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

Kirill Evdokimov†
Department of Economics
Princeton University

August 12, 2010

Abstract

This paper develops a nonparametric generalization of the quasi-differencing method of linear panel data models. A nonparametric panel data model is shown to be identified using three time periods of data. The fixed effects and idiosyncratic errors are not additively separable from the covariates, and hence affect the marginal effects. In contrast to the existing literature the structural function is allowed to vary over time in an arbitrary fashion.

The paper also obtains nonparametric identification results for a nonparametric panel transformation and a multiple spell duration models. The later result substantially extends Honore's (1993, *Review of Economic Studies*) result by relaxing the assumption of multiplicative separability of the unobserved heterogeneity in the specification of the duration function.

*This paper is a revision of my paper "Further Results on Identification of Panel Data Models with Unobserved Heterogeneity" and the second chapter of my dissertation. I am very grateful to Joseph Altonji, Donald Andrews, Xiaohong Chen, James Heckman, Ilze Kalnina, Taisuke Otsu, Peter Phillips, Edward Vytlacil and especially Yuichi Kitamura for their insightful comments and suggestions. I also thank the participants of North American Summer Meeting of the Econometric Society, SITE Summer Workshop on Advances in Nonparametric Econometrics, Stats in the Chateau Summer School, Cowles Summer Conference 2010, and econometric seminars at Yale for their comments. All errors are mine. Financial support of Cowles Foundation via Carl Arvid Anderson Prize is gratefully acknowledged.

†E-mail: kevdokim@princeton.edu.

1 Introduction

This paper considers the problem of nonparametric identification of panel data models with unobserved heterogeneity. A new method of nonparametric identification is proposed, which can be seen as a nonparametric generalization of the quasi-differencing approach to panel data models. In addition, a new identification result for multiple spell duration models is obtained.

Consider the following panel transformation model with unobserved heterogeneity:

$$\Lambda_t(Y_{it}, X_{it}) = m(X_{it}, \alpha_i) + U_{it}, \quad i = 1, \ldots, n, \;\; t = 1, \ldots, T \geq 3, \tag{T}$$

where $Y_{it} \in \mathbb{R}$ and $X_{it} \in \mathcal{X} \subset \mathbb{R}^p$ are observed outcome and covariates. Scalar random variables α_i and U_{it} are unobserved and represent, respectively, individual-specific heterogeneity and idiosyncratic disturbance. The transformation function Λ_t 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. The functions $\Lambda_t(\cdot)$ and $m(\cdot)$ are unknown and are modelled nonparametrically. It is assumed that functions $\Lambda_t(\cdot)$ and $m(\cdot)$ are strictly increasing in Y_{it} and α_i, respectively. The unobserved heterogeneity α_i is allowed to be correlated with the covariates, but is assumed to be independent of the disturbances U_{it}, conditional on the covariates. The disturbance terms U_{it} are assumed to be independent across time, conditional on the covariates and to satisfy the standard mean restriction $E[U_{it}|X_{it}] = 0$. The goal is to identify $\Lambda_t(\cdot)$, $m(\cdot)$, and the (conditional) distributions of α_i and U_{it}.

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

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

$$Y_{it} = g_t(X_{it}, \alpha_i + U_{it}), \tag{P}$$

which is a nonparametric panel data model with nonseparable unobserved heterogeneity. In this model, the derivative $\partial g_t(x, \nu) / \partial x$ depends on ν; thus the model allows for heterogeneous marginal effects that depend on $\nu_{it} = \alpha_i + U_{it}$. The structural function $g_t(\cdot)$ is allowed to vary over time in an arbitrary way. Note that the shocks v_{it} are correlated across time through α_i.[1]

Second, the transformation model (T) can be used for duration analysis (see Horowitz, 1996, for a list of applications of transformation models). Assume that the disturbances U_{it} are

[1] Note that a special case of model (P) is a transformation model extension of the nonparametric panel model of Porter (1996)

$$\lambda_t(Y_{it}) = \varphi_t(X_{it}) + \alpha_i + U_{it},$$

where $\varphi_t(\cdot)$ is an unknown regression function, and $\lambda_t(\cdot)$ is a strictily increasing unknown trasformation function. This model corresponds to specifying $g_t(x, v) = \lambda_t^{-1}(\varphi_t(x) + v)$ in (P).

independent of the covariates and have cumulative distribution function (CDF) $F_{U_{it}|X_{it}}(u|x) = 1 - \exp(-e^u)$. Then, model (T) can be interpreted as a mixed proportional hazard (MPH) model where $\exp(\Lambda_t(y))$ is the integrated baseline hazard, $Y_{it} \geq 0$ is the length of spell t of the individual i, X_{it} is the vector of covariates (that are constant across period of observation), and α_i is the unobserved heterogeneity that may correlate with the covariates X_{it}. The hazard rate takes the form

$$\theta_t(y, x, \alpha) = h_t(y, x)\,\gamma(x, \alpha), \tag{D}$$

where $h_t(y,x) \equiv (\partial \Lambda_t(y,x)/\partial y)\exp(\Lambda_t(y,x))$ and $\gamma(x,\alpha) \equiv \exp(-m(x,\alpha))$.[2] Honoré (1993) considers identification of a multiple spell duration model where its hazard rate has a more restrictive form: $\widetilde{\theta}_t(y, x, \widetilde{\alpha}) = h_t(y,x)\,\widetilde{\alpha}$, i.e. it assumes that the unobserved heterogeneity α enters the specification of the hazard rate multiplicatively. Thus, the specification (D) generalizes Honoré's model. The non-multiplicative effect of α may be important for economic modelling. 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 α_i (e.g., skill). However, it is not possible to model the dependence of the CDF on α_i using a simple multiplicative specification, since the range of the CDF must be $[0,1]$. In contrast, hazard rate specification (D) allows modelling the unknown CDF that depends on the unobserved heterogeneity α_i.

Third, the model (T) itself can be seen as a generalization of the model in Evdokimov (2008). That paper studies the model

$$Y_{it} = m(X_{it}, \alpha_i) + U_{it}, \tag{1}$$

which corresponds to $\Lambda_t(y,x) \equiv y$ in model (T). The general model (T) relaxes the assumption of additive separability of the idiosyncratic disturbance U_{it} in model (1).

This paper demonstrates that model (T) can be identified nonparametrically using a panel data with three or more time periods ($T \geq 3$). The identification method developed in this paper can be seen as a nonparametric generalization of quasi-differencing (see Chamberlain, 1984). The main idea of the identification strategy is easy to demonstrate using model (P). Consider first a linear panel model without covariates $Y_{it} = \gamma_t \cdot (\alpha_i + U_{it})$, where $\gamma_t \neq 0$ are *scalars*. Then, the following moment restriction identifies the ratio γ_1/γ_2:

$$E\left[\left(\gamma_1^{-1} Y_{i1} - \gamma_2^{-1} Y_{i2}\right) Y_{i3}\right] = 0. \tag{2}$$

Now consider a nonlinear panel model $Y_{it} = g_t(\alpha_i + U_{it})$ without covariates, where $g_t(\cdot) : \mathbb{R} \to \mathbb{R}$ are *strictly increasing functions*. Suppose some strictly increasing functions $G_t(\cdot)$,

[2] One needs to impose some normalizations to uniquely identify functions $h_t(y,x)$ and $\gamma(x,\alpha)$. See Section 4 for details.

$t \in \{1, 2\}$ satisfy the following independence condition

$$G_1(Y_{i1}) - G_2(Y_{i2}) \perp Y_{i3}. \tag{3}$$

Then, as proved in Section 3 below, functions $G_t(\cdot)$ must be equal to the inverse functions $g_t^{-1}(\cdot)$ up to a location and scale normalization. Note that condition (2) is a mean independence quasi-difference restriction. Condition (3) can be seen as a stronger version of (2), since it requires full independence. Strengthening the condition of quasi-differencing to full independence thus allows nonparametric identification of the strictly increasing function $g_t(\cdot)$. Model (T) with covariates can be identified by the same independence restriction (3), which holds conditional on the values of covariates X_i.

Let us conclude the introduction with a discussion on 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), Horowitz (1996), and Jacho-Chavez, Lewbel, and Linton (2006). Chiappori and Komunjer (2008) rely on 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) lists many examples of applications of multiple spell duration models. Horowitz and Lee (2004) and Lee (2008) provide estimation procedures for a semiparametric version of Honoré's (1993) panel data duration model. Also, Abrevaya (1999) estimates linear index coefficients in a general semiparametric fixed-effects panel transformation model.

The literature on nonparametric analysis of panel data is growing very rapidly. Recent contributions include Arellano and Bonhomme (2008), Altonji and Matzkin (2005), Bester and Hansen (2007), Chernozhukov, Fernandez-Val, Hahn, and Newey (2008), Chernozhukov, Fernandez-Val, and Newey (2009), Cunha, Heckman, and Schennach (2007), Evdokimov (2008), Graham and Powell (2008), Graham, Hahn, and Powell (2009), and Hoderlein and White (2009).

Cunha, Heckman, and Schennach (2007) focus on linear models, though they also present an interesting and potentially powerful high-level completeness condition for nonparametric identification in panel data models. Implications of their result to the identification of the model in equation (P) deserve further exploration. It appears, however, that verifying their completeness conditions for the model (P) may require very strong moment assumptions; see d'Haultfoeuille (2006). The method of this paper only requires bounded second moments.

The rest of the paper is organized as follows. For expositional clarity, nonparametric identification of model (P) is established first. Then, identification of the the transformation model (T) and of the duration model (D) are considered in Sections 3 and 4, respectively. Section 5 conlcludes. Technical proofs are collected in the Appendix.

2 Identification of Panel Model (P)

Model (P) requires some normalization in order to identify its structural elements. A location and a scale normalizations are necessary since the mean of α_i and variance of α_i and U_{it} are not restricted. The following assumption is a convenient normalization:

Assumption 1. *For some $\overline{x} \in \mathcal{X}$ and a constant $y_0 > 0$, assume that $g_2(\overline{x}, 0) = 0$ and $g_2(\overline{x}, 1) = y_0$.*

Other normalizations are possible, for instance one can assume that $g_2(\overline{x}, 0) = 0$ and $\partial g_2(\overline{x}, 0)/\partial v = 1$. Identification of the model does not require that the covariates X_{it} have common support across t. Moreover, their supports do not even need to overlap. Define the set $\mathcal{X}_1 = \{x_1 \in \mathcal{X} : f_{X_{i1},X_{i2}}(x_1, \overline{x}) > 0\}$, i.e. the set $\mathcal{X}_1$ is the support of X_{i1} given the event that $X_{i2} = \overline{x}$ (the value of $\overline{x}$ should be chosen appropriately, so that the set $\mathcal{X}_1$ is not empty). Here and everywhere below for discrete components of X_{it} the density is taken with respect to the counting measure. For each $x_1 \in \mathcal{X}_1$ define $x_3(x_1)$ to be any $x_3 \in \mathcal{X}$ such that $f_{X_{i1},X_{i2},X_{i3}}(x_1, \overline{x}, x_3(x_1)) > 0$ (such a point x_3 always exists due to the definition of the set $\mathcal{X}_1$). Denote $X_i = (X_{i1}, X_{i2}, X_{i3})$ and $U_i = (U_{i1}, U_{i2}, U_{i3})$. Define the event $\mathcal{G}(x_1) = \{X_i = (x_1, \overline{x}, x_3(x_1))\}$.

Assumption 2. *$T = 3$ and $\{Y_i, X_i, U_i, \alpha_i\}$ is a random sample. Also:*

(i) *$g_t(x, v) : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ are Borel measurable functions for all $t \in \{1, 2, 3\}$;*

(ii) *$g_t(x, v)$ are strictly increasing in v for all $x \in \mathcal{X}$ and $t \in \{1, 2, 3\}$;*

(iii) *U_{i1}, U_{i2}, U_{i3}, and α_i are mutually independent, conditional on the event $\mathcal{G}(x_1)$ for all $x_1 \in \mathcal{X}_1$;*

(iv) *$E[U_{it}|\mathcal{G}(x_1)] = 0$ for all $x_1 \in \mathcal{X}_1$ and all $t \in \{1, 2\}$;*

(v) *$E[|\alpha_i| + |U_{i1}| + |U_{i2}| + |U_{i3}| \, |\mathcal{G}(x_1)] < \infty$ for all $x_1 \in \mathcal{X}_1$;*

(vi) *the conditional distribution of U_{it} is absolutely continuous with respect to Lebesgue measure, and such that $f_{U_{it}}(u|\mathcal{G}(x_1)) > 0$ for all $u \in \mathbb{R}$, $x_1 \in \mathcal{X}_1$, and $t \in \{1, 2\}$;*

(vii) *for all $x_1 \in \mathcal{X}_1$ the set of points $\left\{s \in \mathbb{R} : \phi_{U_{i3}}(s|\mathcal{G}(x_1)) \neq 0\right\}$ is everywhere dense;*[3]

[3] $\phi_{U_{it}}(s|\mathcal{G}(x_1))$ is the conditional characteristic function U_{it}, given the event $\mathcal{G}(x_1)$; it is defined as $\phi_{U_{it}}(s|\mathcal{G}(x_1)) = E[\exp(isU_{it})|\mathcal{G}(x_1)]$, where $i = \sqrt{-1}$.

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

Assumption 2(ii) is important, since it ensures invertibility of function $g_t(x_t, v)$ in the second argument. The independence Assumption 2(iii) is strong; however, some independence assumptions are usually needed for nonparametric identification of unknown functions with nonseparable unobservables. Assumption 2(iv) is standard. Assumption 2(v) is weak and ensures continuity of the conditional characteristic functions of α_i and U_{it}. Full support Assumption 2(vi) is imposed to simplify the presentation of results, but is not essential for the identification strategy; see Remark 2 below. Assumption 2(vii) is technical and is very weak; all standard distributions satisfy this assumption. Assumption 2(viii) implies that the conditional distribution of the unobserved heterogeneity α_i is continuous at least in some small neighborhood. This assumption is necessary; when α_i has discrete distribution the model is not identified. Note that the researcher does not need to know the value of α_0 for identification or estimation.

Theorem 1. *Suppose Assumptions 1 and 2 hold. Suppose some Borel measurable functions* $G_1(x_1, y)$ *and* $G_2(\overline{x}, y)$ *(i) are strictly increasing in* y *for all* $y \in S_{Y_{it}}(\mathcal{G}(x_1))$ *and for all* $x_1 \in \mathcal{X}_1$*, where* $S_{Y_{it}}(\mathcal{G}(x_1))$ *is the conditional support of* Y_{it}*, given* $\mathcal{G}(x_1)$*; (ii) satisfy* $E\left[|G_1(x_1, Y_{i1})|^2 + |G_2(\overline{x}, Y_{i2})|^2 |\mathcal{G}(x_1)\right] < \infty$ *for all* $x_1 \in \mathcal{X}_1$*; and (iii) for all* $x_1 \in \mathcal{X}_1$ *satisfy the condition*

$$G_1(x_1, Y_{i1}) - G_2(\overline{x}, Y_{i2}) \perp Y_{i3}|\mathcal{G}(x_1). \tag{4}$$

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*)*

$$\begin{aligned} g_1(x_1, v) &= G_1^{-1}(x_1, [G_2(\overline{x}, y_0) - G_2(\overline{x}, 0)]\, v + G_2(\overline{x}, 0) + \Delta_{G,1}(x_1)) \ \text{and} \\ g_2(\overline{x}, v) &= G_2^{-1}(\overline{x}, [G_2(\overline{x}, y_0) - G_2(\overline{x}, 0)]\, v + G_2(\overline{x}, 0)), \end{aligned}$$

where $\Delta_{G,1}(x_1) = E[G_1(x_1, Y_{i1}) - G_2(\overline{x}, Y_{i2})|\mathcal{G}(x_1)]$*.*

The above theorem provides a constructive method of identification for the function $g_1(x_1, v)$ (and for the function $g_2(x_2, v)$ at the point $x_2 = \overline{x}$), by presenting the conditional independence restriction (4), which identifies the inverse (in the second argument) functions of $g_1(x_1, v)$ (and $g_2(\overline{x}, v)$). In particular, conditional independence restriction can be readily transformed into a set of conditional moment equalities for the purpose of estimation. Functions $g_2(x, v)$ and $g_3(x, v)$ can be identified by switching the roles of Y_{i1}, Y_{i2}, and Y_{i3}. No further normalization assumptions beyond Assumption 1 are needed.

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 conditioning on $\mathcal{G}(x_1)$ as well as the arguments

x_1, x, and x_3 of functions G_t and g_t. Define function $\overline{G}_t(\cdot) = G_t(g_t(\cdot))$ and note that the following statements are equivalent:

$$\begin{aligned} G_1(Y_{i1}) - G_2(Y_{i2}) &\perp Y_{i3} \iff \\ \overline{G}_1(\alpha_i + U_{i1}) - \overline{G}_2(\alpha_i + U_{i2}) &\perp g_3(\alpha_i + U_{i3}) \iff \\ \overline{G}_1(\alpha_i + U_{i1}) - \overline{G}_2(\alpha_i + U_{i2}) &\perp \alpha_i + U_{i3} \iff \\ \overline{G}_1(\alpha_i + U_{i1}) - \overline{G}_2(\alpha_i + U_{i2}) &\perp \alpha_i, \end{aligned}$$

where the second line follows by the definition of function $\overline{G}_t(\cdot)$ and equation (P), the third line follows from Assumption 2(i)-(ii), and the fourth line follows from Lemma 1 in the Appendix and Assumption 2(iii), (v), and (vii).

Define the function

$$\varkappa(\alpha, u_1, u_2) = \overline{G}_1(\alpha + u_1) - \overline{G}_2(\alpha + u_2) - \left[\overline{G}_1(\alpha_0 + u_1) - \overline{G}_2(\alpha_0 + u_2)\right], \tag{5}$$

where α_0 is taken from Assumption 2(viii). The above chain of equivalent conditions together with Assumption 2(iii) implies that $\varkappa(\alpha_i, U_{i1}, U_{i2}) \perp \alpha_i$. Note also that $\varkappa(\alpha_0, u_1, u_2) = 0$ for all $(u_1, u_2) \in \mathbb{R} \times \mathbb{R}$.

Then, the proof of the theorem in the Appendix demonstrates that $\varkappa(\alpha, u_1, u_2) = 0$ for (Lebesgue) almost all points $(\alpha, u_1, u_2) \in B_{\varepsilon_0}(\alpha_0) \times \mathbb{R} \times \mathbb{R}$. This result follows from the following equalities:

$$E\left[\varkappa^2(\alpha_i, U_{i1}, U_{i2})\right] = \lim_{r \searrow 0} E\left[\varkappa^2(\alpha_i, U_{i1}, U_{i2}) \,|\alpha_i \in B_r(\alpha_0)\right] = 0,$$

where the first equality holds because of the independence between $\varkappa(\alpha_i, U_{i1}, U_{i2})$ and α_i, while the second equality is shown to follow from $\varkappa(\alpha_0, u_1, u_2) \equiv 0$. The above equation implies that $\varkappa(\alpha, u_1, u_2) = 0$ for almost all $(\alpha, u_1, u_2) \in B_{\varepsilon_0}(\alpha_0) \times \mathbb{R} \times \mathbb{R}$.

Moreover, since u_1 and u_2 vary independently, one obtains that for almost all $(\alpha, u_t) \in B_{\varepsilon_0}(\alpha_0) \times \mathbb{R}$ and $t \in \{1, 2\}$, the equality $\overline{G}_t(\alpha + u_t) - \overline{G}_t(\alpha_0 + u_t) = c(\alpha)$ holds for some function $c(\cdot)$ that does not depend on u_1 or u_2. The proof demonstrates that the function $\eta(\xi) \equiv c(\xi + \alpha_0)$ satisfies Cauchy's functional equation. Since the function $\eta(\xi)$ is measurable it must be linear. This, in turn, implies that at almost all points v the function $\overline{G}_t(v)$ is linear, i.e. $\overline{G}_t(v) = c_{0t} + \overline{c}v$. Then, from the definition of the function $\overline{G}_t(v)$ it follows that $g_t(v) = G_t^{-1}(c_{0t} + \overline{c}v)$ and Assumptions 1 and 2(iv) can be used to determine the constants c_{0t} and $\overline{c}$.

Remark 1. *Note that Theorem 1 imposes mild smoothness restrictions on the function* $g_t(x, v)$*. In particular, no differentiability assumptions are made.*

Remark 2. *Full support Assumption 2(vi) is only made for simplicity of notation. It is easy to relax it. For instance, let* $S_{\alpha_i}(\mathcal{G}(x_1)) \subset \mathbb{R}$ *be an open set such that the derivative* $\partial F_{\alpha_i}(\alpha|\mathcal{G}(x_1))/\partial\alpha$ *exists and is positive for all* $\alpha \in S_{\alpha_i}(\mathcal{G}(x_1))$*. Also, denote* $S_{U_{i1}}(x_1)$ *to be the conditional support of* U_{i1}*, given* $X_{i1} = x_1$*. Then, following the proof of Theorem 1, it is easy to show that* $g_1(x_1, v)$ *is identified for all* $v \in \{\alpha + u_1 : \alpha \in S_{\alpha_i}(\mathcal{G}(x_1))$ *and* $u_1 \in S_{U_{i1}}(x_1)\}$.

Remark 3. *The identification strategy relies heavily on that* α_i *having a continuous distribution. When* α_i *has degenerate distribution (i.e.* $P\{\alpha_i = const\} = 1$*) the identification method fails, because the independence condition 4 holds for all functions* $G_1(\cdot)$ *and* $G_2(\cdot)$*. However, this is not a problem. When one finds that (in the population)* Y_{i1}*,* Y_{i2}*, and* Y_{i3} *are conditionally independent one should conclude that there is no endogeneity problem. In that case the analysis can be performed for each time period separately and the one period model becomes the model of Matzkin (2003).*

One can also identify the conditional distribution of α_i and U_{it}. Take any $(x_1, x_2) \in \mathcal{X}^2$ and denote the event $\mathcal{G} = \{(X_{i1}, X_{i2}) = (x_1, x_2)\}$ to shorten the notation.

Corollary 2. *Suppose that: (i) functions* $g_1(x_1, v)$ *and* $g_2(x_2, v)$ *are identified for all* $v \in \mathbb{R}$*; (ii)* $\phi_{U_{it}}(s|\mathcal{G}) \neq 0$ *for all* $s \in \mathbb{R}$*; (iii)* α_i*,* U_{i1}*, and* U_{i2} *are mutually independent, conditional on* $\mathcal{G}$*; (iv)* $E[|\alpha_i| + |U_{i1}| + |U_{i2}| \,|\mathcal{G}] < \infty$*; and (v)* $E[U_{it}|\mathcal{G}] = 0$ *for* $t \in \{1, 2\}$*. Then, the conditional (on* $\mathcal{G}$*) distributions of* α_i*,* U_{i1}*, and* U_{i2} *are identified.*

The proof of the corollary consists of two steps. First, notice that conditional on the event $\mathcal{G}$, the joint distribution of vector $(\alpha_i + U_{i1}, \alpha_i + U_{i2})'$ is identified, since $(\alpha_i + U_{i1}, \alpha_i + U_{i2}) = \left(g_1^{-1}(x_1, Y_{i1}), g_2^{-1}(x_2, Y_{i2})\right)$, where $g_t^{-1}(x, \cdot)$ denotes the inverse function of $g_t(x, \cdot)$. Then, a lemma of Kotlarski (1967) identifies the distributions of α_i, U_{i1}, and U_{i2} using their conditional independence. The unconditional distributions of α_i, U_{i1}, and U_{i2} can be obtained by integrating the conditional distributions over x_1 and x_2 (note that the density of (X_{i1}, X_{i2}) is identified directly from data).

3 Identification of the Nonparametric Transformation Model (T)

As explained in the Introduction, model (T) can be seen as a combination of model (P) and the model (1) considered in Evdokimov (2008). This section first describes the identification method and then provides the formal result.

The identification strategy consists of two main steps. First, one considers the model conditional on the event $\{X_{i1} = X_{i2} = x\}$, which implies that $m(X_{i1}, \alpha_i) = m(X_{i2}, \alpha_i) = m(x, \alpha_i)$. Then, the logic of the analysis of model (P) applies; the conditional independence restriction

$$\widetilde{\Lambda}_1(x, Y_{i1}) - \widetilde{\Lambda}_2(x, Y_{i2}) \perp Y_{i3} | \{X_{i1} = X_{i2} = x, X_{i3} = x_3\} \quad (6)$$

holds if and only if function $\widetilde{\Lambda}_t(x, y)$ is equal to functions $\Lambda_t(x, y)$ up to a location and scale normalization. In this independence restriction it is important to condition on the event that the values of the covariates in time periods 1 and 2 are equal (i.e. on the event $\{X_{i1} = X_{i2} = x\}$); this allows $m(X_{i1}, \alpha_i)$ and $m(X_{i2}, \alpha_i)$ to cancel out when the transformation functions $\widetilde{\Lambda}_t(x, y)$ are equal to the true functions $\Lambda_t(x, y)$. Note that the above independence restriction also includes conditioning on $X_{i3} = x_3$. This is because the analysis of model (P) requires the independence between α_i and U_{it}, which is ensured by the Assumption 3(iii) below, conditional on $X_{i3} = x_3$. Thus, functions $\Lambda_t(x, y)$ can be identified by Theorem 2 applied conditional on the event $\{X_{i1} = X_{i2} = x\}$. Note that this requires the assumption that $f_{(X_{i1}, X_{i2})}(x, x) > 0$ for all $x \in \mathcal{X}$ (then one can always find $x_3 = x_3(x)$ such that $f_{X_{i1}, X_{i2}, X_{i3}}(x, x, x_3(x)) > 0$ holds).

Once functions $\Lambda_t(x, y)$ for $t \in \{1, 2\}$ are identified, one can denote $\widetilde{Y}_{it} = \Lambda_t(Y_{it}, X_{it})$ and consider the model $\widetilde{Y}_{it} = m(X_{it}, \alpha_i) + U_{it}$, $t \in \{1, 2\}$, which is the model (1), i.e. the model considered in Evdokimov (2008). Then, Assumptions 3(iii), (vii), (viii), and (x) below ensure that the identification method of that paper identifies the structural function $m(x, \alpha)$ and the conditional distributions of α_i and U_{it}. The identification strategy for model (1) consists of three steps. First, one considers the distribution of vector $(\widetilde{Y}_{i1}, \widetilde{Y}_{i2})$, conditional on the event $\{X_{i1} = X_{i2} = x\}$, and applies Kotlarski's (1967) lemma to identify the conditional distribution of U_{it}. Then, using conditional deconvolution, one obtains the conditional distributions of $m(x, \alpha_i)$ and $m(\overline{x}, \alpha_i)$ given the event $\{X_{i1} = x, X_{i2} = \overline{x}\}$. Finally, a nonparametric quantile version of within-variation identifies the structural function $m(x, \alpha)$; see Evdokimov (2008) and the proof of Theorem 4 below for details.

Therefore, the assumptions needed to identify model (T) are a combination of the assumptions 1 and 2 and the assumptions imposed in Evdokimov (2008). For all $(x_1, x_2) \in \mathcal{X}^2$, define the event $\mathcal{G}(x_1, x_2) = \{X_i = (x_1, x_2, x_3(x_1, x_2))\}$ where $x_3(x_1, x_2)$ is defined in Assumption 3(viii) below.

Assumption 3. *$T = 3$ and $\{Y_i, X_i, U_i, \alpha_i\}$ is a random sample. Also:*

(i) *functions $\Lambda_t(x, v) : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ and $m(x, \alpha) : \mathcal{X} \times \mathbb{R} \to \mathbb{R}$ are Borel measurable for all $t \in \{1, 2, 3\}$;*

(ii) *for all $x \in \mathcal{X}$ and $t \in \{1, 2, 3\}$ functions $\Lambda_t(x, v)$ and $m(x, \alpha)$ are continuous and strictly increasing in v and α, respectively;*

(iii) *$f_{U_{it}|X_{it}, \alpha_i, X_{i(-t)}, U_{i(-t)}}(u_t | x_t, \alpha, x_{(-t)}, u_{(-t)}) = f_{U_{it}|X_{it}}(u_t | x)$ for all $(u_t, x_t, \alpha, x_{(-t)}, u_{(-t)}) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X}^2 \times \mathbb{R}^2$ and $t \in \{1, 2, 3\}$;*[4]

[4] Index $(-t)$ stands for "other than t" time periods.

(iv) $E\left[U_{it}|\mathcal{G}\left(x_1,x_2\right)\right]=0$ *for all* $(x_1,x_2)\in\mathcal{X}^2$ *and all* $t\in\{1,2\}$*;*

(v) $E\left[\left|m\left(x_t,\alpha_i\right)\right|+\left|U_{it}\right|\left|\mathcal{G}\left(x_1,x_2\right)\right.\right]<\infty$ *for all* $(x_1,x_2)\in\mathcal{X}^2$ *and* $t\in\{1,2,3\}$*;*

(vi) $f_{U_{it}}\left(u|\mathcal{G}\left(x_1,x_2\right)\right)>0$ *for all* $u\in\mathbb{R}$*,* $(x_1,x_2)\in\mathcal{X}^2$*, and* $t\in\{1,2\}$*;*

(vii) $\phi_{U_{it}}\left(s|\mathcal{G}\left(x_1,x_2\right)\right)\neq 0$ *for all* $s\in\mathbb{R}$*,* $(x_1,x_2)\in\mathcal{X}^2$*, and* $t\in\{1,2,3\}$*;*

(viii) $f_{X_{i1},X_{i2}}\left(x_1,x_2\right)>0$ *for all* $(x_1,x_2)\in\mathcal{X}^2$*; also define* $x_3\left(x_1,x_2\right)$ *to be any* $x_3\in\mathcal{X}$ *such that* $f_{X_{i1},X_{i2},X_{i3}}\left(x_1,x_2,x_3\left(x_1,x_2\right)\right)>0$*;*

(ix) α_i *has a continuous distribution, conditional on the event* $\mathcal{G}\left(x_1,x_2\right)$ *for all* $(x_1,x_2)\in\mathcal{X}^2$*;*

(x) $S_{\alpha_i}\left\{(X_{i1},X_{i2})=(x,\overline{x})\right\}=S_{\alpha_i}\left\{X_{i1}=x\right\}$ *for all* $x\in\mathcal{X}$*, where* $S_{\alpha_i}\{\vartheta\}$ *is the support of* α_i*, conditional on the event* ϑ*.*

Assumption 4. *The following normalizations hold: (i) for some fixed* $\overline{y}_a$*,* $\overline{y}_b$*,* $\overline{\Lambda}_a$*,* $\overline{\Lambda}_b$*,* $\overline{y}_a<\overline{y}_b$*, and* $\overline{\Lambda}_a<\overline{\Lambda}_b$*, it holds that* $\Lambda_2\left(\overline{y}_a,x\right)=\overline{\Lambda}_a$ *and* $\Lambda_2\left(\overline{y}_b,x\right)=\overline{\Lambda}_b$ *for all* $x\in\mathcal{X}$*; (ii)* $m\left(\overline{x},\alpha\right)=\alpha$ *for all* $\alpha\in\mathbb{R}$*.*

Assumption 3 implies that Assumption 2 of the previous section and Assumptions ID and CRE(i), (iii) in Evdokimov (2008) are satisfied. Assumption 4 is a normalization and implies that the normalization Assumption 1 above and Assumption CRE(ii) in Evdokimov (2008) hold. Note that in this model one needs to impose location and scale normalizations on function $\Lambda_t\left(x,\cdot\right)$ at each point x for some t, since the distribution of U_{it} and the functional form of $m\left(x,\alpha\right)$ are not restricted. Assumptions 3(iii), (vii), (viii), and (x) are stronger than the assumptions made for identification of model (P). Assumption 3(iii) is satisfied, for example, if $U_{it}=\sigma_t\left(X_{it}\right)\xi_{it}$, where random variables ξ_{it} are i.i.d. and independent of X_i and α_i. Thus, Assumption 3(iii) in particular permits contemporaneous conditional heteroskedasticity. Assumption 3(viii) implies that $f_{(X_{i1},X_{i2})}\left(x,x\right)>0$ for all $x\in\mathcal{X}$. This is needed to ensure that one can condition on the event $\{X_{i1}=X_{i2}=x\}$. The following holds:

Theorem 3. *Suppose Assumptions 3 and 4 hold. Then functions* $\Lambda_t(x,y)$*,* $m(x,\alpha)$*,* $f_{\alpha_i|X_{it}}(\alpha|x)$ *and* $f_{U_{it}|X_{it}}(u|x)$ *in the model (T) are identified for all* $x\in\mathcal{X}$*,* $y\in\mathbb{R}$*,* $\alpha\in S_{\alpha_i}\{X_{i1}=x\}$*,* $u\in\mathbb{R}$*, and* $t\in\{1,2\}$*.*

The main steps of the proof of the theorem are described above and the proof itself is omitted.

4 Duration Model

This section explains how to combine the identification strategies of Honoré (1993) and Evdokimov (2008) to obtain identification of the MPH model (D) using data on two spells per individual.

As described in the Introduction, model (T) can be seen as a duration model with multiple spells when $Y_{it} \geq 0$. Then, Theorem 3 can be used to identify this duration model. Generally, the corresponding hazard rate does not have the mixed proportional form.

A special case of this general duration model with $F_{U_{it}|X_{it}}(u|x) = 1 - \exp(-e^u)$ yields a duration model of mixed proportional hazard type. Indeed, in this case specification (T) implies the following conditional survival function for an individual with the observed covariates $X_{it} = x$ and the unobserved heterogeneity $\alpha_i = \alpha$:

$$\begin{aligned}
\overline{F}_{Y_{it}|X_{it},\alpha_i}(y|x,\alpha) &= P(Y_{it} > y|X_{it} = x, \alpha_i = \alpha) \\
&= P(\Lambda_t(Y_{it}, x) > \Lambda_t(y, x)\,|X_{it} = x, \alpha_i = \alpha) \\
&= P(U_{it} > \Lambda_t(y, x) - m(x, \alpha)) \\
&= \exp(-\exp(\Lambda_t(y, x) - m(x, \alpha))).
\end{aligned}$$

Thus, the integrated hazard equals

$$\begin{aligned}
H_t(y|X_{it} = x, \alpha_i = \alpha) &= \int_0^y \left\{ \frac{\partial \Lambda_t(\zeta, x)}{\partial y} \exp(\Lambda_t(\zeta, x)) \right\} \exp(-m(x, \alpha))\, d\zeta \qquad (7) \\
&= \int_0^y h_t(\zeta, x)\, \gamma(x, \alpha)\, d\zeta,
\end{aligned}$$

yielding the hazard rate specification (D). Here function $\gamma(x, \alpha)$ is assumed to be strictly increasing in α. In addition, it is assumed that $h_t(y, x) > 0$ and $\gamma(x, \alpha) > 0$ for all $y > 0$, x and α.

The lengths of the spells Y_{i1} and Y_{i2} are assumed to be independent of each other, conditional on $X_i = (X_{i1}, X_{i2})$ and α_i. However, Y_{i1} and Y_{i2} may be correlated conditional on the observables X_i only, due to the correlation through α_i. It should also be noted that in this model covariates X_{it} are assumed to vary between the spells, but are constant during each spell.

The MPH duration model with the hazard rate (D) can be identified using data on just two spells per individual. In contrast, the identification results of the previous section required three periods of data. The reason is that the theorems in the previous sections make no parametric assumptions about the conditional distribution of U_{it}, while the duration model (D) implies that U_{it} have a known distribution that has particular properties.

The identification strategy for this model consists of two steps. First, identify the functions

$h_t(y,x)$ using the method of Honoré (1993) (instead of relying on the independence condition (6)). Identification of functions $h_t(y,x)$ implies identification of functions $\Lambda_t(y,x)$. Second, one can denote $\widetilde{Y}_{it} = \Lambda_t(Y_{it},x)$ and write $\widetilde{Y}_{it} = m(X_{it},\alpha_i) + U_{it}$. The logic of Evdokimov (2008) then identifies the function $\gamma(x,\alpha)$, although some changes in the proof are necessary.

These ideas are formalized in the two assumptions and the theorem below:

Assumption 5. *$T = 2$ and $\{Y_{i1}, Y_{i2}, X_i = (X_{i1}, X_{i2}), U_{i1}, U_{i2}, \alpha_i\}$ is a random sample. Also:*

(i) *functions $h_t(y,x) : \mathbb{R} \times \mathcal{X} \to \mathbb{R}$ and $\gamma(x,\alpha) : \mathcal{X} \times (0,\infty) \to (0,\infty)$ are Borel measurable for all $t \in \{1,2\}$;*

(ii) *for all $x \in \mathcal{X}$ and $t \in \{1,2\}$, functions $h_t(y,x)$ and $\gamma(x,\alpha)$ are continuous and strictly increasing in y and α, respectively;*

(iii) *$f_{U_{it}|X_{it},\alpha_i,X_{i(-t)},U_{i(-t)}}\left(u_t|x_t,\alpha,x_{(-t)},u_{(-t)}\right) = e^u \exp(-e^u)$ for all $\left(u_t,x_t,\alpha,x_{(-t)},u_{(-t)}\right) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X} \times \mathbb{R}$ and $t \in \{1,2\}$;*[5]

(iv) *$E\left[|\ln(\gamma(x_t,\alpha_i))|\,|X_i = (x_1,x_2)\right] < \infty$ for all $(x_1,x_2) \in \mathcal{X}^2$ and $t \in \{1,2\}$;*

(v) *$f_{X_{i1},X_{i2}}(x_1,x_2) > 0$ for all $(x_1,x_2) \in \mathcal{X}^2$;*

(vi) *α_i has a continuous distribution, conditional on the event $X_i = (x_1,x_2)$ for all $(x_1,x_2) \in \mathcal{X}^2$;*

(vii) *$S_{\alpha_i}\{(X_{i1},X_{i2}) = (x,\overline{x})\} = S_{\alpha_i}\{X_{i1} = x\}$ for all $x \in \mathcal{X}$, where $S_{\alpha_i}\{\vartheta\}$ is the support of α_i, conditional on the event ϑ.*

Assumption 6. *Suppose that the following normalizations hold: (i) $h_2(\overline{y},x) = \overline{h}$ for some $\overline{y} \geq 0$, $\overline{h} \geq 0$, and all $x \in \mathcal{X}$; (ii) $\gamma(\overline{x},\alpha) = \alpha$ for some $\overline{x} \in \mathcal{X}$ and for all $\alpha \in S_{\alpha_i}\{X_{i1} = x\}$.*

The distributional Assumption 5(iii) is standard for duration analysis, see for example Horowitz (1996). Assumption 5(v) can be weakened, but it is important, since it ensures within-variation that identifies the function $\gamma(x,\alpha)$. The support Assumption 5(vii) is the same as Assumption CRE(iii) in Evdokimov (2008); it ensures that the "extra" conditioning on $\overline{x}$ does not reduce the support of α_i. This is needed to use the nonparametric quantile within-variation for all $\alpha \in S_{\alpha_i}\{X_{i1} = x\}$, and not only for $\alpha \in S_{\alpha_i}\{(X_{i1},X_{i2}) = (x,\overline{x})\}$; see Evdokimov (2008) for further details and a discussion of the assumption.

Theorem 4. *Suppose Assumptions 5 and 6 hold. Then, the mixed proportional model with hazard rate (D) is identified. In particular, functions $\Lambda_t(x,y)$, $m(x,\alpha)$, $f_{\alpha_i|X_{it}}(\alpha|x)$ and $f_{U_{it}|X_{it}}(u|x)$ are identified for all $x \in \mathcal{X}$, $y \in \mathbb{R}$, $\alpha \in \mathbb{R}$, $u \in \mathbb{R}$, and $t \in \{1,2\}$.*

The proof of the theorem is given in the Appendix.

[5] Note that the characteristic function of the distribution with CDF $F(u) = 1 - \exp(-e^u)$ is everywhere nonvanishing.

5 Conclusion

This paper proposes a new identification method for nonparametric panel data models. Applications of the method include identification of a panel transformation model and a panel model with time-varying structural function and nonseparable unboserved heterogeneity. Another important application is identification of a mixed proportional hazard model. In contrast to the existing literature, the unobserved heterogeneity enters the hazard rate non-multiplicatively and nonseparably, which may be important for some economic models.

Theorems 1 and 3 provide conditional independence restrictions (4) and (6), which identify the structural functions of interest. These conditional independence restrictions can be readily written as a set of conditional moment equalities that can be used to estimate the structural functions. The estimation procedure is a topic of future research.

6 Appendix

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) $E[|A| + |B|] < \infty$, and (c) the characteristic function $\phi_U(s)$ of U, is nonvanishing on an everywhere dense set. 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 $\phi_A(s_1)$ and $\phi_B(s_B)$ are bounded and continuous due to (b). Hence, using (c) we obtain $\phi_A(s_1)\,\phi_B(s_B) = \phi_{A,B}(s_1, s_B)$ for all $(s_1, s_B) \in \mathbb{R}^2$. ■

Proof of Theorem 1. 1. As explained in the main text, the conditioning on $\mathcal{G}_x$ will be made implicit and x_t variable will be omitted in notation. Also the main text shows that condition (4) implies $\varkappa(\alpha_i, U_{i1}, U_{i2}) \perp \alpha_i$.

Then, for any $r > 0$ we obtain $E\left[\varkappa^2(\alpha_i, U_{i1}, U_{i2})\,|\alpha_i \in B_r(\alpha_0)\right] = E\left[\varkappa^2(\alpha_i, U_{i1}, U_{i2})\right] = c_\varkappa$, where the first equality follows from the independence of $\varkappa(\alpha_i, U_{i1}, U_{i2})$ and α_i.

2. We are now going to prove that $c_\varkappa = 0$. Note that for all $\epsilon > 0$ there exists $M(\epsilon) > 0$ such that for all $r > 0$

$$E\left[\varkappa^2(\alpha_i, U_{i1}, U_{i2})\, 1\{|U_{i1}| > M(\epsilon) \text{ or } |U_{i2}| > M(\epsilon)\}|\alpha_i \in B_r(\alpha_0)\right] < \epsilon,$$

which follows from the fact that the conditional expectation does not depend on r due to independence of $\varkappa_1^2(\alpha_i, U_{i1}, U_{i2})$ and α_i and from the finiteness of the expectation $E[\varkappa_1^2(\alpha_i, U_{i1}, U_{i2})]$. For any $\epsilon > 0$ and $r \in (0, \varepsilon_0)$ define the set $\Xi_{r,M(\epsilon)} \equiv B_r(\alpha_0) \times [-M(\epsilon), M(\epsilon)] \times [-M(\epsilon), M(\epsilon)]$, where ε_0 is defined in Assumption 2(viii).

The function $\varkappa(\alpha, u_1, u_2)$ is measurable, hence by Lusin's theorem, for any $\epsilon > 0$ there is a compact set $\overline{\Xi}^\epsilon_{M(\epsilon)} \subset \Xi_{\varepsilon_0, M(\epsilon)}$, such that $Leb\left(\Xi_{\varepsilon_0, M(\epsilon)} \backslash \overline{\Xi}^\epsilon_{M(\epsilon)}\right) < \epsilon$ and $\varkappa(\alpha, u_1, u_2)$ is continuous on $\overline{\Xi}^\epsilon_{M(\epsilon)}$ (and hence uniformly continuous on $\overline{\Xi}^\epsilon_{M(\epsilon)}$).[6] Thus, there is a $\delta(\epsilon)$ such that $\left|\varkappa^2(\alpha, u_1, u_2) - \varkappa^2(\widetilde{\alpha}, u_1, u_2)\right| < \epsilon$ for all α, $\widetilde{\alpha}$, u_1, and u_2, such that $|\alpha - \widetilde{\alpha}| < \delta(\epsilon)$, $(\alpha, u_1, u_2) \in \overline{\Xi}^\epsilon_{M(\epsilon)}$, and $(\widetilde{\alpha}, u_1, u_2) \in \overline{\Xi}^\epsilon_{M(\epsilon)}$. For any set $\Omega \subset B_{\varepsilon_0}(\alpha_0) \times \mathbb{R} \times \mathbb{R}$ denote the event $1_\Omega = 1\{(\alpha_i, U_{i1}, U_{i2}) \in \Omega\}$. Due to Assumptions 2(vi) and (viii) there is a constant C_1 such that for any measurable set $A \subset B_{\varepsilon_0}(\alpha_0)$

$$E\left[1_{\Xi_{\varepsilon_0, M(\epsilon)} \backslash \overline{\Xi}^\epsilon_{M(\epsilon)}} \varkappa^2(\alpha_i, U_{i1}, U_{i2})\,|\alpha_i \in A\right] < C_1 \epsilon.$$

Note that $\varkappa^2(\alpha_0, u_1, u_2) \equiv 0$. For all $r \in (0, \varepsilon_0)$ denote $\overline{\Xi}^\epsilon_{r,M(\epsilon)} = \overline{\Xi}^\epsilon_{M(\epsilon)} \cap \Xi_{r,M(\epsilon)}$. Then

$$\begin{aligned}
& c_\varkappa \\
=\ & E\left[\varkappa^2(\alpha_i, U_{i1}, U_{i2})\,|\alpha_i \in B_{\delta(\epsilon)}(\alpha_0)\right] \\
=\ & E\left[1_{\left(B_{\delta(\epsilon)}(\alpha_0) \times \mathbb{R} \times \mathbb{R}\right)} \varkappa^2(\alpha_i, U_{i1}, U_{i2})\,|\alpha_i \in B_{\delta(\epsilon)}(\alpha_0)\right] \\
=\ & E\left[\left(1_{\left(B_{\delta(\epsilon)}(\alpha_0) \times \mathbb{R} \times \mathbb{R}\right) \backslash \Xi_{\delta(\epsilon), M(\epsilon)}} + 1_{\Xi_{\delta(\epsilon), M(\epsilon)} \backslash \overline{\Xi}^\epsilon_{\delta, M(\epsilon)}} + 1_{\overline{\Xi}^\epsilon_{\delta, M(\epsilon)}}\right) \varkappa^2(\alpha_i, U_{i1}, U_{i2}) \Big| \alpha_i \in B_{\delta(\epsilon)}(\alpha_0)\right] \\
\leq\ & (1 + C_1)\epsilon + E\left[1_{\overline{\Xi}^\epsilon_{\delta, M(\epsilon)}} \left|\varkappa^2(\alpha_i, U_{i1}, U_{i2}) - \varkappa^2(\alpha_0, U_{i1}, U_{i2})\right|\,|\alpha_i \in B_{\delta(\epsilon)}(\alpha_0)\right] \\
\leq\ & (1 + C_1)\epsilon + E\left[\sup_{\alpha, u_1, u_2 \in \overline{\Xi}^\epsilon_{\delta, M(\epsilon)}} \left|\varkappa^2(\alpha, u_1, u_2) - \varkappa^2(\alpha_0, u_1, u_2)\right| \Bigg| \alpha_i \in B_{\delta(\epsilon)}(\alpha_0)\right] \\
\leq\ & (2 + C_1)\epsilon.
\end{aligned}$$

Since $\epsilon > 0$ can be taken arbitrarily small and $c_\varkappa \geq 0$ we conclude that $c_\varkappa = 0$.

3. Thus, we proved that $E\left[\varkappa^2(\alpha_i, U_{i1}, U_{i2})\right] = 0$ for almost all $\alpha \in B_{\varepsilon_0}(\alpha_0)$,which implies

[6] For any set A, $Leb(A)$ denotes its Lebesgue measure.

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 $t \in \{1, 2\}$

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

where $c(\alpha)$ is a measurable function that does not depend on u_1 or u_2.

4. Consider any 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

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

where $\widetilde{u}_1 = u_1 + \widetilde{\alpha} - \alpha_0$ and the third and fifth equalities follow from the definition of $c(\alpha)$. Define function $\eta(\cdot) = c(\cdot + \alpha_0)$, then one obtains

$$\begin{aligned}
\eta(\alpha - \alpha_0) &= \eta(\alpha - \widetilde{\alpha}) + \eta(\widetilde{\alpha} - \alpha_0) \text{ or, equivalently,} \\
\eta(\xi_a + \xi_b) &= \eta(\xi_a) + \eta(\xi_b), \qquad (8)
\end{aligned}$$

where $\xi_a = \alpha - \widetilde{\alpha}$ and $\xi_b = \widetilde{\alpha} - \alpha_0$. Thus, the last equality (8) holds for almost all ξ_a and ξ_b such that $\max\{|\xi_a|, |\xi_b|, |\xi_a + \xi_b|\} < \varepsilon_0$.

Equation (8) is 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, $c(\alpha) = \overline{c}(\alpha - \alpha_0)$ for almost all α. Thus, for all $t \in \{1, 2\}$, almost all $\alpha \in B_{\varepsilon_0}(\alpha_0)$ and almost all $u_t \in \mathbb{R}$:

$$\overline{G}_t(\alpha + u_t) - \overline{G}_t(\alpha_0 + u_t) = \overline{c}(\alpha - \alpha_0)$$

This implies that the function $\overline{G}_t(\cdot)$ is itself linear (at almost all points). For any $t \in \{1, 2\}$ consider arbitrary v_a and v_b. 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

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

where the second equality holds for almost all v_a and v_b. Thus, $\overline{G}_t(v) = c_{0t} + \overline{c}v$ for almost all points $v \in \mathbb{R}$. Note that $\overline{c} > 0$ since $\overline{G}_t(v)$ are strictly increasing. Note also that $E\left[\overline{G}_1(\alpha_i + U_{i1}) - \overline{G}_2(\alpha_i + U_{i2})\right] = E\left[c_{01} - c_{02} + \overline{c}(U_{i1} - U_{i2})\right] = c_{01} - c_{02}$, hence $c_{01} = c_{02} + E\left[\overline{G}_1(Y_{i1}) - \overline{G}_2(Y_{i2})\right]$.

5. Finally, one obtains $g_t(v) = G_t^{-1}(c_{0t} + \overline{c}v)$ for almost all v. Now, use normalizations imposed in Assumption 1 to determine the constants c_{0t} and $\overline{c}$: $c_{02} = \overline{G}_2(0) = G_2(0)$, $\overline{c} = \overline{G}_2(1) - \overline{G}_2(0) = G_2(y_0) - G_2(0)$, which concludes the proof. ■

Proof of Corollary 2. As explained in the main text, the conditional distribution of vector $(\alpha_i + U_{i1}, \alpha_i + U_{i2})'$ given the event $\mathcal{G}$ is identified. Then, conditions (ii)-(v) of the corollary ensure that Lemma 1 of Evdokimov (2008) applies and identifies the conditional distributions of α_i, U_{i1}, and U_{i2}. ■

Proof of Theorem 4. Write (using the notation of this paper) equation (2) of Honoré (1993), but with the hazard function specified by (D). That is, assuming $T = 2$, write the joint survival function of Y_{i1} and Y_{i2}, conditional on the values of the observed covariates, as

$$\overline{F}_{Y_{i1},Y_{i2}|X_{i1},X_{i2}}(y_1,y_2|x_1,x_2) = \int e^{-\int_0^{y_1}h_1(\zeta_1,x_1)\gamma(x_1,\alpha)d\zeta_1 - \int_0^{y_2}h_2(\zeta_2,x_2)\gamma(x_2,\alpha)d\zeta_2}dF_{\alpha_i|X_{i1},X_{i2}}(\alpha|x_1,x_2).$$

Similar to Honoré (1993), for any $x \in \mathcal{X}$ the ratio

$$\rho(y_1,y_2,x) = \frac{\partial\overline{F}(y_1,y_2|x,x)/\partial y_1}{\partial\overline{F}(y_1,y_2|x,x)/\partial y_2} = \frac{h_1(y_1,x)}{h_2(y_2,x)}.$$

is identified for all $(y_1,y_2) \in \mathbb{R}^2$ and $x \in \mathcal{X}$, since $\overline{F}(y_1,y_2|x,x)$ is identified directly from data. Thus, $h_1(y_1,x)$ is identified using Assumption 6(i) as: $h_1(y_1,x) = \overline{h}\rho(y_1,\overline{y},x)$. Hence, the function $h_2(y_2,x)$ is also identified. Then, one can obtain functions $\Lambda_t(y,x)$ via integration: $\Lambda_t(y,x) = \ln\left(\int_0^y h_t(s,x)\,ds + C_t(x)\right)$ for some $C_t(x) \geq 0$ and all $y \in [0,+\infty)$, $x \in \mathcal{X}$, and $t \in \{1,2\}$. Moreover, $C_t(x) \equiv 0$ since $\overline{F}_{Y_{it}|X_{it},\alpha_i}(0|x,\alpha) = 1$ for all x and α and hence $\exp(\Lambda_t(0,x)) = 0$ should hold for all x.

Now, define $\widetilde{Y}_{it} = \Lambda_t(Y_{it}, X_{it})$ and following Evdokimov (2008) note that for any $x \in \mathcal{X}$

the conditional characteristic function of $\widetilde{Y}_{it}$ can be written as

$$\begin{aligned}\phi_{\widetilde{Y}_{i1}}\left(s|\left(X_{i1},X_{i2}\right)=\left(x,\overline{x}\right)\right) &= \phi_{m(x,\alpha_i)}\left(s|X_i=\left(x,\overline{x}\right)\right)\phi_{U_{i1}}\left(s\right) \text{ and} \\ \phi_{\widetilde{Y}_{i2}}\left(s|\left(X_{i1},X_{i2}\right)=\left(x,\overline{x}\right)\right) &= \phi_{m(\overline{x},\alpha_i)}\left(s|X_i=\left(x,\overline{x}\right)\right)\phi_{U_{i2}}\left(s\right),\end{aligned}$$

where $\phi_{U_{it}}(s)$ do not depend on X_i and are known due to Assumption 5(iii). Moreover, $\phi_{U_{it}}(s) \neq 0$ for all s. Thus, we identify the conditional characteristic functions $\phi_{m(x,\alpha_i)}(s|X_i=(x,\overline{x}))$ and $\phi_{m(\overline{x},\alpha_i)}(s|X_i=(x,\overline{x}))$. Identification of these characteristic functions is equivalent to identification of the corresponding distributions. Therefore, we identify the distributions of $-m(x,\alpha_i)$ and $-m(\overline{x},\alpha_i)$, conditional on the event $X_i=(x,\overline{x})$. Note that function $-m(x,\alpha)$ is strictly increasing in α for all x due to Assumption 5(ii). Then, one obtains for all $a \in (0,\infty)$ that

$$\begin{aligned}\exp\left(Q_{-m(x,\alpha_i)}\left(F_{-m(\overline{x},\alpha_i)}\left(\ln\left(a\right)\right)\right)\right) &= Q_{\exp(-m(x,\alpha_i))}\left(F_{\exp(-m(\overline{x},\alpha_i))}\left(a\right)\right) \\ &= Q_{\gamma(x,\alpha_i)}\left(F_{\gamma(\overline{x},\alpha_i)}\left(a\right)\right) \\ &= \gamma\left(x, Q_{\alpha_i}\left(F_{\alpha_i}\left(a\right)\right)\right) \\ &= \gamma\left(x,a\right),\end{aligned}$$

where the first equality follows by the property of quantiles, the second equality follows by the definition of $\gamma(x,\alpha)$, the third equality follows by the property of quantiles and Assumption 6(ii), and the last equality follows by Assumption 5(vi). Thus, using Assumption 5(vii) the function $\gamma(x,\alpha)$ is identified for all x and all $\alpha \in S_{\alpha_i}\{X_i=(x,\overline{x})\} = S_{\alpha_i}\{X_{i1}=x\}$, which concludes the proof. ■

References

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

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 (2008): "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," Working Paper, CEMFI.

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

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. (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.

Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2008): "Identification and estimation of marginal effects in nonlinear panel models," CeMMAP working papers CWP25/08, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Chernozhukov, V., I. Fernandez-Val, and W. K. Newey (2009): "Quantile and Average 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 (2007): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," Discussion paper, The University of Chicago.

d'Haultfoeuille, X. (2006): "On the Completeness Condition in Nonparametric Instrumental Problems," Working Papers 2006-32, Centre de Recherche en Economie et Statistique.

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

Graham, B., J. Hahn, and J. Powell (2009): "A Quantile Panel Data Regression Model," Working Paper, UC Berkeley.

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

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

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

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.

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.

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.

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

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.

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.