Separating The Impact Of Macroeconomic Variables and Global Frailty In Event Data.

James Lewis Wolter*
Department of Economics, University of Oxford
INET, University of Oxford
Oxford-Man Institute, University of Oxford

First Version: June, 2011
This Version: January, 2013

Abstract

Global frailty is an unobserved macroeconomic variable. In event data contexts, this unobserved variable is assumed to impact the hazard rate of event arrivals. Attempts to identify and estimate the path of frailty are complicated when observed macroeconomic variables also impact hazard rates. It is possible that the impact of the observed macro variables and global frailty can be confused and identification can fail. In this paper I show that, under appropriate assumptions, the path of global frailty and the impact of observed macro variables can both be recovered. This approach differs from previous work in that I do not assume frailty follows a specific stochastic process form. Previous studies identify global frailty by assuming a stochastic form and using a filtering approach. However, chosen stochastic forms are arbitrary and can potentially lead to poor results. The method in this paper shows how to recover frailty without these assumptions. This can serve as a model check to filtering approaches. The methods are applied to simulations and an application to corporate default.

1 Introduction

The hazard rates of economic events frequently depend on macroeconomic conditions. Just a few examples include mortgage default, corporate default, retirement, investment decisions and labor market decisions. Because of this, we may include macro variables as covariates in hazard analysis. The inclusion of these variables leads to dependence, as all observations at a particular calendar time share the common realized value. Dependence of defaults[1] may also result from correlation between observation-specific covariates. These types of variables and macroeconomic variables are often correlated as well. Economic hazard models should account for all of these potential dependencies.

*This paper is based on the first chapter of my Ph.D. dissertation at Yale University. I thank Don Andrews, Xiaohong Chen, Darrell Duffie, Peter Phillips, Neil Shephard and seminar participants at Barclays, Cambridge, Chicago Booth - Econometrics & Statistics, Federal Reserve Board, George Washington, J.P. Morgan, Mannheim, Oxford, Oxford-Man Institute, UPenn, Stanford GSB - Finance, Swedish Riksbank, Wisconsin and Yale. This research was partially funded by an Anderson Fellowship from the Cowles Foundation and Yale University. All remaining errors are my own.

[1]Throughout this paper, I will refer to random times which have corresponding hazard rates as "default" times.

In some hazard situations, even after accounting for a realistic set of covariates, the character of the observed defaults is difficult to explain. For example, corporate defaults are clustered around recessions or financial crises. If the clustering is too severe, hazard models may capture the realized data poorly. Essentially, time-dependent model misspecification is present. Some evidence of the failure of standard hazard models in this context is given in Das, Duffie, Kapadia and Saita (2007).

In order to account for the observed clustering, Duffie, Eckner, Horel and Saita (2009) introduce a global latent risk factor into a standard hazard model. This risk factor is similar to an unobserved macroeconomic covariate. I will call this variable global frailty.[2] At time periods where frailty is high, there is clustering in defaults beyond what can be explained by the observed covariates. When frailty is low, defaults are suppressed. This additional time-varying element improves model fit. In any specific application it is unlikely available data will ever contain all the elements that are relevant. In addition, parsimony may lead us to restrict the set of covariates used. As a result, this model extension is widely applicable.

Estimation of models with global frailty has received increasing attention. See Creal, Schwaab, Koopman and Lucas (2012); Giesecke and Schwenkler (2012); Azizpour, Giesecke and Schwenkler (2011); Koopman, Lucas and Schwaab (2011) and Duffie et al. (2009). All of these papers take a filtering approach to estimation. Frailty is assumed to follow a certain parametrized stochastic process. Knowledge of this parametric form is then used in estimation. The form of the frailty process contributes to the expectation of likelihoods or other criterion functions. The form also contributes to identification. Our ability to separate out elements of the model is dependent on how global frailty is assumed to propagate. In all of the previously mentioned papers, simple forms are chosen. This is done, at least partly, in order to minimize difficulties related to estimation.

A natural question arises: what if the global frailty does not follow the assumed form? Most of the previously cited papers are concerned with modeling corporate default. In this application, the time span of empirical analysis is usually around 30 years. It seems very reasonable to question the stability of the unobserved frailty factor over this time period. Indeed, almost by nature, this unobservable will take unexpected forms. If frailty is well behaved, highly correlated variables can likely be included in the model, rendering it superfluous. One of the reasons frailty is interesting is that we may be surprised, as in mortgage or corporate default situations around the financial crisis of 2008. This can produce unstable frailty paths which may not be well represented by simple models. In particular, frailty may not be well represented by models which produce well behaved expected criterion functions conducive to the filtering approach. Additionally, when filtering is used, allowing for more complicated forms of global frailty will make identification more difficult, raise model selection and validation issues and increase computational intensity of implementation.

In this paper, I investigate the econometric analysis of hazard models whose observations are dependent as a result of global frailty, macro covariates and correlated observation-specific variables. However, I take a different approach to modeling frailty than previous papers. I do not assume that global frailty follows a stochastic process with a particular form. I only assume that frailty is a deterministic, strictly positive function of time with a few weak regulatory conditions. In this sense, the model of frailty presented is

[2] Duffie et al. (2009) call this frailty. However, the term frailty is used in the statistics and medical literature to mean what economists call unobserved heterogenaity. I use global frailty to avoid confusion. Throughout the paper, the term frailty means global frailty.

more flexible than previous approaches. The objective is to determine under what conditions we can identify and estimate global frailty with minimal assumptions on its path. This will allow us to recover frailty paths which have forms that are not obvious prior to the analysis.

Throughout the paper, I will assume a specific semiparametric structure for hazard functions. Following much of the literature, a Cox proportional hazard form is assumed. In this specification, frailty takes the place of the baseline hazard. When observations are at risk over different calendar time intervals, their baseline hazards will be different. Their baseline hazard will correspond to the portion of the frailty path coincident with the calendar time over which they are at risk. This is the standard notion of global frailty in the literature mentioned above. The assumption of a Cox proportional hazard form is not crucial to the results. This can likely be relaxed, but is a standard model and a good starting point.

As mentioned above, in many empirical situations, a critical aspect of realistic models is allowing for macroeconomic variables in the hazard functions. This element, along with global frailty, produces a fundamental identification issue. Over any time interval, what portion of the hazard rate can be explained by the observed macroeconomic variables and what portion by the unobservable frailty? How can we separate out the impact of the observed macroeconomic variables from the unobserved macroeconomic variable (i.e. frailty)? Sampling more observation in the time interval does not resolve this problem as this provides no additional sampling of macroeconomic covariates. This is not a trivial issue. In the sequel, I provide counterexamples where we are unable to separate these two elements. Moreover, this is not an issue specific to hazards. Similar types of counterexamples occur in other models. Identification of global frailty in the presence of macroeconomic variables appears to be a general problem not specific to particular modeling approaches. This is discussed fully in Section 3 below.

In order to identify and estimate hazard functions in this case, two important elements are needed: (1) a specific type of sampling scheme and (2) a judicious choice of estimation approach which utilizes that scheme. For the sampling scheme, a cross section of observations is not sufficient. If all observations are observed over the same calendar time interval, then each observation is impacted by the same portion of the common processes[3]. Therefore, the average of likelihoods used in estimation will converge to the *conditional* expected likelihood, conditional on the realization of the macroeconomic variables. As will be discussed below, this causes the situation to be unidentified. We are able to trade off the impact of the frailty and the macro variables, the result being the same maximum in the conditional expected likelihood. Therefore, the impact of macro variables cannot be separated from global frailty. This is the econometric manifestation of confusing the impacts of the observed macro variables and global frailty.

The identification issue is handled in the filtering case by assuming a stochastic form for the frailty. This additional structure faciliates separating the two influences on the hazard rates. The expectation of the criterion function is taken using the assumed stochastic structure on the unobservable. As we want to avoid making assumptions on the underlying global frailty, this is not an option in our analysis. In the sequel, I show that identification and consistency are still possible without any additional stochastic form assumptions on the global frailty.

Identification and consistency fail in the cross sectional case because there is effectively no sampling of macro covariates. Every observation is impacted by the same path of these common variables. Fortunately, in many situations of interest, a cross section is not what is observed. Instead, observations begin

[3]Throughout this paper, I will refer to any global variables (such as macroeconomic variables) as common variables.

at various calendar times. For example, in the mortgage default case, mortgages as signed at various dates instead of all at once. This variation will be used in estimation to solve the problem described above.

The proposed sampling scheme assumes that observations potentially default over a fixed time interval. This can be interpreted as the life of a contract. It is assumed that when we gather more observations, they begin at increasingly large calendar times. By sampling observations at risk of default over different blocks of calendar time, we effectively sample from the common process. As the starting times increase to infinity, we have observations impacted by the entire path of the macro variables.

Alone, this type of sampling is not sufficient. We make few assumptions on the frailty path. As a result, observations which do not overlap with a particular calendar time t give us no information about the frailty path at t. In order to asymptotically recover the path, we need the number of observations overlapping every calendar time to increase to infinity.

An appropriate sampling scheme for this situation incorporates both the above requirements. The starting times of observations should increase to infinity *and* for any calendar time, the number of observations which overlap that time should increase to infinity. These two assumptions work together to overcome the identification and consistency problems described above. As we will see in the sequel, this sampling allows us to replace the conditional expected likelihoods with full expected likelihoods.

The second element needed is an appropriate choice of estimation strategy when we have observations corresponding to the above outlined sampling scheme. Below I provide a counterexample which shows that, for the sampling scheme described above and a seemingly reasonable estimation approach, consistency fails badly. The problem is related to identification between observed macro variables and frailty. An asymptotic version of this counterexample is possible in many situations. As a result, additional assumptions are required.

The key to consistency is restricting the form estimates of the frailty path can take given the current amount of data. As more data is gathered, these restrictions are relaxed. In the limit, the restrictions disappear entirely, allowing our estimates to be very flexible. This is a basic principle of the sieve estimation approach taken below. The complexity of the frailty estimate is balanced against how much of the macro variables we observe. If, in some sense, the amount of the macro variables observed grows faster than this complexity, we are able to separate the two elements asymptotically. The number of observations overlapping each calendar time must also be controlled.

Given an appropriate sampling scheme, I derive conditions under which the covariate coefficients and the frailty path can be estimated consistently. These consistent estimates give us the impact of both frailty and the observed macro variables. In addition, the current value of frailty is a by-product of estimation. This is important for extensions to forecasting applications.

The estimator is an extension of the point process likelihood approach taken in Karr (1987). Karr (1987) examines sieve estimation of the baseline hazard in a proportional hazard model. In that work, the observations are assumed *i.i.d.* and the covariates' coefficients are assumed known. There is also no global frailty. In the sequel, I first extend Karr (1987) to a simple cross sectional case with no frailty or macroeconomic variables. The proof in this situation gives us all the tools we need for the full model. Then, we use the same likelihood approach to estimate the impact of macro covariates and global frailty.

The cross sectional results derived are of independent interest. The vast majority of semiparametric

estimators of proportional hazard models use partial likelihood approaches (See Martinussen and Scheike (2010)). These methods estimate the integrated baseline hazard instead of the baseline hazard itself. The only paper I am aware of which uses the full point process likelihood in semiparametric estimation is Karr (1987). In this paper, I also use the full likelihood.

The approach is related to the literature on hazard models with dependence across observations. There are surprisingly few papers on this topic. One specification which has received attention is cluster analysis. In this situation, covariates are allowed to be dependent within groups, while the groups are independent. See Martinussen and Scheike (2010), Hougaard (2000) or Aalen, Borgan and Gjessing (2010) for an overview of these types of methods. This assumption is not sufficient for our purposes. Hazard models using common macroeconomic variables imply dependence between all observations. In economic situations, there may be no natural way to place observations into groups with statistically independent covariates across groups. There are a few other papers related to this work focusing on spatial correlation between observations. See Bastos and Gamerman (2006); Li and Lin (2005); Banerjee and Dey (2005) and Li and Ryan (2002). There has been much less done in this area than with the clustering approach.

The previously mentioned global frailty literature can also be seen as dependent hazard models. Many of these papers do not model hazards for individual observations but use grouping assumptions. Either all events are described by a single point process or observations are divided into groups described by point processes (or similar models). Many of these papers also model a notion of contagion. This can be easily added to our model but is left off for simplicity.

To examine the performance of the estimation approach, simulation studies are conducted. The estimates preform well even in relatively small sample sizes. Estimates are somewhat sensitive to the specific sieve space chosen. However, this is a general problem in the sieve approach. Larger sample sizes than those used in simulations will mitigate this problem.

In an empirical application of these methods, I examine the global frailty path in Moody's corporate default data. The results suggest that frailty may follow a complicated path in this situation. In particular, the level of frailty experienced surrounding the 2008 crisis dwarfs levels in the previous three decades. This suggests that frailty is not necessarily a stable stochastic process as assumed in previous studies. For example, if the estimates of frailty prior to the 2008 crisis were used to characterize how defaults would proceed in the following years, the results would preform very badly. There is no reason there cannot be another large event which dwarfs our current experience as of 2012. This should cause us to proceed with caution when interpreting filtering results. In particular, these potential issues should be kept in mind when interpreting forecasts.

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 introduces point process likelihoods. Here, I present the estimation approach for recovering global frailty. Section 4 shows simulation results examining the performance of the estimation approach. An empirical application to corporate default is also presented. Section 5 concludes. Some results and proofs are presented in the Appendix.

2 Models and Preliminaries

Here I describe how the random times used in this paper are constructed. Special attention is given to incorporating common covariates across observations and global frailty. In addition, dependence between observation-specific variables is allowed for. Below, the random times are shown to have a martingale structure. This structure is needed to facilitate estimation in later sections.

Our sampling scheme has four important elements for each observation: (1) a set of covariate stochastic processes $X^i(t)$ which are specific to each observation i; (2) a set of covariate stochastic processes which are common $Y(t)$; (3) a calendar time G^i at which observation i is "born" or becomes at risk of default and (4) a global frailty element which will be elaborated on below.

2.1 Basic Sampling Structure

Observations are indexed by $i \in \mathbb{N}$. Observations have a deterministic constant G^i which corresponds to the calendar time at which they begin to be at risk. This is the date at which, for example, a mortgage was signed. Each observation is at risk over a fixed time interval of length T. The period over which they are at risk corresponds to the calendar time interval $\left[G^i, G^i + T\right]$. These calendar time intervals are allowed to overlap.

The time interval $\left[G^i, G^i + T\right]$ may be contractually specified, such as the duration of a loan. The situation may also have a natural time interval. For example, a model for school dropout. Another possibility is that observations are censored after a specified duration. This bounded time interval assumption is made for technical reasons, but likely can be relaxed.

Let $\left\{ X^i(t) \middle| t \in [0,T] \right\}$ be d covariate stochastic processes specific to each observation. These variables could correspond to housing values, accounting information, etc. $X^i(t)$ for $t \in [0,T]$ corresponds to the value of these covariates over the calendar time interval $\left[G^i, G^i + T\right]$. Defining $X^i(t)$ on $[0,T]$ instead of $\left[G^i, G^i + T\right]$ is done for notational simplicity. We make the important additional assumption on $X^i(t)$ that its paths are left-continuous with right-hand-limits (càglàd for short)[4]. Assume the distribution of the variables $X^i(t)$ has support contained in the compact set $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d$ for each $t \in [0,T]$. It should be emphasized that we are not making any stationarity assumptions on $X^i(s)$ within observations. We only assume the support of the covariates at each time is contained in $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d$.

In addition, there is another set of j covariate processes $Y(t)$ which correspond to macro variables. These variables represent any covariates which have a single global value at each calendar time. For example US Treasury rates or the GDP growth rate. $Y(t)$ is assumed to be stationary with càglàd paths and to have support contained in the compact set $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_j$ for each $t \in [0,\infty)$. For simplicity, the support of all covariate processes can be thought of as $[0,1]^{d+j}$. The covariate processes $Y(t)$ are common to all observations in that, for each observation i, the portion of $Y(t)$ corresponding to the calendar time the observation is at risk $\left[G^i, G^i + T\right]$ affects the hazard rate for that observation.

To sum up, the relevant covariates for observation i are

$$\left\{ Z^i(t) = \left(X^i(t), Y\left(G^i + t\right) \right) \middle| t \in [0,T] \right\}.$$

[4] Càglàd paths implies that the processes $\left(X^i(t), Y(t)\right)$ used below are predictable, an important technical property for our results. This follows from, for example, Protter (2005) pg. 102.

We will use the notation $Z^i(t)$ in the sequel.

As stated, the starting dates for the observations G^i are very general. More assumptions will be needed to achieve consistent estimators. This is put off until later, as the martingale results presented in the next subsection hold for arbitrary G^i. For now, we only assume $G^i \geq 0$.

2.2 Construction of Random Events and Martingale Structure

The construction of the random times used throught the paper follows Wolter (2012b). I give a brief outline here and refer the reader to that work for more elaboration. Recall that we have defined X_t^i on $[0,T]$ instead of $\left[G^i, G^i+T\right]$. In what follows, $\alpha^i(\cdot)$ will be the hazard functions with covariates taken as arguments. The setup allows different observations to have different $\alpha^i(\cdot)$. We make the following assumption,

(A1): $\alpha^i : [0,T] \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ are continuous functions such that

$$\alpha^i\left(t, X_t^i, Y_{t+G^i}\right) = h^i(t) \exp\left(\beta' Z^i(t)\right)$$

Where β are parameters shared by all observations and $h^i(t)$ can vary across observations. We also assume

$$\begin{aligned} \inf_{(t,x,y)\in[0,T]\times\mathcal{X}\times\mathcal{Y}} \alpha^i(t,x,y) &= \underline{C} > 0, \\ \sup_{(t,x,y)\in[0,T]\times\mathcal{X}\times\mathcal{Y}} \alpha^i(t,x,y) &= \overline{C} < \infty. \end{aligned}$$

for all $i \in \mathbb{N}$.

In our frailty model below, sampling of observations will be in two dimensions. Assumption (A1) easily extends to this case using the index $i, j \in \mathbb{N} \times \mathbb{N}$. We assume (A1) throughout the paper.

Random times τ_i are defined as

$$\begin{aligned} \Gamma_t^i &= \int_0^t h^i(t) \exp\left(\beta' Z^i(t)\right) ds \\ \tau_i &= \inf\left\{ t \in \mathbb{R}_+ \middle| \Gamma_t^i \geq \eta_i \right\} \end{aligned}$$

where η_i is an independent standard exponentially distributed random variable. The η_i variables are independent of all covariates and each other. Notice that the portions of the processes $Y(t)$ and $X^i(t)$ corresponding to $\left[G^i, G^i+T\right]$ are what is used in the definition of Γ_t^i.

In this model, the functions $\alpha^i(\cdot)$ are hazard functions under the traditional definition. There are some technicalities involved with this statement. For a more rigorous result, see Fleming and Harrington (1991) Theorem 4.2.1. Despite some technical qualifications, it is clear what we are interested in estimating. $\alpha^i(\cdot)$ are the objects of interest. The choice of standard exponential distributions for η_i is not arbitrary. This is needed for $\alpha^i(\cdot)$ to be the hazard rate. When the covariates are *i.i.d.*, this model has the same distributions as all *i.i.d.* hazard models in the literature.

Write $N^i(t) = \mathbf{1}_{\{\tau_i \leq t\}}$ for the process that indicates default. Define also,

$$\Lambda_t^i = \int_0^t h^i(t) \exp\left(\beta' Z^i(t)\right) \mathbf{1}\{\tau_i \geq s\} ds,$$

$$M_t^i = N_t^i - \Lambda_t^i.$$

The asymptotic results that follow depend on M_t^i being continuous-time martingales. This is true in great generality.

We assume the covariates $\left(X_t^i, Y_{t+G^i}\right)$ and the random times τ_i are observed. Conditionally independent censoring can be easily added to the model. This type of censoring is independent of τ_i, conditional on the covariates. See Martinussen and Scheike (2010) or Anderson et al. (1994) for an overview. I do not include censoring in what follows for notational simplicity. All results in the sequel follow with conditionally independent censoring and properly adjusted estimators.

2.3 Starting Dates and Global Frailty

I now describe our assumptions on the starting dates G^i and baseline hazards $h^i(t)$. Initially, we will assume a cross section with $G^i = 0$ for all observations. The baseline hazard will be the same for all observations as well, $h^i(t) = h(t)$. This is a simple preliminary case which is the building block of more complicated models. Asymptotics will hold using the obvious sampling approach.

In the cross sectional case, we assume there are no macroeconomic covariates. Macroeconomic covariaes in this situation can cause loss of identification. Because each observation will have the same values for the macro variables, there can be a trade-off where we increase β and reduce the baseline hazard $h(t)$, resulting in the same hazard rate. This corresponds to confusing the influence of observed macro variables and frailty in describing the hazard rate of default. We are able to overcome this identification problem by assuming an appropriate sampling scheme and estimation strategy. This is described below.

I next describe our full model, which contains macroeconomic variables and global frailty. This model was initially proposed as a way to better explain the observed clustering of corporate defaults. Corporate defaults are clustered around periods of financial crisis and recession. This clustering is difficult to explain with standard hazard models. Das et al. (2007) provide some statistical evidence that hazard models with common baseline hazard functions are rejected when using data on US corporate default. A solution proposed in Duffie et al. (2009) is essentially the global frailty model described below. In this model, observations at risk over different calendar time intervals have different baseline hazard functions. This is allowed for in such a way that clustering observed in real data is better captured. There is no reason to restrict the use of this model to corporate default. Any situation where changing macroeconomic conditions are important could potentially use this notion of frailty.

One interpretation of this model is that time-dependent model misspecification is present. The changes in the baseline hazard across time capture this. However, the literature has favored the interpretation that frailty represents a global unobserved risk factor impacting all hazard rates.

In the global frailty model, both G^i and $h^i(t)$ will differ across observations. For any finite sample of data, there are only a finite number of starting times. We use this fact to define sampling of observations in two dimensions, indexing each observation by $j, i \in \mathbb{N} \times \mathbb{N}$. Here, j indexes which ordered starting

time G^j an observation begins at. i indexes the number of observations which begin at G^j.

We assume that G^j each equal one of a countable set of potential starting times , with $G^j < G^{j+1}$. These starting times can be irregularly spaced. However, for notational simplicity, in the remainder of the paper I will assume they are equidistant from each other. This can be written as, $G^j = j\delta$ where $\delta > 0$, $j \in \mathbb{N}$. This assumption naturally groups observations into blocks, where a block is all observations starting at the same time. Our sampling will assume that the number of observations beginning at each G^j approaches infinity. I call this setup block sampling.

I now define global frailty and incorporate it into a Cox proportional hazard model. Global frailty is a strictly positive real valued function $h_0(t)$ defined on $[0, \infty)$. This is similar to a macro variable except it is not observed. The frailty function takes the place of the baseline hazard in the block sampling case. For observation ji, the baseline hazard is the portion of the frailty path corresponding to the calendar time interval $\left[G^j, G^j + T\right]$. This is the calendar time interval over which the observation is at risk. As a result, observations at risk over different calendar time intervals have different baseline hazards. This setup rules out standard estimators, such as those in Andersen and Gill (1982) or Martinussen and Scheike (2010), because they require all observations to have the same baseline hazard.

I write the baseline hazard corresponding to the jth block as h_0^j, where $h_0^j = h_0\left(G^j + t\right)$. The hazard rate for observation ji can now be written as

$$h_0\left(G^j + t\right) \exp\left\{\beta_0' Z^{ji}(t)\right\} \mathbf{1}\left\{\tau_{ji} \geq s\right\}, \tag{1}$$

or

$$h_0^j(t) \exp\left\{\beta_0' Z^{ji}(t)\right\} \mathbf{1}\left\{\tau_{ji} \geq s\right\}.$$

Random times with this hazard rate can easily be constructed as in Section 2.2 above.[5]

One interpretation of this model is that time-dependent model misspecification is present. The changes in the baseline hazard across time and the global nature of the frailty path suggest this. Another interpretation is that the frailty path represents the residual hazard which is left after the impact of the observed variables is accounted for. However, the literature has favored the interpretation that frailty represents a global unobserved risk factor impacting all hazard rates.

As mentioned in the introduction, there is an identification problem in our model. The issue involves separating out the impact of observed macroeconomic variables from frailty, which is an unobserved macro variable. This will be discussed fully in the next section. Here, I simply discuss the requirements of the sampling scheme needed to overcome this issue.

In order for the identification problem to be overcome, we require a type of two dimensional sampling. To recover the impact of the macroeconomic variables, we need observations starting at increasingly large calendar times. This corresponds to increasingly large $G^j = j\delta$. In order to recover the frailty path at any given calendar time t, we need the number of observations ji with $t \in \left[G^j, G^j + T\right]$ to increase to infinity. Because of these two requirements, we must observe blocks of observations at increasingly large G^j *and* the number of observations starting at each G^j must increase to infinity.

[5] It can be shown, with a proof almost identical to Lemma 1, that this set up has a martingale structure defined on $[0, \infty]$. The processes M_t^{ji} are defined to be zero before G^j. They have the hazard rate (1) on the interval $\left[G^j, G^j + T\right]$. After $G^j + T$, $M_t^{ji} = M_{G^j+T}^{ji}$.

I now incorporate the requirements outlined in the previous paragraph into our sampling scheme. The sampling will be indexed by n. Assume n observations per G^j. Additionally, we assume that the researcher only sees the first $k(n)$ blocks of observations, $k(\cdot)$ being a function of $\mathbb{N}$. $k(n) \to \infty$ as $n \to \infty$. The relevant covariates include observation specific covariates $X^{ji}(t)$ and common covariates $Y(t)$. Write $Z^{ji}(t) = \left(X^{ji}(t), Y\left(G^j + t\right)\right)$. I emphasize that adjacent blocks are allowed to overlap (i.e. $\left[G^j, G^j + T\right]$ and $\left[G^{j+1}, G^{j+1} + T\right]$ can intersect).

The sampling scheme is stringent, requiring a large number of observations per block. This is needed for technical reasons.[6] Situations where observations start to be at risk at regular intervals are more likely to satisfy the required conditions. Examples might include observations with weakly, monthly or quarterly reported start times (i.e. school cohorts, etc). The assumption that there are the same number of observations per block is made only for notational simplicity. Asymptotics will hold provided, for example, all blocks retain the current absolute differences in number of observations.

Certain financial defaults fit well into our setup. For example, in the case of mortgages, the number of subprime and Alt-A mortgage originations in the US in 2003 was 1,385,598 and in 2005 was 3,015,434 according to Mayer, Pence and Sherlund (2009). Excluding weekends, the average number of mortgage originations per day was 5,309 and 11,553 respectively. An empirical example presented in Section 4 applies our methods to corporate defaults in the Moody's database. The data size is comparably small with slightly more than 9000 observations. However, simulations conducted in Section 4 suggest that estimates preform well under these conditions.

The assumption that many observations must start at the same time can likely be relaxed. A minimal requirement is that a large number of observations are at risk of default over any fixed calendar time. As we will see in the sequel, it is obvious how to adjust the estimators in this paper to account for more general cases. Even though the blocking assumption may be questionable in certain cases, it is more realistic than a cross section assumption when observations start at different calendar times.

3 Point Process Likelihood Estimation: The Global Frailty Case

In this section, I present estimators which are able to separate the impact of macroeconomic and observation-specific variables from global frailty. A point process likelihood approach is used to estimate the hazard model using sieves. This is initially done in a simple cross sectional context. The results from this case are then extended to our full model. Our approach is similar to that of Karr (1987).

There is already a large literature on likelihood estimation of point processes. Almost all estimators focus on the partial likelihood approach when estimation is semiparametric. A classic reference is Anderson and Gill (1982). See Martinussen and Scheike (2010) or Andersen et al. (1994) for a review of partial likelihood methods. The full likelihood will be used below.

As partial likelihood methods are so widespread, I mention why they are not used here. Normally, partial likelihood estimators have many nice properties. They have been used to derive consistent and asymptotically normal estimators of β_0. These estimators are asymptotically efficient in the *i.i.d.* case (see Andersen et al. (1994)). Additionally, partial likelihood methods result in an estimator of the

[6] We derive estimation from a point process likelihood criterion function. For asymptotics to work, we need our sampled likelihood to converge to the expected likelihood in the limit. Because of the way our global frailty model is specified, this requires a large number of observations starting at the same starting data.

integrated baseline hazard $\Lambda_0(t) = \int_0^t h_0(s)\,ds$ which satisfies a functional CLT. This converges at a $\sqrt{n}$ rate.

Despite these advantages, there are several reasons why the partial likelihood approach is not optimal for our global frailty situation. Previous methods cannot handle the case where observations have different baseline hazards. This is exactly the situation we are interested in. Depending on which calendar time interval $\left[G^i, G^i + T\right]$ an observation is at risk, their baseline hazard will be determined by the portion of the global frailty function $h_0(s)$ (defined on $[0, \infty)$) which overlaps with that calendar time. As different observations have different baseline hazards, current partial likelihood methods fail. In this section, methods are developed which directly estimate $h_0(s)$ accounting for the different baseline hazards.

Another issue is that partial likelihood methods estimate the integral of the baseline hazard $\int_0^t h_0(s)\,ds$ instead of $h_0(s)$. If the main interest is in β_0, this is not an issue. If global frailty is the topic of analysis, an estimate of $h_0(s)$ is of primary interest. Ad hoc transforms of the previously presented estimator $\widehat{\Lambda}(t)$ may not preserve efficiency properties. There is little theoretical research on this topic. It is possible to transform $\widehat{\Lambda}(t)$ to estimate $h_0(t)$. This will slow down the rate of convergence. See Anderson et al. (1994) pg. 507 for an example using kernels.

Finally, there is the issue of dependence. In finite samples, dependence always impacts estimators. While the estimators based on partial likelihood suggest strong first-order asymptotic properties, these may not manifest themselves in finite samples when dependence is present. The additional transform needed to recover the function $h_0(t)$ from its integral estimate will further deteriorate results. Direct estimation of $h_0(t)$ may preform better in finite samples and equally as well asymptotically.

For these reasons, we develop our full likelihood estimation approach. With this method, $h_0(t)$ is estimated directly using sieves. The fact that there are different baseline hazards is accounted for in the estimation. Additionally, we utilize the type of data sampling outlined in Section 2 to overcome the identification problem of separating macroeconomic influences from global frailty.

3.1 Point Process Likelihood and Identification

In this subsection, the point process likelihood used in our estimation approach is presented. I then discuss at length the identification issues raised when trying to both estimate the impact of observed macroeconomic variables and recover the path of global frailty. What is possible to recover will differ depending on our sampling assumptions.

First, a few preliminary assumptions.

(**A2**): $Z^{ji}(t)$ are assumed to be càglàd.

(**A3**): Assume that each observation $Z^{ji}(t)$ has the same functional distribution. Specifically, let $X^{ji}(t+)$ and $Y\left(G^j + t+\right)$ be the right continuous versions of these processes[7]. Assume, for each ji

$$\left\{ Z^{ji}(t+) = \left(X^{ji}(t+), Y\left(G^j + t+\right)\right) \middle| \, t \in [0,T] \right\}$$

[7] The right continuous version of a process is defined as $X(t+) = \lim_{s \downarrow t} X(s)$ for each $t \in \mathbb{R}$. The value of $X(t+)$ at T can be defined arbitrarily. If $X(t)$ is cáglád, this process is cádlág.

has the same functional distribution.[8,9]

(**A4**): For all $t \in [0,T]$, $Z^{ji}(t) = \left(X_s^{ji}, Y_{G^j+s}\right)$ has support contained in a compact rectangle normalized to be $[0,1]^q$, where q is the number of covariates.

(**A5**): β_0 is contained in a compact rectangle $[a_1,b_1] \times \cdots \times [a_q,b_q]$. The true underlying coefficients β_0 are the same for each observation.

Note that I have written these assumptions allowing for both macro and observation-specific variables $\left(X_s^i, Y_{G^i+s}\right)$. I will also refer to these assumptions when we have ruled out macro variables, with obvious adjustments to the conditions.

(A2)-(A5) are mostly uncontroversial. (A3) implies that the expectations of log-likelihoods with the same baseline hazards are the same for all observations. (A3) also makes stationarity of the macro variables a likely assumption. It is possible to relax this, but for simplicity we do not. (A2), (A4) and (A5) are all needed for technical reasons. (A5) is a primitive assumption of our model. It is difficult to relax this condition and still be able to separate out the elements we are interested in.

Let $\mathcal{H}$ be a set of strictly positive, cáglád functions on $[0,T]$. We will discuss more specific assumptions on this function space below. (A5) is in force throughout the paper, so $\beta \in [a_1,b_1] \times \cdots \times [a_q,b_q]$. The following expression is a point process likelihood for the random defaults described above:

$$\frac{d\widetilde{P}}{dP}(h,\beta) = \begin{cases} \exp\left\{\int_0^T \left[1 - h(s)\exp\left(\beta' Z^i(s)\right)\mathbf{1}\left\{\tau_i \geq s\right\}\right] ds\right\} & \tau_i > T \\ h(\tau_i)\exp\left(\beta' Z^i(\tau_i)\right)\exp\left\{\int_0^T \left[1 - h(s)\exp\left(\beta' Z^i(s)\right)\mathbf{1}\left\{\tau_i \geq s\right\}\right] ds\right\} & \tau_i \leq T \end{cases}.$$

The log-likelihood is

$$\log\left[\frac{d\widetilde{P}}{dP}(h,\beta)\right] = \int_0^T \log\left[h(s)\exp\left(\beta' Z^i(s)\right)\right] dN_s^i + \int_0^T \left[1 - h(s)\exp\left(\beta' Z^i(s)\right)\mathbf{1}\left\{\tau_i \geq s\right\}\right] ds.$$

This likelihood gives a change of measure from a standard Poisson process. See Brémaud (1981) section IV.2 for more on point process likelihoods.

We will assume the following condition related to identification. Here, τ_i are the random times constructed in Section 2.

(A6): For each $(h,\beta) \in \mathcal{H} \times [a_1,b_1] \times \cdots \times [a_q,b_q]$, $h(s)\exp\left(\beta' Z^i(s)\right)\mathbf{1}\left\{\tau_i \geq s\right\}$ has a unique functional distribution.[10]

(A6) is a standard type of high level assumption required for identification. If two values in the set $\mathcal{H} \times [a_1,b_1] \times \cdots \times [a_q,b_q]$ produced the same distribution of the hazard function, $h(s)\exp\left(\beta' Z^i(s)\right)\mathbf{1}\left\{\tau_i \geq s\right\}$,

[8] Specifically, assume the observations i,j have the same distribution in the Skorokhod space. See Billingsley (1999) or Jacod and Shiryaev (2003) for more on the Skorokhod space. We need to consider right continuous versions of the above processes because the Skorokhod space is defined on $D[0,T]$, the space of all right continuous paths with left hand limits on $[0,T]$. We had previously assumed $Z^{ji}(s)$ has càglàd paths. Because the relevant processes are used in integrals, making them right continuous does not affect those integrals. Consideration of the right continuous versions of the processes is purely for the convenience of using the Skorohod space for functional distributions.

[9] Càglàd processes can only have a countable number of discontinuities. These are the values that change when we make the process right continuous. See Ethier and Kurtz (1986) pg.116 Lemma 5.1.

[10] There is a similar caveat here reguarding functional distributions as described in the footnotes to assumption (A3).

there would be no way of identifying the correct values. As we will see in the sequal, this assumption does not resolve the entire identification problem in this situation.

By standard arguments (see, for example, van der Vaart (1998) Lemma 5.35) the expected likelihood is uniquely maximized at (h_0, β_0) when the observed point process has the hazard $h_0(s)\exp\left(\beta_0' Z^i(s)\right)\mathbf{1}\{\tau_i \geq s\}$ (and τ_i is constructed using (h_0, β_0)). This is because each choice of $(h, \beta) \in \mathcal{H} \times [a_1, b_1] \times \cdots \times [a_q, b_q]$ corresponds to a unique intensity process $h(s)\exp\left(\beta' Z^i(s)\right)\mathbf{1}\{\tau_i \geq s\}$ in a distributional sense. This implies a unique change of measure. Specifically,

$$H(h, \beta) = \mathbb{E}_{(h_0,\beta_0)}\left\{\log\left[\frac{d\widetilde{P}}{dP}(h, \beta)\right]\right\},$$

$$H(h_0, \beta_0) > H(h, \beta) \qquad (h, \beta) \neq (h_0, \beta_0).$$

This is all that is needed if there are no macroeconomic variables impacting the hazard. As we gather more data in a cross section, the average likelihood will converge to its expectation. This expectation has a unique maxima as described above. Under appropriate regulatory assumptions, consistency would follow using a standard maximum likelihood approach.

When macroeconomic variables are present, the situation is more complicated. In a cross section, we do not sample an increasing amount of these variables. As a result, an average of the likelihoods converges to a *conditional* expected likelihood. Let $\mathcal{Y}_T = \sigma\{Y_t | 0 \leq t \leq T\}$.

$$\begin{aligned}
J(h, \beta) &= \mathbb{E}\left\{\log\left[\frac{d\widetilde{P}}{dP}(h, \beta)\right]\middle|\, \mathcal{Y}_T\right\} \\
&= \int_0^1 \log\left[h(s)\exp\left(\beta^{1\prime} Y(s)\right)\right] h_0(s)\exp\left(\beta_0^{1\prime} Y(s)\right)\mathbb{E}\left[\exp\left(\beta_0^{2\prime} X^i(s)\right)\mathbf{1}_{\{\tau_i \geq s\}}\middle|\, \mathcal{Y}_T\right] ds \\
&\quad + \int_0^1 h_0(s)\exp\left(\beta_0^{1\prime} Y(s)\right)\mathbb{E}\left[\left(\beta^{2\prime} X^i(s)\right) h_0(s)\exp\left(\beta_0^{2\prime} X^i(s)\right)\mathbf{1}_{\{\tau_i \geq s\}}\middle|\, \mathcal{Y}_T\right] ds \\
&\quad - \int_0^1 h(s)\exp\left(\beta^{1\prime} Y(s)\right)\mathbb{E}\left[\exp\left(\beta^{2\prime} X^i(s)\right)\mathbf{1}_{\{\tau_i \geq s\}}\middle|\, \mathcal{Y}_T\right] ds.
\end{aligned}$$

Notice that the values h and β^1 always enter $J(h, \beta)$ in the form $h(s)\exp\left(\beta^{1\prime} Y(s)\right)$. Because of this, we are able to trade off $h(s)$ and β^1. The result can be loss of identification. The following counterexample shows the potential pitfalls of estimating h when macro variables are present.

Counterexample 1 Assume we are in a cross sectional context, therefore $G^i = 0$ for all i. We allow the X_t^i to have any distribution satisfying our previous assumptions. Assume Y_t is piecewise constant, updating its values every $\rho > 0$ units of time. Recall these processes must be cáglád. Assume the underlying global frailty to be piecewise constant and cáglád as well, updating at the same times as Y_t. If we know $h_0(t)$ follows this form, a reasonable choice in estimating the frailty would be a parametric piecewise constant specification. Assume we estimate the global frailty using the likelihood presented above. We estimate $h_0(t)$ and β_0 by maximizing a sum of likelihoods over β and the parametrized piecewise constant cáglád function representing $h_0(t)$ in estimation. Because h and β^1 always enter $J(h, \beta)$ in the form $h(s)\exp\left(\beta^{1\prime} Y(s)\right)$, in this setup, there does not exist a unique maximum of $J(h, \beta)$. No matter which values we choose for (h, β), we can always increase

β^1 and reduce $h(s)$ to produce the same process $h(s)\exp\left(\beta^{1\prime}Y(s)\right)$. There are therefore an infinite number of potential maxima. These values will be far from each other in any reasonable metric.

This is not just a conditional expected likelihood issue. In the sum of likelihoods with the observed data, we can make the same type of trade off. In this piecewise constant case, there does not even exist a unique maximum of the likelihood criterion function. There are many choices of (h, β) which give the exact same maximum value. Implementation of the estimator does not make sense.

The problem with this piecewise constant counterexample is still present in our block sampling setup presented in Section 2. Regardless of how many blocks of data we observe, there is still this type of trade off in realized criterion function and conditional expected likelihoods. For any choice of (h, β) in the criterion function used for estimation, there exists an infinite number of other choices for (h, β) which give the same value. Again, these values are far from each other in any reasonable metric. We will further discuss this counterexample in the sequel.

In Counterexample 1, we effectively have no separate identification of the elements of interest. The trade off between h and β trades off between the impacts of the frailty and macro variables respectively. As a result, we produces the same hazard function for various choices of (h, β), conditional on the realization of $Y(t)$. We cannot tell what portion of the hazard rate is the result of the macro variables and what portion is the result of frailty.

This type of identification failure is not restricted to the above counterexample. Our goal is to flexibly recover the path $h_0(t)$. If we assume h_0 is cáglád (or continuous) and the macro variables have piecewise constant paths, we still have identification problems without additional restrictions. This is because our nonparametric estimates can becoming increasingly close to piecewise constant functions in the limit, allowing for the same type of trade-off as before. Similar problems can happen in other situations. The case of piecewise constant covariates is simply the most likely to be encountered.

Another approach is to arbitrarily restrict the form that h_0 can take. We may be unwilling to do this because economic theory suggest no such restriction, or because the frailty path may have an unexpected form we wish to uncover. For example, restricting the derivatives of h_0 can solve the above mentioned problem. However, it is not obvious what level to restrict the derivatives at. Additionally, the underlying frailty may in fact have discontinuities, as it does in Counterexample 1. In the sequel, I will show how to use the sampling scheme presented above to produce consistent estimates under minimal assumptions on $h_0(t)$.

The key to our identification and estimation approach is the difference between conditional expected likelihoods and full expected likelihoods. Note that, if the macroeconomic variables are assumed to be random, a full expected likelihood can be defined, despite the fact that a simple cross section will only converge to the conditional version. It is possible that the full expected likelihood has a unique maxima while the conditional likelihood does not. This full likelihood will then give identification and a basis for consistency. The question becomes, how do we use sampling to get the full expected likelihood instead of the conditional one?

The answer comes from our sampling scheme. By observing an increasing number of blocks, we observe an increasing amount of the macro variables. We use this additional information about macro covariates

to "fill in" the expectations, giving the full likelihood. This works in tandem with the increasing number of observations within blocks, allowing us to recover both the impact of the macroeconomic variables and the underlying frailty. Recall that the full expected likelihoods $H(h, \beta)$ have unique maxima by assumption (A6).

The use of the full expected likelihoods is one reason for assuming random macroeconomic variables. Without this randomness, we cannot define these objects, and the consistency proof fails. Another approach sometimes taken is to simply condition our the macro variables. Here, we crucially need the additional structure.

In what follows, we only assume the path $h_0(t)$ is cáglád with a few additional weak conditions. Notably, this does not rule out estimation of Counterexample 1. Indeed, the estimation approach presented below is able to identify and consistently estimate the frailty path in this case. This is for reasons discussed above. The full expected likelihood can have a unique maxima while the conditional expected likelihood does not. The same type of phenomena can arise in other situations. This is fully discussed in the sequel.

Loss of identification based on a trade-off between frailty and macro variables is not restricted to our hazard case. Consider the following counterexample using a discrete time approach.

Counterexample 2 Take the standard probit model presented in Newey and McFadden (1994). This model can easily be put into a panel setup, where the likelihood describes the probability of an event happening across observation times. The likelihoods will have the form

$$f(z|\,\theta_z, \theta_t) = \Phi\left(\theta_t + z_t'\theta_z\right)^{\tau_i}\left[1 - \Phi\left(\theta_t + z_t'\theta_z\right)\right]^{1-\tau_i}.$$

Here, Φ is the standard normal CDF. z_t represents macro variables and θ_t represents the global frailty. θ_t is indexed by t because it varies across observation times in the panel. In this model, we can always trade off the θ_t and the θ_z to produce the same value of the likelihood. It does not matter how large the number of observed periods T is or how large the number of observations at any time t is. There never is a unique maximum in this model. We do not have identification between macro variables and global frailty. Indeed, we face the same problem as in Counterexample 1.

In the continuous time model, we will defeat this problem by restricting the choices $\widehat{h}$ (the estimate of $h_0(t)$) can take, given a specific amount of data. As we gather more data, the restrictions on our choices for $\widehat{h}$ weaken. In the limit, these restrictions disappear and the underlying function can be very flexible. In the sequel, we will consider controlling the first derivative of $\widehat{h}$. Greater amounts of data allow us to increase the maximum first derivative. In the limit, the possible first derivative approaches infinity, removing the restriction.

3.2 Cross Sectional Estimation

In this subsection, I analyze the cross sectional context without macro variables. A cross section corresponds to $G^i = 0$. As shown above, this allows for easy identification with a point process likelihood as the basis of estimation. The trade off between frailty and macro variables is not present. Let Θ_n be

a space of functions depending on n, the number of observations. More specifics on the required sieve spaces Θ_n are given below.

Define our estimator as

$$Q_n\left(\widehat{h},\widehat{\beta}\right) \geq \sup_{\beta,h\in\Theta_n} Q_n\left(h,\beta\right) + o_p\left(1\right),$$

where the criterion function is

$$Q_n\left(h,\beta\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^1 \log\left[h\left(s\right)\exp\left(\beta' Z^i\left(s\right)\right)\right] dN_s^i + \int_0^1 \left[1 - h\left(s\right)\exp\left(\beta' Z^i\left(s\right)\right)\mathbf{1}_{\{\tau_i \geq s\}}\right] ds\right].$$

We make the high level assumption (A7) on the covariate processes. This assumption is presented in the Appendix. Below, a number of examples of covariates which satisfy (A7) are given. Note that scaling of the processes may be required to fit the supports into $[0,1]^q$, but this is not a substantive issue. There are many other covariate processes that satisfy (A7). As piecewise constant variables are the most important in applications, these are the focus below. Essentially any piecewise constant covariates will satisfy the needed assumption.

I note in Example 2 that (A7) is a natural functional extension of a standard continuity assumption on covariates that do not vary through time.

Example 1 *Let ξ^i be a continuously distributed random variable with compact support. Define the covariate process as*

$$Z^i\left(s\right) = \xi^i + ct,$$

where $c \in \mathbb{R}$. This process satisfies assumption (A7). Note that we can choose $c = 0$. Here, the covariate process reduces to the static case. Therefore, assumption (A7) can be seen as a natural extension of a standard continuity assumption made when covariates are time-invariant. If ξ^i have a discrete distribution, assumption (A7) holds as well.

Example 2 *Let $N^i\left(s\right)$ be a point process constructed as in Section 2.2, but where there are an infinite number of ordered random times. Each random time is represented by ϕ_j. Let the covariate process have an initial distribution ξ_0^i with support $[0,1]$. At each ϕ_j, a new value is drawn ξ_j^i with support $[0,1]$. This distribution may be dependent on all previous draws $\xi_{j-1}^i, \ldots, \xi_0^i$ and/or the positions of the previous ϕ_j. We assume all ξ_j^i have continuous distributions. The covariate process is*

$$Z^i\left(s\right) = \xi_0^i \mathbf{1}_{\{0\leq s\leq \phi_1\}} + \sum_{j=1}^{\infty} \xi_j^i \mathbf{1}_{\left\{\phi_j < s \leq \phi_{j+1}\right\}}.$$

If, $0 < \mathbb{P}\left\{\phi_j \in (a,b)\right\} < 1$ for all j and all intervals (a,b) contained in $[0,1]$, then Assumption (A7) holds. Weaker assumptions are possible, but have more involved statements. Another simple example is covariates which only change values at fixed times, such as weekly or quarterly. If the updates are continuously distributed on $[0,1]$, (A7) is satisfied.

Example 3 *Assume $Z^i\left(t\right)$ is composed of d covariate processes, each of which has a form given in the previous examples. Assume that over any time interval, there is a positive probability that one of the*

covariate processes has a discontinuity while none of the others do. Similarly, we assume that over any time interval there is a positive probability of no discontinuity. If g is a bounded continuous function, then

$$g\left(Z^i(t)\right)$$

satisfies (A7). Another possibility is that all covariates $Z^i(t)$ are piecewise constant and update at the same times with continuous distributions. This also satisfies (A7).

I now present the consistency result.

Theorem 4 *We make Assumptions (A1)-(A7). Choose a sequence of sieve spaces Θ_n satisfying the following conditions. Let there exist a sequence $h_n \in \Theta_n$ such that $h_n \to^{L^1} h_0$. $h_0 \subset \overline{\mathcal{H}}$ where all functions in $\overline{\mathcal{H}}$ are cáglád and bounded above and below by known fixed constants $C_{\min}$, $C_{\max}$. Assume further that for $h \in \Theta_n$*

$$\begin{aligned} C_{\min} &\leq h \leq C_{\max}, & (2)\\ \left|h'\right| &\leq K_n, & (3) \end{aligned}$$

where $K_n = O\left(n^{1/4-\eta}\right)$ for a small $\eta > 0$ and $K_n \to \infty$. Define,

$$p = 1/\left(\frac{3}{4} + \eta\right).$$

For the system of σ-fields

$$\mathcal{K}_l^m = \vee_{i=l}^m \sigma\left\{\eta_i, X_t^i \middle| 0 \leq t \leq T\right\}, \quad (4)$$

assume the α-mixing coefficients[11] *satisfy,*

$$\sum_{n>0} n^{p-2}\alpha(n) < \infty. \quad (5)$$

Then

$$\widehat{\beta} \to \beta_0$$
$$\widehat{h} \to^{L^1} h_0$$

$\mathbb{P}_{\alpha_0} - a.s.$

Proof. See Appendix. ■

Comments: 1. The theorem is a consistency result. This is a first-step toward other asymptotics, such as asymptotic normality of $\widehat{\beta}$ or functionals of $\widehat{h}$. These further results are the topic of ongoing research. See Chen (2007), (2011) for comprehensive treatments of large sample sieve methods.

2. The proof of the result shows that, if a sequence (h_n, β_n) satisfies

$$H(h_0, \beta_0) - H(h_n, \beta_n) \to 0,$$

[11] See Davidson (1994) for a definition of α-mixing.

where $H(h,\beta)$ is the expected log-likelihood, then (h_n,β_n) converges (i.e. $h_n \to^{L^1} h_0$ and $\beta_n \to \beta_0$). This result is equivalent to showing the situation is well-posed and the expected log-likelihood is identifiably unique (See Chen (2007),(2011)). This result depends on the assumed structure of the covariate paths made in (A7). Convergence of $\left(\widehat{h},\widehat{\beta}\right)$ then follows by showing

$$H(h_0,\beta_0) - H\left(\widehat{h},\widehat{\beta}\right) \to 0,$$

almost surely. This is clearly not possible in Counterexample 1. In that case, $\widehat{h}$ can be far from h_0 in an L^1 sense, even asymptotically. Therefore, consistency is impossible.

3. In our estimation procedure, we allow the first derivative of the sieve spaces to grow as we gather more data. This happens as $K_n \to \infty$. This is done because we may be unwilling to bound the first derivative on the underlying path h_0. We allow for arbitrarily large derivatives in our estimates in the limit. Another possibility is that h_0 has discontinuities. In this case, our estimator still converges in L^1. The estimate will become increasingly steep around the jumps, becoming discontinuous in the limit.

4. For n observations, the sieve space containing the largest number of potential functions for approximation is the set of all functions satisfying (2)-(3). However, this does not have an obvious set of basis functions. In applications, a sieve with know basis functions whose coefficients can be constrained to satisfy (2)-(3) will be used. Theoretically, we would simply take the intersection of a sieve space with (2)-(3). In applications, there will be issues of implementation depending on the chosen basis functions. There will also be issues of bias-variance trade off in finite samples.

5. The ability to control the first derivative of the sieve space is important for our result. Cardinal B-Splines, for example, can do this easily. See de Boor (2001) or Chui (1992) for more on Cardinal B-Splines. Chen (2007) contains a lengthy discussion of potential sieve space choices.

3.3 Global Frailty and Likelihood Estimation

I now use similar likelihood methods to estimate the frailty model. Assume the relevant covariates are $Z^{ij}(t) = \left(X_t^{ij}, Y_{G^j+t}\right)$. The same assumptions as before will be made. However, in the current frailty case, each block of observations has a different expected likelihood because each block has a different baseline hazard.

We define our estimator similarly as in Section 3.2. Here, Θ_n^j are spaces of functions increasing in size with n. Each block has its own sieve space Θ_n^j in the estimation. Define our estimator as,

$$Q_n\left(\widehat{h},\widehat{\beta}\right) \geq \sup_{\beta,h^j\in\Theta_n^j, j=1,\ldots,k(n)} Q_n(h,\beta) + o_p(1)$$

where the criterion function is

$$Q_n(h,\beta) = \sum_{j=1}^{k(n)} \frac{1}{n} \sum_{i=1}^{n} \left[\begin{array}{c} \int_0^1 \log\left[h^j(s)\exp\left(\beta' Z^{ji}(s)\right)\right] dN_s^{ji} \\ + \int_0^1 \left[1 - h^j(s)\exp\left(\beta' Z^{ji}(s)\right)\mathbf{1}_{\{\tau_{ji}\geq s\}}\right] ds \end{array} \right]. \tag{6}$$

This is a sum of the likelihoods over the blocks $k(n)$. Portions of the global frailty h_0 which overlap in different calendar time blocks $\left[G^i, G^i+T\right]$ are restricted to be the same function in estimation. For

example, if $[0,T]$ and $\left[\frac{1}{2}T,\frac{3}{2}T\right]$ are calendar time blocks, then the estimate corresponding to the first block over $\left[\frac{1}{2}T,T\right]$ must be the same function as the estimate corresponding to the second block over the same interval. As we can allow for the sieve spaces Θ_n^j to differ across blocks, this is not an issue in implementation. The estimate of h_0 from this criterion function is an estimate of the entire frailty path over the observed blocks.

The proof of Theorem 5 in the previous section relies on the fact that, if

$$H\left(h_0,\beta_0\right)-H\left(\widehat{h},\widehat{\beta}\right)\to 0,$$

almost surely, then the components $\left(\widehat{h},\widehat{\beta}\right)$ converge to the true values almost surely. In the frailty case, the expectation of the criterion function is now

$$\sum_{j=1}^{k(n)}H^j\left(h^j,\beta\right),$$

where each term in the sum is the expected likelihood of the corresponding block of observations. The expected likelihoods are now indexed by j because they differ across blocks. By similar arguments, if

$$\sum_{j=1}^{k(n)}\left\{H^j\left(h_0^j,\beta_0\right)-H^j\left(\widehat{h}^j,\widehat{\beta}\right)\right\}\to 0,$$

then the components $\left(\widehat{h}^j,\widehat{\beta}\right)$ must all converge to their true values. We can then reconstruct the frailty path by piecing together the components $\widehat{h}^j$. This is what is done in the proof of Theorem 7 below.

What is interesting about this setup is that, if we look at a block of observations individually, the average of the likelihoods does not converge to the expected likelihood. It converges to the conditional expected likelihood, where conditioning is on the realization of the macro variables. This is for reasons discussed in Section 3.1. However, the expected likelihood for each block still exists and can be used to prove identification and consistency. By observing more blocks as $n\to\infty$, we sample more of the common variables, and this information can be used to fill in the expected likelihoods. We get identification and consistency in the limit. For any fixed finite set of blocks, identification and consistency fail.

Here is the formal result.

Theorem 5 *Assume (A1)-(A7). More specifically, we assume (A6) for each block of observations. Choose a sequence of sieve spaces Θ_n^j satisfying the following condition. For all $j\in\mathbb{N}$, let there exist a sequence $h_n^j\in\Theta_n^j$ such that $h_n^j\to^{L^1}h_0^j(t)$. For all $j\in\mathbb{N}$, $h_0^j\subset\overline{\mathcal{H}}$ where $\overline{\mathcal{H}}$ is a space of cáglád functions bounded above and below by known fixed constants $C_{\min}$, $C_{\max}$. Assume further that for $h\in\Theta_n^j$,*

$$\begin{aligned} C_{\min} &\leq h\leq C_{\max}, &(7)\\ \left|h'\right| &\leq K_n. &(8)\end{aligned}$$

Let $k(n)$ and K_n be chosen such that the conditions (48)-(51) in the Appendix are satisfied. additionally, $n\to\infty$, $k(n)\to\infty$, $K_n\to\infty$ and $K_n\left(n/k(n)\right)^{1/2}=o(1)$. Let there exist $\widetilde{h}_n^j\in\Theta_n^j$ for each j (where

functions with overlapping calendar times must agree on the overlapping intervals) such that

$$\sum_{j=1}^{k(n)} \left\{ H^j \left(h_0^j, \beta_0 \right) - H^j \left(\widetilde{h}_n^j, \beta_0 \right) \right\} \to 0. \tag{9}$$

Then

$$\widehat{\beta} \to \beta_0,$$

$$\widehat{h}^j \to^{L^1} h_0^j,$$

for all j, $\mathbb{P}_{\alpha_0} - a.s.$

Proof. See Appendix. ■

Remarks: 1. Condition (9) requires that the best possible choice of function from the sieve spaces Θ_n^j approximate h_0^j fast enough to counteract the increasing number of blocks $k(n)$. Whether or not this is satisfied will depend on the chosen sieve spaces. (9) is trivially satisfied in the cross sectional case. With a cross section, the number of terms in the sum is one and does not grow.

2. n is the number of observations per block. The choice of K_n controls the complexity of the sieve space by controlling how large the first derivatives of functions in Θ_n^j can be. The choice of $k(n)$ controls how much of the macroeconomic variables we observe by describing how many blocks of observations we observe. Both K_n and $k(n)$ approach infinity as $n \to \infty$.

3. Restrictions on K_n are what allow us to avoid the problems of Counterexample 1. This also prevents similar situations from occurring asymptotically. For example, it is possible that K_n increases "too quickly", and that $k(n)$ increases "too slowly", allowing an asymptotic version of Counterexample 1 to arise. In this situation, for a given n, sieve functions are allowed to have first derivatives that are too large. This can lead to steep paths which approach piecewise constant functions too quickly. The frailty does not have to be piecewise constant for these problems to occur. It is a general issue when estimating global frailty in the presence of macroeconomic variables.

4. Fixing the complexity of the sieves, increasing the number of observed blocks improves our estimates of β. This is because all terms in our criterion function (6) contain β. However, if we do not increase the complexity of the sieve spaces we will never achieve consistency. $\left(\widehat{h}^j, \widehat{\beta}\right)$ will converge to a best approximation of the true values from a set of potential approximations. This set is restricted by what forms the sieve spaces take. In order to achieve consistency, we must increase the complexity of the sieve spaces Θ_n^j with n.

In our estimator, this is done by restricting how the terms n, $k(n)$ and K_n interact. We assume,

$$\left(k(n)/n\right)^{1/2} K_n = o(1).$$

For this to hold, n must increase to infinity at a faster rate than $k(n)$. In addition, $\left(k(n)/n\right)^{1/2}$ must also decrease to zero at a faster rate than K_n increases to infinity.

In words, to achieve consistency we must gather more blocks quickly enough, and the first derivatives of our sieve choices must increase slowly enough, that we cannot make an asymptotic trade-off similar to Counterexample 1. Effectively, our information about the macroeconomic processes is increasing so fast that bad behavior of our sieve estimates is ruled out. If we allow the sieve spaces Θ_n^j to become complex

too quickly compared to $k(n)$, this can overwhelm the additional information about the macro variables provided by observing an increasing number of blocks. Inconsistency will follow.

5. Similar restrictions on how the three elements K_n, $k(n)$ and n interact are made in (48)-(51) in the Appendix. The assumptions (48)-(51) are high level. To satisfy these conditions, we must consider the dependence between X^{ij} within blocks and across blocks. In addition, the temporal dependence of $Y(t)$ and its dependence with the processes $X^{ij}(t)$ must be accounted for.

In the point process likelihood approach, the needed conditions (48)-(51) are a natural extension of those used in the cross sectional case. However, exact specification of the required dependence that facilitates these conditions is not obvious. One issue is that there is no simple ordering of the sample. Dependence has to be controlled in two dimensions, within blocks and across blocks. As almost sure convergence is required and we are dealing with an array structure, almost sure convergence of arrays will likely be necessary. See Liebscher (1996). I leave exact characterization of the needed dependence to future research.

6. The results produce an estimate of the current value of frailty. This is of interest for current decision making or forecasting contexts.

7. Another possibility for estimation is that each block of observations has a different baseline hazard, but there is no restriction on the baseline hazards across blocks. This would allow for cohort effects where different blocks have totally unrelated baseline hazards. A potential example of this situation is hazard analysis of mortgage defaults in the aftermath of the 2008 financial crisis. If for unobserved reasons, potentially related to falsified information in mortgage applications, mortgages originated at different times had very different baseline hazards, this approach would be appropriate. With slight adjustments, consistency will hold under the same assumptions made in Theorem 6.

8. In economic contexts, mixed proportional hazard models have received a large amount of attention. See Van den Berg (2001). A mixed proportional hazard model with this type of frailty is not identified. This is because a scaling of $h_0(s)$ is required for identification (again, see Van den Berg (2001)). As $h_0(s)$ is an unknown path, it is impossible to have such a scaling.

4 Simulations and Empirical Application

In this section, I first present results from a simulation study. The goal is to show that the estimation approach preforms reasonably well, even in relatively small samples. Second, I implement our estimation approach on a data set of the corporate default experience of Moody's rated firms over a 30 year period. The global frailty over this period is estimated while accounting for a number of macroeconomic variables. The results suggest that frailty may not follow a simple stochastic form.

The simulation design assumes we observe the window of calendar time $[0, 5]$. This will correspond to five years in our empirical application. The hazard rates are impacted by a single piecewise constant macro covariate which updates every $1/12$ units of calendar time. The macro variable is assumed to take the following form,

$$\begin{aligned} Y_t &= \phi + (\gamma - Y_{t-1})\,\xi_t, \\ Y_0 &= \phi\xi_0. \end{aligned}$$

Here, the ξ_t are standard normal random variables. Throughout the simulations, we choose the values $\phi = 1$ and $\gamma = 0.3$. Other values were tried with similar results. This form was assumed for simplicity.

Observations are constructed to have a Cox proportional hazard form as in the text above. Therefore, the hazard functions take the form

$$h_0(t) \exp\{\beta_1 Y_t\} \mathbf{1}_{\{\tau_i \geq t\}}.$$

For all simulations it is assumed that $\beta_1 = 0$. Other values were tried with similar results. Several functions $h_0(t)$ were tried as well. In particular, simulations are provided assuming

$$\begin{aligned} h_0(t) &= 0.3, \\ h_0(t) &= 0.3 + t^2/125, \\ h_0(t) &= 0.2 + (0.008)\left(-t^2 + 5t\right) \\ h_0(t) &= 0.2 + (0.016)\left(-t^2 + 5t\right). \end{aligned}$$

Estimation is conducted by drawing 100 observations, each assumed to start at time $G^i = 0$. It is observed when (or if) these observations default over the time interval $[0, 5]$. $h_0(t)$ and β_1 are estimated semiparametrically using the likelihood approach presented above. This simulation is replicated 500 times.

In the likelihood, $h_0(t)$ is estimated using Cardinal B-Splines of order 3. The basic spline function used is

$$B_3(x) = \frac{1}{2}\left[\max(0, x)\right]^2 - \frac{3}{2}\left[\max(0, x-1)\right]^2 + \frac{3}{2}\left[\max(0, x-2)\right]^2 - \frac{1}{2}\left[\max(0, x-3)\right]^2.$$

This function is hill-shaped and has support $[0, 3]$. Seven of these functions are used to estimate $h_0(t)$. Their supports are translated to start at the calendar times $-2, -1, 0, 1, 2, 3, 4$. We estimate coefficients for each of these translated functions. Therefore, in estimation we maximize over eight parameters, seven spline function coefficients and one for the macro variable.

Panel 1 presents the estimation results for the choices $h_0(t) = 0.3$ and $h_0(t) = 0.3 + t^2/125$. In the graphs, the solid line is the true function while the dashed line is an average of our estimates over the 500 replications. The symmetric dotted lines around the average estimated function represent two standard deviations. This is computed using the 500 replication sample.

The average functions do well in these cases. They closely follow the true $h_0(t)$ which is well within the two standard deviation bands. The distribution of the estimates $\widehat{\beta}_1$ are presented to the right of the results for $h_0(t)$. These also preform well. Without further theory deriving asymptotic confidence intervals, little more can be said about the performance.

Panel 2 presents similar results for the choices $h_0(t) = 0.2 + (0.008)\left(-t^2 + 5t\right)$ and $h_0(t) = 0.2 + (0.016)\left(-t^2 + 5t\right)$. These are given in the first two rows of Panel 2. The performance of the estimator is relatively poor. The average of the estimate functions does not have the shape of the true function. Additionally, the true function is very close to the two standard deviation bands in one case. The estimates of β_1 preform similarly to those in Panel 1. It seems that estimates of β_1 are less sensitive to the underlying specification than the estimates of $h_0(t)$.

There are a number of potential reasons these estimators have more issues than the other functions. First, it may be that our choice of sieve space has trouble representing these functions. Second, the number of observations may be too small. Finally, it should be noted that in this setup, even if the number of observations approaches infinity we will not get consistency. This is because we are assuming all observations start at time zero. As we argued above, because the covariate is a macro variable, we need observations through time to get consistency.

In order to investigate this further, additional simulations were conducted. Simulations with larger sample sizes or with *i.i.d.* data did not meaningfully change these results. However, notice how in all the previously presented simulations, the two standard deviation bands widen at the beginning and end of the time interval. This is because we have spline functions with support $[-2, 1]$ and $[4, 6]$ in the estimation. These functions only overlap the relevant interval with one third of their support. This allow for much more freedom in estimation toward the beginning and end of the samples. Essentially, the bias-variance trade off is skewed more toward variance than in other places in the sample. This shows up in the confidence intervals. Indeed, at the beginning and end of $[0, 5]$ the bands have a similar form as the basic Cardinal B-Spline function $B_3(x)$ at the first and last third of its support $[0, 3]$.

The final row in Panel 2 presents the same simulations as above with $h_0(t) = 0.2 + (0.008)\left(-t^2 + 5t\right)$. However, the spline functions starting at -2 and 4 are removed from the estimation. As can be seen from the results, the average of the estimates follows more closely the form of the underlying functions. As fewer basis functions are used in estimation, the estimates are less able to capture the exact form of $h_0(t)$. On balance, it appears that this estimator preforms better than the previous one.

It seems that estimation is somewhat sensitive to the exact sieve spaces we choose. This is usually the case in sieve estimation approaches. It should also be noted that 100 observations is a small sample for sieve estimation methods. Our empirical example has over 90 times as much data. Additional data will mitigate the high variance problems we observe at the edges of the time interval. Given the relatively small amount of data, the simulations suggest that our estimation approach preforms well.

I now present an application of these methods to corporate default. Data on corporate defaults is taken from Moody's Default and Recovery Database. This data set has been extensively examined empirically. Just a sampling of previous citations includes Das et al. (2007); Duffie, Saita and Wang (2007); Duffie et al. (2009); Lando and Nielsen (2010); Azizpour, Giesecke and Kim (2011); Creal et al. (2011); Giesecke and Kim (2011a,b); Koopman et al. (2011) and Azizpour et al. (2012).

All Moody's rated firms in the industrial category from Jan 1, 1982 to Dec 31, 2011 are used in this analysis. Additionally, any default dates or other censoring are recorded. Firms are assumed censored if they leave the sample for any reason other than default. This comes to a total of 9264 firms with 1604 defaults in the sample. Macro covariates were also collected in order to estimate their impact on corporate hazard rates. Following Lando and Nielsen (2010), the following macro covariates were used:

1. Trailing 1-year return on the S&P500.
2. 3-month US Treasury rate.
3. Trailing 1-year change in US industrial production index.
4. Spread between 10-year and 1-year Treasury rate.

5. 10-year and 1-year Treasury spread minus 2-year and 1-year Treasury spread.

The first four variables are the same as those used in Lando and Nielsen (2010). They are drawn from CRSP and the US Federal Reserve Board. The fifth covariate is new and meant to capture how much interest rate spreads change over different time horizons. This variable is derived from US Treasury data found on the US Federal Reserve Board website.

Unlike in some previous research, firm specific data is not used here. Matching firms in Moody's database to other databases containing firm specific covariates significantly reduces the number of observations. This is because many observations do not have matching firm specific data in available databases. I chose to keep the additional observations by not using firm specific data. This is important for interpreting the estimates. Exact comparisons with some previous research is not possible. Given the size of data sets from previous studies which utilize firm specific data, our simulations suggest that our estimation approach would preform well even with those reduced sample sizes.

As proposed in Duffie et al. (2009), it is likely that corporate default hazard rates contain an element of global frailty. In order to investigate this, I implement the frailty estimation approach of Section 3 on the data set outlined in the previous paragraphs. Some of the needed assumptions are questionable. For example, it is definitely not the case that a large number of observations start at any calendar time in the observed interval. However, with 9264 observations there is a large number of observations at risk of default at any time in the observed interval. As noted above, a minimal requirement for estimating global frailty is that each calendar time is covered by a large number of observations.

The sieves used are again cardinal B-splines. The global frailty was estimates with B-splines having support corresponding to three years in the 30 year interval. The basis functions were assumed to have supports starting at 1981 through 2009 with a new basis function starting at each year in that interval. This is analogous to the simulation study. Notice that we have removed the spline functions starting at 1980 and 2010 in accordance with our findings from simulations. This will involve 29 coefficients for the underlying frailty. Added to this will be five macro variables, for a total of 34 parameters to be estimated.

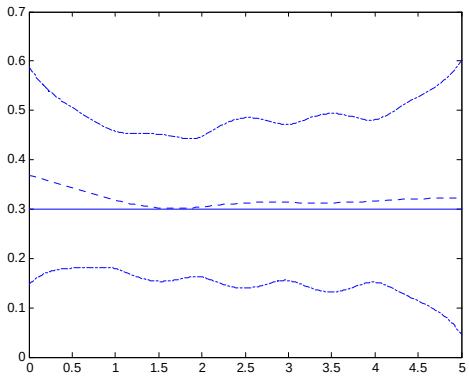
First, the method was conducted assuming there were no covariates. Therefore, only a measure of frailty was produced. The results of this are presented in Figure 1. The measure of frailty broadly corresponds to the default history over this period. Figure 2 gives the number of defaults per year over the same interval. This is encouraging as this is what we would expect. The estimate is similar to the number of defaults divided by the number of firms which are at risk.

Next, the estimation approach is taken while accounting for the five macro covariates described above. The results for the estimated β's are given in Table 1. The estimates are all reasonable. As we would expect, the coefficients on the one year S&P 500 return and Industrial Production Index are negative. The higher the returns the less likely defaults are. Similarly, the more productive industrial production is, the less likely firms will default. The three interest rate variables are difficult to interpret because their influences are clearly related. Taken together, these estimates are not unreasonable.

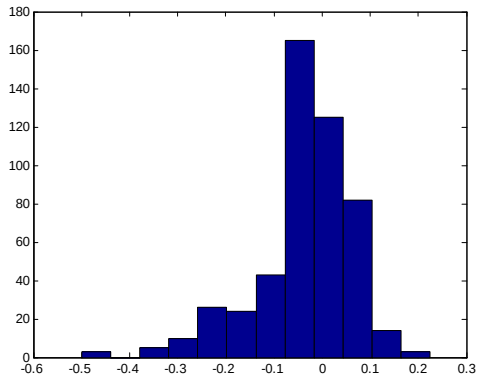
Finally, Figure 3 contains our estimate of the frailty while allowing for macro covariates. The scale of this frailty estimate is smaller than the one estimated with no covariates. This is expected as the observed variables are describing part of the hazard rates. Even after the covariates are accounted for, recent experience seems to depart substantially from the earlier part of the observed time interval.

Toward the beginning of the time interval, the estimate of frailty is very small but positive. This seems unrealistic and represents possible downward bias in the estimate. However, the same downward bias effect would also be impacting the right end of the time interval. Notice that this issue does not appear in the estimate with no covariates. On balance, it seems the results are approximate but reasonable.

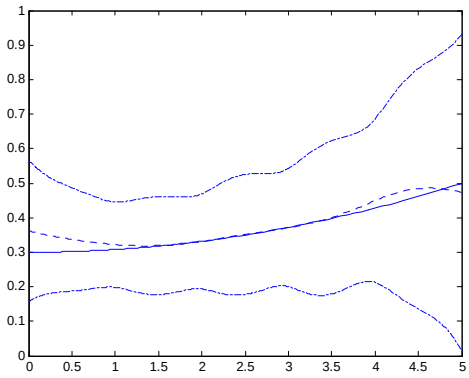
A simple mean-reverting process may have difficulty capturing these results. Even if a simple process can capture this, it is not obvious which form should be used. For example, in Duffie et al. (2009) the frailty is assumed to be inside the exponential. Is this a better choice than putting the diffusion outside? It is not clear before estimation. After estimation, you have a data snooping problem. The uncertainty about the form stochastic global frailty takes raises problems related to identification, model selection and data snooping. All these issues cast doubt on the validity of forecasts based on filtering approaches.



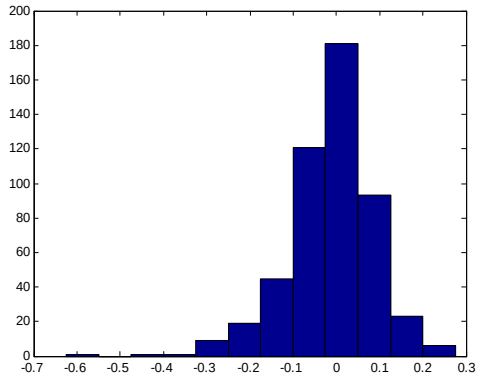
$h_0(t) = 0.3$



$\beta_1 = 0$

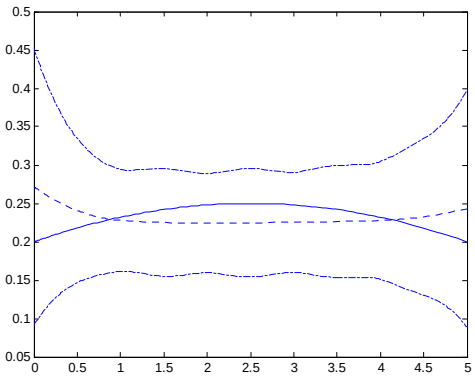


$h_0(t) = 0.3 + t^2/125$

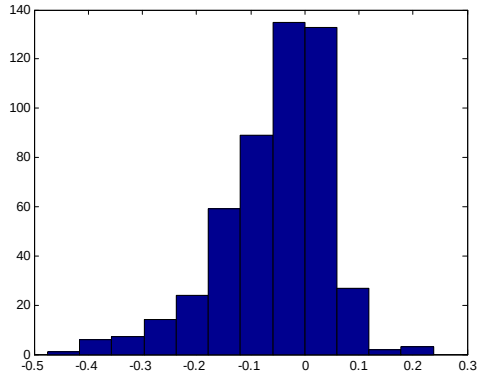


$\beta_1 = 0$

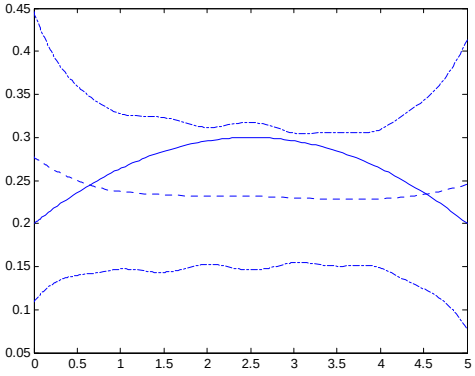
Panel 1.– Estimates of $h_0(t)$ and β_1 from Monte Carlo.



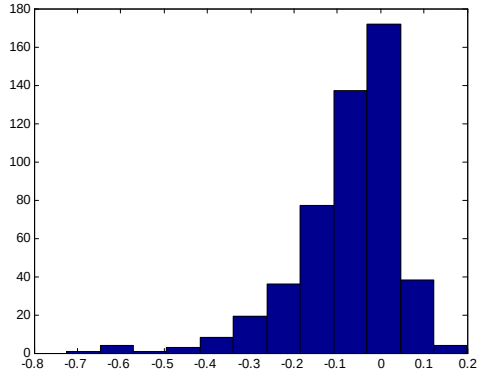
$h_0(t) = 0.2 + (0.008)(-t^2 + 5t)$



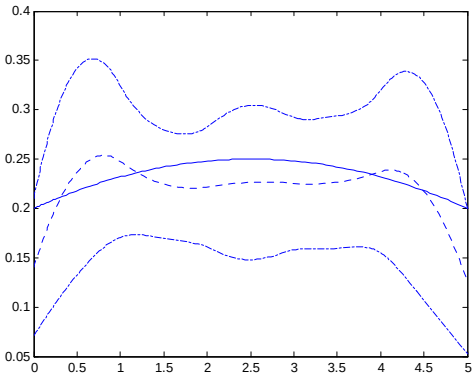
$\beta_1 = 0$



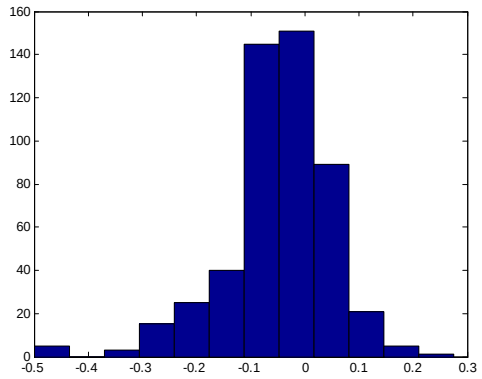
$h_0(t) = 0.2 + (0.016)(-t^2 + 5t)$



$\beta_1 = 0$



$h_0(t) = 0.2 + (0.008)(-t^2 + 5t)$



$\beta_1 = 0$

Panel 2.– Estimates of $h_0(t)$ and β_1 from Monte Carlo.

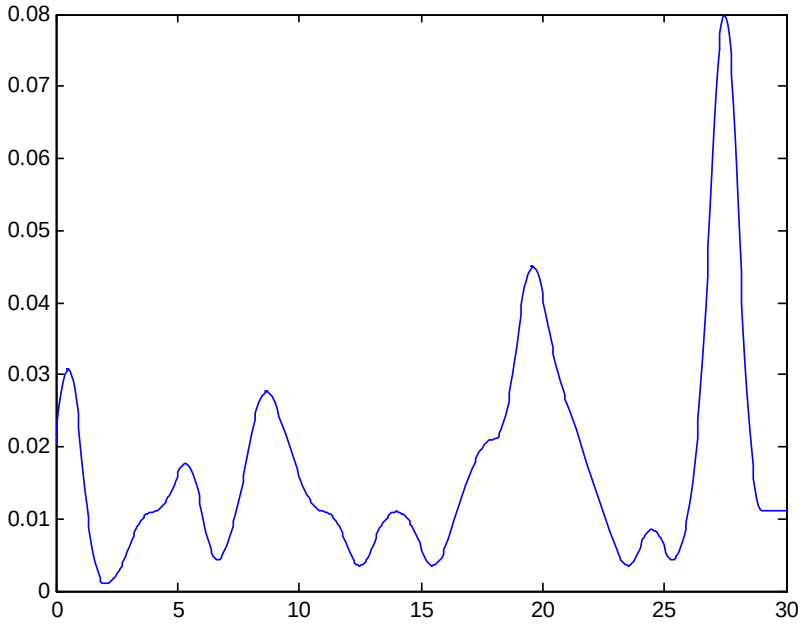


Figure 1: Assuming no covariates, this is the estimate of the underlying frailty in the Moodys data set from Jan 1, 1982 to Dec 31, 2011. Assuming B-splines starting at -1 through 28 by years.

macro variables:	*betas*
1-year S&P500 return	-2.3
3-month US Treasury rate	0.4
Industrial production index	-0.1
10-year vs 1-year spread	4.0
10-year-1-year vs 2-year-1-year spread	-4.5

Table 1: Beta Estimates

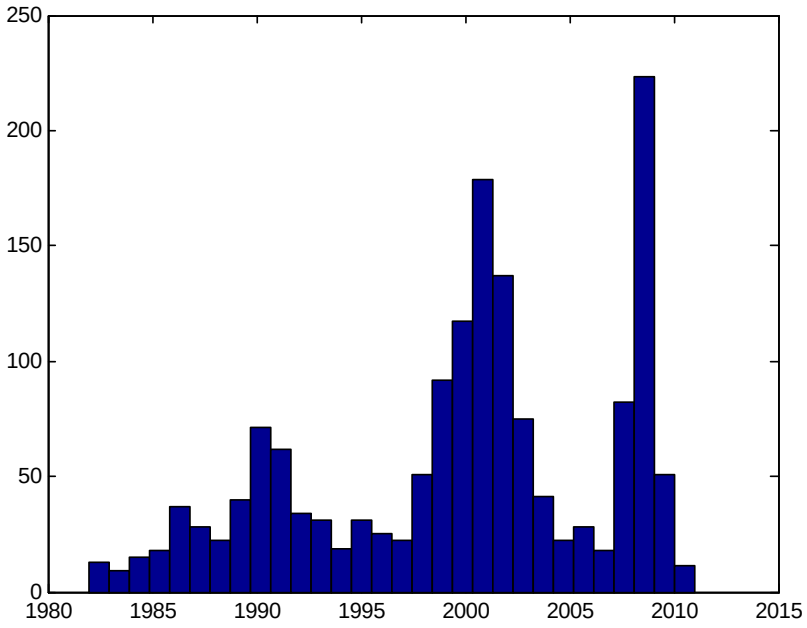


Figure 2: Number of defaults per year in the Moodys data set from Jan 1, 1982 to Dec 31, 2011.

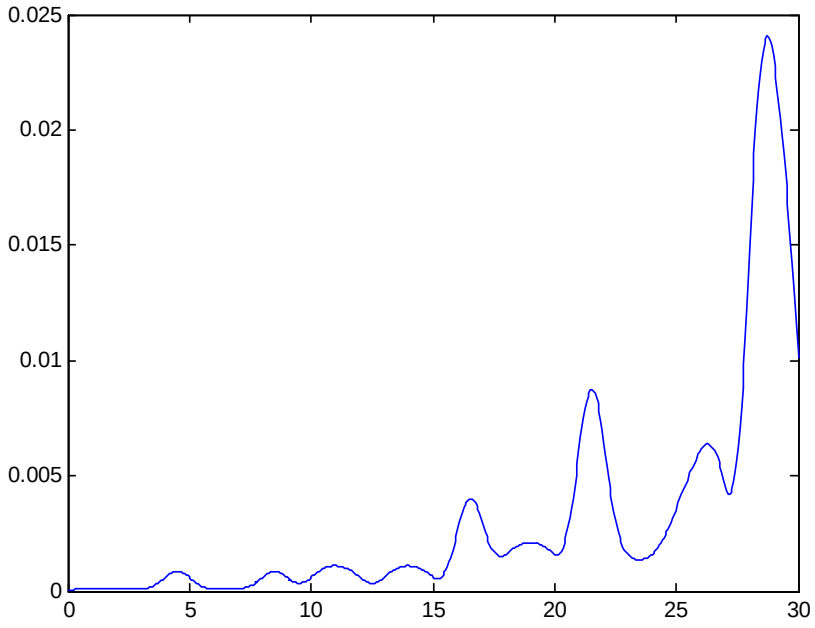


Figure 3: Assuming the five macro covariates, This is the estimate of the underlying frailty in the Moodys data set from Jan 1, 1982 to Dec 31, 2011. Assuming B-splines starting at -1 through 28 by years.

5 Conclusion

In this paper, hazard models with a number of dependence properties between observations are presented. These models are designed to capture realistic correlation between random economic events. Three types of dependence are allowed for. First, correlation between covariates that are specific to observations. For example, housing prices in a model of mortgage default. Second, hazard rates are allowed to depend on macroeconomic covariates. This type of covariate captures dependence between observations resulting from changing global economic conditions. Examples include the GDP growth rate, unemployment rate or three-month US treasury rate. Finally, the model allows for a global unobserved macroeconomic covariate. We refer to this as global frailty. This element is in some sense a residual, capturing the remaining impacts on the hazard rate after observed covariates are accounted for.

When estimating hazard rates which are impacted by both observed macroeconomic variables and an unobserved global frailty, there is a basic identification problem. The impact of the observed and unobserved macro variables can be confused. If we are not careful about our assumptions, the situation can be unidentified and estimation will be inconsistent. I show that, with an appropriate type of sampling, the impact of these two elements can be separated and consistent estimation achieved.

Simulation studies for our estimation approach reveal that the methods preform reasonably well in finite samples. However, the results may be sensitive to the chosen sieve space in small samples. Finally, these methods are applied to Moody's corporate default data. After accounting for several macro variables, I recover the underlying global frailty for Moody's rated firms over the thirty year period from 1982 to 2012. The results indicate that the underlying frailty may not be well represented by the simple stochastic processes previously assumed.

6 Appendix

Some required conditions on the covariates $Z^i(t)$ are presented here. We normalize $[0,T]$ to $[0,1]$ throughout the appendix.

(**A7**): Assume the following for each observation $i \in \mathbb{N}_0$. Assume for each $t \in [0,1]$, $Z^i(t)$ has support contained in a compact rectangle normalized to be $[0,1]^q$ for ease of notation. Assume

$$\mathbb{E}\left[\mathbf{1}_{\{\tau_i \geq t\}} \middle| Z^i(s), 0 \leq s \leq 1\right] \geq M^{\min}(t) > 0 \qquad \forall t \in (0,1], \text{ a.s.} \tag{10}$$

for some deterministic function $M^{\min}(t)$ on $t \in [0,1]$. Let $S \subset D^q[0,1]$ and give S the relative Skorokhod topology. Assume

$$\overline{\mathbb{P}}\left\{Z^i(t+) \in S\right\} = 1.$$

For all $x(t) \in S$ and all $\gamma > 0$, there exists $x'(t) \in S$ such that

$$\sup_{t \in [0,1]} \left\| x(t) - x'(t) \right\| < \gamma. \tag{11}$$

For all $x_0(t) \in S$, for all $k \in \{1, \ldots, q\}$ there exists $x_0(t) + \epsilon(t) \in S$ such that , $\epsilon^i(t) = 0$ for $i \neq j$, $t \in [0,1]$ and $\epsilon^k(t) > 0$ or $\epsilon^k(t) < 0$ on some interval with positive Lebesgue measure contained in

$[0,1]$, finally $\epsilon^k(t) = 0$ for t outside of that interval. Either (1.) For all $x_0 \in S$ and all $\epsilon > 0$,

$$\overline{\mathbb{P}}\left\{\omega |\, Z^i(t+)(\omega) \in S, d\left(Z^i(t+)(\omega), x_0\right) < \epsilon\right\} > 0,$$

$$\exists x_0' \in S,\ x_0' \neq x_0 \qquad s.t. \quad d\left(x_0, x_0'\right) < \epsilon,$$

where $d(\cdot,\cdot)$ is the Skorokhod metric. Or (2.) S consists of a finite number of paths x, each with positive probability.

Condition (10) will always be satisfied because we can construct a point process with hazard

$$\left[\inf_{v\in[0,1]} h_0(v)\right](s) \inf_{\beta,z}\left[\exp\left(\beta' z\right)\right] \mathbf{1}_{\{\tau_i \geq t\}}$$

which will correspond to an $M^{\min}(t)$ that satisfies the condition. We include this assumption for completeness.

The consistency results proven below are similar to verifying the conditions of Theorem 3.1 in Chen (2007). However, we have used specifics of our hazard case to rule out ill-posed situations. Most of the required conditions of Chen (2007) are satisfied with our assumptions.

Lemma 6 *Let f_1, f_2, $\{f_n\}$ be bounded strictly positive functions defined on $[0,1]$. If*

$$\int_0^1 \left[\frac{f_n}{f_1} - 1 - \log\left(\frac{f_n}{f_1}\right)\right] f_1 f_2 ds \to 0,$$

then

$$f_n \to^{L^1} f.$$

Proof. Note that the function $f(x) = x - 1 - \log(x)$ on the interval $(0,\infty)$ is uniquely minimized at 1 where its value is 0. The result follows easily. This results is used in Karr (1987) and Grenander (1981). ■

The following lemma implies our estimation problem is well-posed and identifiably unique.

Lemma 7 *Under assumptions (A1)-(A7), if*

$$H(h_0, \beta_0) - H(h_n, \beta_n) \to 0,$$

then

$$h_n \to^{L^1} h_0,$$

$$\beta_n \to \beta_0.$$

Proof. By a similar manipulation as in Karr's (1987) proof of Theorem 3.3,

$$
\begin{aligned}
& H\left(h_0,\beta_0\right)-H\left(h_n,\beta_n\right) \\
= \;& \mathbb{E}\left\{\int_0^1\left[h_n\left(s\right)\exp\left(\beta_n'Z^i\right)-h_0\left(s\right)\exp\left(\beta_0'Z^i\right)\right]\mathbf{1}_{\{\tau_i\geq s\}}ds\right\} \\
& -\mathbb{E}\left\{\int_0^1\log\left[\frac{h_n\left(s\right)\exp\left(\beta_n'Z^i\right)}{h_0\left(s\right)\exp\left(\beta_0'Z^i\right)}\right]dN_s^i\right\} \\
= \;& \mathbb{E}\left\{\int_0^1\left[\frac{h_n\left(s\right)\exp\left(\beta_n'Z^i\right)}{h_0\left(s\right)\exp\left(\beta_0'Z^i\right)}-1\right]h_0\left(s\right)\exp\left(\beta_0'Z^i\right)\mathbf{1}_{\{\tau_i\geq s\}}ds\right\} \\
& -\mathbb{E}\left\{\int_0^1\log\left[\frac{h_n\left(s\right)\exp\left(\beta_n'Z^i\right)}{h_0\left(s\right)\exp\left(\beta_0'Z^i\right)}\right]h_0\left(s\right)\exp\left(\beta_0'Z^i\right)\mathbf{1}_{\{\tau_i\geq s\}}ds\right\} \\
= \;& \mathbb{E}\left\{\int_0^1\left[\frac{h_n\left(s\right)\exp\left(\beta_n'Z^i\right)}{h_0\left(s\right)\exp\left(\beta_0'Z^i\right)}-1-\log\left(\frac{h_n\left(s\right)\exp\left(\beta_n'Z^i\right)}{h_0\left(s\right)\exp\left(\beta_0'Z^i\right)}\right)\right]h_0\left(s\right)\exp\left(\beta_0'Z^i\right)\mathbf{1}_{\{\tau_i\geq s\}}ds\right\} \\
= \;& \mathbb{E}\left\{\mathbb{E}\left[\begin{array}{c}\int_0^1\left[\frac{h_n(s)\exp\left(\beta_n'Z^i\right)}{h_0(s)\exp(\beta_0'Z^i)}-1-\log\left(\frac{h_n(s)\exp\left(\beta_n'Z^i\right)}{h_0(s)\exp(\beta_0'Z^i)}\right)\right] \\ \times h_0\left(s\right)\exp\left(\beta_0'Z^i\right)\mathbf{1}_{\{\tau_i\geq s\}}ds\end{array}\;\middle|\; Z^i\left(t\right),\ t\in[0,1]\right]\right\} \\
= \;& \mathbb{E}\left\{\begin{array}{c}\int_0^1\left[\frac{h_n(s)\exp\left(\beta_n'Z^i\right)}{h_0(s)\exp(\beta_0'Z^i)}-1-\log\left(\frac{h_n(s)\exp\left(\beta_n'Z^i\right)}{h_0(s)\exp(\beta_0'Z^i)}\right)\right] \\ \times h_0\left(s\right)\exp\left(\beta_0'Z^i\right)\mathbb{E}\left[\mathbf{1}_{\{\tau_i\geq s\}}\middle|\, Z^i\left(t\right),\ t\in[0,1]\right]ds\end{array}\right\} \\
\geq \;& \mathbb{E}\left\{\begin{array}{c}\int_0^1\left[\frac{h_n(s)\exp\left(\beta_n'Z^i\right)}{h_0(s)\exp(\beta_0'Z^i)}-1-\log\left(\frac{h_n(s)\exp\left(\beta_n'Z^i\right)}{h_0(s)\exp(\beta_0'Z^i)}\right)\right] \\ \times h_0\left(s\right)\exp\left(\beta_0'Z^i\right)M^{\min}\left(s\right)ds\end{array}\right\} \qquad (12)
\end{aligned}
$$

Note that we can convert the covariate processes in (12) to their right continuous cádlág versions without changing the expectation. This is because of the integral in the expectation and that cáglád processes can only have a countable number of discontinuities. From now on in the proof, we have changed $Z^i\left(s\right)$ to $Z^i\left(s+\right)$ and therefore can deal with the Skorokhod space. This allows us to exploit assumption (A7). By assumption, (12) converges to zero. Lemma 7 implies that, for a fixed x, if

$$\int_0^1\left[\frac{h_n\left(s\right)\exp\left(\beta_n'x\right)}{h_0\left(s\right)\exp\left(\beta_0'x\right)}-1-\log\left(\frac{h_n\left(s\right)\exp\left(\beta_n'x\right)}{h_0\left(s\right)\exp\left(\beta_0'x\right)}\right)\right]h_0\left(s\right)\exp\left(\beta_0'x\right)M^{\min}\left(s\right)ds\to 0, \qquad (13)$$

then

$$h_n\left(s\right)\exp\left(\beta_n'x\right)\to^{L^1}h_0\left(s\right)\exp\left(\beta_0'x\right). \qquad (14)$$

Assume (14) does not hold for an open ball in S around a path $x_0\in S$. This implies (13) does not hold for this set. By assumption (A7), this open ball has positive probability. Because of the positive probability of Z^i having a realization in this set, (12) would fail to converge to zero if the assumption is true because the function (13) is continuous as a mapping from $D\left[0,T\right]$ to $\mathbb{R}$. So (14) can not fail on an open ball in S. Therefore, (14) must hold for $x\in D$ where D is a dense set of paths in S with the relative Skorokhod topology.

A consequence is that $\sup_n \int |h_n(s)|\, ds$ is bounded. This holds because β is restricted to a compact interval. If it did not hold, (14) would fail at all paths $x(s) \in [0,1]^d$.

$$\begin{aligned}\int \left|h_n(s)\exp\left(\beta_n' x(s)\right) - h_0(s)\exp\left(\beta_0' x(s)\right)\right| ds &\geq \int \left|h_n(s)\exp\left(\beta_n' x(s)\right)\right| ds - \int \left|h_0(s)\exp\left(\beta_0' x(s)\right)\right| ds \\ &\geq C\int |h_n(s)|\, ds - \int \left|h_0(s)\exp\left(\beta_0' x(s)\right)\right| ds.\end{aligned}$$

Similarly, $\int |h_n(s)|\, ds \nrightarrow 0$ because if this happened $h_n(s)\exp\left(\beta_n' x\right) \to^{L^1} 0$ which can not happen because $h_0(s)\exp\left(\beta_0' x\right)$ is strictly positive.

Assume there exists a path $x_0 \in S$ such that (14) fails. Because (14) must hold on a dense set, by assumption (A7), for any $\gamma > 0$ there exists a $x' \in S$ which satisfied (14) such that $x_0(s) - x'(s) = \xi(s)$,

$$\sup_{t\in[0,1]} \|\xi(t)\| < \gamma$$

.

$$\begin{aligned}&\int \left|h_n(s)\exp\left(\beta_n' x_0(s)\right) - h_0(s)\exp\left(\beta_0' x_0(s)\right)\right| ds \\ =\ &\int \left| \begin{array}{c} h_n(s)\exp\left(\beta_n' x_0(s)\right) - h_n(s)\exp\left(\beta_n' x_0(s)\right)\exp\left(\beta_n'\xi(s)\right) \\ +h_n(s)\exp\left(\beta_n' x_0(s)\right)\exp\left(\beta_n'\xi(s)\right) - h_0(s)\exp\left(\beta_0' x_0(s)\right)\exp\left(\beta_0'\xi(s)\right) \\ +h_0(s)\exp\left(\beta_0' x_0(s)\right)\exp\left(\beta_0'\xi(s)\right) - h_0(s)\exp\left(\beta_0' x_0(s)\right) \end{array} \right| ds \\ \leq\ &\int \left|h_n(s)\exp\left(\beta_n' x_0(s)\right) - h_n(s)\exp\left(\beta_n' x_0(s)\right)\exp\left(\beta_n'\xi(s)\right)\right| ds \\ &+\int \left|h_n(s)\exp\left(\beta_n' x_0(s)\right)\exp\left(\beta_n'\xi(s)\right) - h_0(s)\exp\left(\beta_0' x_0(s)\right)\exp\left(\beta_0'\xi(s)\right)\right| ds \\ &+\int \left|h_0(s)\exp\left(\beta_0' x_0(s)\right)\exp\left(\beta_0'\xi(s)\right) - h_0(s)\exp\left(\beta_0' x_0(s)\right)\right| ds \\ =\ &\int \left|h_n(s)\exp\left(\beta_n' x_0(s)\right)\right| \left|1-\exp\left(\beta_n'\xi(s)\right)\right| ds + o(1) \\ &+\int \left|h_0(s)\exp\left(\beta_0' x_0(s)\right)\right| \left|\exp\left(\beta_0'\xi(s)\right) - 1\right| ds \\ \leq\ &C^1 \sup_{\beta,s} \left|1-\exp\left(\beta'\xi(s)\right)\right| + o(1) \\ &+C^2 \sup_s \left|\exp\left(\beta_0'\xi(s)\right) - 1\right|.\end{aligned}$$

The first term is bounded by a constant C^1 because $\sup_n \int |h_n(s)|\, ds$ is bounded and β is in a compact interval. Because we may choose $\xi(s)$ such that (14) holds for any $\gamma > 0$, for any $\eta > 0$ we can choose $\xi(s)$ such that these exists an N such that

$$\int \left|h_n(s)\exp\left(\beta_n' x_0\right) - h_0(s)\exp\left(\beta_0' x_0\right)\right| ds < \eta$$

for all $n \geq N$. Therefore, (14) holds for $x_0(s)$ and therefore (14) holds for all $x \in S$.

Because (14) holds for all $x \in S$, then if $\beta_n \to \beta_0$ this implies $h_n(s) \to^{L^1} h_0(s)$. We can see this from the following Taylor expansion and manipulations,

$$\begin{aligned}
& \int \left| h_n(s) \exp\left(\beta_n' x_0(s)\right) - h_0(s) \exp\left(\beta_0' x_0(s)\right) \right| ds \\
= & \int \left| h_n(s) \left[\exp\left(\beta_0' x_0(s)\right) + \sum_{i=1}^{d} x_0^i(s) \exp\left(c' x_0(s)\right) \left(\beta_n^i - \beta_0^i\right) \right] - h_0(s) \exp\left(\beta_0' x_0(s)\right) \right| ds \\
\geq & \int \left| h_n(s) \exp\left(\beta_0' x_0(s)\right) - h_0(s) \exp\left(\beta_0' x_0(s)\right) + h_n(s) \sum_{i=1}^{d} x_0^i(s) \exp\left(c' x_0(s)\right) \left(\beta_n^i - \beta_0^i\right) \right| ds \\
\geq & \int \left| \left| \left[h_n(s) - h_0(s)\right] \exp\left(\beta_0' x_0(s)\right) \right| - \left| h_n(s) \sum_{i=1}^{d} x_0^i(s) \exp\left(c' x_0(s)\right) \left(\beta_n^i - \beta_0^i\right) \right| \right| ds \\
\geq & \left| \begin{array}{c} \int \left| \left[h_n(s) - h_0(s)\right] \exp\left(\beta_0' x_0(s)\right) \right| ds \\ - \int \left| h_n(s) \sum_{i=1}^{d} x_0^i(s) \exp\left(c' x_0(s)\right) \left(\beta_n^i - \beta_0^i\right) \right| ds \end{array} \right| \qquad (15)
\end{aligned}$$

We have proven that (15) converges to zero. If $h_n(s) \nrightarrow^{L^1} h_0(s)$ and $\beta_n \to \beta_0$ we have a contradiction because, as we showed above, $\sup_n \int |h_n(s)| \, ds$ is bounded.

Note that (15) must converge to zero for any $x_0 \in S$ and for $x_0(s) + \epsilon(s) \in S$ where $\epsilon(s)$ perturbs only one covariate as outlined in the assumptions. We can define such a perturbation for each covariate by the theorem assumptions. Above we have proven

$$\int \left| h_n(s) \exp\left(\beta_n' x_0(s)\right) - h_0(s) \exp\left(\beta_0' x_0(s)\right) \right| ds \to 0, \qquad (16)$$

and

$$\int \left| h_n(s) \exp\left(\beta_n' x_0(s)\right) \exp\left(\beta_n' \epsilon(s)\right) - h_0(s) \exp\left(\beta_0' x_0(s)\right) \exp\left(\beta_0' \epsilon(s)\right) \right| ds \to 0. \qquad (17)$$

We now use a Taylor expansion of the term $\exp\left(\beta_n' \epsilon(s)\right)$ around β_0 in (17).

$$\begin{aligned}
& \int \left| \begin{array}{c} h_n(s) \exp\left(\beta_n' x_0(s)\right) \left[\exp\left(\beta_0' \epsilon(s)\right) + \epsilon^i(s) \exp\left(c' \epsilon(s)\right) \left(\beta_n^i - \beta_0^i\right) \right] \\ -h_0(s) \exp\left(\beta_0' x_0(s)\right) \exp\left(\beta_0' \epsilon(s)\right) \end{array} \right| ds \\
= & \int \left| \begin{array}{c} \left[h_n(s) \exp\left(\beta_n' x_0(s)\right) - h_0(s) \exp\left(\beta_0' x_0(s)\right) \right] \exp\left(\beta_0' \epsilon(s)\right) \\ +h_n(s) \exp\left(\beta_n' x_0(s)\right) \epsilon^i(s) \exp\left(c' \epsilon(s)\right) \left(\beta_n^i - \beta_0^i\right) \end{array} \right| ds \\
\geq & \int \left| \begin{array}{c} \left| \left[h_n(s) \exp\left(\beta_n' x_0(s)\right) - h_0(s) \exp\left(\beta_0' x_0(s)\right) \right] \exp\left(\beta_0' \epsilon(s)\right) \right| \\ - \left| h_n(s) \exp\left(\beta_n' x_0(s)\right) \epsilon^i(s) \exp\left(c' \epsilon(s)\right) \left(\beta_n^i - \beta_0^i\right) \right| \end{array} \right| ds \\
\geq & \left| \begin{array}{c} \int \left| \left[h_n(s) \exp\left(\beta_n' x_0(s)\right) - h_0(s) \exp\left(\beta_0' x_0(s)\right) \right] \exp\left(\beta_0' \epsilon(s)\right) \right| ds \\ - \int \left| h_n(s) \exp\left(\beta_n' x_0(s)\right) \epsilon^i(s) \exp\left(c' \epsilon(s)\right) \left(\beta_n^i - \beta_0^i\right) \right| ds \end{array} \right| \qquad (18)
\end{aligned}$$

(16) shows that the first term in (18) converges to zero. Therefore, because $\int |h_n(s)| \, ds \nrightarrow 0$ and by the definition of $\epsilon(s)$, $\beta_n^i \to \beta_0^i$. Because we can define an appropriate $\epsilon(s)$ for each covariate by the assumptions, we have $\beta_n \to \beta_0$. As a result, $h_n(s) \to^{L^1} h_0(s)$. ■

Theorem 9, stated and proven below, implies Theorem 5 presented in Section 3.2. We prove this implication later in this appendix.

Theorem 8 *We make Assumptions (A1)-(A7). Choose a sequence of sieve spaces Θ_n satisfying the following conditions. Let there exist a sequence $h_n \in \Theta_n$ such that $h_n \to^{L^1} h_0$. $h_0 \subset \overline{\mathcal{H}}$ where all functions in $\overline{\mathcal{H}}$ are cáglád. Assume further that for $h \in \Theta_n$*

$$\begin{aligned} C_{\min}^n &\leq h \leq C_{\max}^n \quad (19) \\ \left|\frac{h'}{h}\right| &\leq K_n. \quad (20) \end{aligned}$$

In addition, the constants $C_{\min}^n$, $C_{\max}^n$ and K_n must satisfy the following $\mathbb{P}_{\alpha_0} - a.s.$:

$$K_n \int_0^1 \left| \frac{1}{n} \sum_{i=1}^n \begin{bmatrix} \mathbb{E}\left\{ \int_0^s h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right\} \\ -\int_0^s h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \end{bmatrix} \right| ds \to 0, \quad (21)$$

$$C_{\max}^n \sup_\beta \left(\int_0^1 \left| \frac{1}{n} \sum_{i=1}^n \left\{ \mathbb{E}\left[\exp\left(\beta' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq s\}} \right] - \exp\left(\beta' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq s\}} \right\} \right| ds \right) \to 0 \quad (22)$$

$$|\log(C_{\min}^n)| \vee |\log(C_{\max}^n)| \sup_\beta \left| \begin{array}{c} \frac{1}{n} \sum_{i=1}^n \left(\beta' Z^i(1)\right) \int_0^1 h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \\ -\mathbb{E}\left[\left(\beta' Z^i(1)\right) \int_0^1 h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right] \end{array} \right| \to 0 \quad (23)$$

$$\sup_\beta \left| \frac{1}{n} \sum_{i=1}^n \int_0^1 \left[\beta' Z^i(s)\right] dN_s^i - \mathbb{E}\left(\int_0^1 \left[\beta' Z^i(s)\right] dN_s^i \right) \right| \to 0 \quad (24)$$

Assume $\overline{C}_n = Cn^{-1/4+\eta}$ for a small $\eta > 0$, where C is an arbitrary constant. Assume

$$1 \Big/ \left(|\log(C_{\min}^n)| \vee |\log(C_{\max}^n)| + \sup_{\beta, x} \left|\beta' x\right| \right) = \overline{C}_n, \quad (25)$$

and

$$\frac{1}{K_n} = \overline{C}_n. \quad (26)$$

Then

$$\widehat{\beta} \to \beta_0$$

$$\widehat{h} \to^{L^1} h_0$$

$\mathbb{P}_{\alpha_0} - a.s.$

Proof (Theorem 9). I will show that $H(h_0, \beta_0) - H\left(\widehat{h}, \widehat{\beta}\right) \to 0$ almost surely and the result follows from Lemma 8. Again we follow Karr (1987) Theorem 3.3. Define $\widetilde{\alpha} = \left(\widetilde{h}, \beta_0\right)$ where

$$\int \left|\widetilde{h} - h_0\right| ds \leq \inf_{h \in \Theta_n} \int |h - h_0| \, ds + o(1).$$

The minima need not be unique because we are dealing with L^1.

$$\begin{aligned} H(\alpha_0) - H(\widehat{\alpha}) &= H(\alpha_0) - H(\widetilde{\alpha}) \\ &\quad + H(\widetilde{\alpha}) - Q_n(\widetilde{\alpha}) \\ &\quad + Q_n(\widetilde{\alpha}) - Q_n(\widehat{\alpha}) \\ &\quad + Q_n(\widehat{\alpha}) - H(\widehat{\alpha}) \\ &\leq o(1) \\ &\quad + H(\widetilde{\alpha}) - Q_n(\widetilde{\alpha}) \\ &\quad + o(1) \\ &\quad + Q_n(\widehat{\alpha}) - H(\widehat{\alpha}) \end{aligned} \tag{27}$$

If we show the second and fourth terms in (27) converge to zero a.s., then $H(\alpha_0) - H(\widehat{\alpha}) \to 0$ a.s. and therefore $\widehat{\alpha} \to \alpha_0$. Note that the third line in (27) is $o(1)$ provided the other lines are $o(1)$. This is because $\widehat{\alpha}$ is chosen to maximize $Q_n(\alpha)$ and $0 \leq H(\alpha_0) - H(\widehat{\alpha})$. Consider the fourth term

$$\begin{aligned} & Q_n(\widehat{\alpha}) - H(\widehat{\alpha}) \\ \leq\ & \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \log\left[\widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\right] dN_s^i + \int_0^1 \left[1 - \widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\mathbf{1}_{\{\tau_i \geq s\}}\right] ds \\ & - \mathbb{E}\left[\int_0^1 \log\left[\widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\right] dN_s^i + \int_0^1 \left[1 - \widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\mathbf{1}_{\{\tau_i \geq s\}}\right] ds\right] \\ =\ & \mathbb{E}\left[\int_0^1 \widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\mathbf{1}_{\{\tau_i \geq s\}} ds\right] - \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\mathbf{1}_{\{\tau_i \geq s\}} ds \\ & + \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \log\left[\widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\right] dN_s^i - \mathbb{E}\left[\int_0^1 \log\left[\widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\right] dN_s^i\right]. \end{aligned} \tag{28}$$

Where expectations are taken w.r.t. the true underlying distribution. Using integration by parts and the properties of log, (28) becomes

$$
\begin{aligned}
& (28) \\
= \; & \mathbb{E}\left[\int_0^1 \widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\mathbf{1}_{\{\tau_i\geq s\}}ds\right] - \frac{1}{n}\sum_{i=1}^n \int_0^1 \widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\mathbf{1}_{\{\tau_i\geq s\}}ds \\
& -\frac{1}{n}\sum_{i=1}^n \int_0^1 N^i(s)\frac{\widehat{h}'(s)}{\widehat{h}(s)}ds + \log\left[\widehat{h}(1)\exp\left(\widehat{\beta}' Z^i(1)\right)\right]N^i(1) - 0(a.s.) \\
& -\mathbb{E}\left[-\int_0^1 N^i(s)\frac{\widehat{h}'(s)}{\widehat{h}(s)}ds\right] - \mathbb{E}\left[\log\left[\widehat{h}(1)\exp\left(\widehat{\beta}' Z^i(1)\right)\right]N^i(1)\right] \\
& \frac{1}{n}\sum_{i=1}^n \int_0^1 \left[\beta' Z^i(s)\right]dN_s^i - \mathbb{E}\left[\int_0^1 \left[\beta' Z^i(s)\right]dN_s^i\right] \\
= \; & \mathbb{E}\left[\int_0^1 \widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\mathbf{1}_{\{\tau_i\geq s\}}ds\right] && (29) \\
& -\frac{1}{n}\sum_{i=1}^n \int_0^1 \widehat{h}(s)\exp\left(\widehat{\beta}' Z^i(s)\right)\mathbf{1}_{\{\tau_i\geq s\}}ds \\
& -\frac{1}{n}\sum_{i=1}^n \int_0^1 N^i(s)\frac{\widehat{h}'(s)}{\widehat{h}(s)}ds && (30) \\
& +\frac{1}{n}\sum_{i=1}^n \int_0^1 \left[\frac{\widehat{h}'(s)}{\widehat{h}(s)}\int_0^s h_0(t)\exp\left(\beta_0' Z^i(t)\right)\mathbf{1}_{\{\tau_i\geq t\}}dt\right]ds \\
& +\frac{1}{n}\sum_{i=1}^n \log\left[\widehat{h}(1)\exp\left(\widehat{\beta}' Z^i(1)\right)\right]N^i(1) && (31) \\
& -\mathbb{E}\left[\log\left[\widehat{h}(1)\exp\left(\widehat{\beta}' Z^i(1)\right)\right]N^i(1)\right] \\
& +\mathbb{E}\left[\int_0^1 N^i(s)\frac{\widehat{h}'(s)}{\widehat{h}(s)}ds\right] && (32) \\
& -\frac{1}{n}\sum_{i=1}^n \int_0^1 \left[\frac{\widehat{h}'(s)}{\widehat{h}(s)}\int_0^s h_0(t)\exp\left(\beta_0' Z^i(t)\right)\mathbf{1}_{\{\tau_i\geq t\}}dt\right]ds \\
& \frac{1}{n}\sum_{i=1}^n \int_0^1 \left[\beta' Z^i(s)\right]dN_s^i - \mathbb{E}\left[\int_0^1 \left[\beta' Z^i(s)\right]dN_s^i\right]. && (33)
\end{aligned}
$$

After this expansion, we need to consider the absolute value of (29) + (30) + (31) + (32) + (33). Therefore, we consider the absolute value of (29)-(33) individually as an upper bound. (33) converges to

zero almost surely by assumption (24).

$$\begin{aligned}
|(29)| &= \left| \frac{1}{n} \sum_{i=1}^{n} \left(\int_0^1 \widehat{h}(s) \left\{ \mathbb{E}\left[\exp\left(\widehat{\beta}' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq s\}} \right] - \exp\left(\widehat{\beta}' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq s\}} \right\} ds \right) \right| \\
&\leq \left| \int_0^1 \left(\widehat{h}(s) \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbb{E}\left[\exp\left(\widehat{\beta}' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq s\}} \right] - \exp\left(\widehat{\beta}' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq s\}} \right\} \right) ds \right| \\
&\leq C_{\max}^n \int_0^1 \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbb{E}\left[\exp\left(\widehat{\beta}' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq s\}} \right] - \exp\left(\widehat{\beta}' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq s\}} \right\} \right| ds \\
&\leq C_{\max}^n \sup_{\beta} \left(\int_0^1 \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbb{E}\left[\exp\left(\beta' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq s\}} \right] - \exp\left(\beta' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq s\}} \right\} \right| ds \right).
\end{aligned}$$

$$\begin{aligned}
|(30)| &= \left| \frac{1}{n} \sum_{i=1}^{n} \int_0^1 N^i(s) \frac{\widehat{h}'(s)}{\widehat{h}(s)} ds - \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left[\frac{\widehat{h}'(s)}{\widehat{h}(s)} \int_0^s h_0(t) \exp\left(\beta_0' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right] ds \right| \\
&= \left| \int_0^1 \left[\frac{\widehat{h}'(s)}{\widehat{h}(s)} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(N^i(s) - \int_0^s h_0(t) \exp\left(\beta_0' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right) \right\} \right] ds \right| \\
&\leq K_n \int_0^1 \left| \frac{1}{n} \sum_{i=1}^{n} \left(N^i(s) - \int_0^s h_0(t) \exp\left(\beta_0' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right) \right| ds \\
&\leq K_n \sup_{s \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} \left(N^i(s) - \int_0^s h_0(t) \exp\left(\beta_0' Z^i(s) \right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right) \right|. \qquad (34)
\end{aligned}$$

The term in the supremum of (34) is a martingale by arguments given in Section 2.2. As noted in Karr (1987), this term is subject to the Burkholder inequality and therefore can be bounded in probability. A similar term will arise again, so we wait to give a specific bound.

$$\begin{aligned}
|(31)| &= \left| \frac{1}{n} \sum_{i=1}^{n} \log\left[\widehat{h}(1) \exp\left(\widehat{\beta}' Z^i(1) \right) \right] N^i(1) - \mathbb{E}\left[\log\left[\widehat{h}(1) \exp\left(\widehat{\beta}' Z^i(1) \right) \right] N^i(1) \right] \right| \\
&\leq \left| \frac{1}{n} \sum_{i=1}^{n} \left\{ \begin{array}{c} \log\left[\widehat{h}(1) \exp\left(\widehat{\beta}' Z^i(1) \right) \right] \\ \times \left(N^i(1) - \int_0^1 h_0(t) \exp\left(\beta_0' Z^i(t) \right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right) \end{array} \right\} \right| \qquad (35) \\
&\quad + \left| \begin{array}{c} \frac{1}{n} \sum_{i=1}^{n} \log\left[\widehat{h}(1) \exp\left(\widehat{\beta}' Z^i(1) \right) \right] \int_0^1 h_0(t) \exp\left(\beta_0' Z^i(t) \right) \mathbf{1}_{\{\tau_i \geq t\}} dt \\ -\mathbb{E}\left[\log\left[\widehat{h}(1) \exp\left(\widehat{\beta}' Z^i(1) \right) \right] N^i(1) \right] \end{array} \right|. \qquad (36)
\end{aligned}$$

We handle the terms (35) and (36) separately

$$
\begin{aligned}
|(35)| &= \left|\frac{1}{n}\sum_{i=1}^{n}\log\left[\widehat{h}\left(1\right)\exp\left(\widehat{\beta}'Z^{i}\left(1\right)\right)\right]\left(N^{i}\left(1\right)-\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right)\right| \\
&\leq \left(\left|\log\left(C_{\min}^{n}\right)\right|\vee\left|\log\left(C_{\max}^{n}\right)\right|+\sup_{\beta,z}\left|\beta'z\right|\right) \\
&\quad\times\left|\frac{1}{n}\sum_{i=1}^{n}\left(N^{i}\left(1\right)-\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right)\right| \\
&\leq \left(\left|\log\left(C_{\min}^{n}\right)\right|\vee\left|\log\left(C_{\max}^{n}\right)\right|+\sup_{\beta,z}\left|\beta'z\right|\right) \qquad (37) \\
&\quad\times\sup_{s}\left|\frac{1}{n}\sum_{i=1}^{n}\left(N^{i}\left(s\right)-\int_{0}^{s}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right)\right|.
\end{aligned}
$$

$$
\begin{aligned}
|(36)| &= \left|\begin{array}{c}\frac{1}{n}\sum_{i=1}^{n}\log\left[\widehat{h}\left(1\right)\exp\left(\widehat{\beta}'Z^{i}\left(1\right)\right)\right]\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt \\ -\mathbb{E}\left[\log\left[\widehat{h}\left(1\right)\exp\left(\widehat{\beta}'Z^{i}\left(1\right)\right)\right]N^{i}\left(1\right)\right]\end{array}\right| \\
&= \left|\begin{array}{c}\frac{1}{n}\sum_{i=1}^{n}\log\left[\widehat{h}\left(1\right)\exp\left(\widehat{\beta}'Z^{i}\left(1\right)\right)\right]\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt \\ -\mathbb{E}\left[\log\left[\widehat{h}\left(1\right)\exp\left(\widehat{\beta}'Z^{i}\left(1\right)\right)\right]\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right]\end{array}\right| \\
&\leq \left(\left|\log\left(C_{\min}^{n}\right)\right|\vee\left|\log\left(C_{\max}^{n}\right)\right|\right) \\
&\quad\times\left|\begin{array}{c}\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\beta}'Z^{i}\left(1\right)\right)\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt \\ -\mathbb{E}\left[\left(\widehat{\beta}'Z^{i}\left(1\right)\right)\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right]\end{array}\right| \\
&\leq \left(\left|\log\left(C_{\min}^{n}\right)\right|\vee\left|\log\left(C_{\max}^{n}\right)\right|\right) \qquad (38) \\
&\quad\times\sup_{\beta}\left|\begin{array}{c}\frac{1}{n}\sum_{i=1}^{n}\left(\beta'Z^{i}\left(1\right)\right)\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt \\ -\mathbb{E}\left[\left(\beta'Z^{i}\left(1\right)\right)\int_{0}^{1}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right]\end{array}\right|.
\end{aligned}
$$

$$
\begin{aligned}
|(32)| &= \left|\mathbb{E}\left[\int_{0}^{1}N^{i}\left(s\right)\frac{\widehat{h}'\left(s\right)}{\widehat{h}\left(s\right)}ds\right]-\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{1}\left[\frac{\widehat{h}'\left(s\right)}{\widehat{h}\left(s\right)}\int_{0}^{s}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right]ds\right| \\
&= \left|\begin{array}{c}\mathbb{E}\left[\int_{0}^{1}\left[\frac{\widehat{h}'(s)}{\widehat{h}(s)}\int_{0}^{s}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right]ds\right] \\ -\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{1}\left[\frac{\widehat{h}'(s)}{\widehat{h}(s)}\int_{0}^{s}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right]ds\end{array}\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{1}\frac{\widehat{h}'\left(s\right)}{\widehat{h}\left(s\right)}\left[\begin{array}{c}\mathbb{E}\left\{\int_{0}^{s}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right\} \\ -\int_{0}^{s}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\end{array}\right]ds\right| \\
&\leq K_{n}\int_{0}^{1}\left|\frac{1}{n}\sum_{i=1}^{n}\left[\begin{array}{c}\mathbb{E}\left\{\int_{0}^{s}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\right\} \\ -\int_{0}^{s}h_{0}\left(t\right)\exp\left(\beta_{0}'Z^{i}\left(t\right)\right)\mathbf{1}_{\{\tau_{i}\geq t\}}dt\end{array}\right]\right|ds. \qquad (39)
\end{aligned}
$$

As outlined in Section 2, the terms

$$M_t^n = \sum_{i=1}^{n} \left(N^i(s) - \int_0^s h_0(t) \exp\left(\beta_0' Z^i(t)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right) \tag{40}$$

are martingales. Therefore, the Burkholder inequality can be used to bound the supremum of (40). See Karr (1987) for the relevant statement of Burkholder's inequality and Dellacherie and Meyer (1980) for a more general statement. For any $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}\left\{ \frac{1}{n\overline{C}_n} \sup |M_t^n| > \epsilon \right\} &\leq \frac{1}{\left(n\overline{C}_n \epsilon\right)^4} \mathbb{E}\left\{ \sup |M_t^n|^4 \right\} \\ &\leq \frac{1}{\left(n\overline{C}_n \epsilon\right)^4} C_2 \mathbb{E}\left\{ [M^n]_1^2 \right\} \end{aligned} \tag{41}$$

The first inequality of (41) comes from Markov's inequality, the second from Burkholder's inequality. The quadratic covariation map $(A, B) \to [A, B]$ is bilinear (Protter (2005) pg. 66). Therefore

$$[M^n]_t = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \left[M^i, M^j\right]_t \right)$$

The semimartingales M^i are "quadratic pure jump" (Protter (2005) pg. 70-71). By Protter (2005) Theorem 28 pg. 75, this implies $\left[M^i, M^j\right]_1 = 0$ a.s. because jumps in both processes only happen at the same time with probability zero. We note that $\left[M^i\right]_1 = N_1^i$ (Protter (2005) pg. 70). Therefore,

$$[M^n]_1 = \sum_{i=1}^{n} N_1^i.$$

Because $0 \leq N_1^i \leq 1$,

$$\begin{aligned} \mathbb{E}\left\{ [M^n]_1^2 \right\} &= \sum_{i=1}^{n} \mathbb{E}\left\{ \left(N_1^i\right)^2 \right\} + 2 \sum_{i \neq j} \mathbb{E}\left\{ N_1^i N_1^j \right\} \\ &\leq n + 2n(n-1) \end{aligned}$$

This implies

$$\frac{1}{\left(n\overline{C}_n \epsilon\right)^4} C_2 \mathbb{E}\left\{ [M^n]_1^2 \right\} = O\left(n^2 \overline{C}_n^4 \epsilon^4 \right)$$

Therefore, we need $n\overline{C}_n^2 \to \infty$ and $\overline{C}_n = Cn^{-1/4+\eta}$. Above, $\overline{C}_n$ takes the following two values:

$$\begin{aligned} \overline{C}_n &= 1 \Big/ \left(\left|\log\left(C_{\min}^n\right)\right| \vee \left|\log\left(C_{\max}^n\right)\right| + \sup_{\beta, x} \left|\beta' x\right| \right), \\ \overline{C}_n &= \frac{1}{K_n}. \end{aligned}$$

As in Karr (1987), a Borel-Cantelli argument gives

$$\frac{1}{n\overline{C}_n}\sup|M_t^n| \to 0, \qquad \mathbb{P}_{\alpha_0} - a.s.$$

Therefore, by the assumptions of the theorem, $|Q_n(\widehat{\alpha}) - H(\widehat{\alpha})| \to 0$. Notice that throughout the proof the exact value of $\widehat{\alpha}$ was irrelevant and the results hold for an arbitrary sequence $\alpha_n \in \Theta_n$ under the assumptions on Θ_n. Therefore, $|H(\widetilde{\alpha}) - Q_n(\widetilde{\alpha})| \to 0$ and by (27), $H(\alpha_0) - H(\widehat{\alpha}) \to 0$. By Lemma 8,

$$\widehat{\beta} \to \beta_0$$

$$\widehat{h} \to^{L^1} h_0$$

$\mathbb{P}_{\alpha_0} - a.s.$ ■

The bound facilitated by the Burkholder inequality in the previous proof is crude. In one instance, we bound a martingale at time $t = 1$ by its supremum over the interval $t \in [0, 1]$, then apply the Burkholder inequality. In another, we bound the integral of a martingale over $t \in [0, 1]$ by its supremum over the same interval, and again apply the Burkholder inequality. These bounds can likely be improved upon. This could improve the choice of the sequences K_n, $C_{\min}^n$ and $C_{\max}^n$.

The next lemma gives a preliminary result to establishing mixing conditions. The proof is obvious and omitted.

Lemma 9 *Let $f(x, y, t)$ be a bounded continuous function on $\mathbb{R}^{d+j+1}$. Define $W^1 = \int_0^T f(X_s, Y_{G^i+s}, s)\, ds$ and $W^2 = \int_0^T f(X_s, Y_{G^i+s}, s)\, \mathbf{1}_{\{\tau_i \geq s\}} ds$. Then*

$$\sigma\left(W^1\right), \sigma\left(W^2\right) \subset \sigma\left\{\eta_i, X^i(s), Y\left(G^i + s\right) \middle| \, 0 \leq s \leq T\right\}.$$

Lemma 10 allows us to convert mixing conditions on the underlying covariate processes into mixing conditions on the integrals we use for estimation. This will facilitate verification of various conditions in the proof of Theorem 5.

Proof (Theorem 5). Fixed bounds $C_{\min}$ and $C_{\max}$ on the underlying function h_0 simplifies the conditions (19)-(26) in Theorem 9. The α-mixing assumption implies that the simpler versions of (19)-(26) are satisfied and the result follows. Recall $\epsilon > 0$ and ϵ small. Define

$$p = 1/\left(\frac{3}{4} + \epsilon\right)$$

For each fixed s, let

$$B_s^{i,1} = \int_0^s h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt$$

satisfy the mixing condition

$$\sum_{n>0} n^{p-2} \alpha(n) < \infty.$$

For all fixed $\beta \in [a_1, b_1] \times \cdots \times [a_q, b_q]$ and all fixed s, let the variables

$$B_s^{i,2} = \exp\left(\beta' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq s\}}$$

$$B^{i,3} = \left(\beta' Z^i(1)\right) \int_0^1 h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt$$

$$B^{i,4} = \int_0^1 \left[\beta' Z^i(s)\right] dN_s^i$$

satisfy the mixing condition

$$\sum_{n>0} n^{-1} \alpha(n) < \infty. \tag{42}$$

Because of Lemma 10 and the assumed mixing conditions on the underlying covariates, these α-mixing conditions are true. As in Karr (1987), we apply a Marcinkiewicz-Zygmund type strong law of large numbers to show (21)-(24) are satisfied. The Marcinkiewicz-Zygmund strong law we use comes from Rio (1995). This strong law allows for dependence between and within covariate processes in our application. Note that the random variables $B_s^{i,1}$, $B_s^{i,2}$, $B^{i,3}$ and $B^{i,4}$ are bounded. This simplifies the results in Rio (1995) as discussed in that paper.

As a result of the specified strong law, there exists a set Ω_0 of probability 1 such that, for a countable dense set $\overline{S} \subset [0,1]$ if $\overline{s} \in \overline{S}$ the following holds for all $\omega \in \Omega_0$:

$$K_n \left| \frac{1}{n} \sum_{i=1}^n \left[\mathbb{E}\left\{ \int_0^{\overline{s}} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right\} - \int_0^{\overline{s}} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right] \right| \to 0. \tag{43}$$

Recall we are assuming $K_n = O\left(n^{1/4-\eta}\right)$. Let $\widetilde{s} \notin \overline{S}$ and $\widetilde{s} + \delta \in \overline{S}$ with δ arbitrarily small. This can always be done because $\overline{S}$ is dense in $[0,1]$.

$$\begin{aligned}
& K_n \left| \frac{1}{n} \sum_{i=1}^n \left[\mathbb{E}\left\{ \int_0^{\widetilde{s}} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right\} - \int_0^{\widetilde{s}} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right] \right| \\
\leq\ & K_n \left| \frac{1}{n} \sum_{i=1}^n \left[\mathbb{E}\left\{ \int_0^{\widetilde{s}} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right\} - \mathbb{E}\left\{ \int_0^{\widetilde{s}+\delta} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right\} \right] \right| \\
& + K_n \left| \frac{1}{n} \sum_{i=1}^n \left[\mathbb{E}\left\{ \int_0^{\widetilde{s}+\delta} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right\} - \int_0^{\widetilde{s}+\delta} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right] \right| \\
& + K_n \left| \frac{1}{n} \sum_{i=1}^n \left[\int_0^{\widetilde{s}+\delta} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt - \int_0^{\widetilde{s}} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right] \right| \\
\leq\ & K_n C^1(\delta) \\
& + \sup_{s \in \overline{S}} \left| K_n \frac{1}{n} \sum_{i=1}^n \left[\mathbb{E}\left\{ \int_0^{s} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right\} - \int_0^{s} h_0(t) \exp\left(\beta_0' Z^i(s)\right) \mathbf{1}_{\{\tau_i \geq t\}} dt \right] \right| \qquad (44) \\
& + K_n C^2(\delta). \\
=\ & K_n C^1(\delta) + o_{a.s.}(1) + K_n C^2(\delta). \qquad (45)
\end{aligned}$$

The supremum term in (44) is $o_{a.s.}(1)$ by a standard bracketing argument using Theorem II.2.2 in Pollard (1984). This theorem's simple proof can be easily modified to incorporate our Marcinkiewicz-Zygmund SLLN. The details are omitted. A sequence δ can be chosen to converge to 0 fast enough so $K_n C^1(\delta) = o(1)$ and $K_n C^2(\delta) = o(1)$. This implies (43) holds for all $s \in [0,1]$. Therefore, condition (21) from Theorem 9 holds. Along with the assumptions in Theorem 5, a simple bracketing argument implies

conditions (22), (23) and (24) in Theorem 9, where these conditions are simplified by the boundedness assumption on h_0. Again, see Pollard (1984) Section II.2 for the required bracketing results. This needs to be coupled with an argument similar to that given above for condition (21). The details are again omitted for brevity. ■

The same decomposition into terms as done in the proof of Theorem 9 can be done with the block sampling case outlined in Section 3. The difference now is we additionally have to sum over $k(n)$. We still must show the terms of the decomposition converge to zero almost surely. The terms handled by the Burkholder inequality can still be handled the same way as the martingale structure is preserved. Provided $K_n \to \infty$ and $K_n = o\left(\left(n/k(n)\right)^{1/2}\right)$, the martingale terms converge to zero almost surely with the same proof. The remaining terms that need to be handled are:

$$K_n \int_0^1 \left| \sum_{j=1}^{k(n)} \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} \mathbb{E}\left\{ \int_0^s h_0^j(t) \exp\left(\beta_0' Z^{ji}(s)\right) \mathbf{1}_{\{\tau_{ji} \geq t\}} dt \right\} \\ - \int_0^s h_0^j(t) \exp\left(\beta_0' Z^{ji}(s)\right) \mathbf{1}_{\{\tau_{ji} \geq t\}} dt \end{bmatrix} \right| ds \to 0 \quad a.s., \tag{46}$$

$$\sup_{\beta} \left(\int_0^1 \left| \sum_{j=1}^{k(n)} \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathbb{E}\left[\exp\left(\beta' Z^{ji}(s)\right) \mathbf{1}_{\{\tau_{ji} \geq s\}} \right] - \exp\left(\beta' Z^{ji}(s)\right) \mathbf{1}_{\{\tau_{ji} \geq s\}} \right\} \right| ds \right) \to 0 \quad a.s., \tag{47}$$

$$\sup_{\beta} \left| \sum_{j=1}^{k(n)} \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} \left(\beta' Z^{ji}(1)\right) \int_0^1 h_0^j(t) \exp\left(\beta_0' Z^{ji}(s)\right) \mathbf{1}_{\{\tau_{ji} \geq t\}} dt \\ -\mathbb{E}\left\{ \left(\beta' Z^{ji}(1)\right) \int_0^1 h_0(t) \exp\left(\beta_0' Z^{ji}(s)\right) \mathbf{1}_{\{\tau_{ji} \geq t\}} dt \right\} \end{bmatrix} \right| \to 0 \quad a.s. \tag{48}$$

$$\sup_{\beta} \left| \sum_{j=1}^{k(n)} \frac{1}{n} \sum_{i=1}^{n} \left[\int_0^1 \left[\beta' Z^{ji}(s)\right] dN_s^{ji} - \mathbb{E}\left[\int_0^1 \left[\beta' Z^{ji}(s)\right] dN_s^{ji} \right] \right] \right| \to 0 \quad a.s. \tag{49}$$

Proof (Theorem 6). The result follows by a slight modification of the proofs of Theorem 9 and Theorem 5. It can be shown that

$$\sum_{j=1}^{k(n)} \left\{ H^j\left(h_0^j, \beta_0\right) - H^j\left(\widehat{h}^j, \widehat{\beta}\right) \right\} \to 0, \tag{50}$$

a.s. and the result follows. The main difference between the proofs is that in (50) there is a sum over the number of blocks $k(n)$. Note that the Burkholder inequality applies to martingales defined on $[0, \infty]$ as in this extension. See Dellacherie and Meyer (1980). ■

References

[1] Aalen, O.O., Ø. Borgan & H.K. Gjessing (2010) *Survival and Event History Analysis.* Springer.

[2] Andersen, P.K., Ø. Borgan, R.D. Gill & N. Keiding (1994) *Statistical Models Based on Counting Processes.* Springer.

[3] Andersen, P.K. & R.D. Gill (1982) Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics* 10 (4), 1100-1120.

[4] Azizpour, S., K. Giesecke & B. Kim (2011) Premia for Correlated Default Risk. *Journal of Economics Dynamics and Control* 35, 1340-1357.

[5] Azizpour, S., K. Giesecke & G. Schwenkler (2011) Exploring the Sources of Default Clustering. Unpublished.

[6] Banerjee, S. & D. K. Dey (2005) Semiparametric Proportional Odds Models for Spatially Correlated Survival Data. *Lifetime Data Analysis* 11, 175-191.

[7] Bastos, L.S. & D. Gamerman (2006) Dynamic Survival Models with Spatial Frailty. *Lifetime Data Analysis* 12, 441-460.

[8] Bielecki, T.R. & M. Rutkowski (2004) *Credit Risk: Modeling, Valuation and Hedging.* Springer.

[9] Billingsley, P. (1999) *Convergence of Probability Measures, 2nd ed.* Wiley.

[10] Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models, Chapter 76 in Handbook of Econometrics, Vol. 6B, 2007, eds. J.J. Heckman and E.E. Leamer, North-Holland.

[11] Chen, X. (2011) Penalized Sieve Estimation and Inference of Semi-Nonparametric Dynamic Models: A Selective Review. Cowles Foundation working paper.

[12] Chen, X., Z. Liao & Y. Sun (2011) Sieve Inference for Weakly Dependent Data. Cowles Foundation working paper.

[13] Chui, C.K. (1992) *An Introduction to Wavelets.* Academic Press.

[14] Creal, D., B. Schwaab, S.J. Koopman & A. Lucas (2011) Observation Driven Mixed-Measurement Dynamic Factor Models with an Application to Credit Risk. Working paper.

[15] Davidson, J. (1994) *Stochastic Limit Theory.* Oxford University Press.

[16] Das, S.R., D. Duffie, N. Kapadia & L. Saita (2007) Common Failings: How Corporate Defaults are Correlated. *Journal of Finance* 62 (1), 93-117.

[17] de Boor, C. (2001) *A Practical Guide to Splines.* Springer.

[18] Dellacherie, C. & P.A. Meyer (1980) *Probabilities and Potential B.* North Holland.

[19] Duffie, D., A. Eckner, G. Horel & L. Saita (2009) Frailty Correlated Default. *Journal of Finance* 64 (5), 2089-2123.

[20] Duffie, D., L. Saita & K. Wang (2007) Multi-Period Corporate Default Prediction With Stochastic Covariates. *Journal of Financial Economics* 83, 635-665.

[21] Ethier, S.N. & T.G. Kurtz (1986) *Markov Processes: Characterization and Convergence.* Wiley.

[22] Fleming, T.R. & D.P. Harrington (1991) *Counting Processes and Survival Analysis*, Wiley.

[23] Giesecke, K. & Kim (2011a) Risk Analysis of Collateralized Debt Obligations. *Operations Research* 59 (1), 32-49.

[24] Giesecke, K. & Kim (2011b) Systemic Risk: What Defaults are Telling Us. *Management Science* 57 (8), 1387-1405.

[25] Giesecke, K. & Schwenkler (2012) Filtered Likelihood for Point Processes. Working paper.

[26] Grenander, U. (1981) *Abstract Inference*. Wiley.

[27] Hougaard, P. (2000) *Analysis of Multivariate Survival Data*. Springer.

[28] Jacod, J. & A.N. Shiryaev (2003) *Limit Theorems for Stochastic Processes, 2nd ed.* Springer.

[29] Karr, A.F. (1987) Multiplicitive Likelihood Estimation in the Multiplicative Intensity Model Via Sieves. *Annals of Statistics* 15 (2), 473-490.

[30] Koopman, S.J., A. Lucas & B. Schwaab (2011) Modeling Frailty-Correlated Defaults Using Many Macroeconomic Covariates. *Journal of Econometrics* 162, 312-325.

[31] Lando, D. & M.S. Nielsen (2010) Correlation in corporate defaults: Contagion or Conditional Independence? *Journal of Financial Intermediation* 19 (3), 355-372.

[32] Li, Y. & X. Lin (2005) Semiparametric Normal Transformation Models for Spatially Correlated Survival Data. Harvard Biostatistics Working Paper.

[33] Li, Y. & L. Ryan (2002) Modeling Spatial Survival Data Using Semiparametric Frailty Models. *Biometrics* 58, 287-297.

[34] Liebscher, E. (1996) Strong Convergence of Sums of α-Mixing Random Variables with Applications to Density Estimation. *Stochastic Processes and their Applications* 65, 69-80.

[35] Martinussen, T. & Scheike T.H. (2010) *Dynamic Regression Models for Survival Data*. Springer.

[36] Mayer, C., K. Pence & S.M. Sherlund (2009) The Rise of Mortgage Defaults. *Journal of Economic Prospectives* 23 (1), 27-50.

[37] Newey, W.K. & D. McFadden (1994) Large Sample Estimation and Hypothesis Testing, Chapter 36 in Handbook of Econometrics, Vol. 4, 1994, eds. R.F. Engle and D.L. McFadden, Elsevier.

[38] Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer-Verlag.

[39] Protter, P.E. (2005) *Stochastic Integration and Differential Equations*. Springer.

[40] Rio, (1995) A Maximal Inequality and Dependent Marcinkiewicz-Zygmund Strong Laws. *Annals of Probability* 23 (2), 918-937.

[41] Van den Berg, G.J. (2001) Duration Models: Specification, Identification and Multiple Durations, chapter 55 in Handbook of Econometrics, Vol. 5, 2001, eds. James J. Heckman and Edward E. Leamer, Elsevier Science.

[42] Van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge.

[43] Wolter, J.L. (2012a) Essays on the Econometrics of Financial Crisis Dynamics. Ph.D dissertation, Dept. Economics, Yale University.

[44] Wolter, J.L. (2012b) Kernel Estimation of Hazard Functions When Observations Have Dependent and Common Covariates. Working Paper, Univ. of Oxford.