ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

Kernel Estimation Of Hazard Functions When Observations Have Dependent and Common Covariates

James Lewis Wolter

Number 761
September 2015

Manor Road Building, Oxford OX1 3UQ

Kernel Estimation Of Hazard Functions When Observations Have Dependent and Common Covariates

James Lewis Wolter*

Department of Economics, University of Oxford

Oxford-Man Institute

September 24, 2015

Abstract

We propose a hazard model where dependence between events is achieved by assuming dependence between covariates. This model allows for correlated variables specific to observations as well as macro variables which all observations share. This setup better fits many economic and financial applications where events are not independent. Nonparametric estimation of the hazard function is then studied. Kernel estimators proposed in Nielsen and Linton (1995, *Annals of Statistics*) and Linton, Nielsen and Van de Geer (2003, *Annals of Statistics*) are shown to have similar asymptotic properties compared with the *i.i.d.* case. Mixing conditions ensure the asymptotic results follow. These results depend on adjustments to bandwidth conditions. Simulations are conducted which verify the impact of dependence on estimators. Bandwidth selection accounting for dependence is shown to improve performance. In an empirical application, trade intensity in high-frequency financial data is estimated.

*Correspondence to: Eagle House, Eagle Works, Walton Well Road, OX2 6ED Oxford, United Kingdom. Tel.: +44 1865 616632; E-mail address: james.wolter@oxford-man.ox.ac.uk. JLE Classification: C13, C14, C51. Keywords: hazard estimation, correlated events, dependent covariates, common covariates, kernel estimation.

1 Introduction

Almost all research on hazard models has focused on parametric and semiparametric cases assuming *i.i.d.* observations. In econometrics, it was recognized early on that unobserved heterogeneity was an important aspect of these specifications. This lead to the mixed proportional hazard model. Lancaster (1990) and van den Berg (2001) summarize much of the early work in this area. Hausman and Woutersen (2014a,b) are recent contributions giving up-to-date summaries of the literature.[1]

While not given as much attention, the nonparametric case has been studied in a parallel literature. Nonparametric estimation of hazard models with observed covariates is well developed in the *i.i.d.* case. In this research, kernel methods have been a frequently used approach. Much of this line of inquiry culminated in the papers of Nielsen and Linton (1995) and Linton, Nielsen and Van de Geer (2003) (hereafter NL and LNV). NL derived asymptotics for fully nonparametric estimators allowing exposure-time to be a covariate. In LNV, the hazard function is restricted to be additive or multiplicative, but otherwise allowed to be nonparametric. Estimators are proposed and a number of asymptotic results are derived. In particular they show that, by restricting the form of the hazard function, the rate of convergence is greatly improved. For a comprehensive overview of previous work in this area see the aforementioned papers.

In a number of economic and financial hazard situations, the assumption that the observed covariates are independent across observations is questionable or clearly false. For example, when considering mortgage default, variables such as interest rates or housing values are correlated across mortgage holders. For corporate default applications, accounting variables are correlated across members of the same industry. In addition, macroeconomic variables such as US Treasury rates or GDP growth rates can be relevant. These covariates have the same realization across observations. Their only variability is through time. If relevant covariates include macro variables, observations cannot satisfy the *i.i.d.* assumption required in the previous literature.

The most common way previous studies have allowed for dependence across observations in hazard models is with a clustering approach. Events are allowed to be correlated within groups of

[1] Papers include Elbers and Ridder (1982); Heckman and Singer (1984); Han (1987); Han and Hausman (1990); Honoré (1990), (1993a,b); Meyer (1990); Ridder (1990); Hahn (1994); Ishwaran (1996); Horowitz (1999); Woutersen (2000), (2002); Bijwaard and Ridder (2002), (2009); Chen (2002); Ridder and Woutersen (2003); Horowitz and Lee (2004); Frederiksen, Honoré and Hu (2007) and Honoré and Hu (2010).

observations, but the groups are assumed to be independent. See Martinussen and Scheike (2010) for an overview. Macro variables eliminate the possibility of a clustering approach. This is because they make it impossible to separate observations into independent groups. Relatively few hazard papers have considered situations with more extensive correlation than the clustering case. The only examples the author is aware of involve spatial correlation. See Henderson, Shimakura and Gorst (2002); Li and Ryan (2002); Banerjee, Wall and Carlin (2003); Banerjee and Dey (2005); Li and Lin (2005); Hennerfeind, Brezger and Fahrmeir (2006); Bastos and Gamerman (2006); Zhao, Hanson and Carlin (2009); Zhao and Hanson (2011) and Lawson, Choi and Zhang (2014) among others. Most of these citations introduce dependence through spatially correlated unobserved heterogeneity[2]. Observed variables are assumed to be independent across observations. Li and Lin (2005) is an exception. All of these papers estimate parametric or semiparametric models.

The first contribution of this paper is to propose a specific construction of correlated random events. The proposed construction makes clear how dependence in events is related to observed covariates. This connection allows application of methods used in the *i.i.d.* case to the dependent case. In the proposed construction, dependence is assumed directly on the covariates. This in turn determines dependence between random events. The model is not spatial. Indeed, one of the main goals of this research is to incorporate macro covariates as described above. Dependence through time is an important aspect of our model not captured in spatial situations.

Compared with the *i.i.d.* case, this setup better fits many economic situations. The model is particularly well suited for analysis of credit events when observations overlap with financial crises. During these episodes, many important macroeconomic and financial variables persistently take on values rarely seen in tranquil periods. This implies these covariates are dependent through time as assumed in the proposed model. These variables are frequently used as measures of macroeconomic conditions by decision making agents. This requires they are accounted for in hazard analysis. Covariates with these properties include yield curve variables, VIX, the TED spread, Moody's corporate default spread and aggregate consumption, investment, income or GDP growth rates. In addition, empirical analysis often uses variables which are cross-sectionally correlated. This is another aspect of our model. Relevant credit situations include mortgage default, corporate

[2] Unobserved heterogeneity is referred to as frailty in many of these papers.

default, bank lending and corporate exercise of credit lines.

Using the proposed model, asymptotic results for estimators considered in NL and LNV are derived. Dependence between random events is controlled using mixing and martingale conditions. This control is then utilized to derive asymptotics. Most of the results from NL and LNV are extended to the dependence case. Compared with the *i.i.d.* situation, the rate at which the bandwidth converges to zero must be slowed down for similar results to hold. This adjustment changes confidence intervals, uniform convergence rates and asymptotic variance estimates.[3]

With the proposed setup, many previous papers can be extended to the dependence case. In this paper we trace the consequences of dependence for kernel approaches. A similar exercise could be undertaken with much of the hazard literature. This work is also related to a large body of research on panel data with dependence. This dependence can manifest itself cross-sectionally, through time or spatially. The literature is too large to comprehensively overview here. We only mention a few areas related to this paper. The first area is panel data models with common factors. Recent citations include Pesaran (2006); Kapetanios and Pesaran (2007); Chudik, Pesaran and Tosetti (2011); Pesaran and Tosetti (2011); Su and Jin (2012) and Jin and Su (2013). See Chudik and Pesaran (2013) for a survey. A second area is panel data models with spatial dependence. Recent work includes Kapoor, Kelejian and Prucha (2007); Yu, de Jong and Lee (2008); Lee and Yu (2010a,b), (2014); Su and Jin (2010); Parent and LeSage (2012); Su (2012); Baltagi, Egger and Pfaffermayr (2013); Baltagi, Fingleton and Pirotte (2014); Su and Yang (2015). A third area is panel data models with common shocks as in Kuersteiner and Prucha (2013) who build on the results of Andrews (2005). A final area is dependent discrete choice. See Robinson (1982); Poirier and Ruud (1988); de Jong and Woutersen (2011) and Hahn and Kuersteiner (2011). Many of these papers assume more complicated forms of dependence than the present work. Ideas from this literature can likely be used to extend our hazard results in various directions. We leave these extensions to future research.

Another application of our results is estimation of trade intensity in high-frequency financial

[3] This paper considers the local constant estimation case. There is a related literature for local linear estimation. See Nielsen (1998); Nielsen and Tanggaard (2001); Mammen, Martínez-Miranda, Nielsen and Sperlich (2011) and Gámiz, Janys, Martínez-Miranda and Nielsen (2013). Similarly to NL and LNV, the local linear approach has not been modified to account for dependence. This line of research would be an extension of the results that are presented below.

data. Several papers have shown that an asset's quotes can partially explain trading rates (see, for example, Hall and Hautsch (2007)). It is also possible information from related assets influences this rate. A commonly used variable is depth imbalance. In an empirical application of our results, the hazard rate of trade arrival for silver futures is estimated. The depth imbalance of both gold and silver futures are used as covariates. We show both these covariates have substantial autocorrelation. Therefore, this application fits into the framework described in this paper. It is not obvious a priori which shape the hazard function will take. This makes kernel estimation an attractive approach.

The remainder of the paper is organized as follows. In Section 2, we propose the construction of random events that is used throughout the paper. In Section 3, nonparametric asymptotic results related to NL are derived assuming covariates are dependent. Section 3.1 presents results related to LNV where the functional form of the hazard is restricted to be additive or multiplicative. Section 4 conducts a number of simulations which examine the effect of dependence on finite sample properties of the relevant estimators. Section 5 presents our application to trade intensity. Section 6 concludes. All proofs are presented in the Appendix or Online Supplement.

2 Model of Random Events

This section describes the random events used throughout the sequel. Special attention is given to incorporating macro covariates across observations. First, the covariate processes assigned to observations are presented. Then these processes are used to construct random times. The hazard function arises naturally in the setup. The construction is presented without censoring. General forms of censoring can be added with few complications.[4] The notation of NL and LNV is used where possible.

2.1 Sampling

Observations are indexed by $i \in \mathbb{N}$. Each observation i has a deterministic constant $G^i \in [0, \infty)$ which is the calendar-time it starts to be at risk of default. These represent, for example, the

[4] See the Online Supplement for more details.

date a mortgage is signed. Each observation is at risk over a fixed time of length T. T is fixed throughout the paper and does not change in the asymptotics. An observation's period at risk corresponds to the calendar-time interval $[G^i, G^i + T]$. These intervals are allowed to overlap.

Let $\{X^i(t)| t \in [0,T]\}$ be d covariate processes specific to each observation.[5] X_t^i for $t \in [0,T]$ corresponds to the value of these covariates over the calendar-time interval $[G^i, G^i + T]$. Defining X_t^i on $[0,T]$ instead of $[G^i, G^i + T]$ is done for notational simplicity. We make the additional assumption that X_t^i has paths which are left-continuous with right-hand-limits (càglàd). The càglàd assumption implies that the processes X_t^i are predictable, an important technical property for our results.[6] The distribution of the variables X_t^i has support equal to the compact rectangle $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_d$ for each $t \in [0,T]$. The exposure-time t will also be a covariate in the sequel. The exposure-time is how long a current duration has been active (e.g., how long since a mortgage has been signed).

In addition, there is a set of j covariate processes W_t. These covariates represent macro variables such as the GDP growth rate. W_t is assumed to have càglàd paths and compact rectangular support $\mathcal{W} = \mathcal{W}_1 \times \cdots \times \mathcal{W}_j$ for each $t \in [0,\infty)$. The time index for these covariates corresponds to calendar-time $[0,\infty)$. For simplicity, the support of all covariate processes can be thought of as $[0,T] \times [0,1]^{d+j}$. W_t are common in that, for each i, the portion of W_t corresponding to $[G^i, G^i + T]$ affects the hazard rate for that observation.

To sum up, the relevant covariates for observation i are

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

In order for asymptotic results to hold, some regularity on the starting times G^i is needed. We make the following assumption.

(G) $G^i \leq G^{i+1}$ and $G^i \to \infty$ as $i \to \infty$.

Under condition (G), observations are ordered by their starting times G^i. For example, if

[5] It is standard notation in the stochastic processes literature to write continuous-time processes $X^i(t)$ as X_t^i. In what follows we freely change between this equivalent notation when using any continuous-time stochastic process.

[6] See Jacod and Shiryaev (2003, p.16) for a definition of predictability.

observations represent mortgages, mortgages signed at earlier times are ordered before mortgages signed later. Sampling increases as time goes to infinity. As more data is gathered, mortgages with increasingly large G^i are observed. That $G^i \to \infty$ as $i \to \infty$ is what gives us effective sampling of the common processes. Asymptotic approximations in this setup follow because, as $i \to \infty$, observations are influenced by portions of the common covariates W_t which start at arbitrarily large calendar times. Notice we are not assuming two dimensional sampling as in panel data models. The asymptotics do not allow an increasingly large number of observations at any given calendar time. Condition (G) rules this out. Examples which satisfy (G) are given in the sequel.

2.2 Construction of Random Times

Our construction of random times follows Bielecki and Rutkowski's (2004) (hereafter BR) Example 9.1.5. See BR for a comprehensive overview of random times and their martingale properties. In what follows, $\psi(\cdot)$ is the hazard function with covariates taken as arguments. We make the following assumption:

(B1) $\psi : [0,T] \times \mathcal{X} \times \mathcal{W} \to \mathbb{R}$ is a continuous function such that,

$$\begin{aligned} \inf_{(t,x,w)\in[0,T]\times\mathcal{X}\times\mathcal{W}} \psi(t,x,w) &= \underline{C} > 0, \\ \sup_{(t,x,w)\in[0,T]\times\mathcal{X}\times\mathcal{W}} \psi(t,x,w) &= \overline{C} < \infty. \end{aligned}$$

We assume (B1) throughout the paper. Random times τ_i are defined as

$$\Gamma_t^i = \int_0^t \psi\left(s, X_s^i, W_{s+G^i}\right) ds, \tag{1}$$

$$\tau_i = \inf\left\{ t \in \mathbb{R}_+ \,|\, \Gamma_t^i \geq \eta_i \right\}, \tag{2}$$

where η_i is an independent, standard exponentially distributed random variable. The η_i variables are independent of all covariates and each other. Note the portion of Z_t^i corresponding to $[G^i, G^i + T]$ is used in the definition of Γ_t^i.

The usual definition of the hazard rate for a random time τ is:

$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{P}\left\{ t \leq \tau < t + \Delta t | \, t \leq \tau ; Z\left(s+\right), s \in [0,t] \right\} = \psi\left(Z\left(t+\right)\right).^{7} \tag{3}$$

This is the starting point for most papers in the literature. In the proposed setup, $\psi(\cdot)$ is the hazard function as in (3).[8][9] When Z_t^i are *i.i.d.*, (1)-(2) is equivalent to models studied in *i.i.d.* hazard papers (with the restriction (B1)). In this paper, Z_t^i can be dependent across observations. Our setup is called the usual or canonical construction in BR.

It is required that η_i are standard exponential because a hazard function gives the Poisson rate of arrival for an event. Poisson rates are closely related to the exponential distribution. BR Section 6.6 discusses this connection in more detail. Other choices for the distribution of η_i do not lead to the standard definition in the literature (3). Studying the case where η_i is not standard exponential would be interesting, but this takes us away from the established *i.i.d.* literature and raises technical issues. In principal, it is possible to test this assumption with an estimator of the hazard function and a time change of the observed events. See Daley and Vere-Jones (2003) Section 7.4 for more details. In applications, a transformation of the covariates Z_t^i may improve alignment with this requirement.

In the dependence case studied in this paper, we make the modeling assumption that each η_i is independent. Therefore, we assume all dependence happens through Z_t^i. This simplifies a number of technical conditions and is required for the results that follow. See BR Section 9.1.2. for more details on the dependent case.

[7] The notation $Z(t+)$ means the right-hand-limit of $Z(t)$ in time. This is standard notation in the continuous-time stochastic processes literature.

[8] See Fleming and Harrington's (1991) Theorem 4.2.1 and BR's Section 6.5.

[9] We also need the weak condition that the filtration generated by $Z(s+)$ is right-continuous. There is no avoiding the adjustment from $Z(s)$ to $Z(s+)$ in (3). This is clear from Fleming and Harrington's (1991) Theorem 4.2.1. It could be assumed $Z(s)$ has paths which are right-continuous with left-hand-limits (càdlàg). Then we would have to replace $Z(t)$ with $Z(t-)$ (the left-hand-limit of the process) in the construction. If this is done, we can replace $Z(t+)$ with $Z(t)$ in (3). The author feels that what is presented is less confusing in the majority of the paper.

Define $Y_t^i = \mathbf{1}(t \leq \tau_i)$, the indicator that default has not occurred.[10] Define also $N_t^i = \mathbf{1}(\tau_i \leq t)$, the indicator of default. Consider the following processes:

$$\Lambda_t^i = \int_0^t \psi\left(s, X_s^i, W_{s+G^i}\right) Y_s^i ds,$$

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

Our asymptotic results depend on M_t^i being continuous-time martingales. The needed martingale structure is verified in the Online Supplement. The adjustments needed for conditionally independent censoring are also discussed there.

A few comments about this setup are in order.

1.) The stochastic processes (X_t^i, W_{t+G^i}) are exogenous. The events do not impact these basic processes, they follow from them. For example, in a corporate default context, assume W_t is the GDP growth rate. It follows that GDP growth cannot be impacted by corporate defaults. This restriction is important to consider when applying or interpreting the results that follow.

2.) The construction allows us to easily control dependence between random times N_t^i. Because N_t^i are directly constructed from Z_t^i and η_i, N_t^i inherit the dependence structure of these variables. This is used in deriving asymptotic results.

3 Kernel Estimation

In this section, kernel estimators are derived which extend the results of Nielsen and Linton (1995) and Linton, Nielsen and Van de Geer (2001) to the dependent case.

Let k be a continuous one-dimensional probability density and $k_b(\cdot) = \frac{1}{b}k(\cdot/b)$ for some bandwidth $b > 0$. $K(u) = \Pi_{l=1}^{d+j+1} k(u_l)$ where $u = (u_1, \ldots, u_{d+j+1})$ and $K_b(u) = \Pi_{l=1}^{d+j+1} k_b(u_l)$. Product kernels and a single bandwidth are used throughout. We additionally make assumption (G). Recall the notation $Z_t^i = (t, X_t^i, W_{t+G^i})$. Pointwise results in the sequel assume an interior

[10] This is made left-continuous to preserve predictability.

point $z=(t,x,w)$ in the support $[0,T]\times\mathcal{X}\times\mathcal{W}$. The following notation is used:

$$
\begin{aligned}
K_b\left(z-Z_s^i\right) &= \frac{1}{b}k\left(\frac{t-s}{b}\right) \\
&\quad \times\frac{1}{b^d}k\left(\frac{x_1-X^{i1}(s)}{b}\right)\cdots k\left(\frac{x_d-X^{id}(s)}{b}\right) \\
&\quad \times\frac{1}{b^j}k\left(\frac{w_1-W^1\left(s+G^i\right)}{b}\right)\cdots k\left(\frac{w_j-W^j\left(s+G^i\right)}{b}\right).
\end{aligned}
$$

As in NL, the hazard rate $\psi(t,x,w)$ is estimated by,

$$
\widehat{\psi}(t,x,w)=\frac{\frac{1}{n}\sum_{i=1}^n\int_0^T K_b\left(z-Z_s^i\right)dN_s^i}{\frac{1}{n}\sum_{i=1}^n\int_0^T K_b\left(z-Z_s^i\right)Y_s^i ds}=\frac{\widehat{o}(z)}{\widehat{e}(z)}.
$$

Define,

$$
\psi^*(t,x,w)=\frac{\frac{1}{n}\sum_{i=1}^n\int_0^T K_b\left(z-Z_s^i\right)\psi\left(Z_s^i\right)Y_s^i ds}{\frac{1}{n}\sum_{i=1}^n\int_0^T K_b\left(z-Z_s^i\right)Y_s^i ds},
$$

and decompose $(\widehat{\psi}-\psi)$ as,

$$
(\widehat{\psi}-\psi)(z)=(\widehat{\psi}-\psi^*)(z)+(\psi^*-\psi)(z)=\frac{V_n(z)}{\widehat{e}(z)}+\frac{B_n(z)}{\widehat{e}(z)}.
$$

Here,

$$
\begin{aligned}
\widehat{e}(z) &= \frac{1}{n}\sum_{i=1}^n\int_0^T K_b\left(z-Z_s^i\right)Y_s^i ds, \qquad (4)\\
V_n(z) &= \frac{1}{n}\sum_{i=1}^n\int_0^T K_b\left(z-Z_s^i\right)dM_s^i, \\
B_n(z) &= \frac{1}{n}\sum_{i=1}^n\int_0^T K_b\left(z-Z_s^i\right)\left[\psi\left(Z_s^i\right)-\psi(z)\right]Y_s^i ds.
\end{aligned}
$$

We will also need the additional notation:

$$\begin{aligned}
\widehat{\sigma}_z^2 &= \frac{1}{\widehat{e}^2(z)} \frac{b^{d+j+1}}{n} \sum_{i=1}^{n} \int_0^T K_b^2\left(z - Z_s^i\right) dN_s^i, \\
\mathcal{K}_z &= \frac{b^{d+j+1}}{n} \sum_{i=1}^{n} \int_0^T K_b^2\left(z - Z_s^i\right) \psi\left(Z_s^i\right) Y_s^i ds, \\
\mathcal{H}_z &= \frac{b^{2(d+j+1)}}{n^2} \sum_{i=1}^{n} \int_0^T K_b^4\left(z - Z_s^i\right) \psi\left(Z_s^i\right) Y_s^i ds.
\end{aligned} \tag{5}$$

The idea behind the estimator $\widehat{\psi}(z)$ is simple. If we compare the number of events τ_i that happen when Z_t^i is near z, to the frequency Z_t^i is near z, we get an idea of the arrival rate $\psi(z)$. $\widehat{\psi}(z)$ is a just a kernel version of this comparison. The numerator measures how frequently events happen when the covariates are near z. The denominator measures how frequently the covariates are near z. We derive exact asymptotic results for $\widehat{\psi}(z)$ to rigorously describe this approximation. A number of regulatory conditions are needed for the asymptotics that follow.

(B2) Assume Z_t^i has the same distribution for each i.

(B3) Each observation i has the same mean functional:

$$y(t) = \mathbb{E}\left[Y^i(t)\right],$$

and conditional distribution function:

$$F_t(x, w) = \mathbb{P}\left[X^i(t) \le x, W\left(t + G^i\right) \le w \middle| Y^i(t) = 1\right].$$

This distribution is assumed to have a corresponding density $f_t(x, w)$ with support equal to $\mathcal{X}_1 \times \cdots \times \mathcal{X}_d \times \mathcal{W}_1 \times \cdots \times \mathcal{W}_j$, the support assumed above.

(K2) The kernel k has support $[-1, 1]$, is symmetric about 0 and is of order r, that is $\int_{-1}^{1} k(u) u^j du = 0$, $j = 0, \ldots, r-1$, and $\mu_r(k) = \int_{-1}^{1} k(u) u^r du \in (0, \infty)$, where $r \ge 2$ is an even integer. The kernel is also $r-1$ times continuously differentiable on $[-1, 1]$ with Lipschitz remainder;

that is, there exists a finite constant k_{lip} such that $\left|k^{(r-1)}(u) - k^{(r-1)}(u')\right| \leq k_{lip}|u-u'|$ for all u, u'. Finally, $k^{(j)}(\pm 1) = 0$ for $j = 0, \ldots, r-1$.

Conditions (B2) and (B3) are similar to assumptions in LNV. These requirements are easily satisfied by a stationarity assumption on the common covariates W_t. Condition (K2) puts restrictions on our choice of kernel. (K2) is assumption (A4) from LNV and is reproduced here for convenience. (K2) makes it possible $\widehat{\psi}(z)$ is negative at some values. This is a standard problem when using kernel methods. The probability this happens vanishes in the limit under the assumptions given below.

Theorem 1 below follows theorem 1 in NL. Additional high level assumptions are necessary because of dependence. Later in this section, assumptions are discussed which imply the conditions (6)-(9). We define

$$e(z) = f_t(x, w)\, y(t),$$

which is used throughout the sequel. Recall that $d + j + 1$ is the number of covariates in the model.

Theorem 1 *Assume (G), (B1)-(B3) and (K2). (S):* $e(z) > 0$ *on a neighborhood of* z*;* $\psi, e \in C^r$ *in a neighborhood of* z *where* $r \geq 2$*. (B):* $nb^{d+j+1} \to \infty$ *and* $b \to 0$*. Define the constant* $\kappa_2 = \int_{-1}^{1} k^2(v)\, dv$ *and the following constant which depends on* z*:*

$$C(z) = \mu_r(k) \sum_{i=1}^{d+j+1} \left\{ \frac{1}{(r-1)!} \frac{\partial^{(r-1)}\psi(z)}{\partial z_i^{(r-1)}} \frac{\partial e(z)}{\partial z_i} + \frac{1}{r!} e(z) \frac{\partial^r \psi(z)}{\partial z_i^r} \right\}.$$

Assume

$$\widehat{e}(z) - \mathbb{E}[\widehat{e}(z)] \to^p 0, \tag{6}$$

$$\mathcal{K}_z - \mathbb{E}[\mathcal{K}_z] \to^p 0, \tag{7}$$

$$\mathcal{H}_z - \mathbb{E}[\mathcal{H}_z] \to^p 0, \tag{8}$$

$$b^{-r}\{B_n(z) - \mathbb{E}[B_n(z)]\} \to^p 0. \tag{9}$$

Then the following results hold:

$$n^{1/2}b^{(d+j+1)/2}\left[\widehat{\psi}(z)-\psi^{*}(z)\right]\Rightarrow N\left[0,\kappa_{2}^{d+j+1}\frac{\psi(z)}{e(z)}\right],\tag{10}$$

$$b^{-r}\left[\psi^{*}(z)-\psi(z)\right]\rightarrow^{p}\frac{C(z)}{e(z)},\tag{11}$$

$$\widehat{\sigma}_{z}^{2}\rightarrow^{p}\sigma_{z}^{2}\equiv\kappa_{2}^{d+j+1}\frac{\psi(z)}{e(z)}.\tag{12}$$

In particular, if we choose the bandwidth such that $b\sim n^{-1/(d+j+1+2r)}$*, then*

$$n^{1/2}b^{(d+j+1)/2}\left[\widehat{\psi}(z)-\psi(z)\right]\Rightarrow N\left[\frac{C(z)}{e(z)},\kappa_{2}^{d+j+1}\frac{\psi(z)}{e(z)}\right].\tag{13}$$

Both the conditions (B) and (S) are similar to assumptions made in NL and LNV. Condition (B) restricts the bandwidth and is unchanged from the *i.i.d.* case. The asymptotic variance in (10) is also equivalent to the *i.i.d.* case. However, in order for conditions (6)-(9) to hold when dependence is present additional restrictions on the bandwidth are needed. This is discussed below. Changing the bandwidth affects confidence intervals. The asymptotic results above follow (regardless of dependence in the covariates) as long as the conditions hold.

The assumptions (6)-(9) each state that a particular array of mean zero random variables converges to zero in probability. All of these arrays are derived from the underlying covariates Z_t^i and random thresholds η_i. As a result, an assumption on the dependence of Z_t^i determines the dependence properties for all these arrays. The approach to verifying the conditions in Theorem 1 is to make mixing assumption on the underlying variables Z_t^i and η_i. These mixing assumptions transfer their properties onto the sums $\widehat{e}(z)$, $B_n(z)$, $\mathcal{K}_z$ and $\mathcal{H}_z$. Once their expectations are subtracted off, the rows of these arrays are sums of mean zero random variables which satisfy mixing conditions. Finally, we show (6)-(9) with a Bernstein inequality based on the mixing properties.

In this program, we assume α-mixing and a specific Bernstein inequality for concreteness. These choices seem to be the most natural and standard that can facilitate asymptotics. Many results that follow could be derived using slightly different technical tools.[11] Recall the definition of α-mixing coefficients for a sequence of random variables D_t, $t \in \mathbb{N}$.[12] The random variables D_t can be used to define the following σ-fields:

$$\mathcal{F}_i^j = \sigma \left\{ D_t \middle| t = i, \ldots, j; i \leq j \right\}.$$

The α-mixing coefficients are:

$$\alpha\left(\overline{n}\right) = \sup_{k \in \mathbb{N}} \sup \left\{ \left| P\left(A \cap B\right) - P\left(A\right) P\left(B\right) \right| ; A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+\overline{n}}^{\infty} \right\},$$

where the second sup is over all sets A, B such that $A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+\overline{n}}^{\infty}$. In the results that follow, we use mixing conditions on the system of σ-fields defined by the covariates:

$$\mathcal{H}_l^m = \vee_{i=l}^m \sigma \left\{ X^i\left(s\right), W\left(u + G^i\right) \middle| 0 \leq s, u \leq T \right\}.^{13}$$

The results that follow will be stated for general α-mixing rates. Any setup which satisfies the assumed conditions will yield the asymptotic results. However, in order to better guide applied researches, some examples which satisfy the required conditions are presented here. The starting point for these examples is a continuous-time process F_t defined on $t \in [0, \infty)$ and assumed to be α-mixing. The α-mixing rates of commonly used processes can be found in Carrasco and Chen (2002) and Chen, Hansen and Carrasco (2010). These papers give many potential choices for F_t.

[11] The assumed α-mixing conditions are somewhat restrictive. We can easily construct processes which do not satisfy α-mixing. For example:

$$W_t = \begin{cases} 1 + v & [t] \text{ even} \\ v & [t] \text{ odd} \end{cases},$$

where v is uniformly distributed on $[0, 1]$. Here, $[\cdot]$ gives the largest integer less than the argument. W_t is not α-mixing and the assumptions in this paper do not hold. For the nonparametric case, it is clear the hazard function can only be recovered at the two realized points $1 + v$ and v. However, if $\psi(z)$ is restricted to be parametric, in many cases it is possible to recover the hazard function for all values of W_t in its support. This example highlights the trade-offs faced in estimating hazard models. W_t can fail to be α-mixing and the hazard can still be recovered. However, this may require restricting the form of $\psi(z)$. I thank a referee for suggesting this example.

[12] See Davidson (1994) for more details.

[13] Here, the covariates for each observation produce a σ-field $\sigma\left\{ X^i\left(s\right), W\left(u + G^i\right) \middle| 0 \leq s, u \leq T \right\}$. $\mathcal{H}_l^m$ is the smallest σ-field which contains all the σ-fields for observations i such that $m \leq i \leq l$. This is a rigorous way of saying all the information from covariates corresponding to observations $m \leq i \leq l$ is contained in $\mathcal{H}_l^m$. The variables η_i do not affect the mixing rate because of independence.

Discrete-time processes can be imbedded in continuous-time while preserving mixing conditions. The mixing results presented in these examples are proven in the Online Supplement.

Example 2 (Only Macro Covariates) *Assume* $W_t = F_t$ *and that we have no covariates* X_t^i. *Let* F_t *have* α*-mixing coefficients such that* $\overline{\alpha}(t) \leq C_1 \exp(-ct)$ *for some* $C_1 > 0$ *and* $c > 0$. *If* $G^i = \left[\frac{i}{a}\right]$ *for some integer* $a > 0$, *then the sigma fields* $\mathcal{H}_l^m$ *have* α*-mixing coefficients such that* $\alpha(n) \leq C_2 n^{-\xi}$ *for some* $C_2 > 0$ *and* $\xi > 0$.

Example 3 (Macro Covariates Plus Idiosyncratic Variables) *Assume* $W_t = F_t$ *and* X_t^i *are independent of* W_t *and i.i.d. Let* F_t *have* α*-mixing coefficients such that* $\overline{\alpha}(t) \leq C_1 \exp(-ct)$ *for some* $C_1 > 0$ *and* $c > 0$. *If* $G^i = \left[\frac{i}{a}\right]$ *for some integer* $a > 0$, *then the sigma fields* $\mathcal{H}_l^m$ *have* α*-mixing coefficients such that* $\alpha(n) \leq C_2 n^{-\xi}$ *for some* $C_2 > 0$ *and* $\xi > 0$.

Example 4 (Macro Covariates Plus Factor Based Variables) *Assume* F_t *has two sets of processes* F_t^1 *and* F_t^2. *Assume* M_t^i *are i.i.d. processes which are independent of* F_t. *Let* $W_t = F_t^1$ *and* $X_t^i = F_{t-G^i}^2 + M_t^i$.[14] *Let* F_t *have* α*-mixing coefficients such that* $\overline{\alpha}(t) \leq C_1 \exp(-ct)$ *for some* $C_1 > 0$ *and* $c > 0$. *If* $G^i = \left[\frac{i}{a}\right]$ *for some integer* $a > 0$, *then the sigma fields* $\mathcal{H}_l^m$ *have* α*-mixing coefficients such that* $\alpha(n) \leq C_2 n^{-\xi}$ *for some* $C_2 > 0$ *and* $\xi > 0$.

All results that follow can be stated for general α-mixing rates. Results for two specific rates commonly found in the literature are presented. These rates are polynomial and exponential decay in the mixing coefficients. The examples given above have polynomial mixing. The bandwidth restrictions required for Proposition 5 are now presented.

(M1) The system of σ-fields $\mathcal{H}_l^m$ has an α-mixing rate $\alpha(n)$ such that for some $C > 0$ and $\xi > 0$,

$$\alpha(n) \leq C n^{-\xi}.$$

Additionally, for some $0 < \widetilde{\gamma} < 1$,

$$\begin{aligned} b^{d+j+1+2r-2} n^{\widetilde{\gamma}} &\rightarrow \infty, \\ b^{-(d+j+1+r)/2} n^{\widetilde{\gamma}} \alpha\left(n^{(1-\widetilde{\gamma})}\right) &= C b^{-(d+j+1+r)/2} n^{-(1-\widetilde{\gamma})\xi+\widetilde{\gamma}} \rightarrow 0. \end{aligned} \tag{14}$$

[14] The adjustment for X_t^i is needed because it is defined on $t \in [0, T]$.

(M2) The system of σ-fields $\mathcal{H}_l^m$ has an α-mixing rate $\alpha(n)$ such that for some $C > 0$ and $c > 0$,

$$\alpha(n) \leq C \exp(-cn).$$

Additionally, for some $0 < \widetilde{\gamma} < 1$,

$$\begin{aligned} b^{d+j+1+2r-2} n^{\widetilde{\gamma}} &\rightarrow \infty, \\ b^{-(d+j+1+r)/2} n^{\widetilde{\gamma}} \alpha\left(n^{(1-\widetilde{\gamma})}\right) &= C b^{-(d+j+1+r)/2} n^{\widetilde{\gamma}} \exp\left(-cn^{1-\widetilde{\gamma}}\right) \rightarrow 0. \end{aligned} \tag{15}$$

Proposition 5 *Assume (G), (B1)-(B3) (K2), (S) and (B). If either (M1) or (M2) hold, then the conditions (6)-(9) of Theorem 1 hold.*

In order to have an asymptotically normal estimator with finite asymptotic bias, the requirements for the bandwidth are $n^{1/2} b^{(d+j+1)/2} \leq b^{-r}$ and either (M1) or (M2). Both (M1) and (M2) require $b^{(d+j+1+2r-2)} n^{\widetilde{\gamma}} \rightarrow \infty$. Comparing this to the analogous condition in the *i.i.d.* case $b^{(d+j+1+2r-2)} n \rightarrow \infty$,[15] we see that the rate at which $b \rightarrow 0$ must be slowed down. The *i.i.d.* case does not require a second term such as (14) or (15) as the mixing coefficients are zero. The extent to which the bandwidth must be slowed down depends on which values of $\widetilde{\gamma}$ satisfy (M1) or (M2). This depends on the rate at which the mixing coefficients decay. In the exponential mixing case, the bandwidths can be chosen arbitrarily close to those of the *i.i.d.* case as the exponential term in (15) will dominate the other terms. When dependence is stronger as in (M1), $\widetilde{\gamma}$ cannot be chosen arbitrarily close to 1. (14) forces us to choose $\widetilde{\gamma}$ differently, which changes the available bandwidths satisfying $b^{(d+j+1+2r-2)} n^{\widetilde{\gamma}} \rightarrow \infty$.

This is a theme throughout our results. Exponential mixing allows for essentially the same bandwidth choices as the *i.i.d.* case whereas polynomial mixing substantively changes our options. There is an important caveat to these asymptotic results. The term (15) (and others like it that follow) is used to bound a probability in the proofs. When $\widetilde{\gamma}$ is close to 1, while (15) holds asymp-

[15] In the analogous *i.i.d.* result presented in NL (their Theorem 1), it appears that the condition $b^{(d+j+1+2r-2)} n \rightarrow \infty$ is needed to prove their result (b) while their theorem only states $b^{(d+j+1)} n \rightarrow \infty$. This follows from the use of Chebyshev's inequality in their proof. The stronger condition is needed to control the bias term. Once this slightly stronger bandwidth condition is assumed, their Theorem 1 holds as stated. NL later assumes this stronger condition in their Theorem 2.

totically, this term will converge slowly. The result can be poor finite sample approximations. A smaller $\widetilde{\gamma}$ (and correspondingly slower b) can give the same asymptotic results with better finite sample properties.

We now present a uniform convergence rate for the dependent case. This requires some additional assumptions. We again present conditions for the polynomial and exponential mixing cases.

(U) Assume $y(t)$ and $f_t(x,w)$ are continuous on $[0,T]$ for all $(x,w) \in \mathcal{X}\times\mathcal{W}$. $\psi(\cdot)$ and $e(\cdot)$ are $r \geq 2$ times continuously differentiable at all points $z \in [0,T]\times\mathcal{X}\times\mathcal{W}$. $\inf_{(t,x,w)\in[0,T]\times\mathcal{X}\times\mathcal{W}}[e(z)] > 0$.

(I) Let I be any compact rectangle such that $I \subset [0,T]\times\mathcal{X}\times\mathcal{W}$ and I is an interior set of $[0,T]\times\mathcal{X}\times\mathcal{W}$.

(M1B) The system of σ-fields $\mathcal{H}_l^m$ has an α-mixing rate $\alpha(n)$ such that for some $C > 0$ and $\xi > 0$,

$$\alpha(n) \leq Cn^{-\xi}.$$

Additionally, for some $0 < \gamma < 1$,

$$\begin{aligned} b^{(d+j+1)}n^{\gamma} &\to \infty, \\ \frac{n^{9/4\gamma}}{b^{(d+j+1)/4}\sqrt{\log n}}\alpha\left(n^{1-\gamma}\right) &= C\frac{n^{9/4\gamma-(1-\gamma)\xi}}{b^{(d+j+1)/4}\sqrt{\log n}} \to 0. \end{aligned}$$

(M2B) The system of σ-fields $\mathcal{H}_l^m$ has an α-mixing rate $\alpha(n)$ such that for some $C > 0$ and $c > 0$,

$$\alpha(n) \leq C\exp(-cn).$$

Additionally, for some $0 < \gamma < 1$,

$$\begin{aligned} b^{(d+j+1)}n^{\gamma} &\to \infty, \\ \frac{n^{9/4\gamma}}{b^{(d+j+1)/4}\sqrt{\log n}}\alpha\left(n^{1-\gamma}\right) &= C\frac{n^{9/4\gamma}}{b^{(d+j+1)/4}\sqrt{\log n}}\exp\left(-cn^{1-\gamma}\right) \to 0. \end{aligned}$$

Theorem 6 *Make the assumptions (G), (B1)-(B3), (K2), (B), (U) and (I). If (M1B) or (M2B) hold then the following holds:*

$$\sup_{z \in I} \left| \widehat{\psi}(z) - \psi(z) \right| = O(b^r) + O_p \left\{ \sqrt{\frac{\log n}{n^\gamma b^{(d+j+1)}}} \right\} \tag{16}$$

The assumptions in Theorem 6 contain the same assumptions LNV use in the *i.i.d.* case (their $(A1), (A3)$ and $(A4)$). We further restrict the mixing conditions and bandwidth rate to deal with dependence. The strength of mixing conditions determines which bandwidth choices and values of γ are possible. This is a similar adjustment as required for Proposition 5.

Comparing our result with Lemma 3 from LNV, we see that dependence slows down the uniform convergence rate by replacing $nb^{(d+j+1)}$ with $n^\gamma b^{(d+j+1)}$. This influences the rate in two ways. First through the choice of the bandwidth and second from the difference in forms of $nb^{(d+j+1)}$ and $n^\gamma b^{(d+j+1)}$. The strength of the mixing conditions determines the final rate of convergence by restricting γ and b. The stronger the mixing condition the slower the uniform rate of convergence. For an exponential rate we can get arbitrarily close to the *i.i.d.* case. A caveat to this statement similar to the one discussed above holds here. The polynomial case will not allow the same rate as the *i.i.d.* case.

3.1 The Curse of Dimensionality

Throughout the paper, we face a standard curse of dimensionality problem as the number of covariates increases. As in LNV, if we restrict the form of the hazard function we can greatly improve the rate of convergence. This follows provided additional bandwidth assumptions hold. Previous results in this section are prerequisites. They are used extensively in the proof of Theorem 7 presented below.

Specifically, we assume the hazard is either additively or multiplicatively separable:

$$\psi(z) = c_A + \sum_{l=1}^{d+j+1} g_l(z_l), \tag{17}$$

$$\psi(z) = c_M \prod_{l=1}^{d+j+1} h_l(z_l), \tag{18}$$

where c_A and c_M are constants. The individual functions in (17)-(18) are not separately identified. An additional probability measure over the compact rectangle I is needed in order to identify them. Let Q be an arbitrary cdf with all probability on I and marginal cdfs $Q_l(z_l) = (\infty, \ldots, \infty, z_l, \infty, \ldots, \infty)$ and $Q_{-l}(z_{-l}) = (z_1, \ldots, z_{l-1}, \infty, z_{l+1}, \ldots, z_{d+j+1})$. The functions in (17)-(18) are identified by assuming,

$$\int g_l(z_l)\, dQ_l(z_l) = 0, \tag{19}$$

or

$$\int h_l(z_l)\, dQ_l(z_l) = 1, \tag{20}$$

for all $l = 1, \ldots, d+j+1$. For simplicity we assume Q has the following form.

(B4) Let I be any compact rectangle such that $I \subset [0,T] \times \mathcal{X} \times \mathcal{W}$ and I is an interior set of $[0,T] \times \mathcal{X} \times \mathcal{W}$. Q is the uniform distribution on I. Therefore, It has density equal to $\frac{1}{q_1} \cdots \frac{1}{q_{d+j+1}}$ where q_l is the length of the compact interval corresponding to variable l.

Under (B4), Q satisfies assumption (A2) from LNV, which contains the conditions on Q required for their asymptotic results. The following results could be extended to more complicated probability measures. Note that $\int \psi(z)\, dQ(z) = c_A$ and $\int \psi(z)\, dQ(z) = c_M$ in each model respectively. We write c generically when the specific model is unimportant.

The following definitions are made:

$$\begin{aligned}
\psi_{Q_{-j}}(z_j) &= \int \psi(z)\, dQ_{-j}(z_{-j}), \\
\psi^A_{Q_{-j}}(z_j) &= \psi_{Q_{-j}}(z_j) - c = g_j(z_j), \\
\psi^M_{Q_{-j}}(z_j) &= \frac{\psi_{Q_{-j}}(z_j)}{c} = h_j(z_j),
\end{aligned}$$

and define the corresponding estimators:

$$\begin{aligned}
\widehat{c} &= \int \widehat{\psi}(z)\, dQ(z), \\
\widehat{\psi}_{Q_{-j}}(z_j) &= \int \widehat{\psi}(z)\, dQ_{-j}(z_{-j}), \\
\widehat{\psi}^A_{Q_{-j}}(z_j) &= \widehat{\psi}_{Q_{-j}}(z_j) - \widehat{c}, \\
\widehat{\psi}^M_{Q_{-j}}(z_j) &= \frac{\widehat{\psi}_{Q_{-j}}(z_j)}{\widehat{c}}, \\
\widehat{\psi}_A(z) &= \sum_{k=1}^{d+j+1} \widehat{\psi}^A_{Q_{-k}}(z_k) + \widehat{c}, \\
\widehat{\psi}_M(z) &= \widehat{c} \prod_{k=1}^{d+j+1} \widehat{\psi}^M_{Q_{-k}}(z_k).
\end{aligned}$$

In order to derive asymptotic results, we again make assumptions on the bandwidth and α-mixing rates. These additional conditions are technical and involved, but similar to (M1)-(M2) and (M1B)-(M2B) previously given. The same pattern arises where stronger dependence requires bandwidths to converge more slowly to zero. The additional assumptions (MB) and (MC) are presented and discussed in the Appendix.

Theorem 7 *Make the assumptions (G), (B1)-(B4), (K2), (B), (U), (I) and the identification assumption (19) or (20). Assume (MB), (MC) and* $n^{1/(2r+1)}b \to C$ *where* $0 \le C < \infty$*. If* $C = 0$*, then there is no asymptotic bias in the distributions that follow. Under these assumptions there exist functions* $m_j(\cdot)$, $v_j(\cdot)$ *which are bounded and continuous on* I_j *such that for all* $z_j \in I_j$

$$(nb)^{1/2}\left(\widehat{\psi}_{Q_{-j}}(z_j) - \psi_{Q_{-j}}(z_j)\right) \Rightarrow N\left[m_j(z_j), v_j(z_j)\right].$$

$v_j(z_j)$ *is equal to*

$$v_j(z_j) = \kappa_2 \int_{I_{-j}} \frac{\psi(z) \frac{1}{(q_{-j})^2}}{e(z)} dz_{-j}.$$

If we compute $\widehat{c}$ *using bandwidths such that* $\widehat{c} - c = O_p\left(n^{-1/2}\right)$, *then*

$$(nb)^{1/2} \left(\widehat{\psi}^A_{Q_{-j}}(z_j) - g_j(z_j)\right) \Rightarrow N\left[m_j(z_j), v_j(z_j)\right], \tag{21}$$

$$(nb)^{1/2} \left(\widehat{\psi}^M_{Q_{-j}}(z_j) - h_j(z_j)\right) \Rightarrow N\left[m_j(z_j)/c, v_j(z_j)/c^2\right]. \tag{22}$$

Finally

$$(nb)^{1/2} \left(\widehat{\psi}_A(z) - \psi_A(z)\right) \Rightarrow N\left[m_A(z), v_A(z)\right], \tag{23}$$

$$(nb)^{1/2} \left(\widehat{\psi}_M(z) - \psi_M(z)\right) \Rightarrow N\left[m_M(z), v_M(z)\right]. \tag{24}$$

Where, for the additive case $m_A(z) = \sum_{k=1}^{d+j+1} m_k(z_k)$ *and* $v_A(z) = \sum_{k=1}^{d+j+1} v_k(z_k)$. *For the multiplicative case* $m_M(z) = \sum_{k=1}^{d+j+1} m_k(z_k) s_k(z_{-k})$ *and* $v_M(z) = \sum_{k=1}^{d+j+1} v_k(z_k) s_k^2(z_{-k})$ *where* $s_k(z_{-k}) = \prod_{i \neq k} \psi_{Q_{-i}}(z_i)/c^{d+j+1}$.

Note that if $b = n^{-1/(2r+1)}$, then $(nb)^{1/2} = n^{r/(2r+1)}$. This corresponds to the optimal rate derived in Stone (1980) when the problem is one-dimensional.

4 Simulations

In this section, a number of simulation studies are conducted to test the performance of the proposed estimators under dependence. We assume basic underlying covariates which take the form,

$$\overline{W}_t = \omega + \alpha \overline{W}_{t-1} (\epsilon_t)^2 + \beta_1 \overline{W}_{t-1} + \beta_2 \overline{W}_{t-2} + \beta_3 \overline{W}_{t-3},$$

where ϵ_t are *i.i.d.* standard normal shocks. The values $\omega = 1$, $\alpha = 0.05$, $\beta_1 = \beta_2 = 0.3$ and $\beta_3 = 0.1$ are used throughout. This process is stationary and α-mixing with exponential decay (see Carrasco and Chen (2002) proposition 12). $10,000$ observations are burned in, so the process starts at its unconditional distribution. $\overline{W}_t$ takes almost all of its values in the range $[3.5, 5.5]$, but does have some larger realizations. The process is rescaled as follows to keep values within the range:

$$W_t = 3.5 + 2\overline{F}\left(\overline{W}_t\right).$$

where $\overline{F}$ is the cdf of $N(4.2, 0.25)$. The hazard specification given above is in continuous-time while W_t is discrete. W_t is made continuous by assuming piecewise constant paths with updates given by the discrete-time process. In the simulations we discretize several aspects of the setup. This is discussed in detail in the Online Supplement.

The specifications used assume two independent processes, W_t^1 and W_t^2, each with the form given above. These processes are macro variables which all observations share. The hazard functions take three different forms: $\psi_{11}(w_1, w_2) = \lambda_1(w_1)\lambda_1(w_2)$, $\psi_{12}(w_1, w_2) = \lambda_1(w_1)\lambda_2(w_2)$ and $\psi_{22}(w_1, w_2) = \lambda_2(w_1)\lambda_2(w_2)$ where λ_1 and λ_2 are as follows:

$$\begin{aligned} \lambda_1(w) &= -\frac{1}{2\sqrt{5}}(w - 3.4)(w - 5.6), \\ \lambda_2(w) &= \frac{1}{\sqrt{5}}\exp((0.1)w). \end{aligned}$$

The three hazard specifications are shown in Figure 1. Each has a very different shape, which allows us to examine the performance of the estimators in various situations.

Observations are at risk over intervals $[G^i, G^i + T]$ where $T = 10$. The processes W_t^1 and W_t^2 are assumed to update every 1/10 unit of time. This results in 100 different realized values over $[G^i, G^i + T]$. For all simulations the number of observation is $n = 1000$. In order to see the effects of dependence, we vary the starting dates. Asymptotic results given above are based on α-mixing rates which are theoretically the same regardless of our choice of G^i (assuming they satisfy (G)). However, as the simulations that follow show, this entirely asymptotic perspective misses something. The performance of the estimators is strongly influenced by our choice of G^i. Intuitively, if the observations are further apart in calendar-time, they are less dependent. Rigorously, observations with different starting times but the same underlying processes will have the same mixing rates but different constants. This idea is explored in the simulations by organizing the $n = 1000$ starting times in different ways.

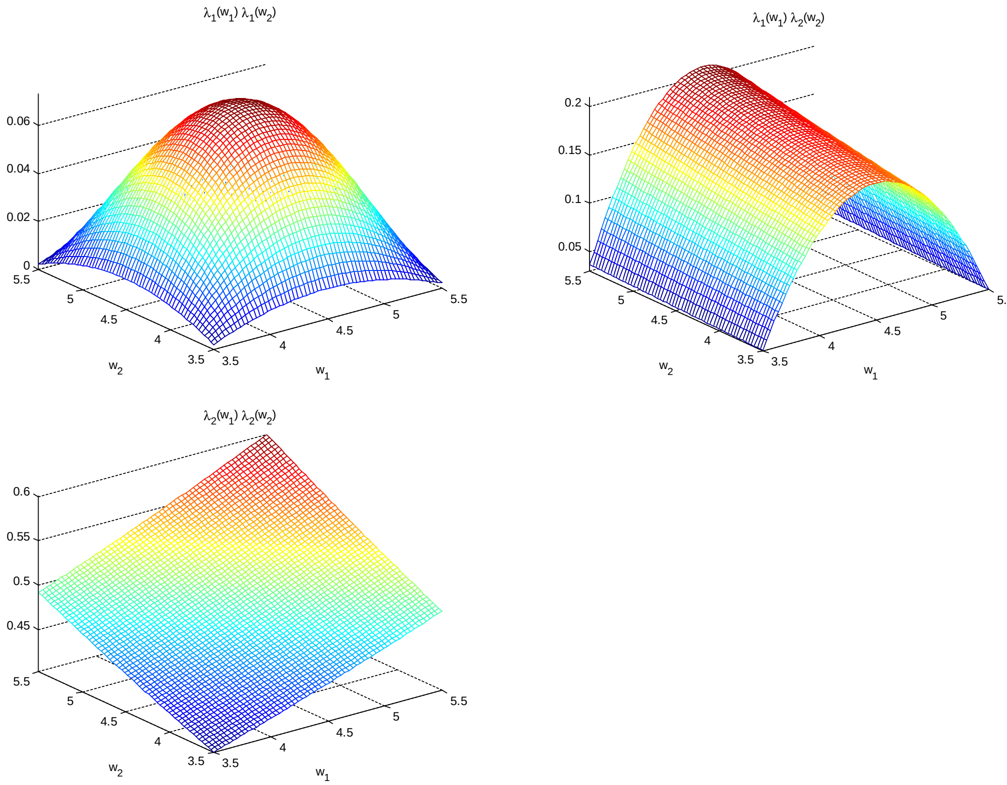

Figure 1: Hazard Specifications

Specifications are based on m groups ordered by $\pi = 1, \ldots, m$. Each of these groups contains $n_m = n/m$ observations. Observations $i = 1, \ldots,\ n/m$ are in group $\pi = 1$, observations $i = n/m + 1, \ldots, 2\,(n/m)$ are in group $\pi = 2$ and so on. All observations in group π are assumed to start at the calendar time $(\pi - 1) * 10$. This implies that, as one group finishes its exposure interval $[(\pi - 1) * 10, \pi * 10]$, the next group begins. As m becomes larger, dependence is reduced. This is because observations are influenced by realizations of the macro covariates which are further apart in calendar-time. We see more of W_t^i when m is larger. Simulations are conducted for the values $m = 1, 2, 5, 10, 20$. As a baseline, we show results when the covariates are $i.i.d.$ We expect performance of our estimators to become more like this baseline as m is increased. α-mixing coefficients corresponding to our sampling scheme die out faster when m is larger. This makes the dependence closer to the $i.i.d.$ case.

The chosen kernel is,

$$k\left(x\right) = \left[\max\left(0, 1 - x^2\right)\right]^{4.1} \left\{ \left(\frac{9}{8} - \frac{15}{8}x^2\right) + \frac{9}{2} * \frac{144}{\left(384\right)^2} * \left(1680x^4 - 1440x^2 + 144\right) \right\}.$$

This is a fourth order kernel constructed using Legendre polynomials and the Gegenbauer basis weighting scheme. See Tsybakov (2009) section 1.2.2 for details. The kernel satisfies assumption (K2). A problem for any kernel estimator is how to choose the bandwidth. Cross validation is a standard approach. See Gámiz et al. (2013) for an overview of the literature. In the simulations that follow, bandwidths are chosen using cross validation as in Gámiz et al. (2013).

In regression cases, previous literature has shown that cross validation preforms poorly in the presence of dependence. Bandwidth choices based on AIC-type information criteria preform better. In nonparametric cases, this requires defining the degrees or freedom (also called the effective number of parameters) which generalizes the number of parameters in a parametric model. See Li, Su and Xu (2015) for a summary.

Following this previous literature, bandwidths chosen using a degrees of freedom (hereafter df) adjustment are also considered. Methods developed for regression models cannot be directly used in hazard cases because the notion of a residual is different. Furthermore, the same definition of df does not apply. In the sequel we consider the df definition:

$$l_{\varrho}(b) = \begin{cases} \frac{\frac{1}{b}k(0)}{\sum_{i=1}^{n}\int_0^T K_b\left(Z^{\varrho}_{\tau_{\varrho}} - Z^i_s\right)Y^i_s ds} & \text{if } \tau_{\varrho} \in [G^{\varrho}, G^{\varrho}+T] \\ 0 & \text{else} \end{cases},$$

$$H(b) = \sum_{\varrho=1}^{n} l_{\varrho}(b).$$

See Loader (1999) for a general discussion of df in kernel smoothing applications. Bandwidths are chosen by minimizing:

$$CV(b) + \frac{2(H(b)+1)}{n - H(b) - 2}, \tag{25}$$

where $CV(b)$ is the criterion function from cross validation. This form is taken from Hurvich, Simonoff and Tsai (1998).

Four estimates are produced for each specification. The first two are $\widehat{\psi}^{CV}(z)$ and $\widehat{\psi}^{CV}_M(z)$. $\widehat{\psi}^{CV}(z)$ is the estimator with bandwidth chosen by cross validation. $\widehat{\psi}^{CV}_M(z)$ is the corresponding multiplicative estimator. Similarly, $\widehat{\psi}^{DF}(z)$ and $\widehat{\psi}^{DF}_M(z)$ are the corresponding estimators

	$\widehat{\psi}^{CV}(z)$						$\widehat{\psi}_M^{CV}(z)$					
$m - n_m$	0.25	0.50	0.75	0.90	0.95	0.99	0.25	0.50	0.75	0.90	0.95	0.99
i.i.d.	0.030	0.037	0.048	0.063	0.075	0.115	0.024	0.027	0.031	0.037	0.041	0.051
20-50	0.030	0.037	0.047	0.065	0.079	0.122	0.024	0.028	0.031	0.038	0.042	0.053
10-100	0.032	0.040	0.054	0.084	0.145	0.599	0.024	0.028	0.034	0.044	0.051	0.081
5-200	0.033	0.044	0.071	0.204	0.552	4.678	0.026	0.030	0.036	0.052	0.061	1.331
2-500	0.039	0.076	0.310	1.016	2.070	14.502	0.027	0.034	0.056	0.131	0.293	7.391
1-1000	0.066	0.210	0.596	1.559	2.695	13.364	0.032	0.045	0.080	0.200	0.512	9.596
	$\widehat{\psi}^{DF}(z)$						$\widehat{\psi}_M^{DF}(z)$					
	0.25	0.50	0.75	0.90	0.95	0.99	0.25	0.50	0.75	0.90	0.95	0.99
i.i.d.	0.029	0.034	0.041	0.051	0.061	0.102	0.025	0.028	0.032	0.037	0.041	0.051
20-50	0.028	0.033	0.040	0.054	0.065	0.110	0.024	0.028	0.031	0.037	0.041	0.059
10-100	0.029	0.035	0.045	0.060	0.073	0.253	0.025	0.029	0.033	0.039	0.043	0.071
5-200	0.030	0.036	0.049	0.105	0.283	1.810	0.025	0.029	0.034	0.043	0.057	0.282
2-500	0.035	0.048	0.161	0.751	1.540	15.967	0.027	0.033	0.046	0.091	0.203	21.160
1-1000	0.050	0.157	0.550	1.894	4.758	30.771	0.032	0.045	0.079	0.316	1.491	95.691

Table 1: $\lambda_1(w_1)\lambda_1(w_2)$, Quantiles of the Supremum of Absolute Difference

with bandwidth chosen by minimizing (25). Note that all three of our hazard specifications are multiplicative. This suggests $\widehat{\psi}_M^{CV}(z)$ and $\widehat{\psi}_M^{DF}(z)$ can potentially produce improvements.

The performance of the estimators is measured by two different metrics. The first is the supremum of the absolute difference between our estimator and the true hazard function. The second metric is the average absolute difference between our estimator and the true hazard function. The results for the two metrics are similar. We present the supremum case here and the second metric in the Online Supplement.

Tables 1-3 display quantiles of the supremum metric for each of the four estimators. Each set of quantiles was produced from 500 realizations of the above specifications. For all hazard functions and all estimators there is a clear pattern in the results. As m increases, the performance of the estimators improves. This improvement is sometimes large, especially in the quantiles representing the tails of the distribution. It is undeniable that increasing dependence in the specification leads to deterioration in performance. Note also that, as expected, when m increases the performance approaches that of the *i.i.d.* benchmark.

There are two other notable trends in Tables 1-3. First, the multiplicative estimators $\widehat{\psi}_M^{CV}(z)$ and $\widehat{\psi}_M^{DF}(z)$ are dramatically superior to the basic estimators $\widehat{\psi}^{CV}(z)$ and $\widehat{\psi}^{DF}(z)$. This is true for all three hazard functions. The improvement is greater when m is smaller. These estimators

	$\widehat{\psi}^{CV}(z)$						$\widehat{\psi}_M^{CV}(z)$					
$m - n_m$	0.25	0.50	0.75	0.90	0.95	0.99	0.25	0.50	0.75	0.90	0.95	0.99
i.i.d.	0.083	0.099	0.128	0.170	0.202	0.260	0.061	0.070	0.084	0.096	0.107	0.129
20-50	0.084	0.103	0.138	0.181	0.214	0.377	0.061	0.071	0.083	0.098	0.113	0.162
10-100	0.086	0.109	0.145	0.217	0.340	1.558	0.064	0.073	0.088	0.106	0.131	0.209
5-200	0.091	0.125	0.221	0.529	2.048	17.297	0.066	0.079	0.100	0.159	0.268	5.471
2-500	0.110	0.197	0.783	2.496	6.552	24.398	0.072	0.091	0.138	0.340	0.711	11.481
1-1000	0.162	0.525	1.521	3.835	6.883	76.162	0.086	0.115	0.182	0.443	1.430	68.012
	$\widehat{\psi}^{DF}(z)$						$\widehat{\psi}_M^{DF}(z)$					
	0.25	0.50	0.75	0.90	0.95	0.99	0.25	0.50	0.75	0.90	0.95	0.99
i.i.d.	0.077	0.093	0.114	0.149	0.179	0.376	0.064	0.074	0.085	0.097	0.110	0.179
20-50	0.079	0.095	0.114	0.147	0.170	0.283	0.064	0.073	0.085	0.100	0.111	0.137
10-100	0.083	0.098	0.126	0.176	0.214	1.931	0.066	0.075	0.088	0.103	0.114	0.211
5-200	0.084	0.106	0.145	0.283	0.975	9.049	0.065	0.076	0.094	0.123	0.165	1.369
2-500	0.100	0.142	0.464	1.900	5.187	32.287	0.075	0.091	0.122	0.234	0.541	24.610
1-1000	0.137	0.333	1.109	3.364	10.422	155.90	0.085	0.113	0.163	0.424	1.445	197.63

Table 2: $\lambda_2(w_1)\lambda_1(w_2)$, Quantiles of the Supremum of Absolute Difference

	$\widehat{\psi}^{CV}(z)$						$\widehat{\psi}_M^{CV}(z)$					
$m - n_m$	0.25	0.50	0.75	0.90	0.95	0.99	0.25	0.50	0.75	0.90	0.95	0.99
i.i.d.	0.080	0.119	0.154	0.230	0.328	0.905	0.066	0.099	0.133	0.159	0.185	0.330
20-50	0.093	0.130	0.161	0.291	0.417	6.193	0.076	0.110	0.138	0.166	0.212	1.381
10-100	0.101	0.131	0.161	0.316	0.477	13.758	0.082	0.112	0.143	0.175	0.250	3.052
5-200	0.095	0.129	0.174	0.471	1.593	18.388	0.076	0.108	0.145	0.203	0.380	4.528
2-500	0.098	0.135	0.222	2.338	8.916	57.119	0.086	0.124	0.161	0.413	0.992	17.760
1-1000	0.117	0.150	0.958	8.144	22.510	122.78	0.099	0.137	0.209	0.761	2.498	154.64
	$\widehat{\psi}^{DF}(z)$						$\widehat{\psi}_M^{DF}(z)$					
	0.25	0.50	0.75	0.90	0.95	0.99	0.25	0.50	0.75	0.90	0.95	0.99
i.i.d.	0.082	0.109	0.134	0.159	0.191	8.839	0.074	0.100	0.124	0.140	0.151	1.398
20-50	0.082	0.109	0.130	0.149	0.166	0.443	0.076	0.103	0.123	0.142	0.154	0.212
10-100	0.083	0.108	0.138	0.162	0.213	0.853	0.078	0.101	0.126	0.145	0.156	0.299
5-200	0.085	0.113	0.135	0.164	0.264	22.293	0.080	0.105	0.128	0.150	0.168	5.321
2-500	0.094	0.122	0.153	0.355	5.183	22.540	0.086	0.115	0.142	0.188	0.564	3.353
1-1000	0.099	0.126	0.180	3.662	8.065	45.783	0.092	0.116	0.149	0.488	0.987	9.334

Table 3: $\lambda_2(w_1)\lambda_2(w_2)$, Quantiles of the Supremum of Absolute Difference

smooth out extreme points in $\widehat{\psi}^{CV}(z)$ and $\widehat{\psi}^{DF}(z)$ leading to better results in the supremum metric. Second, $\widehat{\psi}^{DF}(z)$ and $\widehat{\psi}_M^{DF}(z)$ improve over the corresponding cross validation estimators for most quantiles. The gains can be large. This is not only a result of dependence. Even in the *i.i.d.* specifications the performance is superior. However, when m is small the adjustment can lead to poorer performance at the largest quantiles. On balance, the DF estimators produce substantial improvements. Based on these results, our preferred estimation procedures are $\widehat{\psi}^{DF}(z)$ and $\widehat{\psi}_M^{DF}(z)$. Quantiles for the bandwidths are presented in the Online Supplement. Bandwidths chosen with (25) are on average larger than cross validation bandwidths. The superior performance of these estimators is in keeping with the theory derived above. Asymptotic results show that bandwidths should converge to zero more slowly in the presence of dependence.

The definition of df used above is ad hoc and others are possible. In addition, the form of (25) was designed for regression models. In spite of this, the adjusted estimators outperform cross validation by a wide margin. We see these results as suggestive that more theoretically grounded bandwidth selection procedures along similar lines should be developed. This would involve using the asymptotic results derived above. A full theoretical analysis of data driven approaches to bandwidth selection in the dependence case is beyond the scope of this paper.

5 Estimating Trade Intensity

With the rise of high-frequency trading, financial institutions buying or selling large orders frequently use algorithms which break up these trades into a large number of smaller transactions in order to minimize price impact. These algorithms influence the limit-order-book by changing the number of bids and asks. In this environment, one of the most important variables for describing the arrival rate of trades is depth imbalance. Depth imbalance is simply the number of ask orders on the limit-order-book minus the number of bid orders. When there is pressure to sell contracts, depth imbalance will become more positive. Similarly, when there is pressure to buy contracts it becomes more negative.

In an empirical application of our results, we estimate the hazard rate of trade arrival for futures contracts. The hazard rate of silver futures is estimated using depth imbalance of both gold and

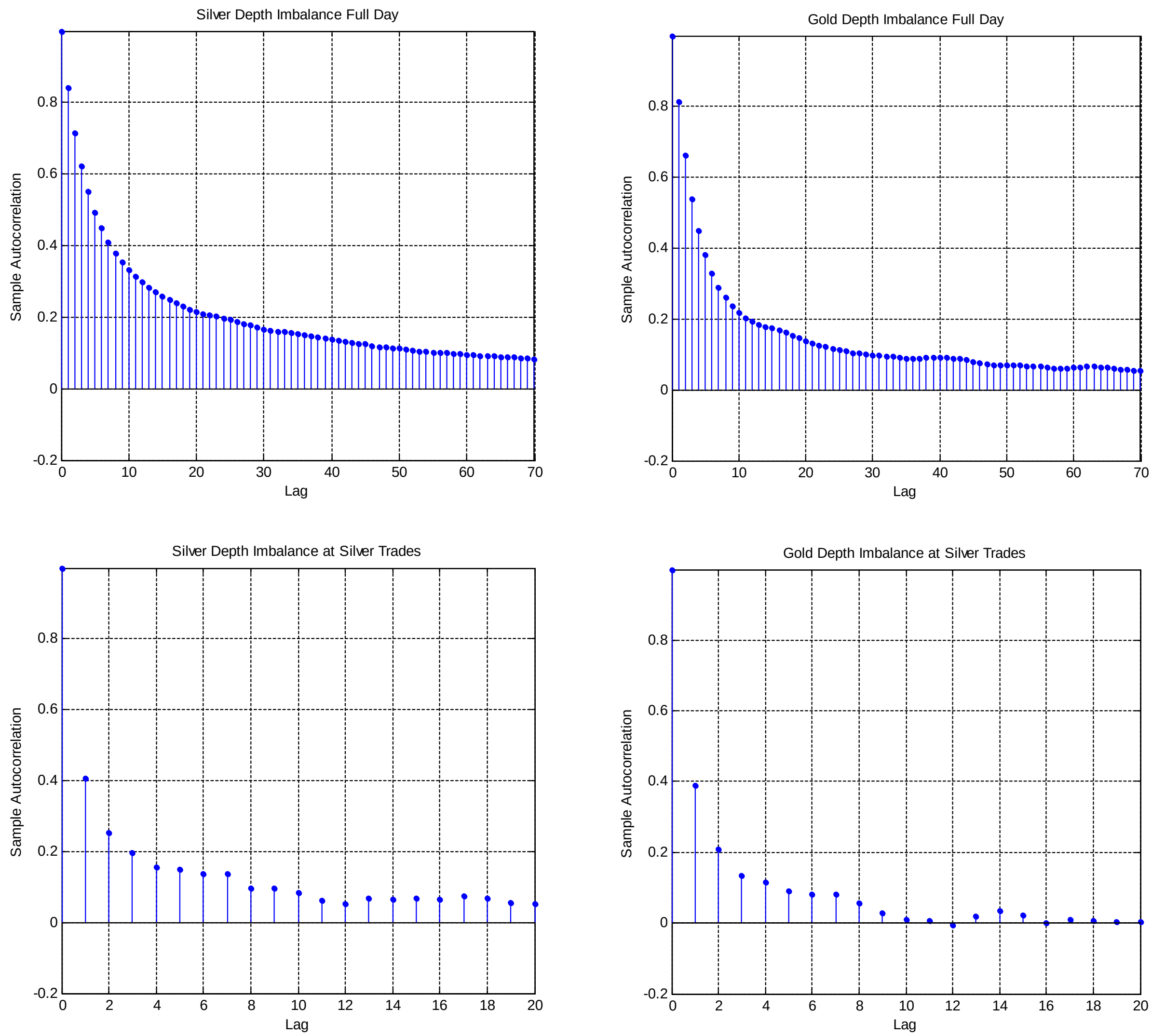

Figure 2: Autocorrelations for standardized log depth imbalance of gold and silver futures; all observations and at silver trades

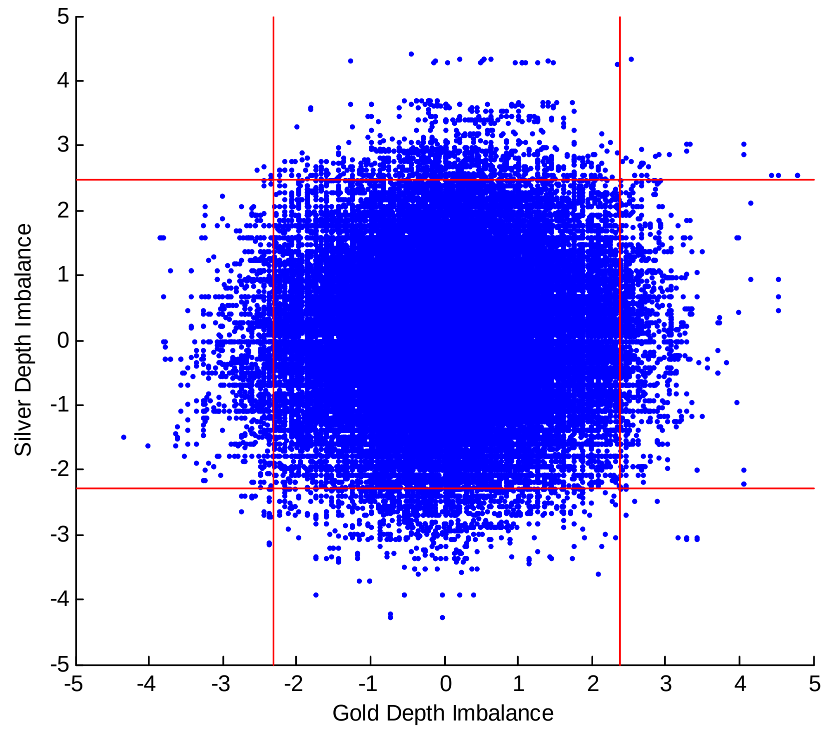

Figure 3: Gold and silver futures' standardized log depth imbalance throughout Dec 11, 2012

silver futures as covariates. Because gold and silver are related by economic fundamentals, the goal is to describe how trading pressure in a related market influences trading in silver futures.

For estimation we use high-frequency data taken from a single day; Dec 11, 2012. The futures examined are traded on the COMEX exchange and the data comes from the Thomson Reuters Tick History (TRTH) database. The hours 7:25-12:25 are used as they are the most liquid, corresponding to open outcry trading. Trades happening in the same millisecond are aggregated and treated as one trade. After this aggregation there are over 6,000 trade events. In this context, G^i corresponds to the end of the previous duration τ_{i-1} (i.e., when a trade happens, the next duration starts). This imposes additional structure which is not exploited in our theoretical results. How best to leverage this structure is left to future research. The results that follow preform well and fit with the simulation evidence. This should ease any concerns. The upper bound T is the length of the entire day. Clearly, no duration can exceed this.

Our definition of depth imbalance is log number of asks minus log number of bids. These variables are standardized by their empirical standard deviation. Logs are taken to attenuate the effect of outliers. Depth imbalance changes frequently between trades. For silver it changes over 50,000 times within the day and for gold over 66,000 times. Figure 2 plots two sets of estimated autocorrelations for depth imbalance. The first set uses as observation times whenever either of the two depth imbalance variables change their value. The second set uses as observation times whenever silver futures trade. There is positive autocorrelation for a substantial number of lags in both gold and silver. These covariates are clearly not *i.i.d.* across trades. Figure 3 plots all observations of gold and silver depth imbalances throughout the day. The vertical and horizontal lines mark the 1% and 99% quantiles for gold and silver respectively. The box inside these lines is where the hazard rate is estimated.

Figure 4 shows our baseline estimate $\widehat{\psi}^{DF}(z)$. The time scale used is in seconds. The bandwidth chosen is 1.7, whereas the cross validation bandwidth is 1.1. This is in keeping with the simulation evidence where the adjusted estimators tend to have higher bandwidths. It is clearly seen that depth imbalance for gold impacts the trading rate in silver futures. When both covariates are large and have the same sign, the estimated hazard visibly increases. When depth imbalance for silver is negative and gold is positive, there is a similar increase in the hazard rate. This increase is smaller when pressure is in opposite directions. Finally, when depth imbalance for silver is positive and gold is negative, the trading rate is low.

The multiplicative estimator $\widehat{\psi}_M^{DF}(z)$ was also computed and is displayed in Figure 5. The shape has substantially changed from the previous case. Now the highest points happen when depth imbalance is positive for gold. The spike when both covariates are negative has been smoothed out.

It is not clear the hazard rate in this situation is multiplicative. Given the amount of data available and the fact that we are estimating a two dimensional hazard, it is reasonable to conclude $\widehat{\psi}^{DF}(z)$ is reliable. In cases with a higher number of variables or less data, the multiplicative estimator will be essential. If for no other reason, this is true because of the curse of dimensionality. More variables would exacerbate the estimation issues raised when covariates are dependent.

There are many other covariates of interest in analyzing trade intensity. Depth imbalance sums

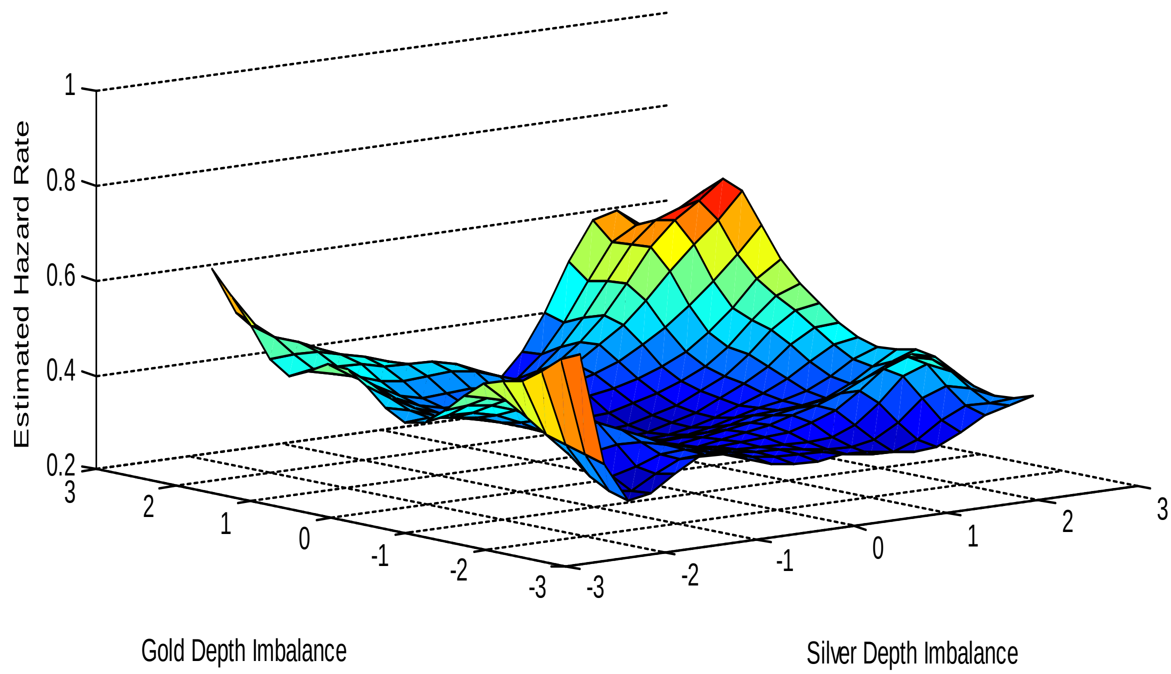

Figure 4: Hazard estimate, $\psi^{DF}(z)$

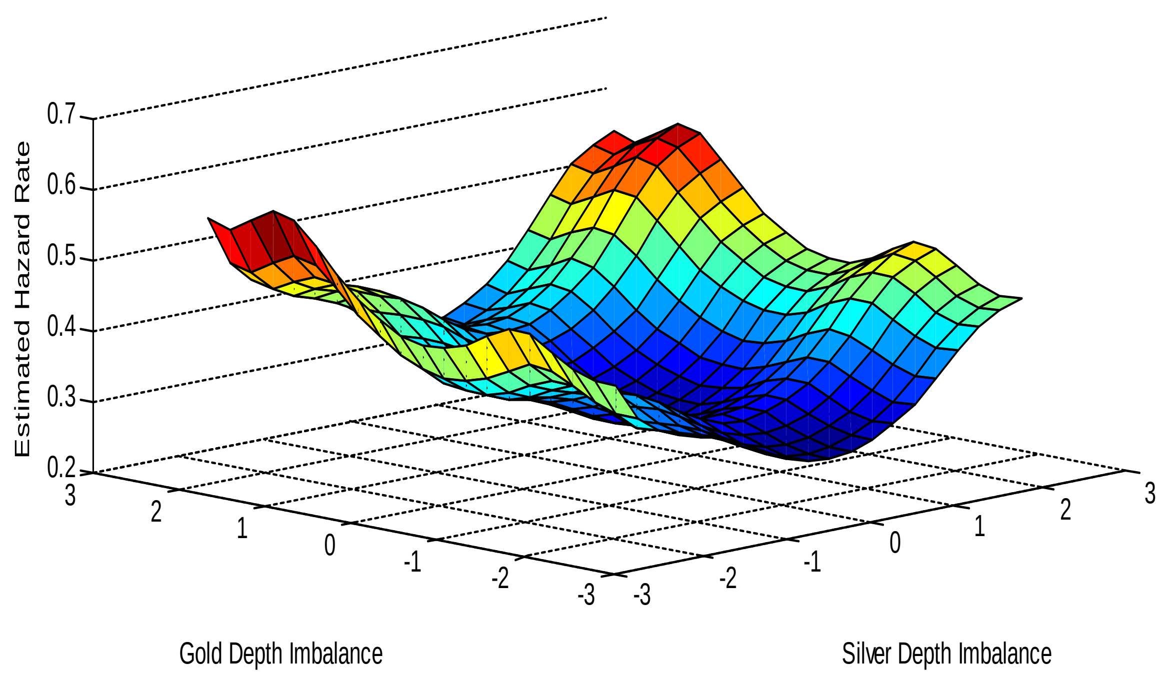

Figure 5: Multiplicative hazard estimate, $\psi_M^{DF}(z)$

up the number of contracts that can be traded at any price. The limit-order-book contains more information than this measure. It may be possible to use the number of bids and asks at various prices to derive better covariates describing the hazard rate. It is also possible to incorporate depth imbalance from any number of other covariates. The results shown here could be extended in many different ways. We leave considerations for higher dimensional cases to future work.

6 Conclusion

In this paper, asymptotic results were derived for nonparametric kernel estimation of hazard functions when observations have dependent covariates. The results show that the basic approach taken in Nielsen and Linton (1995) and Linton et al. (2003) can be extended to the dependent case. Asymptotics require a number of restrictions to bandwidth rates. The required rates are determined by the level of dependence. Compared to the *i.i.d.* case, the rate at which the bandwidth converges to zero must be slowed down. This adjustment changes confidence intervals, uniform convergence rates and asymptotic variance estimates.

Several simulations were conducted examining the finite sample performance of estimators when macroeconomic covariates influence hazard rates. Our results indicate that greater dependence results in poorer estimates in at least two different metrics. Using the multiplicative structure of a hazard function in estimation produces significantly better results. This structure helps mitigate the negative effects of dependence. The simulations suggest that more sophisticated bandwidth selection procedures that account for dependence can improve performance. Further developing these methods is a natural topic for future research.

The methods developed in this paper were applied to high-frequency financial data. The hazard rate of trade arrival for silver futures contracts was estimated. Depth imbalance of both gold and silver futures were used as covariates describing this arrival rate. The estimates show how conditions in a related market impact the hazard. When both assets have a large depth imbalance the hazard rate is highest. Extension of these methods to other high-frequency trading contexts is another natural topic for future research.

7 Appendix

The next lemma gives a preliminary result to establishing mixing conditions on the summands in (6)-(9). The idea is simple: if we known $(\eta_i, X^i(s), W(s+G^i))$, then we know functions of these variables. The Lemma just states this rigorously. The proof is tedious and omitted.

Lemma 8 *Let $f(x,w,t)$ be a bounded continuous function on $\mathbb{R}^{d+j+1}$. Define*

$$F^1 = \int_0^T f(X_s, W_{s+G^i}, s)\, ds,$$

and

$$F^2 = \int_0^T f(X_s, W_{s+G^i}, s)\, Y_s^i ds.$$

Then

$$\sigma\left(F^1\right), \sigma\left(F^2\right) \subset \sigma\left\{\widetilde{\eta}_i, \eta_i, X^i(s), W\left(u+G^i\right)\Big|\, 0 \le s, u \le T\right\}. \tag{26}$$

Lemma 8 allows us to convert joint mixing conditions on the underlying covariate processes into mixing conditions on the integrals we use for estimation. This will facilitate verification of the conditions in Proposition 5. Each of the arrays used in Proposition 5 will inherit the joint mixing conditions of the underlying processes Z_t^i, where mixing is in the dimension $i = 1, \ldots, n$. Therefore, (6)-(9) can all be established with a single joint mixing condition on the covariates. Directly assuming mixing conditions for the appropriate arrays will also verify (6)-(9).

Proof (Theorem 1). Once the martingale CLT in Lemma 0.2 (from the Online Supplement) is verified and the assumptions (6)-(9) hold, the restriction of NL's Theorem 1 to *i.i.d.* observations may be relaxed. Under (B2)-(B3), the expectations in the assumptions (6)-(9) are the same as in the *i.i.d.* case. Therefore, these expectations converge to the same values as those derived in NL's Theorem 1. NL use Chebyshev's inequality to show (6)-(9). However, once Lemma 0.2 is available, as long as (6)-(9) hold, the results (10)-(12) continue to hold using the same proof as NL's Theorem 1. In addition, the fact that

$$\sum_{i=1}^{n} \int_0^t \frac{b^{d+j+1}}{n} K_b^2\left(x - Z_s^i\right) dM_s^i,$$

is a continuous-time martingale in t justifies the use of Lenglart's inequality in the proof (see Shorack and Wellner (1986, p.892-893)). This holds in the dependent case using the information structure outlined in the Online Supplement. Finally, the result (11) is modified from a similar result in NL by allowing for r times continuous differentiability of ψ and e. This allows for Taylor expansions with more terms. The term b^{-r} replaces the term b^{-2} which is found in NL. This type of modification is done in LNV. Details are omitted. ■

Proof (Proposition 5). This holds by Lemma 8 and an application of the Bernstein inequality in Bosq's (1998) Theorem 1.3 (2). The basic structure of this proof will be applied several other times in the sequel. We show this for $\widehat{e}(z)$. Similar proofs hold for the other cases. We first consider the cases (6)-(8). (9) requires a different argument. Let

$$\widehat{e}(z) = \frac{1}{n}\sum_{i=1}^{n} \zeta_{n,i} - \mathbb{E}\left[\zeta_{n,i}\right],$$

where

$$\zeta_{n,i} = \int_0^T K_b\left(z - Z_s^i\right) Y_s^i ds,$$

and we define

$$\zeta_{n,i}^c = \zeta_{n,i} - \mathbb{E}\left[\zeta_{n,i}\right].$$

This array is α-mixing with bounded elements given our assumptions. The terms $\zeta_{n,i}^c$ are bounded by $b^{-(d+j+1)}C$. In another case, we have a bound $n^{-1}b^{-2(d+j+1)}C$ for some constant C. Under our assumptions, the first case is stricter because $nb^{(d+j+1)} \to \infty$. Only results for this case are presented, while the other case easily follows. We must provide a bound for the variance of the sums

$$\sum_{i=1}^{n^*} \zeta_{n,i}^c,$$

where n^* is chosen below and $n^* \neq n$.

$$var\left(\sum_{i=1}^{n^*} \zeta_{n,i}^c\right) \leq \sum_{i=1}^{n^*}\sum_{j=1}^{n^*} \mathbb{E}\left\{\zeta_{n,i}\zeta_{n,j}\right\}.$$

By an application of the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}\left\{\zeta_{n,i}\zeta_{n,j}\right\} &\leq \left\{\int\int_0^T K_b^2\left(z-v\right)e\left(v\right)dv\right\}^{1/2}, \\ &\times\left\{\int\int_0^T K_b^2\left(z-v\right)e\left(v\right)dv\right\}^{1/2}.\end{aligned}$$

By a change of variable argument, this is bounded by $b^{-(d+j+1)}O\left(1\right)$. Therefore:

$$var\left(\sum_{i=1}^{n^*}\zeta_{n,i}^c\right) \leq \left(n^*\right)^2 b^{-(d+j+1)}O\left(1\right). \tag{27}$$

This will be used below. Now we define the parameters used in the Bernstein inequality. These are simplified from the conditions needed in Bosq (1998) by ignoring constants and rounding which will not matter for the final asymptotic rate. Throughout the sequel, we ignore constants and rounding which are irrelevant asymptotically. Define the following values used in Bosq's (1998) Theorem 1.3 (2):

$$\begin{aligned}\epsilon &= \lambda, \\ p &= n^{1-\widetilde{\gamma}}, \\ q &= n^{\widetilde{\gamma}},\end{aligned}$$

where $0<\widetilde{\gamma}<1$ and λ is a constant. Using these values and the bound (27) on the variance, we can define the following term which is also used in Bosq (1998):

$$\begin{aligned}v^2\left(q\right) &= 2p^{-2}\left(p\right)^2 b^{-(d+j+1)}O\left(1\right)+\frac{1}{2}b^{-(d+j+1)}C\lambda, \\ &= b^{-(d+j+1)}O\left(1\right)+b^{-(d+j+1)}C.\end{aligned}$$

Applying the inequality gives

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\zeta_{n,i}^c \geq \lambda\right), \\ \leq\ & C_1\exp\left[-C_2\lambda^2 n^{\widetilde{\gamma}}b^{(d+j+1)}\right]+C_3 n^{\widetilde{\gamma}}b^{-(d+j+1)/2}\alpha\left(n^{(1-\widetilde{\gamma})}\right),\end{aligned} \tag{28}$$

where $\alpha(\cdot)$ is the alpha mixing coefficients from the array. By assumption, $n^{\widetilde{\gamma}} b^{(d+j+1)} \to \infty$ and this makes the first term converge to zero. For (M1) and (M2), the second term is

$$n^{-(1-\widetilde{\gamma})\xi+\widetilde{\gamma}} b^{-(d+j+1)/2},$$

or

$$C_3 b^{-(d+j+1)/2} n^{\widetilde{\gamma}} \exp\left(n^{1-\widetilde{\gamma}}\right),$$

respectively. (M1) and (M2) each assume the corresponding term converges to zero. Therefore:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \zeta_{n,i}^{c} \geq \lambda\right) \to 0,$$

and the result follows.

We now adjust the above arguments to get a b^r convergence rate for (9). Define:

$$\begin{aligned}\sum_{i=1}^{n} X_{n,i} &= \sum_{i=1}^{n} \int_0^T K_b\left(z - Z_s^i\right)\left[\psi\left(Z_s^i\right) - \psi(z)\right] Y_s^i ds, \\ &\quad - \mathbb{E}\left\{\sum_{i=1}^{n} \int_0^T K_b\left(z - Z_s^i\right)\left[\psi\left(Z_s^i\right) - \psi(z)\right] Y_s^i ds\right\}.\end{aligned}$$

We need to bound the variance of sums as was done above:

$$\begin{aligned}& var\left\{\sum_{i=1}^{n^*} \int_0^T K_b\left(z - Z_s^i\right)\left[\psi\left(Z_s^i\right) - \psi(z)\right] Y_s^i ds\right\}, \\ \leq\ & \sum_{j=1}^{n^*}\sum_{i=1}^{n^*} \mathbb{E}\left\{\begin{array}{c} \int_0^T K_b\left(z - Z_s^i\right)\left[\psi\left(Z_s^i\right) - \psi(z)\right] Y_s^i ds \\ \times \int_0^T K_b\left(z - Z_s^i\right)\left[\psi\left(Z_s^i\right) - \psi(z)\right] Y_s^i ds \end{array}\right\}. \qquad (29)\end{aligned}$$

Again, by an application of the Cauchy-Schwarz inequality, we get:

$$\begin{aligned}
& \mathbb{E}\left\{\begin{array}{c} \int_0^T K_b\left(z-Z_s^i\right)\left[\psi\left(Z_s^i\right)-\psi\left(z\right)\right]Y_s^i ds \\ \times\int_0^T K_b\left(z-Z_s^i\right)\left[\psi\left(Z_s^i\right)-\psi\left(z\right)\right]Y_s^i ds \end{array}\right\}, \\
\leq & \left\{\int\int_0^T \left\{K_b\left(z-v\right)\left[\psi\left(v\right)-\psi\left(z\right)\right]\right\}^2 e\left(v\right)dv\right\}^{1/2}, \qquad (30) \\
& \times\left\{\int\int_0^T \left\{K_b\left(z-v\right)\left[\psi\left(v\right)-\psi\left(z\right)\right]\right\}^2 e\left(v\right)dv\right\}^{1/2}.
\end{aligned}$$

By a Taylor expansion and change of variables (as in the LN proof of their Theorem 1), we get that (30) is $b^2/b^{d_z}O(1)$. This implies that (29) is $(n^*)^2 b^2/b^{d_z}O(1)$. Here, we choose the following values used in Bosq's (1998) Theorem 1.3 (2):

$$\begin{aligned}
\epsilon &= b^r\lambda, \\
p &= n^{1-\widetilde{\gamma}}, \\
q &= n^{\widetilde{\gamma}},
\end{aligned}$$

where $0<\widetilde{\gamma}<1$ and λ is a constant. This and our new bound on the variance gives us:

$$\begin{aligned}
v^2\left(q\right) &= 2p^{-2}\left(p\right)^2 b^2/b^{(d+j+1)}O\left(1\right)+\frac{1}{2}b^{-(d+j+1)}b^r\lambda, \\
&= b^2/b^{(d+j+1)}O\left(1\right)+b^{-d+j+1}b^r\lambda C.
\end{aligned}$$

Putting this all together in the exponential inequality, we get

$$\begin{aligned}
& \mathbb{P}\left(\left|\sum_{i=1}^n X_{n,i}\right|\geq n\left(b^r\lambda\right)\right), \qquad (31) \\
\leq & \; 4\exp\left(\frac{-b^{2r}\lambda^2 n^{\widetilde{\gamma}}}{8\left\{2b^2/b^{(d+j+1)}O\left(1\right)+C\frac{1}{b^{(d+j+1)}}b^r\lambda\right\}}\right), \\
& +22\left(1+\frac{4b^{-(d+j+1)}}{b^r\lambda}\right)^{1/2} n^{\widetilde{\gamma}}\alpha\left(n^{1-\widetilde{\gamma}}\right),
\end{aligned}$$

where $\alpha(\cdot)$ is the α-mixing rate of the array. We now bound the first term:

$$
\begin{aligned}
&4\exp\left(\frac{-b^{2r}\lambda^2 n^{\widetilde{\gamma}}}{8\left\{2b^2/b^{(d+j+1)}O(1)+C\frac{1}{b^{(d+j+1)}}b^r\lambda\right\}}\right),\\
\leq\ &4\exp\left(\frac{-b^{2r}\lambda^2 n^{\widetilde{\gamma}}}{8*2*\max\left\{2b^2/b^{(d+j+1)}O(1),C\frac{1}{b^{(d+j+1)}}b^r\lambda\right\}}\right).
\end{aligned}
$$

For this to hold, the following two things must hold:

$$b^{d+j+1+2r-2}n^{\gamma}\to\infty,$$

$$b^{d+j+1+r}n^{\gamma}\to\infty.$$

Because $r\geq 2$, the first condition implies the second. By our assumptions, $b^{d+j+1+2r-2}n^{\gamma}\to\infty$.

We need $q=n^{\widetilde{\gamma}}$ with $\widetilde{\gamma}<1$ because the mixing term requires this. Here we bound the second term in the exponential inequality given above:

$$
\begin{aligned}
&22\left(1+\frac{4b^{-(d+j+1)}}{b^r\lambda}\right)^{1/2} n^{\widetilde{\gamma}}\alpha\left(n^{1-\widetilde{\gamma}}\right),\\
\leq\ &22\left(Cb^{-(d+j+1+r)/2}\right)n^{\widetilde{\gamma}}\alpha\left(n^{1-\widetilde{\gamma}}\right)\to 0.
\end{aligned}
$$

For (M1) and (M2), this term is

$$n^{-(1-\widetilde{\gamma})\xi+\widetilde{\gamma}}b^{-(d+j+1+r)/2},$$

and

$$C_3 b^{-(d+j+1+r)/2}n^{\widetilde{\gamma}}\exp\left(n^{1-\widetilde{\gamma}}\right),$$

respectively. (M1) and (M2) each assume the corresponding term converges to zero. Therefore:

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}X_{n,i}\right|\geq n\left(b^r\lambda\right)\right)\to 0,$$

and the result follows. ■

Proof (Theorem 6). See Online Supplement. ■

Here the required bandwidth conditions for Theorem 7 are presented. In what follows, we drop our practice of writing out the conditions for the specific mixing cases of exponential and polynomial decay. The results are written for arbitrary mixing rates. Both the exponential and polynomial cases can be recovered from the stated conditions.

(MB) The system of σ-fields $\mathcal{H}_l^m$ has an α-mixing rate $\alpha(n)$ such that, for some $0 < \gamma < 1$,

$$\begin{aligned} b^{(d+j+1)} n^{\gamma} &\to \infty, \\ \frac{n^{9/4\gamma}}{b^{(d+j+1)/4}\sqrt{\log n}} \alpha\left(n^{1-\gamma}\right) &\to 0. \end{aligned} \tag{32}$$

$$\begin{aligned} n^{1/2} b^{2r-(d+j+1)/2} &\to 0, \\ n^{1/2-\gamma/2} \log^{1/2}(n)\, b^{r-(d+j+1)} &\to 0, \\ n^{1/2-\gamma} \log(n)\, b^{-3(d+j+1)/2} &\to 0. \end{aligned} \tag{33}$$

$$\begin{aligned} \frac{\log n}{n^{\gamma} b^{2(d+j+1)+1}} &\to 0, \\ \frac{n^{(13/4)\gamma}}{b^{5(d+j+1)/4} \log^{3/2} n} \alpha\left(n^{1-\gamma}\right) &\to 0. \end{aligned} \tag{34}$$

(MC) The system of σ-fields $\mathcal{H}_l^m$ has an α-mixing rate $\alpha(n)$ such that, for some $0 < \underline{\gamma} < 1$, $0 < \overline{\gamma} < 1$

$$\begin{aligned} n^{\overline{\gamma}} b^2 &\to \infty, \\ \frac{n^{\overline{\gamma}}}{b^{1/2}} \alpha\left(n^{1-\overline{\gamma}}\right) &\to 0. \end{aligned} \tag{35}$$

$$\begin{aligned} n^{\underline{\gamma}-1/2} b^{1/2} &\to \infty, \\ \frac{n^{1/4+\underline{\gamma}}}{b^{1/4}} \alpha\left(n^{1-\underline{\gamma}}\right) &\to 0. \end{aligned} \tag{36}$$

The conditions are divided into groups which are used in specific parts of the proof. In particular, (32) implies the conditions (M1B) or (M2B) depending on the assumed strength of the mixing conditions. (MB) and (MC) imply certain restrictions on the bandwidth choice needed for Theorem 7 to hold. (33) implies that $\gamma > 1/2$ and $r > d + j + 1$. This fact shows how smoothness of the hazard function helps facilitate the results. In what follows, we will assume the bandwidth is chosen as $b = Cn^{-1/(2r+1)}$. If this is the case, as long as we choose $\gamma > 1/2$, $\overline{\gamma} > 1/2$ and $\underline{\gamma} > 1/2$, then most of the conditions in (MB) and (MC) will hold with a large enough r. The exceptions are those involving the mixing rate $\alpha(\cdot)$. In the exponential mixing case these will also hold with no further assumptions. When mixing has polynomial decay these restrictions will depend on the coefficient ξ.

Next, we present the proof of Theorem 7. We follow the proof of Theorem 1 from LNV, focusing on the modifications needed for dependence. First, define $\widehat{e}_{-i}(z)$ as $\widehat{e}(z)$ with observation i excluded. Define $K_{-j}(z)$ as $K(z)$ with the jth kernel excluded. In general, a subscript of $-i$ or $-j$ indicates that the ith or jth element has been excluded from the term. This notation is clarified when needed. Presented here is all the (extensive) notation:

$$\begin{aligned}
\left(\widehat{\psi} - \psi\right)(z) &= \frac{(V_n(z) + B_n(z))}{\widehat{e}(z)}, \\
V_n(z) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^T K_b\left(z - Z_s^i\right) dM_s^i, \\
B_n(z) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^T K_b\left(z - Z_s^i\right)\left[\psi\left(Z_s^i\right) - \psi(z)\right] Y_s^i ds,
\end{aligned}$$

$$\begin{aligned}
\left(\widehat{\psi}_{Q_{-j}} - \psi_{Q_{-j}}\right)(z_j) &= V_{Q_{-j}}(z_j) + B_{Q_{-j}}(z_j), \\
V_{Q_{-j}}(z_j) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^T H_i^{(n)}(z_j, s)\, dM_s^i, \\
H_i^{(n)}(z_j, s) &= \int_{I_{-j}} \frac{K_b\left(z - Z_s^i\right)}{\widehat{e}(z)} dQ_{-j}(z_{-j}), \\
B_{Q_{-j}}(z_j) &= \int_{I_{-j}} \frac{B_n(z)}{\widehat{e}(z)} dQ_{-j}(z_{-j}),
\end{aligned}$$

$$
\begin{aligned}
\widetilde{h}_i^{(n)}(z_j,s) &= \int_{I_{-j}} \frac{W_{ni}(z,s)}{e(z)} dQ_{-j}(z_{-j}), \\
\widehat{h}_i^{(n)}(z_j,s) &= \int_{I_{-j}} \frac{W_{ni}(z,s)}{\widehat{e}(z)} dQ_{-j}(z_{-j}), \\
\ddot{h}_i^{(n)}(z_j,s) &= \int_{I_{-j}} \frac{W_{ni}(z,s)}{\widehat{e}_{-i}(z)} dQ_{-j}(z_{-j}), \\
\overline{h}_i^{(n)}(z_j,s) &= \frac{1}{(nb)^{1/2}} k\left(\frac{z_j - Z_j^i(s)}{b}\right) \frac{q_{-j}^2\left(Z_{-j}^i(s)\right)}{e\left(z_j, Z_{-j}^i(s)\right)}, \\
W_{ni}(z,s) &= \frac{b^{1/2}}{n^{1/2}} K_b\left(z - Z_s^i\right),
\end{aligned}
$$

$$
(nb)^{1/2}\widetilde{V}_{Q_{-j}}(z_j) = \sum_{i=1}^n \int_0^T \widetilde{h}_i^{(n)}(z_j,s)\, dM^i(s).
$$

Proof (Theorem 7). We wish to show

$$
(nb)^{1/2}\left(\widehat{\psi}_{Q_{-j}}(x) - \psi_{Q_{-j}}(x)\right)(x_j) \Rightarrow N\left[m_j(z_j), v_j(z_j)\right].
$$

Note the decomposition

$$
(nb)^{1/2}\left(\widehat{\psi}_{Q_{-j}}(x) - \psi_{Q_{-j}}(x)\right)(x_j) = (nb)^{1/2}\left(V_{Q_{-j}}(z_j) + B_{Q_{-j}}(z_j)\right).
$$

The proof is in three steps. First, prove

$$
(nb)^{1/2}\widetilde{V}_{Q_{-j}}(z_j) \Rightarrow N\left(0, v_j(z_j)\right).
$$

Second, prove

$$
(nb)^{1/2}\left\{\widetilde{V}_{Q_{-j}}(z_j) - V_{Q_{-j}}(z_j)\right\} \to^p 0.
$$

Finally, prove

$$
(nb)^{1/2} B_{Q_{-j}}(z_j) \to^p m_j(z_j).
$$

The first step follows from Lemma 0.6 in the Online Supplement. For the second step, we note that:

$$\begin{aligned} & (nb)^{1/2}\left|\widetilde{V}_{Q_{-j}}(z_j)-V_{Q_{-j}}(z_j)\right|, \\ \leq & \left|\sum_{i=1}^{n}\int_0^T \widehat{h}_i^{(n)}(z_j,s)\,dM_s^i-\sum_{i=1}^{n}\int_0^T \ddot{h}_i^{(n)}(z_j,s)\,dM_s^i\right|, \\ & +\left|\sum_{i=1}^{n}\int_0^T \ddot{h}_i^{(n)}(z_j,s)\,dM_s^i-\sum_{i=1}^{n}\int_0^T \widetilde{h}_i^{(n)}(z_j,s)\,dM_s^i\right|. \end{aligned}$$

Each of these components converge to zero in probability by Lemmas 0.7 and 0.8 from the Online Supplement. This implies $(nb)^{1/2}V_{Q_{-j}}(z_j)\Rightarrow N(0,v_j(z))$. The third step is proven in Lemma 0.9 in the Online Supplement. (21) and (22) follow directly from these results.

Finally, we need to prove (23) and (24). As in LNV, we need to analyze

$$cov\left[\sum_{i=1}^{n}\int_0^T \widetilde{h}_i^{(n)}(z_j,s)\,dM_s^i,\sum_{i=1}^{n}\int_0^T \widetilde{h}_i^{(n)}(z_k,s)\,dM_s^i\right]. \tag{37}$$

Note

$$\mathbb{E}\left[\sum_{i=1}^{n}\int_0^T \widetilde{h}_i^{(n)}(z_j,s)\,dM_s^i\right]=0,$$

by linearity and the fact that elements of the sum are mean zero martingales (see Fleming and Harrington's (1991) Theorem 2.4.4). Recall from the Online Supplement, $\left\langle M^i,M^l\right\rangle_s=0$ for $i\neq l$. If follows again from Fleming and Harrington's (1991) Theorem 2.4.4 that for $i\neq l$,

$$\begin{aligned} & \mathbb{E}\left[\int_0^T \widetilde{h}_i^{(n)}(z_j,s)\,dM_s^i\times\int_0^T \widetilde{h}_l^{(n)}(z_k,s)\,dM_s^l\right], \\ = & \mathbb{E}\left[\int_0^T \widetilde{h}_i^{(n)}(z_j,s)\,\widetilde{h}_l^{(n)}(z_k,s)\,d\left\langle M^i,M^l\right\rangle_s\right], \\ = & 0. \end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\left\{\sum_{i=1}^{n}\int_{0}^{T}\widetilde{h}_{i}^{(n)}\left(z_{j},s\right)dM_{s}^{i}\right\}\left\{\sum_{i=1}^{n}\int_{0}^{T}\widetilde{h}_{i}^{(n)}\left(z_{k},s\right)dM_{s}^{i}\right\}\right], \\
= &\sum_{i=1}^{n}\mathbb{E}\left\{\int_{0}^{T}\left[\widetilde{h}_{i}^{(n)}\left(z_{j},s\right)\right]^{2}d\Lambda_{s}^{i}\right\}. \qquad (38)
\end{aligned}$$

Under our assumptions, (38) is the same in the dependent and independent cases. So (37) is the same in both the dependent and independent cases. Therefore, the proof given in LNV still holds in the dependent situation. This proves (23) and (24) in the dependent case. ■

References

[1] Andrews, D.W.K., 2005. Cross-section regression with common shocks. Econometrica 73 (5), 1551-1585.

[2] Baltagi, B.H., Egger, P., Pfaffermayr, M., 2013. A generalized spatial panel data model with random effects. Econometric Reviews, 32, 650-685.

[3] Baltagi, B.H., Fingleton, B., Pirotte, A., 2014. Estimating and forecasting with a dynamic spatial panel data model. Oxford Bulletin of Economics and Statistics 76 (1), 112-138.

[4] Banerjee, S., Dey, D. K., 2005. Semiparametric proportional odds models for spatially correlated survival data. Lifetime Data Analysis 11, 175-191.

[5] Banerjee, S., Wall, M.M., Carlin, B.P., 2003. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. Biostatistics 4, 123-142.

[6] Bastos, L.S., Gamerman, D., 2006. Dynamic survival models with spatial frailty. Lifetime Data Analysis 12, 441-460.

[7] Bielecki, T.R., Rutkowski, M., 2004. Credit risk: modeling, valuation and hedging. Springer-Verlag, Berlin.

[8] Bijwaard, G., Ridder, G., 2002. Efficient estimation of the semi-parametric mixed proportional hazard model. Mimeo.

[9] Bijwaard, G., Ridder, G., 2009. A simple GMM estimator for the semi-parametric mixed proportional hazard. Mimeo

[10] Bosq, D., 1998. Nonparametric statistics for stochastic processes. Springer-Verlag, New York.

[11] Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18, 17-39.

[12] Chen, S., 2002. Rank estimation of transformation models. Econometrica 70, 1683-1696.

[13] Chen, X., Hansen, L.P., Carrasco, M., 2010. Nonlinearity and temporal dependence. Journal of Econometrics 155, 155-169.

[14] Chudik, A., Pesaran, M.H., 2013. Large panel data models with cross-sectional dependence: a survey. Working Paper.

[15] Chudik, A., Pesaran, M.H., Tosetti, E., 2011. Weak and strong cross-section dependence and estimation of large panels. Econometrics Journal 14, C45-C90.

[16] Daley, D.J., Vere-Jones, D., 2003. An introduction to the theory of point processes, vol. I: elementary theory and methods, 2nd ed. Springer Science+Business Media, New York.

[17] De Jong, R.M., Woutersen, T., 2011. Dynamic time series binary choice. Econometric Theory 27, 673-702.

[18] Elbers, C., Ridder, G., 1982. True and spurious duration dependence: the identifiability of the proportional hazard model. Review of Economic Studies 49, 402-409.

[19] Fleming, T.R., Harrington, D.P., 1991. Counting processes and survival analysis. Wiley, Hoboken.

[20] Frederiksen, A., Honoré, B.E., Hu, L., 2007. Discrete time duration models with group-level heterogeneity. Journal of Econometrics 141, 1014-1043.

[21] Gámiz, M.L., Janys, L., Martínez-Miranda, M.D., Nielsen, J.P., 2013. Bandwidth selection in marker dependent kernel hazard estimation. Working Paper.

[22] Hahn, J., 1994. The efficiency bound of the mixed proportional hazard model. Review of Economic Studies 61, 607-629.

[23] Hahn, J., Kuersteiner, G., 2011. Bias reduction for dynamic nonlinear panel models with fixed effects. Econometric Theory 27, 1152-1191.

[24] Hall, A.D., Hautsch, N., 2007. Modelling the buy and sell intensity in a limit order book market. Journal of Financial Markets 10, 249-286.

[25] Han, A.K., 1987. Non-parametric analysis of a generalized regression model, the maximum rank correlation estimator. Journal of Econometrics 35, 303-316.

[26] Han, A.K., Hausman, J.A., 1990. Flexible parametric estimation of duration and competing risk models. Journal of Applied Econometrics 1, 1-28.

[27] Hausman, J.A., Woutersen, T., 2014a. Estimating a semi-parametric duration model without specifying heterogeneity. Journal of Econometrics 178, 114-131.

[28] Hausman, J.A., Woutersen, T., 2014b. Estimating the derivative function and counterfactuals in duration models with heterogeneity. Econometric Reviews 33 (5-6), 472-496.

[29] Heckman, J.J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52, 271-320.

[30] Henderson, R., Shimakura, S., Gorst, D., 2002. Modeling spatial variation in leukemia survival data. Journal of the American Statistical Association 97, 965-972.

[31] Hennerfeind, A., Brezger, A., Fahrmeir, L., 2006. Geoadditive survival models. Journal of the American Statistical Association 101, 1065-1075.

[32] Honoré, B.E., 1990. Simple estimation of a duration model with unobserved heterogeneity. Econometrica 58, 453-473.

[33] Honoré, B.E., 1993a. Identification results for duration models with multiple spells. Review of Economic Studies 60, 241-246.

[34] Honoré, B.E., 1993b. Identification results for duration models with multiple spells or time-varying regressors. Northwestern Working Paper.

[35] Honoré, B.E., Hu, L., 2010. Estimation of a transformation model With truncation, interval observation and time-varying covariates. Econometric Journal 13, 127-144.

[36] Horowitz, J.L., 1999. Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. Econometrica 67, 1001-1028.

[37] Horowitz, J.L., Lee, S., 2004. Semiparametric estimation of a panel data proportional hazards model with fixed effects. Journal of Econometrics 119, 155-198.

[38] Hurvich, C.M., Simonoff, J.S., Tsai, C., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. Journal of the royal statistical society B 60 (2), 271-293.

[39] Ishwaran, H., 1996. Identifiability and rates of estimation for scale parameters in location mixture models. Annals of Statistics 24 , 1560-1571.

[40] Jacod, J., Shiryaev, A.N., 2003. Limit theorems for stochastic processes, 2nd ed. Springer-Verlag, Berlin.

[41] Jin, S., Su, L., 2013. A nonparametric poolability test for panel data models with cross section dependence. Econometric Reviews 32 (4), 469-512.

[42] Kapetanios, G., Pesaran, M.H., 2007. Alternative approaches to estimation and inference in large multifactor panels: small sample results with an application to modelling of asset returns, in: Phillips, G., Tzavalis, E. (Eds.), The refinement of econometric estimation and test procedures: finite sample and asymptotic analysis. Cambridge University Press, Cambridge, pp 282-318.

[43] Kapoor, M., Kelejian, H.H., Prucha, I.R., 2007. Panel data models with spatially correlated error components. Journal of Econometrics 140, 97-130.

[44] Kuersteiner, G.M., Prucha, I.R., 2013. Limit theory for panel data models with cross sectional dependence and sequential exogeneity. Journal of Econometrics 174, 107-126.

[45] Lancaster, T., 1990. The econometric analysis of transition data. Cambridge University Press, Cambridge.

[46] Lawson, A.B., Choi, J., Zhang, J., 2014. Prior choice in discrete latent modeling of spatially referenced cancer survival. Statistical Methods in Medical Research 23 (2), 183-200.

[47] Lee, L.F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154, 165-185.

[48] Lee, L.F., Yu, J., 2010. Some recent developments in spatial panel data models. Regional Science and Urban Economics 40, 255-271.

[49] Lee, L.F., Yu, J., 2014. Efficient GMM estimation of spatial dynamic panel data models with fixed effects. Journal of Econometrics 180, 174-197.

[50] Li, Y., Lin, X., 2005. Semiparametric normal transformation models for spatially correlated survival data. Harvard Biostatistics Working Paper.

[51] Li, Y., Ryan, L., 2002. Modeling spatial survival data using semiparametric frailty models. Biometrics 58, 287-297.

[52] Li, Y., Su, L., Xu, Y., 2015. A combined approach to the inference of conditional factor models. Journal of Business and Economic Statistics 33 (2), 203-220.

[53] Linton, O.B., Nielsen, J.P., Van de Geer, S., 2003. Estimating multiplicative and additive hazard functions by kernel methods. Annals of Statistics 31 (2), 464-492.

[54] Loader, C., 1999. Local regression and likelihood. Springer-Verlag, New York.

[55] Mammen, E., Martínez-Miranda, M.D., Nielsen, J.P., Sperlich, S., 2011. Do-validation for kernel density estimation. Journal of the American Statistical Association 106, 651-660.

[56] Martinussen, T., Scheike, T.H., 2010. Dynamic regression models for survival data. Springer Science+Business Media, New York.

[57] Meyer, B.D., 1990. Unemployment insurance and unemployment spells. Econometrica 58, 757-782.

[58] Nielsen, J.P., 1998. Marker dependent kernel hazard estimation from local linear estimation. Scandinavian Actuarial Journal 2, 113-124.

[59] Nielsen, J.P., Linton, O.B., 1995. Kernel estimation in a nonparametric marker dependent hazard model. Annals of Statistics 23 (5), 1735-1748.

[60] Nielsen, J.P., Tanggaard, C., 2001. Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics 28, 675-698.

[61] Parent, O., LeSage, J.P., 2012. Spatial dynamic panel data models with random effects. Regional Science and Urban Economics 42, 727-738.

[62] Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74 (4), 967-1012.

[63] Pesaran, M.H., Tosetti, E., 2011. Large panels with common factors and spatial correlation. Journal of Econometrics 161, 182-202.

[64] Poirier, D., Ruud, P. 1988. Probit with dependent observations. Review of Economic Studies 55, 593-614.

[65] Ridder, G., 1990. The non-parametric identification of generalized failure time models. Review of Economic Studies 57, 167-182.

[66] Ridder, G., Woutersen, T., 2003. The singularity of the information matrix of the mixed proportional hazard model. Econometrica 71, 1579-1589.

[67] Robinson, P.M., 1982. On the asymptotic properties of estimators of models containing LDV. Econometrica 50, 27-41.

[68] Shorack, G.R., Wellner, J.A., 1986. Empirical processes with applications to statistics. SIAM, Philadelphia.

[69] Stone, C.J., 1980. Optimal rates of convergence for nonparametric estimators. Annals of Statistics 8, 1348-1360.

[70] Su, L., 2012. Semiparametric GMM estimation of spatial autoregressive models. Journal of Econometrics 167, 543-560.

[71] Su, L., Jin, S. 2010. Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. Journal of Econometrics 157, 18-33.

[72] Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. Journal of Econometrics 169, 34-47.

[73] Su, L., Yang, Z., 2015. QML estimation of dynamic panel data models with spatial errors. Journal of Econometrics 185, 230-258.

[74] Tsybakov, A.B., 2009. Introduction to nonparametric estimation. Springer Science+Business Media, New York.

[75] Van den Berg, G.J., 2001. Duration models: specification, identification and multiple duration. in: Heckman, J.J., Leamer, E. (Eds.), Handbook of econometrics, Vol. 5. Elsevier Science, Amsterdam, pp. 3381-3460.

[76] Woutersen, T., 2000. Estimators for panel duration data with endogenous censoring. UWO Working Paper.

[77] Woutersen, T., 2002. Robustness against incidental parameters. UWO Working Paper.

[78] Yu, J., de Jong, R., Lee, L.F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146, 118-134.

[79] Zhao, L., Hanson, T. E., 2011. Spatially dependent polya tree modeling for survival data. Biometrics 67, 391-403.

[80] Zhao, L., Hanson, T. E., Carlin, B.P., 2009. Mixtures of polya trees for flexible spatial frailty survival modeling. Biometrika 96, 263-276.