INAUGURAL-DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der Ruprecht-Karls-Universität
Heidelberg

vorgelegt von

MSc (Mathematical Sciences) Alexander Günther Kreiß

aus Hamburg

Tag der mündlichen Prüfung:

Local Maximum Likelihood Estimation of Time Dependent Parameters in Dynamic Interaction Networks

Betreuer: Prof. Dr. Enno Mammen
Prof. Dr. Wolfgang Polonik

Acknowledgements

I would very much like to thank my supervisor Enno Mammen for being always available for discussions and helping me through the project. By providing suggestions and financial support, he gave me lots of great opportunities for attending conferences and meeting people, thereby he significantly participated in creating a beneficial and exciting time for me.

In this context, I would also like to thank Wolfgang Polonik for being available for advice and inviting me to Davis, California. Moreover, Carsten Jentsch was a great support through many beneficial and pleasant conversations.

Finally, numerous people I met before and during the time in Heidelberg were a great source of help and happiness. Especially, I would like to mention my colleagues and friends at the universities in Heidelberg and Mannheim who created a friendly and productive atmosphere.

I was part of the Research Training Group RTG 1953 (funded by the Deutsche Forschungsgemeinschaft) which I'm thankful for because it provided many good opportunities for discussions and networking in Heidelberg and Mannheim through seminars and workshops. For carrying out the computations in Section 4.5.3, I acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1134-1 FUGG.

Zusammenfassung

Wir interessieren uns für das Verhalten von vernetzten Akteuren in Abhängigkeit von Kovariaten, welche Informationen über jedes Paar von Akteuren enthalten. Hierbei können zwei Akteure durch Interaktionen miteinander in Kontakt treten. Wir modellieren diese Interaktionen durch Methoden der Survival Analysis: Für jedes Paar von Akteuren beobachten wir einen Sprungprozess, der die Anzahl der Interaktionen zwischen diesen beiden Akteuren sowie deren Zeitpunkte kodiert. Die Intensitätsfunktionen dieser Zählprozesse modellieren wir in parametrischer Art und Weise abhängig von den Konvariaten. Wir erlauben dabei, dass der Parameter eine zeitlich veränderliche Funktion ist, dadurch wird das Modell nicht-parametrisch. Wir untersuchen das asymptotische Verhalten eines lokalen maximum Likelihood Schätzers. Dies beinhaltet punktweise Asymptotik der geschätzten Parameterfunktion sowie einer L^2 Teststatistik. Um die mathematische Analyse durchzuführen, stellen wir drei Ideen vor, mit denen die Abhängigkeiten in einem Netzwerk beschrieben werden können. Diese liefern Möglichkeiten um Kovarianzen abzuschätzen und Konzentrationsungleichungen zu beweisen. Dies könnte auch unabhängig von unserem konkreten Kontext interessant sein. Die theoretische Betrachtung wird durch eine Anwendung auf einen realen Datensatz über Leihfahrräder vervollständigt.

Abstract

In the present thesis we are interested in modelling the behaviour of actors in a network in dependence of explanatory variables which give information about every pair of actors. The behaviour is here expressed in interactions which the actors may cast amongst each other. Our model is based on a survival analysis idea: We assume that the interaction times between any two actors are encoded in a counting process such that we observe a counting process for any pair of actors. The intensity functions of these counting processes are then assumed to depend on the covariates in a certain parametric way. We allow that the parameters are time dependent functions, thereby the model becomes non-parametric. We present a rigorous analysis of the asymptotics of a non-parametric estimator based on a local likelihood approach. This includes point-wise asymptotics of the estimated parameter curves as well as asymptotics for an L^2-type test statistic.

In order to carry out the mathematical analysis of these terms we introduce three ideas to handle the complex dependence structure on the network. These provide different tools for handling covariances and proving concentration inequalities which might be of independent interest.

The theoretical analysis is complemented with an application to real-world data: We investigate the impact of different network quantities on a bike sharing network.

Contents

1 Introduction

The statistical analysis of network data has recently become very popular and it is a fast growing field. In particular since the emergence of big computing power and social media, a huge volume of network data has become available (see e.g. the Stanford Large Network Dataset Collection (SNAP) or the Koblenz Network Collection (Konect)). It is of interest to draw conclusions from this type of data about, e.g. social (Jackson (2008); Newman (2010)) or economical behaviour (Brownlees et al. (2018); Diebold and Yilmaz (2014)). Introductions to the field of network data analysis, both from an applied and mathematical standpoint, are for example the books Jackson (2008); Newman (2010); Kolaczyk (2009, 2017).

The type of data is typically comprised of a network which in turn is a set of vertices and edges. We understand the vertices as actors and the edges as channels along which the actors can interact. In a specific example, e.g. a social media setting, the vertices could be users and we place an edge between users if they befriend each other. However, edges might also have a more abstract meaning. Consider the following example of a a rental bike network. In there, the rental bike stations form the vertices and we will place an edge between two bike stations, on a certain day, if at least one person took a bike tour between these two stations at one of the previous two days, i.e., we place an edge between them if there is regular traffic. We elaborate on this in the data analysis section. As both examples suggest, the network structure may change over time, i.e., edges between actors can emerge and dissolve during the observation period. We assume that we observe the network structure and the interactions between the actors. Moreover, we observe a set of covariates for each pair of actors. In the social media example, we could observe for each pair of users the number of common friends, the number of interactions in the past or an indicator which indicates if both users are in the same age class. We give more examples for covariates in the rental bike network: The number of bike rides between two bike stations in the days before, the number of bike stations where people go to from either or both of these two bike stations. These are just examples, and the exact choice of covariates depends on the available data and the exact setting to which the model is applied. The question we are interested in, is to model the influence of the covariates on the interactions. In other words, we want to regress the events on the covariates.

The modelling approach is via counting processes: We assume that for every pair of actors, we observe a counting process which counts the interactions between these two actors. The intensity functions of these counting processes are specified in the model as depending on the covariates and a parameter function. The interpretation of the parameter function is that it quantifies the impact of the covariates on the intensity function over time. Our aim is to apply methods from survival analysis in order to make inference about the covariates. We particularly focussed on two points: Firstly, the influence the covariates have may change over time. Secondly, neighbouring actors and hence the interactions among them are highly correlated. To accommodate for the first requirement our precise modelling idea is to formulate an interpretable parametric model where we replace the constant parameter by a time dependent function. This generalisation is in the same spirit as for example in some non-stationary time series models (cf. Dahlhaus (1997)). Hence, we need non-parametric estimation techniques. In this thesis we choose local-likelihoods (cf. Tibshirani and Hastie (1987) or specifically for survival analysis Hjort (1993)). As part of the asymptotic analysis we also study an L^2-type test statistic (analogue to e.g. Härdle and Mammen (1993)) for testing for a constant parameter function. In order to take care of the second requirement, we need to specify interpretable dependence assumptions on the data. Finding these assumptions is, next to executing the rigorous analysis of the estimator, a big part of the thesis.

1.1 Literature Review

The idea of studying a network of actors by means of bivariate relations (i.e., by the behaviour of pairs) dates back at least to Katz and Proctor (1959). More recently, in social sciences a related class of models are the so called stochastic actor oriented models (cf. Snijders et al. (2010); Snijders (2001)) in which the formation of ties is driven by actor and pair specific effects (also called covariates). Here the idea is that the formation of ties is driven by actor specific decisions. The actors can base their decisions on whether to form a tie or not on: A) Their direct environment they are able to perceive, B) Personal interests and C) External factors. Thus, the covariates can be based on network quantities (e.g., the number of friends an actor has, how many common friends two actors have,...), on personal properties (e.g., gender, age, employment status,...) or on exogenous quantities (e.g., the weather, the current value of currency exchange rates,...). One can also carry these ideas over to relational event data as introduced in Butts (2008). In contrast to before we consider no longer the formation of ties (like friendships, trust relationships,...) but instead the occurrence of single events (like phone calls, liking a post, email sending,...). In this context we also mention Stadtfeld and Block (2017) who consider similar modelling ideas.

Another approach for modelling dynamic networks is by using exponential random graph models, so called ERGMS (see e.g. Frank and Strauss (1986)), which were initially suggested for static networks. Extensions by Hanneke et al. (2010) and Krivitsky and Handcock (2014) allow for modelling of dynamic networks. Further possibilities are dynamic stochastic block models (cf. Ho et al. (2001); Yang et al. (2011)), continuous-times Markov Models (cf. Wasserman (1980)), Bayes modelling using latent factors (cf. Durante and Dunson (2014)), dynamic infinite relation models (cf. Ishiguro et al. (2010)), dynamic Markov random fields (cf. Kolar and Xing (2009)). See also the overview article Goldeberg et al. (2010).

Mathematically, Perry and Wolfe (2013) are very close to this thesis, however in their modelling they do not use time dependent parameter functions. Moreover, they study a different asymptotic: They let the time tend to infinity while we study networks growing in size on a fixed observation period.

1.2 Contribution

In terms of modelling the biggest innovation of this work is to add time dependence of the parameter function into the model as it is studied in Perry and Wolfe (2013). Including this time dependence seems to be a reasonable increase in modelling flexibility because the parameters of interest occur as weights of the covariates in the intensity function and we believe that they may change for two reasons: Firstly, it is plausible that, e.g., in winter or summer, the actors in the network react differently to certain changes in their environment. Secondly, it could also happen that the scale of the covariates changes as the network evolves which requires an adjustment of the weights.

We provide a rigorous mathematical analysis of the large sample properties (i.e., when the number of actors tends to infinity) of the local maximum likelihood estimator and an L^2-type test statistic. To this end we need to formulate weak dependence assumptions on the behaviour of the actors in the network to ensure sufficiently different observations. Making these assumptions mathematically precise was one of the biggest challenges when writing this thesis.

1.3 Structure of Thesis

The structure of this thesis is as follows. In Chapter 2, we will collect basic results and notation about stochastic processes in general, counting processes in particular and briefly about networks. In this section most results were already established and we refer to standard literature for proofs (we give references specific for each result in the next section). Afterwards in Chapter 3, we will introduce the three concepts for quantifying time varying dependence on a network which we will use in the proofs of the main results of this thesis which are stated in Section 4.1 followed by the assumptions we impose in Section 4.4. The theoretical presentation is illustrated by a real world data example using bike sharing networks in Chapter 4.5. The Proofs of the main results are given in Chapter 5. We will finish with some concluding remarks in Section 6. The model and Theorem 4.2 together with the bike example have been published in Kreiß et al. (2017). The remaining theory is being prepared for publication.

2 Preliminaries and Notation

In this chapter, we introduce the basics of counting processes, stochastic integration and networks which we will apply later in the thesis. The main aim of this section is to provide all necessary notation and probability theoretic results which we use later in the statistical network analysis. Particularly, it will be of interest how Stieltjes and Itô Integration relate. This will be important because the Stieltjes Integral is defined pathwise, and thus it allows simple calculations, while the Itô Integral is a more abstract object but has martingale properties. As all results here are standard, we will mostly just provide references for proofs. The counting process Section is mainly based on Andersen et al. (1993) and for the stochastic integration, we refer to Cohen and Elliott (2015) and Protter (2005).

2.1 Stochastic Processes and Integration

We start by talking about stochastic processes and their properties. Let $\mathcal{T} \subseteq [0, \infty)$ be an interval which has zero as smallest element and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ be a filtered probability space, i.e., for $s \leq t \in \mathcal{T}$ we have σ-fields $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$. And let $X : \mathcal{T} \to \mathbb{R}$ be a stochastic process, i.e., $X_t : \Omega \to \mathbb{R}$ is a random variable for every $t \in \mathcal{T}$. In order to emphasize the randomness of X, we will sometimes write $X_t(\omega)$, where $\omega \in \Omega$ is a random element. If $E(\omega)$ is a logical statement whose result (*true* or *false*) depends on ω, we will denote in slight abuse of notation the event that E is *true* by $\{E\} := \{\omega \in \Omega : E(\omega) \text{ is true}\}$. Continuing with this slight but usual sloppiness, we write $\mathbb{P}(E)$ for $\mathbb{P}(\{E\})$. If $\mathbb{P}(E) = 1$ we say that E holds almost surely or for short *a.s.*

We call X *adapted* if X_t is measurable with respect to $\mathcal{F}_t$ for all $t \in \mathcal{T}$. An adapted process with $\mathbb{E}(|X_t|)$ finite for all t and with $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ a.s. for $s < t$ is called a *martingale*. As usual X_t is a sub-martingale if $\mathbb{E}(X_t|\mathcal{F}_s) \geq X_s$. We say that another process Y is a *modification* of X if for every t, we have $X_t = Y_t$ a.s. If the stronger statement

$$X_t(\omega) = Y_t(\omega) \text{ for all } t$$

holds almost surely. Then we call X and Y indistinguishable.

We will later deal with counting processes which are explicitly given in a form which already exhibits useful continuity properties. Nevertheless, for completeness, we give conditions under which martingales have a nice modification. We say that the filtration fulfils the usual conditions (les conditions habituelles, cf. Andersen et al. (1993)) if

$$\mathcal{F}_t = \bigcap_{r>t} \mathcal{F}_r \text{ for all } t \qquad \text{(Right Continuiuty)} \qquad (2.1)$$

$$\text{For all } A \subseteq B \in \mathcal{F} : \mathbb{P}(B) = 0 \Rightarrow A \in \mathcal{F}_0 \qquad \text{(Completeness)}$$

We say that X is *cadlag* (a.s.) if the function $t \mapsto X_t(\omega)$ is cadlag (a.s.). A function is cadlag if it is continuous from the right and has limits from the left. The main result is then Corollary 5.1.9 in Cohen and Elliott (2015):

Theorem 2.1. *If $(\mathcal{F}_t)_t$ is right continuous (c.f. (2.1)), then every martingale admits a cadlag modification.*

We mention already at this point the concept of localizing. A stochastic process X is called a local martingale if there is a sequence of stopping times T_n, i.e., T_n are random variables such that $\{T_n \leq t\} \in \mathcal{F}_t$ for all t, with $T_n \to \infty$ a.s. as $n \to \infty$ and $X_{t \wedge T_n}$ is a martingale. Here, $x \wedge y := \min(x, y)$ denotes the minimum of two numbers $x, y \in \mathbb{R}$. Analogously, we say that X_t exhibits a certain property locally if the process $X_{t \wedge T_n}$ exhibits this property.

Next, we introduce the concept of predictability which is going to play an important role later on.

Definition 2.2. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}}, \mathbb{P})$ be a filtered probability space. We call a process X* predictable *with respect to $\mathcal{F}_t$ if the function $(\omega, s) \mapsto X_s(\omega)$ as a function on $\Omega \times \mathcal{T}$ is measurable with respect to the σ-field Σ_p generated by all left-continuous and with respect to $\mathcal{F}_t$ adapted processes. If the filtration $\mathcal{F}_t$ is clear from the context, we just say that X is predictable.*

In order to understand a bit better what this definition means, we define the σ-field of events strictly before a stopping time. Let therefore S be a stopping time and define

$$\mathcal{F}_{S-} = \sigma\left(\mathcal{F}_0 \cup \{A \cap \{t < S\} : A \in \mathcal{F}_t, t \in \mathcal{T}\}\right),$$

where for a collection of sets $\mathcal{X}$ we denote by $\sigma(\mathcal{X})$ the σ-field generated by $\mathcal{X}$. We note that for any given time point $s \in \mathcal{T}$ the deterministic and constant stopping time $S \equiv s$ is indeed a stopping time and

$$\mathcal{F}_{s-} = \sigma\left(\bigcup_{t<s} \mathcal{F}_t\right).$$

In this sense the value of X_S can be determined by using knowledge available only before S. It can be shown (according to Andersen et al. (1993)) that a process X is predictable if and only if X_S is measurable with respect to $\mathcal{F}_{S-}$ for all stopping times S.

We continue this review by talking about stochastic integration. It is therefore necessary that we restrict to processes with paths for which an integral can be defined almost surely (i.e., we require measurability of the paths). More precisely we call a process X measurable with respect to a σ-field Σ on $\Omega \times \mathcal{T}$ if the function $(\omega, s) \mapsto X_s(\omega)$ is measurable with respect to Σ. If $\Sigma = \mathcal{F} \times \mathcal{B}$ where $\mathcal{B}$ denotes the Borel σ-field, then we just call X measurable. Such processes have exactly the desired property: Every measurable process X has almost surely measurable paths, i.e., $t \mapsto X_t(\omega)$ is Borel measurable for almost all ω (cf. Cohen and Elliott (2015) Exercise 3.4.7).

The ultimate aim of this chapter is to define a notion of stochastic integration which is so simple that we can easily work with it but which is also sophisticated enough such that useful results are available. Later we will only be concerned with processes which arise in the context of counting processes (see the next section), therefore it is practical to make use of a property that all these processes share: They are of *bounded variation* which we will introduce next.

Definition 2.3. *We say that a measurable process X is of* bounded variation *if it is cadlag and for any $t \in \mathcal{T}$ there exists an almost surely finite random variable $C(t)$, i.e., $C(t) < \infty$ a.s., such that for any increasing, deterministic sequence $(t_i)_{i\in\mathbb{N}} \subseteq \mathcal{T}$*

$$\sum_{i=1}^{\infty} |X_{t_{i+1}} - X_{t_i}| < C(t) \ a.s.$$

Combining the results in Elstrodt (2011), we find that every right-continuous monotonically increasing function $f : [0, \infty) \to [0, \infty)$ induces a measure μ_f on $\mathcal{B}$ with $\mu((a, b]) = f(b) - f(a)$. Technically it is possible to extend the measure μ_f to a σ-field which contains all subsets of μ_f null sets but the Borel σ-field is sufficient for our purposes. For a Borel measurable function $g : [a, b] \to \mathbb{R}$ we define the Stieltjes-Integral $\int_a^b g(x)df(x) := \int_a^b g(x)d\mu_f(x)$ where the integral on the right is a regular Lebesgue integral. Finally, if f is a function of bounded variation (we use here Definition 2.3 for the deterministic stochastic process f), we can write it as sum $f = f_1 - f_2$, where f_1, f_2 are both monotonically increasing and cadlag. Thus, they define measures μ_{f_1}, μ_{f_2} as described above. The signed measure $\mu_f := \mu_{f_1} - \mu_{f_2}$ is then uniquely defined (even though f_1 and f_2 are not unique, c.f. Cohen and Elliott (2015) Theorem 1.7.22). In order to get a unique decomposition of the signed measure μ_f into two unsigned measures, we apply the Jordan-Hahn decomposition of μ_f (cf. Cohen and Elliott (2015), Lemma 1.7.6). The Jordan-Hahn decomposition comprises two unique mutually singular measures μ_f^+, μ_f^- (i.e. there is a set $P \subseteq \mathbb{R}$ such that $\mu_f^+(P) = \mu_f^-(\mathbb{R} \setminus P) = 0$) with $\mu_f = \mu_f^+ - \mu_f^-$. We define $\int_a^b g(x)df(x) := \int_a^b g(x)d\mu_f^+(x) - \int_a^b g(x)d\mu_f^-(x)$. Finally, we denote by $|\mu_f| := \mu_f^+ + \mu_f^-$, the variation measure and define for the integral $\int_a^b g(x)d|f|(x) := \int_a^b g(x)d|\mu|(x)$.

By collecting all previous considerations we see that the following definition is a reasonable definition of a stochastic integral.

Definition 2.4. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ be a filtered probability space and let A be measurable stochastic processes and let X have bounded variation as defined in Definition 2.3. For $a, b \in \mathcal{T}$ and $a < b$, we call the random variable*

$$\omega \mapsto \int_a^b A_s(\omega)dX_s(\omega) = \int_a^b A_s dX_s$$

the Lebesgue-Stieltjes Integral *or just* integral *of A with respect to X.*

Note that the integral in the above definition is allowed to be $\pm\infty$. As we define the integral as a Lebesgue integral, the Lebesgue-Stieltjes Integral may be computed by

Lebesgue-type approximations, i.e., by the limit of finite sums. We collect some useful properties of the Lebesgue-Stieltjes Integral in the following Lemma which also ensures that we have defined a proper random process (this is Cohen and Elliott (2015) Lemma 8.1.3, Remark 8.1.4 and Lemma 8.16).

Lemma 2.5. *Let A be measurable process, let X, Y have bounded variation and let λ be a random variable.*

1. *The stochastic process $t \mapsto \int_0^t A_s dX_s$ is measurable itself if it exists for all t.*
2. *If A, X are both adapted to the filtration and cadlag, then $t \mapsto \int_0^t A_s dX_s$ is adapted and cadlag as well.*
3. *The integral is linear in the integrator, i.e.,*

$$\int_0^T A_s d(X_s + \lambda Y_s) = \int_0^t A_s dX_s + \lambda \int_0^t A_s dY_s.$$

Furthermore, we mention at this point that the notation of the stochastic Lebesgue-Stieltjes Integral is different from the famous Itô Integral. For example, for the Lebesgue-Stieltjes notation we do not require that X has martingale properties. On the one hand, the Lebesgue-Stieltjes Integral is very easy to work with because the definition is almost surely path-wise. On the other hand, we do not know that the integral is a martingale in the integration limits. Happily, we can make use of the best of both worlds in our specific setting. Before we can formulate a result about this, we need a bit more notation.

In an analogue fashion to the Lebesgue-Stieltjes integral in Definition 2.4, we understand the random variable $\int_a^b A_s d|X|_s$. We say that an adapted process X of bounded variation has integrable variation if $\mathbb{E}\left(\sup_{t\in\mathcal{T}} \int_0^t d|X|_s\right) < +\infty$. We say that an adapted and monotonically increasing process A is integrable if $\mathbb{E}(\sup_{t\in\mathcal{T}} A_t) < +\infty$. With these definition we can formulate the Doob-Meyer decomposition (Cohen and Elliott (2015) Theorem 9.2.7):

Theorem 2.6. *Let X be a right continuous local sub-martingale. Then there is a unique locally integrable, predictable, monotonically increasing, right-continuous process A with $A_0 = 0$ such that*

$$M := X - A$$

is a local martingale. I.e, there is a sequence of stopping times T_n such that $M_{t\wedge T_n}$ is a martingale and $A_{t\wedge T_n}$ is an integrable, monotonically increasing process. The process A is also called compensator *of X.*

In order to formulate a result about the relation of Lebesgue-Stieltjes and Itô Integration (and for other purposes), we need the concept of quadratic variation. Let M be a right-continuous martingale and assume additionally that $\mathbb{E}(\sup_{t\in\mathcal{T}} |M_t|^2) < +\infty$, then we call M square integrable. For those martingales, M^2 is a sub-martingale and by the Doob-Meyer Decomposition (Theorem 2.6), there is a unique locally integrable, monotonically increasing, right-continuous process $\langle M \rangle$ with $\langle M \rangle_0 = 0$ such that $M^2 - \langle M \rangle$ is a local martingale. We call $\langle M\rangle$ the *predictable quadratic variation of* M. For two martingales M and N, we define the covariation as $\langle M,N\rangle := \frac{1}{2}\left(\langle M+N\rangle - \langle M\rangle - \langle N\rangle\right)$. Note that

$$MN - \langle M,N\rangle = \frac{1}{2}\left((M+N)^2 - M^2 - N^2\right) - \langle M,N\rangle$$

is a local martingale.

For any square integrable martingale M, the process $\langle M\rangle$ is by definition monotonically increasing and right-continuous. As a consequence we may integrate with respect to it in the Lebesgue-Stieltjes sense (because all increasing stochastic processes are of bounded variation). We say that a predictable process H is integrable with respect to M if

$$\mathbb{E}\left((H_0M_0)^2 + \int_{\mathcal{T}} H_s^2 d\langle M\rangle_s\right) < +\infty.$$

We have now all notation available which we need to formulate a result about equality of the Stieltjes and Itô Integral (other than the Itô Integral itself, but as we will only work with the Stieltjes Definition later, we do not introduce it here, for an introduction see Chapter 12 in Cohen and Elliott (2015)). The following result is Theorem 12.2.8 in Cohen and Elliott (2015).

Theorem 2.7. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ be a filtered probability space and let $H, M : \mathcal{T} \to \mathbb{R}$ be stochastic processes such that M is a square integrable martingale of integrable variation and H is predictable and integrable with respect to M. Moreover, we assume that $\mathbb{E}\left[\int_{\mathcal{T}} |H_s| d|M|_s\right] < \infty$. Then the stochastic integral in Stieltjes sense (cf. Definition 2.4) and Itô sense are indistinguishable as processes.*

The two main reasons, why this result will be so useful are firstly, that the Itô Integrals understood as processes in the integration limits are square integrable martingales and secondly, because of that we may apply Itô's Formula which we formulate next. Because the integrals are martingales, we can compute their quadratic predictable variation (Corollary 12.2.4 in Cohen and Elliott (2015)). If M is a square integrable martingale and H is integrable with respect to M, then

$$\left\langle \int_0^t H_s dM_s \right\rangle = \int_0^t H_s^2 d\langle M\rangle_s.$$

In contrast to the predictable quadratic variation, we also define the optional quadratic variation for a cadlag square integrable local martingale M as $[M]_t := M_t^2 - 2\int_0^t M_{s-}dM_s$, where $M_{s-} := \lim_{r\to s, r<s} M_r$. Here the integral is to be understood in an Itô sense because we do not impose any assumption on the variation of M. It is then also clear that $M^2 - [M]$ is a martingale. Denote moreover by $\Delta M_t := M_t - M_{t-}$ the jump size of M at t, then $\Delta[M]_t = (\Delta M_t)^2$. The optional quadratic covariation of two square integrable martingales M and N is defined via the polarization formula $[M,N] := \frac{1}{2}([M+N] - [M] - [N])$. From this definition it is not obvious that $[M]$ is a cadlag, increasing process, but one can prove that it is (cf. Theorem 22 in Chapter II.6 in Protter (2005)). Hence, we can integrate with respect to $[M]$ in the Stieltjes sense. Finally, we

can compute the optional quadratic variation of a stochastic integral (in Itô sense) by (cf. Theorem 29, Chapter II.6 in Protter (2005))

$$\left[\int_0^t H_s dX_s\right]_t = \int_0^t H_s^2 d[X]_s.$$

We remark at this point that we have an explicit formula for the optional quadratic variation, while the predictable quadratic variation is implicitly given by the Doob-Meyer-Decomposition (Theorem 2.6). The following lemma relates both types of quadratic variations and provides a helpful tool for finding the predictable quadratic variation in our case.

Lemma 2.8. *Let X be a cadlag square integrable local martingale. Then $\langle X\rangle$ is the compensator of $[X]$ in the Doob-Meyer-Decomposition of $[X]$.*

Proof. Note that $[X]_t$ is an increasing process and hence a local sub-martingale. By Theorem 2.6 there is a unique right-continuous, predictable and increasing process A_t such that $[X]_t - A_t$ is a local martingale. Thus,

$$X_t^2 - A_t = X_t^2 - [X]_t + [X]_t - A_t$$

is a martingale as a sum of two martingales. But as the compensator in Theorem 2.6 is unique, we conclude $\langle X\rangle = A$. □

With this notation we will now formulate Itô's Formula for martingales with jumps (Theorem 14.2.4 in Cohen and Elliott (2015)). Here X is to be understood as the cadlag modification of X.

Theorem 2.9. *Let X be a n-dimensional vector of square integrable martingales $X = (X^1, ..., X^n)$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Then,*

$$\begin{aligned} f(X_t) = f(X_0) &+ \sum_{i=1}^n \int_{(0,t]} \partial_i f(X_{s-}) dX_s^i + \frac{1}{2}\sum_{i,j=1}^n \int_{(0,t]} \partial_{ij} f(X_{s-}) d[X^i, X^j]_s \\ &+ \sum_{0<s\le t} \left(f(X_s) - f(X_{s-}) - \sum_{i=1}^n \partial_i f(X_{s-})\Delta X_s^i - \frac{1}{2}\sum_{i,j=1}^n \partial_{ij} f(X_{s-})\Delta X_s^i \Delta X_s^j \right). \end{aligned}$$

The above equality means that the processes to the left and to the right are indistinguishable.

We remark here that by continuity of f the difference $f(X_s) - f(X_{s-})$ is only different from zero, if $\Delta X_s \neq 0$, i.e., if X has a jump at time s. So the sum in the formula above is really a sum over all jumps of X. Since X is cadlag, it has only countably many jumps and the sum is well defined.

Next, we introduce Lenglart's Inequality which shows us how a martingale may be controlled by using the quadratic variation. We state in the following a slight adaptation of the original version as it is provided in Lenglart (1977).

Lemma 2.10. *Let X be a non-negative, right-continuous local sub-martingale and denote by A its compensator from Theorem 2.6. Then it holds for all finite stopping times $S > 0$ and all $c, d > 0$ that*

$$\mathbb{P}\left(\sup_{t\in[0,S]} M_t \geq c\right) \leq \frac{1}{c}\mathbb{E}\left(A_S \wedge d\right) + \mathbb{P}\left(A_S \geq d\right).$$

In the following parts of the thesis, we will apply Lenglart's Inequality mostly in the following form which is close to Andersen et al. (1993). The following is an easy corollary to the previous lemma.

Corollary 2.11. *Let M be a locally square integrable, right-continuous martingale and denote by $\langle M\rangle$ it's compensator.*

1. *For all $T, c, d > 0$ we have*

$$\mathbb{P}\left(\sup_{t\in[0,T]} |M_t| \geq c\right) \leq \frac{d}{c^2} + \mathbb{P}\left(\langle M\rangle_T \geq d\right).$$

2. *For all $T > 0$ it is true that*

$$\langle M\rangle_T \xrightarrow{\mathbb{P}} 0 \implies \sup_{t\in[0,T]} |M_t| \xrightarrow{\mathbb{P}} 0.$$

Lastly, we will present in this Section our main tool for finding the asymptotic distributions later: Rebolledo's Martingale Central Limit Theorem. It is known that a Brownian Motion is the only continuous Gaussian process with a certain covariance structure. This is used to formulate a martingale central limit theorem in the following. We state here the version of the theorem as Theorem II.5.1 in Andersen et al. (1993), the original work is Rebolledo (1980).

Let $M^n = (M_1^n, ..., M_k^n)$ be a vector of sequences of locally square integrable martingales on an interval $\mathcal{T}$. For $\varepsilon > 0$ we denote by M_ε^n a vector of locally square integrable martingales that contain all jumps of components of M^n which are larger in absolute value than ε, i.e., $M_i^n - M_{\varepsilon,i}^n$ is a local square integrable martingale for all $i = 1, ..., k$ and $|\Delta M_i^n - \Delta M_{\varepsilon,i}^n| \leq \varepsilon$. Furthermore, we denote by $\langle M^n\rangle := \left(\langle M_i^n, M_j^n\rangle\right)_{i,j=1,...,k}$ the $k \times k$ matrix of quadratic covariations.

Moreover, we denote by M a multivariate, continuous Gaussian martingale with $\langle M\rangle_t = V_t$, where $V : \mathcal{T} \to \mathbb{R}^{k\times k}$ is a continuous deterministic $k \times k$ positive semi-definite matrix valued function on $\mathcal{T}$ such that its increments $V_t - V_s$ are also positive semi-definite for $s \leq t$, then $M_t - M_s \sim \mathcal{N}(0, V_t - V_s)$ is independent of $(M_r : r \leq s)$. Given such a function V, such a Gaussian process M always exists. We can now formulate the central limit theorem for martingales.

Theorem 2.12. *Let $\mathcal{T}_0 \subseteq \mathcal{T}$. Assume that for all $t \in \mathcal{T}_0$ as $n \to \infty$*

$$\langle M^n\rangle_t \xrightarrow{\mathbb{P}} V_t$$

$$\langle M_\varepsilon^n \rangle \xrightarrow{\mathbb{P}} 0.$$

Then

$$M_t^n \xrightarrow{d} M_t$$

as $n \to \infty$ *for all* $t \in \mathcal{T}_0$. *Moreover,* $[M^n]_t \xrightarrow{\mathbb{P}} V_t$ *as* $n \to \infty$.

We remark that the instances of $\langle M^n \rangle$ and $[M^n]$ may be exchanged in the above theorem. We will use the theorem only in the simple situation $\mathcal{T}_0 = \{T\}$.

2.2 Counting Processes

In this section we introduce the main tool which we will apply in order to model the network data: Counting Processes. Most of the theoretic background has been presented in Section 2.1, so most of this section will consist of introducing the quantities from before in the counting process setting.

Definition 2.13. *Let* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ *be a filtered probability space, where* $\mathcal{T} \subseteq \mathbb{R}$ *is an interval with smallest element* $0 \in \mathcal{T}$. *Assume that* $(\mathcal{F}_t)$ *is right continuous (cf. 2.1). We call a multivariate process* $N = (N_1, ..., N_k)$, $k \in \mathbb{N}$, *on* $\mathcal{T}$ *a* counting process *if it is adapted to* $(\mathcal{F}_t)_{t\in\mathcal{T}}$, $N_i(0) = 0$ *for all* $i = 1, ..., k$, *it has paths which are piecewise constant and cadlag, it has jumps of size* $+1$ *only, no two components jump at the same time and* $N_i(t)$ *is almost surely finite for all* $t \in \mathcal{T}$ *and* $i = 1, ..., k$.

If not indicated differently, we will from now on always assume $\mathcal{T} = [0, T]$ for some $T > 0$. Note that $N(t) := \sum_{i=1}^k N_i(t)$ is also a counting process in the above sense because no two processes jump at the same time. With the convention $\inf \emptyset := \infty$, we define the stopping time

$$T_n := \inf\{t \in \mathcal{T} : N(t) \geq n\}.$$

These times T_n are precisely the jump locations of $N(t)$. We illustrate the concept with a small example.

Example 2.14. *Assume* n *people who can call each other and denote by* $t_m^{i,j}$ *the time point of the* m*-th call from person* i *to person* j. *The process* $N = (N_{i,j} : i, j \in \{1, ..., n\})$ *with* $N_{i,j}(t) := \sum_{m=1}^\infty \mathbb{1}(t_m^{i,j} \leq t)$ *is a counting process in the above sense. The set* $\{T_m : m \in \mathbb{N}\}$ *is identical with* $\left\{t_m^{i,j} : i, j \in \{1, ..., n\},\, m \in \mathbb{N}\right\}$.

So counting processes may be understood as a way of recording event times and the time points T_n are exactly the times of these events because the events happen exactly at the jumps of the counting process. Complementary to T_n, one sometimes introduces markers J_n which tell which process jumps at time T_n. But we do not need this concept here.

Additionally to the interpretation, the stopping times T_n have also a mathematical use: Since $t \mapsto N_i(t \wedge T_n)$ (where $x \wedge y := \min(x, y)$) are increasing processes, they are sub-martingales and hence N_i is a right-continuous local sub-martingale for $i =$

$1, ..., k$ with localizing sequence T_n (note that $N_i(t \wedge T_n) \leq n$ is bounded). So we may apply Theorem 2.6 (the Doob-Meyer-Decomposition) and obtain non-decreasing, right-continuous, predictable processes Λ_i with $\Lambda_i(0) = 0$ such that the processes $M_i := N_i - \Lambda_i$ are local martingales. It can be shown that the localizing sequence for M_i can be chosen to be T_n (cf. Andersen et al. (1993)). But we will later discuss only the case in which we do not need a localizing sequence because the processes are martingales. Since Λ_i is increasing, it is of bounded variation and it is almost everywhere differentiable but it may have jumps. If it has no jumps, we call its derivative the intensity function. We summarize this in the next definition.

Definition 2.15. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}}, \mathbb{P})$ be a filtered probability space and let N be a counting process. The unique predictable, non-decreasing, right-continuous processes Λ_i with $\Lambda_i(0) = 0$ and $M_i := N_i - \Lambda_i$ is a local martingale (they exist because of the Doob-Meyer-Decomposition, Theorem 2.6), are called* cumulated intensity processes *of the counting processes N_i with respect to $\mathcal{F}_t$. We write $\Lambda = (\Lambda_1, ..., \Lambda_k)$. If Λ_i is absolutely continuous, i.e., if there are predictable stochastic processes λ_i such that $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$, we call λ_i the* intensity function *of N_i with respect to $\mathcal{F}_t$. We write $\lambda = (\lambda_1, ..., \lambda_k)$. If the filtration $(\mathcal{F}_t)_{t \in \mathcal{T}}$ is clear from the context, we omit the* with respect to *part in the name.*

Note that Λ (and thus also λ if it exists) depends on the filtration $\mathcal{F}_t$. If a counting process N_i has finite expectations and an intensity function which is left continuous and has limits from the right, we have the relation (cf. Andersen et al. (1993) Chapter II.4)

$$\lim_{h \to 0} \frac{1}{h} \mathbb{E}(N_i(t+h) - N_i(t) | \mathcal{F}_t) = \lambda_i(t+),$$

where $\lambda_i(t+) := \lim_{s \to t, s > t} \lambda_i(s)$ denotes the right limit of λ_i at t. Under stronger assumptions it holds even

$$\lim_{h \to 0} \mathbb{P}(N_i(t+h) - N_i(t) = 1 | \mathcal{F}_t) = \lambda_i(t+).$$

In this sense $\lambda_i(t)$ determines the probability of a jump a time t given the past. In particular, a larger intensity at time t means that a jump is more likely at time t then for a lower intensity (conditional on the past).

Next, we are interested in properties of stochastic integrals with respect to counting processes and their martingales. To this end, let N be a univariate counting process with cumulated intensity function Λ. We firstly note that by definition N is right-continuous and increasing as well as Λ, they are both cadlag. Thus, both are of bounded variation. Hence, $M = N - \Lambda$ is of bounded variation too. As a consequence the Lebesgue-Stieltjes Integral (cf. Section 2.1) is defined with respect to all three of them. Next, we note that the stochastic integral with respect to N is actually a sum. We can compute the Lebesgue-Stieltjes Integral of any function $f : \mathcal{T} \to \mathbb{R}$ over any measurable set $A \subseteq \mathcal{T}$ to be

$$\int_A f(t)dN(t) = \sum_{n=1}^{\infty} f(T_n) \mathbb{1}(T_n \in A).$$

We study the properties of stochastic integrals with respect to a counting process martingale M a bit further in the case if N has intensity function λ. Most importantly we note that for an adapted stochastic process A, we have by Lemma 2.5 that $t \mapsto \int_0^t A_s dM_s$ is again an adapted process. Moreover, by using the linearity in the integrator (cf. Lemma 2.5) the mapping

$$F : t \mapsto \int_0^{t-} A_s dM_s := \lim_{r \to t, r<t} \int_0^r A_s dM_s = \lim_{r \to t, r<t} \left(\int_0^r A_s dN_s - \int_0^r A_s \lambda(s) ds \right)$$

is well defined because the integral with respect to N is a piecewise constant function with countably many jumps and the second integral is continuous in the integration limits. It is then clear that F is an adapted and left-continuous process. Thus, F is predictable in the sense of Definition 2.2. In order to formulate a result for functions defined on two variables, we need a new definition of predictability.

Definition 2.16. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ be a filtered probability space and $\mathcal{P}$ be a collection of real-valued stochastic processes defined on a set $\mathcal{D}$. Denote by $\sigma(\mathcal{P})$ the smallest σ-field on $\Omega \times \mathcal{D}$ such that all processes in $\mathcal{P}$ are measurable as functions of $\Omega \times \mathcal{D}$. Let now $\mathcal{P}_{pp}$ denote the set of all processes $X : \Omega \times \mathcal{T}^2$ with: X can be written as $X_{t,s}(\omega) = g_t(\omega) h_s(\omega) f(t,s)$, where $g, h : \Omega \times [0,T] \to \mathbb{R}$ are two stochastic processes and g is predictable and h is adapted both with respect to $\mathcal{F}_t$, while $f : \mathcal{T}^2 \to \mathbb{R}$ is a Lebesgue-Borel measurable, deterministic function. We call the σ-field*

$$\Sigma_{pp} := \sigma\left(\mathcal{P}_{pp}\right)$$

the partially-predictable *σ-field with respect to $\mathcal{F}_t$. A real-valued stochastic process φ defined on $\mathcal{T}^2$ is called* partially-predictable *with respect to $\mathcal{F}_t$, if it is measurable with respect to Σ_{pp}. If the filtration is clear from the context, we won't specifically mention it in the notation.*

Partially-predictable processes exhibit the following useful property. The proof works analogously to the proof of Lemma 2.5 (see references there), however it is not identical and so we give the proof here for completeness.

Lemma 2.17. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathcal{T}}, \mathbb{P})$ be a filtered probability space and N be a counting process with intensity function λ with respect to $\mathcal{F}_t$. Let φ be a real-valued stochastic process defined on $\mathcal{T}^2$. If φ is partially-predictable with respect to $\mathcal{F}_t$ (cf. Definition 2.16), then the process*

$$t \mapsto \int_0^{t-} \varphi(t,r) dM_r$$

is predictable with respect to $\mathcal{F}_t$.

Proof. We apply a monotone class argument. Let $\mathcal{H}$ be the set of all bounded real-valued stochastic processes φ defined on $\mathcal{T}^2$ such that

$$F_{N,\varphi}(t) := \int_0^{t-} \varphi(t,r) dN_r \text{ and } F_{\lambda,\varphi}(t) := \int_0^{t-} \varphi(t,r)\lambda(r) dr$$

are predictable. Note that for these processes, also the integral with respect to M is predictable by linearity of the integral. It is clear that $\mathcal{H}$ contains all constant functions $\varphi \equiv \varphi_0 \in \mathbb{R}$ because $F_{N,\varphi_0}(t) = \varphi_0 N(t-)$ and $F_{\lambda,\varphi_0}(t) = \varphi_0 \int_0^{t-} \lambda(r)dr$ both of which are left continuous and adapted. Thus, they are predictable. Let now $(\varphi_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}$ be a uniformly bounded, monotonically increasing sequence, i.e., $\varphi_n \leq \varphi_{n+1}$ with $\varphi_n \to \varphi$ point-wise as $n \to \infty$. Our aim is to prove that $\varphi \in \mathcal{H}$. Hence, we have to prove that for every stopping time S, it holds that $F_{N,\varphi}(S)$ and $F_{\lambda,\varphi}(S)$ are measurable with respect to $\mathcal{F}_{S-}$ (cf. Discussion after Definition 2.2). We note firstly that by definition F_{N,φ_n} and F_{λ,φ_n} are predictable. Moreover, since intensities are non-negative, the random variables $F_{N,\varphi_n}(S)$ and $F_{\lambda,\varphi_n}(S)$ converge monotonically increasingly towards $F_{N,\varphi}$ and $F_{\lambda,\varphi}$ respectively. Hence, both $F_{N,\varphi}(S)$ and $F_{\lambda,\varphi}(S)$ are measurable with respect to $\mathcal{F}_{S-}$ and hence $\varphi \in \mathcal{H}$ (it is clear that φ is bounded). Very similar arguments may be applied for a general sequence $(\varphi_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}$ with $\varphi_n \to \varphi$ uniformly.

Note finally that for every process $\varphi \in \mathcal{P}_{pp}$ (cf. Definition 2.16) with $\varphi(t,r) = g_t \cdot h_r \cdot f(t,r)$ where g is predictable, h is adapted and f is deterministic, we have that

$$F_{N,\varphi}(t) = g_t \cdot \int_0^{t-} h_r f(t,r) dN_r \text{ and } F_{\lambda,\varphi}(t) = g_t \cdot \int_0^{t-} h_r f(t,r)\lambda(r)dr$$

are both products of two predictable functions (the integrals are predictable because we may approximate the deterministic function by a continuous function), thus they are predictable. Hence, $\mathcal{P}_{pp} \subseteq \mathcal{H}$. Furthermore, $\mathcal{P}_{pp}$ is closed under multiplication. Hence, we may apply the monotone class theorem (see e.g. Cohen and Elliott (2015), Theorem 7.4.1) to conclude that $\mathcal{H}$ contains all bounded function which are measurable with respect to $\Sigma_{pp} = \sigma(\mathcal{P}_{pp})$.

In order to carry the result over to arbitrary processes φ which are measurable with respect to Σ_{pp}, we apply again a convergence result: Let φ be measurable with respect to Σ_{pp} and non-negative, then $\varphi_M := \varphi \cdot \mathbb{1}(\varphi \leq M)$ is bounded and measurable with respect to Σ_{pp}. Thus it is contained in $\mathcal{H}$. Moreover, $\varphi_M \to \varphi$ as $M \to \infty$ increasingly and hence by the same arguments as before $F_{N,\varphi}$ and $F_{\lambda,\varphi}$ are predictable. Lastly, for an arbitrary φ, write $\varphi = \varphi_+ - \varphi_-$ with $\varphi_+ = \varphi\mathbb{1}(\varphi \geq 0)$ and $\varphi_- = -\varphi\mathbb{1}(\varphi < 0)$ and apply the above arguments again separately to the non-negative functions φ_+ and φ_-. □

Before continuing with our study of the stochastic integrals, we note the following result on the jump size of the compensator.

Lemma 2.18. *Let N be a counting process with cumulated intensity process Λ. Then $\Delta\Lambda \in [0,1]$*

Proof. We give the proof as outlined in Andersen et al. (1993). Note that Λ is by definition increasing and hence $\Delta\Lambda \geq 0$. Moreover, we see that

$$\int_0^t \mathbb{1}(\Delta\Lambda_r > 1)dM_r = \sum_{r\leq t} \mathbb{1}(\Delta\Lambda_r > 1)(\Delta N_r - \Delta\Lambda_r) \leq 0,$$

because $\Delta N_r \in \{0,1\}$. On the other hand $\mathbb{1}(\Delta\Lambda_t > 1)$ is a predictable process and hence the above integral has expectation zero. We conclude that $\mathbb{1}(\Delta\Lambda_t > 1) = 0$. □

Next, we study the properties of M a bit further. Note that for multivariate counting processes $N = (N_1, ..., N_n)$ each component is locally bounded because $N_i(t \wedge T_n)$ are bounded by n. Moreover, the cumulated intensities $\Lambda_i(t)$ for $i = 1, ..., n$ are increasing and hence also locally bounded. This means that the processes $M_i(t)$ are locally bounded as well and hence, the M_i are local square-integrable martingales.

In order to show what the predictable and optional quadratic variation are, we follow the same route as in Andersen et al. (1993), Chapter II.4. Note that by the integration by parts formula for the Lebesgue-Stieltjes integral, we have $2\int_0^t X_{s-}dX_s = X_t^2 - \sum_{s\le t}(\Delta X_s)^2$. Using this, we can compute the optional variation process of the counting process martingale $M = N - \Lambda$, for which we have an explicit formula because

$$
\begin{aligned}
[M]_t = M_t^2 - 2\int_0^t M_{s-}dM_s &= \sum_{s\le t}(\Delta N_s - \Delta\Lambda_s)^2 \\
&= N_t - 2\sum_{s\le t}\Delta\Lambda_s \cdot \Delta N_s + \sum_{s\le t}(\Delta\Lambda_s)^2 \\
&= N_t - 2\int_0^t \Delta\Lambda_s dN_s + \int_0^t \Delta\Lambda_s d\Lambda_s \\
&= N_t - \int_0^t \Delta\Lambda_s dN_s - \int_0^t \Delta\Lambda_s dM_s,
\end{aligned}
$$

where we used that $\Delta N_s \in \{0, 1\}$ and hence $(\Delta N_s)^2 = \Delta N_s$, moreover Λ is increasing and hence $\Delta\Lambda_s \ne 0$ only for countably many s. We have defined Λ as the compensator of N and hence for $A_t := \Lambda_t - \int_0^t \Delta\Lambda_s d\Lambda_s$ we get that $[M]_t - A_t$ is a local martingale. Moreover, $A_t = \int_0^t(1 - \Delta\Lambda_s)d\Lambda_s$ is increasing (cf. Lemma 2.18) and predictable. Thus, by Lemma 2.8, we get for two counting processes N and $\widetilde{N}$ with respective compensators Λ and $\widetilde{\Lambda}$ and martingales M and $\widetilde{M}$

$$
\begin{aligned}
\langle M\rangle_t &= \Lambda_t - \int_0^t \Delta\Lambda_s d\Lambda_s \\
\langle M, \widetilde{M}\rangle_t &= -\int_0^t \Delta\Lambda_s d\widetilde{\Lambda}_s.
\end{aligned}
$$

Note for the predictable quadratic covariation that $N + \widetilde{N}$ is a counting process with compensator $\Lambda + \widetilde{\Lambda}$. In particular, for the case where Λ and $\widetilde{\Lambda}$ are absolutely continuous, i.e., they admit intensity functions, we obtain

$$
\begin{aligned}
\langle M\rangle_t &= \int_0^t \lambda(s)ds & [M]_t &= N_t \\
\langle M, \widetilde{M}\rangle &= 0 & [M, \widetilde{M}]_t &= 0. \qquad (2.2)
\end{aligned}
$$

Finally, we want to discuss under which circumstances the counting process martingale is an actual martingale and we do not have to worry about any localizing sequences. Moreover, we may apply then Theorem 2.7 to all counting process martingale integrals and treat them as Stieltjes or Itô integrals depending on the situation. In general it is not so easy to determine whether a local martingale is a martingale. But here we can make use of the very specific form of the stochastic processes M, N and Λ.

Lemma 2.19. *Let $\mathcal{T} = [0, T]$ and let N be a uni-variate counting process as in Definition 2.13 with cumulated intensity function Λ. Assume that $\mathbb{E}(N(T)) < +\infty$ as well as $\mathbb{E}(\Lambda(T)) < +\infty)$. Then $M := N - \Lambda$ is a martingale on $[0, T]$.*

Proof. Both processes N and Λ are increasing and hence $\mathbb{E}(M(t)) < +\infty$ for all $t \in [0, T]$. So we just need to show the integral property of the martingale: Let $s < t$ then by applying monotone convergence, the fact that $T_n \to \infty\, a.s.$ and the assumption of bounded expectations we get

$$\begin{aligned}
\mathbb{E}(M(t)|\mathcal{F}_s) &= \mathbb{E}(\lim_{n\to\infty} N(t \wedge T_n) - \Lambda(t \wedge T_n)|\mathcal{F}_s) \\
&= \lim_{n\to\infty} \mathbb{E}(N(t \wedge T_n|\mathcal{F}_s) - \lim_{n\to\infty} \mathbb{E}(\Lambda(t)|\mathcal{F}_s) \\
&= \lim_{n\to\infty} \mathbb{E}(M(t \wedge T_n|\mathcal{F}_s) \\
&= \lim_{n\to\infty} M(s \wedge T_n) = M(s).
\end{aligned}$$

□

The last theoretic prerequisite we need is the likelihood function for counting processes. We assume that $N = (N_1, ..., N_k)$ is a (multivariate) counting process and $(\mathbb{P}_\theta)_{\theta\in\Theta}$ is a collection of probability measures where Θ is an index set. Assume that the process N has intensity function $\lambda(\theta, t)$ with respect to $\mathbb{P}_\theta$. The likelihood for observations in the interval $[0, T]$ is then proportional to (cf. Andersen et al. (1993), Chapter II.7)

$$\sum_{i=1}^{k} \int_0^T \log \lambda_i(\theta, t) dN_i(t) - \int_0^T \lambda_i(\theta, t) dt. \tag{2.3}$$

2.3 Networks

In order to give a precise mathematical formulation of our statistical results about networks we use the notation from graph theory. However, in this thesis we will use the words graph and network equivalently. Moreover, we will not distinguish simple graphs from regular graphs and consider only simple graphs instead (without saying it explicitly in the notation every time).

Definition 2.20. *A* network *or a* graph *is a pair $G = (V, E)$ where V is a finite set and $E \subseteq V \times V$ with $(v, v) \notin E$ for all $v \in V$. The elements of V are called* vertices *or* actors *and the elements of E are called* edges *or* connections*. Elements of V will be typically denoted by $v, v_1, v_2, ...$ and elements of E will be typically denoted by $i, j, i_1, j_1, i_2, j_2, ...$. We will often identify G and E, i.e., we write $i \in G$ and mean $i \in E$. Moreover, we will understand $i \in G$ as index of an edge and also write e_i for the set of the two incident vertices of i. G is called* complete *if $E = V \times V \setminus \{(v, v) : v \in V\}$.*

If we have given a network $G = (V, E)$, we would visualize it by drawing a dot for each actor in V and draw an arrow from actor $v_1 \in V$ to actor $v_2 \in V$ if $(v_1, v_2) \in G$. If an arrow has two heads we usually omit both of them. We distinguish now networks where all connections have both arrows.

Definition 2.21. *A network $G = (V, E)$ is called* undirected *if $(v_1, v_2) \in E \Leftrightarrow (v_2, v_1) \in E)$ for all $v_1, v_2 \in V$. Otherwise it is called* directed*.*

For the asymptotic statistical analysis of networks we will always consider sequences of networks.

Definition 2.22. *A sequence of directed (undirected) networks $(G_n = (V_n, E_n))_{n\in\mathbb{N}}$ is a sequence of directed (undirected) networks $G_n = (V_n, E_n)$ with $V_n = \{1, ..., n\}$. Denote by r_n the number of connections in G_n, i.e., for directed networks $r_n := |G_n| = |E_n|$ and for undirected networks $r_n := \frac{|G_n|}{2} = \frac{|E_n|}{2}$.*

Our interpretation of a network will be that every vertex represents an actor who can interact with other actors along the connections. As we usually assume that a priori all actors can interact with all other actors, it is natural to have complete networks. But by choosing G_n differently it is possible to exclude certain interactions. However, we will introduce the possibility that pairs of actors can become active or inactive at different times. So we consider actually two networks: One which deterministically and un-dynamically states which actor can interact with each other (this is G_n), and another random and dynamic network which states which pairs actors are at the moment perceptive for interactions (we will introduce this process later). For simplicity of presentation, we will usually think about sequences of complete networks for G_n in the remainder of the thesis, i.e., for directed networks we will have $r_n = n(n-1)$ and for undirected networks $r_n = \frac{n(n-1)}{2}$. Moreover, we observe the interactions over a time period $[0, T]$ for some $T > 0$. Interactions comprise for us two quantities $(i = (v_1, v_2), t)$: The edge i with sender v_1 and receiver v_2 and a time point $t \in [0, T]$, t is called the interaction time. In particular, we do not allow interactions with several recipients. Note that for undirected networks, the notions of receiver and sender are somewhat arbitrary. Undirected interactions should therefore be more understood as interactions between two actors where none of the actors can be identified as sender or receiver. Denote by $\mathcal{E}_i$ the random set of all interaction times for the edge i. The event times are collected in counting processes

$$N^*_{n,i}(t) := \sum_{t_i \in \mathcal{E}_i} \mathbb{1}(t_i \leq t).$$

Although we allow a priori that interactions are possible between any two actors, it is realistic to assume that interactions between certain actors are so rare that they are not of interest, e.g. because of too far geographical distance. Therefore, we allow for a process $C_{n,i} : [0, T] \to \{0, 1\}$ for all $i \in G_n$ which indicates if the edge i is active at time t (if $C_{n,i}(t) = 1$) or not. This is the second network process mentioned in the previous paragraph. We need this distinction because we need a notion of objects we observe but at the same time we want to allow that certain objects do not contribute (but we do not know this in advance, it is part of the observation). We also introduce counting processes $N_{n,i}$ which count only those interactions which happened along active interactions:

$$N_{n,i}(t) := \int_0^t C_{n,i}(s) dN^*_{n,i}(s).$$

Along with these data, we also collect further information for each pair of actors $i = (v_1, v_2)$. We call this additional information *covariates* and denote it by $X_{n,i}(t) \in \mathbb{R}^q$. These covariate vectors may vary over time but they keep their dimension. Our ultimate aim is to understand how the covariates and the interactions on active connections are related and how this relation develops over time. We summarize everything in the definition of an *interaction network*.

Definition 2.23. *An* interaction network *is a sequence of directed or undirected networks $(G_n)_{n\in\mathbb{N}}$ together with the following three types of processes (all defined on the time interval $[0,T]$):*

1. *Counting processes $\left(N^*_{n,i}\right)_{i\in G_n, n\in\mathbb{N}}$,*
2. *$\{0,1\}$ valued processes $(C_{n,i})_{i\in G_n, n\in\mathbb{N}}$,*
3. *$\mathbb{R}^q$ valued processes $(X_{n,i})_{i\in G_n, n\in\mathbb{N}}$.*

Moreover, for every $n \in \mathbb{N}$ there is a filtration $(\mathcal{F}^n_t)_{t\in[0,T]}$ such that for all $n \in \mathbb{N}$ all $N_{n,i}$, $i \in G_n$, are adapted and all $C_{n,i}$ and $X_{n,i}$, $i \in G_n$, are predictable with respect to $(\mathcal{F}^n_t)_{t\in[0,T]}$.

We denote by r_n the number of connections in G_n and characterize an interaction network usually by its observable quantities $(C_{n,i}, X_{n,i}, N_{n,i})_{i\in G_n}$, where

$$N_{n,i}(t) := \int_0^t C_{n,i}(s) dN^*_{n,i}(s).$$

We are not too much concerned about the existence of a filtration as required in the definition above. One possibility would be to assume that $C_{n,i}$ and $X_{n,i}$ are continuous from the left and let $\mathcal{F}^n_t := \sigma(N^*_{n,i}(s), X_{n,i}(s), C_{n,i}(s) : i \in G_n, s \le t)$ be the filtration generated by the processes $(N^*_{n,i}, X_{n,i}, C_{n,i})$ for all $i \in G_n$.

In the rental-bike network example from the beginning we would consider a bike ride from station v_1 to station v_2 an interaction on the edge $i = (v_1, v_2)$. Thus, $N^*_{n,i}$ would encode the number of bike rides. We said that we're only interested in the bike rides on regularly used bike stations, so we set $C_{n,i}(t)$ only during those times to one where the destination i has shown regular activity in the past days. Thus $N_{n,i}$ is encoding the number of bike rides between v_1 and v_2 only during active times. Lastly, the vector $X_{n,i}(t)$ would contain the extra information about i which we listed in the introduction.

3 Describing Dependence on Dynamic Networks

In this chapter we will discuss three concepts for quantifying dependence in networks, namely *Asymptotic Uncorrelation*, *Momentary-m-Dependence* and *β-Mixing*. In order to define a realistic notion of weak dependence we want to allow that information is transported through observed and unobserved random mechanisms. Thus, the way the actors influence each other is random and time dependent itself. We illustrate this idea by informally discussing the three concepts.

Asymptotic Uncorrelation is mainly based on the idea that in undirected, interchangeable networks, i.e., where relabelling the vertices does not change the joint distribution of the whole network, the joint distribution of two edges depends only on the number of common vertices. We then make statements about the average covariance behaviour of pairs of edges. We will see that assumptions of this type will provide an easy method to bound variances of sums of random variables including network variables. Assumptions of this type will be sufficient to prove an asymptotic normality result for a non-parametric estimator at a fixed time point t_0 (cf. Thereom 4.2). In order to prove results about the global behaviour of the estimator (cf. Theorem 4.3 we require more involved independence statements (like exponential inequalities). These in turn need stronger concepts of dependence on a network. We present here *Momentary-m-Dependence* and *β-Mixing* as such concepts. We will begin here with an informal introduction.

On the one hand we believe that at a time point t_0, the near future of two actors given the entire history up to time t_0 is independent of the near future of two other actors, provided that both pairs of actors are separated at time t_0 (by separated we mean that their distance, which has to be defined, is large). This means, the only information flow between two at time t_0 separated pairs of actors is through the past (where they were possibly not separated). We will formalize this idea later under the name *Momentary-m-Dependence*. We will also quantify how strong this information flow is without knowledge of the past. This will be done by using *β-Mixing* coefficients. We emphasize here that both concepts are not nested in the sense that one would imply the other, they concern different scenarios: Momentary-m-Dependence concerns the question which additional knowledge about the future given the past does or does not affect the future of a given pair of actors. β-Mixing quantifies the effect the behaviour of a certain pair of actors has on the behaviour of another pair of actors in a certain time period without any extra knowledge.

To make these concepts precise we need to introduce some structure on the network which we do in the following. Recall that $G_n = (V_n, E_n)$ is a graph with vertices $V = \{1, ..., n\}$ and edges E_n. Now, we consider a sequence of random variables

$(C_{n,i}, X_{n,i}, N_{n,i})_{i\in G_n}$ which form an interaction network in the sense of Definition 2.23. We extend Definition 2.23 by adding a distance function.

Definition 3.1. *An interaction network $(C_{n,i}, X_{n,i}, N_{n,i})_{i\in G_n}$ on $[0,T]$ is called* structured interaction network *if there is a distance function d_t^n such that:*

1. *The processes $N_{n,i}$ have intensity functions $C_{n,i}(t)\lambda_{n,i}(t)$ with respect to $\mathcal{F}_t^n$ (cf. Section 2.2).*
2. *The triples $(C_{n,i}(t), X_{n,i}(t), N_{n,i}(t))_{i\in G_n}$ are identically distributed.*
3. *For all $t \in [0,T]$ and all $n \in \mathbb{N}$, the function $d_t^n : G_n \times G_n \to [0,\infty]$ is a metric and $d_t^n(i,j)$ is predictable with respect to $\mathcal{F}_t^n$ for all $i,j \in G_n$.*

Remark 3.2. *We already introduce the counting processes here because this is the setting in which we will apply the results. However, we could make an analogue definition for general random variables. In the following we will stick to the specific setting or be more general according to our needs.*

Remark 3.3. *For two edges $i,j \in G_n$, the distance $d_t^n(i,j)$ shall reflect how strongly dependent the edges i and j are: A short distance implies dependence while a large distance implies only weak dependence. From a modelling perspective, we emphasize that the distance function d_t^n is a very abstract object but it does not need to be known. Therefore it is a powerful instrument for describing relations between the actors.*

Example 3.4. *Suppose we are interested in modelling a messenger system with the following specifications in our model:*

- *The actors are the users of the system.*
- *We impose no a priori restrictions, i.e., we consider a directed, complete network.*
- *A directed interaction on $i = (v_1, v_2)$ is user v_1 sending a message to user v_2.*

Edges $i = (v_1, v_2)$ are always considered directed because v_2 has no physical possibility to make v_1 not send a message to v_2. Therefore, it makes sense to assume that only the sender is ultimately responsible for an interaction.

The covariates $X_{n,i}$ comprise different markers describing the relation between v_1 and v_2, for example:

- *Age or gender proximity of the users*
- *Proximity in geographic location*
- *Number of real life interactions*
- *Interactions between them and other people in the past*

For a realistic model we need to allow that interactions on $i = (v_1, v_2)$ *and* $j = (v_3, v_4)$ *are not independent because users might tell each other to contact other users or share information outside of the messaging system. But at the same time it is realistic to assume that the dependence is strong if and only if the actors* (v_1, v_2) *and* (v_3, v_4) *are close. We model this structure by assuming that at each time point we can quantify the strength of the connection between any two actors* v_1 *and* v_2 *in a way which is predictable with respect to the filtration* $(\mathcal{F}_t^n)_{t\in[0,T]}$ *(e.g. when it is determined by the covariates):*

- *If* v_1 *and* v_2 *are strongly dependent, the edge* $i = (v_1, v_2)$ *gets a low weight.*
- *If* v_1 *and* v_2 *are weakly dependent,* i *gets a large weight (if they are independent, we give it infinite weight).*

We define now $d_t^n(i, j)$*, the distance between two edges* i *and* j*, as the weight of the path with the lowest weight connecting the senders of* i *and* j *(note that we consider a complete network and hence there is always a path, it might have infinite weight though). The distance function can be interpreted as follows:*

- $d_t^n(i, j)$ *= "large" means that short paths have very high (maybe even infinite) weights, indicating that there is almost no direct influence. On the other hand it might happen that there is a very long path with low weights per edge (i.e., through very closely related actors), but then it is believable that the information is dispersed quickly and again* i *and* j *are only very weakly dependent.*
- $d_t^n(i, j)$ *= "small" means that our model must allow that* i *and* j *are strongly dependent.*

We will keep this example in mind when we make this idea mathematically precise.

3.1 Asymptotic Uncorrelation

Let $G_n = (V_n, E_n)$ be an undirected network and let $(Z_i)_{i\in G_n}$ be a set of random variables indexed by the edges. We assume that $(Z_i)_{i\in G_n}$ is jointly exchangeable in the vertices, i.e.: Let $\sigma : V_n \to V_n$ be a permutation of the vertex set. For an edge $i = (v_1, v_2) \in G_n$, we denote by $\sigma(i) := (\sigma(v_1), \sigma(v_2))$ the permuted edge. The network is called exchangeable if $(Z_i)_{i\in G_n}$ and $(Z_{\sigma(i)})_{i\in G_n}$ have the same joint distribution for all permutations of the vertex set σ. In particular, this means that Z_i and Z_j are identically distributed if there is a permutation σ such that $i = \sigma(j)$. Note that this is always the case because we have defined networks as having no loops, i.e., for $i = (v, v')$ we always assume $v \neq v'$. Next, we want to describe under which circumstances two pairs (Z_{i_1}, Z_{j_1}) and (Z_{i_2}, Z_{j_2}), for $i_1, j_1, i_2, j_2 \in G_n$ have the same distribution. Let therefore $\kappa(i, j) := |e_i \cap e_j| \in \{0, 1, 2\}$ be the number of common vertices of i and j (recall that for $i = (v, v')$, we defined $e_i = \{v, v'\}$ as the set containing the incident vertices). The following lemma is easy to prove.

Lemma 3.5. *Let* $i_1, j_1, i_2, j_2 \in G_n$*. There is a permutation of the vertices* σ *such that* $(i_1, j_1) = (\sigma(i_2), \sigma(j_2))$ *if and only if* $\kappa(i_1, j_1) = \kappa(i_2, j_2)$*.*

Proof. Let $i = (v_{i_1}, v'_{i_1})$ and analogously for i_2, j_1 and j_2. Let σ be a permutation such that $(i_1, j_1) = (\sigma(i_2), \sigma(j_2))$. Then,

$$\kappa(i_1, j_1) = \kappa(\sigma(i_2), \sigma(j_2)) = |\{\sigma(v_{i_2}), \sigma(v'_{i_2})\} \cap \{\sigma(v_{j_2}), \sigma(v'_{j_2})\}| = |e_{i_2} \cap e_{j_2}| = \kappa(i_2, j_2).$$

If $\kappa(i_1, j_1) = \kappa(i_2, j_2)$ we can easily construct σ with $(i_1, j_1) = (\sigma(i_2), \sigma(j_2))$ just by mapping the corresponding vertices onto each other and letting the other vertices unchanged. □

We obtain the following corollary which characterizes when two pairs of random variables are identically distributed.

Corollary 3.6. *Let $(Z_i)_{i\in G_n}$ be a sequence of exchangeable random variables indexed by the edges of a network G_n. For any vertices $i_1, j_1, i_2, j_2 \in G_n$, we have that (Z_{i_1}, Z_{j_1}) and (Z_{i_2}, Z_{j_2}) are identically distributed if $\kappa(i_1, j_1) = \kappa(i_2, j_2)$.*

As an application we can find simplified expression for the variances in undirected, complete networks.

Corollary 3.7. *For all $n \in \mathbb{N}$, let $G_n = (V_n, E_n)$ be undirected and complete networks and assume that $(Z_{n,i})_{i\in G_n}$ are interchangeable and square integrable. Recall that $r_n = |G_n| = \frac{n(n-1)}{2}$ is the number of edges. Then, for pairwise different vertices $v_1, v_2, v_3, v_4 \in V_n$,*

$$\begin{aligned}
&\mathit{Var}\left(\frac{1}{r_n}\sum_{i\in G_n} Z_{n,i}\right)\\
=&\frac{1}{r_n^2}\sum_{i\in G_n}\mathit{Var}(Z_{n,i}) + \frac{1}{r_n^2}\sum_{\substack{i,j\in G_n\\ \kappa(i,j)=1}}\mathit{Cov}(Z_{n,i}, Z_{n,j}) + \frac{1}{r_n^2}\sum_{\substack{i,j\in G_n\\ \kappa(i,j)=0}}\mathit{Cov}(Z_{n,i}, Z_{n,j})\\
=&r_n^{-1}\cdot \mathit{Var}(Z_{n,v_1v_2}) + O\left(r_n^{-\frac{1}{2}}\right)\mathit{Cov}(Z_{n,v_1v_2}, Z_{n,v_2v_3}) + O\left(1\right)\mathit{Cov}(Z_{n,v_1v_2}, Z_{n,v_3v_4}).
\end{aligned}$$

For the proof of this corollary we just need to think about the number of terms in each sum. It is an easy combinatorial exercise to find that their sizes are of the order $r_n, r_n^{\frac{3}{2}}$ and r_n^2 respectively.

If we were to assume that $Z_{n,i}$ and $Z_{n,j}$ are uncorrelated when $i \neq j$, the covariances vanish and we see that $\mathrm{Var}\left(\frac{1}{r_n}\sum_{i\in G_n} Z_{n,i}\right) \to 0$ as $n \to \infty$ if $r_n^{-1}\mathrm{Var}(Z_{n,v_1v_2}) \to 0$ as $n \to \infty$. The latter is not a very strong assumption while the former assumption (uncorrelation) shall be avoided as motivated in the beginning of this chapter. However, as motivated by the last corollary we do not need the asymptotic uncorrelation for all edges, we need it only for edges i and j with $\kappa(i,j) = 0$. For edges i and j with $\kappa(i,j) \geq 1$ we merely need that the covariances do not grow too fast. Thus, we make an assumption on the average behaviour of disjoint edges. For $\kappa(i,j) = 0$, we argue that $\mathrm{Cov}(Z_{n,v_1v_2}, Z_{n,v_3v_4}) \to 0$ is a reasonable assumption, because we believe that most edges are separated and do not strongly influence each other.

When we replace $\frac{1}{r_n}$ by $\frac{1}{\sqrt{r_n}}$ in the above corollary, we need to impose stronger conditions in order to achieve

$$\operatorname{Var}\left(\frac{1}{\sqrt{r_n}}\sum_{i\in G_n} Z_{n,i}\right) = O(1).$$

However, we may split the sum again and impose different assumptions on the covariances depending on the value of $\kappa(i,j)$. And again for edges i and j with a higher value of $\kappa(i,j)$, we may impose less strict assumptions. This is a realistic behaviour because $\kappa(i,j)$ reflects our intuitive understanding of how dependent two edges are.

As the formulation of this property depends on the exact application, we do not go into detail here. But we keep this intuition in mind when discussing specific assumptions of this type later.

3.2 Momentarily m-Dependent Networks

Later in our application to non-parametric kernel estimates, we will use kernels with bandwidth h to localize a likelihood. This will result in squares of sums of stochastic integrals of these kernels with respect to martingales. Expanding this square yields a double sum of double integrals with non-predictable integrands. More formally, we will encounter terms of the form

$$\frac{1}{r_n}\sum_{j_1,j_2\in G_n}\int_0^T\int_{t-2h}^{t-}\varphi_{n,j_1j_2}(t,r)dM_{n,j_2}(r)dM_{n,j_1}(t), \tag{3.1}$$

where $h\to 0$ and φ_{n,j_1j_2} averages the covariate functions $X_{n,i}$ around the time points t and r over an interval of length h. More precisely, φ_{n,j_1j_2} is a random (but bounded) function which depends on the processes $X_{n,j_1}(t)$, $X_{n,j_2}(r)$ and $(X_{n,i}(s),\ s\in[t,t+2h])$ as well as $(C_{n,i}(s),\ s\in[t,t+2h])$ for all $i\neq j_1,j_2$, and $M_{n,i}(t)=N_{n,i}(t)-\int_0^t\lambda_{n,i}(s)ds$ is the martingale associated with the counting process $N_{n,i}$. The main issue here is that the processes $\varphi_{n,j_1j_2}(t,r)$ with $r<t$ are not predictable with respect to $\mathcal{F}_t^n$ and so we cannot apply martingale results. If we could apply such results, we would use that the martingales are uncorrelated and the inner integration is of order h in order to show that the above term is negligible. But, conveniently, the processes φ_{n,j_1j_2} have a later described *leave-something-out* structure that allows us to apply a very similar technique as in Mammen and Nielsen (2007) in order to still obtain results about the convergence of the previously mentioned average.

Remark 3.8. *Throughout the remainder of this section we will use the notion of Stieltjes and Itô Integration interchangeably when possible (cf. Section 2.1). In particular, when φ is not predictable, we will understand $\int_0^T\varphi(t)dM_{n,i}(t)$ as a Stieltjes integral which is defined path-wise. Thus, no predictability of φ is required. If φ is predictable we can understand the same integral as Itô Integral and use its properties.*

We begin by defining what we mean by Momentary-m-Dependence. For a set $J\subseteq G_n$ of edges, let $d_s^n(i,J):=\min\{d_s^n(i,j):\ j\in J\}$ be the distance of i to J at time s.

Definition 3.9. *A structured interaction network* $(C_{n,i}, X_{n,i}, N_{n,i})_{i\in G_n}$ *with filtration* $(\mathcal{F}_t^n)_{t\in[0,T]}$ *and distance* d^n *is said to be* momentarily-m-dependent *for* $m \in [0,\infty)$*, if*

$$\begin{aligned}
&\forall n \in \mathbb{N},\, \forall t_0 \in [0, T-h],\, \forall J \subseteq G_n :\\
&\quad (C_{n,j}(t), X_{n,j}(t), N_{n,j}(t))_{j\in J,\, t\in[t_0,t_0+h]} \text{ is cond. independent of}\\
&\quad \sigma\Big((C_{n,i}(r), X_{n,i}(r), N_{n,i}(r)) \cdot \mathbb{1}(d_s^n(i,J) \geq m) :\\
&\qquad\quad s \leq t_0, r \leq s + 6h,\, i \in G_n\Big)\\
&\quad \text{given } \mathcal{F}_{t_0}^n.
\end{aligned}$$

Remark 3.10. *The choice of* $6h$ *in Definition 3.9 is so that we can use it directly in the kernel estimator setting.* h *is here the bandwidth of the kernel.*

Let us interpret this definition in view of the example given in Remark 3.4. Assume the network is momentarily m-dependent. Let $J = \{j_0\}$ in the above definition, then information about an edge i in the time interval $[t_0, t_0 + 6h]$ is not informative for the behaviour of edge j_0 on the time interval $[t_0, t_0 + h]$ provided that we know already the past up to time t_0 and that i and j_0 have distance at least m at time t_0. We illustrate this in Figure 3.1: The horizontal axis is time and the vertical axis is distance. The two lines correspond to two edges i and j and the distance between these two lines represents the distance between the edges i and j. Dots on the lines indicate events on the respective edge.

The two gray areas in the future (next to the line at t_0) stand for the information of the processes of i on the interval $[t_0, t_0 + h]$ and the processes of j on the interval $[t_0 + t_0 + h]$. Under the assumption of momentary-m-dependence these two are conditionally independent given the information up to time t_0. So there is no direct *information flow* between these two areas. However, they are not unconditionally independent because we can infer from the gray area in the future of j on its past when i and j where close, such that we can infer on the past of i which is informative about its future. But if we already know the past, then additional knowledge of the future of j is independent of the future of i.

In order to work with momentary-m-dependent networks, we will introduce two augmentations of the filtration $\mathcal{F}_t^n$. Generally, when extending filtrations, we have more predictable processes and fewer martingales. More precisely, let $(\mathcal{F}_t)_{t\in\mathcal{T}}$ be a filtration and let A be a predictable process and let M be a martingale, both with respect to $\mathcal{F}_t$. Let furthermore $(\mathcal{G}_t)_{t\in\mathcal{T}}$ be a filtration which extends $(\mathcal{F}_t)_{t\in\mathcal{T}}$ (that is, $\mathcal{F}_t \subseteq \mathcal{G}_t^n$ for all $t \in \mathcal{T}$). Then, A is still predictable with respect to $\mathcal{G}_t$ but M might not be a martingale with respect to $\mathcal{G}_t$. On the other hand, for processes A' and M' which are predictable and a martingale with respect to $\mathcal{G}_t$, respectively, it holds that M' is a martingale with respect to $\mathcal{F}_t$ but A' might not be predictable with respect to $\mathcal{F}_t$. Hence, predictability and martingale properties react in opposite ways to a change of information. In the following definition we introduce two extensions of $\mathcal{F}_t^n$ and one of which is the exact right

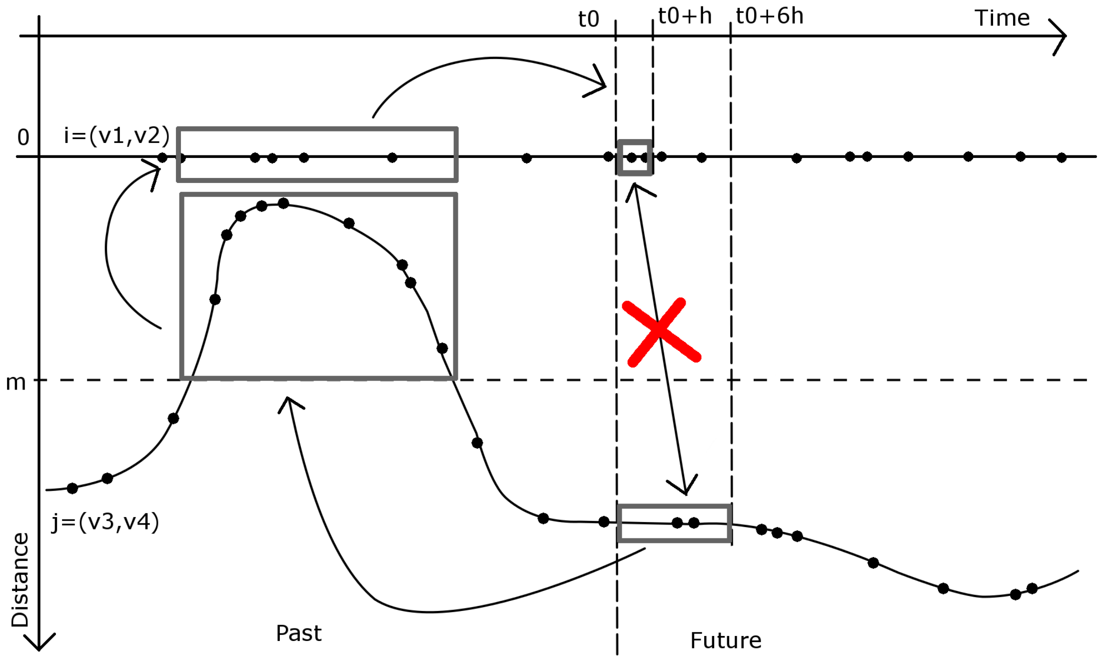


Figure 3.1: Graphical illustration of the information flow in an m-dependent network. In the figure time is progressing from left to right. The two lines represent the two edges i and j, the dots on the lines indicate events on the respective edge. The distance of the two lines is representative for the distance of i and j. The arrows show which areas my influence each other directly or not.

trade-off: Certain processes become predictable with respect to the extension while certain other processes remain martingales (this will be the content of the following Lemma 3.12).

For two σ-fields $\mathcal{A}$ and $\mathcal{B}$ we denote by $\mathcal{A} \vee \mathcal{B}$ the σ-field which is generated by the union of $\mathcal{A}$ and $\mathcal{B}$.

Definition 3.11. *Let $(C_{n,i}, X_{n,i}, N_{n,i})_{i \in G_n}$ be a structured interaction network with filtration $(\mathcal{F}_t^n)_{t \in [0,T]}$ and distance d^n. For a subset $J \subseteq G_n$ define*

$$\mathcal{F}_t^{n,J,m} := \mathcal{F}_t^n \vee \sigma\Big([C_{n,i}(r), X_{n,i}(r), N_{n,i}(r)] \, \mathbb{1}(d_s^n(i,J) \geq m) : \\ s \leq t, r \leq s + 6h, i \in G_n \Big).$$

We call $\mathcal{F}_t^{n,J,m}$ the long-sighted leave-J-out filtration. *In contrast to the* long-sighted leave-J-out filtration, *we also define a* short-sighted leave-J-out filtration $\widetilde{\mathcal{F}}_{I,t}^{n,J,m}$ *for $I \subseteq J$ by*

$$\begin{aligned} \widetilde{\mathcal{F}}_{I,t}^{n,J,m} := {} & \sigma(X_{n,i}(\tau) : \, i \in I, \tau \leq t) \\ & \vee \sigma([C_{n,i}(r), X_{n,i}(r), N_{n,i}(r)] \, \mathbb{1}(\forall j \in J : d_s^n(i,j) \geq m) : \\ & s \leq \max(0, t - 4h), r \leq s + 6h, i \in G_n). \end{aligned}$$

Denote further for any edge $j \in G_n$ and $m \in [0, \infty)$

$$j(m,t) := \{ i \in G_n : \, d_t^n(i,j) \geq m \}.$$

Functions which are measurable with respect to $\widetilde{\mathcal{F}}_{I,t}^{n,J,m}$ will be called of leave-m-out *type.*

Note that the difference between the long and short sighted leave-J-out filtration is given by which future knowledge is considered. Again, the choices of multiples of h are made such that we can apply the results directly to our situation without having notation clutter. Also, it holds that $\mathcal{F}_t^{n,J,m} \supseteq \widetilde{\mathcal{F}}_t^{n,J,m}$. We can now make the earlier mentioned property of the long-sighted leave-J-out filtration precise. Namely, that the counting processes stay counting processes and in particular their martingales are still martingales:

Lemma 3.12. *We consider a structured momentarily-m-dependent interaction network. For $J \subseteq G_n$, the processes $(N_{n,j}(t))_{j \in J}$ form a multivariate counting process with respective intensity functions $(C_{n,j}(t) \lambda_{n,j}(t))_{j \in J}$ with respect to $\mathcal{F}_t^{n,J,m}$. This means in particular that $(M_{n,j}(t))_{j \in J}$ (the counting process martingales) are martingales with respect to $\mathcal{F}_t^{n,J,m}$.*

Proof. Let $t > 0$ and $t' \in [t, t+h]$, then by definition and assumption

$$\mathbb{E}(M_{n,j}(t') | \mathcal{F}_t^{n,J,m}) = \mathbb{E}(M_{n,j}(t') | \mathcal{F}_t^n) = M_{n,j}(t).$$

This implies the assertion. □

We are now able to prove the two main results of this section. When dealing with the double sum in (3.1), we split the double sum in the diagonal sum (where $j_1 = j_2$) which is a single sum and the double sum over $j_1 \neq j_2$. The first main result (Proposition 3.13) is a simple statement about sums over one index, we will apply this to the diagonal sum as described above. For the second sum (the double sum) we need a more difficult result (Proposition 3.15) which is more complicated to prove.

Proposition 3.13. *Let $\varphi_{n,i} : [0,T] \to \mathbb{R}$ for $n \in \mathbb{N}, i \in G_n$ be random functions (not necessarily predictable). Let furthermore $\widetilde{\varphi}_j^I : [0,T] \to \mathbb{R}$ for $j \in I \subseteq G_n$ and $|I| \leq 2$ be of leave-m-out type, i.e., predictable with respect to $\widetilde{\mathcal{F}}_{j,t}^{n,I,m}$ for all $j \in I \subseteq G_n$ and $|I| \leq 2$. Then we have*

$$
\begin{aligned}
&\mathbb{E}\left(\left(\sum_{i\in G_n}\int_0^T \varphi_{n,i}(t)dM_{n,i}(t)\right)^2\right)\\
&\leq \sum_{i\in G_n}\int_0^T \mathbb{E}\left(\widetilde{\varphi}_{n,i}^i(t)^2 C_{n,i}(t)\lambda_{n,i}(\theta_0,t)\right)dt\\
&\quad + 2\sum_{i,j\in G_n}\mathbb{E}\left(\int_0^T \widetilde{\varphi}_{n,i}^{ij}(t)dM_{n,i}(t)\int_0^T\left(\varphi_{n,j}(r)-\widetilde{\varphi}_{n,j}^{ij}(r)\right)dM_{n,j}(r)\right)\\
&\quad + \sum_{i,j\in G_n}\mathbb{E}\left(\int_0^T\left(\varphi_{n,i}(t)-\widetilde{\varphi}_{n,i}^{ij}(t)\right)dM_{n,i}(t)\int_0^T\left(\varphi_{n,j}(r)-\widetilde{\varphi}_{n,j}^{ij}(r)\right)dM_{n,j}(r)\right).
\end{aligned}
$$

Proof. The proof is almost exactly along the lines of Mammen and Nielsen (2007) but it is not identical and we give it here for completeness. We see at first that

$$
\begin{aligned}
&\mathbb{E}\left(\left(\sum_{i\in G_n}\int_0^T \varphi_{n,i}(t)dM_{n,i}(t)\right)^2\right)\\
&= \sum_{i,j\in G_n}\mathbb{E}\left(\int_0^T\int_0^T \widetilde{\varphi}_{n,i}^{ij}(t)\widetilde{\varphi}_{n,j}^{ij}(r)dM_{n,i}(t)dM_{n,j}(r)\right) \qquad (3.2)\\
&\quad + 2\sum_{i,j\in G_n}\mathbb{E}\left(\int_0^T\int_0^T \widetilde{\varphi}_{n,i}^{ij}(t)\left(\varphi_{n,j}(r)-\widetilde{\varphi}_{n,j}^{ij}(r)\right)dM_{n,i}(t)dM_{n,j}(r)\right)\\
&\quad + \sum_{i,j\in G_n}\mathbb{E}\left(\int_0^T\int_0^T\left(\varphi_{n,i}(t)-\widetilde{\varphi}_{n,i}^{ij}(t)\right)\cdot\left(\varphi_{n,j}(r)-\widetilde{\varphi}_{n,j}^{ij}(r)\right)dM_{n,i}(t)dM_{n,j}(r)\right).
\end{aligned}
$$

We use now that $\widetilde{\varphi}_{n,i}^{ij}$ and $\widetilde{\varphi}_{n,j}^{ij}$ are both predictable with respect to $\mathcal{F}_t^{n,ij,m}$ and that $M_{n,i}$ and $M_{n,j}$ are both martingales with respect to the same filtration. Hence, we obtain

$$
(3.2) = \sum_{i\in G_n}\int_0^T \mathbb{E}\left(\widetilde{\varphi}_{n,i}^i(t)^2 C_{n,i}(t)\lambda_{n,i}(\theta_0,t)\right)dt
$$

and the statement follows. ∎

The following result allows a similar statement about double sums. We assume that the functions φ_{n,j_1j_2} may be approximated by functions $\widetilde{\varphi}^I_{n,j_1j_2}$ where $j_1, j_2 \in G_n$ and $I \subseteq G_n$ with $|I| \leq 4$. Later, the functions φ_{n,j_1j_2} will be averages of a certain stochastically bounded quantity over all edges (thus there is hope that the average is stochastically bounded too), and the functions $\widetilde{\varphi}^I_{n,j_1j_2}$ will be the same average but excluding edges in I from the sum. The the approximation error is roughly of order $\frac{|I|}{r_n}$ which converges to zero fast because $|I| \leq 4$. The assumptions of the following proposition should be seen under this light.

Remark 3.14. *We will denote the negation operator for events by $\neg$. So if A is an event then $\neg A$ denotes its negation. For a random variable X and a number x_0 we have for example*

$$\mathbb{1}(X \geq x_0) = 1 - \mathbb{1}(\neg X \geq x_0) = 1 - \mathbb{1}(X < x_0).$$

Proposition 3.15. *Let $m \in [0, \infty)$ and let $(C_{n,i}, X_{n,i}, N_{n,i})_{i \in G_n}$ be a momentarily m-dependent structured interaction network with filtration $(\mathcal{F}^n_t)_{t \in [0,T]}$ and distance d^n. Assume that we have random functions $\varphi_{n,j_1j_2} : [0,T] \times [0,T] \to \mathbb{R}$ which are not necessarily predictable with respect to $\mathcal{F}^n_t$. Let further $\widetilde{\varphi}^I_{n,j_1j_2}(t,r)$ be random functions which are of leave-m-out type in the following sense: The functions $\widetilde{\varphi}^I_{n,j_1j_2}$ are partially-predictable with respect to $\widetilde{\mathcal{F}}^{n,I,m}_{j_1j_2,t}$ for all $\{j_1, j_2\} \subseteq I \subseteq G_n$, $|I| \leq 4$.*

Assume that $\widetilde{\varphi}^I_{n,j_1j_2}$ approximates φ_{n,j_1j_2} in the following sense:

$$\frac{1}{r_n} \sum_{\substack{j_1,j_2 \in G_n \\ j_1 \neq j_2}} \int_0^T \int_{t-2h}^{t-} \left(\varphi_{n,j_1j_2}(t,r) - \widetilde{\varphi}^{j_1j_2}_{n,j_1j_2}(t,r) \right) dM_{n,j_2}(r) dM_{n,j_1}(t) = o_P(1) \tag{3.3}$$

$$\begin{aligned}
&\mathbb{E}\Bigg(\frac{1}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \int_0^T \int_{t-2h}^{t-} \left(\widetilde{\varphi}^{j_1j_2}_{n,j_1j_2}(t,r) - \widetilde{\varphi}^{j_1j_2j_1'j_2'}_{n,j_1j_2} \right)(t,r) dM_{n,j_2}(r) \\
&\quad \times \int_0^T \int_{t-2h}^{t-} \left(\widetilde{\varphi}^{j_1'j_2'}_{n,j_1'j_2'}(t,r) - \widetilde{\varphi}^{j_1j_2j_1'j_2'}_{n,j_1j_2}(t,r) \right) dM_{n,j_2'}(r) dM_{n,j_1'}(t) \Bigg) = o(1)
\end{aligned} \tag{3.4}$$

$$\begin{aligned}
&\frac{2}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\Bigg[\int_0^T \int_{t-2h}^{t-} \left(\widetilde{\varphi}^{j_1j_2}_{n,j_1j_2}(t,r) - \widetilde{\varphi}^{j_1j_2j_1'j_2'}_{n,j_1j_2}(t,r) \right) dM_{n,j_2}(r) \\
&\times \int_t^{t+2h} \int_{\xi-2h}^{\xi-} \widetilde{\varphi}^{j_1j_2j_1'j_2'}_{n,j_1'j_2'}(\xi,\rho) dM_{n,j_2'}(\rho) dM_{n,j_1'}(\xi) \mathbb{1}(\neg j_1', j_2' \in j_1(m,t)) dM_{n,j_1}(t) \Bigg] = o(1)
\end{aligned} \tag{3.5}$$

$$\frac{1}{r_n^2} \sum_{\substack{j_1,j_2 \in G_n \\ j_1 \neq j_2}} \int_0^T \int_{t-2h}^{t-} \mathbb{E}\Bigg[\widetilde{\varphi}^{j_1j_2}_{n,j_1j_2}(t,r)^2 C_{n,j_1}(t) \lambda_{n,j_1}(t) C_{n,j_2}(r) \lambda_{n,j_2}(r)$$

$$\times \mathbb{1}(j_2 \in j_1(m, t-2h))\Bigg] drdt = o(1) \tag{3.6}$$

$$\frac{1}{r_n^2} \sum_{\substack{j_1, j_2 \in G_n \\ j_1 \neq j_2}} \sum_{\substack{j_2' \in G_n \\ j_2' \neq j_2}} \int_0^T \mathbb{E}\Bigg[\int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1 j_2}^{j_1 j_2 j_2'}(t,r) dM_{n,j_2}(r) \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1 j_2'}^{j_1 j_2 j_2'}(t,r') dM_{n,j_2'}(r')$$

$$\times C_{n,j_1}(t) \lambda_{n,j_1}(t) \mathbb{1}(\neg j_2, j_2' \in j_1(m, t-2h))\Bigg] dt = o(1) \tag{3.7}$$

Then

$$\frac{1}{r_n} \sum_{\substack{j_1, j_2 \in G_n \\ j_1 \neq j_2}} \int_0^T \int_{t-2h}^{t-} \varphi_{n,j_1 j_2}(t,r) dM_{n,j_2}(r) dM_{n,j_1}(t) \xrightarrow{\mathbb{P}} 0. \tag{3.8}$$

Proof. The idea of the proof is to translate the convergence statement about $\varphi_{n,j_1 j_2}$ to statements about $\widetilde{\varphi}_{n,j_1 j_2}^I$. This will be useful because the latter are partially predictable with respect to the short sighted filtration. Since we have certain processes which are martingales with respect to the short sighted filtration (cf. Lemma 3.12) we can make use of martingale properties of the Itô Integral. For the first step, we see that the asymptotic behaviour of (3.8) is the same as the sum over the leave-m-out approximations, i.e.,

$$\begin{aligned}
&(3.8) \\
=&\frac{1}{r_n} \sum_{\substack{j_1, j_2 \in G_n \\ j_1 \neq j_2}} \int_0^T \int_{t-2h}^{t-} \varphi_{n,j_1 j_2}(t,r) - \widetilde{\varphi}_{n,j_1 j_2}^{j_1 j_2}(t,r) dM_{n,j_2}(r) dM_{n,j_1}(t) &(3.9)\\
&+ \frac{1}{r_n} \sum_{\substack{j_1, j_2 \in G_n \\ j_1 \neq j_2}} \int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1 j_2}^{j_1 j_2}(t,r) dM_{n,j_2}(r) dM_{n,j_1}(t) &(3.10)
\end{aligned}$$

and (3.9) converges to zero by assumption (3.3). Hence, we only have to study (3.10). $\widetilde{\varphi}_{n,j_1 j_2}^{j_1 j_2}(t,r)$ is partially-predictable with respect to the filtration $\mathcal{F}_t^{n,j_1 j_2,m}$ and, by the assumption of Momentary m-Dependence (c.f. Definition 3.9 and Lemma 3.12), $M_{n,j}$ is a martingale with respect to $\mathcal{F}_t^{n,J,m}$ for all $J \subseteq G_n$ with $j \in J$. We will use this observation in order to prove that (3.10) converges to zero in probability by applying Markov's Inequality:

$$\begin{aligned}
&\mathbb{E}((3.10)^2) \\
= &\frac{1}{r_n^2} \sum_{\substack{j_1, j_2, j_1', j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\Bigg[\int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1 j_2}^{j_1 j_2}(t,r) dM_{n,j_2}(r) dM_{n,j_1}(t) \\
&\qquad \times \int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1' j_2'}^{j_1' j_2'}(t,r) dM_{n,j_2'}(r) dM_{n,j_1'}(t) \Bigg]
\end{aligned}$$

$$
\begin{aligned}
= & \frac{1}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\Bigg[\int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2}(t,r) - \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2}(r) dM_{n,j_1}(t) \\
& \qquad \times \int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1'j_2'}^{j_1'j_2'}(t,r) - \widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2'}(r) dM_{n,j_1'}(t) \Bigg] \qquad (3.11) \\
+ & \frac{2}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\Bigg[\int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2}(t,r) - \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2}(r) dM_{n,j_1}(t) \\
& \qquad \times \int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2'}(r) dM_{n,j_1'}(t) \Bigg] \qquad (3.12) \\
+ & \frac{1}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\Bigg[\int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2}(r) dM_{n,j_1}(t) \\
& \qquad \times \int_0^T \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2'}(r) dM_{n,j_1'}(t) \Bigg] \qquad (3.13)
\end{aligned}
$$

We will treat the terms (3.11)-(3.13) separately. Note, that in contrast to (3.8), all of the above expressions contain only the approximations with their predictability property. We will show in the following how this is useful.

(3.11) converges to zero by assumption (3.4).

In order to see that (3.12) converges to zero, we note firstly that the two stochastic integrals in (3.12) (with respect to $M_{n,j_1}(t)$ and $M_{n,j_1'}(t)$) are martingales with respect to the correct leave-m-out filtrations (namely $\mathcal{F}_t^{n,j_1,m}$ and $\mathcal{F}_t^{n,j_1',m}$, respectively). Although these two filtrations are in general not the same, we can make use of the fact that the leave-m-out filtrations allow future knowledge. Define furthermore for Lebesgue sets $A \subseteq \mathbb{R}$

$$
\begin{aligned}
\bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) &:= \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2}(t,r) - \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) \\
I_A(j_1,j_2,j_1',j_2') &:= \int_{A\cap[0,T]} \int_{t-2h}^{t-} \bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2}(r) dM_{n,j_1}(t), \\
J_A(j_1,j_2,j_1',j_2') &:= \int_{A\cap[0,T]} \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2'}(r) dM_{n,j_1'}(t).
\end{aligned}
$$

Note that $M_{n,j_1'}$ and $M_{n,j_2'}$ are adapted with respect to all leave-m-out filtrations. Since $\widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r)$ is partially-predictable with respect to $\mathcal{F}_t^{n,j_1j_2j_1'j_2',m}$, we get that $t \mapsto \int_{t-2j}^{t-} \widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2'}(r)$ is predictable (cf. Lemma 2.17) and as a consequence, $t \mapsto J_{[0,t)}(j_1,j_2,j_1',j_2')$ is predictable as well with respect to $\mathcal{F}_t^{n,j_1j_2j_1'j_2',m}$.

With these definitions we have (if $\alpha > \beta$ we define $(\alpha, \beta] := \emptyset$)

(3.12)

$$=\frac{2}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\left[\int_0^T \int_{t-2h}^{t-} \bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2}(r) \cdot J_{[0,T]}(j_1,j_2,j_1',j_2') dM_{n,j_1}(t)\right]$$

$$=\frac{2}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\left[\int_0^T \int_{t-2h}^{t-} \bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2}(r) \cdot J_{[t,t+2h]}(j_1,j_2,j_1',j_2') dM_{n,j_1}(t)\right] \tag{3.14}$$

$$+\frac{2}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\left[\int_0^T \int_{t-2h}^{t-} \bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2}(r) \cdot J_{[0,t)}(j_1,j_2,j_1',j_2') dM_{n,j_1}(t)\right] \tag{3.15}$$

$$+\frac{2}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\left[\int_0^T \int_{t-2h}^{t-} \bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r) dM_{n,j_2}(r) \cdot J_{(t+2h,T]}(j_1,j_2,j_1',j_2') dM_{n,j_1}(t)\right] \tag{3.16}$$

We show that this is $o(1)$ by considering the tree lines separately. Recall therefore that $j_1(m,t) = \{i \in G_n : d_t^n(i, j_1) \geq m\}$ is the set of edges which are further away than m from j_1 at time t.

For (3.14), we see by definition of the leave-m-out approximations that

$$J_{[t,t+2h]}(j_1,j_2,j_1',j_2')\mathbb{1}(j_1', j_2' \in j_1(m,t))$$

is predictable with respect to $\mathcal{F}_t^{n,j_1,m}$ because: The integrators $M_{n,j_1'}$ and $M_{n,j_2'}$ in $J_{[t,t+2h]}(j_1,j_2,j_1',j_2')$ are only considered up to time at most $t+2h$ and $\mathcal{F}_t^{n,j_1,m}$ contains information up to and including time $t+6h$ for processes which are at time t at least of distance m to j_1. Now, the integrand in $J_{[t,t+2h]}(j_1,j_2,j_1',j_2')$ needs $X_{j_1'}(\tau)$ and $X_{j_2'}(\tau)$ for $\tau \leq t+2h$ which is well included in $\mathcal{F}_t^{n,j_1,m}$ by the same arguments. By assumption $\widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r)$ is partially-predictable with respect to $\widetilde{\mathcal{F}}_{j_1'j_2',t}^{n,j_1j_2j_1'j_2',m}$. Hence, $J_{[t,t+2h]}(j_1,j_2,j_1',j_2')\mathbb{1}(j_1', j_2' \in j_1(m,t))$ is predictable with respect to $\mathcal{F}_t^{n,j_1,m}$ by Lemma 2.17 because $\widetilde{\mathcal{F}}_{j_1'j_2',t+2h}^{n,j_1j_2j_1'j_2',m} \subseteq \mathcal{F}_q^{n,j_1,m}$ for a $q < t$. Let therefore $s \leq t-2h$ and $r \leq s+6h$, then $[C_{n,i}(r), X_{n,i}(r), N_{n,i}(r)] \cdot \mathbb{1}(d_s^n(i, \{j_1,j_2,j_1',j_2'\}) \geq m)$ is measurable with respect to $\mathcal{F}_q^{n,j_1,m}$ for $q = t-2h$ because $s \leq q$. Hence, $J_{[t,t+2h]}(j_1,j_2,j_1',j_2')\mathbb{1}(j_1', j_2' \in j_1(m,t))$ is predictable with respect to $\mathcal{F}_t^{n,j_1,m}$. Moreover, M_{n,j_1} is a martingale with respect to $\mathcal{F}_t^{n,j_1,m}$ by momentary m-dependence. Hence,

(3.14)

$$
\begin{aligned}
=&\frac{2}{r_n^2}\sum_{\substack{j_1,j_2,j_1',j_2'\in G_n\\ j_1\neq j_2,j_1'\neq j_2'}}\mathbb{E}\Bigg[\int_0^T\int_{t-2h}^{t-}\bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)J_{[t,t+2h]}(j_1,j_2,j_1',j_2')\\
&\qquad\times\mathbb{1}(j_1',j_2'\in j_1(m,t))dM_{n,j_1}(t)\Bigg]\\
&+\frac{2}{r_n^2}\sum_{\substack{j_1,j_2,j_1',j_2'\in G_n\\ j_1\neq j_2,j_1'\neq j_2'}}\mathbb{E}\Bigg[\int_0^T\int_{t-2h}^{t-}\bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)J_{[t,t+2h]}(j_1,j_2,j_1',j_2')\\
&\qquad\times(1-\mathbb{1}(j_1',j_2'\in j_1(m,t)))dM_{n,j_1}(t)\Bigg]\\
=&\frac{2}{r_n^2}\sum_{\substack{j_1,j_2,j_1',j_2'\in G_n\\ j_1\neq j_2,j_1'\neq j_2'}}\mathbb{E}\Bigg[\int_0^T\int_{t-2h}^{t-}\bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)J_{[t,t+2h]}(j_1,j_2,j_1',j_2')\\
&\qquad\times\mathbb{1}(\neg j_1',j_2'\in j_1(m,t))dM_{n,j_1}(t)\Bigg].
\end{aligned}
$$

The last part is $o(1)$ by assumption (3.5).

In (3.15), we see that $J_{[0,t)}(j_1,j_2,j_1'j_2')$ is predictable with respect to $\mathcal{F}_t^{n,j_1,m}\supseteq\mathcal{F}_t^{n,j_1j_2j_1'j_2',m}$. Thus, we conclude by using that M_{n,j_1} is a martingale with respect to $\mathcal{F}_t^{n,j_1,m}$ (with analogue arguments as in the first case):

$$(3.15)=0.$$

For (3.16), we note firstly that

$$
\begin{aligned}
&\int_0^T\int_{t-2h}^{t-}\bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)\cdot J_{(t+2h,T]}(j_1,j_2,j_1',j_2')dM_{n,j_1}(t)\\
=&\int_0^T\int_{t-2h}^{t-}\bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)\\
&\quad\times\int_0^T\int_{\xi-2h}^{\xi-}\widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(\xi,\rho)dM_{n,j_2'}(\rho)\mathbb{1}(\xi>t+2h)dM_{n,j_1'}(\xi)dM_{n,j_1}(t)\\
=&\int_0^T\int_{\xi-2h}^{\xi-}\widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(\xi,\rho)dM_{n,j_2'}(\rho)\\
&\quad\times\int_0^T\int_{t-2h}^{t-}\bar{\Phi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)\mathbb{1}(t<\xi-2h)dM_{n,j_1}(t)dM_{n,j_1'}(\xi)\\
=&\int_0^T\int_{\xi-2h}^{\xi-}\widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(\xi,\rho)dM_{n,j_2'}(\rho)\cdot I_{[0,\xi-2h)}(j_1,j_2,j_1',j_2')dM_{n,j_1'}(\xi).
\end{aligned}
$$

Now, we can play a similar game: This time, $M_{n,j_1'}$ is a martingale with respect to $\mathcal{F}_\xi^{n,j_1',m}$. Furthermore, $I_{[0,\xi-2h)}(j_1,j_2,j_1',j_2')$ requires knowledge of $M_{n,j_1}(\tau)$, $M_{n,j_2}(\tau)$, $X_{n,j_1}(\tau)$ and $X_{n,j_2}(\tau)$ for $\tau<\xi-2h$ which is included in $\mathcal{F}_\xi^{n,j_1',m}$ as well as knowledge of $[N_{n,i}(r),X_{n,i}(r),C_{n,i}(r)]\cdot\mathbb{1}(d_s^n(i,\{j_1,j_2\})\geq m)$ for $s\leq\xi-6h$ and $r\leq s+6h$, i.e., $r\leq\xi$ which is again included in $\mathcal{F}_\xi^{n,j_1',m}$. Hence, $\xi\mapsto I_{[0,\xi-2h)}(j_1,j_2,j_1',j_2')$ is predictable with respect to $\mathcal{F}_\xi^{n,j_1',m}$. Hence, the integrand of (3.16) is a martingale and we obtain

$$(3.16)=0.$$

Thus, we have shown that $(3.12)=o(1)$.

Finally, we consider (3.13). Therefore note firstly that $\tilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)$ and $\tilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r)$ are both partially-predictable with respect to $\mathcal{F}_t^{n,j_1j_2j_1'j_2',m}$. Moreover, M_{n,j_1}, M_{n,j_2}, $M_{n,j_1'}$ and $M_{n,j_2'}$ are all martingales with respect to $\mathcal{F}_t^{n,j_1j_2j_1'j_2',m}$. Hence,

$$\int_{t-2h}^{t-}\tilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)$$

is also a predictable function in t and

$$\int_0^T\int_{t-2h}^{t-}\tilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)dM_{n,j_1}(t)$$

is a martingale. The same holds when M_{n,j_1} and M_{n,j_2} are replaced by $M_{n,j_1'}$ and $M_{n,j_2'}$. Hence, for $j_1\neq j_1'$

$$\begin{aligned}&\mathbb{E}\Bigg[\int_0^T\int_{t-2h}^{t-}\tilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2}(r)dM_{n,j_1}(t)\\&\qquad\times\int_0^T\int_{t-2h}^{t-}\tilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_1'j_2'}(t,r)dM_{n,j_2'}(r)dM_{n,j_1'}(t)\Bigg]\\=\;&0.\end{aligned}$$

For $j_1=j_1'$ we will apply firstly a martingale result to compute the covariance of the two stochastic integrals (first equality below), in the third equality below we employ a similar technique as in the computations for (3.12): Note that $C_{n,j_1}(t)\lambda_{n,j_1}(t)\mathbb{1}(j_2,j_2'\in j_1(m,t-2h))$ is measurable with respect to $\mathcal{F}_{t-2h}^{n,j_2j_2',m}$, additionally M_{n,j_2} and $M_{n,j_2'}$ are martingales with respect to $\mathcal{F}_t^{n,j_2j_2',m}$. Hence,

$$\begin{aligned}&\mathbb{E}\Bigg[\int_0^T\int_{t-2h}^{t-}\tilde{\varphi}_{n,j_1j_2}^{j_1j_2j_2'}(t,r)dM_{n,j_2}(r)dM_{n,j_1}(t)\\&\qquad\times\int_0^T\int_{t-2h}^{t-}\tilde{\varphi}_{n,j_1j_2'}^{j_1j_2j_2'}(t,r)dM_{n,j_2'}(r)dM_{n,j_1}(t)\Bigg]\end{aligned}$$

$$
\begin{aligned}
&= \int_0^T \mathbb{E}\Bigg[\int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_2'}(t,r)dM_{n,j_2}(r) \\
&\qquad \times \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2'}^{j_1j_2j_2'}(t,r')dM_{n,j_2'}(r')C_{n,j_1}(t)\lambda_{n,j_1}(t)\Bigg]dt \\
&= \int_0^T \mathbb{E}\Bigg[\int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_2'}(t,r)dM_{n,j_2}(r) \\
&\qquad \times \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2'}^{j_1j_2j_2'}(t,r')C_{n,j_1}(t)\lambda_{n,j_1}(t)\mathbb{1}(j_2,j_2' \in j_1(m,t-2h))dM_{n,j_2'}(r')\Bigg]dt \\
&+ \int_0^T \mathbb{E}\Bigg[\int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_2'}(t,r)dM_{n,j_2}(r) \\
&\qquad \times \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2'}^{j_1j_2j_2'}(t,r')dM_{n,j_2'}(r')C_{n,j_1}(t)\lambda_{n,j_1}(t)\mathbb{1}(\neg j_2,j_2' \in j_1(m,t-2h))\Bigg]dt
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{1}(j_2=j_2') \int_0^T \int_{t-2h}^{t-} \mathbb{E}\Bigg[\widetilde{\varphi}_{n,j_1j_2}^{j_1j_2}(t,r)^2 C_{n,j_1}(t)\lambda_{n,j_1}(t)C_{n,j_2}(r)\lambda_{n,j_2}(r) \\
&\qquad \times \mathbb{1}(j_2 \in j_1(m,t-2h))]\Bigg]drdt \\
&+ \int_0^T \mathbb{E}\Bigg[\int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_2'}(t,r)dM_{n,j_2}(r) \\
&\qquad \times \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2'}^{j_1j_2j_2'}(t,r')dM_{n,j_2'}(r')C_{n,j_1}(t)\lambda_{n,j_1}(t)\mathbb{1}(\neg j_2,j_2' \in j_1(m,t-2h))\Bigg]dt
\end{aligned}
$$

So we may rewrite

$$
\begin{aligned}
&(3.13) \\
&= \frac{1}{r_n^2} \sum_{\substack{j_1,j_2\in G_n \\ j_1\neq j_2}} \int_0^T \int_{t-2h}^{t-} \mathbb{E}\Bigg[\widetilde{\varphi}_{n,j_1j_2}^{j_1j_2}(t,r)^2 C_{n,j_1}(t)\lambda_{n,j_1}(t)C_{n,j_2}(r)\lambda_{n,j_2}(r) \\
&\quad \times \mathbb{1}(j_2 \in j_1(m,t-2h))\Bigg]drdt \\
&+ \frac{1}{r_n^2} \sum_{\substack{j_1,j_2\in G_n \\ j_1\neq j_2}} \sum_{\substack{j_2'\in G_n \\ j_2'\neq j_2}} \int_0^T \mathbb{E}\Bigg[\int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_2'}(t,r)dM_{n,j_2}(r) \\
&\qquad \times \int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2'}^{j_1j_2j_2'}(t,r')dM_{n,j_2'}(r')C_{n,j_1}(t)\lambda_{n,j_1}(t)\mathbb{1}(\neg j_2,j_2' \in j_1(m,t-2h))\Bigg]dt
\end{aligned}
$$

By the assumptions (3.6) and (3.7) we conclude

$$(3.13) = o(1).$$

Thus we have finally shown that (3.10) $\xrightarrow{\mathbb{P}} 0$ and hence the proof is complete. $\square$

3.3 Mixing Networks

So far we have discussed conditional independence. But it is also useful to talk about unconditional independence. We argued in the previous section (see e.g. the discussion of Figure 3.1) that, from a modelling point of view, we might feel uncomfortable to assume unconditional independence. But, we probably feel comfortable to assume that far apart actors influence each other very weakly. We include this aspect in the model by imposing mixing assumptions. The setting is as in the previous subsection: Consider a structured interaction network $(C_{n,i}, X_{n,i}, N_{n,i})_{i\in G_n}$ with filtration $(\mathcal{F}_t^n)_{t\in[0,T]}$ and distance function d^n. Let $(Z_i)_{i\in G_N}$ be a set of random variables indexed by the edges with values in a space $\mathcal{X}$. Ultimately, we are interested in proving inequalities of Bernstein type (cf. the following Proposition 3.18), i.e.,

$$\mathbb{P}\left(\sum_{i\in G_n} f(Z_i) - \mathbb{E}(f(Z_i)) \geq t\right) \leq \exp\left(-\frac{t^2}{C\cdot t + D}\right), \tag{3.17}$$

where $f : \mathcal{X} \to \mathbb{R}$ is a function (the constants C and D depend on f). We intend to prove this type of inequality by applying the grouping technique for mixing random variables (cf. Rio (2017) and Doukhan (1994)). Such exponential inequalities will become useful when we study the global behaviour of the non-parametric estimator (cf. Theorem 4.3). In contrast to classical β-mixing arguments our treatment is based on a time-varying and random distance function d_t^n. In the network context we consider, it is very natural to assume that actors change their dependencies. Therefore, we consider the distance d_t^n as time dependent. However, we start with a review of important results and definitions about β-Mixing which we need later. Like in Rio (2017), we define the general β-Mixing coefficient as follows.

Definition 3.16. *Let $\mathcal{A}$ and $\mathcal{B}$ be two σ-fields in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by $\mathbb{P}_{\mathcal{A}\otimes\mathcal{B}}$ the unique measure on $(\Omega\times\Omega, \mathcal{A}\otimes\mathcal{B})$ with the property that for any $A \in \mathcal{A}$ and $B \in \mathcal{B}$: $\mathbb{P}_{\mathcal{A}\otimes\mathcal{B}}(A\times B) = \mathbb{P}(A\cap B)$. Denote by $\mathbb{P}_{\mathcal{A}}$ and $\mathbb{P}_{\mathcal{B}}$ the restrictions of $\mathbb{P}$ to $\mathcal{A}$ and $\mathcal{B}$ respectively. Denote finally by $\mathbb{P}_{\mathcal{A}}\otimes\mathbb{P}_{\mathcal{B}}$ the unique measure on $(\Omega\times\Omega, \mathcal{A}\otimes\mathcal{B})$ with $(\mathbb{P}_{\mathcal{A}}\otimes\mathbb{P}_{\mathcal{B}})(A\times B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A\in\mathcal{A}$ and $B\in\mathcal{B}$. The β-Mixing coefficient between $\mathcal{A}$ and $\mathcal{B}$ is then defined by*

$$\beta(\mathcal{A},\mathcal{B}) := \sup_{C\in\mathcal{A}\otimes\mathcal{B}} \left|\mathbb{P}_{\mathcal{A}\otimes\mathcal{B}}(C) - (\mathbb{P}_{\mathcal{A}}\otimes\mathbb{P}_{\mathcal{B}})(C)\right|.$$

For two random variables X and Y defined on the same probability space we define

$$\beta(X,Y) := \beta(\sigma(X),\sigma(Y)).$$

1	2	1	2	1
3	4	3	4	3
1	2	1	2	1
3	4	3	4	3
1	2	1	2	1

Figure 3.2: Partition of the two-dimensional plane: Blocks of the same type have the same number.

The β-Mixing coefficient will be interesting for us through the following lemma which is also taken from Rio (2017) (Lemma 5.1 therein).

Lemma 3.17. *Let $\mathcal{A}$ be a σ-field in $(\Omega, \mathcal{F}, \mathbb{P})$ and let X be a random variable with values in a Polish space $\mathcal{X}$. Let δ be a random variable with uniform distribution over $[0,1]$ which is independent of the σ-field generated by $\mathcal{A}$ and X. Then, there exists a random variable X^* which has the same law as X and which is independent of $\mathcal{A}$, such that $\mathbb{P}(X \neq X^*) = \beta(\mathcal{A}, \sigma(X))$. Furthermore, X^* is measurable with respect to the σ-field generated by $\mathcal{A}$ and (X, δ).*

We also require, as a general result, the Bernstein inequality which is taken from Giné and Nickl (2016).

Proposition 3.18. *Let X_i, $i = 1, ..., n$ be a sequence of independent, centred random variables such that there are numbers c and σ_i such that for all k*

$$\mathbb{E}(|X_i|^k|) \leq \frac{k!}{2}\sigma_i^2 c^{k-2}.$$

Set $\sigma^2 := \sum_{i=1}^n \sigma_i^2$, $S_n := \sum_{i=1}^n X_i$. Then, for all $t \geq 0$

$$\mathbb{P}(S_n \geq t) \leq \exp\left(-\frac{t^2}{2(\sigma^2 + ct)}\right).$$

The idea is now to group the random variables Z_i in blocks which have a large distance amongst each other, then we can apply Lemma 3.17 to these blocks to obtain independent copies of the blocks. Thus, we split the edges into blocks and assign each block one of

$\mathcal{K}$ types. Two vertices from different blocks of the same type shall be far apart. More formally:

Definition 3.19. *Let $\Delta > 0$, $t \in [0,T]$, $\mathcal{K}, n, m \in \mathbb{N}$ and $k \in \{1, ..., \mathcal{K}\}$. We call the sets $G^t(k,m,\Delta) \subseteq G_n$ a Δ-partitioning of the network at time t (note that we omit n in the notation) if*

1. $(k,m) \neq (k',m') \Rightarrow G^t(k,m,\Delta) \cap G^t(k',m',\Delta) = \emptyset$,
2. $\bigcup_{k=1}^{\mathcal{K}} \bigcup_m G^t(k,m,\Delta) = G_n$,
3. *For* $k \in \{1, ..., \mathcal{K}\}$ *and* $m \neq m'$: $i \in G^t(k,m,\Delta)$, $j \in G^t(k,m',\Delta) \Rightarrow d_t^n(i,j) \geq \Delta$.

Remark 3.20. *The distance function d_t^n is only a theoretical construct (cf. Remark 3.3). In the same way the partitioning is also only assumed to exist but we do not require knowledge of it.*

In order to illustrate that such a partitioning may exist, we give two examples on how to find it. For both examples, consider the two-dimensional plane in Figure 3.2 which we split in blocks (squares) of side length Δ and each block is assigned one of four types. In Figure 3.2 all blocks of the same type $k \in \{1, ..., 4\}$ have been assigned the same number. It is clear that the distance between two points taken from two different blocks of the same type k is at least Δ. We assign numbers $\{1, 2, 3, ...\}$ to all blocks of the same type in Figure 3.2 such that we can speak of the m-th block of type k.

The idea is now to assign to each edge random, two-dimensional coordinates. Then, the edges can be plotted in the two-dimensional plane and can be partitioned accordingly. We make this precise in the next example.

Example 3.21. *Let $e, e' \in G_n$ be two arbitrary edges. For any $n \in \mathbb{N}, t \in [0,T]$ and $i \in G_n$ we call $(d_t^n(i,e), d_t^n(i,e'))$ the coordinates of i at time t. Let $G^t(k,m,\Delta)$ for $k = 1,2,3,4$ and $m \in \mathbb{N}$ comprise all edges i with coordinates lying in the m-th block of type k in Figure 3.2.*

Note that above we construct the partitioning for each time point t individually. Hence, the choice of the reference edges e and e' may depend on time as well. At every time point we may choose new reference edges e and e' which are most suitable. It is not clear what a good choice of reference edges is. It would be desirable if the assigned coordinates of the other edges are spread out through the plane. Two reference edges e and e' which are very close would probably not be a good choice because then we would expect that for any other edge $i \in G_n$ we have $d_t^n(i,e) \approx d_t^n(i,e')$. Hence, all coordinates $(d_t^n(i,e), d_t^n(i,e'))$ would basically only lie on the diagonal of the plane. Therefore we would try to choose edges which are far apart as reference edges. Such a random choice of reference edges is not contradicting the definition of a Δ-partitioning because Δ-partitionings are allowed to be random.

That we produce indeed a Δ-partitioning in the above example (for every choice of reference edges) is ensured by the following Lemma.

Lemma 3.22. *Let $\Delta > 0$ be given. The sets $G^t(k, m, \Delta)$ defined in Example 3.21 form a Δ-partitioning of the network in the sense of Definition 3.19.*

Proof. That $G^t(k, m, \Delta)$ and $G^t(k', m', \Delta)$ are disjoint for $(k, m) \neq (k', m')$ is obvious because the distance between two edges and hence their coordinates are unique. That the sets $G^t(k, m, \Delta)$ are exhaustive is also clear.

Let $i, j \in G_n$ and denote by $(q, r) := (d_t^n(i, e), d_t^n(i, e'))$ and $(q', r') := (d_t^n(j, e), d_t^n(j, e'))$ their respective coordinates. Then we obtain by the triangle inequality

$$\begin{aligned} q' = d_t^n(e, j) &\leq d_t^n(e, i) + d_t^n(i, j) = q + d_t^n(i, j) \\ q = d_t^n(e, i) &\leq d_t^n(e, j) + d_t^n(j, i) = q' + d_t^n(i, j), \end{aligned}$$

which yields $d_t^n(i, j) \geq |q - q'|$. Analogously, we obtain $d_t^n(i, j) \geq |r - r'|$. The third condition in Definition 3.19 follows then immediately if we notice that by definition for $m \neq m'$, $i \in G^t(k, m, \Delta)$ and $j \in G^t(k, m', \Delta)$ implies that $|q - q'|, |r - r'| \geq \Delta$. □

Additionally to Example 3.21, we provide another method of how to equip edges with two-dimensional coordinates. Once we have these coordinates, we can proceed in the same way as in Example 3.21. In this second example we intend to obtain the two dimensional coordinates via multidimensional scaling.

Example 3.23. *Every pair of edges (i, j) is at time t equipped with a distance $d_t^n(i, j)$. Our aim is to plot for every edge $i \in G_n$ a point $p(i)$ in the two-dimensional plane such that for any two edges $i, j \in G_n$ the Euclidean distance between $p(i)$ and $p(j)$ equals $d_t^n(i, j)$, i.e., the distance of i and j in the graph. This is exactly the task of Multidimensional Scaling (MDS) (cf. Cox and Cox (1994)). So the result of an MDS algorithm could be used to plot the edges in the two-dimensional plane and group them together as in Example 3.21.*

Remark 3.24. *Suppose that the projection is indeed possible, i.e., that the Euclidean distance between $p(i)$ and $p(j)$ equals $d_t^n(i, j)$. Then, it is clear that we really produced a Δ-partitioning (this is by Definition of the partitioning of the plane, see the discussion before Example 3.21 and Figure 3.2). However, multidimensional scaling techniques usually only produce an approximate projection, i.e., the Euclidean distance between $p(i)$ and $p(j)$ is only approximately equal to $d_t^n(i, j)$. So we possible do not obtain a Δ-partitioning. However, due to the discrete structure of the partitioning, there are probably only a few edges which violate the third Δ-partitioning condition in Definition 3.19 and we can possibly correct for them. We will not pursue this issue any further. Instead we will assume that we have an exact Δ-partitioning.*

We define now the mixing coefficients on the network based on general Δ-partitions. In particular for the following discussion it does not matter how the Δ-partitioning was created.

Definition 3.25. *Let $(Z_i)_{i \in G_n}$ be a sequence of random variables indexed by the edges, let $\Delta > 0$ be given and let $G^t(k, m, \Delta)$ be a Δ-partition of the network as in Definition*

3.19. For every time point t and edges $i \in G_n$, we define

$$I_{n,i}^{k,m,t}(\Delta) := \mathbb{1}(i \in G^t(k,m,\Delta)),$$

the indicator function which checks if i belongs to the m-th block of type k at time t. Group the Z_i based on the partition $G^t(k,m,\Delta)$, i.e.,

$$U_{k,m}^{n,t}(\Delta) := \sum_{i\in G_n} Z_i \cdot I_{n,i}^{k,m,t}(\Delta) - \mathbb{E}(Z_i \cdot I_{n,i}^{k,m,t}(\Delta)).$$

Then we define the β-Mixing coefficient which depends on the graph partitioning $G^t(k,m,\Delta)$ (which we do not indicate in the notation) via:

$$\beta_t(\Delta) := \max_{\substack{M,M'\subseteq\mathbb{N}, M\cap M'=\emptyset \\ k\in\{1,\dots,\mathcal{K}\}}} \beta\left(\left[U_{k,m}^{n,t}(\Delta)\right]_{m\in M}, \left[U_{k,m'}^{n,t}(\Delta)\right]_{m'\in M'}\right).$$

Remark 3.26. *In applications, the random variables Z_i will depend on a time point $t_0 \in [0,T]$. So it will be the case that for t close to t_0 the β-Mixing coefficients at time t will be small while they might be large for t far away from t_0 (in fact, for t far away from t_0 we will sometimes have $Z_i = 0$ and hence the β-Mixing coefficients will equal 1).*

The following result is the main result of this section (inspired by Doukhan (1994)).

Proposition 3.27. *Let $(C_{n,i}, X_{n,i}, N_{n,i})_{i\in G_n}$ be a structured interaction network with filtration $(\mathcal{F}_t^n)_{t\in[0,T]}$ and distance d^n. Consider furthermore a set of random variables $(Z_i)_{i\in G_n}$ indexed by the edges. With the same notation as in Definition 3.25 assume that there is a time point t such that there is a Δ-partitioning such that for all $\rho \in \mathbb{N}$ with $\rho \geq 2$ and all $k \in \{1,...,\mathcal{K}\}$ and $m \in \{1,...,r_n\}$*

$$\mathbb{E}(|U_{k,m}^{n,t}(\Delta)|^\rho) \leq \frac{\rho!}{2} E_{k,m}\sigma^2 \cdot (E_k C)^{\rho-2},$$

for some numbers $\sigma^2, E_{k,m}, E_k$ and C with $|E|_n := \sum_{k=1}^{\mathcal{K}} \sum_{m=1}^{r_n} E_{k,m} < +\infty$ (note that all these numbers may depend on Δ). Then,

$$\mathbb{P}\left(\sum_{i\in G_n} (Z_i - \mathbb{E}(Z_i)) \geq x\right) \leq \sum_{k=1}^{\mathcal{K}} \exp\left(-\frac{|E|_n^{-1}\sum_{m=1}^{r_n} E_{k,m}x^2}{2(|E|_n\sigma^2 + E_k Cx)}\right) + \beta_t(\Delta)\cdot \mathcal{K} r_n.$$

Proof. With the definitions as in Definition 3.25 we obtain (because $\sum_k \sum_m I_{n,i}^{k,m,t}(\Delta) = 1$)

$$\begin{aligned}\sum_{i\in G_n} (Z_i - \mathbb{E}Z_i) &= \sum_k \sum_m \sum_{i\in G_n} Z_i I_{n,i}^{k,m,t}(\Delta) - \mathbb{E}(Z_i I_{n,i}^{k,m,t}(\Delta)) \\ &= \sum_{k=1}^{\mathcal{K}} \sum_{m=1}^{r_n} U_{k,m}^{n,t}(\Delta). \qquad (3.18)\end{aligned}$$

In order to reduce notation, we omit (Δ) when talking about $U^{n,t}_{k,m}(\Delta)$. By Lemma 3.17 we can construct sequences $U^*_{k,m}$ as follows: We assume that the σ-field $\mathcal{F}^n_t$ is rich enough to allow for independent extra random variables $\delta_{k,m}$ which are uniformly distributed on $[0,1]$ and which are independent amongst each other and of everything else. The construction is the same for every k, so we only construct the sequence $U^*_{1,m}$, all other sequences $U^*_{k,m}$ for $k \geq 2$ are constructed analogously. Define $U^*_{1,1} := U_{1,1}$. For $m \geq 2$ there is by Lemma 3.17 a function f_m such that $U^*_{1,m} := f_m(U_{1,1}, ..., U_{1,m-1}, \delta_{1,m}, U_{1,m})$ has the same distribution as $U_{1,m}$, is independent of $U_{1,1}, ..., U_{1,m-1}$ and

$$\mathbb{P}(U_{1,m} \neq U^*_{1,m}) = \beta\left(\left(U_{1,1}, ..., U_{1,m-1}\right), U_{1,m}\right) \leq \beta_t(\Delta).$$

To sum it up, we have sequences $U^*_{k,m}$ with

1. For any k and any fixed $R \in \mathbb{N}$, $\left(U^*_{k,m}\right)_{r=1,...,R}$ is a sequence of independent random variables.
2. $U^*_{k,m}$ and $U_{k,m}$ have the same distribution.
3. For all $k = 1, ..., \mathcal{K}$: $\mathbb{P}\left(\exists m \in \{1, ..., r_n\} : U_{k,m} \neq U^*_{k,m}\right) \leq r_n \cdot \beta_t(\Delta)$.

Denote by R_k the random number of blocks $U_{k,m}$ of type k which exist, i.e., such that for $m > R_k$ we have $U_{k,m} = 0$. So we obtain by (3.18) for any $x \geq 0$ and any sequence $(\alpha_k)_{k=1,...,\mathcal{K}}$ with $\sum_{k=1}^{\mathcal{K}} \alpha_k = 1$ and $\alpha_k \geq 0$:

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i\in G_n} Z_i - \mathbb{E}(Z_i) \geq x\right) \\
= \quad &\mathbb{P}\left(\sum_{k=1}^{\mathcal{K}}\sum_{m=1}^{R_k} U_{k,m} \geq x\right) \\
\leq \quad &\mathbb{P}\left(\sum_{k=1}^{\mathcal{K}}\sum_{m=1}^{R_k} U^*_{k,m} \geq x\right) + \mathbb{P}\left(\exists k \in \{1, ..., \mathcal{K}\}, m \in \{1, ..., r_n\} : U_{k,m} \neq U^*_{k,m}\right) \\
\leq \quad &\mathbb{P}\left(\sum_{k=1}^{\mathcal{K}}\sum_{m=1}^{R_k} U^*_{k,m} \geq x\right) + \sum_{k=1}^{\mathcal{K}} \mathbb{P}(\exists m \in \{1, ..., r_n\} : U_{k,m} \neq U^*_{k,m}) \\
\leq \quad &\sum_{k=1}^{\mathcal{K}} \mathbb{P}\left(\sum_{m=1}^{R_k} U^*_{k,m} \geq \alpha_k \cdot x\right) + \beta_t(\Delta, v, v') \cdot \mathcal{K} r_n
\end{aligned} \tag{3.19}$$

Note that $R_k \leq r_n$ and that $U_{k,m} = 0$ for $m > R_k$. Furthermore, for every k the sequence $U^*_{k,m}$ is a sequence of independent random variables. Moreover, by definition $\mathbb{E}(U_{k,m}) = 0$. So, the assumptions of Proposition 3.18 are fulfilled with $\sigma_m^2 := E_{k,m}\sigma^2$ and $c := E_k C$. So we can estimate the first part of (3.19) by

$$\mathbb{P}\left(\sum_{m=1}^{R_k} U^*_{k,m} \geq \alpha_k \cdot x\right)$$

$$\leq \quad \exp\left(-\frac{\alpha_k^2 x^2}{2\left(\sum_{m=1}^{r_n} E_{k,m}\sigma^2 + E_k C\cdot\alpha_k x\right)}\right). \tag{3.20}$$

When we chose $\alpha_k = |E|_n^{-1}\sum_{m=1}^{r_n} E_{k,m}$, we obtain by combining the equalities (3.19) and (3.20),

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i\in G_n}(Z_i - \mathbb{E}(Z_i)) \geq x\right)\\
\leq\ & \sum_{k=1}^{\mathcal{K}} \exp\left(-\frac{\alpha_k^2 x^2}{2\left(\sum_{m=1}^{r_n} E_{k,m}\sigma^2 + E_k C\alpha_k x\right)}\right) + \beta_t(\Delta)\cdot\mathcal{K}r_n\\
\leq\ & \sum_{k=1}^{\mathcal{K}} \exp\left(-\frac{|E|_n^{-1}\sum_{m=1}^{r_n} E_{k,m}x^2}{2\left(|E|_n\sigma^2 + E_k C x\right)}\right) + \beta_t(\Delta)\cdot\mathcal{K}r_n
\end{aligned}$$

□

This proposition can be utilized for processes on structured interaction networks with the following property.

Definition 3.28. *A structured interaction network $(C_{n,i}, X_{n,i}, N_{n,i})_{i\in G_n}$ with filtration $(\mathcal{F}_t^n)_{t\in[0,T]}$ and distance d^n together with a sequence of stochastic process on $[0,T]$, $(Z_{n,i})_{i\in G_n}$, indexed by the edges (i.e., for every $i \in G_n$, $Z_{n,i}$ is a stochastic process on $[0,T]$) is called* β-mixing *at time t if the following conditions on the quantities defined in Definition 3.25 hold:*

For $a \in \mathbb{R}$ and $\Delta_n := a\cdot\log n$, there is a Δ_n-partitioning and positive numbers $\sigma^2, c_1, c_2, c_3 \in \mathbb{R}_{>0}$ and $E_{k,m}^{n,t}, E_k^{n,t} \in \mathbb{R}$ for $k = 1,..,\mathcal{K}$ and $m = 1,...,r_n$, such that

1. $\forall k \in \{1,...,\mathcal{K}\}, m \in \{1,...,r_n\} : \mathbb{E}(|U_{k,m}^{n,t}(\Delta_n)|^\rho) \leq \frac{\rho!}{2} E_{k,m}^{n,t}\sigma^2\cdot\left(E_k^{n,t}c_1\right)^{\rho-2}$
2. *For $|E|_{n,t} := \sum_{k=1}^{\mathcal{K}}\sum_{m=1}^{r_n} E_{k,m}^{n,t}$, it holds that*

$$\forall k = 1,...,\mathcal{K}: \quad \frac{1}{|E|_{n,t}}\sum_{m=1}^{r_n} E_{k,m}^{n,t} \geq c_2 \qquad \textit{and}$$

$$\forall k = 1,...,\mathcal{K}: \quad \sqrt{\frac{\left(E_k^{n,t}\right)^2 \log|E|_{n,t}}{|E|_{n,t}}} \leq c_3.$$

Combining the definitions and results of this section, we can prove the following lemma which provides an exponential inequality.

Lemma 3.29. *Let $(C_{n,i}, X_{n,i}, N_{n,i})_{i\in G_n}$ be a structured interaction network with filtration $(\mathcal{F}_t^n)_{t\in[0,T]}$ and distance d^n which is β-mixing at time t with respect to random* *variables* $(Z_{n,i})_{i\in G_n}$ *indexed by the edges in the sense of Definition 3.28. Then, for any* $x > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{|E|_{n,t}}\sum_{i\in G_n}(Z_{n,i}-\mathbb{E}(Z_{n,i}))\geq x\cdot\sqrt{\frac{\log|E|_{n,t}}{|E|_{n,t}}}\right)\\
\leq\;& \mathcal{K}|E|_{n,t}^{-\frac{c_2\cdot x^2}{2(\sigma^2+c_1c_3x)}}+\beta_t(\Delta_n)\cdot\mathcal{K}r_n.
\end{aligned}\tag{3.21}$$

Proof. The proof of (3.21) is an immediate consequence of Proposition 3.27 together with the assumptions:

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{|E|_{n,t}}\sum_{i\in G_n}(Z_{n,i}-\mathbb{E}(Z_{n,i}))\geq x\cdot\sqrt{\frac{\log|E|_{n,t}}{|E|_{n,t}}}\right)\\
\leq\;& \mathbb{P}\left(\sum_{i\in G_n}(Z_{n,i}-\mathbb{E}(Z_{n,i}))\geq x\cdot\sqrt{\log|E|_{n,t}\cdot|E|_{n,t}}\right)\\
\leq\;& \sum_{k=1}^{\mathcal{K}}\exp\left(-\frac{c_2x\log|E|_{n,t}\cdot|E|_{n,t}}{2\left(|E|_{n,t}\sigma^2+E_k^{n,t}c_1x^2\cdot\sqrt{\log|E|_{n,t}\cdot|E|_{n,t}}\right)}\right)+\beta_t(\Delta_n)\cdot\mathcal{K}r_n\\
\leq\;& \mathcal{K}\exp\left(-\frac{c_2x^2\log|E|_{n,t}}{2\left(\sigma^2+c_1c_3x\right)}\right)+\beta_t(\Delta_n)\cdot\mathcal{K}r_n.
\end{aligned}$$

□

4 Model Formulation and Theoretic Results

In this chapter we present which type of data we model, how we intend to model it and what we can prove about the model. The model we introduce is very closely related to stochastic actor models (cf. Snijders (2001); Butts (2008)) and has been studied, e.g., in Perry and Wolfe (2013). A similar model was applied in Butts (2008). At first, in Section 4.1, we will introduce the statistical model and its interpretation followed by the main results of this thesis in Sections 4.2 and 4.3. The assumptions which are necessary to prove the main results are presented and discussed in Section 4.4. The proofs of the results are deferred to Chapter 5. We will also illustrate how the model may be applied and how the results may be used for inference by investing a real world data set from bike sharing data in Washington D.C. in Section 4.5.

4.1 Modelling of Interaction Networks

We express the relation between the covariates $X_{n,i}$ and the interactions on active edges in an interaction network (cf. Definition 2.23) by assuming that the counting processes $N_{n,i}$ have intensity functions

$$C_{n,i}(t) \cdot \lambda(t, \theta_0(t), X_{n,i}(t))$$

with respect to $\mathcal{F}_t^n$ as defined in Section 2.2. They depend on the covariates and a parameter function $\theta_0 : [0,T] \to \Theta$, where Θ is some state space (later we will have $\Theta \subset \mathbb{R}^q$), through a known, deterministic link function $\lambda : [0,T] \times \Theta \times \mathbb{R}^q \to [0,\infty)$. For ease of notation we will sometimes omit the covariates in the notation and just write

$$\lambda_{n,i}(\theta_0(t), t) = \lambda(t, \theta_0(t), X_{n,i}(t)).$$

Note firstly that by definition $C_{n,i}$ and $X_{n,i}$ are both predictable with respect to $\mathcal{F}_t^n$ and thus $C_{n,i}\lambda_{n,i}$ is indeed a proper intensity function. The parameter function θ_0 describes the relation between the intensity and the covariates at different times. So θ_0 is the object we would like to estimate and perform inference about. We illustrate this in an example.

Example 4.1. *Assume a company's management is interested in how well the communication among their employees is working. The company is very international and has employees around the world so that their primary means of communication are phone calls. Several employees work together in a group on the same project. It is hereby possible that one employee works in several groups and that employees are assigned to new projects and leave old projects. Moreover, each project has group leaders and employees* *can have different roles in different groups. Although it happens that employees who work in different groups call each other, the management is primarily interested in the communication within a project. Their particular interest is finding out if the communication among group leaders and group members is one-sided or mutual and if this has changed since all employees have been equipped with new phones.*

This question fits in the above described framework: The actors are the employees and the connection between two actors is active as long as they work in the same group. The status of an edge changes then with time as the employees change their group assignments. The phone calls are considered as interactions (more precisely, the begin of a phone call). As covariates $X_{n,i}(t)$ *we put the group hierarchy, i.e., the covariate for* $i = (v_1, v_2)$ *is coding the four possibilities: both are group members, both are group leaders,* v_1 *is leader and* v_2 *is member or* v_2 *is leader and* v_1 *is member. Note that the covariates may change here also over time as the employees can change their roles. We consider now*

$$\lambda(t, \theta_0(t), X_{n,i}(t)) = \exp(\theta_0(t)^T X_{n,i}(t)).$$

Thus, $\theta_0(t)$ *quantifies the importance of the different covariates. It is hence of interest to compare the estimates of* $\theta_0(t)$ *for time points* t *before and after the introduction of new phones.*

It was already shortly mentioned in the example above, but we would like to emphasize again that the fact that an edge between two actors is inactive does not exclude the possibility of interactions between them. It rather means that interactions between them are not of interest for the particular application and hence they are not captured in the modelling.

An estimator of the parameter function θ_0 at a given point $t_0 \in [0, T]$ can be obtained by maximizing the following local log-likelihood function over $\mu \in \Theta$ which is obtained by localizing the log-likelihood for counting processes for a constant parameter (cf. Section 2.2 and line (2.3)) at time t_0 by means of a kernel K (i.e. K is non-negative and integrates to one)

$$\begin{aligned}\ell_n(\mu, t_0) := &\sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{t - t_0}{h}\right) \log \lambda(t, \mu, X_{n,i}(t)) dN_{n,i}(t) \\ &- \int_0^T \sum_{i \in G_n} \frac{1}{h} K\left(\frac{t - t_0}{h}\right) C_{n,i}(t) \lambda(t, \mu, X_{n,i}(t)) dt, \end{aligned} \tag{4.1}$$

where $h = h_n$ is the bandwidth. Note that in (4.1) interactions which happened close to time t_0 get a large weight while those which happened earlier or later get a low weight (hence the term local log-likelihood). We point out that $C_{n,i}(t) = 0 \implies \Delta N_{n,i}(t) = 0$ (there are no interactions if the intensity equals zero). So we do not have to worry about taking the logarithm of zero. By localizing, we hope that the maximizer of the local log-likelihood is a good estimator of $\theta_0(t_0)$ for any fixed $t_0 \in [0, T]$. Thus, we define the corresponding local maximum likelihood estimator as

$$\hat{\theta}_n(t_0) = \arg\max_{\theta \in \Theta} \ell_n(\theta, t_0).$$

In order to understand the form of the local log-likelihood in (4.1) better, we rewrite it by using the Stieltjes integral notation from Section 2.1 and the notion of the cumulated intensity $\Lambda_{n,i}(t) := \int_0^t \lambda_{n,i}(s)ds$ as well as the counting process martingale $M_{n,i}(t) := N_{n,i}(t) - \Lambda_{n,i}(t)$,

$$\begin{aligned}
&\ell_n(\mu, t_0) \\
&= \sum_{i \in G_n} \left(\int_0^T \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \log \lambda(t, \mu, X_{n,i}(t)) dN_{n,i}(t) \right. \\
&\qquad\qquad \left. - \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) C_{n,i}(t) \lambda(t, \mu, X_{n,i}(t)) dt \right) \\
&= \sum_{i \in G_n} \left(\int_0^T \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \log \lambda(t, \mu, X_{n,i}(t)) dM_{n,i}(t) \right. \\
&\quad + \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) C_{n,i}(t) \\
&\qquad \left. \times \left[\log \lambda(t, \mu, X_{n,i}(t)) \cdot \lambda(t, \theta_0(t), X_{n,i}(t)) - \lambda(t, \mu, X_{n,i}(t))\right] dt \right).
\end{aligned}$$

Now we see that the local log-likelihood splits in a martingale and in a non-martingale part (strictly speaking the first part is a local martingale but we ignore this difference for the following heuristic). So in this sense, we can understand the situation we are facing as a regular regression situation: In a standard kernel regression situation we observe signal plus Gaussian noise and build up the log-likelihood (which is then a sum), localize it around the time of interest t_0 and maximize the likelihood, which yields then the standard Nadaraya-Watson kernel estimator. The above expression corresponds exactly to this likelihood. In the counting process world we observe $N = \Lambda + M$, where Λ (the cumulated intensity function) corresponds to the signal, while the martingale M corresponds to the noise. As in the kernel regression setting, we hope that the noise is averaged out and $\ell_n(\mu, t_0)$ is basically only the second part in the equality above which in turn we hope to average to its expectation (assume that all $X_{n,i}$ are identically distributed, the expectations below exist and Fubini may be applied and recall that $r_n = |G_n|$)

$$\begin{aligned}
&\frac{1}{r_n} \mathbb{E}\left(\ell_n(\mu, t_0)\right) \\
=&\mathbb{E}\left(\int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) C_{n,1}(t) \right. \\
&\qquad \left. \times \left[\log \lambda(t, \mu, X_{n,1}(t)) \cdot \lambda(t, \theta_0(t), X_{n,1}(t)) - \lambda(t, \mu, X_{n,1}(t))\right] dt \right) \\
=& \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) p_n(t)
\end{aligned}$$

$$\times \mathbb{E}\left(\log \lambda(t,\mu,X_{n,1}(t)) \cdot \lambda(t,\theta_0(t),X_{n,1}(t)) - \lambda(t,\mu,X_{n,1}(t)) \middle| C_{n,1}(t)=1\right) dt$$
$$= \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) p_n(t) g(\mu,t) dt,$$

where $p_n(t) := \mathbb{P}(C_{n,1}(t)=1)$ and

$$g_n(\mu,t)$$
$$:= \mathbb{E}\left(\log \lambda(t,\mu,X_{n,1}(t)) \cdot \lambda(t,\theta_0(t),X_{n,1}(t)) - \lambda(t,\mu,X_{n,1}(t)) \middle| C_{n,1}(t)=1\right). \tag{4.2}$$

It is easy to see that for any fixed t the function $\mu \mapsto g_n(\mu,t)p_n(t)$ is maximized by the choice $\mu = \theta_0(t)$. So we hope that the estimator $\hat{\theta}_n(t_0)$ which maximizes the local log-likelihood is a good estimator of $\theta_0(t_0)$ for any fixed $t_0 \in [0,T]$. The first main result of this thesis gives an affirmative answer to this question.

4.2 Asymptotic Normality of local log-likelihood estimator

For the first main theorem, we become more specific in terms of the intensity function. We will assume $\Theta \subseteq \mathbb{R}^q$ (recall that q is the dimension of the covariate vector $X_{n,i}$) and $\lambda(t,\mu,X) := \exp(\mu^T X)$, i.e., we have for the intensities the following Cox-type form

$$C_{n,i}(t)\lambda_{n,i}(\theta_0(t),t) = C_{n,i}(t)\lambda(t,\theta_0(t),X_{n,i}(t)) = C_{n,i}(t)\exp\left(\theta_0(t)^T X_{n,i}(t)\right). \tag{4.3}$$

This form of the intensity function is particularly nice to interpret because in the exponential we find a weighted sum of the covariates and the parameters which we will estimate are exactly the weights. Hence, the parameters may be interpreted as the *impact* a covariate has on the overall intensity.

The model description so far was pretty general and it can be adapted to all types of networks. For simplicity of notation and exposition we restrict here to the case of undirected, complete networks, i.e., $r_n = \frac{n(n-1)}{2}$ where n is the number of actors. The methodology can be generalized in a straight forward way to directed, complete networks where $r_n = n(n-1)$. However, in this case the asymptotic uncorrelation type assumptions need to be reformulated. Also a generalisation to arbitrary networks with arbitrary r_n is possible but requires more careful reformulation of the asymptotic uncorrelation assumptions. So let G_n be a complete, undirected network from now on.

Theorem 4.2. *Suppose that Assumptions (A1)-(A7) from Section 4.4 hold for a point $t_0 \in (0,T)$. Then, with probability tending to one, the derivative of the local log-likelihood function $\theta \mapsto \ell_n(\theta,t_0)$ has a root $\hat{\theta}_n(t_0)$, satisfying*

$$\sqrt{l_n h}\left(\hat{\theta}_n(t_0) - \theta_0(t_0) + h^2\left[\frac{1}{2}\Sigma^{-1}v - B_n\right]\right) \to N\left(0, \int_{-1}^1 K(u)^2 \mathrm{d}u\ \Sigma^{-1}\right)$$

for $n \to \infty$ with

$$l_n := r_n \mathbb{P}(C_{n,1}(t_0)=1), \text{ where } r_n = |G_n|$$

$$
\begin{aligned}
v &:= \int_{-1}^{1} K(u)u^2 \mathrm{d}u \cdot \partial_\theta \partial_{t^2} g(\theta_0(t_0), t_0), \\
\Sigma &:= -\partial_{\theta^2} g(\theta_0(t_0), t_0), \\
\gamma_{n,i}(s) &:= (1 - C_{n,i}(t_0))C_{n,i}(s), \\
\tau_{n,i}(\theta, s) &:= X_{n,i}(s)X_{n,i}(s)^T \exp(\theta^T X_{n,i}(s)), \\
B_n &:= \frac{1}{l_n} \sum_{i,j=1}^{n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \frac{\gamma_{n,i}(s)}{h} \tau_{n,i}(\theta_0(s), s)\theta_0'(t_0) \frac{t_0 - s}{h} \mathrm{d}s
\end{aligned}
$$

and g_n was defined in (4.2). If, in addition, $\frac{|E_{t_0}|}{l_n} \xrightarrow{\mathbb{P}} 1$, then l_n can be replaced by $|E_{t_0}|$, where $|E_{t_0}|$ denotes the size of the set E_{t_0}.

The proof of this result is presented in Section 5.1. As motivated before the statement of the theorem, we are in a very similar situation to regular kernel regression. This is reflected by the fact that, up to the bias term, Theorem 4.2 looks very similar to a standard asymptotic normality result in kernel regression. The rate of the convergence is $l_n \cdot h$ where h is the bandwidth and $l_n = r_n \mathbb{P}(C_{n,1}(t_0) = 1)$ is the expected number of observations: Recall that n is the number of actors and r_n is the number of possible pairs which contribute to the observations with probability $\mathbb{P}(C_{n,12}(t_0) = 1)$. The variance of the normal distribution has the standard form of a kernel integral times the inverse of the Fisher information matrix. The bias is (as usual) of order h^2 but comprises two terms: The first one has again a standard form involving the second derivative of the objective function. The second part B_n, however, is a result of the dynamic network setting. We discuss B_n in the next paragraph.

Note firstly that the factor $\gamma_{n,i}(s)$ is equal to zero if the connection i does not change its status $C_{n,i}$ (present or not-present) in the time interval $[s, t_0]$. In particular in a network which is not changing over time, $\gamma_{n,i}(s) = 0$ for all (i, j) and all $s \in [0, T]$ and hence $B_n = 0$. So the bias is induced by the dynamics of the network. B_n is also vanishing if $\theta_0'(t_0) = 0$, i.e., if we estimate a local extremal point of θ this part of the bias vanishes (but the first part of the bias remains). Let us show that $B_n = O_P(1)$ such that the bias as a whole is indeed of order h^2. It holds that

$$
\begin{aligned}
&\mathbb{E}(|B_n|) \\
\leq\ & \frac{1}{l_n} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \mathbb{E}\left(\frac{\gamma_{n,i}(s)}{h} \|\tau_{n,i}(\theta_0(s), s)\|\right) \|\theta_0'(t_0)\| \frac{|t_0 - s|}{h} ds \\
=\ & \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \frac{\mathbb{P}(C_{n,1}(t_0) = 0, C_{n,1}(s) = 1)}{h\mathbb{P}(C_{n,1}(t_0) = 1)} \frac{|t_0 - s|}{h} ds \cdot \|\theta_0'(t_0)\| \\
& \times \sup_{s \in U_h} \mathbb{E}\left[\|\tau_{n,1}(\theta_0(s), s)\| \Big| C_{n,1}(s) = 1, C_{n,1}(t_0) = 1\right].
\end{aligned}
$$

This is $O(1)$ by Assumptions (A6) (boundedness of the fraction of probabilities), (A4) (boundedness of the expectation) and (A2) (kernel is supported on $[-1, 1]$). Hence, we get that $B_n = O_P(1)$. In Section 4.4 we discuss the assumptions with examples. In order to see that B_n is asymptotically not vanishing, note that the only quantity in the expression above which can converge to zero is the fraction of probabilities. We discuss this in more detail after Assumption (A6) so we give here just a heuristic: Let $s \in [t_0 - h, t_0]$ be given, then

$$\mathbb{P}(C_{n,1}(t_0) = 1,\, C_{n,1}(s) = 1) = \mathbb{P}(C_{n,1}(t_0) = 0 |\, C_{n,1}(s) = 1) \cdot \mathbb{P}(C_{n,1}(s) = 1).$$

If the conditional distribution of the time at which an edge changes its status given that it is active at time s has a density, then, it is clear that $\mathbb{P}(C_{n,1}(t_0) = 0,\, C_{n,1}(s) = 1)$ behaves like $h \cdot \mathbb{P}(C_{n,1}(s) = 1)$ because $t_0 - s \leq h$. As $\frac{\mathbb{P}(C_{n,1}(s)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)} \approx 1$ we can only have convergence to zero if $h^{-1}\mathbb{P}(C_{n,1}(t_0) = 0,\, C_{n,1}(s) = 1) \to 0$, i.e., if the density of the random time point of status change is 0 at s. This is not possible if we assume a decreasing density, i.e., later status changes are less likely (which would e.g. be the case for an exponential distribution).

4.3 Asymptotics of Test for Constant Parameter

Now we are interested in testing whether the parameter function θ_0 is indeed time varying, i.e., we want to study the test problem

$$\mathrm{H}_0 : \theta_0 \equiv \text{ const.} \quad \text{vs.} \quad \mathrm{H}_1 : \theta_0 \text{ is time varying.}$$

We suggest therefore the following test statistic similar to Härdle and Mammen (1993)

$$T_n := \int_0^T \|\hat{\theta}_n(t_0) - \bar{\theta}_n\|^2 \bar{p}_n(t_0) w(t_0) dt_0,$$

where w is a non-negative weight function with $\operatorname{supp} w \subseteq [\delta, T - \delta]$ for $\delta > 0$ and $\bar{\theta}_n$ is a suitable estimator in a model that assumes a constant parameter function (e.g. the maximum likelihood estimator). Furthermore, $p_n(t_0) := \mathbb{P}(C_{n,i}(t_0) = 1)$ is just an abbreviation and $\bar{p}_n(t_0) := \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) p_n(s) ds$ is the smoothed version of p_n (we will impose assumptions later which imply that i in these definitions is arbitrary, cf. (B1)). On H_0, the parameter function θ_0 is constant and we denote its value by θ_0 as well. Before continuing with the main result about T_n, we note that in contrast to Härdle and Mammen (1993), we know in advance that we estimate a constant function (because we operate on H_0). Therefore we can directly compare with the estimate and we do not require additional smoothing in the test statistic. For the statement of the following theorem we define

$$\Sigma(t, \theta) := \mathbb{E}\left(\partial_\theta^2 \log \lambda_{n,1}(t, \theta) \cdot \lambda_{n,1}(\theta_0, t) - \partial_\theta^2 \lambda_{n,1}(\theta, t) | C_{n,1}(t_0) = 1\right)$$

and set on H_0 the abbreviation $\Sigma(t) := \Sigma(t, \theta_0)$. We can formulate the following general result about the asymptotics of the test statistic. Note that for the main body of the proof of the following theorem, the Cox-type model (4.3) is not necessary. Thus, we formulate the result more general than Theorem 4.2. However, when we want to prove the detailed assumptions of this main body (cf. Section 4.4), we will have to assume the Cox-Model.

Theorem 4.3. *Under the Assumptions (B1)-(B7), stated in Section 4.4, on H_0*

$$r_n h^{\frac{1}{2}} T_n - h^{-\frac{1}{2}} A_n \xrightarrow{d} N(0, B),\ n \to \infty,$$

where in the general case A_n is defined in Statement 4.5 and B is defined in Statement 4.6. In the case of the Cox-type model (cf. (4.3)), we have

$$A_n := \frac{1}{r_n} \sum_{i \in G_n} \int_0^T X_{n,i}(s)^T \int_\delta^{T-\delta} \frac{1}{h} K\left(\frac{s-t_0}{h}\right)^2 \Sigma^{-T}(t_0)\Sigma^{-1}(t_0) \frac{w(t_0)}{\bar{p}_n(t_0)} dt_0 X_{n,i}(s) dN_{n,i}(s)$$

$$B := 4K^{(4)} \int_0^T \mathit{trace}\left(\left(P(t)\Sigma^{-T}(t)\Sigma^{-1}(t)\right)^2\right) w^2(t) dt$$

and

$$K^{(4)} := \int_0^2 \left(\int_{-1}^1 K(v)K(u+v)dv\right)^2 du,$$
$$P^{r_1,r_2}(t) := \mathbb{E}\left(X_{n,j}^{r_1}(t) X_{n,j}^{r_2}(t) \lambda_{n,j}(t) \middle| C_{n,j}(t) = 1\right).$$

Note that A_n can be approximated by using a plug in estimator for Σ and B can be approximated by Statement 4.6.

The proof of this theorem is presented in Section 5.2.

Remark 4.4. *Analogously as in Theorem 4.2 the rate of convergence should intuitively be influenced by the probability of observing an edge $p_n(t_0)$. However, on first sight it is not appearing the the result. In contrast to Theorem 4.2, in Theorem 4.3 we make a statement about all estimators $\hat{\theta}_n(t_0)$ for all time points $t_0 \in \operatorname{supp} w$. So the probabilities $p_n(t_0)$ appear in the test statistic in the integral, thus each time point receives a different weight. This is why the $p_n(t_0)$ is not appearing directly in the rate of convergence of T_n.*

From an applied point of view, we can use Theorem 4.2 in order interpret results from the estimation procedure. Theorem 4.3 can be used in order to show that a time dependent parameter function $\theta_0(t_0)$ is important in order to represent the data.

4.4 Assumptions

In this section we present and discuss assumptions sufficient for proving Theorems 4.2 and 4.3. The Assumptions (A1)-(A7) will be used for Theorem 4.2 and the (B) and (C) assumptions for Theorem 4.3. We emphasize that Theorem 4.3 requires on the one hand stronger assumptions because we have to handle the estimator at all time points simultaneously, on the other hand it is easier because we assume a constant parameter function. Therefore none of the sets of assumptions is contained in the other. We present the assumptions therefore separately and present connections between the two sets when talking about the (B) assumptions.

Our assumptions do not specify the dynamics of the covariates $X_{n,i}(t)$ and of the censoring variable $C_{n,i}(t)$. Instead of this, we assume that the stochastic behaviour of these variables stabilizes for $n \to \infty$. Assumption (A1) is specific to our setting and it states our general understanding of the dynamics, while assumptions (A2), (A3) and (A5) are standard. Assumption (A4) guarantees that the covariates are well behaved and can be found similarly in Perry and Wolfe (2013). Finally, (A6) and (A7) specifically describe the dependence situation in our context. They quantify how we make the idea mathematically precise that while the network grows the actors get further and further apart and hence influence each other less and less. In the following we firstly state an assumption and then discuss its meaning and the intuition behind it.

(A1) Exchangeability
For every n and for any $t \in [t_0 - h, t_0 + h]$, the joint distribution of $(C_{n,i}(t), X_{n,i}(t))$ is identical for all edges $i \in G_n$. Furthermore, for any $s, t \in [t_0 - h, t_0 + h]$, the conditional distribution of the covariate $X_{n,i}(t)$ given that $C_{n,i}(s) = 1$, has a density $f_{s,t}(y)$ with respect to a measure μ on $\mathbb{R}^q$, and this conditional distribution does not depend on i and n. We use the shorthand notation f_s for $f_{s,s}$. Finally, it holds that: $n \to \infty$, $h \to 0, l_n h \to \infty$, and $l_n h^5 = O(1)$.

Recall that, l_n is the effective sample size at time t_0 (see discussion after Theorem 4.2). With this in mind, the assumptions on the bandwidth are standard. The most restrictive assumption in (A1) is that the conditional distribution of $X_{n,i}(t)$, given $C_{n,i}(s) = 1$, does not depend on $i \in G_n$. Observe that this holds if the array of $(C_{n,i}, X_{n,i})_i$ is jointly exchangeable in $i = (v_1, v_2)$ for any fixed n, i.e., if relabelling the vertices does not change the joint distribution of all $X_{n,i}$. The additional assumption that the conditional distribution of $X_{n,i}(t)$, given $C_{n,i}(s)$, does not change with n is not very restrictive, because it is natural to assume that the distribution depends only on the local structure of the network. For instance, if we assume that a fixed vertex v has only a bounded number of close interaction partners v' while the network grows, then it is natural to assume that the local structure given by the interacting partners does not undergo major changes for $n \to \infty$ (we discuss after assumption (A6) in more detail how this scenario gets incorporated in the $C_{n,i}$). We make this additional assumption mainly to avoid stating lengthy technical assumptions allowing to interchange the order of differentiation and integration at several places in the proof.

We add some standard assumptions on the kernel.

(A2) Kernel Order
The kernel K is positive and supported on $[-1, 1]$, and it satisfies $\int_{-1}^{1} K(u)\mathrm{d}u = 1$, $\int_{-1}^{1} K(u)u\mathrm{d}u = 0$ and $\max_{-1 \le u \le 1} K(u) < \infty$.

The next condition makes smoothness assumptions for the parameter curve θ_0 and the density $f_s(y)$.

(A3) Smoothness of Parameter

The parameter space Θ is compact and convex. Let $\tau := \sup_{\theta\in\Theta} \|\theta\| < \infty$. The parameter function $\theta_0(t)$ takes values in Θ, and, in a neighbourhood of t_0, it is twice continuously differentiable. The value $\theta_0(t_0)$ lies in the interior of Θ.

We continue with some tail conditions on $f_s(y)$ and its derivatives. They are fulfilled if, e.g., the covariates are bounded.

(A4) Moment Conditions

For μ-almost all y (see (A1) for a definition of μ), the density $f_s(y)$ is twice continuously differentiable in s. For an open neighbourhood U of t_0 and $U_h := [t_0 - h, t_0 + h]$ it holds for all edges $i, j \in G_n$

$$\int \sup_{s\in U} \Big\{ \big(1 + \|y\| + \|y\|^2 + \|y\|^3\big) |f_s(y)| + \big(1 + \|y\| + \|y\|^2\big) \left|\frac{\mathrm{d}}{\mathrm{d}s} f_s(y)\right| + (1 + \|y\|) \left|\frac{\mathrm{d}^2}{\mathrm{d}s^2} f_s(y)\right| + \|y\|^2 \cdot f_{s,t_0}(y) \Big\} \cdot \exp(\tau \cdot \|y\|)\mathrm{d}\mu(y) < \infty, \tag{4.4}$$

$$\sup_{s,t\in U_h} \mathbb{E}\Big(\|X_{n,i}(s)\|^2 \cdot \|X_{n,j}(t)\|^2 \cdot e^{\tau(\|X_{n,i}(s)\|+\|X_{n,j}(t)\|)} \Big| C_{n,i}(t_0) = 1,\, C_{n,j}(t_0) = 1\Big) = O(1), \tag{4.5}$$

For $k \in \{2, 3\}$:

$$\sup_{s\in U_h} \mathbb{E}\left(\|X_{n,1}(s)\|^k e^{\tau\|X_{n,1}(s)\|} \Big| C_{n,1}(s) = 1, C_{n,1}(t_0) = 0\right) = O(1), \tag{4.6}$$

$$\mathbb{E}\Big(\sup_{s\in U_h} \big[\|X_{n,1}(s)\| + \|X_{n,1}(s)\|^2 + \|X_{n,1}(s)\|^3 + \|X_{n,1}(s)\|^4\big] \cdot e^{\tau\|X_{n,1}(s)\|} \Big| C_{n,1}(s) = 1\Big) < +\infty. \tag{4.7}$$

The next assumption guarantees identifiability.

(A5) Identifiability

For any $n \in \mathbb{N}$, $\theta^T X_{n,1}(t_0) = 0$ a.s. (w.r.t. f_{t_0}) implies that $\theta = 0$.

The following assumption addresses the asymptotic behaviour of the distributions of the processes $C_{n,i}(t)$. In particular, for t in a neighbourhood of t_0, the assumptions address asymptotic stability of the marginal distributions of these processes, and also a certain kind of asymptotic independence of $C_{n,i}$ and $C_{n,j}$ for $|e_i \cap e_j| = 0$ (recall that for an edge i, we denote by e_i the set of the two incident vertices).

(A6) Asymptotic Uncorrelation I

For $w(u) = K(u)$ and $w(u) = K^2(u)/\int K^2(v)\mathrm{d}v$ it holds that

$$\int_{-1}^{1} w(u) \frac{\mathbb{P}(C_{n,1}(t_0 + uh) = 1)}{\mathbb{P}(C_{n,1}(t_0) = 1)} \mathrm{d}u \to 1 \tag{4.8}$$

for $n \to \infty$. For

$$A_{n,i,j} := \int_{-1}^{1}\int_{-1}^{1} w(u)w(v)\frac{\mathbb{P}(C_{n,i}(t_0+uh)=1, C_{n,j}(t_0+vh)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)^2}\mathrm{d}u\mathrm{d}v,$$

we assume that

$$A_{n,i,j} = \begin{cases} o(n^2) & \text{for} \quad |e_i \cap e_j| = 2, \\ o(n) & \text{for} \quad |e_i \cap e_j| = 1, \\ 1+o(1) & \text{for} \quad |e_i \cap e_j| = 0. \end{cases} \tag{4.9}$$

Furthermore, it holds that

$$\int_0^T \frac{1}{h}K\left(\frac{s-t_0}{h}\right)\frac{\mathbb{P}(C_{n,1}(t_0)=0,\, C_{n,1}(s)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)}ds = O(h), \tag{4.10}$$

and for edges with $|e_i \cap e_j| \le 1$

$$\frac{\mathbb{P}(C_{n,i}(t_0)=1, C_{n,j}(t_0)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)^2} = O(1). \tag{4.11}$$

Note firstly that, due to the localization of our likelihood function, all time dependence happens only locally around the target time t_0. Condition (4.8) appears reasonable in the regime of asymptotics in the size of the network: Consider, for instance a dynamic social media network, and assume, for example, that we consider data from a certain geographic region. One might assume that during the night the number of active pairs, i.e. the pairs with $C_{n,i} = 1$, is lower than during the day, and we expect that there will be a gradual decrease between, e.g., 8pm and 11pm. This time window does not get narrower when n increases and hence a slow change of the distribution seems to be a reasonable assumption. Assumption (4.10) holds for example in the following model: Assume that in the previous example communications between pairs are ended at $\delta_0 := 8pm$ plus a certain random time $\delta_{n,i}$, i.e., $C_{n,i}(t) = \mathbb{1}(t \le \delta_0 + \delta_{n,i})$. In this case, the ratio of probabilities in (4.10) becomes

$$\frac{\mathbb{P}(C_{n,i}(t_0)=0, C_{n,j}(s)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)} = \frac{\mathbb{P}(\delta_{n,i} \in [s-\delta_0, t_0-\delta_0))}{\mathbb{P}(\delta_{n,i} \ge t_0-\delta_0)}.$$

Since we are using a localizing kernel, the length of the interval $[s-\delta_0, t_0-\delta_0)$ is of the order h, and if $\delta_{n,i}$ has a density, then (4.10) holds.

If we assume that relabelling the vertices does not change the joint distribution of the whole process (i.e., if we assume exchangeability). Then, the joint distribution of two edges i and j depends only on $|e_i \cap e_j|$. Hence, it is very natural to distinguish the three regimes $|e_i \cap e_j| \in \{0, 1, 2\}$. This pattern will appear again in the next Assumption (A7). Let us for the moment consider $C_{n,i}$ that are constant over time. Then, in (4.9), in the case $|e_i \cap e_j| = 2$, the assumption is satisfied because $\frac{\mathbb{P}(C_{n,i}=1, C_{n,j}=1)}{\mathbb{P}(C_{n,i}=1)^2} = \mathbb{P}(C_{n,1} = 1)^{-1} = o(n^2)$ according to Assumption (A1).

We discuss the remaining cases for the uniform configuration model. In the uniform configuration model all vertices have (approximately) the same pre-defined degree κ, and we assume that the $C_{n,i}$ are created as follows: Equip each vertex $v = 1, ..., n$ with κ edge stubs, and create edges by randomly pairing the stubs. After that, discard redundant edges and self-loops. If two vertices v and v' are connected after this process, set $C_{n,i} = 1$ for $i = (v, v')$. We use the same heuristics as e.g. in Newman (2010), Chapter 13.2, to compute the probability of edges. Fix v and v', then for any fixed edge stub of v there are $\kappa n - 1$ stubs left to pair with, κ of which belonging to vertex v'. Hence, the probability of connecting to v' is given by $\frac{\kappa^2}{\kappa n - 1}$ as there are κ edge stubs from v as well. Approximating $\kappa n - 1$ by κn, as n gets large, we obtain the following probabilities: Let $i_1, j_1, j_2 \in G_n$ be edges with $|e_{i_1} \cap e_{j_1}| = 1$ and $|e_{i_1} \cap e_{j_2}| = 0$, then

$$\begin{aligned} \mathbb{P}(C_{n,i_1} = 1) &\approx \frac{\kappa}{n} \\ \mathbb{P}(C_{n,i_1} = 1, C_{n,j_1} = 1) = \mathbb{P}(C_{n,i_1} = 1 | C_{n,j_1} = 1) \cdot \mathbb{P}(C_{n,j_1} = 1) &\approx \frac{\kappa(\kappa - 1)}{n^2} \\ \mathbb{P}(C_{n,i_1} = 1, C_{n,j_2} = 1) = \mathbb{P}(C_{n,i_1} = 1 | C_{n,j_2} = 1) \cdot \mathbb{P}(C_{n,j_2} = 1) &\approx \frac{\kappa^2}{n^2}. \end{aligned}$$

We see now that also in the cases $|e_i \cap e_j| \leq 1$, the assumptions (4.9) and (4.11) hold.

The next assumption involves $\theta_{0,n}$, defined as the maximizer of

$$\theta \mapsto \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h}\right) g(\theta, s) ds, \tag{4.12}$$

where g is defined in (4.15) (note that this is the same as g_n in (4.2) but where we used the specific form of the intensity in (4.3) as well as assumption (A1) which says that the conditional distribution is independent of n and hence we write g instead of g_n). We show later that $\theta_{0,n}$ is uniquely defined, and that $\theta_{0,n}$ is close to $\theta_0(t_0)$ (see Lemma 5.3 and Proposition 5.5, respectively). Define furthermore

$$\begin{aligned} \tau_{n,i}(\theta, s) &:= X_{n,i}(s) X_{n,i}(s)^T \exp(\theta^T X_{n,i}(s)), && (4.13) \\ g(\theta, t) &:= \mathbb{E}\left[\theta^T X_{n,i}(t) \exp(\theta_0(t)^T X_{n,i}(t)) \right. && (4.14) \\ &\qquad \left. - \exp(\theta^T X_{n,i}(t)) | C_{n,i}(t) = 1\right] \\ &= \int_{\mathbb{R}^q} \left(\theta^T y e^{\theta_0(t)^T y} - e^{\theta^T y}\right) f_t(y) \mathrm{d}\mu(y), && (4.15) \\ f_{n,1}(\theta, s, t | i, j) &:= \mathbb{E}(\tau_{n,i}(\theta, s) \tau_{n,j}(\theta, t) | C_{n,i}(s) = 1, C_{n,j}(t) = 1), \\ f_2(\theta, t) &:= \mathbb{E}(\tau_{n,i}(\theta, t) | C_{n,i}(t) = 1) = -\partial_{\theta^2} g(\theta, t), \\ r_{n,i}(s) &:= C_{n,i}(s) X_{n,i}(s) \left(e^{\theta_0(s)^T X_{n,i}(s)} - e^{\theta_{0,n}^T X_{n,i}(s)}\right) \\ &\qquad - \partial_\theta g(\theta_{0,n}, s). \end{aligned}$$

Note that, by Assumption (A1), f_2 and g do not depend on i and n.

(A7) Asymptotic Uncorrelation II

We assume that $f_{n,1}$ depends on i and j only through $|e_i \cap e_j|$. Moreover, we assume that, for all sequences $\theta_n \to \theta_0(t_0)$ and $u, v \in [-1, 1]$, it holds that $f_{n,1}(\theta_n, t_0 + uh, t_0 + vh, i, j)$ converges to a value that depends only on $|e_i \cap e_j|$. We denote this limit by $f_1(\theta_0(t_0), |e_i \cap e_j|)$, and assume that

$$f_1(\theta_0(t_0), 0) = f_2(\theta_0(t_0), t_0)^2. \tag{4.16}$$

For $r_{n,i}(s)$, we assume that, with $\rho_{n,ij}(u, v) := r_{n,i}(t_0+uh)r_{n,j}(t_0+vh)$ and for $|e_i \cap e_j| = 0$,

$$\iint\limits_{[-1,1]^2} K(u)K(v)\mathbb{E}(\rho_{n,ij}(u, v)|C_{n,i}(t_0) = 1, C_{n,j}(t_0) = 1)\mathrm{d}u\mathrm{d}v = o\big(\big(l_n h\big)^{-1}\big). \tag{4.17}$$

Assumption (A7) specifies in which sense the covariates are asymptotically uncorrelated. For motivating (A7) build a graph $\mathcal{G}$ with vertices $1, ..., n$ and for $v_1, v_2 \in \{1, ..., n\}$, $i = (v_1, v_2)$ is an edge if $C_{n,i}(t_0) = 1$. Denote by $d_\mathcal{G}$ the distance function between edges on $\mathcal{G}$ (adjacent edges have distance 0). In the same heuristic as explained after Assumption (A6), this graph is very large (asymptotics over the number of vertices) and sparse (n vertices and of order n edges, regardless of possible directedness), because every vertex is incident to at most κ edges. In this scenario, the number of pairs of edges i and j for which $d_\mathcal{G}(i, j) = d$ is of order $(\kappa - 1)^d \cdot n$, and there are of order n^2 many pairs of edges in total. Let now A_i be arbitrary, centred random variables indexed by the edges of $\mathcal{G}$. We make the assumption that A_i is influenced equally by all A_j with j being adjacent to i. In mathematical terms, we formulate this assumption as $\mathbb{E}(A_iA_j|d_\mathcal{G}(i, j) = d) \approx C \cdot \kappa^{-d}$. Then, we obtain for non-adjacent edges i and j

$$\begin{aligned}
\mathbb{E}(A_iA_j) &= \sum_{d=1}^{\infty} \mathbb{P}(d_\mathcal{G}(i, j) = d) \cdot \mathbb{E}(A_iA_j|d_\mathcal{G}(i, j) = d) \\
&\approx \sum_{d=1}^{\infty} \frac{n(\kappa - 1)^d}{n^2} C \cdot \kappa^{-d} \\
&= \frac{C}{n}(\kappa - 1),
\end{aligned}$$

which converges to zero after being multiplied with $l_n h \approx nh$ (in this case because $\mathbb{P}(C_{n,1}(t) = 1) \approx \frac{1}{n}$). Because, in (4.16) and (4.17), we consider only expectations conditional on $C_{n,i}(t) = 1$, we can think of $A_{n,i}$ being the random variables $\tau_{n,i}$ (a centred version of it) or $r_{n,i}$ and the expectations in the above heuristic are conditional expectations conditionally the respective conditions in (4.16) and (4.17). This serves as motivation for these two assumptions. Moreover, unconditionally, $\tau_{n,i}$ and $\tau_{n,j}$ (and $r_{n,i}$ and $r_{n,j}$) do not need to be uncorrelated.

We present now the assumptions for Theorem 4.3. In order to keep an independent treatment of Theorems 4.2 and Theorem 4.3, some of the following assumptions are similar to the Assumptions (A1)-(A7).

(B1) Modelling Assumptions

1. *The conditional distribution of $(X_{n,i}(s), N_{n,i}(s))$ given $C_{n,i}(s) = 1$ is independent of n and i. Also $(C_{n,i}, X_{n,i}, N_{n,i})$ are identically distributed. In particular, $p_n(t_0) := \mathbb{P}(C_{n,i}(t_0) = 1)$ is well defined.*

2. *$\lambda_{n,i}(\theta, s)$ have either the Cox-form as in (4.3) or if not they are almost surely twice continuously differentiable in θ for any time s.*

3. *Let $\bar{p}_n := \int_0^T \bar{p}_n(s)ds$. Then the estimator $\bar{\theta}_n$ fulfils*
$$\|\bar{\theta}_n - \theta_0\| = O_P\left(\frac{1}{\sqrt{r_n \bar{p}_n}}\right).$$

4. *The function*
$$t \mapsto \mathbb{E}\left(X_{n,i}(t)^2 \lambda_{n,i}(\theta_0, t) \middle| C_{n,i}(t) = 1\right)$$
is uniformly bounded.

Part 1 of this assumption is very similar to (A1). For part 2 we remark that we do not assume here the specific Cox form of the intensity as in (4.3). But we require that $\lambda_{n,i}(\theta, s)$ depends differentiably on the parameter θ (for the Cox Model this holds). Part 3 holds for example for the maximum likelihood estimator as introduced in Chapter VI.1.2. in Andersen et al. (1993). Finally, part 4 holds for example if the covariates are bounded. We will argue after Assumption (C1), 2 that this is a reasonable assumption.

(B2) Weak-Dependence

Statements 4.1 to 4.12 hold.

This assumption is discussed in more detail later when we impose dependence assumptions on the network.

(B3) Boundary Cut-Off

The weight function w is bounded and supported on $\mathbb{T} \subseteq [\delta, T - \delta]$ for some $\delta > 0$.

The main role of the weight function is to cut-off boundary problems with the kernel. Therefore it is reasonable to bound its support and to assume that it is bounded.

(B4) Kernel and Bandwidth

1. *The bandwidth h relates to the effective number of observations $r_n p_n$, where $p_n := \inf_{t_0 \in \mathbb{T}} p_n(t)$ as follows*
$$\frac{\sqrt{r_n p_n} \cdot h}{(\log r_n)^{\frac{3}{2}}} \to \infty \text{ and } h(\log r_n)^2 \to 0.$$

2. *The kernel $K : [-1,1] \to \mathbb{R}$ is supported on $[-1,1]$ and Hoelder continuous with exponent α_K and constant H_K, i.e., $|K(x) - K(y)| \leq H_K \cdot |x-y|^{\alpha_K}$. As a consequence it is bounded by a constant which we also denote by K.*

The conditions on the bandwidth in the first part of (B4) are for example true when $h \approx (p_n r_n)^{-\frac{1}{5}}$ is the asymptotically optimal bandwidth choice in most one-dimensional regression contexts (cf. Tsybakov (2009)), so they are standard for this type of problem. The Hoelder continuity of the kernel is a mild assumption which avoids technical problems later. For most simple kernels like Epanechnikov or a triangular kernel it is true.

(B5) Exhaustiveness of Θ

1. *The true parameter lies in a bounded open set Θ. We denote the bound by τ.*
2. *With probability tending to one, we have for all $t_0 \in \mathbb{T}$ that $\hat{\theta}(t_0) \in \Theta$.*

The first part of (B5) is not restrictive because we study the behaviour on the null hypothesis which is $\theta_0 \equiv$ constant. Hence, the assumption holds when Θ is large enough. For the second part, note that the maximum likelihood-estimator $\hat{\theta}_n(t_0)$ is constructed by applying the Newton-Kantorovich Theorem (Theorem 5.9) which guarantees that the estimator lies in the interior of Θ (under the assumptions (A1)-(A7)).

(B6) Smoothness of $\bar{p}_n(t_0)$

The function $t_0 \mapsto \frac{1}{\bar{p}_n(t_0)}$ is Hoelder continuous with exponent α_p and on n depending constant $H_{n,p} \geq 1$, i.e., $\left|\frac{1}{\bar{p}_n(t_0)} - \frac{1}{\bar{p}_n(t_0')}\right| \leq H_{n,p} \cdot |t_0 - t_0'|^{\alpha_p}$ for all $t_0, t_0' \in \mathbb{T}$ and all $n \in \mathbb{N}$.

In the assumption (B6) we make a statement about the speed of the smoothed change in the sparsity of the network. Note that $\bar{p}_n(t) \geq p_n := \inf_{s\in\mathbb{T}} p_n(s)$ and thus we may directly conclude from Assumption (B4), 2 that

$$\left|\frac{1}{\bar{p}_n(t_0)} - \frac{1}{\bar{p}_n(t_1)}\right| \leq \frac{1}{p_n^2} |\bar{p}_n(t_1) - \bar{p}_n(t_0)| \leq \frac{H_K}{p_n^2 h^{1+\alpha_K}} |t_1 - t_0|^{\alpha_K}.$$

Note that we allow that the Hoelder constant depends on n, only the exponent needs to be fixed. As an alternative we may assume smoothness of $p_n(t)$ directly as we illustrate in the following by using the same example as explained after assumption (A6). We assume that during the day n people are each interacting with a fixed number k of people, i.e., $p_n(t) \approx \frac{k}{n}$ for $t \in [8\text{am}, 8\text{pm}]$. Furthermore, we imagine that during the night, the activity decreases and $p_n(t) \approx \frac{1}{n^{\frac{3}{2}}}$ for $t \in [10\text{pm}, 6\text{am}]$. So there is a two hour transit between the two levels of sparsity $\frac{1}{n}$ and $\frac{1}{n^{\frac{3}{2}}}$. Outside of the transit area we assume that the sparsity level is constant. Hence, in terms of continuity, the only critical region is the transit area. We assume that the transit is smooth in the sense that it is formed by a third order polynomial. In order to get easier numbers, we simplify the situation further and assume that f is a function with $f(t) = \frac{1}{n}$ for $t \leq 0$, $f(t) = \frac{1}{n^{\frac{3}{2}}}$ for $t \geq 1$ and

$$f(t) = 2\left(\frac{1}{n} - \frac{1}{n^{\frac{3}{2}}}\right)t^3 - 3\left(\frac{1}{n} - \frac{1}{n^{\frac{3}{2}}}\right)t^2 + \frac{1}{n}$$

for $t \in (0,1)$. Such an f has the same transit (up to constants) as a $p_n(t)$ as described before. We can check that f is differentiable with bounded derivative, hence it is Lipschitz continuous and we may compute the Lipschitz constant Lip_n (the bound of the derivative) to be of order $\frac{1}{n}$, i.e., $|f(s)-f(t)| \leq \mathrm{Lip}_n|s-t|$. From this we may finally conclude

$$\left|\frac{1}{f(s)} - \frac{1}{f(t)}\right| \leq \frac{|f(s)-f(t)|}{f(s)f(t)} \leq n^3\mathrm{Lip}_n \approx n^2|s-t|.$$

Hence, here we have $H_{n,p} \approx n^2$. In Statement 4.10 we make statements about the size of $H_{n,p}$. But generally it will not be a problem if $H_{n,p} \approx n^\alpha$ for some $\alpha \in \mathbb{R}$. This, we have just motivated.

Recall for the next assumption that

$$\Sigma(t,\theta) := \mathbb{E}\left(\partial_\theta^2 \log \lambda_{n,1}(t,\theta) \cdot \lambda_{n,1}(\theta_0,t) - \partial_\theta^2 \lambda_{n,1}(\theta,t) | C_{n,1}(t) = 1\right).$$

(B7) Invertibility of Fisher-Information

It holds that, $\Sigma(t_0,\theta_0)$ is invertible and there are $M,\rho > 0$ such that for every $t_0 \in \mathbb{T}$, every matrix X with $\|X - \Sigma(t_0,\theta_0)\| < \rho$ is invertible and $\|X^{-1}\| \leq M$.

Moreover, $\Sigma(\theta_0,t)$ is continuously differentiable in t with bounded derivative, i.e., there is a constant $D > 0$ such that

$$\sup_{t_0\in\mathbb{T}} \|\partial_t \Sigma(t_0,\theta_0)\| \leq D.$$

In (B7) we assume that the Fisher Information is invertible. This is a classical assumption. If we assume furthermore, as in Assumption (A4), a smooth change in the distributions of the covariates, then $t_0 \mapsto \Sigma(t_0,\theta_0)$ is a continuous function. Moreover, taking the inverse of a matrix is a continuous operation and the set of invertible matrices is open. Thus, $t_0 \mapsto \Sigma(t_0,\theta_0)^{-1}$ is a continuous function on the compact set $\mathbb{T}$ and hence the set $\{\Sigma(t_0,\theta_0)^{-1} : t_0 \in \mathbb{T}\}$ is a compact subset of the open set of all invertible matrices. This implies the first requirement of Assumption (B7).

The second assumption in (B7) concerning the derivatives of $\Sigma(t_0,\theta_0)$ is not restrictive because $\Sigma(s,\theta_0)$ does not depend on n.

We present now the Lemmas which we referred to in (B2). We formulate them without assumptions because Theorem 4.3 holds whenever the assumptions (B1) - (B7) and the following lemmas are true. The assumptions (B1), (B3) - (B7) seem to be very basic and unavoidable while the following lemmas may be provable also in different scenarios. After the lemmas we will present a situation in which they can be proven.

Statement 4.1. *ℓ_n and P_n are twice continuously differentiable in θ and differential and integral can be interchanged.*

Statement 4.2. *The function*

$$\Sigma(s,\theta) := \mathbb{E}\left[\partial_\theta^2 \log \lambda_{n,i}(\theta,s) \cdot \lambda_{n,i}(\theta_0,s) - \partial_\theta^2 \lambda_{n,i}(\theta,s) \middle| C_{n,i}(s) = 1\right]$$

is Lipschitz continuous in θ for every fixed s, i.e., there is $\gamma_\Sigma(s)$ such that for all $s \in \mathbb{T}$ and $\theta_1, \theta_2 \in \Theta$

$$|\Sigma(s,\theta_1) - \Sigma(s,\theta_2)| \leq \gamma_\Sigma(s)\|\theta_1 - \theta_2\|.$$

Additionally, for $\gamma_\Sigma(s)$ it holds

$$\sup_{t_0 \in \mathbb{T}} \frac{1}{h} \int_0^T K\left(\frac{s-t_0}{h}\right) \frac{p_n(s)}{\bar{p}_n(t_0)} \gamma_\Sigma(s) ds = O_P(1).$$

Statement 4.3. *Fubini can be applied to*

$$\frac{1}{h^{\frac{1}{2}} r_n} \sum_{i,j \in G_n} \int_0^T \int_0^T \int_\delta^{T-\delta} \frac{1}{h} K\left(\frac{s-t_0}{h}\right) K\left(\frac{t-t_0}{h}\right) \times \frac{\partial_\theta \lambda_{n,i}(\theta_0,s)^T}{\lambda_{n,i}(\theta_0,s)} \Sigma(t_0)^{-T} \Sigma(t_0)^{-1} \frac{\partial_\theta \lambda_{n,j}(\theta_0,t)}{\lambda_{n,j}(\theta_0,t)} \frac{w(t_0)}{\bar{p}_n(t_0)} dt_0 dM_{n,i}(s) dM_{n,j}(t)$$

and

$$\frac{1}{h^{\frac{1}{2}} r_n} \sum_{i,j \in G_n} \int_0^T \int_0^T f_{n,ij}(s,t) \mathbb{1}_{t \leq s} dM_{n,i}(s) dM_{n,j}(t).$$

Statement 4.4. *For $g_{n,i}(s) = h^{-\frac{1}{2}} \int_0^{s-} f_{n,ii}(s,t) dM_{n,i}(t)$, where $f_{n,ii}$ is defined in (5.40), we have*

$$\frac{1}{r_n^2} \sum_{i \in G_n} \mathbb{E}\left(\int_0^T g_{n,i}(s)^2 C_{n,i}(s) \lambda_{n,i}(\theta_0,s) ds\right) \to 0$$

as $n \to \infty$.

Statement 4.5. *There is a sequence A_n such that for $n \to \infty$*

$$\frac{1}{h^{\frac{1}{2}} r_n} \sum_{i \in G_n} \int_0^T f_{n,ii}(s,s) dN_{n,i}(s) - h^{-\frac{1}{2}} A_n \xrightarrow{\mathbb{P}} 0.$$

Statement 4.6. *It holds*

$$\frac{4}{hr_n^2} \sum_{i \in G_n} \int_0^T \sum_{\substack{j \in G_n \\ j \neq i}} \tau_{n,ij}(s)^2 C_{n,i}(s) \lambda_{n,i}(\theta_0,s) ds \xrightarrow{\mathbb{P}} B$$

and

$$\frac{4}{hr_n^2} \sum_{i \in G_n} \int_0^T \sum_{\substack{j_1,j_2 \in G_n \\ j_1,j_2 \neq i \\ j_1 \neq j_2}} \tau_{n,ij_1}(s) \tau_{n,ij_2}(s) C_{n,i}(s) \lambda_{n,i}(\theta_0,s) ds \xrightarrow{\mathbb{P}} 0,$$

where $\tau_{n,ij}$ is defined in (5.44). In the case of a Cox-type model (cf. (4.3)), we have

$$B := 4K^{(4)} \int_0^T \text{trace}\left(\left(P(t)\Sigma^{-T}(\theta_0,t)\Sigma^{-1}(\theta_0,t)\right)^2\right) w^2(t) dt,$$

where

$$K^{(4)} := \int_0^2 \left(\int_{-1}^1 K(v)K(u+v)dv \right)^2 du,$$
$$P^{r_1,r_2}(t) := \mathbb{E}\left(X_{n,j}^{r_1}(t) X_{n,j}^{r_2}(t) \lambda_{n,j}(t) \middle| C_{n,j}(t) = 1 \right).$$

Statement 4.7. *For any* $\varepsilon > 0$

$$\frac{2}{h^{\frac{1}{2}} r_n} \sum_{i \in G_n} \int_0^T \mathbb{1}\left(\left| \frac{2}{h^{\frac{1}{2}} r_n} \sum_{\substack{j \in G_n \\ j \neq i}} \tau_{n,ij}(s) \right| > \varepsilon \right) \sum_{\substack{j \in G_n \\ j \neq i}} \tau_{n,ij}(s) dM_{n,i}(s) \xrightarrow{\mathbb{P}} 0$$

Statement 4.8. *There is a sequence* B_n *with* $B_n = O_P(1)$*, such that for all* $t_0 \in \mathbb{T}$

$$\left\| \left[\frac{1}{r_n \bar{p}_n(t_0)} \ell_n''(\theta_0, t_0) \right]^{-1} \right\| \leq B_n.$$

Statement 4.9. *There is a sequence* K_n *with* $K_n = O_P(1)$ *such that for all* θ_1, θ_2 *and* $t \in \mathbb{T}$

$$\frac{1}{r_n \bar{p}_n(t_0)} \left\| \ell_n''(\theta_1, t) - \ell_n''(\theta_2, t) \right\| \leq K_n \cdot \|\theta_1 - \theta_2\|.$$

Statement 4.10. *The functions* $\widetilde{H}_{n,i}(s, \theta)$ *(defined in the beginning of the proof of Proposition 5.13) are Hoelder continuous for every* s *in* θ *with exponent* p *and (random) constant* $\gamma_{n,i}(s)$*, i.e., for all* $\theta_1, \theta_2 \in \Theta : |\widetilde{H}_{n,i}(s, \theta_1) - \widetilde{H}_{n,i}(s, \theta_2)| \leq \gamma_{n,i}(s) \cdot |\theta_1 - \theta_2|^p$*. Moreover, there exists* $k_0 \in \mathbb{N}$ *such that for* $\delta_n := \sqrt{\frac{\log r_n p_n}{r_n p_n \cdot h}}$

$$\begin{aligned}
\frac{1}{r_n} \sum_{i \in G_n} \int_0^T \|\partial_\theta \log \lambda_{n,i}(\theta_0, t)\| \cdot C_{n,i}(t) \lambda_{n,i}(\theta_0, t) dt &= O_P\left(\frac{n^{k_0 \cdot \min(\alpha_K, \alpha_p/2)} \sqrt{h p_n \log n}}{\left(H_K + K\sqrt{H_{n,p} p_n}\right) r_n} \right) \\
\sup_{\theta \in \Theta} \frac{1}{r_n} \sum_{i \in G_n} \int_0^T \|\widetilde{H}_{n,i}(s, \theta)\| ds &= O_P\left(p_n h n^{k_0 \alpha_K} \delta_n \right) \\
\sup_{\theta \in \Theta, t_0 \in \mathbb{T}} \frac{1}{r_n} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h} \right) \|\widetilde{H}_{n,i}(s, \theta)\| ds &= O_P\left(\frac{n^{k_0 \alpha_p} \delta_n}{H_{n,p} h^{\alpha_p}} \right) \\
\sup_{t_0 \in \mathbb{T}} \frac{1}{r_n} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h} \right) \gamma_{n,i}(s) ds &= O_P\left(n^{k_0 p} p_n \delta_n \right)
\end{aligned}$$

Statement 4.11. *Let* k_0 *be as in Statement 4.10, then there is* $C > 0$ *such that*

$$\sup_{t_0 \in T_{n,k_0}} \mathbb{P}\left(\left\| \frac{\ell_n'(\theta_0, t_0)}{r_n \sqrt{\bar{p}_n(t_0)}} \right\| \geq C \sqrt{\frac{\log r_n}{r_n h}} \right) = o\left(h^{-1} n^{-k_0} \right).$$

Statement 4.12. *Let k_0 be as in Statement 4.10, then there is $C > 0$ such that*

$$\mathbb{P}\left(\sup_{(t_0,\theta)\in T_{n,k_0}}\left|\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}\int_0^T\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\widetilde{H}_{n,i}(s,\theta)ds\right|>C\cdot\sqrt{\frac{\log r_np_n}{r_np_n\cdot h}}\right)\to 0.$$

In Section 5.3 we prove propositions which exactly show under which of the following assumptions ((C1)-(C8)) the above statements hold. The assumptions are crucially based on the three dependence concepts introduced in Chapter 3. We begin by a set of very basic assumptions.

(C1) Smooth Structure

1. *$\lambda_{n,i}(\theta,t) = \exp(\theta^TX_{n,i}(t))$.*
2. *Assume that all covariates are continuous up to jumps and bounded, i.e., there is $\hat{K} > 0$ such that $\|X_{n,i}(s)\| \leq \hat{K}$ for all $i \in G_n$ and $s \in [0,T]$. Let $\Lambda > 0$ be such that $\exp\left(\theta^TX_{n,i}(s)\right) \leq \Lambda$ for all $\theta \in \Theta$, $i \in G_n$ and $s \in [0,T]$.*
3. *We assume that the function $t \mapsto \frac{1}{p_n(t)}$ is Hoelder continuous with exponent α_p and constant $H_{n,p}$.*
4. *Assume that $r_np_n \geq n^\psi$ for some $\psi > 0$. Moreover, the quantities $H_{n,p}, p_n$ and h behave like a power of n.*

Assumption (C1) imposes a lot of structure by assuming the Cox-type model in part 1. As we will see, we often consider derivatives of the log of the intensity function which becomes then particularly easy. Other than this very practical mathematical use, the Cox-type intensities have the easy interpretation as discussed after (4.3). The boundedness assumption in part 2 seems very strong but a bounded intensity function just means that the expected number of events in a shrinking interval converges to zero. This is a very realistic assumption if we consider events which cost the caster of the event some time, e.g., sending messages to another person. Even the most famous person will physically not be able to send more than a fixed number of messages in a fixed time interval, and so we may assume that the number of sent messages is proportional to the considered time interval. The continuity assumption ensures that Riemann integrals exist and measurability properties remain valid. This assumption seems not very restrictive. Part 3 is a stronger version of (B6). But as we discussed after (B6), it is still a reasonable assumption. Finally, part 4 ensures that the sparsity p_n is not decaying too fast. If we were to assume that $p_n = o(n^{-2})$, we would see that $r_np_n \to 0$ and hence we had effectively no edges to observe events. Clearly, asymptotic estimation is impossible in such a situation. So the assumption of polynomial decay in p_n is reasonable. The polynomial behaviour of h is standard in the kernel smoothing literature (cf. Tsybakov (2009)). Keeping this in my mind, together with the discussion after Assumption (B6), we see that $H_{n,p}$ behaves like $p_n^{-2}h^{-(1+\alpha_K)}$ justifying the polynomial behaviour of $H_{n,p}$.

In the next set of assumptions we refine the continuity statements from (A4).

(C2) Time-Continuity of Fisher-Information

1. *The function* $t \mapsto \mathbb{E}(X_{n,j}(t)X_{n,j}(t)^T\lambda_{n,j}(t)|C_{n,j}(t)=1)$ *is uniformly continuous.*
2. *The functions* $t \mapsto \Sigma(\theta_0,t)$, $t \mapsto \Sigma^{-1}(t,\theta_0)$ *and* $t \mapsto w(t)$ *are uniformly continuous.*

Like in (A4) (but without using the density) we require smooth behaviour of the distribution of the covariates over time. Because of Assumption (B1), 1 we have all both functions above are independent of n and hence the uniform continuity is actually a consequence of the continuity.

The next assumption is also a continuity assumption. We need two definitions, firstly,

$$\xi_{n,j}(t) := X_{n,j}(t)X_{n,j}(t)^T C_{n,j}(t)\lambda_{n,j}(\theta_0,t).$$

Moreover, define $p_n^*(t)$ such that

$$\frac{1}{p_n^*(t)} := \int_{-1}^{1} \frac{1}{\bar{p}_n(t-hv)}dv.$$

We want to motivate the next assumption before stating it. We assume that the random function $\xi_{n,i}(t)$ is continuous up to jumps. By the boundedness assumption in (C1), 2 we have that the height of the jumps is bounded by a constant $\iota > 0$. We assume furthermore that the continuous part of $\xi_{n,j}$ is for each j uniformly continuous with random modulus of continuity $\mathcal{C}_\xi$. Note that we do not indicate in the notation that $\mathcal{C}_\xi$ depends on n. This motivates the following assumptions.

(C3) Model Continuity

By Assumption (C1), 4, we may choose $k_X > 0$ *so large such that*

$$\frac{1}{p_n^2}\left(H_{n,p}^{-\frac{1}{\alpha_p}} n^{-k_X}\right)^{\alpha_K}, \ \frac{1}{p_n}n^{-k_X\alpha_p} = o(1).$$

Assume additionally that $k_X > 0$ *is so large such that the following holds.*

1. *Denote by* $\mathcal{C}_E$ *the modulus of continuity of*

$$t \mapsto \mathbb{E}\left(X_{n,j}(t)^T X_{n,j}(t)\lambda_{n,j}(t)\middle|C_{n,j}(t)=1\right)$$

   which exists by Assumption (C2), 1. Assume that $\frac{1}{p_n}\mathcal{C}_E\left(H_{n,p}^{-\frac{1}{\alpha_p}} \cdot n^{-k_X}\right) = o(1)$.

2. *Assume that there is a random function* $\mathcal{C}_\xi : \mathbb{R} \to \mathbb{R}$ *and a constant* $\iota > 0$ *such that*

$$|\xi_{n,j}(s) - \xi_{n,j}(t)| \le \mathcal{C}_\xi(|s-t|) + (\textit{Number of jumps of } \xi_{n,j} \textit{ on } [s,t]) \cdot \iota$$

   and

$$\frac{\mathcal{C}_\xi\left(n^{-k_X}\right)}{p_n^2} = o_P(1).$$

   Assume additionally that the number of jumps is small when the interval is sufficiently short, i.e.,

$$\sup_{s,t:|s-t|\le n^{-k_X}} \frac{1}{r_n p_n p_n(t+v)} \sum_{j\in G_n} (\textit{Number of jumps of } \xi_{n,j} \textit{ on } [s,t]) = o_P(1).$$

3. *By the Assumptions (C2), 2 and (B6) the functions* $t \mapsto \Sigma(t, \theta_0)^{-1}, w$ *and* $\frac{1}{\bar{p}_n(t)}$ *are uniformly continuous. Assume additionally that* p_n *is uniformly continuous as well. Then,* $t \mapsto \frac{1}{p_n^*(t)}$ *and* $\Xi : (t,s) \mapsto \left(\Sigma^{-T}\Sigma^{-1}\right)(t, \theta_0)\frac{w(t)p_n(s)}{\bar{p}_n(t)}$ *are uniformly continuous too. We denote the modulus of continuity of* Ξ *by* $\mathcal{C}$ *(note that both depend on* n*) and assume that* $\frac{1}{p_n}\mathcal{C}\left(2H_{n,p}^{-\frac{1}{\alpha_P}} n^{-k_X}\right) \to 0$.

We have motivated that the estimation in part 2 is reasonable before the statement of the assumption. The assumption on the moduli of continuity are also realistic because $H_{n,p}$ behaves like a power of n, cf. Assumption (C1), 4 and we may choose k_X as large as we want. It is furthermore realistic to assume that the number of jumps of $\xi_{n,j}$ is linked to jumps in the counting processes. These have bounded intensity functions and therefore the expected number of jumps in an interval is proportional to the interval length. Therefore it is to be expected that the average in part 2 converges to zero as required.

The next set of assumptions concerns the dependence between the interactions and covariates of different pairs of actors. We begin by formulating the mixing property. Let therefore $a > 0$ and define $\Delta_n := a \log n$. Consider for all $t_0 \in [0, T]$ a Δ_n-partition $G^{t_0}(k, m, \Delta_n)$ and define for $c_3 > 0$

$$
\begin{aligned}
\bar{p}_n(t_0) &:= \int_{t_0-h}^{t_0+h} \frac{1}{h} K\left(\frac{s-t_0}{h}\right) p_n(s)ds \\
\bar{p}_n^{k,m}(t_0) &:= \int_{t_0-h}^{t_0+h} \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \mathbb{E}\left[I_{n,1}^{k,m} C_{n,1}(s)\right] ds \\
E_{k,m}^{n,t_0} &:= r_n \cdot \bar{p}_n^{k,m}(t_0) \\
E_k^{n,t_0} &:= \sqrt{\frac{n\bar{p}_n(t_0)}{\log n\bar{p}_n(t_0)}} \cdot c_3 \\
S_k &:= \max_{m=1,\dots,r_n} \sum_{i \in G_n} I_{n,i}^{k,m} \\
\Gamma_n^{t_0} &:= \mathbb{1}\left(\frac{S_k^2 \cdot \log r_n \bar{p}_n(t_0)}{r_n \bar{p}_n(t_0)} \le c_3^2,\ S_k\sqrt{h} \ge 1\right).
\end{aligned}
$$

Note that $E_{k,m}^{n,t_0}$ is by definition the expected size of the m-th block of type k. E_m^{n,t_0} is defined in the largest possible way that the second requirement in the second part in Definition 3.28 holds true. The intuition behind E_k^{n,t_0} is that it is the expected size of the largest block of type k. In this light $\Gamma_n^{t_0}$ is the indicator function that the actual partitioning behaves properly.

(C4) β-Mixing I

Let $a > q(k_0 + 1) + 2$ *and such that*

$$n^{-a} \cdot \left(\left(h^{-1} + n^{k_X}\right) \cdot H_{n,p}^{\frac{1}{\alpha_P}} + h^{-2} p_n^{-\frac{1}{\alpha_K}}\right) r_n \to 0.$$

Assume that $\beta_{t_0}(\Delta_n) \leq \alpha \cdot \exp(-\Delta_n)$, where β_{t_0} is the β-Mixing coefficient as introduced in Definition 3.25 for the following choices of $Z_{n,i}$ (for all $t_0 \in [0,T]$) and α does not depend on $Z_{n,i}$

$$\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right) H_{n,i}(s,\theta)\Gamma_n^{t_0} ds,$$
$$\frac{1}{h}\int_0^T K\left(\frac{t-t_0}{h}\right) X_{n,i}(t) dM_{n,i}(t)\cdot\Gamma_n^{t_0},$$
$$\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right) \Gamma_n^{t_0} C_{n,i}(s) ds.$$

Assume furthermore that there is $c_2 > 0$ such that for all $k \in \{1, ..., \mathcal{K}\}$

$$\frac{1}{r_n \cdot \bar{p}_n(t_0)} \sum_{m=1}^{r_n} E_{k,m}^{n,t_0} \geq c_2. \tag{4.18}$$

Assume also that $\Gamma_n^{t_0}$ is so likely such that

$$\sup_{t\in[0,T]} \mathbb{E}\left(\left|\Gamma_n^t - 1\right| \middle| C_{n,j}(t) = 1\right) \to 0$$

and

$$\sup_{t_0\in[0,T]} \mathbb{P}(\Gamma_n^{t_0} = 0)$$
$$= \min\left\{O\left(\frac{\log r_n p_n}{r_n}\right), o\left(\frac{1}{hn^{k_0(q+1)}}\right), o\left(\left(\left(h^{-1} + n^{k_X}\right)\cdot H_{n,p}^{\frac{1}{\alpha_p}} + h^{-2} p_n^{-\frac{1}{\alpha_K}}\right)^{-1}\right)\right\}.$$

We motivated the idea behind the mixing assumptions in Section 3.3. The random variables we apply it here to depend only on one edge and the function $\Gamma_n^{t_0}$. That means that the mixing property is only required if the partitioning of the network is reasonable. However, we assume too that the probability that the partitioning is reasonable is overwhelmingly large. The inequality in (4.18) means that we assume that the percentage of the edges which are on average contained in the blocks of type k is never negligible, i.e., that no block type is obsolete, a plausible assumption. We also tacitly assume that the number of block types $\mathcal{K}$ is the same for all time points and does not change with n. This assumption might be disputable. It reflects the idea that the network geometry is staying the same while the network size is increasing.

Before going to a substantially different set of assumptions, we unfortunately have to assume that it is possible to define a partitioning of the network also with respect to another normalisation. Define to this end:

$$p_n^*(t) := \left(\int_{-1}^1 \frac{1}{\bar{p}_n(t+uh)} du\right)^{-1}$$

$$p_n^{*,k,m}(t) := p_n^*(t)\mathbb{E}(I_{n,i}^{k,m}).$$

The quantities $E_{k,m}^{*,n,t}$, $E_k^{*,n,t}$, S_k^* and $\Gamma_n^{*,t}$ are defined analogously to before but where $\bar{p}_n$ and $\bar{p}_n^{k,m}$ are replaced with p_n^* and $p_n^{*,k,m}$ respectively. Using these quantities we formulate analogue assumptions to (C4).

(C4*) β-Mixing II

Assume that $\Gamma_n^{*,t}$ is so likely such that

$$H_{n,p}^{\frac{1}{\alpha_p}} n^{k_X} \cdot \sup_{t\in[0,T]} \mathbb{P}(\Gamma_n^{*,t} = 0) \to 0,$$

$$\sup_{t\in[0,T]} \mathbb{E}\left(\left|\Gamma_n^{*,t} - 1\right| \Big| C_{n,j}(t) = 1\right) \frac{p_n(t)}{p_n^*(t)} \to 0.$$

Assume furthermore that there is $c_2 > 0$ such that for all $k \in \{1, ..., \mathcal{K}\}$

$$\frac{1}{r_n \cdot p_n^*(t_0)} \sum_{m=1}^{r_n} E_{k,m}^{*,n,t_0} \geq c_2.$$

Let $\Delta_n = a \log n$ with $a > 0$ such that $n^a r_n \to 0$. Finally, assume that $\beta_{t_0}(\Delta_n) \leq \alpha \cdot \exp(-\Delta_n)$, where β_{t_0} is the β-Mixing coefficient as introduced in Definition 3.25 for $Z_i = \xi_{n,i}(t_0 + vh)\Gamma_n^{t_0}$. α does not depend on $t \in [0, T]$ and $v \in [0, 2]$).

The next assumption asserts momentary m-dependence as in Section 3.2.

(C5) Momentary-m-Dependence

The network is momentary-m-dependent in the sense of Definition 3.9.

The next set of assumptions is concerned with the existence and size of hubs. We define therefore what we mean by a hub. Let m be the constant from the m-dependence Assumption (C5) and let $A \subseteq G_n$ and $F > 0$ be arbitrary. We define then for all $t \in [0, T]$

$$K_m^A(t) := \sup_{k\in A} \sum_{i\in G_n} \sup_{u\in[0,2]} C_{n,i}(t + uh)\mathbb{1}(d_{t-4h}^n(i, k) \leq m)$$

$$H_{UB}^A \geq \sup_{k\in A} \sup_{t\in[0,T]} \mathbb{1}(K_m^k(t) > F)$$

$$N_{UB} = \sum_{i\in G_n} H_{UB}^i$$

$$A_n(t) := \sum_{i\in G_n} \sup_{u\in[0,2]} C_{n,i}(t + uh).$$

Intuitively speaking, $A_n(t)$ is the number of active edges at time t and for $k_0 \in G_n$ we have that $K_m^{k_0}(t)$ is the number of active edges which are closer to k_0 than m around time t. Then $K_m^A(t)$ is the largest of these numbers for all edges in A. We then call an edge $k_0 \in G_n$ a *hub of size F at time t* if $K_m^{k_0}(t) > F$. Then, H_{UB}^A is the number of edges in

A which at any time point during the observation point become a hub of size F. Lastly, N_{UB} is the number of all edges that become a hub of size F at some point. We think about this in the following way: Think of the example of a social media setting where every edge represents the connection between two people. In the works Golder et al. (2007); Huberman et al. (2008) it is argued that in social media most of the *friendships* between users are actually inactive in the sense that they do not interchange messages. This underpins the very much believable idea that every actor has only close contact to a bounded number of people. Having close contact means in our formulae that their distance is less than m. That means that most people interact with not more than, say F people, regardless of the size of the network. Thus, if one edge exceeds the threshold of F, we call it a hub. In the following assumptions we assume that the number of hubs is small. Note that $K_m^{G_n}(t)$ denotes the size of the largest hub at time t. We impose now the following assumptions on these quantities which look clumsy on first sight but we will discuss them afterwards.

(C6) Hub Behaviour

Assume that there is $F > 0$ such that H_{UB}^A is measurable with respect to $\mathcal{F}_0^n$ for all $A \subseteq G_n$. Moreover, for any $\varepsilon > 0$ there is $F_0 > 0$ such that for all $n \in \mathbb{N}$

$$\mathbb{P}\left(\forall t \in [0,T] : K_m^{G_n}(t) > F_0 K_m^{G_n}(t-4h)\right) < \varepsilon. \tag{4.19}$$

Moreover, we assume that

$$\sup_{j_1,j_2 \in G_n} \sup_{t \in [0,T]} \mathbb{E}\left(K_m^{G_n}(t-4h) H_{UB}^{j_1 j_2} \middle| \sup_{r \in [t-2h,t]} C_{n,j_2}(r) \cdot C_{n,j_1}(t) = 1 \right) = O(1) \tag{4.20}$$

$$\mathbb{E}\left(H_{UB}^{G_n} \left(\sup_{t\in[0,T]} K_m^{G_n}(t) \right)^4 \right) = O(1) \tag{4.21}$$

We assume also the following convergences:

$$\sup_{t\in[0,T]} \frac{1}{p_n^*(t)} \mathbb{E}\left(\int_{t-2h}^{t} d|M_{n,j_2}|(r) \cdot K_m^{j_2}(t+2h) \right) = O(1) \tag{4.22}$$

$$\sup_{t\in[0,T]} \frac{1}{r_n p_n^*(t)^2} \mathbb{E}\left(\int_{t-2h}^{t} d|M_{n,j_2}|(r) \cdot K_m^{G_n}(t+2h)(N_{UB} + r_n H_{UB}^{j_2}) \right) = o(1) \tag{4.23}$$

$$\begin{aligned} &\frac{1}{r_n^2} \sum_{\substack{j_1,j_2 \in G_n \\ j_1 \neq j_2}} \mathbb{E}\Bigg(\int_0^T \int_{t-2h}^{t-} \frac{1}{p_n^*(t)} d|M_{n,j_2}|(r) \cdot \frac{A_n(t)}{r_n p_n^*(t)} K_m^{j_1}(t) \times \sup_{\xi \in [t,t+2h]} \frac{A_n(\xi)}{r_n p_n(\xi)} \\ &\quad \times \sup_{\substack{j_1',j_2' \in G_n \\ j_1' \neq j_2'}} \int_t^{t+2h} \int_{\xi-2h}^{\xi-} \frac{1}{p_n(\xi)} d|M_{n,j_2'}|(\rho) d|M_{n,j_1}|(\xi) d|M_{n,j_1}|(t) \Bigg) = o(1) \end{aligned} \tag{4.24}$$

$$\frac{1}{r_n^2} \sum_{\substack{j_1,j_2,j_1' \in G_n \\ j_1 \neq j_2}} \mathbb{E}\Bigg(\int_0^T \int_{t-2h}^{t-} \frac{1}{r_n p_n^*(t)^2} K_m^{G_n}(t)(N_{UB} + r_n H_{UB}^{j_1'}) d|M_{n,j_2}|(r)$$

$$\times \sup_{\xi\in[t,t+2h]} \frac{A_n(\xi)}{r_n p_n^*(\xi)} \times \sup_{\substack{j_1',j_2'\in G_n\\ j_1'\neq j_2'}} \int_t^{t+2h} \int_{\xi-2h}^{\xi-} \frac{1}{p_n(\xi)} d|M_{n,j_2'}|(\rho) d|M_{n,j_1}|(\xi) d|M_{n,j_1}|(t)\Bigg)$$
$$= o(1) \tag{4.25}$$

$$\frac{1}{r_n^2} \sum_{\substack{j_2,j_2'\in G_n\\ j_2\neq j_2'}} \int_0^T \mathbb{E}\Bigg[\int_{t-2h}^{t-} \frac{1}{p_n^*(t)} d|M_{n,j_2}|(r) \int_{t-2h}^{t-} \frac{1}{p_n^*(t)} d|M_{n,j_2'}|(r')$$
$$\times \left(\frac{A_n(t)}{r_n p_n^*(t)}\right)^2 \left(K_m^{j_2}(t+2h) + K_m^{j_2'}(t+2h)\right)\Bigg] dt = o(1) \tag{4.26}$$

$$\int_0^T \frac{p_n(t)}{r_n p_n^*(t)^2} \mathbb{E}\left(\int_{t-2h}^{t-} K_m^{G_n}(t-4h) H_{UB}^j d|M_{n,j}|(r) \lambda_{n,j}(t) \middle| C_{n,j}(t)=1\right) dt = o(1) \tag{4.27}$$

$$\frac{1}{r_n^2} \sum_{i,j\in G_n} \mathbb{E}\Bigg(\int_0^T \frac{1}{r_n p_n^*(t)^2} \left(F + K_m^{G_n}(t) H_{UB}^{ij}\right) d|M_{n,i}|(t)$$
$$\times \int_0^T \frac{1}{r_n p_n^*(t)^2} \left(F + K_m^{G_n}(t) H_{UB}^{ij}\right) d|M_{n,j}|(t)\Bigg) = O(1) \tag{4.28}$$

$$\frac{1}{r_n^2} \sum_{i,j\in G_n} \mathbb{E}\left(\int_0^T \frac{A_n(t)}{r_n p_n^*(t)^2} d|M_{n,i}|(t) \cdot \int_0^T \frac{1}{r_n p_n^*(t)^2} \left(F + K_m^{G_n}(t) H_{UB}^{ij}\right) d|M_{n,j}|(t)\right) = o(1) \tag{4.29}$$

The main problem in posing these assumptions is that several terms appear together in an expectation and we cannot handle them together. However, it is always clear that the expectations of the terms separately behave as required. We observe to this end firstly that $A_n(t)$ is the number of active edges and behaves thus like $r_n \cdot p_n(t)$. If we assume asymptotic uncorrelation as in Section 3.1 then also $\mathbb{E}\left(\left(\frac{A_n(t)}{r_n p_n(t)}\right)^2\right)$ behaves like a constant. If we additionally assume that $p_n^*(t)$ and $p_n(t)$ do not differ too much all expressions in (C6) involving $A_n(t)$ are bounded. Moreover, we always have

$$\mathbb{E}\left(\int_{t-2h}^{t} dM_{n,j}(r)\right) = \int_{t-2h}^{t} p_n(r) \mathbb{E}\left(\lambda_{n,j}(\theta_0, r) | C_{n,j}(r) = 1\right) dr.$$

Thus we see that $\int_{t-2h}^t dM_{n,j}(r)$ can always compensate for one $\frac{1}{p_n(t)}$ or $\frac{1}{p_n^*(t)}$.

More discussion is required for the terms which involve K_m^A, H_{UB}^A and N_{UB}. As motivated before the statement of (C6), we assume that everybody is interacting only with a bounded number of actors actively and thus $K^j_m(t)$ is bounded by F for most $j \in G_n$. There are only very very few edges j for which $K^j_m(t)$ exceeds the threshold F. Thus, we assume that the expectation of $K^j_m(t)$ is bounded. We assume also that the ability of an edge to exceed this threshold is determined before the observation period, thus justifying the assumption that H^A_{UB} is measurable with respect to $\mathcal{F}^n_0$. So the fact whether an edge can become a hub or not is pre-determined and does not change during the observation period. However, the size of a hub $K^{G_n}_m(t)$ may change during the observation period. We just assume in (4.19) that this change is sufficiently slowly. Note moreover that we allow the hub size $K^{G_n}_m(t)$ to be very large, so it might even be reasonable to assume that it's constant. But we assume that existence of hubs is so rare that the number of edges which are contained in a hub, i.e., $K^{G_n}_m(t) \cdot N_{UB}$, can be controlled by the effective number of active edges $r_n p_n(t)$. This is reasonable if we assume that there is only a bounded number of hubs (maybe even only one or two) whose size is proportional to the number of active edges. Recall that a hub in our definition is an edge that is strongly and significantly connected to many other edges (up to the order of the number of all active edges). It is intuitive to assume that there are not so many edges with this property. Lastly, $r_n H^j_{UB}$ is assumed to behave like N_{UB}, in expectation this is easy to see and we assume that it holds here in more complicated senses too.

When we sum in (C6) over $j_1, j_2 \in G_n$, we also want to use asymptotic uncorrelation properties which we exemplify in the following assumption.

(C7) Asymptotic Uncorrelation

Let $i_0, j_0 \in G_n$ be two edges with $\nu := |e_{i_0} \cap e_{j_0}| \leq 1$ and assume that

$$\sup_{t_0 \in [0,T]} \iint_{[0,t]^2} \frac{1}{h^2} K\left(\frac{s-t_0}{h}\right) K\left(\frac{r-t_0}{h}\right) \frac{\left|Cov\left(C_{n,j_0}(r) I^{k,m}_{n,j_0}, C_{n,i_0}(s) I^{k,m}_{n,i_0}\right)\right|}{\bar{p}^{k,m}_n(t_0)} dsdr$$
$$= \begin{cases} O\left(\frac{1}{n^2}\right) & , \quad |e_{i_0} \cap e_{j_0}| = 0 \\ O\left(\frac{1}{n}\right) & , \quad |e_{i_0} \cap e_{j_0}| = 1 \end{cases} \tag{4.30}$$

Assume furthermore that

$$\sup_{t_0 \in [0,T], \theta \in \Theta} n^{2-\nu} \iint_{[0,T]^2} \frac{1}{h^2} K\left(\frac{s-t_0}{h}\right) K\left(\frac{r-t_0}{h}\right) \frac{\mathbb{P}(C_{n,i_0}(s) I^{k,m}_{n,i_0} C_{n,j_0}(r) I^{k,m}_{n,j_0} = 1)}{\bar{p}^{k,m}_n(t_0)}$$
$$\times \mathbb{E}\Bigg(\left[\widetilde{\lambda}_{n,i_0}(\theta,s)\Gamma^{t_0}_n - \mathbb{E}(\widetilde{\lambda}_{n,i_0}(\theta,s)\Gamma^{t_0}_n | C_{n,i_0}(s) I^{k,m}_{n,i_0} = 1)\right]$$
$$\times \left[\widetilde{\lambda}_{n,i_0}(\theta,s)\Gamma^{t_0}_n - \mathbb{E}(\widetilde{\lambda}_{n,j_0}(\theta,r)\Gamma^{t_0}_n | C_{n,j_0}(r) I^{k,m}_{n,j_0} = 1)\right] \Big| C_{n,i_0}(s) I^{k,m}_{n,i_0} C_{n,j_0}(r) I^{k,m}_{n,j_0} = 1 \Bigg)$$
$$= O(1) \tag{4.31}$$

$$\sup_{t_0 \in [0,T]} n^{2-\nu} \frac{1}{h^2} \iint_{[0,T]^2} K\left(\frac{s-t_0}{h}\right) K\left(\frac{t-t_0}{h}\right)$$
$$\times \frac{Cov(C_{n,i_0}(s) I^{k,m}_{n,i_0} \Gamma^{t_0}_n, C_{n,j_0}(t) I^{k,m}_{n,j_0} \Gamma^{t_0}_n)}{\bar{p}^{k,m}_n(t_0)} dsdt = O(1) \tag{4.32}$$

$$\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}} \int_0^T \frac{1}{r_n^2 p_n^*(t)^2}\mathbb{P}\left(\sup_{r\in[t-2h,t]} C_{n,j_2}(r)\cdot C_{n,j_1}(t)=1\right)dt = O(1) \tag{4.33}$$

$$\frac{1}{r_n p_n^{*,k,m}(t)}\sum_{i,j\in G_n} Cov\left(\xi_{n,i}(t+vh)\Gamma_n^{*,t}I_{n,i}^{k,m,t}(\Delta_n),\xi_{n,j}(t+vh)\Gamma_n^{*,t}I_{n,j}^{k,m,t}(\Delta_n)\right)=O(1) \tag{4.34}$$

These assumptions in (C7) are really asymptotic uncorrelation assumptions as earlier imposed and described in Section 3.1.

The last set of assumptions is concerned again with $A_n(t)$ but this time without involvement of the hubs. So the same heuristic as explained in connection with assumption (C6) can be applied here. The only new statement will be (4.39) which uses

$$\mathcal{N}_n := \sup_{j\in G_n, t\in[0,T]} N_{n,j}([t-2h,t)) + 2h\Lambda,$$

the largest number of interactions an edge can have in an interval of length h.

(C8) Edge Stability

We assume that for all $j\in G_n$:

$$\int_0^T \mathbb{E}\left(\frac{A_n(t)}{r_n p_n^*(t)}\cdot\int_{t-2h}^{t-} d|M_{n,j}|(r)\middle| C_{n,j}(t)=1\right)dt = o(1) \tag{4.35}$$

$$\int_0^T \frac{p_n(t)}{r_n p_n^*(t)^2}\mathbb{E}\left(\left(\frac{A_n(t)}{r_n p_n^*(t)}\right)^2\middle| C_{n,j}(t)=1\right)dt = O(1) \tag{4.36}$$

$$\frac{1}{r_n^2}\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}} \int_0^T\int_{t-2h}^t \mathbb{E}\left[\left(\frac{A_n(t)}{r_n p_n^*(t)}\right)^2\cdot\frac{C_{n,j_1}(t)C_{n,j_2}(r)}{p_n^*(t)^2}\right]drdt = o(1) \tag{4.37}$$

$$\int_0^T \frac{p_n(t)}{r_n p_n^*(t)^2}\mathbb{E}\left(\int_{t-2h}^{t-} d|M_{n,j}|(r)\lambda_{n,j}(t)\middle| C_{n,j}(t)=1\right)dt = o(1) \tag{4.38}$$

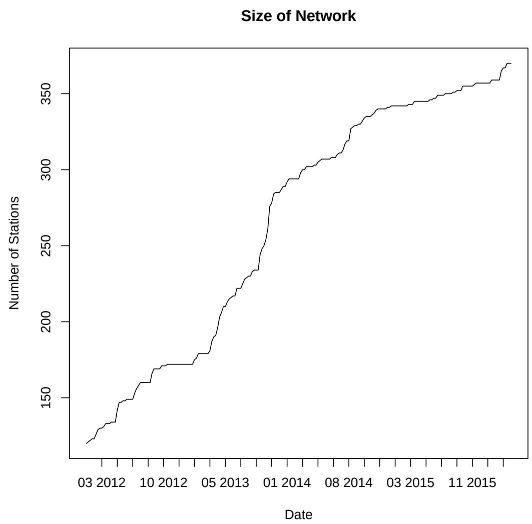
$$\mathbb{E}\left(\left(\sup_{n\in\mathbb{N}}\frac{\mathcal{N}_n}{r_n}\sum_{j=1}^{r_n}\int_0^T \frac{A_n(t-2h)}{r_n p_n^*(t)}\cdot\frac{C_{n,j}(t)}{p_n^*(t)}d|M_{n,j}|(t)\right)^4\right) < +\infty \tag{4.39}$$

As already mentioned, for assumptions (4.35)-(4.38) we can apply the same heuristic as before. For assumption (4.39), we note firstly that it is reasonable to think of $\mathcal{N}_n$ as a random variable which has finite moments because most of the edges will not carry many interactions in a short time interval. For the remaining part of (4.39), the same intuition applies as before as this is just a variation on A_n.

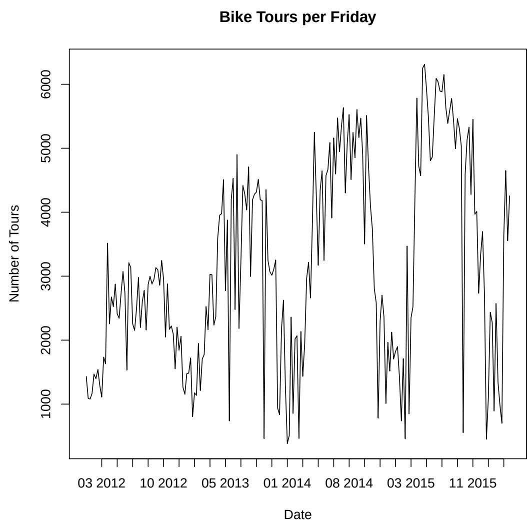
4.5 Application to Bike Data

We intend to illustrate the finite sample performance of our estimation procedure described above, by considering the Capital Bikeshare (CB) Performance Data, publicly available at *http://www.capitalbikeshare.com/system-data*. This data describes the usage of the CB-system at Washington D.C. from Jan. 2012 to March 2016. Using this data, we construct a network as follows. Each bike station v will become a node in our network, and edges between two stations v, v' will be active depending on whether a bike was rented at station v and returned to station v' (or vice versa) at the same day. So, in our analysis, rentals over several days have been ignored. We ignore the direction of travel as well because we aggregate over days and assume that directed effects cancel out (most riders go one way in the morning and the other way in the evening).

It should be noted that, while we believe that this example serves as a serious and interesting illustration of our proposed method, it is not meant to be a full-fledged analysis of bike sharing performance. In particular, for computational and coding simplicity, we ignored that bike stations might be full or empty and thus prohibiting certain bike rides. Also, the authors' personal bike sharing experience is that entirely empty or full bike stations are not encountered too often, and so the hope is that the bias induced by ignoring this effect is negligible.



(a) Shows the number of available bike stations



(b) Number of tours undertaken per Friday

Figure 4.1: Simple descriptive statistics of the bike data

Figure 4.1 shows some summary statistics of the data. In Figure 4.1a, we see the number of available bike stations, which is strongly increasing. Figure 4.1b shows the number of bike tours on Fridays. An obvious periodicity is visible: The cycling activity is much lower in winter than during summer.

4.5.1 Modelling Approach

We aim at modelling the bike sharing activities on Fridays, seen on the right panel of Figure 4.1. We choose Fridays because we believe that there is a difference between different days, so we wanted to concentrate on one day, Friday. We use the event modelling approach as introduced in Section 4.1, where *event* here means that a bike is rented at station v and returned at station v', or vice versa. We will also refer to this event as a tour between v and v'. In order to reduce computational complexity to a minimum (fitting the model takes several minutes on a standard laptop), we assume that the covariates change only at midnight and stay constant over the day. Furthermore, we estimate the time-varying parameter function θ only for one time point per day, namely 12pm noon. The next paragraph contains more details.

Since we do not consider any asymptotics here, we omit the index n. Time t is measured in hours of consecutive Fridays. So, if k is the current week, and r is the time on Friday (in 24h), then $t := (k-1)\cdot 24 + r$. Thus, with $r_t := (t \mod 24)$, the quantity $k_t := \frac{t-r_t}{24} + 1$ gives the week the time point t falls into. The processes $N_i(t)$, counting the number of tours on $i = (v, v')$ on Fridays, are modelled as counting processes with intensities $\lambda_i(\theta_0(t), t) := \alpha(t)\exp(\theta_0(t)^T X_i(k_t)) \cdot C_i(k_t)$. The covariate vector $X_i(k_t)$ and the censoring indicator $C_i(k_t)$ will be defined later. Note that both of them depend on k_t only, i.e., on the current week, and not on the actual time on the Friday under consideration. The function α is 24-periodic and integrates to one over a period, i.e., $\alpha(t) = \alpha(t+24)$ and $\int_t^{t+24}\alpha(s)ds = 1$. The function α is introduced to model the reasonable assumption that the activity varies during the day. Suppose now, that our target is the estimation of the parameter vector $\theta_0(t_0)$ with $t_0 = (k_{t_0} - 1)24 + r_0$ and $r_0 = 12$, say. We choose a piecewise constant kernel K with $K((24k+x)/h) = K(24k/h)$, for all $k \in \mathbb{N}$ and $0 \leq x < 24$. Substituting in these choices of the intensity and the kernel to the log-likelihood (4.1), we see that our maximum likelihood estimator maximizes the function

$$\theta \mapsto \sum_{k=0}^{k_T} K_\kappa(k - k_{t_0})\theta^T X_i(k) \int_{k\cdot 24}^{(k+1)\cdot 24} dN_i(t) \\ - \sum_{k=0}^{k_T} K_\kappa(k - k_{t_0}) \exp(\theta^T X_i(k)) C_i(k),$$

where $\int_{k\cdot 24}^{(k+1)\cdot 24} dN_i(t)$ gives the number of tours on i on the Friday in week k, and where $K_\kappa(k) = K(k/\kappa)$ with $\kappa = h/24$. In our empirical analysis, we chose $K_\kappa(k)$ as triangle weights with support $\{-\kappa, ..., \kappa\}$ and considered only integer choices of the bandwidth κ. We will choose $\kappa = 12$ as the result of a bandwidth selection procedure based on one-sided cross-validation which is discussed in Section 4.5.4.

We explain now the choice of our covariate vector X_i. Denote by $\Delta_i(k, d)$ the number of tours on i on day d in week k, where $d = 4$ means Monday and $d = 7$ refers to Thursday (for us the week starts on Fridays, i.e. Friday is $d = 1$). For $r \in (0, 1)$, we encode the activity on i in week k as $A_{i,k} = (1-r)\sum_{d=4}^{7} r^{7-d}\Delta_i(k, d)$ (mind the limits of the summation - Fridays are not included). In our simulations, we chose $r = 0.8$ (this choice is somewhat arbitrary, and a full study of the data would include investigating the sensitivity of the parameter estimate on the choice of r as well as a data driven choice. We do not attempt to do this here). For further covariates, we construct a network $G(k)$, for every week k, by connecting v and v' if and only if there was at least one tour on the Friday in that week (note that these networks are auxiliary networks for constructing the covariates and that they are different from G_n). For $i = (v_1, v_2)$, we denote by $I_{i,k}$ the number of common neighbours of $v_1 i$ and v_2 in the graph $G(k)$. We let $d_{v,k}$ be the degree of node v in $G(k)$, $T_{i,k}$ the number of tours on i on the Friday in the k-th week, and $T_{i,k,k-1} = (T_{i,k} + T_{i,k-1})/2$ the average number of tours on the two Fridays in weeks k and $k-1$. Finally we collect everything in the covariate vector: Let $i = (v_1, v_2)$

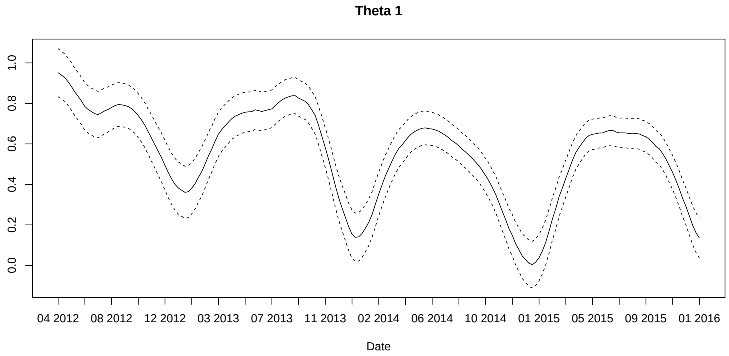
$$X_i(k) \;:=\; \Big(1, A_{i,k-1}, I_{i,k-1}, \max(d_{v_1,k-1}, d_{v_2,k-1}), T_{i,k-1,k-2}, \mathbb{1}(T_{i,k-1,k-2} = 0)\Big)^T.$$

The censoring indicator function C_i is defined to be equal to zero, if there was no tour on i in the last four weeks. In summary, we estimate a total of six parameter curves, corresponding to the effects of six covariates in our model (the term neighbour refers to adjacent vertices in the network $G(k)$):

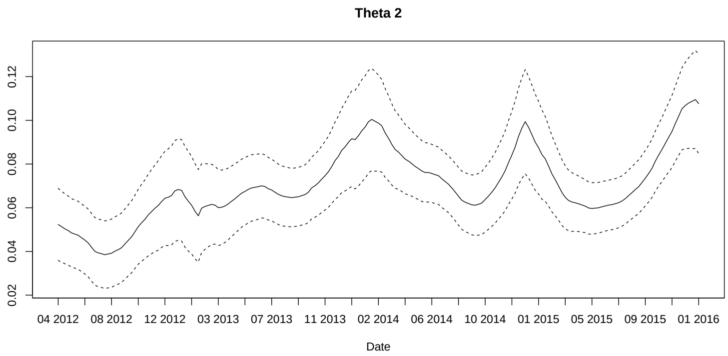
- $\theta_1(t) \quad \triangleq \quad$ baseline
- $\theta_2(t) \quad \triangleq \quad$ activity between stations on previous week-days
- $\theta_3(t) \quad \triangleq \quad$ common neighbours of stations
- $\theta_4(t) \quad \triangleq \quad$ popularity of station, measured by degrees
- $\theta_5(t) \quad \triangleq \quad$ activity between stations on two previous Fridays
- $\theta_6(t) \quad \triangleq \quad$ inactivity between stations on two previous Fridays

4.5.2 Estimation Results

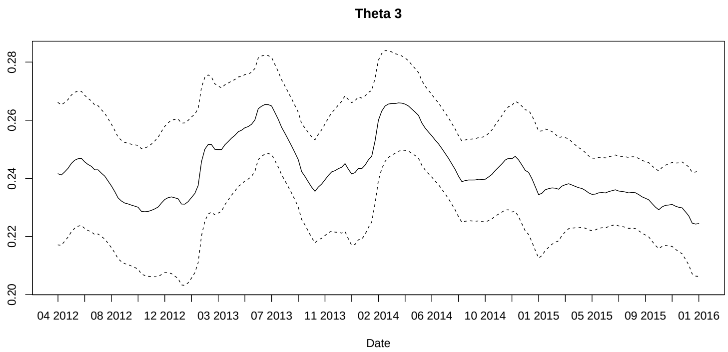
The resulting estimated parameter curves are shown in Figures 4.2 and 4.3. In all six parameter curves in Figures 4.2 and 4.3, we observe a clearly visible seasonality. Looking at Figure 4.2b, we see that importance of the activity in the week (Monday to Thursday) is higher during the winter months than in the summer. A plausible interpretation for this might be that the weather in winter is more persistent, e.g., when there is snow it is likely to remain for a while. Hence, the behaviour of the opportunist cyclists is probably more predictable in winter. Probably, only few people keep using a bike, regardless of the weather. This makes the activity in the week a better predictor.



(a) Estimate and confidence bands for baseline weight

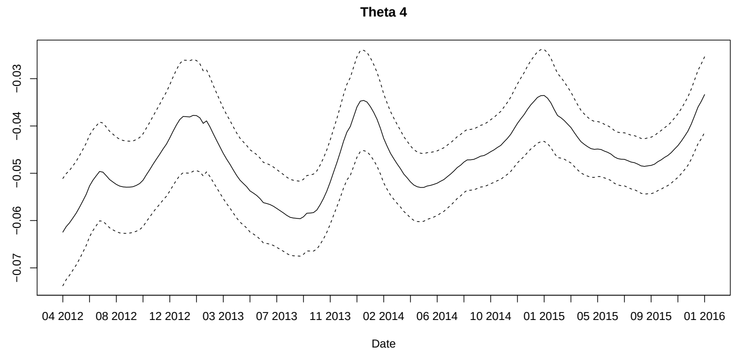


(b) Estimate and confidence bands for previous week-day activity between stations weight

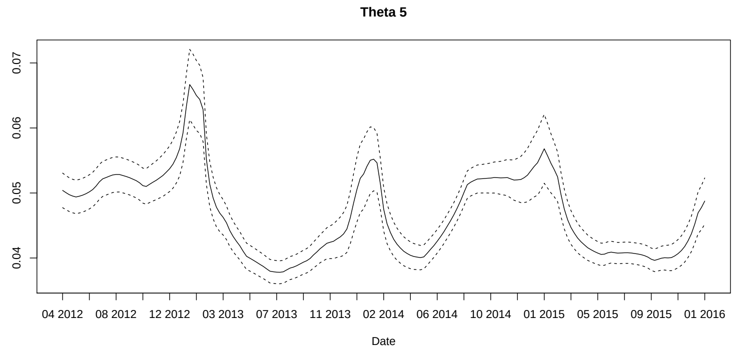


(c) Estimate and confidence bands for common neighbours weight

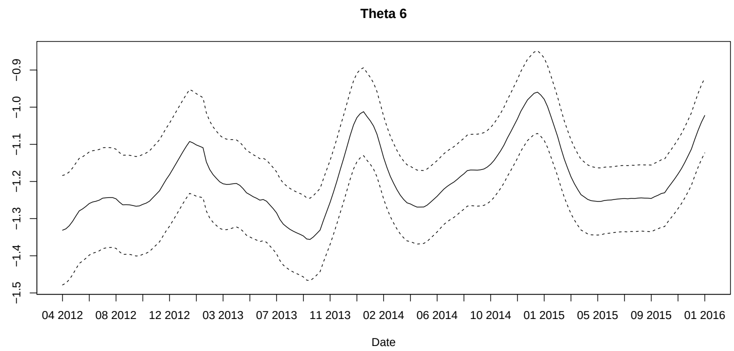
Figure 4.2: Estimates of $\theta_1(t)$, $\theta_2(t)$ and $\theta_3(t)$. The dotted lines indicate (point-wise) 99% confidence regions (plus minus 2.58 times the asymptotic standard deviation).



(a) Estimate and confidence bands for maximal degree weight



(b) Estimate and confidence bands for previous Friday activity between stations weight



(c) Estimate and confidence bands for inactivity on previous Fridays weight

Figure 4.3: Estimates of $\theta_4(t)$, $\theta_5(t)$ and $\theta_6(t)$. The dotted lines indicate (point-wise) 99% confidence regions (plus minus 2.58 times the asymptotic standard deviation).

Figure 4.2c shows that the number of common neighbours always has a significant positive effect on the hazard. This reflects the empirical finding that observed networks cluster more than totally random networks (e.g. Jackson (2008)).

The influence of the popularity of the involved bike stations is investigated in Figure 4.3a (measured by the degree of the bike station). Interestingly, it always has a significant negative impact. The size of the impact is higher in the summer months, which again supports the hypothesis that in summer the behaviour of the network as a whole appears *more random* than in winter. But still, the negative impact is a bit unforeseen. This finding can be interpreted as the observed network having no hubs. Another reason for this effect might be, that stations can only host a fixed number of bikes: If a station v is empty, no new neighbours can be formed. A similar saturation effect happens if a lot of bikes arrive at station v. Moreover, it is plausible that effects caused by the degrees are already included in 4.2b, as well as in Figure 4.3b. They show the effect of the bike rides on the days immediately preceding the current Friday, and the effect of the average number of bike tours on the last two Fridays, respectively. In Figure 4.3b, we observe a similar behaviour as in Figure 4.2b (even more pronounced): In summer the predictive power of the tours on the last two Fridays is significantly lower than in winter, underpinning the theory that the destinations in summer tend to be based on more spontaneous decisions. Finally, in Figure 4.3c, we observe that no bike tours on the last two Fridays between a given pair of stations always has a significant negative impact on the hazard. Again a very plausible finding.

Feeding the estimation results in the test statistic T_n and applying Theorem 4.3 yields for the centred and standardized test statistic (i.e., the following value should be asymptotically standard normal distributed on the hypothesis that the parameter function is a constant) a value of roughly 26. This is highly unlikely and therefore providing strong evidence that a time varying parameter function is an appropriate choice. As an estimate for the constant parameter we have chosen here the average over the estimated time-varying parameter function.

4.5.3 Goodness of Fit Considerations

The computations in this section involve simulations which are computationally more demanding than the model fitting. All calculations have been executed on the BwForCluster (cf. Acknowledgement). In stochastic network analysis, a central strand of research is concerned with the question of whether characteristics observed in real networks can be adequately mimicked by stochastic network models. Important characteristics are degree distribution, clustering coefficient and diameter (these and other characteristics can be found in Jackson (2010) Chapter 2.2). The *degree distribution* of a network states how many vertices with a certain degree (i.e., the number of neighbours) are present in the network for all degrees. The *diameter* of a network is the longest among the shortest path between two vertices in the network. Typically, in observed networks the diameter is much smaller than the number of vertices (cf. Jackson (2010)). The *clustering coefficient* is the number of complete triangles (triples of vertices which are completely connected) divided by the number of incomplete triangles (triples of vertices with at least two edges). Note that every complete triangle is also incomplete, hence the clustering coefficient is between zero and one. The clustering coefficient can be understood as the empirical probability that vertices are connected given that there is a third vertex to which both are connected. It has been reported (cf. Jackson (2010)), that in observed networks this number is usually significantly higher than in an Erdös-Rényi network, where the presence of edges are i.i.d. random variables.

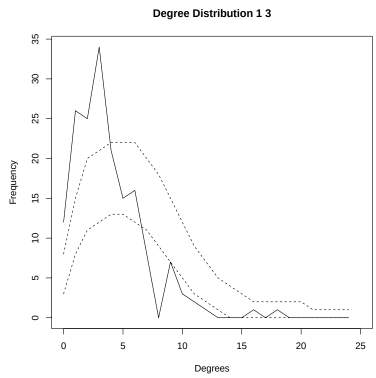
As in Zafarani et al. (2014), Chapter 4, we visually compare these three characteristics with a typical network produced by our model to see if the model produces networks similar to the real-world network. In order to see how much our fitted model is able to capture these characteristics, we have simulated 3840[1] networks corresponding to three randomly chosen days (7th December 2012, 18th April 2014 and 10th July 2015), by using the network model with the fitted parameters of the corresponding day. We then compared the simulated three characteristics on these three days to the ones observed in the real-world networks (this way of assessing the goodness of fit is also used in Hunter et al. (2008)). The heuristic justification underlying this approach is, that, if considered jointly, these three characteristics are able to discriminate between a range of different types of networks (see also Jackson (2010); Zafarani et al. (2014))

In our analysis, we consider sub-networks defined by the popularity of their edges: For given values $0 \leq l_1 < l_2 \leq \infty$, the network is constructed by placing an edge between a pair of nodes $i = (v_1, v_2)$, if the number of tours between v_1 and v_2 falls between l_1 and l_2. Different ranges of l_1 and l_2 are considered. The idea is to consider the network of low frequented tours (for $l_1 = 1$ and $l_2 = 3$) up to the network of highly frequented tours (for $l_1 = 10$ and $l_2 = \infty$).

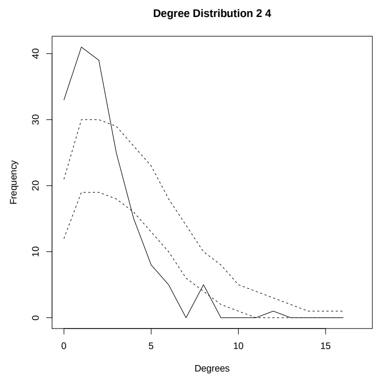
We begin with presenting the results for the degree distribution on 7th December 2012. Figure 4.4 shows the simulated degree distributions for six different choices of l_1 and l_2. The dotted lines indicate 10% and 90% quantiles of the simulated graphs, and the solid line shows the true degree distribution. We see that, in all six cases, the approximation behaves reasonably accurate, in particular if one takes into account that we did not specifically aim at reproducing the degree distributions. The plots show that the largest degree of the simulated networks and the observed network lie not too far from each other, and the overall shape of the degree distribution is captured well. It should also be noted that we used only six covariates, whereas in other related empirical work much higher dimensional models have been used, see e.g. the discussions in Perry and Wolfe (2013).

We present next the results for diameter and clustering coefficient on 7th of December 2012. Figure 4.5 shows the histograms of the simulated diameter in the different regimes. We see that, in 4.5e (as before in Figure 4.4e), the simulation and the reality appear to coincide nicely. In other words, for a moderate number of tours our model seems to fit well. It is interesting to note that our model performs differently in the different regimes suggesting that edges with different activity have to be modelled differently. Finally, in Figure 4.6, we see the histograms of the simulated clustering coefficients. The true value

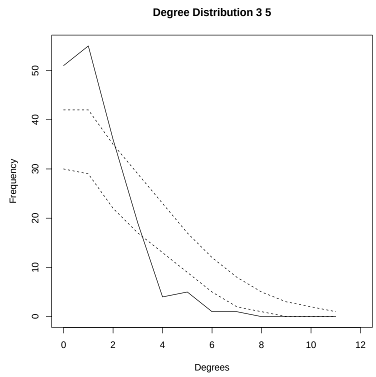
[1] We chose to simulate 3840 networks, because we had 32 cores available, and on each of the cores we ran 120 predictions, which could be done in reasonable time.



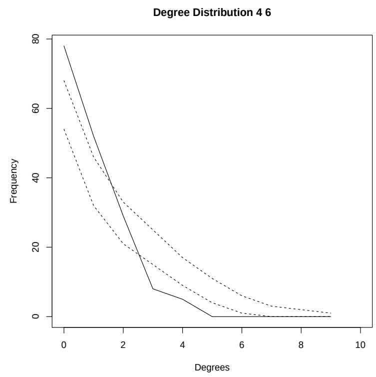
(a) Only edges with tour frequency between one and three



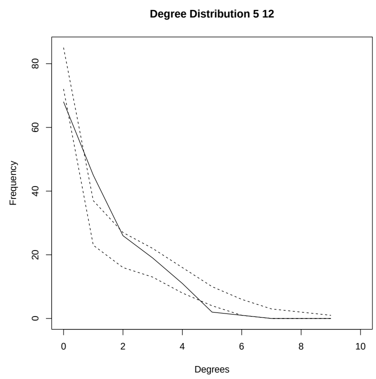
(b) Only edges with tour frequency between two and four



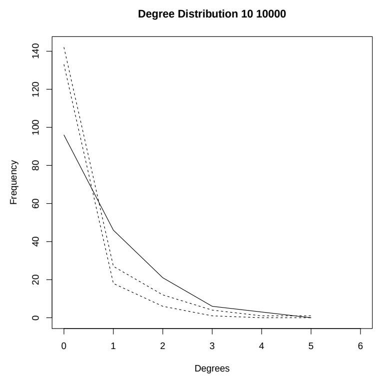
(c) Only edges with tour frequency between three and five



(d) Only edges with tour frequency between four and six



(e) Only edges with tour frequency between five and twelve



(f) Only edges with tour frequency larger than ten

Figure 4.4: Simulated degree distributions of sub-networks with different tour frequencies (see individual caption) for 7th December 2012. Dotted lines show 10% and 90% quantiles of simulations and solid line shows true distributions.

in the corresponding regime is shown in the titles of the plots. Overall, the performance appears reasonable. In particular, in Figure 4.6d the histogram is nicely centred around the true value. Interestingly, the performance in the fifth regime ($l_1 = 5$ and $l_2 = 12$), shown in Figure 4.6e, is not as good as the others. One explanation for this might be that different covariates are needed here.

In Figure 4.13a we see one simulated network compared to the true graph. The color of the edges determine how many tours happened relative the the other edges: The lowest 25% of the edges are coloured green, the next 25% yellow, then orange and the highest 25% of edges are coloured red. Due to the integral value of the activity it is not the case that exactly 25% of the edges are green and so on. The size of the vertices is relative to their degree. We see that the model is able to find the *important* (i.e. high degree) vertices. For the edges we see that some red edges are at wrong places. But generally the vertices with high profile edges are recognized. The remaining graphs in Figure 4.13 show the same comparison for the two other dates (18th April 2014 and 10th July 2015) under consideration. And we see that the results are similar.

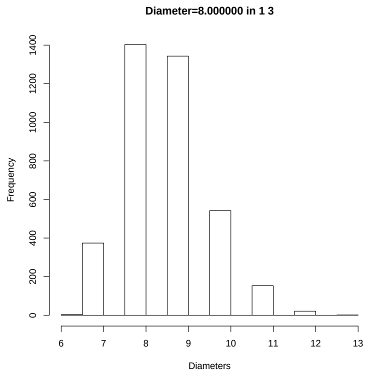
Figures 4.7 till 4.12 show the results of the corresponding simulations for the other two dates (18th April 2014 and 10th July 2015). Overall the results are similar. It should be pointed out that even though the model is not able to reproduce every feature perfectly accurate, the simulated network features are still visually appearing close to the true observation. This becomes more obvious if we remind ourselves that only six covariates were used.

4.5.4 Bandwidth Choice

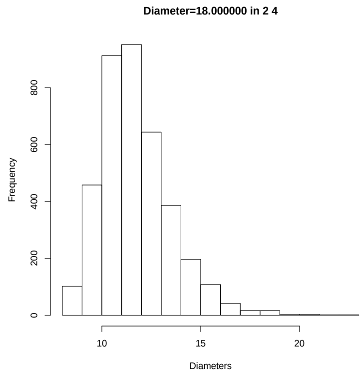
We give firstly a quick overview of our method for bandwidth selection before we back it up with a more detailed, however heuristic explanation. Under our assumptions that the covariates stay constant over the day, it makes sense to consider only integral bandwidths (here one day has length one). In order to choose the bandwidth, we apply a *one-sided cross validation* (OSCV) (cf. Hart and Yi (1998); Mammen et al. (2011)) approach: To choose the bandwidth, we calculate a local linear estimate of the parameter function with a one-sided kernel $K_{+,\kappa}(k) = K_\kappa(k)\mathbb{1}(k < 0)$. For all values of κ, we use this estimate to predict the number of bike tours on all edges. Note that we may really talk about a prediction because by using a one-sided kernel, we only take past observations into account to compute the estimate. Finally this estimate is compared with the observation by computing the empirical mean squared error. This is done for all non-censored edges. The results for different bandwidths are shown in Figure 4.14. The prediction error of the model decreases, until we reach the bandwidth $\kappa = 23$. In OSCV one now makes use of the fact that the ratio of asymptotically optimal bandwidths of two kernel estimators with different kernels, K and L is equal to

$$\rho := \left[\frac{\int K^2(u)\mathrm{d}u \left(\int u^2 L(u)\mathrm{d}u\right)^2}{\int L^2(u)\mathrm{d}u \left(\int u^2 K(u)\mathrm{d}u\right)^2}\right]^{1/5}.$$

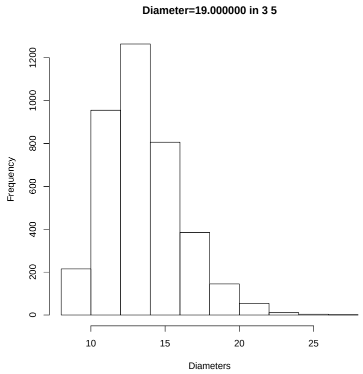
For a triangular kernel, and its one-sided version, we get $\rho \approx 1.82$. The OSCV bandwidth is given by dividing 23 by ρ which yields bandwidth roughly twelve (here we also consider



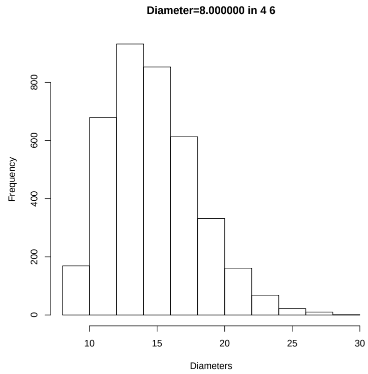
(a) Only edges with tour frequency between one and three



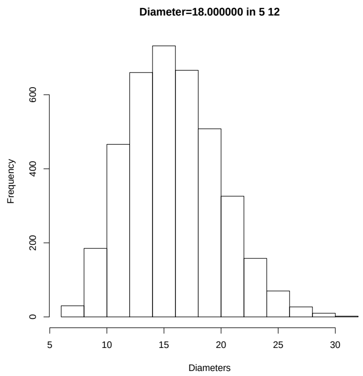
(b) Only edges with tour frequency between two and four



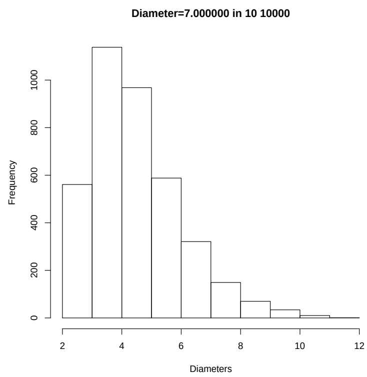
(c) Only edges with tour frequency between three and five



(d) Only edges with tour frequency between four and six

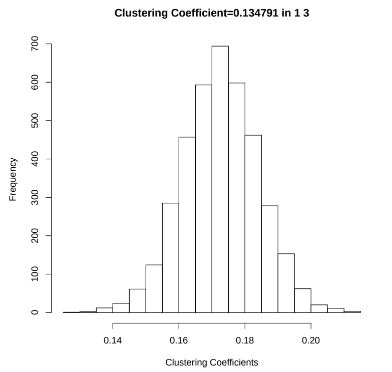


(e) Only edges with tour frequency between five and twelve

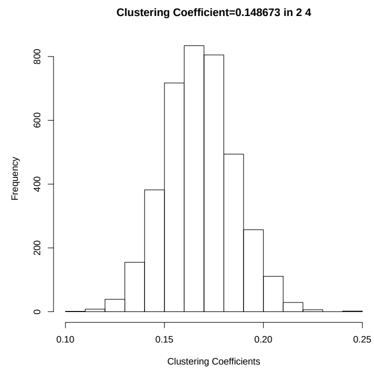


(f) Only edges with tour frequency larger than ten

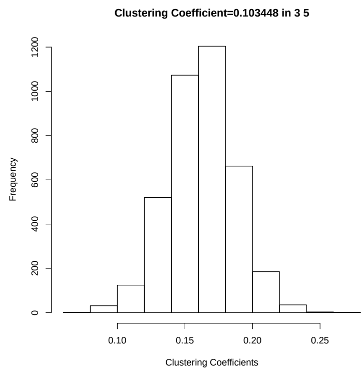
Figure 4.5: Histograms of diameters of the graphs which arise by taking different edges into account (see individual caption) from simulations for 7th December 2012. In the title of the plot the observed value is shown.



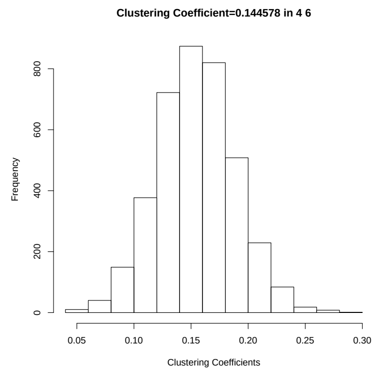
(a) Only edges with tour frequency between one and three



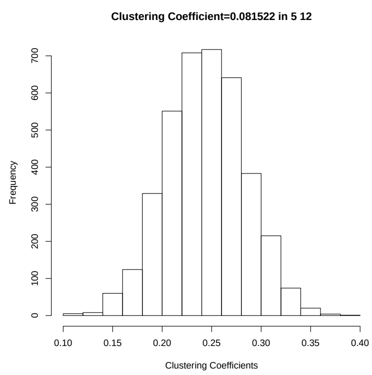
(b) Only edges with tour frequency between two and four



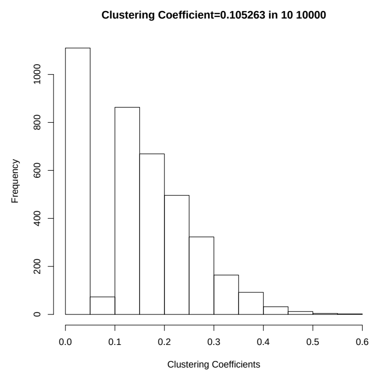
(c) Only edges with tour frequency between three and five



(d) Only edges with tour frequency between four and six

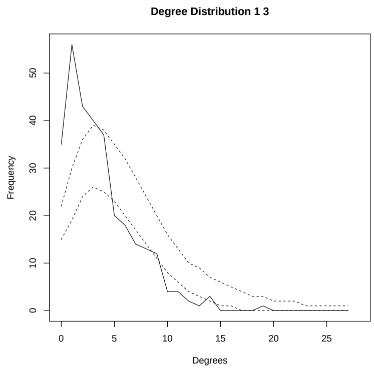


(e) Only edges with tour frequency between five and twelve

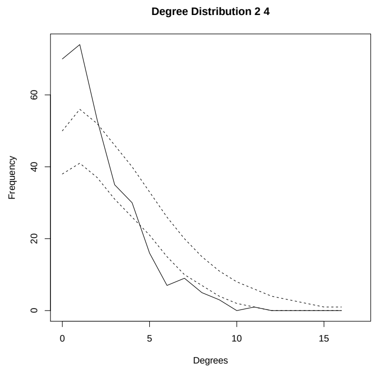


(f) Only edges with tour frequency larger than ten

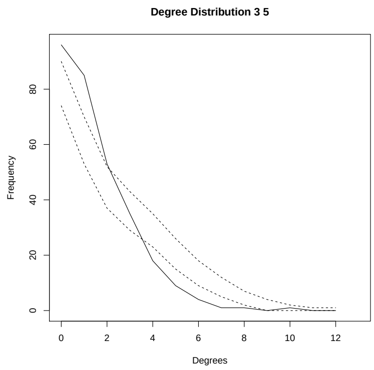
Figure 4.6: Histograms of clustering coefficients of the graphs which arise by taking different edges into account (see individual caption) from simulations for 7th December 2012. In the title of the plot the observed value is shown.



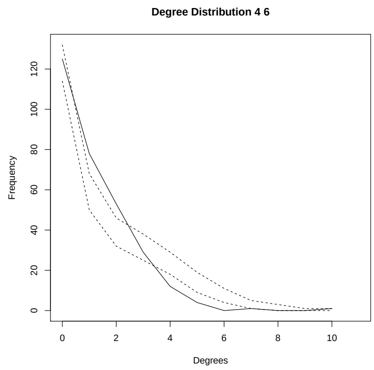
(a) Only edges with tour frequency between one and three



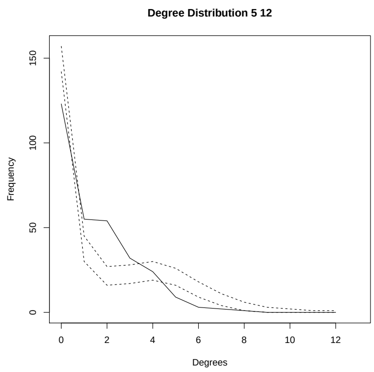
(b) Only edges with tour frequency between two and four



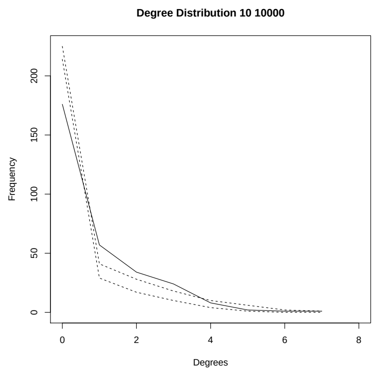
(c) Only edges with tour frequency between three and five



(d) Only edges with tour frequency between four and six

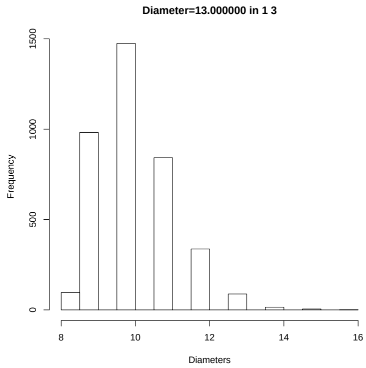


(e) Only edges with tour frequency between five and twelve

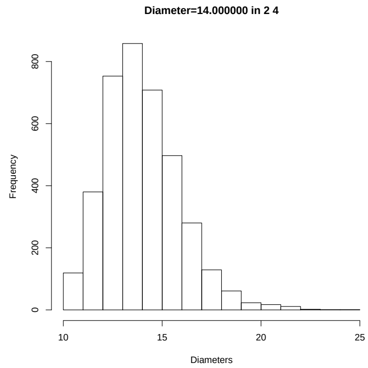


(f) Only edges with tour frequency between larger than ten

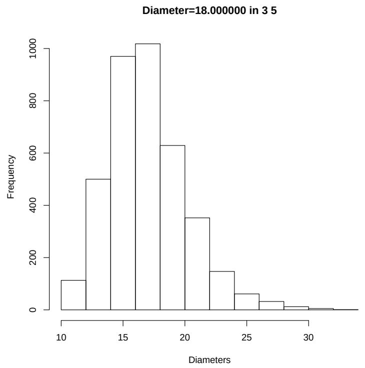
Figure 4.7: Degree distributions of the graphs which arise by taking different edges into account (see individual caption) from simulations for 18th April 2014. Dotted lines show 10% and 90% quantiles of simulations and solid line shows true distributions.



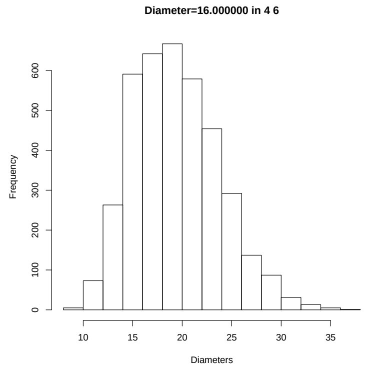
(a) Only edges with tour frequency between one to three



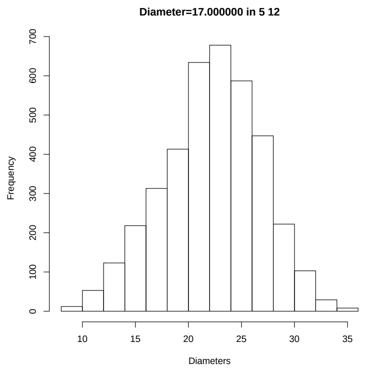
(b) Only edges with tour frequency between two to four



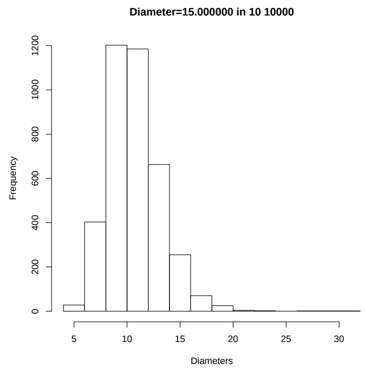
(c) Only edges with tour frequency between three to five



(d) Only edges with tour frequency between four to six

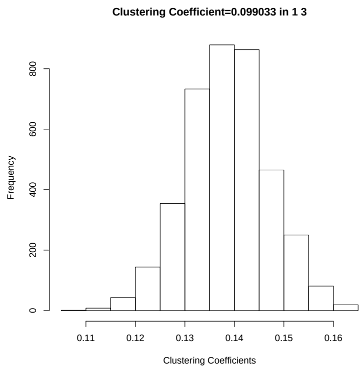


(e) Only edges with tour frequency between five to twelve

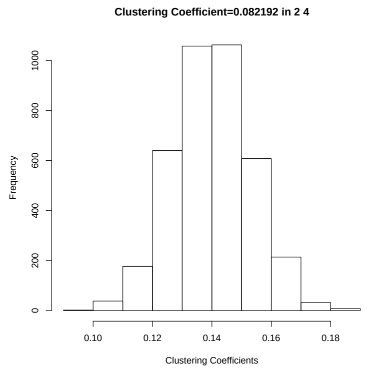


(f) Only edges with tour frequency more than ten

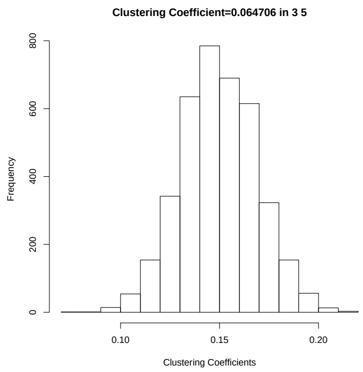
Figure 4.8: Histograms of diameters of the graphs which arise by taking different edges into account (see individual caption) from simulations for 18th April 2014. In the title of the plot the observed value is shown.



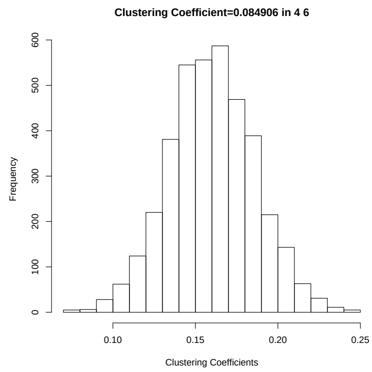
(a) Only edges with tour frequency between one to three



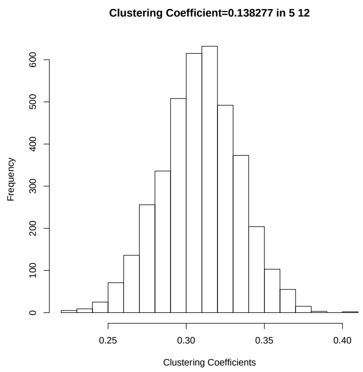
(b) Only edges with tour frequency between two to four



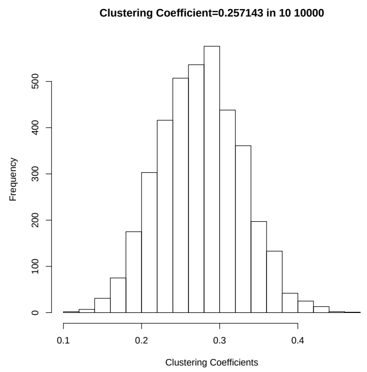
(c) Only edges with tour frequency between three to five



(d) Only edges with tour frequency between four to six

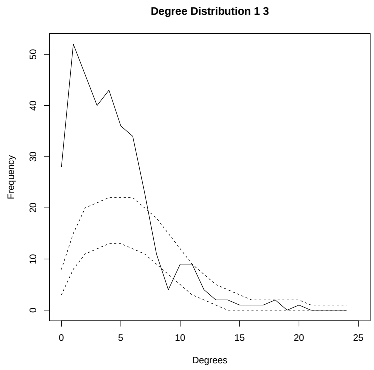


(e) Only edges with tour frequency between five to twelve

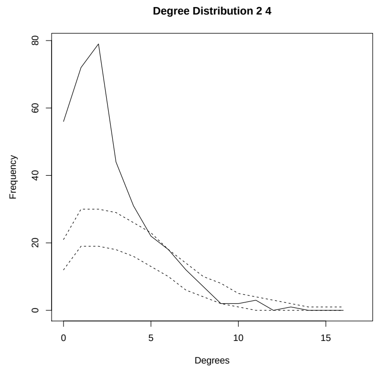


(f) Only edges with tour frequency more than ten

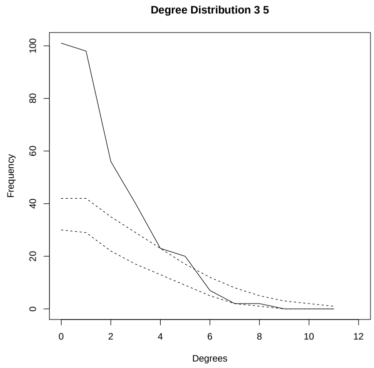
Figure 4.9: Histograms of clustering coefficients of the graphs which arise by taking different edges into account (see individual caption) from simulations for 18th April 2014. In the title of the plot the observed value is shown.



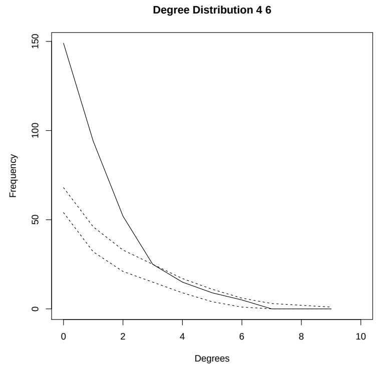
(a) Only edges with tour frequency between one to three



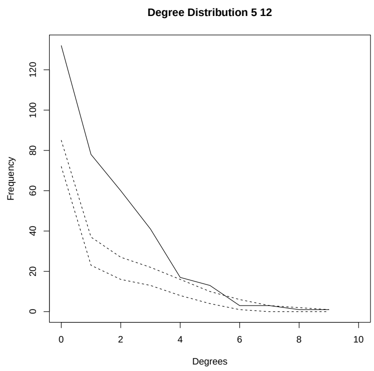
(b) Only edges with tour frequency between two to four



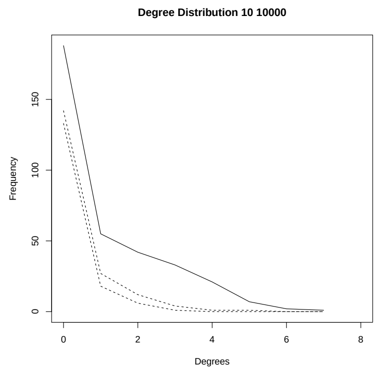
(c) Only edges with tour frequency between three to five



(d) Only edges with tour frequency between four to six

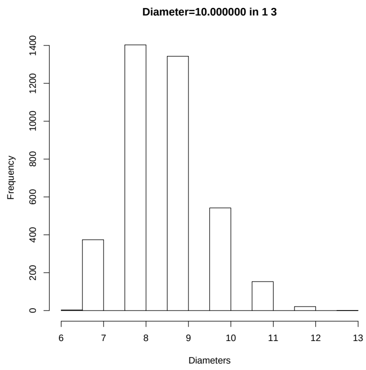


(e) Only edges with tour frequency between five to twelve

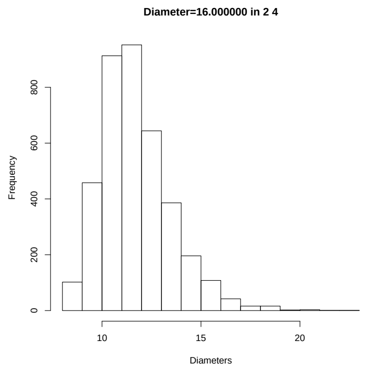


(f) Only edges with tour frequency more than ten

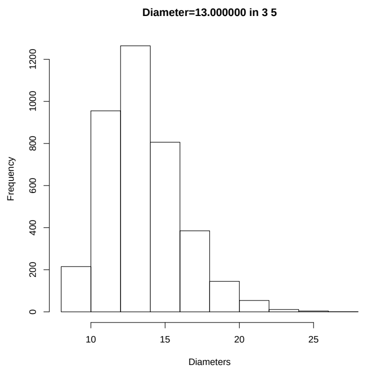
Figure 4.10: Degree distributions of the graphs, which arise by taking different tour frequencies into account (see individual caption) from simulations for 10th July 2015. Dotted lines show 10% and 90% quantiles of simulations and solid line shows true distributions.



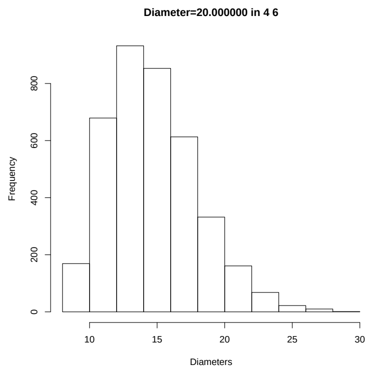
(a) Only edges with tour frequency between one to three



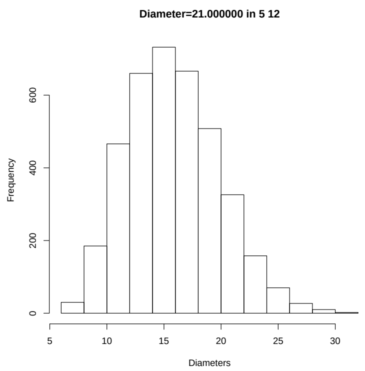
(b) Only edges with tour frequency between two to four



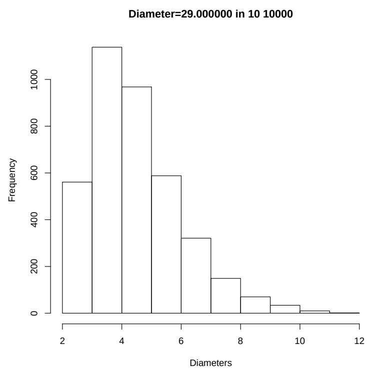
(c) Only edges with tour frequency between three to five



(d) Only edges with tour frequency between four to six

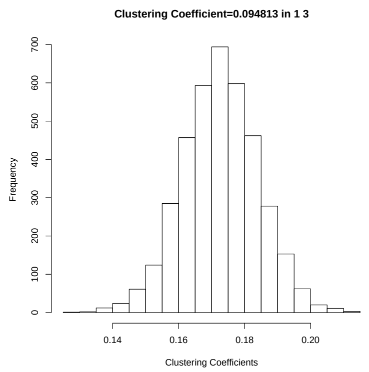


(e) Only edges with tour frequency between five to twelve

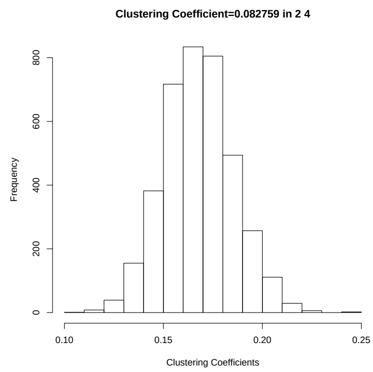


(f) Only edges with tour frequency more than ten

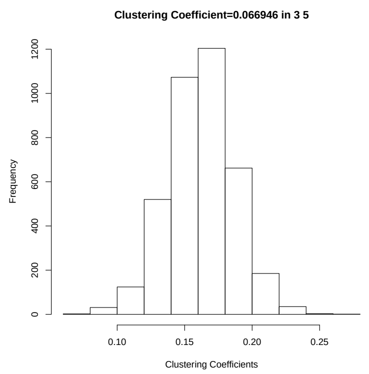
Figure 4.11: Histograms of diameters of the graphs which arise by taking different edges into account (see individual caption) from simulations for 10th July 2015. In the title of the plot the observed value is shown.



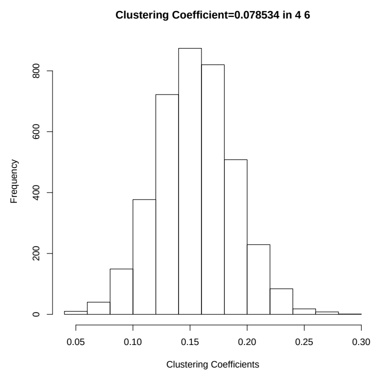
(a) Only edges with tour frequency between one to three



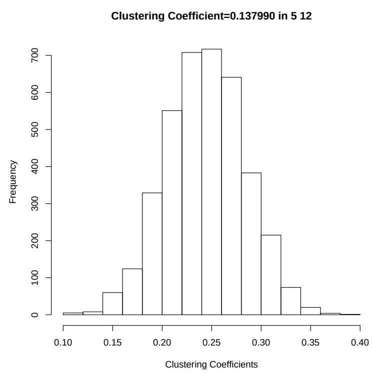
(b) Only edges with tour frequency between two to four



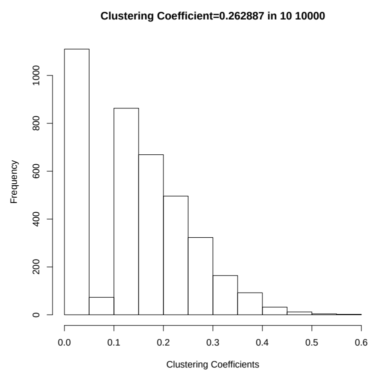
(c) Only edges with tour frequency between three to five



(d) Only edges with tour frequency between four to six



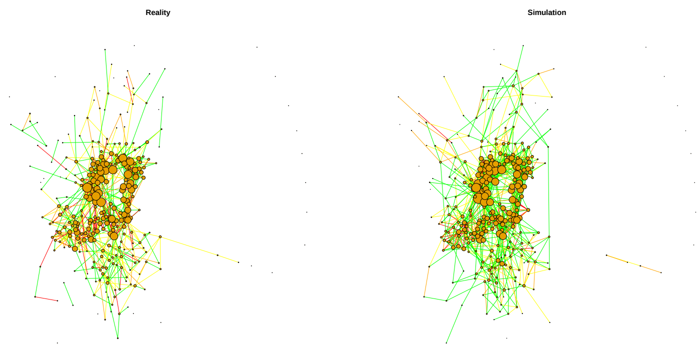
(e) Only edges with tour frequency between five to twelve



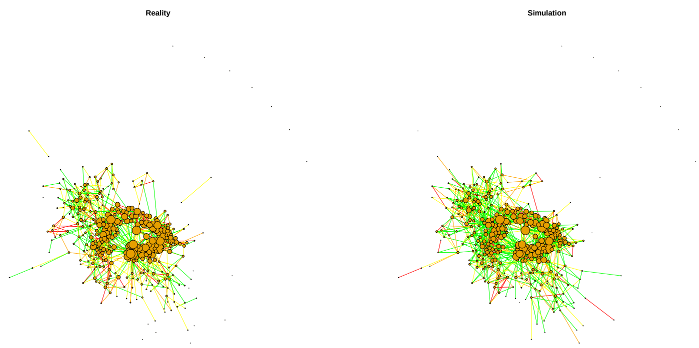
(f) Only edges with tour frequency more than ten

Figure 4.12: Histograms of clustering coefficients of the graphs which arise by taking different edges into account (see individual caption) from simulations for 10th July 2015. In the title of the plot the observed value is shown.

(a) 12th December 2012



(b) 18th April 2014



(c) 10th July 2015

Figure 4.13: Compares one simulated graph with the true observation.

integral bandwidths only).

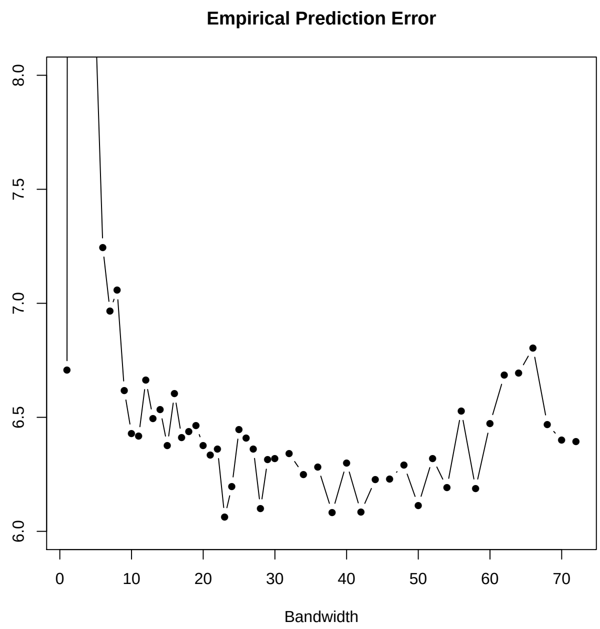


Figure 4.14: Mean Squared Prediction Error for different bandwidths.

We describe the approach described in the first paragraph now in more detail. Let K and L be two kernels fulfilling the assumptions in Section 4.4 and denote by $\hat{\theta}_K(t_0)$ and $\hat{\theta}_L(t_0)$ the maximum likelihood estimators using K and L respectively. Then, by Theorem 4.2, we get that asymptotically the bias and the variance of the estimators can be written as

$$\begin{aligned} bias(\hat{\theta}_K) &= h^2 \int_{-1}^{1} K(u) u^2 du \cdot C_1 \\ var(\hat{\theta}_L) &= \frac{1}{l_n h} \int_{-1}^{1} K(u)^2 du \cdot C_2 \end{aligned}$$

where the constants C_1 and C_2 depend on the true parameter curve θ_0 and the time t_0 but not on the kernel. Hence, the corresponding expressions for $\hat{\theta}_L(t_0)$ can be found, just by replacing every K with an L. The decomposition of the asymptotic mean squared error in squared bias plus variance yields the following asymptotically optimal bandwidths h_K and h_L , minimizing the asymptotic mean squared error:

$$h_K := \left(\frac{1}{l_n} \cdot \frac{\int_{-1}^{1} K(u)^2 du}{\left[\int_{-1}^{1} K(u) u^2 du \right]^2} \cdot \frac{C_1}{4 C_2} \right)^{\frac{1}{5}}.$$

Again, the corresponding expression for h_L can be found by replacing every K by L. So the following formula, known from kernel estimation, holds also true in our setting

$$h_K = \left(\frac{\int_{-1}^{1} K(u)^2 du}{\left[\int_{-1}^{1} K(u)u^2 du\right]^2} \cdot \frac{\left[\int_{-1}^{1} L(u)u^2 du\right]^2}{\int_{-1}^{1} L(u)^2 du} \right)^{\frac{1}{5}} h_L = \rho \cdot h_L. \tag{4.40}$$

This means that knowledge of the bandwidth which minimizes the mean squared error for kernel L, implies knowledge of the bandwidth which minimizes the mean squared error using kernel K. Ultimately, we want to use a triangular kernel $K(u) = (1+u)\mathbb{1}_{[-1,0)}(u) + (1-u)\mathbb{1}_{[0,1]}(u)$. In order to find the bandwidth h_K for this kernel, we want to apply cross-validation. As proposed in Hart and Yi (1998) one-sided cross validation is an attractive method for the case of time series data. Here, one-sided means that we apply cross validation to a kernel L which is only supported on the past $[-1, 0]$. In order to avoid a bias, we use the one-sided kernel together with local linear approximation. The following heuristic derivation motivates this choice.

Firstly, in our regular maximum likelihood setting, we maximize, over $\mu \in \Theta$, the expression (we use a finite (and thus approximate) Taylor Series expansion)

$$\begin{aligned}
&\sum_{0<t\leq T} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \sum_{i\in G_n} \Delta N_{n,i}(t) \mu^T X_{n,i}(t) \\
&\qquad - \int_0^T \sum_{i\in G_n} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) C_{n,i}(t) e^{\mu^T X_{n,i}(t)} dt \\
\approx\ &\sum_{0<t\leq T} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \sum_{i\in G_n} \Delta N_{n,i}(t) \mu^T X_{n,i}(t) \\
&- \int_0^T \sum_{i\in G_n} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) C_{n,i}(t) e^{\theta_0(t_0)^T X_{n,i}(t)} \\
&\qquad \times \left(1 + (\mu - \theta_0(t_0))^T X_{n,i}(t) + \frac{1}{2}\left[(\mu - \theta_0(t_0))^T X_{n,i}(t)\right]^2\right) dt.
\end{aligned}$$

Deriving this expression with respect to μ, setting the derivative equal to zero, and rearranging terms, yields (to save space we use here a fraction, although the denominator is a matrix)

$$\begin{aligned}
&\hat{\theta}_K(t_0) - \theta_0(t_0) \\
\approx\ &\frac{\sum_{i\in G_n} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \Delta N_{n,i}(t) X_{n,i}(t) - \int_0^T \frac{1}{h} K\left(\frac{t-t_0}{h}\right) C_{n,i}(t) X_{n,i} e^{\theta_0(t_0)^T X_{n,i}(t)} dt}{\sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{t-t_0}{h}\right) X_{n,i}(t) X_{n,i}(t)^T e^{\theta_0(t_0)^T X_{n,i}(t)} dt}.
\end{aligned}$$

Using the notation $y_1 := \mathbb{E}(X_{n,i}(t_0) e^{\theta_0(t_0)^T X_{n,i}(t_0)} | C_{n,i}(t_0) = 1) \cdot \mathbb{P}(C_{n,i}(t_0) = 1)$ and $y_2 := \mathbb{E}(X_{n,i}(t_0) X_{n,i}(t_0)^T e^{\theta_0(t_0)^T X_{n,i}(t_0)} | C_{n,i}(t_0) = 1) \cdot \mathbb{P}(C_{n,i}(t_0) = 1)$, we obtain the approximation

$$\hat{\theta}_K(t_0) - \theta_0(t_0) \approx \frac{\sum_{i\in G_n} \sum_{0<t\leq T} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \Delta N_{n,i}(t) X_{n,i}(t) - y_1}{y_2}. \tag{4.41}$$

Now define the local linear estimator $\hat{\theta}_{LC,K}(t_0)$, with respect to a kernel K, as the value of μ_0 maximizing the following expression over $(\mu_0, \mu_1) \in \Theta^2$:

$$\sum_{0<t\leq T} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \sum_{i\in G_n} \Delta N_{n,i}(t)[\mu_0 + \mu_1(t-t_0)]^T X_{n,i}(t) \\ - \int_0^T \sum_{i\in G_n} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) e^{[\mu_0+\mu_1(t-t_0)]^T X_{n,i}(t)} dt.$$

Using the same approximations as in the usual kernel estimation setting, and deriving the resulting approximate likelihood, we obtain

$$\hat{\theta}_{LC,K} - \theta_0(t_0) \\ \approx \frac{\sum_{i\in G_n} \sum_{0<t\leq T} \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \frac{M_2 - \frac{t-t_0}{h} M_1}{M_2 - M_1^2} \Delta N_{n,i}(t) X_{n,i}(t) - y_1}{y_2},$$

where $M_k := \int_{-1}^1 u^k K(u)du$. The previous computations were just a heuristic. But nevertheless, the similarity between the previous display and (4.41) suggests that the local linear estimator $\hat{\theta}_{LC,K}$ using the kernel K is actually just a regular kernel estimator $\hat{\theta}_L$ with kernel function

$$L(u) = K(u) \frac{M_2 - uM_1}{M_2 - M_1^2}. \tag{4.42}$$

This aligns with results about kernel estimation, as, for example, stated in Mammen et al. (2011). It can be easily computed that this new kernel is of order one, i.e., $\int uL(u)du = 0$, even though the original kernel was not. Hence, knowledge of the optimal bandwidth for the local linear estimator using the kernel K implies knowledge of the optimal bandwidth for any other order one kernel by means of (4.40). Taking the same route as in Hart and Yi (1998), the selector for the bandwidth $\hat{h}_K$ for the triangular kernel K is the following: Let $K^*(u) := 2K(u)\mathbb{1}_{[-1,0]}(u)$ denote the one sided version of K.

1. Find a bandwidth $\hat{h}_L$ for the local linear estimator $\hat{\theta}_{LC,K^*}$ based on the kernel K^* via cross validation (since we use a one-sided kernel, this step is also called one-sided cross-validation. We will make it more precise later).

2. Compute $\hat{h}$ by using (4.40) with L defined as in (4.42) but with K replaced by K^*.

For the one-sided cross-validation in step 1, we minimize, in our bike share data analysis, the following function in h

$$\frac{1}{k_T} \sum_{k=0}^{k_T} \frac{1}{|L(k)|} \sum_{i\in L(k)} \frac{\left| e^{\hat{\theta}^{(-k)}_{LC,K^*}(k)^T X_i(k)} - Z_i(k) \right|^2}{e^{\hat{\theta}^{(-k)}_{LC,K^*}(k)^T X_i(k)}}, \tag{4.43}$$

where k_T was the number of weeks (recall that we assume the covariates to remain constant over a day, and that we only consider Fridays), i.e., k refers to the k-th Friday in the dataset. $L(k)$ refers to the set of edges i on which there was a bike tour on Friday k, $Z_i(k)$ is the true number of bike tours observed on i on Friday k. Finally, $\hat{\theta}^{(-k)}_{LC,K^*}(k)$ is the local linear estimator with respect to the kernel K^* based on all but the k-th Friday. Since K^* is left-sided, this really means the estimator is based on Fridays $0, ..., k-1$, and hence the term one-sided cross-validation. The intensities are the theoretical values of the expectation of the number of bike tours if the model is correct. So we compute the squared difference with the true number of bike rides and divide by the estimated intensity, where we only take the non-censored edges into account.

In Figure 4.14, we had displayed results for different bandwidths h of (4.43). The prediction error of the fit decreases, until the bandwidth is equal to 23. Afterwards the prediction error stays roughly the same and starts to increase when the bandwidth reaches a full year (52 weeks). This may be explained by a periodicity with a period of approximately one year. If one uses 23 as minimal value we get as asymptotic optimal bandwidth 23 divided by ρ which is approximately 12. Here, following Step 2 of the above described procedure, we use that ρ is approximately equal to 1.82 for triangular kernels.

5 Proofs

Recall the following notation from the previous chapters. $l_n = r_n \mathbb{P}(C_{n,1}(t_0) = 1)$, where $r_n = |G_n|$ is the number of possible edges. Here we always assume either a directed or an undirected complete graph, i.e., $r_n = n(n-1)$ (directed case) or $r_n = \frac{n(n-1)}{2}$ (undirected case). The processes $N_{n,i}$ are counting processes with intensity given by $C_{n,i}(t) \cdot \lambda_{n,i}(\theta_0(t), t)$. We can decompose these counting processes as (cf. Chapter 2.2)

$$N_{n,i}(t) = M_{n,i}(t) + \int_0^t C_{n,i}(s) \cdot \lambda_{n,i}(\theta_0(s), s) ds, \tag{5.1}$$

where $M_{n,i}$ is a local, square integrable martingale. We use this decomposition of the counting processes in order to decompose the likelihood and its derivatives. Let $P_n(\theta)$ be defined as

$$P_n(\theta) := \frac{1}{l_n h} \sum_{i \in G_n} \int_0^T K\left(\frac{s-t_0}{h}\right) C_{n,i}(s) \left[\log \lambda_{n,i}(\theta, s) \cdot \lambda_{n,i}(\theta_0(s), s) - \lambda_{n,i}(\theta, s)\right] ds \tag{5.2}$$

$$= \frac{1}{l_n h} \sum_{i \in G_n} \int_0^T K\left(\frac{s-t_0}{h}\right) C_{n,i}(s) \big[\theta^T X_{n,i}(s) \exp(\theta_0(s)^T X_{n,i}(s)) - \exp(\theta^T X_{n,i}(s))\big] ds. \tag{5.3}$$

Note that we do not make the dependence of $P_n(\theta)$ on t_0 explicit in the notation. Using $P_n(\theta)$, we can write

$$\frac{1}{l_n}\ell(\theta, t_0) = \frac{1}{l_n h} \sum_{i \in G_n} \int_0^T K\left(\frac{t-t_0}{h}\right) \theta^T X_{n,i}(t) dM_{n,i}(t) \ + \ P_n(\theta), \tag{5.4}$$

$$\frac{1}{l_n} \cdot \frac{\partial}{\partial \theta}\ell(\theta, t_0) = \frac{1}{l_n h} \sum_{i \in G_n} \int_0^T K\left(\frac{t-t_0}{h}\right) X_{n,i}(t) dM_{n,i}(t) + P_n'(\theta), \tag{5.5}$$

$$\frac{1}{l_n} \cdot \frac{\partial^2}{\partial \theta^2}\ell(\theta, t_0) = P_n''(\theta). \tag{5.6}$$

5.1 Proof of Theorem 4.2

Recall that $\theta_{0,n}$ is defined as the maximizer of $\theta \mapsto \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) g(\theta, s) ds$, where g is defined in (A7). Note that the function g does not depend on n, see Assumption (A1). Lemma 5.3 shows that $\theta_{0,n}$ is uniquely defined. The value $\theta_{0,n}$ is the deterministic counterpart of the random quantity $\widetilde{\theta}_n(t_0)$ that is defined as the solution of $P_n'(\widetilde{\theta}_n(t_0)) = 0$.

The existence of the latter is considered in Proposition 5.8.

Lemma 5.1. *We have*

$$\begin{aligned} & \theta^T y \exp(\theta_0(s)^T y) - \exp(\theta^T y) \\ \leq \quad & \theta_0(s)^T y \exp(\theta_0(s)^T y) - \exp(\theta_0(s)^T y). \end{aligned}$$

Equality holds, if and only if, $\theta_0(s)^T y = \theta^T y$*. In particular,* $\theta_0(s)$ *is the unique maximizer of* $\theta \mapsto g(\theta, s)$*.*

Proof. Note that, for arbitrary $y \in \mathbb{R}$,

$$\frac{d}{dx}(xe^y - e^x) = e^y - e^x$$

implies that the differentiable function $x \mapsto xe^y - e^x$ has the unique maximizer $x = y$. This also implies the second statement of the lemma. □

In all lemmas and propositions of this section, we assume that Assumptions (A1)–(A7) hold.

Fact 5.2. *For* $j \in \{0, 1, 2\}, k \in \{0, 1, 2, 3\}$*, with* $j + k \leq 3$*, the partial derivatives of order* j *of the function* $g(\theta, s)$ *with respect to* s*, and of order* k *with respect to* θ*, exist, for* t *in a neighbourhood of* t_0*, and* $\theta \in \Theta$*. The partial derivatives can be calculated by interchanging the order of integration and differentiation in (4.15). For* $\theta \in \Theta$ *and* s *in a neighbourhood of* t_0*, all these partial derivatives of* $g(\theta, s)$ *are absolutely bounded. For the calculation of the first two derivatives of* g *with respect to* θ*, differentiation and application of the expectation operator can be interchanged in (4.14). The matrix* Σ *is invertible.*

Proof. The statement of this fact follows immediately from (4.4) of Condition (A4). Note that the functions θ_0, θ_0' and θ_0'' are absolutely bounded in a neighbourhood of t_0. This holds because these functions are continuous in a neighbourhood of t_0, see (A3). □

Lemma 5.3. *For* n *large enough,* $\theta_{0,n}$ *is uniquely defined, and it holds that* $\theta_{0,n} \to \theta_0(t_0)$ *as* $n \to \infty$*.*

Proof of Lemma 5.3. From Fact 5.2, we know that $\partial_t g(\theta, t)$ is absolutely bounded for t in a neighbourhood of t_0 and $\theta \in \Theta$. This implies that

$$\theta \mapsto \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h}\right) g(\theta, s) ds$$

converges to $g(\theta, t_0)$, uniformly for $\theta \in \Theta$. Because $\partial_{\theta^2} g(\theta, t)$ is negative definite, this implies the statement of the lemma. □

Lemma 5.4. *With* $\Sigma_n = -\int_{-1}^{1} K(u) \int_0^1 \partial_{\theta^2} g(\theta_0(t_0) + \alpha(\theta_{0,n} - \theta_0(t_0)), t_0 + uh) \mathrm{d}\alpha \mathrm{d}u$*, we have*

$$\Sigma_n \to \Sigma \text{ as } n \to \infty.$$

Moreover, the sequence

$$v_n = 2 \int_{-1}^{1} K(u) \int_0^1 (1-\alpha) \frac{d^2}{dt^2} \partial_\theta g(\theta_0(t_0), t_0 + (1-\alpha)uh) u^2 \mathrm{d}\alpha \mathrm{d}u$$

is bounded, and it holds that $v_n \to v$*, as* $n \to \infty$*.*

Proof. Using Lemmas 5.3 and Fact 5.2, we conclude that the integrand

$$\partial_{\theta^2} g(\theta_0(t_0) + \alpha(\theta_{0,n} - \theta_0(t_0)), t_0 + uh) \to \partial_{\theta^2} g(\theta_0(t_0), t_0)$$

(note that $u \in [-1, 1]$ and $\alpha \in [0, 1]$). The first statement of the lemma follows by an application of Lebesgue's Dominated Convergence Theorem, and the fact that $\partial_{\theta^2} g$ is bounded as a continuous function on a compact set. The second statement of the lemma follows similarly. □

Proposition 5.5. *We have, for* $t_0 \in (0, T)$*,*

$$\theta_{0,n} = \theta_0(t_0) + h^2 \Sigma^{-1} v + o(h^2).$$

Proof of Proposition 5.5. Since $\theta_0(s)$ maximizes $\theta \mapsto g(\theta, s)$ (cf. Lemma 5.1), we have $\partial_\theta g(\theta_0(s), s) = 0$. Furthermore, by definition of $\theta_{0,n}$, we have

$$\int_0^T K\left(\frac{s - t_0}{h}\right) \partial_\theta g(\theta_{0,n}, s) ds = 0.$$

Having observed that, we compute, for h small enough,

$$\begin{aligned}
0 &= \frac{1}{h} \int_0^T K\left(\frac{s - t_0}{h}\right) \partial_\theta g(\theta_{0,n}, s) ds \\
&= \int_{-1}^{1} K(u) \partial_\theta g(\theta_{0,n}, t_0 + uh) \mathrm{d}u \\
&= \int_{-1}^{1} K(u) \Big[\partial_\theta g(\theta_0(t_0), t_0 + uh) \\
&\qquad + \int_0^1 \partial_{\theta^2} g(\theta_0(t_0) + \alpha(\theta_{0,n} - \theta_0(t_0)), t_0 + uh) \mathrm{d}\alpha (\theta_{0,n} - \theta_0(t_0)) \Big] \mathrm{d}u \\
&= \int_{-1}^{1} K(u) \partial_\theta g(\theta_0(t_0), t_0 + uh) \mathrm{d}u + \Sigma_n(\theta_{0,n} - \theta_0(t_0)). \qquad (5.7)
\end{aligned}$$

Σ_n converges to the invertible matrix Σ by Lemma 5.4. The first integral is of order h^2. This follows by a Taylor expansion in the time parameter:

$$\int_{-1}^{1} K(u) \partial_\theta g(\theta_0(t_0), t_0 + uh) \mathrm{d}u$$

$$\begin{aligned}
&= \int_{-1}^{1} K(u) \Big[\partial_\theta g(\theta_0(t_0), t_0) + \frac{d}{dt} g_\theta(\theta_0(t_0), t_0) uh + \\
&\qquad \int_0^1 (1-\alpha) \frac{\mathrm{d}^2}{\mathrm{d}t^2} \partial_\theta g(\theta_0(t_0), t_0 + (1-\alpha)uh) \mathrm{d}\alpha u^2 h^2 \Big] \mathrm{d}u \\
&= \frac{1}{2} h^2 v_n.
\end{aligned}$$

By Lemma 5.4, v_n is bounded. Thus, together with (5.7), we have established

$$\theta_{0,n} = \theta_0(t_0) - (\Sigma_n^{-1} - \Sigma^{-1} + \Sigma^{-1}) \tfrac{1}{2} h^2 v_n = \theta_0(t_0) - \tfrac{1}{2} h^2 \Sigma^{-1} v_n - \tfrac{1}{2} h^2 (\Sigma_n^{-1} - \Sigma^{-1}) v_n.$$

The statement of the proposition now follows from $v_n \to v$. □

Lemma 5.6. *We have*

$$P_n'(\theta_{0,n}) \xrightarrow{\mathbb{P}} 0. \tag{5.8}$$

For any $k, l \in \{1, ..., q\}$, it holds that

$$P_n''(\theta_{0,n}) \xrightarrow{\mathbb{P}} -\Sigma. \tag{5.9}$$

Moreover,

$$\sup_{k,l,r,\theta} \left| \frac{\partial^2}{\partial\theta_k \partial\theta_l} P_n'^{(r)}(\theta) \right| = O_P(1), \tag{5.10}$$

where $P_n'^{(r)}$ denotes the r-th component of P_n', the supremum runs over $k, l, r \in \{1, ..., q\}$, and $\theta \in \Theta$.

Proof. We start by showing that $P_n'(\theta_{0,n}) = o_P(1)$. This holds, if $\mathbb{E}(\|P_n'(\theta_{0,n})\|) = o(1)$. Define $\rho_{n,i}(\theta, s) := \|X_{n,i}(s)\| \cdot \left|\exp(\theta_0(s)^T X_{n,i}(s)) - \exp(\theta^T X_{n,i}(s))\right|$. By positivity of $\rho_{n,i}(\theta, s)$, we may apply Fubini's Theorem, and thus we compute

$$\begin{aligned}
&\mathbb{E}(\|P_n'(\theta_{0,n})\|) \\
&\leq \frac{1}{l_n} \sum_{i \in G_n} \int_{-1}^{1} K(u) \mathbb{E}\left(C_{n,i}(t_0 + uh) \rho_{n,i}(\theta_{0,n}, t_0 + uh)\right) \mathrm{d}u \\
&= \frac{1}{l_n} \sum_{i \in G_n} \int_{-1}^{1} K(u) \mathbb{P}(C_{n,i}(t_0 + uh) = 1) \\
&\qquad \times \mathbb{E}\left(\rho_{n,i}(\theta_{0,n}, t_0 + uh) | C_{n,i}(t_0 + uh) = 1\right) \mathrm{d}u.
\end{aligned}$$

The expectation in the integral expression can be bounded by applying a Taylor expansion:

$$\begin{aligned}
&\mathbb{E}\left(\rho_{n,i}(\theta_{0,n}, s_u) | C_{n,i}(s_u) = 1\right) \\
&\leq \mathbb{E}\Bigg(\int_0^1 \exp\left([\theta_0(s_u) - \alpha \cdot (\theta_0(s_u) - \theta_{0,n})]^T X_{n,i}(s_u) \right) \mathrm{d}\alpha
\end{aligned}$$

$$\times \|X_{n,i}(s_u)\|^2 \bigg| C_{n,i}(s_u) = 1 \bigg) \cdot \|\theta_0(s_u) - \theta_{0,n}\|,$$

where $s_u = t_0 + uh$. Now, by (4.7) in Assumption (A4), the expectation in the last upper bound is bounded by a constant C, uniformly in $u \in [-1,1]$. Using $\sup_{u\in[-1,1]} \|\theta_0(t_0 + uh) - \theta_{n,0}\| = o(1)$, we obtain

$$\begin{aligned}
& \mathbb{E}(\|P_n'(\theta_{0,n})\|) \\
\leq\ & \frac{1}{l_n} \sum_{i\in G_n} \int_{-1}^{1} K(u)\mathbb{P}(C_{n,i}(t_0+uh)=1)\mathrm{d}u \cdot C \cdot \sup_{v\in[-1,1]} \|\theta_0(t_0+vh) - \theta_{0,n}\| \\
=\ & C \cdot \mathbb{P}(C_{n,i}(t_0)=1)^{-1} \cdot \int_0^1 K(u)\mathbb{P}(C_{n,i}(t_0+uh)=1)\mathrm{d}u \cdot o(1) \\
=\ & o(1),
\end{aligned}$$

where the last equality is a consequence of (4.8). This shows (5.8).

We now show (5.9). With $\partial_{\theta^2} g(\theta_{0,n}, s) = -\mathbb{E}(\tau_{n,i}(\theta_{0,n}, s)|C_{n,i}(s)=1)$, Fact 5.2 gives

$$\mathbb{E}(P_n''(\theta_{0,n})) = -\frac{1}{l_n h} \sum_{i\in G_n} \int_0^T K\left(\frac{s-t_0}{h}\right) \mathbb{P}(C_{n,i}(s)=1)\mathbb{E}(\tau_{n,i}(\theta_{0,n},s)|C_{n,i}(s)=1) ds.$$

For (5.9), it suffices to show:

$$P_n''(\theta_{0,n}) - \mathbb{E}(P_n''(\theta_{0,n})) = o_P(1), \tag{5.11}$$
$$\mathbb{E}(P_n''(\theta_{0,n})) + \Sigma = o(1). \tag{5.12}$$

For the proof of (5.12), we note that with $a_n(u) = \frac{\mathbb{P}(C_{n,1}(t_0+uh)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)}$,

$$\begin{aligned}
& \mathbb{E}(P_n''(\theta_{0,n})) + \Sigma \\
=\ & \int_{-1}^{1} K(u) \left[a_n(u)\partial_{\theta^2} g(\theta_{0,n}, t_0+uh) - \partial_{\theta^2} g(\theta_0(t_0), t_0)\right] \mathrm{d}u \\
=\ & \int_{1}^{1} K(u)a_n(u) \left[\partial_{\theta^2} g(\theta_{0,n}, t_0+uh) - \partial_{\theta^2}(\theta_0(t_0), t_0)\right] \mathrm{d}u \\
& \quad + \partial_{\theta^2} g(\theta_0(t_0), t_0) \int_{-1}^{1} K(u)(a_n(u)-1)\mathrm{d}u \\
=\ & o(1).
\end{aligned}$$

Here we use (4.8), and $\theta_{0,n} - \theta_0(t_0) = o(1)$ (see Proposition 5.5).

For the proof of (5.11), we write $K_{h,t_0}(s) := K\left(\frac{s-t_0}{h}\right)$ and

$$\begin{aligned}
& P_n''(\theta_{0,n}) - \mathbb{E}(P_n''(\theta_{0,n})) \\
=& \tfrac{1}{l_n h} \sum_{i\in G_n} \int_0^T K_{h,t_0}(s) \left[-C_{n,i}(s)\tau_{n,i}(\theta_{0,n}, s) + \mathbb{P}(C_{n,i}(s)=1)\partial_{\theta^2} g(\theta_{0,n}, s)\right] ds.
\end{aligned}$$

We will apply Markov's inequality to show that this term converges to zero. When squaring the above sum, we can split the resulting double sum over $i,j \in G_n$ into three parts, depending on whether $|e_i \cap e_j| = 0, 1$ or 2. Thus we have to show that the following three sequences converge to zero:

$$\mathbb{E}\left(\frac{1}{l_n^2 h^2} \sum_{i \in G_n} \bar{\kappa}_{n,i}(\theta_{0,n})^2\right) = o(1), \tag{5.13}$$

$$\mathbb{E}\left(\frac{1}{l_n^2 h^2} \sum_{\substack{i,j \in G_n \\ \text{sharing one vertex}}} \bar{\kappa}_{n,i}(\theta_{0,n}) \bar{\kappa}_{n,j}(\theta_{0,n})\right) = o(1), \tag{5.14}$$

$$\mathbb{E}\left(\frac{1}{l_n^2 h^2} \sum_{\substack{i,j \in G_n \\ \text{sharing no vertex}}} \bar{\kappa}_{n,i}(\theta_{0,n}) \bar{\kappa}_{n,j}(\theta_{0,n})\right) = o(1), \tag{5.15}$$

where $\kappa_{n,i}(\theta_{0,n}, s) := -C_{n,i}(s)\tau_{n,i}(\theta_{0,n}, s) + \mathbb{P}(C_{n,i}(s) = 1)\partial_{\theta^2} g(\theta_{0,n}, s)$, and $\bar{\kappa}_{n,i}(\theta_{0,n}) := \int_0^T K\left(\frac{s-t_0}{h}\right) \kappa_{n,i}(\theta_{0,n}, s) ds$. Now note that

$$\begin{aligned}
& \mathbb{E}\big(\bar{\kappa}_{n,i}(\theta_{0,n}) \bar{\kappa}_{n,j}(\theta_{0,n})\big) \\
= & \int_{-1}^{1} \int_{-1}^{1} K(u)K(v) \mathbb{E}\big(\kappa_{n,i}(\theta_{0,n}, t_0 + uh) \kappa_{n,j}(\theta_{0,n}, t_0 + vh)\big) \mathrm{d}u \mathrm{d}v,
\end{aligned}$$

and that the sum in (5.13) has $r_n = O(n^2)$ terms, (5.14) comprises $O(r_n^{\frac{3}{2}}) = O(n^3)$ terms, and finally (5.15) has $O(r_n^2) = O(n^4)$ terms (these orders are true for both: directed and undirected networks). Thus, it is sufficient to show that

$$\begin{aligned}
& \int_{-1}^{1} \int_{-1}^{1} K(u)K(v) \frac{\mathbb{E}\big(\kappa_{n,i}(\theta_{0,n}, t_0 + uh) \kappa_{n,j}(\theta_{0,n}, t_0 + vh)\big)}{\mathbb{P}(C_{n,1}(t_0) = 1)^2} \mathrm{d}u \mathrm{d}v \\
& \qquad = \begin{cases} o(n^2) & \text{for} \quad |e_i \cap e_j| = 2 \\ o(n) & \text{for} \quad |e_i \cap e_j| = 1 \\ o(1) & \text{for} \quad |e_i \cap e_j| = 0. \end{cases}
\end{aligned} \tag{5.16}$$

For the proof of (5.16), we note that

$$\mathbb{E}\big(\kappa_{n,i}(\theta_{0,n}, t_0 + uh) \kappa_{n,j}(\theta_{0,n}, t_0 + vh)\big) = T_{n,1}(u,v) - T_{n,2}(u,v),$$

where

$$\begin{aligned}
T_{n,1}(u,v) &= \mathbb{P}(C_{n,i}(t_0 + uh) = 1, C_{n,j}(t_0 + vh) = 1) \\
&\qquad \times f_{n,1}(\theta_{0,n}, t_0 + uh, t_0 + vh | i,j), \\
T_{n,2}(u,v) &= \mathbb{P}(C_{n,i}(t_0 + uh) = 1) \mathbb{P}(C_{n,j}(t_0 + vh) = 1) \\
&\qquad \times f_2(\theta_{0,n}, t_0 + uh) f_2(\theta_{0,n}, t_0 + vh).
\end{aligned}$$

We get

$$\begin{aligned}\mathbb{P}(C_{n,1}(t_0)=1)^{-2}&\int_{-1}^{1}\int_{-1}^{1}K(u)K(v)T_{n,2}(u,v)\mathrm{d}u\mathrm{d}v\\&=\left[\int_{-1}^{1}K(u)a_n(u)f_2(\theta_{0,n},t_0+uh)\mathrm{d}u\right]^2\to f_2(\theta_0(t_0),t_0)^2,\end{aligned}\tag{5.17}$$

where, again, (4.8) and continuity of $f_2(\theta,t)=-\partial_{\theta^2}g(\theta,t)$ has been used. Furthermore, we have that

$$\begin{aligned}&\mathbb{P}(C_{n,1}(t_0)=1)^{-2}\int_{-1}^{1}\int_{-1}^{1}K(u)K(v)T_{n,1}(u,v)\mathrm{d}u\mathrm{d}v\\&=\int_{-1}^{1}\int_{-1}^{1}K(u)K(v)\frac{\mathbb{P}(C_{n,i}(t_0+uh)=1,C_{n,j}(t_0+vh)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)^2}\\&\quad\times\left(f_{n,1}(\theta_{0,n},t_0+uh,t_0+vh|i,j)-f_1(\theta_0(t_0),|e_i\cap e_j|)\right)\mathrm{d}u\mathrm{d}v\\&\qquad+\int_{-1}^{1}\int_{-1}^{1}K(u)K(v)\frac{\mathbb{P}(C_{n,i}(t_0+uh)=1,C_{n,j}(t_0+vh)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)^2}\\&\qquad\quad\times f_1(\theta_0(t_0),|e_i\cap e_j|)\mathrm{d}u\mathrm{d}v\end{aligned}$$

$$\begin{cases}=o(n^2) & \text{for}\quad |e_i\cap e_j|=2\\ =o(n) & \text{for}\quad |e_i\cap e_j|=1\\ \to f_1(\theta_0(t_0),0)=f_2(\theta_0(t_0),t_0)^2 & \text{for}\quad |e_i\cap e_j|=0\end{cases}\tag{5.18}$$

by Assumptions (4.9) and (4.16). From (5.17) and (5.18), we obtain (5.16). This shows(5.9).

For the proof of (5.10), we calculate a bound for the expectation of the absolute value of the third derivative of P_n. With $s=t_0+uh$, it holds

$$\begin{aligned}&\mathbb{E}\left(\sup_{k,l,r,\theta}\left|\frac{\partial^2}{\partial\theta_k\partial\theta_l}P_n^{\prime(r)}(\theta)\right|\right)\\&\leq\frac{1}{\mathbb{P}(C_{n,1}(t_0)=1)}\int_{-1}^{1}K(u)\mathbb{P}(C_{n,1}(s)=1)\\&\qquad\qquad\times\mathbb{E}\left(\|X_{n,1}(s)\|^3e^{\tau\,X_{n,1}(s)\|}\middle|\,C_{n,1}(s)=1\right)\mathrm{d}u,\end{aligned}$$

where (4.4) has been used to get that the order of differentiation and integration can be interchanged and where Fubini could be used because all involved terms are non-negative. The upper bound for the expectation in the integral expression is bounded by Assumptions (4.4) and (4.8). This shows (5.10). ☐

Lemma 5.7. *It holds that*

$$\frac{1}{l_nh}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)C_{n,i}(t_0)$$

$$\times\left[C_{n,i}(s)X_{n,i}(s)\left(e^{\theta_0(s)^T X_{n,i}(s)} - e^{\theta_{0,n}^T X_{n,i}(s)}\right) - \partial_\theta g(\theta_{0,n}, s)\right] ds$$

$$= o_P\left(\frac{1}{\sqrt{l_n h}}\right). \tag{5.19}$$

With B_n from Theorem 4.2, we have

$$\frac{1}{l_n h}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)(1 - C_{n,i}(t_0)$$

$$\times C_{n,i}(s)X_{n,i}(s)\left(e^{\theta_0(s)^T X_{n,i}(s)} - e^{\theta_{0,n}^T X_{n,i}(s)}\right) ds)$$

$$= h^2 \cdot B_n + o_P(h^2). \tag{5.20}$$

Proof. In this proof, we use the shorthand notation $K_{h,t_0}(s) = \frac{1}{h}K\left(\frac{z-t_0}{h}\right)$. We begin with proving (5.20). Denote for vectors $a, b \in \mathbb{R}^q$ by $[a, b]$ the connecting line between a and b. Note firstly that by a Taylor series application for a random (depending on $X_{n,i}(s)$) intermediate value $\theta^*(s) \in [\theta_0(s), \theta_{0,n}]$

$$e^{\theta_0(s)^T X_{n,i}(s)} - e^{\theta_{0,n}^T X_{n,i}(s)}$$

$$= X_{n,i}(s)^T e^{\theta^*(s)^T X_{n,i}(s)} \cdot (\theta_0(s) - \theta_{0,n}). \tag{5.21}$$

Hence, we obtain

$$\frac{1}{l_n}\sum_{i\in G_n}\int_0^T K_{h,t_0}(s)(1 - C_{n,i}(t_0))C_{n,i}(s)$$

$$\times X_{n,i}(s)\left(e^{\theta_0(s)^T X_{n,i}(s)} - e^{\theta_{0,n}^T X_{ni}(s)}\right) ds$$

$$= \frac{1}{l_n}\sum_{i\in G_n}\int_0^T K_{h,t_0}(s)(1 - C_{n,i}(t_0))C_{n,i}(s)X_{n,i}(s)X_{n,i}(s)^T$$

$$\times\ e^{\theta^*(s)^T X_{n,i}(s)} \cdot (\theta_0(s) - \theta_0(t_0) + \theta_0(t_0) - \theta_{0,n})\mathrm{d}s \tag{5.22}$$

We decompose (5.22) into two terms by splitting $\theta_0(s) - \theta_0(t_0) + \theta_0(t_0) - \theta_{0,n} = (\theta_0(s) - \theta_0(t_0)) + (\theta_0(t_0) - \theta_{0,n})$. For the second part we obtain, by using that the parameter space is bounded by τ and convex (A3), use also Fubini in the second line and rewrite as a conditional expectation in the last line

$$\mathbb{E}\left(\left\|\frac{1}{l_n}\sum_{i\in G_n}\int_0^T K_{h,t_0}(s)(1 - C_{n,i}(t_0))C_{n,i}(s)X_{n,i}(s)X_{n,i}(s)^T\right.\right.$$

$$\left.\left.\times\ e^{\theta^*(s)^T X_{n,i}(s)} \cdot (\theta_0(t_0) - \theta_{0,n})\mathrm{d}s\right\|\right)$$

$$\leq \int_0^T K_{h,t_0}(s)\mathbb{E}\left(\tfrac{(1-C_{n,1}(t_0))C_{n,1}(s)}{\mathbb{P}(C_{n,1}(t_0)=1)}\left\|X_{n,1}(s)\right\|^2 e^{\tau\|X_{n,1}(s)\|}\right) ds \tag{5.23}$$

$$
\begin{aligned}
&\qquad\qquad \times\ \|\theta_0(t_0)-\theta_{0,n}\| \\
&= \int_0^T K_{h,t_0}(s)\tfrac{\mathbb{P}(C_{n,1}(t_0)=0,\,C_{n,1}(s)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)} \\
&\qquad \times \mathbb{E}\left(\|X_{n,1}(s)\|^2 e^{\tau\|X_{n,1}(s)\|}\middle| C_{n,1}(s)=1,\, C_{n,1}(t_0)=0\right) ds \qquad (5.24)\\
&\qquad\qquad \times \|\theta_0(t_0)-\theta_{0,n}\|, \\
&= O(h^3) \qquad (5.25)
\end{aligned}
$$

where the last equality holds, because by assumption (4.10) the first factor is $O(h)$, the second factor is uniformly bounded by (4.6) and $\|\theta_{0,n}-\theta_0(t_0)\| = O(h^2)$ by Proposition 5.5. We now discuss the second term of the split of (5.22). Recall therefore the definitions of $\gamma_{n,i}(s)$ and $\tau_{n,i}(\theta,s)$ from Theorem 4.2 and (4.13), respectively. Applying the above and using that $\theta_0(s)-\theta_0(t_0) = \theta_0'(t^*)(s-t_0)$ for an appropriate point $t^* \in [t_0,s]$, we obtain

$$
\begin{aligned}
(5.22) &= h^2\Big(\tfrac{1}{l_n}\sum_{i\in G_n}\int_0^T K_{h,t_0}(s)\tfrac{\gamma_{n,i}(s)}{h}X_{n,i}(s)X_{n,i}(s)^T \\
&\qquad \times\ \ e^{\theta^*(s)^T X_{n,i}(s)}\tfrac{\theta_0'(t^*)(t_0-s)}{h}\mathrm{d}s\Big) + o_P(h^2) \\
&= h^2\Big(\tfrac{1}{l_n}\sum_{i\in G_n}\int_0^T K_{h,t_0}(s)\tfrac{\gamma_{n,i}(s)}{h}\tau_{n,i}(\theta_0(s),s)\tfrac{\theta_0'(t_0)(t_0-s)}{h}\mathrm{d}s \\
&+ \tfrac{1}{l_n}\sum_{i\in G_n}\int_0^T K_{h,t_0}(s)\tfrac{\gamma_{n,i}(s)}{h}\tau_{n,i}(\theta_0(s),s)\tfrac{(\theta_0'(t^*)-\theta_0'(t_0))(t_0-s)}{h}\mathrm{d}s \qquad (5.26)\\
&\quad + \tfrac{1}{l_n}\sum_{i\in G_n}\int_0^T K_{h,t_0}(s)\tfrac{\gamma_{n,i}(s)}{h}X_{n,i}(s)X_{n,i}(s)^T \\
&\qquad \times\left(e^{\theta^*(s)^T X_{n,i}(s)} - e^{\theta_0(s)^T X_{n,i}(s)}\right)\tfrac{\theta_0'(t^*)(t_0-s)}{h}\mathrm{d}s\Big) \qquad (5.27)\\
&+ o_P(h^2).
\end{aligned}
$$

Hence, we need to prove that (5.26) and (5.27) are $o_P(1)$ (these lines individually without the leading h^2 from the first line) and we are done with the proof. K is supported on $[-1,1]$ and hence $s \in U_h := [t_0-h, t_0+h]$. Moreover, continuity of θ_0' yields $\sup_{s\in U_h}\frac{(\theta_0(t^*)-\theta_0'(t_0)(t_0-s)}{h} \to 0$. Hence, we can show $(5.26) = o_P(1)$ by similar arguments which lead to (5.25). For (5.27) we apply apply Taylor again to get for another intermediate point $\theta^{**}(s) \in [\theta_0(s), \theta^*(s)]$

$$
e^{\theta^*(s)^T X_{n,i}(s)} - e^{\theta_0(s)^T X_{n,i}(s)} = X_{n,i}(s)^T e^{\theta^{**}(s)^T X_{n,i}(s)}(\theta^*(s)-\theta_0(s)).
$$

Now arguments are again similar to the ones leading to (5.25), we just have to use the power three part in (4.6) and the fact that $\sup_{s\in U_h}\|\theta^*(s)-\theta_0(s)\| \le \sup_{s\in U_h}\|\theta_0(s)-\theta_{0,n}\|$ which converges to zero by continuity of θ and Proposition 5.5. This concludes the proof of (5.20).

To prove (5.19), we have to show that

$$\frac{1}{\sqrt{l_n h}} \sum_{i,j} \int_0^T K\left(\frac{s-t_0}{h}\right) C_{n,i}(t_0) r_{n,i}(s) ds = o_P(1),$$

where $r_{n,i}(s)$ was defined in Assumption (A7). By an application of Markov's inequality, this holds if

$$\frac{h}{l_n} \sum_{i,j \in G_n} \int_{-1}^{1} \int_{-1}^{1} K(u)K(v)\mathbb{P}(C_{n,i}(t_0) = 1, C_{n,j}(t_0) = 1)$$
$$\times \mathbb{E}(r_{n,i}(t_0 + uh) r_{n,j}(t_0 + vh) | C_{n,i}(t_0) = 1, C_{n,j}(t_0) = 1) \mathrm{d}u\mathrm{d}v = o(1).$$

We show this similarly as in the proof of Lemma 5.6 by splitting the sum in three sums corresponding to $|e_i \cap e_j| = 2, 1$, or 0. The corresponding sums have $O(n^2), O(n^3)$ and $O(n^4)$ terms, respectively. Before going through these three cases, we note that equations (5.21) and (4.5) imply that $\sup_{u,v \in [-1,1]} \mathbb{E}(r_{n,i}(t_0 + uh) r_{n,j}(t_0 + vh) | C_{n,i}(t_0) = 1, C_{n,j}(t_0) = 1) = O(h^2)$. This rate holds for all i and j. Now we get for the sum over edges i, j with $|e_i \cap e_j| = 2$ the bound

$$h \frac{\mathbb{P}(C_{n,1}(t_0) = 1)}{\mathbb{P}(C_{n,1}(t_0) = 1)} \int_{-1}^{1} \int_{-1}^{1} K(u)K(v) O(h^2) \mathrm{d}u\mathrm{d}v = o(1).$$

For the sum over edges with $|e_i \cap e_j| = 1$, we get the following bound from (4.11)

$$nh \mathbb{P}(C_{n,i}(t_0) = 1) \frac{\mathbb{P}(C_{n,i}(t_0) = 1, C_{n,j}(t_0) = 1)}{\mathbb{P}(C_{n,i}(t_0) = 1)^2} O(h^2) = O(1) \cdot \frac{l_n}{n} O(h^3).$$

Observing that $\frac{l_n h^3}{n} = \frac{l_n^{3/5} (h^5)^{3/5} l_n^{2/5}}{n} = O(\frac{l_n^{2/5}}{n}) = O(n^{-1/5} \mathbb{P}(C_{n,1}(t_0) = 1)^{2/5}) = o(1)$, the bound is of order $o(1)$.

By using (4.11) and (4.17), we get the following bound for the sum over edges with $|e_i \cap e_j| = 0$:

$$l_n h \frac{\mathbb{P}(C_{n,i}(t_0) = 1, C_{n,j}(t_0) = 1)}{\mathbb{P}(C_{n,i}(t_0) = 1)^2}$$
$$\iint_{[-1,1]^2} K(u)K(v) \mathbb{E}(r_{n,i}(t_0 + uh) r_{n,j}(t_0 + vh) | C_{n,i}(t_0) = 1, C_{n,j}(t_0) = 1) \mathrm{d}u\mathrm{d}v$$
$$= o(1).$$

This concludes the proof of (5.19). □

Proposition 5.8. *With probability tending to one, the equation $P_n'(\theta) = 0$ (cf. equation (5.3)) has a solution $\widetilde{\theta}_n(t_0)$, which has the property*

$$\widetilde{\theta}_n(t_0) = \theta_{0,n} + h^2 \cdot B_n + o_P\left(\frac{1}{\sqrt{l_n h}}\right) + o_P(h^2).$$

To prove this proposition, we will make use of the following theorem, see Deimling (1985):

Theorem 5.9. *(Newton-Kantorovich Theorem) Let $R(x) = 0$ be a system of equations where $R : D_0 \subseteq \mathbb{R}^q \to \mathbb{R}$ is a function defined on D_0. Let R be differentiable and denote by R' its first derivative. Assume that there is an x_0 such that all expressions in the following statements exist and such that the following statements are true*

1. $||R'(x_0)^{-1}|| \leq B$,
2. $||R'(x_0)^{-1}R(x_0)|| \leq \eta$,
3. $||R'(x) - R'(y)|| \leq K||x - y||$ *for all* $x, y \in D_0$,
4. $r := BK\eta \leq \frac{1}{2}$ *and* $\Omega_* := \{x : ||x - x_0|| < 2\eta\} \subseteq D_0$.

Then there is $x^* \in \Omega_*$ with $R(x^*) = 0$ and

$$||x^* - x_0|| \leq 2\eta \text{ and } ||x^* - (x_0 - R'(x_0)^{-1}R(x_0))|| \leq 2r\eta.$$

Proof of Proposition 5.8. We show that $P_n'(\theta)$ has a root by using Theorem 5.9. Lemma 5.6 gives that $P_n'(\theta_{0,n}) \xrightarrow{\mathbb{P}} 0$ and $P_n''(\theta_{0,n}) \xrightarrow{\mathbb{P}} -\Sigma$. Since Σ is invertible we also have that the sequence of random variables $B_n := ||P_n''(\theta_{0,n})^{-1}||$ is well-defined (for large n) and that it is of order $O_P(1)$. Thus we also have $\eta_n := ||P_n''(\theta_{0,n})^{-1}P_n'(\theta_{0,n})|| = o_P(1)$. For the Lipschitz continuity of P_n'' we bound the partial derivatives of P_n'' by Lemma 5.6. Hence we conclude that every realization of P_n'' is Lipschitz continuous with (random) Lipschitz constant $K_n = O_P(1)$. Combining everything, we get that $r_n := B_n K_n \eta_n = o_P(1)$. Thus with probability tending to one we have $r_n \leq \frac{1}{2}$, and hence the Newton-Kantorovich Theorem tells us that with probability tending to one the equation $P_n'(\theta) = 0$ has a solution $\widetilde{\theta}_n(t_0)$ with the property that

$$\|\widetilde{\theta}_n(t_0) - \theta_{0,n}\| \leq 2\eta_n = o_P(1).$$

To prove the asserted rate, we have to investigate η_n further. We note first that since $P_n''(\theta_{0,n})^{-1}$ is stochastically bounded, the rate of η_n is determined by the rate with which $P_n'(\theta_{0,n})$ converges to zero. To find this rate we observe that every summand of $P_n'(\theta_{0,n})$ has expectation zero conditionally on $C_{n,i}(s) = 1$:

$$\int_0^T K\big(\tfrac{s-t_0}{h}\big)\mathbb{E}\big[C_{n,i}(s)X_{n,i}(s)\big(e^{\theta_0(s)^T X_{n,i}(s)} - e^{\theta_{0,n}^T X_{n,i}(s)}\big)\big|C_{n,i}(s) = 1\big]ds$$
$$= \int_0^T K\big(\tfrac{s-t_0}{h}\big)\partial_\theta g(\theta_{0,n}, s)ds = 0$$

by the assumption that $\theta_{0,n}$ maximizes $\theta \mapsto \int_0^T K\left(\frac{s-t_0}{h}\right) g(\theta, s)ds$. So, in $P_n'(\theta_{0,n})$, we can subtract $C_{n,i}(t_0)\int_0^T K\left(\frac{s-t_0}{h}\right)\partial_\theta g(\theta_{0,n}, s)ds$ from every summand without changing anything, i.e.,

$$P_n'(\theta_{0,n})$$

$$
\begin{aligned}
=&\frac{1}{l_n h}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)\Big[C_{n,i}(s)X_{n,i}(s)\left(e^{\theta_0(s)^T X_{n,i}(s)}-e^{\theta_{0,n}^T X_{n,i}(s)}\right)\\
&\qquad\qquad\qquad\qquad\qquad\qquad -C_{n,i}(t_0)\partial_\theta g(\theta_{0,n},s)\Big]ds\\
=&\frac{1}{l_n h}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)C_{n,i}(t_0)\Big[C_{n,i}(s)X_{n,i}(s)\Big(e^{\theta_0(s)^T X_{n,i}(s)}\\
&\qquad\qquad\qquad\qquad\qquad\qquad -e^{\theta_{0,n}^T X_{n,i}(s)}\Big)-\partial_\theta g(\theta_{0,n},s)\Big]\,ds\\
&+\frac{1}{l_n h}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)(1-C_{n,i}(t_0))C_{n,i}(s)X_{n,i}(s)\\
&\quad\times\left(e^{\theta_0(s)^T X_{n,i}(s)}-e^{\theta_{0,n}^T X_{n,i}(s)}\right)ds.
\end{aligned}
$$

By Lemma 5.7, this term is equal to $h^2\cdot B_n+o_P\left(\frac{1}{\sqrt{l_n h}}\right)+o_P(h^2)$, which concludes the proof of Proposition 5.8. □

Lemma 5.10. *For $k,l\in\{1,...,q\}$, we have that*

$$
\begin{aligned}
&\frac{1}{l_n h}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)^2 X_{n,i}^{(l)}(s)X_{n,i}^{(k)}(s)C_{n,i}(s)\exp(\theta_0(s)^T X_{n,i}(s))ds\\
&\xrightarrow{\mathbb{P}}\int_{-1}^1 K(u)^2\mathrm{d}u\ \Sigma_{k,l}
\end{aligned}
\tag{5.28}
$$

and

$$
\begin{aligned}
&\frac{1}{l_n h}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)^2\|X_{n,i}(s)\|^2\mathbb{1}\left(\frac{1}{\sqrt{l_n h}}\left\|K\left(\frac{s-t_0}{h}\right)X_{n,i}(s)\right\|>\varepsilon\right)\\
&\qquad\times C_{n,i}(s)\exp(\theta_0(s)^T X_{n,i}(s))ds\xrightarrow{\mathbb{P}}0.
\end{aligned}
\tag{5.29}
$$

Moreover, it holds that

$$
\frac{1}{l_n}\partial_\theta^2\ell(\widetilde{\theta}_n(t_0),t_0)=P_n''(\widetilde{\theta}_n(t_0))\xrightarrow{\mathbb{P}}-\Sigma. \tag{5.30}
$$

Proof. The proof of (5.28) follows by using similar arguments as in the proof of Lemma 5.6, with $\theta_{0,n}$ replaced by $\theta_0(s)$, and with K replaced by K^2.

For the proof of claim (5.29), we calculate the expectation of the left hand side of (5.29). Because the integrand is positive, we can apply Fubini, and we get that the expectation is equal to

$$
\begin{aligned}
&\int_0^T\mathbb{E}\left[\mathbb{1}\left(\frac{1}{\sqrt{l_n h}}\left\|K\left(\frac{s-t_0}{h}\right)X_{n,1}(s)\right\|>\varepsilon\right)\|X_{n,1}(s)\|^2\right.\\
&\left.\times\exp\left(\theta_0(s)^T X_{n,1}(s)\right)\Big|\,C_{n,1}(s)=1\right]\frac{1}{h}K\left(\frac{s-t_0}{h}\right)^2\frac{\mathbb{P}(C_{n,1}(s)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)}\mathrm{d}s
\end{aligned}
$$

$$
\begin{aligned}
\leq\ & \frac{1}{\varepsilon} \cdot \frac{1}{\sqrt{l_n h}} \int_{-1}^{1} K^3(u) \frac{\mathbb{P}(C_{n,1}(t_0+uh)=1)}{\mathbb{P}(C_{n,1}(t_0)=1)} \\
& \mathbb{E}\left(\|X_{n,1}(t_0+uh)\|^3 e^{\tau \|X_{n,1}(t_0+uh)\|} \middle| C_{n,1}(t_0+uh)=1 \right) \mathrm{d}u \\
=\ & O\left(\frac{1}{\sqrt{l_n h}}\right) = o(1).
\end{aligned}
$$

Here we use (4.8), $\max_{-1\leq u\leq 1} K(u) < \infty$ and (4.7). This shows (5.29).

To see (5.30), we show that

$$
P_n''(\theta_{0,n}) - P_n''(\widetilde{\theta}_n(t_0)) = o_P(1). \tag{5.31}
$$

This then implies (5.30) because of (5.9).

By using exactly the same arguments as in the proof of Lemma 5.7, we obtain

$$
e^{\theta_{0,n}^T X_{n,i}(s)} - e^{\widetilde{\theta}_n(t_0)_n^T X_{n,i}(s)} \leq \|X_{n,i}(s)\| e^{\tau \|X_{n,i}(s)\|} \cdot \|\theta_{0,n} - \widetilde{\theta}_n(t_0)\|.
$$

This gives

$$
\begin{aligned}
& P_n''(\theta_{0,n}) - P_n''(\widetilde{\theta}_n(t_0)) \\
=& \frac{1}{l_n h} \sum_{i\in G_n} \int_0^T K\left(\frac{s-t_0}{h}\right) C_{n,i}(s) X_{n,i}(s) X_{n,i}(s)^T \\
& \qquad \times \left(e^{\widetilde{\theta}_n(t_0)^T X_{n,i}(s)} - e^{\theta_{0,n}^T X_{n,i}(s)} \right) ds \\
\leq & \frac{1}{l_n h} \sum_{i\in G_n} \int_0^T K\left(\frac{s-t_0}{h}\right) C_{n,i}(s) \|X_{n,i}(s)\|^3 e^{\tau \|X_{n,i}(s)\|} ds \times \|\theta_{0,n} - \widetilde{\theta}_n(t_0)\|.
\end{aligned}
$$

The expectation of the first factor is bounded because of assumptions (4.8) and (4.7). Furthermore, the second term is of order $o_P(1)$ by Proposition 5.8. Thus, the product is of order $o_P(1)$. This shows (5.31) and concludes the proof of (5.30). □

Proposition 5.11. *With probability tending to one, $\partial_\theta \ell_n(\theta, t_0) = 0$ has a solution $\hat{\theta}_n(t_0)$, and*

$$
\sqrt{l_n h} \cdot (\hat{\theta}_n(t_0) - \widetilde{\theta}_n(t_0)) \xrightarrow{d} N\Big(0, \int_{-1}^{1} K^2(u) \mathrm{d}u\ \Sigma^{-1}\Big).
$$

Proof of Proposition 5.11. The proof is based on modifications of arguments used in the asymptotic analysis of parametric counting process models, see e.g. the proof of Theorem VI.1.1 on p. 422 in Andersen et al.(1993). Define

$$
U^l(\theta) \quad := \quad h \partial_{\theta_l} \ell_n(\theta, t_0), \qquad l = 1, \ldots, q,
$$

and let $U_t^l(\theta)$ be defined as $U^l(\theta)$, but with t being the upper limit of the integral in (4.1), (i.e., $U^l(\theta) = U_T^l(\theta)$). Furthermore, we write $U(\theta) = (U^1(\theta), ..., U^q(\theta))$, and the vector $U_t(\theta)$ is defined analogously. In the first step of the proof, we will show that

$$
\frac{1}{\sqrt{l_n h}} U_T(\widetilde{\theta}_n(t_0)) \xrightarrow{d} N\Big(0, \int_{-1}^{1} K^2(u) \mathrm{d}u\ \Sigma\Big). \tag{5.32}
$$

For the local, square integrable martingale $M_{n,i}$ defined in (5.1), it holds that $M_{n,i}$ and $M_{n,j}$ are orthogonal, meaning that $< M_{n,i}, M_{n,j} >_t = 0$ if $i \neq j$, i.e. the predictable covariation process is equal to zero (cf. Section 2.2, in particular (2.2)). For the predictable variation process of $M_{n,i}$, we have

$$< M_{n,i} >_t = \int_0^t C_{n,i}(s) \exp(\theta_0(s)^T X_{n,i}(s)) ds. \tag{5.33}$$

By definition of $\widetilde{\theta}_n(t_0)$, see the statement of Proposition 5.8, we have that (write $K_{h,t_0}(s) := K\left(\frac{s-t_0}{h}\right)$)

$$\begin{aligned}
&U_t^l(\widetilde{\theta}_n(t_0))\\
&= \sum_{i \in G_n} \int_0^t K_{h,t_0}(s) X_{n,i}^{(l)}(s) dN_{n,i}(s) \qquad (5.34)\\
&\qquad - \int_0^t K_{h,t_0}(s) C_{n,i}(s) X_{n,i}^{(l)}(s) \exp(\widetilde{\theta}_n(t_0)^T X_{n,i}(s)) ds\\
&= \sum_{i \in G_n} \int_0^t K_{h,t_0}(s) X_{n,i}^{(l)}(s) dM_{n,i}(s)\\
&+ \int_0^t K_{h,t_0}(s) C_{n,i}(s) X_{n,i}^{(l)}(s) \left(\exp(\theta_0(s)^T X_{n,i}(s)) - \exp(\widetilde{\theta}_n(t_0)^T X_{n,i}(s))\right) ds\\
&= \sum_{i \in G_n} \int_0^t K_{h,t_0}(s) X_{n,i}^{(l)}(s) dM_{n,i}(s).
\end{aligned}$$

So $\widetilde{\theta}_n(t_0)$ was chosen such that the non-martingale part of $\partial_\theta \ell(\widetilde{\theta}_n(t_0), t_0)$ vanishes. Now, we want to apply Rebolledo's Martingale Convergence Theorem, see e.g. Theorem II.5.1 in Andersen et al.(1993). This theorem implies (5.32), provided a Lindeberg condition (5.29) holds, and

$$\left\langle \frac{1}{\sqrt{l_n h}} U_t^k(\widetilde{\theta}_n(t_0)), \frac{1}{\sqrt{l_n h}} U_t^l(\widetilde{\theta}_n(t_0)) \right\rangle_T \xrightarrow{\mathbb{P}} \int_{-1}^1 K^2(u) \mathrm{d}u \; \Sigma_{kl}(t_0). \tag{5.35}$$

To verify (5.35), first note that (5.33) and (5.28) imply finiteness of

$$\frac{1}{l_n h} \sum_{i \in G_n} \int_0^t K_{h,t_0}(s)^2 \left(X_{n,i}^{(l)}(s)\right)^2 d\langle M_{n,i} \rangle_s,$$

with probability tending to one. Note that Lemma 5.10 is formulated with $t = T$, but the integral is finite also for $t < T$ simply because the integrand is non-negative. From now on we assume the above integral is finite. The process

$$\frac{1}{\sqrt{l_n h}} \sum_{i \in G_n} \int_0^t K_{h,t_0}(s) X_{n,i}^{(l)}(s) dM_{n,i}(s)$$

is a local square integrable martingale, cf. Theorem 2.7 and the discussion afterwards. Since the martingales $M_{n,i}$ are orthogonal, and by using Lemma 5.10, the predictable covariation satisfies

$$\begin{aligned}
&\Big\langle \frac{1}{\sqrt{l_n h}} U_t^k(\widetilde{\theta}_n(t_0)), \frac{1}{\sqrt{l_n h}} U_t^l(\widetilde{\theta}_n(t_0))\Big\rangle_T \\
= \ & \frac{1}{l_n h} \sum_{i \in G_n} \int_0^T K_{h,t_0}(s)^2 X_{n,i}^{(k)}(s) X_{n,i}^{(l)}(s) C_{n,i}(s) \exp(\theta_0(s)^T X_{n,i}(s)) ds \\
\xrightarrow{\mathbb{P}} \ & \int_{-1}^1 K^2(u)\mathrm{d}u \ \Sigma_{kl}(t_0).
\end{aligned}$$

This shows (5.35), and concludes the proof of (5.32).

We now show that

$$||\sqrt{l_n h} \cdot (\widetilde{\theta}_n(t_0) - \hat{\theta}_n(t_0)) - \sqrt{l_n h} Z_n|| \xrightarrow{\mathbb{P}} 0, \tag{5.36}$$

where

$$Z_n = P_n''(\widetilde{\theta}_n(t_0))^{-1} \frac{1}{l_n h} U(\widetilde{\theta}_n(t_0)).$$

We want to apply the Newton-Kantorovich Theorem 5.9 with $R(\theta) := R_n(\theta) := \frac{1}{l_n h} U(\theta)$ and $x_0 := \widetilde{\theta}_n(t_0)$. To this end, define

$$B_n := \|R_n'(\widetilde{\theta}_n(t_0))^{-1}\| = \left\| P_n''\left(\widetilde{\theta}_n(t_0)\right)^{-1} \right\|.$$

From Lemma 5.10, we know that $P_n''(\widetilde{\theta}_n(t_0))$ converges and is invertible for n large enough, and thus $B_n = O_P(1)$. Now let

$$\eta_n := \|R_n'(\widetilde{\theta}_n(t_0))^{-1} R_n(\widetilde{\theta}_n(t_0))\| = \|Z_n\|.$$

Results (5.30) and (5.32) imply that $\eta_n = o_P(1)$. Next, notice that P_n'' has a Lipschitz constant K_n that is bounded by the maximum of the third derivative of P_n. According to (5.10), this maximum is bounded, and we obtain $K_n = O_P(1)$. Hence, $r_n = B_n K_n \eta_n = o_P(1)$. Now, Theorem 5.9 implies that, with probability converging to one, there is $\hat{\theta}_n(t_0)$ such that $U(\hat{\theta}_n(t_0)) = 0$ and

$$||\hat{\theta}_n(t_0) - \widetilde{\theta}_n(t_0)|| \leq 2\eta_n \xrightarrow{\mathbb{P}} 0.$$

To obtain the asymptotic distribution of $\hat{\theta}_n(t_0)$, we note that, by (5.32) and (5.30),

$$\sqrt{l_n h} \cdot Z_n \xrightarrow{d} N(0, \int_{-1}^1 K^2(u)\mathrm{d}u \ \Sigma^{-1}). \tag{5.37}$$

Thus it holds that $\sqrt{l_n h} \cdot Z_n = O_P(1)$, and as a consequence we get $\sqrt{l_n h} \cdot \eta_n = O_P(1)$. Using the second statement of the Newton-Kantorovich Theorem 5.9, we obtain

$$||\sqrt{l_n h} \cdot (\widetilde{\theta}_n(t_0) - \hat{\theta}_n(t_0)) - \sqrt{l_n h} Z_n|| \leq \sqrt{l_n h} \cdot 2 r_n \eta_n = o_P(1).$$

Thus $\sqrt{l_n h} \cdot (\widetilde{\theta}_n(t_0) - \hat{\theta}_n(t_0))$ and $\sqrt{l_n h} \cdot Z_n$ have the same limit distribution. Because of (5.37) this implies the statement of the proposition. □

Proof of Theorem 4.2 Combining Propositions 5.8 and 5.11, and applying Slutzky's Lemma, we obtain by the assumptions on the bandwidth h in (A1)

$$\sqrt{l_n h}\left(\hat{\theta}_n(t_0) - \theta_{0,n}(t_0) - h^2 \cdot B_n\right) \to N\left(0, \int_{-1}^{1} K^2(u)du\, \Sigma^{-1} A \Sigma^{-1}\right).$$

With Proposition 5.5 this proves Theorem (4.2).

5.2 Proof of Theorem 4.3

For the proof we need the following auxiliary propositions.

Proposition 5.12. *With the same assumptions and definitions as in Theorem 4.3 we have*

$$\sup_{t_0 \in \mathbb{T}} \left\| \frac{1}{r_n \sqrt{\bar{p}_n(t_0)}} \ell_n'(\theta_0, t_0) \right\| = O_P\left(\sqrt{\frac{\log r_n}{r_n h}}\right)$$

Proposition 5.13. *Assume the same assumptions as in Theorem 4.3. For any choice of $\theta_1^*(t_0), ..., \theta_q^*(t_0) \in [\theta_0, \hat{\theta}(t_0)]$ (where for $a, b \in \mathbb{R}^q$ we denote by $[a, b]$ the connecting line between a and b), define the matrix*

$$\ell_n^*(t_0) := \begin{pmatrix} \ell''_{n,1\cdot}(\theta_1^*(t_0), t_0) \\ \vdots \\ \ell''_{n,q\cdot}(\theta_q^*(t_0), t_0) \end{pmatrix},$$

where $\ell''_{n,r\cdot}$ denotes for $r \in \{1, ..., q\}$ the r-th line of the second derivative of ℓ_n with respect to θ. The matrix $\ell_n^*(t_0)$ concentrates around $\Sigma(\theta_0, t_0)$ (cf. Statement 4.2), i.e.,

$$\sup_{t_0 \in \mathbb{T}} \left\| \frac{1}{r_n \bar{p}_n(t_0)} \ell_n^*(t_0) - \Sigma(t_0, \theta_0) \right\| = O_P\left(\sqrt{\frac{\log r_n}{r_n p_n \cdot h}} + h\right),$$

with $p_n := \inf_{t_0 \in \mathbb{T}} p_n(t_0)$. Furthermore, $\ell_n^*(t_0)$ is invertible and

$$\sup_{t_0 \in \mathbb{T}} \left\| \left[\frac{1}{r_n \bar{p}_n(t_0)} \ell_n^*(t_0)\right]^{-1} - \Sigma(t_0, \theta_0)^{-1} \right\| = O_P\left(\sqrt{\frac{\log r_n}{r_n p_n \cdot h}} + h\right).$$

Proposition 5.14. *Let $\bar{p}_n := \int_0^T \bar{p}_n(s)ds$. We assume that we are on H_0 and denote the true parameter by θ_0. Then*

$$\sqrt{r_n h} \int_0^T \left(\hat{\theta}_n(t_0) - \theta_0\right) \frac{\bar{p}_n(t_0)}{\sqrt{\bar{p}_n}} w(t_0) dt_0 = o_P(1).$$

The proofs of these propositions are deferred to the end.

Proof of Theorem 4.3. We note firstly that we may replace the estimator $\bar{\theta}_n$ in the test statistic with θ_0 because for

$$T_{0,n} := \int_0^T \left\|\hat{\theta}_n(t_0) - \theta_0\right\|^2 \bar{p}_n(t_0)w(t_0)dt_0$$

it holds that

$$\begin{aligned} r_n h^{\frac{1}{2}} T_n =& r_n h^{\frac{1}{2}} T_{0,n} \\ & r_n h^{\frac{1}{2}} \int_0^T \left\|\bar{\theta}_n - \theta_0\right\|^2 \bar{p}_n(t_0)w(t_0)dt_0 \\ & + 2r_n h^{\frac{1}{2}} \int_0^T \left(\hat{\theta}_n(t_0) - \theta_0\right)^T \left(\bar{\theta}_n - \theta_0\right) \bar{p}_n(t_0)w(t_0)dt_0. \end{aligned}$$

By the Assumption (B1), 3 and Proposition 5.14, we see that the last two lines may be asymptotically neglected and hence the limiting distributions of $r_n h^{\frac{1}{2}} T_n$ and $r_n h^{\frac{1}{2}} T_{0,n}$ are identical and we study the behaviour of $r_n h^{\frac{1}{2}} T_{0,n}$ in the following.

By assumption (B5), 2, $\hat{\theta}(t_0) \in \Theta$ with high probability and hence $\ell_n'(\hat{\theta}(t_0), t_0) = 0$ on this event (the derivative exists by Statement 4.1). As we are concerned with convergence of the distribution of $\hat{\theta}(t_0)$, we can restrict to this event. By a Taylor expansion there are $\theta_r^*(t_0)$ which lie on the connecting line between $\hat{\theta}(t_0)$ and θ_0 such that

$$0 = \ell_{n,r}'(\hat{\theta}, t_0) = \ell_{n,r}'(\theta_0(t_0), t_0) + \ell_{n,r\cdot}''(\theta_r^*(t_0), t_0)(\hat{\theta}(t_0) - \theta_0),$$

where $\ell_{n,r}'$ is the r-th component of the gradient (with respect to θ) of ℓ_n and $\ell_{n,r\cdot}''$ denotes the r-th row of the Hessian Matrix of ℓ_n with respect to θ. Define

$$\ell_n^*(t_0) := \begin{pmatrix} \ell_{n,1\cdot}''(\theta_1^*(t_0), t_0) \\ \vdots \\ \ell_{n,q\cdot}''(\theta_q^*(t_0), t_0) \end{pmatrix}.$$

In this notation we have (use also Proposition 5.13 and Assumption (B7))

$$\begin{aligned} & 0 = \ell_n'(\hat{\theta}(t_0), t_0) = \ell_n'(\theta_0, t_0) + \ell_n^*(t_0)(\hat{\theta}(t_0) - \theta_0) \\ \Leftrightarrow \quad & \hat{\theta}(t_0) - \theta_0 = -\left[\ell_n^*(t_0)\right]^{-1} \cdot \ell_n'(\theta_0, t_0). \end{aligned} \tag{5.38}$$

Using this expansion and by applying Propositions 5.12 and 5.13, we obtain (note also that by Assumption (B3), w is supported on $[\delta, T-\delta]$ and bounded)

$$\begin{aligned} T_{0,n} &= \int_\delta^{T-\delta} \left\|\hat{\theta}(t_0) - \theta_0\right\|^2 w(t_0)\bar{p}_n(t_0)dt_0 \\ &= \int_\delta^{T-\delta} \left\| \left[\frac{1}{r_n \bar{p}_n(t_0)} \ell_n^*(t_0)\right]^{-1} \cdot \frac{1}{r_n \bar{p}_n(t_0)} \ell_n'(\theta_0, t_0) \right\|^2 w(t_0)\bar{p}_n(t_0)dt_0 \\ &= \int_\delta^{T-\delta} \left\| \Sigma(t_0, \theta_0)^{-1} \cdot \frac{1}{r_n \bar{p}_n(t_0)} \ell_n'(\theta_0, t_0) \right. \end{aligned}$$

$$
\begin{aligned}
&+\left(\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n^*(t_0)\right]^{-1}-\Sigma(t_0,\theta_0)^{-1}\right)\cdot\frac{1}{r_n\bar{p}_n(t_0)}\ell_n'(\theta_0,t_0)\Bigg\|^2\\
&\times w(t_0)\bar{p}_n(t_0)dt_0\\
=&\int_\delta^{T-\delta}\left\|\Sigma(t_0,\theta_0)^{-1}\cdot\frac{1}{r_n\bar{p}_n(t_0)}\ell_n'(\theta_0,t_0)\right\|^2 w(t_0)\bar{p}_n(t_0)dt_0\\
&+O_P\left(\frac{\log r_n}{r_n h}\left(\sqrt{\frac{\log r_n}{r_n p_n h}}+h+\left(\sqrt{\frac{\log r_n}{r_n p_n h}}+h\right)^2\right)\right)\\
=&\int_\delta^{T-\delta}\left\|\Sigma(t_0,\theta_0)^{-1}\cdot\frac{1}{r_n\bar{p}_n(t_0)}\ell_n'(\theta_0,t_0)\right\|^2 w(t_0)\bar{p}_n(t_0)dt_0\\
&+O_P\left(\frac{\log r_n}{r_n h}\left(\sqrt{\frac{\log r_n}{r_n p_n h}}+h\right)\right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&r_n h^{\frac{1}{2}} T_{0,n}\\
=&\ r_n h^{\frac{1}{2}}\int_\delta^{T-\delta}\left\|\Sigma(t_0,\theta_0)^{-1}\cdot\frac{1}{r_n\bar{p}_n(t_0)}\ell_n'(\theta_0,t_0)\right\|^2 w(t_0)\bar{p}_n(t_0)dt_0\\
&+O_P\left(\frac{(\log r_n)^{\frac{3}{2}}}{\sqrt{r_n p_n}\cdot h}+\left(h(\log r_n)^2\right)^{\frac{1}{2}}\right),
\end{aligned}
$$

where the O_p part is $o_P(1)$ by Assumption (B4), 1 on the bandwidth. Thus, for the asymptotic considerations, we have to only investigate the first part. On the hypothesis that the true parameter function is constant, we can compute that $\widetilde{\theta}=\theta_0$ by noting that $\log x\cdot y-x\le\log y\cdot y-y$ for all $x,y>0$. Hence, by definitions of $\widetilde{\theta}(t_0)$ and P_n (cf. Proposition 5.8 and (5.2), respectively), we have $P_n'(\theta_0,t_0)=0$ (P_n is differentiable by Statement 4.1). By Assumption (B1), 2, $\lambda_{n,i}$ is differentiable and we denote by $\lambda_{n,i}^\theta$ the derivative of $\lambda_{n,i}$ with respect to θ. Denote furthermore $\Sigma:=\Sigma(t_0,\theta_0)$ for ease of notation. We compute, by Fubini (Statement 4.3)

$$
\begin{aligned}
&r_n h^{\frac{1}{2}}\int_\delta^{T-\delta}\left\|\Sigma^{-1}\cdot\frac{1}{r_n\bar{p}_n(t_0)}\ell_n'(\theta_0,t_0)\right\|^2 w(t_0)\bar{p}_n(t_0)dt_0\\
=&\ r_n h^{\frac{1}{2}}\int_\delta^{T-\delta}\left\|\Sigma^{-1}\cdot\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}\int_0^T\frac{1}{h}K\left(\frac{t-t_0}{h}\right)\frac{\lambda_{n,i}^\theta(\theta_0,t)}{\lambda_{n,i}(\theta_0,t)}dM_{n,i}(t)\right\|^2\\
&\times w(t_0)\bar{p}_n(t_0)dt_0\\
=&\ \frac{1}{h^{\frac{1}{2}}r_n}\sum_{i,j\in G_n}\int_0^T\int_0^T\int_\delta^{T-\delta}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)K\left(\frac{t-t_0}{h}\right)\\
&\times\frac{\lambda_{n,i}^\theta(\theta_0,s)^T}{\lambda_{n,i}(\theta_0,s)}\Sigma^{-T}\Sigma^{-1}\frac{\lambda_{n,j}^\theta(\theta_0,t)}{\lambda_{n,j}(\theta_0,t)}\frac{w(t_0)}{\bar{p}_n(t_0)}dt_0dM_{n,i}(s)dM_{n,j}(t).
\end{aligned}
\tag{5.39}
$$

Let now

$$\begin{aligned}
& f_{n,ij}(s,t) \\
:= \quad & \frac{\lambda^{\theta}_{n,i}(\theta_0,s)^T}{\lambda_{n,i}(\theta_0,s)} \times \int_{\delta}^{T-\delta} \frac{1}{h} K\left(\frac{s-t_0}{h}\right) K\left(\frac{t-t_0}{h}\right) \\
& \times \Sigma^{-T}\Sigma^{-1} \frac{w(t_0)}{\bar{p}_n(t_0)} dt_0 \times \frac{\lambda^{\theta}_{n,j}(\theta_0,t)}{\lambda_{n,j}(\theta_0,t)}.
\end{aligned} \tag{5.40}$$

Note that $f_{n,ij}(s,t) = f_{n,ji}(t,s)$. Then (in the second line we used Fubini again, by Statement 4.3, and the third equality is not term-wise the same but for the whole sum)

$$\begin{aligned}
& (5.39) \\
= & \frac{1}{h^{\frac{1}{2}} r_n} \sum_{i,j\in G_n} \int_0^T \int_0^T f_{n,ij}(s,t) \left(\mathbb{1}_{t<s} + \mathbb{1}_{t>s} + \mathbb{1}_{t=s}\right) dM_{n,i}(s) dM_{n,j}(t) \\
= & \frac{1}{h^{\frac{1}{2}} r_n} \sum_{i,j\in G_n} \left[\int_0^T \int_0^{s-} f_{n,ij}(s,t) dM_{n,j}(t) dM_{n,i}(s) \right. \\
& \qquad + \int_0^T \int_0^{t-} f_{n,ij}(s,t) dM_{n,i}(s) dM_{n,j}(t) \\
& \qquad \left. + \int_0^T \int_{\{s\}} f_{n,ij}(s,t) dM_{n,j}(t) dM_{n,i}(s) \right] \\
= & \frac{1}{h^{\frac{1}{2}} r_n} \sum_{i\in G_n} \left[2\int_0^T \int_0^{s-} f_{n,ii}(s,t) dM_{n,i}(t) dM_{n,i}(s) + \int_0^T \int_{\{s\}} f_{n,ii}(s,t) dM_{n,i}(t) dM_{n,i}(s) \right]
\end{aligned} \tag{5.41}$$

$$\begin{aligned}
& + \frac{1}{h^{\frac{1}{2}} r_n} \sum_{\substack{i,j\in G_n \\ i\neq j}} \left[2\int_0^T \int_0^{s-} f_{n,ij}(s,t) dM_{n,j}(t) dM_{n,i}(s) \right. \\
& \qquad \left. + \int_0^T \int_{\{s\}} f_{n,ij}(s,t) dM_{n,j}(t) dM_{n,i}(s) \right]
\end{aligned} \tag{5.42}$$

We will consider lines (5.41) and (5.42) separately. We start with line (5.41) and in there, we start with the second integral: Note that the martingales $M_{n,i}$ have jumps of height exactly one at those positions where the counting processes $N_{n,i}$ jump (this is because we assume a continuous integrated intensity process). Hence we have

$$\int_{\{s\}} f_{n,ii}(s,t) dM_{n,i}(t) = \mathbb{1}_{\Delta N_{n,i}(s)=1} f_{n,ii}(s,s), \tag{5.43}$$

and furthermore

$$\int_0^T \int_{\{s\}} f_{n,ii}(s,t) dM_{n,i}(t) dM_{n,i}(s)$$

$$= \int_0^T \mathbb{1}_{\Delta N_{n,i}(s)=1} f_{n,ii}(s,s) dM_{n,i}(s) = \int_0^T f_{n,ii}(s,s) dN_{n,i}(s).$$

Using the above equality, we obtain

$$\begin{aligned}(5.41) \\ = \quad & \frac{1}{h^{\frac{1}{2}} r_n} \sum_{i \in G_n} \left[2 \int_0^T \int_0^{s-} f_{n,ii}(s,t) dM_{n,i}(t) dM_{n,i}(s) + \int_0^T f_{n,ii}(s,s) dN_{n,i}(s) \right]\end{aligned}$$

The first sum is a sum of uncorrelated martingales and so it will converge to zero in probability by an application of Markov's inequality: Denote by $g_{n,i}(s)$ a predictable function, then in general $\mathbb{E}\left(\int_0^T g_{n,i}(s) dM_{n,i}(s) \right) = 0$ and for $i \neq j$

$$\mathbb{E}\left(\int_0^T g_{n,i}(s) dM_{n,i}(s) \cdot \int_0^T g_{n,j}(s) dM_{n,j}(s) \right) = 0.$$

So we get for any $\varepsilon > 0$

$$\begin{aligned} & \mathbb{P}\left(\left| \frac{1}{r_n} \sum_{i \in G_n} \int_0^T g_{n,i}(s) dM_{n,i}(s) \right| > \varepsilon \right) \\ \leq \quad & \frac{1}{\varepsilon^2} \frac{1}{r_n^2} \sum_{i \in G_n} \mathbb{E}\left(\int_0^T g_{n,i}(s)^2 C_{n,i}(s) \lambda_{n,i}(\theta_0, s) ds \right). \end{aligned}$$

When letting $g_{n,i}(s) = h^{-\frac{1}{2}} \int_0^{s-} f_{n,ii}(s,t) dM_{n,i}(t)$, we have by Statement 4.4, that the above converges to zero. Moreover, we know that

$$\frac{1}{h^{\frac{1}{2}} r_n} \sum_{i \in G_n} \int_0^T f_{n,ii}(s,s) dN_{n,i}(s) - h^{-\frac{1}{2}} A_n \xrightarrow{\mathbb{P}} 0,$$

as n tends to infinity by Statement 4.5. Combining these considerations yields

$$(5.41) = o_p(1) + h^{-\frac{1}{2}} A_n.$$

Next we consider (5.42). Firstly, we note, using an analogue of (5.43), that the second integral in (5.42) equals zero because the two martingales $M_{n,i}$ and $M_{n,j}$ never jump simultaneously because $i \neq j$. To investigate the first integral we simplify notation by defining

$$\tau_{n,ij}(s) := \int_0^{s-} f_{n,ij}(s,t) dM_{n,j}(t). \tag{5.44}$$

Then $\tau_{n,ij}$ are predictable functions and so we find that

$$(5.42) = \frac{1}{h^{\frac{1}{2}} r_n} \sum_{i \in G_n} \int_0^T 2 \left(\sum_{\substack{j \in G_n \\ j \neq i}} \tau_{n,ij}(s) \right) dM_{n,i}(s)$$

is a martingale in T. Our aim is to show convergence to a normal distribution by using Rebolledo's martingale central limit theorem. To this end, we need to prove the convergence of the variation towards a deterministic quantity and the jump parts of the process converge to zero. We start with the quadratic variation (note that $M_{n,i}$ and $M_{n,j}$ are uncorrelated whenever $i \neq j$):

$$\begin{aligned}
&\left\langle \frac{1}{h^{\frac{1}{2}} r_n} \sum_{i\in G_n} \int_0^T 2 \left(\sum_{\substack{j\in G_n \\ j\neq i}} \tau_{n,ij}(s) \right) dM_{n,i}(s) \right\rangle \\
= &\frac{4}{hr_n^2} \sum_{i\in G_n} \int_0^T \left(\sum_{\substack{j\in G_n \\ j\neq i}} \tau_{n,ij}(s) \right)^2 C_{n,i}(s)\lambda_{n,i}(\theta_0, s)ds \\
= &\frac{4}{hr_n^2} \sum_{i\in G_n} \int_0^T \sum_{\substack{j_1,j_2\in G_n \\ j_1,j_2\neq i}} \tau_{n,ij_1}(s)\tau_{n,ij_2}(s) C_{n,i}(s)\lambda_{n,i}(\theta_0, s)ds \\
= &\frac{4}{hr_n^2} \sum_{i\in G_n} \int_0^T \sum_{\substack{j\in G_n \\ j\neq i}} \tau_{n,ij}(s)^2 C_{n,i}(s)\lambda_{n,i}(\theta_0, s)ds \\
&+ \frac{4}{hr_n^2} \sum_{i\in G_n} \int_0^T \sum_{\substack{j_1,j_2\in G_n \\ j_1,j_2\neq i \\ j_1\neq j_2}} \tau_{n,ij_1}(s)\tau_{n,ij_2}(s) C_{n,i}(s)\lambda_{n,i}(\theta_0, s)ds \\
\xrightarrow{\mathbb{P}} &\ B,
\end{aligned}$$

by Statement 4.6. Now the jump process (the process which contains all jumps of size greater than or equal to $\varepsilon > 0$) is given by (note that no two martingales jump at the same time)

$$\frac{2}{h^{\frac{1}{2}} r_n} \sum_{i\in G_n} \int_0^T \mathbb{1}\left(\left| \frac{2}{h^{\frac{1}{2}} r_n} \sum_{\substack{j\in G_n \\ j\neq i}} \tau_{n,ij}(s) \right| > \varepsilon \right) \sum_{\substack{j\in G_n \\ j\neq i}} \tau_{n,ij}(s) dM_{n,i}(s),$$

which converges to zero by Statement 4.7. Hence, by Rebolledo's martingale central limit theorem (cf. Theorem 2.12)

$$(5.42) \xrightarrow{d} N(0, B)$$

and the statement of the theorem is shown. □

Proof of Proposition 5.12. Let $\delta_n := \sqrt{\frac{\log r_n}{r_n h}}$, denote by $T_{n,k}$ the grid

$$T_{n,k} := \left\{ \frac{j}{hn^k} : j \in \mathbb{N}, \frac{j}{hn^k} \in [0, T] \right\}$$

and let $\pi_{n,k}(t)$ be the projection of $t \in [0,T]$ on $T_{n,k}$. Then $\#T_{n,k} \leq (T+1) \cdot hn^k$ and $|t - \pi_{n,k}(t)| \leq \frac{1}{hn^k}$. Using this projection we can estimate

$$\begin{aligned}
&\mathbb{P}\left(\left|\sup_{t_0 \in \mathbb{T}} \frac{1}{r_n\sqrt{\bar{p}_n(t_0)}} \ell_n'(\theta_0, t_0)\right| \geq C\delta_n\right) \\
\leq\ & \mathbb{P}\Bigg(\sup_{t_0 \in \mathbb{T}} \left|\frac{\ell_n'(\theta_0, t_0)}{r_n\sqrt{\bar{p}_n(t_0)}} - \frac{\ell_n'(\theta_0, \pi_{n,k}(t_0))}{r_n\sqrt{\bar{p}_n(\pi_{n,k}(t_0))}}\right| \\
&\quad + \sup_{t_0 \in \mathbb{T}} \left|\frac{\ell_n'(\theta_0, \pi_{n,k}(t_0))}{r_n\sqrt{\bar{p}_n(\pi_{n,k}(t_0))}}\right| \geq C\delta_n\Bigg) \\
\leq\ & \mathbb{P}\left(\sup_{\substack{t_0, s_0 \in \mathbb{T} \\ |s_0 - t_0| \leq hn^{-k}}} \left|\frac{\ell_n'(\theta_0, t_0)}{r_n\sqrt{\bar{p}_n(t_0)}} - \frac{\ell_n'(\theta_0, s_0)}{r_n\sqrt{\bar{p}_n(s_0)}}\right| \geq \frac{C}{2}\delta_n\right) && (5.45) \\
&+ \mathbb{P}\left(\sup_{t_0 \in T_{n,k}} \left|\frac{\ell_n'(\theta_0, t_0)}{r_n\sqrt{\bar{p}_n(t_0)}}\right| \geq \frac{C}{2}\delta_n\right). && (5.46)
\end{aligned}$$

We have to prove that both (5.45) and (5.46) converge to zero. We start with (5.45). Denote therefore $g_{n,i}(t, t_0) = K\left(\frac{t-t_0}{h}\right) \partial_\theta \log \lambda_{n,i}(\theta_0, t)$, then

$$\ell_n'(\theta_0, t_0) = \sum_{i \in G_n} \int_0^T \frac{1}{h} g_{n,i}(t, t_0) dM_{n,i}(t)$$

because $P_n'(\theta_0, t_0) = 0$ and Statement 4.1. Then we get

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\substack{t_0, s_0 \in \mathbb{T} \\ |s_0 - t_0| \leq hn^{-k}}} \left|\frac{\ell_n'(\theta_0, t_0)}{r_n\sqrt{\bar{p}_n(t_0)}} - \frac{\ell_n'(\theta_0, s_0)}{r_n\sqrt{\bar{p}_n(s_0)}}\right| \geq \frac{C}{2}\delta_n\right) \\
\leq &\mathbb{P}\left(\sup_{\substack{t_0, s_0 \in \mathbb{T} \\ |s_0 - t_0| \leq hn^{-k}}} \frac{1}{r_n h} \sum_{i \in G_n} \left|\int_0^T \frac{g_{n,i}(t, t_0)}{\sqrt{\bar{p}_n(t_0)}} dM_{n,i}(t) - \int_0^T \frac{g_{n,i}(t, s_0)}{\sqrt{\bar{p}_n(s_0)}} dM_{n,i}(t)\right| \geq \frac{C}{2}\delta_n\right) \\
\leq &\mathbb{P}\left(\frac{1}{r_n h} \sum_{i \in G_n} \int_0^T \sup_{\substack{t_0, s_0 \in \mathbb{T} \\ |s_0 - t_0| \leq hn^{-k}}} \left|\frac{g_{n,i}(t, t_0)}{\sqrt{\bar{p}_n(t_0)}} - \frac{g_{n,i}(t, s_0)}{\sqrt{\bar{p}_n(s_0)}}\right| dN_{n,i}(t) \geq \frac{C}{4}\delta_n\right) && (5.47) \\
+ &\mathbb{P}\left(\frac{1}{r_n h} \sum_{i \in G_n} \int_0^T \sup_{\substack{t_0, s_0 \in \mathbb{T} \\ |s_0 - t_0| \leq hn^{-k}}} \left|\frac{g_{n,i}(t, t_0)}{\sqrt{\bar{p}_n(t_0)}} - \frac{g_{n,i}(t, s_0)}{\sqrt{\bar{p}_n(s_0)}}\right| C_{n,i}(t) \lambda_{n,i}(\theta_0, t) dt \geq \frac{C}{4}\delta_n\right). \\
& && (5.48)
\end{aligned}$$

For (5.47) we apply Lenglart's inequality (cf. Corollary 2.11) to obtain for any choice of $c^* > 0$

$$\mathbb{P}\left(\frac{1}{r_n h}\sum_{i\in G_n}\int_0^T \sup_{\substack{t_0,s_0\in\mathbb{T}\\|s_0-t_0|\le hn^{-k}}}\left|\frac{g_{n,i}(t,t_0)}{\sqrt{\bar{p}_n(t_0)}}-\frac{g_{n,i}(t,s_0)}{\sqrt{\bar{p}_n(s_0)}}\right| dN_{n,i}(t) \ge \frac{C}{4}\delta_n\right)$$
$$\le \frac{c^*}{C} + \mathbb{P}\left(\frac{1}{r_n h}\sum_{i\in G_n}\int_0^T \sup_{\substack{t_0,s_0\in\mathbb{T}\\|s_0-t_0|\le hn^{-k}}}\left|\frac{g_{n,i}(t,t_0)}{\sqrt{\bar{p}_n(t_0)}}-\frac{g_{n,i}(t,s_0)}{\sqrt{\bar{p}_n(s_0)}}\right| C_{n,i}(t)\lambda_{n,i}(\theta_0,t)dt \ge \frac{c^*}{4}\delta_n\right).$$

If we restrict to $c^* < C$ we obtain furthermore

$$(5.47) + (5.48)$$
$$\le \frac{c^*}{C} + 2\mathbb{P}\left(\frac{1}{r_n h}\sum_{i\in G_n}\int_0^T \sup_{\substack{t_0,s_0\in\mathbb{T}\\|s_0-t_0|\le hn^{-k}}}\left|\frac{g_{n,i}(t,t_0)}{\sqrt{\bar{p}_n(t_0)}}-\frac{g_{n,i}(t,s_0)}{\sqrt{\bar{p}_n(s_0)}}\right| C_{n,i}(t)\lambda_{n,i}(\theta_0,t)dt \ge \frac{c^*}{4}\delta_n\right). \tag{5.49}$$

Using the Assumptions (B4), 2 and (B6) (note that for any $x, y \ge 0$, $|\sqrt{x}-\sqrt{y}| \le \sqrt{|x-y|}$ implies that $\sqrt{\frac{1}{\bar{p}_n(t_0)}}$ is Hoelder continuous with exponent $\frac{\alpha_p}{2}$ and constant $\sqrt{H_{n,p}}$), we get by using $\sup_{t_0\in\mathbb{T}}\frac{1}{\sqrt{\bar{p}_n(t_0)}} \le \frac{1}{\sqrt{p_n}}$

$$\begin{aligned}
&\left\|\frac{g_{n,i}(t,t_0)}{\sqrt{\bar{p}_n(t_0)}}-\frac{g_{n,i}(t,s_0)}{\sqrt{\bar{p}_n(s_0)}}\right\| \\
=\ & \|\partial_\theta \log\lambda_{n,i}(\theta_0,t)\|\cdot\left|\frac{1}{\sqrt{\bar{p}_n(t_0)}}K\left(\frac{t-t_0}{h}\right)-\frac{1}{\sqrt{\bar{p}_n(s_0)}}K\left(\frac{t-s_0}{h}\right)\right| \\
\le\ & \|\partial_\theta \log\lambda_{n,i}(\theta_0,t)\|\cdot\left[\frac{1}{\sqrt{\bar{p}_n(t_0)}}\left|K\left(\frac{t-t_0}{h}\right)-K\left(\frac{t-s_0}{h}\right)\right|\right. \\
&\qquad\left.+K\left(\frac{s-s_0}{h}\right)\left|\frac{1}{\sqrt{\bar{p}_n(t_0)}}-\frac{1}{\sqrt{\bar{p}_n(s_0)}}\right|\right] \\
\le\ & \|\partial_\theta \log\lambda_{n,i}(\theta_0,t)\|\left[\frac{1}{\sqrt{p_n}}\cdot H_K n^{-k\alpha_K}+K\cdot\sqrt{H_{n,p}}n^{-k\cdot\frac{\alpha_p}{2}}\right] \\
\le\ & \|\partial_\theta \log\lambda_{n,i}(\theta_0,t)\|\cdot\frac{1}{\sqrt{p_n}}\left[H_K+K\sqrt{H_{n,p}p_n}\right]n^{-k\cdot\min(\alpha_K,\alpha_p/2)},
\end{aligned}$$

where K is the bound on the kernel from Assumption (B4), 2. So we get

$$\frac{1}{r_n h}\sum_{i\in G_n}\int_0^T \sup_{\substack{t_0,s_0\in\mathbb{T}\\|s_0-t_0|\le hn^{-k}}}\left\|\frac{g_{n,i}(t,t_0)}{\sqrt{\bar{p}_n(t_0)}}-\frac{g_{n,i}(t,s_0)}{\sqrt{\bar{p}_n(s_0)}}\right\| C_{n,i}(t)\lambda_{n,i}(\theta_0,t)dt$$

$$\leq \frac{1}{r_n h} \sum_{i \in G_n} \int_0^T \|\partial_\theta \log \lambda_{n,i}(\theta_0, t)\| C_{n,i}(t) \lambda_{n,i}(\theta_0, t) dt$$
$$\times \frac{1}{\sqrt{p_n}} \left[H_K + K\sqrt{H_{n,p}p_n} \right] n^{-k \cdot \min(\alpha_K, \alpha_p/2)}.$$

Hence, we get that (5.49) is small, because by Statement 4.10 we can choose $k = k_0$ such that for large enough c^* the probability is small for all $n \in \mathbb{N}$ and then we can choose C large enough such that the whole expression is small. Then, also (5.45) is small, for this good choice $k = k_0$ which we keep fixed from now on.

Let us now turn to (5.46). Here we take the supremum over a finite set and so we can estimate by applying union bound and Statement 4.11 for $C > 0$ large enough

$$\overset{(5.46)}{\leq} \#T_{n,k_0} \cdot \sup_{t_0 \in T_{n,k_0}} \mathbb{P}\left(\left| \frac{\ell_n'(\theta_0, t_0)}{r_n \sqrt{\bar{p}_n(t_0)}} \right| \geq \frac{C}{2} \delta_n \right) \to 0.$$

□

For the proof of Proposition 5.13 we require an auxiliary lemma.

Lemma 5.15. *Under the same assumptions as in Theorem 4.3 we have*

$$\sup_{t_0 \in \mathbb{T}} \left\| \theta_0 - \hat{\theta}_{t_0} \right\| = O_p\left(\sqrt{\frac{\log r_n}{r_n p_n h}} \right).$$

Proof. This result is Proposition 5.8 but uniformly. The proof is therefore entirely analogue, we just have to to make sure that Kantorovich's Theorem can be applied uniformly. By Statements 4.8 and 4.9 we have that for any choice of $t_0 \in [\delta, T - \delta]$

$$\left\| \left[\frac{1}{r_n \bar{p}_n(t_0)} \ell_n''(\theta_0, t_0) \right]^{-1} \right\| \leq B_n$$
$$\frac{1}{r_n \bar{p}_n(t_0)} \left\| \ell_n''(\theta_1, t_0) - \ell_n''(\theta_2, t_0) \right\| \leq K_n \cdot \|\theta_1 - \theta_2\|,$$

where $B_n, K_n = O_P(1)$. And therefore, use also Proposition 5.12, there is η_n such that (recall also that $\frac{\sqrt{p_n}}{\sqrt{\bar{p}_n(t_0)}} \leq 1$ for all $t_0 \in \mathbb{T}$)

$$\eta_n := \sup_{t_0 \in \mathbb{T}} \left\| \left[\frac{1}{r_n \bar{p}_n(t_0)} \ell_n''(\theta_0, t_0) \right]^{-1} \frac{1}{r_n \bar{p}_n(t_0)} \ell_n'(\theta_0, t_0) \right\| = O_P\left(\sqrt{\frac{\log r_n}{p_n r_n h}} \right).$$

Hence, we can apply Kantorovich's Theorem (cf. Theorem 5.9) for all $t_0 \in \mathbb{T}$ with the same choice of B_n, K_n and η_n as above. Thus, there is $\hat{\theta}(t_0)$ such that for all t_0

$$\left\| \theta_0 - \hat{\theta}(t_0) \right\| \leq 2\eta_n = O_P\left(\sqrt{\frac{\log r_n}{r_n p_n h}} \right).$$

This yields the statement. □

Proof of Proposition 5.13. Define and recall

$$
\begin{aligned}
H_{n,i}(s,\theta) &:= \left[\partial_\theta^2 \log \lambda_{n,i}(\theta,s)\cdot\lambda_{n,i}(\theta_0,s) - \partial_\theta^2\lambda_{n,i}(\theta,s)\right] C_{n,i}(s),\\
\Sigma(s,\theta) &= \mathbb{E}\left[H_{n,1}(s,\theta)|C_{n,1}(s)=1\right]\\
\widetilde{H}_{n,i}(s,\theta) &:= H_{n,i}(s,\theta)-\Sigma(s,\theta)p_n(s).
\end{aligned}
$$

Note that the choice of index 1 in the definition of Σ is arbitrary by assumption (B1), 1. Note firstly that

$$
\left\|\frac{1}{r_n\bar{p}_n(t_0)}\ell_n^*(t_0)-\Sigma(t_0)\right\|^2 \le \sum_{r=1}^{q}\left\|\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_r^*(t_0),t_0)-\Sigma(t_0)\right\|^2. \tag{5.50}
$$

Since q doesn't vary in n, it is enough to consider each term in the sum on the right hand side above separately. In order to reduce notation, we do not indicate which intermediate value $\theta_r^*(t_0)$ we consider and write $\theta^*(t_0)$ instead. Now, we can separate the problem as follows:

$$
\begin{aligned}
&\left\|\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta^*(t_0),t_0)-\Sigma(t_0)\right\|\\
\le\ &\left\|\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}\int_0^T\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\left[H_{n,i}(s,\theta^*(t_0))-\Sigma(s,\theta^*(t_0))p_n(s)\right]ds\right\|\\
&+\left\|\frac{1}{r_n}\sum_{i\in G_n}\int_0^T\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\left[\Sigma(s,\theta^*(t_0))\frac{p_n(s)}{\bar{p}_n(t_0)}-\Sigma(t_0)\right]ds\right\|\\
\le\ &\sup_{\theta\in\Theta}\left\|\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}\int_0^T\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\left[H_{n,i}(s,\theta)-\Sigma(s,\theta)p_n(s))\right]ds\right\| && (5.51)\\
&+\left\|\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)\left(\Sigma(s,\theta^*(t_0))-\Sigma(s,\theta_0)\right)\frac{p_n(s)}{\bar{p}_n(t_0)}ds\right\| && (5.52)\\
&+\left\|\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)\left(\Sigma(s,\theta_0)-\Sigma(t_0,\theta_0)\right)\frac{p_n(s)}{\bar{p}_n(t_0)}ds\right\| && (5.53)\\
&+\left\|\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)\Sigma(t_0,\theta_0)\left(\frac{p_n(s)}{\bar{p}_n(t_0)}-1\right)ds\right\|. && (5.54)
\end{aligned}
$$

We note firstly that (5.54) $=0$. Moreover, after taking the sup over all t_0, the convergence rate of line (5.52) equals $O_P\left(\sqrt{\frac{\log r_n}{r_n p_n h}}\right)$, because of the Lipschitz continuity of Σ in Statement 4.2 and Lemma 5.15 (recall that $\theta^*(t_0)$ is the intermediate value between $\hat{\theta}(t_0)$ and θ_0 in Taylor's Formula). The expression in (5.53) can be handled by the differentiability assumption in (B7) together with a Taylor expansion in the time parameter.:

$$
\sup_{t_0\in[0,T]}\left\|\int_0^T\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\left(\Sigma(s,\theta_0)-\Sigma(t_0,\theta_0)\right)\frac{p_n(s)}{\bar{p}_n(t_0)}ds\right\|
$$

$$\leq \sup_{t_0\in[0,T]} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \|\partial_t \Sigma\|_\infty \cdot |s-t_0| \frac{p_n(s)}{\bar{p}_n(t_0)} ds$$
$$\leq h \cdot \|\partial_t \Sigma\|_\infty,$$

where we used in the last step that the kernel is supported on $[-1,1]$ and hence $|s-t_0| \leq h$. So (5.53) is of order h.

To deal with the first expression, line (5.51), we let $\delta_n := \sqrt{\frac{\log r_n p_n}{r_n p_n \cdot h}}$ and $C > 0$ and denote by (here k_0 is the constant from Statement 4.10)

$$T_{n,k_0} := \left\{ \left(\frac{j}{hn^{k_0}}, \frac{j_1}{n^{k_0}}, ..., \frac{j_q}{n^{k_0}} \right) \in \mathbb{T} \times \Theta : \, j, j_1, ..., j_q \in \mathbb{Z} \right\}$$

a discrete grid covering $\mathbb{T} \times \Theta$ and denote by $\pi_{n,k_0} : \mathbb{T} \times \Theta \to T_{n,k_0}$ the corresponding projection. Then, we have

$$\mathbb{P}\left(\sup_{\substack{t_0\in\mathbb{T} \\ \theta\in\Theta}} \left\| \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \widetilde{H}_{n,i}(s,\theta) ds \right\| > C\delta_n \right)$$

$$\leq \mathbb{P}\left(\sup_{\substack{t_0\in\mathbb{T},\theta\in\Theta \\ |t_0-t_0'|\leq hn^{-k_0}, \\ \|\theta_1-\theta_2\|\leq n^{-k_0}}} \left\| \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \widetilde{H}_{n,i}(s,\theta_1) ds \right.\right.$$
$$\left.\left. - \frac{1}{r_n \bar{p}_n(t_0')} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0'}{h}\right) \widetilde{H}_{n,i}(s,\theta_2) ds \right\| > \frac{C}{2}\delta_n \right) \tag{5.55}$$

$$+\mathbb{P}\left(\sup_{(t_0,\theta)\in T_{n,k_0}} \left| \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \widetilde{H}_{n,i}(s,\theta) ds \right| > \frac{C}{2}\delta_n \right) \tag{5.56}$$

In order to show that (5.55) converges to zero, we note that for $|t_0-t_0'| \leq hn^{-k_0}$, $|\theta_1-\theta_2| \leq n^{-k_0}$, we get by assumptions (B4), 2 and (B6)

$$\frac{1}{r_n} \sum_{i\in G_n} \int_0^T \left\| \frac{1}{\bar{p}_n(t_0)} \cdot \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \widetilde{H}_{n,i}(s,\theta_1) - \frac{1}{\bar{p}_n(t_0')} \cdot \frac{1}{h} K\left(\frac{s-t_0'}{h}\right) \widetilde{H}_{n,i}(s,\theta_2) \right\| ds$$
$$\leq \frac{1}{r_n} \sum_{i\in G_n} \int_0^T \frac{1}{\bar{p}_n(t_0)} \cdot \frac{1}{h} \left| K\left(\frac{s-t_0}{h}\right) - K\left(\frac{s-t_0'}{h}\right) \right| \cdot \|\widetilde{H}_{n,i}(s,\theta_1)\|$$
$$+ \left| \frac{1}{\bar{p}_n(t_0)} - \frac{1}{\bar{p}_n(t_0')} \right| \cdot \frac{1}{h} K\left(\frac{s-t_0'}{h}\right) \cdot \|\widetilde{H}_{n,i}(s,\theta_1)\|$$
$$+ \frac{1}{\bar{p}_n(t_0')} \cdot \frac{1}{h} K\left(\frac{s-t_0'}{h}\right) \left\| \widetilde{H}_{n,i}(s,\theta_1) - \widetilde{H}_{n,i}(s,\theta_2) \right\| ds$$

$$\leq \frac{H_K}{p_n} \cdot \frac{1}{h} n^{-k_0 \alpha_K} \frac{1}{r_n} \sum_{i \in G_n} \int_0^T \|\widetilde{H}_{n,i}(s,\theta_1)\| ds$$
$$+ H_{n,p} h^{-\alpha_p} n^{-k_0 \alpha_p} \cdot \frac{1}{r_n} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0'}{h}\right) \|\widetilde{H}_{n,i}(s,\theta_1)\| ds$$
$$+ \frac{1}{p_n} \cdot \frac{1}{r_n} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0'}{h}\right) \gamma_{n,i}(s) n^{-k_0 p} ds,$$

which has the correct order by Statement 4.10.

(5.56) converges to zero by Statement 4.12.

To prove that inversion preserves the rate, we denote

$$X_n(t_0) := \frac{1}{r_n \bar{p}_n(t_0)} \ell_n^*(t_0).$$

Then

$$X_n(t_0)^{-1} - \Sigma(t_0,\theta_0)^{-1} = X_n(t_0)^{-1} \left(\Sigma(t_0,\theta_0) - X_n(t_0)\right) \Sigma(t_0,\theta_0)^{-1}.$$

Since we assume in (B7) that the two inverses above are uniformly bounded, we find that the difference of the inverses has the same rate as the difference of the matrices (without inverse). The inverse of $X_n(t_0)$ exists with large probability because we have convergence towards $\Sigma(t_0,\theta_0)$ which lies well within the set of invertible matrices (cf. Assumption (B7)). $\square$

Proof of Proposition 5.14. To begin with, we note that we may work element-wise because the dimension remains fixed during the asymptotics. Therefore in the following $\ell_n'(\theta_0,t_0)$, $\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n^*(t_0)\right]^{-1}$ and $\Sigma^{-1}(t_0)$ should be understood as one entry from the corresponding matrix or vector. For ease of notation we do not indicate which entry we consider, however we consider always the same entry of $\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n^*(t_0)\right]^{-1}$ and $\Sigma^{-1}(t_0)$. We use firstly the Taylor expansion from (5.38) and the Cauchy-Schwarz Inequality to get

$$\begin{aligned}
&\left|\sqrt{r_n h} \int_0^T \left(\hat{\theta}_n(t_0) - \theta_0\right) \frac{\bar{p}_n(t_0)}{\sqrt{\bar{p}_n}} w(t_0) dt_0\right| \\
\leq &\left|\int_0^T \left(\Sigma^{-1}(t_0) - \left[\frac{1}{r_n \bar{p}_n(t_0)} \ell_n^*(t_0)\right]^{-1}\right) \cdot \sqrt{\frac{h}{r_n \bar{p}_n}} \ell_n'(\theta_0,t_0) w(t_0) dt_0\right| \\
&+ \left|\int_0^T \Sigma^{-1}(t_0) \cdot \sqrt{\frac{h}{r_n \bar{p}_n}} \ell_n'(\theta_0,t_0) w(t_0) dt_0\right| \\
\leq &\left(\int_0^T \left(\Sigma^{-1}(t_0) - \left[\frac{1}{r_n \bar{p}_n(t_0)} \ell_n^*(t_0)\right]^{-1}\right)^2 w(t_0) dt_0 \cdot \int_0^T \frac{h}{r_n \bar{p}_n} \ell_n'(\theta_0,t_0)^2 w(t_0) dt_0\right)^{\frac{1}{2}}
\end{aligned} \tag{5.57}$$

$$+\left|\int_0^T \Sigma^{-1}(t_0)\cdot\sqrt{\frac{h}{r_n\bar{p}_n}}\ell_n'(\theta_0,t_0)w(t_0)dt_0\right|. \tag{5.58}$$

We show now that (5.57) and (5.58) are both $o_P(1)$. We begin with (5.57). Let $\varepsilon,\eta>0$ be arbitrary, then for any $C>0$

$$\begin{aligned}
&\mathbb{P}\left((5.57)>\varepsilon\right)\\
\leq&\mathbb{P}\left(\int_0^T \frac{h}{r_n\bar{p}_n}\ell_n'(\theta_0,t_0)^2w(t_0)dt_0>\frac{\varepsilon^2}{C^2}\right)\\
&+\mathbb{P}\left(\int_0^T\left(\Sigma^{-1}(t_0)-\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n^*(t_0)\right]^{-1}\right)^2w(t_0)dt_0>C^2\right).
\end{aligned} \tag{5.59}$$

By Markov's Inequality and the fact that the counting process martingales are uncorrelated we obtain for $h\leq\frac{\delta}{2}$ (recall that supp$w\subseteq[\delta,T-\delta]$ by Assumption (B3))

$$\begin{aligned}
&\mathbb{E}\left(\int_0^T\frac{h}{r_n\bar{p}_n}\ell_n'(\theta_0,t_0)^2w(t_0)dt_0\right)\\
=&\frac{h}{r_n\bar{p}_n}\sum_{i\in G_n}\int_0^T\int_0^T\frac{1}{h^2}K\left(\frac{t-t_0}{h}\right)^2\mathbb{E}\left(X_{n,i}(t)^2C_{n,i}(t)\lambda_{n,i}(\theta_0,t)\right)dtw(t_0)dt_0\\
\leq&\int_h^{T-h}\int_\delta^{T-\delta}\frac{1}{h}K\left(\frac{t-t_0}{h}\right)^2w(t_0)dt_0\mathbb{E}\left(X_{n,i}(t)^2\lambda_{n,i}(\theta_0,t)\middle|C_{n,i}(t)=1\right)\frac{p_n(t)}{\bar{p}_n}dt\\
\leq&C^*
\end{aligned}$$

by Assumption (B1), 4 and the fact that $\int_h^{T-h}p_n(t)dt\cdot\bar{p}_n^{-1}\leq1$. By Proposition 5.13 we find that for all $C>0$ and thus in particular $C=\frac{\varepsilon\sqrt{\eta}}{\sqrt{2C^*}}$ it holds for n large enough that

$$\mathbb{P}\left(\int_0^T\left(\Sigma^{-1}(t_0)-\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n^*(t_0)\right]^{-1}\right)^2w(t_0)dt_0>C^2\right)\leq\frac{\eta}{2}.$$

Now, by using all previous considerations we may estimate by using (5.59) for n large enough

$$\mathbb{P}((5.57)>\varepsilon)\leq\frac{C^2}{\varepsilon^2}\mathbb{E}\left(\int_0^T\frac{h}{r_n\bar{p}_n}\ell_n'(\theta_0,t_0)^2w(t_0)dt_0\right)+\frac{\eta}{2}\leq\eta.$$

Since $\varepsilon,\eta>0$ were chosen arbitrarily, we have shown that $(5.57)=o_P(1)$.

We continue with (5.58). This is easier to handle because $\Sigma^{-1}(t_0)$ is deterministic and thus in particular predictable: Let $\varepsilon>0$ be arbitrary

$$\begin{aligned}
&\mathbb{P}\left((5.58)>\varepsilon\right)\\
\leq&\frac{h}{\varepsilon^2r_n\bar{p}_n}\mathbb{E}\left(\left(\sum_{i\in G_n}\int_0^T\int_0^T\Sigma^{-1}(t_0)\frac{1}{h}K\left(\frac{t-t_0}{h}\right)w(t_0)dt_0X_{n,i}(t)dM_{n,i}(t)\right)^2\right)
\end{aligned}$$

$$= \frac{h}{\varepsilon^2 r_n \bar{p}_n} \sum_{i \in G_n} \int_0^T \mathbb{E}\left(\left(\int_0^T \frac{1}{h} K\left(\frac{t-t_0}{h}\right) \Sigma^{-1}(t_0) w(t_0) dt_0 \right)^2 X_{n,i}(t)^2 C_{n,i}(t) \lambda_{n,i}(\theta_0, t) \right) dt$$

$$\le \frac{hM^2}{\varepsilon^2 \bar{p}_n} \int_0^T \mathbb{E}\left(X_{n,i}(t)^2 \lambda_{n,i}(\theta_0, t) \middle| C_{n,i}(t) = 1 \right) p_n(t) dt.$$

Since $h \to 0$, this converges to zero by Assumption (B1), 4 and thus also (5.58) $= o_P(1)$. And this is what we wanted to prove. □

5.3 Proofs of Statements 4.1 - 4.12

In this Section we will use Fubini, i.e., interchange the order of integration and/or expectation because we assume in Assumption (C1), 2 that the covariates and thus all functions which appear in our context are bounded. Thus all joint integrals exist and we may compute them iteratively in any order. We will not explicitly refer to Assumption (C1), 2 every time we do this.

Proposition 5.16. *Suppose that Assumptions (C1), 1 and 2 hold. Then, Statement 4.1 holds true.*

Proof. By the form of the Cox intensity function in Assumption (C1), 1 the integrands in ℓ_n and P_n are are twice continuously differentiable with respect to θ. Integral and differential in P_n may be interchanged by boundedness of the integrand, c.f. Assumption (C1), 2. Note for ℓ_n that the stochastic integral is comprised of a regular integral and an integral with respect to the counting process which is actually a sum. For the sum integral and differential may obviously be interchanged and for the regular integral, the boundedness guarantees it. □

Proposition 5.17. *Suppose that Assumptions (C1), 1 and 2 and (B5), 1 hold. Then, Statement 4.2 holds true.*

Proof. We have by Assumption (C1), 1

$$\Sigma(\theta, s) = \mathbb{E}\left[X_{n,i}(s) X_{n,i}(s)^T e^{\theta^T X_{n,i}(s)} \middle| C_{n,i}(s) = 1 \right]$$

and hence a Taylor approximation in the argument of the exponential function yields the desired Lipschitz continuity with constant $\gamma_\Sigma(s) \equiv \hat{K}^3 e^{\tau \hat{K}}$, where $\hat{K}$ is the bound on X and τ is the bound on θ from Assumptions (C1), 2 and (B5), 1 respectively. The second assertion is then an easy consequence. □

Proposition 5.18. *Suppose that Assumptions (C1), 2, (B5), 1 and (B7) hold. Then, Statement 4.3 holds true.*

Proof. We note that the stochastic integrals can be understood as an integral with respect to the counting process (which is really a sum) minus a regular integral. The sums may be interchanged without trouble. For the regular integrals we may apply Fubini because the intensities are bounded by Assumptions (C1), 2 and (B5), 1. Furthermore, $\Sigma(t_0)^{-1}$ exists and is bounded by Assumption (B7). □

Proposition 5.19. *Suppose that Assumptions (C1), 2, (B1), 1, (B3) (B4), 2 and (B7) hold. Then, Statement 4.4 holds true.*

Proof. Since everything is identically distributed by Assumption (B1), 1, we can estimate by the Assumptions (C1), 2 and (B4), 2 and (B7). Moreover, the kernel integrates to one and hence we may also apply Jensen's Inequality as follows

$$
\begin{aligned}
&\frac{1}{r_n^2}\sum_{i\in G_n}\mathbb{E}\left(\int_0^T g_{n,i}(s)^2 C_{n,i}(s)\lambda_{n,i}(\theta_0,s)ds\right)\\
\leq&\frac{\Lambda}{r_n}\int_0^T\frac{1}{h}\mathbb{E}\left(\left(\int_0^{s-} f_{n,ii}(s,t)dM_{n,i}(t)\right)^2\right)ds\\
=&\frac{\Lambda}{r_n}\int_0^T\frac{1}{h}\int_0^s\mathbb{E}\left(f_{n,ii}(s,t)^2 C_{n,i}(t)\lambda_{n,i}(t)\right)dtds\\
\leq&\frac{\hat{K}^4\Lambda^2}{r_n}\sup_{t_0}\|\Sigma^{-T}(t_0)\Sigma^{-1}(t_0)\|^2\\
&\times\int_0^T\frac{1}{h}\int_0^s\left(\int_\delta^{T-\delta}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)K\left(\frac{t-t_0}{h}\right)\frac{w(t_0)}{\bar{p}_n(t_0)}dt_0\right)^2 p_n(t)dtds\\
\leq&\frac{\hat{K}^4\Lambda^2M^2}{r_n}\int_0^T\frac{1}{h}\int_0^T\int_\delta^{T-\delta}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)K\left(\frac{t-t_0}{h}\right)^2\frac{w(t_0)^2p_n(t)}{\bar{p}_n(t_0)^2}dt_0dtds\\
\leq&\frac{\hat{K}^4\Lambda^2M^2}{r_np_n}\cdot\frac{1}{h}\int_0^T\int_\delta^{T-\delta}K\left(\frac{t-t_0}{h}\right)^2\frac{w(t_0)^2p_n(t)}{\bar{p}_n(t_0)}dt_0dt\\
\leq&\frac{\hat{K}^4\Lambda^2M^2T}{r_np_n}\sup_{t_0\in\mathbb{T}}\int_0^T\frac{1}{h}K\left(\frac{t-t_0}{h}\right)\frac{p_n(t)}{\bar{p}_n(t_0)}dt\cdot w(t_0)
\end{aligned}
$$

But this converges to zero by boundedness of w, cf. (B3). □

Proposition 5.20. *Suppose that Assumption (C1), 1 holds. Then, Statement 4.5 holds true.*

Proof. Since $\lambda_{n,i}(\theta,s)=\exp(\theta^T X_{n,i}(s))$ we find that $\frac{\partial_\theta\lambda_{n,i}(\theta_0,s)}{\lambda_{n,i}(\theta_0,s)}=X_{n,i}(s)$ is independent of θ_0 and thus computable. Hence, we can compute

$$
\frac{1}{r_n}\sum_{i\in G_n}\int_0^T f_{n,ii}(s,s)dN_{n,i}(s) \tag{5.60}
$$

and thus we may choose $A_n=$ (5.60) and Statement 4.5 holds trivially. □

Proposition 5.21. *Suppose that Assumptions (B4), 2 and (B7) as well as (C1), (C2), (C3), (C4*), (C5), (C6) (measurability and (4.19)-(4.27)), in (C7) equations 4.33 and (4.34)) and (C8) hold. Then, Statement 4.6 holds true.*

Proof. We begin by recalling that $\tau_{n,ij}(s) = \int_0^{s-} f_{n,ij}(s,t)dM_{n,j}(t)$ with

$$f_{n,ij}(s,t) = X_{n,i}(s)^T \int_\delta^{T-\delta} \frac{1}{h} K\left(\frac{s-t_0}{h}\right) K\left(\frac{t-t_0}{h}\right) \Sigma(t_0)^{-T}\Sigma(t_0)^{-1}\frac{w(t_0)}{\bar{p}_n(t_0)}dt_0 X_{n,j}(t).$$

By substituting $u = \frac{s-t_0}{h}$ we obtain

$$\begin{aligned}
&f_{n,ij}(s,t)\\
=&X_{n,i}(s)^T \int_{\frac{s-T+\delta}{h}}^{\frac{s-\delta}{h}} K(u)K\left(\frac{t-s}{h}+u\right)\Sigma^{-T}(s-uh)\Sigma(s-uh)^{-1}\frac{w(s-uh)}{\bar{p}_n(s-uh)}duX_{n,j}(t).
\end{aligned}$$

Denote finally

$$\widetilde{f}_n(s,t) := \int_{\frac{s-T+\delta}{h}}^{\frac{s-\delta}{h}} K(v)K\left(\frac{t-s}{h}+v\right)\Sigma^{-T}(s-vh)\Sigma(s-vh)^{-1}\frac{w(s-vh)}{\bar{p}_n(s-vh)}dv.$$

Note firstly that $\widetilde{f}_n$ is not random and secondly that for $t < s-2h$ the above expression equals zero because we assume that the kernel is supported on $[-1,1]$ (cf. Assumption (B4), 2). So we obtain for $\tau_{n,ij}$:

$$\tau_{n,ij}(s) = X_{n,i}(s)^T \int_{s-2h}^{s-} \widetilde{f}_n(s,t)X_{n,j}(t)dM_{n,j}(t) \tag{5.61}$$

Now, let us consider the second asserted convergence. For ease of notation we define $\lambda_{n,i}(s) := \lambda_{n,i}(\theta_0, s)$. By using the representation of $\tau_{n,ij}$ in (5.61) we obtain

$$\begin{aligned}
&\frac{4}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T \tau_{n,ij_1}(s)\tau_{n,ij_2}(s)C_{n,i}(s)\lambda_{n,i}(s)ds\\
=&\frac{4}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T X_{n,i}(s)^T\int_{s-2h}^{s-}\widetilde{f}_n(s,t)X_{n,j_1}(t)dM_{n,j_1}(t)\\
&\times\left(\int_{s-2h}^{s-}\widetilde{f}_n(s,t)X_{n,j_2}(t)dM_{n,j_2}(t)\right)^T X_{n,i}(s)C_{n,i}(s)\lambda_{n,i}(s)ds
\end{aligned}$$

The integrals in the previous display are over vector-valued integrands and are to be understood element-wise. In order to study the behaviour of these integrals, we write the product of the two integrals as a sum. The equation from before continues

$$= \frac{4}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T X_{n,i}(s)^T\int_{s-2h}^{s-}\widetilde{f}_n(s,t)X_{n,j_1}(t)dM_{n,j_1}(t)$$

$$\times \int_{s-2h}^{s-} X_{n,j_2}^T(t)\widetilde{f}_n(s,t)^T(t)dM_{n,j_2}(t)X_{n,i}(s)C_{n,i}(s)\lambda_{n,i}(s)ds$$

$$= \frac{4}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T\int_0^T\int_0^T \mathbb{1}_{t<s}\mathbb{1}_{t\geq s-2h}\mathbb{1}_{r<s}\mathbb{1}_{r\geq s-2h}X_{n,i}(s)^T\widetilde{f}_n(s,t)X_{n,j_1}(t)$$

$$\times X_{n,j_2}(r)^T\widetilde{f}_n(s,r)^TX_{n,i}(s)C_{n,i}(s)\lambda_{n,i}(s)dsdM_{n,j_2}(r)dM_{n,j_1}(t)$$

$$= \frac{4}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T\int_0^T\int_{\max(t,r)}^{\min(t,r)+2h} X_{n,i}(s)^T\widetilde{f}_n(s,t)X_{n,j_1}(t)$$

$$\times X_{n,j_2}(r)^T\widetilde{f}_n(s,r)^TX_{n,i}(s)C_{n,i}(s)\lambda_{n,i}(s)dsdM_{n,j_2}(r)dM_{n,j_1}(t)$$

$$= \frac{4}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T\int_0^T(\mathbb{1}_{r<t}+\mathbb{1}_{r>t})\int_{\max(t,r)}^{\min(t,r)+2h} X_{n,i}(s)^T\widetilde{f}_n(s,t)X_{n,j_1}(t)$$

$$\times X_{n,j_2}(r)^T\widetilde{f}_n(s,r)^TX_{n,i}(s)C_{n,i}(s)\lambda_{n,i}(s)dsdM_{n,j_2}(r)dM_{n,j_1}(t)$$

Note that we do not need the indicator $\mathbb{1}_{t=r}$ because $j_1\neq j_2$ and hence martingales M_{n,j_1} and M_{n,j_2} will not jump simultaneously almost surely. We continue (for the second equality interchange the roles of j_1 and j_2 as well as the roles of t and r

$$=\frac{4}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T\int_0^{t-}\int_{\max(t,r)}^{\min(t,r)+2h} X_{n,i}(s)^T\widetilde{f}_n(s,t)X_{n,j_1}(t)$$

$$\times X_{n,j_2}(r)^T\widetilde{f}_n(s,r)^TX_{n,i}(s)C_{n,i}(s)\lambda_{n,i}(s)dsdM_{n,j_2}(r)dM_{n,j_1}(t)$$

$$+\frac{4}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T\int_0^{r-}\int_{\max(t,r)}^{\min(t,r)+2h} X_{n,i}(s)^T\widetilde{f}_n(s,t)X_{n,j_1}(t)$$

$$\times X_{n,j_2}(r)^T\widetilde{f}_n(s,r)^TX_{n,i}(s)C_{n,i}(s)\lambda_{n,i}(s)dsdM_{n,j_1}(t)dM_{n,j_2}(r)$$

$$=\frac{8}{hr_n^2}\sum_{i\in G_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1,j_2\neq i\\ j_1\neq j_2}}\int_0^T\int_{t-2h}^{t-}\int_t^{r+2h} X_{n,i}(s)^T\widetilde{f}_n(s,t)X_{n,j_1}(t)$$

$$\times X_{n,j_2}(r)^T\widetilde{f}_n(s,r)^TX_{n,i}(s)C_{n,i}(s)\lambda_{n,i}(s)dsdM_{n,j_2}(r)dM_{n,j_1}(t)$$

$$=\frac{8}{r_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}}\int_0^T\int_{t-2h}^{t-}\int_0^{\frac{r-t}{h}+2}\frac{1}{r_n}\sum_{\substack{i\in G_n\\ i\neq j_1,j_2}}X_{n,i}(t+uh)^T\widetilde{f}_n(t+uh,t)X_{n,j_1}(t)X_{n,j_2}(r)^T$$

$$\times \widetilde{f}_n(t+uh,r)^T X_{n,i}(t+uh)C_{n,i}(t+uh)\lambda_{n,i}(t+uh)dudM_{n,j_2}(r)dM_{n,j_1}(t)$$
$$=\frac{8}{r_n}\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}}\int_0^T\int_{t-2h}^{t-}\varphi_{n,j_1j_2}(t,r)dM_{n,j_2}(r)dM_{n,j_1}(t). \tag{5.62}$$

Where we have introduced for ease of notation for any set of indices $I\subseteq\{1,...,r_n\}$ the distance to a set $d_t^n(i,I):=\min\{d_t^n(i,j):\, j\in I\}$ and

$$\widetilde{\varphi}^I_{n,j_1j_2}(t,r)$$
$$:=\frac{1}{r_n}\int_0^{\frac{r-t}{h}+2}\sum_{i\neq j_1,j_2}X_{n,i}(t+uh)^T\widetilde{f}_n(t+uh,t)X_{n,j_1}(t)X_{n,j_2}(r)^T\widetilde{f}_n(t+uh,r)^T$$
$$\times X_{n,i}(t+uh)C_{n,i}(t+uh)\lambda_{n,i}(t+uh)\mathbb{1}(d^n_{t-4h}(i,I)\geq m)du.$$

Note that $\varphi_{n,j_1j_2}(t,r):=\widetilde{\varphi}^{\emptyset}_{n,j_1j_2}(t,r)$ equals exactly the inner integral in (5.62) (because $d^n_{t-4h}(i,\emptyset)=\infty$). Furthermore the integrand is partially-predictable with respect to $\widetilde{\mathcal{F}}^{n,I,m}_{j_1j_2,t}$ for all u because it has the product structure from Definition 2.16. Summing and integrating do not harm the measurability (note that the variable integration limits may interpreted as multiplication with a deterministic function). Thus, the functions $\widetilde{\varphi}^I_{n,j_1j_2}(t,r)$ are partially-predictable with respect to $\widetilde{\mathcal{F}}^{n,I,m}_{j_1j_2,t}$. So in order to prove that (5.62) converges to zero, it is enough to check the conditions (3.3)-(3.7) in Proposition 3.15. We do this in the following.

Define $p_n^*(t)$ such that

$$\frac{1}{p_n^*(t)}:=\int_{-1}^1\frac{1}{\bar{p}_n(t-hv)}dv.$$

Then, there is a constant $C>0$ such that for all $r\in[t-2h,t]$ and $t\in[\delta,T-\delta]$ (note that then $\frac{r-t}{h}\leq 0$)

$$|\widetilde{f}_n(t+uh,r)|$$
$$\leq M^2\int_{-1}^1K(v)K\left(\frac{r-t}{h}+v-u\right)\frac{w(t+h(u-v))}{\bar{p}_n(t+h(u-v))}dv$$
$$\leq M^2K\|w\|_\infty\int_{-1}^1\frac{1}{\bar{p}_n(t-h\nu))}d\nu$$
$$\leq C\frac{1}{p_n^*(t)}. \tag{5.63}$$

By the assumption of bounded covariates (Assumption (C1), 2), we get for any index sets $I,J\subseteq G_n$ and edges $j_1,j_2\in G_n$ and $r\in[t-2h,t]$ (i.e. $\frac{r-t}{h}\in[-2,0]$), that for $C^*:=\hat{K}^4\Lambda C^2$ (which does not depend on r,t,j_1,j_2,I and J)

$$\sup_{j_1,j_2}\left|\widetilde{\varphi}^I_{n,j_1j_2}(t,r)-\widetilde{\varphi}^J_{n,j_1j_2}(t,r)\right|$$
$$\leq\frac{C^*}{r_np_n^*(t)^2}\int_0^{\frac{r-t}{h}+2}\sum_{i\in G_n}C_{n,i}(t+uh)\left|\mathbb{1}(d^n_{t-4h}(i,I)\geq m)-\mathbb{1}(d^n_{t-4h}(i,J)\geq m)\right|du$$

$$\leq \frac{C^*}{r_n p_n^*(t)^2} \int_0^{\frac{r-t}{h}+2} \sum_{i\in G_n} C_{n,i}(t+uh) \sum_{k\in I\Delta J} \mathbb{1}(d^n_{t-4h}(i,k)\leq m)du$$

$$\leq \frac{2C^*}{r_n p_n^*(t)^2} \sum_{k\in I\Delta J} \left(\sum_{i\in G_n} \sup_{u\in[0,2]} C_{n,i}(t+uh)\mathbb{1}(d^n_{t-4h}(i,k)\leq m)\mathbb{1}(K^k_m(t)\leq F) \right.$$

$$\left. + \sum_{i\in G_n} \sup_{u\in[0,2]} C_{n,i}(t+uh)\mathbb{1}(d^n_{t-4h}(i,k)\leq m)\mathbb{1}(K^k_m(t) > F)\right)$$

$$\leq \frac{2C^*|I\Delta J|}{r_n p_n^*(t)^2} \left(\cdot F + K^{G_n}_m(t) H^{|I\Delta J|}_{UB}\right) \tag{5.64}$$

where for $A, B \subseteq G_n$, we denote by $A\Delta B := A \setminus B \cup B \setminus A$ the symmetric difference of A and B and recall

$$K^A_m(t) := \sup_{k\in A} \sum_i \sup_{u\in[-6,2]} C_{n,i}(t+uh)\mathbb{1}(d^n_{t-4h}(i,k)\leq m)$$

$$H^A_{UB} \geq \sup_{k\in A} \sup_{t\in[0,T]} \mathbb{1}(K^k_m(t) > F)$$

be the (random) number of friends at distance m an edge $k \in A$ can have and the indicator function whether and edge $k\in A$ has the potential to be a hub, respectively. Note that by Assumption (C6), H^A_{UB} is measurable with respect to $\mathcal{F}^n_0$. We will frequently use that $\lambda_{n,i}(t)$ is bounded by $C_{n,i}(t)\Lambda$, where Λ is the constant from Assumption (C1), 2. We denote $|M_{n,i}|(t) := \lambda_{n,i}(t) + N_{n,i}(t)$. Now, we show the conditions of Proposition 3.15. For condition (3.3) we get by applying the estimate (5.64) for any $\varepsilon > 0$ and any $F_0 > 0$ and F as in Assumption (C6)

$$\mathbb{P}\left(\left|\frac{1}{r_n}\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{r_n} \int_0^T \int_{t-2h}^{t-} \varphi_{n,j_1j_2}(t,r) - \tilde{\varphi}^{j_1j_2}_{n,j_1j_2}(t,r)dM_{n,j_2}(r)dM_{n,j_1}(t)\right| > \varepsilon\right)$$

$$\leq \mathbb{P}\left(\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}} \int_0^T \int_{t-2h}^{t-} \frac{4C^*}{r_n^2 p_n^*(t)^2}\left(F + K^{G_n}_m(t) H^{j_1j_2}_{UB}\right) d|M_{n,j_2}|(r)d|M_{n,j_1}|(t) > \varepsilon\right) \tag{5.65}$$

$$\leq \mathbb{P}\left(\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n} \int_0^T \int_{t-2h}^{t-} \frac{4C^*}{r_n^2 p_n^*(t)^2}\left(F + F_0 K^{G_n}_m(t-4h) H^{j_1j_2}_{UB}\right) d|M_{n,j_2}|(r)d|M_{n,j_1}|(t) > \varepsilon\right)$$

$$+ \mathbb{P}\left(\forall t\in[0,T] : K^{G_n}_m(t) > F_0 K^{G_n}_m(t-4h)\right)$$

By Assumption (C6) in (4.19), the last probability is small for all $n\in\mathbb{N}$, when F_0 is large enough. Keep this F_0 fixed and continue with the term in the first probability.

This term converges to zero by an application of Markov's Inequality:

$$\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}} \mathbb{E}\left(\int_0^T \int_{t-2h}^{t-} \frac{4C^*}{r_n^2 p_n^*(t)^2}\left(F+F_0K_m^{G_n}(t-4h)H_{UB}^{j_1j_2}\right) d|M_{n,j_2}|(r)d|M_{n,j_1}|(t)\right)$$

$$\leq \sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}} \int_0^T \frac{12C^*\Lambda}{r_n^2p_n^*(t)^2}\mathbb{E}\left(\int_{t-2h}^{t} d|M_{n,j_2}|(r)\cdot C_{n,j_1}(t)\cdot\left(F+F_0K_m^{G_n}(t-4h)H_{UB}^{j_1j_2}\right)\right)dt$$

$$= \sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}} \int_0^T \frac{12C^*\Lambda}{r_n^2p_n^*(t)^2}\mathbb{E}\Bigg(\mathbb{E}\left(\int_{t-2h}^{t} d|M_{n,j_2}|(r)\Big|\mathcal{F}_{t-2h}^{n,j_2,m}\right)$$

$$\times C_{n,j_1}(t)\mathbb{1}(d_{t-2h}^n(j_1,j_2)\geq m)\cdot\left(F+F_0K_m^{G_n}(t-4h)H_{UB}^{j_1j_2}\right)\Bigg)dt$$

$$+\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}} \int_0^T \frac{12C^*\Lambda}{r_n^2p_n^*(t)^2}\mathbb{E}\Bigg(\int_{t-2h}^{t} d|M_{n,j_2}|(r)\cdot C_{n,j_1}(t)$$

$$\times \mathbb{1}(d_{t-2h}^n(j_1,j_2)< m)\cdot\left(F+F_0K_m^{G_n}(t-4h)H_{UB}^{j_1j_2}\right)\Bigg)dt$$

$$\leq \sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}} \int_0^T h\frac{24C^*\Lambda^2}{r_n^2p_n^*(t)^2}\mathbb{E}\left(\sup_{r\in[t-2h,t]} C_{n,j_2}(r)\cdot C_{n,j_1}(t)\cdot\left(F+F_0K_m^{G_n}(t-4h)H_{UB}^{j_1j_2}\right)\right)dt$$

$$+\sum_{j_2\in G_n} \int_0^T \frac{12C^*\Lambda}{r_n^2p_n^*(t)^2}\mathbb{E}\Bigg(\int_{t-2h}^{t} d|M_{n,j_2}|(r)$$

$$\times\left(FK_m^{j_2}(t+2h)+F_0K_m^{G_n}(t-4h)(N_{UB}+r_nH_{UB}^{j_2})\right)\Bigg)dt,$$

where $N_{UB}:=\sum_{i\in G_n} H_{UB}^i$ is the number of hubs. These two terms converge to zero by Assumption (C6) in (4.20), (4.22) and (4.23), together with the Assumption (C7) in (4.33). Thus we have shown that (5.65) converges to zero and (3.3) follows.

We continue with (3.4): We use (5.64) in order to obtain

$$\Bigg|\frac{1}{r_n^2}\sum_{\substack{j_1,j_2,j_1'j_2'=1\\ j_1\neq j_1',j_2\neq j_2'}}^{r_n} \int_0^T\int_{t-2h}^{t-} \widetilde{\varphi}_{n,j_1j_2}^{j_1j_2}-\widetilde{\varphi}_{n,j_1j_2}^{j_1j_2j_1'j_2'}dM_{n,j_2}(r)dM_{n,j_1}(t)$$

$$\times\int_0^T\int_{t-2h}^{t-}\widetilde{\varphi}_{n,j_1'j_2'}^{j_1'j_2'}(t,r)-\widetilde{\varphi}_{n,j_1'j_2'}^{j_1j_2j_2'j_2'}(t,r)dM_{n,j_2'}(r)dM_{n,j_1'}(t)\Bigg|$$

$$\leq \frac{1}{r_n^2}\sum_{\substack{j_1,j_2,j_1'j_2'=1\\ j_1\neq j_1',j_2\neq j_2'}}^{r_n}\int_0^T \frac{4C^*}{p_n^*(t)^2}\left(F+K_m^{G_n}(t)H_{UB}^{j_1'j_2'}\right)\int_{t-2h}^{t-} d|M_{n,j_2}|(r)d|M_{n,j_1}|(t)$$

$$\times\frac{1}{r_n^2}\int_0^T \frac{4C^*}{p_n^*(t)^2}\left(F+K_m^{G_n}(t)H_{UB}^{j_1j_2}\right)\int_{t-2h}^{t-} d|M_{n,j_2'}|(r)d|M_{n,j_1'}|(t)$$

$$\leq \frac{1}{r_n^2}\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{r_n}\int_0^T \frac{4C^*}{p_n^*(t)^2}F\int_{t-2h}^{t-} d|M_{n,j_2}|(r)d|M_{n,j_1}|(t)$$

$$\times\frac{1}{r_n^2}\sum_{\substack{j_1',j_2'\\ j_1'\neq j_2'}}^{r_n}\int_0^T \frac{4C^*}{p_n^*(t)^2}F\int_{t-2h}^{t-} d|M_{n,j_2'}|(r)d|M_{n,j_1'}|(t)$$

$$+\frac{1}{r_n^2}\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{r_n}\int_0^T \frac{4C^*}{p_n^*(t)^2}\int_{t-2h}^{t-} d|M_{n,j_2}|(r)d|M_{n,j_1}|(t)\times H_{UB}^{G_n}\sup_{t\in[0,T]}K_m^{G_n}(t)$$

$$\times\frac{1}{r_n^2}\sum_{\substack{j_1',j_2'\\ j_1'\neq j_2'}}^{r_n}\int_0^T \frac{4C^*}{p_n^*(t)^2}\int_{t-2h}^{t-} d|M_{n,j_2'}|(r)d|M_{n,j_1'}|(t)\times H_{UB}^{G_n}\sup_{t\in[0,T]}K_m^{G_n}(t)$$

$$+\frac{2}{r_n^2}\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{r_n}\int_0^T \frac{4C^*}{p_n^*(t)^2}F\int_{t-2h}^{t-} d|M_{n,j_2}|(r)d|M_{n,j_1}|(t)$$

$$\times\frac{1}{r_n^2}\sum_{\substack{j_1',j_2'\\ j_1'\neq j_2'}}^{r_n}\int_0^T \frac{4C^*}{p_n^*(t)^2}\int_{t-2h}^{t-} d|M_{n,j_2'}|(r)d|M_{n,j_1'}|(t)\times H_{UB}^{G_n}\sup_{t\in[0,T]}K_m^{G_n}(t)$$

Other than $H_{UB}^{G_n}\cdot\sup_{t\in[0,T]}K_m^{G_n}(t)$, all expression above may be bounded by the expression in the probability in (5.65), which we have just established to be $o_P(1)$. By applying the Cauchy-Schwarz Inequality, we conclude also the L^1 convergence under additional use of Assumption (C8) in (4.39) and $\mathbb{E}\left(H_{UB}^{G_n}\left(\sup_{t\in[0,T]}K_m^{G_n}(t)\right)^4\right)=O(1)$ (cf. Assumption (C6) in (4.21)).

For showing condition (3.5) we define the random number of active edges as

$$A_n(t):=\sum_{i\in G_n}\sup_{u\in[0,2]}C_{n,i}(t+uh)$$

and use this to estimate for all $j_1,j_2\in G_n$, all $I\subseteq G_n$ and all $r,t\in[0,T]$

$$\left|\widetilde{\varphi}_{n,j_1j_2}^I(t,r)\right|\leq\frac{2C^*}{r_np_n^*(t)^2}A_n(t). \tag{5.66}$$

Using this estimate together with the estimate in (5.64), we obtain

$|(3.5)|$

$$
\begin{aligned}
\leq & \frac{2}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\Bigg(\int_0^T \int_{t-2h}^{t} \frac{4C^*}{r_n p_n^*(t)^2} \left(F + K_m^{G_n}(t) H_{UB}^{j_1 j_2} \right) d|M_{n,j_2}|(r) \\
& \times \int_t^{t+2h} \int_{\xi-2h}^{\xi-} \frac{4C^*}{r_n p_n^*(\xi)^2} A_n(\xi) d|M_{n,j_2'}|(\rho) d|M_{n,j_1'}|(\xi) \mathbb{1}(\neg j_1', j_2' \in j_1(m,t)) d|M_{n,j_1}|(t) \Bigg) \\
\leq & \frac{2}{r_n^2} \sum_{\substack{j_1,j_2,j_1',j_2' \in G_n \\ j_1 \neq j_2, j_1' \neq j_2'}} \mathbb{E}\Bigg(\int_0^T \int_{t-2h}^{t} \frac{4C^*}{r_n p_n^*(t)^2} \left(F + K_m^{G_n}(t) H_{UB}^{j_1 j_2} \right) d|M_{n,j_2}|(r) \\
& \times \sup_{\xi \in [t,t+2h]} \frac{4C^* A_n(\xi)}{r_n p_n^*(\xi)} \cdot \sup_{\substack{j_1',j_2' \in G_n \\ j_1' \neq j_2'}} \int_t^{t+2h} \int_{\xi-2h}^{\xi-} \frac{1}{p_n^*(\xi)} (\rho) d|M_{n,j_2'}|(\rho) d|M_{n,j_1'}|(\xi) \\
& \quad \times \sup_{\rho \in [t-2h,t+2h]} C_{n,j_2'}(\rho) \sup_{\xi \in [t,t+2h]} C_{n,j_1'}(\xi) \mathbb{1}(\neg j_1', j_2' \in j_1(m,t)) d|M_{n,j_1}|(t) \Bigg). \qquad (5.67)
\end{aligned}
$$

At this point we split the the sum in the first line and consider one expression with F and one with $K_m^{G_n}(t) H_{UB}^{j_1 j_2}$. In the first case with F, we will use

$$
\begin{aligned}
& \sup_{\rho \in [t-2h,t+2h]} C_{n,j_2'}(\rho) \sup_{\xi \in [t,t+2h]} C_{n,j_1'}(\xi) \sum_{j_1',j_2' \in G_n} \mathbb{1}(\neg j_1', j_2' \in j_1(m,t)) \\
= & \sum_{j_1' \in G_n} \sup_{\xi \in [t,t+2h]} C_{n,j_1'}(\xi) \mathbb{1}(j_1' \in j_1(m,t)) \times \sum_{j_2' \in G_n} \sup_{\rho \in [t-2h,t+2h]} C_{n,j_2'}(\rho) \mathbb{1}(j_2' \notin j_1(m,t)) \\
& + \sum_{j_1' \in G_n} \sup_{\xi \in [t,t+2h]} C_{n,j_1'}(\xi) \mathbb{1}(j_1' \notin j_1(m,t)) \times \sum_{j_2' \in G_n} \sup_{\rho \in [t-2h,t+2h]} C_{n,j_2'}(\rho) \mathbb{1}(j_2' \in j_1(m,t)) \\
& + \sum_{j_1' \in G_n} \sup_{\xi \in [t,t+2h]} C_{n,j_1'}(\xi) \mathbb{1}(j_1' \notin j_1(m,t)) \times \sum_{j_2' \in G_n} \sup_{\rho \in [t-2h,t+2h]} C_{n,j_2'}(\rho) \mathbb{1}(j_2' \notin j_1(m,t)) \\
= & 3 A_n(t) K_m^{j_1}(t).
\end{aligned}
$$

Now we continue with our main equality chain:

$$
\begin{aligned}
& (5.67) \\
\leq & \frac{2}{r_n^2} \sum_{\substack{j_1,j_2 \in G_n \\ j_1 \neq j_2}} \mathbb{E}\Bigg(\int_0^T \int_{t-2h}^{t-} \frac{4C^* F}{r_n p_n^*(t)^2} d|M_{n,j_2}|(r) \\
& \times \sup_{\xi \in [t,t+2h]} \frac{4C^* A_n(\xi)}{r_n p_n^*(\xi)} \times \sup_{\substack{j_1',j_2' \in G_n \\ j_1' \neq j_2'}} \int_t^{t+2h} \int_{\xi-2h}^{\xi-} \frac{1}{p_n^*(\xi)} d|M_{n,j_2'}|(\rho) d|M_{n,j_1}|(\xi) \\
& \qquad \times 3 A_n(t) K_m^{j_1}(t) d|M_{n,j_1}|(t) \Bigg)
\end{aligned}
$$

$$+\frac{2}{r_n^2}\sum_{\substack{j_1,j_2,j_1'\in G_n\\ j_1\neq j_2}}\mathbb{E}\Bigg(\int_0^T\int_{t-2h}^{t-}\frac{4C^*}{r_np_n^*(t)^2}K_m^{G_n}(t)(N_{UB}+r_nH_{UB}^{j_1'})d|M_{n,j_2}|(r)$$

$$\times\sup_{\xi\in[t,t+2h]}\frac{4C^*A_n(\xi)}{r_np_n^*(\xi)}\times\sup_{\substack{j_1',j_2'\in G_n\\ j_1'\neq j_2'}}\int_t^{t+2h}\int_{\xi-2h}^{\xi-}\frac{1}{p_n^*(\xi)}d|M_{n,j_2'}|(\rho)d|M_{n,j_1}|(\xi)d|M_{n,j_1}|(t)\Bigg).$$

Both expressions converge to zero by assumption (C6) in (4.24) and (4.25).

The indicator function in (3.6) is not significantly shortening the sum and hence we just ignore it. Moreover, we use the bound from (5.66) to obtain

$$\begin{aligned}
&(3.6)\\
\leq&\frac{\Lambda^2}{r_n^2}\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}}\int_0^T\int_{t-2h}^{t}\mathbb{E}\left[\frac{4(C^*)^2}{r_n^2p_n^*(t)^4}A_n(t)^2C_{n,j_1}(t)C_{n,j_2}(r)\right]drdt\\
\leq&\frac{\Lambda^2}{r_n^2}\sum_{\substack{j_1,j_2\in G_n\\ j_1\neq j_2}}\int_0^T\int_{t-2h}^{t}\mathbb{E}\left[\left(\frac{2C^*A_n(t)}{r_np_n^*(t)}\right)^2\cdot\frac{C_{n,j_1}(t)C_{n,j_2}(r)}{p_n^*(t)^2}\right]drdt
\end{aligned}$$

This converges to zero by Assumption (C8) in (4.37).

For (3.7) we finally use again that for every fixed choice of j_2, j_2' we get

$$\sum_{j_1=1}^n C_{n,j_1}(t)\mathbb{1}(\neg j_2,j_2'\in j_1(m,t-2h))\leq K_m^{j_2}(t+2h)+K_m^{j_2'}(t+2h).$$

Thus, we obtain together with (5.66)

$$\begin{aligned}
&|(3.7)|\\
\leq&\frac{\Lambda}{r_n^2}\sum_{\substack{j_2,j_2'\in G_n\\ j_2\neq j_2'}}\int_0^T\mathbb{E}\Bigg[\int_{t-2h}^{t-}\frac{2C^*}{r_np_n^*(t)^2}A_n(t)d|M_{n,j_2}|(r)\\
&\times\int_{t-2h}^{t-}\frac{2C^*}{r_np_n^*(t)^2}A_n(t)d|M_{n,j_2'}|(r')\sum_{j_1\in G_n}C_{n,j_1}(t)\mathbb{1}(\neg j_2,j_2'\in j_1(m,t))\Bigg]dt\\
=&\frac{\Lambda}{r_n^2}\sum_{\substack{j_2,j_2'\in G_n\\ j_2\neq j_2'}}\int_0^T\mathbb{E}\Bigg[\int_{t-2h}^{t-}\frac{1}{p_n^*(t)}d|M_{n,j_2}|(r)\int_{t-2h}^{t-}\frac{1}{p_n^*(t)}d|M_{n,j_2'}|(r')\\
&\times\left(\frac{2C^*A_n(t)}{r_np_n^*(t)}\right)^2\left(K_m^{j_2}(t+2h)+K_m^{j_2'}(t+2h)\right)\Bigg]dt.
\end{aligned}$$

Assumption (C6) in (4.26) ensures convergence of this expression to zero.

To prove the second assumption in Statement 4.6 we will apply very similar techniques as before. In fact, we can use almost exactly the same steps with $j_2 = j_1$ we have taken in order to arrive at (5.62) with one exception: At some point we said that we can ignore the indicator function $\mathbb{1}_{t=r}$ because $j_1 \neq j_2$, this is not true now and we need to take care of this. We obtain

$$\begin{aligned}
&\frac{4}{hr_n^2}\sum_{i\in G_n}\int_0^T \sum_{\substack{j\in G_n\\ j\neq i}} \tau_{n,ij}(s)^2 C_{n,i}(s)\lambda_{n,i}(s)ds \\
=\;& \frac{8}{r_n}\sum_{j\in G_n}\int_0^T\int_{t-2h}^{t-}\varphi_{n,j}(t,r)dM_{n,j}(r)dM_{n,j}(t) \qquad (5.68)\\
&+\frac{4}{r_n}\sum_{j\in G_n}\int_0^T \varphi_{n,j}(t,t)\cdot\Delta M_{n,j}(t)dM_{n,j}(t), \qquad (5.69)
\end{aligned}$$

where we used the abbreviation $\varphi_{n,j}(r,t) := \varphi_{n,jj}(r,t)$. We prove that (5.68) converges to zero in probability by applying similar techniques as before. We start by approximating $\varphi_{n,j}$ by its measurable approximation $\widetilde{\varphi}^j_{n,j}$:

$$\begin{aligned}
&(5.68)\\
=&\frac{8}{r_n}\sum_{j\in G_n}\int_0^T\int_{t-2h}^{t-}\varphi_{n,j}(t,r)-\widetilde{\varphi}^j_{n,j}(t,r)dM_{n,j}(r)dM_{n,j}(t) \qquad (5.70)\\
&+\frac{8}{r_n}\sum_{j\in G_n}\int_0^T\int_{t-2h}^{t-}\widetilde{\varphi}^j_{n,j}(t,r)dM_{n,j}(r)dM_{n,j}(t). \qquad (5.71)
\end{aligned}$$

We now use again the approximation (5.64) and obtain by using martingale properties and Markov's Inequality in exactly the same way as in (5.65)

$$\begin{aligned}
&\mathbb{P}(|(5.70)|>\varepsilon)\\
\leq&\frac{1}{\varepsilon}\mathbb{E}\Bigg(\frac{8}{r_n}\sum_{j\in G_n}\int_0^T\int_{t-2h}^{t-}\frac{2FC^*}{r_np_n^*(t)^2}d|M_{n,j}|(r)d|M_{n,j}|(t)\\
&+\frac{8}{r_n}\sum_{j\in G_n}\int_0^T\int_{t-2h}^{t-}\frac{2C^*F_0}{r_np_n^*(t)^2}K_m^{G_n}(t-4h)H^j_{UB}d|M_{n,j}|(r)d|M_{n,j}|(t)\Bigg)\\
&+\mathbb{P}\left(\forall t\in[0,T]:\, K_m^{G_n}(t)>F_0\cdot K_m^{G_n}(t-4h)\right)\\
=&\frac{8}{\varepsilon}\int_0^T\frac{2FC^*p_n(t)}{r_np_n^*(t)^2}\mathbb{E}\left(\int_{t-2h}^{t-}d|M_{n,j}|(r)\lambda_{n,j}(t)\middle|C_{n,j}(t)=1\right)dt\\
&+\frac{8}{\varepsilon}\int_0^T\frac{2C^*F_0p_n(t)}{r_np_n^*(t)^2}\mathbb{E}\left(\int_{t-2h}^{t-}K_m^{G_n}(t-4h)H^j_{UB}d|M_{n,j}|(r)\lambda_{n,j}(t)\middle|C_{n,j}(t)=1\right)dt\\
&+\mathbb{P}\left(\forall t\in[0,T]:\, K_m^{G_n}(t)>F_0\cdot K_m^{G_n}(t-4h)\right)
\end{aligned}$$

These expressions are small by a good choice of F and then for n large enough by Assumptions (C8) in (4.38) and (C6) in (4.27).

For (5.71), we note that $\widetilde{\varphi}^{j}_{n,j}$ is predictable with respect to $\mathcal{F}^{n,j,m}_t$. Moreover, by (5.66), $|\varphi^{j}_{n,j}(t,r)| \leq \frac{2C^*}{r_n p^*_n(t)^2} A_n(t)$. And thus we get again by Lemma 3.12 that

$$\begin{aligned}
&\mathbb{P}(|(5.71)| > \varepsilon)\\
\leq&\frac{16\Lambda}{\varepsilon r_n}\sum_{j\in G_n}\int_0^T \mathbb{E}\left(\int_{t-2h}^{t-}|\varphi^{j}_{n,j}(t,r)|d|M_{n,j}|(r)C_{n,j}(t)\right)dt\\
\leq&\frac{32C^*\Lambda}{\varepsilon r_n^2}\sum_{j\in G_n}\int_0^T \mathbb{E}\left(\frac{1}{p^*_n(t)^2}\int_{t-2h}^{t-}d|M_{n,j}|(r)\cdot A_n(t)\cdot C_{n,j}(t)\right)dt\\
\leq&\frac{32C^*\Lambda}{\varepsilon}\int_0^T \mathbb{E}\left(\frac{A_n(t)}{r_n p^*_n(t)}\cdot\int_{t-2h}^{t-}d|M_{n,j}|(r)\middle| C_{n,j}(t)=1\right)dt.
\end{aligned}$$

This converges to zero by Assumption (C8) in (4.35).

We study now the convergence behaviour of (5.69). Note firstly that

$$\begin{aligned}
&(5.69)\\
=&\frac{4}{r_n}\sum_{j\in G_n}\int_0^T \varphi_{n,j}(t,t)dM_{n,j}(t) &(5.72)\\
&+\frac{4}{r_n}\sum_{j\in G_n}\int_0^T \varphi_{n,j}(t,t)C_{n,j}(t)\lambda_{n,j}(t)dt &(5.73)
\end{aligned}$$

The first part, (5.72), converges to zero by an application of Proposition 3.13. We get by said Proposition

$$\begin{aligned}
&\mathbb{E}\left(\left(5.72\right)^2\right)\\
=&\frac{16\Lambda}{r_n^2}\sum_{i\in G_n}\int_0^T \mathbb{E}\left(\widetilde{\varphi}^{i}_{n,i}(t,t)^2 C_{n,i}(t)\right)dt &(5.74)\\
&+\frac{32}{r_n^2}\sum_{i,j\in G_n}\mathbb{E}\left(\int_0^T \widetilde{\varphi}^{ij}_{n,i}(t,t)dM_{n,i}(t)\cdot\int_0^T\left(\varphi_{n,j}(t,t)-\widetilde{\varphi}^{ij}_{n,j}(t,t)\right)dM_{n,j}(t)\right) &(5.75)\\
&+\frac{16}{r_n^2}\sum_{i,j\in G_n}\mathbb{E}\left(\int_0^T\left(\varphi_{n,i}(t,t)-\widetilde{\varphi}^{ij}_{n,i}(t,t)\right)dM_{n,i}(t)\cdot\int_0^T\left(\varphi_{n,j}(t,t)-\widetilde{\varphi}^{ij}_{n,j}\right)dM_{n,j}(t)\right). &(5.76)
\end{aligned}$$

We apply estimates (5.64) and (5.66) to show that the three lines above converge to zero. We have

$$(5.74)$$

$$
\begin{aligned}
&=\frac{16\Lambda}{r_n^2}\sum_{i\in G_n}\int_0^T \mathbb{E}\left(\left(\frac{2C^*A_n(t)}{r_np_n^*(t)^2}\right)^2 \middle| C_{n,i}(t)=1\right)p_n(t)dt\\
&=16\Lambda\int_0^T \frac{p_n(t)}{r_np_n^*(t)^2}\mathbb{E}\left(\left(\frac{2C^*A_n(t)}{r_np_n^*(t)}\right)^2 \middle| C_{n,i}(t)=1\right)dt
\end{aligned}
$$

which converges to zero by Assumption (C8) in (4.36). We continue with (5.75) and (5.76) to get

$$
\begin{aligned}
&(5.75)\\
\le&\frac{32}{r_n^2}\sum_{i,j\in G_n}\mathbb{E}\left(\int_0^T\frac{2C^*A_n(t)}{r_np_n^*(t)^2}d|M_{n,i}|(t)\cdot\int_0^T\frac{4C^*}{r_np_n^*(t)^2}\left(F+K_m^{G_n}(t)H_{UB}^{ij}\right)d|M_{n,j}|(t)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&(5.76)\\
\le&\frac{32}{r_n^2}\sum_{i,j\in G_n}\mathbb{E}\Bigg(\int_0^T\frac{4C^*}{r_np_n^*(t)^2}\left(F+K_m^{G_n}(t)H_{UB}^{ij}\right)d|M_{n,i}|(t)\\
&\qquad\times\int_0^T\frac{4C^*}{r_np_n^*(t)^2}\left(F+K_m^{G_n}(t)H_{UB}^{ij}\right)d|M_{n,j}|(t)\Bigg).
\end{aligned}
$$

Both converge to zero by Assumption (C8). And we conclude that (5.72) converges to zero.

So we have left to prove convergence of (5.73) which we do as follows: Denote by superscripts entries of the vectors or matrices, i.e., $X_{n,i}^{r_1}(t)^2$ refers to the square of the r_1-th entry of $X_{n,i}(t)$ and $\widetilde{f}_n^{r_1,r_2}(t+uh,t)$ refers to the entry in row r_1 and column r_2 of the matrix $\widetilde{f}_n(t+uh,t)$. Then

$$
\begin{aligned}
&\frac{4}{r_n}\sum_{j\in G_n}\int_0^T\varphi_{n,j}(t,t)C_{n,j}(t)\lambda_{n,j}(t)dt\\
=&\frac{4}{r_n}\sum_{j\in G_n}\int_0^T\frac{1}{r_n}\int_0^2\sum_{\substack{i\in G_n\\ i\neq j}}X_{n,i}(t+uh)^T\widetilde{f}_n(t+uh,t)X_{n,j}(t)X_{n,j}(t)^T\\
&\qquad\times\widetilde{f}_n(t+uh,t)X_{n,i}(t+uh)C_{n,i}(t+uh)\lambda_{n,i}(t+uh)du\,C_{n,j}(t)\lambda_{n,j}(t)dt\\
=&\frac{4}{r_n}\sum_{j\in G_n}\int_0^T\frac{1}{r_n}\int_0^2\sum_{\substack{i\in G_n\\ i\neq j}}\left[X_{n,i}(t+uh)^T\widetilde{f}_n(t+uh,t)X_{n,j}(t)\right]^2\\
&\qquad\times C_{n,i}(t+uh)\lambda_{n,i}(t+uh)du\,C_{n,j}(t)\lambda_{n,j}(t)dt\\
=&4\sum_{r_1,r_2,r_1',r_2'=1}^{q}\int_0^T\int_0^2\frac{1}{r_n}\sum_{i\in G_n}X_{n,i}^{r_1}(t+uh)X_{n,i}^{r_1'}(t+uh)\widetilde{f}_n^{r_1,r_2}(t+uh,t)
\end{aligned}
$$

$$
\begin{aligned}
&\quad\times \frac{1}{r_n}\sum_{j\in G_n} X_{n,j}^{r_2}(t)X_{n,j}^{r_2'}(t)\widetilde{f}_n^{r_1',r_2'}(t+uh,t)C_{n,j}(t)\lambda_{n,j}(t)C_{n,i}(t+uh)\lambda_{n,i}(t+uh)dudt\\
&-\frac{4}{r_n}\sum_{r_1,r_2,r_1',r_2'=1}^{q}\int_0^T\int_0^2\frac{1}{r_n}\sum_{i\in G_n} X_{n,i}^{r_1}(t+uh)X_{n,i}^{r_1'}(t+uh)\widetilde{f}_n^{r_1,r_2}(t+uh,t)\widetilde{f}_n^{r_1',r_2'}(t+uh,t)\\
&\quad\times X_{n,i}^{r_2}(t)X_{n,i}^{r_2'}(t)C_{n,i}(t)\lambda_{n,i}(t)C_{n,i}(t+uh)\lambda_{n,i}(t+uh)dudt\\
=&4\sum_{\substack{r_1,r_2=1\\ r_1',r_2'=1}}^{q}\int_0^T\int_0^2\frac{1}{r_n}\sum_{i\in G_n}\Delta_{n,i}^{r_1r_1',r_1r_2}(t+uh,t,t+uh)\cdot\frac{1}{r_n}\sum_{j\in G_n}\Delta_{n,j}^{r_2r_2',r_1'r_2'}(t,t,t+uh)dudt\\
&-\frac{4}{r_n}\sum_{r_1,r_2,r_1',r_2'=1}^{q}\int_0^T\int_0^2\frac{1}{r_n}\sum_{i\in G_n}\Delta_{n,i}^{r_1r_1',r_1r_2}(t+uh,t,t+uh)\Delta_{n,i}^{r_2r_2',r_1'r_2'}(t,t,t+uh)dudt,
\end{aligned}
\tag{5.77}
$$

where for all $a,b,c,d\in\{1,...,q\}$

$$
\begin{aligned}
\Delta_{n,j}^{ab,cd}(\tau,t,s) &:= X_{n,j}^{a}(\tau)X_{n,j}^{b}(\tau)\widetilde{f}_n^{c,d}(s,t)C_{n,j}(\tau)\lambda_{n,j}(\tau)\\
\widetilde{\Delta}_{n,j}^{ab,cd}(\tau,t,s) &:= \Delta_{n,j}^{ab,cd}(\tau,t,s)-\mathbb{E}\left(\Delta_{n,j}^{ab,cd}(\tau,t,s)\right)\\
&= \widetilde{f}_n^{c,d}(s,t)\Big[X_{n,j}^{a}(\tau)X_{n,j}^{b}(\tau)C_{n,j}(\tau)\lambda_{n,j}(\tau)\\
&\qquad -\mathbb{E}\left(X_{n,j}^{a}(\tau)X_{n,j}^{b}(\tau)C_{n,j}(\tau)\lambda_{n,j}(\tau)\right)\Big].
\end{aligned}
$$

We keep this in mind and prove now for all $r_1,r_1',r_2,r_2'\in\{1,...,q\}$

$$
\sup_{t\in[0,T],u,v\in[0,2]}\left|\frac{1}{r_n}\sum_{i\in G_n}\widetilde{\Delta}_{n,i}^{r_1r_1',r_1r_2}(t+vh,t,t+uh)\right| = o_P(1) \tag{5.78}
$$

via exponential inequality techniques. Since the argument is the same for all indices, we omit r_1,r_1',r_2,r_2' in the notation. Let therefore $\mathbb{T}_n$ denote a grid of $[0,T]\times[0,2]^2$ with mesh $H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}$ (where k_X is as in Assumption (C3)) and let (t^*,u^*,v^*) be the projection of $(t,u,v)\in[0,T]\times[0,2]^2$ onto $\mathbb{T}_n$, i.e., $\|(t,u,v)-(t^*,u^*,v^*)\|\le H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}$. We obtain

$$
\sup_{t\in[0,T],u,v\in[0,2]}\left|\frac{1}{r_n}\sum_{i\in G_n}\widetilde{\Delta}_{n,i}(t+vh,t,t+uh)\right|
$$

$$
\le \sup_{t\in[0,T],u,v\in[0,2]}\left|\frac{1}{r_n}\sum_{i\in G_n}\left(\widetilde{\Delta}_{n,i}(t+vh,t,t+uh)-\widetilde{\Delta}_{n,i}(t^*+v^*h,t^*,t^*+u^*h)\right)\right| \tag{5.79}
$$

$$
+\sup_{t\in[0,T],u,v\in[0,2]}\left|\frac{1}{r_n}\sum_{i\in G_n}\widetilde{\Delta}_{n,i}(t^*+v^*h,t^*,t^*+u^*h)\right|. \tag{5.80}
$$

For (5.79) we note that by Assumption (C3), 3, we have that

$$\begin{aligned}
&\Bigg| \left(\Sigma^{-T}\Sigma^{-1}\right)(t+h(u-\nu))\frac{w(t+h(u-\nu))}{p_n^*(t+h(u-\nu))}p_n(t+vh) \\
&\qquad\qquad - \left(\Sigma^{-T}\Sigma^{-1}\right)(t^*+h(u^*-\nu))\frac{w(t^*+h(u^*-\nu))}{p_n^*(t^*+h(u^*-\nu))}p_n(t^*+v^*h)\Bigg| \\
\leq& \mathcal{C}\left(|(t+h(u-\nu),t+vh)-(t^*+h(u^*-\nu),t^*+v^*h)|\right) \\
=& \mathcal{C}\left(|(t-t^*+h(u-u^*),t-t^*+h(v-v^*))|\right) \\
\leq& \mathcal{C}\left(2H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}\right).
\end{aligned}$$

This estimation holds as well if we had just considered one entry of the matrix $\Sigma^{-T}\Sigma^{-1}$. Thus, we can estimate by boundedness of w and Σ^{-1} and by choice of k_X

$$\begin{aligned}
&\left|p_n(t+vh)\widetilde{f}_n^{r_1r_2}(t+uh,t)-p_n(t^*+v^*h)\widetilde{f}_n^{r_1r_2}(t^*+u^*h,t^*)\right| \\
\leq& \int_{-1}^{1} K(\nu)K(u+\nu)d\nu\cdot\mathcal{C}\left(2H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}\right)d\nu \\
&\qquad + \int_{-1}^{1} K(\nu)w(t^*+h(u^*-\nu))p_n(t^*+v^*h)d\nu\frac{H_K}{p_n}|u-u^*|^{\alpha_K}M^2d\nu \\
\leq& K\mathcal{C}\left(2H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}\right)+O(1)\cdot\frac{1}{p_n}\left(H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}\right)^{\alpha_K}=o(p_n).
\end{aligned}$$

Recall $\xi_{n,i}(t)=X_{n,i}(t)X_{n,i}(t)C_{n,i}(t)\lambda_{n,i}(t)$. We find now for (5.79) by boundedness of the covariates (note that we omit also indices on $X_{n,j}$ and take them as univariate here)

$$\begin{aligned}
&\left|\frac{1}{r_n}\sum_{i\in G_n}\left(\widetilde{\Delta}_n(t+vh,t,t+uh)-\widetilde{\Delta}_{n,i}(t^*+v^*h,t^*,t^*+u^*h)\right)\right| \\
=&\Bigg|\left(p_n(t+uh)\widetilde{f}_n(t+uh,t)-p_n(t^*+u^*h)\widetilde{f}_n(t^*+u^*h,t^*)\right) \\
&\qquad\times\left(\frac{\xi_{n,i}(t+vh)}{p_n(t+vh)}-\mathbb{E}\left(\frac{\xi_{n,i}(t+vh)}{p_n(t+vh)}\right)\right) \\
&\quad + p_n(t^*+u^*h)\widetilde{f}_n(t^*+u^*h,t^*) \\
&\qquad\times\left(\frac{\xi_{n,i}(t+vh)}{p_n(t+uh)}-\frac{\xi_{n,i}(t^*+v^*h)}{p_n(t^*+u^*h)}-\mathbb{E}\left(\frac{\xi_{n,i}(t+vh)}{p_n(t+uh)}-\frac{\xi_{n,i}(t^*+v^*h)}{p_n(t^*+u^*h)}\right)\right)\Bigg| \\
\leq&\left(K\mathcal{C}\left(2H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}\right)+O(1)\cdot\frac{1}{p_n}\left(H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}\right)^{\alpha_K}\right)\cdot\frac{\hat{K}^2\Lambda}{p_n} \\
&\quad + K^2M^2\|w\|_\infty\frac{1}{p_n}\cdot\left(\frac{\xi_{n,i}(t+vh)}{p_n(t+vh)}-\frac{\xi_{n,i}(t^*+v^*h)}{p_n(t^*+v^*h)}-\mathbb{E}\left(\frac{\xi_{n,i}(t+vh)}{p_n(t+vh)}-\frac{\xi_{n,i}(t^*+v^*h)}{p_n(t^*+v^*h)}\right)\right).
\end{aligned}$$

The first line converges to zero as just discussed and for the second line, we make the following considerations. It holds that

$$\begin{aligned}
&\frac{1}{r_n p_n}\sum_{j\in G_n}\Bigg[\frac{1}{p_n(t+vh)}X_{n,j}(t+vh)X_{n,j}(t+vh)C_{n,j}(t+vh)\lambda_{n,j}(t+vh)\\
&\qquad -\frac{1}{p_n(t^*+v^*h)}X_{n,j}(t^*+v^*h)X_{n,j}(t^*+v^*h)C_{n,j}(t^*+v^*h)\lambda_{n,j}(t^*+v^*h)\Bigg] \qquad (5.81)\\
&+\frac{1}{p_n}\mathbb{E}\Bigg(X_{n,j}(t+vh)X_{n,j}(t+vh)\frac{C_{n,j}(t+vh)}{p_n(t+vh)}\lambda_{n,j}(t+vh)\\
&\qquad -X_{n,j}(t^*+v^*h)X_{n,j}(t^*+v^*h)\frac{C_{n,j}(t^*+v^*h)}{p_n(t^*+v^*h)}\lambda_{n,j}(t^*+v^*h)\Bigg).
\end{aligned}$$

The expectation behaves well because we assume uniform continuity of it in Assumption (C2), 1 and the choice of k_X in Assumption (C3), 1. For (5.81) we have (use the continuity Assumptions (C3), 2 and (C1), 3)

$$\begin{aligned}
&\sup_{t\in[0,T],v\in[0,2]}|(5.81)|\\
\leq&\sup_{t\in[0,T],v\in[0,2]}\frac{1}{p_n r_n}\sum_{j\in G_n}\Bigg[\left|\frac{1}{p_n(t+vh)}-\frac{1}{p_n(t^*+v^*h)}\right|\hat{K}^2\Lambda\\
&\quad+\frac{1}{p_n(t+vh)}\Big(\mathcal{C}_\xi(|t-t^*|+h|v-v^*|)\\
&\qquad+(\text{Number of jumps of } \xi_{n,j} \text{ on } [t+vh,t^*+v^*h])\cdot\iota\Big)\Bigg]\\
\leq&\frac{H_{n,p}}{p_n}\left|H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}\right|^{\alpha_p}\hat{K}^2\Lambda+\frac{1}{p_n^2}\mathcal{C}_\xi(n^{-k_X})\\
&\quad+\sup_{t\in[0,T],v\in[0,2]}\ \sup_{s:|t-s|\leq n^{-k_X}}\frac{1}{r_n p_n p_n(t+v)}\sum_{j\in G_n}(\text{Number of jumps of } \xi_{n,j} \text{ on } [s,t])\cdot\iota.
\end{aligned}$$

By choice of the mesh $H_{n,p}^{-\frac{1}{\alpha_p}}n^{-k_X}$ and the choice of k_X in Assumption (C3), 2 everything converges to zero. Hence, (5.79) $= o_P(1)$.

Lastly we need to discuss (5.80). To this end, we apply a standard union bound technique together with Lemma 3.29. We can estimate when noting the sup in (5.80) is actually only taken over $\mathbb{T}_n$ that for every $\varepsilon>0$ by (5.63) (recall $\xi_{n,j}=X_{n,j}X_{n,j}C_{n,j}\lambda_{n,j}$)

$$\begin{aligned}
&\mathbb{P}\left((5.80)>\varepsilon\right)\\
=&\mathbb{P}\left(\sup_{(t,u,v)\in\mathbb{T}_n}\left|\widetilde{f}_n(t+uh,t)\right|\cdot\frac{1}{r_n}\left|\sum_{j\in G_n}\left[\xi_{n,j}(t+vh)-\mathbb{E}\left(\xi_{n,j}(t+vh)\right)\right]\right|>\varepsilon\right)
\end{aligned}$$

$$
\begin{aligned}
\leq & \mathbb{P}\left(\sup_{(t,u,v)\in\mathbb{T}_n} \frac{1}{r_n p_n^*(t)} \left| \sum_{j\in G_n} \Big[\xi_{n,j}(t+vh) - \mathbb{E}\left(\xi_{n,j}(t+vh)\right) \Big] \right| > \frac{\varepsilon}{C} \right) \\
\leq & \#\mathbb{T}_n \cdot \sup_{(t,u,v)\in\mathbb{T}_n} \mathbb{P}\left(\frac{1}{r_n p_n^*(t)} \sum_{j\in G_n} \Big[\xi_{n,j}(t+vh) - \mathbb{E}\left(\xi_{n,j}(t+vh)\right) \Big] > \frac{\varepsilon}{C} \right) \\
& + \#\mathbb{T}_n \cdot \sup_{(t,u,v)\in\mathbb{T}_n} \mathbb{P}\left(\frac{1}{r_n p_n^*(t)} \sum_{j\in G_n} \Big[\xi_{n,j}(t+vh) - \mathbb{E}\left(\xi_{n,j}(t+vh)\right) \Big] < -\frac{\varepsilon}{C} \right).
\end{aligned}
$$

We will see that the two lines above work completely analogously and hence, we continue only with the first line. The prove for second line is then identical, we just have to replace $\xi_{n,j}$ by $-\xi_{n,j}$. We will also replace now $\frac{\varepsilon}{C}$ by ε for notational convenience. We have

$$
\begin{aligned}
& \#\mathbb{T}_n \cdot \sup_{(t,u,v)\in\mathbb{T}_n} \mathbb{P}\left(\frac{1}{r_n p_n^*(t)} \sum_{j\in G_n} \Big[\xi_{n,j}(t+vh) - \mathbb{E}\left(\xi_{n,j}(t+vh)\right) \Big] > \varepsilon \right) \\
\leq & \#\mathbb{T}_n \cdot \sup_{(t,u,v)\in\mathbb{T}_n} \Bigg[\mathbb{P}(\Gamma_n^{*,t} = 0) && (5.82) \\
& + \mathbb{P}\left(\frac{1}{r_n p_n^*(t)} \sum_{j\in G_n} \Big[\xi_{n,j}(t+vh)\Gamma_n^{*,t} - \mathbb{E}\left(\xi_{n,j}(t+vh)\Gamma_n^{*,t}\right) \Big] > \varepsilon \right) && (5.83) \\
& + \mathbb{P}\left(\frac{1}{p_n^*(t)} \mathbb{E}\left(\xi_{n,j}(t+vh)\left(\Gamma_n^{*,t} - 1\right)\right) > \varepsilon \right) \Bigg]. && (5.84)
\end{aligned}
$$

We have that line (5.82) converges to zero by Assumption (C4*). Note next that the expression in the probability in (5.84) is actually deterministic and hence the probability equals either zero or one. We have by assumption (C1), 2 that

$$
\begin{aligned}
& \sup_{t\in[0,T],v\in[0,2]} \mathbb{E}\left(\frac{1}{p_n^*(t)} \xi_{n,j}(t+vh)\left(\Gamma_n^{*,t} - 1\right) \right) \\
\leq & \sup_{t\in[0,T]} \hat{K}^2 \Lambda \mathbb{E}\left(\left|\Gamma_n^{*,t} - 1\right| \Big| C_{n,j}(t) = 1 \right) \frac{p_n(t)}{p_n^*(t)} \to 0
\end{aligned}
$$

as $n \to \infty$ by Assumption (C4*). Thus for n large enough, we have that (5.84) equals zero. Finally, for line (5.83), we apply Lemma 3.29 to

$$Z_i = \xi_{n,i}(t+vh)\Gamma_n^t.$$

The second part of Definition 3.28 is fulfilled by the Assumptions in (C4*). We need to check the moment bound in the first part of Definition 3.28. We get for $\rho \geq 2$ (note the definition of $\Gamma_n^{*,t}$)

$$\mathbb{E}\left(\left| U_{k,m}^{n,t}(\Delta_n) \right|^\rho \right)$$

$$
\begin{aligned}
=&\mathbb{E}\left(\left|\sum_{j\in G_n}\xi_{n,j}(t+vh)\Gamma_n^{*,t}I_{n,j}^{k,m,t}(\Delta_n)-\mathbb{E}\left(\xi_{n,j}(t+vh)\Gamma_n^{*,t}I_{n,j}^{k,m,t}(\Delta_n)\right)\right|^{\rho}\right)\\
\leq&\mathbb{E}\left(\left|\sum_{j\in G_n}\xi_{n,j}(t+vh)\Gamma_n^{*,t}I_{n,j}^{k,m,t}(\Delta_n)-\mathbb{E}\left(\xi_{n,j}(t+vh)\Gamma_n^{*,t}I_{n,j}^{k,m,t}(\Delta_n)\right)\right|^{2}\right)\\
&\times\left(2\hat{K}^2\Lambda\cdot E_k^{*,n,t}\right)^{\rho-2}\\
\leq&\sigma^2E_{k,m}^{*,n,t}\cdot\left(2\hat{K}^2\Lambda\cdot E_k^{*,n,t}\right)^{\rho-2}
\end{aligned}
$$

by Assumption (C7) in (4.34) because

$$
\begin{aligned}
&\frac{1}{E_{k,m}^{*,n,t}}\mathrm{Var}\left(\sum_{j\in G_n}\xi_{n,j}(t+vh)\Gamma_n^{*,t}I_{n,j}^{k,m,t}(\Delta_n)\right)\\
=&\frac{1}{r_np_n^{*,k,m}(t)}\sum_{i,j\in G_n}\mathrm{Cov}\left(\xi_{n,i}(t+vh)\Gamma_n^{*,t}I_{n,i}^{k,m,t}(\Delta_n),\xi_{n,j}(t+vh)\Gamma_n^{*,t}I_{n,j}^{k,m,t}(\Delta_n)\right)
\end{aligned}
$$

Thus, we may apply Lemma 3.29 with $c_1:=2\hat{K}^2\Lambda$. Choose x so large such that

$$
H_{n,p}^{\frac{1}{\alpha}}n^{k_X}\cdot\sup_{(t,u,v)\in\mathbb{T}_n}\left(r_np_n^*(t)\right)^{-\frac{c_2x^2}{2(\sigma^2+c_1c_3x)}}\to0
$$

for $n\to\infty$ (this is possible by Assumption (C1), 4). We obtain therefore by Assumption (C4*) because here $|E|_{n,t}=\sum_{k,m}E_{k,m}^{*,n,t}=p_n^*(t)$ for n so large such that $x\cdot\sqrt{\frac{\log|E|_{n,t}}{|E|_{n,t}}}\leq\varepsilon$

$$
\begin{aligned}
&(5.83)\\
\leq&\#\mathbb{T}_n\cdot\sup_{(t,u,v)\in\mathbb{T}_n}\mathbb{P}\Bigg(\frac{1}{|E|_{n,t}}\sum_{j\in G_n}\left[\xi_{n,j}(t+vh)\Gamma_n^{*,t}-\mathbb{E}\left(\xi_{n,j}(t+vh)\Gamma_n^{*,t}\right)\right]\\
&\qquad>x\cdot\sqrt{\frac{\log|E|_{n,t}}{|E|_{n,t}}}\Bigg)\\
\leq&\#\mathbb{T}_n\cdot\sup_{(t,u,v)\in\mathbb{T}_n}\left(\mathcal{K}\cdot|r_np_n^*(t)|^{-\frac{c_2x^2}{2(\sigma^2+c_1c_3x)}}+\beta_t(\Delta_n)\mathcal{K}r_n\right)\\
&\to0
\end{aligned}
$$

because $\#\mathbb{T}_n\approx H_{n,p}^{\frac{1}{\alpha_p}}n^{k_X}$. This was the last piece for establishing (5.78). Because of that we can continue to compute (5.73). We note firstly that by Assumption (C1), 2 and (5.63)

$$
\frac{1}{r_n}\Delta_{n,i}^{r_2r_2',r_1r_2'}\leq\frac{1}{r_np_n^*(t)}C\hat{K}^2\Lambda\to0
$$

as $n \to \infty$. Thus, the second line in (5.77) converges to zero and the limit of (5.77) is the same as

$$\begin{aligned}
&4\sum_{r_1,r_2,r_1',r_2'=1}^{q}\int_0^T\int_0^2 \mathbb{E}\left(\Delta_{n,j}^{r_1r_1',r_1r_2}(t+uh,t,t+uh)\right)\cdot\mathbb{E}\left(\Delta_{n,j}^{r_2r_2',r_1'r_2'}(t,t,t+uh)\right)dudt\\
\to&4\sum_{r_1,r_2,r_1',r_2'=1}^{q}K^{(4)}\int_0^T P^{r_1,r_1'}(t)P^{r_2,r_2'}(t)[\Sigma^{-T}\Sigma^{-1}]^{r_1,r_2}(t)[\Sigma^{-T}\Sigma^{-1}]^{r_1',r_2'}(t)w(t)^2dt\\
=&4K^{(4)}\int_0^T \operatorname{trace}\left(\left(P(t)\Sigma^{-T}(t)\Sigma^{-1}(t)\right)^2\right)w^2(t)dt,
\end{aligned}$$

by the continuity assumptions in (B7), (C2), 2 and (C2), 1 and where

$$\begin{aligned}
K^{(4)} &:= \int_0^2\left(\int_{-1}^1 K(v)K(u+v)dv\right)^2du\\
P^{r_1,r_2} &:= \mathbb{E}\left(X_{n,j}^{r_1}(t)X_{n,j}^{r_2}(t)\lambda_{n,j}(t)\middle|C_{n,j}(t)=1\right).
\end{aligned}$$

This proves the statement. □

Proposition 5.22. *Suppose that Statement 4.6 holds (cf. Proposition 5.21). Then, Statement 4.7 holds true.*

Proof. For any given $\varepsilon > 0$ we obtain, by applying Lenglart's Inequality as in Corollary 2.11 and simply taking the sup

$$\begin{aligned}
&\frac{4}{hr_n^2}\sum_{i\in G_n}\int_0^T \mathbb{1}\left(\left\|\frac{2}{h^{\frac{1}{2}}r_n}\sum_{j\neq i}\tau_{n,ij}(s)\right\|>\varepsilon\right)\left(\sum_{j\neq i}\tau_{n,ij}(s)\right)^2\lambda_{n,i}(s)ds\\
\leq&\mathbb{1}\left(\sup_{s\in[0,T],i\in G_n}\left\|\frac{2}{h^{\frac{1}{2}}r_n}\sum_{j\neq i}\tau_{n,ij}(s)\right\|>\varepsilon\right)\frac{4}{hr_n^2}\sum_{i\in G_n}\int_0^T\left(\sum_{j\neq i}\tau_{n,ij}(s)\right)^2\lambda_{n,i}(s)ds.
\end{aligned}$$

Statement 4.6 is stating that the second part is converging and hence it is sufficient to prove that the indicator function is converging to zero in probability which is equivalent of proving uniform convergence in probability (uniform in i and s) of

$$\frac{1}{h^{\frac{1}{2}}r_n}\sum_{j\neq i}\tau_{n,ij}(s)$$

to zero. We are going to employ Lemma 5.28. To this end note firstly that $\tau_{n,ij}(s)$ has the following structure

$$\tau_{n,ij}(s) = X_{n,i}(s)^T\int_0^{s-}\widetilde{f}_n(s,t)X_{n,j}(t)dM_{n,j}(t),$$

where

$$\widetilde{f}_n(s,t) := \int_\delta^{T-\delta} \frac{1}{h}K\left(\frac{s-t_0}{h}\right) K\left(\frac{t-t_0}{h}\right) \Sigma(t_0)^{-T}\Sigma(t_0)^{-1}\frac{w(t_0)}{\bar{p}_n(t_0)}dt_0.$$

We can simplify the expression by interchanging the integrals, taking the norm inside and using all boundedness properties from Assumptions (C1), 2 and (B7):

$$\begin{aligned}
&\sup_{s\in[0,T],i\in G_n} \frac{1}{h^{\frac{1}{2}}r_n}\left\|\sum_{j\neq i}\tau_{n,ij}(s)\right\| \\
=&\sup_{s\in[0,T],i\in G_n} h^{\frac{1}{2}}\int_\delta^{T-\delta} \|X_{n,i}(s)\|\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\left\|\Sigma(t_0)^{-T}\Sigma(t_0)^{-1}\right\| w(t_0) \\
&\qquad\times\frac{1}{r_n\bar{p}_n(t_0)}\sum_{j\neq i}\int_0^{s-}\frac{1}{h}K\left(\frac{t-t_0}{h}\right)\|X_{n,j}(t)\|d|M_{n,j}|(t)dt_0. \\
\leq& h^{\frac{1}{2}}\hat{K}M^2\sup_{s\in[0,T]}\int_\delta^{T-\delta}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)w(t_0) \\
&\qquad\times\frac{1}{r_n\bar{p}_n(t_0)}\sum_{j\in G_n}\int_0^T\frac{1}{h}K\left(\frac{t-t_0}{h}\right)\|X_{n,j}(t)\|d|M_{n,j}|(t)dt_0 \\
\leq& h^{\frac{1}{2}}\hat{K}M^2\|w\|_\infty \\
&\qquad\times\sup_{t_0\in\mathbb{T}}\frac{1}{r_n\bar{p}_n(t_0)}\sum_{j\in G_n}\int_0^T\frac{1}{h}K\left(\frac{t-t_0}{h}\right)\|X_{n,j}(t)\|dM_{n,j}(t) \qquad (5.85)\\
&+h^{\frac{1}{2}}2\hat{K}^2M^2\Lambda\|w\|_\infty \\
&\qquad\times\sup_{t_0\in\mathbb{T}}\frac{1}{r_n\bar{p}_n(t_0)}\sum_{j\in G_n}\int_0^T\frac{1}{h}K\left(\frac{t-t_0}{h}\right)C_{n,i}(t)dt. \qquad (5.86)
\end{aligned}$$

Now, (5.85) $= o_P(1)$ because the expression in (5.85) is the same as in $\ell_n'(\theta_0,t_0)$ but with $X_{n,j}(t)$ replaced by $\|X_{n,j}(t)\|$. Moreover, all mixing properties valid for $X_{n,j}(t)$ hold for $\|X_{n,j}(t)\|$ as well and, of course, $\|X_{n,j}(t)\|$ is also bounded. Thus, we may repeat the proof of Proposition 5.12 and all subsidiary results word by word and (5.85) converges to zero in probability. We also have that (5.86) is $O_P\left(h^{\frac{1}{2}}\right) = o_P(1)$ by (5.94). Hence, we have shown that

$$\sup_{s\in[0,T],i\in G_n}\frac{1}{h^{\frac{1}{2}}r_n}\left\|\sum_{j\neq i}\tau_{n,ij}(s)\right\| = o_P(1)$$

and this finalizes the proof of the statement. □

Proposition 5.23. *Suppose that Assumptions (B4), 2, (B6) and (B7) as well as (C1), 1, 2 and 4, (C2), 2, (C4) and* (4.30) *and* (4.31) *in (C7) hold. Then, Statement 4.8 holds true.*

Proof. We will prove shortly that $\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(t_0,\theta_0)$ converges uniformly (in t_0) to $\Sigma(t_0,\theta_0)$. So let us assume this for the moment. And define the event A_n by

$$A_n := \left\{\forall t_0 \in [\delta, T-\delta] : \left\|\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0) - \Sigma(t_0,\theta_0)\right\| \le \rho\right\},$$

where ρ is the same as in Assumption (B7). So we have that $\mathbb{P}(A_n) \to 1$. On A_n, we find

$$\begin{aligned}
&\left\|\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0)\right]^{-1} - \Sigma(t_0,\theta_0)^{-1}\right\| \\
\le &\left\|\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0)\right]^{-1}\right\| \cdot \left\|\Sigma(t_0,\theta_0) - \frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0)\right\| \cdot \left\|\Sigma(t_0,\theta_0)^{-1}\right\|.
\end{aligned}$$

By Assumption (B7), we know that $\left\|\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0)\right]^{-1}\right\|$ and $\left\|\Sigma(t_0,\theta_0)^{-1}\right\|$ are both bounded by M and hence we may conclude

$$\sup_{t_0\in[\delta,T-\delta]}\left\|\left[\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0)\right]^{-1}\right\| = O_P(1)$$

as required. So let us prove the uniform convergence of $\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0)$ to $\Sigma(\theta_0,t_0)$. Denote therefore by T_n a grid of $\mathbb{T}$ with mesh $h\cdot\min\left(H_{n,p}^{-\frac{1}{\alpha_p}}, h\cdot p_n^{\frac{1}{\alpha_K}}\right)$ and let for $t_0\in\mathbb{T}$ be t_0^* be the projection of t_0 on T_n, i.e., we have $|t_0-t_0^*| \le h\cdot\min\left(H_{n,p}^{-\frac{1}{\alpha_p}}, h\cdot p_n^{\frac{1}{\alpha_K}}\right)$. Then we obtain

$$\begin{aligned}
&\sup_{t_0\in\mathbb{T}}\left\|\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0) - \Sigma(\theta_0,t_0)\right\| \\
\le &\sup_{t_0\in\mathbb{T}}\left\|\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0) - \frac{1}{r_n\bar{p}_n(t_0^*)}\ell_n''(\theta_0,t_0^*)\right\| + \sup_{t_0\in\mathbb{T}}\|\Sigma(\theta_0,t_0)-\Sigma(\theta_0,t_0^*)\| \qquad (5.87)\\
&+\sup_{t_0\in\mathbb{T}}\left\|\frac{1}{r_n\bar{p}_n(t_0^*)}\ell_n''(\theta_0,t_0^*) - \Sigma(\theta_0,t_0^*)\right\|. \qquad (5.88)
\end{aligned}$$

The second sup in (5.87) is converging to zero because $\sup_{t_0\in\mathbb{T}}|t_0-t_0^*| \to$ as $n\to\infty$ and by uniform continuity of $t\mapsto\Sigma(\theta_0,t)$ (cf. Assumption (C2), 2) we conclude that also $\sup_{t_0\in\mathbb{T}}\|\Sigma(\theta_0,t_0)-\Sigma(\theta_0,t_0^*)\| \to 0$ as $n\to\infty$. To prove that the first part of (5.87) is $o_P(1)$, we see that for $|t_0-t_0^*| \le h\cdot\min\left(H_{n,p}^{-\frac{1}{\alpha_p}}, h\cdot p_n^{\frac{1}{\alpha_K}}\right) \le h$ we can employ the Hoelder continuity from Assumptions (B4), 2 and (B6) (use all bounds from Assumption (C1), 2 and recall that kernels are supported on $[-1,1]$ by Assumption (B4), 2)

$$\left\|\frac{1}{r_n\bar{p}_n(t_0)}\ell_n''(\theta_0,t_0) - \frac{1}{r_n\bar{p}_n(t_0^*)}\ell_n''(\theta_0,t_0^*)\right\|$$

$$
\begin{aligned}
\leq & \frac{\hat{K}^2 e^{\tau\cdot\hat{K}}}{r_n} \sum_{i\in G_n} \int_0^T \left| \frac{1}{h\cdot\bar{p}_n(t_0)} K\left(\frac{s-t_0}{h}\right) - \frac{1}{h\cdot\bar{p}_n(t_0^*)} K\left(\frac{s-t_0^*}{h}\right)\right| ds \\
\leq & \hat{K}^2 e^{\tau\cdot\hat{K}} \left(\int_{t_0-2h}^{t_0+2h} \frac{1}{h\bar{p}_n(t_0)} \left| K\left(\frac{s-t_0}{h}\right) - K\left(\frac{s-t_0^*}{h}\right)\right| ds + \left|\frac{1}{\bar{p}_n(t_0)} - \frac{1}{\bar{p}_n(t_0^*)}\right|\right) \\
\leq & \hat{K}^2 e^{\tau\hat{K}} \left(\frac{4H_K}{h^{\alpha_K} p_n} \cdot |t_0 - t_0^*|^{\alpha_K} + H_{n,p}\cdot |t_0-t_0^*|^{\alpha_p}\right) \\
\leq & \hat{K}^2 e^{\tau\hat{K}} \left(4H_K h^{\alpha_K} + h^{\alpha_p}\right) \to 0
\end{aligned}
$$

as $n\to\infty$. Finally, we have to prove that (5.88) is also $o_P(1)$. To this end, we firstly note that the sup is actually only taken over T_n because we only consider t_0^*. So we apply a standard union bound technique to get the sup out of the probability and we include $\Gamma_n^{t_0}$: Let $x>0$ and recall the Definition of $H_{n,i}(s,\theta) = -C_{n,i}(s)X_{n,i}(s)X_{n,i}(s)^T e^{\theta^T X_{n,i}(s)}$ from the proof of Proposition 5.13. Then,

$$
\begin{aligned}
& \mathbb{P}\left(\sup_{t_0\in\mathbb{T}} \left\| \frac{1}{r_n\bar{p}_n(t_0^*)} \ell_n''(\theta_0, t_0^*) - \Sigma(\theta_0, t_0^*)\right\| > x\right) \\
=& \mathbb{P}\left(\sup_{t_0\in T_h} \left\| \frac{1}{r_n\bar{p}_n(t_0)} \ell_n''(\theta_0, t_0) - \Sigma(\theta_0, t_0)\right\| > x\right) \\
=& \# T_n \cdot \sup_{t_0\in\mathbb{T}} \mathbb{P}\left(\left\| \frac{1}{r_n\bar{p}_n(t_0)} \ell_n''(\theta_0, t_0) - \Sigma(\theta_0, t_0)\right\| > x\right) \\
\leq & \# T_n \cdot \sup_{t_0\in\mathbb{T}} \Bigg[\mathbb{P}\left(\Gamma_n^{t_0} = 1\right) & (5.89) \\
& \mathbb{P}\Bigg(\Bigg\| \frac{1}{r_n\bar{p}_n(t_0)} \sum_{i\in G_n} \frac{1}{h} \int_0^T K\left(\frac{s-t_0}{h}\right) \\
& \times \left(H_{n,i}(s,\theta_0)\Gamma_n^{t_0} - \mathbb{E}\left(H_{n,i}(s,\theta_0)\Gamma_n^{t_0}\right)\right) ds \Bigg\| > \frac{x}{4}\Bigg) & (5.90) \\
& + \mathbb{P}\left(\left\| \frac{1}{\bar{p}_n(t_0)} \cdot \frac{1}{h} \int_0^T K\left(\frac{s-t_0}{h}\right) \mathbb{E}\left(H_{n,i}(s,\theta_0)\left(\Gamma_n^{t_0} - 1\right)\right) ds \right\| \geq \frac{x}{4}\right) & (5.91) \\
& + \mathbb{P}\left(\left\| \frac{1}{h} \int_0^T K\left(\frac{s-t_0}{h}\right) \Sigma(s,\theta_0) \frac{p_n(s)}{\bar{p}_n(t_0)} ds - \Sigma(\theta_0, t_0)\right\| \geq \frac{x}{2}\right)\Bigg]. & (5.92)
\end{aligned}
$$

Now, line (5.89) converges to zero by Assumption (C4) because

$$
\# T_n \leq T\cdot\left(h^{-1}\cdot H_{n,p}^{\frac{1}{\alpha_p}} + h^{-2} p_n^{-\frac{1}{\alpha_K}}\right),
$$

the probability in line (5.91) is either zero or one because the term inside is deterministic. But by Lemma 5.29 it converges to zero and hence the probability equals zero for n large enough (regardless of the size of T_n). Similarly, the probability in line (5.92) equals zero for n large enough because

$$\begin{aligned}&\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)\Sigma(s,\theta_0)\frac{p_n(s)}{\bar{p}_n(t_0)}-\Sigma(\theta_0,t_0)\\=&\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)(\Sigma(s,\theta_0)-\Sigma(t_0,\theta_0))\frac{p_n(s)}{\bar{p}_n(t_0)}ds=(5.53)=o(1).\end{aligned}$$

Finally, line (5.90) will be treated by applying Lemma 3.29 to

$$Z_{n,i}:=-\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)H_{n,i}(s,\theta_0)\Gamma_n^{\theta_0}ds.$$

Note therefore firstly that we may work element-wise because we can estimate the norm from above by the 1-norm and consider each term separately (note that the dimension of the covariates is not increasing). Thus, we may pretend for the following that $H_{n,i}(s,\theta_0)$ is a number rather than a matrix. Moreover, we can repeat the following proof word by word for $-H_{n,i}(s,\theta_0)$ and thus we may consider $H_{n,i}(s,\theta_0)$ instead of $|H_{n,i}(s,\theta)|$. In order to apply Lemma 3.29, we need to fulfil the assumptions from Definition 3.28. The first part is most importantly Statement 5.30 together with the following inequality: Let $\rho\geq 2$ and recall that $|Z_{n,i}|\leq\hat{K}^2\Lambda\Gamma_n^{t_0}$. Moreover, on $\Gamma_n^{t_0}$ we have $\sum_{i\in G_n}I_{n,i}^{k,m}\leq S_k\leq E_k^{n,t_0}$. Thus,

$$\begin{aligned}&\mathbb{E}\left(\left|U_{k,m}^{n,t_0}\right|^\rho\right)\\\leq&\mathbb{E}\left(\left|U_{k,m}^{n,t_0}\right|^2\cdot\left|\sum_{i\in G_n}(Z_{n,i}I_{n,i}^{k,m}-\mathbb{E}(Z_{n,i}I_{n,i}^{k,m}))\right|^{\rho-2}\right)\\\leq&\mathbb{E}\left(\left|U_{k,m}^{n,t_0}\right|^2\right)\cdot\left|2\hat{K}^2\Lambda\cdot E_k^{n,t_0}\right|^{\rho-2}.\end{aligned}$$

Now, Statement 5.30 yields that there is a $\sigma^2>0$ such that the first part of Definition 3.28 holds true with $c_1=2\hat{K}^2\Lambda$. The second part holds by Assumption (C4). Note that $|E|_{n,t_0}=r_n\bar{p}_n(t_0)\to\infty$, then for any $x,c>0$ there is $N\in\mathbb{N}$ such that for $n\geq N$, $\frac{x}{4}\geq c\sqrt{\frac{\log|E|n,t_0}{|E|_{n,t_0}}}$. So we get by Lemma 3.29

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}(Z_{n,i}-\mathbb{E}(Z_{n,i}))\geq\frac{x}{4}\right)\\\leq&\mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}(Z_{n,i}-\mathbb{E}\left(Z_{n,i}\right))\geq c\cdot\sqrt{\frac{\log|E|_{n,t_0}}{|E|_{n,t_0}}}\right)\\\leq&\mathcal{K}\exp\left(-\log r_n\bar{p}_n(t_0)\cdot\frac{c_2\cdot c^2}{2\left(\sigma^2+2\hat{K}\Lambda c_3\cdot c\right)}\right)+\beta_{t_0}(\Delta_n)\cdot\mathcal{K}r_n.\end{aligned}$$

By Assumption (C4), we find that $\beta_{t_0}(\Delta_n) \cdot \mathcal{K} r_n \cdot \# T_n$ converges to zero. Also we see that for c chosen large enough, the exponential above converges faster to zero than $\# T_n$ (cf. Assumption (C1), 4). In total we have thus shown that (5.90) converges to zero. Thus, we have finally shown

$$\sup_{t_0 \in \mathbb{T}} \left\| \frac{1}{r_n p_n(t_0)} \ell_n''(\theta_0, t_0) - \Sigma(\theta_0, t_0) \right\| \to 0$$

in probability as $n \to \infty$. This finalizes the proof of the Statement. □

Proposition 5.24. *Suppose that Assumptions (B6), (C1), 1 and 2, (C4) and (C7), (4.32) hold. Then, Statement 4.9 holds true.*

Proof. We note firstly that by Assumption (C1), 2 for all $t_0 \in \mathbb{T}$ and all $\theta_1, \theta_2 \in \Theta$ we can estimate by a Taylor approximation

$$\begin{aligned} &\left\| X_{n,i}(s) X_{n,i}(s)^T C_{n,i}(s) \left(e^{\theta_1^T X_{n,1}(s)} - e^{\theta_2^T X_{n,i}(s)} \right) \right\| \\ \leq & \hat{K}^3 e^{\tau \hat{K}} \|\theta_1 - \theta_2\|. \end{aligned} \tag{5.93}$$

Hence, we obtain for all $t_0 \in \mathbb{T}$ and all $\theta_1, \theta_2 \in \Theta$

$$\begin{aligned} &\frac{1}{r_n \bar{p}_n(t_0)} \left\| \ell_n''(\theta_1, t_0) - \ell_n''(\theta_2, t_0) \right\| \\ \leq & \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h}\right) C_{n,i}(s) \left\| X_{n,i}(s) X_{n,i}(s)^T \right\| \cdot \left| e^{\theta_1^T X_{n,i}(s)} - e^{\theta_2^T X_{n,i}(s)} \right| ds \\ \leq & \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h}\right) C_{n,i}(s) ds \cdot \hat{K}^3 e^{\tau \hat{K}} \cdot \|\theta_1 - \theta_2\|. \end{aligned}$$

Hence, we can choose $K_n := \sup_{t_0 \in \mathbb{T}} \hat{K}^3 e^{\tau \hat{K}} \cdot \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h}\right) C_{n,i}(s) ds$ which is $O_P(1)$ if

$$\sup_{t_0 \in \mathbb{T}} \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h}\right) C_{n,i}(s) ds = O_P(1). \tag{5.94}$$

So let us prove this. Denote therefore by T_n a grid with mesh $h \cdot \min\left\{ H_{n,p}^{-\frac{1}{\alpha_P}}, h \cdot p_n^{\frac{1}{\alpha_K}} \right\}$ which covers $\mathbb{T}$. For a given time $t_0 \in \mathbb{T}$ we denote by $t_0^* \in T_h$ the closest element of T_n to t_0, i.e., $|t_0 - t_0^*| \leq h \cdot \min\left\{ H_{n,p}^{-\frac{1}{\alpha_P}}, h \cdot p_n^{\frac{1}{\alpha_K}} \right\}$. Now we split the sup over an uncountable set as usual in a sup over close elements and a sup over a finite set:

$$\sup_{t_0 \in \mathbb{T}} \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s - t_0}{h}\right) C_{n,i}(s) ds$$

$$\leq \sup_{t_0\in\mathbb{T}} \int_0^T \frac{1}{r_n h}\left[\frac{1}{\bar{p}_n(t_0)}K\left(\frac{s-t_0}{h}\right) - \frac{1}{\bar{p}_n(t_0^*)}K\left(\frac{s-t_0^*}{h}\right)\right]\sum_{i\in G_n} C_{n,i}(s)ds$$
$$+ \sup_{t_0\in\mathbb{T}} \int_0^T \frac{1}{r_n\bar{p}_n(t_0^*)h}K\left(\frac{s-t_0^*}{h}\right)\sum_{i\in G_n} C_{n,i}(s)ds$$
$$\leq \sup_{t_0\in\mathbb{T}} \int_0^T \frac{1}{r_n h}\left|\frac{1}{\bar{p}_n(t_0)} - \frac{1}{\bar{p}_n(t_0^*)}\right| K\left(\frac{s-t_0}{h}\right)\sum_{i\in G_n} C_{n,i}(s)ds \tag{5.95}$$
$$+ \sup_{t_0\in\mathbb{T}} \int_0^T \frac{1}{r_n\bar{p}_n(t_0^*)h}\left|K\left(\frac{s-t_0}{h}\right) - K\left(\frac{s-t_0^*}{h}\right)\right|\sum_{i\in G_n} C_{n,i}(s)ds \tag{5.96}$$
$$+ \sup_{t_0\in\mathbb{T}} \int_0^T \frac{1}{r_n\bar{p}_n(t_0^*)h}K\left(\frac{s-t_0^*}{h}\right)\sum_{i\in G_n} C_{n,i}(s)ds. \tag{5.97}$$

We apply the Hoelder continuity of $\frac{1}{\bar{p}_n(t)}$(cf. Assumption (B6)) together with $|t_0 - t_0^*| \leq hH_{n,p}^{-\frac{1}{\alpha_p}}$ in order to see that

$$(5.95) \leq \sup_{t_0\in\mathbb{T}} \int_0^T \frac{1}{h}K\left(\frac{s-t_0}{h}\right)ds \cdot H_{n,p}|t_0 - t_0^*|^{\alpha_p} \leq h^{\alpha_p}.$$

And hence $(5.95) = O(1)$. Since the kernel is supported on $[-1, 1]$, we see that

$$\left|K\left(\frac{s-t_0}{h}\right) - K\left(\frac{s-t_0^*}{h}\right)\right| \leq 2K\mathbb{1}(s \in [t_0 - h, t_0 + h] \cup [t_0^* - h, t_0^* + h]).$$

As $|t_0 - t_0^*| \leq h^2 \cdot p_n^{\frac{1}{\alpha_K}} \leq h$, we find that

$$\begin{aligned}(5.96) \leq& \sup_{t_0\in\mathbb{T}} \int_{t_0-2h}^{t_0+2h} \frac{H_K}{\bar{p}_n(t_0^*)h}\frac{|t_0 - t_0^*|^{\alpha_K}}{h^{\alpha_K}}ds\\ \leq& \frac{4H_K}{p_n}p_n h^{\alpha_K}\\ \leq& 4H_K h^{\alpha_K} \to 0\end{aligned}$$

as $n \to \infty$. Finally, we show now the proof for (5.97). We begin with extracting the sup outside the probability and inserting $\Gamma_n^{t_0}$:

$$\begin{aligned}&\mathbb{P}((5.97) > x)\\ =&\mathbb{P}\left(\sup_{t_0\in\mathbb{T}} \frac{1}{r_n\bar{p}_n(t_0^*)h}\int_0^T K\left(\frac{s-t_0^*}{h}\right)\sum_{i\in G_n} C_{n,i}(s)ds > x\right)\\ \leq&\#T_n \cdot \sup_{t_0\in T_n} \mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)h}\int_0^T K\left(\frac{s-t_0}{h}\right)\sum_{i\in G_n} C_{n,i}(s)ds > x\right)\\ \leq&\#T_n \cdot \sup_{t_0\in T_n} \Bigg[\mathbb{P}(\Gamma_n^{t_0} = 0)\end{aligned}$$

$$+\mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)h}\int_0^T K\left(\frac{s-t_0}{h}\right)\sum_{i\in G_n} C_{n,i}(s)\Gamma_n^{t_0}ds > x\right)\Bigg]$$

$$\leq \#T_n \cdot \sup_{t_0\in T_n}\Bigg[\mathbb{P}(\Gamma_n^{t_0}=0) \tag{5.98}$$

$$+\mathbb{P}\Bigg(\frac{1}{r_n\bar{p}_n(t_0)h}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)$$

$$\times\left(C_{n,i}(s)\Gamma_n^{t_0}-\mathbb{E}(C_{n,i}(s)\Gamma_n^{t_0})\right)ds > \frac{x}{2}\Bigg) \tag{5.99}$$

$$+\mathbb{P}\left(\frac{1}{\bar{p}_n(t_0)h}\int_0^T K\left(\frac{s-t_0}{h}\right)\mathbb{E}(C_{n,1}(s)\Gamma_n^{t_0})ds > \frac{x}{2}\right)\Bigg]. \tag{5.100}$$

For line (5.100), we note that the expression in the probability is deterministic and that it can be bounded from above by $\frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)\frac{p_n(s)}{\bar{p}_n(t_0)}ds = 1$ and hence the probability equals zero for x large enough. Line (5.98) converges to zero by Assumption (C4) because $\#T_n = O\left(h^{-1}\cdot H_{n,p}^{\frac{1}{\alpha_p}} + h^{-2}p_n^{-\frac{1}{\alpha_K}}\right)$. For line (5.99), we will apply Lemma 3.29 to

$$Z_{n,i} = \frac{1}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)\Gamma_n^{t_0}C_{n,i}(s)ds.$$

We need to fulfil the requirements in Definition 3.28. The second part is true by Assumption (C4) but we have to prove the moment bound in the first part of Definition 3.28. By definition of $\Gamma_n^{t_0}$, we obtain for natural numbers $\rho\geq 2$

$$\begin{aligned}
&\mathbb{E}\left(\left|U_{k,m}^{n,t_0}\right|^{\rho}\right)\\
\leq&\mathbb{E}\left(\left(\sum_{i\in G_n}\frac{1}{h}\int_0^T K\left(\frac{s-t}{h}\right)(\Gamma_n^{t_0}C_{n,i}(s)I_{n,i}^{k,m}-\mathbb{E}(\Gamma_n^{t_0}C_{n,i}(s)I_{n,i}^{k,m}))ds\right)^2\right)\\
&\times\left(2E_k^{n,t}\right)^{\rho-2}\\
\leq&\left(\sum_{\substack{i,j\in G_n\\|e_i\cap e_j|=0}}+\sum_{\substack{i,j\in G_n\\|e_i\cap e_j|=1}}+\sum_{\substack{i,j\in G_n\\|e_i\cap e_j|=2}}\right)\frac{1}{h^2}\iint_{[0,T]^2}K\left(\frac{s-t_0}{h}\right)K\left(\frac{t-t_0}{h}\right)\\
&\qquad\times\mathrm{Cov}(C_{n,i}(s)I_{n,i}^{k,m}\Gamma_n^{t_0},C_{n,j}(t)I_{n,j}^{k,m}\Gamma_n^{t_0})dsdt\\
&\times\left(2E_k^{n,t}\right)^{\rho-2}.
\end{aligned}$$

In the case $|e_i\cap e_j|=2$, i.e., if $i=j$, we see that

$$\sum_{i\in G_n}\frac{1}{h^2}\iint_{[0,T]^2}K\left(\frac{s-t_0}{h}\right)K\left(\frac{t-t_0}{h}\right)$$

$$
\begin{aligned}
&\times Cov(C_{n,i}(s)I_{n,i}^{k,m}\Gamma_n^{t_0}, C_{n,j}(t)I_{n,j}^{k,m}\Gamma_n^{t_0})dsdt\\
\leq &\frac{r_n}{h}\int_0^T K\left(\frac{s-t_0}{h}\right)\mathbb{E}\left(I_{n,1}^{k,m}C_{n,1}(s)\right)ds\\
= &E_{k,m}^{n,t_0}
\end{aligned}
$$

In the other two cases, we quote Assumption (C7): Let $\nu \in \{0,1\}$ and let $i_0, j_0 \in G_n$ with $|e_{i_0} \cap e_{j_0}| = \nu$. Note that the number of terms in the sum $\sum_{\substack{i,j\in G_n\\ |e_i\cap e_j|=\nu}}$ is smaller than $n^{4-\nu}$. Then,

$$
\begin{aligned}
&\frac{1}{E_{k,m}^{n,t_0}}\sum_{\substack{i,j\in G_n\\ |e_i\cap e_j|=\nu}}\frac{1}{h^2}\iint_{[0,T]^2}K\left(\frac{s-t_0}{h}\right)K\left(\frac{t-t_0}{h}\right)\\
&\qquad\qquad\times \mathrm{Cov}\left(C_{n,i}(s)I_{n,i}^{k,m}\Gamma_n^{t_0}, C_{n,j}(t)I_{n,j}^{k,m}\Gamma_n^{t_0}\right)dsdt\\
\leq &\frac{n^{4-\nu}}{r_n}\frac{1}{h^2}\iint_{[0,T]^2}K\left(\frac{s-t_0}{h}\right)K\left(\frac{t-t_0}{h}\right)\frac{\mathrm{Cov}\left(C_{n,i_0}(s)I_{n,i_0}^{k,m}\Gamma_n^{t_0}, C_{n,j_0}(t)I_{n,j_0}^{k,m}\Gamma_n^{t_0}\right)}{\bar{p}_n^{k,m}(t_0)}dsdt
\end{aligned}
$$

which is bounded by Assumption (C7) in (4.32). Hence, the first part of Definition 3.28 is fulfilled with $c_1 = 2$ and for some $\sigma > 0$. So we may apply Lemma 3.29 to the probability in (5.99) (note that here $|E|_{n,t_0} = r_n\bar{p}_n(t_0)$).

$$
\begin{aligned}
&(5.99)\\
=&\mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}Z_{n,i} > x\right)\\
\leq&\mathcal{K}\cdot\exp\left(-\log r_n\bar{p}_n(t_0)\cdot\frac{c_2x^2\frac{r_n\bar{p}_n(t_0)}{\log r_n\bar{p}_n(t_0)}}{2\left(\sigma^2+2c_3x\cdot\sqrt{\frac{r_n\bar{p}_n(t_0)}{\log r_n\bar{p}_n(t_0)}}\right)}\right)+\beta_{t_0}(\Delta_n)\cdot\mathcal{K}r_n\\
\leq&\mathcal{K}\cdot\exp\left(-\frac{c_2x^2\sqrt{r_n\bar{p}_n(t_0)}}{2\left(\sigma^2+2c_3x\right)}\right)+\beta_{t_0}(\Delta_n)\cdot\mathcal{K}r_n.
\end{aligned}
$$

By Assumption (C4), we see that the above expression converges to zero because $\#T_h = O\left(h^{-1}\cdot H_{n,p}^{\frac{1}{\alpha_p}} + h^{-2}p_n^{-\frac{1}{\alpha_K}}\right)$ and this completes the proof. □

Proposition 5.25. *Suppose that Assumption (C1) 4 holds. Then, Statement 4.10 holds true.*

Proof. We begin by proving the Hoelder continuity of $\widetilde{H}_{n,i}(s,\theta)$ in θ. In fact, $\widetilde{H}_{n,i}(s,\theta)$ is even Lipschitz continuous because we can apply (5.93) to obtain for $s \in [0,T]$ and $\theta_1, \theta_2 \in \Theta$

$$
\left\|\widetilde{H}_{n,i}(s,\theta_1) - \widetilde{H}_{n,i}(s,\theta_2)\right\|
$$

$$\leq \left\| X_{n,i}(s)X_{n,i}(s)^T C_{n,i}(s) \left(e^{\theta_1^T X_{n,1}(s)} - e^{\theta_2^T X_{n,i}(s)} \right) \right\|$$
$$+ \mathbb{E} \left(\left\| X_{n,i}(s)X_{n,i}(s)^T C_{n,i}(s) \left(e^{\theta_1^T X_{n,1}(s)} - e^{\theta_2^T X_{n,i}(s)} \right) \right\| \right)$$
$$\leq 2\hat{K}^3 e^{\tau \hat{K}} \cdot \|\theta_1 - \theta_2\|.$$

And hence we may choose $\gamma_{n,i}(s) := 2\hat{K}^3 e^{\tau \hat{K}}$. For the rates which we want to prove, we note firstly that all terms on the left are bounded and so we just have to choose k_0 so large such that the terms on the right tend to infinity which is always possible by Assumption (C1), 4. □

Proposition 5.26. *Suppose that Assumptions (C1), 1 and 2 and (C4) hold. Then, Statement 4.11 holds true.*

Proof. By employing Lemma 5.28 the proof of this result is fairly straight forward. Let c^{**} be the constant such that $\|y\| \leq c^{**} \cdot \|y\|_1$ for all $y \in \mathbb{R}^q$ where $\|.\|$ and $\|.\|_1$ denote the Euclidean- and the 1-Norm, respectively. We have

$$\mathbb{P}\left(\left\| \frac{\ell_n'(\theta_0, t_0)}{r_n \sqrt{\bar{p}_n(t_0)}} \right\| \geq C \cdot \sqrt{\frac{\log r_n}{r_n h}} \right)$$
$$\leq \mathbb{P}\left(\left\| \frac{\ell_n'(\theta_0, t_0)}{r_n \bar{p}_n(t_0)} \right\| \geq \frac{C}{qc^{**}\sqrt{h}} \cdot qc^{**} \sqrt{\frac{\log r_n \bar{p}_n(t_0)}{r_n \bar{p}_n(t_0)}} \right). \tag{5.101}$$

Since
$$\ell_n'(\theta_0, t_0) = \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{t - t_0}{h}\right) X_{n,i}(t) dM_{n,i}(t),$$
we can directly apply Lemma 5.28 and obtain

$$(5.101)$$
$$\leq 2q \left(\mathcal{K} \exp\left(-\frac{c_2 \frac{C^2}{q^2 (c^{**})^2}}{2K\hat{K}A\left(\Lambda A + \sqrt{\frac{\Lambda}{2}} c_3 \cdot \frac{C}{qc^{**}}\right)} \cdot \log n p_n(t_0 - h) \right) \right.$$
$$\left. + \beta_{t_0}(\Delta_n) \cdot \mathcal{K} r_n + \mathbb{P}(\Gamma_n^{t_0} = 0) \right).$$

We see that for a sufficiently largely chosen C the first term decreases faster as hn^{k_0}. Moreover, by Assumption (C4), $\beta(\Delta_n)$ decrease fast enough too. Finally, $\mathbb{P}(\Gamma_n^{t_0} = 0)$ decreases fast enough as well by the same assumption. □

Proposition 5.27. *Suppose that Assumptions (C1), 1, 2 and 4, (C4) and (C7),* (4.30) *and* (4.31) *hold. Then, Statement 4.12 holds true.*

Proof. Let $\delta_n := \sqrt{\frac{\log r_n p_n}{r_n p_n \cdot h}}$. We begin with extracting the sup from the probability by a standard union bound argument during which we include $\Gamma_n^{t_0}$

$$\begin{aligned}
&\mathbb{P}\Bigg(\sup_{(t_0,\theta)\in T_{n,k_0}} \Bigg| \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \\
&\qquad \times (H_{n,i}(s,\theta) - \mathbb{E}(H_{n,i}(\theta,s)))\, ds \Bigg| > C\delta_n \Bigg) \\
\leq &\#T_{n,k_0} \cdot \sup_{(t_0,\theta)\in T_{n,k_0}} \\
&\mathbb{P}\Bigg(\Bigg| \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \\
&\qquad \times (H_{n,i}(s,\theta) - \mathbb{E}(H_{n,i}(\theta,s)))ds \Bigg| > C\delta_n \Bigg) \\
\leq &\#T_{n,k_0} \cdot \sup_{(t_0,\theta)\in T_{n,k_0}} \Bigg[\mathbb{P}(\Gamma_n^{t_0} = 0) \\
&+ \mathbb{P}\Bigg(\Bigg| \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \\
&\qquad \times (H_{n,i}(s,\theta)\Gamma_n^{t_0} - \mathbb{E}(H_{n,i}(\theta,s)))ds \Bigg| > C\delta_n \Bigg) \Bigg] \\
\leq &\#T_{n,k_0} \cdot \sup_{(t_0,\theta)\in T_{n,k_0}} \Bigg[\mathbb{P}(\Gamma_n^{t_0} = 0) \qquad (5.102) \\
&+ \mathbb{P}\Bigg(\Bigg| \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \\
&\qquad \times (H_{n,i}(s,\theta)\Gamma_n^{t_0} - \mathbb{E}(H_{n,i}(\theta,s)\Gamma_n^{t_0}))ds \Bigg| > \frac{C}{2}\delta_n \Bigg) \qquad (5.103) \\
&+ \mathbb{P}\Bigg(\Bigg| \frac{1}{r_n \bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \\
&\qquad \times \mathbb{E}(H_{n,i}(\theta,s)(\Gamma_n^{t_0} - 1))ds \Bigg| > \frac{C}{2}\delta_n \Bigg) \Bigg]. \qquad (5.104)
\end{aligned}$$

By Statement 5.29, (5.104) has the correct rate even after taking the sup (note that (5.104) is zero or one because the expression inside is deterministic, so we require here that it holds for all time points). The probability (5.102) converges to zero by Assumption (C4) because $\#T_{n,k_0} = O\left(hn^{k_0(q+1)}\right)$. For (5.103), we intent to use Lemma 3.29

with

$$Z_{n,i} := \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) H_{n,i}(s,\theta)\Gamma_n^{t_0} ds.$$

Define to this end

$$\Sigma_{n,i}^{k,m}(s,\theta) \quad := \quad \mathbb{E}(H_{n,i}(s,\theta) I_{n,i}^{k,m}\Gamma_n^{t_0}).$$

The conditions on $E_{k,m}^{n,t_0}$ and E_k^{n,t_0} in part 2 of Definition 3.28 hold by Definition and Assumption (C4). We need to prove the moment bound, which holds by the boundedness Assumption (C1), 2 and definition of $\Gamma_n^{t_0}$: Let k,m be given and let $\rho \geq 2$ be a natural number. Then

$$\begin{aligned}
&\mathbb{E}\left(\left|U_{k,m}^{n,t_0-h}(\Delta_n)\right|^\rho\right) \\
=&\mathbb{E}\left(\left|\sum_{i\in G_n}\int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) H_{n,i}(s,\theta) I_{n,i}^{k,m}\Gamma_n^{t_0} - \Sigma_{n,i}^{k,m}(s,\theta) ds\right|^\rho\right] \\
\leq&\mathbb{E}\left(\left|\sum_{i\in G_n}\int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) H_{n,i}(s,\theta) I_{n,i}^{k,m}\Gamma_n^{t_0} - \Sigma_{n,i}^{k,m}(s,\theta) ds\right|^2\right. \\
&\left.\times\left(\int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \hat{K}^2\Lambda\left(S_k\Gamma_n^{t_0} + \mathbb{E}(S_k\Gamma_n^{t_0})\right) ds\right)^{\rho-2}\right) \\
\leq&\mathbb{E}\left(\left|\sum_{i\in G_n}\int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) H_{n,i}(s,\theta) I_{n,i}^{k,m}\Gamma_n^{t_0} - \Sigma_{n,i}^{k,m}(s,\theta) ds\right|^2\right) \cdot \left(2\hat{K}^2\Lambda E_k^{n,t_0}\right)^{\rho-2} \\
\leq&\sigma^2 E_{k,m}^{n,t_0} \cdot \left(2\hat{K}^2\Lambda E_k^{n,t_0}\right)^{\rho-2},
\end{aligned}$$

by Statement 5.30. So condition 1 in Definition 3.28 is fulfilled as well. Moreover, $|E|_{n,t_0-h} = r_n\bar{p}_n(t_0) \geq r_n \geq n^\psi$ by Assumption (C1), 4. Hence, by Lemma 3.29, there is a C large enough such that the probability in (5.103) decays faster than $\#T_{n,k_0} = O\left(hn^{k_0(q+1)}\right)$ grows. Moreover, by choice of $\Delta_n = a\log n$, we find by Assumption (C4) that $\beta(\Delta_n)$ converges also quicker to zero than $\#T_{n,k_0}$ grows to infinity and this proves the claim. □

5.3.1 Further Supporting Lemmas

The following Lemma provides the inequality necessary for proving Statements 4.11 and 4.12.

Lemma 5.28. *Suppose that (C1), 1 and 2 and (C4) hold. Recall the following definitions: Λ is the bound on the intensity function, K the bound on the kernel and $\hat{K}$ the bound on the covariates. Let furthermore $c_3 > 0$ such that:*

$$E_{k,m}^{n,t_0} \quad := \quad r_n \int_{t_0-h}^{t_0+h} \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \mathbb{E}\left[I_{n,1}^{k,m} C_{n,1}(s)\right] ds$$

$$\begin{aligned}
\bar{p}_n(t_0) &:= \int_{t_0-h}^{t_0+h} \frac{1}{h} K\left(\frac{s-t_0}{h}\right) p_n(s)ds \\
E_k^{n,t_0} &:= \sqrt{\frac{r_n \bar{p}_n(t_0)}{\log r_n \bar{p}_n(t_0)}} \cdot c_3 \\
\sigma^2 &:= \frac{1}{h} \Lambda A^2 K \hat{K} \\
c_1 &:= \sqrt{\frac{\Lambda}{2h}} K \hat{K} A \\
I_{n,i}^{k,m} &:= \mathbb{1}\Big(i \in G^{t_0-h}(k,m,\Delta_n)\Big) \\
S_{k,m} &:= \sum_{i \in G_n} \int_{t_0-h}^{t_0+h} \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \mathbb{E}\left(I_{n,i}^{k,m} C_{n,i}(s) \Big| \mathcal{F}_{t_0-h}^n\right) ds \\
S_k &:= \max_{m=1,\ldots,r_n} \sum_{i \in G_n} I_{n,i}^{k,m} \\
\Gamma_n^{t_0} &:= \mathbb{1}\left(\frac{S_k^2 \cdot \log r_n \bar{p}_n(t_0)}{r_n \bar{p}_n(t_0)} \leq c_3^2, S_k \sqrt{h} \geq 1\right)
\end{aligned}$$

Assume for all $k \in \{1, ..., \mathcal{K}\}$

$$\frac{1}{r_n \bar{p}_n(t_0)} \sum_{m=1}^{r_n} E_{k,m}^{n,t_0} \geq c_2.$$

Let furthermore $A > 0$ be so large such that

$$A \geq \max\left\{\sqrt{\hat{K}}, \hat{K}, \frac{1}{K}, \sqrt{2^{\frac{3}{2}} \sqrt{\Lambda} \frac{\hat{K}}{K}}\right\}, \ \frac{1}{A} \cdot \exp\left(\frac{\sqrt{2}}{A\sqrt{\Lambda}}\right) \leq 1. \tag{5.105}$$

Then it holds

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{r_n \bar{p}_n(t_0)} \left\| \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) X_{n,i}(s) dM_{n,i}(s) \right\| \geq xqc^{**} \sqrt{\frac{\log r_n \bar{p}_n(t_0)}{r_n \bar{p}_n(t_0)}}\right) \\
\leq\ & 2q\mathcal{K} \left[r_n p_n(t_0)\right]^{-\frac{c_2 x^2}{2(\sigma^2 + c_1 c_3 x)}} + 2q\beta_{t_0}(\Delta_n) \cdot \mathcal{K} r_n + 2q\mathbb{P}(\Gamma_n^{t_0} = 0),
\end{aligned}$$

where q is the dimension of the covariate and c^{**} is the constant for which $\|y\| \leq c^{**}\|y\|_1$ for all $y \in \mathbb{R}^q$ and $\|.\|$ and $\|.\|_1$ are the Euclidean and 1-Norm respectively.

Proof. We remark firstly that it is sufficient to consider univariate covariates, because (denote by $X_{n,i}^r$ the r-th entry of $X_{n,i}$ for $r = 1, ..., q$)

$$\mathbb{P}\left(\frac{1}{r_n \bar{p}_n(t_0)} \left\| \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) X_{n,i}(s) dM_{n,i}(s) \right\| \geq xqc^{**} \sqrt{\frac{\log r_n \bar{p}_n(t_0)}{r_n \bar{p}_n(t_0)}}\right)$$

$$\le \sum_{r=1}^{q} \mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)} \left| \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) X_{n,i}^r(s) dM_{n,i}(s) \right| \ge x\sqrt{\frac{\log r_n\bar{p}_n(t_0)}{r_n\bar{p}_n(t_0)}} \right)$$
$$\le \sum_{r=1}^{q} \mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) X_{n,i}^r(s) dM_{n,i}(s) \ge x\sqrt{\frac{\log r_n\bar{p}_n(t_0)}{r_n\bar{p}_n(t_0)}} \right)$$
$$+ \mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) (-X_{n,i}^r(s)) dM_{n,i}(s) \ge x\sqrt{\frac{\log r_n\bar{p}_n(t_0)}{r_n\bar{p}_n(t_0)}} \right).$$

Since $-X_{n,i}^r$ is a covariate with the exact same properties as $X_{n,i}$ (in particular predictability with respect to $\mathcal{F}_t^n$ and boundedness by $\hat{K}$, cf. Assumption (C1), 2), it is sufficient to assume (for simplicity of notation) that $X_{n,i}$ is univariate and to prove that

$$\mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)} \sum_{i\in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) X_{n,i}(s) dM_{n,i}(s) \ge x\sqrt{\frac{\log r_n\bar{p}_n(t_0)}{r_n\bar{p}_n(t_0)}} \right)$$
$$\le \mathcal{K}\left[r_n p_n(t_0)\right]^{-\frac{c_2x^2}{2(\sigma^2+c_1c_3x)}} + \beta_{t_0}(\Delta_n)\cdot \mathcal{K} r_n + \mathbb{P}(\Gamma_n^{t_0}=0). \tag{5.106}$$

The main idea of the proof is to apply Lemma 3.29 to the correct structured interaction network (in the sense of Definition 3.1). Define to this end

$$\widetilde{F}_{n,i}(s) := \frac{1}{h} K\left(\frac{s-t_0}{h}\right) X_{n,i}(s)\cdot \Gamma_n^{t_0}$$
$$F_{n,i}^{k,m}(s) := \widetilde{F}_{n,i}(s)\cdot I_{n,i}^{k,m}$$

and

$$Z_{n,i}(t) := \int_0^t \widetilde{F}_{n,i}(s) dM_{n,i}(s).$$

Note that both, $\widetilde{F}_{n,i}(s)$ and $F_{n,i}^{k,m}(s)$, are predictable processes because they are deterministically equal to zero for $s \le t_0 - h$ and the sets $t \mapsto G^t(k,m,\Delta_n)$ are predictable with respect to $\mathcal{F}_t^n$. Hence, $Z_{n,i}(t)$ is a martingale. We are going to prove that $(Z_{n,i}(T))_{i=1,\dots,r_n}$ is a mixing network process as defined in Definition 3.28. By martingale properties

$$\mathbb{E}\left(\mathbb{E}\left(\int_0^T F_{n,i}(s) dM_{n,i}(s) \middle| \mathcal{F}_{t_0-h}\right)\right) = 0$$

and we may apply Lemma 3.29 if the conditions in Definition 3.28 are satisfied. Note furthermore that for each edge i there is exactly one block (k,m) to which i belongs, i.e., such that $i \in G^{t_0-h}(k,m,\Delta_n)$. Thus, we have

$$|E|_{n,t_0-h} = \sum_{k=1}^{\mathcal{K}} \sum_{m=1}^{r_n} E_{k,m}^{n,t_0}$$
$$= r_n \int_{t_0-h}^{t_0+h} \mathbb{E}(C_{n,i}(s)) ds = r_n\bar{p}_n(t_0).$$

Condition 2 in Definition 3.28 is now clear by definition of E_k^{n,t_0} and Assumption (C4).

The main part of this proof is now to prove condition 1. Note therefore firstly that

$$\mathbb{E}\left(\int_0^T \widetilde{F}_{n,i}(s)dM_{n,i}(s)I_{n,i}^{k,m}\right) = \mathbb{E}\left(\underbrace{\mathbb{E}\left(\int_0^T \widetilde{F}_{n,i}(s)dM_{n,i}(s)\middle|\mathcal{F}_{t_0-h}^n\right)}_{=0} I_{n,i}^{k,m}\right) = 0.$$

Hence, we need to show

$$\begin{aligned}
&\mathbb{E}\left(\left|U_{k,m}^{n,t_0-h}(v,v',\Delta_n)\right|^\rho\right)\\
=&\mathbb{E}\left(\left|\sum_{i\in G_n}\int_0^t F_{n,i}^{k,m}(s)dM_{n,i}(s)\right|^\rho\right)\\
\leq&\frac{\rho!}{2}\cdot E_{k,m}^{n,t_0}\sigma^2\cdot\left(E_k^{nt_0-h}c_1\right)^{\rho-2}.
\end{aligned} \tag{5.107}$$

The idea of the proof is to prove a recursion inequality for the moments of stochastic integrals by applying Itô's Formula and then using induction. Note that $F_{n,i}^{k,m}(s) = 0$ for $s \notin [t_0 - h, t_0 + h]$. Therefore, (5.107) holds trivially for $t \leq t_0 - h$ and it holds for $t \geq t_0+h$ when it holds for $t = t_0+h$. Hence, we can restrict to the case $t \in [t_0-h, t_0+h]$. We will use the forms of the optional variation processes of the martingale $M_{n,i}$ and the stochastic integrals with respect to it given in Section 2.1.

For $\rho \geq 2$ we have that the function $f_\rho(x) := |x|^\rho$ is twice continuously differentiable and hence also $\widetilde{f}_\rho(x_1, ..., x_m) := f_\rho(x_1 + ... + x_m)$ is twice continuously differentiable. So by the multivariate Itô Formula for semi martingales with jumps given in Theorem 2.9 and the fact that with probability one no two counting processes jump at the same time, we obtain for $\rho \geq 2$:

$$\begin{aligned}
&\left|\sum_{i\in G_n}\int_0^t F_{n,i}^{k,m}(\tau)dM_{n,i}(\tau)\right|^\rho\\
=&\ \widetilde{f}_\rho\left(\int_0^t F_{n,1}^{k,m}(\tau)dM_{n,1}(\tau), ..., \int_0^t F_{n,n}^{k,m}(\tau)dM_{n,n}(\tau)\right)\\
=&\ \sum_{i\in G_n}\int_0^t \partial_i\widetilde{f}_\rho\left(\int_0^{s-} F_{n,1}^{k,m}(\tau)dM_{n,1}(\tau), ..., \int_0^{s-} F_{n,n}^{k,m}(\tau)dM_{n,n}(\tau)\right)F_{n,i}^{k,m}(s)dM_{n,i}(s)\\
&+\frac{1}{2}\sum_{i,j\in G_n}\int_0^t \partial_{ij}\widetilde{f}_\rho\left(\int_0^{s-} F_{n,1}^{k,m}(\tau)dM_{n,1}(\tau), ..., \int_0^{s-} F_{n,n}^{k,m}(\tau)dM_{n,n}(\tau)\right)\\
&\qquad F_{n,i}^{k,m}(s)F_{n,j}^{k,m}(s)d[M_{n,i}, M_{n,j}(s)](s)\\
&+\int_0^t \widetilde{f}_\rho\left(\int_0^s F_{n,1}^{k,m}(\tau)dM_{n,1}(\tau), ..., \int_0^s F_{n,n}^{k,m}(\tau)dM_{n,n}(\tau)\right)\\
&\qquad-\widetilde{f}_\rho\left(\int_0^{s-} F_{n,1}^{k,m}(\tau)dM_{n,1}(\tau), ..., \int_0^{s-} F_{n,n}^{k,m}(\tau)dM_{n,n}(\tau)\right)
\end{aligned}$$

$$
\begin{aligned}
& -\sum_{i\in G_n} \partial_i \widetilde{f}_\rho \left(\int_0^{s-} F_{n,1}^{k,m}(\tau) dM_{n,1}(\tau), ..., \int_0^{s-} F_{n,n}^{k,m}(\tau) dM_{n,n}(\tau) \right) F_{n,i}^{k,m}(s) \Delta N_{n,i}(s) \\
& -\frac{1}{2} \sum_{i,j}^{n} \partial_{ij} \widetilde{f}_\rho \left(\int_0^{s-} F_{n,1}^{k,m}(\tau) dM_{n,1}(\tau), ..., \int_0^{s-} F_{n,n}^{k,m}(\tau) dM_{n,nn}(\tau) \right) \\
& \qquad \times F_{n,i}^{k,m}(s) F_{n,j}^{k,m}(s) \Delta N_{n,i}(s) \Delta N_{n,j}(s) d\left(\sum_{r\in G_n} N_{n,r} \right)(s) \\
= & \sum_{i\in G_n} \int_0^t f'_\rho \left(\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) F_{n,i}^{k,m}(s) dM_{n,i}(s) \\
& +\frac{1}{2} \sum_{i\in G_n} \int_0^t f''_\rho \left(\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) F_{n,i}^{k,m}(s)^2 dN_{n,i}(s) \\
& + \int_0^t f_\rho \left(\sum_{r\in G_n} \int_0^{s} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) - f_\rho \left(\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) \\
& - \sum_{i\in G_n} f'_\rho \left(\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) F_{n,i}^{k,m}(s) \Delta N_{n,i}(s) \\
& -\frac{1}{2} \sum_{i\in G_n} f''_\rho \left(\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) F_{n,i}^{k,m}(s)^2 \Delta N_{n,i}(s)\, d\left(\sum_{r\in G_n} N_{n,r} \right)(s) \\
= & \sum_{i\in G_n} \int_0^t f'_\rho \left(\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) F_{n,i}^{k,m}(s) dM_{n,i}(s) \\
& + \int_0^t f_\rho \left(\sum_{r\in G_n} \int_0^{s} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) - f_\rho \left(\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) \\
& - f'_\rho \left(\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \right) \sum_{i\in G_n} F_{n,i}^{k,m}(s) \Delta N_{n,i}(s) d\left(\sum_{r\in G_n} N_{n,r} \right)(s) \\
=: & \ (*)
\end{aligned}
$$

Note now that

$$
\begin{aligned}
& \sum_{r\in G_n} \int_0^{s} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) - \sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau) \\
= & \sum_{r\in G_n} F_{n,r}^{k,m}(s) \Delta N_{n,r}(s).
\end{aligned}
$$

Hence, (*) contains a Taylor series expansion of f_ρ around the point

$$
\sum_{r\in G_n} \int_0^{s-} F_{n,r}^{k,m}(\tau) dM_{n,r}(\tau)
$$

and we continue:

$$
\begin{aligned}
&\overset{(*)}{=} \quad \sum_{i\in G_n}\int_0^t f_\rho'\left(\sum_{r\in G_n}\int_0^{s-} F_{n,r}^{k,m}(\tau)dM_{n,r}(\tau)\right) F_{n,i}^{k,m}(s)dM_{n,i}(s)\\
&\qquad + \int_0^t \frac{1}{2} f_\rho''\left(\sum_{r\in G_n}\int_0^{s-} F_{n,r}^{k,m}(\tau)dM_{n,r}(\tau) + \Delta(s)\right)\\
&\qquad\qquad \times \left(\sum_{r\in G_n} F_{n,r}^{k,m}(s)\Delta N_{n,r}(s)\right)^2 d\left(\sum_{r\in G_n} N_{n,r}\right)(s)\\
&=: \quad (**),
\end{aligned}
$$

where $\Delta(s) \in \left[0, \sum_{r\in G_n} F_{n,r}^{k,m}(s)\Delta N_{n,r}(s)\right]$. Since only one of the counting processes jumps at a time, we obtain $|\Delta(s)| \le K_h$ with $K_h := \frac{1}{h}K\hat{K}$ and continue by using again that no two processes jump at the same time:

$$
\begin{aligned}
&\overset{(**)}{=} \quad \sum_{i\in G_n}\int_0^t f_\rho'\left(\sum_{r\in G_n}\int_0^{s-} F_{n,r}^{k,m}(\tau)dM_{n,r}(\tau)\right) F_{n,i}^{k,m}(s)dM_{n,i}(s)\\
&\qquad + \sum_{i\in G_n}\int_0^t \frac{1}{2} f_p''\left(\sum_{r\in G_n}\int_0^{s-} F_{n,r}^{k,m}(\tau)dM_{n,r}(\tau) + \Delta(s)\right) F_{n,i}^{k,m}(s)^2 dN_{n,i}(s)\\
&\le \quad \sum_{i\in G_n}\int_0^t f_\rho'\left(\sum_{r\in G_n}\int_0^{s-} F_{n,r}^{k,m}(\tau)dM_{n,r}(\tau)\right) F_{n,i}^{k,m}(s)dM_{n,i}(s)\\
&\qquad + \sum_{i\in G_n}\int_{t_0-h}^t \frac{1}{2} f_\rho''\left(\left|\sum_{r\in G_n}\int_0^{s-} F_{n,r}^{k,m}(\tau)dM_{n,r}(\tau)\right| + K_h\right)\\
&\qquad \times \frac{1}{h}K\left(\frac{s-t_0}{h}\right) K_h\hat{K} I_{n,i}^{k,m}\Gamma_n^{t_0} dN_{n,i}(s),
\end{aligned}
$$

where we used in the last line that $F_{n,i}(s) = 0$ when $t \le t_0 - h$. Now, the integrand is predictable and we can apply the expectation on both sides, to obtain a recursion formula: Use that for $x \ge 0$ we have $f_\rho''(x) = \rho(\rho-1)f_{\rho-2}(x)$ to get

$$
\begin{aligned}
&\mathbb{E}\left(\left|\sum_{i\in G_n}\int_0^t F_{n,i}^{k,m}(\tau)dM_{n,i}(\tau)\right|^\rho \middle| \mathcal{F}_{t_0-h}^n\right)\\
&\le \sum_{i\in G_n}\int_{t_0-h}^t \frac{1}{2}\rho(\rho-1)\mathbb{E}\Bigg(\left(\left|\sum_{r\in G_n}\int_0^{s-} F_{n,r}^{k,m}(\tau)dM_{n,r}(\tau)\right| + K_h\right)^{\rho-2}\\
&\qquad \times \frac{1}{h}K\left(\frac{s-t_0}{h}\right) K_h\hat{K} I_{n,i}^{k,m}\Gamma_n^{t_0}\lambda_{n,i}(s)\Bigg|\mathcal{F}_{t_0-h}^n\Bigg)ds
\end{aligned}
$$

$$\leq \int_{t_0-h}^{t} \frac{1}{2}\rho(\rho-1)K_h\hat{K}\Lambda\mathbb{E}\Bigg(\left(\left|\sum_{r\in G_n}\int_0^{s-} F_{n,r}^{k,m}(s)dM_{n,r}(s)\right| + K_h\right)^{\rho-2}$$
$$\times \sum_{i\in G_n}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)C_{n,i}(s)I_{n,i}^{k,m}\Gamma_n^{t_0}\Bigg|\mathcal{F}_{t_0-h}^n\Bigg)ds.$$

Define $Z^{k,m}(t) = \sum_{i\in G_n}\int_0^t F_{n,i}^{k,m}(\tau)dM_{n,i}(\tau)$ to summarize the previous inequality chain in the following recursion formula: For $\rho \geq 2$ it holds almost surely

$$\begin{aligned}
&\mathbb{E}\left(\left|Z^{k,m}(t)\right|^{\rho}\Big|\mathcal{F}_{t_0-h}^n\right)\\
\leq&\frac{1}{2}\int_{t_0-h}^{t}\rho(\rho-1)K_h\hat{K}\Lambda\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\\
&\times\mathbb{E}\left(\left(|Z^{k,m}(s-)|+K_h\right)^{\rho-2}\sum_{i\in G_n}I_{n,i}^{k,m}C_{n,i}(s)\Big|\mathcal{F}_{t_0-h}^n\right)\Gamma_n^{t_0}ds.
\end{aligned}\tag{5.108}$$

By uniting the (countably many) exception sets of measure zero, these inequalities hold for all $\rho \geq 2$ and all $t \in [0,T]\cap\mathbb{Q}$ on the same set of measure one. Since both sides are continuous from the right by Theorem 2.1, we also have it for all $t\in[0,T]$ on the same set of measure one. Taking now limits from the left and repeating the same argument with continuity from the left, we obtain the same result for $Z^{k,m}(t-)$ on the left hand side also on the same set of measure one.

We are going to prove now via induction that almost surely (on the same set of measure one)

$$\mathbb{E}\left(|Z^{k,m}(t)|^{\rho}\Big|\mathcal{F}_{t_0-h}^n\right) \leq \frac{\rho!}{2}S_{k,m}\Lambda K_h A^{\rho}\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{\rho-2}\Gamma_n^{t_0}.\tag{5.109}$$

We begin with the induction start: For $\rho=2$, (5.108) gives for all $t\in[t_0-h,t_0+h]$

$$\mathbb{E}\left(|Z^{k,m}(t)|^2\Big|\mathcal{F}_{t_0-h}^n\right)\leq K_h\hat{K}\Lambda S_{k,m}\cdot\Gamma_n^{t_0}\leq S_{k,m}\Lambda K_h A^2\cdot\Gamma_n^{t_0},$$

where the last inequality holds by choice of A in (5.105) and because $t\in[t_0-h,t_0+h]$. Hence, the induction start is complete and we continue with the induction step. Assume that (5.109) holds for all powers $2\leq p\leq\rho$ and all $t\in[t_0-h,t_0+h]$ and show that it holds for $\rho+1$ and all $t\in[t_0-h,t_0+h]$ as well. We use first (5.108), then the binomial theorem and finally the induction hypothesis (5.109) for powers greater than one:

$$\begin{aligned}
&\mathbb{E}\left(|Z^{k,m}(t)|^{\rho+1}\Big|\mathcal{F}_{t_0-h}^n\right)\\
\leq&\frac{1}{2}\int_{t_0-h}^{t}(\rho+1)\rho K_h\hat{K}\Lambda\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\\
&\times\mathbb{E}\left((|Z^{k,m}(s-)|+K_h)^{\rho-1}\sum_{i\in G_n}I_{n,i}^{k,m}C_{n,i}(s)\Big|\mathcal{F}_{t_0-h}^n\right)ds\cdot\Gamma_n^{t_0}
\end{aligned}$$

$$
\begin{aligned}
\leq & \frac{1}{2}\int_{t_0-h}^{t}(\rho+1)\rho K_h\hat{K}\Lambda\sum_{p=0}^{\rho-1}\binom{\rho-1}{p}\frac{1}{h}K\left(\frac{s-t_0}{h}\right) \\
& \quad \times\mathbb{E}\left(|Z^{k,m}(s-)|^{\rho-1-p}\sum_{i\in G_n}I_{n,i}^{k,m}C_{n,i}(s)\Big|\mathcal{F}_{t_0-h}^n\right)K_h^p ds\cdot\Gamma_n^{t_0} \\
\leq & \frac{K_h\hat{K}\Lambda}{2}(\rho+1)\rho\Bigg[\sum_{p=0}^{\rho-3}\frac{(\rho-1)!}{2p!}S_{k,m}\Lambda K_h\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{\rho-3-p}A^{\rho-1-p}K_h^pS_k \\
& \qquad +(\rho-1)\int_{t_0-h}^{t_0+h}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\mathbb{E}\left(|Z^{k,m}(s-)|\sum_{i\in G_n}I_{n,i}^{k,m}C_{n,i}(s)\Big|\mathcal{F}_{t_0-h}^n\right)K_h^{\rho-2}ds \\
& \qquad +K_h^{\rho-1}S_{k,m}\Bigg]\cdot\Gamma_n^{t_0}
\end{aligned}
\tag{5.110}
$$

Recall that $S_k=\max_{m=1,\dots,r_n}\sum_{i\in G_n}I_{n,i}^{k,m}\geq\sum_{i=1}I_{n,i}^{k,m}C_{n,i}(s)$ for all k and m as well as for all s, moreover S_k is measurable with respect to $\mathcal{F}_{t_0-h}^n$. Hence, we may estimate

$$
\begin{aligned}
& \int_{t_0-h}^{t_0+h}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\mathbb{E}\left(\left|Z^{k,m}(s-)\right|\sum_{i\in G_n}I_{n,i}^{k,m}C_{n,i}(s)\Big|\mathcal{F}_{t_0-h}^n\right)ds \\
\leq & \int_{t_0-h}^{t_0+h}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\sum_{i\in G_n}\mathbb{E}\left(\int_{t_0-h}^{t_0+h}\frac{1}{h}K\left(\frac{\tau-t_0}{h}\right)I_{n,i}^{k,m}\hat{K}d|M_{n,i}|(\tau)\Big|\mathcal{F}_{t_0-h}^n\right)S_kds \\
= & \int_{t_0-h}^{t_0+h}\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\sum_{i\in G_n}\int_{t_0-h}^{t_0+h}\frac{1}{h}K\left(\frac{\tau-t_0}{h}\right)2\Lambda\hat{K}\mathbb{E}\left(I_{n,i}^{k,m}C_{n,i}(\tau)\Big|\mathcal{F}_{t_0-h}^n\right)S_kds \\
= & 2\Lambda\hat{K}S_{k,m}S_k.
\end{aligned}
$$

Using this estimation we continue with the main inequality chain

$$
\begin{aligned}
& (5.110) \\
\leq & \frac{K_h\hat{K}\Lambda}{2}(\rho+1)\rho\Bigg[\sum_{p=0}^{\rho-3}\frac{(\rho-1)!}{2p!}S_{k,m}\Lambda K_h\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{\rho-3-p}A^{\rho-1-p}K_h^pS_k \\
& \qquad +(\rho-1)K_h^{\rho-2}2\Lambda\hat{K}S_{k,m}S_k \\
& \qquad +K_h^{\rho-1}S_{k,m}\Bigg]\cdot\Gamma_n^{t_0} \\
= & \frac{(\rho+1)!}{2}S_{k,m}\Lambda K_hA^{\rho+1}\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{\rho-1}\Gamma_n^{t_0} \\
& \times\frac{1}{A}\cdot\left[\sum_{p=0}^{\rho-3}\frac{1}{2p!}\cdot\Lambda K_h\hat{K}\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{-2-p}A^{-1-p}K_h^pS_k\right.
\end{aligned}
$$

$$+\frac{1}{(\rho-2)!}K_h^{\rho-2}2\Lambda\hat{K}^2 S_k A^{-\rho}\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{-\rho+1}$$

$$\left.+\frac{1}{(\rho-1)!}K_h^{\rho-1}\hat{K}A^{-\rho}\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{-\rho+1}\right]$$

At this point, we see that we're obviously done with the induction step if $\Gamma_n^{t_0} = 0$. Hence, we only need to show that the above is lesser than or equal to (5.109) on the event $\Gamma_n^{t_0} = 1$. This, in turn, we may conclude if the second part above is smaller than or equal to one (on the event $\Gamma_n^{t_0} = 1$). This is the case because we have chosen A appropriately and because $h \leq 1$ and $S_k\sqrt{h} \geq 1$ (and thus also $S_k \geq 1$) on $\Gamma_n^{t_0}$:

$$\frac{1}{A}\cdot\left[\sum_{p=0}^{\rho-3}\frac{1}{2p!}\cdot\Lambda K_h\hat{K}\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{-2-p}A^{-1-p}K_h^p S_k\right.$$

$$+\frac{1}{(\rho-2)!}K_h^{\rho-2}2\Lambda\hat{K}^2 S_k A^{-\rho}\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{-\rho+1}$$

$$\left.+\frac{1}{(\rho-1)!}K_h^{\rho-1}\hat{K}A^{-\rho}\left(S_k\sqrt{\frac{h\Lambda}{2}}K_h\right)^{-\rho+1}\right]$$

$$\frac{1}{A}\cdot\left[\sum_{p=0}^{\rho-3}\frac{1}{p!}\left(S_k\sqrt{\frac{h\Lambda}{2}}A\right)^{-p}\cdot\frac{1}{S_k K A}\right.$$

$$+\frac{1}{(\rho-2)!}\left(S_k\sqrt{\frac{h\Lambda}{2}}A\right)^{-\rho+2}\frac{2^{\frac{3}{2}}}{KA^2}\sqrt{h\Lambda}\hat{K}$$

$$\left.+\frac{1}{(\rho-1)!}\left(S_k\sqrt{\frac{h\Lambda}{2}}A\right)^{-\rho+1}\hat{K}A^{-1}\right]$$

$$\leq\frac{1}{A}\sum_{p=0}^{\infty}\frac{1}{p!}\left(S_k\sqrt{\frac{h\Lambda}{2}}A\right)^{-p}$$

$$=\frac{1}{A}\exp\left(\frac{\sqrt{2}}{AS_k\sqrt{h\Lambda}}\right)\leq\frac{1}{A}\exp\left(\frac{\sqrt{2}}{A\sqrt{\Lambda}}\right)\leq 1$$

and the induction is complete. To finalize the proof, we compute the expectation of $S_{k,m}S_k^{\rho-2}$. Note that on $\Gamma_n^{t_0} = 1$, $S_k \leq c_3\cdot\sqrt{\frac{r_n\bar{p}_n(t_0)}{\log r_n\bar{p}_n(t_0)}} = E_k^{n,t_0}$

$$\mathbb{E}\left(S_{k,m}S_k^{\rho-2}\Gamma_n^{t_0}\right)\leq\mathbb{E}(S_{k,m})\cdot\left(E_k^{n,t_0}\right)^{\rho-2}\leq E_{k,m}^{n,t_0}\cdot\left(E_k^{n,t_0}\right)^{\rho-2}.$$

Taking expectations on both sides of (5.109) and together with the previous line, we obtain

$$\mathbb{E}\left(|Z^{k,m}(T)|^\rho\right)\leq\frac{\rho!}{2}E_{k,m}^{n,t_0}\widetilde{K}K_h A^\rho\left(E_k^{n,t_0}\sqrt{\frac{h\Lambda}{2}}K_h\right)^{\rho-2}.$$

Hence, we have a mixing interaction network and we can apply Lemma 3.29 to get

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)X_{n,i}(s)dM_{n,i}(s)\geq x\cdot\sqrt{\frac{\log r_n\bar{p}_n(t_0)}{r_n\bar{p}_n(t_0)}}\right)\\
\leq\ &\mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}\int_0^T K\left(\frac{s-t_0}{h}\right)X_{n,i}(s)dM_{n,i}(s)\geq x\cdot\sqrt{\frac{\log r_n\bar{p}_n(t_0)}{r_n\bar{p}_n(t_0)}},\Gamma_n^{t_0}=1\right)\\
&\quad+\mathbb{P}(\Gamma_n^{t_0}=0)\\
\leq\ &\mathbb{P}\left(\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}Z_{n,i}(T)\geq x\cdot\sqrt{\frac{\log r_n\bar{p}_n(t_0)}{r_n\bar{p}_n(t_0)}}\right)+\mathbb{P}(\Gamma_n^{t_0}=0)\\
\leq\ &\mathcal{K}(r_n\bar{p}_n(t_0))^{-\frac{c_2\cdot x^2}{2(\sigma^2+c_1c_3x)}}+\beta_t(\Delta_n)\cdot\mathcal{K}r_n+\mathbb{P}(\Gamma_n^{t_0}=0).
\end{aligned}$$

□

Lemma 5.29. *Suppose (C1), 2 and (C4) hold. In the context of Statement 4.12, we have*

$$\sup_{\substack{t_0\in\mathbb{T}\\ \theta\in\Theta}}\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}\int_0^T\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\mathbb{E}(H_{n,i}(s,\theta)(\Gamma_n^{t_0}-1))ds=O\left(\sqrt{\frac{\log r_np_n}{r_np_n\cdot h}}\right).$$

Proof. We can estimate by boundedness of the covariates in Assumption (C1), 2 that $|H_{n,i}(s,\theta)|\leq\Lambda\hat{K}^2C_{n,i}(s)$ for all θ. Hence, we get by applying the Cauchy-Schwartz Inequality as well as Jensen's inequality

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i\in G_n}\mathbb{E}(H_{n,i}(s,\theta)(\Gamma_n^{t_0}-1))\right|\\
\leq&\Lambda\hat{K}^2\cdot\mathbb{E}\left(|\Gamma_n^{t_0}-1|\cdot\frac{1}{r_n}\sum_{i\in G_n}C_{n,i}(s)\right)\\
\leq&\Lambda\hat{K}^2\sqrt{\mathbb{P}(\Gamma_n^{t_0}=0)}\cdot\sqrt{\mathbb{E}\left(\left(\frac{1}{r_n}\sum_{i\in G_n}C_{n,i}(s)\right)^2\right)}\\
\leq&\Lambda\hat{K}^2\sqrt{\mathbb{P}(\Gamma_n^{t_0}=0)}\cdot\sqrt{p_n(s)}.
\end{aligned}$$

Applying this estimate to the expression we are interested in yields (apply again Jensen's Inequality):

$$\begin{aligned}
&\sqrt{\frac{r_np_n\cdot h}{\log r_np_n}}\left|\frac{1}{r_n\bar{p}_n(t_0)}\sum_{i\in G_n}\int_0^T\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\mathbb{E}(H_{n,i}(s,\theta)(\Gamma_n^{t_0}-1))ds\right|\\
\leq&\sqrt{\frac{r_np_n\cdot h}{\log r_np_n}}\frac{\Lambda\hat{K}^2}{\bar{p}_n(t_0)}\int_0^T\frac{1}{h}K\left(\frac{s-t_0}{h}\right)\sqrt{\mathbb{P}(\Gamma_n^{t_0}=0)p_n(s)}ds
\end{aligned}$$

$$\leq \Lambda \hat{K}^2 \sqrt{\frac{r_n h \cdot \mathbb{P}(\Gamma_n^{t_0} = 0)}{\log r_n p_n}} \cdot \sqrt{\frac{p_n}{\bar{p}_n(t_0)}} \cdot \sqrt{\int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \frac{p_n(s)}{\bar{p}_n(t_0)} ds}.$$

By construction $\frac{p_n}{\bar{p}_n(t_0)} \leq 1$ and the remaining expressions are $O(1)$ even after taking $\sup_{t_0,\theta}$ by Assumption (C4). This proves the claim. □

Lemma 5.30. *Suppose (C1), 1 and 2 and* (4.30) *and* (4.31) *in (C7) are true. In the context of Statement 4.12, there is* $\sigma^2 > 0$ *such that*

$$\mathbb{E}\left(\left|\sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) H_{n,i}(s,\theta) I_{n,i}^{k,m} \Gamma_n^{t_0} - \Sigma_{n,i}^{k,m}(s,\theta) ds\right|^2\right) \leq \sigma^2 E_{k,m}^{n,t_0}.$$

Proof. Let $\widetilde{\lambda}_{n,i}(s,\theta) = X_{n,i}(s) X_{n,i}(s)^T \exp(\theta^T X_{n,i}(s))$, then $H_{n,i}(\theta,s) = \widetilde{\lambda}_{n,i}(\theta,s) C_{n,i}(s)$. Note furthermore that $\Sigma_{n,i}^{k,m}(s,\theta) = \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,s) C_{n,i}(s) I_{n,i}^{k,m} \Gamma_n^{t_0})$. Then we obtain

$$\begin{aligned}
&\left|\sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \left[\widetilde{\lambda}_{n,i}(\theta,s) C_{n,i}(s) I_{n,i}^{k,m} \Gamma_n^{t_0} - \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,s) C_{n,i}(s) I_{n,i}^{k,m} \Gamma_n^{t_0})\right] ds\right|^2 \\
=&\Bigg(\sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \\
&\quad \times C_{n,i}(s) I_{n,i}^{k,m} \left[\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} - \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} | C_{n,i}(s) I_{n,i}^{k,m} = 1)\right] ds \\
&\quad + \sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} | C_{n,i}(s) I_{n,i}^{k,m} = 1) \\
&\quad \times \left[C_{n,i}(s) I_{n,i}^{k,m} - \mathbb{P}(C_{n,i}(s) I_{n,i}^{k,m} = 1)\right] ds\Bigg)^2 \\
\leq& 2\Bigg(\sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) C_{n,i}(s) I_{n,i}^{k,m} \\
&\quad \times \left[\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} - \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} | C_{n,i}(s) I_{n,i}^{k,m} = 1)\right] ds\Bigg)^2 \qquad (5.111) \\
&\quad + 2\Bigg(\sum_{i \in G_n} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} | C_{n,i}(s) I_{n,i}^{k,m} = 1) \\
&\quad \times \left[C_{n,i}(s) I_{n,i}^{k,m} - \mathbb{P}(C_{n,i}(s) I_{n,i}^{k,m} = 1)\right] ds\Bigg)^2 \qquad (5.112)
\end{aligned}$$

We consider now both parts above separately and show that their expectations behave as required. We start with (5.112):

$\mathbb{E}((5.112))$

$$= \sum_{i,j \in G_n} \iint_{[0,T]^2} \frac{1}{h^2} K\left(\frac{s-t_0}{h}\right) K\left(\frac{r-t_0}{h}\right)$$
$$\times \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} | C_{n,i}(s) I_{n,i}^{k,m} = 1) \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,r)\Gamma_n^{t_0} | C_{n,j}(r) I_{n,j}^{k,m} = 1)$$
$$\times \mathbb{E}\left((C_{n,i}(s) I_{n,i}^{k,m} - \mathbb{P}(C_{n,i}(s) I_{n,i}^{k,m} = 1)) \cdot (C_{n,j}(r) I_{n,j}^{k,m} - \mathbb{P}(C_{n,j}(r) I_{n,j}^{k,m} = 1)) \right) dsdr$$
$$= \left(\sum_{\substack{i,j \in G_n \\ |e_i \cap e_j| = 2}} + \sum_{\substack{i,j \in G_n \\ |e_i \cap e_j| = 1}} + \sum_{\substack{i,j \in G_n \\ |e_i \cap e_j| = 0}} \right) S_{n,ij},$$

where $S_{n,ij}$ is just the summand corresponding to the indices i, j. The number of summands in each sum is given by $r_n = O(n^2)$, $r_n \cdot 2(n-2) = O(n^3)$ and $r_n \cdot (r_n - 1) = O(n^4)$, respectively. In the situation $|e_i \cap e_j| = 2$, i.e., $i = j$, the sum comprises r_n many terms and is clearly bounded after dividing by $E_{k,m}^{n,t_0}$) because $|\widetilde{\lambda}_{n,i}(\theta,s)| \leq \hat{K}^2\Lambda$ (cf. Assumption (C1), 2) and

$$\frac{1}{E_{k,m}^{n,t_0}} \sum_{\substack{i,j \in G_n \\ |e_i \cap e_j| = 2}} S_{n,ij}$$
$$\leq \frac{r_n \cdot \hat{K}^4 \Lambda^2}{E_{k,m}^{n,t_0}} n \iint_{[0,T]^2} \frac{1}{h^2} K\left(\frac{s-t_0}{h}\right) K\left(\frac{r-t_0}{h}\right) \mathbb{E}\left(C_{n,1}(s) I_{n,1}^{k,m} C_{n,1}(r) \right) dsdr$$
$$\leq \frac{\hat{K}^4 \Lambda^2 \cdot r_n}{E_{k,m}^{n,t_0}} \int_0^T \frac{1}{h} K\left(\frac{s-t_0}{h}\right) \mathbb{E}\left(I_{n,1}^{k,m} C_{n,1}(s) \right) ds = \hat{K}^4 \Lambda.$$

In the other two situations, we write

$$\frac{S_{n,ij}}{E_{k,m}^{n,t_0}}$$
$$\leq \iint_{[0,T]^2} \frac{\hat{K}^4 \Lambda^2}{h^2} K\left(\frac{s-t_0}{h}\right) K\left(\frac{r-t_0}{h}\right)$$
$$\times \frac{\left| \mathrm{Cov}\left(C_{n,j}(r) I_{n,j}^{k,m}, C_{n,i}(s) I_{n,i}^{k,m} \right) \right|}{r_n \bar{p}_n^{k,m}(t_0)} dsdr$$

which is bounded by Assumption (C7) in (4.30). Thus, part (5.112) is behaving correctly. Part (5.111) is handled in a very similar fashion:

$$\mathbb{E}((5.111))$$
$$= \sum_{i,j \in G_n} \iint_{[0,T]^2} \frac{1}{h^2} K\left(\frac{s-t_0}{h}\right) K\left(\frac{r-t_0}{h}\right) \mathbb{P}(C_{n,i}(s) I_{n,i}^{k,m} C_{n,j}(r) I_{n,j}^{k,m} = 1)$$
$$\times \mathbb{E}\Bigg(\left[\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} - \mathbb{E}(\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} | C_{n,i}(s) I_{n,i}^{k,m} = 1) \right]$$

$$\times \left[\widetilde{\lambda}_{n,i}(\theta,s)\Gamma_n^{t_0} - \mathbb{E}(\widetilde{\lambda}_{n,j}(\theta,r)\Gamma_n^{t_0}|C_{n,j}(r)I_{n,j}^{k,m}=1)\right] \Big| C_{n,i}(s)I_{n,i}^{k,m}C_{n,j}(r)I_{n,j}^{k,m}=1\Bigg).$$

Exactly analogue to before, we split the double sum into the three regimes and argue separately. In the case $|e_i \cap e_j| = 2$, i.e., $i = j$, we argue exactly as before, namely applying boundedness of the covariates in Assumption (C1), 2. In the other two cases we apply Assumption (C7) in the following way: Let $\nu \in \{0,1\}$ and let i_0, j_0 be two edges with $|e_{i_0} \cap e_{j_0}| = \nu$, denote also by $S'_{n,ij}$ the summand corresponding to the indices i, j in the above sum. Then

$$\frac{1}{E_{k,m}^{n,t_0}} \left| \sum_{\substack{i,j \in G_n \\ |e_i \cap e_j| = \nu}} S'_{n,ij} \right|$$

$$\leq \frac{n^{4-\nu}}{r_n} \iint_{[0,T]^2} \frac{1}{h^2} K\left(\frac{s-t_0}{h}\right) K\left(\frac{r-t_0}{h}\right) \frac{\mathbb{P}(C_{n,i_0}(s)I_{n,i_0}^{k,m}C_{n,j_0}(r)I_{n,j_0}^{k,m}=1)}{\bar{p}_n^{k,m}(t_0)}$$

$$\times \mathbb{E}\Bigg(\left[\widetilde{\lambda}_{n,i_0}(\theta,s)\Gamma_n^{t_0} - \mathbb{E}(\widetilde{\lambda}_{n,i_0}(\theta,s)\Gamma_n^{t_0}|C_{n,i_0}(s)I_{n,i_0}^{k,m}=1)\right]$$

$$\times \left[\widetilde{\lambda}_{n,j_0}(\theta,s)\Gamma_n^{t_0} - \mathbb{E}(\widetilde{\lambda}_{n,j_0}(\theta,r)\Gamma_n^{t_0}|C_{n,j_0}(r)I_{n,j_0}^{k,m}=1)\right] \Big| C_{n,i_0}(s)I_{n,i_0}^{k,m}C_{n,j_0}(r)I_{n,j_0}^{k,m}=1\Bigg).$$

By Assumption (C7) in (4.31) these terms are bounded. □

6 Conclusion and Outlook

Although the specific results of Theorems 4.2 and 4.3 are in line with kernel regression results (cf. Hjort (1993) and Härdle and Mammen (1993), respectively), it has been shown in this thesis that it is possible to adapt these results to the more complex dependence situation of network indexed data. In Theorem 4.2 we saw that a change in the sparsity of the network induces a new bias which is not present in the regular kernel regression setting. However, this innovation might be based on the fact that we allow, by introducing the indicator $C_{n,i}$, that individuals become part of the estimation or drop out. It might be interesting to study such scenarios also in other kernel regression situations.

For proving Theorem 4.2 we introduced asymptotic uncorrelation as dependence measure. It is intuitively based on the assumption of joint exchangeability, i.e., relabelling the vertices does not change the distribution: The joint distribution of two edges i and j depends only on $|e_i \cap e_j|$. It is then intuitive that in the case $|e_i \cap e_j| = 0$ the processes on the two edges i and j are on average almost uncorrelated. This is because we assume that such edges i and j are, conditional on the exact structure of the network, actually far apart in most cases. This idea is then made more precise in the mixing and m-dependence notations. These are directly based on a distance measure. We formulate dependence then based on this distance. These measures are appealing because they directly reflect our intuition about the relations in a network and hence, imposing such assumptions seems very reasonable in real-world applications. This is particularly useful for the β-Mixing assumption because it quickly provides us with a useful exponential inequality (cf. Lemma 3.29).

It could be interesting to see if these dependence measures can be applied also to other estimation strategies or other models where the data is indexed by edges or vertices in a network. As one example, we mention here projection estimators. Here we would, instead of the localized log-likelihood in (4.1), consider the complete likelihood and a penalty term. Let therefore Θ_p be a set of functions, e.g., let $b_1, b_2, ...$ be a basis of the set of all functions $\theta : [0, T] \to \mathbb{R}^q$ which shall be considered as parameter functions and define $\Theta_p = \left\{ \sum_{i=1}^k \mu_i b_i : k \in \mathbb{N}, \mu_1, ..., \mu_k \in \mathbb{R} \right\}$. The estimator $\hat{\theta}_p$ would then for a penalty function $\text{pen} : \Theta_p \to [0, \infty)$ be defined as

$$\begin{aligned}\hat{\theta}_p := \operatorname*{argmin}_{\theta \in \Theta_p} - \Bigg(& \sum_{i \in G_n} \int_0^T \log \lambda(t, \theta(t), X_{n,i}(t) dN_{n,i}(t) \\ & - \int_0^T \sum_{i \in G_n} C_{n,i}(t) \lambda(t, \theta(t), X_{n,i}(t)) dt \Bigg) + \text{pen}(\theta).\end{aligned}$$

Such a model could also be extended to considering different bases and parameters for different entries of the parameter function. In particular it could be possible to estimate some parameter curves as constant and others as time dependent.

Bibliography

P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes.* Springer, 1993.

C. Brownlees, E. Nualart, and Y. Sun. Realized networks. *Journal of Applied Econometrics, to appear*, 2018.

Carter T. Butts. A relational event framework for social action. *Sociological Methodology*, 38(1):155–200, 2008.

Samuel N. Cohen and Robert J. Elliott. *Stochastic Calculus and Applications.* Birkhäuser, 2015.

Trevor F. Cox and Michael A. A. Cox. *Multidimensional Scaling.* Chapman and Hall, London, 1994.

R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25 (1):1–37, 02 1997.

Klaus Deimling. *Nonlinear Functional Analysis.* Springer, New York, 1985.

F. Diebold and K. Yilmaz. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134, 2014.

Paul Doukhan. *Mixing.* Lecture notes in statistics. Springer, New York ; Berlin ; Heidelberg [u.a.], 1994.

D. Durante and D. Dunson. Nonparametric bayes dynamic modelling of relational data. *Biometrika*, 101:883–898, 2014.

J. Elstrodt. *Maß- und Integrationstheorie.* Springer, 2011.

Ove Frank and David Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986.

Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models.* Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, New York, 2016.

A. Goldeberg, A.X. Zheng, S.E. Fienberg, and E.M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2:129–233, 2010.

S.A. Golder, D. M. Wilkinson, and B.A. Huberman. Rhythms of social interaction: Messaging within a massive online network. In C. Steinfield, B.T. Pentland, M. Ackerman, and N. Contractor, editors, *Communities and Technologies*. Springer, London, 2007.

Steve Hanneke, Wenjie Fu, and Eric P. Xing. Discrete temporal models of social networks. *Electron. J. Statist.*, 4:585–605, 2010.

Jeffrey D. Hart and Seongbaek Yi. One-sided cross-validation. *Journal of the American Statistical Association*, 93(442):620–631, 1998.

N. L. Hjort. Dynamic likelihood hazard rate estimation. 1993. URL `http://hdl.handle.net/10852/47750`.

Q. Ho, L. Song, and E.P. Xing. Evolving cluster mixed-membership blockmodel for time-evolving networks. *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 2001.

W. Härdle and E. Mammen. Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, 21(4):1926–1947, 12 1993.

B.A. Huberman, D. M. Romero, and F. Wu. Social networks that matter: Twitter under the microscope. *arxiv*, 2008. URL `https://arxiv.org/abs/0812.1045`.

David R Hunter, Steven M Goodreau, and Mark S Handcock. Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481):248–258, 2008.

Katsuhiko Ishiguro, Tomoharu Iwata, Naonori Ueda, and Joshua B. Tenenbaum. Dynamic infinite relational model for time-varying relational data analysis. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 919–927. Curran Associates, Inc., 2010.

M. O. Jackson. *Social and Economic Networks.* Princeton University Press, 2010.

Matthew O. Jackson. *Social and Economic Networks.* Princeton University Press, Princeton, 2008.

Leo Katz and Charles H. Proctor. The concept of configuration of interpersonal relations in a group as a time-dependent stochastic process. *Psychometrika*, 24(4):317–327, Dec 1959.

Eric Kolaczyk. *Statistical Analysis of Network Data.* Springer, New York, 2009.

Eric D. Kolaczyk. *Topics at the Frontier of Statistics and Network Analysis: (Re)Visiting the Foundations.* SemStat Elements. Cambridge University Press, 2017.

M. Kolar and E. P. Xing. Sparsistent estimation of time-varying markov random fields. *arxiv*, 2009. URL `https://arxiv.org/abs/0907.2337`.

A. Kreiß, E. Mammen, and W. Polonik. Nonparametric inference for continuous-time event counting and link-based dynamic network models. *ArXiv e-prints*, May 2017.

Pavel N. Krivitsky and Mark S. Handcock. A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1): 29–46, 2014.

E. Lenglart. Relation de domination entre deux processus. *Ann. Inst. Henri Poincaré*, 13:171–179, 1977.

Enno Mammen and Jens Perch Nielsen. A general approach to the predictability issue in survival analysis with applications. *Biometrika*, 94(4):873–892, 2007.

Enno Mammen, María Dolores Martínez Miranda, Jens Perch Nielsen, and Stefan Sperlich. Do-validation for kernel density estimation. *Journal of the American Statistical Association*, 106(494):651–660, 2011.

M. E. J. Newman. *Networks - An Introduction.* Oxford University Press, Oxford, 2010.

Patrick O. Perry and Patrick J. Wolfe. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(5):821–849, 2013.

P. E. Protter. *Stochastic Integration and Differential Equations.* Springer, 2005.

R. Rebolledo. Central limit theorems for local martingales. *Z. Wahrsch. verw. Geb.*, 51: 269–286, 1980.

Emmanuel Rio. *Asymptotic Theory of Weakly Dependent Random Processes.* Probability Theory and Stochastic Modelling ; 80. Springer, Berlin, Heidelberg, 2017.

Tom A. B. Snijders. The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–395, 2001.

Tom A.B. Snijders, Gerhard G. van de Bunt, and Christian E.G. Steglich. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44 – 60, 2010. Dynamics of Social Networks.

Christoph Stadtfeld and Per Block. Interactions, actors, and time: Dynamic network actor models for relational events. *Sociological Science*, 4(14):318–352, 2017. ISSN 2330-6696.

Robert Tibshirani and Trevor Hastie. Local likelihood estimation. *Journal of the American Statistical Association*, 82(398):559–567, 1987.

Alexandre Tsybakov. *Introduction to Nonparametric Estimation.* Springer Series in Statistics. Springer, 2009.

Stanley Wasserman. Analyzing social networks as stochastic processes. *Journal of the American Statistical Association*, 75(370):280–294, 1980.

Tianbao Yang, Yun Chi, Shenghuo Zhu, Yihong Gong, and Rong Jin. Detecting communities and their evolutions in dynamic social networks—a bayesian approach. *Machine Learning*, 82(2):157–189, Feb 2011.

R. Zafarani, M. A. Abbasi, and H. Liu. *Social Media Mining*. Cambridge University Press, 2014.