Econometric Theory, **37**, 2021, 1238–1266.
doi:10.1017/S0266466620000420

A MULTIPLEX INTERDEPENDENT DURATIONS MODEL*

ZHONGJIAN LIN
Emory University

RUIXUAN LIU
Emory University

We propose a multiplex interdependent durations model and study its empirical content. The model considers an empirical stopping game of multiple agents making optimal timing decisions with incomplete information. We characterize the unique Bayesian Nash equilibrium of the stopping game in a system of simultaneous equations involving the conditional distribution of each duration with a moderate strategic interaction condition. The system of nonlinear simultaneous equations allows us to obtain constructive identification results of the interaction effects and other nonparametric model primitives. We propose two consistent semiparametric estimation methods based on different parameterizations of modeling components with right-censored duration data.

1. INTRODUCTION

The study of strategic interaction effects has received considerable attention in economics. Models with continuous (Manski, 1993; Lee, 2007; Bramoullé, Djebbari, and Fortin, 2009) or discrete (Brock and Durlauf, 2001a; Aguirregabiria and Mira, 2007; Bajari et al., 2010) decision variables have been applied to many fields, including financial economics, industrial organization, labor economics, and socioeconomics. The work of de Paula (2009) and Honoré and de Paula (2010) are among the first to bring strategic interactions explicitly into the optimal timing decisions of economic agents. It is evident in de Paula (2009) and Honoré and de Paula (2010, 2018) that the strategic interaction effect captures a notable externality and explains interesting timing coordination behaviors.

Complementing Honoré and de Paula (2010), we propose a multiple durations model involving strategically interacting agents under an incomplete information structure. Each duration variable is formally defined by the threshold-crossing rule, motivated from the generic solution of an optimal stopping game (Honoré and

*This paper benefited greatly from insightful comments and suggestions provided by co-editor, Sokbae Lee, and two reviewers. We are grateful to Jaap Abbring, Jeffrey Campbell, Yingyao Hu, Qi Li, Xiaodong Liu, Essie Maasoumi, Áureo de Paula, Yuya Sasaki, Tiemen Woutersen, and Haiqing Xu for many thoughtful discussions. We also thank the audience at Tilburg University, Johns Hopkins University, University of Colorado, Boulder, and the ES 2017 European Meeting. All errors remaining are our own. Address correspondence to Zhongjian Lin, Department of Economics, Emory University, 1602 Fishburne Drive, GA 30307, USA; e-mail: zhongjian.lin@emory.edu.

© The Author(s), 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

de Paula, 2010; Abbring, 2012). The interdependence among multiple durations in our model is generated directly by the strategic feedback among agents rather than by the dependence of unobserved heterogeneity, which is in contrast with the standard multivariate mixed proportional hazards model (Lancaster, 1979). Compared with Honoré and de Paula (2010), who propose a bivariate interdependent durations model with *complete* information, we consider an *incomplete* information game and allow for more than two interacting agents.

To formalize the idea, consider K economic agents making optimal timing decisions regarding durations $T_1, \ldots, T_K$, taking into account the strategic interactions of each other. All agents (and the econometrician) observe a common set of d-dimensional covariates $X = (X_1, \ldots, X_d)$ and they make decisions about when to switch from the initial state to the alternative one. The resulting durations are characterized simultaneously by the following threshold-crossing rules:

$$
\begin{aligned}
T_1 &= \inf\left\{t : \Lambda_1(t)\phi_1(X) + \frac{\delta_1}{K-1}\sum_{k\neq 1}\mathbb{E}[1\{T_k \leq t\}|X] > \varepsilon_1\right\},\\
T_2 &= \inf\left\{t : \Lambda_2(t)\phi_2(X) + \frac{\delta_2}{K-1}\sum_{k\neq 2}\mathbb{E}[1\{T_k \leq t\}|X] > \varepsilon_2\right\},\\
&\vdots\\
T_K &= \inf\left\{t : \Lambda_K(t)\phi_K(X) + \frac{\delta_K}{K-1}\sum_{k\neq K}\mathbb{E}[1\{T_k \leq t\}|X] > \varepsilon_K\right\},
\end{aligned}
\tag{1}
$$

where model primitives consist of the kth agent's monotone trend function Λ_k, the covariate effect ϕ_k, the distribution of unobserved threshold ε_k, and the interaction parameter δ_k, for $k \in \mathcal{K} \equiv \{1, \ldots, K\}$. Our model assumes that the utility flow of switching to the alternative state is monotonically increasing with respect to t as encoded by $\Lambda_k(t)$ and the direct effect from covariates X either amplifies (or dampens) the magnitude of $\Lambda_k(t)$ through the multiplicative function $\phi_k(X)$. The unobservable threshold ε_k represents the static sunk cost associated with the transition. The covariates and unobservable terms are incorporated here by extending the generalized accelerated failure time (GAFT) model as in Ridder (1990), due to the fact that the model boils down to GAFT model in the absence of interaction effects; that is, when $\delta_k = 0$ for $k \in \mathcal{K}$. The extra term $\frac{\delta_j}{K-1}\sum_{k\neq j}\mathbb{E}[1\{T_k \leq t\}|X]$ is new in our model and captures the strategic interaction effect. We follow the incomplete information paradigm since it is (the average of) the expected stopping time of other agents, not the exact stopping time that affects the decision rules (Manski, 1991; Ahn and Manski, 1993; Bajari et al., 2010). In accordance with the aforementioned works, we assume the thresholds are mutually independent and they follow a common unknown distribution $F_\varepsilon(\cdot)$.

The construction of our model is motivated by the empirical study of network externalities and technology adoption in Gowrisankaran and Stavins (2004), who analyze the adoption of the automated clearinghouse (ACH) electronic payments

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

system in different markets for a group of banks. The complementarity of the adoption decisions arises here, due to the fact that parties involved in a payment transaction have to agree on the method of payment and their financial institutions have to coordinate technologies. In particular, if a bank decides to adopt the ACH electronic payment technology, this benefits other banks that already use the technology because they can then directly exchange payments with an additional institution (Gowrisankaran and Stavins, 2004). The main empirical analysis of Gowrisankaran and Stavins (2004) builds on a dynamic panel logit model (see Sect. 3 of Gowrisankaran and Stavins, 2004). Our framework offers a continuous-time stopping game model in which banks within a network simultaneously choose when to adopt the ACH technology based on the observable bank-specific covariates, the preferences of their customers, and the adoption decisions of other banks. Our model captures the essential feature of the empirical problem in Gowrisankaran and Stavins (2004) as it provides the simultaneity in the equilibrium adoption decisions under the strategic interactions between banks. For applied researchers and policy makers, it is important to quantify the interdependent effect and there may be a need for policy interventions depending on the magnitude of such externalities. We use the example to illustrate model primitives and assumptions we make for the multiplex interdependent durations model below in Section 2.

We obtain an explicit characterization of the stopping game in equation (1), assuming the complementarity of the stopping action; that is, $\delta_k \geq 0$ for $k \in \mathcal{K}$. Many economic phenomena are naturally modeled as games of incomplete information (Morris and Shin, 2003) with complementarity among players (Vives, 2005). In the current setting, the complementary interaction in the optimal switching decision means that a higher expected proportion of others who have already stopped would induce the agent to stop earlier rather than later. The main consequence of the complementarity assumption is that the simultaneous threshold-crossing rules in equation (1) lead to a system of simultaneous equations for the conditional distributions of T_k's. Under mild restrictions, we obtain the uniqueness of the conditional distributions given structural parameters in the model, resorting to a contraction mapping argument (Lin and Xu, 2017; Xu, 2018). Given the system equations of conditional distributions, our identification analysis embeds interesting generalizations of Elbers and Ridder (1982) and Heckman and Singer (1984a) concerning the strategic interaction effects.

The main contribution of this paper is to combine the empirical games for binary choice models (Manski, 1991; Ahn and Manski, 1993; Bajari et al., 2010) with the identification techniques developed for the transformation models in order to determine the empirical content of the interdependent durations model in equation (1). Our simultaneous equations naturally extend the single-equation transformation model by incorporating strategic interaction effects. Moreover, referring to each equation separately, the mathematical structure within the unknown link function (Horowitz and Mammen, 2007) involves both multiplicative and additive terms, which makes existing identification results inapplicable as they only handle

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

either a purely additive or multiplicative form (Ekeland, Heckman, and Nesheim, 2004; Horowitz and Mammen, 2007; Jacho-Chavez, Lewbel, and Linton, 2010; Abbring and Ridder, 2015; Chiappori, Komunjer, and Kristensen, 2015). Thus, a separate analysis is required. Considering the explicit structure of the simultaneous equations, we propose two consistent semiparametric estimation methods based on different parameterizations of modeling components, which extend the scope of Ahn and Manski (1993) to multiple durations models.

It is generally known that the incomplete information game is often more tractable than its complete information counterpart (Morris and Shin, 2003; Aradillas-Lopez, 2010; Bajari et al., 2011; Khan and Nekipelov, 2018). Within our context, there are several substantial advantages when compared with the specification and assumptions in Honoré and de Paula (2010). First of all, it is straightforward to consider the optimal timings of more than two agents, whereas it seems prohibitively complicated to extend the model in Honoré and de Paula (2010) to multiplex settings. Second, we allow for different covariate effects $\phi_k(\cdot)$ and trend functions $\Lambda_k(\cdot)$—which coincides with the cumulative hazard functions in the absence of interaction effects—for different players. The latter one is certainly a desirable property, as in the econometric duration analysis, the (cumulative) hazard function encodes the genuine state dependence and is of critical importance (Elbers and Ridder, 1982). In contrast, both the covariate effect ϕ and trend function Λ are assumed to be common for both players in Honoré and de Paula (2010). Third, we provide primitive conditions to establish uniqueness of the Bayesian Nash game, which distinguishes from Honoré and de Paula (2010) where the multiple equilibria arise. The uniqueness guarantees the identification of the conditional distribution of each duration and also facilitates the identification of model primitives. Last but not least, our identification results circumvent the exclusion restrictions from covariates in Honoré and de Paula (2010), so that the set of covariates X includes observable characteristics for all players.

Despite the continuing interest in applied works for simultaneous equations models of multiple durations, the literature is relatively sparse. Lancaster (1985) is among the first who studied the simultaneous equations of an unemployment duration derived from search theory for a single agent, but the simultaneity is caused by the accepted wage rather than another duration variable. Lillard (1993) considers a parametric model that belongs to a triangular system instead of the simultaneous equation and has demonstrated its practical utility. Abbring and van den Berg (2003) study a model where the direct impact of one duration variable on another characterizes the duration-type treatment effect. The central theme in Abbring and van den Berg (2003) is to separate out the duration treatment effect from the selection bias due to unobserved heterogeneity. This framework is extended by Drepper and Effraimidis (2015) to model interactions in multiple duration outcomes. Our paper adds not only to the literature in nonlinear simultaneous equations models for limited dependent variables originating from Heckman (1978), but especially to the recent literature on interdependent duration models initiated by de Paula (2009) and Honoré and de Paula (2010, 2018).

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

The rest of the paper is organized as follows. We present the econometric model for the multiplex interdependent durations in Section 2, where the uniqueness of the Bayesian Nash equilibrium is shown. In Section 3, we obtain the semiparametric identification of the structural components in the interdependent durations model. We also present two straightforward estimation procedures of our model with different parameterized model primitives and prove their consistency properties in Section 4. We conclude in Section 5. The proofs of main theorems are collected in Appendix A and more technical lemmas are delegated to Appendix B.

2. MULTIPLEX INTERDEPENDENT DURATIONS

Equation (1) can be viewed as a generic solution of optimal stopping time problems and it extends the discrete choice model to the continuous-time setup (see Abbring, 2012). We take the same stand as in Honoré and de Paula (2010) regarding the players' strategies, so that agents condition their strategies only based on the calendar time and commit to their plans at the beginning of the game. This constitutes the main feature of many interesting theoretical game models. For example, in the "silent timing game," timing decisions must be made well before the action begins. Its empirical application ranges from high-tech market entry decisions to the choice of movie release dates (Park and Smith, 2008). It is worth pointing out that we assume all players start the game at the same time. In practice, players may enter the game at different times. Our model can be adapted to accommodate this situation when all players enter the game before any actual switching happens and the adaption is to include a (negative) constant indicating the passed time.

Referring to each agent's stopping behavior, the decision at time t depends on one's own utility flow $\Lambda_k(t)\phi_k(X)$ and the belief of the average action taken by others conditional on a set of covariates X, $\frac{\delta_k}{K-1}\sum_{j\neq k}\mathbb{E}[1\{T_j \leq t\}|X]$. The monotone trend $\Lambda_k : R^+ \to R^+$ is a nondecreasing function, which coincides with the baseline hazard function for $k \in \mathcal{K}$ in the absence of interaction effects (hereafter, we use the terminologies interchangeably). The covariate effect $\phi_k(X)$ acts multiplicatively on the trend function and is assumed to be non-negative. A standard parameterization used in the Cox regression sets $\phi_k(X) = \exp(X^\top \beta_k)$ with the regression coefficient β_k (this is not needed for our identification analysis). Considering the strategic interaction effect, it is the expected average behavior of others that affects one's stopping decision action, which is standard in the econometric analysis (Brock and Durlauf, 2001a, 2001b) and is consistent with the empirical analysis of conformity (Bernheim, 1994) or of herd behavior (Banerjee, 1992).

We elaborate on the model primitives and their use in the empirical application of banks' adoption of ACH payment systems studied by Gowrisankaran and Stavins (2004). First of all, our multiplex durations setting fits the scenario well as there are often more than two banks making adoption decisions in different (geographically separated) markets. The observable covariates X collect bank-specific variables; for example, the measure of concentration, bank assets, assets

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

squared, and deposits. The monotone trend $\Lambda_k(t)$ encodes the notion that the ACH payment system is indeed more efficient than the use of checks over time. The documented network externalities—that a customer's value from using ACH payments— increases in the number of customers at other banks that are using ACH payments, motivates our assumption that $\delta_k \geq 0$ for all k. The latent thresholds ε_k represent the sunk-cost related to the technology adoption and are assumed to be independent. Note that the main identification argument of the externalities and the estimation of the dynamic panel logit model in Gowrisankaran and Stavins (2004) also relies on the independence of the latent error terms.

Without additional restrictions, it remains quite a challenge to pin down the empirical content of the model merely based upon equation (1). The crucial assumption that delivers an elegant characterization in our context is the complementarity of the strategic interaction; that is, $\delta_k \geq 0, k \in \mathcal{K}$. Intuitively, the complementarity implies that one's stopping shall induce others to switch earlier rather than later. Besides the aforementioned example in Gowrisankaran and Stavins (2004), numerous empirical applications naturally belong to this category (Vives, 2005). This timing complementarity among players makes the time-varying part nondecreasing with respect to t in the utility flow, so that

$$\Lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k}\mathbb{E}[1\{T_j \leq t\}|x],$$

is a nondecreasing function of t for any x. Thus, if we know T_k is less than or equal to t, then it must be the case that the time-varying part is above or equal to the threshold variable at time t; that is,

$$\{T_k \leq t\} \Leftrightarrow \left\{\varepsilon_k \leq \Lambda_k(t)\phi_k(X) + \frac{\delta_k}{K-1}\sum_{j\neq k}\mathbb{E}[1\{T_j \leq t\}|X]\right\}. \tag{2}$$

The tractability of a monotone trend also resides in more complicated models; see Liu (2020) for an example involving some latent stochastic processes.

Define $G_k(t|X) = \mathbb{P}(T_k \leq t|X)$ and write its probability density function (p.d.f.) as $g_k(t|X)$ for $k \in \mathcal{K}$. Let $f_\varepsilon(\cdot)$ denote the p.d.f. of the threshold ε. Considering the equivalence in equation (2), we have

$$\begin{aligned}
G_1(t|X) &= F_\varepsilon\left[\Lambda_1(t)\phi_1(X) + \frac{\delta_1}{K-1}\sum_{k\neq 1}G_k(t|X)\right],\\
G_2(t|X) &= F_\varepsilon\left[\Lambda_2(t)\phi_2(X) + \frac{\delta_2}{K-1}\sum_{k\neq 2}G_k(t|X)\right],\\
&\vdots\\
G_K(t|X) &= F_\varepsilon\left[\Lambda_K(t)\phi_K(X) + \frac{\delta_K}{K-1}\sum_{k\neq K}G_k(t|X)\right],
\end{aligned} \tag{3}$$

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

so that the interacting effects generate a system of simultaneous equations of the conditional distributions, $G_k(t|X), k \in \mathcal{K}$. To the best of our knowledge, this feature of our model is new in duration analyses. Interestingly, the distribution of the unobserved heterogeneity term directly becomes the unknown link function (Horowitz and Mammen, 2007), resembling the classical binary choice model.

There is a close connection of our simultaneous equations to the standard hazard-based models in duration analyses. Namely, our model boils down to the additive-multiplicative hazards model, when the thresholds are i.i.d. draws from the unit exponential distribution. In this case, we get

$$G_k(t|X) = 1 - \exp\Big[- \Lambda_k(t)\phi_k(X) - \frac{\delta_k}{K-1} \sum_{j \neq k} G_j(t|X) \Big], \quad k = 1, \ldots, K. \tag{4}$$

Let $g_k(t|X)$ be the conditional density function of $G_k(t|X)$. Then, we have the following representation of the conditional hazard density:

$$\lambda_k(t|X) \equiv \frac{g_k(t|X)}{1 - G_k(t|X)} = \lambda_k(t)\phi_k(X) + \delta_k W_k(t), \tag{5}$$

with

$$W_k(t) \equiv \frac{1}{K-1} \sum_{j \neq k} g_j(t|X), \tag{6}$$

for the kth agent, $k \in \mathcal{K}$. Given the baseline hazard rate function $\lambda_k(t)$, the direct effect of covariates is modeled via $\phi_k(x)$, which causes a proportional change of $\lambda_k(t)$. The influence from the (expected) behavior of other players is now encoded as the additional time-varying regressor $W_k(t)$, so that $\delta_k W_k(t)$ measures the excess risk (Martinussen and Scheike, 2007) on top of $\lambda_k(t)\phi_k(X)$. The hazard function specified by equation (5) belongs to the general additive-multiplicative hazard model in Lin and Ying (1995) and Martinussen and Scheike (2002), which encompasses both Aalen's additive hazards model and Cox's proportional hazards model. We refer readers to Martinussen and Scheike (2007) for a more comprehensive discussion on such models.

The simultaneous system of equations in equation (3) characterizes the Bayesian Nash equilibrium (BNE) for our stopping game. Multiplicity of the equilibria is not uncommon for an empirical game model, especially for a *complete* information game. Tamer (2003) obtains the partial identification results and interval estimation of structural parameters of an empirical game with complete information. Xiao (2018) and Aguirregabiria and Mira (2019) use different approaches to address the multiplicity in incomplete information games. We refer interested readers to Tamer (2010) for a comprehensive discussion on the partial identification methods in empirical game studies. In the literature, three types of solutions are proposed concerning the equilibrium (or equilibria) in empirical game models. First, one can explicitly allow for the multiplicity and resort to the partial identification and bounds analysis as demonstrated by Tamer (2003), Chernozhukov, Hong, and Tamer (2007), and Xiao (2018). Second, one can assume the data is drawn from a

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

single equilibrium, no matter how many equilibria there are, as exemplified by Bajari et al. (2010). Finally, one can impose enough regularity assumptions to enforce a unique equilibrium as done by Brock and Durlauf (2001a), Horst and Scheinkman (2006), Lin and Xu (2017), and Xu (2018). In this paper, we follow the last stream to impose the moderate social influence (MSI) assumption on the interaction parameters $\delta_k, k \in \mathcal{K}$, so that we obtain the unique BNE.

We make the following assumptions of the covariate effects, the distribution of unobserved threshold, and the interaction effects.

Assumption 2.1. The heterogeneity terms $(\varepsilon_1, \ldots, \varepsilon_k)$ are independent of covariates X. Furthermore, $\varepsilon_1, \ldots, \varepsilon_k$ are independently and identically distributed with the c.d.f. $F_\varepsilon(\cdot)$. $F_\varepsilon(\cdot)$ is absolutely continuous with respect to the Lebesgue measure with a p.d.f. denoted by $f_\varepsilon(\cdot)$.

Assumption 2.2. The support of the heterogeneity term's density function f_ε is $(0, +\infty)$ and f_ε is uniformly bounded from above; that is, the supnorm $\| f_\varepsilon(\cdot) \|_\infty$ is less than some positive finite constant B.

Assumption 2.3. We have absolutely continuous covariates X and the effect ϕ_k has a uniformly bounded and continuous derivative for $k \in \mathcal{K}$. Also, $\phi_k \geq 0$ for $k \in \mathcal{K}$.

Assumption 2.4. The baseline hazard function Λ_k is absolutely continuous with a non-negative derivative function λ_k and $\lim_{t\to\infty} \Lambda_k(t) = +\infty$ for $k \in \mathcal{K}$.

Assumption 2.5. (MSI) The interdependences are moderate in the sense that $0 \leq \delta_k < \frac{1}{B}, k \in \mathcal{K}$.

The following remarks are regarding the modeling assumptions.

Remark 2.1. The mutual independence assumption between thresholds is arguably strong and is consistent with Bajari et al. (2010). Heuristically speaking, the interdependence of our model is driven by the strategic interaction effect rather than the dependence among thresholds. There are notable works relaxing the independence assumption in the binary choice games, as done by Aradillas-Lopez (2010), Wan and Xu (2014), Xu (2014), and Liu, Vuong, and Xu (2017). However, the extension to our model is challenging and will be pursued in the future.

Remark 2.2. Assumption 2.2 requires the p.d.f. of ε_k to be bounded. For specific parametric error terms, the value of the upper bound B can be easily determined. For instance, if we draw the error term from the standard exponential distribution, then $B = 1$. In Assumption 2.2, we set B as a general value, which is combined with our Assumption 2.5 to guarantee the unique BNE.

Remark 2.3. Assumption 2.5 is a primary condition that restricts the size of the interaction effects, motivated by the moderate social influence (MSI) assumption in Glaeser and Scheinkman (2003) and Horst and Scheinkman (2006). Similar interpretative assumptions are also imposed in the social interactions literature with

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

logistic type models; see Brock and Durlauf (2001a, 2001b), Durlauf and Ioannides (2010), Lin and Xu (2017), Xu (2018), Hu and Lin (2020), and Jackson, Lin, and Yu (2020). Furthermore, it is worth pointing out that Assumption 2.5 is a sufficient condition for the uniqueness and not the necessary condition. Therefore, its failure in some applications does not necessarily lead to multiple equilibria.

We now present the main theoretical result in this section.

THEOREM 2.1. *Under our assumptions, the following statements hold: (i) For any* (t,x)*, there exists a unique solution* $(G_1(t|x),\ldots,G_K(t|x))$ *for the simultaneous equation system defined in equation (3). (ii) The solution* $G_k(t|x)$ *is absolutely continuous with respect to both* t *and* x *for any* $k \in \mathcal{K}$*. (iii) The solution* $G_k(t|x)$ *is a monotonically increasing function with respect to* t *for any* x *and* $k \in \mathcal{K}$*.*

Remark 2.4. The proof of Theorem 2.1 is different from the existing literature, which mainly focuses on discrete choices models (see Brock and Durlauf, 2001a, 2001b; Durlauf and Ioannides, 2010; Bajari et al., 2010). In those works, the uniqueness of the BNE boils down to whether one can obtain the conditional choice probabilities as the unique fixed point. In our setting the BNEs concern different conditional distributions as functions of t. As shown in the proof of Theorem 2.1,[1] we first argue for the existence and uniqueness of the BNE at a fixed point t for any given x. The second step is to show the conditional distributions resulting from different (t,x) are well defined c.d.f., so that they are monotone and satisfy the usual boundary restriction; that is, $\lim_{t\to 0} G_k(t|x) = 0$ and $\lim_{t\to\infty} G_k(t|x) = 1$.

Remark 2.5. Referring to Brock and Durlauf (2001a, 2001b), there may be multiple equilibria when social interactions are strong. In the logistic model of Brock and Durlauf (2001a), all the equilibria can be characterized as roots of a certain function on a proper interval of the real line and that provides a simple (numerical) device to determine the number of multiple equilibria with given observed group characteristics and any possible values of parameters in the model; see Prop. 2 in Brock and Durlauf (2001a). For incomplete information games, the recent work of Xiao (2018) also explores the finite equilibria condition in discrete choices models. However, the sharp characterization of Brock and Durlauf (2001a) relies heavily on the binary logistic structure and is not readily extended to our case as the number of solutions of our system equations in general will be difficult to determine. Throughout the rest of our paper, we maintain the MSI condition and the resulting uniqueness equilibrium in our incomplete information game.

3. IDENTIFICATION RESULTS

In this section, we address the identification issue of the structural components based on the system of equations for the conditional distributions in equation (3).

[1] In principle, one can also establish the result by some high-level functional fixed point theorem. Nonetheless, our current proof is quite elementary and has a closer connection with the existing literature for discrete choices models.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

The model primitives are the finite dimensional interaction effects $\{\delta_k\}_{k=1}^K$, as well as the nonparametric components including $(\Lambda_k(\cdot), \phi_k(\cdot))_{k\in\mathcal{K}}$ and the unknown distribution function $F_\varepsilon(\cdot)$. Given the individual duration $(T_k)_{k\in\mathcal{K}}$, the covariates $X = (X_1, \ldots, X_d)$, and the uniqueness of our BNE, we have access to the conditional distribution functions $G_k(t|x)$ for all $k \in \mathcal{K}$.

Recall that the covariate X is a random vector of dimension d. Referring to the absolute continuity of conditional distributions with respect to covariates, we denote $\nabla_l G_k(t|x) = \frac{\partial G_k(t|x)}{\partial x_l}, k \in \mathcal{K}$, which is expressible as

$$\nabla_l G_k(t|x) \equiv \frac{\partial G_k(t|x)}{\partial x_l} = f_\varepsilon\Big[\Lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k} G_j(t|x)\Big] \cdot \Big(\Lambda_k(t)\frac{\partial \phi_k(x)}{\partial x_l} + \frac{\delta_k}{K-1}\sum_{j\neq k}\nabla_l G_j(t|x)\Big), \tag{7}$$

for all $k \in \mathcal{K}$, and $l = 1, \ldots, d$.

It is clear from equation (3) that structural components cannot be identified without a proper scale or power normalization even in the absence of interacting terms because the model would then boil down to the generalized accelerated failure time model as in Ridder (1990) and Abbring and Ridder (2015), so that the nonidentifiability issue discussed therein is applicable to our setting too. Also, considering the extra complications involving the interaction term, we opt for making sufficient normalizations to get a clean point identification result, instead of trying to find all observationally equivalent pairs.

Assumption 3.1. The distribution of unobserved thresholds satisfies the following regular varying condition at the right tail:

$$\lim_{t\to\infty}\frac{f_\varepsilon(ut)}{f_\varepsilon(t)} = u^{-(1+\eta)}, \tag{8}$$

with some known $\eta > 0$.

Assumption 3.2. There exists a point $\tilde{x}$ and a set of unknown constants $c_1, \ldots, c_k$, such that $\phi_k(\tilde{x}) = c_k$. There is a limit point x^* of the covariates' support, such that $\lim_{x\to x^*}\phi_k(x) = 0$ for all k. Moreover, the following normalization condition holds at some point x_0:

$$\frac{\partial\phi_k(x_0)/\partial x_l}{\phi_k^{1+\eta}(x_0)} = 1, \tag{9}$$

for all $k \in \mathcal{K}$ and $l = 1, \ldots, d$.

Assumption 3.3. We assume that there exists $\bar{x}$, such that

$$\nabla_1 G_k\Big(G_k^{-1}(s|\bar{x})\Big|\bar{x}\Big) \neq 0, \tag{10}$$

for all $s \in (0, 1)$ and $k \in \mathcal{K}$.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Remark 3.1. Both Assumptions 3.1and 3.2 are about the normalization conditions that we impose. The regular variation in Assumption 3.1 is satisfied by many distributions for positive η; for example, stable distribution, Pareto distribution, Burr distribution, and log logistic distribution. The imposition of a known tail index $(1+\eta)$ follows from the key insight in Heckman and Singer (1984a). To be more concrete, if one parameterizes the heterogeneity term to follow a Pareto distribution, the free parameter would be the scale parameter and the tail index η is assumed to be known. A Pareto distribution of the heterogeneity term can be generated by the standard mix proportional hazards model with a gamma frailty as in Lancaster (1979) and Horowitz (1999). The normalization conditions about the functions of interest and their derivatives in Assumption 3.2 regarding the covariate effects are indispensable for nonparametric identification and are assumed in the literature in one way or another; see Ekeland, Heckman, and Nesheim (2004), Jacho-Chavez et al. (2010), and Chiappori et al. (2015). The required limit point x^* is implied by a large support condition on at least one of coordinates of X, for example, when the standard specification $\phi_k(x) = \exp(x^\top \beta_k)$ is adopted. In this case, the limit point x^* is at the boundary.

Remark 3.2. We want to emphasize that the tail regularity condition in Assumption 3.1 is consistent with the boundedness condition of f_ε in Assumption 2.2 because the former only specifies how fast (slow) the density function approaches zero at the right tail. To further elaborate on this point, we consider the following Log-Logistic distribution of ε with

$$F_\varepsilon(u) = \frac{(u/a)^\eta}{1+(u/a)^\eta}, \quad u>0, a>0, \eta>0, \tag{11}$$

where a is the unknown scale parameter and η is the known shape parameter. Its probability density function (p.d.f.) is

$$f_\varepsilon(u) = \frac{(\eta/a)(u/a)^{\eta-1}}{(1+(u/a)^\eta)^2}. \tag{12}$$

It is straightforward to verify that the p.d.f. has a unique mode at $a\left(\frac{\eta-1}{\eta+1}\right)^{1/\eta}$ when $\eta \geq 1$, which leads to a finite maximum of the p.d.f. as well. Therefore, the uniform boundedness required by Assumption 2.2 translates into the compactness of the parameter space for the positive scale parameter a.

Remark 3.3. Assumption 3.3 is a technical assumption that is used to identify the marginal quantile function of ε by solving an *inhomogeneous* differential equation, following the strategy in Elbers and Ridder (1982). In particular, Elbers and Ridder (1982) identify (the inverse of) the Laplace transform of the heterogeneity term in the mixed proportional hazards model by solving a *homogeneous* differential equation, cf. equations (11)–(15) therein. In comparison, we face an inhomogeneous differential equation where the extra term is due to the interaction effect. Intuitively speaking, Assumption 3.3 requires that the covariates affect the

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

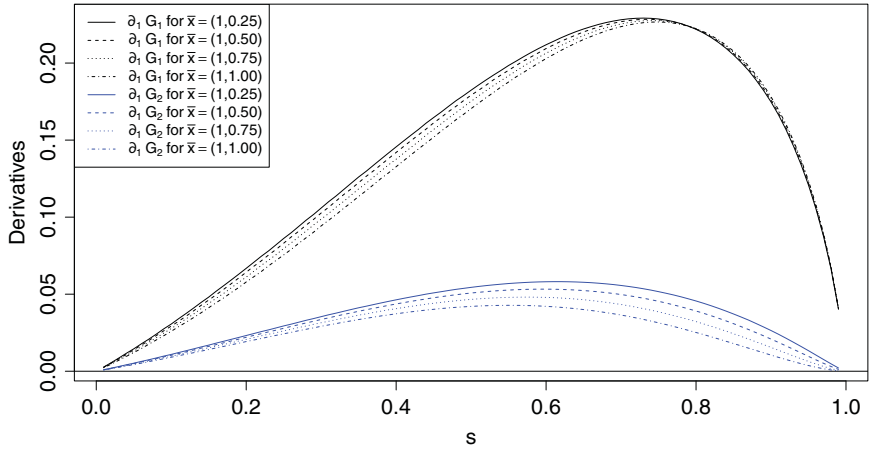


FIGURE 1. A Numerical illustration of Assumption 3.3.

conditional distributions in the sense that the partial derivative with respect to the covariate is nonzero along the line $G_k^{-1}(s|\bar{x})$ for some given point $\bar{x}$. We provide a numerical illustration of this condition with two players. In particular, we set $\Lambda(t) = t^2$ and $\phi_k(X) = exp(1 + X_k)$ for $k = 1, 2$. For each realization of $(X_1, X_2) = (x_1, x_2)$ and t, we numerically solve the conditional distributions and their partial derivatives with respect to x_1 through fixed point algorithms in equations (3) and (7). Both conditional distributions and the corresponding derivative functions are evaluated over a dense grid along $t \in \mathcal{T} \equiv (0.001, \ldots, 10)$. Figure 1 plots $\nabla_1 G_1(G_1^{-1}(s|\bar{x}))$ and $\nabla_1 G_2(G_2^{-1}(s|\bar{x}))$ for $\bar{x}_1 = 1$ and $\bar{x}_2 \in \{0.25, 0.50, 0.75, 1.00\}$ in Figure 1. Apparently, all the curves in the display are above the horizontal line.

Our identification method proceeds in two steps. First, we make use of the simultaneous equations of the conditional density functions to identify the (ratio of) strategic interaction parameters. In the second step, we work separately with each conditional distribution to identify ϕ_k, Λ_k and F_ε.

THEOREM 3.1. *Under Assumptions 2.1 to 2.5 and 3.1 to 3.3, the interaction effects $\delta_1, \ldots, \delta_K$ are identifiable up to a scale normalization.*

Remark 3.4. There is an interesting connection of the proof of Theorem 3.1 with eqn. (i) in Thm. 1 of Kortram et al. (1995), where they also identify the relative strength of the regression function (when only discrete covariates are available) via the limit of a certain ratio. In theory, one can estimate the ratio δ_k/δ_K by

$$\frac{\widehat{g}_k(0|x_n)\sum_{j\neq K}\widehat{g}_j(0|x_n)}{\widehat{g}_K(0|x_n)\sum_{j\neq k}\widehat{g}_j(0|x_n)}, \tag{13}$$

based on some kernel-type conditional density estimates $g_j(\cdot|\cdot)$ and a properly chosen sequence x_n that converges to x^*. Such an estimation strategy is used by

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Fermanian (2001) in a different context for estimating a nonparametric competing risks model, see equation (4.18) therein. The asymptotic analysis in Fermanian (2001) is quite complicated. In our Section 4, we present two straightforward semiparametric estimation methods imposing parametric assumptions on certain parts of the model primitives.

From now on, we normalize[2] $\delta_K = 1$ so that all $\delta_1, \ldots, \delta_{K-1}$ are identified in lieu of our Theorem 3.1. We focus on the identification of Λ_k, ϕ_k for $k \in \mathcal{K}$ and F_ε in the remaining part of this section.

THEOREM 3.2. *Suppose Assumptions 2.1 to 2.5 and 3.1 to 3.3 hold. Then the structural components, $\Lambda_k(t), \phi_k(x), k \in \mathcal{K}$, and $F_\varepsilon(\cdot)$ are point identified.*

Remark 3.5. In all empirical analyses involving durations, the censoring issue is a salient feature. Our identification results still hold when the censoring variable C_k is conditionally independent of T_k given X, so that one can identify the conditional distributions $G_k(t|x)$ for each $k \in \mathcal{K}$. In fact, our estimation methods are constructed explicitly accounting for the presence of censoring with right-censored duration data $(T_1 \wedge C_1, \ldots, T_K \wedge C_K, D_1, \ldots, D_K)$, where the censoring indicators stand for $D_k = 1\{T_k \leq C_k\}$ for $k \in \mathcal{K}$.

Remark 3.6. The multiplicative separability represented by $\Lambda_k(t)\phi_k(x)$ is standard in many econometric duration models (see Elbers and Ridder, 1982; Heckman and Singer, 1984a; Ridder, 1990; Abbring and Ridder, 2015) and the more general transformation models of Ekeland et al. (2004); Jacho-Chavez et al. (2010); Chiappori et al. (2015). When $\Lambda_k(t)\phi_k(x)$ is relaxed to some general nonseparable function $\Lambda_k(t,x)$, one possible path that leads to identification is to work with a known distribution F_ε for the latent heterogeneity. Under similar assumptions imposed in our Theorem 3.1, one first identifies the interaction effects $\delta_1, \ldots, \delta_K$. Thereafter, we get

$$\Lambda_k(t,x) = F_\varepsilon^{-1} \circ G_k(t|x) - \delta_k \sum_{j \neq k} G_j(t|x). \tag{14}$$

We leave a thorough investigation with an unknown distribution F_ε to future work.

Remark 3.7. The set of normalization conditions is designed to address the identification of general nonparametric functions $\phi_k(x)$ for $k \in \mathcal{K}$. On one hand, if one adopts the standard parameterization as in the Cox regression by setting $\phi_k(x) = \exp(x^\top \beta_k)$, then the identification of the coefficients β_k can be achieved under less stringent normalizations. In this case, one gets

$$(x - x_0)^\top \beta_k = \log \left[\lim_{t \to \infty} \frac{\nabla_1 G_k(t|x)}{\nabla_1 G_k(t|x_0)} \right] \Big/ \eta, \tag{15}$$

[2] In Assumption 2.5, we require the size of δ_K to be less than $1/B$, where B is the upper bound of the heterogeneity's density function. In cases where the error distribution under consideration entails some $B > 1$, one may choose a different normalization for δ_K.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

along the line of our proof of Theorem 3.2. Therefore, as long as the covariates are not in a proper linear subspace of R^d, β_k is identifiable without assuming equation (9). On the other hand, if $\log \phi_k(x)$ contains additional nonlinear terms such as higher-order polynomials or interaction terms of covariates, then equation (9) becomes indispensable.

4. SEMIPARAMETRIC ESTIMATION

The estimation of our model is based on a random sample $(V_{ik}, D_{ik}; X_i)_{i=1}^n$ of (V_k, D_k, X), where the duration T_k is subject to random censoring by C_k, so that one observes $V_k = \min(T_k, C_k)$ and the censoring indicator $D_k = 1\{T_k \leq C_k\}$. Both of our estimation procedures can be seen as natural variants of Ahn and Manski (1993) applied to multiple durations models. The first approach concerns the model specified by equation (5), in which the thresholds follow the unit exponential distribution. Given the explicit additive-multiplicative structure of the hazard rate functions, we adapt the estimation method from Lin and Ying (1995) with additional nonparametrically estimated time-varying regressors. Our second approach leaves the distribution of ε_k unspecified, however, we parameterize the trend functions $\Lambda_k(t)$ similar to the strategy in Heckman and Singer (1984b), where the cumulative baseline hazard function is parameterized. The sieve extremum estimation developed here has some interesting connections with the NPMLE in Heckman and Singer (1984a), as we also utilize the compactness of the space for cumulative distribution functions (c.d.f.) in showing the consistency property.

The following notations facilitate the presentation of our estimation. Let $N_{ik}(t)$ and $Y_{ik}(t)$ be the counting and at-risk processes (Andersen et al., 1993), respectively:

$$N_{ik}(t) = 1\{V_{ik} \leq t\} D_{ik}, \quad Y_{ik}(t) = 1\{V_{ik} > t\}, \quad k = 1, \ldots, K.$$

From now on, we adopt the standard parameterization for the covariate effect by setting $\phi_k(x) = \exp(x^\top \beta_k)$ for all k and assume that the covariates X are linearly independent. The unknown true parameter values are denoted by $(\beta_{0k}, \delta_{0k})$ for each k. Let $\mathcal{X}$ be a compact set in the interior of the support for covariates X. For the Kth agent, we simply take $\delta_K = 1$, which is consistent with our identification strategy.

Our estimation procedures require preliminary nonparametric estimates of the conditional distribution and density functions, for which we adopt the kernel-type estimators. The following definition for the generalized kernel suits our purpose for handling the (left) boundary point of the duration near time zero and the trimmed compact set $\mathcal{X}$ within the covariates' support.

Definition 4.1. Let $K_b(\cdot, \cdot)$ denote a univariate generalized kernel with properties $K_b(u, t) = 0$ if $u > t$ or $u < t - 1$. For all $t \in [0, 1]$,

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

$$b^{-(j+1)}\int_{t-1}^{t} u^j K_b(u,t)du = \begin{cases} 1, & \text{if } j=0, \\ 0, & \text{if} 1 \le j \le r-1. \end{cases}$$

We call $K_b(\cdot,\cdot)$ a univariate generalized kernel of order r.

An explicit construction of generalized kernels can be found in Mueller (1991). It is sufficient to take a second-order generalized kernel for proving the consistency property. For multivariate covariates, we denote $\mathbb{K}_b(x,\bar{x})$ to be the product kernel; that is, $\mathbb{K}_b(x,\bar{x}) = \prod_{l=1}^{d} K_b(x_l,\bar{x}_l)$. Also, denote

$$\begin{aligned}\mathbb{N}_{nk}(t|x) &= \frac{1}{nb_n^d}\sum_{i=1}^{n} 1\{V_{ik} \le t\} D_{ik}\mathbb{K}_{b_n}(x,X_i), \mathbb{Y}_{nk}(t|x) \\ &= \frac{1}{nb_n^d}\sum_{i=1}^{n} 1\{V_{ik} \ge t\}\mathbb{K}_{b_n}(x,X_i), k=1,\ldots,K.\end{aligned}$$

Then the conditional distribution function is estimated by

$$\widehat{G}_k(t|x) \equiv 1 - \exp\left(-\int_0^t \frac{d\mathbb{N}_{nk}(s|x)}{\mathbb{Y}_{nk}(s|x)}\right), \tag{16}$$

and the associated conditional density is its convolution with a generalized kernel:

$$\widehat{g}_k(t|x) \equiv \int K_{b_n}(t,s)d\widehat{G}_k(s|x). \tag{17}$$

The following assumptions are maintained to establish the uniform consistency of our first-stage estimation.

Assumption 4.1. The censoring variable C_k is independent of the duration T_k conditional on X for all $1 \le k \le K$.

Assumption 4.2. There exists a finite constant τ, such that

$$\mathbb{P}\{V_k > \tau | X = x\} \ge \xi_1 > 0,$$

for all $1 \le k \le K$ and all $x \in \mathcal{X}$. Also, the marginal density of X satisfies $\inf_{x\in\mathcal{X}} f_X(x) \ge \xi_2 > 0$.

Assumption 4.3. The bandwidth b_n of the kernel function satisfies the following restrictions: $b_n \to 0$, $nb_n^d/\log(n) \to 0$, and $nb_n^{2r+2} \to 0$.

Assumption 4.4. Referring to equation (A.17), there exists an envelope function $H_k(V_k,X)$, such that $\sup_{s\in[0,\tau],\beta,\delta} \left|Y_{jk}(s)\psi_{jl,k}(s;\beta,\delta)\right| \le H_k(V_k,X)$ and it satisfies the following integrability condition:

$$\mathbb{E}|H_k(V_k,X)| < +\infty, \quad k=1,\ldots,K. \tag{18}$$

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

LEMMA 4.1. *Under Assumptions 4.1 to 4.4, we have the following uniform consistency for the conditional c.d.f. and p.d.f. using the kernel-type estimates:*

$$\sup_{t\in[0,\tau],x\in\mathcal{X}} \left|\widehat{G}_k(t|x) - G_k(t|x)\right| = o_p(1), \tag{19}$$

and

$$\sup_{t\in[0,\tau],x\in\mathcal{X}} |\widehat{g}_k(t|x) - g_k(t|x)| = o_p(1). \tag{20}$$

4.1. Estimation with Exponential Thresholds

There exists a straightforward estimation strategy of our model with parameterized thresholds following the unit exponential distribution. Strictly speaking, the exponential threshold does not satisfy the tail condition in Assumption 3.1. Nonetheless, one can easily obtain the point identification result given the equivalence of our model to the additive-multiplicative model from Lin and Ying (1995) or Martinussen and Scheike (2002). It is convenient to introduce the following notations related to the counting processes. The counting processes $N_{1k}(t),\ldots,N_{nk}(t)$ are adapted to a filtration $\{\mathcal{F}_t : t \in [0,\tau]\}$ satisfying the usual conditions in Andersen et al. (1993). Associated with each $N_{ik}(t)$ is a covariate process $(X_i, W_{ik}(t))$ that is $\mathcal{F}_t$ predictable. Thus,

$$M_{ik}(t) \equiv N_{ik}(t) - \int_0^t Y_{ik}(s)\left[\delta_{0k}W_{ik}(s) + \lambda_{0k}(s)\exp(X_i^\top\beta_{0k})\right]ds \tag{21}$$

is an $\mathcal{F}_t$-martingale. We also denote

$$M_{ik}(t;\beta,\delta) \equiv N_{ik}(t) - \int_0^t Y_{ik}(s)\left[\delta W_{ik}(s) + \lambda_{0k}(s)\exp(X_i^\top\beta)\right]ds, \tag{22}$$

so that $M_{ik}(t) = M_{ik}(t;\beta_{0k},\delta_{0k})$.

In the construction of our estimating equations, we choose the $(d+1)$-dimensional score vector $S_{ik}(\cdot;\beta)$ to be $S_{ik}(t;\beta) = (W_{ik}(t)e^{-X_i^\top\beta}, X_i^\top)^\top \times 1\{X_i \in \mathcal{X}\}$; see Sect. 2.4 on p. 1719 of Lin and Ying (1995). Thereafter, we denote the scores based on the generated regressors by $\widehat{S}_{ik}(t;\beta) = (\widehat{W}_{ik}(t)e^{-X_i^\top\beta}, X_i^\top)^\top \times 1\{X_i \in \mathcal{X}\}$ after plugging in the kernel estimates of $\widehat{W}_{ik}(t) \equiv \frac{1}{K-1}\sum_{j\neq k}\widehat{g}_j(t|X_i)$. For the second step, the finite dimensional parameter is estimated by solving the estimating equation:

$$\widehat{U}_{n,k}(\beta,\delta) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{\widehat{S}_{ik}(s;\beta) - \frac{\sum_{j=1}^n Y_{jk}(s)\widehat{S}_{jk}(s;\beta)e^{X_j^\top\beta}}{\sum_j Y_{jk}(s)e^{X_j^\top\beta}}\right\} \left[dN_{ik}(s) - Y_{ik}(s)\delta\widehat{W}_{ik}(s)ds\right], \tag{23}$$

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

so that $\widehat{\beta}_k, \widehat{\delta}_k$ satisfies $\widehat{U}_{n,k}(\widehat{\beta}_k, \widehat{\delta}_k) = 0$. The cumulative baseline hazard function is then estimated by the Breslow-type estimator (Andersen et al., 1993):

$$\widehat{\Lambda}_k(t) = \int_0^t \frac{\sum_{j=1}^n dN_{jk}(s) - Y_{jk}(s)\widehat{\delta}_k \widehat{W}_{jk}(s)ds}{\sum_{j=1}^n Y_{jk}(s)e^{X_j^\top \widehat{\beta}_k}}. \tag{24}$$

Denote

$$\bar{S}_{nk}(s;\beta) = \frac{\sum_{j=1}^n Y_{jk}(s)S_{jk}(s;\beta)e^{X_j^\top \beta}}{\sum_{j=1}^n Y_{jk}(s)e^{X_j^\top \beta}}.$$

The rationale behind those estimating equations is that they can be approximations of the following martingale when evaluated at the true unknown parameter values:

$$U_{n,k}(\beta_{0k}, \delta_{0k}) \equiv \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{S_{ik}(s;\beta_{0k},\delta_{0k}) - \bar{S}_{nk}(s;\beta_{0k},\delta_{0k})\right\} dM_{ik}(s), \tag{25}$$

which has a zero mean. The following theorem establishes the consistency property.

THEOREM 4.1. *Suppose Assumptions 4.1 to 4.4 hold and the thresholds are independent draws for the unit exponential distribution. Also, assume that* $\sup_{t\in[0,\tau]} \lambda_{0k}(t) < \infty$ *and* $\int_0^\tau \mathbb{E}[Y_k(t)]\{1+\lambda_{0k}(t)\}dt < \infty$. *Then, for all large n, the root* $\widehat{\beta}_k, \widehat{\delta}_k$ *for the estimating equation* $U_{n,k}(\beta,\delta)$ *exists with probability tending to one and is a consistent estimator of* $(\beta_{0k}, \delta_{0k})$. *Also, the estimator* $\widehat{\Lambda}_k(t)$ *is uniformly (for all* $t \in [0,\tau]$*) consistent for the cumulative baseline hazard function* $\Lambda_{0k}(t)$.

4.2. Estimation with Arbitrary Thresholds

In this part, we allow an arbitrary distribution function of the threshold but parameterize the baseline hazard function as $\Lambda_k(t) = \Lambda_k(t;\gamma_k)$ for $k \in \mathcal{K}$. Also, we assume the parameter γ_k is identifiable given the identification of the baseline hazard function $\Lambda_k(t)$; that is, $\Lambda_k(t;\gamma_k) = \Lambda_k(t;\tilde{\gamma}_k)$ for any t implies that $\gamma_k = \tilde{\gamma}_k$. Denote the collection of unknown parameters by $\theta_k = (\beta_k, \delta_k, \gamma_k, F_\varepsilon(\cdot))$. The convergence in the space of c.d.f. $\mathcal{S}$ is defined relative to the metric:

$$\| F_1 - F_2 \| \equiv \int |F_1(u) - F_2(u)|e^{-|u|}du. \tag{26}$$

It is well known that the distance in equation (26) metrizes the weak convergence for distribution functions and the space $\mathcal{S}$ is compact under this topology. This compactness result is also explored by Cosslett (1983), Heckman and Singer (1984b), and Matzkin (1991) in establishing the consistency results for NPMLEs in different contexts.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Denote the unknown parameter space by $\Theta \equiv \mathbb{B} \times \Delta \times \Gamma \times \mathcal{S}$, with $\beta_k \in \mathbb{B}$, $\delta_k \in \Delta$, $\gamma \in \Gamma$, and $F_\varepsilon \in \mathcal{S}$. Denote the corresponding metric by

$$d(\theta, \theta_{0k}) = |\beta - \beta_{0k}| + |\delta - \delta_{0k}| + |\gamma - \gamma_{0k}| + \parallel F - F_\varepsilon \parallel .$$

The corresponding sieve space is $\Theta_{J_n} \equiv \mathbb{B} \times \Delta \times \Gamma \times \mathcal{S}_{J_n}$. Recall the estimated conditional distribution function is $\widehat{G}_k(t|x)$. We consider the sieve extremum estimator with the following criterion function:

$$Q_{n,k}(\theta) \equiv \int \left| \widehat{G}_k(t|x) - F_k\Big[\Lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k}\widehat{G}_j(t|x)\Big]\right| d\mu(t,x), \tag{27}$$

and the resulting estimates

$$\widehat{\theta}_k \equiv (\widehat{\beta}_k, \widehat{\delta}_k, \widehat{\gamma}_k, \widehat{F}_k) = \arg\min_{\theta_k \in \Theta_{J_n}} Q_{n,k}(\theta), \tag{28}$$

for some given positive and finite measure $\mu(t,x)$ that is supported on $[0,\tau] \times \mathcal{X}$.

Assumption 4.5. The parameter spaces $\mathbb{B}$, Δ, and Γ are all compact.

Assumption 4.6. The parameterization of $\Lambda_k(t;\gamma_k)$ is continuous with respect to γ_k.

Assumption 4.7. The sieve space $\mathcal{S}_J$ is nested as $\mathcal{S}_J \subset \mathcal{S}_{J+1} \subset \mathcal{S}$ for all $J \geq 1$ and there exists a sequence $\pi_J F_\varepsilon \in \mathcal{S}_J$, such that $\parallel F_\varepsilon - \pi_J F_\varepsilon \parallel \to 0$ as $J \to \infty$. Also, the sieve spaces $\mathcal{S}_J$ are compact under the metric defined in equation (26).

The desired sequence of sieve spaces $\mathcal{S}_{J_n}$ should be able to approximate any cumulative distribution function accurately. For a particular approximating sieve space, we take the monotone spline wavelet sieve on p. 5578 of Chen (2007):

$$\mathcal{S}_{J_n} = \left\{ F_{J_n}(u) = \sum_{m=-\infty}^{\infty} \alpha_m 2^{J_n/2} B_\upsilon(2^{J_n}u - m + \frac{\upsilon}{2}), \quad \alpha_m \leq \alpha_{m+1} \right\}, \tag{29}$$

where $B_\upsilon(\cdot)$ is the cardinal B-spline of order $\upsilon \geq 1$; see Sect. 2.1 of Dechevsky and Penev (1998).

THEOREM 4.2. *Suppose that Assumptions 4.1 to 4.7 hold, then* $d(\widehat{\theta}_k, \theta_{0k}) \to 0$ *for all* $k \in \mathcal{K}$ *in probability, when* $J_n \to \infty$ *as* $n \to \infty$.

Remark 4.1. The main complication to allow for nonparametric baseline cumulative hazard functions $\Lambda_{0k}(t)$ is due to the fact that $\Lambda_{0k}(t)$ are not bounded, so that the key compactness assumption on the underlying functional space is no longer satisfied. More technical proofs from Sect. 3 of Chen (2007) are required to incorporate the nonparametric baseline hazard functions and we leave this for future investigation.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

5. CONCLUSION

This paper proposes a multiplex interdependent durations model with an incomplete information structure and complementarity, complementing the study of Honoré and de Paula (2010). The structure is explored for the characterization of the Bayesian Nash equilibrium and the semiparametric identification with minimal assumptions. Our study entails the synthesis of duration analysis and empirical discrete game. The empirical content of our model is expressed via the continuous-time simultaneous equations of conditional distributions for each duration outcome. We establish the uniqueness of the Bayesian Nash Equilibrium by resorting to a moderate social influence condition. We obtain the identification of the finite dimensional parameters capturing the interaction effects and all nonparametric model primitives. The identification results extend the techniques for the single-equation transformation model to the simultaneous equations incorporating strategic interaction effects. Because the sampling information is explicitly presented by the system of conditional distributions in our model, this opens the possibility of a flexible semiparametric estimation with right-censored duration data.

We discuss two interesting questions that are worthwhile for future investigation. First, our model and other econometric duration models restrict agents to commit to the timing decision from the beginning, whereas it is certainly desirable to allow agents to re-optimize continuously in reaction to time-varying shocks in more complicated settings. However, it is challenging to incorporate such dynamic learning into a tractable empirical model. Second, the proposed estimators follow our identification strategy closely and they depend on pilot kernel smoothing estimation of the conditional distributions. Some more refined estimation procedures that can scale up to the modern high dimensional data would be appealing to applied researchers.

APPENDIX

A. Proofs of Main Results

Proof of Theorem 2.1. Part (i) For fixed (t,x), we consider the following map from $[0,1]^K$ to $[0,1]^K$:

$$\Psi(G_1(t|x),\ldots,G_K(t|x))$$

$$=\left(F_\varepsilon\Big[\Lambda_1(t)\phi_1(X)+\frac{\delta_1}{K-1}\sum_{k\neq 1}G_k(t|X)\Big],\ldots,F_\varepsilon\Big[\Lambda_K(t)\phi_K(X)+\frac{\delta_K}{K-1}\sum_{k\neq K}G_k(t|X)\Big]\right),$$

where the existence of solutions $(G_1(t|x),\ldots,G_K(t|x))$ is guaranteed by the Brouwer's fixed point theorem for the map Ψ. Now suppose there are two sets of solutions, denoted by $G^1=\left(G_1^1(t|x),\ldots,G_K^1(t|x)\right)$ and $G^2=\left(G_1^2(t|x),\ldots,G_K^2(t|x)\right)$, which are two different sets of conditional distributions evaluated at t and x, respectively. Let $G^+=(G_1^+(t|x),\ldots,G_K^+(t|x))$

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

be the intermediate values. For the kth agent and any (t,x), we have

$$|G_k^1(t|x) - G_k^2(t|x)| = \sum_{j\neq k} f_\varepsilon\Big(\Lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k} G_j^+(t|x)\Big)\cdot\frac{\delta_k}{K-1}|G_j^1(t|x) - G_j^2(t|x)|,$$
$$\leq \max_{j\in\mathcal{K}} \| f_\varepsilon \|_\infty \delta_k |G_j^1(t|x) - G_j^2(t|x)|,$$
$$< \max_{j\in\mathcal{K}} |G_j^1(t|x) - G_j^2(t|x)|. \tag{A.1}$$

Taking the maximal over the left-hand side (l.h.s.), we have

$$\max_{k\in\mathcal{K}} |G_k^1(t|x) - G_k^2(t|x)| < \max_{j\in\mathcal{K}} |G_j^1(t|x) - G_j^2(t|x)|,$$

which is a contradiction. Therefore, equation (3) delivers a unique solution.

Part (ii) We prove in Appendix B that the solution $\Big(G_1(t|x),\ldots,G_K(t|x)\Big)$ is globally Lipschitz continuous with respect to both t and x, which is basically inherited from the smoothness assumption on $\Big(\phi_1(x),\ldots,\phi_K(x)\Big)$ and $\Big(\Lambda_1(t),\ldots,\Lambda_K(t)\Big)$. Therefore, we get the desired absolute continuity that is implied by the Lipschitz condition.

Part (iii) Given the almost everywhere differentiability justified in Part (ii), we have the conditional density functions defined by

$$g_k(t|x) \equiv \frac{\partial G_k(t|x)}{\partial t} = f_\varepsilon\Big[\Lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k} G_j(t|x)\Big]$$
$$\cdot\Big(\lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k} g_j(t|x)\Big), k\in\mathcal{K}. \tag{A.2}$$

We prove Part (iii) in Lemma B.2 by contradiction. Specifically, we show that if $\min_k g(t|x) < 0$ for some (t,x), then $\min_k g(t|x)$ necessarily diverges to negative infinity, which contradicts the global Lipschitz property proved in Part (ii). The non-negative density function justifies that the solution in equation (3) is a monotone nondecreasing function. Furthermore, the boundary restriction $G_k(0|x) = 0$ for any x and k is clearly consistent with the system. Finally, $\lim_{t\to\infty} G_k(t|x) = 1$ follows from the fact that $\lim_{t\to\infty}\Lambda_k(t) = \infty$ and $\lim_{t\to\infty} F_\varepsilon(t) = 1$. In sum, the unique solution of the simultaneous equations leads to a set of well-defined conditional cumulative distribution functions. □

Proof of Theorem 3.1. We break the proof into two parts. First, we construct observationally equivalent pairs to demonstrate that interaction effects are not identified in the absence of scale normalization. We then identify the ratio of interaction effects $\delta_k/\delta_K, k\neq K$ by assuming $\delta_K \neq 0$.

Regarding the nonidentifiability issue, consider two pairs, $(\Lambda_k(t), F_\varepsilon(u), \delta_k)$ and $(\sigma\Lambda_k(t), F_\varepsilon(u/\sigma), \sigma\delta_k)$, for an arbitrary positive constant σ and any choice of covariate effects $\phi_k(x)$, $k\in\mathcal{K}$. Obviously, these two pairs lead to an observationally equivalent system of conditional distributions. Thereafter, the best possible result is to point identify the ratio $\delta_k/\delta_K, k\neq K$ by assuming $\delta_K\neq 0$.

Taking the ratio of g_k and g_K and letting $t=0$, we obtain

$$\frac{g_k(0|x)}{g_K(0|x)} = \frac{\lambda_k(0)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k} g_j(0|x)}{\lambda_K(0)\phi_K(x) + \frac{\delta_K}{K-1}\sum_{j\neq K} g_j(0|x)}. \tag{A.3}$$

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Utilizing the identification-at-limit strategy, we get

$$\lim_{x\to x^*}\phi_k(x_k)=\lim_{x\to x^*}\phi_K(x_K)=0,\ \text{for some } x^*. \tag{A.4}$$

Thus, it becomes clear that

$$\frac{\delta_k}{\delta_K}=\lim_{x\to x^*}\frac{g_k(0|x)\cdot\sum_{j\neq K}g_j(0|x)}{g_K(0|x)\cdot\sum_{j\neq k}g_j(0|x)}, \tag{A.5}$$

where the right-hand side of the equation can be identified from observables. □

Proof of Theorem 3.2. We consider the nonparametric structural components for the kth agent. Taking the normalization point x_0, one starts with the partial derivative with respect to the lth coordinate of covariates:

$$\begin{aligned}
&\frac{\nabla_l G_k(t|x)}{\nabla_l G_k(t|x_0)}\\
&=\frac{f_\varepsilon\Big(\Lambda_k(t)\phi_k(x)+\frac{\delta_k}{K-1}\sum_{j\neq k}G_j(t|x)\Big)\cdot\Big(\Lambda_k(t)\partial\phi_k(x)/\partial x_l+\frac{\delta_k}{K-1}\sum_{j\neq k}\nabla_l G_j(t|x)\Big)}{f_\varepsilon\Big(\Lambda_k(t)\phi_k(x_0)+\frac{\delta_k}{K-1}\sum_{j\neq k}G_j(t|x_0)\Big)\cdot\Big(\Lambda_k(t)\partial\phi_k(x_0)/\partial x_l+\frac{\delta_k}{K-1}\sum_{j\neq k}\nabla_l G_j(t|x_0)\Big)},\\
&=\frac{f_\varepsilon\Big(\Lambda_k(t)\phi_k(x)+\frac{\delta_k}{K-1}\sum_{j\neq k}G_j(t|x)\Big)\cdot\Big(\partial\phi_k(x)/\partial x_l+\frac{\delta_k}{(K-1)\Lambda_k(t)}\sum_{j\neq k}\nabla_l G_j(t|x)\Big)}{f_\varepsilon\Big(\Lambda_k(t)\phi_k(x_0)+\frac{\delta_k}{K-1}\sum_{j\neq k}G_j(t|x_0)\Big)\cdot\Big(\partial\phi_k(x_0)/\partial x_l+\frac{\delta_k}{(K-1)\Lambda_k(t)}\sum_{j\neq k}\nabla_l G_j(t|x_0)\Big)}.
\end{aligned} \tag{A.6}$$

Given the boundedness of $\nabla_l G_j(t|x)$, one obtains

$$\lim_{t\to\infty}\frac{\sum_{j\neq k}\nabla_l G_j(t|x)}{\Lambda_k(t)}=0, \tag{A.7}$$

utilizing $\lim_{t\to\infty}\Lambda_k(t)=\infty$. Referring to the regular variation of the heterogeneity term on the tail, one gets

$$\frac{\partial\phi_k(x)/\partial x_l}{\phi_k^{1+\eta}(x)}\Big/\frac{\partial\phi_k(x_0)/\partial x_l}{\phi_k^{1+\eta}(x_0)}=\lim_{t\to\infty}\frac{\nabla_l G_k(t|x)}{\nabla_l G_k(t|x_0)}. \tag{A.8}$$

Thus, we have identified all the partial derivatives of the function $\phi_k^{\eta}(x)$; that is, $\partial\phi_k(x)/\partial x_l\times\phi_k^{-1-\eta}(x)$ for all $k\in\mathcal{K}$. A simple integration leads to the identification of $\phi_k(x)$ utilizing the normalization at point $\tilde{x}$, referring to Assumption 3.2.

Once we have identified δ_k and ϕ_k, we proceed by identifying the quantile function $Q_\varepsilon(\cdot)$ of the threshold ε following the strategy in Elbers and Ridder (1982). The extra complication for our case is that, due to the interaction terms, one faces an inhomogeneous linear differential equation instead of the homogeneous one seen in Elbers and Ridder (1982). It follows that

$$\Lambda_k(t)=\frac{Q_\varepsilon[G_k(t|x)]-\frac{\delta_k}{K-1}\sum_{j\neq k}G_j(t|x)}{\phi_k(x)},\quad\forall x. \tag{A.9}$$

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Differentiating with respect to the first coordinate x_1, we obtain

$$\Big[Q'_\varepsilon[G_k(t|x)]\nabla_1 G_k(t|x) - \frac{\delta_k}{K-1}\sum_{j\neq k}\nabla_1 G_j(t|x)\Big]\phi_k(x) - \Big[Q_\varepsilon[G_k(t|x)] - \frac{\delta_k}{K-1}\sum_{j\neq k} G_j(t|x)\Big]\partial\phi_k(x)/\partial x_1 = 0. \tag{A.10}$$

For some fixed $\bar{x}$, we make a change of variable by letting $G_k(t|\bar{x}) = s$. This yields an inhomogeneous linear differential equation for $Q_\varepsilon(s)$ as follows:

$$Q'_\varepsilon(s) + \tilde{P}(s)Q_\varepsilon(s) = \tilde{L}(s), \tag{A.11}$$

where

$$\tilde{P}(s) = -\frac{\partial\phi_k(\bar{x})/\partial x_1}{\nabla_1 G_k[G_k^{-1}(s|\bar{x})|\bar{x}]\phi_k(\bar{x})}, \tag{A.12}$$

$$\tilde{L}(s) = \frac{\delta_k\phi_k(\bar{x})\sum_{j\neq k}\nabla_1 G_j[G_k^{-1}(s|\bar{x})|\bar{x}] - \delta_k\partial\phi_k(\bar{x})/\partial x_1\sum_{j\neq k} G_j[G_k^{-1}(s|\bar{x})|\bar{x}]}{(K-1)\nabla_1 G_k[G_k^{-1}(s|\bar{x})|\bar{x}]\phi_k(\bar{x})}. \tag{A.13}$$

Referring to the solution for eqn. (A.11) in Lemma B.3, one obtains:

$$Q_\varepsilon(s) = e^{-\int_0^s \tilde{P}(t)dt}\Big[\int_0^s e^{-\int_0^t \tilde{P}(u)du}\tilde{L}(t)dt\Big]. \tag{A.14}$$

Note that the undetermined constant in the general solution is pinned down with $C = 0$, because $Q_\varepsilon(0) = 0$. Therefore, $F_\varepsilon(\cdot)$ is point identified by inverting its quantile function $Q_\varepsilon(\cdot)$. Finally, $\Lambda_k(t)$ is identified from equation (A.9). □

Proof of Theorem 4.1. The crux of our proof is to show that the generated regressor $\widehat{W}_{ik}(t)$ can be replaced by $W_{ik}(t)$ in order to account for the estimation of the conditional p.d.f. in the time-varying regressor. One can then make use of the results in Lin and Ying (1995) or Martinussen and Scheike (2002) to establish the consistency property. First, we show that

$$\frac{1}{n}\sum_{i=1}^n\int_0^t \widehat{S}_{ik}(s;\beta)\big[Y_{ik}(s)\delta\widehat{W}_{ik}(s)ds\big] = \frac{1}{n}\sum_{i=1}^n\int_0^t S_{ik}(s;\beta)\left[Y_{ik}(s)\delta W_{ik}(s)ds\right] + o_p(1), \tag{A.15}$$

and

$$\frac{1}{n}\sum_{j=1}^n Y_{jk}(s)\widehat{S}_{jk}(s;\beta,\delta)e^{X_j^\top\beta} = \frac{1}{n}\sum_{j=1}^n Y_{jk}(s)S_{jk}(s;\beta,\delta)e^{X_j^\top\beta} + o_p(1), \tag{A.16}$$

uniformly over (t,β,δ). We only prove equation (A.16) due to the similarity. Referring to the specific form of $\widehat{S}_{jk}$, we only need to verify equation (A.16) for the first coordinate of $\widehat{S}_{jk}$ as that is the part involving the estimated conditional p.d.f. Hence, the first coordinate of the l.h.s. of equation (A.16) becomes

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

$$\frac{1}{n}\sum_{j=1}^{n} Y_{jk}(s)\delta \widehat{W}_{jk}(s)e^{-X_j^\top \beta}e^{X_j^\top \beta} = \frac{1}{n}\sum_{j=1}^{n} Y_{jk}(s)\delta W_{jk}(s) + \frac{1}{n^2}\sum_{1\le j,l\le n} Y_{jk}(s)\psi_{jl,k}(s;\beta,\delta) + o_p(1),$$

where

$$\psi_{jl,k}(t;\beta,\delta) = \frac{\delta}{K-1}\sum_{m\neq k}\int K_{b_n}(t,s)\xi_{l,m}(ds,X_j), \tag{A.17}$$

by the uniform linear representation in Lemma B.4. Then by the uniform law of large numbers (ULLN) for the V-process in Thm. 5.2.9 of de la Pena and Gine (1999), we get

$$\frac{1}{n^2}\sum_{j,l} Y_{jk}(s)\psi_{jl,k}(s;\beta,\delta) = o_p(1).$$

More specifically, one verifies that the kernel (of the U-statistic) related to the function $Y_{jk}(s)\psi_{jl,k}(s;\beta,\delta)$ is of VC-subgraph (de la Pena and Gine, 1999). Following a similar argument as in Appendix C.3 of Fan and Liu (2018), we get the ULLN for the corresponding *U-process*. Also, the moment condition for the envelope function required by Thm. 5.2.9 of de la Pena and Gine (1999) is satisfied given our assumption. Thereafter, the desired ULLN for the corresponding *V-process* follows from combining Thms. 5.2.2 and 5.2.9 of de la Pena and Gine (1999). Based on equations (A.15) and (A.16), we arrive at

$$\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left\{\frac{\sum_{j=1}^n Y_{jk}(s)\widehat{S}_{jk}(s;\beta,\delta)e^{X_j^\top\beta}}{\sum_j Y_{jk}(s)e^{X_j^\top\beta}}\right\}\left[Y_{ik}(s)\delta\widehat{W}_{ik}(s)ds\right] = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \bar{S}_{nk}(s;\beta,\delta)\left[Y_{ik}(s)\delta W_{ik}(s)ds\right] + o_p(1). \tag{A.18}$$

The combination of equations (A.15) and (A.18) gives

$$\sup_{\beta,\delta}\left|\widehat{U}_{n,k}(\beta,\delta) - U_{n,k}(\beta,\delta)\right| = o_p(1), \tag{A.19}$$

where

$$U_{n,k}(\beta,\delta) = \frac{1}{n}\sum_{i=1}^{n}\left[\int_0^\tau \left\{S_{ik}(s;\beta,\delta) - \bar{S}_{nk}(s;\beta,\delta)\right\}\left[dN_{ik}(s) - Y_{ik}(s)\delta W_{ik}(s)ds\right]\right].$$

Then, the estimating equation $U_{n,k}(\beta,\delta)$ at the population level has a unique root at the true parameter value (β_{0k},δ_{0k}) and the desired consistency result follows from the conclusion in Cor. 2.3 of Lin and Ying (1995). In the same vein, we have

$$\widehat{\Lambda}_k(t) = \int_0^t \frac{\sum_{j=1}^n dN_{jk}(s) - Y_{jk}(s)\delta_{0k}W_{jk}(s)ds}{\sum_{j=1}^n Y_{jk}(s)e^{X_j^\top\beta_{0k}}} + o_p(1),$$

for the baseline cumulative hazard function. Hence, its consistency follows from Thm. 3.1 of Lin and Ying (1995). □

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Proof of Theorem 4.2. It is straightforward to see that the limit of the sample criterion function $Q_{n,k}(\theta)$ is

$$Q_k(\theta) \equiv \int \left| G_k(t|x) - F_\varepsilon\left[\Lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k} G_j(t|x)\right]\right| d\mu(t,x). \tag{A.20}$$

The proof essentially verifies Cond. 3.1″, 3.2, 3.4, and 3.5(i) in Chen (2007), which give rise to the consistency result via Thm. 3.1 of Chen (2007). First, the compactness required in Conditions 3.1″ and 3.4 therein are satisfied in our context, because the parameter spaces for the finite dimensional parameters are compact and the space for the c.d.f. is compact under the norm $\|\cdot\|$ in equation (26).

Regarding the uniform convergence of the sample criterion function over the sieve space required by Condition 3.5(i), we get

$$Q_{n,k}(\theta) = \int \left| G_k(t|x) - F_{J_n}\left[\Lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k} \widehat{G}_j(t|x)\right]\right| d\mu(t,x) + o_p(1), \tag{A.21}$$

$$= \int \left| G_k(t|x) - F_{J_n}\left[\Lambda_k(t)\phi_k(x) + \frac{\delta_k}{K-1}\sum_{j\neq k} G_j(t|x)\right]\right| d\mu(t,x) + o_p(1), \tag{A.22}$$

based on the uniform consistency of the conditional c.d.f. and the continuity of the c.d.f. F_{J_n} in the sieve space; see Dechevsky and Penev (1998). In other words, we have

$$\sup_{\theta\in\Theta_{J_n}} \left|Q_{n,k}(\theta) - Q_k(\theta)\right| = o_p(1), \tag{A.23}$$

for any $J_n \geq 1$. The approximation error due to the use of the sieve is determined by the space $\mathcal{S}_{J_n}$. The order of the sieve approximation error is bounded up by $O(2^{-2J_n})$ as in Dechevsky and Penev (1998), which satisfies Cond. 3.2 of Chen (2007). □

B. Auxiliary Lemmas and Proofs

LEMMA B.1. *Under our assumptions, the solution of the simultaneous equation $G_k(t|x)$ is absolutely continuous with respect to both t and x, for any $k \in \mathcal{K}$.*

Proof. We first consider the absolute continuity with respect to x. For any x_1, x_2 in the support of covariates, we have

$$\begin{aligned}
&|G_k(t|x_1) - G_k(t|x_2)| \\
&\leq \|f_\varepsilon\|_\infty \left[\Lambda_k(t)|\phi_k(x_1) - \phi_k(x_2)| + \frac{\delta_k}{K-1}\sum_{j\neq k}\left|G_j(t|x_1) - G_j(t|x_2)\right|\right] \\
&\leq \|f_\varepsilon\|_\infty \max_{1\leq j\leq K}\Lambda_j(t)|\phi_j(x_1) - \phi_j(x_2)| + \|f_\varepsilon\|_\infty \delta_k \max_{1\leq j\leq K}\left|G_j(t|x_1) - G_j(t|x_2)\right|.
\end{aligned}$$

Utilizing the assumption that $\|f_\varepsilon\|_\infty \delta_k < 1$ for any k, we obtain

$$(1 - \delta_k\|f_\varepsilon\|_\infty)\max_{1\leq j\leq K}\left|G_j(t|x_1) - G_j(t|x_2)\right| \leq \|f_\varepsilon\|_\infty \max_{1\leq j\leq K}\Lambda_j(t)|\phi_j(x_1) - \phi_j(x_2)|.$$

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Thus, by the assumed Lipschitz continuity of $\phi_j(\cdot)$ functions for $j = 1, \ldots, K$, we get the Lipschitz continuity of our conditional c.d.f. with respect to x. Therefore, one gets the desired absolute continuity, such that $\frac{\partial}{\partial x} G_k(t|x)$ is defined almost everywhere for all t and $k = 1, \ldots, K$.

A complete analogous argument gives the Lipschitz continuity of the conditional c.d.f. with respect to t, which in turn guarantees the legitimate conditional p.d.f. $g_k(t|x)$ for $k = 1, \ldots, K$. □

LEMMA B.2. *Under our assumptions, the solution of simultaneous equations $G_k(t|x)$ is a monotonically increasing function with respect to t, for any x and $k \in \mathcal{K}$.*

Proof. We show that the conditional p.d.f. are all non-negative; that is, $\min_{1\le k\le K} g_k(t|x) \ge 0$ for all t and x. We prove this by contradiction. Suppose there exists some (t_0, x_0) and $k_0 \in \{1, \ldots, K\}$, such that $g_{k_0}(t_0|x_0) < 0$. Without loss of generality, we take $k_0 = \arg\ \min_{1\le k\le K} g_k(t_0|x_0)$. Given that $g_{k_0}(t_0|x_0) < 0$, a quick inspection of the formula in equation (A.2) leads to

$$\lambda_{k_0}(t_0)\phi_{k_0}(x_0) + \frac{\delta_{k_0}}{K-1}\sum_{j\neq k_0} g_j(t_0|x_0) < 0,$$

which implies that

$$g_{k_0}(t_0|x_0) = \min_{1\le j\le K} g_j(t_0|x_0) \le \min_{j\neq k_0} g_j(t_0|x_0) \le \frac{1}{K-1}\sum_{j\neq k_0} g_j(t_0|x_0) < -\frac{1}{\delta_{k_0}}\lambda_{k_0}(t_0)\phi_{k_0}(x_0).$$

We refer to the formula in equation (A.2),

$$\begin{aligned} g_{k_0}(t_0|x_0) = & f_\varepsilon\Big[\Lambda_{k_0}(t_0)\phi_{k_0}(x_0) + \frac{\delta_{k_0}}{K-1}\sum_{j\neq k} G_j(t_0|x_0)\Big] \\ & \cdot\Big(\lambda_{k_0}(t_0)\phi_{k_0}(x_0) + \frac{\delta_{k_0}}{K-1}\sum_{j\neq k_0} g_j(t_0|x_0)\Big), \end{aligned}$$

which gives rise to

$$\frac{1}{K-1}\sum_{j\neq k_0} g_j(t_0|x_0) < -\left(\frac{1}{\delta_{k_0}\,\|f_\varepsilon\,\|_\infty} + \frac{1}{\delta_{k_0}}\right)\lambda_{k_0}(t_0)\phi_{k_0}(x_0),$$

and

$$g_{k_0}(t_0|x_0) = \min_{1\le j\le K} g_j(t_0|x_0) < -\left(\frac{1}{\delta_{k_0}\,\|f_\varepsilon\,\|_\infty} + \frac{1}{\delta_{k_0}}\right)\lambda_{k_0}(t_0)\phi_{k_0}(x_0). \tag{B.1}$$

At the Nth iteration, one has the following bound for any natural number N:

$$\min_{1\le j\le K} g_j(t_0|x_0) < -\sum_{n=1}^{N}\frac{1}{(\delta_{k_0}\,\|f_\varepsilon\,\|_\infty)^n}\lambda_{k_0}(t_0)\phi_{k_0}(x_0) - \frac{1}{\delta_{k_0}}\lambda_{k_0}(t_0)\phi_{k_0}(x_0). \tag{B.2}$$

Apparently, $\min_{1\le k\le K} g_k(t_0|x_0) \to -\infty$ by letting $N \to +\infty$, because $\frac{1}{\delta_k\|f_\varepsilon\|_\infty} > 1$ for all k. This contradicts the global Lipschitz continuity of the conditional c.d.f. with respect to t for any x. □

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

LEMMA B.3. *For functions $y(s)$, $a(s)$, and $b(s)$, suppose the following inhomogenous linear differential equation holds:*

$$y'(s) + a(s)y(s) = b(s), \tag{B.3}$$

where $y'(s)$ stands for the first-order derivative. The general solution is

$$y(s) = e^{-\int_0^s a(u)du}\left[\int_0^s e^{-\int_0^u a(v)dv} b(u)du + C\right], \tag{B.4}$$

for some constant C; see Sect. 2.6 in Coddington (2012).

We restate some necessary definitions and terminologies in Chap. 5 of de la Pena and Gine (1999). Let $\mathcal{F}$ be the class of functions and $L_2(Q)$ be the L_2-norm defined by a probability measure Q. For any probability measure Q, let $N(\varepsilon, \mathcal{F}, L_2(Q))$ be the minimal number of balls of radius ε needed to cover the class $\mathcal{F}$. An envelope function of a functional class $\mathcal{F}$ is a function F, such that $|f(x)| \leq F(x)$ for all x and $f \in \mathcal{F}$. The covering number of a VC-subgraph functional class $\mathcal{F}$ with index v admits the following upper bound:

$$N(\delta, \mathcal{F}, L_2(Q)) \lesssim \delta^{-2v}. \tag{B.5}$$

The following lemma states the linear representation of the kernel estimators for the conditional probability density functions. Its proof is essentially the same as Thm. 4.1 from Fan and Liu (2018) under the independence censoring scheme (so that C_k is independent of T_k conditional on covariates X for each k), while eliminating the bias terms.

LEMMA B.4. *Under Assumptions 4.1 to 4.4, we have the following uniform linear representation for the conditional p.d.f.:*

$$\widehat{g}_k(t|x) - g_k(t|x) = \frac{1}{n}\sum_{i=1}^{n}\int K_{b_n}(t,s)\xi_{i,k}(ds,x) + o_p(1). \tag{B.6}$$

for any $t \in [0, \tau]$ and $x \in \mathcal{X}$, where

$$\begin{aligned}\xi_{i,k}(s,x) = \frac{1}{f_X(x)S_{T_k}(s|x)}\mathbb{K}_{b_n}(X_i,x)\Bigg[&\int_0^s S_{V_k}^{-2}(v|x)\left[I\left(V_{i,k}\leq v\right) - F_{V_k}(v|X_i)\right] \\ &\times dF_{V_k,D=1}(v|x) + S_{V_k}^{-1}(s|x)\left[I\left(V_{i,k}\leq s, D_i = 1\right) - F_{V_k,D=1}(s|X_i)\right] \\ &- \int_0^s S_{V_k}^{-2}(v|x)\left[I\left(V_{i,k}\leq v, D_i = 1\right) - F_{V_k,D=1}(v|X_i)\right]dF_{V_k}(v|x)\Bigg],\end{aligned}$$

which has a zero mean for any s and x.

REFERENCES

Abbring, J.H. (2012) Mixed hitting-time models. *Econometrica* 80(2), 783–819.

Abbring, J.H. & G. Ridder (2015) Regular variation and the identification of generalized accelerated failure-time models. *Econometric Theory* 31(6), 1229–1248.

Abbring, J.H. & G.J. van den Berg (2003) The nonparametric identification of treatment effects in duration models. *Econometrica* 71(5), 1491–1517.

Aguirregabiria, V. & P. Mira (2007) Sequential estimation of dynamic discrete games. *Econometrica* 75(1), 1–53.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Aguirregabiria, V. & P. Mira (2019) Identification of games of incomplete information with multiple equilibria and common unobserved heterogeneity. *Quantitative Economics* 10(4), 1659–1701.
Ahn, H. & C.F. Manski (1993) Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations. *Journal of Econometrics* 56(3), 291–321.
Andersen, P., O. Borgan, R. Gill, & N. Keiding (1993) *Statistical Models Based on Counting Processes*. Springer.
Aradillas-Lopez, A. (2010) Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics* 157(2), 409–431.
Bajari, P., J. Hahn, H. Hong, & G. Ridder (2011) A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models. *International Economic Review* 52(3), 807–824.
Bajari, P., H. Hong, J. Krainer, & D. Nekipelov (2010) Estimating static models of strategic interactions. *Journal of Business & Economic Statistics* 28(4), 469–482.
Banerjee, A.V. (1992) A simple model of herd behavior. *The Quarterly Journal of Economics* 107(3), 797–817.
Bernheim, D.B. (1994) A theory of conformity. *Journal of Political Economy* 102(5), 841–877.
Bramoullé, Y., H. Djebbari, & B. Fortin (2009) Identification of peer effects through social networks. *Journal of Econometrics* 150(1), 41–55.
Brock, W.A. & S.N. Durlauf (2001a) Discrete choice with social interactions. *The Review of Economic Studies* 68(2), 235–260.
Brock, W.A. & S.N. Durlauf (2001b) Interactions-based models. *Handbook of Econometrics* 5, 3297–3380.
Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics* 6, 5549–5632.
Chernozhukov, V., H. Hong, & E. Tamer (2007) Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75(5), 1243–1284.
Chiappori, P.A., I. Komunjer, & D. Kristensen (2015) Nonparametric identification and estimation of transformation models. *Journal of Econometrics* 188(1), 22–39.
Coddington, E.A. (2012) *An Introduction to Ordinary Differential Equations*. Courier Corporation.
Cosslett, S.R. (1983) Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* 51, 765–782.
de la Pena, V. & E. Gine (1999) *Decoupling: From Dependence to Independence*. Springer.
de Paula, Á. (2009) Inference in a synchronization game with social interactions. *Journal of Econometrics* 148(1), 56–71.
Dechevsky, L. & S. Penev (1998) On shape-preserving wavelet estimators of cumulative distribution functions and densities. *Stochastic Analysis and Applications* 16(3), 423–462.
Drepper, B. & G. Effraimidis (2015) Social Interaction Effects in Duration Models. Working Paper.
Durlauf, S.N. & Y.M. Ioannides (2010) Social interactions. *Annual Review of Economics* 2(1), 451–478.
Ekeland, I., J. Heckman, & L. Nesheim (2004) Identification and estimation of hedonic models. *Journal of Political Economy* 112(1), 60–109.
Elbers, C. & G. Ridder (1982) True and spurious duration dependence: The identifiability of the proportional hazard model. *The Review of Economic Studies* 49(3), 403–409.
Fan, Y. & R. Liu (2018) Partial identification and inference in censored quantile regression. *Journal of Econometrics* 206, 1–38.
Fermanian, J.D. (2001) Nonparametric estimation of competing risks models with covariates. *Journal of Multivariate Analysis* 85, 156–191.
Glaeser, E. & J.A. Scheinkman (2003) Nonmarket interactions. *Advances in Economics and Econometrics* 1, 339–370.
Gowrisankaran, G. & J. Stavins (2004) Network externalities and technology adoption: Lessons from electronic payments. *RAND Journal of Economics* 35(2), 260–276.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Heckman, J.J. (1978) Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4), 931–959.
Heckman, J.J. & B. Singer (1984a) The identifiability of the proportional hazard model. *The Review of Economic Studies* 51(2), 231–241.
Heckman, J.J. & B. Singer (1984b) A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52, 271–320.
Honoré, B.E. & Á. de Paula (2010) Interdependent durations. *The Review of Economic Studies* 77(3), 1138–1163.
Honoré, B.E. & Á. de Paula (2018) A new model for interdependent durations. *Quantitative Economics* 9(3), 1299–1333.
Horowitz, J. (1999) Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica* 67(5), 1001–1028.
Horowitz, J. & E. Mammen (2007) Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Annals of Statistics* 35(6), 2589–2619.
Horst, U. & J.A. Scheinkman (2006) Equilibria in systems of social interactions. *Journal of Economic Theory* 130, 44–77.
Hu, Y. & Z. Lin (2020) Binary Choice with Misclassification and Social Interactions, with an Application to Peer Effects in Attitude. Working paper.
Jacho-Chavez, D., A. Lewbel, & O. Linton (2010) Identification and nonparametric estimation of a transformed additively separable model. *Journal of Econometrics* 156(2), 392–407.
Jackson, M., Z. Lin, & N. Yu (2020) Adjusting for Peer-Influence in Propensity Scoring When Estimating Treatment Effects. Working paper.
Khan, S. & D. Nekipelov (2018) Information structure and statistical information in discrete response models. *Quantitative Economics* 9(2), 995–1017.
Kortram, R.A., A.C.M. van Rooij, A.J. Lenstra, & G. Ridder (1995) Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica* 49(3), 269–281.
Lancaster, T. (1979) Econometric methods for the duration of unemployment. *Econometrica* 47(4), 939–956.
Lancaster, T. (1985) Simultaneous equations models in applied search theory. *Journal of Econometrics* 28(1), 113–126.
Lee, L.F. (2007) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140(2), 333–374.
Lillard, L.A. (1993) Simultaneous equations for hazards: Marriage duration and fertility timing. *Journal of Econometrics* 56(1-2), 189–217.
Lin, D. & Z. Ying (1995) Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Annals of Statistics* 23(5), 1712–1734.
Lin, Z. & H. Xu (2017) Estimation of social-influence-dependent peer pressures in a large network game. *The Econometrics Journal* 20(3), 86–102.
Liu, N., Q. Vuong, & H. Xu (2017) Rationalization and identification of binary games with correlated types. *Journal of Econometrics* 201(2), 249–268.
Liu, R. (2020) A competing risks model with time-varying heterogeneity and simultaneous failure. *Quantitative Economics* 11(2), 535–577.
Manski, C.F. (1991) Nonparametric estimation of expectations in the analysis of discrete choice under uncertainty. *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, vol. 5, pp. 259–275. Cambridge University Press.
Manski, C.F. (1993) Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60(3), 531–542.
Martinussen, T. & T.H. Scheike (2002) A flexible additive multiplicative hazard model. *Biometrika* 89(2), 283–298.
Martinussen, T. & T.H. Scheike (2007) *Dynamic Regression Models for Survival Data*. Springer Science & Business Media.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press

Matzkin, R.L. (1991) Semiparametric estimation of monotone and concave utility functions for polychotomous choice models. *Econometrica* 59, 1315–1327.
Morris, S. & H.S. Shin (2003) Global games: Theory and applications. *Advances in Economics and Econometrics* 56–114.
Mueller, H. (1991) Smooth optimum kernel estimators near endpoints. *Biometrika* 78, 521–530.
Park, A. & L. Smith (2008) Caller number five and related timing games. *Theoretical Economics* 3, 231–256.
Ridder, G. (1990) The non-parametric identification of generalized accelerated failure-time models. *The Review of Economic Studies* 57(2), 167–181.
Tamer, E. (2003) Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies* 70(1), 147–165.
Tamer, E. (2010) Partial identification in econometrics. *Annual Review of Economics* 2(1), 167–195.
Vives, X. (2005) Complementarities and games: New developments. *Journal of Economic Literature* 43(2), 437–479.
Wan, Y. & H. Xu (2014) Semiparametric identification of binary decision games of incomplete information with correlated private signals. *Journal of Econometrics* 182(2), 235–246.
Xiao, R. (2018) Identification and estimation of incomplete information games with multiple equilibria. *Journal of Econometrics* 203(2), 328–343.
Xu, H. (2014) Estimation of discrete games with correlated types. *The Econometrics Journal* 17(3), 241–270.
Xu, H. (2018) Social interactions in large networks: A game theoretical approach. *International Economic Review* 59(1), 257–284.

https://doi.org/10.1017/S0266466620000420 Published online by Cambridge University Press