Nonparametric Identification of a Nonlinear Panel Model with Application to Duration Analysis with Multiple Spells*

Kirill Evdokimov†

Department of Economics

Princeton University

This version: February 4, 2011

Abstract

This paper presents two sets of results on nonparametric identification. First, a nonparametric generalization of the quasi–differencing method is developed. A nonparametric panel data model is shown to be identified using three time periods of data. An explicit characterization of the structural function is obtained. The fixed effects and idiosyncratic errors are not separable from the covariates and hence affect the marginal effects. The structural function is allowed to vary over time in an arbitrary fashion. In addition, a new nonparametric panel transformation model is introduced and is shown to be identified.

The first result is then used to establish nonparametric identification of several duration models with multiple spells. The existing results are substantially extended by allowing for the nonseparability of the unobserved heterogeneity and the covariates in the specification of the hazard rate. As an important consequence, the paper demonstrates identification of a multiple state duration model that treats unobserved heterogeneity as a fixed effect, rather than as a random effect, as has been done in previous studies. Identification of duration models with multivariate unobserved heterogeneity and censoring is also established.

Keywords: *duration model, generalized accelerated failure model, multiple spells, nonlinear panel, nonparametric identification.*

*I am very grateful to Jaap Abbring, Joseph Altonji, Donald Andrews, Xiaohong Chen, James Heckman, Ilze Kalnina, Simon Lee, Taisuke Otsu, Peter Phillips, Edward Vytlacil, and especially Bo Honoré and Yuichi Kitamura for their insightful comments and suggestions. I also thank the participants at the North American Summer Meeting of the Econometric Society 2009, SITE Summer Workshop on Advances in Nonparametric Econometrics, Stats in the Chateau Summer School, Cowles Summer Conference 2010, Advancing Applied Econometrics Conference, Greater NY Econometrics Colloquium 2010, EC2, and the econometric seminar at Yale for their valuable comments. This paper supersedes my earlier paper "Further Results on Identification of Panel Data Models with Unobserved Heterogeneity." All errors are mine. Financial support from the Cowles Foundation at Yale University and from the Gregory C. Chow Econometric Research Program at Princeton University is gratefully acknowledged.

†Kirill Evdokimov, Princeton University, Department of Economics, Fisher Hall 001, Princeton, NJ 08544; tel: +16092580161, fax: +16092586419, e-mail: `kevdokim@princeton.edu`.

1 Introduction

The importance of taking latent individual characteristics into account is widely recognized by the empirical economic analysis. Having multiple observations per individual, such as in panel and in multiple spell duration models, greatly enhances researchers' ability to deal with individual unobserved heterogeneity. The recent literature has recognized that standard models impose strong and often unrealistic assumptions on the data. For example, in panel models additive separability of unobserved heterogeneity implies homogeneity of the marginal effects, a property that is hard to justify in economic models. In duration models, multiplicative separability of the unobserved heterogeneity is similarly restrictive for economic modeling. In addition, parametric specification of models may result in model misspecification and lead to inaccurate inference about the objects of interest.

To overcome these issues the literature considers nonlinear and nonparametric models. Serious difficulties arise. For example, in panel models standard first differencing yields biased and inconsistent estimates when the unobserved heterogeneity (fixed effect) is not additively separable. A solution would be to use data with a large number of observations per individual. However, many microeconometric panel and duration datasets contain just a few time periods and hence new methods of analysis are needed.

This paper contributes to the literature on panel and duration data models in two ways. First, this paper concerns itself with a nonparametric generalization of the quasi–differencing idea. The paper presents a new nonparametric panel data model with nonseparable unobserved heterogeneity and demonstrates its nonparametric identification using panel data with three or more time periods. The effect of the unobserved heterogeneity on the outcome is allowed to vary over time. The paper not only establishes identification of the structural functions, but also provides an explicit characterization of these functions.

The paper's second contribution consists of an analysis of duration models with multiple spells. The quasi–differencing method can be used to identify a panel extension of the Generalized Accelerated Failure Time (GAFT) model of Ridder (1990), as well as its special case, the Mixed Proportional Hazard model. Honoré (1993) demonstrates that duration data with multiple spells per an individual has strong identification power. As discussed by Van den Berg (2001), multiple spell duration models yield more reliable inference than the more traditional single spell duration models. Multiple spells permit more robust inference on duration models and, in particular, allow identification when the unobserved heterogeneity and observed covariates are dependent. Extending the existing literature, the duration models considered in this paper allow the effect of the unobserved heterogeneity to differ between spells with different covariates. Another interpretation of the obtained results is that the paper shows nonparametric identification of a multiple state duration model, where the effect of the unobserved heterogeneity depends on the state (or the values of the covariates), yet the unobserved heterogeneity can be arbitrarily correlated with the covariates. Such models can be used for empirical analysis in many applications; several examples are listed below.

The results of this paper apply equally to the analysis of static panel data and grouped data. Panel data contain a large number of observations on individuals i, such as persons or firms, where for each individual the data are recorded for several time periods $j = 1, \ldots, J$.[1] Grouped data contain records on a large number of groups i (e.g., families), where each group contains J individual members. To simplify the discussion, in this paper I use the panel data terminology.

Consider the following panel transformation model with unobserved heterogeneity:

$$\Lambda_j (Y_j, X_j) = m (X_j, A) + U_j, \;\; j = 1, \ldots, J \geq 3, \tag{1}$$

where $Y_j \in \mathbb{R}$ is the observed outcome and $X_j \in \mathcal{X} \subset \mathbb{R}^p$ are the observed covariates. The scalar random variables A and U_j are unobserved and represent individual-specific heterogeneity and idiosyncratic disturbance, respectively.[2] The functions $\Lambda_j (\cdot)$ and $m (\cdot)$ are unknown and are modeled nonparametrically. It is assumed that the functions $\Lambda_j (\cdot)$ and $m (\cdot)$ are strictly increasing in Y_j and A, respectively. The transformation function $\Lambda_j (\cdot)$ is allowed to depend on the value of the covariate, which is more general than what is usually allowed in the analysis of transformation models.[3,4] The unobserved heterogeneity A is allowed to be arbitrarily correlated with the covariates but is assumed to be independent of the disturbances U_j, conditional on the covariates. The disturbance terms U_j are assumed to be independent over j, conditional on the covariates. The goal is to identify $\Lambda_j (\cdot)$, $m (\cdot)$, and the distribution of A and U_j conditional on the covariates.

To justify interest in model (1), consider the following special cases and interpretations of this model.

First, take $m (x, \alpha) \equiv \alpha$ and denote $g_j (x, \nu) \equiv \Lambda_j^{-1} (\nu, x)$, where $\Lambda_j^{-1} (\nu, x)$ is the inverse of $\Lambda_j (y, x)$ in the first argument. Then, equation (1) can be written as

$$Y_j = g_j (X_j, A + U_j), \tag{2}$$

which is a nonparametric panel data model with nonseparable unobserved heterogeneity. In this model, the derivative $\partial g_j (x, \nu) / \partial x$ depends on ν; thus the model allows for heterogeneous marginal effects that depend on $V_j = A + U_j$. The structural function $g_j (\cdot)$ is allowed to vary over time in an arbitrary way. Note that the shocks V_j are correlated across time through A. Obviously, (2) is a generalization of the linear model $Y_j = \beta_j' X_j + \gamma_j A + U_j$ (with $U_j = \gamma_j \widetilde{U}_j$).[5]

[1] Time periods are indexed by j instead of the more traditional index t, because symbols t and T are reserved to represent the lengths of duration spells.

[2] Here the subscript i is suppressed; $Y_j \equiv Y_{ij}$, $X_j \equiv X_{ij}$, $U_j \equiv U_{ij}$, and $A \equiv A_i$, $i = 1, 2, \ldots$

[3] Define $\Lambda_j^{-1} (\cdot, x)$ to be the inverse of the function $\Lambda_j (\cdot, x)$ in the first argument. Then equation (1) can also be written as $Y_j = \Lambda_j^{-1} (m (X_j, A) + U_j, X_j)$, which does not contain covariates on the left-hand side but appears more cumbersome.

[4] See (Horowitz 1996) for a review of the literature and a list of applications of transformation models.

[5] Another special case of model (2) is the transformation model extension of the nonparametric panel model of Porter (1996)

$$\lambda_j (Y_j) = \varphi_j (X_j) + A + U_j,$$

where $\varphi_j (\cdot)$ is an unknown regression function, and $\lambda_j (\cdot)$ is a strictly increasing unknown transformation function. This model corresponds to specifying $g_j (x, v) = \lambda_j^{-1} \left(\varphi_j (x) + v \right)$ in (2).

In addition, the transformation model (1) has a close connection to the duration analysis. Model (1) can be seen as a Panel Generalized Accelerated Failure Time (PGAFT) model, which extends the GAFT model of Ridder (1990). A very important special case of this model is the Mixed Proportional Hazard (MPH) multiple spell duration model.[6] Multiple spell duration analysis has been used in a wide variety of empirical applications. These in particular include studies of employment and unemployment durations (Heckman and Borjas, 1980; Flinn and Heckman, 1983; Bonnal, Fougere, and Serandon, 1997), durations of the life span of siblings or twins in development and health economics (Guo and Rodriguez, 1992; Hougaard, Harvald, and Holm, 1992), birth intervals (Newman and McCulloch, 1984; Heckman, Hotz, and Walker, 1985), recurrences of an illness or tumor (Wei, Lin, and Weissfeld, 1989; Lin, Sun, and Ying, 1999), and intervals between purchasing a product (Gonul and Srinivasan, 1993; Allenby, Leone, and Jen, 1999); see Van den Berg (2001) for other examples. In many of these applications, it is reasonable to expect the effect of the unobserved heterogeneity to vary across spells with different states or covariates.[7] For instance, the effect of the unobserved ability on the durations of employment and unemployment may differ non-proportionally; similarly, the effect of the unobserved fertility on the age at first birth and the time between the birth of the first and the second child are likely to differ non-proportionally. At the same time, it is often not plausible to assume independence between the unobserved heterogeneity and the observed covariates. This paper provides an econometric framework for analysis of such duration models.

For example, assume that the disturbances U_j are independent of the covariates and have a cumulative distribution function (CDF) $F_{U_j|X_j}(u|x) = 1 - \exp(-e^u)$. Then, model (1) can be interpreted as an MPH model, where $T_j \equiv Y_j \geq 0$ is the length of spell j, X_j is the vector of covariates (that are constant over the duration of a spell but may vary between spells), and A is the unobserved heterogeneity that can be arbitrarily correlated with the covariates, i.e., is a fixed effect.[8] The hazard rate has the form:

$$\theta_{T_j|X_j,A}(t|x,\alpha) = h_j(t,x)\,\gamma(x,\alpha), \tag{3}$$

where $h_j(\cdot)$ and $\gamma(\cdot)$ are positive functions, and $\gamma(x,\alpha)$ is strictly increasing in α for all x.[9] Honoré (1993) considers identification of a multiple spell duration model with the hazard rate of a more

[6] Duration distributions are usually studied in terms of their hazard rates. The hazard rate function of the continuous positive random variable $T_j \equiv Y_j$ is defined as

$$\theta_{T_j|X_j,A}(t|x,\alpha) = \lim_{\epsilon \searrow 0} P\{t \leq T_j < t+\epsilon | T_j \geq t, X_j = x, A = \alpha\}/\epsilon.$$

[7] A duration model with the hazard rate $\theta_{Y_j}(t_j|x_j,\alpha) = h_j(t_j,x_j)\,\alpha$ implies that the ratio of hazard rates $\theta_{Y_j}(t_j|x_j,\alpha)/\theta_{Y_k}(t_k|x_k,\alpha)$ does not depend on α. In other words, hazard rates corresponding to different states and/or covariate values are proportional. This assumption appears too restrictive since it requires the unobserved heterogeneity to have a similar effect on the hazard rates corresponding to different states (e.g., employment and unemployment).

[8] See Graham and Powell (2008) for a discussion of the "fixed effect" terminology.

[9] Here $h_j(y,x) \equiv (\partial\Lambda_j(y,x)/\partial y)\exp(\Lambda_j(y,x))$ and $\gamma(x,\alpha) \equiv \exp(-m(x,\alpha))$. One needs to impose some normalizations to uniquely identify functions $h_j(y,x)$ and $\gamma(x,\alpha)$. See Section 3.2 for details.

restrictive form $\widetilde{\theta}_{T_j|X_j,A}(t|x,\widetilde{\alpha}) = h_j(t,x)\,\widetilde{\alpha}$; i.e., the unobserved heterogeneity α is only allowed to enter the specification of the hazard rate multiplicatively. Specification (3) has a more general functional form that may be useful for economic modelling.[10] For instance, duration specifications implied by structural search models usually contain a (conditional) CDF as one of the elements (e.g., CDF of wage offers), and this CDF may depend on the unobserved heterogeneity A (e.g., skill). However, it is not possible to model the dependence of the CDF on A using a multiplicative specification, since the range of the CDF must be $[0,1]$. In contrast, hazard rate specification (3) allows modelling the unknown CDF that depends on the unobserved heterogeneity A.[11]

In addition, identification of model (3) has important implications for the analysis of duration models with multiple states, e.g., Flinn and Heckman (1982, 1983) and the other examples listed above. This paper demonstrates that such models can be identified treating the unobserved heterogeneity as a fixed effect, rather than as a random effect, as was done in previous studies. To fix ideas, consider a simple labor market model with two individual states K_j, (e)mployment and (u)nemployment. The often used approach consists of estimating a one-factor multiple state duration model

$$\theta_{T_j|X_j,K_j,A}(t|x,k,\alpha) = z_k(t)\exp\left\{\delta_k' x + \gamma_k \alpha\right\}, \tag{4}$$

where x and the state $k \in \{e,u\}$ are the covariates, α is a *random* effect, and $z_k(\cdot)$ and $(\delta_k,\gamma_k)_{k\in\{e,u\}}$ are the function and parameters to estimate. Identification of a nonparametric MPH version of the model (4) follows from Theorem 2 in Honoré (1993) under the assumption of independence between the covariates and the unobserved heterogeneity A. However, the semiparametric model (4) is a special case of model (3). Therefore, model (4) can be identified and estimated allowing for an arbitrary correlation between the covariates and the unobserved heterogeneity, i.e., treating A as a *fixed* effect. Permitting dependence between the covariates and the unobserved heterogeneity may be important for the empirical analysis.

Specification (3) is only one of the Mixed Proportional Hazard models considered in the paper. Several duration models are proposed, and the difference between the assumptions identifying the models is discussed. Although the PGAFT model requires $J = 3$ spells per individual for identification, it is shown that one of the considered MPH models is identified using data on only $J = 2$ spells.

This paper demonstrates that model (1) can be nonparametrically identified using panel data with at least three time periods. The identification method developed in this paper can be seen as a nonparametric generalization of quasi–differencing (e.g., see Chamberlain, 1984). The main idea of the identification strategy is easy to demonstrate using model (2). Consider first a linear panel model without covariates $Y_j = \gamma_j \cdot (A + U_j)$, where $\gamma_j \neq 0$ are *scalars*, $E[AU_j] = 0$ for all j, and

[10] The conditional methods of Chamberlain (1985) and Ridder and Tunali (1999) do not identify the effect of covariates in (3), because in this specification the covariates x are not separable from t and α.

[11] Van den Berg (2001) carefully investigates economic implications of separability assumptions in duration models and presents examples where those do and do not hold.

U_j are serially uncorrelated. Then, the following moment restriction identifies the ratio γ_1/γ_2:

$$E\left[\left(\gamma_1^{-1}Y_1 - \gamma_2^{-1}Y_2\right)Y_3\right] = 0. \tag{5}$$

Now consider a nonlinear panel model $Y_j = g_j(A + U_j)$ without covariates, where $g_j(\cdot) : \mathbb{R} \to \mathbb{R}$ are *strictly increasing functions*. Assume U_1, U_2, U_3, and A are mutually independent. Suppose some strictly increasing functions $G_j(\cdot)$, $j = 1, 2$, satisfy the following independence condition

$$G_1(Y_1) - G_2(Y_2) \perp Y_3. \tag{6}$$

Obviously, the above condition holds if one takes the functions $G_j(\cdot)$ to be the inverse functions $g_j^{-1}(\cdot)$. Interestingly, Section 2 shows that no other strictly increasing functions $G_j(\cdot)$ can satisfy condition (6); i.e., functions $G_j(\cdot)$ *must* be equal to the inverse functions $g_j^{-1}(\cdot)$ up to a location and scale normalization.[12] Note that condition (5) is a mean independence quasi-difference restriction. Condition (6) can be seen as a stronger version of (5), since it requires full independence. Strengthening the condition on the unobservables from mean to full independence thus allows nonparametric identification of the strictly increasing functions $g_j(\cdot)$. Model (2) with covariates can be identified by the same independence restriction (6), which holds conditional on the values of covariates X.

The identifying independence restriction (6) is intuitive. An advantage of this restriction is that it can be readily employed to estimate the functions $G_j(\cdot)$ and $g_j(\cdot)$. This is discussed in Section 2.2 below.

The quasi–differencing identification approach does not require the structural functions $g_j(\cdot)$ and the distributions of U_j to be the same over time, because it does not rely on time homogeneity for identification. The assumption of time homogeneity has been previously used for identification of nonparametric panel data models. For example, Chernozhukov, Fernandez-Val, Hahn, and Newey (2010), Evdokimov (2008), Graham and Powell (2008), and Hoderlein and White (2009) allow additive and/or multiplicative time effects in their models, but otherwise require the structural relationship between the covariates, unobserved heterogeneity, and outcomes to be stable (time homogeneous) over time. The assumption of time homogeneity is powerful and relevant in many economic applications. Yet, in some scenarios this assumption may fail, especially when the panel covers a substantial period of calendar time. The nonparametric quasi–differencing method does not rely on the time homogeneity assumption, thus extending the existing literature.

That quasi–differencing does not require the time homogeneity assumption comes at a price. Even in the linear model, quasi–differencing requires at least three time periods of data. At the same time, the standard first–differencing method in linear panels, as well as the methods of Chernozhukov, Fernandez-Val, Hahn, and Newey (2010), Evdokimov (2008), Graham and Powell (2008), and Hoderlein and White (2009), obtain results using just two time periods.

Two more issues are addressed in this paper. Duration data often suffer from censoring. Section

[12] One also needs to normalize one of the functions $g_j(\cdot)$ to be strictly increasing or strictly decreasing.

4.1 shows that the analysis of the multiple spell duration models above can be carried out under the common assumptions on the mechanism of duration censoring.

Finally, the paper establishes identification of models with multivariate unobserved heterogeneity. For example, Section 4.2 establishes identification of a duration model with the following hazard rate

$$\theta_{T_j|X_j,W_j,B}\left(t|x,w,\beta\right) = h_j\left(t,x\right)\gamma\left(x,\beta' w\right), \tag{7}$$

where X_j and W_j are observable covariates that do not contain common elements. The random time-invariant vector $B \in \mathbb{R}^q$ represents the multivariate unobserved heterogeneity. The elements of B are allowed to be arbitrarily correlated with each other and with the covariates; i.e., they are treated as fixed effects. Covariates W_j can be seen as factor-loadings. The hazard rate (7) is a generalization of the hazard rate (3) that allows for multivariate unobserved heterogeneity. Identification of the nonparametric model (7) implies identification of a corresponding semiparametric duration model with multivariate unobserved heterogeneity. Examples and details on identification are presented in Section 4.2.

Let us conclude the introduction with a discussion of the related literature. Transformation models have been used in econometrics and statistics at least since the seminal work of Box and Cox (1964). In the cross-section settings, numerous papers have exploited the condition of independence between an observable covariate and the unobservable error term for nonparametric identification and estimation of the transformation function; an incomplete list includes Han (1987), Ridder (1990), Abbring and Ridder (2010), Horowitz (1996), and Jacho-Chavez, Lewbel, and Linton (2006). Chiappori and Komunjer (2008) use a related completeness condition for identification. In contrast to these papers, the nonparametric quasi–differencing identification method exploits the conditional independence between the unobservable individual-specific effects and the idiosyncratic disturbances.

Van den Berg (2001) provides an excellent review of duration models, including the multiple spell and multiple state duration models.[13] Horowitz and Lee (2004) and Lee (2008) provide estimation procedures for semi-nonparametric versions of Honoré's (1993) multiple spell duration model. Abbring and den Berg (2003) consider identification of treatment effects in duration models, while Honoré and de Paula (2010) consider multiple spell duration models with strategic interactions. Khan and Tamer (2007) and Woutersen (2000) develop estimators for analysis of censored duration models. Abrevaya (1999) provides an estimator of the linear index coefficients in a general semiparametric fixed-effects panel transformation model.

Holtz-Eakin, Newey, and Rosen (1988), Chamberlain (1992), Wooldridge (1997), Blundell, Griffith, and Windmeijer (2002), and Bonhomme (2010) consider several semiparametric generalizations of the idea of quasi–differencing, while Graham and Powell (2008) study a related correlated random coefficient model with nonparametric specification of the coefficients. This paper extends the idea of quasi–differencing in a different and fully nonparametric way.

The literature on nonlinear and nonparametric analysis of panel data is growing very rapidly.

[13] Van den Berg (2001) calls some of these models multivariate mixed proportional hazard models.

Arellano and Honoré (2001) and Arellano and Hahn (2006) review the literature on semiparametric nonlinear panel data models. Recent contributions on nonparametric panel data models include Arellano and Bonhomme (2009), Altonji and Matzkin (2005), Bester and Hansen (2007), Chernozhukov, Fernandez-Val, Hahn, and Newey (2010), Cunha, Heckman, and Schennach (2010), Evdokimov (2008), Hoderlein and White (2009), and Hu and Shum (2008). The identification strategies of these papers differ substantially from the nonparametric quasi–differencing method proposed here. Moreover, none of these papers considers panel transformation models or multiple spell duration models such as the PGAFT and MPH models discussed above.

The rest of the paper is organized as follows. For clarity of exposition, I first establish the nonparametric identification of model (2) and of the corresponding MPH duration model in Sections 2.1 and 2.2. Then, identification of the transformation model (1) and of the duration model (3) is considered in Sections 3.1 and 3.2. Sections 4.1 and 4.2 demonstrate identification of censored duration models and models with multivariate unobserved heterogeneity. Section 5 concludes.

In what follows the notation $Y \equiv (Y_1, \ldots, Y_J)$, $X \equiv (X_1, \ldots, X_J)$, and $U = (U_1, \ldots, U_J)$ will often be used.

2 Nonparametric Quasi–Differencing

2.1 Identification of Panel Model (2)

Identification of model (2) does not require that the covariates X_j have common support across j. Moreover, their supports do not even need to overlap. Fix an $\overline{x}$ in the support of X_2 and define the set $\mathcal{X}_1 = \{x_1 \in \mathcal{X} : f_{X_1,X_2}(x_1, \overline{x}) > 0\}$; i.e., the set $\mathcal{X}_1$ is the support of X_1 given the event that $X_2 = \overline{x}$. The value of $\overline{x}$ should be chosen appropriately, so that the set $\mathcal{X}_1$ is not empty; for instance, one may take $\overline{x}$ to be a mode of the distribution of X_2. Here and everywhere below f is the density function if X_j is continuously distributed, and a probability mass function if X_j has a discrete distribution.[14] For each $x_1 \in \mathcal{X}_1$, define $\widetilde{x}_3(x_1)$ to be an $x_3 \in \mathcal{X}$ such that $f_{X_1,X_2,X_3}(x_1, \overline{x}, \widetilde{x}_3(x_1)) > 0$ (such a point x_3 always exists due to the definition of the set $\mathcal{X}_1$). Correspondingly, define the event $\mathcal{G}(x_1) = \{X = (x_1, \overline{x}, \widetilde{x}_3(x_1))\}$.

Model (2) requires some normalizations in order to identify its structural elements. Both a location and a scale normalization are necessary, since the mean of A and the variances of A and U_j are not restricted. The following assumption is a convenient normalization:

Assumption 1. $g_2(\overline{x}, 0) = 0$ *and* $g_2(\overline{x}, 1) = 1$.[15]

[14] More generally, f_{X_j} is the Radon-Nikodym derivative of the probability measure for X_j with respect to a product measure of Lebesgue (for the continuously distributed components of X_j) and counting (for the discrete components of X_j) measures.

[15] This normalization implicitly assumes that $f_{Y_2|X_2}(0|\overline{x}) > 0$ and $f_{Y_2|X_2}(1|\overline{x}) > 0$. This implicit assumption can be trivially relaxed in this and the following sections at the expense of introducing more notation. One can always find $\overline{y}_a$ and $\overline{y}_b$, such that $\overline{y}_a < \overline{y}_b$, $f_{Y_2|X_2}(\overline{y}_a|\overline{x}) > 0$ and $f_{Y_2|X_2}(\overline{y}_b|\overline{x}) > 0$. Then one could normalize $g_2(\overline{x}, 0) = \overline{y}_a$ and $g_2(\overline{x}, 1) = \overline{y}_b$.

Other normalizations are possible; for instance, one can assume that $g_2(\overline{x}, 0) = 0$ and $\partial g_2(\overline{x}, 0)/\partial v = 1$.

Assumption 2. *$J = 3$, $\{Y, X, U, A\}$ are random, and Y is generated according to (2). In addition:*

(i) *the functions $g_j(x, v)$ are strictly increasing in v for all x and $j = 1, 2, 3$;*

(ii) *U_1, U_2, U_3, and A are mutually independent, conditional on $\mathcal{G}(x_1)$ for all $x_1 \in \mathcal{X}_1$;*

(iii) *$E[U_j|\mathcal{G}(x_1)] = 0$ for all $x_1 \in \mathcal{X}_1$ and $j = 1, 2$;*

(iv) *the conditional distribution of U_j is absolutely continuous with respect to Lebesgue measure, and $f_{U_j}(u|\mathcal{G}(x_1)) \equiv f_{U_j|X_1,X_2,X_3}(u|x_1, \overline{x}, \widetilde{x}_3(x_1)) > 0$ for all $u \in \mathbb{R}$, $x_1 \in \mathcal{X}_1$, and $j = 1, 2$;*

(v) *for all $x_1 \in \mathcal{X}_1$ the set of points where the conditional characteristic function $\phi_{U_3}(s|\mathcal{G}(x_1))$ of U_3 is nonzero (i.e., the set $\{s \in \mathbb{R} : \phi_{U_3}(s|\mathcal{G}(x_1)) \neq 0\}$) is everywhere dense in $\mathbb{R}$;*[16]

(vi) *for each $x_1 \in \mathcal{X}_1$, there exist constants $C_{\alpha 1} > 0$, $C_{\alpha 2} > 0$, $\alpha_0 \in \mathbb{R}$, and $\varepsilon_0 > 0$, such that the conditional cumulative distribution $F_A(\alpha|\mathcal{G}(x_1))$ is differentiable for all $\alpha \in B_{\varepsilon_0}(\alpha_0) = \{\alpha \in \mathbb{R} : |\alpha - \alpha_0| < \varepsilon_0\}$, and $C_{\alpha 1} < \partial F_A(\alpha|\mathcal{G}(x_1))/\partial\alpha < C_{\alpha 2}$ for all $\alpha \in B_{\varepsilon_0}(\alpha_0)$.*

(vii) *the functions $g_j(x_j, v)$, $f_{U_j|X_1,X_2,X_3}(u|x_1, x_2, x_3)$, $F_{A|X_1,X_2,X_3}(\alpha|x_1, x_2, x_3)$, and $\widetilde{x}_3(x_1)$ are everywhere continuous in the continuous components of x_j for all $\alpha \in \mathbb{R}$, $u \in \mathbb{R}$, and $j = 1, 2, 3$.*

A discussion of these assumptions is in order. Assumption 2(i) ensures the invertibility of the function $g_j(x_j, v)$ in the second argument. Note that the functions $g_j(x, v)$ do not need to be continuous in v. The independence Assumption 2(ii) is strong; however, some independence assumptions are usually needed for nonparametric identification of unknown functions with nonseparable unobservables. Moreover, this assumption is naturally satisfied by the duration model in Section 2.2. Location restriction of Assumption 2(iii) is standard. The full support Assumption 2(iv) is imposed to simplify the presentation of the results but is not essential for the identification strategy; see Remark 1 below. Assumption 2(v) is technical and is very weak; all standard distributions satisfy this assumption; see also the discussion below. Assumption 2(vi) implies that the conditional distribution of the unobserved heterogeneity A is continuous at least in some small neighborhood. This assumption is necessary; when A has a discrete distribution, the model is not identified; see Remark 2 below. Note that the researcher does not need to know the values of α_0, ε_0, $C_{\alpha 1}$, or $C_{\alpha 2}$ for identification or estimation. Finally, Assumption 2(vii) is needed only when the covariates X_j contain continuously distributed components. In this case the conditioning event $\mathcal{G}(x_1)$ has probability zero and a continuity assumption such as Assumption 2(vii) is needed.[17]

[16] For any event ϑ, $\phi_{U_j}(s|\vartheta)$ is the conditional characteristic function U_j, given ϑ, and is defined as $\phi_{U_j}(s|\vartheta) = E[\exp(isU_j)|\vartheta]$, where $i = \sqrt{-1}$.

[17] It is straightforward to allow the functions to be almost everywhere continuous. In this case the functions $g_j(x, v)$ are identified at all points of continuity in x.

Theorem 1. *Suppose Assumptions 1 and 2 hold. Suppose some functions $G_1(x_1, y)$ and $G_2(\overline{x}, y)$ (i) are strictly increasing in y for all y and all $x_1 \in \mathcal{X}_1$, and (ii) for all $x_1 \in \mathcal{X}_1$ satisfy the condition*

$$G_1(x_1, Y_1) - G_2(\overline{x}, Y_2) \perp Y_3 | \mathcal{G}(x_1). \tag{8}$$

Then the following equalities hold for all $x_1 \in \mathcal{X}_1$ and (Lebesgue) almost all points $v \in \mathbb{R}$ (in particular, for all points of continuity in v)[18]

$$\begin{aligned} g_1(x_1, v) &= G_1^{-1}(x_1, [G_2(\overline{x}, 1) - G_2(\overline{x}, 0)] v + G_2(\overline{x}, 0) + \Delta_{G,1}(x_1)) \text{ and} \\ g_2(\overline{x}, v) &= G_2^{-1}(\overline{x}, [G_2(\overline{x}, 1) - G_2(\overline{x}, 0)] v + G_2(\overline{x}, 0)), \end{aligned}$$

where $\Delta_{G,1}(x_1) = E[G_1(x_1, Y_1) - G_2(\overline{x}, Y_2) | \mathcal{G}(x_1)]$.

This theorem demonstrates identification of the function $g_1(x_1, \cdot)$ for all $x_1 \in \mathcal{X}_1$ (and of the function $g_2(x_2, \cdot)$ at the point $x_2 = \overline{x}$). Suppose the random vector (Y, X) was generated by model (2) and the conditions of the theorem hold. The theorem then establishes that for any such random vector (Y, X), there exist *unique* functions $g_1(\cdot)$ and $g_2(\overline{x}, \cdot)$ that can generate vector (Y, X) in model (2). Indeed, suppose two different sets of functions $\{g_1(\cdot), g_2(\overline{x}, \cdot)\}$ and $\{\widetilde{g}_1(\cdot), \widetilde{g}_2(\overline{x}, \cdot)\}$ satisfying Assumptions 1 and 2 correspond to the same distribution of (Y, X). Define functions $G_1(x_1, y) \equiv \widetilde{g}_1^{-1}(x_1, y)$ and $G_2(\overline{x}, y) \equiv \widetilde{g}_2^{-1}(\overline{x}, y)$, where $\widetilde{g}_j^{-1}(\cdot)$ is the inverse function of $\widetilde{g}_j(\cdot)$ in the second argument, which exists because $\widetilde{g}_j(\cdot)$ satisfies Assumption 2(i). Notice that $G_2(\overline{x}, 0) = 0$ and $G_2(\overline{x}, 1) = 1$ because $\widetilde{g}_2(\cdot)$ satisfies the normalization Assumption 1. In addition, for these functions $G_j(\cdot)$ the above-defined $\Delta_{G,1}(x_1)$ satisfies $\Delta_{G,1}(x_1) \equiv 0$ for all $x_1 \in \mathcal{X}_1$. Moreover, the functions $G_1(\cdot)$ and $G_2(\overline{x}, \cdot)$ satisfy the independence condition (8) by construction. Thus, all the conditions of Theorem 1 are satisfied and hence the conclusion of the theorem holds. Remember that $\{g_1(\cdot), g_2(\overline{x}, \cdot)\}$ generates the same distribution of (Y, X). Thus for all $x_1 \in \mathcal{X}_1$ we have

$$\begin{aligned} g_1(x_1, v) &= G_1^{-1}(x_1, v) = \widetilde{g}_1(x_1, v) \text{ and} \\ g_2(\overline{x}, v) &= G_2^{-1}(\overline{x}, v) = \widetilde{g}_2(\overline{x}, v), \end{aligned}$$

where in each line the first equality holds for Lebesgue almost all points $v \in \mathbb{R}$ by Theorem 1 and the second equality holds by construction of $G_j(\cdot)$. This shows that if two sets of structural functions $\{g_1(\cdot), g_2(\overline{x}, \cdot)\}$ and $\{\widetilde{g}_1(\cdot), \widetilde{g}_2(\overline{x}, \cdot)\}$ generate the same distribution of vector (Y, X) in model (2), then $g_1(x_1, v) = \widetilde{g}_1(x_1, v)$ and $g_2(\overline{x}, v) = \widetilde{g}_2(\overline{x}, v)$ for almost all $v \in \mathbb{R}$.

The functions $g_2(\cdot)$ and $g_3(\cdot)$ can be identified by switching the roles of Y_1, Y_2, and Y_3.

The proof of the theorem is given in the Appendix. Here I briefly discuss the idea of the proof. First, fix any $x_1 \in \mathcal{X}_1$ and suppress the conditioning on $\mathcal{G}(x_1)$ as well as the arguments x_1, x, and x_3 of functions G_j and g_j. Define function $\overline{G}_j(\cdot) = G_j(g_j(\cdot))$ and note that the following

[18] Obviously, if the functions are assumed to be right- or left-continuous they are identified everywhere.

statements are equivalent:

$$\begin{aligned} G_1(Y_1) - G_2(Y_2) &\perp Y_3 \iff \\ \overline{G}_1(A+U_1) - \overline{G}_2(A+U_2) &\perp g_3(A+U_3) \iff \\ \overline{G}_1(A+U_1) - \overline{G}_2(A+U_2) &\perp A+U_3 \iff \\ \overline{G}_1(A+U_1) - \overline{G}_2(A+U_2) &\perp A, \end{aligned}$$

where the second line follows from the definition of function $\overline{G}_j(\cdot)$ and equation (2), the third line follows from the strict monotonicity of function $g_3(\cdot)$ (Assumption 2(i)), and the fourth line follows from Lemma 1 in the Appendix, which uses Assumptions 2(ii,v).

We would like to show that only linear functions $\overline{G}_1(\cdot)$ and $\overline{G}_2(\cdot)$ with the same slopes can satisfy the above independence condition. This will imply that functions $G_j(\cdot)$ are the inverse functions of $g_j(\cdot)$ up to a location and scale normalization. Consider the function

$$\varkappa(\alpha, u_1, u_2) = \overline{G}_1(\alpha+u_1) - \overline{G}_2(\alpha+u_2) - \left[\overline{G}_1(\alpha_0+u_1) - \overline{G}_2(\alpha_0+u_2)\right], \tag{9}$$

where α_0 is defined in Assumption 2(vi). The above chain of equivalent conditions together with Assumption 2(ii) implies that $\varkappa(A, U_1, U_2) \perp A$. Note also that $\varkappa(\alpha_0, u_1, u_2) = 0$ for all $(u_1, u_2) \in \mathbb{R} \times \mathbb{R}$ by construction.

Take any $\epsilon > 0$ and note that

$$P\left[|\varkappa(A, U_1, U_2)| > \epsilon\right] = \lim_{r \searrow 0} P\left[|\varkappa(A, U_1, U_2)| > \epsilon | A \in (\alpha_0 - r, \alpha_0 + r)\right] = 0,$$

where the first equality holds because of the independence between $\varkappa(A, U_1, U_2)$ and A, while the proof of the second equality is lengthy and uses Luzin's theorem (e.g., Folland, 1999) and that $\varkappa(\alpha_0, u_1, u_2) \equiv 0$. Then, $\varkappa(\alpha, u_1, u_2) = 0$ for almost all $(\alpha, u_1, u_2) \in B_{\varepsilon_0}(\alpha_0) \times \mathbb{R} \times \mathbb{R}$, since $\epsilon > 0$ can be taken arbitrarily small.

Thus, the left-hand side of (9) is zero for almost all $(\alpha, u_j) \in B_{\varepsilon_0}(\alpha_0) \times \mathbb{R}$ and $j = 1, 2$. We can rewrite (9) as

$$\overline{G}_1(\alpha+u_1) - \overline{G}_1(\alpha_0+u_1) = \overline{G}_2(\alpha+u_2) - \overline{G}_2(\alpha_0+u_2).$$

Since u_1 and u_2 vary independently it must be that

$$\overline{G}_j(\alpha+u_j) - \overline{G}_j(\alpha_0+u_j) = c(\alpha)$$

holds for some function $c(\cdot)$ that does not depend on u_j. It is now intuitively clear that the only way u_j may cancel out on the left-hand side above for all α is when $\overline{G}_j(\cdot)$ is a linear function. More formally, note that $c(\alpha_0) = 0$ and define function $\eta(\xi) \equiv c(\xi + \alpha_0)$. The proof in the Appendix demonstrates that the function $\eta(\cdot)$ satisfies Cauchy's functional equation. The only solution of Cauchy's equation in the class of measurable functions is the linear function. Thus, $\eta(\cdot)$ and $\overline{G}_j(\cdot)$ must be linear, i.e. $\overline{G}_j(v) = c_{0j} + \overline{c}v$. Then, from the definition of the function $\overline{G}_j(v)$ it follows that $g_j(v) = G_j^{-1}(c_{0j} + \bar{c}v)$ and Assumptions 1 and 2(iii) can be used to determine the constants c_{0j} and $\bar{c}$.

Remark 1. *The full support Assumption 2(iv) is made only for simplicity of notation. It is straightforward to relax it. Let $H_A(\mathcal{G}(x_1)) \subset \mathbb{R}$ be an open set such that the derivative $\partial F_A(\alpha|\mathcal{G}(x_1))/\partial\alpha$ exists and is positive for all $\alpha \in H_A(\mathcal{G}(x_1))$. Also, let $S_{U_1}(x_1)$ be the conditional support of U_1, given $X_1 = x_1$. Then, following the proof of Theorem 1, it is straightforward to show that $g_1(x_1, v)$ is identified for all $v \in \{\alpha + u_1 : \alpha \in H_A(\mathcal{G}(x_1))$ and $u_1 \in S_{U_1}(x_1)\}$.*

Remark 2. *The identification strategy relies on A having a nondegenerate distribution. When A has a degenerate distribution (i.e., $P\{A = const\} = 1$) the identification method fails, because the independence condition 8 holds for all functions $G_1(\cdot)$ and $G_2(\cdot)$. However, this is not a problem, since it is easy to detect and handle. Conditional on covariates, the dependence between Y_1, Y_2, and Y_3 comes only from A. When Y_1, Y_2, and Y_3 are independent, conditional on covariates, one should conclude that there is no common source of heterogeneity; i.e., there is no A in the model. In that case the analysis can be performed separately for each time period, and each one period model is the model studied in Matzkin (2003).*

Note that the identification argument of Theorem 1 fails if A has a nondegenerate but discrete distribution. Example 1 in the Appendix demonstrates that the independence restriction (8) does not identify functions $g_j(\cdot)$ when A has a discrete distribution. Arguably, economists are inclined to think of unobserved heterogeneity as being continuously distributed, when present.[19]

Remark 3. *The restrictions that identify the model are different from the completeness condition. The model is shown to be identified even when the characteristic functions of U_j have real zeros (possibly even an infinite number). However, it is well known (e.g., see Mattner, 1993) that Y_j is not complete (not even bounded complete) for A in the model $Y_j = g_j(A + U_j)$ when the characteristic function of U_j has zeros; hence, an identification strategy based on completeness conditions would fail to identify the model in this case.*

Moreover, a simple sufficient condition for Assumption 2(v) exists. When the density of U_3 has tails that are no thicker than exponential, Assumption 2(v) holds; see page 3 of Paley and Wiener (1934) and also (d'Haultfoeuille forthcoming).[20] *This sufficient condition is restrictive. The advantage of this condition is that it has an easy interpretation; in a number of economic applications researchers may have some intuition or an economic model that implies U_j having thin tails (or even bounded support).*[21]

To identify the functions $g_j(x, v)$ for all j, one exchanges the roles of time indices j and repeatedly applies Theorem 1. No further normalizations beyond Assumption 1 are needed; once

[19] Econometric models with discrete unobserved heterogeneity are often used in applied research. However, this discrete distribution is usually seen as an approximation for the true underlying continuous distribution.

[20] Formally, Assumption 2(v) holds if there exist positive constants c_1 and c_2, such that for large u, $f_{U_3}(u|\mathcal{G}(x_1)) < c_1 \exp(-c_2 u)$. In this case $\phi_{U_3}(s|\mathcal{G}(x_1))$ is an analytic function, and hence Assumption 2(v) holds.

[21] In contrast, I am not aware of any condition that would imply that the characteristic function of a random variable has no real zeros and would have an economic interpretation.

function $g_1(\cdot)$ is identified, it serves as a source of normalizations for the functions $g_2(\cdot)$ and $g_3(\cdot)$. Note that the first use of Theorem 1 identifies function $g_1(x,\cdot)$ for all $x \in \mathcal{X}_1$, but $\mathcal{X}_1$ depends on the choice of $\overline{x}$ and may not include the whole support of X_1. Then one may need to apply Theorem 1 several times, alternating the roles of time indices, to obtain identification of $g_1(x,\cdot)$ for all x in the support of X_1.[22] Formally, suppose all the components of X_j are continuously distributed and define $\widetilde{\mathcal{X}}$ to be the set of all interior points of $\{(x_1,x_2,x_3) : f_{X_1,X_2,X_3}(x_1,x_2,x_3) > 0\}$. Assume that $\widetilde{\mathcal{X}}$ is a connected set. Then one can identify the structural functions $\{g_1(x_1,\cdot), g_2(x_2,\cdot), g_3(x_3,\cdot)\}$ for all points (x_1,x_2,x_3) in the closure of $\widetilde{\mathcal{X}}$.[23] When X_j contains discrete covariates, a similar statement holds, but its formulation requires additional notation and is not presented to save space.

The conditional distributions of A and U_j are also identified. Take any $(x_1,x_2) \in \mathcal{X}^2$ and define the event $\mathcal{G} = \{(X_1,X_2) = (x_1,x_2)\}$ to shorten the notation. First, notice that conditional on the event $\mathcal{G}$, the joint distribution of vector $(A+U_1, A+U_2)'$ is identified, since $(A+U_1, A+U_2) = \left(g_1^{-1}(x_1,Y_1), g_2^{-1}(x_2,Y_2)\right)$, where $g_j^{-1}(x,\cdot)$ denotes the inverse function of $g_j(x,\cdot)$. Then, an extension of a lemma of Kotlarski (1967) identifies the distributions of A, U_1, and U_2 using their conditional independence. The proof of the following Corollary is given in the Appendix.

Corollary 2. *Suppose that (i) the functions $g_1(x_1,v)$ and $g_2(x_2,v)$ are identified for almost all $v \in \mathbb{R}$, (ii) A, U_1, and U_2 are mutually independent, conditional on $\mathcal{G}$, and (iii) $E[U_2|\mathcal{G}] = 0$. Moreover, suppose that one of the following conditions holds:*

(iv,a) $E[|A| + |U_1| + |U_2|\,|\mathcal{G}] < \infty$*; and for all* $s \in \mathbb{R}$ *and* $j = 1,2$*, if* $\phi_{U_j}(s|\mathcal{G}) = 0$ *then* $\partial\phi_{U_j}(s|\mathcal{G})/\partial s \neq 0$*;*

(iv,b) *there exist positive constants c_1 and c_2, such that $f_{U_j}(u|\mathcal{G}) < c_1 \exp(-c_2 u)$ for large u and $j = 1,2$.*

Then the distributions of A, U_1, and U_2, conditional on $\mathcal{G}$, are identified.

Conditions (ii) and (iii) have already been imposed by Assumption 2. Condition (iv,a) is technical and very weak and holds for all standard distributions, including normal, log-

[22] Consider the following example. Suppose $X_j \in \mathbb{R}$ and the joint support of $(X_1,X_2,X_3) = \left\{(\xi_1,\xi_2) : \xi_1^2 + \xi_2^2 \leq 1\right\} \times \{0\}$; i.e., the joint support of (X_1,X_2) is a ball of radius 1 in $\mathbb{R}^2$ with the center at the origin, and for simplicity X_3 is degenerate, $X_3 = 1$ (this degeneracy does not violate any of the assumptions made above). Suppose the researcher chooses $\overline{x} = 1/2$. Then, $\mathcal{X}_1$ defined in the beginning of this section is $\mathcal{X}_1 = \left[-\sqrt{3}/2, \sqrt{3}/2\right]$, and hence Theorem 1 identifies $g_1(x_1,v)$ for all $x_1 \in \left[-\sqrt{3}/2, \sqrt{3}/2\right]$ and $v \in \mathbb{R}$. Now we can identify $g_2(\cdot)$. For any $x_2 \in [-1,1]$ we have $f_{X_1,X_2}(0,x_2) > 0$; hence, we can use the independence condition such as (8), but with conditioning on the event $\{X_1 = 0, X_2 = x_2\}$ and identify $g_2(x_2,v)$ for all $x_2 \in [-1,1]$ and $v \in \mathbb{R}$. No further normalizations are needed, since $g_1(0,v)$ has already been identified. Then, one may observe that $f_{X_1,X_2}(x_1,0) > 0$ for all $x_1 \in [-1,1]$, and use the independence condition like (8), conditional on the event $\{X_1 = x_1, X_2 = 0\}$, and identify $g_1(x_1,v)$ for all $v \in \mathbb{R}$ and all $x_1 \in [-1,1]$, and not just for $x_1 \in \mathcal{X}_1$. Of course, in this example one could have taken $\overline{x} = 0$ from the beginning and thus could have avoided the need for this chain argument. However, such chaining may be necessary if the support of X has an irregular shape.

[23] If set $\widetilde{\mathcal{X}}$ is not connected, one has to impose separate normalizations on the disconnected parts.

normal, truncated-normal, Student-t, Uniform, Cauchy, and extreme value type-I, among others.[24] Condition (iv,b), as was discussed earlier, is restrictive but has the virtue of having economic interpretation. These conditions are used to extend Kotlarski's lemma; more details are given in Evdokimov and White (2010).

2.2 Duration Model

Let us study the connection between the panel model (2) and the multiple spell duration models. Since functions $g_j(\cdot)$ are invertible, we can write (2) as a transformation model

$$\Lambda_j(Y_j, X_j) = A + U_j, \tag{10}$$

where $\Lambda_j(\cdot)$ is defined to be the inverse function of $g_j(\cdot)$ in the second argument for $j = 1, 2, 3$. This model corresponds to the model (1) with $m(x, \alpha) \equiv \alpha$ for all x. This model can also be interpreted as a multiple spell duration model. Let $T_j \equiv Y_j \geq 0$ be the length of the j-th spell, and consider the conditional survival function for spell T_j with observed covariates $X_j = x$ and the unobserved heterogeneity $A = \alpha$

$$\begin{aligned}
\overline{F}_{T_j|X_j,A}(t|x,\alpha) &= P(T_j > t|X_j = x, A = \alpha) \\
&= P(\Lambda_j(T_j, x) > \Lambda_j(t, x)|X_j = x, A = \alpha) \\
&= P(U_j > \Lambda_j(t, x) - \alpha|X_j = x) \\
&= \overline{F}_{U_j|X_j}(\Lambda_j(t, x) - \alpha|x),
\end{aligned}$$

where as usual $\overline{F}(\cdot) \equiv 1 - F(\cdot)$, the second equality follows from $\Lambda_j(t, x)$ being strictly increasing in t, and the third equality follows from the assumption of independence between U_j and A conditional on covariates X_j that was made earlier.

The above expression shows that the transformation model (1) can be seen as a panel extension of the GAFT model. The hazard rate for this model takes the form

$$\theta_{T_j|X_j,A}(t|x,\alpha) = -\frac{\partial \ln \overline{F}_{T_j|X_j,A}(t|x,\alpha)}{\partial t} = -\frac{\partial \ln \overline{F}_{U_j|X_j}(\Lambda_j(t,x) - \alpha|x)}{\partial t}. \tag{11}$$

As discussed by Ridder (1990), the hazard rate of the GAFT model in general does not have the mixed proportional form. However, an MPH model with the following hazard rate is a special case of (11)

$$\theta_{T_j|X_j,A}(t|x,\alpha) = h_j(t,x)\,\alpha^{r_j(x)} \tag{12}$$

Here $h_j(t, x)$ and $r_j(x)$ are positive functions, i.e., $h_j(t, x) > 0$ and $r_j(x) > 0$ for all $t \geq 0$, x, and $j = 1, 2, 3$.[25] For any functions $h_j(\cdot)$ and $r_j(\cdot)$ the MPH model (12) corresponds to the

[24] Obviously, condition (iv,a) holds when $\phi_{U_j}(\cdot|\mathcal{G})$ is nonvanishing, i.e., when $\phi_{U_j}(s|\mathcal{G}) \neq 0$ for all $s \in \mathbb{R}$. The assumption of the nonvanishing characteristic function of measurement errors is routinely imposed in the analysis of measurement error models.

[25] In this specification we may also include the time period j among the covariates x and not index the functions

transformation model (10) with

$$\begin{aligned}\overline{F}_{U_j|X_j}(u|x) &= \exp\left\{-e^{r_j(x)u}/r_j(x)\right\}, \text{ and} \\ \Lambda_j(t,x) &= \frac{1}{r_j(x)}\ln\left(r_j(x)\int_0^t h_j(\xi,)\,d\xi\right).\end{aligned}$$

The above takes U_j to have conditional extreme value type-I distribution as in Ridder (1990) or Horowitz (1996); however, this distribution is rescaled by $r_j(x)$.

The hazard rate (12) can be equivalently written as

$$\theta_{T_j|X_j,A}(t|x,\alpha) = h_j(t,x)\exp(r_j(x)\widetilde{\alpha}), \qquad \widetilde{\alpha} = \ln(\alpha).$$

This representation illustrates that model (12) allows covariate- and time-dependent factor loadings in the MPH model. Theorem 2 in Honoré (1993) identifies a multiple state duration model where the j-th duration spell has the hazard rate of the form $\theta_{T_j|X_j,V_j}(t|x,v) = \lambda_j(t)\phi_j(x)v$ and the random variables $(V_1,\ldots,V_J)$ do not need to be independent from each other but are assumed to be independent from the covariates $(X_1,\ldots,X_J)$. There are clear trade-offs between this model of Honoré (1993) and model (12). Theorem 2 in Honoré (1993) allows for multivariate unobserved heterogeneity, while the results of this section permit only scalar unobserved heterogeneity. However, the effect of this scalar unobserved heterogeneity varies across duration spells, since the unobserved heterogeneity enters the hazard rate nonmultiplicatively through the term $\exp(r_j(x)\widetilde{\alpha})$, which is a nonparametric generalization of the one-factor model of Flinn and Heckman (1982, 1983).[26] Moreover, in contrast to the framework of Flinn and Heckman and to Theorem 2 of Honoré (1993), the models of this paper allow for arbitrary dependence between the unobserved heterogeneity and the covariates. Section 4.2 below demonstrates that the assumption of scalar unobserved heterogeneity can be relaxed allowing for multiple-factor models.

The following assumption is imposed to identify model (12).

Assumption 3. *For all j, conditional on X_j and A, the length of the spell T_j is independent of the length of the other spells $T_{(-j)}$ and the covariates in the other time periods $X_{(-j)}$.*

This assumption implies that the hazard rate has the form $\theta_{T_j|T_1,\ldots T_{j-1},T_{j+1},\ldots T_J,X_1,\ldots,X_J,A} = \theta_{T_j|X_j,A}$. In the analysis, the covariates X_j are assumed to vary between spells but are considered constant within each spell. Take $\overline{x}$ and $\mathcal{G}(x_1)$ defined in Section 2.1, and consider the following assumption:

Assumption 4. $J=3$, $\{(T_1,T_2,T_3),X,A\}$ *are random,* (T_1,T_2,T_3) *are generated according to the duration model with the hazard rate (12), and for all* $j=1,2,3$*:*

$h(\cdot)$ and $r(\cdot)$ by j.

[26] Although the joint distribution of vector $(V_1,\ldots,V_J)$ is identified by Theorem 2 in Honoré (1993), in practice, researchers often estimate the one-factor model of Flinn and Heckman (1982, 1983) specifying $V_j = a_j + b_j A$, where a_j and b_j are constants that vary over j, and A is a scalar random variable that does not vary over j and is independent of covariates. This paper shows that this model is identified even when the unobserved heterogeneity A and the covariates are dependent.

(i) *$h_j(t,x) \geq 0$ and $r_j(x) > 0$ for all x, $t \geq 0$;*

(ii) *for all $x \in \mathcal{X}$, functions $\lambda_j(\tau,x) = \int_0^\tau h_j(t,x)\,dt$ are defined for all $\tau \geq 0$ and $\lim_{\tau\to\infty} \lambda_j(\tau,x) = \infty$;*

(iii) *for each $x_1 \in \mathcal{X}_1$, there are constants $C_{\alpha 1} > 0$, $C_{\alpha 2} > 0$, $\alpha_0 \in \mathbb{R}$, and $\varepsilon_0 > 0$, such that the conditional cumulative distribution $F_A(\alpha|\mathcal{G}(x_1))$ is differentiable for all $\alpha \in B_{\varepsilon_0}(\alpha_0) = \{\alpha \in \mathbb{R} : |\alpha - \alpha_0| < \varepsilon_0\}$, and $C_{\alpha 1} < \partial F_A(\alpha|\mathcal{G}(x_1))/\partial\alpha < C_{\alpha 2}$ for all $\alpha \in B_{\varepsilon_0}(\alpha_0)$;*

(iv) *normalizations $h_2(1,\overline{x}) = 1$ and $r_2(\overline{x}) = 1$ hold;*

(v) *functions $h_j(t,x)$, $r_j(x)$, $F_{A|X_1,X_2,X_3}(\alpha|x_1,x_2,x_3)$, and $\widetilde{x}_3(x_1)$ are everywhere continuous in the continuously distributed components of x and x_j for all $t \in [0,\infty)$, $\alpha \in [0,\infty)$.*

Assumptions 4(i,ii) ensure that T_j has a proper conditional distribution for all x and α.[27] Assumption 4(iii) is the same as Assumption 2(vi)discussed above and requires that the unobserved heterogeneity have a continuous distribution at least in some small neighborhood. Assumption 4(iv) is a normalization, while Assumption 4(v) is a technical smoothness restriction, similar to Assumption 2(vii). As in Honoré (1993), no assumptions on the moments or tail behavior of the density of the unobserved heterogeneity A are needed. The proof of the following theorem is given in the Appendix:

Theorem 3. *Suppose Assumptions 3 and 4 hold. Then in the duration model with the hazard rate (12) the functions $h_1(t,x)$, $r_1(x)$, and $h_2(t,\overline{x})$ are identified for all $x \in \mathcal{X}_1$ and almost all $t \geq 0$.*

The functions $h_2(\cdot)$, $h_3(\cdot)$, $r_2(\cdot)$, and $r_3(\cdot)$ can be identified by exchanging the roles of time periods j; see also Remark 2.1. Once functions $h_j(\cdot)$ and $r_j(\cdot)$ are identified, we can write the model in the form (2). Then, we can use Corollary 2 to identify the distribution of the unobserved heterogeneity A.

Assumption 4(i) requires $r_j(x) > 0$ so that $\overline{\theta}_{T_j|X_j,A}(t|x,\alpha)$ is strictly increasing in α. One can imagine a multiple state duration model where the unobserved heterogeneity has opposite effects on the spells corresponding to different states. Thus, one may be interested in the model where, say, $r_1(x) < 0$ and $r_2(x) > 0$. In fact, Theorem 3 identifies such a model. The related Theorem 1 requires that $g_j(x,v)$ be strictly increasing in v. As can be immediately seen from the proofs, "strictly increasing" may be replaced with "strictly monotone" and the identification result will hold. One will, of course, need to impose a normalization, such as assume $g_2(\cdot)$ to be strictly increasing in v, but no such restrictions on $g_1(\cdot)$ or $g_3(\cdot)$ will be needed. As a consequence, $r_j(x)$ may be allowed to have different signs for different j, which implies that in the corresponding semiparametric model (4) the signs of the parameters γ_k may vary over k.

Although this paper restricts its attention to the questions of identification, it is worth briefly discussing possible estimation procedures for the models considered. Estimation of the panel

[27] Note that allowing $h_j(t,x)$ to be zero for some t implies that the corresponding function $g_j(\cdot)$ may be discontinuous, but this is allowed by Theorem 1.

model (2) and of the related multiple spell duration model (12) can be based on the independence restriction (8), conditional on covariates. One can use sieves to specify $G_j(\cdot)$, imposing monotonicity restrictions on the sieve space. Then it is easy to specify a set of moment conditions identifying the functions $G_j(\cdot)$. There is a continuum of moment conditions that the independence restriction generates. Thus, there is a continuum of conditional moment restrictions. The work of Chen and Pouzo (2008), although it does not immediately apply, can be extended to obtain the rates of convergence of the estimator.

Importantly, estimating functions $G_j(\cdot)$ from the independence restriction (8) is easier to implement than, for example, the nonparametric maximum likelihood. The procedure does not require specification of the conditional distribution of (U_1, U_2, U_3, A), which makes it not only easier to implement, but also more robust as it leaves no room for misspecification of the distribution of the unobservables.

Fully nonparametric estimation of the duration models may not be viable in many empirical applications due to the limited sample sizes. A semiparametric model such as (4) may be of interest in such cases. Use of the sieve estimation procedure has the advantage that it readily allows adding semiparametric components to the model. Most empirical studies of durations consider discrete covariates, which significantly simplifies the conditional independence restrictions. Suppose that the function $z_k(t)$ in model (4) is written as $z_k(t) \equiv \exp(\zeta_k(t, \beta_k))$, where $\zeta_k(t, \beta_k)$ is a piecewise linear function of t, parameterized by vector β_k. Then, one can obtain an analytical expression for the corresponding transformation function $\Lambda_j(\cdot)$. This function is then used in the independence restriction to estimate the model parameters $(\beta_k, \delta_k, \gamma_k)$.

3 General PGAFT Model

3.1 Identification of the Transformation Model

The strategy of nonparametric identification of model (1) consists of two main steps. First, one considers the model conditional on the event $\{X_1 = X_2 = x\}$, which implies that $m(X_1, A) = m(X_2, A) = m(x, A)$. Then, the logic of the nonparametric quasi–differencing analysis of the previous section applies; the conditional independence restriction

$$\widetilde{\Lambda}_1(Y_1, x) - \widetilde{\Lambda}_2(Y_2, x) \perp Y_3 | \{X_1 = X_2 = x, X_3 = x_3\} \tag{13}$$

holds if and only if functions $\widetilde{\Lambda}_j(y, x)$ are equal to functions $\Lambda_j(y, x)$ up to a location and scale normalization. Importantly, the independence restriction is imposed conditional on the event that the values of the covariates in time periods 1 and 2 are equal (i.e., on the event $\{X_1 = X_2 = x\}$); this allows $m(X_1, A)$ and $m(X_2, A)$ to cancel out when and only when the transformation functions $\widetilde{\Lambda}_j(y, x)$ are equal to the true functions $\Lambda_j(y, x)$, $j = 1, 2$. Note that this argument requires assuming $f_{(X_1,X_2)}(x,x) > 0$ for all $x \in \mathcal{X}$.[28] Note that for any $x \in \mathcal{X}$ one can always choose x_3 such that $f_{(X_1,X_2,X_3)}(x,x,x_3) > 0$ if $f_{(X_1,X_2)}(x,x) > 0$.

Once functions $\Lambda_j(y,x)$ are identified for $j = 1,2$, denote $\widetilde{Y}_j = \Lambda_j(Y_j,X_j)$ and consider the model

$$\widetilde{Y}_j = m(X_j,A) + U_j, \quad j = 1,2. \tag{14}$$

This is the model considered in Evdokimov (2008). The nonparametric distributional first–differencing analysis developed in that paper can now be applied. Since the method is thoroughly studied in Evdokimov (2008), here I present only a concise description of the idea. Essentially, one first uses Corollary 2 to identify the distribution of the idiosyncratic errors U_j. Then, one uses conditional deconvolution to remove the effect of U_j from Y_j and hence obtains the distribution of $m(X_j,A)$ given the covariates. In particular, for any (x_1,x_2) one obtains $F_{m(X_1,A)|X_1,X_2}(\omega|x_1,x_2)$ and $F_{m(X_1,A)|X_1,X_2}(\omega|x_1,x_2)$. Since the fixed effect A is time invariant, the difference between these cumulative distribution functions has to be attributed to the difference between $m(x_1,\cdot)$ and $m(x_2,\cdot)$, which allows identification of the structural function $m(\cdot)$. These arguments are made precise in the proof of Theorem 4.

The assumptions needed to identify model (1) are a strengthened version of Assumptions 1 and 2 and also allow the nonparametric first–differencing analysis. Let $\mathcal{X}$ be the support of X_1. For all $(x_1,x_2) \in \mathcal{X}^2$, define the event $\mathcal{G}(x_1,x_2) = \{X = (x_1,x_2,\widetilde{x}_3(x_1,x_2))\}$, where $\widetilde{x}_3(x_1,x_2)$ is defined in Assumption 6(vii) below.

Assumption 5. *The following normalizations hold: (i)* $\Lambda_2(0,x) = 0$ *and* $\Lambda_2(1,x) = 1$ *for all* $x \in \mathcal{X}$*; (ii) for some fixed* $\overline{x}$, $m(\overline{x},\alpha) = \alpha$ *for all* $\alpha \in \mathbb{R}$.

Assumption 6. $J = 3$, $\{Y,X,U,A\}$ *are random, and* Y *is generated according to (1). In addition:*

(i) *for all* $x \in \mathcal{X}$ *and* $j = 1,2,3$, *the functions* $\Lambda_j(y,x)$ *and* $m(x,\alpha)$ *are strictly increasing in* y *and* α, *respectively;*

(ii) $f_{U_j|X_j,A,X_{(-j)},U_{(-j)}}\left(u_j|x_j,\alpha,x_{(-j)},u_{(-j)}\right) = f_{U_j|X_j}(u_j|x_j)$ *for all* $\left(u_j,x_j,\alpha,x_{(-j)},u_{(-j)}\right) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X}^2 \times \mathbb{R}^2$ *and* $j = 1,2,3$;[29]

(iii) $E[U_j|\mathcal{G}(x_1,x_2)] = 0$ *for all* $(x_1,x_2) \in \mathcal{X}^2$ *and* $j = 1,2$;

(iv) $f_{U_j}(u|\mathcal{G}(x_1,x_2)) > 0$ *for all* $u \in \mathbb{R}$, $(x_1,x_2) \in \mathcal{X}^2$, *and* $j = 1,2$;

(v) *for all* $(x_1,x_2) \in \mathcal{X}^2$, $E[|m(x_1,A)| + |U_1| + |U_2|\,|\mathcal{G}(x_1,x_2)]$ *is bounded, and for all* $s \in \mathbb{R}$ *and* $j = 1,2$, *if* $\phi_{U_j}(s|\mathcal{G}(x_1,x_2)) = 0$ *then* $\partial\phi_{U_j}(s|\mathcal{G}(x_1,x_2))/\partial s \neq 0$;

(vi) *for all* $(x_1,x_2) \in \mathcal{X}^2$ *the set* $\left\{s \in \mathbb{R} : \phi_{U_3}(s|\mathcal{G}(x_1,x_2)) \neq 0\right\}$ *is everywhere dense;*

[28] The above independence restriction also includes conditioning on $X_3 = x_3$. This is because the analysis of model (2) requires the independence between A and U_j, which is ensured by Assumption 6(ii) below conditional on $X_3 = x_3$.

[29] Index $(-j)$ stands for the "other than j" time periods.

(vii) *$f_{X_1,X_2}(x_1,x_2) > 0$ for all $(x_1,x_2) \in \mathcal{X}^2$; also take $\widetilde{x}_3(x_1,x_2)$ to be an $x_3 \in \mathcal{X}$ such that $f_{X_1,X_2,X_3}(x_1,x_2,\widetilde{x}_3(x_1,x_2)) > 0$;*

(viii) *A has a continuous distribution, conditional on $\mathcal{G}(x_1,x_2)$ for all $(x_1,x_2) \in \mathcal{X}^2$;*

(ix) *$S_A\{(X_1,X_2) = (x,\overline{x})\} = S_A\{X_1 = x\}$ for all $x \in \mathcal{X}$, where $S_A\{\vartheta\}$ is the support of A, conditional on the event ϑ;*

(x) *the functions $\Lambda_j(y,x)$, $m(x,\alpha)$, $f_{U_j|X_j}(u|x)$, $f_{A|X_1,X_2,X_3}(\alpha|x_1,x_2,x_3)$, and $\widetilde{x}_3(x_1,x_2)$ are everywhere continuous in the continuous components of x and x_j for all $y \in \mathbb{R}$, $\alpha \in \mathbb{R}$, $u \in \mathbb{R}$, and $j = 1,2,3$.*

Assumption 5 is a normalization; one needs to impose a location and scale normalization on function $\Lambda_j(\cdot,x)$ at each point x for some j, since the distribution of U_j and the functional form of $m(\cdot)$ are not restricted. One also needs a normalization on the function $m(\cdot)$ since the distribution of A is left unrestricted. Assumption 6(ii) requires that $U_j \perp \left(A, X_{(-j)}, U_{(-j)}\right)|X_j$ and is satisfied, for example, if $U_j = \sigma_j(X_j)\xi_j$, where the random variables ξ_j are i.i.d. and independent of X_j, $X_{(-j)}$, and A. Thus, Assumption 6(ii) in particular permits contemporaneous conditional heteroskedasticity. The technical Assumption 6(v) is the same as condition (v,a) of Corollary 2; it is needed to ensure that a Kotlarski type result holds and was discussed earlier. Assumption 6(v) can be replaced by the condition (iv,b) of Corollary 2. Assumption 6(viii) can be weakened, but the continuity of the distribution of A at least in some neighborhood is necessary, as demonstrated by Example 1. Assumptions similar to Assumptions 6(iii,iv,vi,x) were previously imposed to prove identification of model (2) and were discussed in the previous section.

As was discussed above and is demonstrated formally in the proof in the Appendix, to identify the model $\widetilde{Y}_j = m(X_j,A) + U_j$ one uses the nonparametric distributional first–differencing approach and conditions on the event $(X_1,X_2) = (x,\overline{x})$. Assumption 6(ix) ensures that the "extra" conditioning on $\overline{x}$ does not reduce the support of A, so that the nonparametric first–differencing analysis for $\alpha \in S_A\{(X_1,X_2) = (x,\overline{x})\}$ in fact covers all possible values of α, i.e., all $\alpha \in S_A\{X_1 = x\}$. One obtains useful identification results even when Assumption 6(ix) does not hold. Conditioning on the event $\{(X_1,X_2) = (x,\overline{x})\}$ might remove some extreme values of A from consideration, yet one can still use the nonparametric first–differencing for the "not too extreme" quantiles of the unobserved heterogeneity A. Note that Assumption 6(ix) is not needed to identify the functions $\Lambda_j(\cdot)$.

Assumption 6(vii) is a restriction on observables and can be substantially relaxed, although the condition $f_{X_1,X_2}(x,x) > 0$ for all $x \in \mathcal{X}$ still needs to hold. As can be seen from the proof, the value of the structural function $m(x,\cdot)$ is obtained by comparing the CDFs $F_{m(x,A)|X_1,X_2}(w|x,\overline{x})$ and $F_{m(\overline{x},A)|X_1,X_2}(w|x,\overline{x})$ where $\overline{x}$ is our choice of normalization. The condition $f_{X_1,X_2}(x,\overline{x}) > 0$ is needed to perform this comparison. However, if this condition does not hold, it may be possible to find an intermediate point $\widetilde{x}$ such that $f_{X_1,X_2}(\widetilde{x},\overline{x}) > 0$ and $f_{X_1,X_2}(x,\widetilde{x}) > 0$ and first obtain the value of $m(\widetilde{x},\cdot)$ and then use it to obtain $m(x,\cdot)$. One may also use a sequence of such intermediate points. Thus, we would like to assume that the points of $\mathcal{X}$ are connected. For example, if the covariates X_j have a continuous distribution, define $\widetilde{\mathcal{X}}$ to be the set of all interior points of $\{(x_1, x_2) : f_{X_1,X_2}(x_1, x_2) > 0\}$. Assume that this set is a connected set. Then one can identify the structural function $m(x, \cdot)$ for all points x in the closure of $\widetilde{\mathcal{X}}$.

The proof of the following theorem is presented in the Appendix:

Theorem 4. *Suppose Assumptions 5 and 6 hold. Then functions $\Lambda_j(y,x)$, $m(x,\alpha)$, $f_{A|X_j}(\alpha|x)$ and $f_{U_j|X_j}(u|x)$ in the model (1) are identified for all $x \in \mathcal{X}$, $y \in \mathbb{R}$, $\alpha \in S_A\{X_1 = x\}$, $u \in \mathbb{R}$, and $j = 1, 2$.*

Function $\Lambda_3(y, x)$ can be identified by switching the roles of Y_1 and Y_3. Note that model (1) can be seen as the transformation model extension of Evdokimov's (2008) model.

3.2 Duration Model

Similar to model (2), the transformation model (1) can be used for the duration analysis. Following the steps in Section 2.2, it is easy to derive the conditional survival function corresponding to model (1):

$$\overline{F}^{PGAFT}_{T_j|X_j,A}(t|x,\alpha) = \overline{F}_{U_j|X_j}(\Lambda_j(t,x) - m(x,\alpha)|x),$$

as well as the hazard rate function:

$$\theta^{PGAFT}_{T_j|X_j,A}(t|x,\alpha) = -\frac{\partial \ln \overline{F}_{T_j|X_j,A}(t|x,\alpha)}{\partial t} = -\frac{\partial \ln \overline{F}_{U_j|X_j}(\Lambda_j(t,x) - m(x,\alpha)|x)}{\partial t}. \quad (15)$$

These functions are similar to the corresponding functions in Section 2.2, except α is replaced with $m(x, \alpha)$.

Making the standard assumption $F_{U_j|X_j}(u|x) = 1 - \exp(-e^u)$ (e.g., Ridder, 1990; Horowitz, 1996) one obtains the hazard rate (3):

$$\theta_{T_j|X_j,A}(t|x,\alpha) = h_j(t,x)\gamma(x,\alpha),$$

where $h_j(t, x) \equiv (\partial\Lambda_j(t, x)/\partial t)\exp(\Lambda_j(t, x))$ and $\gamma(x, \alpha) \equiv \exp(-m(x, \alpha))$.[30]

It is important to distinguish the following two cases: the case when for a subpopulation of individuals the covariate does not change between spells, and the case when there is no such subpopulation; i.e., the covariates necessarily change between the spells for every individual. To secure identification of the transformation model (1), it was important to impose Assumption 6(vii), which corresponds to the former case. This assumption in particular implies that $f_{X_1,X_2}(x, x) > 0$ for all $x \in \mathcal{X}$.

It turns out that the assumption $f_{X_1,X_2}(x, x) > 0$ for all $x \in \mathcal{X}$ permits identification of model (3) using data on just two duration spells. In contrast, identification of the PGAFT model required

[30] Analogously to Section 2.2, one can also identify the slightly more general MPH model with the hazard rate $\theta_{T_j|X_j,A}(t|x,\alpha) = h_j(t,x)\gamma(x,\alpha)^{r_j(x)}$.

three periods of data. There is no contradiction here. The theorems in the previous sections make no parametric assumptions about the conditional distributions of U_j, while the duration model (3) implies that U_j has a known distribution that has some special properties. Identification of model (3) using two duration spells per individual combines the ideas of Honoré (1993) and Evdokimov (2008) and consists of two main steps. First, the functions $h_j(t,x)$ are identified by conditioning on the event $\{X_1 = X_2 = x\}$ and then using the method of Honoré (1993). Identification of functions $h_j(t,x)$ then implies identification of functions $\Lambda_j(t,x)$. Now, denote $\widetilde{Y}_j = \Lambda_j(T_j, X_j)$ and write $\widetilde{Y}_j = m(X_j, A) + U_j$. The logic of the identification proof in Evdokimov (2008) can then be used to identify the function $m(x,\alpha)$, which corresponds to the function $\gamma(x,\alpha)$. Theorem 5 below establishes identification of the duration model (3) using the following assumption:

Assumption 7. *$J = 2$ and $f_{X_1,X_2}(x_1,x_2) > 0$ for all $(x_1,x_2) \in \mathcal{X}^2$.*

Assumption 7 is similar to Assumption 6(vii) and the ways of relaxing it were considered in the previous section. It is useful to discuss the empirical relevance of this assumption before Theorem 5 is presented. Assumption 7 in particular implies that for a subpopulation of individuals the covariates do not change over time ($f_{X_1,X_2}(x,x) > 0$ for all $x \in \mathcal{X}$). For instance, such an assumption might hold in a study of employment durations T_j, when the covariate X_j is marital status. Alternatively, the assumption may hold if T_j is the age of the j-th twin sister ($j = 1,2$) at the time of the birth of her first child, and the covariate X_j is the number of years of education of the j-th sister.

On the other hand, in many multiple spell or multiple state duration models the assumption $f_{X_1,X_2}(x,x) > 0$ does not hold (here we include the state as one of the covariates X_j). For example, consider a study of fertility (such as Heckman, Hotz, and Walker (1985)) where the spell T_1 denotes the age of the woman at the time of the birth of her first child, and for the subsequent spells, T_j is the time between the births of the j-th and $(j-1)$-th child. In such a study, the number of previously born children $(j-1)$ will be one of the covariates X_j and hence the condition $f_{X_1,X_2}(x,x) > 0$ and Assumption 7 are not satisfied.

Obviously, Assumption 7 does not allow individual age, calendar time, or time index j to be among the covariates X_j. Thus, model (3) requires time-homogeneity of the effect of unobserved heterogeneity, although this effect may be changed by the observed covariates. In contrast, model (12) allows the effect of the unobserved heterogeneity to vary over time. For example, one may expect that the distribution of duration of employment on the first job is intrinsically different from the distribution of duration of employment on all the subsequent jobs.

Both nonparametric models (12) and (3) have the semiparametric model (4) as a special case. However, the two nonparametric models correspond to two different ways of identifying the semiparametric model, depending on the assumptions the researcher is willing to impose on the economic and econometric models.[31]

Theorems 3 and 5 (below) establish identification of the multiple spell duration models (12) and (3), respectively. For model (3), Assumption 7 allows first concentrating on the duration spells

[31] Note that with $J \geq 3$, Theorem 3 allows the factor loading coefficient γ to depend on j.

with repeated covariates (e.g., two unemployment spells). For such spells $\gamma(X_1, A) = \gamma(X_2, A)$ is time-invariant and can be treated as a fixed effect, and hence we identify the functions $h_j(t, x)$ in a way that is reminiscent of identification of the time effects $\eta_j(x)$ for $j > 1$ in the model $Y_j = m(X_j, A) + \eta_j(X_j) + U_j$ as $\eta_j(x) = E[Y_j - Y_1 | X_j = X_1 = x]$ with the normalization $\eta_1(\cdot) \equiv 0$. Once $h_j(t, x)$ are known, the nonparametric within-variation is used to obtain $\gamma(x, A)$. However, the first step of this strategy fails if either the effect of the unobserved heterogeneity varies over time (as in specification (12)), or if Assumption 7 does not hold and it is not possible to consider the spells with repeated covariates, i.e., the subpopulation with $X_j = X_1$. Theorem 3 establishes identification in a different way. Similar to the discussion of identification of model (2) in the introduction, to identify model (12) one seeks to find transformations of the dependent variables that would allow canceling out the effect of unobserved heterogeneity. Of course, the problem of identifying the duration models (3) and (12) is, in fact, harder than identification of standard panel data models. This is because one does not actually observe the conditional hazard rates on the left-hand sides of equations (3) and (12).

The following assumption is used for identification of model (3):

Assumption 8. *$J = 2$, $\{(T_1, T_2), X, A\}$ are random, and (T_1, T_2) are generated according to the duration model with the hazard rate (3). In addition, for all $j = 1, 2$:*

(i) *$h_j(t, x) \geq 0$, $\gamma(x, \alpha) > 0$, and $\gamma(x, \alpha)$ is strictly increasing in α for all $x \in \mathcal{X}$, $t \geq 0$, and $\alpha \geq 0$;*

(ii) *for all $x \in \mathcal{X}$, functions $\lambda_j(\tau, x) = \int_0^\tau h_j(t, x)\, dt$ are defined for all $\tau \geq 0$ and $\lim_{\tau \to \infty} \lambda_j(\tau, x) = \infty$;*

(iii) *A has a continuous distribution, conditional on $\{X = (x_1, x_2)\}$ for all $(x_1, x_2) \in \mathcal{X}^2$;*

(iv) *$S_A\{(X_1, X_2) = (x, \overline{x})\} = S_A\{X_1 = x\}$ for all $x \in \mathcal{X}$, where $S_A\{\vartheta\}$ is the support of A, conditional on the event ϑ;*

(v) *the following normalizations hold: (a) $h_2(1, x) = 1$ for all $x \in \mathcal{X}$; (b) $\gamma(\overline{x}, \alpha) = \alpha$ for some $\overline{x} \in \mathcal{X}$ and for all $\alpha \in S_A\{X_1 = x\}$;*

(vi) *the functions $h_j(t, x)$, $\gamma(x, \alpha)$, and $f_{A|X_1, X_2}(\alpha | x_1, x_2)$ are everywhere continuous in the continuously distributed components of x and x_j for all $t \in [0, \infty)$, $\alpha \in [0, \infty)$, and $j = 1, 2$.*

Assumptions 8(i,ii) are similar to Assumptions 4(i,ii) and ensure that T_j has a proper conditional distribution for all x and α. Assumptions 7 and 8(iii,iv) are the same as Assumptions 6(vii-ix) and were discussed above. Assumption 8(v) is a normalization, while Assumption 8(vi) is a technical smoothness restriction, similar to Assumptions 4(v) and 6(x). The proof of the following theorem is given in the Appendix:

Theorem 5. *Suppose Assumptions 3, 7, and 8 hold. Then, the multiple spell duration model with the hazard rate (3) is identified, i.e., functions $h_j(t, x)$, $\gamma(x, \alpha)$, and $f_{A|X_j}(\alpha | x)$ are identified for all $x \in \mathcal{X}$, $t \in [0, \infty)$, $\alpha \in S_A\{X_1 = x\}$, and $j = 1, 2$.*

As discussed earlier, model (12) is not a special case of model (3); the conditions of Theorem 5 require that $\gamma(x,\alpha)$ be time invariant (does not depend on j) and they impose restrictions on the support of X_j, while Theorem 3 allows the function $r_j(x)$ to vary arbitrarily over j and does not constrain the support of X_j.

4 Extensions

This section considers two extensions of the obtained results. Section 4.1 shows that the censoring can be taken into account, so that the above identification results also apply to the multiple spell duration models when the duration data are censored. Section 4.2 discusses identification of models with multivariate unobserved heterogeneity.

4.1 Censoring

The previous analysis of duration models assumed that the observed durations were uncensored. In practice, one often has to deal with censoring of duration spells. Below, I show that the joint distribution of the uncensored duration spells can be recovered from the joint distribution of the censored duration spells under standard assumptions. The theorems of the previous section then apply to the joint distribution of uncensored spells to obtain identification of the respective multiple spell duration models.

Consider first the case when the duration spells are sequential, such as employment spells, gaps between recurrences of illnesses, or spells between giving birth to children. Assume that the individual is observed for censoring time C; that is, one observes only the length of duration spells that happened to an individual over the time C. Then one observes $L \geq 0$ completed durations, where $T_1 + \cdots + T_L \leq C$, but $T_1 + \cdots + T_L + T_{L+1} > C$, so that the realizations of T_j for $j > k$ are unobserved. Assume that C is random with full support on $\mathbb{R}^+$ and that C is independent of $\{T_j,\ j = 1, \ldots\}$, conditional on the covariates. This assumption has been previously used by Horowitz and Lee (2004), Khan and Tamer (2007), and Lee (2008) for estimation of semiparametric multiple spell duration models.[32] The distribution of C can depend on the covariates, allowing for covariate dependent censoring. In the data the econometrician observes realizations of the random triplet $\left(C, L, \{T_j, X_j\}_{j=1}^{L}\right)$, where $L \geq 0$ is the number of completed duration spells and is such that $T_1 + \ldots + T_L \leq C$. To identify the PGAFT and the MPH duration models of the previous sections we need to identify $F_{T_1,\ldots,T_J|X}(\cdot)$, where $J = 3$ for Theorems 3 and 4, and $J = 2$ for Theorem 5. The conditional cumulative distribution function $F_{C|X}(\cdot)$ is identified directly from

[32] Visser (1996), Wang and Wells (1998), and Lee (2008) note that although conditional on covariates C and T_1 are independent, C and T_2 are not independent given T_1, since T_2 is censored by $(C - T_1)\,1\{C > T_1\}$.

the data. Denote $x^J = (x_1, \ldots, x_J)$. Then, for any x^J, similar to Lee (2008),

$$
\begin{aligned}
& E\left[\left.\frac{1\{C \geq T_1 + \ldots + T_J\} 1\{T_1 \leq t_1, \ldots, T_J \leq t_J\}}{1 - F_{C|X}(T_1 + \ldots + T_J|X)}\right| X = x^J\right] \\
= \;& E\left[E\left[\left.\frac{1\{C \geq T_1 + \ldots + T_J\} 1\{T_1 \leq t_1, \ldots, T_J \leq t_J\}}{1 - F_{C|X}(T_1 + \ldots + T_J|X)}\right| (T_1, \ldots, T_J), X\right]\right| X = x^J\right] \\
= \;& E\left[E\left[1\{T_1 \leq t_1, \ldots, T_J \leq t_J\}| (T_1, \ldots, T_J), X\right]| X = x^J\right] \\
= \;& F_{T_1,\ldots,T_J|X}\left(t_1, \ldots, t_J|x^J\right),
\end{aligned}
$$

where the first and the third equalities follow by the law of iterated expectations, and the second equality follows by the conditional independence between C and $(T_1, \ldots, T_J)$. Thus, we can identify the joint conditional distribution of spells $(T_1, \ldots, T_J)$ and hence can use Theorems 4, 5, and 4 to identify the corresponding duration models.

Now consider the case of parallel duration spells, for instance, life spells of siblings or twins. In this case each of the duration spells T_j is censored by an individual censoring variable C_t. Assume that one observes realizations of the random tuple $\left\{\widetilde{T}_j, X_j, C_j, D_j\right\}_{j=1}^{J}$, where D_j is 1 when the j-th duration spell is censored and is zero otherwise. Here $\widetilde{T}_j = T_j 1\{D_j = 0\} + C_j 1\{D_j = 1\}$. Assume that the censoring times $(C_1, \ldots, C_J)$ are independent of the duration spell lengths $(T_1, \ldots, T_J)$ given the covariates, although $(C_1, \ldots, C_J)$ do not need to be independent of each other or covariates. The joint conditional distribution function $F_{C_1,\ldots,C_J|X}(\cdot)$ is identified directly from data. Then, similar to the above, $F_{T_1,\ldots,T_J|X_1,\ldots,X_J}(t_1, \ldots, t_J|X_1, \ldots, X_J)$ is identified:

$$
\begin{aligned}
& E\left[\left.\frac{1\{C_1 \geq T_1, \ldots, C_J \geq T_J\} 1\{T_1 \leq t_1, \ldots, T_J \leq t_J\}}{1 - F_{C_1,\ldots,C_J|X}(T_1, \ldots, T_J|X)}\right| X = x^J\right] \\
= \;& E\left[E\left[\left.\frac{1\{C_1 \geq T_1, \ldots, C_J \geq T_J\} 1\{T_1 \leq t_1, \ldots, T_J \leq t_J\}}{1 - F_{C_1,\ldots,C_J|X}(T_1, \ldots, T_J|X)}\right| (T_1, \ldots, T_J), X\right]\right| X = x^J\right] \\
= \;& E\left[E\left[1\{T_1 \leq t_1, \ldots, T_J \leq t_J\}| (T_1, \ldots, T_J), X\right]| X = x^J\right] \\
= \;& F_{T_1,\ldots,T_J|X}\left(t_1, \ldots, t_J|x^J\right).
\end{aligned}
$$

The above derivations together with Theorems 3, 4, and 5 establish nonparametric identification of the corresponding multiple spell duration models. It is worth noting that for the semiparametric duration models the expressions above may substantially simplify, as can be seen in Horowitz and Lee (2004), Khan and Tamer (2007), Lee (2008), and Woutersen (2000).

4.2 Multiple Sources of Unobserved Heterogeneity

The results of the previous sections consider models with scalar unobserved heterogeneity. This section demonstrates that these results can be extended to allow for multivariate unobserved heterogeneity as long as it enters the model specification as a single index. Arbitrary dependence between the covariates and the unobserved heterogeneity is permitted. Allowing multivariate unobserved heterogeneity may be particularly important for the duration analysis. This section demonstrates that the identification results of the previous sections can be extended to allow for a multiple factor model.

For concreteness, I will illustrate the method using model (2), although it can be applied to all the other models in the paper. Consider the following model

$$Y_j = g_j \left(X_j, B'W_j + U_j \right) \tag{16}$$

where $X_j \in \mathbb{R}^p$ and $W_j \in \mathbb{R}^q$ are observable time-varying covariates that do not contain common elements.[33] The random time-invariant vector $B \in \mathbb{R}^q$ represents the multivariate unobserved heterogeneity. For instance, B may represent a person's different skills, such as cognitive and non-cognitive abilities. Then, B_r represents the level of skill r of an individual, and W_{jr} may represent the intensity of use of skill r on job j. The elements of B can be arbitrarily correlated with each other and with the covariates (X, W). The goal is to identify the structural functions $h_j(\cdot)$ and $\gamma(\cdot)$, and the distribution of B.

Identification of this model will require that $J \geq \max\{q, 3\}$; i.e., there are at least as many time periods as there are elements of the vector B. Identification of model (16) consists of two main steps. First, one takes some w and considers the event $\{W_1 = W_2 = w\}$. Conditional on this event, one can define univariate unobserved heterogeneity $A = B'W_j = B'w$, which does not vary over $j = 1, 2$. Then, the model becomes the panel model (2) and Theorem 1 can be applied to identify the functions $g_j(\cdot)$. Once these functions are identified, one can define $\widetilde{V}_j = g_j^{-1}(X_j, Y_j)$ and note that $\widetilde{V}_j = B'W_j + U_j$. Then, similar to step 2 of the proof of Theorem 4, one can obtain the joint distribution of $(B'W_1, \ldots, B'W_J)$, given $(X, W) = \left(x^J, w^J\right)$ for any values of $\left(x^J, w^J\right)$, where $W = (W_1, \ldots, W_J)$, $x^J = (x_1, \ldots, x_J)$, and $w^J = (w_1, \ldots, w_J)$. Then, taking the values of w^J so that the rank of the matrix w^J equals q, it is possible to obtain the conditional distribution of B given $(X, W) = \left(x^J, w^J\right)$. To see this, note that identification of the conditional distribution of $(B'W_1, \ldots, B'W_J)$ means that the characteristic function

$$\phi_{(B'w_1,\ldots,B'w_J)|X,W}\left(s_1, \ldots, s_J \middle| x^J, w^J\right) = E\left[\exp\left\{\mathrm{i}\left(s_1 B'w_1 + \ldots + s_J B'w_J\right)\right\} \middle| (X, W) = \left(x^J, w^J\right)\right]$$

is identified for all $(s_1, \ldots, s_J) \in \mathbb{R}^J$. Then, it is easy to see that

$$\begin{aligned} & \phi_{(B'w_1,\ldots,B'w_J)|X,W}\left(s_1, \ldots, s_J \middle| x^J, w^J\right) \\ = \; & E\left[\exp\left(\mathrm{i}B\left(\sum\nolimits_{j=1}^{J} s_j w_j\right)\right)\middle| (X, W) = \left(x^J, w^J\right)\right] \\ = \; & \phi_{B|X,W}\left(\sum\nolimits_{j=1}^{J} s_j w_{j1}, \sum\nolimits_{j=1}^{J} s_j w_{j2}, \ldots, \sum\nolimits_{j=1}^{J} s_j w_{jq} \middle| x^J, w^J\right), \end{aligned}$$

where w_{jk} stands for the k-th component of the vector w_j. Since $Rank\left(w^J\right) = q$, the above equation identifies $\phi_{B|X,W}\left(s_1, \ldots, s_q | x^J, w^J\right)$ for all $(s_1, \ldots, s_q) \in \mathbb{R}^q$. Thus, the conditional density function $f_{B|Z}\left(\beta | x^J, w^J\right)$ is identified.

This method of identification is similar to the proof of the Cramer-Wald device. Independently of this paper, Arellano and Bonhomme (2009) use similar manipulation of the characteristic function of $B'W$ to identify a linear panel model with multivariate unobserved heterogeneity. In contrast to

[33] The vector W_j may include a constant.

this paper, Arellano and Bonhomme (2009) do not consider any nonlinear or duration models.

This strategy identifies the distribution of B, conditional on $W = w^J$, only for w^J that satisfy the restriction that the rank of w^J is equal to q. This explains the need to have at least as many time periods as there are elements of B, i.e., $J \geq q$. When W_j has a continuous distribution and the conditional density $f_{B|X,W}\left(\beta|x^J, w^J\right)$ is continuous in w^J, the identification result holds for all w^J by continuity. When the vector W_j contains discrete components, the ability to consider only w^J with $Rank\left\{w^J\right\} = q$ is restrictive, although natural. For instance, if $W_j = (1, D_j)$, where D_j is a dummy variable, $D_j \in \{0, 1\}$, we identify the distribution of B only for the subpopulation of those who change the covariate at least once over the periods of observation. The distribution of B_2 is not identified for the subpopulation with $D = (0, \ldots, 0)$, since B_2 is never observed for this subpopulation. Similarly, for the subpopulation with $D = (1, \ldots, 1)$, one only identifies the distribution of the sum $B_1 + B_2$. Thus, in the model with multivariate unobserved heterogeneity and discrete covariates, the marginal effects for the whole population may not be identified, although one might be able to obtain bounds for it. This non-identification result for discrete covariates is well known; e.g., see Chamberlain (1982).

This identification method can be used for other models in the paper. The following MPH model, corresponding to model (16), is identified

$$\theta_{T_j|X_j,W_j,B}\left(t|x, w, \beta\right) = h_j\left(t, x\right)\left(\beta' w\right)^{r_j(x,w)}.$$

The PGAFT model (1) can be extended similarly to (16). Such an extension will correspond to the MPH model (7).

5 Conclusion

This paper proposes several new nonparametric panel transformation and multiple spell duration models. These models are identified using a new method of nonparametric identification, which can be seen as a nonparametric generalization of the quasi–differencing idea. This new method allows identification of nonparametric panel data models with time-varying structural functions and nonseparable unobserved heterogeneity. The obtained characterization of the structural functions naturally suggests a way of estimating the model.

An important area of application of these results is duration analysis with multiple spells. This paper introduces the Panel Generalized Accelerated Failure Time model and establishes its nonparametric identification. In addition, several new Mixed Proportional Hazard models are introduced. In contrast to the existing literature, these models allow the unobserved heterogeneity to enter the hazard rate non-multiplicatively and nonseparably.

This paper also opens the possibility of analysis of multiple state duration models when the covariates and the unobserved heterogeneity are dependent.

6 Appendix

In all proofs C is a generic positive constant that may vary between uses. The following lemma is used to prove Theorem 1.

Lemma 1. *Suppose U, A, and B are scalar random variables that satisfy: (a) $U \perp (A,B)$, (b) the characteristic function ϕ_U of U is non-zero on a dense set in $\mathbb{R}$. Then $A \perp B$ if and only if $A+U \perp B$.*

Proof of Lemma 1. When $A+U \perp B$,

$$\phi_{A+U,B}(s_1,s_B) = \phi_{A+U}(s_1)\,\phi_B(s_B) = \phi_A(s_1)\,\phi_U(s_1)\,\phi_B(s_B),$$

where the last equality follows from (a). At the same time, using (a),

$$\phi_{A+U,B}(s_1,s_B) = \phi_{A,B,U}(s_1,s_B,s_1) = \phi_{A,B}(s_1,s_B)\,\phi_U(s_1).$$

Equating the two expressions for $\phi_{A+U,B}(s_1,s_B)$ we obtain

$$\phi_A(s_1)\,\phi_U(s_1)\,\phi_B(s_B) = \phi_{A,B}(s_1,s_B)\,\phi_U(s_1).$$

The characteristic functions $\phi_A(s_1)$ and $\phi_B(s_B)$ are bounded and continuous. Hence, using (b) we obtain $\phi_A(s_1)\,\phi_B(s_B) = \phi_{A,B}(s_1,s_B)$ for all $(s_1,s_B) \in \mathbb{R}^2$, which implies that $A \perp B$.

Proof in the other direction is obvious. ■

Proof of Theorem 1. 1. As explained in the main text, conditioning on $\mathcal{G}(x_1)$ will be made implicit and X_j will be omitted in the notation. When the covariates X_j are discrete this causes no concerns. However, when X_j contains continuously distributed components, the conditioning event $\mathcal{G}(x_1)$ has zero probability and care must be taken when dealing with the expectations, conditional on this event. It can be shown that Assumption 2(vii) is sufficient to guarantee that the manipulations below are valid conditional on $\mathcal{G}(x_1)$. Instead of the event $\mathcal{G}(x_1)$, one can consider a sequence of events $\mathcal{G}^n(x_1) = \left\{X \in \left(x_1 - n^{-1}, x_1 + n^{-1}\right) \times \left(\overline{x} - n^{-1}, \overline{x} + n^{-1}\right) \times \left(\widetilde{x}_3(x_1) - n^{-1}, \widetilde{x}_3(x_1) + n^{-1}\right)\right\}$. These events have a positive probability for any finite n; hence, one can perform all of the analyses below, conditional on $\mathcal{G}^n(x_1)$. As n tends to infinity, the continuity Assumption 2(vii) can be used to establish the results we obtain below. Since this argument is relatively standard while the arguments below are not, for clarity, the proof below proceeds conditioning implicitly on the event $\mathcal{G}(x_1)$.[34] The same comment applies to all the other proofs below. The reader is encouraged to think of the case of discrete covariates X_j (and hence of $\Pr\{\mathcal{G}(x_1)\} > 0$) at first reading.

2. The main text shows that the condition (8) implies

$$\varkappa(A,U_1,U_2) \perp A. \tag{17}$$

[34] For an example of a proof handling such conditioning on a probability zero event, see Section 6.3 in Evdokimov (2008).

For all $r > 0$ define $B_r(\alpha_0) \equiv (\alpha_0 - r, \alpha_0 + r)$Then, for any $r > 0$ and any $\epsilon > 0$ we have

$$P[|\varkappa(A, U_1, U_2)| > \epsilon] = P[|\varkappa(A, U_1, U_2)| > \epsilon | A \in B_r(\alpha_0)].$$

We are going to use the above equality to show that $P[|\varkappa(A, U_1, U_2)| > 0] = 0$. Note that for all $\epsilon > 0$ there exists $M(\epsilon) > 0$ such that for all $r > 0$

$$P[|U_1 + \alpha_0| > M(\epsilon) \text{ or } |U_2 + \alpha_0| > M(\epsilon) | A \in B_r(\alpha_0)] < \epsilon,$$

because the distribution of (U_1, U_2) does not depend on A by Assumption 2(ii).

Define the set $\mathcal{V}_{M(\epsilon)} = [-M(\epsilon) - \varepsilon_0, M(\epsilon) + \varepsilon_0]$, where ε_0 is defined in Assumption 2(vi). For $j = 1, 2$, function $\overline{G}_j(v) : \mathcal{V}_{M(\epsilon)} \to \mathbb{R}$ is measurable; hence, by Luzin's theorem, for any $\epsilon > 0$ there is a measurable compact set $\mathcal{V}^{\epsilon}_{j,M(\epsilon)} \subset \mathcal{V}_{M(\epsilon)}$ such that restriction of the function $\overline{G}_j$ to $\mathcal{V}^{\epsilon}_{j,M(\epsilon)}$, $\overline{G}_j(v) : \mathcal{V}^{\epsilon}_{j,M(\epsilon)} \to \mathbb{R}$, is continuous and $Leb\left(\mathcal{V}_{M(\epsilon)} \backslash \mathcal{V}^{\epsilon}_{j,M(\epsilon)}\right) < \epsilon$.[35] Define the set $\mathcal{V}^{\epsilon}_{M(\epsilon)} = \mathcal{V}^{\epsilon}_{1,M(\epsilon)} \cap \mathcal{V}^{\epsilon}_{2,M(\epsilon)}$ and note that $Leb\left(\mathcal{V}_{M(\epsilon)} \backslash \mathcal{V}^{\epsilon}_{M(\epsilon)}\right) < 2\epsilon$. Moreover, by Assumption 2(iv), there is a function $\rho(\epsilon) : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{\epsilon \searrow 0} \rho(\epsilon) = 0$ such that for all $\alpha \in B_{\varepsilon_0}(\alpha_0)$ and $j = 1, 2$

$$\begin{aligned} P\left[U_j + \alpha \in \mathcal{V}^{\epsilon}_{M(\epsilon)}\right] &\geq P\left[U_j + \alpha \in \mathcal{V}_{M(\epsilon)}\right] - \rho(\epsilon) \\ &= P[|U_j + \alpha_0 + (\alpha - \alpha_0)| < M(\epsilon) + \varepsilon_0] - \rho(\epsilon) \\ &\geq P[|U_j + \alpha_0| < M(\epsilon)] - \rho(\epsilon) \\ &\geq 1 - \epsilon - \rho(\epsilon), \end{aligned} \tag{18}$$

where the first equality and the last inequality follow by the definitions of $\mathcal{V}_{M(\epsilon)}$ and $M(\epsilon)$, respectively. Define the set

$$\Xi^{\epsilon} \equiv \left\{(\alpha, u_1, u_2) : \alpha \in [-\varepsilon_0/2, \varepsilon_0/2], u_1 + \alpha \in \mathcal{V}^{\epsilon}_{M(\epsilon)}, u_2 + \alpha \in \mathcal{V}^{\epsilon}_{M(\epsilon)}\right\},$$

and note that for any $\delta \in (0, \varepsilon_0/2)$

$$\begin{aligned} & P[(A, U_1, U_2) \in \Xi^{\epsilon} | A \in B_{\delta}(\alpha_0)] \\ = \; & E\left[P\left[U_1 + A \in \mathcal{V}^{\epsilon}_{M(\epsilon)}, U_2 + A \in \mathcal{V}^{\epsilon}_{M(\epsilon)} \middle| A\right] \middle| A \in B_{\delta}(\alpha_0)\right] \\ = \; & E\left[P\left[U_1 \in \mathcal{V}^{\epsilon}_{M(\epsilon)} - A\right] P\left[U_2 \in \mathcal{V}^{\epsilon}_{M(\epsilon)} - A\right] \middle| A \in B_{\delta}(\alpha_0)\right] \\ \geq \; & 1 - C(\epsilon + \rho(\epsilon)), \end{aligned} \tag{19}$$

where the first equality follows by the law of iterated expectations, the second equality follows from Assumption 2(ii), and the inequality follows from (18).

The function $\varkappa(\alpha, u_1, u_2)$ is uniformly continuous on Ξ^{ϵ} and $\varkappa(\alpha_0, u_1, u_2)$ for all (u_1, u_2) by definition. Thus, there is a $\delta(\epsilon) > 0$, such that $|\varkappa(\alpha, u_1, u_2)| < \epsilon$ for any such point $(\alpha, u_1, u_2) \in \Xi^{\epsilon}$ that there is a point $(\alpha_0, \widetilde{u}_1, \widetilde{u}_2) \in \Xi^{\epsilon}$ with $\|(\alpha - \alpha_0, u_1 - \widetilde{u}_2, u_2 - \widetilde{u}_2)\| < 2\delta(\epsilon)$. By construction of Ξ^{ϵ}, if $(\alpha, u_1, u_2) \in \Xi^{\epsilon}$ then $(\alpha_0, u_1 + \alpha - \alpha_0, u_2 + \alpha - \alpha_0) \in \Xi^{\epsilon}$. Therefore, $|\varkappa(\alpha, u_1, u_2)| < \epsilon$ for all $(\alpha, u_1, u_2) \in \Xi^{\epsilon} \cap \left(B_{\delta(\epsilon)}(\alpha_0) \times \mathbb{R} \times \mathbb{R}\right)$.

[35] For any set $\mathcal{S}$, let $Leb(\mathcal{S})$ denote its Lebesgue measure.

To shorten the notation, define the event $\mathcal{X}(\epsilon) = \{|\varkappa(A, U_1, U_2)| > \epsilon\}$. Combining the above

$$\begin{aligned} & P[|\varkappa(A, U_1, U_2)| > \epsilon] \\ = \ & P\big[|\varkappa(A, U_1, U_2)| > \epsilon | A \in B_{\delta(\epsilon)}(\alpha_0)\big] \\ \leq \ & P\big[|\varkappa(A, U_1, U_2)| > \epsilon | A \in B_{\delta(\epsilon)}(\alpha_0), (A, U_1, U_2) \in \Xi^{\epsilon}\big] P\big[(A, U_1, U_2) \in \Xi^{\epsilon} | A \in B_{\delta(\epsilon)}(\alpha_0)\big] \\ & + P\big[(A, U_1, U_2) \notin \Xi^{\epsilon} | A \in B_{\delta(\epsilon)}(\alpha_0)\big] \\ \leq \ & 0 + C(\epsilon + \rho(\epsilon)), \end{aligned}$$

where the equality follows from (17), the first inequality is by the law of total probability, and the second inequality follows from the properties of Ξ^{ϵ}, definition of $\delta(\epsilon)$, and (19). Since $\epsilon > 0$ can be taken to be arbitrarily small, we conclude that $P[|\varkappa(A, U_1, U_2)| > 0] = 0$.

The main technical difficulties in the above proof arise because functions $\overline{G}_j(\cdot)$ in general are not continuous and may have an infinite number of discontinuities. However, Luzin's theorem establishes that functions $\overline{G}_j(\cdot)$ behave almost like continuous functions, because they are measurable.

3. Thus, we proved that $P[\varkappa(A, U_1, U_2) = 0] = 1$, which implies that

$$\varkappa(\alpha, u_1, u_2) = 0$$

for (Lebesgue) almost all $(\alpha, u_1, u_2) \in B_{\varepsilon_0}(\alpha_0) \times \mathbb{R} \times \mathbb{R}$. Rewrite this as

$$\overline{G}_1(\alpha + u_1) - \overline{G}_1(\alpha_0 + u_1) = \overline{G}_2(\alpha + u_2) - \overline{G}_2(\alpha_0 + u_2).$$

The left-hand side of the equation does not depend on u_1, while the right-hand side of the equation does not depend on u_2. This implies that for $j = 1, 2$

$$\overline{G}_j(\alpha + u) - \overline{G}_j(\alpha_0 + u) = c(\alpha) \text{ for almost all } (\alpha, u) \in B_{\varepsilon_0}(\alpha_0) \times \mathbb{R}, \tag{20}$$

where $c(\alpha)$ is a measurable function that depends only on α, but does not depend on u_1, u_2, or u.

4. Consider α and $\widetilde{\alpha}$, such that $\alpha \in B_{\varepsilon_0}(\alpha_0)$, $\widetilde{\alpha} \in B_{\varepsilon_0}(\alpha_0)$, $|\widetilde{\alpha} - \alpha| < \varepsilon_0$. Then, for almost all such α, $\widetilde{\alpha}$, and almost all $u_1 \in \mathbb{R}$ the following chain of equalities holds

$$\begin{aligned} c(\alpha) &= \overline{G}_1(\alpha + u_1) - \overline{G}_1(\alpha_0 + u_1) \\ &= \overline{G}_1(\alpha + u_1) - \overline{G}_1(\widetilde{\alpha} + u_1) + \overline{G}_1(\widetilde{\alpha} + u_1) - \overline{G}_1(\alpha_0 + u_1) \\ &= \overline{G}_1(\alpha + u_1) - \overline{G}_1(\alpha_0 + (u_1 + \widetilde{\alpha} - \alpha_0)) + c(\widetilde{\alpha}) \\ &= \overline{G}_1(\alpha + \alpha_0 - \widetilde{\alpha} + \widetilde{u}_1) - \overline{G}_1(\alpha_0 + \widetilde{u}_1) + c(\widetilde{\alpha}) \\ &= c(\alpha + \alpha_0 - \widetilde{\alpha}) + c(\widetilde{\alpha}), \end{aligned} \tag{21}$$

where $\widetilde{u}_1 = u_1 + \widetilde{\alpha} - \alpha_0$, and the third and the fifth equalities follow from the definition of $c(\alpha)$. Define function $\eta(\cdot) = c(\cdot + \alpha_0)$ and note that $c(\cdot)$ and $\eta(\cdot)$ are measurable. Then (21) can be written as

$$\begin{aligned}\eta(\alpha-\alpha_0) &= \eta(\alpha-\widetilde{\alpha})+\eta(\widetilde{\alpha}-\alpha_0) \text{ or, equivalently,}\\ \eta(\xi_a+\xi_b) &= \eta(\xi_a)+\eta(\xi_b),\end{aligned} \tag{22}$$

where $\xi_a = \alpha - \widetilde{\alpha}$ and $\xi_b = \widetilde{\alpha} - \alpha_0$. Equality (22) holds for almost all ξ_a and ξ_b such that $\max\{|\xi_a|, |\xi_b|, |\xi_a + \xi_b|\} < \varepsilon_0$.

Equation (22) is the Cauchy's functional equation. Its only solution in the class of measurable functions is $\eta(\xi) = \overline{c}\xi$ for some constant $\overline{c}$. Thus, we proved that $c(\alpha) = \overline{c}(\alpha - \alpha_0)$ for almost all α. Hence, for $j = 1, 2$, almost all $\alpha \in B_{\varepsilon_0}(\alpha_0)$ and almost all $u_j \in \mathbb{R}$:

$$\overline{G}_j(\alpha + u_j) - \overline{G}_j(\alpha_0 + u_j) = \overline{c}(\alpha - \alpha_0)$$

This implies that the function $\overline{G}_j(\cdot)$ is linear everywhere, except possibly a set of measure zero. For any $j \in \{1, 2\}$ take any $v_a \in \mathbb{R}$ and $v_b \in \mathbb{R}$. Without loss of generality assume that $v_a < v_b$ and take a positive integer k such that $(v_b - v_a)/k < \varepsilon_0$. Then

$$\begin{aligned}\overline{G}_j(v_b) - \overline{G}_j(v_a) &= \sum_{j=1}^{k}\left(\overline{G}_j\left(\frac{j}{k}(v_b - v_a) + v_a\right) - \overline{G}_j\left(\frac{j-1}{k}(v_b - v_a) + v_a\right)\right)\\ &= \sum_{j=1}^{k}\overline{c}\frac{v_b - v_a}{k} = \overline{c}(v_b - v_a),\end{aligned}$$

where the second equality holds for almost all v_a and v_b. Thus, $\overline{G}_j(v) = c_{0j} + \overline{c}v$ for almost all points $v \in \mathbb{R}$. Note that $\overline{c} > 0$ since $\overline{G}_2(v)$ is strictly increasing.

5. Finally, one obtains $g_j(v) = G_j^{-1}(c_{0j} + \overline{c}v)$ for almost all v. Now, we use the normalizations imposed in Assumption 1 to determine the constants c_{0j} and $\overline{c}$. Note that $E\left[\overline{G}_1(A + U_1) - \overline{G}_2(A + U_2)\right] = E[c_{01} - c_{02} + \overline{c}(U_1 - U_2)] = c_{01} - c_{02}$, hence $c_{01} = c_{02} + E\left[\overline{G}_1(Y_1) - \overline{G}_2(Y_2)\right]$. In addition, $0 = g_2(0) = G_2^{-1}(c_{02})$ and $1 = g_2(1) = G_2^{-1}(c_{02} + \overline{c})$; hence, $c_{02} = G_2(0)$ and $\overline{c} = G_2(1) - G_2(0)$, which concludes the proof. ■

Example 1 (of identification failure with discretely distributed A). Consider model (2) without covariates. Suppose $g_j(v) \equiv v$, U_j have nondegenerate normal distributions, A is a Bernoulli random variable with parameter $p = 1/2$, and (U_1, U_2, U_3, A) are mutually independent. We are going to see that in this case there are strictly increasing nonlinear functions $G_j(\cdot)$ that satisfy the identification restriction (8) (or (6), since there are no covariates). The existence of such functions means that the independence condition (8) fails to identify the true structural functions in the model (2) when A has discrete distribution.

Consider functions $G_j(y) = 2\pi y + \sin(2\pi y)$ for $j = 1, 2$. These functions are strictly increasing.

Moreover, these functions satisfy the independence restriction (6) since in this case

$$\begin{aligned}G_1(Y_1) - G_2(Y_1) &= G_1(A+U_1) - G_2(A+U_1) \\ &= 2\pi(U_1 - U_2) + \sin(2\pi A + 2\pi U_1) - \sin(2\pi A + 2\pi U_2) \\ &= 2\pi(U_1 - U_2) + \sin(2\pi U_1) - \sin(2\pi U_2),\end{aligned}$$

where the last equality follows from sine being a periodic function with period 2π and A taking values 0 and 1 only. Thus $G_1(Y_1) - G_2(Y_1)$ is independent of A and hence these functions $G_j(\cdot)$ satisfy the independence condition (6). However, $G_j(\cdot)$ are not linear, while $g_j^{-1}(y) = y$ are. This example demonstrates that Assumption 2(vi) is necessary for the independence condition (8) to be able to identify the functions $g_j(\cdot)$. ■

Proof of Corollary 2. As explained in the main text, the conditional distribution of the vector $(A+U_1, A+U_2)'$ given the event $\mathcal{G}$ is identified. Then, a Kotlarski-type lemma identifies the distributions the conditional distributions of A, U_1, and U_2. Here I use the extensions of Kotlarski's result by Evdokimov and White (2010). Conditions (i-iii) and (iv,a) (or (iv,b)) imply that Lemma 1 (or Lemma 2) of Evdokimov and White (2010) applies and identifies A, U_1, and U_2. ■

Proof of Theorem 3. It is straightforward to check that taking

$$\begin{aligned}\overline{F}_{U_j|X,U,A}(u|(x_1,x_2,x_3),(u_1,u_2,u_3),\alpha) &= \overline{F}_{U_j|X_j}(u_j|x_j) = \exp\left\{-e^{r_j(x_j)(u_j-\bar{\gamma})}/r_j(x_j)\right\}, \\ \Lambda_j(t,x_j) &= \frac{1}{r_j(x_j)}\ln\left(r_j(x_j)\int_0^t h_j(\xi,x_j)\,d\xi\right),\end{aligned}$$

equation (15) implies the hazard rate

$$\widetilde{\theta}_{T_j}(t|x_j,\alpha) = h_j(t,x_j)\left[\exp(-\alpha-\bar{\gamma})\right]^{r_j(x_j)},$$

and redefining the unobserved heterogeneity $\exp\left(-\widetilde{A}-\bar{\gamma}\right) \mapsto A$ we obtain (12). Here $E[U_j|X_j] = 0$ and $\bar{\gamma}$ denotes the Euler–Mascheroni constant.

Assumption 2 holds by the definitions of $\Lambda_j(\cdot)$ and $\overline{F}_{U_j|X,A}(\cdot)$, and by Assumptions 3 and 4(i-iii,v). Thus, Theorem 1 identifies functions $\Lambda_j(\cdot)$ and $\overline{F}_{U_j|X,A}(\cdot)$, and hence functions $h_j(\cdot)$ and $r_j(\cdot)$, up to normalizations, which are provided by Assumption 4(iv). ■

Proof of Theorem 4. **1.** For any $x \in \mathcal{X}$, condition on the event $\mathcal{G}(x,x) = \{X_1 = X_2 = x, X_3 = x_3(x,x)\}$, which is possible due to Assumption 6(vii). Then, the model becomes

$$Y_j = \Lambda_j^{-1}(X_j, \theta + U_j),$$

where $\theta \equiv m(x,A)$, and Theorem 1 identifies the functions $\Lambda_j(y,x)$ using Assumption 5(i). In addition, for all $x \in \mathcal{X}$, $u \in \mathbb{R}$, and $j = 1,2$, Corollary 2 identifies the conditional densities

$$f_{U_j|X_1,X_2,X_3}(u|x,x,x_3(x,x)) = f_{U_j|X_j}(u|x),$$

where the equality follows from Assumption 6(ii).

2. Consider just the first two time periods. Denote $\widetilde{Y}_j = \Lambda_j(Y_j, X_j)$ and note that the distribution of $\widetilde{Y}_j$ is identified. Now we follow Evdokimov (2008); repeatedly using Assumption 6(ii) we obtain

$$\begin{aligned}
\phi_{\widetilde{Y}_1|X_1,X_2}(s|x,\overline{x}) &= E\left[\exp\{\mathrm{i}s\widetilde{Y}_1\}|(X_1,X_2)=(x,\overline{x})\right] \\
&= E\left[\exp\{\mathrm{i}s(m(x,A)+U_1)\}|(X_1,X_2)=(x,\overline{x})\right] \\
&= \phi_{m(x,A)|X_1,X_2}(s|x,\overline{x})\,\phi_{U_1|X_1,X_2}(s|x,\overline{x}) \\
&= \phi_{m(x,A)|X_1,X_2}(s|x,\overline{x})\,\phi_{U_1|X_1}(s|x).
\end{aligned}$$

Conditioning on the event $(X_1, X_2) = (x, \overline{x})$ is possible by Assumption 6(vii). Using Assumption 6(v) we can rewrite the above as

$$\phi_{m(x,A)|X_1,X_2}(s|x,\overline{x}) = \frac{\phi_{\widetilde{Y}_1|X_1,X_2}(s|x,\overline{x})}{\phi_{U_1|X_1}(s|x)},$$

for all points s except the set $\Xi_1(x) = \left\{s \in \mathbb{R} : \phi_{U_1|X_1}(s|x) = 0\right\}$, which contains at most a countable number of elements due to Assumption 6(v). Since the right-hand side is identified, we identify the conditional characteristic function $\phi_{m(x,A)|X_1,X_2}(s|x,\overline{x})$ of $m(x,A)$, given the event $\{(X_1,X_2)=(x,\overline{x})\}$, for all $s \notin \Xi_1(x)$. Moreover, characteristic functions are continuous and $\Xi_1(x)$ is at most countable, hence, by continuity $\phi_{m(x,A)|X_1,X_2}(s|x,\overline{x})$ is identified for all $s \in \mathbb{R}$. In exactly the same way we identify the conditional distribution of $m(\overline{x}, A)$ from

$$\phi_{m(\overline{x},A)|X_1,X_2}(s|x,\overline{x}) = \frac{\phi_{\widetilde{Y}_2|X_1,X_2}(s|x,\overline{x})}{\phi_{U_2|X_2}(s|\overline{x})}.$$

3. Having identified the conditional characteristic functions of $m(x,A)$ and $m(\overline{x},A)$ we also identify the corresponding conditional distributions and conditional quantiles. Notice that

$$\begin{aligned}
Q_{m(x,A)|X_1,X_2}(q|x,\overline{x}) &= m\left(x, Q_{A|X_1,X_2}(q|x,\overline{x})\right), \\
Q_{m(\overline{x},A)|X_1,X_2}(q|x,\overline{x}) &= Q_{A|X_1,X_2}(q|x,\overline{x}),
\end{aligned}$$

where the first equality follows from the monotonicity Assumption 6(i) and the second equality follows from the normalization Assumption 5(ii). Thus, the function $m(x,\alpha)$ is identified, since

$$m(x,\alpha) = Q_{m(x,A)|X_1,X_2}\left(F_{A|X_1,X_2}(\alpha|x,\overline{x})\,|x,\overline{x}\right).$$

The above establishes that the function $m(x,\alpha)$ is identified for all $\alpha \in S_A\{(X_1,X_2)=(x,\overline{x})\}$. Moreover, $m(x,\alpha)$ is identified for all $\alpha \in S_A\{X_1 = x\}$ when Assumption 6(ix) holds. ■

Proof of Theorem 5. Write the joint survival function of T_1 and T_2, for the duration model with the hazard function (3):

$$\overline{F}_{T_1,T_2|X_1,X_2}(t_1,t_2|x_1,x_2) = \int e^{-\int_0^{t_1} h_1(\zeta_1,x_1)\gamma(x_1,\alpha)d\zeta_1 - \int_0^{t_2} h_2(\zeta_2,x_2)\gamma(x_2,\alpha)d\zeta_2} dF_{A|X_1,X_2}(\alpha|x_1,x_2),$$

and note that the left-hand side is identified. Take any $x \in \mathcal{X}$, and consider the event

$\{X_1 = X_2 = x\}$, which is possible by Assumption 7. Similar to Honoré (1993), the ratio

$$\rho(t_1, t_2, x) = \frac{\partial \overline{F}_{T_1,T_2|X_1,X_2}(t_1, t_2|x, x) / \partial t_1}{\partial \overline{F}_{T_1,T_2|X_1,X_2}(t_1, t_2|x, x) / \partial t_2} = \frac{h_1(t_1, x)}{h_2(t_2, x)}.$$

is identified for all $(t_1, t_2) \in \mathbb{R}^+ \times \mathbb{R}^+$ and $x \in \mathcal{X}$. Thus, $h_1(t_1, x)$ is identified by using Assumption 8(v(a)) as $h_1(t_1, x) = \rho(t_1, 1, x)$ for all $t_1 \in \mathbb{R}^+$. Hence, the function $h_2(t_2, x)$ is also identified. Then, one can obtain functions $\Lambda_j(t, x)$ by integration: $\Lambda_j(t, x) = \ln\left(\int_0^t h_j(\xi, x)\, d\xi + C_j(x)\right)$ for some $C_j(x) \geq 0$ and all $t \in [0, +\infty)$, $x \in \mathcal{X}$, and $j = 1, 2$. Moreover, $C_j(x) \equiv 0$ since $\overline{F}_{T_j|X_j,A}(0|x, \alpha) = 1$ for all x and α and hence $\exp(\Lambda_j(0, x)) = 0$ should hold for all x.

As explained in the main text, we can now write the duration model in the form (1) with two time periods, where the functions $\Lambda_j(t, x)$, $j = 1, 2$ were already identified by the previous arguments, and the density of the idiosyncratic errors is taken to be

$$f_{U_j|X_j,A,X_{(-j)},U_{(-j)}}\left(u_j|x_j, \alpha, x_{(-j)}, u_{(-j)}\right) = e^u \exp(-e^u)$$

for all $\left(u_j, x_j, \alpha, x_{(-j)}, u_{(-j)}\right) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X} \times \mathbb{R}$ and $j = 1, 2$. Note that the characteristic function of the idiosyncratic errors U_j is everywhere nonvanishing. Note also that the moment Assumption 6(v) does not need to hold (and is not imposed) for the transformation model (1) corresponding to (3), because we do not use Kotlarski's lemma to identify the distribution of U_j.

Define $\widetilde{Y}_j = \Lambda_j(T_j, X_j)$ and note that for any $x \in \mathcal{X}$ the conditional characteristic function of $\widetilde{Y}_j$ can be written as

$$\begin{aligned} \phi_{\widetilde{Y}_1|X_1,X_2}(s|x, \overline{x}) &= \phi_{m(x,A)|X_1,X_2}(s|x, \overline{x})\, \phi_{U_1}(s) \text{ and} \\ \phi_{\widetilde{Y}_2|X_1,X_2}(s|x, \overline{x}) &= \phi_{m(\overline{x},A)|X_1,X_2}(s|x, \overline{x})\, \phi_{U_2}(s), \end{aligned}$$

where $\phi_{U_j}(s)$ do not depend on X and are known. Moreover, $\phi_{U_j}(s) \neq 0$ for all s. Then, we identify the conditional characteristic functions $\phi_{m(x,A)}(s|X = (x, \overline{x}))$ and $\phi_{m(\overline{x},A)}(s|X = (x, \overline{x}))$. Identification of these characteristic functions is equivalent to identification of the corresponding distributions. Therefore, we identify the distributions of $-m(x, A)$ and $-m(\overline{x}, A)$, conditional on the event $X = (x, \overline{x})$. Note that the function $-m(x, \alpha) \equiv \ln(\gamma(x, \alpha))$ is strictly increasing in α for all x due to Assumption 8(i), and that $\alpha \in (0, \infty)$ due to Assumptions 8(i,v(b)). Then, for all $\alpha \in (0, \infty)$ we obtain

$$\begin{aligned} & \exp\left\{Q_{-m(x,A)|X_1,X_2}\left(F_{-m(\overline{x},A)|X_1,X_2}(\ln\{\alpha\}|x, \overline{x})|x, \overline{x}\right)\right\} \\ = \;& Q_{\exp(-m(x,A))|X_1,X_2}\left(F_{\exp(-m(\overline{x},A))|X_1,X_2}(\alpha|x, \overline{x})|x, \overline{x}\right) \\ = \;& Q_{\gamma(x,A)|X_1,X_2}\left(F_{\gamma(\overline{x},A)|X_1,X_2}(\alpha|x, \overline{x})|x, \overline{x}\right) \\ = \;& \gamma\left(x, Q_{A|X_1,X_2}\left(F_{A|X_1,X_2}(\alpha|x, \overline{x})|x, \overline{x}\right)\right) \\ = \;& \gamma(x, \alpha), \end{aligned}$$

where the first equality follows by the properties of quantiles and CDFs, the second equality follows from $\gamma(x, \alpha) = \exp(-m(x, \alpha))$, the third equality follows by the property of quantiles and the normalization of Assumption 8(v(b)), and the last equality follows by Assumption 8(iii). When Assumption 8(iv) holds, function $\gamma(x, \alpha)$ is identified for all x and all $\alpha \in S_A\{X = (x, \overline{x})\} = S_A\{X_1 = x\}$, which concludes the proof. ■

References

Abbring, J. H., and G. J. V. den Berg (2003): "The Nonparametric Identification of Treatment Effects in Duration Models," *Econometrica*, 71(5), 1491–1517.

Abbring, J. H., and G. Ridder (2010): "A note on the non-parametric identification of generalized accelerated failure-time models," Mimeo, CentER, Department of Econometrics and OR, Tilburg University, Tilburg.

Abrevaya, J. (1999): "Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable," *Journal of Econometrics*, 93(2), 203–228.

Allenby, G. M., R. P. Leone, and L. Jen (1999): "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of the American Statistical Association*, 94(446), pp. 365–374.

Altonji, J. G., and R. L. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73(4), 1053–1102.

Arellano, M., and S. Bonhomme (2009): "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," Working Paper, CEMFI.

Arellano, M., and J. Hahn (2006): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," in *Advances in Economics and Econometrics, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge University Press.

Arellano, M., and B. Honoré (2001): "Panel data models: some recent developments," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5 of *Handbook of Econometrics*, chap. 53, pp. 3229–3296. Elsevier.

Bester, A., and C. Hansen (2007): "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," Manuscript, University of Chicago.

Blundell, R., R. Griffith, and F. Windmeijer (2002): "Individual effects and dynamics in count data models," *Journal of Econometrics*, 108(1), 113 – 131.

Bonhomme, S. (2010): "Functional Differencing," Working Paper, CEMFI.

Bonnal, L., D. Fougere, and A. Serandon (1997): "Evaluating the Impact of French Employment Policies on Individual Labour Market Histories," *Review of Economic Studies*, 64(4), 683–713.

Box, G. E. P., and D. R. Cox (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211–252.

Chamberlain, G. (1982): "Multivariate regression models for panel data," *Journal of Econometrics*, 18(1), 5–46.

——— (1984): "Panel Data," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2 of *Handbook of Econometrics*, chap. 22, pp. 1247–1318. Elsevier Science.

——— (1985): "Heterogeneity, omitted variables, and duration dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J. Heckman, and B. Singer. Cambridge University Press.

——— (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60(3), 567–596.

Chen, X., and D. Pouzo (2008): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments," Cowles Foundation Discussion Papers 1650, Cowles Foundation for Research in Economics, Yale University.

Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2010): "Average and Quantile Effects in Nonseparable Panel Models," Working Paper, MIT.

Chiappori, P.-A., and I. Komunjer (2008): "Correct Specification and Identification of Nonparametric Transformation Models," Discussion paper, Boston College Department of Economics.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

d'Haultfoeuille, X. (forthcoming): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*.

Evdokimov, K. (2008): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Yale University.

Evdokimov, K., and H. White (2010): "An Extension of a Lemma of Kotlarski," Working Paper, Princeton University.

Flinn, C. J., and J. J. Heckman (1982): "Models for the Analysis of Labor Force Dynamics," NBER Working Papers 0857, National Bureau of Economic Research, Inc.

Flinn, C. J., and J. J. Heckman (1983): "Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States?," *Journal of Labor Economics*, 1(1), 28–42.

Folland, G. B. (1999): *Real Analysis: Modern Techniques and Their Applications*. Wiley, 2nd edn.

Gonul, F., and K. Srinivasan (1993): "Consumer Purchase Behavior in a Frequently Bought Product Category: Estimation Issues and Managerial Insights from a Hazard Function Model with Heterogeneity," *Journal of the American Statistical Association*, 88(424), pp. 1219–1227.

Graham, B., and J. Powell (2008): "Identification and Estimation of Irregular Correlated Random Coefficient Models," Working Paper 14469, National Bureau of Economic Research.

Guo, G., and G. Rodriguez (1992): "Estimating a Multivariate Proportional Hazards Model for Clustered Data Using the EM Algorithm, with an Application to Child Survival in Guatemala," *Journal of the American Statistical Association*, 87(420), pp. 969–976.

Han, A. K. (1987): "A non-parametric analysis of transformations," *Journal of Econometrics*, 35(2-3), 191–209.

Heckman, J. J., and G. J. Borjas (1980): "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence," *Economica*, 47(187), 247–83.

Heckman, J. J., V. J. Hotz, and J. R. Walker (1985): "New Evidence on the Timing and Spacing of Births," *American Economic Review*, 75(2), 179–84.

Hoderlein, S., and H. White (2009): "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," Working Paper, Brown University.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56(6), 1371–95.

Honoré, B., and A. de Paula (2010): "Interdependent Durations," *Review of Economic Studies*, 77(3), 1138–1163.

Honoré, B. E. (1993): "Identification Results for Duration Models with Multiple Spells," *Review of Economic Studies*, 60(1), 241–46.

Horowitz, J. L. (1996): "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable," *Econometrica*, 64(1), 103–37.

Horowitz, J. L., and S. Lee (2004): "Semiparametric estimation of a panel data proportional hazards model with fixed effects," *Journal of Econometrics*, 119(1), 155–198.

Hougaard, P., B. Harvald, and N. V. Holm (1992): "Measuring the Similarities Between the Lifetimes of Adult Danish Twins Born Between 1881-1930," *Journal of the American Statistical Association*, 87(417), pp. 17–24.

HU, Y., AND M. SHUM (2008): "Nonparametric identification of dynamic models with unobserved state variables," CeMMAP working papers CWP13/08, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

JACHO-CHAVEZ, D., A. LEWBEL, AND O. LINTON (2006): "Identification and Nonparametric Estimation of a Transformed Additively Separable Model," Boston College Working Papers in Economics 652, Boston College Department of Economics.

KHAN, S., AND E. TAMER (2007): "Partial rank estimation of duration models with general forms of censoring," *Journal of Econometrics*, 136(1), 251–280.

KOTLARSKI, I. (1967): "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal Of Mathematics*, 20(1), 69–76.

LEE, S. (2008): "Estimating Panel Data Duration Models With Censored Data," *Econometric Theory*, 24(05), 1254–1276.

LIN, D., W. SUN, AND Z. YING (1999): "Nonparametric estimation of the gap time distribution for serial events with censored data," *Biometrika*, 86(1), 59–70.

MATTNER, L. (1993): "Some Incomplete But Boundedly Complete Location Families," *Annals of Statistics*, 21(04), 2158–2162.

MATZKIN, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375.

NEWMAN, J. L., AND C. E. MCCULLOCH (1984): "A Hazard Rate Approach to the Timing of Births," *Econometrica*, 52(4), pp. 939–961.

PALEY, R., AND N. WIENER (1934): *Fourier Transforms in the Complex Domain.* Colloq. Publ. Amer. Math. Soc.

PORTER, J. (1996): "Essays in Econometrics," Ph.D. thesis, MIT.

RIDDER, G. (1990): "The Non-parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57(2), 167–81.

RIDDER, G., AND I. TUNALI (1999): "Stratified partial likelihood estimation," *Journal of Econometrics*, 92(2), 193–232.

VAN DEN BERG, G. J. (2001): "Duration models: specification, identification and multiple durations," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5 of *Handbook of Econometrics*, chap. 55, pp. 3381–3460. Elsevier.

VISSER, M. (1996): "Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS," *Biometrika*, 83(3), 507–518.

Wang, W., and M. T. Wells (1998): "Nonparametric Estimation of Successive Duration Times Under Dependent Censoring," *Biometrika*, 85(3), pp. 561–572.

Wei, L. J., D. Y. Lin, and L. Weissfeld (1989): "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions," *Journal of the American Statistical Association*, 84(408), pp. 1065–1073.

Wooldridge, J. M. (1997): "Multiplicative Panel Data Models Without the Strict Exogeneity Assumption," *Econometric Theory*, 13(05), 667–678.

Woutersen, T. (2000): "Estimators for Panel Duration Data with Endogenous Censoring and Endogenous Regressors," Econometric Society World Congress 2000 Contributed Papers 1581, Econometric Society.