On the role of time in nonseparable panel data models

Stefan Hoderlein
Yuya Sasaki

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP15/11

On the Role of Time in Nonseparable Panel Data Models

Stefan Hoderlein* Yuya Sasaki

Boston College Brown University

First: April 9, 2010 Current: March 13, 2011

Abstract

This paper contributes to the understanding of the source of identification in panel data models. Recent research has established that few time periods suffice to identify interesting structural effects in nonseparable panel data models even in the presence of complex correlated unobservables, provided these unobservables are time invariant. A communality of all of these approaches is that they point identify effects only for subpopulations. In this paper we focus on average partial derivatives and continuous explanatory variables. We elaborate on the parallel between time in panels and instrumental variables in cross sections and establish that point identification is generically only possible in specific subpopulations, for finite T. Moreover, for general subpopulations, we provide sharp bounds. Finally, we show that these bounds converge to point identification as T tends to infinity only. We systematize this behavior by comparing it to increasing the number of support points of an instrument. Finally, we apply all of these concepts to the semi-parametric panel binary choice model and establish that these issues determine the rates of convergence of estimators for the slope coefficient.

Keywords: Nonseparable Models, Identification, Panel Data, Semiparametric, Binary Choice.

* Stefan Hoderlein: Boston College, Department of Economics, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, email: stefan_hoderlein@yahoo.com. Yuya Sasaki: Brown University, Department of Economics, Box B, Providence, RI 02912, USA, email: yuya.sasaki@gmail.com. We have received helpful comments from seminar participants at Berkeley, Duke, Hohenheim, New York University, University of North Carolina, North Carolina State, University of Soutern California and Syracuse. The usual disclaimer applies.

1 Introduction

What is the role of time in panel data? The fact that certain individual traits like preferences can often assumed to be time invariant together with the fact that individuals are observed repeatedly opens up the way for a powerful identification principle, whose driving force is akin to the exogenous variation provided by instruments in a cross section. The recent literature has pointed out that only few time periods are necessary to exploit this fact and point identify interesting structural effects among subpopulations (e.g., Honore and Kyriazidou (2000), Altonji and Matzkin (2005), Arellano and Bonhomme (2010), Graham and Powell (2010), Hoderlein and White (2010), Chernozhukov, Fernandez Val, Hahn and Newey (2010, henceforth CFHN)), or partially separable structural functions (Evdokimov, 2010). The question then becomes: what do we gain from observing individuals more often across time?

This paper contributes to the understanding of the role of time in panels with continuous explanatory variables. It develops the notion that repeated observations provide exogenous variation that parallels identification through instrumental variables. From this perspective, few time periods correspond to few support points of the instrument. In cross sectional models with discrete support, point identification is available for structural effects in subpopulations (e.g., "compliers" in the LATE framework) only, and not more than bounds are available for structural effects in the entire population (Chesher (2005), Shaik and Vytlacil (2010)). This point identification for subpopulations only is paralleled by the recent results in panel data models mentioned earlier: Hoderlein and White (2010) concentrate on the effects for "stayers", i.e., the subpopulation for which $X_1 = X_2 = x$, while Arellano and Bonhomme (2010) and Graham and Powell (2010) focus on the population for which (at least) $X_1 \neq X_2$.

In this paper, we establish that these parallels are not accidental. Consider the nonseparable structural panel data model:

$$Y_t = \phi(X_t, A, U_t) \qquad t \in \mathcal{T}(T) := \{1, \ldots, T\} \tag{1.1}$$

where Y_t is an outcome, X_t is an endogenous regressor, A is an individual fixed effect and U_t is an idiosyncratic shock. As econometricians, we observe the joint distribution $F_{Y_1 \cdots Y_T X_1 \cdots X_T}$ of outcomes and explanatory variables across time. With the notation $\beta(x, a, u) = \frac{\partial}{\partial x}\phi(x, a, u)$, the object of interest for identification is the local average response (LAR)

$$\mathbb{E}\left[\beta(X_t, A, U_t)\right| X_t = x\right], \tag{1.2}$$

which is the mean partial effect among the (arbitrary) subpopulation of individuals who chose $X_t = x$ at time t. Since the marginal distribution F_{X_t} of X_t is observable, integration of this object with respect to F_{X_t} will yield the population average partial effect (APE – Chamberlain, 1984; Blundell and Powell, 2003; Wooldridge, 2005) as a byproduct. This parameter is also closely related to Graham and Powell's (2010) correlated random coefficient models, who restrict the model to be linear in X_t. In contrast to this and some of the panel data work mentioned above (e.g., Hoderlein and White (2010)), our conditioning set in (1.2) can encompass the entire population by varying x, provided individuals change their ranks in the X_t distribution over time.

Our parameter (1.2) exactly resembles that of Altonji and Matzkin (2005, AM), and is hence also related to Bester and Hansen (2008), who employ an index restriction closely related in spirit to AM's exclusion (control function) restriction[1]. While the parameter is similar to the one considered by AM, our analysis however shows some marked differences: While AM impose an exchangeability restriction to construct control variables from observed data, we avoid such an assumption at the expense of an alternative assumption that allows to make inter-temporal comparisons, and essentially amounts to a weak joint stationarity requirement on the distribution of correlated unobservables and ranks F_{X_t}. Both assumptions serve the purpose to facilitate inter-temporal comparisons across conditioning sets; since our assumption is different and in tendency weaker, in contrast to AM we obtain only bounds, as well as "irregular" identification.

Our contributions to the literature are as follows: We establish that for finite T, average structural marginal effects like (1.2) are only set identified, and we provide sharp bounds for this effects for an arbitrary subpopulation as well as the entire population. We moreover show that for finite T, point identification happens generically only on a set of zero measure,

[1]Using the notation $\mathbf{X} = (X_1, \cdots, X_T)$, Bester and Hansen (2009) focus on $\mathbb{E}\left[\frac{\partial}{\partial x_t}\mathbb{E}[Y_t \mid A, \mathbf{X}]\middle| \mathbf{X}\right]$, which is rather different from what we consider.

which explains that the results obtained in the literature focused on subpopulations. Third, we describe the process by which partial identification asymptotes to point identification of the effect of interest across the entire population as the number of time periods increases. Finally, we discuss the analogy between panel data models and IV models, and argue in which sense these two apparently different setups can be considered as a single unified paradigm.

While concentrating on a different object, this paper complements also CFHN's (2010) recent seminal developments in nonlinear nonseparable panel data models. While CFHN focus on the effect of discrete explanatory variables, this paper focuses on continuous variables. Moreover, CFHN consider the average effect of a discrete difference (e.g., a classical binary treatment effect), while we consider the LAR given in (1.2), which is more akin to an average random coefficient (see also Hoderlein and Sasaki (2010) for a detailed discussion of this parameter in the cross section case, and Graham and Powell (2010)). However, similarly in spirit to CFHN, we establish that we only obtain partial identification of the effect of interest with fixed T in general, and that the effect of interest becomes point identified as $T \to \infty$.

But do we have to wait until T becomes really infinite to obtain point-identification? While the answer is generally affirmative for finite T, we show that (1.2) is point-identified for some specific points $X_t = x^*$ even in finite T. These (special, but not necessarily large) values x^* are observable from data, hence we can tell which locality admits point identification in finite time. Moreover, the number of such points x^* exhibiting point-identification is increasing in T under certain conditions, illustrating the passage from partial to point identification in another fashion. This passage from partial to point identification as T increases resembles a similar passage observed by Chesher (2005) in the transition from discrete to continuous instrument in the context of cross section data. Indeed, we will provide an unified view on these two paradigms by representing panel data models in a form analogous to two-stage cross section models.

The aforementioned identification principle can be translated into semiparametric binary choice models as a special case. The fact that point-identification is achieved at some specific points $X_t = x^*$ even under finite T leads to identification of the coefficient in a semiparametric binary choice models based on information from a population of measure zero. On the other hand, as $T \to \infty$, we obtain point-identification of semiparametric binary choice models based on information from a population with positive measure. The former implies a nonparametric (i.e., slower than $\sqrt{n}$) rate with finite T, whereas the latter implies a parametric rate of convergence as $T \to \infty$.

This paper is organized as follows: In the following section, we provide the main identification results. We establish partial identification of (1.2) for arbitrary subpopulations, provide sharp bounds, and establish the behavior of these bounds as $T \to \infty$. Moreover, we show point identification for special subpopulations characterized by $F_{X_s}(X_s) = F_{X_t}(X_t)$. In the third subsection, we provide a numerical analysis that illustrates these points, not least the behavior as $T \to \infty$, graphically. In the fourth subsection, we discuss the relationship between panels and instruments in the general case, while we discuss extensions, as well as the leading special case, the semiparametric binary choice model, in the fifth section. Finally, an outlook concludes.

2 Main Identification Results

In this section, we discuss the main identification results. To this end, we first discuss and introduce a set of assumptions which are mostly standard for this literature. We then provide the main results on partial identification, point identification of a set of measure zero, and identification as $T \to \infty$. All of these results are followed by a discussion.

All the identification results of this paper are derived through the single device, the time-variant rank of X_t, defined by

$$V_t := F_{X_t}(X_t) \qquad t \in \mathcal{T}(T).$$

As its expression implies, it plays the role of a control variable. We start out by stating the assumptions on which our identification results are based:

Assumption 1. *Basic Restrictions:*
(ID) U_t *is identically distributed across* $t \in \mathcal{T}(T)$.
(IND) $U_s \perp\!\!\!\perp (X_t, A)$ *for all* $s, t \in \mathcal{T}(T)$.

(ACC) The distribution of X_t is absolutely continuous with a convex support.

(IJD) $(A, V_s) \overset{Law}{=} (A, V_t)$ *for all* $s, t \in \mathcal{T}(T)$.

Remarks: Assumption ID (Identical Distribution) states that idiosyncratic errors are identically distributed across time. This assumption is standard in this literature and appears in CFHN, Graham and Powell (2010), Hoderlein and White (2010), among others. The independence assumption IND (Independence) defines the period specific error to be fully independent across time, which given the nonseparable framework employed is a straightforward generalization of the notion of strict exogeneity provided in the literature. The assumption ACC (Absolute Continuity & Convex support) specifies X_t to be continuously distributed; this guarantees that the distribution function F_{X_t} is bijective on its support.

Assumption IJD (Invariant Joint Distribution) states that the joint distribution of individual fixed effects A and the rank $V_t = F_{X_t}(X_t)$ remains invariant over time. Note that both the marginals of A and V_t are stationary (the latter is always $\mathcal{U}[0,1]$), this assumption only restricts the correlation structure of the joint distribution to be time invariant. While individuals may change their ranks in the X_t distribution over time, the population covariance between fixed effects and ranks is required to be time invariant.

This is a rather weak assumption. In particular, it does *not* require rank invariance, and allows for a dynamic relationship of the rank process, e.g., the process

$$V_{t+1} = \mu_a + \rho(a)(V_t - \mu_a) + W_t \qquad W_t \mid A = a \sim NID(0, s_a^2)$$

where $|\rho(a)| < 1$ for a.e. $a \in \text{supp}(A)$ and $V_1 \mid A = a \sim N(\mu_a, \sigma_a^2)$ will satisfy assumption IJD. Note that the initial V_1 or the dynamic perturbation W_t need not be independent of unobserved fixed effect A in this AR example. More generally, any stationary process (such as the ARMA with certain coefficient restrictions) of V_t conditional on $A = a$ for a.e. a satisfies IJD.

Rank invariance across time $V_t = V_s = V$ is a special case of the assumption IJD; this special case may be plausible in certain economic applications, e.g., the ordering of various individual specific income processes may be the same across time, while the individual processes diverge. In order to further motivate the IJD assumption, we consider a structure that encompasses a wide array of economic model:

Example 1 (Economic Models Satisfying the IJD)**.** *Suppose the choice of quantity X_{it} by agent i at time t takes the form*

$$X_{it} = \psi(W_t, \iota(A_i, B_{it})) \tag{2.1}$$

for Borel-measurable functions ψ and ι, where ψ is strictly increasing in the second argument. None of A_i, B_{it}, or W_t is observable to the econometrician. If $(A_i, B_{is}) \overset{Law}{\equiv} (A_i, B_{it})$ for all s, t, then the process $\{X_{it}\}_t$ satisfies the Assumption IJD. To see this, denote $C_{it} := \iota(A_i, B_{it})$. Then, $(A_i, V_{is}) = (A_i, F_{X_{is}}(X_{is})) = (A_i, F_{C_{is}}(C_{is})) \overset{Law}{\equiv} (A_i, F_{C_{it}}(C_{it})) = (A_i, F_{X_{it}}(X_{it})) = (A_i, V_{it})$.

To give an economic example for such a choice relationship, suppose that X_{it} is chosen by agent i according to the following decision rule:

$$x_{it} = \arg\max_x \left[\int v(x, a_i, u_{it}) dF_{U_t|B_t}(u_{it} \mid b_{it}) - p(w_t) \cdot x \right] \tag{2.2}$$

where v is an utility function, $F_{U_t|B_t}$ is distribution of period specific error u_{it} given agents beliefs b_{it}, and $p(w_t)$ is a time-varying unit cost of treatment choice x_{it}. To fix ideas, think of the integral as the expected benefit of choosing x which depends on the individual "taste" fixed effect a_i, and think of $p(w_t) \cdot x$ as the costs that depend on the price $p(w_t)$ which in turn depends on the macroeconomic environment w_t. To derive an explicit solution, suppose that the objective function v takes the form $v(x, a, u) := x^{1-c}\varphi(a, u)/(1-c)$ with $c \in (0, 1)$. Then, we deduce that

$$x_{it} = p(w_t)^{-1/c} \cdot \underbrace{\left[\int \varphi(a_i, u_{it}) dF_{U_t|B_t}(u_{it} \mid b_{it}) \right]^{1/c}}_{\iota(a_i, b_{it})},$$

which satisfies the expression (2.1). □

While this example is related in spirit to one in Imbens and Newey (2009) in the cross section case, it highlights some communalities and differences between cross sections and panels: First, unlike in cross sections, $(A_i, B_{is}) \overset{Law}{\equiv} (A_i, B_{it})$ is assumed, i.e., the joint distribution of preferences and beliefs is time invariant. In contrast, the distribution of W_t is allowed to vary over time. The source of time variation W_t, e.g., the macroeconomic environment, enters the decision rule (2.2) as an exogenous "cost-shifter." Observe that in cross-section models, exogenous cost-shifters are usually instruments. This highlights a parallel between the role of time in panel data models and the role of instrumental variables in cross section models. This analogy will be more formally discussed in Section 4.

Armed with these assumptions, we can now proceed to the formal analysis. An important component in the subsequent analysis is introduced through the following short-hand notation:

$$\Delta(s,t,x) := \frac{\mathbb{E}\left[Y_s \mid X_s = F_{X_s}^{-1} \circ F_{X_t}(x)\right] - \mathbb{E}\left[Y_t \mid X_t = x\right]}{F_{X_s}^{-1} \circ F_{X_t}(x) - x} \tag{2.3}$$

Note that this object is identified from observed data. It in turn identifies a structural feature as stated in the following lemma.

Lemma 1. *Suppose that Assumption 1 holds. Then*

$$\mathbb{E}\left[\left.\frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x), A, U_t) - \phi(x, A, U_t)}{F_{X_s}^{-1} \circ F_{X_t}(x) - x}\right| X_t = x\right] = \Delta(s,t,x)$$

holds for all $x \in supp(X_t)$ for all $s, t \in \mathcal{T}(T)$ such that $F_{X_s}(x) \neq F_{X_t}(x)$.

This lemma states that the local average difference quotient (left hand side) can be identified by $\Delta(s,t,x)$, which is observable from data. This intermediate result will be used in proving all the main identification results of this paper, i.e., partial identification under $T < \infty$ (Theorem 1), local point identification under $T < \infty$ (Theorem 2), and point identification as $T \to \infty$ (Theorem 3). Due the condition $F_{X_s}(x) \neq F_{X_t}(x)$ of the lemma, we require a nontrivial time-variation in the marginal distributions of X_t as an empirically testable restriction. Note that if there is horizontal shift in cdfs over time - as would for instance be implied if there is proportional income growth - the condition $F_{X_s}(x) \neq F_{X_t}(x)$ holds across the entire population. While we think of this as rather the rule than the exception, it rules out time invariant or stationary X_t.

2.1 Partial Identification under $T < \infty$.

The first theorem of this paper shows that $\Delta(s,t,x)$ can be used to bound on the local average response (LAR) of ϕ. To this end, we define, for $x \in \text{supp}(X_t)$ and $t \in \mathcal{T}(T)$, two time indices

$$\begin{aligned}
\underline{\tau}(T,t,x) &:= \quad \arg\max_{s \in \mathcal{T}(T)} \{F_{X_s}^{-1} \circ F_{X_t}(x) \mid F_{X_s}^{-1} \circ F_{X_t}(x) < x\} \\
\overline{\tau}(T,t,x) &:= \quad \arg\min_{s \in \mathcal{T}(T)} \{F_{X_s}^{-1} \circ F_{X_t}(x) \mid F_{X_s}^{-1} \circ F_{X_t}(x) > x\}
\end{aligned}$$

These time indices specify the time s for which (2.3) forms sharp bounds of (1.2). If we were to identify an individual characterized by $X_t = x'$ with its rank $F_{X_t}(x')$ then it would provide the smallest movement across X between any two periods s, t of an individual with $X_t = x$, both to the left and to the right of x. For these quantities to be well defined there have to be enough time periods, i.e., at any given point x at least one below, and one above. This structure provides the vehicle to employ the following assumption.

Assumption 2. *(Local Curvature):*
There exists an interval I containing x, $F^{-1}_{X_{\underline{\tau}(T,t,x)}} \circ F_{X_t}(x)$, and $F^{-1}_{X_{\overline{\tau}(T,t,x)}} \circ F_{X_t}(x)$ such that the sign of $\frac{\partial^2}{\partial x^2}\phi(x', a, u)$ is invariant for all $x' \in I$ and $[P_{AU_t}]$-a.s. (a, u).

Assumption 2 requires that, given x and a time $t \in \mathcal{T}(T)$, the two "closest" time periods $\underline{\tau}(T, t, x)$ to the left and $\overline{\tau}(T, t, x)$ to the right in terms of the X-distributions are indeed close enough so that the sign of the second derivative of the structural function remains the same within this proximity. There are three special cases that are sufficient for Assumption 2 to hold on the interval I : 1. (Locally) Non-increasing returns/concavity, 2. a (local) linear structure, and, 3. (locally) non-decreasing returns/convexity. These sufficient conditions are often *globally* implied by economic theory in certain applications (e.g., in inter-temporal consumer choice, Carroll and Kimball (1996)), while we only need them to hold in the neighborhood I of smallest changes. Note moreover that it suffices if any of these conditions are satisfied locally.

The bounds which arise in this setup are analyzed in the following theorem:

Theorem 1 (Partial Identification for $T < \infty$)**.** *Let $x \in supp(X_t)$ for some $t \in \mathcal{T}(T)$. Suppose that Assumptions 1 and 2 hold in (1.1). Then we obtain that*

$$L(T, t, x) \leqslant \mathbb{E}\left[\beta(x, A, U_t)\right| X_t = x] \leqslant U(T, t, x)$$

i.e., the local average response (LAR) is set identified with sharp bounds given by

$$\begin{aligned} L(T, t, x) :=& \quad \min\left\{\Delta(\underline{\tau}(T, t, x), t, x), \Delta(\overline{\tau}(T, t, x), t, x)\right\} \qquad \textit{and} \\ U(T, t, x) :=& \quad \max\left\{\Delta(\underline{\tau}(T, t, x), t, x), \Delta(\overline{\tau}(T, t, x), t, x)\right\}. \end{aligned}$$

Figure 1 graphically illustrates the mechanism of Theorem 1. The bottom figure shows that $F_{X_2}^{-1} \circ F_{X_1}(x)$ and $F_{X_3}^{-1} \circ F_{X_1}(x)$ serve as left- and right-counterfactual choice of $X_1 = x$, respectively. Project these left- and right-counterfactual choices of x up onto the top figure to illustrate the counterfactual outcomes $\phi(F_{X_2}^{-1} \circ F_{X_1}(x), a, u)$ and $\phi(F_{X_3}^{-1} \circ F_{X_1}(x), a, u)$, respectively. Under a negative sign for Assumption 2, the left and right difference quotients of ϕ constitute upper and lower bounds of $\beta(x, a)$, respectively. Upper and lower bounds would switch under a positive sign of the second derivative.

If the sign in Assumption 2 is strictly positive or strictly negative, then the inequalities in Theorem 1 are strict, too. On the other hand, if Assumption 2 holds with the second derivative being zero on $I = \mathbb{R}$, then the inequalities in Theorem 1 hold with equalities. In particular, the following point identification result for linear models follows immediately from Lemma 1.

Corollary 1 (Linear Models)**.** *If the structural function is a linear random coefficient model,*

$$\phi(X_t, A, U_t) = \alpha(A, U_t) + \beta(A, U_t) X_t \qquad \text{for all } t \in \mathcal{T}(T),$$

with correlated random coefficients, then Theorem 1 implies point identification, i.e.,

$$\mathbb{E}\left[\beta(x, A, U_t) \middle| X_t = x\right] = \mathbb{E}\left[\beta(A, U_t) \middle| X_t = x\right] = \Delta(s, t, x)$$

for all $x \in supp(X_t)$ for all $s, t \in \mathcal{T}(T)$ such that $F_{X_s}(x) \neq F_{X_t}(x)$.

This result parallels the main identification result of Graham and Powell (2010, GP). In this special case, the LAR is an average random coefficient. An important difference is that GP's identification is based on "movers" that consist of agents with $X_t \neq X_s$ for some $s, t \in \mathcal{T}(T)$, while Corollary 1 is based on "distributional movers" in the sense of $F_{X_s}(x) \neq F_{X_t}(x)$. Frequently, average effects are of interest, in particular the APE of Chamberlain (1984). In the case of continuously distributed X_t, GP's identified parameter is the LAR conditional on pairs (X_t, X_s) with $X_t \neq X_s$. Put differently, in finite samples GP have to exclude observations for which $X_t \cong X_s$, which may be a significant part of the population. In contrast, we can in principle use the entire population, provided $F_{X_s}(x) = F_{X_t}(x)$ holds only on a set of $[P_{X_t}]$-measure zero. Conversely, it may also be the case that F_{X_t} is stationary over time while

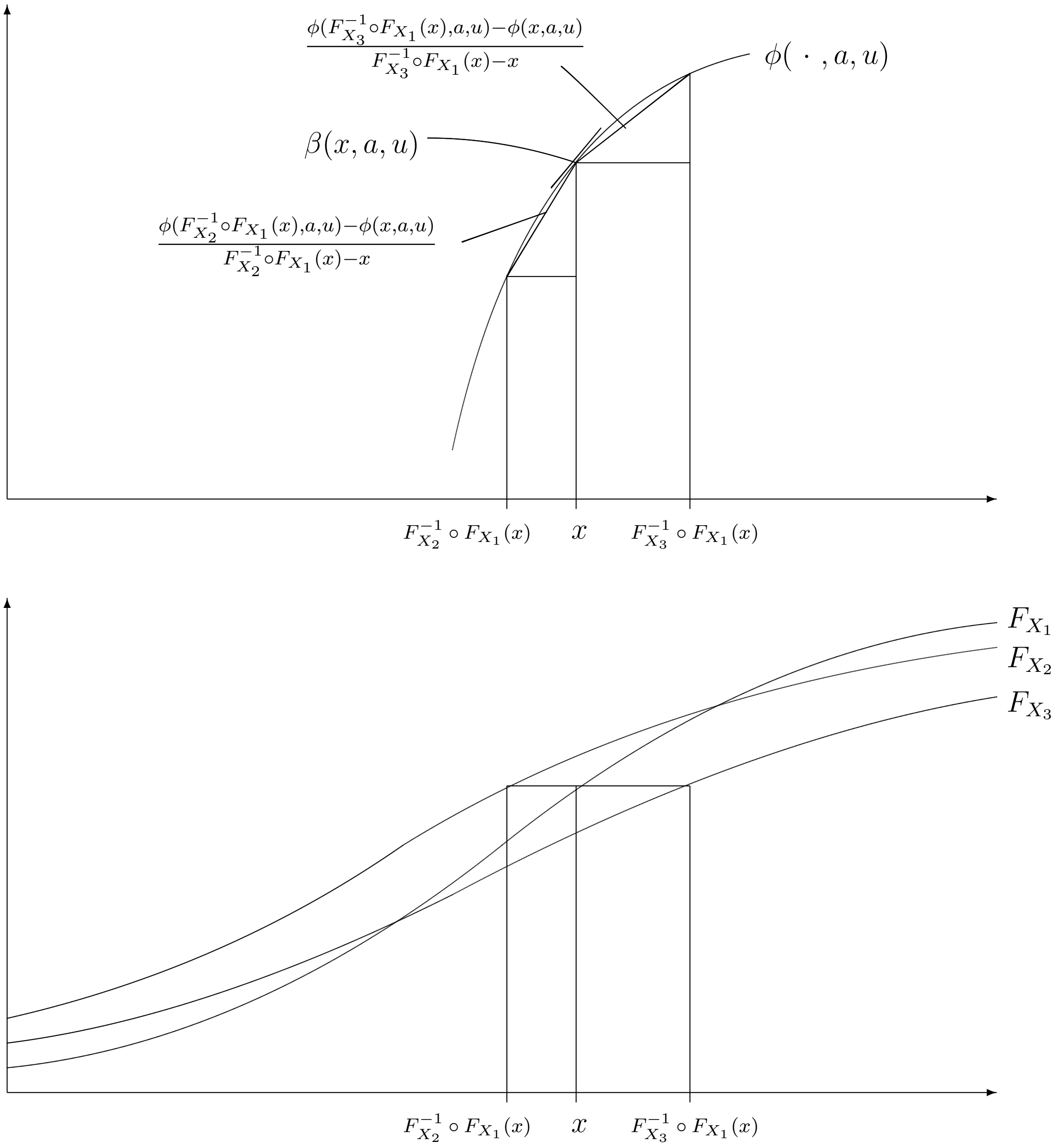

Figure 1: Partial identification under $T < \infty$: difference quotients form bounds of $\beta(x, a, u)$.

$X_{it} \neq X_{is}$ for a large set of individuals, in which case GP's analysis should be preferred. Which method is more useful in this special case hence depends on the specific data configuration, however, our method does apply to any model, not just linear ones.

2.2 Local Point Identification under $T < \infty$.

The last section shows that we can at least partially identify conditional partial effects for $T < \infty$. This section shows that point identification is achieved locally for some x, even if T is finite. However, such a set of points has measure zero. To make this point precise, we require the following assumption:

Assumption 3. *Local Counterfactual: There exist* $s, t \in \mathcal{T}(T)$ *and* x^* *such that* $F_{X_t}(x^*) = F_{X_s}(x^*)$ *and* x^* *is in the closure of the set* $\{x \mid F_{X_t}(x) \neq F_{X_s}(x)\}$*.*

This assumption states that the cdfs of X_s and X_t non-tangentially intersect at x^*. Intuitively, x^* is associated with identical rank in times s and t, but there are points x arbitrarily close to x^* at which ranks are not identical between time periods s and t. With smooth (e.g., real analytic) cdfs, a cluster point of a set of such points x^* satisfying Assumption 3 under $T < \infty$ does not exist, hence only a set of measure zero admits Assumption 3. This observation is important for understanding fundamental limitations in the identification of certain semiparametric panel data models like the fixed effects binary choice model, and the associated impact on the rate of convergence of estimators in these models, cf. Section 5.3.

Next, we invoke a set of regularity conditions:

Assumption 4. *Regularity:*

(i) $[\phi(\tilde{x}, a, u) - \phi(x, a, u)]/(\tilde{x} - x) \to \beta(x^*, a, u)$ *as* $x, \tilde{x} \to x^*$ *uniformly across* (a, u)*.*

(ii) $\beta(x, a, u)$ *is bounded uniformly across* (a, u) *and across* x *in a neighborhood of* x^**.*

(iii) For t*,* $f_{A|X_t}(a \mid x^* + \delta) \to f_{A|X_t}(a \mid x^*)$ *as* $\delta \to 0$ *uniformly across* a*.*

With these sets of assumptions in place, we are now in the position to state our second main theorem:

Theorem 2 (Local Point Identification under $T < \infty$)**.** *Suppose that Assumptions 1 and 4 hold in (1.1). Then, for $s, t \in \mathcal{T}(T)$ and x^* that satisfy Assumption 3, we have*

$$\lim_{x \to x^*} \Delta(s,t,x) = \mathbb{E}\left[\beta(x^*, A, U_t) \middle| X_t = x^*\right].$$

Figure 2 graphically illustrates the mechanism of Theorem 2. Assumption 3 is satisfied at x^* with time periods 1 and 3, as F_{X_1} and F_{X_3} intersect at x^*, but differ in its deleted neighborhood. The bottom figure depicts three elements x', x'', and x''' of a convergent sequence $x \to x^*$. The discrepancy between F_{X_1} and F_{X_3} in the deleted neighborhood of x^* (i.e., excluding x^*), as implied by Assumption 3 allows $x \neq F_{X_3}^{-1} \circ F_{X_1}(x)$ for each of $x = x'$, x'', and x''', as illustrated in the bottom figure. This in turn allows well-defined difference quotients for each of x', x'', and x''' in the top figure. These difference quotient asymptote to the partial effect at x^*.

Just as Corollary 1 paralleled the result of Graham and Powell (2010), this Theorem 2 parallels the main identification result of Hoderlein and White (2010). Their identification is based on 'stayers' that consist of agents with $X_t = X_s$ for some $s, t \in \mathcal{T}(T)$. Similarly, Theorem 2 is based on "distributional stayers" in the sense of $F_{X_s}(x^*) = F_{X_t}(x^*)$ as in Assumption 3. While the results are similar, the identified parameters are different. Our identified parameter is the LAR conditional on a single $X_t = x^*$ that allows for movers, whereas Hoderlein and White's identified parameter is the LAR for any $X_t = x$, but conditional on the subpopulation of stayers $X_t = X_s$.

Also related to our Assumption 3[2] is the identification approach of Evdokimov (2010). He exploits information about the distributions for a partially separable structure conditionally on $X_1 = X_2 = x$, i.e., the stayers. He then point identifies structural functions by the quantile representation of the identified distribution. As we do not assume his partial separability, we reach point identification only at the specific points x^* satisfying Assumption 3, the "distributional stayers."

While point identification under $T < \infty$ is certainly superior to partial identification, it is achieved only at those points x^* satisfying Assumption 3, hence not globally. An interesting

[2] Although the context is different, the idea of using the points x^* satisfying Assumption 3 is also related to the identification approach of Torgovitsky (2010).

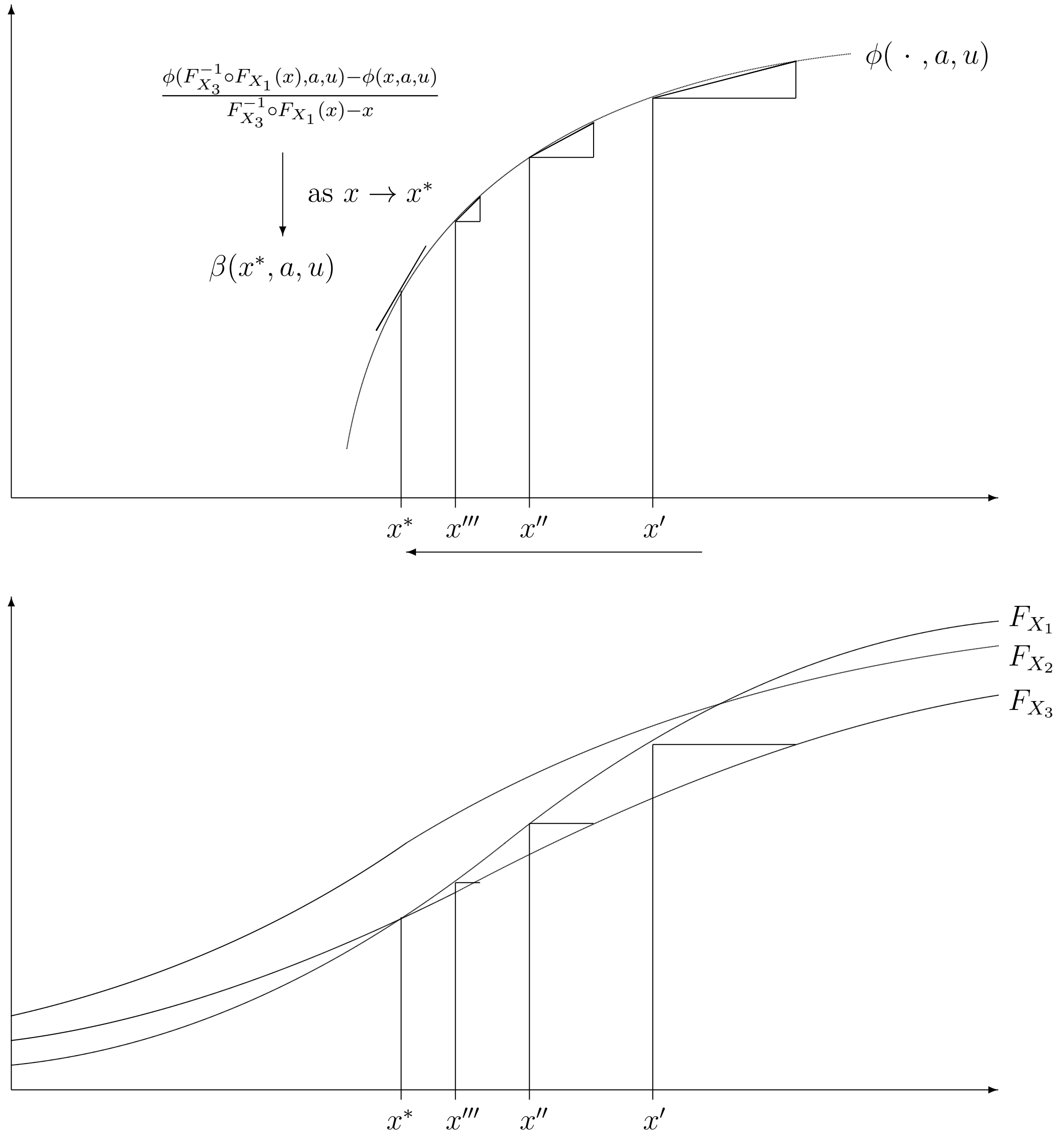

Figure 2: Local point identification under $T < \infty$: difference quotients converge to $\beta(x^*, a, u)$.

aspect is that the number of points x^* satisfying Assumption 3 is clearly increasing in T. That is, as time increases, point identification is achieved at a larger number of locations. Under some assumptions, this insight can be used to characterize the process by which partial identification 'converges' to point identification as $T \to \infty$; it is more formally studied in the following subsection.

2.3 From Partial to Point Identification as $T \to \infty$.

The previous analysis assumed that $\mathcal{T}(T) = \{1, \ldots, T\}$ where $T < \infty$ is fixed. We will now turn to an analysis where $T \to \infty$, which allows us to obtain point identification on a wider range of values. To this end, we invoke the following assumption:

Assumption 5. *Limit Point: Given t, x is a limit point of the sequence* $\left\{F_{X_s}^{-1} \circ F_{X_t}(x)\right\}_{s=1}^{\infty}$.

This assumption intuitively means the following. Suppose that an agent with some unobserved ranks $V_s = v$ chooses $X_t = x$ at some point t in time. Then, a choice of X arbitrarily close to (but not exactly) x will eventually be taken by an agent with similar unobserved characteristics $V_s \approx v$ over the infinite horizon. Assumption 5 can be satisfied for x over a set of positive measure unlike Assumption 3, since the separability of $\mathbb{R}$ allows the set of all the intersection points of $\{F_{X_t}\}_t$ to constitute a dense subset of a continuum. This property is useful since identification over a set of positive measure enables parametric rate of convergence of estimators in certain applications, cf. Section 5.3.

Assumption 6. *Regularity:*
(i) Given x, $[\phi(x+\delta, a, u) - \phi(x, a, u)]/\delta \to \beta(x, a, u)$ as $\delta \to 0$ uniformly across (a, u).
(ii) Given x, $\beta(x, a, u)$ is bounded uniformly across (a, u).

Under this regularity assumption, an application of integration theory yields the following auxiliary property.

Lemma 2. *Suppose that Assumption 6 holds. If $x_j \to x$, then*

$$\mathbb{E}\left[\left.\frac{\phi(x_j, A, U_t) - \phi(x, A, U_t)}{x_j - x}\right| X_t = x\right] \to \mathbb{E}\left[\beta(x, A, U_t)\right| X_t = x\right].$$

Armed with this lemma, it remains to find a sequence $x_j \to x$ across time in order to obtain T-asymptotic point identification of the LAR. The first part of the following theorem states that we can indeed find such a convergent sequence, whose existence is guaranteed by Assumption 5. Once we find such a sequence, Lemmata 1 and 2 together yield the T-asymptotic point identification result. Moreover, the sequence that T-asymptotically point-identify the LAR will turn out to be exactly the sequence of bounds from Theorem 1.

Theorem 3 (Point Identification as $T \to \infty$)**.** *Suppose that Assumptions 1, 5, and 6 hold in (1.1). Given $x \in supp(X_t)$ for some t, at least one of the following two results holds:*

$$(I) \qquad F^{-1}_{X_{\underline{\tau}(T,t,x)}} \circ F_{X_t}(x) \to x \quad as \quad T \to \infty \qquad \textit{and/or}$$
$$(II) \qquad F^{-1}_{X_{\overline{\tau}(T,t,x)}} \circ F_{X_t}(x) \to x \quad as \quad T \to \infty.$$

If (I) is the case, then $\Delta(\underline{\tau}(T,t,x),t,x) \to \mathbb{E}\left[\beta(x,A,U_t)\middle|\, X_t = x\right]$ as $T \to \infty$.
If (II) is the case, then $\Delta(\overline{\tau}(T,t,x),t,x) \to \mathbb{E}\left[\beta(x,A,U_t)\middle|\, X_t = x\right]$ as $T \to \infty$.

Recall that $\Delta(\underline{\tau}(T,t,x),t,x)$ and $\Delta(\overline{\tau}(T,t,x),t,x)$ are the objects that constitute the sharp bounds of the LAR for $T < \infty$ in Theorem 1. Now, this Theorem 3 states that at least one of them converges to the LAR as $T \to \infty$. Moreover, we know which of them will do so since we can observe in the data which of (I) or (II) holds. Figure 3 illustrates the mechanism of this Theorem 3 graphically. As time accumulates, more choices of counter-factual x become available in an infinitesimal neighborhood (bottom figure), hence providing increasingly precise difference quotients (top figure).

Assumption 5 states that a point x will be revisited by an agent with a similar unobserved characteristics over the infinite horizon. Theorem 3 shows T-asymptotic convergence of the bound(s) to the LAR at such points x. This is analogous to one of the results of CFHN where they used the ergodicity assumption to show T-asymptotic convergence of their bounds to their parameters. Table 1 summarizes the required assumptions, advantages, and disadvantages of the three identification theorems.

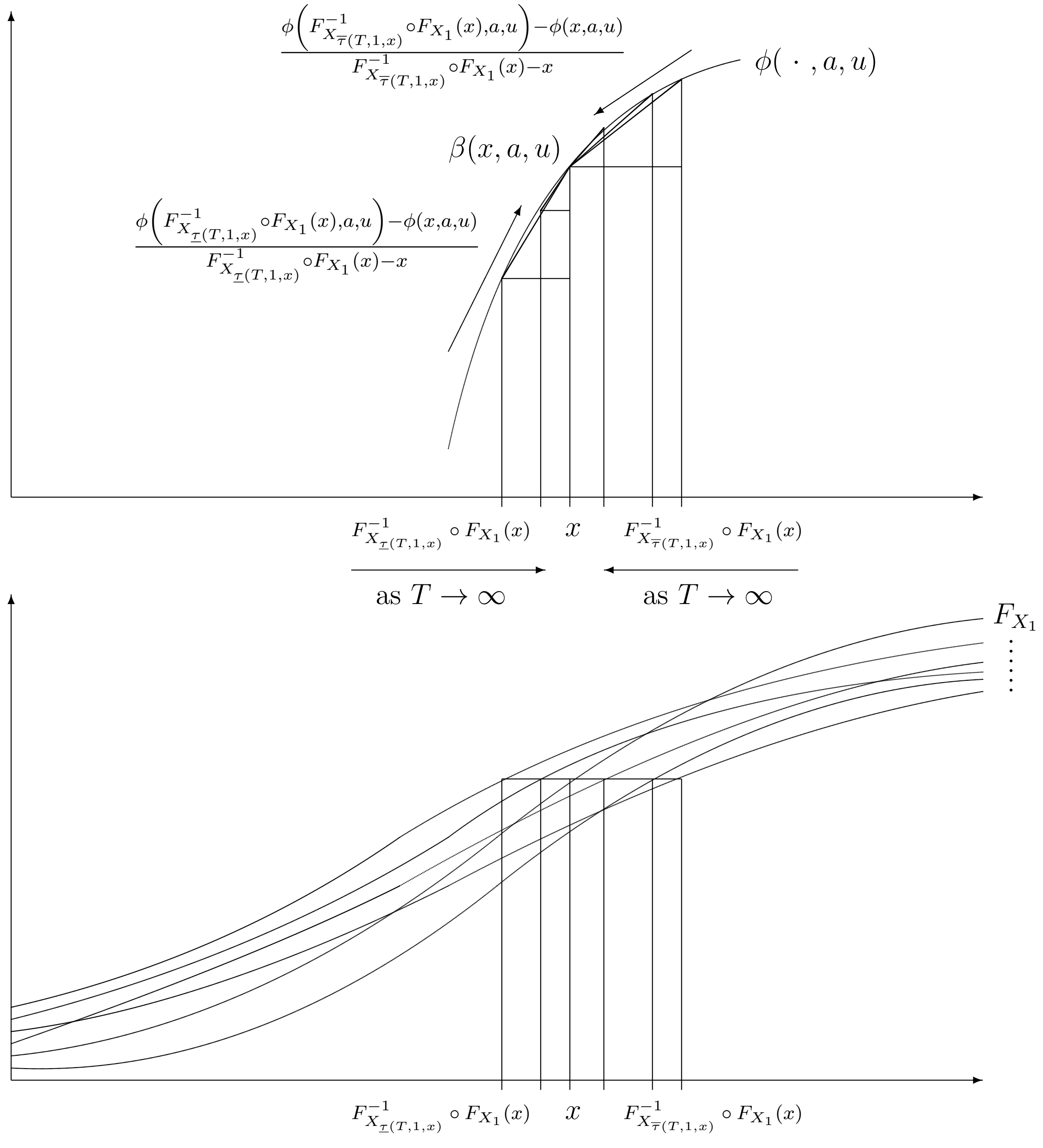

Figure 3: Point identification as $T \to \infty$: left/right difference quotients converge to $\beta(x,a,u)$.

	Time	Identification	Locality	Assumptions
Theorem 1	$T < \infty$	Partial		1 – Basic Restrictions 2 – Sufficient Proximity
Theorem 2	$T < \infty$	Point	Only at x^*	1 – Basic Restrictions 3 – Local Counterfactual 4 – Regularity
Theorem 3	$T \to \infty$	Point		1 – Basic Restrictions 5 – Limit Point 6 – Regularity

Table 1: Summary of identification results for local mean partial effects.

3 Numerical Illustration

In this simulation study we illustrate some of the above results numerically and graphically. A surprising but reappearing fact in this literature is that the bounds become tight very quickly, implying that for our "large T" results to hold we in fact require only few time periods. As such, this simulation study underscores that the results obtained are relevant for applications of even moderate time dimension.

Consider the following data generating process that is compatible with Assumption 1.

$$V \sim \text{Unif}(0,1), \quad A \sim \text{Normal}(V,1), \quad U_t \sim \text{Normal}(0,1),$$
$$X_t := F_{X_t}^{-1}(V), \quad Y_t := \phi(X_t, A, U_t) \quad \text{for each } t$$

Define a nonlinear nonseparable structural function (1.1) by

$$\phi(x,a,u) := \sqrt{x}e^{a+u},$$

which satisfies Assumption 2 globally. A sequence $\{F_{X_t}\}_{t=1}^{\infty}$ consists of the cdf's for uniform distributions with time-varying limits of support. For simplicity, F_{X_t} is the cdf of the law $\text{Unif}(\underline{x}_t, \overline{x}_t)$ for some random process $\{\underline{x}_t, \overline{x}_t\}_{t=1}^{\infty}$ with $0 < \underline{x}_t < \overline{x}_t$.

In this setting, we can numerically calculate the LAR

$$D(t,x) \ := \ \mathbb{E}\left[\beta(x,A,U_t)\right| X_t = x]$$

and its bounds (cf. Theorem 1):

$$\begin{aligned} L(T,t,x) :=& \quad \min\left\{\Delta(\underline{\tau}(T,t,x),t,x), \Delta(\overline{\tau}(T,t,x),t,x)\right\} \\ U(T,t,x) :=& \quad \max\left\{\Delta(\underline{\tau}(T,t,x),t,x), \Delta(\overline{\tau}(T,t,x),t,x)\right\} \end{aligned}$$

In particular, we will study

$$D(1,x) \ = \ \mathbb{E}\left[\beta(x,A,U_t)\right| X_1 = x\right],$$

which is the LAR at x among the subpopulation of individuals who chose $X_t = x$ at time $t = 1$. By Theorem 1, its lower and upper bounds from T-period panel data are

$$\begin{aligned} L(T,1,x) &= \min\left\{\Delta(\underline{\tau}(T,1,x),1,x), \Delta(\overline{\tau}(T,1,x),1,x)\right\} \\ U(T,1,x) &= \max\left\{\Delta(\underline{\tau}(T,1,x),1,x), \Delta(\overline{\tau}(T,1,x),1,x)\right\} \end{aligned}$$

respectively. The identification as $T \to \infty$ is investigated numerically and illustrated in Figure 4.

The top left graph in the figure illustrates the true $D(1,x)$ across $x \in (0,1)$. It is decreasing near $x = 0$, reflecting strong diminishing returns of ϕ. On the other hand, it is increasing near $x = 1$, reflecting positively endogenous choice of x dominating relatively flattened marginal returns of ϕ. The rest of the graphs in the figure illustrates a relationship between the true $D(1,x)$ and its bounds $L(T,1,x)$ and $U(T,1,x)$ as T rises. The bounds $L(T,1,x)$ and $U(T,1,x)$ are missing for values of x at which $\mathcal{T}^-(T,1,x) = \emptyset$ or $\mathcal{T}^+(T,t,x) = \emptyset$. Note that $L(T,1,x)$ indeed appears below $D(1,x)$ and $U(T,1,x)$ indeed appears above $D(1,x)$, thus numerically illustrating Theorem 1. Moreover, for some points x^*, both bounds, $L(T,1,x)$ and $U(T,1,x)$, coincide with the true $D(1,x)$ as early as at $T = 4$. In other words, local point identification is achieved under $T < \infty$, which numerically illustrates Theorem 2. Lastly, the bounds point-wise converge to $D(1,x)$ as T rises, hence numerically illustrating Theorem 3.

4 IV and Panels: Similarities and Differences

Our results on the passage from partial to point identification of panel data models as $T \to \infty$ parallel the analogous passage in cross-sectional IV models as the support of instrumental

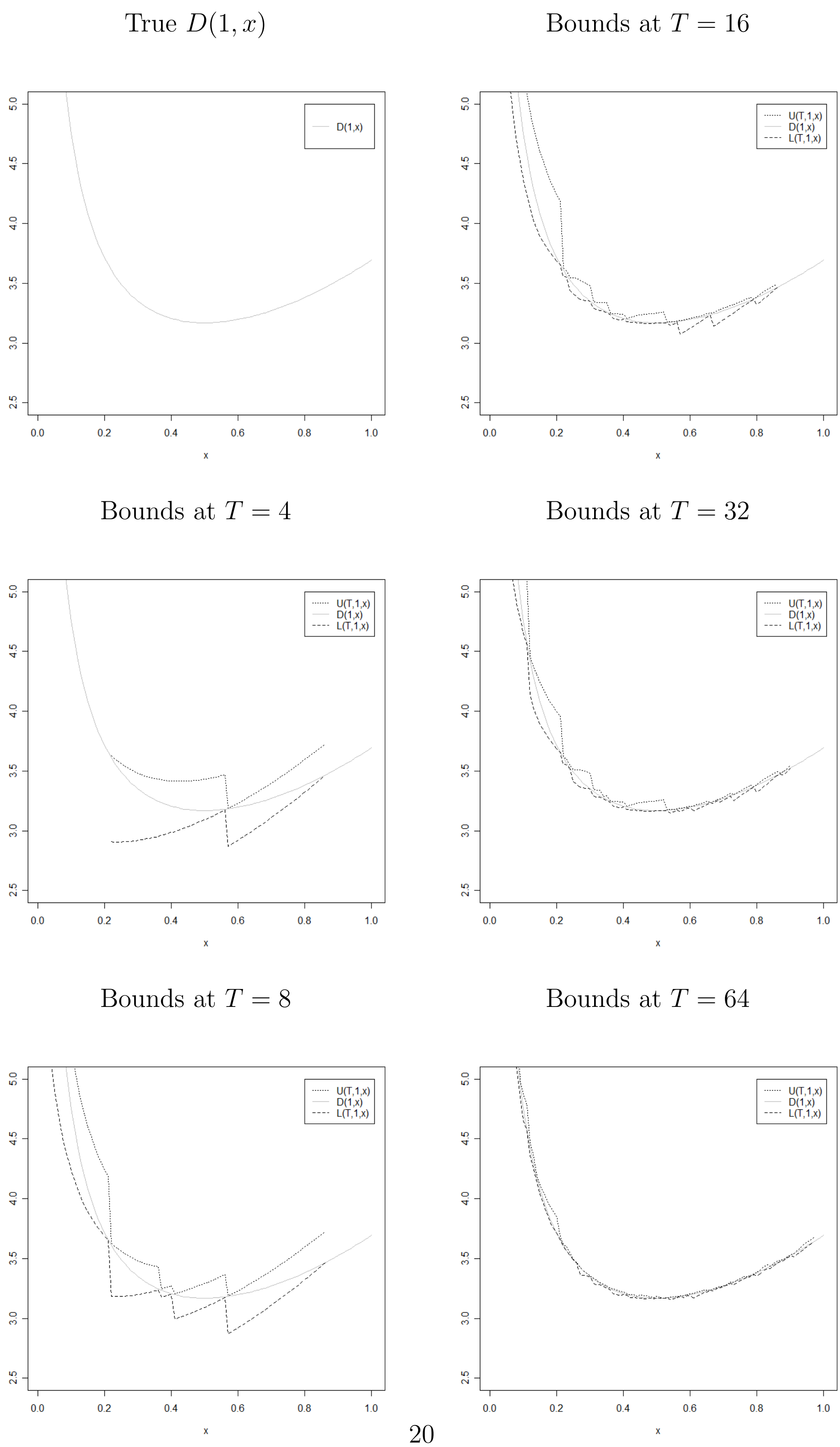

Figure 4: Numerical illustration of transition of identifiability as T rises.

variable becomes richer. Structural functions in IV models are point identified if the support of the instrumental variable is rich, while partial identification prevails otherwise (Chesher, 2005; Sec. 2.6). A richer support of the instrument allows for a finer resolution of the counterfactual variation in the endogenous variable. Similarly, in panel data, a large number of points x^* of non-tangential intersection of F_{X_t}, as characterized in Assumption 3, provides fine resolution of counterfactual variations in x in a neighborhood of x^*. As time increases, under appropriate assumptions the number of such points x^* increases, hence allowing for a better approximation of structural features at an increasing number of locations.

To make a connection with the existing literature on cross-sectional IV models, consider the triangular system with instrumental variable Z, for which we introduce the following nonstandard notation:

$$\begin{cases} Y_z = \eta(X_z, A) \\ X_z = \psi(z, V) \end{cases}$$

where X is continuous, but Z can be discrete and is fixed at the value z. To make the connection to the IV case transparent, we adopt for the following argumentation the assumption that the structural model of the outcome equation contains a scalar unobservable which enters monotonically. In this setup, Chesher (2005; Sec. 2.6) showed that $F^{-1}_{Y|XZ}(a \mid x, z')$ and $F^{-1}_{Y|XZ}(a \mid x, z'')$ form bounds of $\phi\left(x, F^{-1}_{A|V}(a \mid v)\right)$, where $F_{X|Z}(x \mid z') \leq v \leq F_{X|Z}(x \mid z'')$. He further shows that as the support of Z becomes richer, more values of z become available to eventually allow to equate $F_{X|Z}(x \mid z) = v$ for some $z \in \text{supp}(Z)$, which in turn allows point identification of $\phi\left(x, F^{-1}_{A|V}(a \mid v)\right)$ by $F^{-1}_{Y|XZ}(a \mid x, z)$.

Replacing z by time t yields a "restrictive panel data model"

$$\begin{cases} Y_t = \eta(X_t, A) \\ X_t = \psi(t, V) \end{cases} \tag{4.1}$$

which does not contain U_t, the transitory component[3]. In the model 4.1, exogeneity of time t,

[3]This could either be the model from the outset, or it could arise out of a model

$$\begin{cases} \bar{Y}_t = \phi(X_t, A, U_t) \\ X_t = \psi(t, V) \end{cases},$$

i.e., invariance of the two unobservables A and V to changes in t, of course yields exactly the same passage from partial to point identification, except that the label t replaces the label z. That is, $F^{-1}_{Y_t|X_t}(a \mid x)$ and $F^{-1}_{Y_s|X_s}(a \mid x)$ form bounds of $\eta\left(x, F^{-1}_{A|V}(a \mid v)\right)$, where $F_{X_t}(x) \leq v \leq F_{X_s}(x)$. As $T \to \infty$, more values of t become available to eventually achieve $F_{X_t}(x) \cong v$ for some $t \in \mathcal{T}(T)$, hence allowing point identification of $\eta\left(x, F^{-1}_{A|V}(a \mid v)\right)$ by $F^{-1}_{Y_t|X_t}(a \mid x)$ with such t.

However, this naive translation from the cross-sectional IV model is too restrictive. In contrast to instruments, panel data allow us to observe (Y_t, X_t) for any t. Hence, we do not have to assume that the first stage function ψ entails rank-invariance of X_t across time t, we can allow for a time dependent error in both the outcome and the first stage equation. In particular, instead of $X_t = \psi(t, V)$, we can allow for the relationship $X_t = \psi(t, V_t)$, i.e., individuals can have a different rank across time. Thus, including the previous two models, we can distinguish the following three first-stage models:

	Model	First Stage	Control Variable
(I)	Cross-Sectional IV Models	$X_z = \psi(z, V)$	$F_{X\|Z}(X_z, z) = V$
(II)	Restrictive Panel Data with Rank Invariance	$X_t = \psi(t, V)$	$F_{X_t}(X_t) = V$
(III)	General Panel Data with Rank Mobility	$X_t = \psi(t, V_t)$	$F_{X_t}(X_t) = V_t$

where ψ is strictly monotone in the second argument, and the distributions of V_t and V are absolutely continuous. If we normalize w.l.o.g. V_t and V to be $\mathcal{U}(0, 1)$, the associated control variables follow as shown in the last column in the above table. In the models (I) and (II) that are mutually equivalent except for the labels t and z, the control variable is the z- or t-invariant rank V of X. On the other hand, in the most general panel data model (III), the control variable is the time-varying rank V_t of X.

The strong assumption of first-stage rank invariance as in the naive panel data model (II) is not necessary, and the weaker restriction of Assumption 1 (IJD) compatible with model (III) suffices for our identification results. This Assumption 1 (IJD) states that $(A, V_t) \overset{\text{Law}}{=} (A, V_s)$

by integrating over U_t given X_t, A, i.e., $\eta(\xi, a) = \mathbb{E}_{U_t}[\phi(\xi, a, U_t)]$ (hence the exogeneity of the period-specific error U_t, and time invariance of its distribution. For the following argument it does not matter whether we have the original observations $\bar{Y}_t$ or $Y_t = \mathbb{E}[\bar{Y}_t|X_t, A]$) because we integrate over the unobservables anyway.

for all $s, t \in \mathcal{T}(T)$, hence guaranteeing t-invariance of the conditional distribution $F_{A|V_t}$, i.e., we can write $F_{A|V_t} = F_{A|V}$ for all $t \in \mathcal{T}(T)$. Because of this restriction we can use both period s and t information, and

$$F^{-1}_{Y_t|X_t}(a \mid x) \text{ and } F^{-1}_{Y_s|X_s}(a \mid x) \text{ form bounds for } \phi\left(x, F^{-1}_{A|V_t}(a \mid v)\right) \text{ in model (III)}$$

where $F_{X_t}(x) \leq v \leq F_{X_s}(x)$.[4] Similarly, large T point identification holds for (III) as well as it holds for (II) under Assumption 1 (IJD). Since our parameter of interest is the average partial effect under a non-monotonic outcome equation, we use a different approach of partial identification than Chesher (2005) in this paper; however, the bounds retain the logic to control $A \mid V_t = v$ as presented here.

In summary, we have the following observations about similarity and differences between panel data and IV models. The intermediate model (II) bridges the two important models (I) and (III). The models (I) and (II) are essentially the same in structure, except for the different labels t and z. Model (II) is quite restrictive as a panel data model, whereas model (III) admits more general panel data structures. By the rank similarity restriction postulated as Assumption 1 (IJD), however, the model (III) is shown to have the partial and T-asymptotic point identification results equivalent to those in model (II), which in turn is equivalent to the traditional IV models of type (I).

5 Extensions

In this section, we extend the main identification results of Section 2 to models with covariates and semiparametric binary choice models.

[4] Also,

$$F^{-1}_{Y_t|X_t}(a \mid x) \text{ and } F^{-1}_{Y_s|X_s}(a \mid x) \text{ form bounds for } \phi\left(x, F^{-1}_{A|V}(a \mid v)\right) \text{ in model (II)}$$

where $F_{X_t}(x) \leq v \leq F_{X_s}(x)$.

5.1 Extension I: Covariates

While we have suppressed covariates in our analysis in Section 2, similar conclusions carry over even if we had covariates in the structure. Suppose that the structural function ϕ contains a covariate Z_t as an additional argument

$$Y_t = \phi(X_t, Z_t, A, U_t) \qquad t \in \mathcal{T}(T) := \{1, \ldots, T\} \tag{5.1}$$

The object of interest is the conditional mean of the partial effects $\beta_x := \frac{\partial}{\partial x}\phi$. With the notation $V_t := F_{X_t Z_t}(X_t, Z_t)$, consider the following adaptation of the basic restrictions to the new scenario:

Assumption 7. *Basic Restrictions with Endogenous Covariates:*
(ID) U_t is identically distributed across $t \in \mathcal{T}(T)$.
(IND′) $U_s \perp\!\!\!\perp (X_t, Z_t, A)$ for all $s, t \in \mathcal{T}(T)$.
(ACC′) The distribution of $X_t \mid Z_t = z$ is absolutely continuous with a convex support.
(IJD′) $(A, V_s, Z_s) \overset{Law}{=} (A, V_t, Z_t)$ for all $s, t \in \mathcal{T}(T)$.

Note that this set of assumptions is largely comparable to the previous ones. With respect to Z_t, however, the assumptions leave the distribution largely unspecified (i.e., Z_t could be discrete or continuous, time invariant or varying), and also do not restrict the correlation structure with A (other than the mild IJD' assumption). Define a slight extension of Δ as follows.

$$\bar{\Delta}(s, t, x, z) := \frac{\mathbb{E}\left[Y_s \mid X_s = F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z), Z_s = z\right] - \mathbb{E}\left[Y_t \mid X_t = x, Z_t = z\right]}{F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z) - x}$$

With this device, we obtain the following lemma, which is analogous to Lemma 1, except that Z_t now appears as an additional conditioning variable

Lemma 3. *Suppose that Assumption 7 holds. Then*

$$\mathbb{E}\left[\left.\frac{\phi(F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z), z, A, U_t) - \phi(x, z, A, U_t)}{F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z) - x}\right| X_t = x, Z_t = z\right] = \bar{\Delta}(s, t, x, z)$$

holds for all $(x, z) \in supp(X_t, Z_t)$ for all $s, t \in \mathcal{T}(T)$ such that $F_{X_s Z_s}(x, z) \neq F_{X_t Z_t}(x, z)$.

Applying this lemma yields three results analogous to the main identification theorems. First $\bar{\Delta}(s,t,x,z)$ partially identifies the conditional average partial effect $\mathbb{E}[\beta_x(x,z,A,U_t) \mid X_t = x, Z_t = z]$ under $T < \infty$ similarly to Theorem 1. Second, $\bar{\Delta}(s,t,x,z)$ locally point-identify $\mathbb{E}[\beta_x(x,z,A,U_t) \mid X_t = x, Z_t = z]$ under $T < \infty$ similarly to Theorem 2. Third, $\bar{\Delta}(s,t,x,z)$ point-identify $\mathbb{E}[\beta_x(x,z,A,U_t) \mid X_t = x, Z_t = z]$ as $T \to \infty$ similarly to Theorem 3.

5.2 Extension II: Gradients of Multivariate Regressors

Consider again the structural function (5.1) with (X_t, Z_t) as bivariate regressors. In Section 5.1, the only object of interest was the conditional mean of the partial effects $\beta_x := \frac{\partial}{\partial x}\phi$. Now, suppose that we are interested in the conditional mean of the gradient $(\beta_x, \beta_z)'$, where $\beta_z := \frac{\partial}{\partial z}\phi$. Identification of this gradient will require time variation of the marginal distributions of Z_t as well as X_t. Hence, the (IJD$'$) in Assumption 7 is no longer a feasible restriction. We therefore, alter Assumption 7 as follows.

Assumption 8. *Basic Restrictions with Endogenous Covariates:*
(ID) U_t is identically distributed across $t \in \mathcal{T}(T)$.
(IND$'$) $U_s \perp\!\!\!\perp (X_t, Z_t, A)$ and $A \perp\!\!\!\perp (X_t, Z_t) \mid V_t$ for all $s, t \in \mathcal{T}(T)$.
(ACC$'$) The distribution of (X_t, Z_t) is absolutely continuous with a convex support.
(IJD$''$) $(A, V_s) \overset{Law}{=} (A, V_t)$ for all $s, t \in \mathcal{T}(T)$.

Two changes are noticeable: First, the relaxation in IJD$''$, to allow time variation of F_{Z_t}. Second, the new independence restriction $A \perp\!\!\!\perp (X_t, Z_t) \mid V_t$, which is not assumed prior to this subsection. While this is an index sufficiency type of restriction, it is different from the assumption in Altonji and Matzkin (2005), as it restricts the contemporaneous correlations between the regressors and A, and not the correlation between the regressors at different points in time and A. As such, it has more the flavor of a dimension reduction device, and is also used in this purpose, as the following arguments illustrate.

Define a slight extension of Δ as follows.

$$\tilde{\Delta}_x(s,t,x,z) \;:=\; \frac{\mathbb{E}\left[Y_s \mid X_s = F^{-1}_{X_sZ_s}(\cdot,z)\circ F_{X_tZ_t}(x,z), Z_s = z\right] - \mathbb{E}\left[Y_t \mid X_t = x, Z_t = z\right]}{F^{-1}_{X_sZ_s}(\cdot,z)\circ F_{X_tZ_t}(x,z) - x}$$

$$\tilde{\Delta}_z(s,t,x,z) \;:=\; \frac{\mathbb{E}\left[Y_s \mid X_s = x, Z_s = F^{-1}_{X_sZ_s}(x,\cdot)\circ F_{X_tZ_t}(x,z)\right] - \mathbb{E}\left[Y_t \mid X_t = x, Z_t = z\right]}{F^{-1}_{X_sZ_s}(x,\cdot)\circ F_{X_tZ_t}(x,z) - z}$$

With this object, we obtain a lemma which is quite analogous to Lemma 3, except that Assumption 8 replaces Assumption 7.

Lemma 4. *Suppose that Assumption 8 holds. Then*

$$\mathbb{E}\left[\left.\frac{\phi(F^{-1}_{X_sZ_s}(\cdot,z)\circ F_{X_tZ_t}(x,z), z, A, U_t) - \phi(x,z,A,U_t)}{F^{-1}_{X_sZ_s}(\cdot,z)\circ F_{X_tZ_t}(x,z) - x}\right| X_t = x, Z_t = z\right] = \tilde{\Delta}_x(s,t,x,z)$$

$$\mathbb{E}\left[\left.\frac{\phi(x, F^{-1}_{X_sZ_s}(x,\cdot)\circ F_{X_tZ_t}(x,z), A, U_t) - \phi(x,z,A,U_t)}{F^{-1}_{X_sZ_s}(x,\cdot)\circ F_{X_tZ_t}(x,z) - z}\right| X_t = x, Z_t = z\right] = \tilde{\Delta}_z(s,t,x,z)$$

hold for all $(x,z) \in supp(X_t,Z_t)$ for all $s,t \in \mathcal{T}(T)$ such that $F_{X_sZ_s}(x,z) \neq F_{X_tZ_t}(x,z)$.

Because of the relaxed IJD′ assumption, the identification results hold for both X and Z symmetrically. The partial identification result uses the following notations and assumptions.

$$\underline{\tau}_x(T,t,x,z) := \quad \arg\max_{s\in\mathcal{T}(T)}\{F^{-1}_{X_sZ_s}(\cdot,z)\circ F_{X_tZ_t}(x,z) \mid F^{-1}_{X_sZ_s}(\cdot,z)\circ F_{X_tZ_t}(x,z) < x\}$$

$$\overline{\tau}_x(T,t,x,z) := \quad \arg\min_{s\in\mathcal{T}(T)}\{F^{-1}_{X_sZ_s}(\cdot,z)\circ F_{X_tZ_t}(x,z) \mid F^{-1}_{X_sZ_s}(\cdot,z)\circ F_{X_tZ_t}(x,z) > x\}$$

Assumption 9. *Local Curvature:*
There exists an interval I containing elements x, $F^{-1}_{X_{\underline{\tau}_x(T,t,x,z)}Z_{\underline{\tau}_x(T,t,x,z)}}(\cdot,z)\circ F_{X_tZ_t}(x,z)$, and $F^{-1}_{X_{\overline{\tau}_x(T,t,x,z)}Z_{\overline{\tau}_x(T,t,x,z)}}(\cdot,z)\circ F_{X_tZ_t}(x,z)$ such that the sign of $\frac{\partial^2}{\partial x^2}\phi(x',z,a,u)$ is invariant for all $x' \in I$ and $[P_{AU_t}]$-a.s. (a,u).

Theorem 4 (Partial Identification for $T<\infty$)**.** *Let $(x,z)\in supp(X_t,Z_t)$ for some $t\in\mathcal{T}(T)$. Suppose that Assumptions 8 and 9 hold. Then we obtain that*

$$L_x(T,t,x,z) \leqslant \mathbb{E}\left[\beta_x(x,z,A,U_t)\right| X_t = x, Z_t = z] \leqslant U_x(T,t,x,z)$$

i.e., the LAR is set identified with sharp bounds given by

$$L_x(T,t,x,z) := \quad \min\left\{\tilde{\Delta}(\underline{\tau}_x(T,t,x,z),t,x,z), \tilde{\Delta}(\overline{\tau}_x(T,t,x,z),t,x,z)\right\} \qquad \textit{and}$$

$$U_x(T,t,x,z) := \quad \max\left\{\tilde{\Delta}(\underline{\tau}_x(T,t,x,z),t,x,z), \tilde{\Delta}(\overline{\tau}_x(T,t,x,z),t,x,z)\right\}.$$

Symmetrically exchanging the roles of x and z yields sharp bounds for the LAR with respect to z, which is $\mathbb{E}\left[\beta_z(x,z,A,U_t)\right| X_t = x, Z_t = z]$.

The local point identification under finite T uses the following straightforward adaptation of assumptions.

Assumption 10. *Local Counterfactual: There exist two time indices $s,t \in \mathcal{T}(T)$ and (x^*, z^*) such that $F_{X_tZ_t}(x^*,z^*) = F_{X_sZ_s}(x^*,z^*)$, x^* is in the closure of the set $\{x \mid F_{X_tZ_t}(x,z^*) \neq F_{X_sZ_s}(x,z^*)\}$, and z^* is in the closure of the set $\{z \mid F_{X_tZ_t}(x^*,z) \neq F_{X_sZ_s}(x^*,z)\}$.*

Assumption 11. *Regularity:*
(i) $[\phi(\tilde{x},z^*,a,u) - \phi(x,z^*,a,u)]/(\tilde{x}-x) \to \beta_x(x^*,z^*,a,u)$ as $x,\tilde{x} \to x^*$ uniformly across (a,u).
(ii) $\beta(x,z^*,a,u)$ is bounded uniformly across (a,u) and across x in a neighborhood of x^*.
(iii) For t, $f_{A|X_tZ_t}(a \mid x^*+\delta, z^*) \to f_{A|X_tZ_t}(a \mid x^*,z^*)$ as $\delta \to 0$ uniformly across a.

Theorem 5 (Local Point Identification under $T < \infty$)**.** *Suppose that Assumptions 8 and 11 hold. Then, for $s,t \in \mathcal{T}(T)$ and (x^*,z^*) that satisfy Assumption 10, we have*

$$\lim_{x\to x^*} \tilde{\Delta}_x(s,t,x,z^*) = \mathbb{E}\left[\beta_x(x^*,z^*,A,U_t)\right| X_t = x^*, Z_t = z^*] .$$

Symmetrically exchanging the roles of x and z yields local point identification of the LAR, $\mathbb{E}\left[\beta_z(x,z,A,U_t)\right| X_t = x, Z_t = z]$.

The T-asymptotic point identification uses the following modified assumptions.

Assumption 12. *Limit Point:*
Given t and z, x is a limit point of the sequence $\left\{F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z)\right\}_{s=1}^{\infty}$.

Assumption 13. *Regularity: Given (x,z), the following hold.*
(i) $[\phi(x+\delta,z,a,u) - \phi(x,z,a,u)]/\delta \to \beta_x(x,z,a,u)$ as $\delta \to 0$ uniformly across (a,u).
(ii) $\beta_x(x,z,a,u)$ is bounded uniformly across (a,u).

Theorem 6 (Point Identification as $T \to \infty$)**.** *Suppose that Assumptions 8, 12, and 13 hold. Given $(x,z) \in supp(X_t,Z_t)$ for some t, at least one of the following two results holds:*

$$(I) \qquad F^{-1}_{X_{\underline{\tau}_x(T,t,x,z)}Z_{\underline{\tau}_x(T,t,x,z)}}(\cdot,z) \circ F_{X_tZ_t}(x,z) \to x \quad \text{as} \quad T \to \infty \qquad \textit{and/or}$$

$$(II) \qquad F^{-1}_{X_{\overline{\tau}_x(T,t,x,z)}Z_{\overline{\tau}_x(T,t,x,z)}}(\cdot,z) \circ F_{X_tZ_t}(x,z) \to x \quad \text{as} \quad T \to \infty.$$

If (I) is the case, then $\Delta(\underline{\tau}_x(T,t,x,z),t,x,z) \to \mathbb{E}\left[\beta_x(x,z,A,U_t)\right| X_t = x, Z_t = z]$ *as* $T \to \infty$. *If (II) is the case, then* $\Delta(\overline{\tau}_x(T,t,x,z),t,x,z) \to \mathbb{E}\left[\beta_x(x,z,A,U_t)\right| X_t = x, Z_t = z]$ *as* $T \to \infty$. *Symmetrically exchanging the roles of* x *and* z *yields* T*-asymptotic point identification of the LAR,* $\mathbb{E}\left[\beta_z(x,z,A,U_t)\right| X_t = x, Z_t = z]$.

Theorems 4, 5, and 6 are multivariate extensions of Theorems 1, 2, and 3 for the univariate case, respectively. All results extend straightforwardly, taking into account the modification required due to the fact that we are no looking at the gradient, and hence have to invoke modified independence conditions that allow to interpret the changes in the denominator of the local difference quotient in a symmetric fashion. These results will be applied in Section 5.3 in order to identify the semiparametric binary choice models, which is also a model that requires identification of a gradient.

5.3 Extension III: Semiparametric Binary Choice Models

In this section, we consider the semiparametric binary choice panel data model

$$Y_t = \mathbb{1}\{X_t\beta + Z_t\gamma + A + U_t^* \geqslant 0\} \qquad t \in \mathcal{T}(T) := \{1, \ldots, T\} \tag{5.2}$$

with endogenous regressors (X_t, Z_t) and individual fixed effects A. U_t^* denotes idiosyncratic errors at time t. Assuming $\gamma \neq 0$, the objective is to identify the parameter vector (β, γ) up to scale, i.e., to identify β/γ, which is the identifiable part of the vector of coefficients. As it turns out, we will require a ratio of derivatives, which means we have to use the gradient of effects. This implies that we cannot use the less restrictive extension put forward in Section 5.1, and have to employ stronger assumptions, akin to those of Section 5.2. The basic notations carry also over from Section 5.2:

Assumption 14. *Basic Restrictions for Semiparametric Binary Choice Models:*
(IND′) $U_s^* \perp\!\!\!\perp (X_t, Z_t, A)$ *and* $A \perp\!\!\!\perp (X_t, Z_t) \mid V_t$ *for all* $s, t \in \mathcal{T}(T)$.
(ACC) The distribution of (X_t, Z_t) *is absolutely continuous with a convex support.*
(IJD) $(A, V_s) \overset{Law}{=} (A, V_t)$ *for all* $s, t \in \mathcal{T}(T)$.

As in the standard semiparametric binary choice models, we write

$$\mathbb{E}[Y_t \mid X_t, Z_t, A] = \psi(X_t\beta, Z_t\gamma, A)$$

where $\psi(a,b,c) := F_{-U_t^*}(a+b+c)$. Moreover, defining the residual as

$$U_t := Y_t - \mathbb{E}[Y_t \mid X_t, Z_t, A],$$

allows us to construct the panel data model

$$Y_t \equiv \phi(X_t, Z_t, A, U_t) := \psi(X_t\beta, Z_t\gamma, A) + U_t \qquad t \in \mathcal{T}(T) := \{1, \ldots, T\}. \tag{5.3}$$

With this transformation, Assumption 14 implies Assumption 8, except that a counterpart of (ID) is missing. In the current setting (5.3) in which U_t is additively separable, we do not require a counterpart of (ID),[5] hence obtaining the same conclusion as Theorems 4–6 and from Section 5.2, as far as the associated additional assumptions are satisfied.

Theorem 4 implies that $\tilde{\Delta}_x(\underline{\tau}_x(T,t,x,z),t,x,z)$ and $\tilde{\Delta}_x(\overline{\tau}_x(T,t,x,z),t,x,z)$ form bounds of

$$\mathbb{E}\left[\frac{\partial}{\partial x}\phi(x,z,A,U_t)\middle| X_t = x, Z_t = z\right] = \mathbb{E}\left[f_{-U_t^*}(x\beta + z\gamma + A)\middle| X_t = x, Z_t = z\right]\cdot\beta \tag{5.4}$$

Theorem 6 implies that at least one of these bounds converges as $T \to \infty$. Similarly, Theorem 4 implies that $\tilde{\Delta}_z(\underline{\tau}_z(T,t,x,z),t,x,z)$ and $\tilde{\Delta}_z(\overline{\tau}_z(T,t,x,z),t,x,z)$ form bounds of

$$\mathbb{E}\left[\frac{\partial}{\partial z}\phi(x,z,A,U_t)\middle| X_t = x, Z_t = z\right] = \mathbb{E}\left[f_{-U_t^*}(x\beta + z\gamma + A)\middle| X_t = x, Z_t = z\right]\cdot\gamma \tag{5.5}$$

Since taking the ratio of (5.4) to (5.5) yields

$$\frac{\mathbb{E}\left[\frac{\partial}{\partial x}\phi(x,z,A,U_t)\middle| X_t = x, Z_t = z\right]}{\mathbb{E}\left[\frac{\partial}{\partial z}\phi(x,z,A,U_t)\middle| X_t = x, Z_t = z\right]} = \frac{\beta}{\gamma},$$

one may heuristically conjecture that the ratio $\tilde{\Delta}_x(\tau_x,t,x,z)/\tilde{\Delta}_z(\tau_z,t,x,z)$ yields some information about β/γ, in particular that the ratio may allow to somehow bound the ratio β/γ.

[5]The whole point of Assumption 8 (ID) was to make F_{U_t} time invariant (see proof of Lemma 4 in the appendix). However, under the current U_t-separable structure (5.3), U_t can be integrated out in the definition of $\tilde{\Delta}_x$, hence we do not require a counterpart of (ID).

However, closer inspection of the proofs reveal that an unsurmountable obstacle to partial identification of β/γ with finite T is that Theorem 4 requires generally Assumption 9 for ψ. This would in turn imply that the cdf $F_{-U_t^*}$ has no nearby inflection point, and assumption that is difficult to maintain for most cdfs and data configurations. Hence we can unfortunately not provide bounds for β/γ.

However, point identification results do carry over from Theorems 5 and 6 to the current setup since they do not rely on Assumption 9. Identification of β/γ with finite T is formally stated as Corollary 2, and follows from Theorem 5. This result is based only on a population of measure zero, consisting of a null set of the points x^* at which Assumption 10 is satisfied as remarked in Section 2.2. As in Hoderlein and White (2010), even under the stronger set of assumptions invoked in this paper we thus only obtain a nonparametric rate of convergence. Identification of β/γ as $T \to \infty$ using possibly a population of positive measure is stated as Corollary 3, and follows from Theorem 6. We think of this result as transporting the essence of why "bias correction" of $\sqrt{n}$ consistent estimators works for $T \to \infty$. Specifically, the underlying Assumption 12 can be satisfied for x over a set of positive measure since it is feasible that the set of all the intersection points of the cdfs constitute a dense subset of a continuum (as discussed in Section 2.3). Identification over a set of positive measure enables then a parametric rate of convergence for the corresponding estimators.

To state these results formally, we invoke the following assumptions.

Assumption 15. *Regularity:*
(i) $[F_{-U_t^*}(\tilde{e}) - F_{-U_t^*}(e)]/(\tilde{e} - e) \to f_{-U_t^*}(e^*)$ *as* $e, \tilde{e} \to e^*$.
(ii) $f_{-U_t^*}$ *is uniformly bounded.*
(iii) For t, $f_{A|X_t}(a \mid x^* + \delta) \to f_{A|X_t}(a \mid x^*)$ *and* $f_{A|Z_t}(a \mid z^* + \delta) \to f_{A|X_t}(a \mid z^*)$ *as* $\delta \to 0$ *uniformly across* a.

This Assumption 15 implies Assumption 11 with ϕ replaced by ψ. The next local point identification result of β/γ under $T < \infty$ thus follows from Theorem 5.

Corollary 2 (Local Point Identification under $T < \infty$)**.** *Suppose that Assumptions 10, 14, and*

15 hold in (1.1) with $\gamma \neq 0$. Then, for $s, t \in \mathcal{T}(T)$ and (x^*, z^*) that satisfy Assumption 10,

$$\frac{\lim_{x \to x^*} \tilde{\Delta}_x(s, t, x, z^*)}{\lim_{z \to z^*} \tilde{\Delta}_z(s, t, x^*, z)} = \frac{\beta}{\gamma}.$$

Assumption 16. *Regularity:*

(i) $[F_{-U_t^*}(e + \delta) - F_{-U_t^*}(e)]/\delta \to f_{-U_t^*}(e)$ as $\delta \to 0$ uniformly across e.

(ii) $f_{-U_t^*}$ is uniformly bounded.

This Assumption 16 implies Assumption 13 with ϕ replaced by ψ. The next point identification result of β/γ as $T \to \infty$ thus follows from Theorem 6.

Corollary 3 (Point Identification as $T \to \infty$)**.** *Suppose that Assumptions 12, 14, and 16 hold in (5.2) with $\gamma \neq 0$. Then, the parameter ratio β/γ is point-identified in the limit as $T \to \infty$. Specifically, for $(x, z) \in supp(X_t, Z_t)$ for some t, at least one of the following four results holds:*

$$
\begin{aligned}
&(I) && \left(F^{-1}_{X_{\underline{\tau}_x(T,t,x,z)} Z_{\underline{\tau}_x(T,t,x,z)}}(\cdot, z) \circ F_{X_t Z_t}(x, z), F^{-1}_{X_{\underline{\tau}_z(T,t,x,z)} Z_{\underline{\tau}_z(T,t,x,z)}}(x, \cdot) \circ F_{X_t Z_t}(x, z)\right) \to (x, z) \\
&(II) && \left(F^{-1}_{X_{\overline{\tau}_x(T,t,x,z)} Z_{\overline{\tau}_x(T,t,x,z)}}(\cdot, z) \circ F_{X_t Z_t}(x, z), F^{-1}_{X_{\underline{\tau}_z(T,t,x,z)} Z_{\underline{\tau}_z(T,t,x,z)}}(x, \cdot) \circ F_{X_t Z_t}(x, z)\right) \to (x, z) \\
&(III) && \left(F^{-1}_{X_{\underline{\tau}_x(T,t,x,z)} Z_{\underline{\tau}_x(T,t,x,z)}}(\cdot, z) \circ F_{X_t Z_t}(x, z), F^{-1}_{X_{\overline{\tau}_z(T,t,x,z)} Z_{\overline{\tau}_z(T,t,x,z)}}(x, \cdot) \circ F_{X_t Z_t}(x, z)\right) \to (x, z) \\
&(IV) && \left(F^{-1}_{X_{\overline{\tau}_x(T,t,x,z)} Z_{\overline{\tau}_x(T,t,x,z)}}(\cdot, z) \circ F_{X_t Z_t}(x, z), F^{-1}_{X_{\overline{\tau}_z(T,t,x,z)} Z_{\overline{\tau}_z(T,t,x,z)}}(x, \cdot) \circ F_{X_t Z_t}(x, z)\right) \to (x, z)
\end{aligned}
$$

as $T \to \infty$.

If (I) is the case, then $\tilde{\Delta}_x(\underline{\tau}_x(T, t, x, z), t, x, z)/\tilde{\Delta}_z(\underline{\tau}_z(T, t, x, z), t, x, z) \to \beta/\gamma$ as $T \to \infty$.

If (II) is the case, then $\tilde{\Delta}_x(\overline{\tau}_x(T, t, x, z), t, x, z)/\tilde{\Delta}_z(\underline{\tau}_z(T, t, x, z), t, x, z) \to \beta/\gamma$ as $T \to \infty$.

If (III) is the case, then $\tilde{\Delta}_x(\underline{\tau}_x(T, t, x, z), t, x, z)/\tilde{\Delta}_z(\overline{\tau}_z(T, t, x, z), t, x, z) \to \beta/\gamma$ as $T \to \infty$.

If (IV) is the case, then $\tilde{\Delta}_x(\overline{\tau}_x(T, t, x, z), t, x, z)/\tilde{\Delta}_z(\overline{\tau}_z(T, t, x, z), t, x, z) \to \beta/\gamma$ as $T \to \infty$.

Table 2 summarizes trade-offs concerning time requirement, assumptions, and population of information for point-identification of the semiparametric binary choice model (5.2). With the information from the set of points x^* satisfying Assumption 10, Corollary 2 yields point identification for fixed $T < \infty$ as in Manski (1987), but without requiring unbounded support of regressors. The fact that this identification result generally fails to utilize a population of positive measure under $T < \infty$ is not surprising, given the well-known findings of Chamberlain

	Time	Assumptions	Identification	Population of Information
Corollary 2	$T < \infty$	10, 14, 15	Point	Measure Zero
Corollary 3	$T \to \infty$	12, 14, 16	Point	Positive Measure

Table 2: Summary of time requirement, assumptions, identifiability, and population of information for the semiparametric binary choice model.

(2010), who shows that $\sqrt{n}$-consistent estimation of the binary choice model is generically infeasible, and only possible if the distribution of U_t is logistic (see also Lee, 1999; Arellano and Hahn, 2006), in which case a subpopulation of positive measure is available. Availability of only a population of measure zero means that the convergence rate of a sample-analog estimator falls short of $\sqrt{n}$, which parallels Chamberlain's impossibility result in our slightly different setup. As stressed before, Corollary 3 implies that the passage $T \to \infty$ may allow an estimator to utilize information from a population of positive measure under semiparametric conditions.

6 Estimation and Applicability

This paper is largely concerned with identification. In particular, the main results in this paper aim at making the role of time in a panel transparent. As such, the paper contributes to the understanding of the fundamentals of the model. However, both the results on bounds as well as the results on point identification at special x can easily be applied to data. In fact, compared to some results that require regressors from several time periods (e.g., Hoderlein and White (2010), Graham and Powell (2010), CFHN (2010)), since we only use contemporaneous X, i.e., always regress Y_t on X_t only, we actually face less of a "curse of dimensionality" problem. Also, the methods required can be taken from straightforward applications of standard nonparametric procedures, e.g., local polynomials or splines to estimate the respective regressions; for confidence intervals in the partially identified case methods can be taken from the rapidly expanding literature on partial identification. Hence, for the purpose of a concise exposition, we desist from elaborating at great length on the obvious, and focus in this paper on the main innovation. Nevertheless, we would like to emphasize that we think of many of the results we obtain as particularly useful for applications precisely because of the comparably low dimensionality and straightforward structure of all objects involved. This remarks apply in particular to the semiparametric binary choice case, which is a workhorse model of econometrics.

7 Conclusions

Time provides exogenous variation in a panel that allows to estimate structural models of interest, even in the presence of correlation between unobservables and regressors of interest. In our opinion, the key for understanding the role of time lies in the fact that the correlated unobservable is persistent, while the regressor of interest is time varying. If we assume stationarity of the distribution of transitory errors, the model effectively becomes very similar to a triangular IV model. Consequently, we may compare time to an instrument with discrete support.

In this paper, interest centers on marginal effects of a continuous explanatory variable. It is well known in a class of more restrictive models than we consider that discrete instruments provide only partial identification, and that partial identification "asymptotes" to point identification as the number of support points tend to infinity. This paper argues that we should think of time in the very same sense; as time accumulates, bounds shrink until we finally obtain point identification with many time periods. Also, with discrete instruments, only structural effects on subpopulations are identified. The same happens with time: the recent point identification results on subpopulations, e.g., Honore and Kyriazidou (2000), Arellano and Bonhomme (2010), Evdokimov (2010), Graham and Powell (2010), Hoderlein and White (2010), Chernozhukov, Fernandez Val, Hahn and Newey (2010) are in this sense not an accident. As we argue in this paper, when interest centers on local average responses or (average) random coefficients, point identification in subpopulations is generic for finite T. Discrete time hence indeed provides exogenous identifying information, but like a discrete instrument usually not enough to identify structural effects across the entire population. Translated to the semiparametric binary choice model, these issues manifest in different rates of convergence for the slope coefficient. As such, our paper complements the impossibility results of Chamberlain (2010), and provides further insight into the nature of identification of this important class of models. We hope this research spurns further interest in the important issue of understanding time in panels.

A Proofs of Auxiliary Lemmata

A.1 Proof of Lemma 1

Proof. First, note that

$$\mathbb{E}[\phi(x, A, U_t) \mid X_t = x] = \mathbb{E}[Y_t \mid X_t = x]$$

holds for all $x \in \operatorname{supp}(X_t)$ for all $t \in \mathcal{T}(T)$. In order to prove the lemma, it remains to claim that

$$\mathbb{E}[\phi(F_{X_s}^{-1} \circ F_{X_t}(x), A, U_t) \mid X_t = x] = \mathbb{E}[Y_s \mid X_s = F_{X_s}^{-1} \circ F_{X_t}(x)]$$

holds for all $x \in \operatorname{supp}(X_t)$ for all $s, t \in \mathcal{T}(T)$. We use the notation $V_t := F_{X_t}(X_t)$ for all $t \in \mathcal{T}(T)$. But then, Assumption 1 (IJD) implies $(A, V_s) \overset{\text{Law}}{=} (A, V_t)$ for all $s, t \in \mathcal{T}(T)$. With Assumption 1, we obtain

$$\begin{aligned}
\mathbb{E}[\phi(x', A, U_t) \mid X_t = x] &= \int\int \phi(x', a, u) f_{AU_t|X_t}(a, u \mid x) dadu \\
&\overset{\text{(IND)}}{=} \int\int \phi(x', a, u) f_{A|X_t}(a \mid x) f_{U_t}(u) dadu \\
&\overset{\text{(ACC)}}{=} \int\int \phi(x', a, u) f_{A|V_t}(a \mid F_{X_t}(x)) f_{U_t}(u) dadu \\
&\overset{\text{(IJD)}}{=} \int\int \phi(x', a, u) f_{A|V_s}(a \mid F_{X_t}(x)) f_{U_t}(u) dadu \\
&\overset{\text{(ACC)}}{=} \int\int \phi(x', a, u) f_{A|X_s}(a \mid F_{X_s}^{-1} \circ F_{X_t}(x)) f_{U_t}(u) dadu \\
&\overset{\text{(ID)}}{=} \int\int \phi(x', a, u) f_{A|X_s}(a \mid F_{X_s}^{-1} \circ F_{X_t}(x)) f_{U_s}(u) dadu \\
&\overset{\text{(IND)}}{=} \int\int \phi(x', a, u) f_{AU_s|X_s}(a, u \mid F_{X_s}^{-1} \circ F_{X_t}(x)) dadu \\
&= \mathbb{E}\left[\phi(x', A, U_s) \mid X_s = F_{X_s}^{-1} \circ F_{X_t}(x)\right]
\end{aligned}$$

Now, substitute $x' := F_{X_s}^{-1} \circ F_{X_t}(x)$ to obtain

$$\mathbb{E}[\phi(F_{X_s}^{-1} \circ F_{X_t}(x), A, U_t) \mid X_t = x] = \mathbb{E}[Y_s \mid X_s = F_{X_s}^{-1} \circ F_{X_t}(x)].$$

Thus, we have

$$\mathbb{E}[\phi(F_{X_s}^{-1} \circ F_{X_t}(x), A, U_t) - \phi(x, A, U_t) \mid X_t = x] = \mathbb{E}[Y_s \mid X_s = F_{X_s}^{-1} \circ F_{X_t}(x)] - \mathbb{E}[Y_t \mid X_t = x].$$

If $F_{X_s}(x) \neq F_{X_t}(x)$, then Assumption 1 (ACC) guarantees $F_{X_s}^{-1} \circ F_{X_t}(x) - x \neq 0$. Dividing both sides of the above equation by $F_{X_s}^{-1} \circ F_{X_t}(x) - x$ yields the desired result. $\square$

A.2 Proof of Lemma 2

Proof. By Assumption 6 (i), there exists a function λ such that

$$\left\| \frac{\phi(x_j, \cdot, \cdot) - \phi(x, \cdot, \cdot)}{x_j - x} - \frac{\partial}{\partial x}\phi(x, \cdot, \cdot) \right\|_\infty \leqslant \lambda(x_j - x)$$

where $\|\cdot\|_\infty$ is the uniform norm over (a, u), and $\lambda(x_j - x) \to 0$ as $x_j \to x$. Moreover, by Assumption 6 (ii), there exists a positive constant $\eta < \infty$ such that

$$\left\| \frac{\partial}{\partial x}\phi(x, \cdot, \cdot) \right\|_\infty \leqslant \eta$$

Therefore,

$$\left\| \frac{\phi(x_j, \cdot, \cdot) - \phi(x, \cdot, \cdot)}{x_j - x} \right\|_\infty \leqslant \lambda(x_j - x) + \eta$$

Moreover, by Hölder's inequality, we have

$$\begin{aligned} \left\| \frac{\phi(x_j, \cdot, \cdot) - \phi(x, \cdot, \cdot)}{x_j - x} f_{AU|X_t}(\cdot, \cdot \mid x) \right\|_1 &\leqslant \left\| \frac{\phi(x_j, \cdot, \cdot) - \phi(x, \cdot, \cdot)}{x_j - x} \right\|_\infty \left\| f_{AU|X_t}(\cdot, \cdot \mid x) \right\|_1 \\ &\leqslant \lambda(x_j - x) + \eta \end{aligned}$$

where $\|\cdot\|_1$ is the $\mathbb{L}^1$-norm (with respect to the Lebesgue measure). Since $\lambda(x_j - x) \to 0$ as $j \to \infty$, this implies that there is an $\mathbb{L}^1$ dominating function for the sequence

$$\left\{ \frac{\phi(x_j, \cdot, \cdot) - \phi(x, \cdot, \cdot)}{x_j - x} f_{AU|X_t}(\cdot, \cdot \mid x) \right\}_j$$

for large enough j. Therefore, by the Lebesgue Dominated Convergence Theorem, we have

$$\begin{aligned} \mathbb{E}\left[\frac{\phi(x_j, A, U_t) - \phi(x, A, U_t)}{x_j - x} \middle| X_t = x \right] &= \int\int \frac{\phi(x_j, a, u) - \phi(x, a, u)}{x_j - x} f_{AU|X_t}(a, u \mid x) da du \\ &\to \int\int \frac{\partial}{\partial x}\phi(x, a, u) f_{AU|X_t}(a, u \mid x) da du \\ &= \mathbb{E}\left[\frac{\partial}{\partial x}\phi(x, A, U_t) \middle| X_t = x \right]. \end{aligned}$$

as $x_j \to x$. $\square$

A.3 Proof of Lemma 3

Proof. First, note that $\mathbb{E}[\phi(x,z,A,U_t) \mid X_t = x, Z_t = z] = \mathbb{E}[Y_t \mid X_t = x, Z_t = z]$ holds for all $(x,z) \in \operatorname{supp}(X_t, Z_t)$ for all $t \in \mathcal{T}(T)$. In order to prove the lemma, it remains to claim that

$$\begin{aligned}
&\mathbb{E}[\phi(F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), z, A, U_t) \mid X_t = x, Z_t = z] \\
&\qquad = \mathbb{E}[Y_s \mid X_s = F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), Z_t = z]
\end{aligned}$$

holds for all $(x,z) \in \operatorname{supp}(X_t, Z_t)$ for all $s,t \in \mathcal{T}(T)$. We use the notation $V_t := F_{X_tZ_t}(X_t, Z_t)$ for all $t \in \mathcal{T}(T)$. But then, Assumption 8 (IJD′) implies $(A, V_s, Z_s) \overset{\text{Law}}{=} (A, V_t, Z_t)$ for all $s,t \in \mathcal{T}(T)$. With Assumption 8, we obtain

$$\begin{aligned}
& \mathbb{E}[\phi(x', z, A, U_t) \mid X_t = x, Z_t = z] \\
= \quad & \int\int \phi(x', z, a, u) f_{AU_t|X_tZ_t}(a, u \mid x, z) da du \\
\overset{(\text{IND}')}{=} \quad & \int\int \phi(x', z, a, u) f_{A|X_tZ_t}(a \mid x, z) f_{U_t}(u) da du \\
\overset{(\text{ACC}')}{=} \quad & \int\int \phi(x', z, a, u) f_{A|V_tZ_t}(a \mid F_{X_tZ_t}(x,z), z) f_{U_t}(u) da du \\
\overset{(\text{IJD}')}{=} \quad & \int\int \phi(x', z, a, u) f_{A|V_sZ_s}(a \mid F_{X_tZ_t}(x,z), z) f_{U_t}(u) da du \\
\overset{(\text{ACC}')}{=} \quad & \int\int \phi(x', z, a, u) f_{A|X_sZ_s}(a \mid F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), z) f_{U_t}(u) da du \\
\overset{(\text{ID})}{=} \quad & \int\int \phi(x', z, a, u) f_{A|X_sZ_s}(a \mid F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), z) f_{U_s}(u) da du \\
\overset{(\text{IND}')}{=} \quad & \int\int \phi(x', a, u) f_{AU_s|X_sZ_s}(a, u \mid F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), z) da du \\
= \quad & \mathbb{E}\left[\phi(x', A, U_s) \mid X_s = F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), Z_s = z\right]
\end{aligned}$$

Now, substitute $x' := F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z)$ to obtain

$$\begin{aligned}
&\mathbb{E}[\phi(F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), z, A, U_t) \mid X_t = x, Z_t = z] \\
&\qquad = \mathbb{E}[Y_s \mid X_s = F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), Z_t = z].
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\mathbb{E}[\phi(F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), z, A, U_t) - \phi(x, z, A, U_t) \mid X_t = x, Z_t = z] \\
&= \mathbb{E}[Y_s \mid X_s = F^{-1}_{X_sZ_s}(\cdot,z) \circ F_{X_tZ_t}(x,z), Z_s = z] - \mathbb{E}[Y_t \mid X_t = x, Z_t = z].
\end{aligned}$$

If $F_{X_s Z_s}(x,z) \neq F_{X_t X_t}(x,z)$, then Assumption 8 (ACC′) guarantees $F^{-1}_{X_s Z_s}(\cdot,z) \circ F_{X_t Z_t}(x,z) - x \neq 0$. Dividing both sides of the above equation by $F^{-1}_{X_s Z_s}(\cdot,z) \circ F_{X_t Z_t}(x,z) - x$ yields the desired result. $\square$

A.4 Proof of Lemma 4

Proof. First, note that $\mathbb{E}[\phi(x,z,A,U_t) \mid X_t = x, Z_t = z] = \mathbb{E}[Y_t \mid X_t = x, Z_t = z]$ holds for all $(x,z) \in \text{supp}(X_t, Z_t)$ for all $t \in \mathcal{T}(T)$. In order to prove the lemma, it remains to claim that

$$\begin{aligned}
&\mathbb{E}[\phi(F^{-1}_{X_s Z_s}(\cdot,z) \circ F_{X_t Z_t}(x,z), z, A, U_t) \mid X_t = x, Z_t = z] \\
&\quad = \mathbb{E}[Y_s \mid X_s = F^{-1}_{X_s Z_s}(\cdot,z) \circ F_{X_t Z_t}(x,z), Z_t = z]
\end{aligned}$$

holds for all $(x,z) \in \text{supp}(X_t, Z_t)$ for all $s,t \in \mathcal{T}(T)$. We use the notation $V_t := F_{X_t Z_t}(X_t, Z_t)$ for all $t \in \mathcal{T}(T)$. But then, Assumption 8 (IJD′) implies $(A, V_s, Z_s) \overset{\text{Law}}{=} (A, V_t, Z_t)$ for all $s,t \in \mathcal{T}(T)$. With Assumption 8, we obtain

$$\begin{aligned}
& \mathbb{E}[\phi(x',z,A,U_t) \mid X_t = x, Z_t = z] \\
= \quad & \int\int \phi(x',z,a,u) f_{AU_t|X_t Z_t}(a,u \mid x,z) dadu \\
\overset{(\text{IND}')}{=} \quad & \int\int \phi(x',z,a,u) f_{A|X_t Z_t}(a \mid x,z) f_{U_t}(u) dadu \\
\overset{(\text{ACC}')}{=} \quad & \int\int \phi(x',z,a,u) f_{A|V_t Z_t}(a \mid F_{X_t Z_t}(x,z), z) f_{U_t}(u) dadu \\
\overset{(\text{IND}')}{=} \quad & \int\int \phi(x',z,a,u) f_{A|V_t}(a \mid F_{X_t Z_t}(x,z)) f_{U_t}(u) dadu \\
\overset{(\text{IJD}')}{=} \quad & \int\int \phi(x',z,a,u) f_{A|V_s}(a \mid F_{X_t Z_t}(x,z)) f_{U_t}(u) dadu \\
\overset{(\text{IND}')}{=} \quad & \int\int \phi(x',z,a,u) f_{A|V_s Z_s}(a \mid F_{X_t Z_t}(x,z), z) f_{U_t}(u) dadu \\
\overset{(\text{ACC}')}{=} \quad & \int\int \phi(x',z,a,u) f_{A|X_s Z_s}(a \mid F^{-1}_{X_s Z_s}(\cdot,z) \circ F_{X_t Z_t}(x,z), z) f_{U_t}(u) dadu \\
\overset{(\text{ID})}{=} \quad & \int\int \phi(x',z,a,u) f_{A|X_s Z_s}(a \mid F^{-1}_{X_s Z_s}(\cdot,z) \circ F_{X_t Z_t}(x,z), z) f_{U_s}(u) dadu \\
\overset{(\text{IND}')}{=} \quad & \int\int \phi(x',a,u) f_{AU_s|X_s Z_s}(a,u \mid F^{-1}_{X_s Z_s}(\cdot,z) \circ F_{X_t Z_t}(x,z), z) dadu \\
= \quad & \mathbb{E}\left[\phi(x',A,U_s) \mid X_s = F^{-1}_{X_s Z_s}(\cdot,z) \circ F_{X_t Z_t}(x,z), Z_s = z\right]
\end{aligned}$$

Now, substitute $x' := F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z)$ to obtain

$$\begin{aligned}
&\mathbb{E}[\phi(F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z), z, A, U_t) \mid X_t = x, Z_t = z] \\
&\quad = \mathbb{E}[Y_s \mid X_s = F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z), Z_t = z].
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\mathbb{E}[\phi(F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z), z, A, U_t) - \phi(x, z, A, U_t) \mid X_t = x, Z_t = z] \\
&= \mathbb{E}[Y_s \mid X_s = F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z), Z_s = z] - \mathbb{E}[Y_t \mid X_t = x, Z_t = z].
\end{aligned}$$

If $F_{X_s Z_s}(x, z) \neq F_{X_t X_t}(x, z)$, then Assumption 8 (ACC$'$) guarantees $F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z) - x \neq 0$. Dividing both sides of the above equation by $F^{-1}_{X_s Z_s}(\cdot, z) \circ F_{X_t Z_t}(x, z) - x$ yields the desired result. $\square$

B Proofs of Main Results

B.1 Proof of Theorem 1

Proof. First, suppose that Assumption 2 holds with negative sign for $x \in \text{supp}(X_t)$. For notational convenience, we define the two sets of time indices

$$\begin{aligned}
\mathcal{T}^-(T, t, x) &:= \left\{ s \in \mathcal{T}(T) \,\middle|\, F^{-1}_{X_s} \circ F_{X_t}(x) < x,\ F^{-1}_{X_s} \circ F_{X_t}(x) \in I \right\} \\
\mathcal{T}^+(T, t, x) &:= \left\{ s \in \mathcal{T}(T) \,\middle|\, x < F^{-1}_{X_s} \circ F_{X_t}(x),\ F^{-1}_{X_s} \circ F_{X_t}(x) \in I \right\}
\end{aligned}$$

Given these notations, the time indices $\underline{\tau}(T, t, x)$ and $\overline{\tau}(T, t, x)$ are defined as

$$\begin{aligned}
\underline{\tau}(T, t, x) &:= \arg \max_{s \in \mathcal{T}^-(T,t,x)} F^{-1}_{X_s} \circ F_{X_t}(x) \\
\overline{\tau}(T, t, x) &:= \arg \min_{s \in \mathcal{T}^+(T,t,x)} F^{-1}_{X_s} \circ F_{X_t}(x)
\end{aligned}$$

By Assumption 2 with negative sign,

$$\frac{\phi(F^{-1}_{X_s} \circ F_{X_t}(x), A, U_t) - \phi(x, A, U_t)}{F^{-1}_{X_s} \circ F_{X_t}(x) - x} \leqslant \frac{\partial}{\partial x} \phi(x, A, U_t)$$

holds a.s. for all $s \in \mathcal{T}^+(T,t,x)$. Taking $\mathbb{E}[\cdot \mid X_t = x]$ yields

$$\begin{aligned} & \mathbb{E}\left[\left.\frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x), A, U_t) - \phi(x, A, U_t)}{F_{X_s}^{-1} \circ F_{X_t}(x) - x}\right| X_t = x\right] \\ \leqslant\ & \mathbb{E}\left[\left.\frac{\partial}{\partial x}\phi(x, A, U_t)\right| X_t = x\right] = \mathbb{E}\left[\left.\frac{\partial}{\partial x}\phi(x, A, U_t)\right| X_t = x\right] \end{aligned}$$

for all $s \in \mathcal{T}^+(T,t,x)$. But by Lemma 1 and Equation (2.3), it follows that

$$\Delta(s,t,x) \leqslant \mathbb{E}\left[\left.\frac{\partial}{\partial x}\phi(x, A, U_t)\right| X_t = x\right]$$

holds for all $s \in \mathcal{T}^+(T,t,x)$. Similarly, we can show that

$$\mathbb{E}\left[\left.\frac{\partial}{\partial x}\phi(x, A, U_t)\right| X_t = x\right] \leqslant \Delta(s,t,x)$$

holds for all $s \in \mathcal{T}^-(T,t,x)$. The intersection of all these identified regions yields the sharp interval

$$\max_{s \in \mathcal{T}^+(T,t,x)} \Delta(s,t,x) \leqslant \mathbb{E}\left[\left.\frac{\partial}{\partial x}\phi(x, A, U_t)\right| X_t = x\right] \leqslant \min_{s \in \mathcal{T}^-(T,t,x)} \Delta(s,t,x) \tag{B.1}$$

under Assumption 2 with negative sign.

Moreover, under Assumption 2 with negative sign, we have

$$\frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x), A, U_t) - \phi(x, A, U_t)}{F_{X_s}^{-1} \circ F_{X_t}(x) - x} \leqslant \frac{\phi(F_{X_{\overline{\tau}(T,t,x)}}^{-1} \circ F_{X_t}(x), A, U_t) - \phi(x, A, U_t)}{F_{X_{\overline{\tau}(T,t,x)}}^{-1} \circ F_{X_t}(x) - x}$$

a.s. for all $s \in \mathcal{T}^+(T,t,x)$. Taking $\mathbb{E}[\cdot \mid X_t = x]$ and using Lemma 1 and Equation (2.3), we obtain

$$\Delta(s,t,x) \leqslant \Delta(\overline{\tau}(T,t,x),t,x)$$

for all $s \in \mathcal{T}^+(T,t,x)$, thus

$$\max_{s \in \mathcal{T}^+(T,t,x)} \Delta(s,t,x) = \Delta(\overline{\tau}(T,t,x),t,x).$$

Similarly, we can derive

$$\min_{s \in \mathcal{T}^-(T,t,x)} \Delta(s,t,x) = \Delta(\underline{\tau}(T,t,x),t,x).$$

Therefore, the sharp interval (B.1) reduces to

$$\Delta(\overline{\tau}(T,t,x),t,x) \leqslant \mathbb{E}\left[\left.\frac{\partial}{\partial x}\phi(x, A, U_t)\right| X_t = x\right] \leqslant \Delta(\underline{\tau}(T,t,x),t,x)$$

under Assumption 2 with negative sign.

Now, suppose on the other hand that Assumption 2 holds with positive sign. Then, similar lines of argument yield sharp interval

$$\Delta(\underline{\tau}(T,t,x),t,x) \leqslant \mathbb{E}\left[\frac{\partial}{\partial x}\phi(x,A,U_t)\middle| X_t = x\right] \leqslant \Delta(\overline{\tau}(T,t,x),t,x)$$

under Assumption 2 with positive sign.

Treating the two cases of the signs for Assumption 2 together, we obtain the lower bound

$$L(t,x) \quad := \quad \min\left\{\Delta(\underline{\tau}(T,t,x),t,x), \Delta(\overline{\tau}(T,t,x),t,x)\right\}$$

and upper bound

$$U(t,x) \quad := \quad \max\left\{\Delta(\underline{\tau}(T,t,x),t,x), \Delta(\overline{\tau}(T,t,x),t,x)\right\}$$

such that

$$L(t,x) \leqslant \mathbb{E}\left[\frac{\partial}{\partial x}\phi(x,A,U_t)\middle| X_t = x\right] \leqslant U(t,x)$$

under both cases of the signs for Assumption 2. □

B.2 Proof of Theorem 2

Proof. First, we consider consequences of the regularity conditions in Assumption 4. Observe

$$\begin{aligned}
&\int\int\left|\frac{\phi(F_{X_s}^{-1}\circ F_{X_t}(x),a,u)-\phi(x,a,u)}{F_{X_s}^{-1}\circ F_{X_t}(x)-x}\left(f_{AU|X_t}(a,u\mid x)-f_{AU|X_t}(a,u\mid x^*)\right)\right|dadu \\
\overset{\text{A.1 (IND)}}{=}\;&\int\int\left|\frac{\phi(F_{X_s}^{-1}\circ F_{X_t}(x),a,u)-\phi(x,a,u)}{F_{X_s}^{-1}\circ F_{X_t}(x)-x}\left(f_{A|X_t}(a\mid x)-f_{A|X_t}(a\mid x^*)\right)f_U(u)\right|dadu \\
\overset{\text{Hölder}}{\leqslant}\;&\left\|\frac{\phi(F_{X_s}^{-1}\circ F_{X_t}(x),\cdot,\cdot)-\phi(x,\cdot,\cdot)}{F_{X_s}^{-1}\circ F_{X_t}(x)-x}\right\|_\infty\cdot\left\|f_{A|X_t}(a\mid x)-f_{A|X_t}(a\mid x^*)\right\|_\infty\cdot\|f_U\|_1 \\
\overset{A.4(i)(ii)}{=}\;&C\left\|f_{A|X_t}(a\mid x)-f_{A|X_t}(a\mid x^*)\right\|_\infty \overset{A.4(iii)}{\longrightarrow} 0
\end{aligned}$$

as $x \to x^*$. This implies that

$$\begin{aligned}
\int\int\frac{\phi(F_{X_s}^{-1}\circ F_{X_t}(x),a,u)-\phi(x,a,u)}{F_{X_s}^{-1}\circ F_{X_t}(x)-x}\left(f_{AU|X_t}(a,u\mid x)-f_{AU|X_t}(a,u\mid x^*)\right)dadu \\
=: R(s,t,x,x^*) \longrightarrow 0 \qquad \text{as } x^* \to x \qquad \text{(B.2)}
\end{aligned}$$

Moreover, Assumption 4 yields

$$\begin{aligned}
&\left\| \frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x), \cdot, \cdot) - \phi(x, \cdot, \cdot)}{F_{X_s}^{-1} \circ F_{X_t}(x) - x} f_{AU|X_t}(\cdot, \cdot \mid x^*) \right\|_1 \\
\overset{\text{Hölder}}{\leqslant} \quad &\left\| \frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x), \cdot, \cdot) - \phi(x, \cdot, \cdot)}{F_{X_s}^{-1} \circ F_{X_t}(x) - x} \right\|_\infty \cdot \left\| f_{AU|X_t}(\cdot, \cdot \mid x^*) \right\|_1 \overset{A.4(i)(ii)}{<} \infty,
\end{aligned}$$

hence satisfying the condition of the Lebesgue Dominated Convergence Theorem to yield

$$\mathbb{E}\left[\frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x), A, U_t) - \phi(x, A, U_t)}{F_{X_s}^{-1} \circ F_{X_t}(x) - x} \middle| X_t = x^* \right] \longrightarrow \mathbb{E}\left[\frac{\partial}{\partial x} \phi(x^*, A, U_t) \middle| X_t = x^* \right] \quad \text{(B.3)}$$

as $x \to x^*$.

By Assumptions 1 (ACC) and 3, we have a sequence $x_j \to x^*$ such that $F_{X_s}^{-1} \circ F_{X_t}(x_j) - x_j \neq 0$ for all j, but $F_{X_s}^{-1} \circ F_{X_t}(x_j) - x_j \to 0$ as $x_j \to x^*$. Therefore

$$\begin{aligned}
\Delta(s, t, x_j) \quad &\overset{(2.3)}{=} \quad \frac{\mathbb{E}\left[Y_s \mid X_s = F_{X_s}^{-1} \circ F_{X_t}(x_j) \right] - \mathbb{E}\left[Y_t \mid X_t = x_j \right]}{F_{X_s}^{-1} \circ F_{X_t}(x_j) - x_j} \\
&\overset{\text{Lemma 1}}{=} \quad \mathbb{E}\left[\frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x_j), A, U_t) - \phi(x_j, A, U_t)}{F_{X_s}^{-1} \circ F_{X_t}(x_j) - x_j} \middle| X_t = x_j \right] \\
&= \quad \mathbb{E}\left[\frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x_j), A, U_t) - \phi(x_j, A, U_t)}{F_{X_s}^{-1} \circ F_{X_t}(x_j) - x_j} \middle| X_t = x^* \right] + R(s, t, x_j, x^*) \\
&\overset{\text{(B.2) (B.3)}}{\to} \quad \mathbb{E}\left[\frac{\partial}{\partial x} \phi(x^*, A, U_t) \middle| X_t = x^* \right] \quad \text{as} \quad x_j \to x^*.
\end{aligned}$$

□

B.3 Proof of Theorem 3

Proof. As in the proof of Theorem 1, we use the notations

$$\begin{aligned}
\mathcal{T}^-(T, t, x) \quad &:= \quad \left\{ s \in \mathcal{T}(T) \,\middle|\, F_{X_s}^{-1} \circ F_{X_t}(x) < x \right\} \\
\mathcal{T}^+(T, t, x) \quad &:= \quad \left\{ s \in \mathcal{T}(T) \,\middle|\, x < F_{X_s}^{-1} \circ F_{X_t}(x) \right\}
\end{aligned}$$

Given these notations, the time indices $\underline{\tau}(T, t, x)$ and $\overline{\tau}(T, t, x)$ are defined as

$$\begin{aligned}
\underline{\tau}(T, t, x) \quad &:= \quad \arg \max_{s \in \mathcal{T}^-(T,t,x)} F_{X_s}^{-1} \circ F_{X_t}(x) \\
\overline{\tau}(T, t, x) \quad &:= \quad \arg \min_{s \in \mathcal{T}^+(T,t,x)} F_{X_s}^{-1} \circ F_{X_t}(x)
\end{aligned}$$

Note that the 'max' and 'min' are defined since $\mathcal{T}^-(T,t,x)$ and $\mathcal{T}^+(T,t,x)$ are finite for $T<\infty$.

By Assumption 5, there exists a subsequence $\{t_j\}_{j=1}^\infty \subset \mathcal{T}(\infty)$ such that $F^{-1}_{X_{t_j}} \circ F_{X_t}(x) \neq x$ for all j and $F^{-1}_{X_{t_j}} \circ F_{X_t}(x) \to x$ as $j \to \infty$. By the first property ($F^{-1}_{X_{t_j}} \circ F_{X_t}(x) \neq x$ for all j), we have $\{t_j\}_{j=1}^\infty \subset \mathcal{T}^-(\infty,t,x) \cup \mathcal{T}^+(\infty,t,x)$. This implies that at least one of the following two cardinal equalities is true:

$$\begin{array}{lll} (I') & \left|\{t_j\}_{j=1}^\infty \cap \mathcal{T}^-(\infty,t,x)\right| = \infty & \text{and/or} \\ (II') & \left|\{t_j\}_{j=1}^\infty \cap \mathcal{T}^+(\infty,t,x)\right| = \infty. & \end{array}$$

If (I′) is the case, then consider the set $\mathcal{S}^-(T,t,x) := \{t_j\}_{j=1}^\infty \cap \mathcal{T}^-(T,t,x)$. Since $\mathcal{S}^-(\infty,t,x)$ is a sub-subsequence of the subsequence $\{t_j\}_{j=1}^\infty$, the fact that $F^{-1}_{X_{t_j}} \circ F_{X_t}(x) \to x$ as $j \to \infty$ implies that $F^{-1}_{X_s} \circ F_{X_t}(x) \to x$ as $\mathcal{S}^-(\infty,t,x) \ni s \to \infty$. This in turn implies that

$$\max_{s \in \mathcal{S}^-(T,t,x)} F^{-1}_{X_s} \circ F_{X_t}(x) \to x \qquad \text{as } T \to \infty.$$

But by definitions of $\underline{\tau}(T,t,x)$, $\mathcal{S}^-(T,t,x)$, and $\mathcal{T}^-(T,t,x)$, we have

$$\max_{s \in \mathcal{S}^-(T,t,x)} F^{-1}_{X_s} \circ F_{X_t}(x) \leqslant F^{-1}_{X_{\underline{\tau}(T,t,x)}} \circ F_{X_t}(x) \leqslant x$$

for all T. Thus, it follows from the Squeeze Theorem that

$$(I) \qquad F^{-1}_{X_{\underline{\tau}(T,t,x)}} \circ F_{X_t}(x) \to x \qquad \text{as } T \to \infty.$$

If, on the other hand (II′), is the case, then similar lines of argument will show

$$(II) \qquad F^{-1}_{X_{\overline{\tau}(T,t,x)}} \circ F_{X_t}(x) \to x \qquad \text{as } T \to \infty.$$

The rest of the proof follows from Lemmata 1 and 2. Specifically, if (I) is the case, then

$$\begin{aligned} \Delta(\underline{\tau}(T,t,x),t,x) \quad & \overset{(2.3)}{=} \quad \frac{\mathbb{E}\left[Y_{\underline{\tau}(T,t,x)} \mid X_{\underline{\tau}(T,t,x)} = F^{-1}_{X_{\underline{\tau}(T,t,x)}} \circ F_{X_t}(x)\right] - \mathbb{E}\left[Y_t \mid X_t = x\right]}{F^{-1}_{X_{\underline{\tau}(T,t,x)}} \circ F_{X_t}(x) - x} \\ & \overset{\text{Lemma 1}}{=} \quad \mathbb{E}\left[\left.\frac{\phi(F^{-1}_{X_{\underline{\tau}(T,t,x)}} \circ F_{X_t}(x), A, U_t) - \phi(x,A,U_t)}{F^{-1}_{X_{\underline{\tau}(T,t,x)}} \circ F_{X_t}(x) - x}\right| X_t = x\right] \\ & \overset{\text{Lemma 2}}{\to} \quad \mathbb{E}\left[\left.\frac{\partial}{\partial x}\phi(x,A,U_t)\right| X_t = x\right] \quad \text{as} \quad T \to \infty. \end{aligned}$$

Similarly, if (II) is the case, then

$$\Delta(\overline{\tau}(T,t,x),t,x) \to \mathbb{E}\left[\frac{\partial}{\partial x}\phi(x,A,U_t) \middle| X_t = x \right] \quad \text{as} \quad T \to \infty.$$

by analogous steps. □

B.4 Proofs of Theorems 4, 5, and 6

Proof. Theorems 4, 5, and 6 can be proven by adapting the proofs of of Theorems 1, 2, and 3, respectively, by using Lemma 4 instead of Lemma 1. □

B.5 Proof of Corollary 1

Proof. Under the linear random coefficient model

$$\phi(X_t, A, U_t) = \alpha(A, U_t) + \beta(A, U_t)X_t,$$

Lemma 1 obviously reduces to

$$\Delta(s,t,x) = \mathbb{E}\left[\frac{\phi(F_{X_s}^{-1} \circ F_{X_t}(x), A, U_t) - \phi(x, A, U_t)}{F_{X_s}^{-1} \circ F_{X_t}(x) - x} \middle| X_t = x \right] = \mathbb{E}\left[\beta(A,U_t) \middle| X_t = x \right]$$

for all $x \in \text{supp}(X_t)$ for all $s, t \in \mathcal{T}(T)$ such that $F_{X_s}(x) \neq F_{X_t}(x)$. □

B.6 Proof of Corollary 2

Proof. Under the equation (5.3), Assumption 14 implies Assumption 8, and Assumption 15 implies Assumption 11 with ϕ replaced by ψ. Therefore, by Theorem 5, we have

$$\begin{aligned} \lim_{x \to x^*} \tilde{\Delta}_x(s,t,x,z^*) &= \mathbb{E}\left[\frac{\partial}{\partial x}\phi(x^*, z^*, A, U_t) \middle| X_t = x^*, Z_t = z^* \right] \\ &= \mathbb{E}\left[f_{-U_t^*}(x^*\beta + z^*\gamma + A) \middle| X_t = x^*, Z_t = z^* \right] \cdot \beta \end{aligned}$$

and

$$\begin{aligned} \lim_{z \to z^*} \tilde{\Delta}_z(s,t,x^*,z) &= \mathbb{E}\left[\frac{\partial}{\partial z}\phi(x^*, z^*, A, U_t) \middle| X_t = x^*, Z_t = z^* \right] \\ &= \mathbb{E}\left[f_{-U_t^*}(x^*\beta + z^*\gamma + A) \middle| X_t = x^*, Z_t = z^* \right] \cdot \gamma. \end{aligned}$$

Taking the ratio of these two equalities yields

$$\frac{\lim_{x \to x^*} \tilde{\Delta}_x(s,t,x,z^*)}{\lim_{z \to z^*} \tilde{\Delta}_z(s,t,x^*,z)} = \frac{\beta}{\gamma}.$$

$\square$

B.7 Proof of Corollary 3

Proof. Under the equation (5.3), Assumption 14 implies Assumption 8, and Assumption 16 implies Assumption 13 with ϕ replaced by ψ. Therefore, by Theorem 6, at least one of

$$F^{-1}_{X_{\underline{\tau}_x(T,t,x,z)} Z_{\underline{\tau}_x(T,t,x,z)}}(\cdot, z) \circ F_{X_t Z_t}(x,z) \quad \text{or} \quad F^{-1}_{X_{\overline{\tau}_x(T,t,x,z)} Z_{\overline{\tau}_x(T,t,x,z)}}(\cdot, z) \circ F_{X_t Z_t}(x,z)$$

converges to x as $T \to \infty$. Similarly, at least one of

$${}^{-1}_{X_{\underline{\tau}_z(T,t,x,z)} Z_{\underline{\tau}_z(T,t,x,z)}}(x, \cdot) \circ F_{X_t Z_t}(x,z) \quad \text{or} \quad F^{-1}_{X_{\overline{\tau}_z(T,t,x,z)} Z_{\overline{\tau}_z(T,t,x,z)}}(x, \cdot) \circ F_{X_t Z_t}(x,z)$$

converges to z as $T \to \infty$. Therefore, at least one of the cases (I)–(IV) holds. If case (I) holds, then by Theorem 6, we have

$$\begin{aligned} \tilde{\Delta}_x(\underline{\tau}_x(T,t,x,z),t,x,z) &\to \mathbb{E}\left[\frac{\partial}{\partial x}\phi(x,z,A,U_t) \middle| X_t = x, Z_t = z \right] \\ &= \mathbb{E}\left[f_{-U_t^*}(x\beta + z\gamma + A) \middle| X_t = x, Z_t = z \right] \cdot \beta \end{aligned}$$

and

$$\begin{aligned} \tilde{\Delta}_z(\underline{\tau}_z(T,t,x,z),t,x,z) &\to \mathbb{E}\left[\frac{\partial}{\partial z}\phi(x,z,A,U_t) \middle| X_t = x, Z_t = z \right] \\ &= \mathbb{E}\left[f_{-U_t^*}(x\beta + z\gamma + A) \middle| X_t = x, Z_t = z \right] \cdot \gamma. \end{aligned}$$

Taking the ratio of these convergence results, we obtain

$$\tilde{\Delta}_x(\underline{\tau}_x(T,t,x,z),t,x,z) / \tilde{\Delta}_z(\underline{\tau}_z(T,t,x,z),t,x,z) \to \beta/\gamma$$

as $T \to \infty$. Similar arguments show the conclusions for cases (II)–(IV). $\square$

References

Altonji, Joseph G. and Rosa L. Matzkin (2005) "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, Vol. 73, No. 4, pp. 1053-1102.

Arellano, Manuel and Stéphane Bonhomme (2010) "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," Working Paper.

Arellano, Manuel and Jinyong Hahn (2006) "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," Working Paper.

Bester, Alan C. and Christian Hansen (2009) "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," *Journal of Business and Economic Statistics,* Vol. 27, No. 2, pp.235–250.

Blundell, Richard W. and James L. Powell (2003) "Endogeneity in Nonparametric and Semiparametric Regression Models," *Advances in Economics and Econometrics: Theory and Applications II,* M. Dewatripont, L.P. Hansen, S. J. Turnovsky, Eds., pp. 312–357.

Carroll, Christopher D. and Miles S. Kimball (1996) "On the Concavity of the Consumption Function," *Econometrica,* Vol. 64, No. 4, pp. 981–992.

Chamberlain, Gary (1984) "Panel Data," *Handbook of Econometrics,* Z. Griliches & M.D. Intriligator, Eds., Vol. 2, pp. 1247–1318.

Chamberlain, Gary (2010) "Binary Response Models for Panel Data: Identification and Information," *Econometrica,* Vol. 78, Np. 1, pp.159–168.

Chernozhukov, Victor, Iv́an Fernández-Val, Jinyong Hahn, and Whitney Newey (2010) "Average and Quantile Effects in Nonseparable Panel Models," MIT Working Paper.

Chesher, Andrew (2005) "Nonparametric Identification under Discrete Variation," *Econometrica*, Vol. 73, No. 5, pp. 1525–1550.

Evdokimov, Kirill (2010) “Identifcation and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity” Working Paper.

Graham, Bryan S. and James Powell (2010) “Identification and Estimation of ‘Irregular’ Correlated Random Coefficient Models,” Working Paper.

Hoderlein, Stefan and Yuya Sasaki (2010) “Continuous Treatments” Working Paper.

Hoderlein, Stefan and Halbert White (2010) “Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects” Working Paper.

Honoré, Bo E. and Ekaterini Kyriazidou (2000) “Panel Data Discrete Choice Models with Lagged Dependent Variables,” *Econometrica,* Vol. 68, No. 4, pp. 839-4.

Imbens, Guido W. and Whitney K. Newey (2009) “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity” *Econometrica*, Vol. 77, No. 5, pp 1481-1512

Lee, Myoung-Jae (1999) “A Root-N Consistent Semiparametric Estimator for Related-Effect Binary Response Panel Data,” *Econometrica,* Vol. 67, No. 2, pp. 427–433.

Manski, Charles F. (1987) “Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data,” *Econometrica,* Vol. 55, No. 2, pp. 357–362.

Shaikh, Azeem and Edward Vytlacil (2010) “Partial Identification in Triangular Systems of Equations with Binary Dependent Variables,” Working Paper.

Torgovitsky, A (2010) “Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity,” Working Paper.

Wooldridge, Jeffrey M. (2005) “Unobserved Heterogeneity and Estimation of Average Partial Effects,” *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,* D.W.K. Andrews & J.H. Stock, Eds., pp. 27–55.