ABSTRACT

Title of Dissertation: ON TERRORIST ATTACKS AND ESTIMATION METHODS

Pablo Antonio Salinas Macario
Doctor of Philosophy, 2023

Dissertation Directed by: Professor Ingmar R. Prucha
Department of Economics

In my thesis I propose a theoretical model of terrorist attacks and an estimation strategy which I compare to existing methods in the literature. The modeling approach was designed with terrorism in mind, but can be applied to other discrete dynamic decision processes with a latent component and a random payoff variable that is measured when the agent exits a state of waiting. Chapter 1 briefly describes the structure of the thesis.

Chapter 2 provides a literature review of empirical studies of terrorist attacks. The primary focus is the *series hazard* model that estimates the effect of policy interventions on the risk of terrorist attacks. Recent contributions include LaFree et al. (2009), Dugan (2011), Carson (2014) Argomaniz and Vidal-Diez (2015), and Carson (2017). A major limitation of the series hazard approach is that it is unable to evaluate the impact of a policy intervention on the *outcomes* of attacks (e.g., the number of fatalities) even if these are measured during each event.

Chapter 3 introduces the *sequence hazard* model of a terrorist group deciding when to attack. The model links the outcome of terrorist attacks to the choice of when to attack by taking the

amount of time elapsed since the last attack as an input into the planning of the next attack. The agent trades off the desire to improve their attack against the risk that their plans are sabotaged before they are able to carry them out. The sequence hazard model is dynamic because agents take into account the potential size of future attacks when deciding whether or not to attack today. As a consequence, the hazard implied by the sequence approach is non-proportional in time. This distinguishes the sequence hazard model from the proportional hazard assumed by the series (Cox) approach.

The sequence model implies a data generating process for attack outcomes that takes into account the probability the agent attacks. Chapter 3 derives the implied mathematical expectation and variance of attack outcomes which allows researchers to extend the notion of deterrence to allow for the possibility that counterterrorist policies that reduce the frequency of attacks, but increase the expected severity of attacks that do take place. Two types of attack outcomes are considered, a mixed Poisson-beta model for the number of casualties and a mixed Bernoulli-beta model for attack success or failure.

Chapter 4 presents a Monte Carlo study demonstrating the validity of estimating the sequence hazard model by maximum likelihood. In contrast, when the underlying data are generated according to the simple behavioral model presented in Chapter 3, the series hazard fails to estimate the true effect of a policy intervention on the risk of attacks. Moreover, the standard tests fail to reject the null hypothesis that the data are generated according to a proportional model. The simulation implies that if planning time and uncertainty over attack outcomes are important elements in terrorist decision making, then methods of policy evaluation based on the assumption of proportionality may not be appropriate. In contrast, by modeling both the timing of attacks as well as their size, the sequence hazard offers a straightforward way of incorporating terrorist attack outcomes into the analysis of counterterrorism policy.

ON TERRORIST ATTACKS AND ESTIMATION METHODS

by

Pablo Antonio Salinas Macario

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Ingmar R. Prucha, Chair
Professor John Chao
Professor Gerrit-Jan Knaap
Professor Guido M. Kuersteiner
Professor John Wallis

Acknowledgments

I would like to thank my thesis committee members Professors Ingmar R. Prucha, John Chao, Gerrit-Jan Knaap, Guido M. Kuersteiner and John Wallis for their guidance.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ϕ_1	Humbert Type 1 Function
$\mathcal{H}_{a,b}$	Regularized Humbert Type 1 Function
$\mathcal{M}_{a,b}$	Kummer's Confluent Hypergeometric Function
$\mathrm{P}\left(\cdot \mid \mu\right)$	PMF of a Poisson Distribution with Parameter μ
$F\left(\cdot; \alpha, \beta\right)$	CDF of a Beta Distribution with Parameters α and β
$F\left(\cdot; \lambda\right)$	CDF of an Exponential Distribution with Parameter λ
$f\left(\cdot; \alpha, \beta\right)$	PDF of a Beta Distribution with Parameters α and β
$f\left(\cdot; \lambda\right)$	PDF of an Exponential Distribution with Parameter λ
$\overline{F}\left(\cdot; \alpha, \beta\right)$	Survival Function of a Beta Distribution with Parameters α and β
$\overline{F}\left(\cdot; \lambda\right)$	Survival Function of an Exponential Distribution with Parameter λ
$\mathrm{Beta}\left(\alpha, \beta\right)$	Beta function with parameters α and β
$\Gamma\left(\cdot\right)$	Gamma Function
GTD	Global Terrorism Database
CCP	Conditional Choice Probability
JCOP	Joint Choice and Outcome Probability
ETA	Euskadi Ta Askatasuna

Chapter 1: Introduction

This dissertation proposes a new method to study sequences of terrorist attacks by a single terrorist organization. It assumes an agent's decision of when to attack can be represented by an optimal choice problem linking the expected size of an attack to the time spent planning in a waiting state. Because larger attacks require more time to plan, by modeling both the timing of attacks as well as their size, I create a model of terrorist decision making that enables me to more accurately describe the constraints to terrorist activity, and thus potentially their behavior. The behavioral model has an associated econometric model that can used with data on terrorist attacks. A Monte Carlo experiment compares the new sequence hazard model to its reduced-form counterpart, the series hazard model.

Chapter 2 introduces the concept of terrorism and studies on the social and economic impact of the risk of terrorist attacks. The chapter focuses on empirical studies of terrorism that employ a rational agent framework. Methods covered include interrupted time-series approaches as well as a fully parametric model. The rest of the chapter is spent reviewing the series hazard model. As opposed to the model presented in Chapter 3, the series hazard model cannot account for the outcome of attacks even if these are recorded at the time of the event. Chapter 2 ends with a review of studies of dynamic terrorist behavior that focus on the size-frequency trade-off and emphasize the role that planning plays in the production of terrorist attacks. These studies provide the foundation for the sequence haqzard model, an alternative to the series hazard model used by much of the literature.

Chapter 3 presents the sequence hazard model of terrorist attacks and Chapter 4 demonstrates its validity via a Monte Carlo experiment. The sequence hazard is a discrete duration model of the risk of a terrorist attack. The model is defined at the level of the agent and takes the view of a campaign in order to study attacks as part of a sequence. This type of model has been missing in the literature (see Sandler and Enders (2004) and Enders and Sandler (2011)). The model assumes the sequence of observed attacks and waiting times are the outcome of a stochastic decision problem. The agent in the model decides when to attack by balancing two competing imperatives. Terrorist want to invest as much time as possible into the planning of their attack in order to make it as large as possible or increase the probability of a success. On the other hand, they want to attack before their plans are interfered with. In other words, terrorists trade off need to improve the attack against the danger of the attack being disrupted before they are able to carry it out. The model assumes that a researcher does not observe all the state variables relevant to terrorist decision making. These unobservable states are given specific parametric forms that lead to an estimable empirical model. The framework has similarities to Rust (1987) with important distinctions that arise from updating the bus model to apply it to the study of terrorism.

The decision problem presented in Chapter 3 is built around the agent's expectation of the size (or probability of success) of a terrorist attack. The basic model does not use the outcome data and simply models the choice of when to attack. This model can only make predictions about the expected size of future attacks. To make inference about attack outcomes I further impose the assumption that the outcomes of terrorist attacks are believed by the agent to be distributed as a Poisson-beta mixture which results in a likelihood function that takes into account the outcomes of terrorist attacks. The distributional assumptions imply a mean and variance of attack outcome that depend explicitly on the probability of an attack. As a result, the data generating process for the time-series of outcomes observed by the researcher is over-dispersed (the variance is greater than the mean). This is a desirable feature for a model of terrorist attack outcomes since it is a well-known feature of the data. Moreover, the data generating process for outcome takes into account the selection on unobservable states, and in fact, if there is no selection (i.e., the probability of attacking is equal to 1) then the process for attack outcomes is a pure Poisson-beta mixture.

The structural model for the choice of when to attack implies a structural hazard that is different than the type assumed by the series hazard model. The latter is proportional in time meaning it can be written as a multiplicative function of time elapsed since the last event, whereas the structural hazard is shown to be a non-linear function of the waiting time. From the perspective of the model presented in Chapter 3, the series hazard model is misspecified because of the proportional hazard assumption. The effects of misspecification are confirmed via the Monte Carlo experiment presented in Chapter 4.

Chapter 2: Models of Terrorist Risk Literature Review

2.1 Introduction

This chapter provides an overview of the series hazard model and its application to the study of counterterrorism policy. The chapter first defines the concept of terrorism and then discusses research on the impact of terrorist attacks. Next, it outlines the theory of rational terrorism and reviews time-series analyses of terrorist behavior. The chapter then turns to the series hazard model and its application to the study of counterterrorist policies against Euskadi-Ta-Askatasuna (ETA) in Spain.

2.2 Impact of Terrorist Attacks

"One such act may, in a few days, make more propaganda than thousands of pamphlets."

Pyotr Alexeyevich Kropotkin, The Spirit of Revolt, 1880 (Kropotkin (2016))

Terrorists are non-state actors that use attacks or the threat of attacks to draw attention to their religious, ideological, or political goals. Terrorism is 'propaganda by deed', violence intended to intimidate a wide audience with the goal of influencing policy. Terrorism, like war, is a continuation of politics by other means (Clausewitz (1832)). However, terrorism is distinct from full-blown war because it does not involve armies holding territory or engaging in battles. Like guerrillas, terrorist groups avoid direct combat with more powerful adversaries, preferring to blend or hide within a territory rather than control it outright. But unlike guerrillas, terrorists deliberately target the unarmed or defenseless. The use of indiscriminate, seemingly random acts of violence against non-military targets is premeditated and intentional. The ultimate goal is to induce compliance by a fearful population and/or provoke a response from authorities. Thus, “terrorism differs from full-blown war in means and scale, but not in ends” Shughart (2011).

As noted by Enders and Sandler (1995), the presence of a political objective is crucial to the definition of terrorism and is what distinguishes it from purely criminal endeavors like extortion or homicide. This dissertation follows the US State Department’s definition of terrorism, updated after the terrorist attacks of September 11, 2001 (Executive Order 13224) “For the purpose of the Order, “terrorism” is defined to be an activity that (1) involves a violent act or an act dangerous to human life, property, or infrastructure; and (2) appears to be intended to intimidate or coerce a civilian population; to influence the policy of a government by intimidation or coercion; or to affect the conduct of a government by mass destruction, assassination, kidnapping, or hostage-taking.“ [1] This definition accounts for a wide variety of terrorist groups and ideologies.

Terrorism is first mentioned in the historical record by Josephus, a historian and participant of the First Jewish-Roman War from 66 to 73 CE. Josephus wrote about a group numbering a few hundred dagger-wielding assassins known as the Sicarii. The Sicarii’s ideas and tactics placed them on the fringes of a much larger faction of Judean rebels that opposed Roman imperial rule. In “The First Jewish War”, Josephus describes the Sicarii’s campaign of sabotage and targeted killings against Roman occupiers. The Sicarii poisoned the Roman legion’s water supply and murdered Roman officials alongside suspected Jewish collaborators. The attacks were audacious and spectacular, often taking place in broad daylight. When the Romans were ready to counter the threat they would find that the Sicarii had long dispersed among the local population.

As illustrated by the Sicarii’s campaign, terrorism is pervasive because it has proven to be

[1] https://www.state.gov/executive-order-13224/

an effective tactic for a group fighting a more resourceful adversary. The Sicarii's campaign of terror had its intended effect. As anxiety grew among the Romans and their local allies, Roman forces began to retaliate indiscriminately against the local Jewish population. These events served as the backdrop to the first of three armed revolts against Roman rule in the region. The revolt of a distant province against Rome was of course not uncommon, but historians argue the Sicarii's campaign of terror acted as a catalyst in bringing about their preferred political outcome of all out war against the Romans (Horsley (1979), D'Alessio and Stolzenberg (1990)).[2]

The logic of terror pursued by the Sicarii in biblical times has been imitated by armed political outfits across diverse ideological and geographical settings. The Global Terrorism Database (GTD) maintained by START at the University of Maryland tracks over 100,000 terrorist incidents taking place in over 140 countries since 1973 alone.[3] In spite of the number of events and the potential consequences of any given attack, economists have found the macroeconomic effects of terrorism to be usually small, of limited duration, and local in scope (Shughart (2011)). But it is also the case that terrorist attacks have changed the course of history, like the assassination of Archduke Ferdinand of Austria on May 1914 that precipitated World War I. The attacks of 9/11 are another important example. Though the attacks are estimated to have cost US$90 billion, less than 1% of US GDP (Enders and Sandler (2011), Sandler (2014)), they played an important role in subsequent wars that cost upwards of US$3 trillion (Stiglitz and Bilmes (2008)) and led to profound changes to laws and civil liberties both in the US and elsewhere. So terrorists not only impact societies through the immediate aftermath of an attack, but also in how targeted societies respond to the threat of terrorism (McCauley (2006)).

[2] The revolt ended with most of the Sicarii committing mass suicide after Masada fell to the Romans in 73 CE. Remnants of the Sicarii would later go on to form one of the main political parties in Jerusalem.

[3] See https://www.start.umd.edu.

There are three broad channels through which terrorist attacks impose a cost on society: human costs, economic costs, and political costs. Arce (2019) summarizes a breakdown of the deaths and injuries associated with various terrorist tactics such as suicide bombings and mass shootings. His analysis shows that the morbidity from terrorism is relatively low compared to the global burden of disabilities, injuries and disorders. However, when taking a more micro perspective there is evidence that the fear of unpredictable violence induced by terrorist attacks can impose long-lasting human costs. Camacho (2008) uses a data set of approximately 4 million births in Colombia from 1998 to 2003 to study the effect of landmine explosions on birth weight. Low weight at birth has been correlated with poor health later in life, higher infant child mortality, and lower cognitive abilities. Landmines are still used to this day by terrorist groups in Colombia to limit mobility and instill fear in civilian populations. Camacho's study finds that childhood birth weights were significantly negatively impacted by random landmine explosions during a woman's first trimester of pregnancy. The study offers an example of the long term human costs due to the risk terrorism that are hard to quantify and unlikely to be reflected in attack outcome statistics.

The fear of unpredictable violence also generates uncertainty which gives terrorist groups potential leverage over economic outcomes. There are numerous studies that quantify the impact of terrorism on tourism, for example. Tourist areas are an attractive target for terrorists because they present would-be attackers with numerous soft targets that can be dissuaded from visiting a specific location, resulting in economic loss. Moreover, tourists are often citizens of wealthy or powerful nations and attacking them may increase international pressure on a local government. Enders and Sandler (1991) use a VAR to show that between 1970 and 1988 transnational terrorism greatly reduced Spanish tourism. Pizam and Smith (2000) classify major terrorism events around the world from 1985 to 1998 and study their impact on the demand for tourism. They find that a large portion of terrorist acts caused a significant decline in tourism demand lasting from one to six months, with only 50% of the locations recovering within three months or less. They also find that terrorist attacks that resulted in death or injuries had a longer negative effect on tourism than attacks that only damaged property.

The effects of terrorist attacks also spread throughout different industrial sectors of targeted economies. Corbet et al. (2019) estimate the economic impact of twelve terrorist attacks that took place in ten European cities between 2011 and 2018, including three coordinated attacks in Paris in November 2015 that claimed 137 lives. The study focuses specifically on the linkages between tourism and the aviation industry and shows how in the months after each attack, airlines cut capacity and customers bought fewer tickets in spite of significant fare reductions. Lanouar and Goaied (2019) study the number of tourist arrivals and overnight stays in Tunisia from January 2000 to September 2016, a period of time over which the country saw two terrorist attacks on tourist destinations as well as political violence in the wake of Tunisia's Jasmine revolution. The authors estimate a Markov switching model using monthly data and conclude that local terrorist attacks had a stronger impact on Tunisian tourism than international shocks such as the terrorist attacks of 9/11 or the global financial crisis of 2008. The study also find that the negative impact from terrorist attacks outlasted the negative impact from political violence. Moussa and Talbi (2019) further investigate the influence of political uncertainty caused by terrorism in Tunisia. They use data from the Tunisian stock market to estimate an EAGARCH model and find that terrorist attacks negatively influenced the performance and volatility of the Tunisian market index. The effect was found to be particularly strong in the period of political uncertainty that followed the Jasmine revolution.

The two terrorist attacks in Tunisia considered by Lanouar and Goaied (2019) and Moussa and Talbi (2019) had a large economic impact even though the attacks destroyed only a small fraction of the capital stock of the country. Abadie and Gardeazabal (2008) argue that this discrepancy between direct and equilibrium effects can be explained by open economy mechanisms that allow capital to move freely in response to a terrorist event. The authors use a dataset of country-level terrorist risk and foreign direct investment and find that higher levels of terrorist risk are associated with lower levels of foreign direct investment. On average, a one standard deviation increase in the risk of terrorism costs targeted countries about 5% of GDP worth of foreign direct investment.

Terrorism also causes countries to receive lower credit ratings and face higher costs of borrowing. Using a dataset on sovereign risk for 102 countries, Procasky and Ujah (2016) show that small increases in the risk of terrorism led to a downgrading of investor outlook. The effect is more pronounced for developing markets where terrorist activity was found to lower sovereign credit rating by an entire notch. In fact, emerging markets face relatively larger stock market volatility as a result of terrorism (Arin et al. (2008)). The effects of terrorist attacks on stock market prices caused by uncertainty extend to commodity markets as well. Gupta et al. (2017) use a quantile-predictive-regression approach to study gold returns and show that terrorist attacks have predictive value for the lower and particularly the upper quantiles of the conditional distribution of gold returns.

The economic effects of terrorism have been documented for various other economic settings. Abadie and Gardeazabal (2003) study the impact of the Basque conflict on regional GDP growth. The Basque Country was one of the richest regions in Spain in the 1970s in terms of GDP per capita, but by the late 1990's it had dropped to the sixth province out of 17. In the intervening years the region had experienced hundreds of terrorist attacks perpetrated by Euskadi-Ta-Askatasuna

(ETA), a Basque separatist group fighting for the establishment of an independent state. The attacks caused an estimated 523 deaths in the Basque Country alone. The authors propose a control method that compares the Basque Country to a synthetically constructed region without terrorism. Their analysis concludes that the terrorist attacks cost the Basque region about 10 percentage points in per capita GDP.

The costs of terrorism are not only the result of direct violence as in the Basque case. Abadie and Dermisi (2007) explore the effect that *perceived* levels of terrorism have on economic activity using commercial real estate in downtown Chicago as an example. Chicago was under threat of 9/11-style attacks, but no terrorist attack actually took place. The authors use building-level data on vacancy rates from the office real estate market and find that the taller, better-known landmark buildings as well as buildings in their vicinity had the highest increase in vacancies following the shock of the 9/11 attacks. Understanding terrorism risk is important because the mere perception of terrorist violence is enough to cause significant economic disruption

The fear and uncertainty generated by terrorist attacks also influence financial markets. Karolyi and Martell (2010) study 75 terrorist attacks between 1995 and 2002 that targeted publicly traded firms. Firms experienced a negative shock on the day of the attack that amounted to about US$401 million of firm market capitalization per company. Their study found that attacks that took place in countries with wealthier and more democratic societies were associated with larger drops in share prices, in line with the finding in Abadie and Gardeazabal (2008) described above regarding the cost of terrorism and the mobility of capital. Karolyi and Martell's study also find that attacks on human capital (e.g., the kidnapping of a company executive) had larger a negative effect on share prices than did material losses such as bombings of buildings or facilities. One reason may be that investors believe it is harder to recoup losses of human capital. Relatedly, Drakos (2010) tests the hypothesis of whether terrorist activity influences investor confidence using a capital asset pricing model estimated on a sample of 22 countries and he finds that terrorist attacks significantly lower returns on the day of the attack. The effect of terrorist attacks on investor sentiment is controlled for using an indicator found in the GTD database that measures the psychosocial effect of each attack. The study finds that the more severe the psychosocial effect of an attack, the stronger the negative effect on asset returns.

Terrorist attacks have also been shown to have political costs in the sense of influencing the outcome of elections. Berrebi and Klor (2006) study terrorist attacks in Israel between 1990 and 2003 and show that the relative support of right-wing parties increased after heightened levels of terrorism and decreased after periods with few attacks. Another widely studied case involves the bombings of three railway stations in Madrid on March 11th 2004 that resulted in 200 deaths and 200 injured. These were large-scale attacks that involved only a small loss as a fraction of GDP. They are estimated to have cost $US\$289$ million or the equivalent of 0.03% of the GDP of Spain (Buesa et al. (2007), Sandler (2014)). The political implications, however, were significant. The attacks took place just three days before Spanish parliamentary elections were held. Bali (2007) uses individual level survey data to show that the terrorist attacks motivated participation from citizens who would have otherwise not participated in the election. Their analysis also suggests the incumbent government's subsequent handling of the attack led many voters to switch to the opposition. These finding are further strengthened by the work of Montalvo (2011) who employs the synthetic control group method of Abadie and Gardeazabal (2003) using data on individuals who voted before the terrorist attacks took place. Their study also rejects the hypothesis that the winner of the Spanish parliamentary election was not influenced by the terrorist attacks of March 11th.

2.3 Rational Choice Models of Terrorism

2.3.1 Foundations of Rational Terrorism

The oldest known surviving set of laws is the Sumerian Code of Ur-Nammu found on a tablet from circa 2100 BCE. The laws are stated in "*if* x, *then* y" form. For each crime x there is a prescribed penalty y ranging from a payment of a few shekels to capital punishment. Deterrence was based on the premise that the threat of punishment would reduce criminal activity. In economics and criminology, studies of deterrence evaluate the effectiveness of policies designed to dissuade individuals from engaging in criminal behavior. The economic framework for thinking about terrorism is founded on the theory of household production where agents act in self-interest and rationally substitute between productive and criminal activities (Becker (1968), Ehrlich (1973)). The assumption is that individuals or groups engaged in criminal behavior make choices that maximize their utility subject to constraints. The rational choice paradigm has been used to investigate a wide range of criminal behavior, including burglary and shoplifting, car theft (see LaFree et al. (2009)), and also terrorism (Sandler et al. (1983) and Enders and Sandler (1995)).

Studies of deterrence are predicated on the assumption that terrorists make choices on the basis of rational self-interest. The rational approach to terrorism takes the abstract cause that terrorists fight for as given and studies the choices the agent makes in pursuit of their goals. The means and ends of terrorists are often reproachable or repugnant, but as will be shown in Section (2.3.3) there is evidence that terrorists act rationally in pursuit of their goals. Terrorist groups have limited resources yet they are able sustain long term conflicts against much more organized and resourceful opponents. Moreover, the use of terrorism as a tactic is widespread and has been adopted around the globe by terrorist groups of diverse ideologies. [4]

Some of the most shocking terrorist tactics that involve seemingly altruistic or irrational behaviors such as suicide bombings have been shown to be motivated by self-gratifying benefits consistent with a rational choice theory (see Pape (2003), Azam (2005), and Perry and Hasisi (2015)). The behavior of suicide terrorists is driven mostly by anticipated costs and benefits associated with certain actions. Suicide terrorists consider the probability of success when evaluating their opportunities and in general there is no fundamental difference between terrorists and other types of criminals. These findings in particular validate the use of a rational choice framework to study the behavior of terrorist networks that rely heavily on suicide terrorism such as Al-Qai'da (Carson (2017)).

The rational choice theory of deterrence has two levels of inquiry: 1) why people engage in criminal behavior and 2) how do criminals make their choices. The event-based studies of terrorism considered in this dissertation are concerned with the second level of analysis. Terrorists as individuals can derive a pecuniary benefit from terrorism, but can also be rewarded in terms of recognition, achievement, or a sense of belonging. The intention of the terrorist to attack is taken as given and the focus of the sequence hazard is on how the agent chooses to attack over time in order to maximize their reward.

Terrorists face various constraints to their behavior. Policy makers intentionally design mechanisms to influence the decision-making of criminals by reducing the probability of success or the severity of an attack if one takes place. Terrorists can also be apprehended, killed, and

[4]An example of an influential study of terrorism that takes an opposing view, consider De Cataldo Neuburger et al. (1996) who write that "a terrorist does not simply weigh risks against the likelihood of success, as is normally the case, but adds into the equation the abstract value of the cause for which he or she is fighting. This is the reason why traditional notions of deterrence are ineffective against such a subject. Unlike a common criminal, he or she does not act on he basis of a cost-benefit analysis."

additionally face logistical costs due to the illicit nature of their planning. Policies that try to raise the cost to terrorists of engaging in terrorism are a combination of laws (e.g., the Patriot Act) and law enforcement (e.g., air marshals on US domestic flights). These counterterror measures can be passive/defensive (metal detectors, fortifications, harsher penalties for terror-related offenses) or active/offensive (raids, strikes, infiltration of terrorist groups) (Enders and Sandler (1995)). The studies considered in the next section employ interrupted time series techniques to test for evidence of rational terrorist behavior using data on how terrorist attacks before and after such interventions (Enders and Sandler (1993)).

Traditional thinking about deterrence has been extended to consider cases where government policies have the opposite of the intended effect. If deterrence is measured by the net decrease in the incidence or seriousness of future terrorism in a given area of operation, then a backlash effect is defined by a net *increase* (LaFree et al. (2009)). Policies designed to deter terrorism often involve restrictions on liberty or violent, punitive actions, and these policies may cause grievances that motivate terrorists to retaliate, increasing violent behavior. This is known as the "backlash" effect. For example, punitive actions may strengthen the within-group identity of terrorist groups or help with recruitment, particularly if the attempts to control behavior are viewed as illegitimate. Backlash has a weaker theoretical foundation than deterrence because it is not part of the static rational choice model of terrorism in the literature (see Landes (1978), Sandler et al. (1983), Enders and Sandler (1993), Enders and Sandler (1995), Enders and Sandler (2011)). Backlash implies a dynamic process whereby the severity of attacks increase as a response to counterterrorism. Additionally, statistical models of terrorism (including the series hazard model) do not account for the *outcome* of terrorist attacks even though these are observed. As a result they can only make inference on backlash indirectly (i.e., does the frequency of attacks increase or does the overall severity of terrorism increase).

Theories for why backlash occurs at an individual level have been proposed by Dugan et al. (2003) while McCauley (2006) does so specifically for terrorism. LaFree et al. (2009) test a model of backlash and deterrence in the context of Irish Republican Army terrorism and Argomaniz and Vidal-Diez (2015) performs a similar analysis in the context of Basque terrorism in Spain. Both these studies use the series hazard model presented in the following sections. Before presenting the model it is instructive to review two fundamental studies of empirical terrorism.

2.3.2 A Parametric Model of Airline Hijackings

Hsu and Quandt (1976) propose a statistical model for the time-series of hijackings of US-registered aircraft from 1962 to 1972. During this period there were 159 events where an individual or a group of individuals took control of an airplane under the explicit or implied threat of violence. Hijackings became more frequent after 1968, but by 1973 they largely disappeared. The early hijackings were largely political as the perpetrators would divert US aircraft towards Cuba. Over time the hijackings took on more narrowly economic aims. Air pirates known as parajackers would demand ransom money and escape by jumping off the hijacked aircraft mid-flight. Hsu and Quandt's study is the first to make the point that the money spent on airport security is several times larger than any given ransom demand. This fact motivates them to characterize the time series of hijackings with a view to informing prevention policy.

The approach in Hsu and Quandt (1976) is purely statistical in the sense that the model is not informed by the notion of an economic agent operating in a utility maximization framework, nor do they make use of covariates to support their model. However, the timing of changes in the intensity of hijackings in their model are shown to coincide with important policy initiatives and political events that affected the cost and benefit of a hijacking. Hsu and Quandt's study is important because in addition to offering a rigorous statistical examination of the time series of hijackings, it is also the first paper to find empirical evidence suggesting that the hijacking time series responded to changes to hijackers' valuation of hijacking or their probability of getting caught. The latter two are the main ingredients of an economic model of terrorism, starting with Landes (1978). The main take away from Hsu and Quandt's 1976 paper is that the intensity of hijackings is globally non-stationary, possible due to changes in the social and political context that hijackings take place in, but can be a locally homogeneous Poisson. It is worth going over Hsu and Quandt's 1976 framework in detail since it covers concepts closely related to the main model proposed in this dissertation.

The era of US hijackings began May 1, 1961 when a National Airlines twin-engine CV-440 was commandeered and diverted to Havana, Cuba. By December of 1972, 159 such events had take place, with 148 of them occurring between 1968 and 1972. The data display some obvious time-varying features.[5] Hijackings diverted to Cuba fell dramatically after a congressional hearing in February of 1969 was used to inform the public that the treatment of hijackers who landed in Cuba was a far cry from the commonly held perception that it was a hero's welcome. In September 1969 Cuba instituted an immigration reform that introduced prison sentences for hijackers making convictions more likely. Hijackings declined further as federal pre-boarding screenings were installed in airports nationwide. Successful hijackings disappeared almost entirely by 1973. The evolution of the hijackings series suggest that perpetrators responded to changes in

[5]Hsu and Quandt's 1976 data come from the 1968–1972 Chronology of Hijacking of US Registered Aircraft, prepared by the Office of Air Transportation Security, Federal Aviation Administration. The data include the date of the event, the type of flight (scheduled, chartered, private), the boarding point, and the destination. Only the date is used in their analysis.

the value of hijacking (recognition and good treatment in Cuba) and to changes in the probability of apprehension (security measures after 1970s). Hsu and Quandt show that the time-series of hijacking events is non-stationary and they locate statistically significant shifts in the Poisson intensity parameter early and late in the sample.

Hsu and Quandt's analysis proceeds in three stages. In the first stage they explore the notion that the point process underlying the hijackings is homogeneous Poisson. In the second stage they explore departures from the assumption of homogeneity, namely serial dependence and bunching. The presence of either of these would imply that the data generating process for hijackings changes over the sample and is thus non-stationary. In the last part of their analysis they fit a parametric, non-homogeneous Poisson model to the time-series of hijackings. The intensity of hijackings is allowed to gradually rise and then fall in a manner shown to be consistent with the major trends in hijacking prevention policy alluded to above.

The starting point for the analysis is the null hypothesis that the hijacking series is a homogeneous Poisson process. Consider the hijackings that take place inside a given length of time, say, a day. The number of hijackings could be none, one, or several. If the data generating process for the hijackings obeys the postulates of a Poisson distribution (e.g. see pp. 150-151 of Hogg et al. (2013)) then the probability of observing h hijackings in a given length of time is

$$\text{Prob}(\text{number of hijackings} = h) = \frac{e^{\lambda}\lambda^{h}}{h!}$$

where $\lambda > 0$.

From the vantage point of a Poisson distribution, on any given day the number of hijackings observed is the realization of a random variable whose expected value and variance are determined by the intensity parameter λ. This parameter fully describes the distribution. For example, $\lambda = 0.30$ is taken to imply a hijacking occurring on average every 3 days. Hsu and Quandt's entire focus is on how to model λ and how to interpret changes in λ over the sample. If there is imitative behavior among hijackers, for example, λ could increase from one occurrence to the next. Successful security policies could have the opposite effect. But the first step is to posit that λ remains the same throughout the entire sample.

Starting at January 1, 1968 let t_i, $i = 1, 2, \ldots$ represent the time elapsed in terms of days between hijackings so that t_1 is the number of days until the first hijacking, t_2 the number of days between the first and second observed hijacking, and so on. A well known result (e.g. see pp. 156-157 Hogg et al. (2013)) is that if the original process is Poisson, then interoccurrence times t_i will be distributed independently exponential with parameter λ (see Definition 6.1.8 in the Appendix)

$$f(t_i) = \lambda e^{-\lambda t_i}, \text{ for } i = 1, 2, \ldots. \tag{2.1}$$

A second implication of the Poisson assumption is that for a given total number of occurrences, the actual location of the occurrences on the time axis are independently and uniformly distributed over the interval $(0, T)$. Put differently, the quantities $\tau_i = \sum_{j=1}^{i} t_j / T$ are distributed uniformly over the $(0, 1)$ interval. Intuitively, homogeneity of the underlying Poisson distribution implies that the probability of seeing any number of hijackings remains stable throughout the sample, so *a priori*, no location on the time axis should have more hijackings than any other.

The two implications outlined above suggest ways of testing whether the underlying distribution is homogeneous Poisson. Note that investigating whether a process is Poisson is a different exercise from assuming that a process is Poisson and directly estimating λ. For one, the latter necessitates a theory about λ to begin with. To test the first implication that interoccurrence times are independently exponential, Hsu and Quandt use a test derived by Moran (1951). The test uses the mean $\bar{t} = \sum_{i=1}^{n} t_i$ of the interoccurrence times t_i and is approximately distributed $\chi^2(n-1)$ where n is the number of total occurrences (here $n = 148$). The Moran test is asymptotically most powerful for the assumed exponential distribution against the gamma alternative. To test the second implication, Hsu and Quandt use both the Kolmogorov-Smirnov statistic $D = \sup \left|F(x) - S(x)\right|$, where $F(x)$ is the hypothesized cumulative distribution and $S(x)$ is the sample cumulative distribution, and the Cramer-von Mises statistic $W^2 = n \int_0^1 \left[S(X) - F(X)\right]^2 dF(X)$. The union of these three tests cannot be rejected at the 0.2 level which provides preliminary evidence that the series underlying the hijackings may be homogeneous Poisson.

Next Hsu and Quandt explore two potential different departures from homogeneity. The first is the possibility of serial dependence between consecutive interoccurrence times, the second is the possibility of bunching of hijacking incidents which the authors attribute to imitative behavior. They cannot reject the hypothesis of serial independence. Hsu and Quandt validate this finding by examining of the significance of the lagged dependent value coefficient in a first order autoregressive model fitted to the interoccurrence time data by way of a transformation.

Bunching in the hijacking series is apparent if one plots the cumulative number of events against the elapsed time axis. A homogeneous process would have events rarely departing from the 45-degree line. Hsu and Quandt draw attention to moments early and late in the sample where there are clear departures from this uniform pattern. In order to explore the presence of bunching of incidents the authors employ a Durbin-Knott frequency domain decomposition of the Cramer-von Mises statistic through which they are able to reject the null hypothesis that each of the interoccurrence times are drawn from the same distribution.

The data cannot reject the assumption of no serial correlation, however, there is strong statistical evidence of bunching which is also corroborated by visual inspection of the series. The bunching appears to occur between the 300th and 400th day of the sample (the 400th day corresponds roughly to late 1972 when preventive measures were brought into effect). Hsu and Quandt divide the sample period into three mutually exclusive and exhaustive sets within which they assume that the Poisson process is homogeneous. By testing for the equality of the intensity parameter λ the authors conclude that the hijacking series is non-homogeneous Poisson with a higher intensity occurring during the three month period between October 1968 and early February of 1969. The separation of the sample into three non-overlapping groups is checked against a more general detection scheme that uses a moving block procedure that assumes equality of the Poisson intensity parameter λ between each adjacent sample. The result of this more general approach detects a shift in the Poisson parameter roughly before the 400th day and again towards the end end of the sample around 1972 when preventive measured were installed that significantly cut back on hijackings. On evidence, the time-series of hijacking events is non-stationary with a strong possibility of shifts in the intensity parameter located early and late in the sample.

Hsu and Quandt specify a fully parametric non-homogeneous Poisson model to accommodate their findings thus far. Instead of one intensity parameter λ for the entire sample they specify several parameters λ_k indicating the hijacking intensity following the kth occurrence time. The intensity λ_k is modelled using two pieces of a half $cosine$ curve, one increasing and one decreasing, to describe changes in the intensity after each occurrence.

The intensity parameter λ starts off at a level $\lambda = a$ until reaching occurrence K_1 after which the parameter gradually rises to a peak determined by b. After the K_2-th hijacking λ falls from b towards a level c. The parameter θ_i measures the speed of this change with respect to k through the expression $\theta_i(k - K_i)$ for $i = 1, 2$. The boundary condition $\pi/\theta_1 = k - K_i$ for $k > K_i$ specifies how many hijackings it takes to reach the limit point of λ (see Definition 6.1.10 for the details).

The $k = 1, 2, \ldots, 148$ observed hijacking interoccurrence times can be used to estimate parameters a, b, c, d, K_1, K_2, θ_1 and θ_2 described above. Working with interoccurrence times permits use of the exponential distribution written above in (2.1). Taking the product of likelihood contributions and then taking logs gives the log-likelihood function to maximize

$$\mathcal{L}(a, b, c, d, K_1, K_2, \theta_1, \theta_2) = \sum_{k=1}^{148} \log \lambda_k - \sum_{k=1}^{148} \lambda_k t_k.$$

Numerically maximizing the equation above for the sample of hijacking gives that $\hat{a} = 0.049$, $\hat{b} = 0.300$, $\hat{c} = 0.082$, $\hat{K}_1 = 11.24$, $\hat{\theta}_1 = 0.1396$, $\hat{K}_2 = 33.75$, $\hat{\theta}_2 = 3.8410$. The solution implies that hijackings climbed from August 4, 1968 (the 11th incident) through February 3, 1969 (the 33rd incident) and that at peak intensity there was a hijacking every three days ($\lambda = b = 0.3$). The increase after the 11th incident was gradual ($\pi/\theta_1 \approx 23$ occurrences), but the decrease is sharp after K_2 and only took a single occurrence ($\pi/\theta_2 \approx 0.82$).

The non-homogeneous Poisson model is validated by Hsu and Quandt in four ways. First, a non-homogeneous approach is appropriate due to the statistically significant presence of bunching in the series. Second, the authors calculate the conventional log-likelihood ratio statistic which shows a dramatic improvement over the simple Poisson model. Third, when they focus on the sample starting after the 33rd incident, they cannot reject any of the previously outlined tests at the 0.05 level. Fourth, the authors argue that the timing of the increase in intensity coincides with a period where more and more hijackers were diverting flights to Cuba. The beginning of the decline coincides exactly with a Congressional hearing in February, 1969 that made public the fact that hijackers who land in Cuba did not receive a hero's welcome but were instead detained, interrogated for arbitrary lengths of time, and often given hard labor. The authors claim but do not report that applying the analysis to a sample of worldwide hijackings leads to a similar conclusion in terms of the dynamics of hijacking intensity.

The next section reviews studies that explore Hsu and Quandt (1976)'s insights about the hijacking time series via regression models. These models rely explicitly on a rational choice framework and view policy interventions as a potential source of non-homogeneity. The regression methodology allows for the inclusion of covariates that describe how the rewards and costs of hijacking evolved over time. Crucially, it allows for simple statistical tests of the effect that counter terrorist policy interventions had on the frequency of terrorist attacks.

2.3.3 Interrupted Time Series Analyses of Terrorist Attacks

Economic studies of terrorism started with Landes (1978) who applied the Becker (1968) theory of crime and punishment to estimate a deterrence model of airline hijackings. Landes derived an implicit function that related the number of hijackings to variables that influence a terrorist's valuation and constraints to hijacking. As described above, hijackings were a popular terrorist tactic during the 1960s and early 1970s. Landes (1978) uses regression analysis to show that the dramatic reduction in the number of hijackings after 1972 can be attributed to new security measures implemented at airports such as metal detectors and screenings. The use of regression analysis motivated by an economic model of a hijacker allows Landes to formally test the deterrence hypothesis suggested by Hsu and Quandt (1976). Namely, whether or not the data generating process for the intensity of hijackings was influenced by changes in prevention policy and increased security measures.

Landes (1978) extends the hijacking data from Hsu and Quandt (1976) to cover events between 1961 and 1976. After the first U.S.-registered aircraft is hijacked in May of 1961 there were only seven more such incidents until 1967, after which 124 hijacking events took place between 1968 and 1972. Hijackings then dropped in frequency with only ten events taking place after 1973, a pattern observed as well for hijackings taking place outside the U.S.. Unlike Hsu and Quandt (1976), Landes's main focus is a formal test of how *ex-ante* and *ex-post* deterrence efforts can account for the sharp reduction of hijackings in the U.S..[6] Whereas Hsu and Quandt (1976) characterize the raw time-series of hijackings, Landes (1978) seeks to identify a causal effect. Landes pits a deterrence hypothesis against a fad hypothesis that posits that U.S. hijackings were merely an imitative trend that would have lost momentum regardless of any deterrence efforts. To be clear, Hsu and Quandt (1976) also relate the rise and fall in hijacking intensity of their model to actual policy changes, but they do not formally test whether there is a causal effect. Landes' modeling choices reflect this different imperative.

Landes writes a non-parametric utility function for a potential hijacking offender. The utility function has to account for any factors that may influence the difference in returns between legal and illegal activities. A hijacker does not know ex-ante that they will be caught, so their utility

[6]Passenger screenings and pre-boarding checks are examples of *ex-ante* deterrence efforts. *Ex-post* deterrence policies involve the kill or capture of a hijacker once the event is underway. The addition of armed Federal Marshals on U.S.-registered flights and a treaty between the United States and Cuba on February 15, 1973 to cooperate on the extradition and punishment of hijackers are examples of *ex-post* deterrence policies.

involves a lottery over all the relevant possible outcomes. Suppose a hijacker considers taking over an aircraft flying from country i to j (where country j may be identical to i if the terrorist event is a domestic hijacking). The utility function may be written as

$$\bar{U} = (1 - P_a)U(W_j) + P_a P_c U(W_i - S) + P_a(1 - P_c)U(W_j - C) \tag{2.2}$$

where P_a is the agent's estimate of the probability of apprehension occurring in i, P_c is the conditional probability (given apprehension) of conviction and incarceration, and W_j and W_i are the agent's wealth in country i and j. This wealth component also includes any non-pecuniary benefits from hijacking an airplane (say, for those political dissidents headed to Cuba). The first quantity in the utility function is the utility from arriving to location j without being apprehended. If apprehended, the agent faces the chance of being sentenced. The second term is the utility of being sentenced in country i at a cost S. If apprehended but not sentenced, the agent incurs a cost C. An agent's reservation utility is determined by their reward from staying in country i without attempting a hijacking, $U = U(W_i)$. An agent will attempt a hijacking if and only if $\bar{U} \geq U$. A necessary condition for this is that $W_j > W_i$.

The utility function implies that if the probability of apprehension P_a or incarceration P_c increase, or if the costs of incarceration C or sentencing S increase, or if the relative rewards to hijacking $W_j - W_i$ decrease, then $\bar{U}$ will decrease and with it the likelihood of a hijacking taking place. Landes aggregates among all potential hijackers to arrive at an offense function O given by

$$O = O(\bar{P}_a, \bar{P}_c, \bar{S}, \bar{C}, \bar{Z}, \bar{X}) \quad (2.3)$$

where the first four elements are time t averages of the variables specified previously and where $\bar{Z}$ is a vector of the time t average wealth differential between country i and j, and $\bar{X}$ denotes the inclusion of other variables. The analysis presumes that offenders cannot affect the equilibrium probability of being apprehended or incarcerated. The theory predicts that the observed level of offenses is negatively related to the quantities $\bar{P}_a, \bar{P}_c, \bar{S}$ and $\bar{C}$, and positively related to net wealth $\bar{Z}$.

Equation (2.3) above gives Landes a guide to estimating the model. An important component is how to aggregate the left hand side variable (the number of offenses) in a way that is amenable for regression analysis. The right-hand side contains quantities that need to be calculated (e.g., the probability of apprehension) and are potentially endogenous since the empirical probability of apprehension depends on the number of offenses observed. If the intensity of hijackings increases then the probability of apprehension can decrease if the same number of resources are being spread over more offenders. The intensity of hijackings can also decrease if within that same period if more resources are spent on catching offenders. Landes argues that the latter situation is more likely so he includes lagged deterrence variables in the regression to deal with the possibility of the estimates having a downward bias. No hijackings have taken place at the same airport on the same day and only two hijackings have ever taken place on the same day. In practice it is unlikely for the level of offenses to reach a point where the probability of apprehension is decreasing in the number of hijackings.

Before turning to the hijacking data used in the study it is is helpful to point out that nothing so far precluded Landes from adopting an empirical framework that is closer to Hsu and Quandt (1976). Specifically, Landes could model the time t intensity of hijackings λ as depending on observable quantities as outlined by equation (2.3).

$$\lambda = g(\bar{P}_a, \bar{P}_c, \bar{S}, \bar{C}, \bar{Z}, \bar{X})$$

where as in Landes' model the right hand side variables are time t averages and where $g(\cdot) > 0$ is a link function to satisfy positivity of the Poisson parameter. Note that the Poisson process implied by the parameter as specified above is potentially non-homogeneous. The model would be fit to interoccurrence times and the null hypothesis would be that the parameters on the deterrence variables are not significantly different from zero. The likelihood criteria for comparing the restricted and unrestricted models is straightforward. This method potentially dispenses with the practical problems that lead Landes to work with aggregate data. The interoccurrence hijacking time series could be modeled directly. This is the focus of the series hazard model presented in the next section.

Instead Landes explores three different ways of aggregating the raw hijacking time series. The first is to sum the number of hijackings that take place every quarter over the course of the 16 years of data (64 observations). Hijackings occur in around half of the quarters which means that by construction there is a lot of missing sentencing and incarceration rate data. Landes addressed this by filling in missing data with the prediction from quarterly regressions of the data series on the probability of apprehension and the probability of incarceration if apprehended. The quarterly deterrence series may have a large random component because it is calculated from actual apprehensions. If there are not many apprehensions in any given quarter then the measure may depend heavily on features idiosyncratic to the particular hijacking offenses. Landes uses moving averages of quarterly deterrence variables to ameliorate this effect.

A second way of aggregating hijacking data is by studying the time interval measured in days that elapse in between hijacking events. This is closer to what Hsu and Quandt (1976) do using interoccurrence times. Doing so increases the number of observations from 64 records of incidents per quarter spanning 16 years up to 143 observed interoccurrence times, one per observed incident. Landes (and later Dugan (2011)) offers further motivations for taking this perspective. It is worth going through his reasoning since interoccurrence times are the primary focus of both the series and sequence hazard models discussed later.

Suppose the probability of a hijacking taking place on any given day is measured by $p \in (0, 1)$. Then the duration of days between hijackings is equal to 1 if another hijacking takes place immediately (with probability p), or it is equal to 2 if another hijacking takes place after the first day (with probability $(1-p)\,p$), and so on, so that the expected number of days until the next hijacking can be calculated as

$$\begin{aligned}
\mathbb{E}(\text{days until next hijacking}) &= p \cdot 1 + (1-p) \cdot p \cdot 2 + (1-p)^2 \cdot p \cdot 3 + \dots \\
&+ (1-p)^{n-1} \cdot p \cdot n \\
&= p \left[\sum_{i=1}^{n} \frac{\partial\,(1-p)^i}{\partial\,(1-p)} \right] \\
&= 1/p \text{ as } n \to \infty.
\end{aligned}$$

Thus the expected time until the next hijacking is the reciprocal of the probability of a hijacking, and the larger the probability of a hijacking, the sooner one should expect one to take place. If one expects that deterrence efforts decrease the probability of a hijacking, then it follows that the length of time in between hijackings should increase on average.

The downside, according to Landes, is that if deterrence does its job, there is a smaller sample of offenders from which to make inference on. Consider the deterrence methods implemented after 1973 that caused the number of offenses to fall drastically and the probability of apprehension to increase as well. As a result there were fewer observations, the interoccurrence times were longer, only a few hijackings were observed over several years, and in each quarter the probability of apprehension was high. In this scenario it is difficult to elucidate a deterrence effect, and in the limit the number of offenses falls completely. This is a valid observation by Landes in the context of regression analysis, but it could be dealt with by adopting a parametric method closer to Hsu and Quandt (1976) since this observed fall in offenses is something their model can account for. The question then would be then whether or not including deterrence variables in the model for λ noticeably improves the fit of the model over one in which deterrence variables do not play a role in determining the attack intensity.[7]

Landes finds statistically significant evidence of a deterrence effect. The result holds whether the raw data on aircraft hijackings are transformed into quarterly observations, the time interval between successive hijackings, or the number of air carrier flights between successive flight intervals. The probability of apprehension and other deterrence variables lowered the number of quarterly hijackings and increased both the time elapsed and the number of flights

[7]A third method that Landes (1978) uses to aggregate the left hand side variable of hijacking offenses is to count the number of air carrier flights that take place between successful hijackings. This view takes on a similar interpretation as the time interval perspective.

completed between successive hijackings. Variables unrelated to deterrence such as the level of unemployment or personal consumption expenditures were not found to be statistically significant in any of the regressions.

Landes addresses the fad hypothesis for hijackings by including foreign hijackings as a dependent variable in the regression for U.S. hijackings. Not only are the coefficients of the deterrence variables not affected by the inclusion of the foreign hijacking series, the U.S. deterrence variables themselves do not affect foreign hijackings, providing evidence that the deterrence effect previously found for U.S. hijackings is not merely spurious. Landes uses his regression results to forecast that there would have been between 41 and 67 more hijackings between 1973 and 1976 if mandatory screenings had not been instituted and if the probability of apprehension had remained constant and equal to its 1972 value. Finally, using data on the actual costs of screenings implemented nationwide, Landes provides a range for the implied dollar equivalent of the loss to an individual passenger from experiencing a hijacking.

Empirical Analyses of Rational Terrorism

Since Landes' work a number of studies have used a rational choice framework to guide empirical analysis. Enders and Sandler (1993) formally augment Landes' framework to explore how a rational terrorist would optimally respond to counterterror policies by substituting not only between terrorist and non-terrorist activities (say, hijacking and not hijacking), but also between types of terrorist activities (e.g., hijackings vs bombings). They use vector autoregression and intervention analysis to test their model against quarterly data on transnational terrorist events from 1968 to 1988. They find strong evidence that different attack modes act as substitutes of each other, in other words, policies that target one mode of attack may induce an increase in the frequency of another mode. For example, improved security at embassies reduced direct attacks on the buildings, but resulted in an increase in assassinations. Enders and Sandler (1995) study key counterterror policies enacted by the U.S. such as the drastic increase in embassy security in October 1976 (passive) and the retaliatory airstrikes against Libya in 1986 (active) after that country allegedly played a role in the bombing of a discotheque in West Berlin.

Sandler and Enders (2004) and Sandler and Enders (2007) provide a similar overview of methods using updated data series. Dugan (2010) surveys the main autoregressive and interrupted time-series techniques used in empirical work on terrorism. Sandler (2014) provides an up to date summary of studies that use the rational actor model. Their study highlights key areas of research on terrorism, one of which is the study of counterterrorism offenses.

The sequence hazard presented in Chapter 3 assumes that planning time is an important input into the production of the next terrorist attacks. A number of studies consider the role of planning in the context of cycles in the production of terrorist attacks. Enders et al. (1992) applies spectral analysis to a time series of transnational terrorist attacks between 1970 and 1989. All the attack modes show evidence of cyclical behavior. Logistically complex attacks were associated with longer planning cycles (i.e., more quarters in between attacks). Their findings are validated by Enders and Sandler (1999) who extend the transnational terrorism data to 1996. Consistent with a rational choice framework, these empirical studies find that 1) the data are consistent with terrorists responding to binding counterterrorist policies by substituting to attack modes that are less costly and 2) terrorist attacks that are more logistically complex and require more planning (such as hijackings) have longer cycles than less complex attacks such as bombings (Sandler and Enders (2004)).

The first wave of empirical papers on terrorism relied on autoregressive and interrupted time series methods that allowed for covariates and used dummies to represent policy interventions.

Enders and Sandler (2002) used this method to study a time series of fatal transnational terrorist events that occurred between 1970 and 1999. The results of this approach are compared to a regime switching threshold autoregressive (TAR) model which provides a better fit to the fatality series. The model describes a reversion to the mean: during eras with high levels of terrorism groups were unable to sustain above average levels of terrorism, but during low-activity eras, sustained increases in attack frequency were feasible. One reason may be that increased levels of terrorism are not sustainable if the group is constrained by law enforcement or retaliation that followed a previous attack. During periods of low terrorism agents can reallocate their resources which enables them to sustain a higher number of casualties per attack. On the other hand, during periods of intense terrorist activity, groups cannot sustain large attacks and thus quickly revert back to the mean attack size. Enders and Sandler (2005) perform a similar analysis comparing the TAR model to the autoregressive approach and find that allowing for high and low terrorism regimes outperforms the autoregressive approach that only gives an average representation

The majority of the events analyzed in the studies cited so far are transnational terrorist events, but evidence of rational terrorism also applies to domestic terrorism. Berrebi and Lakdawalla (2007) analyze the determinants of terrorism risk in Israel from 1949 to 2004 and find the data are consistent with terrorists rationally selecting targets on a cost-benefit basis. For example, terrorists are more likely to strike cities close to international borders, cities that are symbolic centers of government power, or cities with more densely populated Jewish areas.

A common method for fighting terrorism is the use of targeted killings. For example, in response to a wave of terrorist of attacks in 2002 the government of Israel instituted a policy of targeted killings against terrorists within its occupied territories. A targeted killing is the premeditated use of deliberately lethal force against specific individuals, for example, one not in the custody of government forces. This policy has been implemented as part of the "war of terrorism" waged by the U.S. government in the aftermath of the attacks of September 11, 2001 (Yoo (2011)). The killings are carried out using kill–capture missions or unmanned aerial vehicles. Such a policy can be particularly helpful when terrorist have top-down type of organizational structures. D'Alessio et al. (2014) used an ARIMA design to investigate the effect that disabling Abimael Guzman the leader of the Maoist insurgent group the Shinning Path had on the frequency of attacks the group committed in Peru between January 1979 and December 2009. Their intervention analysis showed that attacks dropped significantly six months after the capture of the insurgent leader. Though arguments have been made in favor of targeted killings in specific circumstances (e.g., Israel (Luft (2003)), prior research has shown that decapitation strikes are ineffective and counterproductive against religious groups such as those found in the jihadist context (Jordan (2009) and Mannes (2008)). In a more recent study of targeted killings of Al-Qai'da operatives in Iraq, Carson (2017) finds that the use of targeted killings had negligible effects or resulted in a backlash (more terrorist attacks).

2.4 Series Hazard Model

2.4.1 Method Overview

In economic studies of terrorism the goal is to detect the effect that a counterterrorist intervention has on terrorist activity. Researchers use interrupted time series methods where the events of interest are aggregated according to a given time frequency (e.g., daily, monthly, or quartlery) and inference is done on the frequency of attacks within that time frame (e.g., Landes (1978)). Terrorist attacks are event based outcomes that are publicly observed so the researcher observes the time an attack take place and the state of the world at the time of the attack. It is possible to use this event-specific information, as well as the time elapsed between events, to inform a model that avoids the temporal aggregation (e.g. Landes (1978)). In a duration model the dependent variable is the time until a discrete event such as death, a strike, accepting a job offer, or in this case, a terrorist attack. In this manner the subtleties of period to period differences are not lost as they would be if aggregated. Many studies have taken advantage of the availability of detailed data on terrorist events to estimate the risk of terrorist actvity after deterrence policies were put in place (Dugan et al. (2005), LaFree et al. (2009)).

In economics, researchers often observe many units that experience a single spell (Lancaster (1979)). Elbers and Ridder (1982) show that the proportional hazard model can identify the effect of time dependence from unobserved heterogeneity if there are explanatory variables in the hazard. If the mean of the mixing distribution is finite then all that is needed to identify the model is that a regressor have two distinct values. Heckman and Singer (1984a) extend the identification results in Elbers and Ridder (1982) to include unobserved unit-specific variables and they relax the finite mean condition. Heckman and Singer (1984b) showed that single spell duration models are sensitive to the choice of mixing distribution and they propose a nonparametric maximum likelihood estimator for the distribution of unobservables (see the handbook chapter by Heckman and Singer (1986)).

The commonly employed series hazard assumes proportionality. Heckman (1991) showed that a hazard is non-parametrically underidentified unless a proportional hazards assumption was imposed which he argued was not justified by economics. Han and Hausman (1990) proposed a flexible parametric proportional hazard model that combined a non-parametric estimator of the baseline hazard with a parametric specification of the model covariates. In practice many studies choose a gamma mixing distribution to model unobservables, Abbring and Van Den Berg (2007) rationalize this choice by characterizing the gamma distribution as the limit of a mixture of exponentials, but there is no strong economic justification for the choice. In contrast, the sequence hazard presented in Chapter 3 is based on an economic model of behavior that results in a multiple-spell hazard that is non-proportional in time.

Hamerle (1988) formulates a stochastic process to study risk across multiple-spells. Honoré (1993) investigates the identifiability of such models, extending the results of Elbers and Ridder (1982) and Heckman and Singer (1984a) to models with lagged duration dependence over multiple spells. He shows that when studying single risk, mutiple-spell data allow for less stringent conditions for identication than do single-spell data. McCall (1996) studied the identifiability of the mixed proportional hazards model with time-varying regressors (see Van den Berg (2001)). Abbring and Van den Berg (2003a) extend the identification of the multiple spell proportional hazard model to a competing risks framework. Abbring and Van den Berg (2003b) study the analysis of treatment effects in multiple-spell data with unobserved heterogeneity that is dependent over time and where regressor do not take necessarily take on two distinct values. They impose the assumption that treated units do not anticipate the treatment so that the timing of events can be informative. Abbring and Heckman (2007) study treatment effects in the context of dynamic discrete choice models.

Studies of terrorism ignore unobserved heterogeneity even when modeling more than a single terrorist group. The approach most commonly used in the literature is the Cox proportional hazard model. Dugan (2011) formalizes the benefits of using duration models in the context of studying terrorism. Specifically, the series hazard model is an extension of the Cox proportional hazard model to multiple observations of a single unit. The series hazard estimates changes in risk for continuing events conditional on the relevant states of the world. These may include the history of past attacks or date-specific characteristics like whether or not a counterterror policy or law is in place. The series hazard model is built around the hazard function

$$\lambda_i \left(t \,|\, \boldsymbol{X}_i\right) = \lambda_0 \left(t\right) \cdot \exp\left(\boldsymbol{X}_i \beta\right) \tag{2.4}$$

where k indexes each individual attack carried out by a single terrorist group. The attack-specific social, political, and deterrence contexts for each event are collected in $\boldsymbol{X}_i$. The quantity $\lambda_0 \left(t\right)$ represents the baseline hazard obtained when $\boldsymbol{X}_i = 0$. That is, if $\boldsymbol{X}_i$ is composed of dummy variables representing policy interventions, then the baseline hazard represents the status quo absent of any policy. Note that $\boldsymbol{X}_i$ can also contain functions of past failures. Let $\mathcal{R}\left(t\right)$ represent the risk set at time t. The risk set includes all the peacetime durations that have yet to end at time t. Following Cox (1972), for a particular attack taking place at time t, conditionally on the risk set $\mathcal{R}\left(t\right)$, the probability that the attack that takes place is the one observed is

$$\exp\left(\boldsymbol{X}_i \beta\right) \Big/ \sum_{l \in \mathcal{R}(t_i)} \exp\left(\boldsymbol{X}_l \beta\right). \tag{2.5}$$

Each of the N attacks contributes this factor so the conditional log likelihood is written

$$\mathcal{L}\left(\boldsymbol{\beta}\right) = \sum_{i=1}^{N} \boldsymbol{X}_i \beta - \sum_{i=1}^{N} \log\left[\sum_{l \in \mathcal{R}(t_i)} \exp\left(\boldsymbol{X}_l \beta\right)\right]. \tag{2.6}$$

Within subject correlation implies that the time until an event takes place may be correlated for multiple events. In single-spell data the spells can be correlated if the agents belong to a group or cluster (e.g., a certain industry in industrial organization, or siblings in the medical literature).

In multiple-spell data where each subject can experience more than one event then the events may be correlated over time. For example, a terrorist group can have a long spell of inaction as they build their attack, launch the attack, and then attack again shortly after.

The original formulation of the proportional hazard model (PHM) is due to Cox (1972). Prentice et al. (1981) is another early reference. Lin and Wei (1989) derived the asymptotic distribution of the maximum partial likelihood estimator of the vector of regressor coefficients when a Cox model is potentially misspecified (see also Box-Steffensmeier and Zorn (2002)). Extensions of the Cox model to clustered or multiple data event are provided in Kelly and Lim (2000). For a survey of PHM for multiple failure see Ezell et al. (2003) and the textbook by Kalbfleisch and Prentice (2011). Willett and Singer (1995) propose a parametric discrete hazard model of multiple spells that does not depend on the assumption of non-proportionality.

The series hazard model has been widely applied to the study of terrorism and other conflict-related phenomena (e.g., active shootings Kissner (2016)). Barros et al. (2006) and Argomaniz and Vidal-Diez (2015) applied series hazard model to the study of ETA counterterrorism. Note that these studies do not necessarily model a single agent. Dugan et al. (2005) studies airline hijackings by multiple perpetrators and uses the series hazard again in Dugan et al. (2008) model attacks by two different Armenian terrorist groups. LaFree et al. (2009) applied the series hazard to the study of terrorism in Northern Ireland and are the first to introduce the idea of a backlash. The series hazard was also employed by Carson (2014) to show that the hazard of terrorist attacks by radical eco-groups overall decreased in response to legal sanctions.

Duration Dependence

The way the hazard rate changes with time is called duration dependence. It is one of the key elements of duration analysis. If $\frac{\partial\lambda(t^*)}{\partial t} = 0$ at $t = t^*$ then there is *no duration dependence*. There is *positive duration dependence* if $\frac{\partial \lambda(t)}{\partial t} > 0$ at $t = t^*$. This means the probability of an attack happening soon increases as more time goes by with no attack taking place. For example, Berrebi and Lakdawalla (2007) found that the risk of a terrorist attack in Israeli cities shows positive duration dependence while less urban locales show negative duration dependence meaning the risk of an attack subsides as time elapses from the last attack. *Negative duration dependence* occurrs if $\frac{\partial \lambda(t)}{\partial t} < 0$ at $t = t^*$.

As previously mentioned, a limitation of series hazard model is that it does not model outcome data even though outcome data is available: "by relying exclusively on series hazard modeling, we are unable to estimate the impact of an intervention on the number of fatalities, even if the deaths were measured during each event." Dugan (2011)-

2.4.2 Model Diagnostics and Tests of Proportionality

The series hazard model is basically identical to the Cox proportional model without recurrent events so all of the same diagnostics of the Cox model apply to the series hazard model. The same tests can be used to test the proportionality of the hazard over time Dugan (2011). The proportional hazard asumption means that the effects do not change with time beyond the parametrized form. Provided the specification is correct, the effect of a covariate does not change over time and adding additional variables add no additoinal exlanatory power. The tests of proportionality are essentially specification test that interact the covariates with the time index and verify whether or not the coefficient on the interacted variable are significantly different from zero (see Cleves et al. (2008)). The most popular test of non-proportionality for the Cox model is based on the Schoenfeld residuals (Schoenfeld (1982)).

2.5 Series Hazard Model Analysis of ETA Counterterrorism

Argomaniz and Vidal-Diez (2015) applies the series hazard model (Dugan (2011)) to a case study of six major counterterrorist initiatives targeting the Basque terrorist group Euskadi Ta Askatasuna from 1977 to 2010. The theoretical framework of LaFree et al. (2009) is used to determine if government interventions resulted in a reduction of terrorist violence (deterrence) or if they caused defiance and retaliation that led to an increase in terrorist violence (backlash). Argomaniz and Vidal-Diez (2015) control for political negotiations and policies that provide alternative explanations for changes in ETA's pattern of attacks. Their analysis points to a mix of deterrent and backlash effects that are next placed in context.

ETA is the most widely studied terrorist group in history and leading up to its dissolution in 2018 it was Europe's largest ethno-nationalist terrorist network. ETA (Basque Homeland and Freedom) was founded on July 31, 1959 by a break-away group from the Basque Nationalist Party's with the purpose of fighting for an independent Basque state. The Basque people are spread across seven provinces, four of which are in northern Spain in the Basque Autonomous Community (BAC) and Navarra. The other three are located in the Pyrenees regions of south-west France. The Basque are culturally distinct from their neighbors, the only ones in Spain and France where the language is not based on Latin. The Basque speak Euskera, the oldest living language spoken in Europe. Euskera is older than any Indo-European language of Europe and its origins are still unknown. The language is spoken by many Basque and still taught in the majority of schools (Barros et al. (2006)).

Throughout its long history ETA has been composed of loosely connected networks: a small hard core of activists aided by several hundred supporters in an around the Basque region. ETA's terror cells were dispersed, non-hierarchical, and at times acted largely independently from each other. The operations were largely financed by extortion, kidnapping, and robbery of wealthy Basque industrialists. The effect of this was flight of human capital and foreign direct investment which in turn lowered GDP per capita in the region. ETA's attacks took place mostly in the Basque country which accounted for around 70 percent of all deaths between 1968 and 1997. Between 1979 and 2007 all but six of the 47 provinces in Spain experienced at least one attack (LaFree et al. (2012)). ETA also struck abroad in France and targeted Spanish interests as far as Mexico, but the Basque region saw 37 times the number of deaths per year per inhabitants than the rest of Spain (Abadie and Gardeazabal (2003)).

ETA's campaign of terror began in 1961 during Franco's regime. The first salvo was a failed attempted derailment of a train carrying veterans of Spain's civil war. They were en route to San Sebastian to celebrate the 25th year anniversary of the war. The first lethal attack took place in 1968 when a national police officer and active member of the Franco's civil guard was killed by ETA operatives. Franco's government retaliated, resulting in mass arrests and the military tribunals of Burgos in 1970 after which ETA averaged no more than two victims in any given year until 1973. In the mid-1970s ETA start carrying out more sophisticated attacks, the most impactful being the assassination of Spanish premiere Carrero Blanco in 1973. After Franco died in 1975 Spain began a 2-year transition to democracy and a move towards a more liberal policy towards the Basque region, culminating in a general amnesty in 1977. ETA attacks continued unabated during this period, averaging about 16 per year until July 1 of 1978 when the Spanish government promulgated a law that gave police new powers of surveillance, arrest, and detention, and also eliminated bail for suspected terrorists. Although enacted by the Socialist Party, the law of 1978 was enforced by a security apparatus inherited from the Franco era which resulted in frequent instances of mistreatment and torture. Argomaniz and Vidal-Diez (2015) investigate whether this law had a net deterrent or backlash effect.

In tandem with the coercive deterrents of the law of 1978, the Spanish government enacted the Guernica Statute on July 19, 1978 which gave the Basque state extensive autonomy over its own affairs, including its own parliament and judiciary, an independent police force, Euskera radio and television, the ability independently to set health policy, and various historic rights recognized in the Spanish constitution. During this time ETA's lethality saw a marked increase. Up until this time, ETA attacks had focused mainly on the Basque Autonomous Region, but after they became more widely dispersed (LaFree et al. (2012)) and between 1978 and 1980 they claimed 235 victims. In response, the left-wing government allied with right-wing extremists and foreign mercenaries in December of 1982 to form the GAL (Anti-terrorist Liberation Group), a paramilitary unit tasked with fighting a dirty war against ETA. The GAL were mainly active between 1983 and 1987. Its covert nature allowed the GAL to operate in French regions that up to this point had served as a haven for terrorist suspects wanted in Spain. GAL operations also increased pressure on the French to act against ETA inside their own borders. This lead to an increase in extraditions and deportations of ETA operatives by French security forces and a drop in the severity of ETA terrorism to 39 deaths on average per year. In 1986 a joint Spanish and French contingent raided the *Sokoa* furniture plant in the Basque region of France and led to the arrest of nearly a dozen ETA members and the capture of a cache of arms and explosives.

With ETA somewhat depleted from the government offensive, in January 1988 all the political parties of the Basque parliament (with the exception of ETA's political wing Harri Batasuna) agreed to the Ajuria-Enea Pact. In it they repudiated ETA's violent tactics, but at the same time signaled a departure from Franco era counterterrorist policies and a reduction in leftover personnel from the era. The pact was an effort to professionalize counterterrorist efforts and reduce instances of mistreatment. In March 1992, as the Olympic games to be hosted by Spain neared and the threat of an ETA attack increased, French and Spanish security forces arrested almost all of ETA's executive leadership found hiding in the town of Bidart, France. This raid spelled the beginning of a slow end for ETA. During the 1990s they averaged 16 deaths per year. A peace conference was held in in Algeria in 1992 that failed to alter the status quo. ETA responded by increasing the profile of their attacks. In April 1995 they carried out a failed assassination attempt on the leader of the Popular Party Jose Aznar and followed that by an attempt on King Juan Carlos in August 1995.[8]

In September of 1998 ETA declared a cease-fire and agreed to the Lizarra pact whereby Basque political parties committed to supporting Basque independence. The peace was short and the cease-fire ended in November 1999, but attacks stopped for a while until restarting again in 2000 when ETA killed 23 people. In 2002 the Socialist Party and right-wing Popular Party joined to ban Herri Batasuna, ETA's political wing. The banning had a profound effect on ETA operations. Attacks occurred at a low intensity until December 2006 when two people where killed and 52 others were injured by a van bomb that exploded at the airport in Madrid. In January of 2011 ETA announced a ceasefire and in 2017 the group gave up its cache of weapons and explosives. ETA was completely dissolved in April 2018.

Previous studies have also investigated the efficacy of the counterterrorist policies against ETA. Barros (2003) uses the VAR framework of Enders and Sandler (1991) to study kidnapping and assassination incidents perpetrated by ETA from 1968 to 2000. Barros (2003) these to a constant,

[8]Aznar came to power May 1996 and remained in power until the Madrid train bombings on March 2004 referred to in Section (2.2).

a lagged value of endogenous variables, and contemproray or lagged values of other variables that control for deterrence and economic or political factors. Their study finds that deterrent efforts such as the number of police or the military expenditure do not have a significant effect on the series of incidents. On the other hand, political factors such as a democratic government in office does induce more terrorist incidents. Economic variables such as a measure of tourist activity (nights slept in hotels) and the level of foreign direct investment had no statistical effect on the level terrorism.

Barros et al. (2006) use a variety of specifications to study the hazard of terrorist attacks committed by ETA in the period 1968 to 2002. The time elapsed between ETA's terrorist attacks are fit using an exponential, Weibull, Cox, and piecewise constant exponential (PCE) and PCE with gamma heterogeneity (HPCE) (Lancaster (1990)). The PCE is like the Cox proportional hazard model presented in Section 2.4.1 only the baseline is allowed to changed at a pre-specified interval. Barros et al. (2006) finds the PCE using 10-day intervals (with a constant hazard after 90 days) outperforms the other methods. The HPCE model is similarly specified, but for the inclusion of a multiplicative factor representing unobserved heterogeneity. Overall, Barros et al. (2006) finds that ETA attacks decreased as a consequence of the political regime of Franco and increased with democratic governments. The Bidart raid that led to the capture of ETA leadership was the most effective deterrent. Political pacts were found to not play a significant role in reducing ETA activity. Sanchez-Cuenca (2009) uses an ARIMA model to study quarterly attack data covering the same time period as Barros et al. (2006) but ending in 2007. In this analysis, again, the repressive tactics of the Franco regime and the joint raid on Bidart were found to be the strongest deterrents to terrorism.

Argomaniz and Vidal-Diez (2015) uses the series hazard extension of the Cox proportional hazard model presented in Chapter (2.4). The hazard is assumed to take the form

$$\lambda_k\left(t \mid data_k\right) = \lambda_0 \exp\left(\beta' Government\, Interventions_k + \gamma' Controls_k\right)$$

where k represents each terrorist attack, λ_0 represents the baseline hazard function with unspecified distribution, and β_i and γ_j are coefficients of the variables used as interventions and controls, respectively. Following LaFree et al. (2009) and Dugan (2011), the set of controls includes the cumulative month to adjust for trends and a measure of attack density. The latter represents the days covered by the three most recent attacks, including the current attack. This measure is intended to explain the expected correlation between attacks. By controlling for the momentum of previous attack, the model can account for the rising tide of ETA attacks after the 1978 law.

The main counter terrorist policy interventions that Argomaniz and Vidal-Diez (2015) focuses on are the anti-terror law of 1978 (July 1, 1978 - October 29, 1980), the dirty war conducted by the GAL (October 17, 1983 - July 24, 1987), the French policy of deportations and extraditions starting January 1, 1984, the Sokoa (November 5, 1986) and Bidart (March 29, 1992) raids, and the banning of Harri Batasuna (August 26, 2002). In addition, Argomaniz and Vidal-Diez control for ceasefires and negotiations that affect the net level of deterrence or backlash from counterterrorist policies. These include the Guernica Statute (October 15, 1979), the Ajuria-Enea Pact (January 12, 1988), and the meetings of Algiers (January 8, 1989), Zurich (November 3, 1998), and Oslo (June 25, 2005). These controls are all measured using dummy variables. A key part of the modeling is that Argomaniz pools all attack modes together meaning assassinations, bombings, and hostage events are all included in the same series.

Argomaniz and Vidal-Diez (2015) find that the cumulative month and density control explain most of the variation in the data. The only variable with a significant coefficient in the model is the French deportation and extradition dummy which has a positive sign implying a backlash effect (more frquent attacks). Relative to the rest of the sample, the harsher approach towards ETA increased the hazard of a terrorist attack between 8.4 and 48.8%. This implies the time elapsed between attacks (peacetime duration) decreased as a result of the policy. Argomaniz and Vidal-Diez note there is a problem of collinearity with the GAL operations. The two interventions are nearly contemporaneous and each coincide with an in increase in terrorist attacks. Unlike previous analysis, the Bidart arrests of 1992 are not found to be a significant deterrent. This may be due to a known limitation of the series hazard model. When a few number of attacks take place during a given period then the effect of a policy is measured at fewer intervals leading to lower statistical variation and largers standard errors.

Chapter 3: Sequential Model of Terrorist Attacks with Non-Proportional Hazard

3.1 Introduction

This chapter proposes a model of sequential decision-making as an alternative to the series hazard model approach when evaluating the effectiveness of counterterrorism policies (LaFree et al. (2009) and Dugan (2011)). A limitation of the series hazard model for risk analysis is that it predicts the incidence of terrorist attacks, but it does not predict attack outcomes even though these are recorded at the time of the events. Additionally, researchers have called for models of terrorist campaigns where attacks form part of a sequence, (e.g., Sandler and Enders (2004), pg 315), but to the best of my knowledge no such model has been suggested to date. This dissertation formulates a dynamic discrete choice model of terrorist decision-making that links an agent's choice of when to attack and the expected attack outcome by taking the time elapsed in between attacks as an input into the production of the next terrorist attack.

As covered in Section 2.4.1, the series hazard model is based on the Cox proportional hazard model: a widely-applicable, semi-parametric, continuous-time method that assumes changes to a policy shift the hazard proportionally to a baseline hazard, irrespective of the shape of the baseline hazard. In contrast, the *sequence* hazard model presented below is a fully parametric model designed specifically to study a terrorist group's choice of when to attack. It makes strong assumptions about the behavior of an agent and relevant probability distributions in order to estimate the impact of a counterterrorist policy on the risk of future attacks as well as their expected size. The sequence hazard model is a simple model of terrorism that, starting from basic fundamentals, does not lead to a proportional hazard model. Rather, in the sequence hazard a policy can have a differential impact over time, and the model does not admit time proportionality as a special case. The implications of the model and the Monte Carlo simulation presented in Chapter 4 raise potential concerns about the use of a proportional hazard in empirical work evaluating the efficacy of counterterrorist policies.

In the series hazard model a counterterrorist policy is categorized as having a deterrent (backlash) effect if it significantly reduces (increases) the hazard of an attack, without taking into account its effect on the outcome of attacks. The sequence hazard model extends the notion of deterrence (backlash) to consider not only the risk of attacks, but also their expected outcomes. Under this formulation, a policy may reduce the frequency of attacks, but also increase the size or probability of success of attacks that do take place.

Understanding the frequency and outcome of terrorist attacks is crucial to evaluating the impact of counterterrorism policies. Terrorists use violence to advertise their cause and intimidate opponents. Larger, more spectacular attacks gain more media attention and have a stronger coercive effect on the target audience. The threat of a severe terrorist attack always looms over campaigns of terror consisting of relatively more frequent, less severe events, such as ETA's in Spain. The literature reviewed in Section 2.2 illustrated how both severe and less severe events can have long-lasting economic effects so it is important to understand how terrorist groups substitute between them and space out their attacks over time.

Empirical analyses of terrorist attacks suggest there is a fundamental trade-off between the size of an attack and the preparation (and thus time) needed to carry it out. Clauset et al. (2007) showed the empirical distribution of the frequency and severity of terrorist attacks worldwide since 1968 are characterized by a power-law distribution. Power laws have been fit to a number of social phenomena and often indicate that some underlying optimization is taking place (Richardson (1948)). The scale-invariance property of a power law implies that attacks with severe outcomes are not fundamentally different from less severe ones, rather they are a scaled up versions of the same. Clauset and Gleditsch (2012) ruled out factors such as network size, experience, geographic location, and ideology as explanations for the severity of attacks, positing instead that event planning time and severity are strongly related through a competition between state actor and terrorist.

The sequence hazard assumes the severity of a terrorist attack is constrained by the need to resource and plan for the attack while being opposed by a state actor. The trade-off between size and frequency of attacks is the result of the decision of an agent maximizing utility over a sequence of of periods over which they can carry out a terrorist attack. The model is dynamic because the agent takes into account the effect of their choice on future rewards from terrorism. The model assumes that terrorists are rewarded by large or successful attacks, but these are costly to plan and the size or probability of success of an attack is influenced by the amount of planning. Moreover, terrorist groups do not prepare their attacks in a vacuum. Attack preparations are subject to sabotage by a law enforcement that seeks to hinder the agent's plans. The cost of attack innovation and unobserved shock to planning capital by law enforcement are both interpreted as state variables assumed to be unobservable to the researcher. Following the standard approach, these unobserved states are integrated out of the agent's choice problem to arrive at an estimator for the probability of attacking.

The sequence hazard takes a theory of the firm perspective of the production of terrorist attacks. Terrorist networks are assumed to act as a single optimizer and any concerns about the size, structure, or information sharing within the network are assumed away. The sequence hazard is a single agent model as there is no strategic interaction between terrorist and a state actor, between terrorist groups, or between the terrorist and the intended target of the political violence. The state actor is merely represented by an iid shock.[1]

The sequence hazard model presented below is simple. Still, it suffices to show that modeling the decision process of terrorist networks to attack, taking into account the severity of the attack, does generally not lead to decision making process which is compatible with the assumption of a proportional hazard as postulated by the series hazard model. This is an important warning to the empirical literature that simply postulates the validity of the series hazard model. A sequence hazard model can also be beneficial in other ways. Using an explicit optimization models allows a researcher to use the full sample of the campaign of terror in estimation. In contrast, the series hazard method throws away observations on days when attacks do not take place. Moreover, for the sequence hazard model the risk of an attack is given by the Conditional Choice Probability (CCP) so the duration dependence can be expressed as derivatives of the CCP which under our parametric assumptions take a convenient functional form. Whether the risk of terrorism and the severity of attacks is increasing, decreasing, or constant in the days following an attack depends on the agent's stage of planning and expectation of future states. The parametric assumptions allow the duration dependence to be expressed in an economically meaningful way as a function of the primitives of the models. The expected change in attack severity can be similarly expressed as a function of the hazard of a shock to the agent's planning. A benefit the sequence hazard is that it allows a model of terrorist behavior to be estimated only using planning time, it does not require specialized information about the terrorist network. At the same time, the sequence

[1]A more complete model could allow for too severe of an attack to invite heightened scrutiny by a state or provoke the all out war that terrorists seek to avoid. A severe attack may also put off potential sympathizers to the terrorist's cause.

hazard also allows for the inclusion of other covariates or state variables if these are available to the researcher. For ETA there is limited but publicly available data on the kill or capture of ETA operatives. Government agencies that monitor terrorist groups of course have access to more detailed data that could be employed by the sequence hazard model.

The sequence hazard approach models attack outcomes as random variables. The model assumes that before deciding whether or not to attack, an agent knows the current state of their attacking plans and thus they form an expectation about likely outcome of an attack (and by implication the cost of further improving it should they choose to wait). The researcher observes a subset of those states of the world, the amount of time elapsed since the last attack, and the history of previous attacks, but does not observe the current shock to the terrorist's plans. The presence of this unobserved shock amplifies the range of outcomes of terrorist attacks that is observed in practice relative to its expected value implied by the agent's belief about the expected size of an attack. This last feature is crucial for estimating the model because it allows the mean and variance of the size of a terrorist attack by a particular terrorist group to be expressed in terms of the observable characteristics of the agents themselves and the areas in which they operate. I discuss the decision making process for two cases, corresponding to two alternative outcome measures terrorist attacks. One measure relates to casualties, etc., from a wide range of possible terrorist attacks, e.g., bombings, armed assault, etc. In this case outcome is modeled to follow a Poisson distribution. The other measure relates to success failure, and in this case outcome is modeled to follow a Bernoulli distribution. In the Poisson case I show that the selection on unobservable states implies the variance is greater than the mean casualties, otherwise known as *overdispersion*. Overdispersion is an important feature of the actual data on terrorist attacks.

Terrorist attacks are only as strong as the weakest link in the preparation, small mistakes or disruptions can drastically impact the outcome of an attack. The production of terrorist attacks is a special kind of process where disruption by a state actor can be fatal to the whole endeavor. One misstep can affect the entirety of the terrorist's plans. Terrorism is like other complex processes (e.g., soccer match outcomes or space shuttle flights - see Kremer (1993) and Anderson and Sally (2013)), where the failure of a single task can cause an acute drop in output. To represent this feature of production, the sequence hazard assumes that shocks to the agent's production of terrorism are multiplicative rather than additive.

Recall that if the agent chooses to wait they may increase the expected size/success of their attack, but they risk being sabotaged before they can carry out their attack. The tension of when to attack becomes particularly stark when the agent's plans are already well underway. For example, computers and communications recovered from the ISIS-affiliated terrorist cell in Brussels in 2016 contain explicit references to the need to attack soon before the plot and cache of unexploded vests was discovered (New York Times, March 2016). The La Rambla van attack in Barcelona in August of 2017 was precipitated by an accidental explosion that killed the ring leader and master bomb-maker before the intended attack could be carried out. In November of 2015 an Islamic State of Iraq and the Levant (ISIL) suicide bomber detonated his vest in the vicinity of Gate D of the Stade de France in Saint-Denis after being turned away at the entrance by event security. The blast killed one civilian, a fraction of the would-be casualties had the bomber been allowed past the gate as they intended or had he waited for a better occasion to use his vest.

Assuming that the decision of when to attack is done on a rational basis, terrorists will only attack when the costs and risk of further improving their attack outweighs the benefits to them in terms of a more impactful attack, perhaps released at a latter date. When the attack does take place, the terrorist cannot control the exact outcome, only the *expected* outcome, which they seek to maximize in every instance.

Overview of Model

A terrorism researcher has data on a single terrorist network

$$\text{Data} = \{a_t, x_t, y_t : t = 1, 2, \cdots, T\}$$

where a_t represents the terrorist's decision variable to attack $(a_t = 1)$, or not to attack, corresponding to $(a_t = 0)$, x_t contains states of the world relevant to terrorist decision-making, y_t measures the outcomes of attacks that take place, and T represents the periods over which we observe the terrorist group. The estimation approach assumes the agent's observed choices a_t are an optimal response to states of the world x_t and that the outcome of an attack is a random variable that depends on action a_t and states x_t.

To link y_t and x_t assume that terrorist attacks are the deliberate choice of an agent that aims to maximize the size or (probability of success) and thus the shock value of their act. The agent is a terrorist organization that carries out a campaign of terror for T periods. The terrorists derive a reward from the size/success of their attack, but larger or more successful attacks require planning. The agent does not know the outcome of the attack prior to carrying it out. The model assumes that the expected size of a terrorist attack is a function of the amount of time the agent spends waiting in a "planning" state. That is, the model assumes that waiting time or the time elapsed in between attacks is in input into the production of a large or successful attack. For simplicity the model assumes that the stock of planning time available is reset to zero each time an attack is carried out. In every period of the campaign the agent must decide whether or not to attack. If the agent decides wait $(a_t = 0)$ then they incur a random cost $\varepsilon_t(0) > 0$ which can be interpreted as the cost of spending one day innovating the attack plans. If instead the agent attacks $(a_t = 1)$ then they are rewarded by the random outcome of the attack y_t, where the expected value of y_t is affected by a shock bounded between 0 and 1. The agent holds beliefs about the expected size of an attack given the invested waiting time and given the size of the shocks. Agents do not attack in every period because attacks take time to plan.

The network does not consider each of the attacks to be an isolated event. Instead, attacks form part of a campaign of terror lasting $t = 1, \ldots, T$ periods for $T = \infty$. The campaign is described by four objects $\left\{\delta, u_a\left(y_t, z_t\right), \mathbb{E}\left[y_t \mid z_t, a_t\right], F_{z_{t+1} \mid z_t, a_t}\left(z_{t+1} \mid z_t, a_t\right)\right\}$ considered primitives in the sense that they are all assumed in advance. The function $u_a\left(y_t, z_t\right)$ represents the utility to a terrorist of staging a terrorist attack with outcome y_t and when the state of the world is described by z_t. The expectation $\mathbb{E}\left[y_t|z_t, a_t\right]$ represents the agent's belief about the expected outcome y_t given action a_t and state z_t. The agent does not know the exact outcome y_t in advance so they make a decision on the basis of this conditional expectation. The scalar $\delta \in (0, 1)$ is the factor by which agents discount future rewards, and $F_{z_{t+1} \mid z_t, a_t}\left(z_{t+1} \mid z_t, a_t\right)$ represents the agent's beliefs about the evolution of future states of the world z_{t+1} given past states of the world and past actions. The model is considered dynamic because agent's decision are linked over time via the belief about attack outcome $\mathbb{E}\left[y_t|z_t, a_t\right]$ and a Markov specification for the transition probabilities $F_{z_{t+1} \mid z_t, a_t}$. Imposing parametric and independence assumptions on the unobservable states ε_t, assuming functionals form for u_a and $\mathbb{E}\left[y_t \mid z_t, a_t\right]$, and specifying $F_{z_{t+1} \mid z_t, a_t}\left(z_{t+1} \mid z_t, a_t\right)$ imply a model that can be estimated from the data.

The outcome of a terrorist attack y_t is unknown *ex-ante* so the agent forms beliefs about y_t. If y_t is meant to denote the number of casualties resulting from an attack, then $y \in \mathbb{Z}_+$ (Poisson). If y_t represents whether an attack was a success or a failure then $y \in \{0, 1\}$ (Bernoulli). Both these types of outcomes are considered in this dissertation. An outcome of interest omitted by this formulation is when y_t is continuous, such as the total playtime that news outlets dedicate to covering the terrorist attack or the dollar amount in damages. Even so, in these cases the data are often discretized or discretization has to occur at some point for estimation.

The rest of this chapter is organized as follows. The next section reviews the economics literature on the estimation of structural models to provide context for the sequence hazard model. Rather than provide an exhaustive review of this large literature, the focus is on motivating the chosen approach of a full solution. In Section 3.2 below I first discuss, to fix ideas, a model for the terrorist's decision making in a static setting. Section 3.2.1 imposes parametric assumptions for the distribution of unobservable state variables and solves for the CCP in this setting. Section 3.2.2 introduces the data generating process for attack outcomes. I also derive the likelihood functions for the choice-based and outcome-based models. Section 3.3 generalizes the model to the dynamic case. It turns out that the dynamic CCP takes the exact same form as in the static case with the exception of an intercept parameter that depends on the agent's expected reward over future states.

3.2 Specification of a Static Discrete Choice Model of Terrorism

3.2.1 Model of Terrorism Risk

The easiest way to explain the model is by starting with the static version of the agent's problem. In the static or myopic model the agent considers each decision period in isolation while in the dynamic model the agent takes the outcome of future periods into account. The static model is obtained by setting $\delta = 0$. States of the world can be further partitioned as $z_t = [x_t, \varepsilon_t]$ where

x_t represents state variables relevant to terrorism decision-making that are observable to both the agent and the researcher (including the number of days elapsed since the previous attack), and $\varepsilon_t = (\varepsilon_t(0), \varepsilon_t(1))'$ represents choice-specific states of the world observable only to the agent. Here, $\varepsilon_t(0) > 0$ represents the cost of spending a day improving the terrorist attack and $\varepsilon_t(1) \in (0,1)$ represents an adverse shock to the agent's plans. Assume that $\varepsilon_t(0)$ is a random variable with expectation and variance both equal to 1 and assume that $\varepsilon_t(1)$ is a random variable with expected value μ and variance (or "precision") σ^2.

Definition 3.2.1 (Agent Preferences). Assume the agent's preferences are defined by

$$u_a(y_t, x_t, \varepsilon_t) = \begin{cases} y_t(x_t, \varepsilon_t(1)) & \text{if } a_t = 1 \\ \theta_0 - \theta_\varepsilon \varepsilon_t(0) & \text{if } a_t = 0 \end{cases} \tag{3.1}$$

where u_a represents the choice-specific utility known up to a set of parameters, $\theta_0 > 0$ is the subjective benefit of waiting which is assumed positive else a terrorist would never be observed to wait, $\varepsilon_t(0) > 0$ represents the cost of improving the attack, θ_ε scales the innovation cost so that it is proportional to the unit of measurement of y_t (and θ_0), x_t represents states of the world relevant to terrorist decision-making, y_t represents the random outcome of the attack, and $\varepsilon_t(1) \in (0,1)$ represents the shock to the agent's planning.

The model assumes the agent's preferences are linear in attack outcome y_t. Later I will impose a specific distribution for $\varepsilon_t(0)$ and a simple scaling condition for y_t in order to identify the model. Utility is time-separable meaning past decisions and attack outcomes do not influence current and future preferences over the size or success of a terrorist attack. The agent's behavior can still exhibit intertemporal substitution effects in response to changes in a state variables (e.g., a counterterrorist policy). The agent is risk neutral. Note that if θ_0 was a function of the data instead of a fixed parameter then for the model to be identified the function would have to depend on state variables not found in x_t. The empirical applications usually do not have enough data to distinguish state variables that affect expected size outcome from those that influence the subjective benefit to the agent of waiting (e.g., a prevailing wage rate). Thus, I assume θ_0 is a fixed scalar parameter.

Assume that attack outcome y_t is related to state variables x_t and shock $\varepsilon_t(1)$ as follows:

Assumption 3.2.1 (Conditional Expectation of Attack Outcome (CEY)).) Attack outcome y_t is a discrete random variable with expectation

$$\mathbb{E}\left[y_t \,|\, a_t = 1, x_t, \varepsilon_t\right] \; = q\left(x_t; \theta_q\right) \cdot \varepsilon_t\left(1\right) \tag{3.2}$$

where $x_t \in X$, $\varepsilon_t(1) \in (0,1)$. Here $q(\cdot)$ is a link function with $q : X \to \mathbb{R}_+$ if y_t is intended to model attack size and $q : X \to (0,1)$ if y_t is intended to model attack success, and θ_q is a vector of parameters of the link function $q(\cdot; \theta_q)$.

The fact that $\varepsilon_t(1) \in (0,1)$ implies that $q(x_t; \theta_q) \cdot \varepsilon_t(1) < q(x_t; \theta_q)$. The best the agent can expect is that $\varepsilon_t(1)$ is close to 1 so their attack plans are nearly intact. Attack plans cannot be carried out to perfection due to friction, for example, the effort of some state actor that seeks to thwart the terrorist plans. The worst outcome for the terrorist is that $\varepsilon_t(1)$ is instead close to 0 meaning the agent's plans in that period have been sabotaged. In this simple model, the agent's plans fully regenerate by the next period. This is unrealistic from the perspective of an empirical application, but imposing a simple "memoryless" structure strengthens the main point of the model that even the simplest low-level assumptions about terrorist behavior result in a hazard that is non-proportional.

Two candidates for the link function are (i) $q(x_t;\theta_q) = \exp(x_t'\theta_q)$ if modeling attack severity and (ii) $q(x;\theta_q) = \exp(x'\theta_q) / [1 + \exp(x_t'\theta_q)]$ if modeling attack success. Functional form (i) is helpful because it is the same used by the semi-parametric series hazard model. If the outcome variable y_t is binary zero-one, then the expected value of y_t is a probability, hence the choice of functional form (ii). Note that $\frac{\partial q(x_t;\theta_q)}{\partial x_t} = \theta_q \cdot q(x_t;\theta_q)$ follows the sign of the coefficient θ_q because it is assumed that $q(x_t;\theta_q) > 0$.

The assumption that $q(x_t;\theta_q) > 0 \; \forall \; x_t \in X$ implies that all attacks are expected to potentially cause harm or be successful, even those with no preparation. This assumption is validated, for example, by the activities of the Irish Republican Army that had the capacity to engage in opportunistic terror (e.g., the impromptu kidnapping of a police officer).

The following outline of the per-period timing of the model summarizes the elements of the static model. The setup reflects the fact that the outcome of an attack is a random variable from the perspective of an agent who is deciding in advance whether or not to carry out the attack.

Assumption 3.2.2 (Per-period Timing (Static)). The per-period timing is as follows

1. the agent observes state variables x_t and $\varepsilon_t = (\varepsilon(0), \varepsilon(1))'$
2. the agent chooses whether to attack $(a_t = 1)$ or wait $(a_t = 0)$
    - if the agent attacks, random outcome $y_t(x_t, \varepsilon_t(1))$ is realized
    - if the agent waits, no attack takes place and the agent derives utility $\theta_0 - \theta_\varepsilon \varepsilon_t(0)$.

Assuming the agent is a rational actor, they will attack if and only if the expected value of attacking is greater or equal to the known value from waiting. The decision to attack is determined by the following inequality:

$$a_t = 1 \iff \mathbb{E}_y \left[u_1 \left(y_t, x_t, \varepsilon_t \right) \mid a_t = 1, x_t, \varepsilon_t \right] \geq u_0 \left(y_t, x_t, \varepsilon_t \right)$$

$$\iff q\left(x_t; \theta_q\right) \cdot \varepsilon_t \left(1\right) \geq \theta_0 - \theta_\varepsilon \varepsilon_t \left(0\right)$$

$$\iff \varepsilon_t \left(1\right) \geq \frac{\theta_0 - \theta_\varepsilon \varepsilon_t \left(0\right)}{q\left(x_t; \theta_q\right)}. \tag{3.3}$$

Let Ω_t denote the state space for the agent consisting of all the elements of their decision problem at time t, $\Omega_t = \{x_t, \varepsilon_t\}$ where x_t describes states of the world observable to both the agent and the researcher and $\varepsilon_t = \left(\varepsilon_t \left(0\right), \varepsilon_t \left(1\right)\right)$' is the vector of choice-specific unobservable observed only by the agent. Specifically, $\varepsilon_t \left(0\right) > 0$ represents the cost to the agent of improving their attack and $\varepsilon_t \left(1\right) \in \left(0, 1\right)$ represents a shock that reduces the expected size or probability of success of a terrorist attack. Let $\Omega_t^- = \{x_t\}$ denote the researcher's information set. Let $S\left(\Omega_t^-\right) = \left\{\varepsilon_t \,|\, \varepsilon_t \left(1\right) \geq \frac{\theta_0 - \theta_\varepsilon \varepsilon_t(0)}{q(x_t;\theta_q)}\right)$ define the set of values of the unobservable state ε_t that induce an agent terrorist with state space Ω_t^- to choose $a_t = 1$. Let $F_{\varepsilon_t \,|\, x_t}$ denote the joint density of ε_t conditional on observable state x_t. The probability of attacking conditional on Ω_t^- is given by

$$\Pr\left(a_t = 1 \,|\, \Omega_t^-\right) = \Pr\left(\varepsilon_t \left(1\right) \geq \frac{\theta_0 - \theta_\varepsilon \varepsilon_t \left(0\right)}{q\left(x_t; \theta_q\right)}\right) \tag{3.4}$$

$$= \int_{S\left(\Omega_t^-\right)} \mathrm{d}F_{\varepsilon_t \,|\, x_t} \tag{3.5}$$

where $\Pr\left(a_t = 0 \,|\, \Omega_t^-\right) = 1 - \Pr\left(a_t = 1 \,|\, \Omega_t^-\right)$.

Equations (3.4) - (3.5) make it clear that the Conditional Choice Probability (CCP) is a function of state variable x_t and that the specific form of the CCP depends on i) the choice of utility function $u_a(y_t, x_t, \varepsilon_t)$, ii) the agent's belief about attack severity/success $\mathbb{E}[y_t \mid a_t, x_t, \varepsilon_t]$ (which is assumed separable in $\varepsilon_t(1)$, and iii) the distribution of unobserved state variables $F_{\varepsilon_t \mid x_t}$. These are elements to be estimated in the model. I assume that $\varepsilon_t(0)$ and $\varepsilon_t(1)$ are independently distributed.

The decision rule derived above (see Equation (3.3)) helps understand the need for a scaling condition on the function $q(\cdot)$ in order to identify the parameters of the model. If the parameters are all scaled by a factor c then the fraction on the right-hand side stays the same unless the following property holds:

Definition 3.2.2 (Scaling Condition). The function $q : X \to \mathbb{R}_{++}$ and has the property that $c \cdot q(x; \theta_q) \neq q(x; c \cdot \theta_q)$.

Next I assume a parametric functional form for the unobserved cost $\varepsilon_t(0)$ and the unobserved shock $\varepsilon_t(1)$.

Distribution of $\varepsilon_t(0)$.

Recall the value of waiting to the agent when the state of the world is Ω_t is given by $\theta_0 - \theta_\varepsilon \varepsilon_t(0)$. I will assume that $\varepsilon_t(0)$ is distributed exponential (1) so that in turn $\theta_\varepsilon \varepsilon_t(0)$ is distributed exponential (λ) with $\lambda = 1/\theta_\varepsilon$ implying that $\mathbb{E}[\theta_\varepsilon \varepsilon_t(0)] = \theta_\varepsilon$ and Var $[\theta_\varepsilon \varepsilon_t(0)] = \theta_\varepsilon^2$. In the sequel I will simply assume that the value to the agent of waiting is $\theta_0 - \varepsilon_t(0)$ and assume that $\varepsilon_t(0)$ is distributed exponential with parameter $\lambda = \frac{1}{\theta_\varepsilon} > 0$. The exponential distribution has the important property that it is memoryless so the truncated expectations take on a simple form which will be important for the dynamic model presented in the sequel. Denote the cdf of the exponential distribution by $F(\cdot;\lambda)$.

Distribution of $\varepsilon_t(1)$.

Assume that $\varepsilon_t(1)$ is distributed as a beta (α,β) random variable on $(0,1)$ where $\alpha>0$ and $\beta>0$ are shape parameters (Definition 6.1.5) in Appendix A). Denote the cdf of this distribution by $F(\cdot;\alpha,\beta)$ (see Definition 6.1.6 in Appendix A). The Beta distribution a flexible way of modeling a random variable bounded by 0 and 1. There is no other *a priori* motivation for this distribution, but it is worth noting that alternative approaches to modeling random variables defined on the unit interval such as the logit-normal, the truncated normal, or the raised cosine do not lead to convenient closed-form solutions for the CCP when combined with the exponential distribution of $\varepsilon_t(0)$ (or most other distributions). Moreover, the cdf of the beta distribution has an inverse that is straightforward to compute and widely available in statistical software packages which proves convenient in the next section.

Assumption BETAEXP for the sequence hazard model was motivated on the one hand by a model for the process of innovations and on the other by the convenience of the beta distribution to model shocks between 0 and 1. Larsen et al. (2012) test the validity of the extreme value type I assumption for the model of engine bus replacement (Rust (1987)) by estimating a more general model using the exponentiated generalized beta family (which contains the extreme value as a special case) and then applying a likelihood ratio test. The key impediment to using more flexible distributions are the fact that the truncated expectation - conditioning on the event $a_t=1$ - do not admit closed-form solutions even though their cdfs have simple closed-form expressions. In contrast, the exponential-beta model formulation for the sequence hazard, like the logit model, has simple expressions for the truncated expectations.[2]

[2]Larsen et al. (2012) estimate the model including the truncated expectations using Gauss-Laguerre quadrature.

As demonstrated by Equation (3.3), identification of the sequence hazard model depends on restrictions to both the link function representing the agent's belief about the expected size of an attack and the scaling of $\varepsilon_t(0)$ through θ_ε. Equation (3.3) also shows that in order to estimate the model it is necessary to normalize the variance of the beta distribution. This is akin to setting the variance in a probit model equal to 1 owing to the fact that the errors can scale up without changing the decision rule.[3] The following re-parametrization recasts the beta (α, β) beta random variable $\varepsilon_t(1)$ as a random variable with expected value μ and variance σ^2 so that σ^2 (and thus the scale of $\varepsilon_t(1)$) can be fixed.

Definition 3.2.3 (Re-parametrization of the beta distribution). Re-parameterize the beta distribution from Definition 6.1.5 (see Ferrari and Cribari-Neto (2004)) setting $\mu = \alpha/(\alpha+\beta)$ and $\nu = \alpha+\beta$ so that $\alpha = \mu\nu$ and $\beta = (1-\mu)\nu$ from which it follows that

$$\mathbb{E}[x] = \mu$$

and

$$\text{Var}[x] = \frac{\text{V}(\mu)}{1+\nu}$$

where $\text{V}(\mu) = \mu(1-\mu)$. The idea is to fix the variance $\text{Var}[x] = \sigma^2$ and represent α and β as a

Doing so is beyond the scope of this dissertation, but the idea can be applied as well to the sequence hazard model. Specifically, the exponential-beta could be estimated via Gauss-Laguerre quadrature and compared to the full or iterated solution to show the validity of the numerical approach. Then the quadrature approach can be used to estimate the more flexible versions of the generalized beta distribution of the second kind taking care to select flexible distributions bounded between 0 and 1 as needed. Then a likelihood ratio test can be applied to compare the simple exponential-beta against the more flexible distributions that are disconnected from a model of terrorist behavior.

[3] Similarly, in the context of a logit dynamic discrete choice model, Rust (1987) sets the variances of the EV1 distribution to a constant and then uses data on actual replacement costs to identify the scale.

function of μ and ν:

$$\nu = \frac{\mu (1-\mu)}{\sigma^2} - 1,$$
$$\alpha = \mu\nu = \mu \left[\frac{\mu (1-\mu)}{\sigma^2} - 1\right],$$
$$\beta = \nu - \alpha = (1-\mu)\,\nu = (1-\mu) \cdot \left[\frac{\mu (1-\mu)}{\sigma^2} - 1\right].$$

Note that for the beta distribution $\sigma^2 < 1$. When $\alpha = \beta = 2$, then $\sigma^2 = \frac{1}{20}$. Thus when estimating the model, fix σ^2 and estimate μ, then use the estimated μ to recover the implied α and β. Figure 3.2.1 shows the density of a re-parametrized beta random variable for different choices of μ with variance fixed at $\sigma^2 = \frac{1}{20}$, along with the implied α and β.

The following are definitions used to characterize the conditional choice probability for the agent's problem.

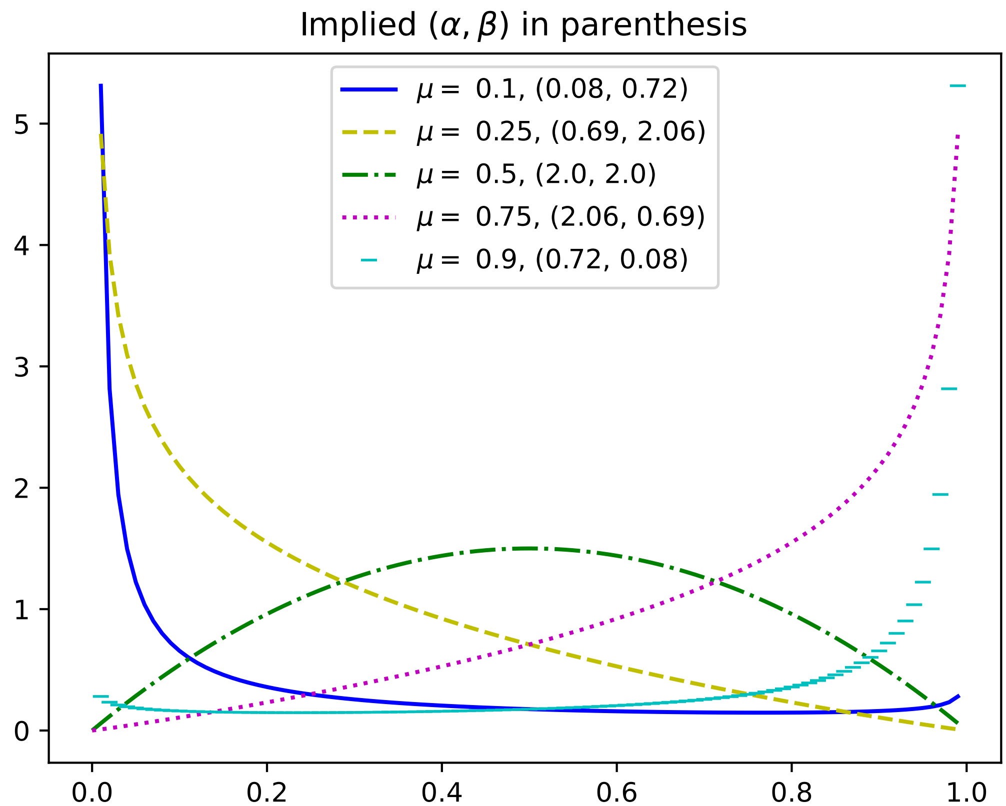
Definition 3.2.4 (Confluent hypergeometric function (Kummer)). Kummer's function is a generalized hypergeometric series

$$\mathcal{M}(a, b, z) = \sum_{n=0}^{\infty} \frac{(a)_n\, z^n}{(b)_n\, n!}$$

where $(a)_n$ represents the rising factorial with $(a)_n = a(a+1)(a+2)\ldots(a+n-1)$ and $(a)_0 = 1$. See 13.1.2 on page 504 of Abramowitz et al. (1964). Use $\mathcal{M}_{a,b}(z)$ as a shorthand for this function. Kummer's function has the property that $\mathcal{M}(a, b, z) = \exp(z)\,\mathcal{M}(b-a, b, -z)$.

Definition 3.2.5 (Confluent Hypergeometric function of two variables (Humbert type 1)). Humbert's type 1 function is a generalized hypergeometric series in two variables is defined for $|\gamma| < 1$

Figure 3.1: Examples of the Beta pdf for Different Values of μ with $\sigma^2 = 1/20$.



by

$$\Phi_1 \left(a, b, c; \gamma, z\right) = \sum_{m,n=0}^{\infty} \frac{(a)_{m+n} (b)_m}{(c)_{m+n} m!n!} \gamma^m z^n.$$

where $(a)_n$ represents the rising factorial with $(a)_n = a\,(a+1)\,(a+2)\ldots(a+n-1)$ and $(a)_0 = 1$. This function is also known as known as a Humbert type 1 function. See Equation (20) on page 225 of Erdélyi (1953). See Humbert (1922) for the original statement.

Definition 3.2.6 (Regularized Type 1 Humbert Series). For $|x| < 1$, $u > 0$, and $v > 0$ define

$$\mathcal{H}_{u,v}(\mu,\gamma) = \frac{1}{B(u,v)} \cdot \frac{\mu^u}{u} \cdot \Phi_1(u, 1-v, u+1; \mu, \gamma)$$

where Beta (u, v) is the Beta function (Definition 6.1.2) and $\Phi_1(\cdot)$ is the Type 1 Humbert function (Definition 3.2.5).

Result 3.2.1 (Conditional Choice Probability (static)). The independence assumption together with the exponential-beta distribution of the choice-specific elements in ε_t imply the CCP defined by equation (3.5) is given by

$$\mathrm{P}\left(a_t = 1 \,|\, \Omega^-\right) = \begin{cases} 1 - F(\theta_0/q_t; \alpha, \beta) + \lambda^{-1} \cdot f(\theta_0; \lambda) \cdot \mathcal{H}_{\alpha,\beta}(\theta_0/q_t, \lambda\theta_0) & \text{if } \theta_0/q_t \in (0,1) \\ \lambda^{-1} \cdot f(\theta_0; \lambda) \cdot \mathcal{M}_{\alpha,\alpha+\beta}(\lambda q_t) & \text{if } \theta_0/q_t \geq 1 \end{cases} \tag{3.6}$$

where $q_t \equiv q(x_t; \theta_q)$, $\mathcal{H}_{a,b}(\cdot,\cdot)$ is the regularized Humbert type 1 function (Definition 3.2.4), $\mathcal{M}_{a,b}()$ is Kummer's confluent hypergeometric function (Definition 3.2.5), $F(\cdot; a, b)$ denotes the cdf of the beta distribution representing the security shocks, and $f(\cdot; \lambda)$ is the pdf of the exponential distribution representing the innovation costs.

Proof. See Appendix Result 6.3.8. □

The Humbert and Kummer functions have appeared previously in the work of Gordy (1998a) who derived equilibrium bidding strategies for common value auctions (see also Gordy (1998b)). As shown in Appendix C Result 6.4.5, the Humbert Type 1 function can be used to represent the survivor function of a beta distributed random variable: $\overline{F}(u; \alpha, \beta) = \mathcal{H}_{\beta,\alpha}(1-u, 0)$ where

$\overline{F}(u; \alpha, \beta) \equiv 1 - F(u; \alpha, \beta)$. Kummer's function is equivalent to the moment generating function of a beta $(a, b-a)$ random variable (see Result 6.3.5 in Appendix C).

The CPP in Equation 3.2.1) is non-proportional in time meaning it cannot be written as a product of two functions, one of which depends exclusively on the time spent in the waiting state. The CCP depends on state variables x_t via the link function $q_t(\cdot)$. The agent has a positive probability of attacking even if the certain return from waiting outweighs the expected size of the attack (e.g., $\theta_0 \geq q_t(x_t; \theta_q)$). In other words, the CCP is saturated (has full support) which means a wide number of solutions are available even if the distributions of unobservable states is not assumed logit or additively-separable.

A case of particular interest for the CCP is when the observable state variable x_t includes the number days elapsed since the last terrorist attack, taken to be the number of days the agent spent in a waiting state preparing their attack. In this case the CCP as a function of the peacetime durations is equivalent to a hazard.

Result 3.2.2 (The CCP is a proper hazard). The CCP defined in Result 3.2.1 is a proper hazard when viewed as a function of time elapsed since the last attack.

Proof. Suppose a terrorist group just launched an attack. Let $M \in \{0, 1, 2, \cdots\}$ represent the number of days elapsed until the next terrorist attack. M is a discrete random variable with cdf

$$\begin{aligned}\Pr(M \leq m-1) &= 1 - \Pr(M > m-1) \\ &= 1 - \prod_{s=0}^{m-1} \Pr(a_s = 0 \mid s)\end{aligned}$$

and probability mass function

$$\text{Pr}(M = m) = \prod_{s=0}^{m-1} \text{Pr}(a_s = 0 \,|\, s) \times \text{Pr}(a_m = 1 \,|\, m).$$

The resulting hazard function for M is

$$\begin{aligned}
\text{Pr}(M = m \,|\, M \geq m) &= \frac{\text{Pr}(M = m)}{\text{Pr}(M \geq m)} \\
&= \frac{\text{Pr}(M = m)}{1 - \text{Pr}(M \leq m - 1)} \\
&= \text{Pr}\left(a_m = 1 \,|\, m\right).
\end{aligned}$$

$\square$

The above result implies that the sequence and series hazard models are comparable and that the condition for duration dependence for the sequence hazard can be expressed in terms of derivatives of the CCP with respect to the waiting time state variable. Let P $(a_t = 1 \,|\, x_t)$ represent the CCP where $\theta = (\theta_0, \theta_q, \alpha, \beta, \lambda)'$ denotes the parameters of the model where, as a reminder, θ_0 is a utility function parameter representing the subjective net benefit to the agent of waiting, θ_q represents the parameters of the link function $q\,(x_t; \theta_q) > 0$ representing the agent's belief about the expected size of an attack, where α and β are the parameters of a beta distribution for the unobserved shock $\varepsilon_t\,(1) \in (0, 1)$ and λ is the parameter of the exponential distribution for the unobserved cost $\varepsilon_t\,(0) > 0$.

The likelihood for the CCP-based model takes the form of a straightforward Bernoulli. The data on attack outcome is not needed to estimate the parameters of the model.

Result 3.2.3 (CCP-based likelihood). The independence assumption implies that for a sample of data on terrorist attacks staged by a given group $\{x_t, a_t\}_{t=0}^{t=T}$ the likelihood function is given by

$$\mathcal{L}(\theta \mid \text{Data}) = \prod_{t=1}^{T} [\text{Pr}(a_t = 0 \mid x_t; \theta)]^{1(a_t=0)} [\text{Pr}(a_t = 1 \mid x_t; \theta)]^{1(a_t=1)} \tag{3.7}$$

where Pr $(a_t = 1 \mid x_t)$ is the CCP defined in Equation (3.2.1).

Proof. The result follows from independence of the unobserved shocks across time. □

In the simple case presented later, x_t is completely deterministic because $x_t = \{w_t\}$ where $w_{t+1} = (1 - a_t) \cdot (w_t + 1)$ where $w_0 > 0$ is generally not known by the researcher. In the case of ETA the founding date of the group is known and could be thought of as a starting point, but in general the model can be estimated past the date of the first attack. The following result proved in the next section simply states that in the static model the risk of a terrorist attack is always increasing following an attack.

Result 3.2.4 (Duration dependence). In the static model the risk of a terrorist attack is always increasing over time.

Proof. See Result 6.3.9 in Appendix C. □

3.2.2 Joint Model of Terrorism Risk and Severity

This section describes how to derive the main elements of the empirical model that makes use of attack outcome data. The existing literature uses only information on the timing of terrorist events, not on the magnitude of events even though these are recorded at the time of the attack. In this section I derive a framework that allows one to use terrorist outcome data to estimating the model.

The idea is that incorporating outcome data can lead to a richer description of terrorist behavior and potentially more accurate predictions of future terrorist activity.

They key step is deriving the expected value of the states that are not observed by the researcher, $\varepsilon_t = (\varepsilon_t(0), \varepsilon_t(1))'$. When forming the expectation the researcher conditions on the fact that the agent attacked ($a_t = 1$) or decided not to attack ($a_t = 0$). The decision rule derived in Equation (3.3) shows how the agent's choice has implications for the random states in ε_t. This is particularly important for $\varepsilon_t(1)$ because it directly impacts y_t, the observed size of the attack. Assumption CEY (see 3.2.1) regarding the expectation of $\varepsilon_t(1)$ is now strengthened to specify its probability distribution. In the case where outcome y_t measures casualties, the decision rule involves a mixture of a Poisson and beta distribution, while for the case where outcome y_t represents the decision to or not to attack, the decision rule involves a mixture of a Bernoulli and beta distribution. Appendix B provides a short primer on the pmf and first and second moments Poisson-beta and Bernoulli-beta mixture distributions.

From the perspective of the agent $\mathbb{E}[y_t \mid a_t = 1, x_t, \varepsilon_t(1)] = q_t \cdot \mathbb{E}[\varepsilon_t(1) \mid a_t = 1, x_t]$. The researcher does not observe the shock $\varepsilon_t(1)$ so

$$\mathbb{E}[\varepsilon_t(1) \mid a_t = 1, x_t] = \frac{\overline{F}(\bar{\varepsilon}_t; \alpha + 1, \beta)}{\mathrm{P}(a_t = 1 \mid x_t)} \cdot \mathbb{E}[\varepsilon_t(1)]$$

where $\bar{\varepsilon}_t = \mathrm{F}^{-1}(1 - \mathrm{P}(a_t = 1 \mid x_t); \alpha, \beta)$, where $F(\cdot; \alpha, \beta)$ is the cdf of the beta distribution (Definition 6.1.6), and $\overline{F}$ denotes the survival function.

Define $M_j(\gamma) \equiv \mathbb{E}\left[\varepsilon_t(1)^j\right] \cdot \overline{F}(\gamma; \alpha + j, \beta)$ to write the moment of $\varepsilon_t(1)$ conditional on $a_t = 1$ more succinctly as $\mathbb{E}[\varepsilon_t(1) \mid a_t = 1, x_t] = M_1(\bar{\varepsilon}_t) / \mathrm{P}(a_t = 1 \mid x_t)$.

The following assumptions strengthen Assumption CEY (3.2.1) because they impose a parametric distribution for the random outcome y_t. The benefit of doing so is that it allows one to model the joint density of the frequency and severity of terrorist attacks and leverage the information in attack outcomes instead of discarding the data.

Assumption 3.2.3 ((Severity) Attack Outcome is a Poisson Random Variable (CEY-POISS)). Attack outcome $y_t \in \mathbb{Z}_+$ is a Poisson distributed random variable with parameter $q\left(x_t;\theta_q\right) \cdot \varepsilon_t\left(1\right) > 0$ (see Definition 6.1.7 in the Appendix).

Note that Poisson (severity) model is equidispersed

$$\mathbb{E}\left[y_t \,|\, a_t = 1, x_t, \varepsilon_t\left(1\right)\right] = \text{Var}\left[y_t \,|\, a_t = 1, x_t, \varepsilon_t\left(1\right)\right] = q\left(x_t;\theta_q\right) \cdot \varepsilon_t\left(1\right).$$

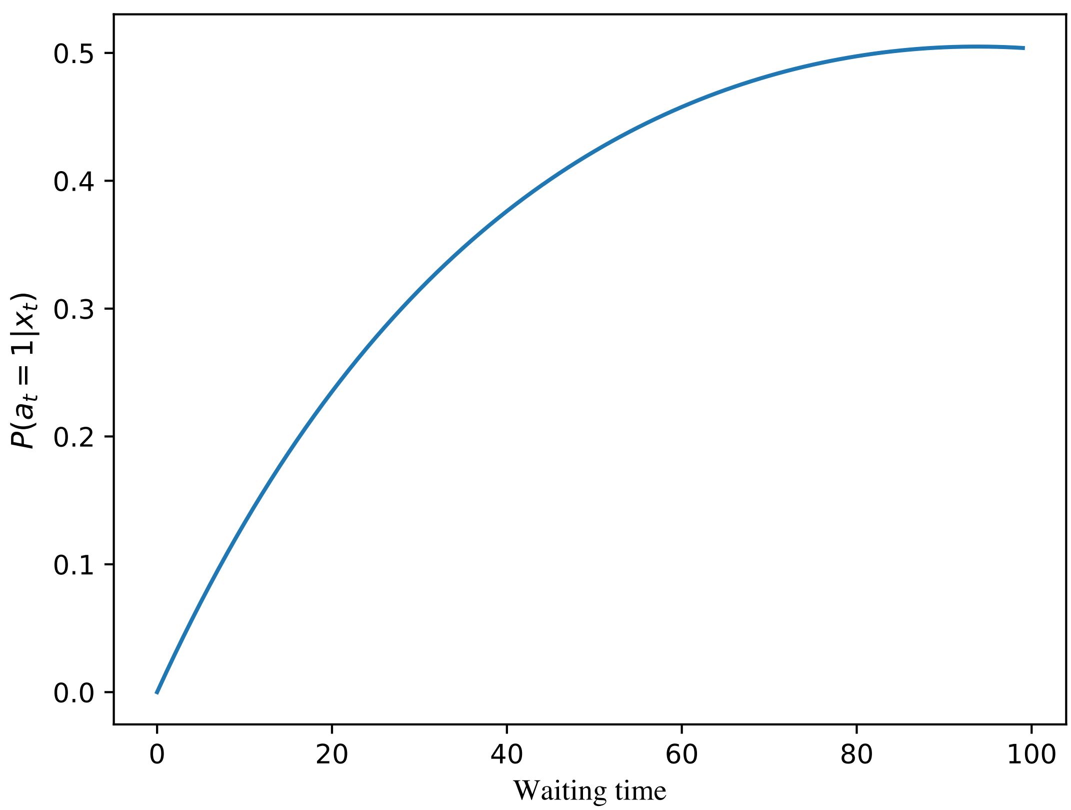
Assumption 3.2.4 ((Success) Attack Outcome is a Bernoulli Random Variable (CEY-BERN)). Attack outcome y_t is a Bernoulli distributed random variable that takes on the value $y_t = 1$ with probability $q\left(x_t;\theta_q\right) \cdot \varepsilon_t\left(1\right) \in (0,1)$ and $y_t = 0$ with probability $1 - q\left(x_t;\theta_q\right) \cdot \varepsilon_t\left(1\right)$.

Note that for the Bernoulli (success) model

$$\begin{aligned}
\mathbb{E}\left[y_t \,|\, a_t = 1, x_t, \varepsilon_t\left(1\right)\right] &= q\left(x_t;\theta_q\right) \cdot \varepsilon_t\left(1\right) \\
\text{Var}\left[y_t \,|\, a_t = 1, x_t, \varepsilon_t\left(1\right)\right] &= q\left(x_t;\theta_q\right) \cdot \varepsilon_t\left(1\right) \cdot \left[1 - q\left(x_t;\theta_q\right) \cdot \varepsilon_t\left(1\right)\right].
\end{aligned}$$

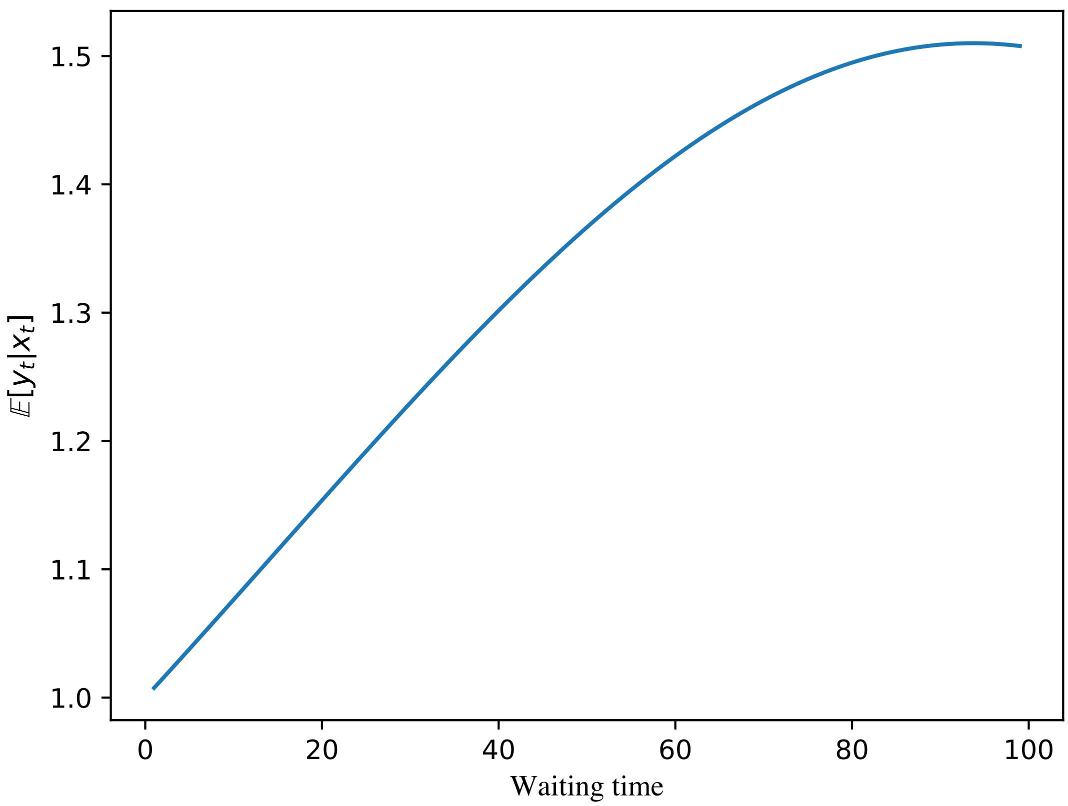
The assumption assumption of a discrete outcome variable y_t applies a number of state variables of interest as continuous state spaces are typically discretized in applied work. For terrorism, many of the state variables of interest are naturally discrete (e.g., attack casualties or attack success or failure). The x_t can contain waiting time which are discrete and can be assumed to be finite, bounded by the largest waiting time found in the sample. The x_t can also contain a binary variable that takes on values 0 or 1 depending on whether a counterterrorism policy is in place. This is the most common representation of a counterterrorist policy in empirical work (e.g., see Barros et al. (2006) and Argomaniz and Vidal-Diez (2015)). Lastly, x_t could also be a bounded integer representing the number of operatives killed or captures, or other measures relevant to terrorist decision-making. Figures (3.2.2) and (3.2.2) illustrate the CCP and joint likelihood functions (defined in Result 3.2.1 and Result 3.2.5, respectively) with $\alpha = \beta = \lambda = 1$, $\theta_0 = 1, \theta_1 = 0.015$, and $\theta_2 = -8 \times 10^{-5}$.

Figure 3.2: Plot of the Conditional Choice Probability of Attacking.



Result 3.2.5 (Joint Choice and Outcome Probability (JCOP)). If the independent exponential-beta assumptions from the CCP hold and in addition Assumption CEY-POISS holds (Definition 3.2.3)

Figure 3.3: Plot of the Expected Attack Size.



so that the outcome of an attack is a Poisson distributed random variable (see Definition 6.1.7 in Appendix A) then the joint probability of the outcome y_t and the choice to attack $a_t = 1$ is given by

$$\mathrm{P}\left(Y_t = y_t, a_t = 1 \,|\, x_t\right) = \mathrm{P}_Y\left(y_t \,|\, q_t\right) \cdot \mathbb{E}\left[\varepsilon_t\left(1\right)^{y_t}\right] \cdot \mathcal{H}_{\beta,\alpha+y_t}\left(\bar{\bar{\varepsilon}}_t, q_t\bar{\bar{\varepsilon}}_t\right). \tag{3.8}$$

where $q_t \equiv q\left(x_t; \theta_q\right)$, $\mathrm{P}_Y\left(\cdot \,|\, q_t\right)$ denotes the pmf of a Poisson random variable Y_t with parameter q_t (see Definition 6.1.7 in Appendix A), recall that $\varepsilon_t\left(1\right)$ is a beta(α, β) distributed random variable so that $\mathbb{E}\left[\varepsilon_t\left(1\right)^{y_t}\right]$ is the y-th moment and equal to B $\left(\alpha + y, \beta\right)/$B $\left(\alpha, \beta\right)$ (see Definition 6.1.3 in

Appendix A) where $\mathrm{B}(\cdot,\cdot)$) and is the Beta function (Definition 6.1.2), $\mathcal{H}_{a,b}(\cdot,\cdot)$ is the Regularized Type 1 Humbert Series (Definition 3.2.5), $\bar{\bar{\varepsilon}} \equiv 1-\bar{\varepsilon}$ where $\bar{\varepsilon}_t = F^{-1}\left(\mathrm{P}\left(a_t=0\,|\,\Omega_t^-\right);\alpha,\beta\right) = F^{-1}\left(1-\mathrm{P}\left(a_t=1\,|\,x_t\right);\alpha,\beta\right)$ so that $\bar{\bar{\varepsilon}}$ is like an "inverse" survivor function, and $F^{-1}(\cdot;\alpha,\beta)$ is the inverse cdf of a beta (α,β) random variable. Note that if $\mathrm{P}(a_t=1\,|\,x_t)=1$ (e.g., agent attacks with certainty) so that $\bar{\varepsilon}_t=0$ and $\bar{\bar{\varepsilon}}_t=1$ then the joint probability in Equation (3.8) takes the form of a pure Poisson-beta mixture (see Equation 6.1 in Appendix B). When $\mathrm{P}(a_t=1\,|\,x_t)=1$ with certainty, then the agent's action $a_t=1$ contains no information about the realization of $\varepsilon_t(1)$ so the underlying process for outcome y_t is a pure Poisson-beta mixture.

If instead the outcome variable is modeled as a Bernoulli random variable representing attack success or failure so Assumption CEY-BERN (see Definition 3.2.4 above) holds, then

$$\mathrm{P}\left(Y=y_t,a_t=1\,|\,x_t\right) = \left[q_t\cdot M_1\left(\bar{\varepsilon}_t\right)\right]^{\mathbf{1}(y=1)}\cdot\left[\mathrm{P}\left(a_t=1\,|\,x_t\right)-q_t\cdot M_1\left(\bar{\varepsilon}_t\right)\right]^{\mathbf{1}(y=0)}$$
$$\text{for } y\in\{0,1\}$$

where recall that

$$M_j(\gamma) = \mathbb{E}\left[\varepsilon_t(1)^j\right]\cdot\overline{F}(\gamma;\alpha,\beta),$$
$$\mathbb{E}\left[\varepsilon_t(1)\right] = \alpha/(\alpha+\beta),$$

and $\bar{\varepsilon}_t = F^{-1}\left(1-\mathrm{P}\left(a_t=1\,|\,x_t\right);\alpha,\beta\right)$ where $F^{-1}(\cdot;\alpha,\beta)$ is the inverse of the beta cdf (Definition 6.1.6). In this case assume $q_t\in(0,1)$.

The first and second moments implied for $Y_t\,|\,x_t$ when the underlying process for Y_t is

Poisson-beta are

$$\mathbb{E}\left[Y_t \,|\, x_t\right] = q_t \cdot M_1\left(\bar{\varepsilon}_t\right) \tag{3.9}$$

$$\text{Var}\left[Y_t \,|\, x_t\right] = \mathbb{E}\left[Y_t \,|\, x_t\right] + q_t^2 \cdot \left(M_2\left(\bar{\varepsilon}_t\right) - M_1\left(\bar{\varepsilon}_t\right)^2\right) \tag{3.10}$$

For the model of attack success where assumption CEY-BERN holds (Definition 3.2.4) so that the outcome of an attack is a Bernoulli distributed random variables then the mathematical expectation is the same as Equation (3.9) only the link function is now restricted to the $(0, 1)$ interval and the variance is given by

$$\text{Var}\left[Y_t \,|\, x_t\right] = \mathbb{E}\left[Y_t \,|\, x_t\right] \cdot \left(1 - \mathbb{E}\left[Y_t \,|\, x_t\right]\right). \tag{3.11}$$

The *intensity* or expected size (or success) of a terrorist attack is increasing/constant/decreasing as a result of a change in scalar state variable x_t whenever

$$\frac{\mathrm{d}}{\mathrm{d}x_i} q\left(x_t; \theta_q\right) \lesseqgtr -\frac{\mathrm{d}}{\mathrm{d}x_i} \log \overline{F}\left(\bar{\varepsilon}_t; \alpha, \beta\right). \tag{3.12}$$

Note that the expression on the right-hand side is the hazard or risk of receiving a shock $\varepsilon_t\left(1\right)$ as bad as one that would induce the agent to wait instead of attacking.

Proof. See the Appendix Results 6.4.12, 6.4.13, and 6.4.14. $\square$

The first and second moments of $Y_t \,|\, x_t$ are functions of x_t via the expected attack size function function $q\left(x_t; \theta_q\right)$ and the quantity $\bar{\varepsilon}_t\left(x_t\right)$ which represents the smallest value of the adverse shock $\varepsilon_t\left(1\right)$ that induces the agent to attack. The expected value of the outcome (3.9) includes the term $\overline{F}\left(\bar{\varepsilon}_t\left(1\right);\alpha,\beta\right) \equiv 1 - F\left(\bar{\varepsilon}_t\left(1\right);\alpha,\beta\right)$ which is the probability of experiencing a shock at least as bad as $\bar{\varepsilon}_t$. If there is no uncertainty about the attack taking place then $\mathrm{P}\left(a_t = 1 \,|\, x_t\right) = 1$ and $\bar{\varepsilon}_t = 0$ so $\overline{F}\left(\bar{\varepsilon}_t\left(1\right);\alpha,\beta\right) = 1$ and the data generating process for $Y_t \,|\, x_t$ is a pure Poisson-beta or Bernoulli-beta process (see Appendix D Results 6.4.12 and 6.4.13, respectively).

The expected intensity of an attack (size or probability of success) increases as a result of a change in state variable x_t whenever the effect of the change on the expected attack size function $q\left(x_t;\theta_q\right)$ is greater than the effect it has on the signal $\varepsilon_t\left(1\right)$ that induces the agent to attack. The left hand side of equation (3.12) is the size of the effect of the change of x_t on the belief the agent has about the size/success of an attack. The right hand side the change in the minimum shock that induces the agent to attack times the inverse Mill's Ratio or hazard (probability) of experiencing a shock of that magnitude given that the agent is attacking.

Result 3.2.6 (JCOP-based likelihood). The independence assumption implies that for a sample of data on terrorist attacks staged by a given group $\{x_t, a_t, y_t\}_{t=0}^{t=T}$ the likelihood function is given by

$$\mathcal{L}\left(\theta \,|\, \text{Data}\right) = \prod_{t=1}^{T} \left[\Pr\left(a_t = 0 \,|\, x_t\right)\right]^{1(a_t=0)} \times \left[\Pr\left(Y_t = y_t, a_t = 1 \,|\, x_t;\theta\right)\right]^{1(a_t=1)} \tag{3.13}$$

In the simple waiting time model where x_t is a non-negative integer then $x_{t+1} = (1 - a_t)\cdot(1 + x_t)$. For example, if y_t is a mixed Poisson-beta process then the contribution to the log-likelihood of a single day of data is

$$(1 - a_t)\cdot\log\left(1 - \mathrm{P}\left(a_t = 1 \,|\, x_t\right)\right) + a_t \cdot \left\{ \log \mathrm{P}_Y\left(y_t \,|\, q_t\right) + \log \mathbb{E}\left[\varepsilon_t\left(1\right)^{y_t}\right] + \log \mathcal{H}_{\beta,\alpha+y_t}\left(1 - \bar{\varepsilon}_t, q_t \cdot \left(\bar{\varepsilon}_t\right)\right) \right\} \tag{3.14}$$

where recall that $\mathbb{E}\left[\varepsilon_t\left(1\right)^{y_t}\right]=\frac{\mathrm{B}(\alpha+y_t,\beta)}{\mathrm{B}(\alpha,\beta)}$ because $\varepsilon_t\left(1\right)\sim$ beta $\left(\alpha,\beta\right)$.

3.3 Specification of a Dynamic Discrete Choice Model of Terrorism

The sequence hazard models a campaign of terror as a succession of discrete time periods $t=0,1,2,\cdots,T$. The agent is a terrorist group that in each period of the campaign chooses either to wait ($a_t=0$) or attack ($a_t=1$). In each period the agent's time-separable utility from choosing alternative a_t is $u_a\left(y_t,x_t,\varepsilon_t\right)$ (see Definition in Section 3.2) where y_t denotes the outcome of a terrorist attack, x_t denotes publicly observable state variables that influence terrorist decision-making (e.g., counter-terror policy, metal detectors, economic indicators, and so on) and the ε_t denote shocks that are privately observed by the agent. Assume the transition dynamics of state variables $[x_t,\varepsilon_y]$ are determined by $F\left(x_{t+1},\varepsilon_{t+1}\,|\,a_t,x_t,\varepsilon_t\right)$. I will show that in the simplest version of the model x_t is represents time elapsed since the last attack which is entirely deterministic $x_{t+1}=\left(1-a_t\right)\cdot\left(x_t+1\right)$ which greatly simplifies the joint cdf.

In each period the agent chooses optimal a_t^* to maximize rewards for the duration of the campaign given the current state of the world $[x_t,\varepsilon_t]$:

$$a_t^*=\underset{a_t\in\{0,1\}}{\arg\max}\,\mathbb{E}\left[\sum_{\tau=0}^{T}\delta^{\tau}u_a\left(y_{t+\tau},x_{t+\tau},\varepsilon_{t+\tau}\right)\,|\,x_t,\varepsilon_t\right]\tag{3.15}$$

where the expectation is taken with respect to the path of states x_t and ε_t as well as the expected outcome of attack y_t in future periods. The discount factor is $\delta\in(0,1)$. The publicly observable state of the world x_t can also contain the history of past attack outcomes. The per-period timing of the dynamic model augments the static model to account for the evolution of observable and unobservable state variables period to period.

Assumption 3.3.1 (Per-period Timing). The per-period timing is as follows

1. the agent observes state variables x_t and $\varepsilon_t = (\varepsilon_t(0), \varepsilon_t(1))'$
2. the agent chooses whether to attack $(a_t = 1)$ or wait $(a_t = 0)$
   - if the agent attacks, random outcome y_t is realized
   - if the agent waits, no attack takes place
3. the period ends and the next period begins
   - next period state variables x_{t+1} and ε_{t+1} are generated according to $F(x_{t+1}, \varepsilon_{t+1} \mid a_t, x_t, \varepsilon_t)$

The terrorist's optimization problem can be represented by the value function $W(x_t, \varepsilon_t)$ that solves the Bellman equation

$$W(x_t, \varepsilon_t) = \max_{a_t \in \{0,1\}} \left\{ u_a(y_t, x_t, \varepsilon_t) + \delta \int W(x_{t+1}, \varepsilon_{t+1})\, \mathrm{d}F(x_{t+1}, \varepsilon_{t+1} \mid a_t, x_t, \varepsilon_t) \right\}. \quad (3.16)$$

The optimization above is greatly simplified by assuming the ε_t are independent across alternatives and independent across time t and by assigning a parametric distribution for the unobserved states.

Assumption 3.3.2 ((EXPBETA) Joint beta and exponential unobservable states). The unobservable states $\varepsilon_t = (\varepsilon_t(0), \varepsilon_t(1))'$ are independently distributed across alternatives where $\varepsilon_t(0) \sim f(\cdot; \alpha, \beta)$ is the pdf of the beta distribution (Definition 6.1.5) with parameters $\alpha > 0$ and $\beta > 0$ and $\varepsilon_t(1) \sim f(\cdot; \lambda)$ is the pdf of the exponential distribution (Definition 6.1.8) with parameter $\lambda > 0$.

Assumption 3.3.3 ((IID) IID Unobservable States). The unobserved state variables ε_t are independently and identically distributed over time with cdf $G(\varepsilon_t)$ with finite first moments and continuous and twice differentiable in ε_t.

The following Conditional Independence (CIX) helps by breaking down the integral into manageable parts.

Assumption 3.3.4 ((CIX) Conditional Independence of Future x). The transition probability of the state variables factors as $F(x_{t+1} \mid a_t, x_t, \varepsilon_t) = F(x_{t+1} \mid a_t, x_t)$.

The assumption above says that conditional on the current choice a_t and observable state variables x_t, the realization of next period's observable state variables x_{t+1} do not depend on current period unobservable state ε_t. Assumptions CI-X and IID together imply that the transition law for future states $[x_t, \varepsilon_t]$ can be broken into two parts:

$$F(x_{t+1}, \varepsilon_{t+1} \mid a_t, x_t, \varepsilon_t) = G(\varepsilon_{t+1}) \cdot F(x_{t+1} \mid x_t, a_t).$$

Recall that in the simple form of the model x_t belongs to a finite set integers representing the amount of days elapsed since the last attack. In that case x_t evolves deterministically and $F(x_{t+1} \mid x_t, a_t)$ is replaced by the recursive equation $x_{t+1} = (1 - a_t) \cdot (w_t + 1)$.

Assumption 3.3.5 ((FIN) Finite support for x_t). The support of observable state variables x_t is discrete and finite where $X = \{x_1, \cdots, x_M\}$ with $|x_m| < \infty$ for $m \in M$ where M denotes the dimension of X.

Next the conditional independence assumption and IID assumptions will be used to express the agent's problem using the integrated value function which is only a function of observable state variable x_t. Specifically, Assumption CIX can be used to integrate both sides of equation (3.16) with respect to unobservable states ε_t and write the agent's maximized reward from the campaign as a function of only the state variables x_t.

$$
\begin{aligned}
V(x_t) &\equiv \int_{\mathcal{E}} W(x_t, \varepsilon_t)\, \mathrm{d}G(\varepsilon_t \,|\, x_t) \\
&= \int_{\mathcal{E}} \max_{a_t \in \{0,1\}} \left\{ u_a(y_t, x_t, \varepsilon_t) + \delta \int_{\mathcal{E}\times\mathcal{X}} W(x_{t+1}, \varepsilon_{t+1})\, \mathrm{d}G(\varepsilon_{t+1} \,|\, x_{t+1})\, \mathrm{d}F(x_{t+1} \,|\, x_t, a_t) \right\} \cdot \mathrm{d}G(\varepsilon_t \,|\, x_t) \\
&= \int_{\mathcal{E}} \max_{a_t \in \{0,1\}} \left\{ u_a(y_t, x_t, \varepsilon_t) + \delta \int_{\mathcal{X}} V(x_{t+1})\, \mathrm{d}F(x_{t+1} \,|\, x_t, a_t) \right\} \cdot \mathrm{d}G(\varepsilon_t \,|\, x_t) .
\end{aligned}
$$

The expression above can be written using operator notation as $V(x_t) = \mathcal{T}(V)(x_t)$ where

$$
\mathcal{T}(V)(x_t) = \int_{\mathcal{E}} \max_{a_t \in \{0,1\}} \left\{ u_a(y_t, x_t, \varepsilon_t) + \delta \int_{\mathcal{X}} V(x_{t+1})\, \mathrm{d}F(x_{t+1} \,|\, x_t, a_t) \right\} \cdot \mathrm{d}G(\varepsilon_t \,|\, x_t) .
$$

where $\Gamma(\cdot)$ on is a contraction mapping and V is the unique fixed point of $V = \Gamma(V)$.

The agent's preferences defined in Definition 3.2.1 can be augmented to reflect the dynamic component. Here, $\bar{u}_a$ updates $u_a(y_t, x_t, \varepsilon_t)$ from Definition 3.2.1 in Section 3.2.1 to include the future value of the campaign after the agent attacks or waits

$$
\bar{u}_a(y_t, x_t, \varepsilon_t) = \begin{cases} y_t(x_t, \varepsilon_t(1)) + \delta \mathbb{E}_t V(0) & \text{if } a_t = 1 \\ \theta_0 - \varepsilon_t(0) + \delta \mathbb{E}_t V(x_{t+1}) & \text{if } a_t = 0 \end{cases}
$$

which can be written more succinctly as $\bar{u}_a = y_t(x_t, \varepsilon_t(1))$ if $a_t = 1$ and $\bar{u}_a = \bar{\theta}_0 - \varepsilon_t(0)$ if $a_t = 0$ where $\bar{\theta}_0 = \theta_0 + \delta \mathbb{E}_t V(x_{t+1}) - \delta \mathbb{E}_t V(0)$ where recall that in the simplest version of the model $x_{t+1} = (1 - a_t) \cdot (x_t + 1)$.

Subtracting $\bar{\theta}_0 - \varepsilon_t(0)$ from both sides of Equation (3.22) gives an integrated Bellman equation of the form (see Result 6.5.1 in Appendix E).

$$
\begin{aligned}
V &= \mathrm{P}(a_t = 1 \mid x_t) \cdot \left[\mathbb{E}[y_t \mid a_t = 1, x_t] - \bar{\theta}_0 + \mathbb{E}[\varepsilon_t(0) \mid a_t = 1, x_t]\right] \\
&= \mathrm{P}(a_t = 1 \mid x_t) \cdot \left[q(x_t; \theta_q) \cdot \mathbb{E}[\varepsilon_t(1) \mid a_t = 1, x_t] - \bar{\theta}_0 + \mathbb{E}[\varepsilon_t(0) \mid a_t = 1, x_t]\right]
\end{aligned}
$$

where the second line substitutes for the expectation of y_t conditional on $a_t = 1$ and x_t given by $q(x_t; \theta_q) \cdot \varepsilon_t(1)$ in assumption CEY (see 3.2.1 from the static model presented in Section 3.2.1).

The following truncated or conditional moments of the choice-specific elements of ε_t follow from the independence assumption and the parametrization of $G(\varepsilon_t)$ as the product of joint beta and exponential cdfs. These moments are used to further simplify the integrated Bellman equation.

Result 3.3.1 (Conditional moments of $\varepsilon_t(0)$ and $\varepsilon_t(1)$ given $a_t = 1$). Assume that the choice-specific elements of ε_t are independent and distributed $\varepsilon_t(0)$ as exponential with parameter λ and $\varepsilon_t(1)$ as beta with parameters α and β. Then

$$
\mathbb{E}[\varepsilon_t(0)^{y_t} \mid a_t = 1, x_t] = \lambda^{-y_t} \cdot [1 - \log \mathrm{P}(a_t = 1 \mid x_t)]^{y_t} \tag{3.17}
$$

$$
\mathbb{E}[\varepsilon_t(1)^{y_t} \mid a_t = 1, x_t] = P(a_t = 1 \mid x_t)^{-1} \cdot M_{y_t}(\bar{\varepsilon}_t) \tag{3.18}
$$

where

$$
M_j(\gamma) = \mathbb{E}\left[\varepsilon_t(1)^j\right] \cdot \overline{F}(\gamma; \alpha + j, \beta), \tag{3.19}
$$

$$
\mathbb{E}\left[\varepsilon_t(1)^j\right] = \mathrm{B}(\alpha + j, \beta) / \mathrm{B}(\alpha, \beta), \tag{3.20}
$$

where $\overline{F}(\gamma; \alpha + j, \beta)$ is the survivor function of the beta $(\alpha + j, \beta)$ distribution, $\mathrm{B}(\cdot, \cdot)$ is the beta function (see Definition 6.1.2 in Appendix A), and

$$\bar{\varepsilon}_t = F^{-1}\left(1 - \mathrm{Pr}\left(a_t = 1 \,|\, x_t\right); \alpha, \beta\right) \tag{3.21}$$

where $\mathrm{Pr}\left(a_t = 1 \,|\, x_t\right)$ denotes the conditional probability of the agent choosing to attack.

Proof. See Appendix Results 6.4.7 and 6.4.8. □

Plugging in for the formula for the conditional moments gives the integrated Bellman

Result 3.3.2 (Integrated Bellman equation).

$$\begin{aligned} V\left(x_t\right) &= q_t \cdot M_1\left(\bar{\varepsilon}_t\right) + \frac{\mathrm{P}\left(a_t = 1 \,|\, x_t\right)}{\lambda}\left[1 - \log \mathrm{P}\left(a_t = 1 \,|\, x_t\right) - \lambda\theta_0\right] \\ &\quad - \mathrm{P}\left(a_t = 1 \,|\, x_t\right) \cdot \delta\left[\mathbb{E}_t V\left(x_{t+1}\right) - \mathbb{E}_t V\left(0\right)\right]. \end{aligned}$$

Proof. The result follows from plugging in the truncated moments from Results 3.17 and 3.18 above and combining that with the expectation of a max operator (see Result 6.5.1 in Appendix E). □

Following Aguirregabiria and Mira (2002) one can exploit the stationary structure of the dynamic programming problem to represent forward time periods with a tilde, e.g., x', and exploit the finiteness of state variables x_t to replace the integral with a sum over discrete states of the world. Note that $V\left(x_t\right)$ is defined by an integral that depends on the distribution of the elements in ε_t as well as the utility function $u_a\left(\cdot\right)$. Thus V is a function of distribution parameters α, β, λ, utility function parameters θ_0 and θ_q, and discount parameter δ.

Representing the dynamic programming solution using a linear operator gives

$$V = \sum_{a \in \{0,1\}} \mathrm{P}(a \mid x) \left\{ \mathbb{E}\left[u_a(y, x, \varepsilon(a)) \mid x, a\right] + \delta \sum_{x'} f(x' \mid x, a) V(x') \right\}. \tag{3.22}$$

Note that the truncated moments of y_t and $\varepsilon_t(0)$ are both functions of $\mathrm{P}(a_t = 1 \mid x_t)$ (recall that $\bar{\varepsilon}_t = F^{-1}(1 - \mathrm{P}(a_t = 1 \mid x_t); \alpha, \beta)$). To emphasize this fact, write $Y(a, P)$ for the conditional expectation of y_t (determined by the truncated moment Equation (3.17)) and write $e(a, P)$ for the conditional expectation of $\varepsilon_t(0)$ (determined by the truncated moment in Equation (3.18)) and let $\bar{e}(a, P) = e(a, P) - \theta_0$. Continuing as per Aguirregabiria and Mira (2002) gives the Bellman equation

$$V = \sum_{a \in \{0,1\}} P(a) * \left[Y(a, P) + \bar{e}(a, P) + \delta F(a) V\right]$$

where V is a $M \times 1$ vector containing the value function $V(x)$ evaluated at the M points, P is the conditional choice probability that $a = 1$, and $Y(1, P)$, $\bar{e}(1, P)$, and $P(a)$ are $M \times 1$ vectors stacking the functions described above. The matrix $F(a)$ is $M \times M$ and contains the transition probabilities for the next period value of x' conditional on current period action a and current state x (see Equation (3.22)).

Hotz and Miller (1993) showed that under additive separability and infinite support of the unobservable shocks ε_t that there is a one-to-one mapping between CCPs and differences in the value functions. The sequence hazard does not admit additive separability, nonetheless it is still possible to apply the inversion property, e.g., apply the nested pseudo-likelihood method of Aguirregabiria and Mira (2002). Kristensen et al. (2015) showed that a one-to-one mapping between the CCPs and value function differences is possible with no additive separability if and only if the agent's payoffs have full support on the Euclidean space $\mathbb{R}^2$ (see also page 251 of Aguirregabiria et al. (2021)).

For completeness, I state an assumption summarizing the agent's preferences in the sequence hazard model for easy comparison to the framework of Rust (1987).

Assumption 3.3.6 ((MM) Multiplicative Model). State variable $\varepsilon_t = (\varepsilon_t(0), \varepsilon_t(1))'$ has unbounded support $\varepsilon_t \in \mathcal{E} = \mathbb{R}_+ \times (0,1)$ and $\varepsilon_t(1)$ enters $u_a(y_t, x_t, \varepsilon_t(a))$ non-additively where $u_a(y_t, x_t, \varepsilon_t(a), \varepsilon_t(a))$ depends only on the a-th component of the 2×1 vector ε_t, $\varepsilon_t(0)$ is a positive random variable, and $\varepsilon_t(1)$ is bounded between 0 and 1.

The assumptions introduced in this section contrasted with the six assumptions that define the conditional logit model from Rust (1987) (see also Aguirregabiria and Mira (2010)), namely, 1) additive separability, 2) iid unobservables, 3) conditional independence of x, 4) conditional independence of the payoff variable, 5) conditional logit assumption, and 6) discrete observable states x. The key differences with the sequence hazard model are that the latter has one of the choice-specific unobservable states enter the utility function non-additively (e.g., see Assumption MM above). Moreover, the conditional independence of the payoff assumption from Rust (1987) is replaced with the CEY conditional expectation assumption meaning the outcome of the payoff variable *is* influenced by an unobservable state, $\varepsilon_t(1)$. The model that uses attack outcomes makes even stronger assumptions about the particular parametric distribution for the outcome variable y_t. The sequence hazard also swaps the assumption that unobserved states are conditionally logit for the BETAEXP assumption that they are exponentially and beta distributed which also implies independence of irrelevant alternatives (see the CLOGIT assumption on pg. 40 of Aguirregabiria and Mira (2010)). Assumptions IID, CEX, and FIN are the same as in Rust (1987), that is, unobservable states are independently identically distributed and, conditional on current state variables, next period observable state variables do not depend on current *unobservable* states. Keeping these assumptions from the standard model is crucial to simplifying the Bellman equation and to estimating the model.

Thus the $\mathbb{E}\left[u\left(\cdot\right) \mid x, a\right]$ from Equation (3.22) are known functions of the CCP. Stacking the X equations as $u\left(a, P\left(a\right)\right)$ and using the $X \times X$ conditional probability matrix $F\left(a\right)$ to represent the transition of state x as a function of action a gives

$$V = \left(I_M - \delta F^U\left(P\right)\right)^{-1}\left\{P * \left(Y\left(a, P\right) + \bar{e}\left(a, P\right)\right)\right\} \tag{3.23}$$

where $F^U\left(P\right) = \sum_{a \in A} \mathrm{P}\left(a\right) * F\left(a\right)$ is induced by P. Compare this to equation (8) from Aguirregabiria and Mira (2002) where the quantity in brackets is $u\left(a\right) + e\left(a, P\right)$ is instead tailored to the extreme value distribution, but is otherwise identical to Equation (3.23). Note that matrix $F\left(a\right)$ takes on a particularly simple form in the simple waiting model. $F\left(0\right)$ is a $X \times X$ matrix with a column of 1s as the last column and $F\left(1\right)$ is a $X \times X$ matrix with 1s in the off diagonal.

The following result collects the properties of the CCP in the dynamic case.

Result 3.3.3 (Conditional Choice Probability (CCP)). In the optimal control problem described by the per-period timing of Definition 3.3.1 and the preferences and assumptions given by MM, IID, CEX, CEY, EXPBETA and FIN (i) the probability of the agent attacking is given by equation (3.24) below, (ii) the CCP in equation (3.24) is equivalent to a hazard, (iii) the CCP is saturated in the sense that $\mathrm{P}\left(a_t = 1 \mid x_t\right) > 0$ for all $x_t \in X$, (iv) the duration dependence is positive for all

$x_t \in X$ and given by

$$\frac{\partial \mathrm{P}\left(a_t=1 \,|\, x_t\right)}{\partial x_{ti}} \lesseqqgtr 0 \iff \begin{cases} q_t' \cdot \mathbb{E}\left[\varepsilon_t\left(1\right)\right] \lesseqqgtr \delta \cdot \mathbb{E}V'\left(x_t\right) \cdot \frac{\mathcal{H}_{\alpha,\beta}\left(\bar{\theta}_0/q_t, \lambda\bar{\theta}_0\right)}{\mathcal{H}_{\alpha+1,\beta}\left(\bar{\theta}_0/q_t, \lambda\bar{\theta}_0\right)} & \text{if } \bar{\theta}_0/q_t \in (0,1) \\ q_t' \cdot \mathbb{E}\left[\varepsilon_t\left(1\right)\right] \lesseqqgtr \delta \cdot \mathbb{E}V'\left(x_t\right) \cdot \frac{\mathrm{P}\left(a_t=1 \,|\, x_t\right)}{\mathrm{P}_{\alpha+1}\left(a_t=1 \,|\, x_t\right)} & \text{if } \bar{\theta}_0/q_t \geq 1 \end{cases} \tag{3.24}$$

where $q_t \equiv q\left(x_t; \theta_q\right)$, $P_{\alpha+1}\left(a_t = 1 \,|\, x_t\right)$ denotes the CCP $\mathrm{P}\left(a_t = 1 \,|\, x_t\right)$ where the parameter α is replaced by $\alpha + 1$, and $\bar{\theta}_0 \equiv \theta_0 + \delta \mathbb{E} \Delta V\left(x_t\right)$ where $\theta_0 > 0$ with $\Delta V\left(x_t\right) = V\left(x_t + 1\right) - V\left(0\right)$.

Proof. The CCPs equivalence to a hazard was proved in the text of Section 3.2.1. For the result on duration dependence see Result 6.3.9 in Appendix C. □

The CCP in the dynamic case is just like the static case but the parameter θ_0 has an extra term representing the dynamic component of the agent's problem. In the dynamic model the CCP enters the likelihood function just as it did in the static case. In the static model the risk of a terrorist attack was always increasing, but in the dynamic model it depends on the slope of the function q_t determining the agent's belief about the expected size of an attack. The probability of an attack is increasing if the change in the expected size/success of the function weighed by the expected adverse shock $\mathbb{E}\left[\varepsilon_t\left(1\right)\right]$ is greater than the change in the discounted value of the campaign weighed by a geometric average of the probabilities of attacking.

3.4 Overview of Dynamic Discrete Choice Models in Economics

Chapter 2 reviewed empirical methods for evaluating the effectiveness of counterterrorism policies and identified the need for a sequential model of terrorist decision-making that takes into account not only the frequency of attacks, but also their severity. In risk analysis one is interested in drawing inference about a specific terrorist group's behavior having observed only a single time series of a terrorist group's attack choices and attack outcomes. There are no random control trials or natural experiments on which to base empirical work. Moreover, there are few sources of detailed information on the inner workings of terrorist groups or the secretive state agencies that try to interdict them. One obvious but as of yet unexplored avenue for empirical work in this setting is the use of structural economic models. Discrete choice and dynamic programming (DP) models have been widely used in economics to describe situation where agents make sequential decisions in the face of uncertainty. They are particularly useful in applications topics like terrorism where researchers observe only a limited set of the state variables relevant to the agent's choices.

Structural models are those in which an agent's decision-making is reflected in the specification of the model. The goal of any DP model is to solve for a decision rule that can be employed in estimating the structural parameters of an agent's preferences and beliefs. In other words, observed choices are viewed as the solution to an optimal control problem that the agent is assumed to be solving. Agents are given explicitly formulated expectations over future states of the world as a function of past choices and past states of the world. Following Blundell (2017), the building blocks of a structural model are (1) the *structural "deep" parameters*: e.g., the utility function parameters; (ii) *underlying mechanisms:* e.g., planning time as an input into the preparation of a terrorist attack; (iii) *policy counterfactuals*: eg., the level of terrorist violence absent a counterterrorist policy. Reiss and Wolak (2007), Keane and Wolpin (2009), and Low and Meghir (2017) review the unifying elements of structural models.

Structural modeling models have been deployed in a wide range of empirical settings. Pakes (1986) proposed a non-stationary, optimal stopping model of the return to innovation to estimate renewal rates of European patents as a function of patent age. The model has a finite horizon and is solved by backward induction. The behavior of patent holding imposes restrictions on the exit from holding a patent. Pakes used aggregate patent expiry data to derive the value functions for the model and then used a simulation to estimate model parameters. Rust (1987) modeled the replacement of bus engines by the Wisconsin DOT as a cost minimization problem using only monthly bus mileage data. Cho (2011) employed a similar framework as Rust to study the replacement of computer mainframes at a large technology firm using data on computer prices and replacement decisions. DP models have also been used to model the career decisions of young men as a function of their age (Keane and Wolpin (1997), Keane and Wolpin (2009), and Keane et al. (2011)). Aguirregabiria and Mira (2010) and Reiss and Wolak (2007) surveys these methods. Amemiya (1996) and Kiefer (1988) are some early examples of structural duration models applied to IO and labor, respectively.

The sequence hazard model is distinguished by the fact that one of the unobservable states enters the agent's problem in non-additively separable manner. The model draws from two related types of dynamic discrete choice models. On the one hand there is the framework introduced by Rust (1987) (see also Rust (1994)) where the choice model consists of an infinite horizon problem with binary controls. The unobserved states enter the utility function as a summed quantities assumed to be conditionally independently drawn as Generalized Extreme Value Type 1 random variables. The combination of parametric assumption, additive separability, and conditional independence allow for a closed form solution for the conditional choice probability that is similar to McFadden et al. (1973) but with the addition of component representing the agent's (discounted) valuation of future utility. Rust proposes a nested algorithm with an outer layer that searches for a minimizer and an internal layer that solves a DP problem for a given draw of the parameters. Crucially, the additive separability avoids multiple integration over unobservable states which greatly reduces the dimensionality of the problem. The method exploits the fact that the derivatives of the CCPs are straightforward to derive and employ in a Newton-style algorithm.

On the other hand there is the framework used by Keane and Wolpin (1997) and others to study career choices. These models usually involve a multinomial choice (e.g., multiple career paths), employ a finite horizon, and are solved by backward induction. In this framework some of the career outcomes are determined by wages which have an additively separable error component at the level of a log-wage regression (as in empirical labor models), but do not enter the agent's problem additively. Because of the multiplicative component of utility and specific parametric assumptions, these models do not admit a closed form solution and are thus solved by simulation and interpolation. Specifically, the value functions themselves are modeled by polynomial approximations and then these are used in a simulation to estimate the model parameters. Keane and Wolpin (2009) and Keane et al. (2011) provide a survey of these models, and Britton and Waltmann (2021) provide an up to date review of the simulation methods used to solve this type of model. For another model that uses a similar multiplicative structure for shocks, but in a non-labor context, see Kasahara (2009).

The two modeling approaches outlined above have relied on a full solution or appealed to simulation. A full solution is attractive because it is efficient, but it is costly to implement, particularly as the dimension of possible state values increases. Starting with Hotz and Miller (1993) a number of authors have proposed iterative procedures as alternatives to a full solution method. Hotz and Miller (1993) exploited the assumption that the unobserved states are distributed Generalized Extreme Value Type 1 to arrive at a closed form expression for the value function as a function of the conditional choice probabilities and showed that under certain conditions this mapping could be inverted which allows one to work in the space of probability functions instead of value functions. This two-step method was further generalized by Aguirregabiria and Mira (2002) and further refined by Kasahara and Shimotsu (2008) who propose initially estimating the CCPs with consistent non-parametric estimates and then apply the Holtz and Miller's method K times until convergence. This addressed the former method's tendency to suffer from finite sample bias when the CCPs were poorly estimated in the first step. But a major limitation of these one or K-step iterative methods remains which is that they only apply if the unobservable state is additive and follows a Generalized Extreme Value Type 1 (GEV) distribution. Kasahara et al. (2011) and Kasahara and Shimotsu (2012) proposed an iterative method that focuses on the value function space rather than the conditional probability space. This allows their method to include setups with non-additive errors and distributional assumptions other than the GEV, but they do not elaborate on any such case. Arcidiacono and Miller (2011) implement unobserved heterogeneity into dynamic models using Expectations Maximization (EM) algorithm.

As shown above, most of the research has been done on models with separative additivity, the same applies to identification results. Rust (1987) showed that the nested fixed point algorithm was unable to estimate the dynamic discount factor β (it is non-parametrically unidentified) so he set it either to 0 (static/myopic) or 0.999. Magnac and Thesmar (2002) later demonstrated that for dynamic discrete choice models the utility function for each action cannot be identified without more information about the distribution of unobserved errors/states and the discount factor. The Monte Carlo experiment conducted in Chapter 4 shows that the sequence hazard model is identified.

The main feature that distinguishes the sequence hazard model from previous single agent DP models is the mixed additivity assumption. In the canonical GEV model originally proposed by McFadden and generalized to the dynamic case by Rust (1987), the unobservable states enter the agent's preferences in additive fashion, one for each decision available to the agent. In contrast, in the sequence hazard model one of the errors enters the agent's problem multiplicatively because it represents a shock that is detrimental to the terrorist's plans implying it is bounded between 0 and 1. Relaxing additive separability is usually difficult and models that do so such as Keane and Wolpin (1997), Keane and Wolpin (2009), Keane et al. (2011) are estimated via simulation and involve polynomial approximations.

The series hazard's parametric assumptions for the distribution of the additive and multiplicative unobservable state lead to a closed form solution for the CCP. However, the CCP makes use of functions for which the derivatives are quite complicated so the model cannot be estimated using the Nested Fixed Point algorithm in Rust (1987). Moreover, in principle the two-step method proposed by Aguirregabiria and Mira (2002) cannot be employed because as conjectured in Kristensen et al. (2015), Aguirregabiria et al. (2021), and Aguirregabiria and Mira (2010) the support for the unobserved state must be unbounded which does not hold for the security shock between 0 and 1. However, the CCP in this specific case is saturated meaning it is positive for all possible states of the world so the Aguirregabiria and Mira (2002) method still applies.

Structural models are designed specifically for the application in mind so in general there are no standardized tests that can be used to validate a structural model. Certain distributional assumptions can be tested against the data such as the evolution of future bus mileage in Rust (1987) and likelihood ratio tests can be used to reject a myopic/static model. Larsen et al. (2012) exploit the fact that the GEV distribution can be nested as a special case of more exponentiated generalized beta distribution. This allows them to use likelihood ratio tests to validate the GEV distribution (fail to reject) against the bus engine replacement data.

In some cases there are government interventions that create treatment groups that can be compared directly to the predictions of a discrete dynamic model such as the case of the PROGRESA conditional cash transfer program implemented by the Mexican government in 1997 in an effort to meet school attendance goals (see Todd and Wolpin (2006), Keane and Wolpin (2009), and Low and Meghir (2017)). No such government interventions are available in the case of terrorism so instead the focus of a model validation exercise is the ability to predict future patterns of attacks and to explain shifts in tactics and attack type as a function of counterterrorist policies (see Chapter 4).

Chapter 4: Comparing the Sequence and Series Hazard Models

4.1 Introduction

This chapter presents a Monte Carlo simulation to compare the sequence hazard model presented in Chapter 3 to the series hazard model presented in Chapter 2.

4.2 Monte Carlo Design

The sequence hazard model is a dynamic discrete choice model formulated very similarly to the Rust (1987) bus engine replacement model with a few key differences owing to the mixed additive exponential-beta formulation. However, the conditional independence assumption still holds so as derived in Section 3.3 the two models have closed-solutions which simplifies the truncated expectation of the unobserved state conditional on attacking. A major drawback of the sequence hazard model is that depends on functions that are more complicated than the exponential function used for the logit bus model. These functions are difficult to calculate precisely at the endpoints and are time consuming to compute. Moreover, the derivatives do not have straightforward formula so the Newton-based nested fixed point algorithm from Rust (1987) cannot be implemented.

The next outline describes the Monte Carlo simulation.

Definition 4.2.1 (Outline of Monte Carlo routine).

1. set the true parameter θ_o
2. draw an $S \times N$ matrix of random errors $\varepsilon = (\varepsilon(0), \varepsilon(1))'$ according to the exponential (λ) distribution for $\varepsilon(0)$ and the beta (α, β) distribution for $\varepsilon(1)$. S is the number of simulations

and N is the sample size

3. iterate the value function from equation (3.22) to solve for $EV_o(\theta_o)$

4. use $EV_o(\theta_o)$ to calculate the CCP, $P_o(\theta_o)$

5. use knowledge of $EV_o(\theta_o)$ and $P_o(\theta_o)$ to simulate a campaign according to the decision rule (Equation (3.1)) using the N random errors for simulation S in step 1

6. start with a trial θ_k and solve the dynamic problem for the EV_k and P_k implied by θ_k and calculate the corresponding likelihood (see Equation 3.2.3))

7. maximize the likelihood to estimate parameter θ_{k+1}

8. repeat until the sequence of $|\theta_{k+1} - \theta_k|$ achieves the desired tolerance

4.3 Monte Carlo Results

Parameter	Role	Values	Note
θ_0	linear	$+1.250$	net benefit of waiting
θ_1	linear	$+0.500$	belief function
θ_2	quadratic	-0.050	belief function
θ_3	policy dummy	-1.000	belief function
λ	innovation cost	$+1.000$	exponential rate
α	security shock	$+2.000$	beta shape 1
β	security shock	$+2.000$	beta shape 2

Table 4.1: Parameters for the Monte Carlo Simulation

Table 4.3 summarizes the parameter choices for the Monte Carlo experiment. The parameters of the expected attack size function $q(x_t; \theta_q)$ for the simulated model were given a quadratic structure to mimic a rise and fall in terrorism risk. The policy dummy was set to -1 to simulate a counterterrorist policy because the agent expects that, all else equal, it lowers the severity of attacks.

The sequence hazard behavioral model of terrorism was used to simulate $S = 250$ datasets for sample sizes $N = 500$, $N = 1000$, and $N = 2000$ representing about a year and a third to five years and a half of data on terrorist attacks. Table 4.2 demonstrates the validity of the CCP and JCOP likelihood approaches presented in Result 3.2.3 in Section 3.2.1 and result 3.2.6 in Section 3.2.2 to estimate the parameters of the model. In the table the variance of estimates decreases as the sample size increase. The JCOP approach appears to do slightly better with less data than the CCP approach which makes sense since it is leveraging more information.

Figure 4.1 illustrates the difference in estimation performance between the sequence hazard and series hazard models. Plotted are the risk or hazard of a terrorist attack with the counterterrorism policy dummy is on and off. The true hazard is found by evaluating the CCP at the true value of the parameters θ_o. The estimated sequence hazard plotted in the figure is the average over $S = 250$ estimated sequential hazards (CCPs) that used the likelihood for the CCP-based model (see Result 3.2.3). In this particular simulation there is a brief initial period where the risk of a terrorist attack increases, but overall the policy has a deterrent effect. The severity or size of attacks before activating the counterterrorist policy dummy is 0.9153 expected casualties per attack. Once implemented, the policy that effectively reduced the probability of an attack from 0.5 to 0.37 (see Figure 4.1) and reduced the average severity to 0.376 casualties per attack.

The series and sequence hazard models are not directly comparable. Thus, there is no "true" value to report for the series hazard estimates and the only quantity estimated is the coefficient on the policy dummy. The estimated series hazard plotted in the figure is the average over the $S = 250$ hazards estimated via the Cox proportional hazard method. The series hazard is unable to

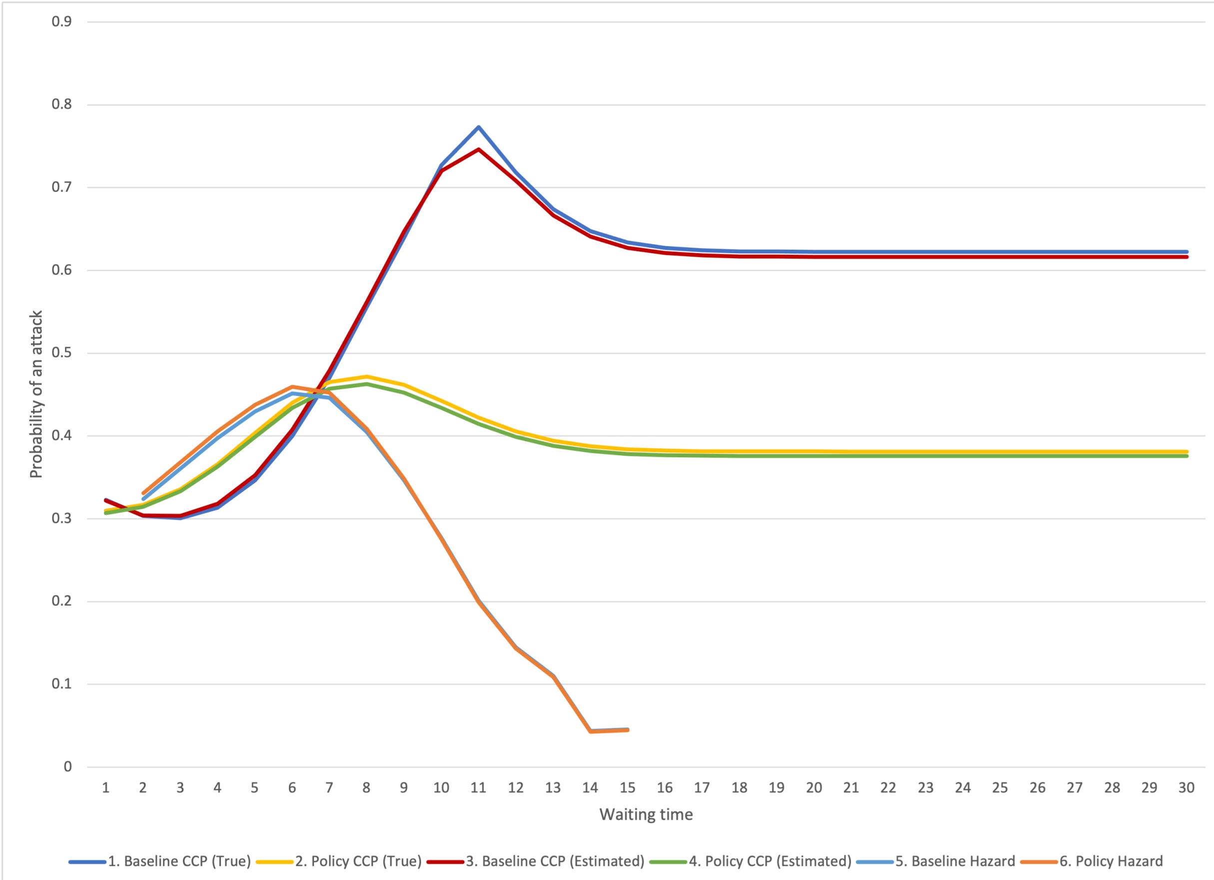


Figure 4.1: Comparing the Sequence and Series Hazards.

capture the true hazard, a finding that is robust to different specifications, including the addition of cumulative past failures as a measure of attack density as suggested by Dugan (2011). The mean hazard ratio over 250 simulations with sample size 1000 was .899 (with sd = 0.100) so the series hazard is able to at least detect that the risk of an attack decreased as a result of the deterrence policy. I implemented various tests of the null of proportional data and found the probability of rejecting the null to be essentially a coin toss.[1]

[1]The tests of proportionality implemented and their respective p-values are Identity: 0.57, Kaplan-Meier: 0.52, Log: 0.48, Logrank: 0.49, Rank: 0.52.

Table 4.2: Monte Carlo Simulation Results

Sample Size		$N = 500$			$N = 1000$			$N = 2000$		
Parameter	True	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
θ_0										
CCP	1.250	1.257	0.047	0.048	1.256	0.043	0.043	1.255	0.024	0.025
JCOP	1.250	1.247	0.044	0.045	1.254	0.042	0.042	1.248	0.023	0.024
θ_1										
CCP	0.500	0.501	0.055	0.055	0.505	0.040	0.040	0.502	0.032	0.032
JCOP	0.500	0.501	0.053	0.053	0.502	0.039	0.039	0.501	0.030	0.031
θ_2										
CCP	-0.050	-0.050	0.004	0.004	-0.050	0.004	0.004	-0.051	0.003	0.003
JCOP	-0.050	-0.051	0.003	0.003	-0.050	0.003	0.003	-0.050	0.002	0.002
θ_3										
CCP	-1.000	-1.008	0.087	0.088	-1.002	0.066	0.066	-1.014	0.062	0.064
JCOP	-1.000	-1.009	0.075	0.077	-1.001	0.054	0.054	-1.008	0.060	0.610
λ										
CCP	1.000	1.003	0.037	0.037	1.000	0.034	0.034	0.998	0.019	0.020
JCOP	1.000	1.002	0.039	0.040	1.000	0.032	0.032	1.003	0.018	0.019
α										
CCP	2.000	2.017	0.173	0.173	1.999	0.169	0.169	2.005	0.101	0.102
JCOP	2.000	2.011	0.134	0.134	2.001	0.163	0.163	2.003	0.098	0.098
β										
CCP	2.000	1.943	0.162	0.171	1.969	0.133	0.136	1.981	0.107	0.109
JCOP	2.000	2.101	0.153	0.162	2.011	0.130	0.134	2.001	0.104	0.105

Chapter 5: Conclusion

Terrorism is a type of political violence where deliberate attacks against unarmed, non-military assets are used as a tactic of intimidation to influence a targeted society. Chapter 2 started off emphasizing the importance of modeling the risk of a terrorist attack by reviewing studies that estimate the human and economic costs of terrorism. The literature review emphasized how the mere risk of terrorism can have long lasting economic consequences even if the outcome of any given terrorist attack costs only a fraction of total GDP. Thus it is important to understand the decision making process of terrorist networks to reduce the severity or prevent further terrorist attacks.

Next the chapter reviewed a parametric (Hsu and Quandt (1976)) and a regression model (Landes (1978)) of airline hijacking. The former introduced the notion that the time series of terrorist events departs from a homogeneous Poisson model and the latter employed a rational terrorism framework to formally test for counterterrorist policies as a source of non-homogeneity in the data. The majority of empirical work on the risk of terrorism has maintained the high level assumption that the hazard of an attack can be described by a proportional hazard model. In this dissertation I introduced a simply stylized model I called the sequence hazard where the decision making process of terrorists regarding the timing and severity of an attack is derived from low level assumptions. I modeled the decision making process of terrorists as an optimal control problem where a longer planning period may lead to an increase in the severity of the attack, but also to a higher chance of disruption of the terrorist network before the attack takes place. I found that the hazard implied by the sequential model is not proportional, in contrast to the high level assumption of proportionality maintained by empirical studies of terrorism risk. This finding is significant in that it suggest the possibility of a fundamental misspecification problem for much of the empirical literature. Estimates based on a misspecified model are typically inconsistent. This will likely lead to a false assessment of the terrorist decision making process and possibly mistakes in measures to counter terrorism.

The parametric assumptions of the sequence hazard allowed me to model both the decision to attack and the interdependent decision on the expected severity of the attack. The existing literature on terrorist activity uses only information on terrorist events, not on the magnitude of events. By building a simple model I characterized the trade-off the terrorist faces: bigger events take more planning and time, but have a bigger impact, which must be weighed against the probability of being detected and stopped before the attack can take place. By considering both the timing of attacks as well as their expected size or severity, I create a model of terrorist decision making that enables me to more accurately describe the constraints to terrorist activity, and thus potentially their behavior

I obtained close-form solutions for both the hazard of an attack as well as the interdependent probability distribution that determines the observed outcome. Obtaining closed form solutions helps generate observations on the decision to attack and attack severity in the Monte Carlo study. In particular, I am able to obtain closed form expression for the truncated expectation function of state variables unobservable to the researcher, a key part of solving the dynamic model.

Two approaches to studying the impact of counterterrorist policies are time series methods and the series hazard duration approach. The series hazard model has been shown to be a valid method of studying the risk of terrorist attacks (Dugan (2011)). A limitation of the series hazard is that it does not model attack outcomes even though these are observed at the time of the event.

Accounting for attack outcomes is an important part of policy analysis because policies aimed at reducing the frequency of attacks may lead to attacks of increased severity. Chapter 3 introduced the sequence hazard model which considers attacks by a single terrorist group as forming part of a campaign of terror. The model links the outcome of terrorist attacks to the choice of when to attack by taking the time elapsed in between attacks as an input into the planning of the next terrorist attacks. The simplest version of the model results in a attack hazard that is non-proportional in time. If in addition one assumes that outcomes are generated as mixed Poisson-beta or Bernoulli-beta then the model implies a data generating process for terrorist attack outcomes that can be taken to the data. The sequence hazard differs from the classic dynamic discrete choice model with additively separable logit unobservable states using instead a exponential-beta approach where one of the unobservable states enters the utility function non-additively. Instead of assuming conditional independence of the payoff variable, the sequence hazard model imposes stronger assumptions about the data generating process for attack outcomes. The model can be estimated in full via maximum likelihood or by using an iterative approach (Aguirregabiria and Mira (2002)).

Chapter 4 presented a Monte Carlo simulation of the sequence hazard model. It demonstrates that the model parameters can be recovered from the data, including the effect of a counterterrorist policy. In contrast, the series hazard model is unable to estimate the true policy effect or detect the non-proportionality of the data via the usual statistical tests. The behavioral model was based on a fundamental trade off faced by all terrorist groups that leads to a probability of attacking that cannot be represented by a proportional hazard. The simulations raise the possibility that proportional hazard models might not be able to deliver accurate estimates of the hazard if the data are generated by such a model.

The theoretical model of sequential choice presented in this dissertation and the implied estimation strategy are not confined to the study of terrorism. The model can also be applied to other situations where agents innovate over time and exit a waiting state with an event that is measured at the time of the event.

Chapter 6: Appendix

6.1 Appendix A: Definitions and Basic Results

Result 6.1.1 (Binomial theorem). The Binomial theorem states that for any two numbers u and v and any non-negative integer n

$$(u+v)^k = \sum_{k=0}^{n} \binom{n}{k} u^{n-k} v^k = \sum_{k=0}^{n} \binom{n}{k} u^k v^{n-k}.$$

Definition 6.1.1 (Gamma function). For $z > 0$ the gamma function is defined as

$$\Gamma(z) = \int_0^\infty s^{z-1} \exp(-s) \, \mathrm{d}s$$

and has the property that $\Gamma(n) = (n-1)!$ for any positive integer n.

Definition 6.1.2 (Beta function). For α and β such that $\alpha > 0$ and $\beta > 0$ the beta function is defined by the integral

$$\mathrm{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} \, \mathrm{d}t$$

and can be represented by the gamma function defined in (Definition 6.1.1) as

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Definition 6.1.3 (Incomplete beta function). For α, β, and z such that $\alpha > 0$, $\beta > 0$, and $z > 0$ the incomplete beta function is defined by the integral

$$\mathrm{B}\left(z;\alpha,\beta\right) = \int_0^z t^{\alpha-1}\left(1-t\right)^{\beta-1}\mathrm{d}t.$$

Definition 6.1.4 (Regularized incomplete beta function). For α, β, and z such that $\alpha > 0$, $\beta > 0$, and $z > 0$ the regularized incomplete beta function is defined by

$$I_z\left(\alpha,\beta\right) = \frac{\mathrm{B}\left(z;\alpha,\beta\right)}{\mathrm{B}\left(\alpha,\beta\right)}$$

where $\mathrm{B}\left(z;\alpha,\beta\right)$ is the incomplete beta function (Definition 6.1.3) and $\mathrm{B}\left(\alpha,\beta\right)$ is the beta function (Definition 6.1.2).

Definition 6.1.5 (Beta probability distribution function). A random variable S distributed beta with parameters $\alpha > 0$ and $\beta > 0$ has probability distribution function given by

$$f\left(s;\alpha,\beta\right) = \frac{s^{\alpha-1}\left(1-s\right)^{\beta-1}}{B\left(\alpha,\beta\right)} \text{ for } s \in \left(0,1\right).$$

Definition 6.1.6 (Beta distribution cumulative density function). The cumulative distribution function of a beta distributed random variable is equivalent to the regularized incomplete beta function (Defintion 6.1.3)

$$F\left(s;\alpha,\beta\right) = \frac{\mathrm{B}\left(s;\alpha,\beta\right)}{\mathrm{B}\left(\alpha,\beta\right)} \equiv I_s\left(\alpha,\beta\right)$$

where $\mathrm{B}\left(s;\alpha,\beta\right)$ is the incomplete beta function (Definition 6.1.3) and $\mathrm{B}\left(\alpha,\beta\right)$ is the beta function

(Definition 6.1.2).

Definition 6.1.7 (Poisson distribution). A random variable Z distributed as a Poisson random variable with parameter $\mu > 0$ has probability mass function

$$\Pr(Z = z; \mu) = \frac{\exp(-\mu)\,\mu^z}{z!} \text{ for } z = 0, 1, 2, \ldots.$$

Definition 6.1.8 (Exponential probability distribution function). A random variable Z distributed exponential with parameter $\lambda > 0$ has probability distribution function

$$f(z; \lambda) = \lambda \exp(-\lambda z) \text{ for } z > 0.$$

Definition 6.1.9 (Exponential cumulative distribution function). A random variable Z distributed exponential with parameter $\lambda > 0$ has cumulative distribution function

$$F(z; \lambda) = 1 - \exp(-\lambda z) \text{ for } z > 0.$$

Result 6.1.2 (Non-centered p-th moment of an exponentially distributed random variable). If Z is distributed as an exponential random variable with parameter $\lambda > 0$ then

$$\mathbb{E}[Z^p] = \lambda^{-p} \cdot \Gamma(p+1).$$

Proof. Substitute for Definition (6.1.8)

$$
\begin{aligned}
\mathbb{E}\left[Z^p\right] &= \int_0^\infty z^p \cdot f\left(z; \lambda, \beta\right) \mathrm{d}x \\
&= \int_0^\infty z^p \cdot \lambda \exp\left(-\lambda z\right) \mathrm{d}z.
\end{aligned}
$$

Let $t = \lambda z$ so that $\mathrm{d}t = \lambda \mathrm{d}z$ and make use of Definition 6.1.1 to write

$$
\lambda^{-p} \int_0^\infty t^p \exp\left(-t\right) \mathrm{d}t = \lambda^{-p} \cdot \Gamma\left(p+1\right).
$$

$\square$

Result 6.1.3 (Non-centered p-th moment of a beta distributed random variable). If S is a beta distributed random variables with parameters $\alpha > 0$ and $\beta > 0$ then

$$
\mathbb{E}\left[S^p\right] = \frac{B\left(\alpha + p, \beta\right)}{B\left(\alpha, \beta\right)}.
$$

where B $(\cdot, \cdot)$ is the beta function (Definition 6.1.2). When $y = 1$, $\mathbb{E}\left[S\right] = \frac{\alpha}{\alpha+\beta}$.

Proof. Recall that $f\left(s; \alpha, \beta\right)$ is the probability density function of a beta distributed random variable (Definition 6.1.5). Substitute for Definition 6.1.5 and calculate directly

$$
\begin{aligned}
\mathbb{E}\left[S^p\right] &= \int_0^1 s^p \cdot f\left(s; \alpha, \beta\right) \mathrm{d}s \\
&= \frac{B\left(\alpha + p, \beta\right)}{B\left(\alpha, \beta\right)} \cdot \int_0^1 f\left(\alpha + p, \beta\right) \mathrm{d}s \\
&= \frac{B\left(\alpha + p, \beta\right)}{B\left(\alpha, \beta\right)}.
\end{aligned}
$$

$\square$

Result 6.1.4 (Poisson-beta mixture distribution). If random R is distributed Poisson with parameter $\mu \cdot s$ where $\mu > 0$ and $s \in (0,1)$ is distributed as a beta (α, β). Then

$$\begin{aligned}
\Pr(R = r) &= \int_0^1 \Pr(Z = r \mid \mu u) \cdot f(u; \alpha, \beta)\, \mathrm{d}u \\
&= \int_0^1 \frac{\exp(-\mu u)(\mu u)^r}{r!} \cdot \frac{u^{\alpha-1}(1-u)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} \mathrm{d}u \\
&= \frac{\mu^r / r!}{\mathrm{B}(\alpha,\beta)} \int_0^1 \exp(-\mu u) \cdot u^{\alpha+r-1}(1-u)^{\beta-1}\, \mathrm{d}u.
\end{aligned}$$

Next use Definition 6.3.3 with $z = -u$, $a = \alpha + r$, $b = \alpha + \beta + r$ to rewrite the integral using Kummer's function

$$\begin{aligned}
\frac{\mu^r}{r!} \frac{\mathrm{B}(\alpha+r,\beta)}{\mathrm{B}(\alpha,\beta)} \mathcal{M}_{\alpha+r,\alpha+\beta+r}(-\mu) &= \frac{\exp(-\mu)\mu^r}{r!} \cdot \frac{\mathrm{B}(\alpha+r,\beta)}{\mathrm{B}(\alpha,\beta)} \cdot \mathcal{M}_{\alpha+r,\alpha+\beta+r}(-\mu) \\
&= \frac{\exp(-\mu)\mu^r}{r!} \cdot \frac{\mathrm{B}(\alpha+r,\beta)}{\mathrm{B}(\alpha,\beta)} \cdot \mathcal{M}_{\beta,\alpha+\beta+r}(\mu) \\
&= \mathrm{P}(r \mid \mu) \cdot \mathbb{E}[s^r] \cdot \mathcal{M}_{\beta,\alpha+\beta+r}(\mu).
\end{aligned}$$

where the second line uses the special property of Kummer functions called Kummer's transformation which states that $\mathcal{M}(a,b,z) = \exp(z)\,\mathcal{M}(b-a,b,z)$ and the last line substitute for the Poisson probability mass function (Definition 6.1.7) and using the formula for the non-centered p-th moment of a beta (α,β) random variable (Definition 6.1.3).

Definition 6.1.10 (Parametric model for hijacking intensity). The intensity of hijacking λ_k defined on page 12 of Hsu and Quandt (1976) is given by

$$\lambda_k = \begin{cases} a + \big[(1 - \cos g_1(k))/2\big](b - a) & \text{for } 0 < k \leq K_2 \\ \\ b + \big[(1 - \cos g_2(k))/2\big](c - b) & \text{for } K_2 + \frac{\pi}{\theta_1} < k \leq n \end{cases}$$

where n is the end of the sample and where for $i = 1, 2$

$$g_i(k) = \begin{cases} \theta_i(k - K_i) & \text{if } 0 \leq \theta_i(k - K_i) \leq \pi \\ \\ 0 & \text{if } \theta_i(k - K_i) < 0 \\ \\ \pi & \text{if } \pi < \theta_i(k - K_i) \end{cases}$$

and $K_1 + \frac{\pi}{\theta_1} \leq K_2$.

6.2 Appendix B: Poisson-beta and Bernoulli-beta Mixtures

The sequence hazard model is based on a mixture distribution. Classic references for compound Poisson distributions are Gurland (1958) and Katti (1966). See also Karlis and Xekalaki (2005). Let Z be a random variable distributed as Poisson with parameter $\mu \cdot s$ (Definition 6.1.7 of Appendix A) with $\mu > 0$ and $S \sim \text{beta}\,(\alpha, \beta)$ (Definition 6.1.5 of Appendix A) then the probability mass function of Z is given by Result 6.4.9 in the Appendix (see also pages 1094-1095 of Sarabia and Gómez-Déniz (2011) for an alternative formulation):

$$\Pr(Z = z) = \text{P}(z \,|\, \mu) \cdot \mathbb{E}\left[S^z\right] \cdot \mathcal{M}_{\beta, \alpha+\beta+z}(\mu) \;\; \text{for } z = 0, 1, 2, \ldots \tag{6.1}$$

where $\mathcal{M}_{a,b}(\cdot)$ is Kummer's function (Definition 3.2.4), $\mathbb{E}[S^p] = \mathrm{B}(\alpha + p, \beta) / \mathrm{B}(\alpha, \beta)$ is the p-th moment of a beta (α, β) distributed random variable where $\mathrm{B}(\alpha, \beta)$ is the Beta function (Definition 6.1.2), and $\mathrm{P}(\cdot \mid \mu)$ is the probability mass function of the Poisson distribution (Definition 6.1.7). The expectation and variance of Z are

$$\begin{aligned}
\mathbb{E}[Z] &= \mu \cdot \mathbb{E}[S] = \mu \cdot \frac{\alpha}{\alpha+\beta} \\
\mathrm{Var}[Z] &= \mu \cdot \mathbb{E}[S] + \mu^2 \cdot \mathrm{Var}[S] = \mu \cdot \frac{\alpha}{\alpha+\beta} + \mu^2 \cdot \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}$$

In the case of a Bernoulli-beta with parameter $\mu \in (0,1)$ (as in the model of attack success/-failure) the probability mass function for $Z = 0, 1$ is given by

$$\Pr(Z = z) = (\mu \cdot \mathbb{E}[S])^{\mathbf{1}(z=1)} \cdot (1 - \mu \cdot \mathbb{E}[S])^{\mathbf{1}(z=0)} \text{ for } z = 0, 1$$

and in this case the expectation of variance of Z are

$$\begin{aligned}
\mathbb{E}[Z] &= \mu \cdot \mathbb{E}[S] \\
\mathrm{Var}[Z] &= \mu \cdot \mathbb{E}[S] \cdot (1 - \mu \cdot \mathbb{E}[S]).
\end{aligned}$$

6.3 Appendix C: Proofs for Section 3.2.1

Result 6.3.1 (Joint probability distribution). Let U and V be independent random variables, one with probability density function $f_U(u)$ and the other with cumulative density function $f_V(v)$.

Then

$$\Pr\left(U < V\right) = \int \left[1 - F_V\left(u\right)\right] f_U\left(u\right) \mathrm{d}u.$$

Proof. Note that $P\left(U < V\right) = 1 - P\left(U \geq V\right) = 1 - P\left(V \leq U\right)$. Calculate directly that

$$\begin{aligned} \Pr\left(V \leq U\right) &= \int \Pr\left(V \leq U \,|\, U = u\right) f_U\left(u\right) \mathrm{d}u \\ &= \int \Pr\left(V \leq u\right) f_U\left(u\right) \mathrm{d}u \\ &= \int F_V\left(u\right) f_U\left(u\right) \mathrm{d}u \end{aligned}$$

where the second line follows from the independence of U and V. $\square$

Result 6.3.2 (Shifted exponential distribution). Let Z_1 be an exponential distributed random variable with parameter $\lambda > 0$ and define $Z_2 = \frac{u - Z_1}{v}$ for $u > 0$ and $v > 0$ then Z_2 has cdf

$$F\left(z_2; \lambda, u, v\right) = \begin{cases} \exp\left(-\lambda\left(u - z_2 v\right)\right) & \text{for } z_2 \in \left(-\infty, u/v\right) \\ 1 & \text{else} \end{cases}$$

Proof. Write $z_1\left(z_2\right) = u - z_2 v$ with $\left|\frac{\partial z_1\left(z_2\right)}{\partial z_2}\right| = v > 0$ and note that $z_1 > 0$ imples $z_2 \in \left(-\infty, u/v\right)$. Recall from Definition 6.1.8 that Z_1 has probability distribution function $f\left(z_1; \lambda\right) = \lambda \exp\left(-\lambda z_1\right)$

so

$$
\begin{aligned}
f(z_2; \lambda, u, v) &= f(z_1(z_2); \lambda) \cdot \left| \frac{\partial z_1(z_2)}{\partial z_2} \right| \\
&= \lambda \exp(-\lambda(u - z_2 v)) \cdot v \\
&= \lambda b \exp(-\lambda(u - z_2 v)).
\end{aligned}
$$

Integrating $f(\cdot; \lambda, u, v)$ from $-\infty$ to z_2 gives

$$
\begin{aligned}
F(z_2; \lambda, u, v) &= \int_{-\infty}^{z_2} \lambda v \exp(-\lambda(u - vt)) \, \mathrm{d}t \\
&= \exp(-\lambda(u - vt)) \Big|_{-\infty}^{z_2} \\
&= \exp(-\lambda(u - z_2 v)) \ \text{ if } z_2 \in (-\infty, u/v)
\end{aligned}
$$

and $F(z_2; \lambda, u, v) = 1$ else. □

Result 6.3.3 (Integral representation of Kummer's function). For $\beta > \alpha > 0$ Kummer's function $\mathcal{M}(\cdot)$ (Definition 3.2.4) can be represented by

$$
M(\alpha, \beta, t) = \frac{\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta - \alpha)} \int_0^1 \exp(tu)\, u^{\alpha - 1} (1 - u)^{\beta - \alpha - 1} \, \mathrm{d}u. \tag{6.2}
$$

See 13.2.1 on page 505 of Abramowitz et al. (1964).

Result 6.3.4 (Integral representation of Humbert Type 1 function). For $\gamma > \alpha > 0$ and $|u| < 1$ the

Humbert type 1 function $\Phi_1(\cdot)$ (Definition 3.2.5) can be represented by

$$\Phi_1(\alpha,\beta,\gamma,u,v) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)} \int_0^1 \exp(vt)\, t^{\alpha-1}(1-t)^{\gamma-\alpha-1}(1-ut)^{-\beta}\,\mathrm{d}t.$$

See page 79 of Humbert (1922).

Result 6.3.5 (Moment generating function of a beta distributed random variable). The moment generating function of the beta (α,β) random variable S with $\alpha > 0$ and $\beta > 0$ (Definition 6.1.5) is given by

$$\mathbb{E}[\exp(S\cdot t)] = \mathcal{M}_{\alpha,\alpha+\beta}(t)$$

where $\mathcal{M}_{\cdot,\cdot}(\cdot,\cdot)$ is Kummer's confluent hypergeometric function (Definition 3.2.4).

Proof. Apply Definition 6.3.3 above substituting $\beta' = \beta+\alpha$ for β. □

Result 6.3.6 (Integral result 1). Let $f(s;\alpha,\beta)$ denote the probability distribution function of a random variable distributed beta (α,β) on $(0,1)$ with $\alpha > 0$ and $\beta > 0$ and let $F(y;\lambda,u,v)$ denote the cdf of a shifted exponential distribution derived in Result 6.3.2. Then

$$\int_0^1 F(t;\lambda,u,v)\cdot f(t;\alpha,\beta)\,\mathrm{d}t = \lambda^{-1} f(u;\lambda)\,\mathcal{M}_{\alpha,\alpha+\beta}(\lambda v)$$

Proof. Substituting directly for the definitions of $F(\cdot;\lambda,u,v)$ and $f(\cdot;\alpha,\beta)$ gives

$$\begin{aligned}\int_0^1 \exp(-\lambda(u-vt))\cdot\frac{t^{\alpha-1}(1-t)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}\mathrm{d}t &= \frac{\exp(-\lambda u)}{\mathrm{B}(\alpha,\beta)}\cdot\int_0^1 \exp(\lambda tv)\,t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t\\ &= \exp(-\lambda u)\cdot\mathcal{M}_{\alpha,\alpha+\beta}(\lambda v)\\ &= \lambda^{-1}f(u;\lambda)\cdot\mathcal{M}_{\alpha,\alpha+\beta}(\lambda v)\end{aligned}$$

where the second line uses the integral representation in Result 6.3.3 substituting $\beta'=\beta+\alpha$ for β and where the last line substitutes for the exponential probability density function (Definition 6.1.8 in Appendix A). □

Result 6.3.7 (Integral result 2). Let $f(s;\alpha,\beta)$ denote the probability distribution function of a random variable distributed beta (α,β) on $(0,1)$ with $\alpha>0$ and $\beta>0$. Recall that $F(\cdot;\lambda,u,v)$ is the cdf of a shifted exponential distribution (see Definition 6.3.2 in Appendix A). For $r\in(0,1)$

$$\int_0^r F(t;\lambda,u,v)\cdot f(t;\alpha,\beta)\,\mathrm{d}t = \lambda^{-1}f(u;\lambda)\,\mathcal{H}_{\alpha,\beta}(r,vr)$$

where $\mathcal{H}_{\alpha,\beta}(\cdot,\cdot)$ is the Regularized Type 1 Humbert Series (Definition 3.2.6).

Proof. Change variables using $p=t/r$ so that $\mathrm{d}t=r\mathrm{d}p$ and $t=0$ implies $p=0$ while $t=r$ implies that $p=1$. Directly applying Definition 6.3.4 gives

$$\begin{aligned}\frac{u^\alpha}{\mathrm{B}(\alpha,\beta)}\int_0^1 \exp(qrp)\,p^{\alpha-1}(1-rp)^{\beta-1}\,\mathrm{d}p &= \frac{u^\alpha}{\mathrm{B}(\alpha,\beta)}\frac{\Gamma(\alpha)\Gamma(1)}{\Gamma(\alpha+1)}\Phi_1(\alpha,1-\beta,\alpha+1;u,qu)\\ &= \frac{1}{\mathrm{B}(\alpha,\beta)}\frac{u^\alpha}{\alpha}\Phi_1(\alpha,1-\beta,\alpha+1;u,qu)\,.\end{aligned}$$

□

Result 6.3.8 (Derivation of the CCP). Let $s \sim$ beta$\left(\alpha, \beta\right)$ and $\varepsilon \sim$ exponential$\left(\lambda\right)$ and assume that $u > 0$ and $v > 0$. Let $Y \equiv \frac{u-\varepsilon}{v}$ and consider the probability $\Pr\left(s \geq \frac{u-\varepsilon}{v}\right) = 1 - \Pr\left(s < Y\right)$. Directly applying Result 6.3.1 to evaluate this probability gives

$$1 - \int_0^1 \left[1 - F\left(t; \lambda, u, v\right)\right] \cdot f\left(t; \alpha, \beta\right) \mathrm{d}t = \int_0^1 F\left(t; \lambda, u, v\right) \cdot f\left(t; \alpha, \beta\right) \mathrm{d}t$$

which is the expectation $\mathbb{E}_s\left[F\left(s; \lambda, u, v\right)\right]$ taken with respect to beta-distributed random variable S. Substituting for the cdf of Z derived in Result 6.3.2 and noting that $F\left(z; \lambda, u, v\right) = 1$ when $z \geq u/v$ then

$$\mathrm{P}\left(s \geq Z\right) = \begin{cases} \int_0^{u/v} F\left(t; \lambda, u, b\right) \cdot f\left(t; \alpha, \beta\right) \mathrm{d}t + \int_{u/v}^1 f\left(t; \alpha, \beta\right) \mathrm{d}z & \text{if } u/v \in (0,1) \\ \int_0^1 F\left(t; \lambda, u, v\right) \cdot f\left(t; \alpha, \beta\right) \mathrm{d}t & \text{if } u/v \geq 1 \end{cases}$$

$$= \begin{cases} 1 - F_{\alpha,\beta}\left(u/v\right) + \lambda^{-1} f\left(u; \lambda\right) \mathcal{H}_{\alpha,\beta}\left(u/v, \lambda u\right) & \text{if } u/v \in (0,1) \\ \lambda^{-1} f\left(u; \lambda\right) \mathcal{M}_{\alpha,\alpha+\beta}\left(\lambda v\right) & \text{if } u/v \geq 1 \end{cases}$$

where $F\left(\cdot; \alpha, \beta\right)$ is the cdf of the beta distribution (Definition 6.1.6), $\mathcal{M}_{\alpha,\alpha+\beta}\left(\cdot\right)$ is Kummer's confluent hypergeometric function (Definition 3.2.4) and $\mathcal{H}_{\alpha,\beta}\left(\cdot,\cdot\right)$ is the Type 1 Humbert Series (Definition 3.2.5).

Result 6.3.9 (Derivative of the CCP). Let $u\left(\gamma\right) > 0$ and $v\left(\gamma\right) > 0$ denote two scalar functions of γ. The CCP derived in Result 6.3.8 has a derivative with respect to γ that is increasing/constant/decreasing according to

$$\frac{\mathrm{dP}\left(s \geq Y\right)}{\mathrm{d}\gamma} \lesseqqgtr 0 \iff \begin{cases} \frac{\mathcal{H}_{\alpha+1,\beta}(u/v,\lambda u)}{\mathcal{H}_{\alpha,\beta}(u/v,\lambda u)} \cdot \mathbb{E}\left[s\right] \lesseqqgtr \frac{u'}{v'} & \text{if } u/v \in (0,1) \\ \frac{CCP_{\alpha+1}}{CCP_{\alpha}} \cdot \mathbb{E}\left[s\right] \lesseqqgtr \frac{u'}{v'} & \text{if } u/v \geq 1 \end{cases}$$

where $\mathbb{E}\left[S\right] = \alpha/\alpha + \beta$, where $CCP_{\alpha+1}$ denotes the CCP evaluated at $\alpha + 1$ instead of α, where $\mathcal{H}_{\cdot,\beta}\left(\cdot,\cdot\right)$ denotes the Humbert type 1 function (Definition 3.2.6), and where u' and v' denote the derivatives of functions u and v with respect to γ.

Proof. If $u/v > 1$, write the CCP as

$$\int_0^1 F\left(t;\lambda,u,v\right) \cdot f\left(t;\alpha,\beta\right) \mathrm{d}t$$

and take derivative directly substituting first for the shifted exponential cdf from Definition 6.3.2

$$\begin{aligned} \frac{\partial CCP}{\partial \gamma} &= \int_0^1 \frac{\partial F\left(t;\lambda,u,v\right)}{\partial \gamma} \cdot f\left(t;\alpha,\beta\right) \mathrm{d}t \\ &= \int_0^1 -\lambda\left(a' - tv'\right) \exp\left(-\lambda\left(a - tv\right)\right) \cdot f\left(t;\alpha,\beta\right) \mathrm{d}t \\ &= \lambda b' \int_0^1 F\left(t;\lambda,u,v\right) \cdot f\left(t;\alpha+1,\beta\right) \mathrm{d}t \cdot \frac{B\left(\alpha+1,\beta\right)}{B\left(\alpha,\beta\right)} - \lambda u' \int_0^1 F\left(t;\lambda,u,v\right) \cdot f\left(t;\alpha,\beta\right) \mathrm{d}t \\ &= \lambda\Big[v' CCP_{\alpha+1} \cdot \mathbb{E}\left[S\right] - u' CCP_{\alpha}\Big] \end{aligned}$$

where the third line uses the integral representation from Result 6.3.6 and the last line substitutes for the expectation of a beta (α, β) random variable. The duration dependence is increasing/decreasing/constant if and only if

$$\frac{CCP_{\alpha+1}}{CCP_{\alpha}} \cdot \mathbb{E}\left[S\right] \lesseqqgtr u'/v'.$$

Proceeding similarly for $u/v \in (0,1)$, write the CCP as

$$\int_0^{u/v} F\left(t;\lambda,u,v\right) \cdot f\left(t;\alpha,\beta\right) \mathrm{d}t + \int_{u/v}^1 f\left(t;\alpha,\beta\right) \mathrm{d}t$$

and taking the derivative with respect to γ,

$$\begin{aligned}
\frac{\partial CCP}{\partial \gamma} &= \left(u/v\right)' F\left(u/v;\lambda,u,v\right) \cdot f\left(u/v;\alpha,\beta\right) \\
&- \int_0^{u/v} \lambda\left(u' - v't\right) \exp\left(-\lambda\left(u - vt\right)\right) \cdot f\left(t;\alpha,\beta\right) \mathrm{d}t \\
&- \left(u/v\right)' \cdot f\left(u/v;\alpha,\beta\right) \\
&= \lambda v' \int_0^{u/v} F\left(t;\lambda,u,v\right) \cdot f\left(t;\alpha+1,\beta\right) \mathrm{d}t \cdot \frac{B\left(\alpha+1,\beta\right)}{B\left(\alpha,\beta\right)} \\
&- \lambda u' \int_0^{u/v} F\left(t;\lambda,u,v\right) \cdot f\left(t;\alpha,\beta\right) \mathrm{d}t \\
&= f\left(u;\lambda\right) \cdot \Big[v'\mathcal{H}_{\alpha+1,\beta}\left(u/v,\lambda u\right) \cdot \mathbb{E}\left[S\right] - u'\mathcal{H}_{\alpha,\beta}\left(u/v,\lambda u\right)\Big]
\end{aligned}$$

where the third line uses the integral representation from Result 6.3.7 and the last line substitutes for the pdf of the exponential distribution with parameter λ and the expectation of a beta (α,β) random variable. The duration dependence is increasing/constant/decreasing if and only if

$$\frac{\mathcal{H}_{\alpha+1,\beta}\left(u/v,\lambda u\right)}{\mathcal{H}_{\alpha,\beta}\left(u/v,\lambda u\right)} \cdot \mathbb{E}\left[S\right] \lesseqqgtr u'/v'.$$

□

6.4 Appendix D: Proofs for Section 3.2.2

Result 6.4.1 (Law of total variance). Let U and V be random variables defined on the same probability space and assume V has finite variance. The variance of V can be written as

$$\text{Var}\left[V\right] = \mathbb{E}\left[\text{Var}\left[V \mid U\right]\right] + \text{Var}\left[\mathbb{E}\left[V \mid U\right]\right].$$

Proof.

$$\begin{aligned}
\text{Var}\left[V\right] &= \mathbb{E}\left[V^2\right] - \left(\mathbb{E}\left[V\right]\right)^2 \\
&= \mathbb{E}\left[\mathbb{E}\left[V^2 \mid U\right]\right] - \left(\mathbb{E}\left[\mathbb{E}\left[V \mid U\right]\right]\right)^2 \\
&= \mathbb{E}\left[\mathbb{E}\left[V^2 \mid U\right] - \left(\mathbb{E}\left[V \mid U\right]\right)^2\right] + \mathbb{E}\left[\mathbb{E}\left[V \mid U\right]\right]^2 - \left[\mathbb{E}\left[\mathbb{E}\left[V \mid U\right]\right]\right]^2 \\
&= \mathbb{E}\left[\text{Var}\left[V \mid U\right]\right] + \text{Var}\left[\mathbb{E}\left[V \mid U\right]\right].
\end{aligned}$$

□

Result 6.4.2 (Integral representation 3). Let $f\left(z; \alpha, \beta\right)$ denote the probability distribution function of a beta $\left(\alpha, \beta\right)$ distributed random variable (Definition 6.1.5). Then

$$\int_r^1 \exp\left(-\mu z\right) \cdot f\left(z; \alpha, \beta\right) \mathrm{d}z = \exp\left(-\mu\right) \cdot \mathcal{H}_{\beta,\alpha}\left(1 - r, \mu\left(1 - r\right)\right)$$

Proof. Change variables using $t = \frac{1-z}{1-r}$ so $z = r$ implies $t = 1$ and $z = 1$ implies $t = 0$. Thus

$(1-r)\,t = 1-z$ or $z = 1-(1-r)\,t$ so $\mathrm{d}z = -(1-r)\,\mathrm{d}t$ and now write the integral as

$$\begin{aligned}
&\int_0^1 \exp\left(-\mu\left[1-(1-r)\,t\right]\right)\left(1-(1-r)\,t\right)^{\alpha-1}\left((1-r)\,t\right)^{\beta-1} \times -(1-r)\,\mathrm{d}t/B(\alpha,\beta) \\
&\qquad = \frac{(1-r)^{\beta}}{B(\alpha,\beta)} \cdot \exp(-\mu) \int_0^1 \exp\left(\mu\,(1-r)\,t\right) t^{\beta-1} \times \left(1-(1-r)\,t\right)^{\alpha-1} \mathrm{d}t
\end{aligned}$$

Now use $a=\beta$, $-b=\alpha-1$, $c=a+1$, $x=(1-r)$, $t=t$, $y=\mu\,(1-r)$ to get

$$\begin{aligned}
&\frac{(1-r)^{\beta}}{B(\alpha,\beta)} \cdot \frac{\Gamma(\beta)\,\Gamma(1)}{\Gamma(\beta+1)} \exp(-\mu)\,\Phi_1\left(\beta, 1-\alpha, \beta+1, 1-r, \mu\,(1-r)\right) \\
&= \exp(-\mu) \cdot \frac{(1-r)^{\beta}}{\beta} \cdot \frac{1}{B(\alpha,\beta)} \Phi_1\left(\beta, 1-\alpha, \beta+1; 1-r, \mu\,(1-r)\right) \\
&= \exp(-\mu) \cdot \mathcal{H}_{\beta,\alpha}\left(1-r, \mu\,(1-r)\right).
\end{aligned}$$

□

Result 6.4.3 (Conditional pdf of a truncated exponentially distributed random variable). Let $\varepsilon \sim \text{exponential}(\lambda)$. Recall the definition of $\mathrm{P}(a=1)$ derived in Result 6.3.8 of Appendix B. The conditional pdf of ε given $a=1$ is

$$f_{\varepsilon\,|\,a=1}(\varepsilon) = \frac{f(\varepsilon;\lambda)}{\mathrm{P}(a=1)} \text{ for } (\underline{\varepsilon}, \infty)$$

where $\underline{\varepsilon} = -\lambda^{-1} \log\left[\mathrm{P}(a=1)\right]$ and $f(\varepsilon;\lambda)$ is the pdf of the exponential distribution with parameter $\lambda > 0$ (Definition 6.1.8).

Proof. Working directly with the cdf method gives $F_{\varepsilon\,|\,a=1}(\nu) = \Pr(\varepsilon \le \nu \,|\, a=1)$. Recall that the event $a = 1$ implies that for $s \sim \text{beta}(\alpha, \beta)$ and $u > 0$ and $v > 0$ that $s \geq \frac{u-\varepsilon}{v}$ so

$$\begin{aligned}
\Pr(\varepsilon \leq \nu \,|\, a = 1) &= \Pr(\varepsilon \leq \nu \,|\, \varepsilon \geq u - sv) \\
&= \int_0^1 \frac{\Pr(u - sv \leq \varepsilon \leq \nu)}{P(a = 1)} f(s; \alpha, \beta)\, \mathrm{d}s \\
&= P(a = 1)^{-1} \int_0^1 [F(\nu; \lambda) - F(a - sv; \lambda)]\, f(s; \alpha, \beta)\, \mathrm{d}s \\
&= P(a = 1)^{-1} \int_0^1 F(\nu; \lambda)\, f(s; \alpha, \beta)\, \mathrm{d}s \\
&- P(a = 1)^{-1} \int_0^1 [1 - \exp(-\lambda(u - sv))]\, f(s; \alpha, \beta)\, \mathrm{d}s \\
&= P(a = 1)^{-1} F(\nu; \lambda) - P(a = 1)^{-1} \int_0^1 [1 - F(s; \lambda, u, v)]\, f(s; \alpha, \beta)\, \mathrm{d}s \\
&= P(a = 1)^{-1} F(\nu; \lambda) - \frac{P(a = 0)}{P(a = 1)} \text{ for } \nu \in (\underline{\varepsilon}, \infty)
\end{aligned}$$

where $\underline{\varepsilon} = -\lambda^{-1} \log \mathrm{P}[(a = 1)]$ is found by evaluating the cdf at the endpoint and setting it equal to 0. Note that $F_{\varepsilon\,|\,a=1}(\underline{\varepsilon}) = 0$ and $F_{\varepsilon\,|\,a=1}(\nu) \to 1$ as $\nu \to \infty$. Taking the derivative of the last line with respect to ν gives the result. Note as well that

$$f_{\varepsilon\,|\,a=0}(\varepsilon) = \frac{f(\varepsilon; \lambda)}{\mathrm{P}(a = 0)} \text{ for } (0, \underline{\varepsilon})$$

where $\underline{\varepsilon} = -\lambda^{-1} \log[\mathrm{P}(a = 1)]$ and $f(\varepsilon; \lambda)$ is the pdf of the exponential distribution with parameter $\lambda > 0$ (Definition 6.1.8). □

Result 6.4.4 (Conditional pdf of a truncated beta distributed random variable). Let $s \sim \text{beta}(\alpha, \beta)$ and recall the definition of $\mathrm{P}(a = 1)$ from Result 6.3.8 in Appendix B or Result 6.3.8 in the main text. The conditional pdf of s given event $a = 1$ is

$$f_{s\,|\,a=1}(s) = \frac{f(s;\alpha,\beta)}{\mathrm{P}(a=1)} \text{ for } (\underline{s},\infty)$$

where $\underline{s} = F^{-1}(\mathrm{P}(a=1);\alpha,\beta)$, $F^{-1}(\cdot;\alpha,\beta)$ is the inverse cdf, and $f(\cdot;\alpha,\beta)$ is the pdf of the beta distribution with parameters $\alpha > 0$ and $\beta > 0$ (Definition 6.1.5).

Proof. Recall the event $a = 1$ implies that for s and $\varepsilon \sim$ exponential (λ) with parameter $\lambda > 0$ that $s \geq \frac{u-\varepsilon}{v}$. Define $Z = \frac{u-\varepsilon}{v}$ and recall Z has cdf and pdf $F(\cdot;\lambda,u,b)$ and $f(\cdot;\lambda,u,v)$, respectively, from Result 6.3.2. Using the cdf method to calculate directly gives

$$\begin{aligned}
F_{s\,|\,a=1}(s) &= \mathrm{Pr}(S \leq s \,|\, a=1) \\
&= \mathrm{Pr}\left(S \leq s \,|\, s \geq \frac{u-\varepsilon}{v}\right) \\
&= \mathrm{Pr}(S \leq s \,|\, S \geq Z) \\
&= \int_{-\infty}^{u/v} \mathrm{Pr}(z \leq S \leq s) f(z;\lambda,u,v)\,\mathrm{d}z/P(a=1) \\
&= \int_{-\infty}^{u/v} [F(s;\alpha,\beta) - F(z;\alpha,\beta)] f(z;\lambda,u,b)\,\mathrm{d}z/P(a=1) \\
&= \int_{-\infty}^{u/v} \frac{F(s;\alpha,\beta) f(z;\lambda,u,v)\,\mathrm{d}z}{P(a=1)} - \int_{-\infty}^{u/v} F(z;\alpha,\beta) f(z;\lambda,u,v)\,\mathrm{d}z/P(a=1) \\
&= \frac{F(s;\alpha,\beta)}{P(a=1)} - \frac{P(a=0)}{P(a=1)} \text{ for } s \in (\bar{s},1)
\end{aligned}$$

where $\bar{s} = F^{-1}(1 - \mathrm{P}(a=1);\alpha,\beta)$. Note that $F_{s\,|\,a=1}(\bar{s}) = 0$ and $F_{s\,|\,a=1}(1) = 1$. Taking the derivative of the last line with respect to s gives the result. $\square$

Result 6.4.5 (Useful relationship). The following relationship holds between the regularized Humbert type 1 function $\mathcal{H}_{a,b}(\cdot,\cdot)$ (Definition 3.2.6) and the survivor function for the beta (α,β)

distribution

$$\overline{\mathrm{F}}(s;\alpha,\beta)=\mathcal{H}_{\beta,\alpha}(1-s,0)$$

where $\overline{\mathrm{F}}(s;\alpha,\beta)\equiv 1-F(s;\alpha,\beta)$.

Proof. Let $f(s;\alpha,\beta)$ represent the probability density function of a beta (α,β) random variable. Working directly with the definition of a survivor function gives that for $u\in(0,1)$,

$$\overline{\mathrm{F}}(u;\alpha,\beta)=\int_u^1 f(s;\alpha,\beta)\,\mathrm{d}s=\int_u^1 \frac{s^{\alpha-1}(1-s)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}s.$$

Change variables using $r=\frac{1-s}{1-u}=\frac{1-s}{\bar{u}}$ where $\bar{u}=1-u$. Then $r(u)=1$ and $r(1)=0$, $s=1-r\bar{u}$, and $\mathrm{d}s=-\bar{u}\mathrm{d}r$. Hold the $\mathrm{B}(\alpha,\beta)$ term and calculate directly to get

$$-\bar{u}\int_0^1 (1-r\bar{u})^{\alpha-1}(r\bar{u})^{\beta-1}\,\mathrm{d}r.$$

Next use the integral representation in Definition 6.3.4 and set $a=\beta$, $b=1-\alpha$, $c=\beta+1$, $x=\bar{u}$, $y=0$, and bring back the $\mathrm{B}(\alpha,\beta)$ to get

$$\begin{aligned}\frac{\bar{u}^{\beta}}{\mathrm{B}(\alpha,\beta)}\cdot\frac{\Gamma(\beta)\Gamma(1)}{\Gamma(\beta+1)}\cdot\Phi_1(\beta,1-\alpha,\beta+1;\bar{u},0) &= \frac{1}{\mathrm{B}(\alpha,\beta)}\cdot\frac{\bar{u}^{\beta}}{\beta}\cdot\Phi_1(\beta,1-\alpha,\beta+1;\bar{u},0)\\ &=\mathcal{H}_{\beta,\alpha}(1-u,0).\end{aligned}$$

□

Result 6.4.6 (Moments of a truncated exponentially distributed random variable). Let $\varepsilon\sim$

exponential (λ). The truncated mean of ε is

$$\int_r^\infty \varepsilon^z f(\varepsilon;\lambda)\,\mathrm{d}\varepsilon = \frac{(\lambda r+1)^z}{\lambda^{z+1}} \cdot f(r;\lambda)$$

where $f(r;\lambda) = \lambda \exp(-r\lambda)$ is the pdf of the exponential distribution (Definition 6.1.8).

Proof.

$$\int_r^\infty \varepsilon^z f(\varepsilon)\,\mathrm{d}\varepsilon = \int_r^\infty \varepsilon^z \lambda \exp(-\lambda\varepsilon)\,\mathrm{d}\varepsilon.$$

Integrate this expression by parts using $u = \varepsilon^z \to \mathrm{d}u = z\varepsilon^{z-1}\mathrm{d}\varepsilon$ and $\mathrm{d}v = \lambda \exp(-\lambda\varepsilon)\,\mathrm{d}\varepsilon \to v = -\exp(-\lambda\varepsilon)$ to arrive at

$$-\varepsilon^z \exp(-\lambda\varepsilon)\Big|_r^\infty + \int_r^\infty z\varepsilon^{z-1} \exp(-\lambda\varepsilon)\,\mathrm{d}\varepsilon.$$

Integrate by parts again using $u = z\varepsilon^{z-1} \to \mathrm{d}u = z(z-1)\,\varepsilon^{z-2}\mathrm{d}\varepsilon$ and $\mathrm{d}v = \exp(-\lambda\varepsilon)\,\mathrm{d}\varepsilon \to v = -\lambda^{-1}\exp(-\lambda\varepsilon)$ to arrive at

$$-\varepsilon^z \exp(-\lambda\varepsilon)\Big|_r^\infty - \lambda^{-1} z\varepsilon^{z-1} \exp(-\lambda\varepsilon)\Big|_r^\infty + \int_r^\infty z(z-1)\,\varepsilon^{z-2}\lambda^{-1} \exp(-\lambda\varepsilon)\,\mathrm{d}\varepsilon.$$

Repeat this step a total z times, collecting the common factor $\exp(-\lambda r)$, and apply the Binomial theorem (Definition 6.1.1) after evaluating at the limits to get

$$
\begin{aligned}
\exp\left(-\lambda r\right)\sum_{j=0}^{z}\frac{\partial r^{z}}{\partial^{(j)}r}\lambda^{-j} &= \exp\left(-\lambda r\right)\sum_{j=0}^{z}\frac{z!}{j!\left(z-j\right)!}r^{z-j}\lambda^{-j} \\
&= \exp\left(-\lambda r\right)\lambda^{-z}\sum_{j=0}^{z}\frac{z!}{j!\left(z-j\right)!}\left(\lambda r\right)^{z-j}\left(1\right)^{j} \\
&= \exp\left(-\lambda r\right)\lambda^{-z}\sum_{j=0}^{z}\binom{z}{j}\left(\lambda r\right)^{z-j}\left(1\right)^{j} \\
&= \frac{\left(\lambda r+1\right)^{z}}{\lambda^{z+1}}\cdot f\left(r;\lambda\right)
\end{aligned}
$$

where the last lines multiplies by λ and substitutes for the pdf of the exponential distribution. $\square$

Result 6.4.7 (Conditional p-th moment of a truncated exponentially distributed random variable). Let $\varepsilon\sim$ exponential (λ) and recall that $\mathrm{P}\left(a=1\right)$ represents the decision to attack derived in Result 6.3.8. The conditional p-th moment of ε given $a=1$ is given by

$$
\mathbb{E}\left[\varepsilon^{p}\,|\,a=1\right]=\lambda^{-p}\cdot\left[1-\log\mathrm{P}\left(a=1\right)\right]^{p}.
$$

Proof. Calculate the conditional expectation directly using the probability density defined in Result 6.4.3

$$
\begin{aligned}
\mathbb{E}\left[\varepsilon^{p}\,|\,a=1\right] &= \int_{u}^{\infty}\varepsilon^{p}\cdot f_{\varepsilon\,|\,a=1}\left(\varepsilon\,|\,a=1\right)\mathrm{d}\varepsilon \\
&= \frac{1}{\mathrm{P}\left(a=1\right)}\int_{u}^{\infty}\varepsilon^{p}f\left(\varepsilon\right)\mathrm{d}\varepsilon \\
&= \frac{1}{\mathrm{P}\left(a=1\right)}\int_{u}^{\infty}\varepsilon^{p}\lambda\exp\left(-\lambda\varepsilon\right)\mathrm{d}\varepsilon
\end{aligned}
$$

where $u=-\lambda^{-1}\log\left(\mathrm{P}\left(a=1\right)\right)$. The integral above was derived in Result 6.4.6 of Appendix D, substituting directly gives

$$\frac{1}{\mathrm{P}(a=1)} \cdot \frac{(\lambda u+1)^p}{\lambda^{p+1}} \cdot f(u;\lambda)$$

and note that $f(u;\lambda) = \lambda \cdot \mathrm{P}(a=1)$ so

$$\begin{aligned}\frac{1}{\mathrm{P}(a=1)} \cdot \frac{(\lambda u+1)^p}{\lambda^{p+1}} \cdot \lambda \cdot \mathrm{P}(a=1) &= \frac{(\lambda u+1)^p}{\lambda^p} \\ &= \left(u+\lambda^{-1}\right)^p \\ &= \left(-\lambda^{-1}\log \mathrm{P}(a=1)+\lambda^{-1}\right)^p \\ &= \left(\lambda^{-1}-\lambda^{-1}\log \mathrm{P}(a=1)\right)^p \\ &= \lambda^{-p}\cdot(1-\log \mathrm{P}(a=1))^p .\end{aligned}$$

□

Result 6.4.8 (Conditional p-th moment of a truncated beta distributed random variable). Let $s \sim$ beta (α,β) and recall that $\mathrm{P}(a=1)$ represents the latent variable function derived in Result 6.3.8. The conditional p-th moment of s given $a=1$ is given by

$$\mathbb{E}\left[S^p \,|\, a=1\right] = \frac{M_p(\bar{s})}{P(a=1)}$$

where recall that $M_j(\gamma) = \mathbb{E}\left[S^j\right] \cdot \bar{F}(\gamma;\alpha+j,\beta)$, $\mathbb{E}\left[S^j\right] = \mathrm{B}(\alpha+j,\beta)/\mathrm{B}(\alpha,\beta)$ and $\bar{s} = F^{-1}(1-\mathrm{P}(a=1);\alpha,\beta)$ is the inverse of the beta cdf (Defintion 6.1.6) and $\overline{\mathrm{F}}(s;\alpha+k,\beta)$ is the survivor function for the beta $(\alpha+k,\beta)$ distribution. Using Definition 6.4.5 the expectation can also be written

$$\mathbb{E}\left[S^p \,|\, a=1\right] = \frac{\mathcal{H}_{\beta,\alpha+p}\left(\bar{s},0\right)}{P\left(a=1\right)} \cdot \mathbb{E}\left[S^p\right].$$

Proof. Working directly with the expectation

$$\begin{aligned}
\mathbb{E}\left[S^p \,|\, a=1\right] &= \int_0^1 s^p f_{s\,|\,a=1}\left(s \,|\, a=1\right) \mathrm{d}s \\
&= \frac{1}{P\left(a=1\right)} \frac{B\left(\alpha+p,\beta\right)}{B\left(\alpha,\beta\right)} \int_{\bar{s}}^1 f\left(s;\alpha+p,\beta\right) \mathrm{d}s \\
&= \frac{1}{P\left(a=1\right)} \cdot \mathbb{E}\left[S^p\right] \cdot \left(1 - F\left(\bar{s};\alpha+p,\beta\right)\right) \\
&= \frac{\bar{F}\left(\bar{s};\alpha+p,\beta\right)}{P\left(a=1\right)} \cdot \mathbb{E}\left[S^p\right]
\end{aligned}$$

and the result follows by substituting for the definition of $M_j\left(\gamma\right)$. □

Result 6.4.9 (Probability mass function for the Poisson-beta distribution). Let Y be a random variable distributed according to a Poisson-beta mixture with parameter $\mu = bs$ for $b \in (0,1)$ where $s \sim$ beta (α,β). The probability mass function for Y is

$$\mathrm{P}\left(Y=y\right) = \mathrm{P}\left(y \,|\, b\right) \cdot \mathbb{E}\left[S^y\right] \cdot \mathcal{M}_{\beta,\alpha+y}\left(b\right)$$

where $\mathbb{E}\left[S\right] = \text{Beta}\left(\alpha+y,\beta\right)/\text{Beta}\left(\alpha,\beta\right)$, $\mathrm{P}\left(y \,|\, b\right)$ is the probability mass function of a Poisson distribution, and $\mathcal{M}_{a,b}\left(\cdot\right)$ is Kummer's function (Definition 3.2.4).

Proof.

$$
\begin{aligned}
\mathrm{P}(Y=y) &= \int_0^1 \mathrm{P}(y \mid bu) \cdot f(u;\alpha,\beta)\,\mathrm{d}u \\
&= \int_0^1 \mathrm{P}(y \mid bu) \cdot f(u;\alpha,\beta)\,\mathrm{d}u \\
&= \int_0^1 \frac{\exp(-bu)\cdot(bu)^y}{y!} \cdot \frac{u^{\alpha-1}(1-u)^{\beta-1}}{B(\alpha,\beta)}\mathrm{d}u \\
&= \frac{b^y}{y!} \cdot \frac{B(\alpha+y,\beta)}{B(\alpha,\beta)} \int_0^1 \exp(-bu)\cdot f(u;\alpha+y,\beta)\,\mathrm{d}u \\
&= \frac{b^y}{y!} \cdot \mathbb{E}\left[S^y\right] \cdot \mathcal{M}_{\alpha+y,\alpha+\beta+y}(-b) \\
&= \frac{b^y}{y!} \cdot \mathbb{E}\left[S^y\right] \cdot \exp(-b) \cdot \mathcal{M}_{\beta,\alpha+\beta+y}(b) \\
&= \mathrm{P}(y \mid b) \cdot \mathbb{E}\left[S^y\right] \cdot \mathcal{M}_{\beta,\alpha+\beta+y}(b)
\end{aligned}
$$

where the line two before last applies the integral representation from Result 6.3.3 (or alternatively Result 6.3.5) and applies Kummer's transformation $\mathcal{M}_{a,b}(y) = \exp(y) \cdot \mathcal{M}_{b-a,b}(-y)$ before substituting for the probability mass function of a Poisson distribution. □

Result 6.4.10 (Joint outcome probability for the Poisson-beta sequence hazard model). Let Y be a random variable distributed according to a Poisson-beta mixture with parameter $\mu = bs$ for $b > 0$ where $s \sim$ beta (α, β). The joint probability mass function is

$$\mathrm{P}(Y=y, a=1) = \mathrm{P}(y \mid b) \cdot \mathbb{E}\left[S^y\right] \cdot \mathcal{H}_{\beta,\alpha+y}(\bar{s}, b\bar{\bar{s}})$$

where $\mathrm{P}(y \mid b)$ is the pmf of a Poisson distribution (Definition 6.1.7), $\mathbb{E}\left[S^y\right] = \mathrm{B}(\alpha+y,\beta)/\mathrm{B}(\alpha,\beta)$, and $\bar{\bar{s}} = 1-\bar{s}$ where $\bar{s} = F^{-1}(1-\mathrm{P}(a=1);\alpha,\beta)$ is the inverse of the beta cdf (Definition 6.1.6) evaluated at $1-\mathrm{P}(a=1)$.

Proof. In light of the definition of conditional probabilities

$$\mathrm{P}(Y = y, a = 1) = \mathrm{P}(Y = y \mid a = 1) \cdot \mathrm{P}(a = 1)$$

Calculating directly,

$$\begin{aligned}
\mathrm{P}(Y = y \mid a = 1) &= \int_0^1 \mathrm{P}(y \mid bu) \cdot f_{s \mid a=1}(u; \alpha, \beta)\, \mathrm{d}u \\
&= \mathrm{P}(a = 1)^{-1} \int_{\bar{s}}^1 \mathrm{P}(y \mid bu) \cdot f(u; \alpha, \beta)\, \mathrm{d}u \\
&= \mathrm{P}(a = 1)^{-1} \int_{\bar{s}}^1 \frac{\exp(-bu)(bu)^y}{y!} \cdot \frac{u^{\alpha-1}(1-u)^{\beta-1}}{B(\alpha, \beta)} \mathrm{d}u \\
&= \frac{b^y/y!}{\mathrm{P}(a = 1)} \cdot \frac{B(\alpha + y, \beta)}{B(\alpha, \beta)} \int_{\bar{s}}^1 \exp(-bu) \cdot f(u; \alpha + y, \beta)\, \mathrm{d}u \\
&= \frac{b^y/y!}{\mathrm{P}(a = 1)} \cdot \mathbb{E}[S^y] \cdot \exp(-b) \cdot \mathcal{H}_{\beta, \alpha+y}(1 - \bar{s}, b \cdot (1 - \bar{s})) \\
&= \frac{1}{\mathrm{P}(a = 1)} \cdot \mathrm{P}(y \mid b) \cdot \mathbb{E}[S^y] \cdot \mathcal{H}_{\beta, \alpha+y}(\bar{\bar{s}}, b\bar{\bar{s}})
\end{aligned}$$

where the line before last uses the integral representation from Result 6.4.2 and where $\bar{\bar{s}} = 1 - \bar{s}$ and $\bar{s} = F^{-1}(1 - \mathrm{P}(a = 1); \alpha, \beta)$. Multiplying through by the factor $\mathrm{P}(a = 1)$ gives

$$\mathrm{P}(Y = y, a = 1) = \mathrm{P}(y \mid b) \cdot \mathbb{E}[S^y] \cdot \mathcal{H}_{\beta, \alpha+y}(\bar{\bar{s}}, b\bar{\bar{s}}).$$

□

Result 6.4.11 (Joint outcome probability for the Bernoulli-beta sequence hazard model). Let Y be a random variable distributed according to a Bernoulli-Beta mixture with probability $\mu = bs$ for

$b \in (0,1)$ where $s \sim \text{beta}(\alpha, \beta)$. The joint distribution is

$$\text{P}(Y = y, a = 1) = [bM_1(\bar{s})]^{\mathbf{1}(y=1)} \cdot [\text{P}(a = 1) - bM_1(\bar{s})]^{\mathbf{1}(y=0)}$$

$$\text{for } y \in \{0, 1\}$$

where $M_j(x) = \mathbb{E}[S^j] \cdot \overline{F}(x; \alpha + j, \beta)$, $\mathbb{E}[S] = \alpha / (\alpha + \beta)$ and $\bar{s} = F^{-1}(1 - \text{P}(a = 1); \alpha, \beta)$ is the inverse of the beta cdf (Definition 6.1.6).

Proof. Note that

$$\text{Pr}(Y = 1 \mid s, a = 1) = bs$$

$$\text{Pr}(Y = 0 \mid s, a = 1) = 1 - bs.$$

In light of the definition of conditional probabilities

$$\text{P}(Y = y, a = 1) = \text{P}(Y = y \mid a = 1) \cdot \text{P}(a = 1).$$

Holding the $\text{P}(a = 1)$ and integrating out the s when for $Y = 1$:

$$\begin{aligned}
\text{Pr}(Y = 1 \mid a = 1) &= b \int_0^1 u \cdot f_{s \mid a=1}(u; \alpha, \beta) \, \mathrm{d}u \\
&= \frac{b}{\text{P}(a = 1)} \cdot \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} \cdot \int_{\bar{s}}^1 f(u; \alpha + 1, \beta) \, \mathrm{d}u \\
&= \frac{b}{\text{P}(a = 1)} \cdot \mathbb{E}[S] \cdot \bar{F}_{\alpha+1,\beta}(\bar{s}) \\
&= \frac{b}{\text{P}(a = 1)} \cdot \mathbb{E}[S] \cdot \mathcal{H}_{\beta,\alpha+1}(1 - \bar{s}, 0)
\end{aligned}$$

so that $\text{P}(Y=1, a=1) = b \cdot \mathbb{E}[S] \cdot \bar{F}_{\alpha+1,\beta}(\bar{s})$. Similarly for when $Y=0$,

$$\begin{aligned}\text{Pr}(Y=0 \,|\, a=1) &= 1 - b\int_0^1 u \cdot f_{s\,|\,a=1}(u;\alpha,\beta)\,\mathrm{d}u \\ &= 1 - \frac{b}{\text{P}(a=1)} \cdot \mathbb{E}[S] \cdot \bar{F}_{\alpha+1,\beta}(\bar{s})\end{aligned}$$

so that $\text{P}(Y=0, a=1) = \text{P}(a=1) - b \cdot \mathbb{E}[S] \cdot \bar{F}_{\alpha+1,\beta}(\bar{s})$. □

Result 6.4.12 (Moments of a truncated Poisson-beta mixture). Let $b > 0$ and let $\overline{F}(s;\alpha+j,\beta)$ denote the cdf of s which is distributed beta (α,β). Recall that $\text{P}(a=1)$ represents the latent variable function derived in Result 6.3.8. Assume the data generating process for an attack outcome is distributed as Poisson (bs). The mathematical expectation and variance of y are given by

$$\begin{aligned}\mathbb{E}[y] &= b \cdot M_1(\bar{s}) \\ \text{Var}[y] &= \mathbb{E}[y] + b^2 \cdot \left(M_2(\bar{s}) - M_1(\bar{s})^2\right)\end{aligned}$$

where $M_j(x) = \mathbb{E}[S^j] \cdot \overline{F}(x;\alpha+j,\beta)$, where $\mathbb{E}[S^j] = \text{B}(\alpha+j,\beta)/\text{B}(\alpha,\beta)$, and where $\bar{s} = F^{-1}(1-\text{P}(a=1);\alpha,\beta)$.

Proof. Let $l_j = \mathbb{E}[S^j \,|\, a=1]$ and define $M_j(x) = \mathbb{E}[S^j] \cdot \overline{F}_{\alpha+1,\beta}(x)$ and let $P \equiv \text{P}(a=1)$ so that $l_j = \frac{M_j(\bar{s})}{P}$. Random variable $y \,|\, a=1, s$ follows a Poisson distribution with parameter bs so

$$\mathbb{E}[y \,|\, a=1, s] = \text{Var}[y \,|\, a=1, s] = bs$$

implying that $\mathbb{E}[y \,|\, a=1] = b \cdot \mathbb{E}[S \,|\, a=1]$ so that $\mathbb{E}[y \,|\, a=1] = b \cdot l_1$. Using Result 6.4.1 write the variance of y conditional on the event $a = 1$ as

$$\begin{aligned}
\text{Var}[y \mid a=1] &= \mathbb{E}_{s \mid a=1}[\text{Var}[y \mid a=1, s]] + \text{Var}_{s \mid a=1}[\mathbb{E}[y \mid a=1, s]] \\
&= \mathbb{E}_{s \mid a=1}[bs] + \text{Var}_{s \mid a=1}[bs] \\
&= b \cdot \mathbb{E}[S \mid a=1] + b^2 \cdot \text{Var}[S \mid a=1] \\
&= b \cdot \mathbb{E}[S \mid a=1] + b^2 \cdot \left(\mathbb{E}\left[S^2 \mid a=1\right] - \mathbb{E}[S \mid a=1]^2\right) \\
&= b \cdot \mathbb{E}[S \mid a=1] + b^2 \cdot \mathbb{E}\left[S^2 \mid a=1\right] - b^2 \cdot \mathbb{E}[S \mid a=1]^2 \\
&= b \cdot l_1 + b^2 \cdot l_2 - b^2 \cdot l_1^2 \\
&= b \cdot l_1 \cdot (1 - b \cdot l_1) + b^2 \cdot l_2.
\end{aligned}$$

Using the same variance formula again to calculate $\text{Var}[y]$ by conditioning on the event a that occurs with probability $P \equiv \text{P}(a = 1)$

$$\begin{aligned}
\text{Var}[y] &= \mathbb{E}_a[\text{Var}[y \mid a]] + \text{Var}_a[\mathbb{E}[y \mid a]] \\
&= \mathbb{E}_a\left[a \cdot \left(b \cdot l_1 \cdot (1 - b \cdot l_1) + b^2 \cdot l_2\right)\right] + \text{Var}_a[abl_1] \\
&= P \cdot \left(b \cdot l_1 \cdot (1 - b \cdot l_1) + b^2 \cdot l_2\right) + P(1-P)(bl_1)^2 \\
&= P \cdot \left(b \cdot \frac{M_1}{P} \cdot \left(1 - b \cdot \frac{M_1}{P}\right) + b^2 \cdot \frac{M_2}{P}\right) + P(1-P)\left(b \cdot \frac{M_1}{P}\right)^2 \\
&= b \cdot M_1 \cdot \left(1 - b \cdot \frac{M_1}{P}\right) + b^2 \cdot M_2 + \frac{(1-P)}{P}(b \cdot M_1)^2 \\
&= b \cdot M_1 - \frac{(b \cdot M_1)^2}{P} + b^2 \cdot M_2 + \frac{(b \cdot M_1)^2}{P} - (b \cdot M_1)^2 \\
&= b \cdot M_1 + b^2 \cdot \left(M_2 - M_1^2\right).
\end{aligned}$$

Note that $P = 1$ implies $\bar{s} = 0$ and $M_j(0) = \mathbb{E}[S^j]$ so the variance formula then becomes

$b \cdot \mathbb{E}[S] + b^2 \cdot \left(\mathbb{E}[S^2] - \mathbb{E}[S]^2\right) = b \cdot \mathbb{E}[S] + b^2 \text{Var}[S]$ as in the case where Y is a pure Poisson-Beta process. $\square$

Result 6.4.13 (Moments of a truncated Bernoulli-beta mixture). Let $b > 0$ and let $\overline{F}(s; \alpha + j, \beta)$ denote the cdf of s which is distributed beta (α, β). Recall that $\text{P}(a = 1)$ represents the latent variable function derived in Result 6.3.8. Assume the data generating process for an attack outcome is distributed as Poisson (bs). The mathematical expectation and variance of y are given by

$$\mathbb{E}[y] = b \cdot M_1(\bar{s})$$

$$\text{Var}[y] = \mathbb{E}[y] \cdot (1 - \mathbb{E}[y])$$

where

$$M_j(x) = \overline{F}(x; \alpha + j, \beta) \cdot \mathbb{E}\left[S^j\right],$$

$$\mathbb{E}\left[S^j\right] = \text{B}(\alpha + j, \beta) / \text{B}(\alpha, \beta),$$

and where $\bar{\varepsilon} = \text{F}^{-1}(1 - \text{P}(a = 1); \alpha, \beta)$.

Proof. Random variable $y \mid a = 1, s$ follows a Bernoulli distribution with parameter bs so

$$\mathbb{E}[y \mid a = 1, s] = bs$$

$$\text{Var}[y \mid a = 1, s] = bs \cdot (1 - bs)$$

implying that $\mathbb{E}[y \mid a=1] = b \cdot \mathbb{E}[S \mid a=1]$. Let $l_j \equiv \mathbb{E}[S^j \mid a=1]$ and write $\mathbb{E}[y \mid a=1] = b \cdot l_1$. Using Result 6.4.1 write the variance of y conditional on the event $a=1$ as

$$
\begin{aligned}
\text{Var}[y \mid a=1] &= \mathbb{E}_{s \mid a=1}[\text{Var}[y \mid a=1, s]] + \text{Var}_{s \mid a=1}[\mathbb{E}[y \mid a=1, s]] \\
&= \mathbb{E}_{s \mid a=1}[bs \cdot (1-bs)] + \text{Var}_{s \mid a=1}[bs] \\
&= \mathbb{E}_{s \mid a=1}\left[bs - q^2 s^2\right] + q^2 \cdot \text{Var}_{s \mid a=1}[S] \\
&= q \cdot \mathbb{E}[S \mid a=1] - q^2 \cdot \mathbb{E}\left[S^2 \mid a=1\right] + q^2 \cdot \text{Var}[S \mid a=1] \\
&= q \cdot l_1 - q^2 \cdot l_2 + q^2 \cdot \left[l_2 - l_1^2\right] \\
&= q \cdot l_1 - q^2 \cdot l_1^2 \\
&= q \cdot l_1 \cdot (1 - q \cdot l_1)\,.
\end{aligned}
$$

Note that $l_j \equiv \mathbb{E}[S^j \mid a=1] = \frac{M_j(\bar{s})}{P}$ where $M_j(x) = \mathbb{E}[S] \cdot \overline{F}_{\alpha+1,\beta}(x)$ using Result 6.4.8. Using the same variance formula again to calculate $\text{Var}[y]$ by conditioning on the event a that occurs with probability $P \equiv \mathrm{P}(a=1)$

$$
\begin{aligned}
\operatorname{Var}[y] &= \mathbb{E}_a\left[\operatorname{Var}[y \mid a]\right] + \operatorname{Var}_a\left[\mathbb{E}[y \mid a]\right] \\
&= \mathbb{E}_a\left[abl_1\left(1-bl_1\right)\right] + \operatorname{Var}_a\left[abl_1\right] \\
&= Pbl_1\left(1-bl_1\right) + P\left(1-P\right)\left(bl_1\right)^2 \\
&= Pb \cdot \frac{M_1}{P}\left(1 - b \cdot \frac{M_1}{P}\right) + P\left(1-P\right)\left(b \cdot \frac{M_1}{P}\right)^2 \\
&= b \cdot M_1\left(1 - b \cdot \frac{M_1}{P}\right) + \frac{\left(1-P\right)}{P} \cdot \left(b \cdot M_1\right)^2 \\
&= b \cdot M_1 - \frac{\left(b \cdot M_1\right)^2}{P} + \frac{\left(b \cdot M_1\right)^2}{P} - \left(b \cdot M_1\right)^2 \\
&= b \cdot M_1 - \left(b \cdot M_1\right)^2 \\
&= b \cdot M_1 \cdot \left(1 - b \cdot M_1\right).
\end{aligned}
$$

Note that $P = 1$ implies $\bar{s} = 0$ and $M_j(0) = \mathbb{E}\left[S^j\right]$ so the variance formula then becomes $b \cdot \mathbb{E}[S] \cdot (1 - b \cdot \mathbb{E}[S])$ as in the case where Y is a pure Bernoulli-Beta process. $\square$

Result 6.4.14 (Change in attack intensity). Let $s \sim \text{beta}(\alpha, \beta)$ and $\varepsilon \sim \text{exponential}(\lambda)$ and let $a > 0$ and $b > 0$. Recall that $\mathrm{P}(a=1)$ represents the latent variable function derived in Result 6.3.8.

$$
\frac{\partial \mathbb{E}[y]}{\partial x_i} \lesseqqgtr 0 \iff \frac{\mathrm{d}}{\mathrm{d}x_i} q\left(x_t; \theta_q\right) \lesseqqgtr -\frac{\mathrm{d}}{\mathrm{d}x_i} \log \bar{F}\left(\bar{\varepsilon}_t; \alpha, \beta\right)
$$

where $\bar{\varepsilon} = \mathrm{F}^{-1}\left(1 - \mathrm{P}(a=1); \alpha, \beta\right)$ is the inverse cdf of the Beta distribution (Definition 6.1.6).

Proof. Recall that

$$\mathbb{E}\left[y\right] = q \cdot \mathbb{E}\left[S\right] \cdot \overline{\mathrm{F}}_{\alpha+1,\beta}\left(\bar{s}\right)$$

where $\mathbb{E}\left[S\right] = \alpha / \left(\alpha + \beta\right)$ and $\overline{\mathrm{F}}\left(\cdot; a, b\right) = 1 - \mathrm{F}\left(\cdot; a, b\right)$. Calculating the derivative directly:

$$\frac{\partial \mathbb{E}\left[y\right]}{\partial x} = b' \cdot \mathbb{E}\left[S\right] \cdot \overline{\mathrm{F}}_{\alpha+1,\beta}\left(\bar{s}\right) - b \cdot \mathbb{E}\left[S\right] \cdot f\left(\bar{s}; \alpha+1, \beta\right) \cdot \frac{\partial \bar{\varepsilon}}{\partial x}$$

noting that the function b satisfies $b' = \theta \cdot b$. Rearranging this expression and applying the definition of a hazard to the right-hand side gives the result. □

6.5 Appendix E: Proofs for Section 3.3

Result 6.5.1 (Expectation of the Bellman max operator). Let ε be a random variable with density $f(\varepsilon)$, and $u_1(.)$ and $u_2(.)$ real valued functions. Then

$$E \max\{u_1(\varepsilon), u_2(\varepsilon)\} = P(a=1) \cdot E[u_1(\varepsilon) \mid a=1) + P(a=0) \cdot E[u_2(\varepsilon) \mid a=0).$$

Proof. Let ε be a random variable with density $f(\varepsilon)$, and $u_1(.)$ and $u_2(.)$ real valued functions.

Then

$$
\begin{aligned}
& E\max\{u_1(\varepsilon),u_2(\varepsilon)\} \\
= \quad & E\left\{u_1(\varepsilon)\mathbf{1}[u_1(\varepsilon)>u_2(\varepsilon)]+u_2(\varepsilon)\mathbf{1}[u_1(\varepsilon)\leq u_2(\varepsilon)]\right\} \\
= \quad & E\left\{u_1(\varepsilon)\mathbf{1}[u_1(\varepsilon)>u_2(\varepsilon)]\right\}+E\left\{u_2(\varepsilon)\mathbf{1}[u_1(\varepsilon)\leq u_2(\varepsilon)]\right\} \\
= \quad & \int_{\{\varepsilon:u_1(\varepsilon)>u_2(\varepsilon)\}} u_1(\varepsilon)f(\varepsilon)d\varepsilon+\int_{\{\varepsilon:u_1(\varepsilon)\leq u_2(\varepsilon)\}} 0f(\varepsilon)d\varepsilon \\
& +\int_{\{\varepsilon:u_1(\varepsilon)\leq u_2(\varepsilon)\}} u_2(\varepsilon)f(\varepsilon)d\varepsilon+\int_{\{\varepsilon:u_1(\varepsilon)>u_2(\varepsilon)\}} 0f(\varepsilon)d\varepsilon \\
= \quad & \int_{\{\varepsilon:u_1(\varepsilon)>u_2(\varepsilon)\}} u_1(\varepsilon)f(\varepsilon)d\varepsilon+\int_{\{\varepsilon:u_1(\varepsilon)\leq u_2(\varepsilon)\}} u_2(\varepsilon)f(\varepsilon)d\varepsilon \\
= \quad & \int_A u_1(\varepsilon)f(\varepsilon)d\varepsilon+\int_{A^c} u_2(\varepsilon)f(\varepsilon)d\varepsilon,
\end{aligned}
$$

with $A=\{\varepsilon:u_1(\varepsilon)>u_2(\varepsilon)\}$. Next observe that

$$
\begin{aligned}
\int_A u_1(\varepsilon)f(\varepsilon)d\varepsilon &= P(A)\int_A u_1(\varepsilon)\frac{f(\varepsilon)}{P(A)}d\varepsilon=P(A)E[u_1(\varepsilon)\mid A), \\
\int_{A^c} u_2(\varepsilon)f(\varepsilon)d\varepsilon &= P(A^c)\int_{A^c} u_2(\varepsilon)\frac{f(\varepsilon)}{P(A^c)}d\varepsilon=P(A^c)E[u_2(\varepsilon)\mid A^c).
\end{aligned}
$$

Now define a random variable

$$
a(\varepsilon)=\begin{cases} 1 & \text{if} \quad u_1(\varepsilon)>u_2(\varepsilon) \\ 0 & \text{if} \quad u_1(\varepsilon)\leq u_2(\varepsilon) \end{cases}
$$

then we can also write $A = \{\varepsilon : a(\varepsilon) = 1\}$ and

$$\begin{aligned} \int_A u_1(\varepsilon) f(\varepsilon) d\varepsilon &= P(a=1) E[u_1(\varepsilon) \mid a=1), \\ \int_{A^c} u_2(\varepsilon) f(\varepsilon) d\varepsilon &= P(a=0) E[u_2(\varepsilon) \mid a=0), \end{aligned}$$

and thus

$$\begin{aligned} E \max\{u_1(\varepsilon), u_2(\varepsilon)\} = \quad & P(a=1) E[u_1(\varepsilon) \mid a=1) \\ + \quad & P(a=0) E[u_2(\varepsilon) \mid a=0). \end{aligned}$$

$\square$

Bibliography

Abadie, Alberto and Sofia Dermisi (2007) "Is Terrorism Eroding Agglomeration Economies in Central Business Districts? Lessons from the Office Real Estate Market in Downtown Chicago."

Abadie, Alberto and Javier Gardeazabal (2003) "The economic costs of conflict: A case study of the Basque Country," *American economic review*, 93 (1), 113–132.

——— (2008) "Terrorism and the world economy," *European economic review*, 52 (1), 1–27.

Abbring, Jaap H and Gerard J Van den Berg (2003a) "The identifiability of the mixed proportional hazards competing risks model," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65 (3), 701–710.

——— (2003b) "The nonparametric identification of treatment effects in duration models," *Econometrica*, 71 (5), 1491–1517.

Abbring, Jaap H and James J Heckman (2007) "Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation," *Handbook of econometrics*, 6, 5145–5303.

Abbring, Jaap H and Gerard J Van Den Berg (2007) "The unobserved heterogeneity distribution in duration analysis," *Biometrika*, 94 (1), 87–99.

Abramowitz, Milton, Irene A Stegun et al. (1964) *Handbook of mathematical functions*, 55: Dover New York.

Aguirregabiria, Victor, Allan Collard-Wexler, and Stephen P Ryan (2021) "Dynamic games in empirical industrial organization," in *Handbook of Industrial Organization*, 4, 225–343: Elsevier.

Aguirregabiria, Victor and Pedro Mira (2002) "Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models," *Econometrica*, 70 (4), 1519–1543.

——— (2010) "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156 (1), 38–67.

Amemiya, Takeshi (1996) "Structural duration models," *Journal of statistical planning and inference*, 49 (1), 39–52.

Anderson, Chris and David Sally (2013) *The numbers game: Why everything you know about soccer is wrong*: Penguin.

Arce, Daniel G (2019) "On the human consequences of terrorism," *Public Choice*, 178 (3), 371–396.

Arcidiacono, Peter and Robert A Miller (2011) "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 79 (6), 1823–1867.

Argomaniz, Javier and Alberto Vidal-Diez (2015) "Examining deterrence and backlash effects in counter-terrorism: The case of ETA," *Terrorism and Political Violence*, 27 (1), 160–181.

Arin, K Peren, Davide Ciferri, and Nicola Spagnolo (2008) "The price of terror: The effects of terrorism on stock market returns and volatility," *Economics Letters*, 101 (3), 164–167.

Azam, Jean-Paul (2005) "Suicide-bombing as inter-generational investment," *Public choice*, 122 (1-2), 177–198.

Bali, Valentina A (2007) "Terror and elections: Lessons from Spain," *Electoral Studies*, 26 (3), 669–687.

Barros, Carlos Pestana (2003) "An intervention analysis of terrorism: The Spanish ETA case," *Defence and Peace Economics*, 14 (6), 401–412.

Barros, Carlos Pestana, José Passos, and Luis A Gil-Alana (2006) "The timing of ETA terrorist attacks," *Journal of Policy Modeling*, 28 (3), 335–346.

Becker, Gary S (1968) "Crime and punishment: An economic approach," *Journal of political economy*, 76 (2), 169–217.

Van den Berg, Gerard J (2001) "Duration models: specification, identification and multiple durations," in *Handbook of econometrics*, 5, 3381–3460: Elsevier.

Berrebi, Claude and Esteban F Klor (2006) "On terrorism and electoral outcomes: Theory and evidence from the Israeli-Palestinian conflict," *Journal of conflict resolution*, 50 (6), 899–925.

Berrebi, Claude and Darius Lakdawalla (2007) "How does terrorism risk vary across space and time? An analysis based on the Israeli experience," *Defence and Peace Economics*, 18 (2), 113–131.

Blundell, Richard (2017) "What have we learned from structural models?" *American Economic Review*, 107 (5), 287–292.

Box-Steffensmeier, Janet M and Christopher Zorn (2002) "Duration models for repeated events," *Journal of Politics*, 64 (4), 1069–1094.

Britton, Jack and Ben Waltmann (2021) "Revisiting the solution of dynamic discrete choice models: Time to bring back Keane and Wolpin (1994)?".

Buesa, Mikel, Aurelia Valiño, Joost Heijs, Thomas Baumert, and Javier Gonzalez Gomez (2007) "The Economic Cost of March 11: Measuring the direct economic cost of the terrorist attack on March 11, 2004 in Madrid," *Terrorism and Political Violence*, 19 (4), 489–509.

Camacho, Adriana (2008) "Stress and birth weight: evidence from terrorist attacks," *American Economic Review*, 98 (2), 511–15.

Carson, Jennifer Varriale (2014) "Counterterrorism and radical eco-groups: A context for exploring the series hazard model," *Journal of Quantitative Criminology*, 30 (3), 485–504.

——— (2017) "Assessing the Effectiveness of High-Profile Targeted Killings in the "War on Terror" A Quasi-Experiment," *Criminology & Public Policy*, 16 (1), 191–220.

Cho, Sung-Jin (2011) "An empirical model of mainframe computer investment," *Journal of Applied Econometrics*, 26 (1), 122–150.

Clauset, Aaron and Kristian Skrede Gleditsch (2012) "The developmental dynamics of terrorist organizations," *PloS one*, 7 (11), e48633.

Clauset, Aaron, Maxwell Young, and Kristian Skrede Gleditsch (2007) "On the frequency of severe terrorist events," *Journal of Conflict Resolution*, 51 (1), 58–87.

Clausewitz, Carl (1832) *On war*: Penguin UK.

Cleves, Mario, William Gould, William W Gould, Roberto Gutierrez, and Yulia Marchenko (2008) *An introduction to survival analysis using Stata*: Stata press.

Corbet, Shaen, John F O'Connell, Marina Efthymiou, Cathal Guiomard, and Brian Lucey (2019) "The impact of terrorism on European tourism," *Annals of Tourism Research*, 75, 1–17.

Cox, David R (1972) "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, 34 (2), 187–202.

D'Alessio, Stewart J and Lisa Stolzenberg (1990) "Sicarii and the Rise of Terrorism," *Studies in Conflict & Terrorism*, 13 (4-5), 329–335.

De Cataldo Neuburger, Luisella, Tiziana Valentini et al. (1996) *Women and terrorism*: Springer.

Drakos, Konstantinos (2010) "Terrorism activity, investor sentiment, and stock returns," *Review of financial economics*, 19 (3), 128–135.

Dugan, Laura (2010) "Estimating effects over time for single and multiple units," *Handbook of quantitative criminology*, 741–763.

——— (2011) "The series hazard model: An alternative to time series for event data," *Journal of Quantitative Criminology*, 27 (3), 379–402.

Dugan, Laura, Julie Y Huang, Gary LaFree, and Clark McCauley (2008) "Sudden desistance from terrorism: the Armenian secret army for the liberation of Armenia and the Justice Commandos of the Armenian Genocide," *Dynamics of Asymmetric Conflict*, 1 (3), 231–249.

Dugan, Laura, Gary LaFree, and Alex R Piquero (2005) "Testing a rational choice model of airline hijackings," *Criminology*, 43 (4), 1031–1065.

Dugan, Laura, Daniel S Nagin, and Richard Rosenfeld (2003) "Exposure reduction or retaliation? The effects of domestic violence resources on intimate-partner homicide," *Law & society review*, 37 (1), 169–198.

D'Alessio, Stewart J, Lisa Stolzenberg, and Dustin Dariano (2014) "Does targeted capture reduce terrorism?" *Studies in Conflict & Terrorism*, 37 (10), 881–894.

Ehrlich, Isaac (1973) "Participation in illegitimate activities: A theoretical and empirical investigation," *Journal of political Economy*, 81 (3), 521–565.

Elbers, Chris and Geert Ridder (1982) "True and spurious duration dependence: The identifiability of the proportional hazard model," *The Review of Economic Studies*, 49 (3), 403–409.

Enders, Walter, Gerald F Parise, and Todd Sandler (1992) "A time-series analysis of transnational terrorism: Trends and cycles," *Defence and Peace Economics*, 3 (4), 305–320.

Enders, Walter and Todd Sandler (1991) "Causality between transnational terrorism and tourism: The case of Spain," *Studies in Conflict & Terrorism*, 14 (1), 49–58.

——— (1993) "The Effectiveness of Antiterrorism Policies: A Vector-Autoregression-Intervention Analysis," *American Political Science Review*, 87 (4), 829–844.

——— (1995) "Terrorism: Theory and applications," *Handbook of defense economics*, 1, 213–249.

——— (1999) "Transnational terrorism in the post–Cold War era," *International Studies Quarterly*, 43 (1), 145–167.

——— (2002) "Patterns of transnational terrorism, 1970–1999: Alternative time-series estimates," *International Studies Quarterly*, 46 (2), 145–165.

——— (2005) "Transnational terrorism 1968-2000: Thresholds, persistence, and forecasts," *Southern Economic Journal*, 71 (3), 467–482.

——— (2011) *The political economy of terrorism*: Cambridge University Press.

Erdélyi, A (1953) "Higher Transcendental Functions vol 1 (New York, MacGraw-Hill) p 103, 224, 238, 253 and 254 Google Scholar Erdélyi A 1953," *Higher transcendental functions*, 2, 188.

Ezell, Michael E, Kenneth C Land, and Lawrence E Cohen (2003) "Modeling multiple failure time data: A survey of variance–corrected proportional hazards models with empirical applications to arrest data," *Sociological Methodology*, 33 (1), 111–167.

Ferrari, Silvia and Francisco Cribari-Neto (2004) "Beta regression for modelling rates and proportions," *Journal of applied statistics*, 31 (7), 799–815.

Gordy, Michael B (1998a) "Computationally convenient distributional assumptions for common-value auctions," *Computational Economics*, 12, 61–78.

——— (1998b) "A generalization of generalized beta distributions."

Gupta, Rangan, Anandamayee Majumdar, Christian Pierdzioch, and Mark E Wohar (2017) "Do terror attacks predict gold returns? Evidence from a quantile-predictive-regression approach," *The Quarterly Review of Economics and Finance*, 65, 276–284.

Gurland, John (1958) “A generalized class of contagious distributions,” *Biometrics*, 14 (2), 229–249.

Hamerle, A (1988) “A dynamic logistic regression model for multiple spell failure time data,” *Biometrical journal*, 30 (4), 471–481.

Han, Aaron and Jerry A Hausman (1990) “Flexible parametric estimation of duration and competing risk models,” *Journal of applied Econometrics*, 5 (1), 1–28.

Heckman, James J (1991) “Identifying the hand of past: Distinguishing state dependence from heterogeneity,” *The American Economic Review*, 81 (2), 75–79.

Heckman, James J and Burton Singer (1986) “Econometric analysis of longitudinal data,” *Handbook of econometrics*, 3, 1689–1763.

Heckman, James and Burton Singer (1984a) “The identifiability of the proportional hazard model,” *The Review of Economic Studies*, 51 (2), 231–241.

——— (1984b) “A method for minimizing the impact of distributional assumptions in econometric models for duration data,” *Econometrica: Journal of the Econometric Society*, 271–320.

Hogg, R.V., J.W. McKean, and A.T. Craig (2013) *Introduction to Mathematical Statistics*, Always learning: Pearson.

Honoré, Bo E (1993) “Identification results for duration models with multiple spells,” *The Review of Economic Studies*, 60 (1), 241–246.

Horsley, Richard A (1979) “The Sicarii: Ancient Jewish” Terrorists”,” *The Journal of Religion*, 59 (4), 435–458.

Hotz, V Joseph and Robert A Miller (1993) “Conditional choice probabilities and the estimation of dynamic models,” *The Review of Economic Studies*, 60 (3), 497–529.

Hsu, Der-Ann and Richard E Quandt (1976) *Statistical analyses of aircraft hijackings and political assassinations*: Univ.

Humbert, Pierre (1922) “IX.—The confluent hypergeometric functions of two variables,” *Proceedings of the Royal Society of Edinburgh*, 41, 73–96.

Jordan, Jenna (2009) “When heads roll: Assessing the effectiveness of leadership decapitation,” *Security Studies*, 18 (4), 719–755.

Kalbfleisch, John D and Ross L Prentice (2011) *The statistical analysis of failure time data*: John Wiley & Sons.

Karlis, Dimitris and Evdokia Xekalaki (2005) “Mixed poisson distributions,” *International Statistical Review/Revue Internationale de Statistique*, 35–58.

Karolyi, G Andrew and Rodolfo Martell (2010) “Terrorism and the Stock Market.,” *International Review of Applied Financial Issues & Economics*, 2 (2).

Kasahara, Hiroyuki (2009) "Temporary Increases in Tariffs and Investment: The Chilean Experience," *Journal of Business & Economic Statistics*, 27 (1), 113–127.

Kasahara, Hiroyuki, Katsumi Shimotsu et al. (2011) "Sequential estimation of dynamic programming models with unobserved heterogeneity."

Kasahara, Hiroyuki and Katsumi Shimotsu (2008) "Pseudo-likelihood estimation and bootstrap inference for structural discrete Markov decision models," *Journal of Econometrics*, 146 (1), 92–106.

——— (2012) "Sequential estimation of structural models with a fixed point constraint," *Econometrica*, 80 (5), 2303–2319.

Katti, SK (1966) "Interrelations among generalized distributions and their components," *Biometrics*, 22 (1), 44–52.

Keane, Michael P, Petra E Todd, and Kenneth I Wolpin (2011) "The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications," in *Handbook of labor economics*, 4, 331–461: Elsevier.

Keane, Michael P and Kenneth I Wolpin (1997) "The career decisions of young men," *Journal of political Economy*, 105 (3), 473–522.

——— (2009) "Empirical applications of discrete choice dynamic programming models," *Review of Economic Dynamics*, 12 (1), 1–22.

Kelly, Patrick J and Lynette L-Y Lim (2000) "Survival analysis for recurrent event data: an application to childhood infectious diseases," *Statistics in medicine*, 19 (1), 13–33.

Kiefer, Nicholas M (1988) "Economic duration data and hazard functions," *Journal of economic literature*, 26 (2), 646–679.

Kissner, Jason (2016) "Are active shootings temporally contagious? An empirical assessment," *Journal of Police and Criminal Psychology*, 31, 48–58.

Kremer, Michael (1993) "The O-ring theory of economic development," *The quarterly journal of economics*, 108 (3), 551–575.

Kristensen, Dennis, Lars Nesheim, and Áureo de Paula (2015) "CCP and the estimation of nonseparable dynamic models,"Technical report, Mimeo, University College London.

Kropotkin, Peter (2016) *The spirit of revolt*: Lazy Bear Press.

LaFree, Gary, Laura Dugan, and Raven Korte (2009) "The impact of British counterterrorist strategies on political violence in Northern Ireland: Comparing deterrence and backlash models," *Criminology*, 47 (1), 17–45.

LaFree, Gary, Laura Dugan, Min Xie, and Piyusha Singh (2012) "Spatial and temporal patterns of terrorist attacks by ETA 1970 to 2007," *Journal of Quantitative Criminology*, 28, 7–29.

Lancaster, Tony (1979) "Econometric methods for the duration of unemployment," *Econometrica: Journal of the Econometric Society*, 939–956.

——— (1990) *The econometric analysis of transition data* (17): Cambridge university press.

Landes, William M (1978) "An economic study of US aircraft hijacking, 1961-1976," *the Journal of Law and Economics*, 21 (1), 1–31.

Lanouar, Charfeddine and Mohamed Goaied (2019) "Tourism, terrorism and political violence in Tunisia: Evidence from Markov-switching models," *Tourism Management*, 70, 404–418.

Larsen, Bradley J, Florian Oswald, Gregor Reich, and Dan Wunderli (2012) "A test of the extreme value type I assumption in the bus engine replacement model," *Economics Letters*, 116 (2), 213–216.

Lin, Danyu Y and Lee-Jen Wei (1989) "The robust inference for the Cox proportional hazards model," *Journal of the American statistical Association*, 84 (408), 1074–1078.

Low, Hamish and Costas Meghir (2017) "The use of structural models in econometrics," *Journal of Economic Perspectives*, 31 (2), 33–58.

Luft, Gal (2003) "The logic of Israel's targeted killing," *Middle East Quarterly*.

Magnac, Thierry and David Thesmar (2002) "Identifying dynamic discrete decision processes," *Econometrica*, 70 (2), 801–816.

Mannes, Aaron (2008) "Testing the snake head strategy: Does killing or capturing its leaders reduce a terrorist group's activity?" *The Journal of International Policy Solutions*, 9.

McCall, Brian P (1996) "The identifiability of the mixed proportional hazards model with time-varying coefficients," *Econometric Theory*, 12 (4), 733–738.

McCauley, Clark (2006) "Jujitsu Politics: Terrorism and Responses to Terrorism.."

McFadden, Daniel et al. (1973) "Conditional logit analysis of qualitative choice behavior."

Montalvo, Jose G (2011) "Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections," *Review of Economics and Statistics*, 93 (4), 1146–1154.

Moran, PAP (1951) "The random division of an interval—Part II," *Journal of the Royal Statistical Society: Series B (Methodological)*, 13 (1), 147–150.

Moussa, Fatma Ben and Mariem Talbi (2019) "Stock Market Reaction to Terrorist Attacks and Political Uncertainty: Empirical Evidence from the Tunisian Stock Exchange," *International Journal of Economics and Financial Issues*, 9 (3), 48.

Pakes, Ariel (1986) "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica: Journal of the Econometric Society*, 755–784.

Pape, Robert A (2003) "The strategic logic of suicide terrorism," *American political science review*, 97 (3), 343–361.

Perry, Simon and Badi Hasisi (2015) "Rational choice rewards and the jihadist suicide bomber," *Terrorism and Political Violence*, 27 (1), 53–80.

Pizam, Abraham and Ginger Smith (2000) "Tourism and terrorism: A quantitative analysis of major terrorist acts and their impact on tourism destinations," *Tourism Economics*, 6 (2), 123–138.

Prentice, Ross L, Benjamin J Williams, and Arthur V Peterson (1981) "On the regression analysis of multivariate failure time data," *Biometrika*, 68 (2), 373–379.

Procasky, William J and Nacasius U Ujah (2016) "Terrorism and its impact on the cost of debt," *Journal of International Money and Finance*, 60, 253–266.

Reiss, Peter C and Frank A Wolak (2007) "Structural econometric modeling: Rationales and examples from industrial organization," *Handbook of econometrics*, 6, 4277–4415.

Richardson, Lewis F (1948) "Variation of the frequency of fatal quarrels with magnitude," *Journal of the American Statistical Association*, 43 (244), 523–546.

Rust, John (1987) "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, 999–1033.

——— (1994) "Structural estimation of Markov decision processes," *Handbook of econometrics*, 4, 3081–3143.

Sanchez-Cuenca, Ignacio (2009) "Analyzing Temporal Variation in the Lethatlity of ETA," *Revista International de Sociologia*.

Sandler, Todd (2014) "The analytical study of terrorism: Taking stock," *Journal of Peace Research*, 51 (2), 257–271.

Sandler, Todd and Walter Enders (2004) "An economic perspective on transnational terrorism," *European Journal of Political Economy*, 20 (2), 301–316.

——— (2007) "Applying analytical methods to study terrorism," *International Studies Perspectives*, 8 (3), 287–302.

Sandler, Todd, John T Tschirhart, and Jon Cauley (1983) "A theoretical analysis of transnational terrorism," *American Political Science Review*, 77 (1), 36–54.

Sarabia, José María and Emilio Gómez-Déniz (2011) "Multivariate Poisson-Beta distributions with applications," *Communications in Statistics-Theory and Methods*, 40 (6), 1093–1108.

Schoenfeld, David (1982) "Partial residuals for the proportional hazards regression model," *Biometrika*, 69 (1), 239–241.

Shughart, William F (2011) "Terrorism in rational choice perspective," in *The handbook on the political economy of war*: Edward Elgar Publishing.

Stiglitz, Joseph E and Linda J Bilmes (2008) *The three trillion dollar war: The true cost of the Iraq conflict*: WW Norton & Company.

Todd, Petra E and Kenneth I Wolpin (2006) "Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility," *American economic review*, 96 (5), 1384–1417.

Willett, John B and Judith D Singer (1995) "It's déjà vu all over again: Using multiple-spell discrete-time survival analysis," *Journal of Educational and Behavioral Statistics*, 20 (1), 41–67.

Yoo, John (2011) "Assassination or targeted killings after 9/11," *NYL Sch. L. Rev.*, 56, 57.