J Quant Criminol (2011) 27:379–402
DOI 10.1007/s10940-010-9127-1

ORIGINAL PAPER

The Series Hazard Model: An Alternative to Time Series for Event Data

Laura Dugan

Published online: 15 December 2010

Abstract An important pursuit by a body of criminological research is its endeavor to determine whether interventions or policy changes effectively achieve their intended goals. Because theories predict that interventions could either improve or worsen outcomes, estimators designed to improve the accuracy of identifying program or policy effects are in demand. This article introduces the series hazard model as an alternative to interrupted time series when testing for the effects of an intervention on event-based outcomes. It compares the two approaches through an example that examines the effects of two interventions on aerial hijacking. While series hazard modeling may not be appropriate for all event-based time series data or every context, it is a robust alternative that allows for greater flexibility in many contexts.

Keywords Hazard modeling · Time series · Event data · Series hazard model

Introduction

An important pursuit by a body of criminological research is its endeavor to determine whether interventions or policy changes effectively achieve their intended goal. Theories such as deterrence and rational choice predict that right-minded would-be offenders will recognize that the adverse consequences of bad behavior far outweigh its benefits, leading them to, instead, choose legal alternatives (Gibbs 1975; Cornish and Clarke 1986; Paternoster 1987; Clarke and Felson 1993). Yet, alternative perspectives predict the opposite outcome. Theories of defiance and backlash predict that attempts to control criminal behavior could be perceived as illegitimate and might actually increase crime (Braithwaite 1989, 2005; Sherman 1993; Dugan et al. 2003; LaFree et al. 2009). These potentially harmful consequences of well-intended policies underscore the importance of accurately detecting intervention effects.

L. Dugan (✉)
University of Maryland, College Park, MD, USA
e-mail: ldugan@umd.edu

Attempts to detect precise intervention effects use estimators generated by a wide range of methodological strategies (for descriptions see Dugan 2010). This article introduces an alternative to one approach, interrupted time series, when testing for the effects of an intervention on event-based outcomes—the series hazard model. Current strategies to analyze event-based data require temporal aggregation that masks important sources of variation found in the details of each event and in the duration between events (Freeman 1989). As Shellman (2008) points out, by relying on country-year aggregations, important day-to-day nuances are lost to the scholar. Given the wealth of detail available to help us more fully understand the dynamics of these activities, it seems inefficient to rely exclusively on methods that are designed to draw conclusions based on tallies. In recent years, with the growing availability and comprehensiveness of terrorist incident databases (LaFree and Dugan 2007), scholars have harnessed this extra variation to estimate changes in the risk of continued activity after implementing interventions designed to reduce offenses (Dugan et al. 2005; LaFree et al. 2009). Yet, to date, no scholar has formally outlined the benefits and validity of adopting the hazard approach to estimate the hazard of multiple events on a single unit. This article formally introduces the series hazard model, by delineating its methodological features and demonstrating its utility as an alternative to time series when analyzing most event-specific crime data.

The series hazard model is an extension of the Cox proportional hazard model that estimates the changes in risk for subsequent events conditioned upon characteristics of past events and other event-specific or date-specific covariates—such as the implementation date for a policy or intervention. While an intuitive modeling strategy, very few researchers have adopted this approach—perhaps because no article has yet been written that demonstrates its validity and usefulness. Research that has used series hazard modeling includes an article by LaFree and colleagues that estimates the effects of six major initiatives by the British government to stop republican violence (LaFree et al. 2009). They found that only Operation Motorman seemed to have the intended effect of reducing the hazard of continued attacks. Three others (internment, criminalization, and targeted assassinations in Gibraltar) appeared to increase terrorist activity, while the other two showed no real change. In earlier research, Dugan et al. (2005) used this method to evaluate the success of three policy efforts to end aerial hijackings, finding, among other things, that once metal detectors were installed the hazard of hijackings dropped and after Cuba passed a law criminalizing aerial hijacking, the risk of hijacked flights to Cuba fell. Simply tightening screening appeared to have no impact on reducing continued hijacking. More recent work by Dugan et al. (2009) estimated the impact of one excessively damaging terrorist attack by the Armenian Secret Army for the Liberation of Armenia (ASALA) on the continued violence by two Armenian terrorist organizations, ASALA and the Justice Commandos of the Armenian Genocide (JCAG). The findings did, indeed, show that attacks by both groups subsided after the deadly attack by ASALA at Paris' Orly Airport.

All of these projects estimate changes in activity for a single unit after one or more interventions or noteworthy events. Currently, the most common analytical strategy for this type of problem is to use some form of time series (e.g., Brandt and Williams 2001; D'Alessio and Stolzenberg 1995; Lewis-Beck 1986; McDowall et al. 1991). While appropriate in many contexts, this article proposes that when the data are event-based, the series hazard model can be used as an alternative to time series. Datasets that chronicle events over time are good candidates for choosing the series hazard model over interrupted time series to estimate intervention effects because the model relies on the variation between those events. This article uses the hijacking data from Dugan et al. (2005) to demonstrate the utility of using series hazard modeling to estimate the effect of two interventions on aerial hijacking. I first estimate the effects for each intervention using interrupted time series, then demonstrate the added value of estimating them using the series method. I conclude by comparing the benefits and weaknesses of each effort.

An Alternative to Interrupted Time Series Analysis

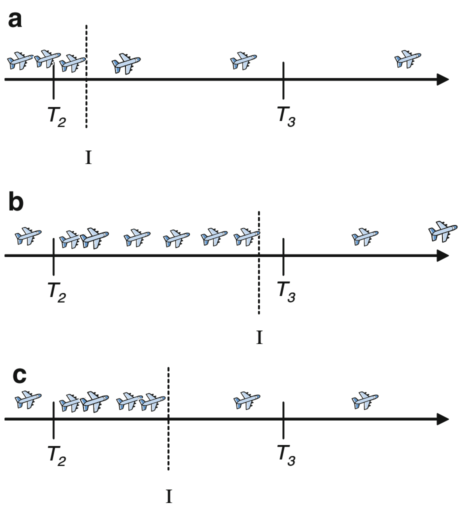
Interrupted time series has long been an important tool for estimating the impact of interventions on changes in an important outcome. For example McDowall et al. (1991) used this method to nullify the claim that civilian firearm ownership deters crime. Work by Stolzenberg and D'Alessio (1994, D'Alessio and Stolzenberg 1995) estimates the impact of the adoption of Minnesota's sentencing guidelines on changes in unwarranted disparity and the judicial use of the jail sanction. Kaminski et al. (1998) found evidence that law enforcement's practice of carrying pepper spray keeps suspects from assaulting police officers. More recently, Pridemore et al. (2007, 2008) used interrupted time series to estimate the effects of large events (the dissolution of the Soviet Union and the terror attacks in Oklahoma City and on September 11, 2001) on homicide an other bad outcomes.

Several variations of time series have been developed to accommodate different data types and address a variety of research questions (Lewis-Beck 1986; Dugan 2010). Basic autoregressive time series allows for multiple independent variables, as it corrects for dependence of the error terms (Greene 2008). Interrupted time series estimates the effects of a single intervention on the time trend using Autoregressive Integrated Moving Average (ARIMA) or other likelihood-based models (Lewis-Beck 1986; McDowall et al. 1980). Vector autoregression examines the relationship between two or more dependent variables (Brandt and Williams 2007).

While these methods all have their strengths, applying interrupted time series to event data also has its limitations. First, statistical power is a direct function of the length of the series and the span of measurement (year, quarter, month, or week). This creates a tension between gaining statistical power and losing stability. We may initially be tempted to aggregate to the smallest unit possible to achieve great statistical power. Yet, when relying on smaller units, events become sparse, possibly destabilizing the model and producing unreliable estimates. Furthermore, different levels of aggregation may produce different findings (Freeman 1989). Important research by Shellman (2004a) demonstrates that conclusions can dramatically change when event data are aggregated to broader units when using vector autoregression. More generally, Alt et al. (2000) show the importance of understanding and avoiding pitfalls of aggregation bias (see also Shellman 2004b).

Second, by aggregating events to a period of time, important distinctions are masked—imposing homogeneity across heterogeneous events. This could be especially problematic if specific characteristics of one event influence the likelihood of subsequent events. For example, a successful hijacking of a large passenger airplane for ransom is likely to inspire more hijackings—such as the highly publicized hijacking by D. B. Cooper in 1971 (Dugan et al. 2005)—increasing the likelihood of subsequent hijacking attempts, whereas the less publicized hijacking of a small private plane may fail to inspire anyone. Furthermore, a devastating terrorist hijacking that kills passengers and crew might dissuade more common criminals from attempting this modality of extortion. When analyzing these cases using interrupted time series, each is recorded as equal, ignoring context specific dependence and accounting only for temporal dependence.

Fig. 1 Sample temporal distributions of hijackings and intervention



Related to this, a third limitation is that important variation in the temporal distribution of events within a time span is lost. For example, two periods with seven events will be indistinguishable in interrupted time series even if the seven events in the first period are clustered near the beginning and those in the second are evenly distributed throughout. The choice to focus on the aggregate time period is essentially imposing an artificial structure on the more interesting distribution of the time between each event (Shellman 2008).

We face a related and fourth limitation that could lead to bias when we ignore the relative timing of events to an intervention within the span. To illustrate, Fig. 1 presents three examples of how the distribution of airline hijackings (the events) around an intervention that is implemented and measured at T_2 might affect analysis if, in fact, the intervention does indeed reduce hijackings. Recall that in interrupted time series each observation is measured with a tally of the total hijackings and an indicator of whether the intervention has been implemented.[1] The precise timing of the events and interventions are obscured during aggregation. In example *a*, the intervention was implemented near the beginning of the period. After aggregation, analysis will likely detect the drop in frequency from T_2 onwards leading to the correct conclusion that the intervention was successful. Note that this conclusion is likely despite the fact that the first hijacking in T_2 actually occurred prior to the intervention.

Example *b* demonstrates a more problematic situation. In this case, the large frequency of hijackings in T_2 actually occurred *prior* to the intervention, but is modeled as if they occurred *after* it. This measurement error will lead to bias, possibly nullifying any detectible intervention effect. In fact, the cluster of hijackings prior to the intervention might have even "caused" the intervention, thus producing simultaneity bias. One obvious precaution is to measure the intervention during the period after its implementation (T_3), or lagging the intervention, which forces all events that are concurrent to the intervention to be measured prior to it. In example *b*, we see that this strategy works well when the intervention is actually implemented near the end of the period. However, if applied to the situation in example *c*, lagging the intervention might help us avoid simultaneity bias, but

[1] While different transfer functions can be used to model the intervention, all rely on an indicator marking the beginning of the intervention period (Cook and Campbell 1979).

the difference between the pre-intervention period (T_2) and post-intervention period (T_3) is diluted because T_2 actually includes some of the actual post-intervention period [I, T_3), biasing any effects toward zero. This could be especially problematic if the actual intervention effect is short-lived and event frequency returns to pre-intervention volume quickly. The consequences of this problem become more obvious if we choose to lag the intervention in example *a*. See Lewis-Beck (1986) for a discussion of this type of measurement error.

This article introduces a relatively new strategy to address these limitations. By extending the Cox proportional hazard model to accommodate repeated events over time for a single unit, we can estimate the effects of an intervention on the hazard of continued events. Like time series, series hazard modeling relies upon the variation in activity for one unit. Unlike time series, it relies on the duration between activities instead of artificially aggregating the activities to multiple time periods.

Ideal Conditions for Series Hazard Modeling

The series hazard model is not appropriate for all research questions or data types. It only works with event data that record discrete incidents for one unit over time and include the exact date of all events so that the dependent variable can be calculated as the duration until the next event. The unit can be a person, an organization, a movement, a locality, or some other entity defined by a unifying condition. Discrete events can be criminal events, terrorist attacks, fatal incidents, hijackings, or any other newsworthy act, as long as it is possible to model the duration between events. For example, Lewis-Beck's (1986) interrupted time series example that examines changes in the number of traffic related fatalities after a 1972 speeding crackdown could be estimated as a series hazard model after converting the dependent variable to the number of days until the next fatal event.

A second criterion is that the events must occur with relative frequency. At the most basic level, because the event is the unit of analysis, rare events will unlikely produce enough statistical power to efficiently estimate parameters. Furthermore, because changes in temporal covariates are only measured during events, rare events could reduce statistical variation making it difficult to detect effects. As an extreme example, if an intervention successfully reduces events to zero, the intervention will appear as a vector of zeros in the data because no events occurred after the intervention. Having said this, if the events become less frequent after the intervention, time-varying covariates that capture the change in the intervention condition could be incorporated into the model, as done with conventional hazard modeling. While this can increase statistical power and document successful interventions, it will also produce missing values for the incident-specific variables since no incident is associated with the additional observations. Depending on the severity of the problem and the importance of the event-specific covariates, one might be able to cleverly interpolate missing values in order to optimize the efficiency of the model while reducing bias.

Introduction to Model

I begin this introduction by reviewing basic hazard modeling. Hazard modeling—which also goes by names such as survival analysis, duration analysis, and event history analysis—was first developed by bio-statisticians to study the timing until a subject dies and has since been applied to a wide range of outcomes (Allison 1995; Ezell et al. 2003). What distinguishes hazard modeling from other methodologies is that the dependent variable measures the time until some discrete event. This event can be virtually instantaneous, such as a death, marriage, arrest, or bombing; or it can be the crossing of a threshold by a more continuous measure, such as the unemployment rate or market price. Regardless of how it is measured, the event must be documented as a discrete unit of time, such as day, week, or month. Thus, a typical analysis includes many subjects, each with some measure of timing until the event occurred. Those subjects who never experienced the event are considered right censored, and can easily be accommodated by the model.

The timing until the event is often discussed within the context of a hazard function. As the duration until the event shortens, the hazard increases. While the hazard function is not a probability, it can be thought of as the probability that an event will occur during a very small interval of time (Allison 1995). Thus, a constant hazard function implies that the subject has the same risk over time. Yet, most hazards are likely to change over time, especially if the event is inevitable, such as death. For inevitable events the hazard function increases over time. Because the hazard function can take any number of forms, specific functions have been derived from known distributions (Box-Steffensmeier and Jones 2004; Allison 1995; Lawless 1982). For example, the constant hazard function is derived from the exponential distribution, and the Gompertz distribution produces a hazard function that when logged is a linear function of time. Other common distributions that produce reasonable hazard functions are the Weibull, log-normal, and log-logistic distributions (Box-Steffensmeier and Jones 2004; Allison 1995; Lawless 1982).

One obvious concern is that by specifying a particular hazard function, we are making an assumption that might compromise our findings if wrong. In essence, imposing distributional assumptions forces structure on the data, which is sometimes erroneous. In 1972, David Cox introduced the proportional hazard model which assumes that the ratio of the hazards from any two individuals is constant (Cox 1972). Thus, by estimating the proportional hazard model using partial likelihood, we avoid relying on distributional assumptions and allow the data to speak for itself.

Cox Proportional Hazard Model

I start by discussing the simplest research scenario where each unit can only experience the event once, hereafter more simply referred to as "failure". The Cox model is estimated according to the order of failures rather than accounting for the exact timing of each failure; and can be noted by the order statistics $t_{(1)} < \cdots < t_{(f)}$, where f is the number of individuals who fail (Lawless 1982; Ezell et al. 2003). The hazard function for the Cox proportional hazard model takes the following form for each unit i:

$$\lambda_i(t|\mathbf{X}_i) = \lambda_0(t)\exp(\mathbf{X}_i\boldsymbol{\beta}), \tag{1}$$

where $\mathbf{X}_i$ is a vector of covariate values for unit i, $\boldsymbol{\beta}$ is a vector of unknown parameters for $\mathbf{X}$ and $\lambda_0(t)$is an unspecified baseline hazard function for all units. In other words, it is the hazard for an individual whose value of $\mathbf{X} = 0$. As Cox (1972) shows, the following partial likelihood function estimates the values of $\boldsymbol{\beta}$ without relying on the specific form of $\lambda_0(t)$:

$$L(\beta) = \prod_{i=1}^{f}\left(\frac{\exp(\mathbf{X}_{(i)}\boldsymbol{\beta})}{\sum_{l\in R_i}\exp(\mathbf{X}_l\boldsymbol{\beta})}\right), \tag{2}$$

where f is the number of units that eventually fail ($n-f$ units are censored); and R_i is the risk set for unit i. The risk set includes all units who have yet to fail. In essence, the likelihood for unit i represents the probability that i fails conditional on someone failing (Allison 1995).

One extension of this model that draws us closer to series hazard modeling allows each unit to experience more than one failure or event. When studying the predictors of events that can repeat, it makes little sense to estimate the timing until only the first event. Yet, the non-independence of recurrent events within the same subject raises methodological concerns that are discussed in the following section.

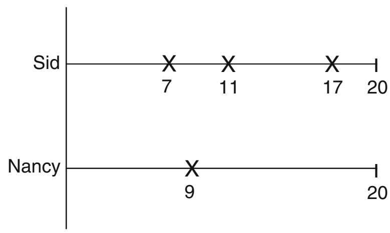
Recurrent Hazard Model

Progress has been made in estimating the hazards of repeated events by others who have estimated extensions of the Cox proportional hazard model that allow for recurrent events or multiple failure times (Ezell et al. 2003; Box-Steffensmeier and Zorn 2002; Kelly and Lim 2000). While recurrent event models still incorporate multiple units, some issues raised in these models apply to series hazard modeling. Kelly and Lim (2000) nicely summarize five Cox-based models and two variant models that accommodate recurrent event data. They introduce four key components that can be used to evaluate modeling choices across different contexts. According to Kelly and Lim (2000), we must consider the risk intervals, baseline hazard, risk set, and within subject correlation in the context of the research. Each is summarized below and its specific relevance to series hazard modeling is addressed.

Risk Intervals

Kelly and Lim (2000) explain that risk intervals refer to the method used to measure time until each recurrent event. For analyses where each subject can only fail once, the duration is typically measured as the number of time units until failure (e.g., days, weeks, or months) or (0, d] where 0 is the time that the subject is first measured—perhaps at the end of a treatment—and d is the time at failure. Figure 2 provides an illustration of the failures of two subjects, Sid and Nancy, over a 20-week-period. Sid failed three times, and Nancy only failed once. Because each subject is able to fail more than once, the risk interval for later failures can operate in one of three ways. First, time can be reset to zero after the first failure ignoring the duration prior to the first failure. *Gap Time* can be depicted by the interval (0, d_2-d_1], where d_2 is the time at second failure and d_1 is the time at first failure. For example, Sid's second risk interval is (0, 4], because $11 - 7 = 4$. Second, time can continue after the first failure without resetting. *Total Time* for the second failure is the same as it would have been had the subject never failed (0, d_2]. Sid's risk set for his second

Fig. 2 The failures of Sid and Nancy



failure is (0, 11]. This measurement is especially useful if the analyst wants to preserve the time scale since the beginning. The third type of risk interval is called *Counting Process*. It combines both Gap Time and Total Time by relying on the same time scale as Total Time and allowing the subject to start over after the first failure like Gap Time. Thus, the time at risk for the second failure is $(d_1, d_2]$, or (7, 11] for Sid (Kelly and Lim 2000; see also Ezell et al. 2003).

For the series hazard model, it makes little sense to preserve the time scale for each failure because it would impose an ordering to each event that is based exclusively on the sequence of failures and not on the risk since the time of the previous event. If the timing of events is relevant to the model, it can easily be incorporated as a covariate in **X**; for example by including a measure of the cumulative number of failures, or by directly measuring the daily, weekly, or monthly count since the beginning of the series. Series hazard modeling is expressly designed to model the time since the last failure until the next failure, which is only measured using Gap Time.

Baseline Hazard

Kelly and Lim (2000) point out that the baseline hazard for subsequent failures can remain the same as that for the first failure, or it can be different. One could easily imagine why the baseline hazard might change after the first failure in more traditional settings. After experiencing a heart attack, the subject's risk of another increases. Or, once arrested, the chance of future arrests increase, now that the subject has been flagged by the criminal justice system. However, within the context of series hazard modeling, it makes little sense to allow the baseline to change after each failure. With only one unit, the only variation is across events. Thus, allowing the baseline to change with each failure is analogous to allowing it to change across units in a more standard hazard model. Instead, we assume one baseline hazard and attempt to model changes through the covariates in **X**.

Risk Set

The risk set is comprised of individuals who are at risk at the time of each failure, depicted by R_i in Eq. 2. When we allow failures to repeat for each individual, the risk set varies depending on how the risk interval is defined because the risk set only includes members of the current interval. According to Kelly and Lim (2000), the risk set can be unrestricted, restricted, or semi-restricted. If it is unrestricted, a subject can be represented multiple times in a risk set. Whether the subject is measured using the gap time, total time, or counting process formulation, the risk set for an event at time *t*, includes all time intervals that have yet to experience their next failure by time *t*, including an earlier or later event for the same subject (Kelly and Lim 2000). For example, because, according to Fig. 2, Sid experienced failures at weeks 7, 11, and 17, and he is censored at week 20, his gap time failures are (0, 7], (0, 4], (0, 6], and (0, 3]. Because his second interval ends at 4 weeks, the risk set for the second failure would include his first three intervals as well as both of Nancy's intervals (0, 9] and (0, 11]. Sid's last interval is excluded from that risk set because it ended in 3 weeks. Similarly, the risk set for Sid's third failure includes his first interval, but not his second or fourth. Had we measured failure using the total time formulation with risk intervals (0, 7], (0, 11], (0, 17], and (0, 20], the risk set Sid's second failure would only include his latter three intervals and Nancy's second because they had yet to occur in real time. Had we measured failure using the counting process formulation with counting process failure intervals at (0, 7], (7, 11], (11, 17], and (17, 20], the risk set for Sid's second failure would only include the interval for Sid's second failure and Nancy's second interval because all others fail to overlap with Sid's second interval. The baseline hazard for an unrestricted risk set is the same for each failure (Kelly and Lim 2000).

Subjects in a restricted risk set can only be included once for each risk interval; and they are only included in the risk set for the *k*th failure if they have already experienced $k-1$ failures (Kelly and Lim 2000). For example, because Nancy experienced only one failure at time 9, her risk set for that failure, regardless of how the interval is formulated, would only include her first failure and the interval for others' whose first failures had not yet occurred at time 9. Thus, Sid would not qualify for Nancy's first risk interval. However, in the restricted set, Nancy would be included in Sid's first and second risk intervals, regardless of formulation because Nancy first failed after Sid's first failure; and Nancy had yet to experience her second failure at the time that Sid failed for a second time. The baseline hazard for each failure is specific to that ordered event (Kelly and Lim 2000).

According to Kelly and Lim (2000), semi-restricted risk sets also have event-specific baseline hazards. These sets also only allow each subject to be included only once for each failure, however, they also allow them to be included in the risk sets for the *k*th failures even if they hadn't experienced their $k-1$ failure by including them as a "dummy interval" (Kelly and Lim 2000). Furthermore, semi-restricted risk sets can only be applied to total time and counting process formulations. In our example, the period after Nancy's first failure can be included as a "dummy" risk interval in the risk set for Sid's third failure. However, this accommodation is only one-way. Sid's first risk interval fails to qualify for Nancy's risk set for her second failure, which is censored.

When considering which risk set is most appropriate to the series hazard model, two reasons make it clear that the unrestricted is best. First, because we are relying on a common baseline hazard and only the unrestricted accommodates this. More practically, because the series model only has one unit with multiple failures, the risk set must include the intervals for the other failures regardless of their order of failure. By only letting each subject only appear once in a risk interval, the series model will not work with either the restricted or semi-restricted risk set. Thus, by using the unrestricted risk set, a failure that has yet to occur by day 12 includes all other intervals with failure times of 12 or more days.

Within Subject Correlation

Clearly when we include repeated measures from the same subject in an analysis, those values will be more correlated than the values across subjects. In fact, if the subjects were selected randomly, we expect the cross-subject correlations to be zero. Ezell et al. (2003) frame these correlations in terms of two much discussed concepts, state dependence and population heterogeneity (Land et al. 1996; Nagin and Paternoster 1991, 2000). State dependence refers to the contagion of behavior within a subject that leads to within subject correlation. In other words, a subject's risk of a current failure is directly affected by his or her experience with past failures (Ezell et al. 2003). Unobserved heterogeneity within the context of hazard modeling explains the within subject correlations as due to the constant, but different hazard rates across units. Some subjects consistently have lower baseline hazard rates while others consistently have higher (Ezell et al. 2003). Regardless of the source of correlation, by ignoring it, the standard errors will be biased downward, possibly leading us to erroneously conclude statistical significance (Box-Steffensmeier and Zorn 2002; Ezell et al. 2003).

The most relevant distinction between the recurrent hazard and series hazard models is that the series approach only models one unit, making efforts to adjust for within subject correlation irrelevant. While, dependence across failures remains in the model, it need not be interpreted as limiting. In fact, the point of the series hazard model is to estimate the source of this dependence through covariates, which is described by Kelly and Lim (2000) as the conditional approach to account for within subject variation.[2] By explicitly modeling the correlation, we can estimate changes in the hazard function due to changes in conditions at the time of each subsequent failure. More practically, we can estimate changes in the failure rate after an intervention.

One strategy to capture the across-failure correlation is to include measures of previous failures as covariates in order to establish conditional independence. Analogous to the lagged dependent variable in time series, one measure could be the time until the most current observation. In fact, Allison (1995) suggests including that measure as a test for dependence across recurrent failures. Other useful measures are the cumulative number of failures to date or the number of months, weeks, or days that have passed since the first failure. Ideally, the measure will also have theoretical relevance. For example, in Dugan et al. (2005) we included a measure of success density, which was a function of the success rate of the three most recent hijackings and the number of days between the first and the third. This functionally measures some of the dependence across failures while testing the predictions of contagion (Holden 1986).

The Series Hazard Model

In practice, the series hazard model is estimated using the Cox proportional hazard model, except that it is estimated across failures instead of subjects. Thus, the hazard function for the series hazard model, shown in Eq. 3, looks almost exactly like the hazard function of Eq. 1, except that now the subscripts represent failure k instead of individual i.

$$\lambda_k(t|\mathbf{X}_k) = \lambda_0(t)\exp(\mathbf{X}_k\boldsymbol{\beta}) \tag{3}$$

Because it is important to directly model the history of previous failures in order to assure conditional independence, Eq. 3 can be expanded to the form shown in Eq. 4; where $\mathbf{Z}$ can be any function, or series of functions of previous failures that are theoretically relevant to the current analysis.

$$\lambda_k(t|\mathbf{X}_k) = \lambda_0(t)\exp(\mathbf{X}_k\boldsymbol{\beta} + \mathbf{Z}_k\gamma) \tag{4}$$

The partial likelihood function that is derived from the series hazard function in Eq. 4 is shown in Eq. 5. Notice that once again, the baseline hazard function $\lambda_0(t)$ drops out, making its specific form irrelevant.

$$L(\boldsymbol{\beta}) = \prod_{k=1}^{f}\left(\frac{\exp\left(\mathbf{X}_{(k)}\boldsymbol{\beta} + \mathbf{Z}_{(k)}\gamma\right)}{\sum_{l\in R_t}\exp(\mathbf{X}_l\boldsymbol{\beta} + \mathbf{Z}_l\gamma)}\right) \tag{5}$$

The matrix $\mathbf{X}_{(k)}$ includes information on the specific failure; the social, political, or policy context at the time of failure (including the intervention profile); and $\mathbf{Z}_{(k)}$ measures

[2] Kelly and Lim (2000) discuss three approaches to account for within subject correlation: conditional, marginal, and random effects. We use the conditional approach when we include time dependent independent variables to capture the structure of dependence across the recurrent failures. For example, the analyst may include a variable that measures the number of previous failures. The marginal approach, also referred to as variance-correction, models the data as if all failures are independent, and then calculates the robust standard errors by using a sandwich estimator (Kelly and Lim 2000; Box-Steffensmeier and Zorn 2002). The random effects approach, also referred to as frailty models, introduces a random covariate into the model which accommodates the dependence across repeated failures (Kelly and Lim 2000).

information about the history of failures to account for dependencies across failures. Because the series hazard model is virtually the same as the Cox proportional hazard model without recurrent events, all of the same diagnostics apply to this model. Most relevant is that the same tests can be used here to test for the proportionality of the hazard over time.

Recall that the failure times are measured in gap time, which means that the clock starts over after each failure. Just like the standard Cox proportional hazard modeling, this model will have tied data—or multiple failures at the same point in time, which produces duplicate failure intervals. Since each duplicate observation will have its distinct set of covariate values, these differences must be resolved prior to estimation. Most statistical packages can easily resolve tied data by either accounting for all possible orderings (exact marginal) or by adopting a multinomial strategy to estimate the likelihood function (exact partial) (Allison 1995). Regardless, these strategies also work well with the series hazard model.[3]

A specific tie that is handled differently is when two or more failures occur on the same day. The default values of the gap time will depend on the ordering each failure in the dataset, which is usually arbitrary. For example, if failures A and B occurred on the same day while failure C occurs three days later and A is listed prior to B, then the gap time from A to B is (0, 0]; and the gap time from B to C is (0, 3]. Standard statistical packages will remove those cases with gap times of (0, 0] before estimating the model. While this may be appropriate for cross-sectional hazard modeling because this indicates immediate failure, for series hazard modeling it arbitrarily selects one case over others based on the order of data entry. Although this type of selection could be random, removing cases does reduce important information and statistical power that can be used to better estimate the model. Information and power need not be lost. The scholar can consider the context of the failures and recode the gap time failures (0, 0] to accurately reflect the next time until failure. For example, the case study to follow estimates the hazard of aerial hijackings across the globe from 1947 to 1985 using a gap time of days until the next hijacking. When two or more hijackings occur on the same day, it is highly likely that they were either a coordinated event or the planners were unaware that other hijackings were also planned for that day. In either of these contexts, the best measure of time until the next hijacking would be the days beyond the current day until the next event. Thus, for the example below all events with gap time failures of (0, 0] were recoded to have the same gap time as the last event listed on their days.[4]

A second issue with the series hazard model is how to estimate intervention effects that change over time. This possibility is accommodated in interrupted time series with a transfer function that measures whether the effect was immediate or gradual and whether the change was permanent or temporary (McDowall et al. 1980). In series hazard modeling, changes in intervention effects can be estimated by allowing the hazard to be a function of time since the beginning of the intervention period. This can easily be accomplished by including a variable that counts the time since the beginning of the study, and then interacting this count variable with an indicator for the intervention period. Similar to its interpretation in interrupted times series, as demonstrated by Lewis-Beck (1986), the interaction will capture any trend that is linearly increasing or decreasing, which is analogous to a gradual or temporary effect measured in interrupted time series (Cleves et al. 2004).

When using this strategy the scholar should keep in mind that the unit of observation is each failure, not each time increment. In other words, a failure must occur in order to

[3] Analysis of high frequency events would likely have much tied data. One strategy to avoid this problem would be to estimate the time until the next N events instead of the subsequent event.

[4] As it turns out, in this example the substantive findings are the same whether the (0, 0] cases are deleted or recoded and retained.

measure a change in a specific period of time. Thus, changes in hazard are not measured by evenly spaced increments as they are when interactions are used in time series. This should not be problematic, because the time counter is measured by evenly spaced increments—such as days, weeks, or months, the estimated interaction effect will accurately measure the variation between failures. However, if failures are too infrequent, variation might be too small to detect any changes in the hazard. For example, if an intervention successfully reduces failures, its dummy variable may be a vector of mostly zeroes making its coefficient difficult to estimate. Simple tabulations can reveal this type of problem.

A related problem might arise if the intervention has no clear end date. For example, while negotiations take place over a discrete period of time, they are expected to promote peace well after the negotiation period ends. Yet, it may be naïve to assume that the effects will extend to the end of the series. LaFree et al. (2009) faced this problem when estimating the effect of several British interventions on republican terrorist attacks in the United Kingdom. The British initiatives of Falls Curfew, the Loughall attack, and the Gibraltar attack were short-lived interventions that were expected to have a relatively lasting impact. To address this issue, the authors set the intervention period for a year, which was long enough to capture any effects and short enough to not overlap with later interventions, and then conducted sensitivity tests to assure the robustness of the findings to the length of the intervention period.[5]

The next section demonstrates the value of modeling repeated failures using series hazard modeling by comparing efforts to estimate the effects of two interventions in worldwide aerial hijacking from 1947 to 1985. I first use interrupted time series and then series hazard modeling.

The Case of Global Aerial Hijackings

In this case study I raise the question of which analytical strategy is better suited to estimate the impact of two interventions designed to reduce the incidence of aerial hijacking, interrupted time series or series hazard modeling. For each method, I estimate the effects of installing metal detectors and adopting a Cuban crime law separately, and then together.

Data

The data include all attempts to hijack an airplane worldwide from 1947 to 1985. These cases were compiled from data supplied by the Federal Aviation Administration (FAA), the Global Terrorism Database (GTD) and the RAND Chronology (see Dugan et al. 2005, for a detailed description of the data). This series begins with the second non-war-related hijacking on July 25, 1947 when three men hijacked a Romanian flight, landing the plane in Turkey, and killing one member of the crew. The first recorded aerial hijacking occurred 16 years earlier in Peru in 1931 when revolutionaries commandeered a small plane to shower Peru with propaganda pamphlets (Arey 1972).[6] After the 16 year hijacking hiatus,

[5] The models were re-estimated after extending the end date by monthly increments for up to 36 months. Graphical displays of the estimates with 95% confidence bounds show that the findings were robust to any end date.

[6] Arey (1972) records the year of this event as 1930, however, current FAA records record it as August 31, 1931.

airplanes began to be hijacked more regularly making the year of the second hijacking a more reasonable starting point for the series. The data end in 1985 because that is the last year that the FAA kept detailed records of each event. Further, because many of the hijackings were motivated by circumstances other than terrorism, the missing details in these cases are also absent from the GTD and RAND data.

The intervention variables in both analyses are recorded as occurring on October 31, 1970 and February 5, 1973 (see FAA 1983). In October of 1970, the Cuban government amended their law making hijacking a crime. This law was timely given the recent increases in flights hijacked to Cuba in the late 1960s (Arey 1972). The second intervention represents a series of changes that were made between January 5th and February 5th, 1973. On January 5, 1973, metal detectors were installed in U.S. airports and, though the dates and times differ substantially, similar devices were gradually introduced to major airports around the world. About a month later, on February 3, 1973, the United States and Cuba signed a Swedish-brokered agreement that defined hijacking as a criminal act in both nations and promised to either return hijackers or put them on trial. Finally on February 5, 1973 the FAA required that local law enforcement officers be stationed at all passenger check points during boarding periods. Because these three intervention dates are close enough to be indistinguishable over the series, I preserve the last date, and refer to it as the implementation of metal detectors for simplicity.

The dependent variable for each analytical method measures in some way, the frequency of hijacking events. For the interrupted time series, the events are aggregated by year, quarter, and month to determine the sensitivity of the results to the temporal unit. Because the frequency of hijacking events has a large right skew, it is logged to make the distribution more normal.[7] The dependent variable for the series hazard model is the number of days until the next hijacking or the gap interval of $(0, d_{k+1}]$. As noted above, when two ore more events occur on the same date producing at least one interval of (0, 0], I replace the all zero interval with that for the last event that is listed on that date. This forces all hijackings that occurred on the same date to be treated as tied data. The exact marginal method is used to resolve ties.

The times series analyses have only one independent variable aside from the two interventions mentioned above. To account for the large change in aerial hijackings beginning in 1968, I include an indicator for the pre-1968 period, which I would expect to be negatively related to hijacking. A linear trend could also be used; however, the patterns shown later in Fig. 3 suggest non-linearity. The series hazard model has more independent variables (see Dugan et al. 2005 for a more complete model). First, because each observation describes a specific event, I include a subset of incident-specific measures. These include success density, which was constructed by combining the success rate and time span of the seven most recent hijackings (Success Density);[8] an indicator of whether the event was motivated by terrorism (Terrorist); an indicator of whether the flight was private

[7] See Brandt and Williams (2001) and Brandt et al. (2000) for introductions to the Poisson AR(p) model and the Poisson exponentially weighted moving average model, which better accommodate non-normal data in time series.

[8] Success density is measured by taking the current and six previous flights, and calculating the proportion of those flights that were successful over the number of days spanning the seven events and then calculating the natural logarithm after adding a value of 0.05 to avoid missing values. Thus, a large success density indicates that most events were successful over a relatively short period. This measure differs from the one used in Dugan et al. (2005), which only used the three most recent hijackings. I found that the one based on 7 flights better demonstrates the model and the natural logarithm is more linearly related to the Martingale residuals (Cleves et al. 2004). Both serve the same substantive purpose.

(Private); and a logged count of days since the previous hijacking (Last Hijacking).[9] This last variable and Success Density each serve both theoretical and instrumental purposes. Theoretically, it may be reasonable to expect the hazard of a third hijacking to be lowered after two events that are relatively close together because security at airports is likely high. And conversely, after a relatively short interval of successful hijackings, a potential perpetrator might be motivated to attempt his or her own. Instrumentally, as noted above, by directly modeling the dependency across subsequent events, the error terms are more likely to be conditionally independent (**Z** in equations 4 and 5).

The first of two control variables is also included to support conditional independence. It measures the depth of the incident into the series. Because I am interested in interacting time with the above interventions, I use a monthly count from January 1947 for this purpose. I also calculated the cumulative number of hijackings and found this to be correlated with the monthly count variable at 0.99, suggesting that the monthly count will adequately capture the cumulative number of events, while remaining exogenous to the model. The second control variable is the pre-1968 indicator to differentiate the less frequent period from the remainder of the series.

Estimating Policy Effect Using Interrupted Time Series

When beginning interrupted time series analysis with incident-level data, one must decide to which time unit to aggregate. Figure 3 displays the count of hijackings aggregated to the year, quarter, and month from the beginning of 1947 to the end of 1985. At one end of the spectrum, we could choose to aggregate to the year—as shown in the top graph, which provides an appealing smooth trend. However, we would be relying on only 39 observations to estimate any effects. Even if the policy was effective, the statistical power would likely be too low to significantly distinguish the effect. On the other end of the spectrum, by using monthly aggregations—as shown in the bottom graph, with 468 observations we would likely have sufficient statistical power to identify any effects, even the more subtle. However, with 45% of the months experiencing no hijackings, data sparseness might make it too difficult to identify any policy impacts. By aggregating to the quarter, we might strike a balance between the two extremes of enough statistical power ($n = 156$) and relatively little sparseness (only 28% of the quarters have no hijackings) allowing us to efficiently estimate any policy effects. For demonstrative purposes, I will apply interrupted time series analysis to all three units.

Also note that prior to 1968 aerial hijackings of flights from U.S. airports were sporadic. After this period there was a preponderance of flights hijacked to Cuba (Arey 1972). To separate this period from the remainder of the cases, I include in the model an indicator of the pre-1968 period. Thus, the error structure of the ARIMA is determined after adjusting for that period, which is also referred to as ARMAX (Autoregressive Moving Average with eXogenous inputs; Stata Press 2005). Fitting an ARMAX model to logged frequency of hijacking, while controlling for the pre-1968 period, I find that the yearly data fit an AR(1) error structure; and the quarterly and monthly data both fit an ARMAX(1,0,1) error structure. After correcting for this, the residuals in all three models are now basically white noise, clearing the way to detect differences in the pre- and post- intervention series (see, Dugan 2010, for a thorough discussion of how to fit ARIMA models).

[9] The natural logarithm of the count of days plus 0.05 is used because it produces a more linear relationship with the Martingale residuals (Cleves et al. 2004).

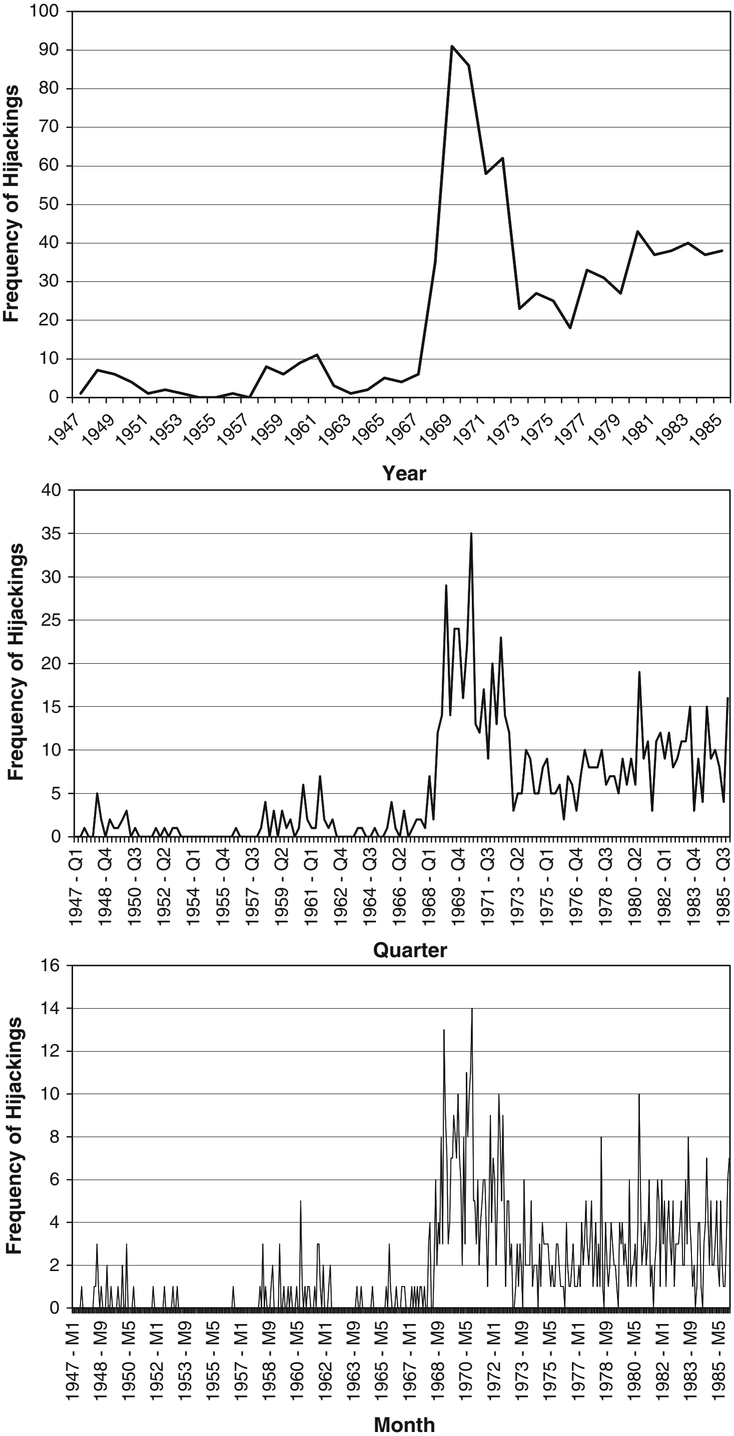


Fig. 3 Yearly, quarterly, and monthly aerial hijackings worldwide 1947–1985

Prior to testing the effects of the 1973 implementation of metal detectors, we first need to decide whether to lag the year, quarter, or month of the implementation (i,e., measure it in the following period). As Fig. 1 demonstrated, by measuring the unit of implementation concurrently in the series we risk confounding the temporal ordering of the independent and dependent variables. For example, in a quarterly series, hijackings in January preceded the February 5, 1973 implementation date. However, if the implementation is measured during the first quarter, January hijackings would be interpreted by the model as following the implementation. By lagging the year, quarter, or month, we impose a temporal ordering that forces the implementation to be measured after to all hijackings during the implementation period. Yet, as discussed earlier, by lagging, we might fail to detect any immediate intervention effects, which is especially problematic if the overall impact is short-termed. Because the first of the policies implemented in 1973 actually began on January 5, and because the date is at the beginning of the year, quarter, and month (making it more akin to example *a* in Fig. 1), I measure the intervention during the period it was implemented, choosing not to lag the intervention (1973, first quarter, February 1973).[10]

I now estimate whether the effect of the metal detector intervention abrupt and permanent, abrupt and temporary, or gradual and permanent (Cook and Campbell 1979; McDowall et al. 1980). When using the quarterly and monthly data I find a gradual, yet permanent effect; and when using the yearly data I find no evidence of any effect. Table 1 presents the coefficient estimates and standard errors for each. Because the dependent variable is logged, we can interpret the coefficient estimate as a change in the percent of hijackings for that variable. Not surprising, the hijackings prior to 1968 are systematically lower regardless of the unit of analysis. However, its magnitude depends on the granularity of the unit, producing smaller effects at more refined levels of aggregation. When examining the effects of metal detectors, we see that regardless of the unit of analysis, the percent of hijackings appears to have dropped after metal detectors are implemented—as denoted by the negative estimates. However, the drop is only statistically significant for the quarterly and monthly analysis, which is unsurprising given the low statistical power of the yearly model. The rate parameter is estimated around 0.67 and 0.83 for the quarterly and monthly analyses, respectively, suggesting that the drop in hijackings after the implementation of metal detectors is more gradual in the monthly series (McDowall et al. 1980). The difference in rate is expected because it simply means that it takes more months than quarters to reach the asymptotic change. When we calculate the asymptotic change, we see that the difference virtually disappears once the implementation drops by 0.69 and 0.64% according to the quarterly and monthly analyses, respectively.[11]

On October 31, 1970 Cuba made hijacking a crime, arresting those who land their hijacked planes on Cuban soil. I first test for the effects of this legislation by including it without the metal detector intervention. Since October 31 is late in the year, near the middle of the quarter, and late in the month, that measure is lagged by one unit to avoid simultaneity bias (see examples *b* and *c* in Fig. 1). Table 2 presents the estimated coefficients for the effects of Cuban Crime on the hazard of logged hijackings. Once again, the yearly model fails to detect any statistically significant relationship between the intervention and hijacking regardless of how it is operationalized, although the directionality is as expected. The effects of Cuban Crime in the quarterly and monthly model are gradual, yet permanent both producing an asymptotic drop of just over 1%.

[10] The results are similar when the intervention is lagged.

[11] The asymptotic change in hijackings was calculated using $\frac{\omega}{(1-\delta)}$, where ω is the coefficient estimate for Metal Detectors and δ is the estimated rate parameter (McDowall et al. 1980).

Table 1 Coefficient estimates for time series models on logged hijackings (SE)

Variable	Yearly ($n = 39$)	Quarterly ($n = 155$)	Monthly ($n = 467$)
Metal detectors	−0.527 (0.908)	−0.227* (0.123)	−0.112** (0.029)
Rate	–	0.673** (0.095)	0.825** (0.038)
Pre-1968	−2.836** (0.545)	−0.893** (0.284)	−0.358** (0.075)
Asymptotic change[a] after Met. Det.	–	−0.694	−0.640
Error structure	ARMAX(1,0,0)	ARMAX(1,0,1)	ARMAX(1,0,1)

* $p < 0.05$; ** $p < 0.01$, one tailed

[a] See note # for this calculation

Table 2 Coefficient estimates for time series models on logged hijackings (SE)

Variable	Yearly ($n = 39$)	Quarterly ($n = 155$)	Monthly ($n = 467$)
Cuban crime	−0.220 (0.948)	−0.295* (0.135)	−0.122** (0.023)
Rate	–	0.731** (0.083)	0.894** (0.023)
Pre-1968	−2.647** (0.646)	−0.845** (0.221)	−0.272** (0.049)
Asymptotic change[a] after Cuba Cr. law	–	−1.097	−1.151
Error structure	ARMAX(1,0,0)	ARMAX(1,0,1)	ARMAX(1,0,1)

* $p < 0.05$; ** $p < 0.01$, one tailed

[a] See note # for this calculation

While these findings are revealing, they fail to inform us about the whether the effects of each intervention would persist had we directly accounted for the other's effect. Yet, the timing of the interventions might be too close to discern independent effects. In fact, if we include both in the yearly model, the Cuban Crime variable would only differ from the Metal Detector by three observations (years), clearly making the model vulnerable to multicollinearity (see Lewis-Beck 1986, for a discussion of multicollinearity in interrupted time series). This issue may be less problematic for the quarterly and monthly data because Cuban Crime differs from Metal Detectors by nine quarters and 27 months, respectively. As expected, when both interventions are included in the models, the findings are inconclusive. The coefficient estimates for Cuban Crime fail to reach statistical significance in the yearly and quarterly models, producing p values greater than 0.80. The estimate for Metal Detectors remains significant in the quarterly equation. However, in the monthly model the pattern of significance reverses. The Cuba Crime estimate reaches statistical significance, but at the cost of the estimate for Metal Detectors. The correlation for those two variables is 0.88 making it unreasonable to confidently draw any conclusions in that model.

Thus, despite detecting a drop in hijacking after each intervention when modeled separately, and because of their temporal proximity, I am unable to isolate the effects of one while conditioning on the other. In other words, it is unclear whether the calculated asymptotic drop in hijackings after the implementation of metal detectors shown in Table 1 is also picking up on the drop in hijackings due to the adoption of the Cuban law 3 years earlier, or vice versa. In the following section, I estimate the effects of metal detectors and the Cuban crime law in the same model using the series hazard modeling method.

Estimating Policy Effect with the Series Hazard Model

When analyzing the data using series hazard modeling we rely on the disaggregated data and, therefore need not worry about levels of aggregation. The model is estimated using the variation between hijacking events, not the frequency of temporally clustered events. These two methods clearly correspond, because those hijackings that fall in the high frequency months will have shorter gap times between events compared to those hijackings that fall in the low frequency months. By using that timing between events as the dependent variable and modeling it using a Cox proportional hazard model or the series hazard model, we can estimate how the adoption of different policies changes the baseline hazard of continued hijacking events. Interactions with a temporal count variable are used to determine if the policy effect is gradual, immediate, temporary, or permanent. Furthermore, because the unit of analysis is the hijacking event, we no longer have to assume that all hijackings are the same. Differences can be directly modeled by including specific characteristics of the current hijacking event that might contribute to changes in the baseline hazard.

Table 3 lists the estimated coefficients for the hazard models using various combinations of policies and interactions. Before turning to the policies, I will first discuss the results for the control variables and the flight characteristics, as a way to assess whether the models are behaving as expected. Turning first to the bottom of the table, we see that regardless of the model, the estimated coefficients for Pre-1968 are significantly negative across all five models. This finding matches what is expected because there were relatively fewer hijackings prior to 1968, implying a lower baseline hazard. After controlling for the earlier years, the findings for models 4 and 5 suggest there may be a slight increasing trend of hijackings that changes after the Cuban crime law is adopted. That trend will be discussed further below.

Turning to the flight characteristics, we see that the only consistently significant characteristic is the density of successes for recent hijackings. This suggests that when the most recent seven hijackings are successful and temporally clustered, the hazard of another hijacking increases—supporting others' hypothesis of hijacking contagion (Holden 1986; Dugan et al. 2005). In contrast, after controlling for success density, there is marginal evidence that when two hijackings occur temporally close to one in other, the hazard of the third is lowered. The two remaining characteristics are surprisingly null. Recent terrorist motivated hijackings fail to dissuade continued hijackings; and hijacked private planes have no influence on the hazard of the next event. Basically, the primary influencing characteristic is the perceived success of recent hijackings.

In light of the multicollinearity issues raised in the interrupted time series models, I first present the hazard models with the two policies modeled alone, with interactions, and then together. Models 1 and 2 present the estimates for the Metal Detector main effect and interaction with monthly count. Given the statistical insignificance of the interaction, the main effect model adequately estimates the Metal Detector effect. By taking the exponent of the estimate, we can conclude that after metal detectors were implemented the hazard of continued hijackings dropped by just over half (h.r. $= \exp(-0.715) = 0.49$; drop $= 1 - 0.49 = 0.51$). When we estimate the change in hazard after the Cuban crime law was adopted in Model 3, we see that the hazard dropped by 34% $(1 - \exp(-0.422))$. However, when I include the interaction of Cuban Crime with the monthly count variable, the main effect is only marginally significant but the hazard drops by one-half of one percent with each subsequent month. Taking a step back, we can see that the relatively flat trend now increases once we include this interaction, which essentially allows for a break in the trend

Table 3 Coefficient estimates for serial hazard models on days until next hijack (SE) ($n = 821$)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Policies					
Metal detectors	−0.715** (0.147)	−0.932** (0.339)	–	–	−0.652** (0.154)
MD × M. count	–	0.001 (0.001)	–	–	–
Cuban crime	–	–	−0.422** (0.161)	0.524† (0.288)	0.305 (0.287)
CC × M. count	–	–	–	−0.005** (0.001)	−0.004** (0.001)
Flight characteristics					
Success density	0.150** (0.038)	0.137** (0.043)	0.122** (0.045)	0.119** (0.045)	0.094* (0.043)
Terrorist	0.021 (0.100)	0.017 (0.100)	−0.041 (0.100)	−0.051 (0.100)	−0.034 (0.100)
Private flight	−0.028 (0.119)	−0.031 (0.119)	−0.080 (0.118)	−0.070 (0.118)	−0.029 (0.119)
Last hijacking	−0.056 (0.036)	−0.057 (0.036)	−0.083* (0.035)	−0.070* (0.035)	−0.051 (0.036)
Control variables					
Monthly count	0.001† (0.000)	−0.000 (0.001)	0.000 (0.000)	0.004** (0.001)	0.005** (0.001)
Pre-1968	−1.807** (0.200)	−1.896** (0.235)	−1.967** (0.208)	−1.453** (0.251)	−1.560** (0.249)

† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$, one tailed

after the Cuban law is implemented. As noted above, with the added flexibility, the series of hijackings show a monthly increase in the hazard of hijacking of about two-fifths of one percent (0.4%), until the Cuban law is passed. At that point, the increase is offset by a drop in the hazard by just a little more (0.5%). A likelihood ratio test ($p = 0.0000$) and a comparison of each model's AIC and BIC statistics all conclude that Model 4 is a better fit than Model 3.[12]

Finally, Model 5 presents the results for the model that estimates both policies together. Despite the relatively high correlation between the two main effects ($r = 0.72$), there is enough independent variation to produce estimates of each policy's marginal contributions to the hazard of continued hijackings.[13] The statistically significant policy effects found in Models 1 and 4 remain significant in Model 5. Finally, a test for proportionality based on the Schoenfeld residuals concludes that the model meets the proportionality assumption (Cleves et al. 2004). In fact, any other diagnostic that is used for hazard modeling can be used with the series hazard model.

Comparing Models

When we compare the substantive findings of the times series and series hazard models we draw the same conclusions. After the Cuban crime law was adopted in 1970 and metal detectors were installed in 1973, the rate of aerial hijackings had slowed down. However, had I relied exclusively on the interrupted time series analysis, I would only have been able to speculate that each policy separately contributed to the decline in hijacking. By including both policies in the final model, I was able to disentangle the specific contribution of each. With the insight gained in these findings, we can now see that the estimates in Tables 1 and 2 were substantively correct, but the magnitudes were likely inflated. Each likely includes the drop in hijackings due to the omission of the other policy.

Having said this, it would be misleading to suggest that interrupted time series can only accommodate single interventions while series hazard modeling can always accommodate many. With sufficient timing between interventions each method can adequately estimate multiple effects (D'Alessio and Stolzenberg 1995; Lewis-Beck and Alford 1980). Furthermore, interventions that are implemented closely in time will fail to offer distinct enough variation to disentangle effects for either series hazard modeling or interrupted time series. As the hijacking example shows, it was impossible to estimate separately the effects of metal detectors (installed on January 5, 1973), the U.S.-Cuban agreement (February 3, 1973), and the presence of law enforcement at checkpoints (beginning on February 5, 1973).

A second weakness of the interrupted time series model is that its covariates are required to be measured at the temporal unit. It is easy enough to acquire data on yearly changes on macro economic, social, or political conditions that might contribute to the changing patterns of hijacking. However, once we refine our data to the quarter or month, exogenous variables become more difficult to obtain. Conversely, because incident-level data lift the restriction to only temporal covariates, event characteristics can easily be

[12] For the restricted model, AIC = 5067.388 and BIC = 5100.361; and for the unrestricted model, AIC = 5054.063 and BIC = 5091.747.

[13] In a separate equation that includes the Metal Detector interaction with Monthly Count, the findings remain virtually the same. A likelihood ratio test concludes that model 5 is a better fit than the more flexible model that includes the Metal Detector interaction ($p = 0.2204$). Furthermore, when comparing model 5 to a more restrictive model that only includes the two policy interaction, model 5 is once again preferred ($p = 0.0018$).

added to the series hazard model.[14] We learned in this second analysis that the risk of continued hijackings does rise after a surge of successful hijackings. We also learned that, at least between 1947 and 1985, the contempt generated by terrorist hijackings fails to dissuade others from attempting to hijack airplanes. In their investigation of Armenian terrorism Dugan et al. (2009) find that the hazard of ASALA attacks drop after unsuccessful attacks. Further, they were able to test their hypotheses that ASALA's choice of targets influenced their hazard of continued attacks. As expected, they found that target choice had very different effects before and after the Orly Airport attack on July 15, 1983. Prior to the Orly attack, ASALA attacks against non-Turkish targets seemed to have perpetuated more terrorist attacks by that organization. However, after the Orly Airport attack, ASALA attacks against non-Turkish targets reduced their momentum. The authors expected this difference because Armenian Diaspora and western nations who were previously sympathetic to ASALA, vehemently denounced the attack. Once ASALA was more vulnerable to dissent, they no longer could attack non-Turkish targets without consequence (Dugan et al. 2009). Only with an incident-level analysis could this hypothesis been tested.

Finally, because we are able to preserve the characteristics of the individual events, if theoretically sound, the sample can be further refined to only include more homogeneous events. Dugan et al. (2005) modeled subsets of this hijacking data to only examine flights that originated from the U.S., those that originated elsewhere, flights diverted to Cuba, terrorist motivated hijackings, and non-terrorist motivated hijackings. Dugan et al. (2009) examined only those global terrorist attacks perpetrated by ASALA and JCAG. Current work by Chen and colleagues (Chen et al. 2008) uses series hazard modeling to estimate the hazard of *any* attack in Israel, while LaFree et al. (2009) restricted attacks in Northern Ireland to only those perpetrated by republican terrorists. In essence, by modeling data using the series hazard method, the researcher gains nearly all of the flexibility that is inherent to modeling incident-level data. One limitation of the series hazard model compared to other individual-level models is that the only variation measured in the dependent variable is the duration until the next event, treating all "next events" the same regardless of their specific characteristics. For example, while we can estimate how the hijacking of a private plane affects the hazard of another hijacking, we cannot estimate how the hijacking of a private plane affects the hazard of another hijacking of a private plane compared to that of a commercial plane; unless we conduct separate analyses on two different datasets. In other words, all of the incident level variables capture the characteristics of the current event (or independent variable), not the subsequent event (or dependent variable). Having said this, this limitation may be addressed by incorporating competing risks into the model (Allison 1995).

Conclusion

In this paper, I introduce an alternative to event-based interrupted time series that is better equipped to accommodate the need to estimate the impact of multiple interventions while preserving detailed information about each event. The series hazard model is intuitively

[14] Temporal covariates can also be included in the series model. In another paper by LaFree et al. (2009), the authors model annual measures of the number of troops in Northern Ireland and the number of reported crimes. Because republican terrorists also combated loyalist terrorist groups we included the monthly total of loyalist attacks from the previous calendar month, which was found to be marginally significant.

appealing because it exploits the variation between events rather than relying on tallies generated from artificial divisions in time. Further it directly models event specific characteristics and any contagion between events through its covariates. Finally, series hazard modeling is most appealing for its simplicity. One can simply run the Cox proportional hazard model on failures over time rather than over units. All estimation and post-estimation procedures for the Cox model apply to the series model.

Series hazard modeling is not appropriate for all time series data. It only works for event specific data that are typically aggregated to a temporal unit. Thus, it is inappropriate for estimating changes in rates, values, or other substantive measures. For example, 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. This method can only estimate the effect of an intervention on the hazard of the next fatal event. However, when a research question asks how a policy or intervention influences the frequency of incidents, series hazard modeling will likely be more precise than interrupted time series because it captures more variation in the dependent variable while controlling for more detail in the independent variables.

Thus far, series hazard modeling has been used to estimate the hazard of global hijacking; republican terrorist attacks in the United Kingdom; global terrorist attacks by ASALA and JCAG; and attacks by Hamas, Fatah, and the Palestinian Islamic Jihad in Israel and the Palestinian territories (Dugan et al. 2005, 2009; LaFree et al. 2009; Chen et al. 2008). Efforts are currently underway to use it to examine patterns of attacks against abortion providers in the U.S. (Bartholomew 2009) and cases of eco-related crimes in the U.S. (Varriale 2009). Because it can directly measure dependence across events, this method is especially useful for modeling events where past events are likely to be known to those who precipitate future events. Contagion is not necessary, however, because the serial hazard model does a fine job in estimating intervention effects across independent events. In sum, when appropriate for the data and context, the series hazard model is a robust alternative to interrupted time series that harnesses the hidden variation in the duration between events and in the details of each event.

References

Allison PD (1995) Survival analysis using the SAS system: a practical guide. SAS Institute Inc, Cary

Alt JE, King G, Signorino CS (2000) Aggregation among binary, count, and duration models: estimating the same quantities from different levels of data. Polit Anal 9:21–44

Arey JA (1972) The sky pirates. Charles Schribner's Sons, New York

Bartholomew B (2009) Fear and darkness: anti-abortion terrorism and the United States Supreme Court. Working Paper, The University of Maryland

Box-Steffensmeier JM, Jones B (2004) Event history modeling. Cambridge University Press, New York

Box-Steffensmeier JM, Zorn C (2002) Duration models for repeated events. J Polit 64:1069–1094

Braithwaite J (1989) Crime, shame and reintegration. Cambridge University Press, Cambridge

Braithwaite J (2005) Pre-empting terrorism. Curr Iss Crim Justice 17:96–114

Brandt PT, Williams JT (2001) A linear poisson autoregressive model: the poisson AR(p) model. Polit Anal 9:164–184

Brandt PT, Williams JT (2007) Multiple time series models. Sage Publications, Inc, Thousand Oaks

Brandt PT, Williams JT, Fordham BO, Pollins B (2000) Dynamic modeling for persistent event-count time series. Am J Polit Sci 44:823–843

Chen X, Minacapelli L, Fishman S, Orehek E, Dechesne M, Segal E, Kruglanski AW (2008) Impact of political and military interventions on terrorist activities in Gaza and the West Bank: a hazard modeling analysis. Poster presented at START Workshop, September 2008. College Park, MD

Clarke RV, Felson M (eds) (1993) Routine activities and rational choice. Transaction Publishers, New Brunswick
Cleves MA, Gould WW, Gutierrez RG (2004) An introduction to survival analysis using Stata College Station. Stata Press, TX
Cook TD, Campbell DT (1979) Quasi-experimentation design & analysis issues for field settings. Houghton Mifflin Company, Boston
Cornish DB, Clarke RV (1986) The reasoning criminal: rational choice perspectives on offending. Springer, New York
Cox D (1972) Regression models and life-tables. J R Stat Soc Ser B (Methodological) 34:187–220
D'Alessio SJ, Stolzenberg L (1995) The impact of sentencing guidelines on jail incarceration in Minnesota. Criminology 33(2):283–302
Dugan L (2010) Estimating effects over time for single and multiple observations. In: Piquero A, Weisburd D (eds) Handbook of quantitative criminology. Springer, New York, pp 741–763
Dugan L, Nagin D, Rosenfeld R (2003) Exposure reduction or retaliation? The effects of domestic violence resources on intimate partner homicide. Law Soc Rev 27:169–198
Dugan L, LaFree G, Piquero AR (2005) Testing a rational choice model of airline hijackings. Criminology 43:1031–1066
Dugan L, Huang JY, LaFree G, McCauley C (2009) Sudden desistance from terrorism: the Armenian secret army for the liberation of Armenia and the Justice Commandos of the Armenian Genocide. Dyn Asymmetr Conflict (forthcoming)
Ezell ME, Land KC, Cohen LE (2003) Modeling multiple failure time data: a survey of variance-correction proportional hazards models with empirical applications to arrest data. Sociol Methods 33:111–167
Federal Aviation Administration (FAA) (1983) Aircraft hijackings and other criminal acts against civil aviation statistical and narrative reports. Office of Civil Aviation Security, Washington, DC
Freeman JR (1989) Systematic sampling, temporal aggregation, and the study of political relationships. Polit Anal 1:61–98
Gibbs JP (1975) Crime, punishment, and deterrence. Elsevier, New York
Greene WH (2008) Econometric analysis, 6th edn. Pearson Prentice Hall, Upper Saddle River
Holden RT (1986) The contagiousness of aircraft hijacking. Am J Sociol 91:874–904
Kaminski RJ, Edwards SM, Johnson JW (1998) The deterrent effects of oleoresin capsicum on assaults against police. Police Quart 1:1–20
Kelly P, Lim LL-Y (2000) Survival analysis for recurrent event data: an application to childhood infectious diseases. Stat Med 19:13–33
LaFree G, Dugan L (2007) Introducing the global terrorism database. Polit Viol Terror 19:181–204
LaFree G, Dugan L, Korte R (2009) Is counter terrorism counterproductive? Northern Ireland 1969–1992. Criminology 47:501–530
Land KC, McCall PL, Nagin DS (1996) A comparison of poisson, negative binomial, and semi-parametric mixed poisson regression models with empirical applications to criminal career data. Sociol Methods Res 24:387–442
Lawless JF (1982) Statistical models and methods for lifetime data wiley series in probabilty and mathematical statistics. Wiley, New York
Lewis-Beck MS (1986) Interrupted time series. In: Berry WD, Lewis-Beck MS (eds) New tools for social scientists. Sage Publications, Beverly Hills, pp 209–240
Lewis-Beck MS, Alford JR (1980) Can government regulate safety? The coal mine example. Am Polit Sci Rev 74:745–756
McDowall D, McCleary R, Meidinger EE, Hay RA (1980) Interrupted time series analysis. Sage Publications, Inc, Thousand Oaks
McDowall D, Lizotte AJ, Wiersema B (1991) General deterrence through civilian gun ownership: an evaluation of the quasi-experimental evidence. Criminology 29:541–559
Nagin D, Paternoster R (1991) On the relationship of past and future participation in delinquency. Criminology 29(2):163–190
Nagin D, Paternoster R (2000) Population heterogeneity and state dependence: state of the evidence and directions for future research. J Quantitat Criminol 16(2):117–144
Paternoster R (1987) The deterrent effect of the perceived certainty and severity of punishment: a review of the evidence and issues. Justice Q 4:173–217
Pridemore WA, Chamlin MB, Cochran JK (2007) An interrupted time-series analysis of Durkheim's social deregulation thesis: the case of the Russian federation. Justice Q 24:271–290
Pridemore WA, Chamlin MB, Trahan A (2008) A test of competing hypotheses about homicide following terrorist attacks: an interrupted time series analysis of September 11 and Oklahoma City. J Quant Criminol 24:381–396

Shellman SM (2004a) Time series intervals and statistical inference: the effects of temporal aggregation on event data analysis. Polit Anal 12:97–104

Shellman SM (2004b) Measuring the intensity of international political events data: two interval-like scales. Int Interact 30:109–141

Shellman SM (2008) Coding disaggregated intrastate conflict: machine processing the behavior of substate actors over time and space. Polit Anal 16:464–477

Sherman LW (1993) Defiance, deterrence, and irrelevance: a theory of the criminal sanction. J Res Crime Delinq 30:445–473

Stata Press (2005) Stata times-series reference manual release 9. Stata Press, College Station

Stolzenberg L, D'Alessio SJ (1994) Sentencing and unwarranted disparity: an empirical assessment of the long-term impact of sentencing guidelines in Minnesota. Criminology 32(2):301–310

Varriale J (2009) The threat of radical environmental and animal rights groups: the role of moral inhibitions in 'Monkey-Wrenching. Dissertation proposal, The University of Maryland