, j \in J/j_{dummy} \tag{4.12}$$

$$x_{ijm}^k = 0 \quad \forall k, i, m, j \notin o_m \tag{4.13}$$

$$x_{ijm}^k \geq 0 \quad \forall i, j, k, m \tag{4.14}$$

$$z_{di} \geq 0 \quad \forall i, d \tag{4.15}$$

$$y_m \in \{0, 1\} \tag{4.16}$$

In the objective function (1), the first term is the total production cost of products in plants, second term is the sum of fixed cost incurred if freight option m is chosen and last term is the total variable cost for transporting x_{ijm}^k in the outbound network. The equations (2) and (3) are constraints for the production capacity in plants and freight carrying capacity. The constraint (4) balances flow at each plant. The equation (5) ensures that demand from every order k is fulfilled. The constraints (6), (7) and (8) maintain network restrictions among

plant, port and customer combinations. The equations $(9) - (13)$ ensure that goods can be transported to dummy port only from dummy plant using dummy freight. (9) ensures that there is no flow between dummy plant and non-dummy port. Similarly, no flow between dummy port and non dummy plant is established using (10). (11) ensures regular freight cannot be used between dummy plant and dummy port. Similarly, (12) prevents using dummy freight for transportation between non dummy ports and plants. The constraint (13) ensures usage of only freight which have same origin as the current port j.

4.3.5 Robustness Analysis

Robustness analysis investigates how much the optimal solution to the original problem violates the constraints of the perturbed problem. We next present a numerical experiment to demonstrate that the optimal solution to the deterministic model might not be a good solution from practical standpoint if production capacity is uncertain.

Numerical experiment

We generate a data sample by randomly sampling 20 orders from *OrderList* and 50 freight options from *FreightRates*, while using other excel tabs entirely. We solve the model using GUROBI and denote the optimal solution by S_d. In table 4.2, first column represents the plants which are used for production in the optimal solution. The second column is the total capacity of these plants. We refer to the available capacity at a plant after production of optimal quantity as slack. The third, fourth and fifth column is the slack in each plant if there is a percentage change in the total capacity. For example, 5% $\downarrow$ for plant 2 is 5% reduction in total capacity from 1380 to 1311 while 5% $\uparrow$ is 5% increase in capacity from 1380 to 1449. We observe that S_d has 3 three binding plant capacity constraint at 0% i.e., entire production capacities of plants 2, 3 and 8 are used as signified by 0 slack. Any decrease in capacity in these plants due to disruption in plant operation would render S_d infeasible to the deterministic model.

Furthermore, resolving the deterministic model with perturbed capacities poses two major challenges. First, the perturbed formulation may be infeasible. As seen in the table, a 5%

Plant	Cap	Slack		
		0%	5% ↓	5% ↑
2	1380	0	0	0
3	10130	0	0	0
5	3850	1422	1422	701
7	2650	231	0	0
8	140	0	0	0
10	1180	247	247	948
12	2090	459	325	202
Dummy	C	C	C−5594	C

Table 4.2: Effect of change in capacity on slack

decrease in production capacity at plants 2, 3 and 8 makes the formulation infeasible as indicated by the usage of the dummy warehouse. In order to have an optimal solution, we need to have an inventory of 5594 products in addition to the production capacity. Second, the optimal solution may require making large logistical changes, which may not be practically feasible. Even in this small data sample, one out of 6 freight option changed(y_{45} to y_{50}) in the optimal solution with the perturbation in the capacity. These two issues amplify as the size of sample data is increased. Also, dramatic operational changes which might not be practically feasible can be seen in case of increase in production capacity. In the next section we propose a stochastic model which incorporates robustness against capacity perturbation while minimizing the changes to the current optimal solution.

4.3.6 Stochastic model formulation

In this section, we consider randomness in production capacity of the plants. In the real world, this capacity disruption could be caused by labor availability, natural disaster or change in regulatory policies.

Assumption

In addition to earlier assumptions, we assume that cap_i is normally distributed with mean μ_i and standard deviation σ_i.

Parameters

δ_m: Tolerance for change in production quantity

δ_l: Tolerance for change in logistics plan

$\overline{z}_{di}$, $\overline{x}^k_{ijm}$, $\overline{y}_m$: Optimal values from deterministic formulation

μ_i: Mean production capacity at plant i

σ_i: Standard deviation of production capacity at plant i

α_i: Service level at plant i

Model

$$\min \sum_i p_i(\sum_d z_{di}) + \sum_m f_m y_m + \sum_m \sum_k \sum_i \sum_j v_m x^k_{ijm} \tag{4.17}$$

$s.t$

$(3) - (16)$

$$|z_{di} - \overline{z}_{di}| \leq \delta_m \quad \forall d, i \tag{4.18}$$

$$|x^k_{ijm} - \overline{x}^k_{ijm}| \leq \delta_l \quad \forall k, i, j, m \tag{4.19}$$

$$\sum_d z_{di} \leq \mu_i + z_{1-\alpha_i}\sigma_i \quad \forall i \tag{4.20}$$

The objective is same as the deterministic model. The constraints $(3) - (16)$ ensure that the new solution is feasible to the deterministic model. The deviation of new solution from the optimal solution of deterministic counterpart is maintained within a tolerance level using $(18) - (19)$. These two constraint give us the flexibility to choose different tolerance level for change in production and logistic plan. The constraint (20) is a deterministic equivalent of the following chance constraint:

$$P(\sum_d z_{di} \leq cap_i) \geq \alpha_i \quad \forall i \tag{4.21}$$

Notice that we can retrieve deterministic model from the stochastic model by setting $\delta_m = \delta_l = \sigma_i = 0$ and $\mu_i = cap_i$.

4.4 Empirical Evaluation

We present the numerical results and key insights from the models in this section.

4.4.1 Data

We randomly sampled 20 orders from data processed *OrderList* and 50 freight options from data processed *FreightRates*. We use all of data from *WhCosts*, *WhCapacities*, *ProductsPerPlant*, *VmiCustomers* and *PlantPorts*. The datasets used for deterministic and stochastic models can be found in Appendix.

4.4.2 Parameters

In this experiment we assume that the $p_{dummy} = 10$ which is 8 times the highest production cost of any plant and $cap_{dummy} = 10^{16}$ to replicate infinite capacity. The dummy freight from dummy warehouse to dummy port has $w_m^{min} = 0$, $w_m^{max} = 10^{16}$, $f_m = 600$ and $v_m = 150$. In stochastic model, we assume $\delta_m = \delta_l = \delta = 1000$, $mu_i = cap_i$, $\sigma_i = cap_i/100$ and $\alpha_i = 0.9$.

4.4.3 Result & Insight: Deterministic model

Table 4.3, 4.4 and 4.5 are the optimal values of z_{di}, y_m and x_{ijm}^k for the deterministic formulation with the chosen data set. The total quantity ordered is 1,41,491 which far exceeds the capacity of plants, 57,910. The excess demand is taken care by the dummy variables. The introduction of dummy variables not only prevents the model from becoming infeasible when demand and supply do not match but also prescribe a strategy overcome the situation. The variable $z_{di_{dummy}}$ indicates the level of inventory of a product that should be maintained in the warehouse to tackle the excess demand. In table 4.3 the required inventory level of different products is presented with the dummy plant.

In the optimal solution, plants 2, 3, 4, 9 and 10 are used for manufacturing products. Freight option 1128, 1342, 1158 and 1228 are used for transporting goods from plant to customers using entry ports 2, 3, 4 and 5. Since x^k_{ijm} can take non integer values, an order can be split among different freights, for example, in table 4.5 order 1447254154 is split between 1128, 1228, 1158 and a dummy freight.

4.4.4 Result & Insight: Stochastic model

In essence, the chance constraint formulation forces a risk averse decision making. If we consider the right hand-side of the constraint (20), any value of $\alpha_i > .5$ leads to a negative z value. Since, we consider $\mu_i = cap_i$, constraint (20) implies that the total production at a plant is bounded by a lower right hand side value compared to the deterministic case. The results presented in table 4.6 and 4.7 are consistent with this rational.

In table 4.6 first column represents the plants which are used for production in the optimal solution. Second column corresponds to the product being manufactured at different plants. S_d and S_p are the solutions to deterministic and stochastic formulations respectively. We solve for two different level of tolerances, $\delta = 10^3$(higher) and $\delta = 10^2$(lower).

In the higher tolerance model, there is more focus on reducing the cost than maintaining closeness to the current production and logistic plans. The optimal cost for deterministic model and higher tolerance model is equal at 16,462. In order to achieve this in stochastic setting, there are some plant product combination where production plans differs by more than 100. For example, products 1691700 and 1689547 in plants 3, 7 and 10 and product 1686762 in plant 12. In table 4.7, high tolerance model also uses an extra freight option, 39.

In low tolerance model, the focus is on staying close to the optimal deterministic plan. It is interesting to see the way production quantities are redistributed among different plants compared to high tolerance model. The production of 1689547 is reshuffled between plant 2,3 and 7. Since, the tolerance is low and capacity has randomness, at-times staying with-in the tolerance is not possible. The usage of dummy warehouse at a penalty cost is the only way of maintaining feasibility as in the last column of table 4.6. The low tolerance model tells us that if we maintain inventory for products 1688587 and 1691700, we can operate

Plant	Product Id	Quantity
2	1688587	303
	1682769	341
	1689547	717
	1689544	19
3	1660897	3804
	1695862	4925
	1660673	860
	1689548	541
4	1696073	779
9	1695862	110
10	1689544	355
Dummy	1699557	1141
	1699794	3507
	1653588	418
	1695862	95287
	1683419	334
	1660673	38
	1700139	26244
	1643057	689
	1682751	344
	1682736	309
	1688290	326
	1688598	275
Others	-	0

Table 4.3: Deterministic model: Optimal z_{di}

Freight option	Chosen(1) or not(0)
1128	1
1342	1
1158	1
1228	1
Dummy	1
Others	0

Table 4.4: Deterministic model: Optimal y_m

very close to the deterministic optimal production and logistic plans.

4.5 Conclusion and future work

In this paper, we provided a deterministic formulation for optimal outbound network routing. We showed the shortcomings of its optimal solution using a numerical experiment and designed a practical robust formulation to handle uncertain production capacity. We numerically studied the optimal solutions for low and high tolerance to deviation from deterministic optimal solution. In our numerical study, we assumed that the tolerance to change in logistic plan is same as to production plan. It could be worth while to study heterogeneous tolerances and its impact on optimal solution.

In this research we isolated randomness in production capacity and proposed a model to incorporate it in decision making. In reality, there are multiple sources of un-certainity such as demand, lead-time, pricing and many more. Random sampling can provide a better framework to tackle these multiple sources of un-certainity. We can create samples by selecting a fixed number of rows randomly from each excel tabs and use deterministic model repeatedly on each sample. Understanding the computational complexity and the accuracy of such a method compared to stochastic programming methods poses an interesting research question. Furthermore, we have assumed production capacity to be normally distributed. It could be worthwhile to generalize the model by using parameterized distribu-

Freight option	Plant	Entry port	Order ID	Quantity
1128	2	3	1447417878	303
	2	3	1447297184	341
	2	3	1447223367	218
	3	4	1447367967	1321
	9	4	1447403031	110
	3	4	1447403031	4925
	3	4	1447245203	860
	3	4	1447296431	541
	3	4	1447184536	2483
	2	3	1447365954	19
1228	10	2	1447365954	268
	2	3	1447223367	499
1158	10	2	1447365954	87
1342	4	5	1447237012	779
Dummy	Dummy	Dummy	1447254154	1141
			1447182403	3507
			1447265598	418
			1447403031	95287
			1447245203	38
			1447379081	26244
			1447373366	689
			1447329944	344
			1447383137	309
			1447203021	326
			1447232384	275
Others	-	-	-	0

Table 4.5: Deterministic model: Optimal x_{ijm}^{k}

Plant	Product Id	S_d	S_p	
			$\delta = 10^3$	$\delta = 10^2$
2	1689548	1373	1430	1398
	1689547			33
3	1688587	498	498	422
	1691700	1320	1516	1220
	1674853	286	286	286
	1657063	319	319	319
	1689547	3200	2867	3239
	1689012	1112	1112	1112
	1690616	285	285	285
	1696747	1242	1242	1242
	1651302	308	308	308
	1696746	594	594	594
	1672034	308	308	308
	1682739	518	518	518
	1668545	303	303	303
	1688285	344	344	344
5	1693453	3150	3099	3250
7	1691700	2650	2456	2616
8	1691700	147	145	145
10	1689547	236	645	240
12	1686762	1514	1624	1501
	1617714	374	374	374
Dummy	1688587			76
	1691700			136
Others	-	0	0	0

Table 4.6: Optimal z_{di} for different tolerance levels in stochastic solution(S_p) compared with deterministic solution(S_d)

Freight option	S_d	S_p	
		$\delta = 10^3$	$\delta = 10^2$
7	1	1	1
12	1	1	1
29	1	1	1
39	0	1	0
44	1	1	1
45	1	1	1
Dummy	0	0	1

Table 4.7: Optimal y_m for different tolerance levels in stochastic solution(S_p) compared with deterministic solution(S_d)

tion based on historical data and also use predictive machine learning methods to estimate future production capacity.

BIBLIOGRAPHY

[1] Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. Testing, voluntary social distancing, and the spread of an infection. *Operations Research*, 72(2):533–548, 2024.

[2] Vishal V Agrawal, L Beril Toktay, and Şafak Yücel. Non-ownership business models for solar energy. *Manufacturing & Service Operations Management*, 2022.

[3] R Ahmad. Introduction to stochastic differential equations, 1988.

[4] Saed Alizamir, Francis de Véricourt, and Peng Sun. Efficient feed-in-tariff policies for renewable energy technologies. *Operations Research*, 64(1):52–66, 2016.

[5] Hunt Allcott, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang. Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of public economics*, 191:104254, 2020.

[6] Linda JS Allen. Stochastic population and epidemic models. *Mathematical biosciences lecture series, stochastics in biological systems*, 1:120–128, 2015.

[7] Fernando Alvarez, David Argente, and Francesco Lippi. A simple planning problem for covid-19 lock-down, testing, and tracing. *American Economic Review: Insights*, 3(3):367–382, 2021.

[8] Nadia Ameli and Nicola Brandt. What impedes household investment in energy efficiency and renewable energy? 2015.

[9] Alexandar Angelus. Distributed renewable power generation and implications for capacity investment and electricity prices. *Production and Operations Management*, 30(12):4614–4634, 2021.

[10] Volodymyr Babich, Ruben Lobel, and Şafak Yücel. Promoting solar panel investments: Feed-in-tariff vs. tax-rebate policies. *Manufacturing & Service Operations Management*, 22(6):1148–1164, 2020.

[11] Norman TJ Bailey. *The mathematical theory of infectious diseases and its applications*. 1975.

[12] Abhijit Banerjee. Inequality and investment. *Massachusetts Institute of Technology. Cambridge, MA*, 2004.

[13] Galen L Barbose, Sydney Forrester, Eric O'Shaughnessy, and Naïm R Darghouth. Residen-

tial solar-adopter income and demographic trends: 2021 update. Technical report, Lawrence Berkeley National Lab., Berkeley,CA, 2021.

[14] Frank M Bass. A new product growth for model consumer durables. *Management science*, 15(5):215–227, 1969.

[15] R Battaglia. State auditor unveils new multi-state collaboration to improve covid-19 data accuracy. link, 2020.

[16] Gary S Becker and Casey B Mulligan. The endogenous determination of time preference. *The Quarterly Journal of Economics*, 112(3):729–758, 1997.

[17] K Breuninger. Gov. cuomo says the 'ultimate resolution' of coronavirus will come with vaccine in 18 months. link, 2020.

[18] Tom Britton. Stochastic epidemic models: a survey. *Mathematical biosciences*, 225(1):24–35, 2010.

[19] David P Brown and David EM Sappington. Designing compensation for distributed solar generation: Is net metering ever optimal? *The Energy Journal*, 38(3), 2017.

[20] Peter Buchen. *An Introduction to exotic option pricing*. CRC Press, 2012.

[21] Siyang Cai, Yongmei Cai, and Xuerong Mao. A stochastic differential equation sis epidemic model with two independent brownian motions. *Journal of mathematical analysis and applications*, 474(2):1536–1550, 2019.

[22] Abraham Charnes and William W Cooper. Chance-constrained programming. *Management science*, 6(1):73–79, 1959.

[23] Fabio Clementi and Mauro Gallegati. Pareto's law of income distribution: Evidence for germany, the united kingdom, and the united states. In *Econophysics of wealth distributions*, pages 3–14. Springer, 2005.

[24] Thomas Dangl. Investment and capacity choice under uncertain demand. *Eur.J. of Op. Res.*, 117(3):415–428, 1999.

[25] Ndiame Diop. Why is reducing energy subsidies a prudent, fair, and transformative policy for indonesia? *World Bank-Economic Premise*, (140):1–6, 2014.

[26] Avinash K Dixit and Robert S Pindyck. *Investment under uncertainty*. Princeton university press, 1994.

[27] DOE. Department of energy guidelines. link, 2021.

[28] Adrian Drăgulescu and Victor M Yakovenko. Evidence for the exponential distribution of

income in the usa. *The European Physical Journal B-Condensed Matter and Complex Systems*, 20(4):585–589, 2001.

[29] Kimon Drakopoulos and Ramandeep S Randhawa. Why perfect tests may not be worth waiting for: Information as a commodity. *Management Science*, 67(11):6678–6693, 2021.

[30] DSIRE. Dsire insight, nc clean energy technology center. link, 2021.

[31] DSIRE. Dsire insight, nc clean energy technology center. link, 2024.

[32] Heidi Bruderer Enzler, Andreas Diekmann, and Reto Meyer. Subjective discount rates in the general population and their predictive power for energy saving behavior. *Energy Policy*, 65:524–540, 2014.

[33] US FDA. Medical device shortages during the covid-19 public health emergency, 2020.

[34] Daniel Joseph Finkenstadt, Robert Handfield, and Peter Guinto. Why the us still has a severe shortage of medical supplies. *Harvard business review*, 9(20):20, 2020.

[35] Max Fisher. R0, the messy metric that may soon shape our lives, explained. *New York Times*, 2020.

[36] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 584(7820):257–261, 2020.

[37] Sydney Forrester, Galen L Barbose, Eric O'Shaughnessy, Naïm R Darghouth, and Cristina Crespo Montañés. Residential solar-adopter income and demographic trends: November 2022 update. Technical report, LBNL,CA, 2022.

[38] Kaiser Family Foundation. State covid-19 data and policy actions. *KFF State Vaccine Rollout*, 2021.

[39] Shana Kushner Gadarian, Sara Wallace Goodman, and Thomas B Pepinsky. Partisanship, health behavior, and policy attitudes in the early stages of the covid-19 pandemic. *Plos one*, 16(4):e0249596, 2021.

[40] K Gelles and George Petras. How ventilators work and why covid-19 patients need them to survive coronavirus. *USA Today*, 2020.

[41] Hélyette Geman and Marc Yor. Bessel processes, asian options, and perpetuities. *Mathematical finance*, 3(4):349–375, 1993.

[42] Kenneth Gillingham, Richard G Newell, and Karen Palmer. Energy efficiency economics and policy. *Annu. Rev. Resour. Econ.*, 1(1):597–620, 2009.

[43] Dan Goldberg. It's going to disappear": Trump's changing tone on coronavirus. *Politico, March*, 17, 2020.

[44] Alison Gray, David Greenhalgh, Liangjian Hu, Xuerong Mao, and Jiafeng Pan. A stochastic differential equation sis epidemic model. *SIAM Journal on Applied Mathematics*, 71(3):876–902, 2011.

[45] Neil A Halpern, Kay See Tan, et al. United states resource availability for covid-19. *Society of Critical Care Medicine*, 3:1–16, 2020.

[46] J Michael Harrison. *Brownian models of performance and control*. Cambridge University Press, 2013.

[47] Jerry A Hausman. Individual discount rates and the purchase and utilization of energy-using durables. *The Bell Journal of Economics*, pages 33–54, 1979.

[48] Jenny Heeter, Kaifeng Xu, and Gabriel Chan. Sharing the sun: Community solar deployment, subscription savings, and energy burden reduction [slides]. Technical report, National Renewable Energy Lab.,CO, 2021.

[49] Teck-Hua Ho, Sergei Savin, and Christian Terwiesch. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management science*, 48(2):187–206, 2002.

[50] Rodney P Jones. Would the united states have had too few beds for universal emergency care in the event of a more widespread covid-19 epidemic? *International Journal of Environmental Research and Public Health*, 17(14):5210, 2020.

[51] Christopher R Knittel and Bora Ozaltun. What does and does not correlate with covid-19 death rates. Technical report, National Bureau of Economic Research, 2020.

[52] RJ Kubiak. Decision making in energy efficiency investments-a review of discount rates and their implications for policy making. *Eceee Industrial Summer Study Proceedings*, 2016.

[53] Sunil Kumar and Jayashankar M Swaminathan. Diffusion of innovations under supply constraints. *Operations research*, 51(6):866–879, 2003.

[54] H Dharma Kwon. Invest or exit? optimal decisions in face of declining profit stream. *Op. Res.*, 58(3):638–649, 2010.

[55] H Dharma Kwon, Wenxin Xu, Anupam Agrawal, and Suresh Muthulingam. Impact of bayesian learning and externalities on strategic investment. *Management Science*, 62(2):550–570, 2016.

[56] Rachid Laajaj. Endogenous time horizon and behavioral poverty trap: Theory and evidence from mozambique. *Journal of Development Economics*, 127:187–208, 2017.

[57] Emily C Lawrance. Poverty and the rate of time preference: evidence from panel data. *Journal of Political Economy*, 99(1):54–77, 1991.

[58] Michael Leachman and Jennifer Sullivan. Some states much better prepared than others for recession. *Center on Budget and Policy Priorities*, 2020.

[59] Maia Martcheva. *An introduction to mathematical epidemiology*, volume 61. Springer, 2015.

[60] Sanjay Mehrotra, Hamed Rahimian, Masoud Barah, Fengqiao Luo, and Karolina Schantz. A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat covid-19. *Naval Research Logistics (NRL)*, 67(5):303–320, 2020.

[61] Sarah Mervosh, Denise Lu, and Vanessa Swales. See which states and cities have told residents to stay at home; 2020. *The New York Times, https://www. nytimes. com/interactive/2020/us/coronavirus-stay-at-home-order. html*, 2020.

[62] Bruce L Miller and Harvey M Wagner. Chance constrained programming with joint constraints. *Operations research*, 13(6):930–945, 1965.

[63] G Mulvihill. Us states share, get creative in hunt for medical supplies. *Ap News*, 2020.

[64] NCSL. National conference of state legislatures brief. link, 2021.

[65] NREL. National renewable energy lab.,co. link, 2021.

[66] NREL. National renewable energy lab.,co. link, 2024.

[67] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science, 2013.

[68] Siobhan Roberts. How to think like an epidemiologist. *New York Times*, 2020.

[69] L Chris G Rogers and Zo Shi. The value of an asian option. *Journal of Applied Probability*, 32(4):1077–1088, 1995.

[70] Kathleen Ronayne. California lends 500 ventilators to 4 states, 2 territories. *US News*, 2020.

[71] M Sanger-Katz, S Kliffs, and A Parlopiano. These places could run out of hospital beds as coronavirus spreads. *New York Times*, 2020.

[72] Alan Sanstad, Michael Hanemann, Maximillian Auffhammer, et al. End-use energy efficiency in a 'post-carbon'california economy: policy issues and research frontiers. *Managing greenhouse gas emissions in California*, pages 6–32, 2006.

[73] Igor Sazonov, Mark Kelbert, and Michael B Gravenor. A two-stage model for the sir outbreak: Accounting for the discrete and stochastic nature of the epidemic at the initial contamination stage. *Mathematical Biosciences*, 234(2):108–117, 2011.

[74] SEIA. Solar energy industries association brief. link, 2021.

[75] Moshe Shaked and J George Shanthikumar. *Stochastic orders*. Springer, 2007.

[76] Wenjing Shen, Izak Duenyas, and Roman Kapuscinski. New product diffusion decisions under supply constraints. *Management Science*, 57(10):1802–1810, 2011.

[77] Albert N Shiryaev. *Optimal stopping rules*, volume 8. Springer Science & Business Media, 2007.

[78] Steven E Shreve et al. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004.

[79] Siddharth Prakash Singh and Alan Scheller-Wolf. That's not fair: Tariff structures for electric utilities with rooftop solar. *Manufacturing & Service Operations Management*, 24(1):40–58, 2022.

[80] Shriya S Srinivasan, Khalil B Ramadi, Francesco Vicario, Declan Gwynne, Alison Hayward, David Lagier, Robert Langer, Joseph J Frassica, Rebecca M Baron, and Giovanni Traverso. A rapidly deployable individualized system for augmenting ventilator capacity. *Science translational medicine*, 12(549):eabb9401, 2020.

[81] D. Steinberg and et al Bielen, D. E. Electrification and decarbonization: exploring us energy use. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2017.

[82] Nur Sunar and Jayashankar M Swaminathan. Net-metered distributed renewable energy: A peril for utilities? *Management Science*, 67(11):6716–6733, 2021.

[83] Deborah A Sunter, Sergio Castellanos, and Daniel M Kammen. Disparities in rooftop photovoltaics deployment in the united states by race and ethnicity. *Nature Sustainability*, 2(1):71–76, 2019.

[84] Ryuta Takashima, Afzal S Siddiqui, and Shoji Nakada. Investment timing, capacity sizing, and technology choice of power plants. *Handbook of Networks in Power Systems I*, pages 303–321, 2012.

[85] Elisabetta Tornatore, Stefania Maria Buccellato, and Pasquale Vetro. Stability of a stochastic sir system. *Physica A: Statistical Mechanics and its Applications*, 354:111–126, 2005.

[86] Kenneth Train. Discount rates in consumers' energy-related decisions: a review of the literature. *Energy*, 10(12):1243–1253, 1985.

[87] Ramon Van Handel. Stochastic calculus, filtering, and stochastic control. *Course notes., URL http://www. princeton. edu/rvan/acm217/ACM217. pdf*, 14, 2007.

[88] Jan Vecer. A new approach for pricing arithmetic average asian options. *J. of Comp. Fin.*, 4(4):105–113, 2001.

[89] Roberto Verzola. Crossing over: the energy transition to renewable electricity. *(No Title)*, 2016.

[90] J Vom Scheidt. Gard, tc, introduction to stochastic differential equations. new york-basel, marcel dekker inc. 1988. xi, 234 pp., 78. – $.isbn0 - -8247 - 7776 - x(pureandappliedmathematics114), 1989$.

[91] Emilia Vynnycky and Richard White. *An introduction to infectious disease modelling*. Oxford university press, 2010.

[92] Worldbank. Worldbank. link, 2021.

[93] WSJ. California's solar-power welfare state:the rich object to a subsidies rollback for solar power. link, 2022.

[94] Luyi Yang, Shiliang Cui, and Zhongbin Wang. Design of covid-19 testing queues. *Production and Operations Management*, 31(5):2204–2221, 2022.

Appendix A

CHAPTER 2

Note: It is assumed that the stochastic process, I_t, which satisfies the given stochastic differential equation, is defined on an underlying probability space $(\Omega, \mathcal{F}, P)$, which is equipped with a filtration F and the stochastic process is adapted to F.

Proposition 1. *If I_{t_0} and I_{t_1} are the number of infections at the start of tracking of infections, t_0, and the number of infections at the start of shutdown, t_1, respectively and, (r_0, s_0) and (r_1, s_1) are the pre-shutdown and post-shutdown rate of infection and standard deviations, then for a given t,*

$$I_t = \begin{cases} I_{t_0} e^{(r_0 - \frac{s_0^2}{2})(t-t_0) + s_0(W_t - W_{t_0})} & t_0 \leq t < t_1 \\ I_{t_1} e^{(r_1 - \frac{s_1^2}{2})(t-t_1) + s_1(W_t - W_{t_1})} & t_1 \leq t \leq t_1 + T \end{cases}$$

Proof. The dynamics of spread of infection at $s \in (t_0, t_1)$ is given by the GBM,

$$dI_s = r_0 I_s ds + s_0 I_s dW_s$$

Let $Y_s = \log I_s$. Using Ito's Lemma

$$\log(I_t) - \log(I_{t_0}) = \int_{t_0}^{t} \frac{1}{I_s} dI_s - \int_{t_0}^{t} \frac{1}{2(I_s)^2} d[I, I]_s$$

Substituting $dI_s = r_0 I_s ds + s_0 I_s dW_s$ and $d[I, I]_s = s_0^2 I_s^2 ds$ in the above equation yields the following desired result

$$\log \frac{I_t}{I_{t_0}} = (r_0 - \frac{s_0^2}{2})(t - t_0) + s_0(W_t - W_{t_0})$$

$$I_t = I_{t_0} e^{(r_0 - \frac{s_0^2}{2})(t-t_0) + s_0(W_t - W_{t_0})}$$

The expression for I_t during post-shutdown period can be derived in similar way by using rate of infection, r_1 and standard deviation, s_1.

□

Proposition 2.

$$P(\widehat{I_T} > m) = 1 - \Phi\left(\frac{\frac{1}{s_i}\log\left(\frac{m}{I_{t_1}}\right) - \frac{1}{s_i}(r_i - \frac{s_i^2}{2})T}{\sqrt{T}}\right) + e^{2\frac{1}{s_i}\log\left(\frac{m}{I_{t_1}}\right)\frac{1}{s_i}(r_i - \frac{s_i^2}{2})}\Phi\left(\frac{-\frac{1}{s_i}\log\left(\frac{m}{I_{t_1}}\right) - \frac{1}{s_i}(r_i - \frac{s_i^2}{2})T}{\sqrt{T}}\right)$$

where $\widehat{I_T} = max_{t_1 \le t' \le t_1+T} I_{t'}$, *m is the preparation level capacity and T is the decision horizon. For* $i \in \{0, 1\}$, r_i *and* s_i *represents the rate and standard deviation of the underlying process during decision horizon.*

Proof. Consider $X_t = \alpha t + W_t$, where α is a given number and W_t is a standard Brownian motion. The distribution of running maximum of X_t is derived using change of measure and reflection property of Brownian motion. Let $\widehat{X_T} = max_{0 \le t \le T} X_t$, then [Corollary 7.2.2, Shreve]

$$P(\widehat{X_T} > b) = 1 - \Phi\left(\frac{b - \alpha T}{\sqrt{T}}\right) + e^{2\alpha b}\Phi\left(\frac{-b - \alpha T}{\sqrt{T}}\right) \tag{A.1}$$

Here, we consider the case in which the decision maker is looking into a decision horizon of lockdown period, T. So, the rate of spread and the standard deviation during T are r_1 and s_1, respectively. A similar argument can be made if the decision horizon is of no lockdown.

Let $\widehat{I_T} = max_{t_1 \le t' \le t_1+T} I_{t'}$. The probability of the maximum infection in the time horizon crossing the preparation level capacity, m, is denoted by $P(\widehat{I_T} > m)$.

$$P(\widehat{I_T} > m) = P(I_{t'} > m : t' \in [t_1, t_1 + T])$$

Using Proposition 1,

$$\begin{aligned} &= P\big(I_t e^{(r_1 - \frac{s_1^2}{2})(t'-t) + s_1 W_{t'-t_1}} > m : t' \in [t, t+T]\big) \\ &= P\left(\frac{1}{s_1}(r_1 - \frac{s_1^2}{2})(t'-t) + W_{(t'-t_1)} > \frac{1}{s_1}\log\left(\frac{m}{I_t}\right) : t' \in [t, t+T]\right) \end{aligned} \tag{A.2}$$

Comparing (1) and (2) yields the desired result such that $b = \frac{1}{s_1}\log(\frac{m}{I_t})$ and $\alpha = \frac{1}{s_1}(r_1 - \frac{s_1^2}{2})$.

□

Proposition 3. *If* $b = \frac{1}{s}\log\left(\frac{m}{I_t}\right)$, $\alpha = \frac{1}{s}(r_1 - \frac{s^2}{2})$ *and* $s_0 = s_1 = s$ *then for a given horizon,* T,

(i) There exists $\widehat{b}$, *such that if* $b > \widehat{b}$, *it is optimal not to shutdown irrespective of the* $\frac{C_e}{C_h}$ *preference level.*

(ii) There exists threshold $\widehat{\frac{C_e}{C_h}}$, *such that if* $\frac{C_e}{C_h} > \widehat{\frac{C_e}{C_h}}$, *it is optimal not to shut down*

(iii) Given the preference level, $\frac{C_e}{C_h}$, *such that* $0 < \frac{C_e}{C_h} < \widehat{\frac{C_e}{C_h}}$, *there exist* b^*, *such that if* $b < b^*$, *it is optimal to shut down.*

(iv) There exists $\widehat{T}$, *such that if* $T < \widehat{T}$, *then it is optimal not to shut down.*

Proof. We know that $r_1 < r_0$, which implies $\alpha_1 < \alpha_0$. Using Proposition 2, Let

$$\begin{aligned}\Delta P(b) = e^{2\alpha_0 b}\Phi\left(\frac{-b-\alpha_0 T}{\sqrt{T}}\right) - \Phi\left(\frac{b-\alpha_0 T}{\sqrt{T}}\right) \\ -e^{2\alpha_1 b}\Phi\left(\frac{-b-\alpha_1 T}{\sqrt{T}}\right) + \Phi\left(\frac{b-\alpha_1 T}{\sqrt{T}}\right)\end{aligned} \tag{A.3}$$

The sum, difference and product of continuous and differentiable functions are continuous and differentiable. This implies $\Delta P(m)$ is continuous with respect to m on $[0,\infty)$ and differentiable on $(0,\infty)$.

(i) Consider an increasing sequence, b_n, such that $b_n \in (0,\infty)$ and $\lim_{n\to\infty} b_n = \infty$. From (3), we observe that $\lim_{n\to\infty}\Delta P(b_n) = 0$. Since $\Delta P(b)$ is a continuous function of b, $\Delta P(\lim_{n\to\infty} b_n) = 0$. This implies there exists a finite N, such that for a given $\frac{C_e}{C_h} > 0$, $|\Delta P(b_n)| < \frac{C_e}{C_h}$ for all $n > N$. This implies that there exists a threshold, $\widehat{b} = b_N$, such that it is optimal not to lockdown if $b > \widehat{b}$. Contextually, it implies that if the capacity is high enough, there is no benefit of a lockdown.

(ii) From (3), we observe that $\Delta P(0) = 0$ and $lim_{b\to\infty}\Delta P(b) = 0$. Using Leibniz integral rule,

$$\begin{aligned}\frac{\partial \Delta P}{\partial \alpha} &= 2be^{2\alpha b}\Phi\left(\frac{-b-\alpha T}{\sqrt{T}}\right) - \sqrt{T}e^{2\alpha b}e^{\frac{-(b+\alpha T)^2}{2T}} + \sqrt{T}e^{\frac{-(b-\alpha T)^2}{2T}} \\ &= 2be^{2\alpha b}\Phi\left(\frac{-b-\alpha T}{\sqrt{T}}\right) \geq 0\end{aligned} \tag{A.4}$$

Since ΔP is increasing in α for a given b and $\alpha_1 < \alpha_0$, $\Delta P(b) \geq 0$. Combining $\Delta P(b) \geq 0$, $\Delta P(0) = 0$ and $\lim_{b\to\infty} \Delta P(b) = 0$ and using Rolle's theorem, it implies there exists a $b_c \in (0, \infty)$ such that $\widehat{\frac{C_e}{C_h}} = \Delta P(b_c)$ is a global maximum. If $\frac{C_e}{C_h} > \widehat{\frac{C_e}{C_h}}$, using the condition for not shutting down in the main text, the optimal policy is not to shut down.

(iii) From (ii), $\Delta P(b_c) = \widehat{\frac{C_e}{C_h}}$ and $\Delta P(b_c) = 0$. $\Delta P(b)$ is continuous with respect to b on $[0, \infty)$. Using Intermediate Value Theorem, there exists $b_1^* \in (0, b_c)$ and $b_2^* \in (b_c, \infty)$ such that $\Delta P(b_1^*) = \Delta P(b_2^*) = \frac{C_e}{C_h}$. Let $B = \{b : \Delta P(b) = \frac{C_e}{C_h}\}$. We know that B is a non empty set, since b_1^* and b_2^* belong to B. Let $b^* = max\{b : b \in B\}$. It can be seen from the definition of b^* that $\{b : \Delta P(b) > \frac{C_e}{C_h}\} \subset [0, b^*]$. The cost at any $b \in B$ is given $C_h P(\widehat{I_T} > b : u_0)$ since the total cost of lockdown is same as no lockdown. The set $\{\widehat{I_T} > b^* : u_0\} \subset \{\widehat{I_T} > b : u_0\}$ for all $b \in B$. Using the monotonicity property of probability measure, we know that $P(\widehat{I_T} > b^* : u_0) \leq P(\widehat{I_T} > b : u_0)$ for all $b \in B$. This implies that the cost, $C_h P(\widehat{I_T} > b : u_0)$, is minimum at b^* for all $b \in B$. If we combine the above arguments, we can conclude that if $b < b^*$, it is optimal to shut down.

We have assumed that $b = \frac{1}{s} \log\left(\frac{m}{I_t}\right)$. Since, log is a monotonic function, for a given m, there exist a one to one mapping between b and I_t. Similarly, for given I_t, there exits a one to one mapping between b and m. This enables us to extend the results to m and I_t. We can say that for given I_t, there exists $\widehat{m}$ such that if $m > \widehat{m}$, it is optimal not to shut down, irrespective of $\frac{C_e}{C_h}$ and under the condition of (iii), there exists a m^*, such that if $m < m^*$, it is optimal to shut down. For a given m, there exists a $I_{\widehat{\tau h}}$ such that if $I_t < I_{\widehat{\tau h}}$, it is optimal not to shut down, irrespective of $\frac{C_e}{C_h}$ and under the condition of (iii), there exists a $I_{\tau h^*}$, such that if $I_t > I_{\tau h^*}$, it is optimal to shut down.

(iv) We observe that (3) tends to 0 as T tends to 0. Repeating the concept used in (i) with a decreasing sequence defined in time, it can be shown that any C_e/C_h is greater than ΔP for $T < \widehat{T}$. This implies that if $T < \widehat{T}$, it optimal not to shut down. □

Proposition 4. *Given $u(x) = cI_t$ and $\phi(x) = C_e - C_h\Delta\widehat{P(I_\tau)}$, there exists I_t^* for a given capacity level, such that it is optimal to keep the state open if the number of infections are less than I_t^*.*

Proof. Let V(x) be the candidate value function. V(x) is a solution to

$$min\{r_0 x\frac{\partial V(x)}{\partial x} + \frac{\sigma_0^2}{2}x^2\frac{\partial^2 V(x)}{\partial x^2} + cx, C_e - C_h\Delta\widehat{P(I_\tau)} - V(x)\} = 0$$

Then we have

$$r_0 x\frac{\partial V(x)}{\partial x} + \frac{\sigma_0^2}{2}x^2\frac{\partial^2 V(x)}{\partial x^2} + cx = 0 \qquad V(x) < C_e - C_h\Delta\widehat{P(x)} \qquad \forall x \in D \tag{A.5}$$

$$r_0 x\frac{\partial V(x)}{\partial x} + \frac{\sigma_0^2}{2}x^2\frac{\partial^2 V(x)}{\partial x^2} + cx \geq 0 \qquad V(x) = C_e - C_h\Delta\widehat{P(x)} \qquad \forall x \in D^c \tag{A.6}$$

First, we solve for V(x) on the continuation region, D, of the stopping time τ^*. We define τ^* as the first exit time of D i.e $\tau^* = \inf\{t \geq 0 : I_t \notin D\}$.

We can guess the solution to be of the form

$$V_s(x) = -\frac{cx}{r_0} + \frac{x^{(1-\frac{2r_0}{s^2})}}{1-\frac{2r_0}{s^2}} + d$$

where d is a constant. We can substitute the guessed valued in the original differential equation on D to check its validity. We assume here that the per infection cost rate is less than C_e otherwise $\tau = 0$ trivially i.e. it is optimal to shutdown the economy if $I_{t_0} \geq 1$. To find the value of x^* and d, we use the principle of smooth fit, which requires that V(x) and V'(x) are continuous at the boundary of D.

$$\lim_{\epsilon\to 0^+} V_s^{'}(x^* - \epsilon) = \lim_{\epsilon\to 0^+} V_s^{'}(x^* + \epsilon)$$

Since $x^* + \epsilon \in D^c$, using (9), $V_s'(x^* + \epsilon) = (C_e - C_h\Delta\widehat{P(x^*)})'$.
This implies that x^* satisfies

$$-\frac{c}{r_0} + \frac{1}{x^{\frac{2r_0}{s^2}}} = -C_h\frac{\partial}{\partial x}\Delta\widehat{P(x)} \tag{A.7}$$

Using proposition 3,

$$-\Delta\widehat{P(x)} = e^{2\alpha_1 \frac{\ln(m/x)}{s}} \Phi\left(\frac{-\frac{\ln(m/x)}{s} - \alpha_1 T}{\sqrt{T}}\right) - \Phi\left(\frac{\frac{\ln(m/x)}{s} - \alpha_1 T}{\sqrt{T}}\right) - e^{2\alpha_0 \frac{\ln(m/x)}{s}} \Phi\left(\frac{-\frac{\ln(m/x)}{s} - \alpha_0 T}{\sqrt{T}}\right) + \Phi\left(\frac{\frac{\ln(m/x)}{s} - \alpha_0 T}{\sqrt{T}}\right)$$

$$-C_h \frac{\partial}{\partial x}\Delta\widehat{P(x)} = C_h\Bigg(- \frac{2\alpha_1}{sx} e^{2\alpha_1 \frac{\ln(m/x)}{s}} \Phi\left(\frac{-\frac{\ln(m/x)}{s} - \alpha_1 T}{\sqrt{T}}\right) + \frac{1}{sx}\sqrt{\frac{2}{\pi T}} e^{-\frac{1}{2T}\left(\frac{\ln(m/x)}{s} - \alpha_1 T\right)^2} + \frac{2\alpha_0}{sx} e^{2\alpha_0 \frac{\ln(m/x)}{s}} \Phi\left(\frac{-\frac{\ln(m/x)}{s} - \alpha_0 T}{\sqrt{T}}\right) - \frac{1}{sx}\sqrt{\frac{2}{\pi T}} e^{-\frac{1}{2T}\left(\frac{\ln(m/x)}{s} - \alpha_0 T\right)^2}\Bigg)$$

For $\left(r_0 - \frac{s^2}{2}\right) \geq 0$ and $x > 0$, $-\frac{c}{r_0} + \frac{1}{x^{\frac{2r_0}{s^2}}}$ is continuous and non-increasing function on x with limit tending to $-\frac{c}{r_0}$ as x tends to infinity. The function $-C_h \frac{\partial}{\partial x}\Delta\widehat{P(x)}$ is continuous and tends to zero as x tends to infinity. Since, $-C_h \frac{\partial}{\partial x}\Delta\widehat{P(x)} < -\frac{c}{r_0} + \frac{1}{x^{\frac{2r_0}{s^2}}}$ in the delta neighborhood of zero, using intermediate value theorem, there exists a x^* such that (7) holds. We can find the value of d using x^* and the following continuity condition of smooth fit.

$$V_s(x^*) = (C_e - C_h\Delta\widehat{P(x^*)})$$

□

Proposition 5. *Assuming* $r_1 - \frac{(s_1)^2}{2} \geq 0$ *and* $I_t \leq m$
(a) $\frac{\partial PrB(I_t,T;r_1,s_1,m)}{\partial r_1} \geq 0$ *(b)* $\frac{\partial PrB(I_t,T;r_1,s_1,m)}{\partial I_t} \geq 0$

Proof. (a) Let $\alpha = \frac{1}{s_i}\log\left(\frac{m}{I_t}\right)$. Using Proposition 2. and Leibniz rule

$$\frac{\partial PrB}{\partial r_1} = \frac{1}{s\sqrt{2\pi T}} e^{\frac{-(\alpha - \frac{1}{s_1}(r_1 - \frac{(s_1)^2}{2})T)^2}{2T}} + \frac{2\alpha}{s} e^{\frac{2\alpha(r_1 - \frac{s^2}{2})}{s}} \Phi\left(\frac{-\alpha - \frac{1}{s_1}(r_1 - \frac{(s_1)^2}{2})T}{\sqrt{T}}\right) - \frac{1}{s\sqrt{2\pi T}} e^{\frac{-(\alpha - \frac{1}{s_1}(r_1 - \frac{(s_1)^2}{2})T)^2}{2T}}$$

Since second term is greater than equal to 0 since $\alpha \geq 0$, (a) holds true. Part (b) can be proven in a similar way. □

Proposition 6. *If I_t^a and I_t^b represent the dynamics of the spread of infection in two states, A and B respectively, then given a time horizon, T,*

$$P(\widehat{I_T^a + I_T^b} > m_a + m_b) \leq P(\widehat{I_T^a} + \widehat{I_T^b} > m_a + m_b) \leq P(\widehat{I_T^a} > m_a) + P(\widehat{I_T^b} > m_b)$$

where $\{\widehat{I_T^a}\}$ and $\{\widehat{I_T^b}\}$ is the maximum number of infections in state A and B respectively during period T, $\widehat{I_T^a + I_T^b}$ is the maximum of the combined number of infections in two states and, m_a and m_b are the preparation level in state A and B respectively.

Proof. We assume that both the stochastic processes are defined in the same probability space. Using the countable sub-additivity of the probability measure,

$$P(\{\widehat{I_T^a} > m_a\} \cup \{\widehat{I_T^b} > m_b\}) \leq P(\{\widehat{I_T^a} > m_a\}) + P(\{\widehat{I_T^b} > m_b\})$$

Since $\{\widehat{I_T^a + I_T^b} > m_a + m_b\} \subset \{\widehat{I_T^a} + \widehat{I_T^b} > m_a + m_b\} \subset \{\widehat{I_T^a} > m_a\} \cup \{\widehat{I_T^b} > m_b\}$, applying the monotonicity property of the probability measure,

$$P(\{\widehat{I_T^a + I_T^b} > m_a + m_b\}) \leq P(\{\widehat{I_T^a} + \widehat{I_T^b} > m_a + m_b\}) \leq P(\{\widehat{I_T^a} > m_a\} \cup \{\widehat{I_T^b} > m_b\})$$

Combing the above two equations yields the desired result.

□

Proposition 7. *For two identical regions, there exists a preparation level threshold $\overline{m_1}$, such that $P(\widehat{I_t^a + I_t^b} > m + m) < P(\widehat{I_t^b} > m) = P(\widehat{I_t^a} > m)$ if $m > \overline{m_1}$.*

Proof. Notice that for identical regions the rate of spread of infection, standard deviation of spread and the initial number of infections are same. The only difference is that the underlying Brownian motions are different and independent. This enables us to say that, at any given time, t, $\widehat{I_t^a}$ and $\widehat{I_t^b}$ are independent and identically distributed random variables. Hence, $\widehat{I_t^a}$ and $\widehat{I_t^b}$ are exchangeable random variables.

Let a and b be two vector of constants such that $a = (1, 1)$ and $b = (0, 2)$. Using Theorem 3.A.35 (Shaked and Shantikumar), $\widehat{I_t^a} + \widehat{I_t^b} \leq_{cx} \widehat{2I_t^b}$, where cx denotes the convex ordering. According to Theorem 3.A.44.(Shaked and Shantikumar), if X and Y are

non-negative random variables, with equal means and $X \leq_{cx} Y$ with distribution function F and G respectively, then $X \leq_{cx} Y$ implies that $S^-(\overline{F} - \overline{G}) = 1$ and sign changes from + to -. Here $S^-(x)$ is the number of sign changes. Since $\widehat{I_t^a} + \widehat{I_t^b}$ and $2I_t^b$ satisfies the above conditions, we can say that there exits a $\overline{m_2}$ such that $P(\widehat{I_t^a} + \widehat{I_t^b} > 2m) < P(2\widehat{I_t^b} > 2m)$ if $m > \overline{m_2}$.

Using the same procedure with I_t^a and I_t^b as the initial exchangeable random variables and using the same a and b vectors, we can claim that there exits a m_2 such that $P(I_t^a + I_t^b > 2m) < P(2I_t^b > 2m)$ if $m > m_2$. Through Proposition 5, we know that $P(I_t^a + I_t^b > 2m) \leq P(\widehat{I_t^a + I_t^b} > 2m) \leq P(\widehat{I_t^a} + \widehat{I_t^b} > 2m)$. This implies that there exits a $\overline{m_1}$ in the interval $[m_2, \overline{m_2}]$ such that $P(\widehat{I_t^a + I_t^b} > m + m) < P(\widehat{I_t^b} > m)$ if $m > \overline{m_1}$.

□

Proposition 8. *Given that r^a, r^b, s^a, s^b and $l \in [0,1]$ are known positive constants, W_t^a and W_t^b are correlated Wiener processes such that $[W^a, W^b]_t = \rho t$, where $\rho \in (-1,1)$, and I_0^a and I_0^b are the initial points, the coupled equations*

$$\begin{aligned} dI_t^a &= r^a((1-l)I_t^a + lI_t^b)dt + s^a I_t^a dW_t^a & I_0^a \\ dI_t^b &= r^b(lI_t^a + (1-l)I_t^b)dt + s^b I_t^b dW_t^b & I_0^b \end{aligned}$$

has a unique positive solution if

$$max\left\{r^a(1-l) - r^b(1-l) + \frac{(s^b)^2}{2}, -r^a(1-l) + r^b(1-l) + \frac{(s^a)^2}{2}\right\} - \frac{\rho s^a s^b}{2} + l(r^a + r^b) > 0$$

Proof. Let $I_t = \begin{bmatrix} I_t^a \\ I_t^b \end{bmatrix}$, $\mu = \begin{bmatrix} r^a(1-l) & r^a l \\ r^b l & r^b(1-l) \end{bmatrix}$, $\sigma = \begin{bmatrix} s^a & 0 \\ 0 & s^b \end{bmatrix}$ and $W_t = \begin{bmatrix} W_t^a \\ W_t^b \end{bmatrix}$. We can rewrite the coupled equation in vector notation as

$$dI_t = \mu I_t dt + \sigma I_t dW_t$$

Since I_t is continuous and, μ and σ are constants, it is straight forward to see that both At-most linear growth and Lipschitz continuity conditions are satisfied. This implies the coupled SDE equations have a unique solution.

In order to show the positivity of the unique solution, we start with defining a sequence of stopping times. Let $\tau_k^x = \inf\{t : I_t^x \notin (1/k, \infty)\}$ and $\tau_k^y = \inf\{t : I_t^y \notin (1/k, \infty)\}$. Let $\tau_k = \tau_k^x \wedge \tau_k^y$. τ_k is a stopping time, since minimum of two stopping times is a stopping time. If $\lim_{k\to\infty} \tau_k = \infty$ a.s, then both I_t^a and I_t^b are positive a.s. We will prove this by contradiction. Let there exist $N > 0$ and $\epsilon > 0$, such that $P(\lim_{k\to\infty} \tau_k \leq N) \geq \epsilon$. This implies that there exits k_1 such that $P\{\tau_k \leq N\} \geq \epsilon$ for $k \geq k_1$.

Consider a continuously differentiable function $V : (0,\infty)\mathrm{x}(0,\infty) \to R^+$. Using the martingale property and Ito's lemma,

$$\mathbb{E}\big[V(I^a_{t\wedge\tau_k}, I^b_{t\wedge\tau_k})\big] = V(I_0^a, I_0^b) + \mathbb{E}\bigg[\int_0^{t\wedge\tau_k} \mathcal{L}V(I_s^a, I_s^b)ds\bigg] \tag{A.8}$$

where $\mathcal{L}$ is the infinitesimal generator of the coupled equations. The derivation of $\mathcal{L}$ is a straight forward application of multi dimensional Ito's lemma.

$$\mathcal{L} = r^a((1-l)x + ly)\frac{\partial}{\partial x} + r^b(lx + (1-l)y)\frac{\partial}{\partial y} + \frac{(s^a)^2x^2}{2}\frac{\partial^2}{\partial x^2} + \frac{(s^b)^2y^2}{2}\frac{\partial^2}{\partial y^2} + \frac{\rho s^a s^b xy}{2}\frac{\partial^2}{\partial y \partial x}$$

Let $V(x,y) = \frac{x}{y} + \frac{y}{x}$.

$$\begin{aligned}\mathcal{L}V(x,y) =& \bigg(r^a(1-l) - r^b(1-l) + \frac{(s^b)^2}{2}\bigg)\frac{x}{y} + \bigg(-r^a(1-l) + r^b(1-l) + \frac{(s^a)^2}{2}\bigg)\frac{y}{x} \\ &+ l(r^a + r^b) - l\bigg(r^a\frac{y^2}{x^2} + r^b\frac{x^2}{y^2}\bigg) - \frac{\rho s^a s^b}{2}\bigg(\frac{x}{y} + \frac{y}{x}\bigg)\end{aligned} \tag{A.9}$$

Since $l \in [0,1]$, $r^a \geq 0$ and $r^b \geq 0$,

$$\begin{aligned}\mathcal{L}V(x,y) \leq& \bigg(r^a(1-l) - r^b(1-l) + \frac{(s^b)^2}{2}\bigg)\frac{x}{y} + \bigg(-r^a(1-l) + r^b(1-l) + \frac{(s^a)^2}{2}\bigg)\frac{y}{x} \\ &+ l(r^a + r^b) - \frac{\rho s^a s^b}{2}\bigg(\frac{x}{y} + \frac{y}{x}\bigg)\end{aligned} \tag{A.10}$$

Let $C = max\{r^a(1-l) - r^b(1-l) + \frac{(s^b)^2}{2}, -r^a(1-l) + r^b(1-l) + \frac{(s^a)^2}{2}\}$

$$\begin{aligned}\mathcal{L}V(x,y) &\leq C\left(\frac{x}{y}+\frac{y}{x}\right)-\frac{\rho s^a s^b}{2}\left(\frac{x}{y}+\frac{y}{x}\right)+l(r^a+r^b)\\ &\leq \left(C-\frac{\rho s^a s^b}{2}\right)\left(\frac{x}{y}+\frac{y}{x}\right)+l(r^a+r^b)\\ &\leq SV(x,y) \qquad \text{(A.11)}\end{aligned}$$

where, $S = C - \frac{\rho s^a s^b}{2} + l(r^a + r^b)$. The last inequality follows since $V(x, y) \geq 2$ for $x > 0$ and $y > 0$.

Assuming $S > 0$ and substituting (11) in (8)

$$\mathbb{E}\big[V(I^a_{t\wedge\tau_k}, I^b_{t\wedge\tau_k})\big] \leq V(I^a_0, I^b_0) + S\mathbb{E}\bigg[\int_0^{t\wedge\tau_k} V(I^a_s, I^b_s)ds\bigg] \leq V(I^a_0, I^b_0) + S\mathbb{E}\bigg[\int_0^{t} V(I^a_{s\wedge\tau_k}, I^b_{s\wedge\tau_k})ds\bigg]$$

Using Gronwall's inequality

$$\mathbb{E}\big[V(I^a_{N\wedge\tau_k}, I^b_{N\wedge\tau_k})\big] \leq V(I^a_0, I^b_0)e^{SN} \qquad \text{(A.12)}$$

(12) shows that $\mathbb{E}\big[V(I^a_{N\wedge\tau_k}, I^b_{N\wedge\tau_k})\big]$ is bounded. We also observe that

$$\begin{aligned}\mathbb{E}\big[V(I^a_{N\wedge\tau_k}, I^b_{N\wedge\tau_k})\big] &\geq \mathbb{E}\big[1_{\{\tau_k\leq N\}}V(I^a_{\tau_k}, I^b_{\tau_k})\big]\\ &\geq \mathbb{E}\big[1_{\{\tau_k\leq N\}}V(I^a_{\tau_k}, I^b_{\tau_k})\big]\\ &\geq \mathbb{E}\big[1_{\{\tau_k\leq N\}\cap\{\tau^x_k<\tau^y_k\}}V(I^a_{\tau_k}, I^b_{\tau_k}) + 1_{\{\tau_k\leq N\}\cap\{\tau^x_k>\tau^y_k\}}V(I^a_{\tau_k}, I^b_{\tau_k})\big]\\ &\geq \mathbb{E}\big[1_{\{\tau_k\leq N\}\cap\{\tau^x_k<\tau^y_k\}}\left(kI^b_{\tau^x_k}+\frac{1}{kI^b_{\tau^x_k}}\right) + 1_{\{\tau_k\leq N\}\cap\{\tau^x_k>\tau^y_k\}}\left(kI^a_{\tau^y_k}+\frac{1}{kI^a_{\tau^y_k}}\right)\big]\\ &\geq k\Bigg(\mathbb{P}(\{\tau_k\leq N\}\cap\{\tau^x_k<\tau^y_k\}\cap\{I^b_{\tau^x_k}\in(1,\infty)\}) + \mathbb{P}(\{\tau_k\leq N\}\cap\{\tau^x_k>\tau^y_k\}\cap\{I^a_{\tau^y_k}\in(1\end{aligned}$$

Since $\mathbb{P}(\{\tau_k \leq N\}\cap\{\tau^x_k < \tau^y_k\}\cap\{I^b_{\tau^x_k} \in (1,\infty)\})+\mathbb{P}(\{\tau_k \leq N\}\cap\{\tau^x_k > \tau^y_k\}\cap\{I^a_{\tau^y_k} \in (1,\infty)\}) > 0$, as k tends to infinity, right side tends to infinity. This implies that $\mathbb{E}\big[V(I^a_{N\wedge\tau_k}, I^b_{N\wedge\tau_k})\big]$ is unbounded, which is a contradiction. We can infer that $lim_{k\to\infty}\tau_k = \infty$. □

Appendix B

CHAPTER 3

Retail cost of electricity per unit per unit time	p_b
Cost of solar subscription per unit capacity	p_{sub}
Fixed cost of installing Roof-top solar	K
Variable cost of installing Rooftop solar per unit capacity	k
Installed solar capacity	c
Duration of solar billing cycle	t_b
Electricity produced by capacity c during t_b	$\eta(c, t_b)$
Subsidy(%) for solar product i	δ_i

Lemma 1. *Given* $\lambda_\delta^* = -\frac{1}{t_b} \ln\left(1 - \frac{(1-\delta_2)p_{sub}c}{(1-\delta_1)(K+kc)}\right)$,

1. *if* $\lambda_\delta^* \geq \overline{\lambda}$ *then* $r^*(\delta) = 0$.
2. *if* $\lambda_\delta^* \leq \underline{\lambda}$ *then* $r^*(\delta) = \infty$.
3. *if* $\underline{\lambda} < \lambda_\delta^* < \overline{\lambda}$ *then there exists* $r^*(\delta) \in (0, \infty)$.

Proof. Consider a function $\Delta(r) := (1-\delta_1)K + kc - (1-\delta_2)\sum_{n=0}^{\infty} e^{-n\lambda(r)t_b} p_{sub}c$.

$$\frac{\partial \Delta(r)}{\partial r} = (1-\delta_2)\frac{e^{-\lambda(r)t_b}}{(1-e^{-\lambda(r)t_b})^2} p_{sub} c t_b \frac{\partial \lambda(r)}{\partial r} \tag{B.1}$$

We notice that $\Delta(r)$ is a continuous decreasing function of r, since $\frac{\partial \lambda(r)}{\partial r} < 0$. This implies that the difference in cost of roof-top installation and subscription solar decreases in r. Next, we discusses 3 possible cases.

Case 1: If $\lambda_\delta^* > \overline{\lambda}$, then

$$-\frac{1}{t_b}\ln\left(1-\frac{(1-\delta_2)p_{sub}c}{(1-\delta_1)(K+kc)}\right) > \overline{\lambda}. \tag{B.2}$$

$$(1-e^{-\overline{\lambda}t_b}) < \frac{(1-\delta_2)p_{sub}c}{(1-\delta_1)(K+kc)}. \tag{B.3}$$

$$(1-\delta_1)(K+kc) - \frac{(1-\delta_2)p_{sub}c}{1-e^{-\overline{\lambda}t_b}} < 0. \tag{B.4}$$

The function Δ is decreasing and continuous in r and λ is decreasing continuous bounded function of r. Since $\lambda(r) \leq \overline{\lambda}$, the following holds for all r.

$$\Delta(r) \leq (1-\delta_1)(K+kc) - (1-\delta_2)\frac{p_{sub}c}{1-e^{-\overline{\lambda}t_b}} < 0 \tag{B.5}$$

and $r^*(\delta) = 0$.

Case 2: If $\lambda_\delta^* < \underline{\lambda}$, then

$$-\frac{1}{t_b}\ln\left(1-\frac{(1-\delta_2)p_{sub}c}{(1-\delta_1)(K+kc)}\right) < \underline{\lambda}. \tag{B.6}$$

$$(1-e^{-\underline{\lambda}t_b}) > \frac{(1-\delta_2)p_{sub}c}{(1-\delta_1)(K+kc)}. \tag{B.7}$$

$$(1-\delta_1)(K+kc) - \frac{(1-\delta_2)p_{sub}c}{1-e^{-\underline{\lambda}t_b}} > 0. \tag{B.8}$$

The function Δ is decreasing and continuous in r and λ is decreasing continuous bounded function of r. Since $\lambda(r) \geq \underline{\lambda}$, the following holds for all r.

$$\Delta(r) \geq (1-\delta_1)(K+kc) - (1-\delta_2)\frac{p_{sub}c}{1-e^{-\underline{\lambda}t_b}} > 0 \tag{B.9}$$

and $r^*(\delta) = \infty$.

Case 3: If $\underline{\lambda} \leq \lambda_\delta^* \leq \overline{\lambda}$, then there exist r' and r'' such that $\Delta(r') < 0$ and $\Delta(r'') > 0$. Using intermediate value theorem, there exist a $r^*(\delta)$ such that $\Delta(r^*(\delta)) = 0$. Using the monotonicity of Δ function we have $\Delta(r) \geq 0$ for $r \leq r^*(\delta)$ and $\Delta(r) \leq 0$ for $r \geq r^*(\delta)$. □

Lemma 2. *Assuming* $\underline{\lambda} < \lambda_\delta^* < \overline{\lambda}$,

$$g(x;r,\delta) = \begin{cases} A(r)x - B(r) + \frac{(1-\delta_2)p_{sub}c}{1-e^{-\lambda(r)t_b}} & , r < r^*(\delta) \\ A(r)x - B(r) + (1-\delta_1)(K+kc) & , r \geq r^*(\delta). \end{cases}$$

where,

$$A(r) = \frac{p_b}{\mu(r)} \frac{(1 - e^{-\mu(r)t_b})}{e^{(\lambda(r)-\mu(r))t_b} - 1}. \tag{B.10}$$

$$B(r) = \frac{p_b}{e^{\lambda(r)t_b} - 1} \eta(c, t_b). \tag{B.11}$$

Proof. Consider the case when $r \geq r^*(\delta)$

$$g(x; r, \delta) = E_x \left[\sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} p_b \left(\int_{(n-1)t_b}^{nt_b} X_s^r \mathtt{d}s - \eta(c, t_b) \right) \right] + (1 - \delta(r))(K + kc). \tag{B.12}$$

$$= \sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} p_b \left(\int_{(n-1)t_b}^{nt_b} E_x \left[X_s^r \right] \mathtt{d}s - \eta(c, t_b) \right) + (1 - \delta(r))(K + kc). \tag{B.13}$$

$$= \sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} p_b \left(\int_{(n-1)t_b}^{nt_b} x e^{\mu(r)s} \mathtt{d}s - \eta(c, t_b) \right) + (1 - \delta(r))(K + kc). \tag{B.14}$$

$$= \sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} p_b \left(\frac{x}{\mu(r)} (e^{n\mu(r)t_b} - e^{(n-1)\mu(r)t_b}) - \eta(c, t_b) \right) + (1 - \delta(r))(K + kc). \tag{B.15}$$

$$= \frac{p_b x}{\mu(r)} \sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} (e^{n\mu(r)t_b} - e^{(n-1)\mu(r)t_b}) - p_b \eta(c, t_b) \sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} + (1 - \delta(r))(K + kc). \tag{B.16}$$

$$= \frac{p_b x}{\mu(r)} (1 - e^{-\mu(r)t_b}) \sum_{n=1}^{\infty} e^{n(\mu(r)-\lambda(r))t_b} - p_b \eta(c, t_b) \sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} + (1 - \delta(r))(K + kc). \tag{B.17}$$

$$= \frac{p_b}{\mu(r)} \frac{(1 - e^{-\mu(r)t_b})}{e^{(\lambda(r)-\mu(r))t_b} - 1} x - \frac{p_b}{e^{\lambda(r)t_b} - 1} \eta(c, t_b) + (1 - \delta(r))(K + kc). \tag{B.18}$$

Now, for the case when $r < r^*(\delta)$,

$$g(x; r, \delta) = E_x \left[\sum_{n=1}^{\infty} e^{-n\lambda(r)t_b} p_b \left(\int_{(n-1)t_b}^{nt_b} X_s^r \mathtt{d}s - \eta(c, t_b) \right) \right] + (1 - \delta(r)) \sum_{n=0}^{\infty} e^{-n\lambda(r)t_b} p_{sub} c. \tag{B.19}$$

$$= \frac{p_b x}{\mu(r)} (1 - e^{-\mu(r)t_b}) \sum_{n=1}^{\infty} e^{n(\mu(r)-\lambda(r))t_b} - p_b \eta(c, t_b) \sum_{n=1}^{\infty} e^{-n\lambda t_b} + (1 - \delta(r)) p_{sub} c \sum_{n=0}^{\infty} e^{-n\lambda(r)t_b}. \tag{B.20}$$

$$= \frac{p_b}{\mu(r)} \frac{(1 - e^{-\mu(r)t_b})}{e^{(\lambda(r)-\mu(r))t_b} - 1} x - \frac{p_b}{e^{\lambda(r)t_b} - 1} \eta(c, t_b) + \frac{(1 - \delta(r)) p_{sub} c}{1 - e^{-\lambda(r)t_b}}. \tag{B.21}$$

□

Lemma 3. *For the given structure of function g and subsidy policy δ,*
(i) there exists a electricity demand threshold for a customer from every income level such that solar adoption is an optimal choice if current electricity consumption is greater than the demand threshold.
(ii) the demand threshold is given by

$$\overline{X}(r,\delta) = \left(\frac{\gamma_1(r)}{\gamma_1(r)-1}\right)\left(\frac{f(r,\delta)-B(r)}{\frac{p_b}{\lambda(r)-\mu(r)}-A(r)}\right) \tag{B.22}$$

where,

$$f(r,\delta) = 1_{\{r\leq r^*(\delta)\}}\left((1-\delta_2)\frac{p_{sub}c}{1-e^{-\lambda(r)t_b}}\right) + 1_{\{r>r^*(\delta)\}}(1-\delta_1)(K+kc) \tag{B.23}$$

Proof. (i) We use Lemma 9 condition (ii) to prove this lemma. For a geometric Brownian motion the infinitesimal generator is given by $\mathcal{L}_{X^r} = \mu(r)x\frac{\partial}{\partial x} + \frac{\sigma^2}{2}x^2\frac{\partial^2}{\partial x^2}$. The net solar adoption cost, g, is given by Lemma 2 and $w(x) = p_b x$. We define $Z(r) := -\lambda(r)g(x;r,\delta) + \mathcal{L}_{X^r}g(x;r,\delta) + w(x)$ and consider following two cases:

Case 1: $r < r^*(\delta)$

Substituting the functional form of different functions,

$$Z(r) = x((\mu(r)-\lambda(r))A(r)+p_b) - \lambda(r)\left(\frac{(1-\delta_2)p_{sub}c}{1-e^{-\lambda(r)t_b}}\right) \tag{B.24}$$

Case 2: $r \geq r^*(\delta)$

Similarly for this case,

$$Z(r) = x((\mu(r)-\lambda(r))A(r)+p_b) - \lambda(r)(1-\delta_1)(K+kc) \tag{B.25}$$

If $(\mu(r')-\lambda(r'))A(r')+p_b > 0$, then there exists a x such that $Z(r') \geq 0$ since other components of $Z(r)$ are bounded.

Proposition 9. $(\mu(r')-\lambda(r'))A(r')+p_b > 0$ *for all* $r' \in (0,\infty)$.

Proof. Using the expression for $A(r')$ from Lemma 2 and rearranging the terms leads to the following expression:

$$(\mu(r') - \lambda(r'))A(r') + p_b = -p_b\left(\frac{\frac{e^{-\mu(r)t_b}-1}{-\mu(r)t_b}}{\frac{e^{(\lambda(r)-\mu(r))t_b}-1}{(\lambda(r)-\mu(r))t_b}} - 1\right) \tag{B.26}$$

$$> 0 \tag{B.27}$$

The last inequality follows because the function $\frac{e^x-1}{x}$ is an increasing function and $\lambda(r) - \mu(r) > -\mu(r)$. □

The terminal cost function is bounded and the function Z in x is monotonic, we can say that $x' = \frac{\lambda(r)\left(\frac{(1-\delta_2)p_{sub}c}{1-e^{-\lambda(r)t_b}}\right)}{(\mu(r)-\lambda(r))A(r)+p_b}$ if $r < r^*(\delta)$ and $x' = \frac{\lambda(r)(1-\delta_1)(K+kc)}{(\mu(r)-\lambda(r))A(r)+p_b}$ if $r \geq r^*(\delta)$ such that $Z(r) \geq 0$ for $x > x'$.

(ii) We have

$$\begin{aligned}
\overline{J}(x, \tau_r; \delta, r) :=& E_x\left[\int_0^{\tau_r} e^{-\lambda(r)s} p_b X_s^r \mathtt{d}s + e^{-\lambda(r)\tau_r} g(X_{\tau_r}^r; \delta, r)\right], \\
\overline{V}(x; \delta, r) :=& \inf_{\tau_r \geq 0} J(x, \tau_r; \delta, r), \\
\tau_r^* :=& \arg\min_{\tau_r \geq 0} J(x, \tau_r; \delta, r).
\end{aligned}$$

We define

$$\mathtt{d}Y_t^r = \begin{bmatrix} \mathtt{d}t \\ \mathtt{d}X_t^r \end{bmatrix} = \begin{bmatrix} 1 \\ \mu(r)X_t^r \end{bmatrix} \mathtt{d}t + \begin{bmatrix} 0 \\ \sigma(r)X_t^r \end{bmatrix} \mathtt{d}W_t^r \qquad Y_0^r = (s, x), \tag{B.28}$$

$$\mathtt{d}Z_t^r = \begin{bmatrix} \mathtt{d}Y_t^r \\ \mathtt{d}U_t^r \end{bmatrix} = \begin{bmatrix} 1 \\ \mu(r)X_t^r \\ e^{-\lambda(r)t} p_b X_t^r \end{bmatrix} \mathtt{d}t + \begin{bmatrix} 0 \\ \sigma(r)X_t^r \\ 0 \end{bmatrix} \mathtt{d}W_t^r \qquad Z_0^r = z = (s, x, u). \tag{B.29}$$

If we let $G(z) = u + e^{-\lambda(r)s} g(x; \delta, r)$, then we can reduce the original problem to an equivalent problem with only terminal reward. Using section 10.3 Oksendal, the characteristic function

of Z_t is given by $\mathcal{L}_Z G = \mathcal{L}_X G + \frac{\partial G}{\partial s} + e^{-\lambda(r)s} p_b x$ where $\mathcal{L}_X$ is the characteristic function of X_t and is given by $\mu(r)x\frac{\partial}{\partial x} + \frac{\sigma^2}{2}x^2\frac{\partial^2}{\partial x^2}$. We denote the continuation region of the above optimal stopping problem by D. We notice that the terminal cost of the optimal stopping problem is bounded above by $K + kc$ and the running cost is an increasing function of X_t^r assuming $p_b > 0$. We can observe that there exists a $\overline{X}(r,\delta)$ such that $(0, \overline{X}(r,\delta)) \subset D$. Next, we derive closed form expression for $\overline{X}(r,\delta)$ and optimal value function using pde derived by applying separation of variables on $\mathcal{L}_Z G = \mathcal{L}_X G + \frac{\partial G}{\partial s} + e^{-\lambda(r)s} p_b x$.

We consider a candidate value function $V(x;r,\delta)$ then V satisfies the following equations,

$$\mathcal{L}V(x;r,\delta) - \lambda(r)V(x;r,\delta) + p_b x = 0 \qquad x \in D, \tag{B.30}$$

$$V(x;r,\delta) = g(x;r,\delta) \qquad x \notin D, \tag{B.31}$$

First we solve for the case $x \in D$. $V(x;r,\delta)$ satisfies the following non-homogeneous Cauchy-Euler differential equation.

$$\mu(r)x\frac{\partial V(x;r,\delta)}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 V(x;r,\delta)}{\partial x^2} - \lambda(r)V(x;r,\delta) + p_b x = 0. \tag{B.32}$$

The solution to the above equation is of the form

$$V(x;r,\delta) = M_1 x^{\gamma_1} + M_2 x^{\gamma_2} + \frac{p_b x}{\lambda(r) - \mu(r)} \qquad V(0;r,\delta) = 0. \tag{B.33}$$

where,

$$\gamma_{1,2}(r) = \frac{-(\mu(r) - \frac{\sigma^2}{2}) \pm \sqrt{(\mu(r) - \frac{\sigma^2}{2})^2 + 2\sigma^2\lambda(r)}}{\sigma^2}. \tag{B.34}$$

We notice that $\gamma_1(r) > 1$ and $\gamma_2(r) < 0$. Since, $\gamma_2 < 0$ and $V(0;r,\delta) = 0$, $M_2 = 0$. We calculate M_1 and $\overline{X}(r,\delta)$ using value matching and smooth pasting technique at the boundary of D.

$$M_1(\overline{X}(r,\delta))^{\gamma_1(r)} + \frac{p_b\overline{X}(r,\delta)}{\lambda(r) - \mu(r)} = g(\overline{X}(r,\delta);\delta,r). \tag{B.35}$$

$$M_1\gamma_1(r)(\overline{X}(r,\delta))^{\gamma_1(r)-1} + \frac{p_b}{\lambda(r) - \mu(r)} = g'(\overline{X}(r,\delta);\delta,r). \tag{B.36}$$

Using the above two equations we know that $\overline{X}(r,\delta)$ satisfies the following equation

$$\frac{(\gamma_1(r)-1)p_b\overline{X}(r,\delta)}{\lambda(r)-\mu(r)} = \gamma_1(r)g(\overline{X}(r,\delta);r,\delta) - \overline{X}(r,\delta)g'(\overline{X}(r,\delta);r,\delta). \tag{B.37}$$

Substituting $g(\overline{X}(r,\delta);r,\delta)$ using Lemma 2 in (B.37) completes the proof.

We can also derive the closed form expression for $V(x;r,\delta)$ by simulatenously solving equation (B.35) and (B.36).

$$V(x;r,\delta) = \frac{(B(r)-f(r,\delta))}{\gamma_1-1}\left(\frac{x}{\overline{X}(r,\delta)}\right)^{\gamma_1(r)} + \frac{p_b x}{\lambda(r)-\mu(r)}$$

In this procedure we have made multiple assumptions. Next, we prove some properties necessary to apply verification theorem.

Proposition 10. $E[\tau_r^*] < \infty$

Proof. Given that X_t^r follows geometric Brownian motion, we know that

$$\log(X^r_{t\wedge\tau_r^*}) = \log(X^r_{t_0}) + \left(\mu(r) - \frac{\sigma^2}{2}\right)(t\wedge\tau_r^* - t_0) + \sigma(W^r_{t\wedge\tau_r^*} - W^r_{t_0})$$

Taking expectation on both sides

$$E[\log(X^r_{t\wedge\tau_r^*})] = \log(X^r_{t_0}) + \left(\mu(r) - \frac{\sigma^2}{2}\right)E[t\wedge\tau_r^* - t_0]$$

Using monotone convergence theorem

$$E[\tau_r^*] = \frac{\lim_{t\to\infty} E[\log(X^r_{t\wedge\tau_r^*})] - \log x}{\mu(r)-\frac{\sigma^2}{2}} + E[t_0] \leq \frac{\log(\overline{X}(r,\delta)) - \log x}{\mu(r)-\frac{\sigma^2}{2}} + E[t_0] < \infty$$

□

Consider a stopping time $\tau_r' = \inf\{t : X_t^r \notin (0,\overline{X}(r,\delta))\}$. Assuming $x < \overline{X}(r,\delta)$, clearly, $\tau_r^* \leq \tau_r'$. Using procedure similar to the proof of previous lemma, we can show that $E[\tau_r'] < \infty$. $V(x;r,\delta)$ is a continuously differentiable function by construction. Using Ito's lemma

$$\begin{aligned}
E[V(X^r_{\tau_r'};r,\delta)|X^r_{\tau_r^*}] =& V(X^r_{\tau_r^*};r,\delta) \\
&+ E\left[\int_{\tau_r^*}^{\tau_r'}\left(\mu(r)x\frac{\partial V(x;r,\delta)}{\partial x} + \frac{\sigma^2}{2}x^2\frac{\partial^2 V(x;r,\delta)}{\partial x^2} - \lambda(r)V(x;r,\delta) + p_b x\right)\mathtt{d}t|X^r_{\tau_r^*}\right] \\
&+ E\left[\int_{\tau_r^*}^{\tau_r'}\sigma\frac{\partial V(x;r,\delta)}{\partial x}\mathtt{d}W_t|X^r_{\tau_r^*}\right]
\end{aligned} \tag{B.38}$$

Proposition 11. $E\left[\int_{\tau_r^*}^{\tau_r'} \sigma(r)\frac{\partial V(x;r,\delta)}{\partial x}\,\mathrm{d}W_t | X^r_{\tau_r^*}\right] = 0$

Proof. We use Corollary 4.8 from Harrison(2013) to prove the above proposition. The corollary states, for a stopping time T, if $Z = \int_0^T X_t dW_t$ such that X_t is bounded in [0,T] and $E[T] < \infty$, then $E[Z] = 0$. From the previous lemma, we have that both $E[\tau_r^*]$ and $E[\tau_r']$ are finite.

$$\frac{\partial V(x;r,\delta)}{\partial x} = M_1\gamma_1 x^{\gamma_1 - 1} + M_2\gamma_2 x^{\gamma_2 - 1} + \frac{p_b}{\lambda(r) - \mu(r)} \tag{B.39}$$

$\frac{\partial V(x;r,\delta)}{\partial x}$ is a continuous function in the compact interval $[\overline{X}(r,\delta), \overline{X}(r,\delta) + \Delta]$, where $\Delta > 0$. This implies that $\frac{\partial V(x;r,\delta)}{\partial x}$ is bounded. Since, $\sigma(r)$ is a known constant and $\frac{\partial V(x;r,\delta)}{\partial x}$ is bounded, we get the desired result using corollary 4.8. □

Proposition 12. $E\left[\int_{\tau_r^*}^{\tau_r'} \left(\mu(r)x\frac{\partial V(x;r,\delta)}{\partial x} + \frac{\sigma^2}{2}x^2\frac{\partial^2 V(x;r,\delta)}{\partial x^2} - \lambda(r)V(x;r,\delta) + p_b x\right)\mathrm{d}t | X^r_{\tau_r^*}\right] \geq 0$ *for* $x \geq \overline{X}(r,\delta)$.

Proof. For $x \geq \overline{X}(r,\delta)$, $V(x;r,\delta) = g(x;r,\delta) = A(r)x - B(r) + f(r,\delta)$. Substituting $\frac{\partial g(x;r,\delta)}{\partial x} = A(r)$ and $\frac{\partial^2 g(x;r,\delta)}{\partial x^2} = 0$ in $\mu(r)x\frac{\partial V(x;r,\delta)}{\partial x} + \frac{\sigma^2}{2}x^2\frac{\partial^2 V(x;r,\delta)}{\partial x^2} - \lambda(r)V(x;r,\delta) + p_b x$ yields

$$(f(r,\delta) - B(r))\left(\left(\frac{x}{\overline{X}(r,\delta)}\right)(\lambda(r) - \mu(r))\frac{\gamma_1(r)}{\gamma_1(r) - 1} - \lambda(r)\right), \tag{B.40}$$

$$\geq (f(r,\delta) - B(r))\left(\frac{\lambda(r) - \mu(r)\gamma_1(r)}{\gamma_1(r) - 1}\right) \tag{B.41}$$

$$\geq 0 \tag{B.42}$$

The second inequality holds as we are considering the case $x \geq \overline{X}(r,\delta)$. We have $f(r,\delta) \geq B(r)$ because demand threshold cannot assume negative value. We can infer from the expression for γ_1 that $\gamma_1 \geq 1$ and with additional simple algebraic manipulations that $\lambda(r) - \mu(r)\gamma_1(r) \geq 0$. □

Using results from Proposition 11 and Proposition 12 in (B.38), we can infer that for $x \geq \overline{X}(r,\delta)$, the following inequality holds

$$E[V(X^r_{\tau_r'};r,\delta)|X^r_{\tau_r^*}] \geq V(X^r_{\tau_r^*};r,\delta) \tag{B.43}$$

The above inequality implies that it is better to adopt solar immediately when the level of demand hits $\overline{X}(r,\delta)$ than to wait. □

Lemma 4. *The probability density function of $\tau_r^*(\delta)$ is given by*

$$f_{\tau_r^*(\delta)}(t;\delta,r) = \frac{a(r,\delta)}{\sqrt{2\pi}t^{3/2}} e^{\frac{-(a(r,\delta)-b(r)t)^2}{2t}} \tag{B.44}$$

where,

$$a(r,\delta) := \frac{1}{\sigma}\log\left(\frac{\overline{X}(r,\delta)}{x}\right) \tag{B.45}$$

$$b(r) := \frac{1}{\sigma}\left(\mu(r) - \frac{\sigma^2}{2}\right) \tag{B.46}$$

$$\tag{B.47}$$

Proof. We first find an expression for $P(\tau_r^* \leq \mathcal{T}|X_0^r = x)$. We define $\tilde{X}_{\mathcal{T}}^r := \max\{X_s^r : 0 \leq s \leq \mathcal{T}\}$. The stochastic process, $\tilde{X}_t^r$, is the running maximum of the demand process X_t^r till time t. We notice that $P(\tau_r^* \leq \mathcal{T}|X_0^r = x) = P(\tilde{X}_{\mathcal{T}}^r \geq \overline{X}(r)|X_0^r = x)$. We have assumed that the dynamics of X_t^r follows geometric Brownian motion. The distribution of the running maximum process is given by

$$P(\tilde{X}_{\mathcal{T}}^r \geq \overline{X}(r)|X_0^r = x) = 1 - \Phi\left(\frac{a(r,\delta) - b(r)\mathcal{T}}{\sqrt{\mathcal{T}}}\right) + e^{2a(r,\delta)b(r)}\Phi\left(\frac{-a(r,\delta) - b(r)\mathcal{T}}{\sqrt{\mathcal{T}}}\right). \tag{B.48}$$

where, $\Phi(\cdot)$ is the CDF of standard normal distribution. Using the following equation gives us the desired result.

$$f_{\tau_r^*(\delta)}(t;x) := \frac{d}{dt}P(\tau_r^*(\delta) \leq t|X_0^r = x). \tag{B.49}$$

□

Lemma 5. *The feasible region of the central planner's problem is a convex set.*

Proof. Notice that the last constraint reduces to $(1-\delta)(K + kc - p_{sub}c) \geq \epsilon$ and is true for all subsidies. We need to study the first 3 constraints. We show that the right hand side of those constraint is a quasi-concave function. The super level sets of quasi-concave

functions are convex.

$$P(\tau^*(\delta) \leq \mathcal{T}|X_0 = x) = \int_0^\infty P(\tau_r^*(\delta) \leq \mathcal{T}|X_0^r = x)p(r)\mathtt{d}r \tag{B.50}$$

$$= \int_0^\infty P(\tilde{X}_{\mathcal{T}}^r \geq \overline{X}(r,\delta)|X_0^r = x)p(r)\mathtt{d}r \tag{B.51}$$

For each r, $P(\tilde{X}_{\mathcal{T}}^r \geq \overline{X}(r,\delta)|X_0^r = x)$ is a non-decreasing function of δ since $\overline{X}(r,\delta)$ is non-increasing in δ. Monotonicity of integrals imply that $P(\tau^*(\delta) \leq \mathcal{T}|X_0 = x)$ is non-decreasing function of δ. Any monotonic function is quasi-concave and the super level set of a quasi-concave functions is convex. We can use similar argument to prove that the set of δ satisfying second and third constraint are also convex. The feasible region for the central planner's problem is the intersection of these finite number of convex sets, hence convex. □

Lemma 6. *$z(\delta)$ is a convex and increasing function of δ.*

Proof. Let

$$\nu(r) := 1_{\{r \leq r^*\}}\left(\frac{p_{sub}c}{1 - e^{-\lambda(r)t_b}}\right) + 1_{\{r > r^*\}}(K + kc) \tag{B.52}$$

Since the integrand in $z(\delta)$ is non-negative, using Tonelli theorem, we can exchange the expectation and the integral.

$$z(\delta) := \int_0^\infty E_x\left[e^{-\lambda(r)\tau_r^*(\delta)}\right]\delta\nu(r)p(r)\mathtt{d}r \tag{B.53}$$

The expression $E_x\left[e^{-\lambda(r)\tau_r^*(\delta)}\right]$ represents the Laplace transform of the adoption time for a given r and δ. Using the expression of the Laplace transform of the stopping time for a geometric Brownian motion, we have

$$E_x\left[e^{-\lambda(r)\tau_r^*(\delta)}\right] = \left(\frac{x}{\overline{X}(r,\delta)}\right)^{l} \tag{B.54}$$

where,

$$l(r) = \frac{\sqrt{\mu(r)^2 + 2\lambda(r)\sigma^2} - \mu(r)}{\sigma^2} \tag{B.55}$$

Substituting (eq: laptau) in (eq: zdeltaconcex)

$$z(\delta) := \int_0^\infty x^{l(r)}\frac{\delta}{(\overline{X}(r,\delta))^l}\nu(r)p(r)\mathtt{d}r \tag{B.56}$$

The integrand in the above integral is a bounded and a continuous function of δ and r. Using dominated convergence theorem we can show that $z(\delta)$ is a continuous function of δ. Next we substitute the expression for $\overline{X}(r,\delta)$ from lemma 3, which leads us to the following two cases.

Case 1: $\overline{X}(r,\delta) > x$

$$z(\delta) = \int_0^\infty x^{l(r)} \frac{\delta}{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)}} \left(\left(\frac{\gamma_1(r)-1}{\gamma_1(r)} \right) \left(\frac{p_b}{\lambda(r)-\mu(r)} - A(r) \right) \right)^{l(r)} \nu(r)p(r)\mathtt{d}r. \tag{B.57}$$

$$\frac{\partial}{\partial\delta} z(\delta)$$
$$= \int_0^\infty x^{l(r)} \frac{\partial}{\partial\delta} \left(\frac{\delta}{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)}} \right) \left(\left(\frac{\gamma_1(r)-1}{\gamma_1(r)} \right) \left(\frac{p_b}{\lambda(r)-\mu(r)} - A(r) \right) \right)^{l(r)} \nu(r)p(r)\mathtt{d}r. \tag{B.58}$$

$$\frac{\partial^2}{\partial\delta^2} z(\delta)$$
$$= \int_0^\infty x^{l(r)} \frac{\partial^2}{\partial\delta^2} \left(\frac{\delta}{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)}} \right) \left(\left(\frac{\gamma_1(r)-1}{\gamma_1(r)} \right) \left(\frac{p_b}{\lambda(r)-\mu(r)} - A(r) \right) \right)^{l(r)} \nu(r)p(r)\mathtt{d}r. \tag{B.59}$$

We observe that for every r

$$\frac{\partial}{\partial\delta} \frac{\delta}{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)}}$$
$$= \frac{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)} + \delta l(r)\nu(r)((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)-1}}{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{2l(r)}} \tag{B.60}$$

$$\geq 0. \tag{B.61}$$

Similarly

$$\frac{\partial^2}{\partial\delta^2} \frac{\delta}{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)}} = \frac{l(r)\nu(r)}{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)+1}}$$
$$+ l(r)\nu(r) \left[\frac{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)+1} + \delta(l(r)+1)((1-\delta)\nu(r) - B(r)\gamma_1(r))^{l(r)}}{((1-\delta)\nu(r) - B(r)\gamma_1(r))^{2l(r)+2}} \right] \tag{B.62}$$

$$\geq 0. \tag{B.63}$$

This implies that $\frac{\partial}{\partial\delta}z(\delta) \geq 0$ and $\frac{\partial^2}{\partial\delta^2}z(\delta) \geq 0$.

Case 2: $\overline{X}(r,\delta) = x$

Substituting $\overline{X}(r,\delta) = x$ in the expression for $z(\delta)$, we can notice that $z(\delta)$ is a linear function of δ, $\frac{\partial}{\partial\delta}z(\delta) \geq 0$ and $\frac{\partial^2}{\partial\delta^2}z(\delta) = 0$. □

Lemma 7. *$r^*(\delta)$ is non-increasing function of δ_1 and a non-decreasing function of δ_2.*

Proof. We know that $\lambda^*_\delta = -\frac{1}{t_b}\ln\left(1 - \frac{(1-\delta_2)p_{sub}c}{(1-\delta_1)(K+kc)}\right)$. It is straight forward to see that $\frac{\partial\lambda^*_\delta}{\partial\delta_1} \geq 0$ and $\frac{\partial\lambda^*_\delta}{\partial\delta_2} \leq 0$. The result follows since $\lambda(r)$ is a decreasing function of r. □

Lemma 8. *The probability density function of τ^* is given by*

$$f_{\tau^*}(t;x,\delta,G,\alpha) = \frac{1}{\sqrt{2\pi}}\frac{1}{(G\alpha)t^{3/2}}\int_0^\infty a(r,\delta)\frac{(r/\alpha)^{\frac{1-G}{G}}}{(1+(r/\alpha)^{1/G})^2}e^{\frac{-(a(r,\delta)-b(r)t)^2}{2t}}\,\mathtt{d}r. \tag{B.64}$$

Proof. The probability density function of τ^* is defined as

$$f_{\tau^*}(t;x,\delta,G,\alpha) := \int_0^\infty f_{\tau^*_r}(t;x,\delta)p(r)\mathtt{d}r.$$

We know from Lemma 4 the closed form expression for $f_{\tau^*_r}(t;x,\delta)$. Substituing the expression for $p(r) = \frac{1}{G\alpha}\frac{(r/\alpha)^{\frac{1-G}{G}}}{(1+(r/\alpha)^{1/G})^2}$ gives us the desired result. □

Lemma 9. *Assuming $g \in \mathcal{C}^2(R)$, $w \in \mathcal{C}^2(R)$ and a subsidy policy δ, if a customer belongs to income level r such that*

(i) $\big(-\lambda(r)g(x;r,\delta) + \mathcal{L}_{X^r}g(x;r,\delta) + w(x;p_b)\big) \geq 0$ *for all x then $\tau^*_r = 0$ a.s.*

(ii) $\big(-\lambda(r)g(x;r,\delta) + \mathcal{L}_{X^r}g(x;r,\delta) + w(x;p_b)\big) < 0$ *for all x then $\tau^*_r = \infty$ a.s.*

Here $\mathcal{L}_{X^r}$ is the infinitesimal generator for the demand process $\{X^r_t\}_{t\geq 0}$.

Proof. We introduce a new Ito's diffusion Z^r_t to define a time homogeneous stopping problem as follows

$$dZ^r_t = \begin{bmatrix} 1 \\ \mu(r,X^r_t) \\ e^{-\lambda(r)t}w(X^r_t) \end{bmatrix}\mathtt{d}t + \begin{bmatrix} 0 \\ \sigma(r,X^r_t) \\ 0 \end{bmatrix}\mathtt{d}W^r_t \qquad\qquad Z_0 = (0,x,0). \tag{B.65}$$

We then define the a new cost function as

$$\tilde{g}(t, x, \overline{w}; r, \delta) = \overline{w} + e^{-\lambda(r)t} g(x; r, \delta), \tag{B.66}$$

and our equivalent optimal stopping problem is

$$V(x; r, \delta) := \inf_{\tau_r \geq 0} E_{Z_0}\big[\tilde{g}(Z_{\tau_r}; r, \delta)\big], \tag{B.67}$$

where $Z_{\tau_r} = (\tau_r, x, \overline{w}_{\tau_r})$.

Next, we write the infinitesimal generator $\mathcal{L}_{Z^r}$ for the process $\{Z_t^r\}_{t\geq 0}$. Let $\phi(z) \in \mathcal{C}^2$, then

$$\mathcal{L}_{Z^r}\phi = \frac{\partial \phi}{\partial t} + \mu(r, x)\frac{\partial \phi}{\partial x} + \frac{\sigma^2(r, x)}{2}\frac{\partial^2 \phi}{\partial x^2} + e^{-\lambda(r)t} w(x)\frac{\partial \phi}{\partial \overline{w}}. \tag{B.68}$$

Using equations (B.67) and (B.68), we get

$$\mathcal{L}_{Z^r}\tilde{g} = \big(- \lambda(r) g(x; r, \delta) + \mu(r, x) g'(x; r, \delta) + \frac{\sigma^2(r, x)}{2} g''(x; r, \delta) + w(x)\big) e^{-\lambda(r)t}. \tag{B.69}$$

$$= \big(- \lambda(r) g(x; r, \delta) + \mathcal{L}_{X^r} g(x; r, \delta) + w(x)\big) e^{-\lambda(r)t}, \tag{B.70}$$

For a given r and δ we consider two cases:

Case 1: If $\big(- \lambda(r) g(x; r, \delta) + \mathcal{L}_{X^r} g(x; r, \delta) + w(x)\big) \geq 0$ for all x, then $\mathcal{L}_{Z^r}\tilde{g} \geq 0$. Since g and w belong to the class of twice differentiable functions, $\tilde{g} \in \mathcal{C}^2(R)$. Using Dynkin formula(Oksendal, Section 10.1) we can say that $\tilde{g}$ is a sub-harmonic function and $\tau_r^* = 0$ i.e customers from income level r adopt solar technology immediately.

Case 2: If $\big(- \lambda(r) g(x; r, \delta) + \mathcal{L}_{X^r} g(x; r, \delta) + w(x)\big) < 0$ for all x, then $\mathcal{L}_{Z^r}\tilde{g} < 0$. This implies that the continuation region is R and $\tau_r^* = \infty$ i.e customers from income level r do not adopt solar technology.

□