Essays in Econometrics of Heterogeneous Agents

by

Yuya Sasaki

B. S., Utah State University, 2002

M. S., Utah State University, 2007

M. A., Brown University, 2008

A Dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Economics at Brown University

PROVIDENCE, RHODE ISLAND

MAY 2012

This dissertation by Yuya Sasaki is accepted in its present form
by the Department of Economics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Frank Kleibergen, Advisor

Recommended to the Graduate Council

Date _____ _____

Eric Renault, Reader

Date _____ _____

Susanne Schennach, Reader

Approved by the Graduate Council

Date _____ _____

Peter M. Weber, Dean of the Graduate School

Vitae

The author was born on June 22nd, 1979 in Tokyo, Japan. He received B.S. from Utah State University in 2002, and M.S. from Utah State University in 2007. He entered Brown University in 2007, and received M.A. in 2008 and Ph.D. in 2012.

Acknowledgements

I am indebted to my advisor, Frank Kleibergen, for continuous support and guidance. I also thank Stefan Hoderlein, Blaise Melly, Eric Renault, and Susanne Schennach for constructive advice throughout my dissertation. Ken Chay, Kaivan Munshi, and Anna Aizer helped me to know how to produce econometric methods which may be of practical use. I benefited from discussion with my colleagues, Toru Kitagawa, Zhaoguo Zhan, Alexei Abrahams, Philipp Ketz, Maria Jose Boccardi, Andrew Elzinga, Bruno Gasperini, and Daniela Scida. I am thankful to my family members for their patience and support.

Contents

List of Tables

List of Figures

Abstract of "Essays in Econometrics of Heterogeneous Agents"

by Yuya Sasaki, Ph.D., Brown University, May 2012

Economic models often involve non-separability between observed and unobserved heterogeneous characteristics of economic agents. This dissertation presents methods of identification, estimation, and inference of nonparametric and nonseparable economic models for cross section and panel data. The first chapter discusses identification and estimation of nonseparable dynamic panel data with non-random dynamic selection. It shows that nonseparable dynamic panel models with endogenous attrition can be identified from six time periods of unbalanced panel data. The principle of constrained maximum likelihood is proposed for consistent estimation. The second chapter discusses identification of average structural partial effects for endogenous nonseparable cross-section models without assuming monotonicity. Nonparametric identification methods are proposed for various first-stage structural and reduced-form assumptions. The third chapter discusses statistical methods of model tests for endogenous nonseparable cross-section models when instruments exhibit discrete variations and the outcome structure is not monotone with respect to unobserved heterogeneity. It shows that the testing method possesses sufficient power even if instruments are discrete and exert only local effects on endogenous choice.

CHAPTER 1

Heterogeneity and Selection in Dynamic Panel Data

1. Introduction

Dynamics, nonseparable heterogeneity, and selection have been separately treated in the panel data literature, in spite of their joint relevance to a wide array of applications. First, common economic variables of interest are modeled to follow dynamics, e.g., assets, income, physical capital and human capital. Second, many economic models entail nonseparable heterogeneity, i.e., an additively separable residual does not summarize abilities, preferences and technologies. Third, most empirical panel data are unbalanced by (self-) selection. Indeed, consideration of these three issues – dynamics, nonseparable heterogeneity, and selection – is essential, but existing econometric methods do not handle them at the same time.

To fill this gap, this paper proposes a set of conditions for identification of dynamic panel data models in the presence of both nonseparable heterogeneity and dynamic selection.[1] Nonparametric point identification is achieved by using information involving either a proxy variable or a slightly longer panel. Specifically, the model is point-identified using $T = 3$ periods of unbalanced panel data and a proxy variable. A special case of this identification result occurs by constructing the proxy variable from three additional periods, i.e., $T = 6$ in total.

[1] In the introductory section of his monograph, Hsiao (2003) particularly picks up heterogeneity and selection as the two major sources of bias in panel data analysis, which motivates the goal of this paper.

For example, consider the dynamics of socio-economic status (SES) and its causal effects on adult mortality.[2] Many unobserved individual characteristics such as genetics, patience and innate abilities presumably affect both SES and survival in non-additive ways, which would incur a bias unless explicitly accounted for. Furthermore, a death outcome of the survival selection induces subsequently missing observations, which may result in a selection bias, often called survivorship bias. The following dynamic panel model with selection accommodates this example:

$$\begin{cases} Y_t = g(Y_{t-1}, U, \mathcal{E}_t) & t = 2, \cdots, T \qquad \text{(Dynamic Panel Model)} \\ D_t = h(Y_t, U, V_t) & t = 1, \cdots, T-1 \qquad \text{(Selection Model)} \\ F_{Y_1 U} & \qquad\qquad\qquad\qquad\quad \text{(Initial Condition)} \end{cases}$$

The first equation models the dynamics of the observed state variable Y_t, such as SES, as a first-order Markov process with unobserved heterogeneity U. The second equation models a binary choice of the selection variable, D_t, such as survival, as a Markov decision process with unobserved heterogeneity U. The initial condition $F_{Y_1 U}$ models the dependence of the initial state Y_1 on unobserved heterogeneity U.[3] The period-specific shocks $(\mathcal{E}_t, V_t)$ are exogenous, leaving the fixed effect U as the only source of endogeneity. The economic agent drops out of the panel upon $D_t = 0$, as in the case of death. Consequently, data is observed in the following manner: (Y_2, D_2) is observed if $D_1 = 1$; (Y_3, D_3) is observed if $D_1 = D_2 = 1$; and so on. Heckman and Navarro (2007) introduced this formulation of dynamic selection.

[2] Among many biological and socioeconomic factors of mortality (Cutler, Deaton, and Lleras-Muney, 2006), the role of SES and economic environments has been investigated by a number of empirical researches (e.g., Ruhm, 2000; Deaton and Paxson, 2001; Snyder and Evans, 2006; Sullivan and von Wachter, 2009a,b).

[3] The distribution $F_{Y_1 U}$ features the *initial conditions problem* for dynamic panel data models. See Wooldridge (2005) and Honoré and Tamer (2006) for discussions on the initial conditions problem in the contexts of nonlinear and binary outcome models. Blundell and Bond (1998) and Hahn (1999) use semiparametric distributions to obtain identifying restrictions and efficiency gain. In applications, the initial condition $F_{Y_1 U}$ together with the function g are important to disentangle spurious state dependence of a long-run outcome (Heckman, 1981a,b).

How can we nonparametrically point identify the nonseparable functions (g, h) and the initial condition F_{Y_1U} under this setup of endogenously unbalanced panel data? Common ways to handle selection include matching and weighting. These approaches, however, presume selection on observables, parametric models, and additively separable models, none of which is assumed in this paper. Even without selection, the standard panel data techniques such as first differencing, demeaning, projection, and moment restrictions do not generally work for nonseparable and nonparametric models.

The literature on nonseparable cross section models proposes constructing auxiliary variables, such as a proxy variable or a control variable, to remove endogeneity (e.g., Garen, 1984; Imbens and Newey, 2009).[4] Likewise, Altonji and Matzkin (2005) show that a control variable can be also constructed from panel data for sibling and neighborhood panels. This paper complements Altonji and Matzkin along two dimensions. First, we show that a proxy variable can be constructed from dynamic panel data, similar to their construction of a control variable from sibling and neighborhood panel data. Second, the proxy variable, akin to a control variable,[5] handles not only nonseparable heterogeneity, but also dynamic selection. We propose a method of using the proxy variable to *nonparametrically difference out* both nonseparable heterogeneity and dynamic selection at the same time.

The nonparametric differencing relies on a nonclassical proxy variable, which we define as a noisy signal of true unobserved heterogeneity with a nonseparable noise. This definition is reminiscent of nonclassical measurement errors (ME).[6] A natural

[4] Chesher (2003) can be also viewed as a control variable method, cf. Imbens and Newey (2009; Theorem 2).

[5] Proxy and control variables are similar in that both of them are correlated with unobserved factors. But they differ in terms of independence conditions: if X denotes an endogenous regressor and U denotes unobserved factors, then a proxy variable Z and a control variable Z' satisfy $Z \perp\!\!\!\perp X \mid U$ and $U \perp\!\!\!\perp X \mid Z'$, respectively.

[6] See Lewbel (2006), Mahajan (2006), Schennach (2007), Hu (2008), Hu and Schennach (2008), and Schennach, Song, and White (2011) for the literature on the nonclassical ME.

approach to identification, therefore, is to adapt the methods used in the nonclassical ME literature to the current context. This paper follows the spectral decomposition approach (e.g., Hu, 2008; Hu and Schennach, 2008) to nonparametrically identify mixture components.[7]

The identification procedure is outlined as follows. First, the method of nonparametric differencing removes the influence of nonseparable heterogeneity and selection. After removing these two sources of bias, the spectral decomposition identifies the mixture component $f_{Y_t|Y_{t-1}U}$, which in turn represents the observational equivalence class of the true nonseparable function g by normalizing the distribution of the exogenous error $\mathcal{E}_t$, following Matzkin (2003, 2007). This sequence yields nonparametric point identification of g from short unbalanced panel data. The selection function h can be similarly identified by a few additional steps of the spectral decomposition and solving integral equations[8] to identify representing mixture components.

2. Background

Selection is of natural interest in panel data analysis because attrition is an issue in most, if not all, panel data sets. While many applications focus on the dynamic model g as the object of primary interest, the selection function h also helps to explain important causal effects in a variety of economic problems. In the SES and mortality example, identification of the survival selection function h allows us to learn about the causal effects of SES on mortality. Generally, the selection function h can be used to model hazards of panel attrition. Examples include (i) school dropout (Cameron and Heckman, 1998; Eckstein and Wolpin, 1999; Belzil and Hansen, 2002; Heckman and Navarro, 2007); (ii) retirement from a job (Stock and Wise, 1990; Rust and Phelan, 1997; Karlstrom, Palme, and Svensson, 2004; French, 2005; Aguirregabiria, 2010; French and Jones, 2011); (iii) replacement of depreciated capital (Rust, 1987)

[7] See Henry, Kitamura, and Salanié (2010) for general identification results for mixture models.

[8] Precisely, they are the Fredholm equations of the first kind. See Carrasco, Florens, and Renault (2007).

and replacement of managers (Brown, Goetzmann, Ibbotson, and Ross, 1992); (iv) sterilization (Hotz and Miller, 1993); (v) exit from markets (Aguirregabiria and Mira, 2007; Pakes, Ostrovsky, and Berry, 2007); (vi) recovery from a disease (Crawford and Shum, 2005); and (vii) death (Contoyannis, Jones, and Rice, 2004; Halliday, 2008). Examples (i)–(v) are particularly motivated by rational hazards formulated in the following structural framework.

EXAMPLE 1 (Optimal Stopping as a Rational Choice of Hazard). Suppose that an economic agent knows her current utility or profit as a function π of state y_t and heterogeneity u. Let v_t^d denote a selection-specific private shock for each choice $d \in \{0, 1\}$, which is known to the agent. She also knows her exit value as a function $\overline{\nu}$ of state y_t and heterogeneity u. Using the dynamic function g, define the value function ν as the fixed point of the Bellman equation

$$\nu(y_t, u) = \mathrm{E}[\max\{\pi(y_t, u) + V_t^1 + \beta \mathrm{E}[\nu(g(y_t, u, \mathcal{E}_{t+1}), u)], \quad \pi(y_t, u) + V_t^0 + \beta\overline{\nu}(y_t, u)\}],$$

where β denotes the rate of time preference. The reduced-form self-selection function h is then defined by

$$h(y_t, u, v_t) := \mathbb{1}\{\underbrace{\beta \mathrm{E}[\nu(g(y_t, u, \mathcal{E}_{t+1}), u)]}_{\text{Continuation value}} - \underbrace{\beta\overline{\nu}(y_t, u)}_{\text{Exit Value}} \geqslant \underbrace{v_t^0 - v_t^1}_{\equiv v_t}\}.$$

The agent decides to exit at time t if $h(Y_t, U, V_t) = 0$. Identification of the reduced form h is important in many applications.[9] Moreover, the reduced form h also reveals the *heterogeneous* conditional choice probability (CCP), $f_{D_t|Y_t U}$, which in turn can be used to recover heterogeneous structural primitives by using the method of Hotz and Miller (1993).[10] □

[9] Counterfactual policy analysis is often possible with reduced-form selection function as a sufficient statistic; see the Marschak's (1953) maxim discussed by Heckman (2000) and Heckman and Vytlacil (2007).

[10] I keep identification of the primitives out of the scope of this paper. Primitives are known to be generally under-identified without additional restrictions (Rust, 1994; Magnac and Thesmar, 2002; Pesendorfer and Schmidt-Dengler, 2008). These features may be more generally treated in the literature of set identification and set inference, e.g., Bajari, Benkard, and Levin (2007) and the follow-up literature.

As this example suggests, nonparametric identification of the heterogeneous CCP follows as a byproduct of our identification results,[11] showing a connection between this paper and the literature on structural dynamic discrete choice models. When attrition, $D_t = 0$, is associated with hazards or ends of some duration, our identification results also entail nonparametric identification of the mixed hazard model and the distribution of unobserved heterogeneity.[12] In this sense, our objective is also related to the literature on duration analysis (e.g., Lancaster, 1979; Elbers and Ridder, 1982; Heckman and Singer, 1984; Honoré, 1990; Ridder, 1990; Horowitz, 1999; Ridder and Woutersen, 2003).

The paper covers three econometric topics, (A) panel data, (B) selection/missing data, and (C) nonseparable models. To show the place of this paper, I briefly discuss these related branches of the literature. Because the field is extensive, the following list is not exhaustive.

(A) and (B): panel data with selection has been discussed from the perspective of (i) a selection model (Hausman and Wise, 1979; Das, 2004), (ii) variance adjustment (Baltagi, 1985; Baltagi and Chang, 1994), (iii) additional data such as refreshment samples (Ridder, 1992; Hirano, Imbens, Ridder, and Rubin, 2001; Bhattacharya, 2008), (iv) matching (Kyriazidou, 1997), (v) weighting (Hellerstein and Imbens, 1999; Moffitt, Fitzgerald, and Gottschalk, 1999; Wooldridge, 2002), and (vi) partial identification (Khan, Ponomareva, and Tamer, 2011). We contribute to this literature by allowing nonseparability in addition to selection/missing data.

Identification of CCP under finite heterogeneous types has been discussed by Magnac and Thesmar (2002) and Kasahara and Shimotsu (2009). Aguirregabiria and Mira (2007) considered market-level unobserved heterogeneity as a variant of their main model. While we focus on identification of heterogeneous CCP, Arcidiacono and Miller (2011) suggested a method of estimating heterogeneous CCP.

[11] Taking the expectation of $h(y, u, \cdot)$ with respect to the distribution of the exogenous error V_t yields the heterogeneous CCP, $f_{D_t|Y_tU}(1 \mid y, u)$ for each (y, u). The heterogeneous CCP is also identified by Kasahara and Shimotsu (2009), which this paper complements by introducing missing observations in data.

[12] The nonparametric mixed hazard model and the marginal distribution F_U of unobserved heterogeneity follow from the identified survival selection function h and the initial condition F_{Y_1U}, respectively.

(A) and (C): nonseparable panel models have been treated with (i) random coefficients and interactive fixed effects (Hsiao, 1975; Pesaran and Smith, 1995; Hsiao and Pesaran, 2004; Graham annd Powell, 2008; Arellano and Bonhomme, 2009; Bai, 2009). (ii) bias reduction (discussed in the extensive body of literature surveyed by Arellano and Hahn, 2005), (iii) identification of local partial effects (Altonji and Matzkin, 2005; Altonji, Ichimura, and Otsu, 2011; Graham and Powell, 2008; Arellano and Bonhomme, 2009; Bester and Hansen, 2009; Chernozhukov, Fernández-Val, Hahn, and Newey, 2009; Hoderlein and White, 2009), (iv) partial identification (Honoré and Tamer, 2006; Chernozhukov, Fernández-Val, Hahn, and Newey, 2010), (v) partial separability (Evdokimov, 2009), and (vi) assumptions of surjective and/or injective operators (Kasahara and Shimotsu, 2009; Bonhomme, 2010; Hu and Shum, 2010; Shiu and Hu, 2011). The paper contributes to this literature by introducing selection/missing data in addition to allowing nonseparability.

Identification of a nonseparable dynamic panel data model is studied by Shiu and Hu (2011) who use independently evolving covariates as auxiliary variables, similar to one of the two identification results of this paper using a proxy as an auxiliary variable. This paper complements Shiu and Hu along two dimensions. First, our identification result using $T = 6$ periods eliminates the need to assume the independently evolving covariates or any other auxiliary variable. Second, we can allow for selection/missing data in addition to dynamics and nonseparability.

3. An Overview

We start out with an informal overview of the identification strategy in this section, followed by formal identification results summarized in Section 4.

Briefly described, the nonparametric differencing method works in the following manner. Let z denote a proxy variable. Observed data A_z, which are contaminated by mixed heterogeneity and selection, can be decomposed as $A_z = B_z C$, where B_z contains model information and C contains the two sources of bias, i.e., heterogeneity and selection. The contaminant holder, C, does not depend on z by an exclusion

restriction. Thus, using two values of z, say $z = 0, 1$, selectively eliminates C by the operator composition $A_1 A_0^{-1} = B_1 C C^{-1} B_0^{-1} = B_1 B_0^{-1}$, without losing the model information B_z. This shows how heterogeneity and selection contained in C are nonparametrically differenced out, and is analogous to the familiar first differencing method which eliminates fixed effects by using two values of t instead of two values of z.

Section 3.1 sketches the identification strategy using $T = 3$ periods of panel data and a proxy variable. An intuition is the following. First, using variations in Y_1 in the equation $y_2 = g(Y_1, U, \mathcal{E}_2)$ involving the first two periods, $t = 1, 2$, we can retrieve information about $(U, \mathcal{E}_2)$ associated with $Y_2 = y_2$. This is comparable to the first stage in the cross section context except for the endogeneity of the first-stage regressor Y_1. The proxy variable, which is correlated with Y_1 only through U, disentangles U and Y_1 to fix the endogeneity. We then use this knowledge about U to identify the heterogeneous dynamic through the equation $Y_3 = g(y_2, U, \mathcal{E}_3)$ involving the latter two periods, $t = 2, 3$, which is comparable to the second stage.

Section 3.2 sketches the identification strategy using $T = 6$ periods without a proxy variable. With six periods, the three consecutive observations, Y_2, Y_3 and Y_4, together constitute a substitute for the proxy. Intuitively, controlling for the adjacent states, Y_2 and Y_4, the intermediate state Y_3 is correlated with (Y_1, Y_5, Y_6) only through the heterogeneity U. This allows Y_3 to serve as a proxy for U, conditionally on Y_2 and Y_4. The constructed proxy identifies both $F_{Y_6|Y_5U}$, which represents the dynamic function g, and the initial condition F_{Y_1U}.

3.1. A Sketch of the Identification Strategy. Consider the model which consists of $(g, h, F_{Y_1U}, \zeta, F_{\mathcal{E}_t}, F_{V_t}, F_W)$ where

$$
(1) \qquad \begin{cases} Y_t = g(Y_{t-1}, U, \mathcal{E}_t) \quad t = 2, \cdots, T & \text{(State Dynamics)} \\ D_t = h(Y_t, U, V_t) \quad t = 1, \cdots, T-1 & \text{(Selection)} \\ F_{Y_1U} & \text{(Initial Condition)} \\ Z = \zeta(U, W) & \text{(Optional: Nonclassical Proxy)} \end{cases}
$$

The observed state variable Y_t, such as SES, follows a first-order Markov process g with nonseparable heterogeneity U. The selection variable D_t, such as survival, follows a Markov decision process h with heterogeneity U. The outcome $D_t = 0$, such as death, indicates attrition after which the counterfactual state variable Y_t becomes unobservable. The distribution F_{Y_1U} of (Y_1, U) models dependence of the initial state Y_1 on unobserved heterogeneity U. The last optional equation models the proxy variable Z as a noisy signal of the true unobserved heterogeneity U with a nonseparable noise variable W.[13] This proxy equation is optional when $T \geqslant 6$, because the three additional periods construct the proxy. The functional relations in (1) together with the following exogeneity assumptions define the econometric model.

$$
\begin{array}{lll}
 & \text{(i)} & \text{Exogeneity of } \mathcal{E}_t\text{:} \quad \mathcal{E}_t \perp\!\!\!\perp (U, Y_1, \{\mathcal{E}_s\}_{s<t}, \{V_s\}_{s<t}, W) \text{ for all } t \geq 2. \\
(2) & \text{(ii)} & \text{Exogeneity of } V_t\text{:} \quad V_t \perp\!\!\!\perp (U, Y_1, \{\mathcal{E}_s\}_{s\leqslant t}, \{V_s\}_{s<t}) \text{ for all } t \geq 1. \\
 & \text{(iii)} & \text{Exogeneity of } W\text{:} \quad W \perp\!\!\!\perp (Y_1, \{\mathcal{E}_t\}_t, \{V_t\}_t).
\end{array}
\tag{2}
$$

This construction of the model leaves the nonseparable fixed effect U as the only source of endogeneity, and is in accordance with the standard assumptions in the panel data literature. We assume that the exogenous shocks, $\mathcal{E}_t$, V_t, and the noise, W, are continuously distributed, and normalize the distributions $F_{\mathcal{E}_t}, F_{V_t}$, and F_W to Uniform$(0, 1)$ so that the model consists of only the four elements (g, h, F_{Y_1U}, ζ).

The nonseparable fixed effect U and the exogenous errors $\mathcal{E}_t$, V_t and W are unobservable by econometricians. Observation of the the state variable Y_t is contingent on self-selection by economic agents. For a three-period panel data, the states are observed according to the rule:

[13] A standard proxy Z is an additively separable function of U and W (cf. Wooldridge, 2001; Ch. 4). Our proxy model ζ allows for nonseparability and nonlinearity to avoid a misspecification bias. One can think of the pair (U, W) as fixed unobserved characteristics, where U is the part that enters the economic model whereas W is the part excluded from these functions (i.e., exclusion restriction). Therefore, W is exogenous by construction.

$$\begin{aligned}&\text{Observe } Y_1.\\ &\text{Observe } Y_2 \quad \text{if } D_1 = 1.\\ &\text{Observe } Y_3 \quad \text{if } D_1 = D_2 = 1.\end{aligned}$$

Consequently, panel data reveals only the specific parts, $F_{Y_2Y_1ZD_1}(\,\cdot,\ \cdot,\ \cdot, 1)$ and $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\ \cdot,\ \cdot,\ \cdot, 1, 1)$, of the joint distributions, but they will become unobservable once a '1' is replaced by a '0' in the slot of D_t.

In the current section, we consider a special case where Y_t, U, and Z are Bernoulli random variables for ease of exposition. This special case conveniently allows to describe the identification strategy by means of matrices instead of operators. The basic idea of the identification strategy for this special case extends to more general cases, as formally stated in Section 4. For this setting, the function g can be represented by a heterogeneous Markov transition probability, $f_{Y_{t+1}|Y_tU}$. Similarly, the selection function h can be represented by the heterogeneous conditional choice probability (heterogeneous CCP), $f_{D_t|Y_tU}$. In this way, the model (g, h, F_{Y_1U}, ζ) in (1) can be represented by the quadruple $(f_{Y_{t+1}|Y_tU},\ f_{D_t|Y_tU},\ f_{Y_1U},\ f_{Z|U})$ of conditional and joint mass functions. Given this statistical representation, identification amounts to that

$$f_{Y_2Y_1ZD_1}(\,\cdot,\ \cdot,\ \cdot, 1)\ \&\ f_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\ \cdot,\ \cdot,\ \cdot, 1, 1) \quad \overset{\text{uniquely}}{\underset{\text{determine}}{\mapsto}} \quad (f_{Y_{t+1}|Y_tU},\ f_{D_t|Y_tU},\ f_{Y_1U},\ f_{Z|U}).$$

The exogeneity in (2) implies the following two conditional independence restrictions:

$$\text{(3)} \quad \text{Exogeneity of } \mathcal{E}_3 \Rightarrow \text{ Markov Property:} \qquad Y_3 \perp\!\!\!\perp (Y_1, D_1, D_2, Z) \mid (Y_2, U)$$

$$\text{(4)} \quad \text{Exogeneity of } W \Rightarrow \text{Redundant Proxy:} \qquad Z \perp\!\!\!\perp (Y_2, Y_1, D_2, D_1) \mid U$$

See Lemma 3 in the appendix for a derivation of the above conditional independence restrictions. The Markov property (3) states that the current state Y_2 and the heterogeneity U are sufficient statistics for the distribution of the next state Y_3. The redundant proxy (4) states that, once the true heterogeneity U is controlled for, the proxy Z is redundant for the model.[14] These independence restrictions derive the

[14] The redundant proxy assumption is stated in terms of conditional moments in the context of linear additively separable models; see Wooldridge (2001), Ch. 4.

following chain of equalities for each y_1, y, y_3, z:

$$\begin{aligned}
&\underbrace{f_{Y_3Y_2Y_1ZD_2D_1}(y_3, y, y_1, z, 1, 1)}_{\text{Observed Data}} = \sum_u f_{Y_3Y_2Y_1ZUD_2D_1}(y_3, y, y_1, z, u, 1, 1) \\
= &\sum_u f_{Y_3|Y_2Y_1ZUD_2D_1}(y_3 \mid y, y_1, z, u, 1, 1) \cdot f_{Z|Y_2Y_1UD_2D_1}(z \mid y, y_1, u, 1, 1) \\
&\cdot f_{Y_2Y_1UD_2D_1}(y, y_1, u, 1, 1) \\
\overset{(*)}{=} &\sum_u \underbrace{f_{Y_3|Y_2U}(y_3 \mid y, u)}_{\text{Model } g} \cdot \underbrace{f_{Z|U}(z \mid u)}_{\text{Model } \zeta} \cdot \underbrace{f_{Y_2Y_1UD_2D_1}(y, y_1, u, 1, 1)}_{\substack{\text{Nonparametric Residual} \\ \text{Involving Selection } D_2 = D_1 = 1 \\ \text{\& Nonseparable Fixed Effect } U}}
\end{aligned} \tag{5}$$

where the last equality $(*)$ follows from (3) and (4). $f_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1)$ on the left-hand side can be observed from data because the slots of D_1 and D_2 contain '1.' The right-hand side consists of three factors, where $f_{Y_3|Y_2U}$ and $f_{Z|U}$ represent g and ζ, respectively. The last factor $f_{Y_2Y_1UD_2D_1}(\,\cdot,\,\cdot,\,\cdot,1,1)$ can be thought of as the *nonparametric residual* of the observed data after extracting the two preceding economic components, g and ζ. This nonparametric residual absorbs the selection, $D_2 = D_1 = 1$, which is a source of selection bias. Moreover, the nonparametric residual also absorbs the nonparametric distribution of the nonseparable fixed effect U, which is a source of endogeneity bias. In other words, the two sources of bias – nonseparable heterogeneity and selection – captured by the nonparametric residual are isolated from the economic models (g, ζ) in the decomposition (5).

For convenience of calculation, we rewrite the equality (5) in terms of matrices as

$$L_{y,z} = P_y Q_z \tilde{L}_y \qquad \text{for each } y \in \mathcal{Y} \text{ and } z \in \mathcal{Z}, \tag{6}$$

where $L_{y,z}$, P_y, Q_z, and $\tilde{L}_y$ are defined as[15]

$$
\begin{aligned}
L_{y,z} &:= \begin{bmatrix} f_{Y_3Y_2Y_1ZD_2D_1}(0,y,0,z,1,1) & f_{Y_3Y_2Y_1ZD_2D_1}(0,y,1,z,1,1) \\ f_{Y_3Y_2Y_1ZD_2D_1}(1,y,0,z,1,1) & f_{Y_3Y_2Y_1ZD_2D_1}(1,y,1,z,1,1) \end{bmatrix} \\
&\qquad\qquad \rightarrow \text{Observed data} \\
P_y &:= \begin{bmatrix} f_{Y_3|Y_2U}(0 \mid y,0) & f_{Y_3|Y_2U}(0 \mid y,1) \\ f_{Y_3|Y_2U}(1 \mid y,0) & f_{Y_3|Y_2U}(1 \mid y,1) \end{bmatrix} \\
&\qquad\qquad \rightarrow \text{Represents model } g \\
Q_z &:= \operatorname{diag}(f_{Z|U}(z \mid 0) \quad f_{Z|U}(z \mid 1))' \\
&\qquad\qquad \rightarrow \text{Represents model } \zeta \\
\tilde{L}_y &:= \begin{bmatrix} f_{Y_2Y_1UD_2D_1}(y,0,0,1,1) & f_{Y_2Y_1UD_2D_1}(y,1,0,1,1) \\ f_{Y_2Y_1UD_2D_1}(y,0,1,1,1) & f_{Y_2Y_1UD_2D_1}(y,1,1,1,1) \end{bmatrix} \\
&\qquad\qquad \rightarrow \text{Residual}
\end{aligned}
$$

In (6), the observed matrix $L_{y,z}$ is decomposed into three factors, P_y, Q_z, and $\tilde{L}_y$. The matrices P_y and Q_z represent the economic models g and ζ, respectively. The matrix $\tilde{L}_y$ contains the remainder as the nonparametric residual, and particularly contains the two sources of bias.

Given the decomposition (6), the next step is to eliminate the nonparametric residual matrix $\tilde{L}_y$ in order to nonparametrically difference out the influence of selection and nonseparable heterogeneity, or to remove biases induced by them. Using the two values of z of the proxy variable, 0 and 1, we form the following composition:

$$
\underbrace{L_{y,1}L_{y,0}^{-1}}_{\text{Observed Data}} = P_yQ_1\tilde{L}_y\tilde{L}_y^{-1}Q_0^{-1}P_y^{-1} = \underbrace{P_y}_{g}\underbrace{Q_1Q_0^{-1}}_{\zeta}\underbrace{P_y^{-1}}_{g}. \tag{7}
$$

The nonparametric residual matrix $\tilde{L}_y$ has been eliminated as desired. Consequently, the observed data on the left hand side is now purely linked to a product of model

[15] These two-by-two matrices follow from the simplifying assumption of the current subsection that Y_t, U, and Z are Bernoulli random variables. In general cases, integral and multiplication operators replace these matrices.

components (g, ζ) without any influence of the selection or the nonseparable heterogeneity.

The composition (7) is valid provided that P_y, Q_z, and $\tilde{L}_y$ are all non-singular. The following rank restrictions guarantee that they are indeed non-singular under the current setting.

(8)	Heterogeneous Dynamics:	$\mathrm{E}[g(y, 0, \mathcal{E}_t)] \neq \mathrm{E}[g(y, 1, \mathcal{E}_t)]$
(9)	Nondegenerate Proxy Model:	$0 < \mathrm{E}[\zeta(u, W)] < 1$ for each $u \in \{0, 1\}$
(10)	No Extinction:	$\mathrm{E}[h(y, u, V_t)] > 0$ for each $u \in \{0, 1\}$
(11)	Initial Heterogeneity:	$\mathrm{E}[U \mid Y_2 = y, Y_1 = 0, D_1 = 1] \neq$ $\mathrm{E}[U \mid Y_2 = y, Y_1 = 1, D_1 = 1]$

Restriction (8) requires that the dynamic model g is a nontrivial function of the unobserved heterogeneity, and implies that the matrix P_y is non-singular.[16] Restriction (9) requires that the proxy model (ζ, F_W) exhibits nondegeneracy, and implies that the matrix Q_0 is non-singular. Restriction (10) requires a positive survival probability for each heterogeneous type $u \in \{0, 1\}$, and hence drives no type U into extinction. Restriction (11) requires that the unobserved heterogeneity is present at the initial observation. The last two restrictions (10) and (11) together imply that the nonparametric residual matrix $\tilde{L}_y$ is non-singular.

Now that the nonparametric residual $\tilde{L}_y$ containing the two sources of bias has gone, it remains to identify the elements of the matrices P_y and Q_z from equation (7). This can be accomplished by showing the uniqueness of eigenvalues and eigenvectors (e.g., Hu, 2008; Kasahara and Shimotsu, 2009). Because the matrix Q_z is diagonal, (7) forms a diagonalization of the observable matrix $L_{y,1}L_{y,0}^{-1}$. The diagonal elements of $Q_1 Q_0^{-1}$ and the columns of P_y are the eigenvalues and the eigenvectors of $L_{y,1}L_{y,0}^{-1}$,

[16] Under the current simplified setting with Bernoulli random variables, a violation of this assumption implies absence of endogeneity in the dynamic model, and thus the dynamic function g would still be identified. However, the other functions are not guaranteed to be identified without this assumption.

respectively. Therefore, $Q_1 Q_0^{-1}$ is identified by the eigenvalues of the observable matrix $L_{y,1} L_{y,0}^{-1}$ without an additional assumption.

On the other hand, identification of P_y follows from the following additional restriction:

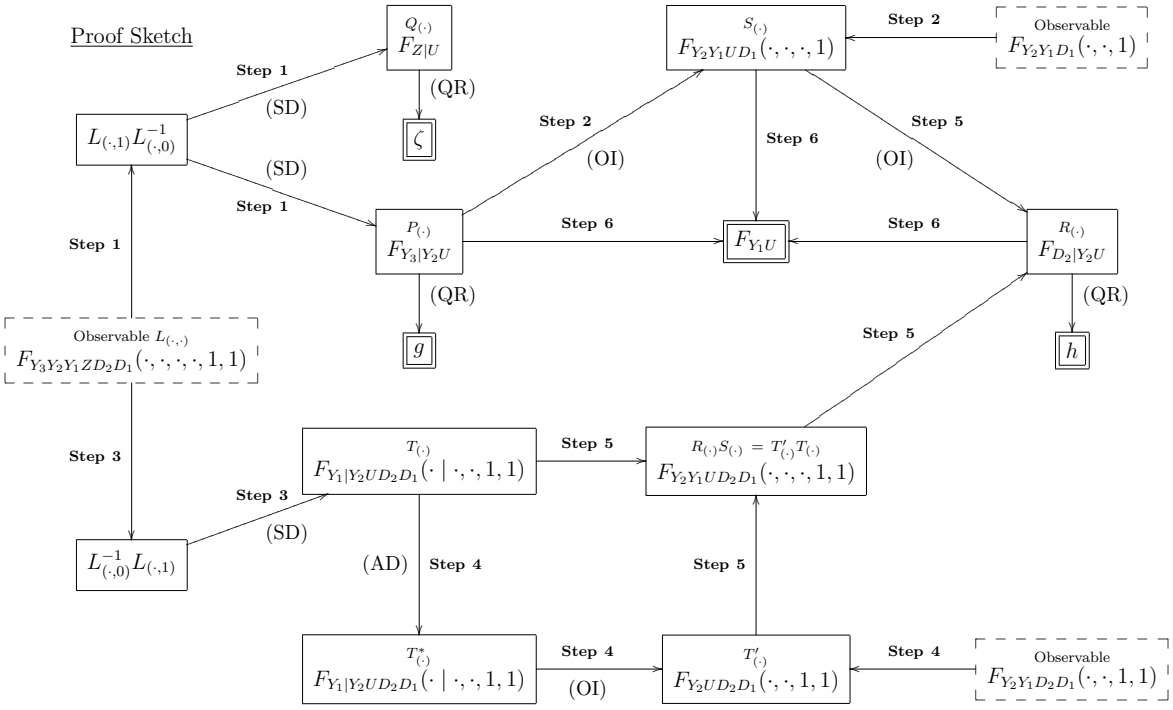
$$\text{(12)} \qquad \text{Relevant Proxy:} \qquad \mathrm{E}[\zeta(0, W)] \neq \mathrm{E}[\zeta(1, W)].$$

This restriction (12) requires that the proxy model ζ is a nontrivial function of the true unobserved heterogeneity on average. It characterizes the relevance of Z as a proxy of U, and implies that the elements of $Q_1 Q_0^{-1}$ (i.e., the eigenvalues of $L_{y,1} L_{y,0}^{-1}$) are distinct. The distinct eigenvalues uniquely determine the corresponding eigenvectors up to scale. Since an eigenvector $[f_{Y_{t+1}|Y_t U}(0 \mid y, u),\ f_{Y_{t+1}|Y_t U}(1 \mid y, u)]'$ is a vector of conditional densities which sum to one, the scale is also uniquely determined. Therefore, P_y and $Q_1 Q_0^{-1}$ are identified by the observed data $L_{y,1} L_{y,0}^{-1}$. The identified eigenvalues[17] take the form of the proxy odds $f_{Z|U}(1 \mid u)/(1 - f_{Z|U}(1 \mid u))$, which in turn uniquely determines the diagonal elements $f_{Z|U}(z \mid u)$ of Q_z for each z. This procedure heuristically shows how the the elements (g, ζ) of the model is identified from endogenously unbalanced panel data.

Remark 1. The general identification procedure consists of six steps. The current subsection presents a sketch of the first step to identify (g, ζ). Five additional steps show that the remaining two elements $(h, F_{Y_1 U})$ of the model are also identified. Figure 3.1 summarizes all the six steps. Section 4 presents a complete identification result.

Discussion 1. This sketch of the identification strategy demonstrates how the proxy handles both selection and nonseparable heterogeneity at the same time. The trick of Equation (5) or (6) is to isolate the selection ($D_1 = D_2 = 1$) and the nonparametric distribution of the nonseparable heterogeneity U into the nonparametric

[17] Even though we obtain real eigenvalues in this spectral decomposition, $L_{y,1} L_{y,0}^{-1}$ need not be symmetric. Note that a Hermitian operator is sufficient for real spectrum, but not necessary. This identification result holds as far as the identifying restrictions are satisfied.



(AD) – Adjoint Operator (OI) – Operator Inversion

(QR) – Quantile Representation (SD) – Spectral Decomposition

FIGURE 1.1. A sketch of the proof of the identification strategy.

residual matrix $\tilde{L}_y$, which in turn is eliminated in Equation (7). Our method thus can be considered as a *nonparametric differencing* facilitated by a proxy variable, nonparametrically differencing out both nonseparable fixed effects and endogeneous selection. This process is analogous to the first differencing method which differences out fixed effects arithmetically. Our nonparametric differencing occurs in the non-commutative group of matrices (generally the group of linear operators), whereas the first differencing occurs in $(\mathbb{R}, +)$. In the non-commutative group, the proxy Z plays the role of selectively canceling out the nonparametric residual matrix $\tilde{L}_y$ while leaving the P_y and Q_z matrices intact. The use of a proxy parallels the classical idea of using instruments as means of removing endogeneity (Hausman and Taylor, 1981). Instrumental variables are useful for additive models because projection (moment restriction) of additive models on instruments removes fixed effects as in Hausman

and Taylor. This projection method is not generally feasible for nonseparable and nonparametric models. Therefore, this paper uses a proxy variable, akin to a control variable,[18] to nonparametrically difference out the nonseparable fixed effect along with selection as argued above. This point is revisited in Section 3.2: Discussion 3.

3.2. A Sketch of the Identification Strategy for $T = 6$. When $T = 6$, we identify the model (g, h, F_{Y_1U}) without using an outside proxy variable or the proxy model ζ. In the presence of an outside proxy, the main identification strategy was to derive the decomposition $L_{y,z} = P_y Q_z \tilde{L}_y$ from which $\tilde{L}_y$ was eliminated (cf. Section 3.1). A similar idea applies to the case of $T = 6$ without an outside proxy.

Again, assume that Y_t, U, and Z follow the Bernoulli distribution for ease of exposition. Let $Z := Y_3$ for notational convenience. Using the exogeneity restriction (3) yields the following decomposition of the observed data.

$$\underbrace{f_{Y_6Y_5Y_4ZY_2Y_1D_5D_4D_3D_2D_1}(y_6, y_5, y_4, \overbrace{z}^{y_3}, y_2, y_1, 1, 1, 1, 1, 1)}_{\text{Observed from Data } \to\ L_{y_5,y_4,z,y_2}} = \sum_u \underbrace{f_{Y_6|Y_5U}(y_6 \mid y_5, u)}_{\text{Model } g\ \to\ P_{y_5}}$$
$$\times \underbrace{f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, z, 1, 1, 1 \mid y_2, u)}_{\text{An Alternative to Proxy Model } \zeta\ \to\ Q_{y_4,z,y_2}} \cdot \underbrace{f_{Y_5|Y_4U}(y_5 \mid y_4, u) \cdot f_{Y_2Y_1UD_2D_1}(y_2, y_1, u, 1, 1)}_{\text{To Be Eliminated } \to\ \tilde{L}_{y_5,y_4,y_2}}$$

This equality can be equivalently written in terms of matrices as $L_{y_5,y_4,z,y_2} = P_{y_5} Q_{y_4,z,y_2} \tilde{L}_{y_5,y_4,y_2}$ for each (y_5, y_4, z, y_2), where the 2×2 matrices are defined as

$$\begin{aligned}
L_{y_5,y_4,z,y_2} &:= [f_{Y_6Y_5Y_4ZY_2Y_1D_5D_4D_3D_2D_1}(i, y_5, y_4, y_3, y_2, j, 1, 1, 1, 1, 1)]_{(i,j)\in\{0,1\}\times\{0,1\}} \\
P_{y_5} &:= \left[f_{Y_6|Y_5U}(i \mid y_5, j)\right]_{(i,j)\in\{0,1\}\times\{0,1\}} \\
Q_{y_4,z,y_2} &:= \operatorname{diag}\left(f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, z, 1, 1, 1 \mid y_2, 0) \;\; f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, z, 1, 1, 1 \mid y_2, 1)\right)' \\
\tilde{L}_{y_5y_4y_2} &:= \left[f_{Y_5|Y_4U}(y_5 \mid y_4, u) \cdot f_{Y_2Y_1UD_2D_1}(y_2, j, i, 1, 1)\right]_{(i,j)\in\{0,1\}\times\{0,1\}}
\end{aligned}$$

Similarly to the case with an outside proxy, varying $z = y_3$ while fixing (y_5, y_4, y_2) eliminates $\tilde{L}_{y_5,y_4,y_2}$ because it does not depend on $z = y_3$. Under rank restrictions,

[18] If X denotes an endogenous regressor and U denotes unobserved factors, then a proxy, a control variable, and an instrument are characterized by $Z \perp\!\!\!\perp X \mid U$, $X \perp\!\!\!\perp U \mid Z$, and $Z \perp\!\!\!\perp U$, respectively. The conditional independence (4) thus characterizes Z as a proxy rather than a control variable or an instrument.

the composition

$$\underbrace{L_{y_5,y_4,1,y_2}L^{-1}_{y_5,y_4,0,y_2}}_{\text{Observed Data}} = P_{y_5}Q_{y_4,1,y_2}\tilde{L}_{y_5y_4y_2}\tilde{L}_{y_5y_4y_2}Q^{-1}_{y_4,0,y_2}P^{-1}_{y_5} = P_{y_5}\underbrace{Q_{y_4,1,y_2}Q^{-1}_{y_4,0,y_2}}_{\text{Diagonal}}P^{-1}_{y_5}$$

yields the eigenvalue-eigenvector decomposition to identify the dynamic model g represented by the matrix P_{y_5}. Five additional steps identify the rest (h, F_{Y_1U}) of the model.

Discussion 2. Why do we need $T = 6$? For convenience of illustration, ignore selection. The arrows in the diagram below indicate the directions of the causal effects. We are interested in g and F_{Y_1U} enclosed by round shapes. First, note that a variation in Y_6 in response to a variation in (Y_5, U) reveals g. Second, a variation in U in response to a variation in Y_1 reveals $F_{U|Y_1}$, hence F_{Y_1U}. We can see from the causal diagram that Y_2, Y_3, and Y_4 are correlated with U, and hence we may conjecture that they could serve as a proxy of U. However, any of them, say $Z := Y_3$, cannot be a genuine proxy because the redundant proxy assumption similar to (4), say $Z \perp\!\!\!\perp Y_5 \mid Z$, would be violated with this choice $Z = Y_3$. That is, even if we control for U, Y_3 is still correlated with Y_5 through the dynamic channel along the horizontal arrows. In order to shut out this dynamic channel, we control the intermediate state Y_4 between Y_3 and Y_5. Using the language of the causal inference literature, we say that (U, Y_4) "d-separates" Y_3 and Y_5 in the causal diagram below, and this d-separation implies the conditional independence restriction $Y_3 \perp\!\!\!\perp Y_5 \mid (U, Y_4)$; see Pearl (2000). Therefore Y_3 is now a genuine proxy of U conditionally on the fixed Y_4 to analyze the dynamic model g. Similarly, we control the intermediate state Y_2 between Y_1 and Y_3 to analyze the initial condition F_{Y_1U}. The causal diagram indicates that (U, Y_2) "d-separates" Y_1 and Y_3, hence $Y_3 \perp\!\!\!\perp Y_1 \mid (U, Y_2)$. This makes Y_3 a genuine proxy of U conditionally on the fixed Y_2 to analyze the initial condition F_{Y_1U}. Controlling for the two adjacent states, Y_2 and Y_4, costs the consecutive three periods (Y_2, Y_3, Y_4) for the constructed proxy model Q_{y_4,z,y_2}. This is an intuition behind the requirement of three additional periods for identification without an outside proxy variable. See Section 10.2 for a formal proof.

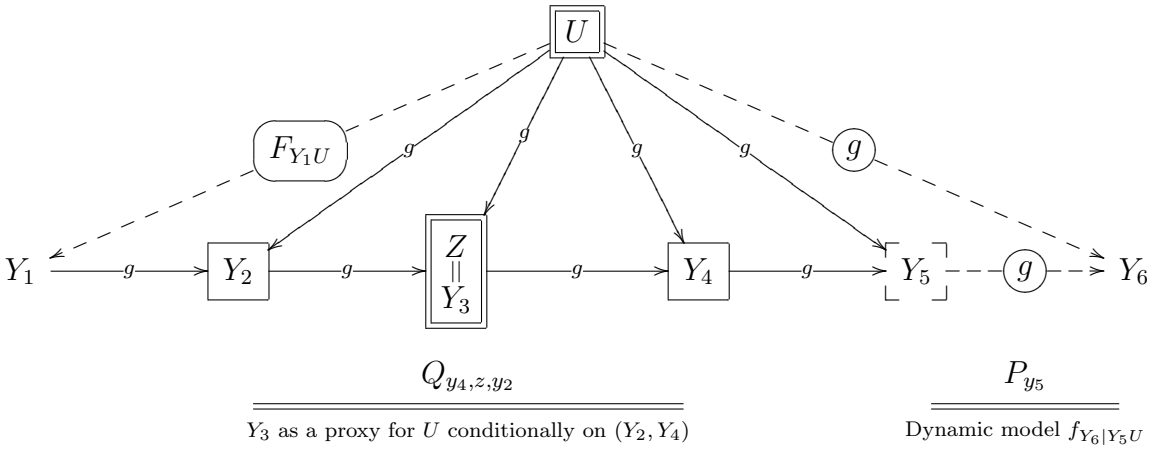


FIGURE 1.2. Causal diagram for six periods.

DISCUSSION 3. The idea of using three additional time periods to construct a proxy variable parallels the well-known idea of using lags as instruments to form identifying restrictions for additively separable dynamic panel data models (e.g., Anderson and Hsiao, 1982; Arellano and Bond, 1991; Ahn and Schmidt, 1995; Arellano and Bover, 1995; Blundell and Bond, 1998; Hahn, 1999). Because projection or moment restriction on instruments is not generally a viable option for nonseparable and nonparametric models, the literature on nonseparable cross section models proposes constructing a proxy variable or a control variable from instruments (Garen, 1984; Florens, Heckman, Meghir, and Vytlacil, 2008; Imbens and Newey, 2009). Altonji and Matzkin (2005) show that a control variable can also be constructed from panel data for sibling and neighborhood panels. This paper proposes constructing a proxy variable from three additional observations of dynamic panel data, similar to Altonji and Matzkin's construction of a control variable from sibling and neighborhood panels. The constructed proxy variable turns out to account for not only nonseparable heterogeneity but also selection as argued above.

4. Identification

This section formalizes the identification result, a part of which is sketched in Section 3.

4.1. Identifying Restrictions. Identification is proved by showing the well-definition of the inverse DGP correspondence

$$(F_{Y_2Y_1ZD_1}(\,\cdot,\,\cdot,\,\cdot,1),\ F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1))\ \mapsto\ (g,h,F_{Y_1U},\zeta,F_{\mathcal{E}_t},F_{V_t},F_W),$$

up to observational equivalence classes represented by a certain normalization of the error distributions $F_{\mathcal{E}_t}$, F_{V_t}, and F_W a la Matzkin (2003). To this end, we invoke the following four restrictions on the set of potential data-generating models.

Restriction 1 (Representation). Each of the functions g, h, and ζ is non-decreasing and càglàd (left-continuous with right limit) in the last argument. The distributions of $\mathcal{E}_t$, V_t, and W are absolutely continuous with convex supports, and each of $\{\mathcal{E}_t\}_t$ and $\{V_t\}_t$ is identically distributed across t.

The weak – as opposed to strict – monotonicity of the functions with respect to idiosyncratic errors accommodates discrete outcomes Y_t, D_t, and Z under absolutely continuous distributions of errors $(\mathcal{E}_t, V_t, W)$. The purpose of Restriction 1 is to construct representations of the equivalence classes of nonseparable functions up to which g, h, and ζ are uniquely determined by the distributions $F_{Y_t|Y_{t-1}U}$, $F_{D_t|Y_tU}$, and $F_{Z|U}$, respectively. The independence restriction stated below in addition to Restriction 1 allows for their quantile representations in particular.

Restriction 2 (Independence).

(i) Exogeneity of $\mathcal{E}_t$: $\quad \mathcal{E}_t \perp\!\!\!\perp (U, Y_1, \{\mathcal{E}_s\}_{s<t}, \{V_s\}_{s<t}, W)$ for all $t \geq 2$.

(ii) Exogeneity of V_t: $\quad V_t \perp\!\!\!\perp (U, Y_1, \{\mathcal{E}_s\}_{s\leqslant t}, \{V_s\}_{s<t})$ for all $t \geq 1$.

(iii) Exogeneity of W: $\quad W \perp\!\!\!\perp (Y_1, \{\mathcal{E}_t\}_t, \{V_t\}_t)$.

In the context of Section 3.1, Restriction 2 (i) and (iii) imply the conditional independence restrictions (3) and (4), respectively. Parts (i) and (ii) impose exogeneity of the idiosyncratic errors $\mathcal{E}_t$ and V_t, thus leaving U as the only source of endogeneity. Part (iii) requires exogeneity of the noise W in the nonseparable proxy model ζ. This means that the unobserved characteristics consist of two parts (U, W) where U is the part that enters the functions g and h, whereas W is the part excluded from

those functions (i.e., exclusion restriction), and hence is exogenous by construction. Part (iii) implies $Z \perp\!\!\!\perp (Y_1, \{\mathcal{E}_t\}_t, \{V_t\}_t) \mid U$, which is similar to the redundant proxy restriction in the classical sense as discussed in Section 3.1: once the true unobserved heterogeneity U is controlled for, the proxy Z is redundant for $(g, h, F_{Y_1,U})$. These independence conditions play the role of decomposing observed data into model components and the nonparametric residual, as we saw through the sketch in Section 3.

Restriction 3 (Rank Conditions). The following conditions hold for every $y \in \mathcal{Y}$:

(i) Heterogeneous Dynamics: the integral operator $P_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$ defined by $P_y\xi(y') = \int f_{Y_3|Y_2U}(y' \mid y, u) \cdot \xi(u)du$ is bounded and invertible.

(ii) Nondegenerate Proxy Model: there exists $\delta > 0$ such that $0 < f_{Z|U}(1 \mid u) \leqslant 1 - \delta$ for all u.

Relevant Proxy: $f_{Z|U}(1 \mid u) \neq f_{Z|U}(1 \mid u')$ whenever $u \neq u'$.

(iii) No Extinction: $f_{D_2|Y_2U}(1 \mid y, u) > 0$ for all $u \in \mathcal{U}$.

(iv) Initial Heterogeneity: the integral operator $S_y : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$ defined by $S_y\xi(u) = \int f_{Y_2Y_1UD_1U}(y, y', u, 1) \cdot \xi(y')dy'$ is bounded and invertible.

Under the special case discussed in Section 3.1, Restriction 3 is equivalent to (8)–(12), by which the dynamic function g and the proxy model ζ were identified in that section. The notion of 'invertibility' depends on the normed linear spaces on which the operators are defined. We use $\mathcal{L}^2$ in order to exploit convenient properties of the Hilbert spaces.[19] A bounded linear operator between Hilbert spaces guarantees existence and uniqueness of its adjoint operator, which of course presumes a pre-Hilbert space structure in particular. Moreover, the invertibility guarantees that the adjoint operator is also invertible, which is an important property used to derive identification of the selection rule h and initial condition F_{Y_1U}. Andrews (2011) shows

[19] Carrasco, Florens, and Renault (2007) review some important properties of operators on Hilbert spaces.

that a wide variety of injective operators between $\mathcal{L}^2$ spaces can be constructed from an orthonormal basis, and that the completeness assumption 'generically' holds.

The first part of the rank condition (ii) requires that the proxy model (ζ, F_W) exhibits nondegeneracy. The second part of the rank condition (ii) requires that Z is a relevant proxy for unobserved heterogeneity U, as characterized by distinct proxy scores $f_{Z|U}(1 \mid u)$ across u. The rank condition (iii) requires that there continue to exist some survivors in each heterogeneous type, hence no type U goes extinct. This restriction is natural because one cannot learn about a dynamic structure of the group of individuals that goes extinct after the first time period.

RESTRICTION 4 (Labeling of U in Nonseparable Models). $u \equiv f_{Z|U}(1 \mid u)$ for all $u \in \mathcal{U}$.

Due to its unobservability, U has neither intrinsic values nor units of measurement. This is a reason for potential non-uniqueness of fully nonseparable functions. The purpose of Restriction 4 is to attach concrete values to unobserved heterogeneity U; see also Hu and Schennach (2008). Restriction 4 is innocuous in nonseparable models in the sense that identification is considered up to observational equivalence $g(y, u, \varepsilon) \equiv g_\pi(y, \pi(u), \varepsilon)$ for any permutation π of $\mathcal{U}$. On the other hand, this restriction is redundant and too stringent for additively separable models, in which U has the same unit of measurement as Y by construction. In the latter case, we can replace Restriction 4 by the following alternative labeling assumption.

Restriction 4′ (Labeling of U in Separable Models). $u \equiv \mathrm{E}[g(y, u, \mathcal{E})] - \tilde{g}(y)$ for all $u \in \mathcal{U}$ and $y \in \mathcal{Y}$ for some function $\tilde{g}$.

This alternative labeling restriction is innocuous for separable models in the sense that it is automatically satisfied by additive models of the form $g(y, u, \varepsilon) = \tilde{g}(y) + u + \varepsilon$ with $\mathrm{E}[\mathcal{E}] = 0$.

4.2. Representation. Nonparametric identification of nonseparable functions is generally feasible only up to some equivalence classes (e.g., Matzkin, 2003, 2007). Representations of these equivalence classes are discussed as a preliminary step toward identification. Restrictions 1 and 2 allow representations of functions g, h, and ζ by normalizing the distributions of the independent errors.

LEMMA 1 (Quantile Representations of Non-Decreasing Càglàd Functions).
(i) Suppose that Restrictions 1 and 2 (i) hold. Then $F_{Y_3|Y_2U}$ uniquely determines g up to the observational equivalence classes represented by the normalization $\mathcal{E}_t \sim$ Uniform$(0,1)$.
(ii) Suppose that Restrictions 1 and 2 (ii) hold. Then $F_{D_2|Y_2U}$ uniquely determines h up to the observational equivalence classes represented by the normalization $V_t \sim$ Uniform$(0,1)$.
(iii) Suppose that Restrictions 1 and 2 (iii) hold. Then $F_{Z|U}$ uniquely determines ζ up to the observational equivalence classes represented by the normalization $W \sim$ Uniform$(0,1)$.

A proof is given in Section 8.1. The representations under these assumptions and normalizations are established by the respective quantile regressions:

$$
\begin{aligned}
g(y,u,\varepsilon) &= F^{-1}_{Y_3|Y_2U}(\varepsilon \mid y,u) := \inf\{y' \mid \varepsilon \leq F_{Y_3|Y_2U}(y' \mid y,u)\} && \forall (y,u,\varepsilon) \\
h(y,u,v) &= F^{-1}_{D_2|Y_2U}(v \mid y,u) := \inf\{d \mid v \leq F_{D_2|Y_2U}(d \mid y,u)\} && \forall (y,u,v) \\
\zeta(u,w) &= F^{-1}_{Z|U}(w \mid u) := \inf\{z \mid w \leq F_{Z|U}(z \mid u)\} && \forall (u,w)
\end{aligned}
$$

The non-decreasing condition in Restriction 1 is sufficient for almost-everywhere equivalence of the quantile representations. Furthermore, we also require the càglàd condition of Restriction 1 for point-wise equivalence of the quantile representations. Given Lemma 1, it remains to show that the observed distributions $F_{Y_2Y_1ZD_1}(\,\cdot,\,\cdot,\,\cdot\,,1)$ and $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1))$ uniquely determine $(F_{Y_3|Y_2U}, F_{D_2|Y_2U}, F_{Z|U})$ as well as F_{Y_1U}.

4.3. The Main Identification Result. Section 4.2 shows that $F_{Y_3|Y_2U}$, $F_{D_1|Y_1U}$, and $F_{Z|U}$ uniquely determine g, h, and ζ, respectively, up to the aforementioned equivalence classes. Therefore, the model (g, h, F_{Y_1U}, ζ) can be identified by showing the well-definition of the inverse DGP correspondence

$$(F_{Y_2Y_1ZD_1}(\cdot,\ \cdot,\ \cdot\ ,1),\ F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\ \cdot,\ \cdot,\ \cdot,1,1)) \mapsto (F_{Y_3|Y_2U}, F_{D_2|Y_2U}, F_{Y_1U}, F_{Z|U}).$$

Lemma 2 (Identification). *Under Restrictions 1, 2, 3, and 4, the quadruple* $(F_{Y_3|Y_2U}, F_{D_2|Y_2U}, F_{Y_1U}, F_{Z|U})$ *is uniquely determined by* $F_{Y_2Y_1ZD_1}(\cdot,\ \cdot,\ \cdot\ ,1)$ *and* $F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\ \cdot,\ \cdot,\ \cdot,1,1)$.

Combining Lemmas 1 and 2 yields the following main identification result of this paper.

Theorem 1 (Identification). *Under Restrictions 1, 2, 3, and 4, the model* (g, h, F_{Y_1U}, ζ) *is identified by* $F_{Y_2Y_1ZD_1}(\cdot,\ \cdot,\ \cdot,1)$ *and* $F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\ \cdot,\ \cdot,\ \cdot,1,1)$ *up to the equivalence classes represented by the normalizations* $\mathcal{E}_t, V_t, W \sim$ *Uniform*$(0,1)$.

A proof of Lemma 2 is given in Section 8.2, and consists of six steps of spectral decompositions, operator inversions (solving Fredholm equations of the first kind), and algebra. Figure 3.1 illustrates how observed data uniquely determines the model (g, h, F_{Y_1U}, ζ) through the six steps. Section 3.1 provided an informal sketch of the proof of the first step among others.

Remark 2. While the baseline model only induces a permanent dropout through $D_t = 0$, we can also allow for an entry through $D_t = 1$. See Section 10.1. This is useful to model entry of firms as well as reentry of female workers into the labor market after child birth. This extension also accommodates general unbalanced panel data with various causes of selection.

Remark 3. Three additional time periods can be used as a substitute for a nonclassical proxy variable. In other words, $T = 6$ periods of panel data alone identifies the model, and a proxy variable is optional. See Section 3.2 and Section 10.2.

REMARK 4. The baseline model consists of the first-order Markov process. Section 10.3 generalizes the baseline result to allow for higher-order lags in the functions g and h. Generally, $\tau + 2$ periods of unbalanced panel data identifies the model with τ-th order Markov process g and $(\tau - 1)$-st order Markov decision rule h. The baseline model is a special case with $\tau = 1$.

REMARK 5. The baseline model only allows individual fixed effects U. Suppose that the dynamic model g_t involves time effects, for example, to reflect common macroeconomic shocks to income dynamics. Then the model $(\{g_t\}_{t=2}^T, h, F_{Y_1U}, \zeta)$ can be identified by $T + 1$ periods of unbalanced panel data from $t = 0$ to $t = T \geqslant 3$. See Section 10.4.

REMARK 6. The rule for missing observations was defined in terms of a lagged selection indicator D_{t-1} which depends on Y_{t-1}. Data may instead be selected based on contemporaneous D_t which depends on Y_t. See Section 10.5. Note that, in the latter case, we may not observe Y_t based on which the data is selected. These two selection criteria reflect the ex ante versus ex post Roy selection processes by rational agents.

REMARK 7. Restriction 3 (i) implies the cardinality relation $|\text{supp}(U)| \leqslant |\text{supp}(Y_t)|$. This cardinality restriction in particular rules out binary Y_t with continuously distributed U. Furthermore, the relevant proxy in Restriction 3 (ii) implies $\dim(U) \leqslant 1$.

REMARK 8. For the result using a proxy variable, the notation appears to suggest that a proxy is time-invariant. However, a time-variant proxy $Z_t = \zeta(U, W_t)$ may also be used as far as W_t satisfies the same independence restriction as W.

5. Estimation

The identification result is derived through six steps of spectral decompositions, operator inversions (solutions to Fredholm equations of the first kind), and algebra, as illustrated in Figure 3.1. A sample-analog or plug-in estimation following all these

steps is practically infeasible. The present section therefore discusses how to turn this six-step procedure into a one-step procedure.

5.1. Constrained Maximum Likelihood. After showing nonparametric identification as in Section 4, one can generally proceed with the maximum likelihood estimation of parametric or semi-parametric sub-models. In our context, however, the presence of missing observations biases the standard maximum likelihood estimator. In this section, we apply the Kullback-Leibler information inequality to translate our main identification result (Lemma 2) into an *identification-preserving* criterion function, which is robust against selection or missing data.

Because the model (g, h, F_{Y_1U}, ζ) is represented by $(F_{Y_t|Y_{t-1}U}, F_{D_t|Y_tU}, F_{Y_1U}, F_{Z|U})$ (see Lemmas 1 and 4), we use $\mathcal{F}$ to denote the set of all the admissible model representations:

$$\mathcal{F} = \{(F_{Y_t|Y_{t-1}U}, F_{D_t|Y_tU}, F_{Y_1U}, F_{Z|U}) \mid (g, h, F_{Y_1U}, \zeta) \text{ satisfies Restrictions 1, 2, 3, and 4}\}.$$

As a consequence of the main identification result, the quadruple for the true model $(F^*_{Y_t|Y_{t-1}U}, F^*_{D_t|Y_tU}, F^*_{Y_1U}, F^*_{Z|U})$ can be characterized by the following criterion, allowing for a one-step plug-in estimator.

Corollary 1 (Constrained Maximum Likelihood). *If the quadruple for the true model $(F^*_{Y_t|Y_{t-1}U}, F^*_{D_t|Y_tU}, F^*_{Y_1U}, F^*_{Z|U})$ is an element of $\mathcal{F}$, then it is the unique solution to*

$$\max_{\left(F_{Y_t|Y_{t-1}U}, F_{D_t|Y_tU}, F_{Y_1U}, F_{Z|U}\right) \in \mathcal{F}} \quad c_1 E\Big[\log \int f_{Y_t|Y_{t-1}U}(Y_2 \mid Y_1, u) f_{D_t|Y_tU}(1 \mid Y_1, u) \\ f_{Y_1U}(Y_1, u) f_{Z|U}(Z \mid u) d\mu(u) \big| \, D_1 = 1\Big] + \\ c_2 E\Big[\log \int f_{Y_t|Y_{t-1}U}(Y_3 \mid Y_2, u) f_{Y_t|Y_{t-1}U}(Y_2 \mid Y_1, u) f_{D_t|Y_tU}(1 \mid Y_2, u) f_{D_t|Y_tU}(1 \mid Y_1, u) \\ f_{Y_1U}(Y_1, u) f_{Z|U}(Z \mid u) d\mu(u) \big| \, D_2 = D_1 = 1\Big]$$

for any $c_1, c_2 > 0$ subject to

$$\int f_{D_t|Y_tU}(1 \mid y_1, u) f_{Y_1U}(y_1, u) d\mu(y_1, u) = f_{D_1}(1) \qquad \textit{and}$$

$$\int f_{Y_t|Y_{t-1}U}(y_2 \mid y_1, u) f_{D_t|Y_tU}(1 \mid y_2, u) f_{D_t|Y_tU}(1 \mid y_1, u) f_{Y_1U}(y_1, u) d\mu(y_2, y_1, u) = f_{D_2D_1}(1, 1).$$

A proof is found in Section 9.1. The sense of uniqueness stated in the corollary is up to the equivalence classes identified by the underlying probability measures. Once a representing model $(F_{Y_t|Y_{t-1}U}, F_{D_t|Y_tU}, F_{Y_1U}, F_{Z|U})$ is parametrically or semi-/non-parametrically specified, the sample analog of the objective and constraints can be formed from observed data. The first term in the objective can be estimated since (Y_2, Y_1, Z) is observed conditionally on $D_1 = 1$. Similarly, the second term can be estimated since (Y_3, Y_2, Y_1, Z) is observed conditionally on $D_2 = D_1 = 1$. All the components in the two constraints are also computable from observed data since $f_{D_1}(1)$ and $f_{D_2D_1}(1, 1)$ are observable.

This criterion is related to the maximum likelihood. The objective consists of a convex combination of expected log likelihoods conditional on survivors. Using this objective alone therefore would incur a survivorship bias. To adjust for the selection bias, the constraints bind the model to correctly predict the observed selection probabilities. Any pair of positive values may be chosen for c_1 and c_2. However, there is a certain choice of these coefficients that makes the constrained optimization problem easier, as discussed in the following remark.

REMARK 9. Solutions to constrained optimization problems like Corollary 1 are characterized by saddle points of the Lagrangean functional. Although it appears easier than the original six-step procedure, this saddle-point problem over a function space is still practically challenging. By an appropriate choice of c_1 and c_2, we can, however, turn this saddle point problem into an unconstrained maximization problem. Let λ_1 and λ_2 denote the Lagrange multipliers for the two constraints in the corollary. Under some regularity conditions (Fréchet differentiability of the objective and constraint functionals, differentiability of the solution to the selection probability, and the regularity of the solution for the constraint functionals), the choice of $c_1 = \Pr(D_1 = 1)$ and $c_2 = \Pr(D_2 = D_1 = 1)$ guarantees $\lambda_1^* = \lambda_2^* = 1$ at the optimum (see Section 9.2). With this knowledge of the values of λ_1^* and λ_2^*, the solution to the problem in the corollary can now be characterized by a maximum rather than a saddle point. This fact is useful both for implementation of numerical solution methods and

for availability of the existing large sample theories of parametric, semiparametric, and nonparametric M-estimators.

REMARK 10. In case of using $T = 6$ periods of unbalanced panel data instead of a proxy variable, a similar one-step criterion to Corollary 1 can be derived. See Section 10.2.

5.2. An Estimator. The six steps of the identification strategy do not admit a practically feasible plug-in estimator. On the other hand, Corollary 1 and Remark 9 together yield the standard M-estimator by the sample analog. We decompose the model set as $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3 \times \mathcal{F}_4$, where $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$, and $\mathcal{F}_4$ are sets of parametric or semi-/non-parametric models for $f_{Y_t|Y_{t-1}U}$, $f_{D_t|Y_tU}$, f_{Y_1U}, and $f_{Z|U}$, respectively. Accordingly, we denote an element of $\mathcal{F}$ by $f = (f_1, f_2, f_3, f_4)$ for brevity. With this notation, Corollary 1 and Remark 9 imply that the estimator of the true model f_0 can be characterized by a solution $\hat{f}$ to the maximization problem:

$$\max_{f \in \mathcal{F}_{k(n)}} \frac{1}{n} \sum_{i=1}^{n} l(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}; f)$$

for some sieve space $\mathcal{F}_{k(n)} = \mathcal{F}_{1,k_1(n)} \times \mathcal{F}_{2,k_2(n)} \times \mathcal{F}_{3,k_3(n)} \times \mathcal{F}_{4,k_4(n)} \subset \mathcal{F}$, where

$$\begin{aligned}
l(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}; f) &:= \mathbb{1}\{D_{i1} = 1\} \cdot l_1(Y_{i2}, Y_{i1}, Z_i; f) \\
&\quad + \mathbb{1}\{D_{i2} = D_{i1} = 1\} \cdot l_2(Y_{i3}, Y_{i2}, Y_{i1}, Z_i; f) - l_3(f) - l_4(f), \\
l_1(Y_{i2}, Y_{i1}, Z_i; f) &:= \log \int f_1(Y_{i2} \mid Y_{i1}, u) f_2(1 \mid Y_{i1}, u) f_3(Y_{i1}, u) f_4(Z_i \mid u) d\mu(u), \\
l_2(Y_{i3}, Y_{i2}, Y_{i1}, Z_i; f) &:= \log \int f_1(Y_{i3} \mid Y_{i2}, u) f_1(Y_{i2} \mid Y_{i1}, u) f_2(1 \mid Y_{i2}, u) f_2(1 \mid Y_{i1}, u) \\
&\quad \times f_3(Y_{i1}, u) f_4(Z_i \mid u) d\mu(u), \\
l_3(f) &:= \int f_2(1 \mid y_1, u) f_3(y_1, u) d\mu(y_1, u), \qquad \text{and} \\
l_4(f) &:= \int f_1(y_2 \mid y_1, u) f_2(1 \mid y_2, u) f_2(1 \mid y_1, u) f_3(y_1, u) d\mu(y_2, y_1, u).
\end{aligned}$$

Note that Y_{i3} and Y_{i2} may be missing in data, but the interactions with the indicators $\mathbb{1}\{D_{i1} = 1\}$ and $\mathbb{1}\{D_{i2} = D_{i1} = 1\}$ allow the expression $l(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}; f)$ to make sense even if they are missing.

Besides the identifying Restrictions 1, 2, 3, and 4 for the model set $\mathcal{F}$, we require additional technical assumptions, stated in the appendix for brevity of exposition, to guarantee a well-behaved estimator in large samples.

PROPOSITION 1 (Consistency). *Suppose that $\mathcal{F}$ satisfies Restrictions 1, 2, 3, 4, and the assumptions under Section 9.2. If, in addition, Assumptions 1, 2, and 3 in Section 9.3 restrict the model set $\mathcal{F}$, choice of the sieves $\{\mathcal{F}_{k(n)}\}_{n=1}^{\infty}$, and the data $F_{Y_3Y_2Y_1ZD_2D_1}$, then $\left\|\hat{f} - f_0\right\| = o_p(1)$ holds, where this norm $\|\cdot\|$ is defined in Section 9.3.*

The estimator can also be adapted to semi-parametric and parametric sub-models which can be more relevant for empirical analysis. Section 9.4 introduces a semi-parametric estimator and its asymptotic distribution. Parametric models may be estimated with the standard M-estimation theory.

5.3. Monte Carlo Evidence. This section shows Monte Carlo evidence to evaluate the estimation method proposed in this paper. The endogenously unbalanced panel data of $N = 1,000$ and $T = 3$ are generated using the following DGP:

$$\begin{cases} Y_t = \alpha_1 Y_{t-1} + U + \mathcal{E}_t & \mathcal{E}_t \sim \text{Normal}(0, \alpha_2) \\ D_t = \mathbb{1}\{\beta_0 + \beta_1 Y_t + \beta_2 U + V_t \geq 0\} & V_t \sim \text{Logistic}(0,1) \\ F_{Y_1 U} & (Y_1, U) \sim \text{Normal}\left(\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \begin{pmatrix} \gamma_3^2 & \gamma_3\gamma_4\gamma_5 \\ \gamma_3\gamma_4\gamma_5 & \gamma_4^2 \end{pmatrix}\right) \\ Z = \mathbb{1}\{\delta_0 + \delta_1 U + W \geq 0\} & W \sim \text{Normal}(0,1) \end{cases}$$

Monte Carlo simulation results of the constrained maximum likelihood estimation are displayed in the first four rows of Table 1.1. The first row shows simulated distributions of parameter estimates by the fully-parametric estimation using the true model. The inter-quartile ranges capture the true parameter value of 0.5 without suffering from attrition bias. The second row shows simulated distributions of parameter estimates by semiparametric estimation, where the distribution of F_{Y_1U} is assumed to be semiparametric with normality of the conditional distribution $F_{U|Y_1}$. The third

True Parameter Values: $\alpha_1 = \alpha_2 = \beta_0 = \beta_1 = \beta_2 = 0.5$

		Dynamic Model		Hazard Model		
	Percentile	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
Parametric CMLE	75%	0.556	0.519	0.673	0.855	1.293
	50%	0.502	0.502	0.517	0.523	0.569
	25%	0.454	0.483	0.403	0.169	−0.182
Semi-parametric* CMLE	75%	0.566	0.524	0.758	0.882	1.212
	50%	0.513	0.508	0.555	0.523	0.475
	25%	0.465	0.489	0.428	0.153	−0.288
Semi-parametric** CMLE	75%	0.558	—	—	0.589	
	50%	0.459	—	—	0.418	0.500
	25%	0.368	—	—	0.204	(Fixed)
Semi-parametric*** CMLE	75%	0.686	—	—	1.049	
	50%	0.436	—	—	0.719	0.500
	25%	0.271	—	—	0.403	(Fixed)
Semi-parametric† 1st Step	75%	0.585	—	—	—	—
	50%	0.495	—	—	—	—
	25%	0.385	—	—	—	—
Arellano-Bond	75%	0.464	—	—	—	—
	50%	0.412	—	—	—	—
	25%	0.352	—	—	—	—
Fixed-Effect Logit	75%	—	—	—	−0.134	—
	50%	—	—	—	−0.287	—
	25%	—	—	—	−0.441	—
Random-Effect Logit	75%	—	—	—	0.793	—
	50%	—	—	—	0.729	—
	25%	—	—	—	0.672	—

$*$ The distribution of F_{Y_1U} is semi-parametric.

$**$ The distributions of $\mathcal{E}_t$ and V_t are nonparametric.

$***$ The distribution of F_{Y_1U} is semi-parametric, and the distributions of $\mathcal{E}_t$ and V_t are nonparametric.

† The distribution $(\mathcal{E}_t, V_t, Y_1, U)$ and the functions g and h are nonparametric.

TABLE 1.1. MC-simulated distributions of parameter estimates.

row shows simulated distributions of parameter estimates by semiparametric estimation, where the distributions of $\mathcal{E}_t$ and V_t are assumed to be unknown. The fourth row shows simulated distributions of parameter estimates by semiparametric estimation combining the above two semiparametric assumptions. While the medians are slightly off the true values, the inter-quartile ranges again capture the true parameter value of 0.5.

If one is interested in only the dynamic model g, then the sample-analog estimation of the first step in the six-step identification strategy can be used instead of the constrained maximum likelihood. With the notations from Section 3.1, minimizing $\rho(\hat{L}_{y,1}\hat{L}_{y,0}^{-1},\ P_y(\alpha)Q_1Q_0^{-1}P_y^{-1}(\alpha))$ for some divergence measure or a metric ρ yields the first-step estimation. Using the square-integrated difference for ρ, the fifth row of Table 1.1 shows a semiparametric first-step estimation for the dynamic model g. The interquartile range of the MC-simulated distribution indeed captures the true value of 0.5, and it is reasonably tight for a semiparametric estimator. But why is the interquartile range of this first-step estimator tighter than those of the CMLE in the first four rows, despite the greater degree of nonparametric specification? Recall that the first-step estimation uses only two semi-/non-parametric functions (g, ζ) because the other elements have been nonparametrically differenced out. This is to be contrasted with the CMLE, which uses all the four semi-/non-parametric functions (g, h, F_{Y_1U}, ζ). The first-step estimator therefore uses less sieve spaces than the CMLE, and incurs smaller mean square errors in finite sample.

If there were no missing observations from attrition, existing methods such as Arellano and Bond (1991) would consistently estimate α_1. Similarly, because V_t follows the logistic distribution, the fixed-effect logit method (which is the only $\sqrt{N}$-consistent binary response estimator in particular; Chamberlain, 2010) would consistently estimate β_1 if the counterfactual binary choice of dynamic selection were observable after attrition. However, missing data from attrition causes these estimators to be biased as shown in the bottom three rows of Table 1.1. Observe that the fixed effect logit estimator is not only biased, but the sign is even opposite to the truth. This fact

evidences that ignorance of selection could lead to a misleading result, even if the true parametric and distributional model is known.

6. Empirical Illustration: SES and Mortality

6.1. Background. A large number of biological and socio-economic elements help to explain mortality (Cutler, Deaton, and Lleras-Muney, 2006). Among others, measures of socioeconomic status (SES) including earnings, employment, and income are important, yet puzzling as presumable economic determinants of mortality. The literature has reached no consensus on the sign of the effects of these measures of SES on mortality. On one hand, higher SES seems to play a malignant role. For example, at the macroeconomic unit of observations, recessions reduce mortality (Ruhm, 2000). For another example, higher social security income induces higher mortality (Snyder and Evans, 2006). On the other hand, higher SES has been reported to play a protective role. Deaton and Paxson (2001) and Sullivan and von Wachter (2009a) show that higher income reduces mortality. Job displacement, which results in a substantial drop in income, induces higher and long-lasting mortality (Sullivan and von Wachter, 2009b). The apparent discrepancy of the signs may be reconciled by the fact that these studies consider different sources of income and different units of observations.

A major concern in empirical analysis is the issue of endogeneity. Design-based empirical analysis often provides a solution. However, while non-labor income may allow exogenous variations to facilitate natural and quasi-experimental studies (e.g., Snyder and Evans, 2006), labor outcome is often harder to control exogenously. An alternative approach is to explicitly control the common factors that affect both SES and mortality. Education, in particular, is an important observable common factor, e.g., Lleras-Muney (2005) reports causal effects of education on adult mortality. Controlling for this common factor may completely remove the effect of income on mortality. For example, Adams, Hurd, McFadden, Merrill, and Ribeiro (2003) show that income conditional on education is not correlated with health.

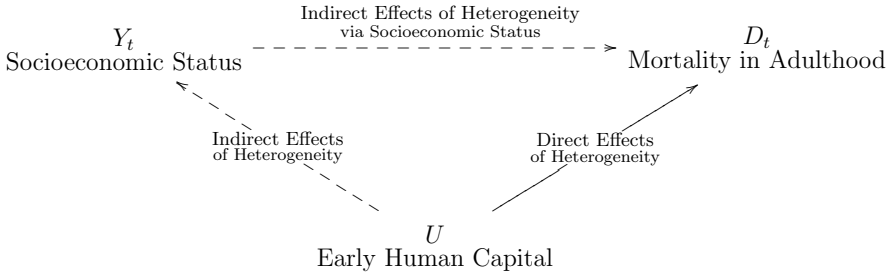


FIGURE 1.3. Causal relationships among early human capital, socioeconomic status, and mortality in adulthood.

Education may play a substantial role, but it may not be the only common factor that needs to be controlled for. A wide variety of early human capital (HC) besides those reflected on education is considered to affect SES and/or adult health in the long run. Case, Fertig, and Paxson (2005) report long-lasting direct and indirect effects of childhood health on health and well-being in adulthood. Maccini and Yang (2009) find that the natural environment at birth affects adult health. Almond and Mazumder (2005) and Almond (2006) show evidence that HC acquired in utero affects long-run health. Early HC could contain a wide variety of categories of HC, such as genetic expression, acquired health, knowledge, and skills, all of which develop in an interactive manner with inter-temporal feedbacks during childhood (e.g., Heckman, 2007; Cunha and Heckman, 2008; Cunha, Heckman, and Schennach, 2010).

Failure to control for these components of early HC would result in identification of "spurious dependence" (e.g., Heckman, 1981ab, 1991). Early HC may directly affect adult mortality via the development of chronic conditions in childhood. Early HC may also affect earnings, which may in turn affect adult mortality, as illustrated below.

Identification of these two competing causal effects of Y_t and U on D_t, or distinction between the two channels in the above diagram, requires to control for the unobserved heterogeneity of early HC. Unlike education, however, most components of early HC are unobservable from the viewpoint of econometricians. Suppose that early HC develops into fixed characteristics by adulthood. How can we control for these heterogeneous characteristics? Because of the strong cross-section correlation

between SES and these heterogeneous characteristics, a variation over time is useful to disentangle their competing effects on mortality, e.g., Deaton and Paxson (2001) and Sullivan and von Wachter (2009a). I extend these ideas by treating early HC as a fixed unobserved heterogeneity to be distinguished from time-varying observed measures of SES. To account for both nonseparable heterogeneity and the survivorship bias, I use the econometric method developed in this paper.

6.2. Empirical Model. Sullivan and von Wachter (2009b) show elaborate evidence on the malignant effects of job displacement on mortality, carefully ruling out the competing hypothesis of selective displacement. Sullivan and von Wachter (2009a), focusing on income as a measure of SES, find that there are protective effects of higher SES on mortality whereas there is no or little evidence of causal effects of unobserved attributes such as patience on mortality. Using the econometric methods developed in this paper, I attempt to complement these analyses by explicitly modeling unobserved heterogeneity and survival selection.

The following econometric model represents the causal relationship described in the above diagram.

$$\begin{cases} (i) & Y_{it} = g(Y_{i,t-1}, U_i, \mathcal{E}_{it}) & \text{SES Dynamics} \\ (ii) & D_{it} = h(Y_{it}, U_i, V_{it}) & \text{Survival Selection} \\ (iii) & F_{Y_1 U} & \text{Initial Condition} \\ (iv) & Z_i = \zeta(U_i, W_i) & \text{Nonclassical Proxy} \end{cases}$$

where Y_{it}, D_{it}, and U_i denote SES, survival, and unobserved heterogeneity, respectively. As noted earlier, the heterogeneity U reflects early human capital (HC) acquired prior to the start of the panel data, which play the role of sustaining employment dynamics in model (i). This early HC may include acquired and innate abilities, knowledge, skills, patience, diligence, and chronic health conditions, which may affect the survival selection (ii) as well as the income dynamics. The initial condition

(iii) models a statistical summary of the initial observation of SES that has developed cumulatively and dependently on the early HC prior to the first observation by econometrician.

For this empirical application, we consider the model in which all the random variables are binary as in Section 3. Specifically, Y_{it} indicates that individual i is (0) unemployed or (1) employed, D_{it} indicates that individual i is (0) dead or (1) alive, and U_i indicates that individual i belongs to (0) type I or (1) type II. Several proxy variables are used for Z_i as means of showing robustness of empirical results. The heterogeneous type U does not yet have any intrinsic meaning at this point, but it turns out empirically to have a consistent meaning in terms of a pattern of employment dynamics as we will see in Section 6.4.

Besides unobserved heterogeneity, other main sources of endogeneity in analysis of SES and mortality are cohort effects and age effects. In parametric regression analysis, one can usually control for these effects by inserting additive dummies or polynomials of age. Since additive controls are infeasible for our setup of nonseprable models, we implement the econometric analysis for each bin of age categories in order to mitigate the age and cohort effects.

6.3. Data. The NLS Original Cohorts: Older Men consist of 5,020 individuals aged 46–60 as of April 1, 1966. The subjects were surveyed annually or biennially starting in 1966. Attrition is frequent in this panel data. In order for the selection model to exactly represent the survival selection, we remove those individuals with attrition due to reasons other than death.

It is important to rule out competing hypotheses that obscure the credibility of our empirical results. For example, health as well as wealth is an important factor of retirement deicision (Bound, Stinebrickner, and Waidmann, 2010). It is not unlikely that individuals who have chosen to retire from jobs for health problems subsequently die. If so, we would erroneously impute death to voluntary retirements. To eliminate this confounding factor, we consider the subsample of individuals who reported that

health problems do not limit work in 1971. Furthermore, we also consider the subsample of individuals who died from acute diseases such as heart attacks and strokes, because workers dying unexpectedly from acute diseases are less likely to change labor status before a sudden death than those who die from cancer or diabetes. Death certificates are used to classify causes of deaths to this end.

Recall that the econometric methods presented in this paper offer two paths of identification. One is to use a panel of $T = 3$ with a nonclassical proxy variable, and the other is to use a panel of $T = 6$ without a proxy. While the the survey was conducted at more than six time points, the list of survey years do not exhibit equal time intervals (1966, 67, 68, 69, 71, 73, 75, 76, 78, 80, 81, 83, and 90). None of annual or biennial sequences consist consecutive six periods from this anomalistic list of years. Therefore we choose the method of proxy variables. Because one of the proxy variables is collected only once in 1973, we need to set T=1 or T=2 to year 1973 in order to satisfy the identifying restriction. We thus set $T = 2$ to year 1973 to exploit a larger size of data, hence using the three-period data from years 71, 73, and 75 in our analysis. The subjects are aged 51–65 in 1971, but we focus on the younger cohorts not facing the retirement age.

We use height, mother's occupation, and father's occupation, as potential candidates for proxy variables. Height reflects health investments in childhood (Schultz, 2002). Mother's education and father's occupation reflect early endowments and investments in human capital in the form of intergenerational inheritance; e.g., Currie and Moretti (2003) show evidence of intergenerational transmission of human capital. We use these three proxies to aim to show robustness of our empirical results.

6.4. Empirical Results. Table 1.2 summarizes estimates of the first-order process of employment dynamics and the conditional two-year survival probabilities using height as a proxy variable. The top and bottom panels correspond to younger cohorts (aged 51–54 in 1971) and older cohorts (aged 55–58 in 1971), respectively. The left and right columns correspond to Type I ($U_i = 0$) and Type II ($U_i = 1$), respectively. These unobserved types exhibit a consistent pattern: off-diagonal elements

Birth year cohorts 1917–1920 (aged 51–54 in 1971)

$N = 822$

Type I ($U = 0$)

54.2%

Markov Matrix		Y_{t-1} = 0	Y_{t-1} = 1
Y_t	0	0.930	0.128
	1	0.070	0.872

2-Year Survival Probability

Y_t	0	0.899 (0.044)
	1	1.000 (0.000)
H_0: equal probability		
p-value = 0.021**		

Type II ($U = 1$)

45.8%

Markov Matrix		Y_{t-1} = 0	Y_{t-1} = 1
Y_t	0	1.000	0.025
	1	0.000	0.975

2-Year Survival Probability

Y_t	0	0.878 (0.149)
	1	0.999 (0.038)
H_0: equal probability		
p-value = 0.445		

Birth year cohorts 1913–1916 (aged 55–58 in 1971)

$N = 727$

Type I ($U = 0$)

53.9%

Markov Matrix		Y_{t-1} = 0	Y_{t-1} = 1
Y_t	0	0.954	0.162
	1	0.046	0.838

2-Year Survival Probability

Y_t	0	0.912 (0.036)
	1	1.000 (0.000)
H_0: equal probability		
p-value = 0.013**		

Type II ($U = 1$)

46.1%

Markov Matrix		Y_{t-1} = 0	Y_{t-1} = 1
Y_t	0	1.000	0.066
	1	0.000	0.934

2-Year Survival Probability

Y_t	0	0.890 (0.107)
	1	0.983 (0.042)
H_0: equal probability		
p-value = 0.416		

TABLE 1.2. Model estimates with height as a proxy.

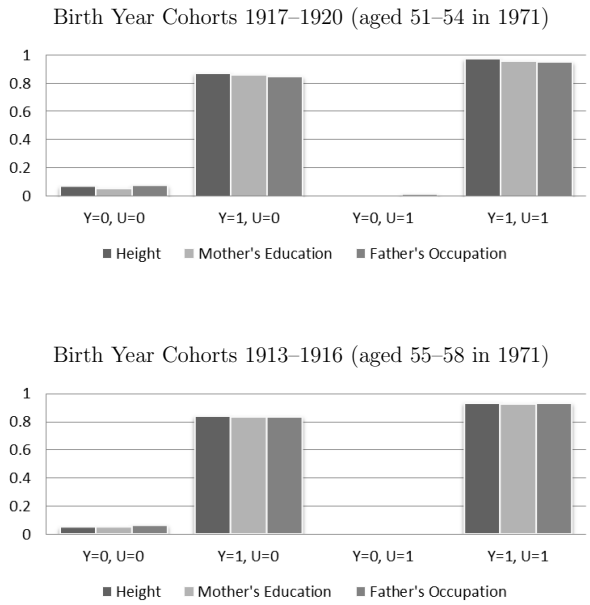


FIGURE 1.4. Markov probabilities of employment in the next two years.

of the employment Markov matrices for Type I dominate those of Type II. In other words, Type I and Type II can be characterized as movers and stayers, respectively. In view of the survival probabilities in the top panel (young cohorts), we find that individuals almost surely stay alive as far as they are employed. On the other hand, the two-year survival probabilities drop by about 10% if individuals are unemploed. While the data indicates statistical significance of their difference only for Type I, the magnitudes of differences in the point estimates are almost identical between the two types. The same qualitative pattern persists in the older cohorts.

To show a robustness of this baseline result, we repeat this estimation using the other two proxy variables, mother's education and father's occupation. Figure 1.4 graphs estimates of Markov probabilities of employment. The shades in the bars indicate different proxy variables used for estimation. We see that these point estimates are robust across the three proxy variables, implying that a choice of a particular proxy does not lead to an irregular result in favor of certain claims. Table 1.5 graphs estimates of conditional two-year survival probabilities. Again, the point estimates are robust across the three proxy variables.

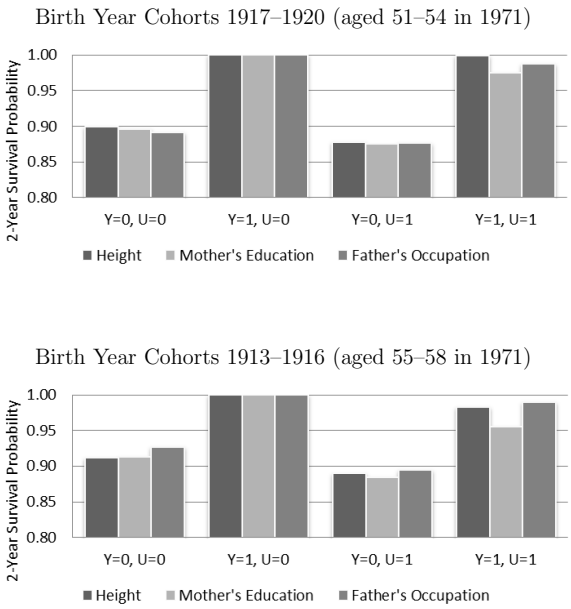


FIGURE 1.5. Conditional survival probabilities in the next two years.

As mentioned earlier, selective or voluntary retirement is a potential source of bias. To rule out this possibility, we consider two subpopulations: 1. those individuals who reported that health problems do not limit their work in 1971; and 2. those individuals who eventually died from acute diseases. Figures 1.6 and 1.7 show estimates for the first subpopulation. Figures 1.8 and 1.9 show estimates for the second subpopulation. Again, robustness across the three proxies persists, and the qualitative pattern remains the same as the baseline result. The relatively large variations in the estimates for the second subpopulation is imputed to small sample sizes due to the limited availability of death certificates from which we identify causes of deaths.

In summary, we obtain the following two robust results. First, accounting for unobserved heterogeneity and survivorship bias as well as voluntary retirements, employment status has protective effects on survival selection. This reinforces the results of Sullivan and von Wachter (2009b). Second, there is no evidence of the effects of unobserved attributes on survival selection, since the conditional survival probabilities are almost the same between type I and type II. This is in accord with the claim

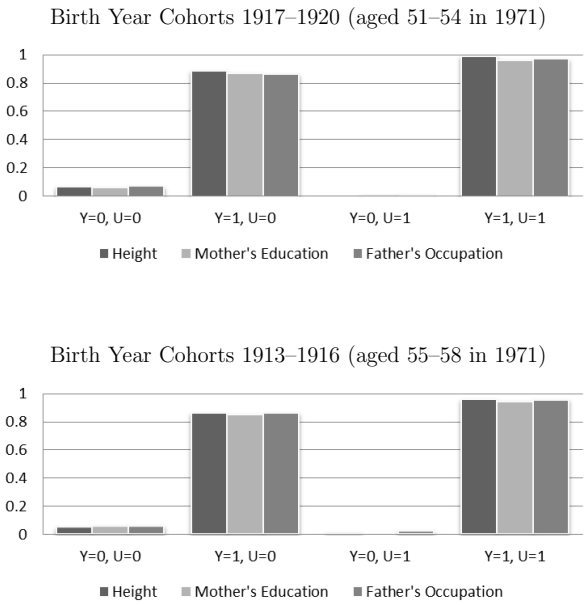


FIGURE 1.6. Markov probabilities of employment in the next two years among the subpopulation of individuals who reported health problems that limit work in 1971.

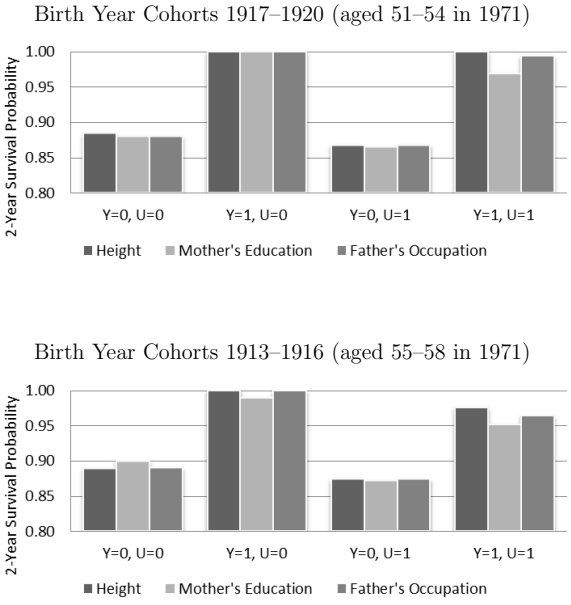


FIGURE 1.7. Conditional survival probabilities in the next two years among the subpopulation of individuals who reported health problems that limit work in 1971.

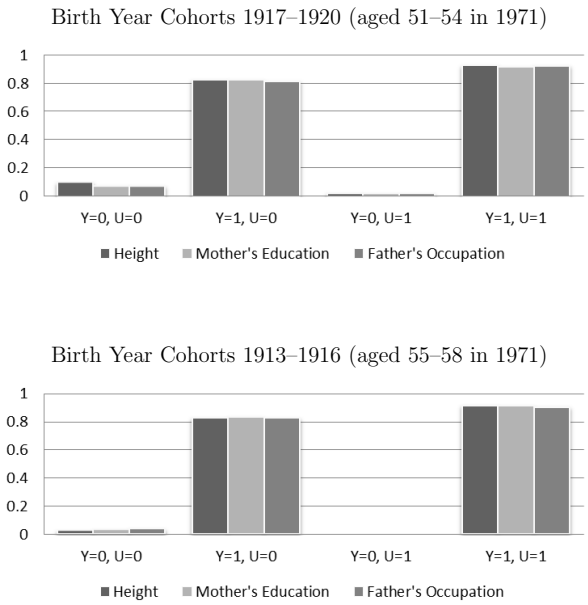


FIGURE 1.8. Markov probabilities of employment in the next two years among the subpopulation of individuals who eventually died from acute diseases according to death certificates.

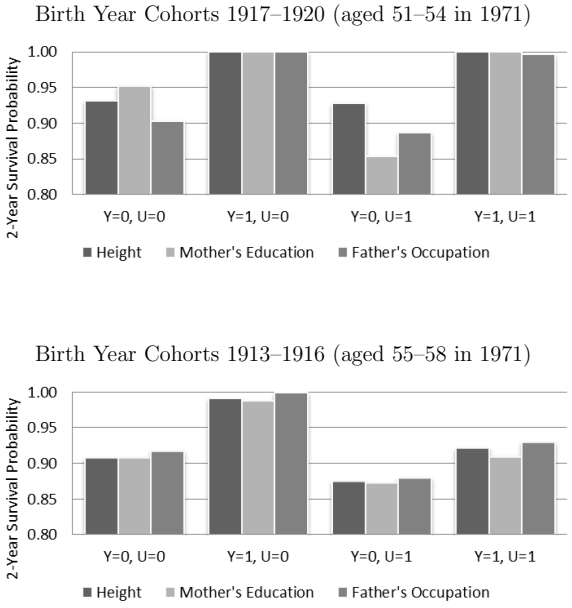


FIGURE 1.9. Conditional survival probabilities in the next two years among the subpopulation of individuals who eventually died from acute diseases according to death certificates.

of Sullivan and von Wachter (2009a), who deduce that lagged SES has little effect on mortality conditionally on the SES of immediate past.

Using the estimated Markov model g and the estimated initial condition F_{Y_1U}, we can simulate the counterfactual employment rates assuming that all the individuals were to remain alive throughout the entire period. Figure 1.10 shows actual employment rates (black lines) and counterfactual employment rates (grey lines) for each cohort category for each proxy variable. I again remark that the qualitative patterns are the same across three proxy variables for each cohort category. Not shockingly, if it were not for deaths, the counterfactual employment rates would have been even lower than what we observed from actual data. In other words, deaths of working age population are saving the actual figures of employment rates to look higher.

7. Summary

This paper proposes a set of nonparametric restrictions to point-identify dynamic panel data models by nonparametrically differencing out both nonseparable heterogeneity and selection. Identification requires either $T = 3$ periods of panel data and a proxy variable or $T = 6$ periods of panel data without an outside proxy variable. As a consequence of the identification result, the constrained maximum likelihood criterion follows, which corrects for selection and allows for one-step estimation. Monte Carlo simulations are used to evidence the effectiveness of the estimators. In the empirical application, I find protective effects of employment on survival selection, and the result is robust.

8. Appendix: Proofs for Identification

8.1. Lemma 1 (Representation).

Proof. (i) First, we show that there exists a function $\bar{g}$ such that $(\bar{g}, \text{Uniform}(0,1))$ is observationally equivalent to $(g, F_{\mathcal{E}})$ for any $(g, F_{\mathcal{E}})$ satisfying Restrictions 1 and 2. By the absolute continuity and the convex support in Restriction 1, $F_{\mathcal{E}}$ is invertible. Hence, we can define $h := F_{\mathcal{E}}^{-1}$. Now, define $\bar{g}$ by $\bar{g}(y, u, \cdot\,) := g(y, u, \cdot\,) \circ h^{-1}$ for

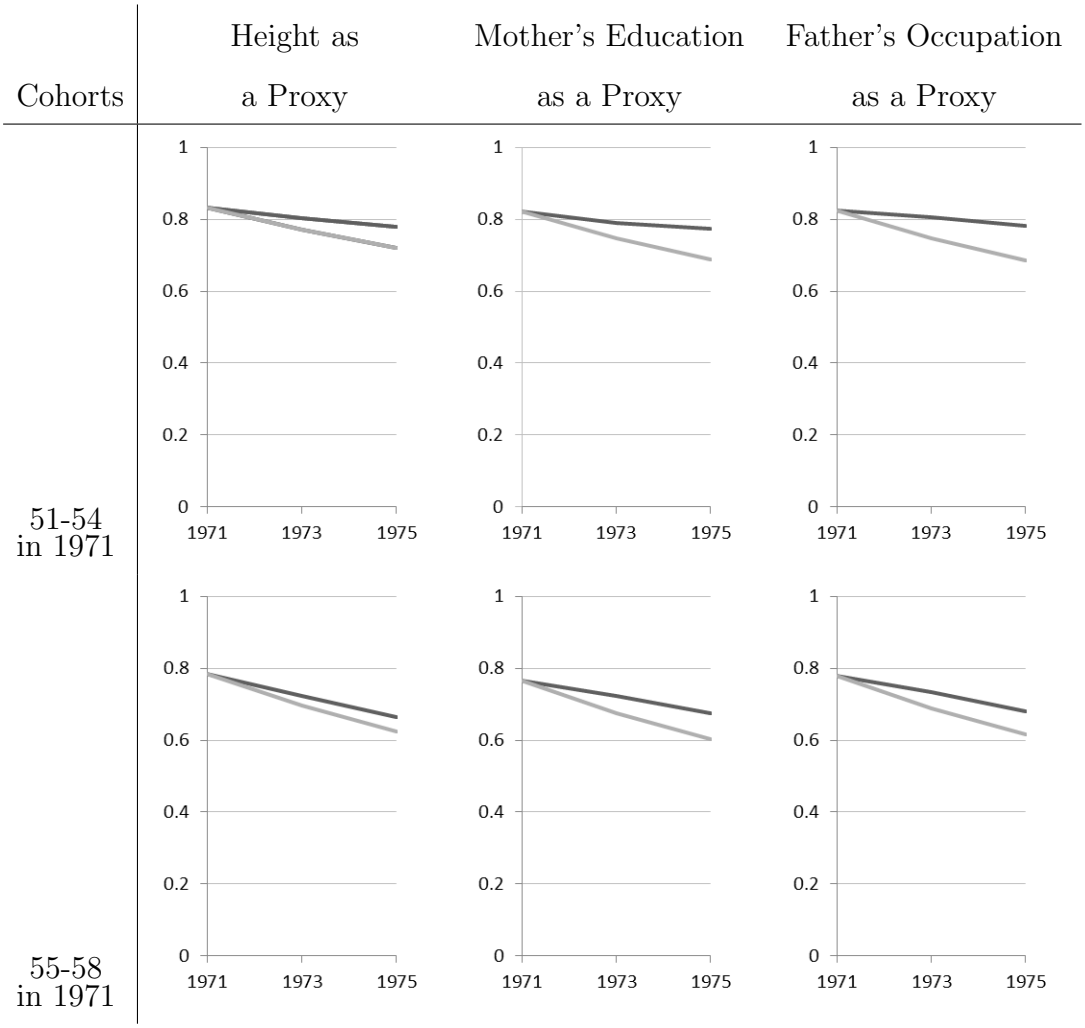


FIGURE 1.10. Counterfactual simulations.

each (y, u). Note that, under Restriction 2, $(\bar{g}, F_{h(\mathcal{E})})$ is observationally equivalent to $(g, F_{\mathcal{E}})$ by construction. However, we have $h(\mathcal{E}) \sim \text{Uniform}(0, 1)$ by the definition of h. It follows that $(\bar{g}, \text{Uniform}(0, 1))$ is observationally equivalent to $(g, F_{\mathcal{E}})$.

In light of the previous paragraph, we can impose the normalization $\mathcal{E}_t \sim \text{Uniform}(0, 1)$. Let $\Lambda(y, u, \varepsilon)$ denote the set $\Lambda(y, u, \varepsilon) = \{y' \in g(y, u, (0, 1)) \mid \varepsilon \leq F_{Y_3|Y_2U}(y' \mid y, u)\}$, where $g(y, u, (0, 1))$ denotes the set $\{g(y, u, \varepsilon) \mid \varepsilon \in (0, 1)\}$. I claim that $g(y, u, \varepsilon) = \inf \Lambda(y, u, \varepsilon)$.

First, we note that $g(y, u, \varepsilon) \in \Lambda(y, u, \varepsilon)$. To see this, calculate

$$\begin{aligned} F_{Y_3|Y_2U}(g(y, u, \varepsilon) \mid y, u) &= \Pr(g(y, u, \mathcal{E}_3) \leq g(y, u, \varepsilon) \mid Y_2 = y, U = u) \\ &= \Pr(g(y, u, \mathcal{E}_3) \leq g(y, u, \varepsilon)) \geq \Pr(\mathcal{E}_3 \leq \varepsilon) = \varepsilon, \end{aligned}$$

where the first equality follows from $Y_3 = g(y, u, \mathcal{E}_3)$ given $(Y_2, U) = (y, u)$, the second equality follows from Restriction 2 (i), the next inequality follows from the non-decrease of $g(y, u, \cdot)$ by Restriction 1 together with monotonicity of the probability measure, and the last equality is due to $\mathcal{E}_t \sim U(0, 1)$. This shows that $\varepsilon \leq F_{Y_3|Y_2U}(g(y, u, \varepsilon) \mid y, u)$, hence $g(y, u, \varepsilon) \in \Lambda(y, u, \varepsilon)$.

Second, I show that $g(y, u, \varepsilon)$ is a lower bound of $\Lambda(y, u, \varepsilon)$. Let $y' \in \Lambda(y, u, \varepsilon)$. Since g is non-decreasing and càglàd (left-continuous) in the third argument by Restriction 1, we can define $\varepsilon' := \max\{\varepsilon \in (0, 1) \mid g(y, u, \varepsilon) = y'\}$. But then,

$$\begin{aligned} F_{Y_3|Y_2U}(y' \mid y, u) &= \Pr(g(y, u, \mathcal{E}_3) \leq y' \mid Y_2 = y, U = u) \\ &= \Pr(g(y, u, \mathcal{E}_3) \leq y') \;=\; \varepsilon', \end{aligned}$$

where the first equality follows from $Y_3 = g(y, u, \mathcal{E}_3)$ given $(Y_2, U) = (y, u)$, the second equality follows from Restriction 2 (i), and the last equality follows from the definition of ε' together with the non-decrease of $g(y, u, \cdot)$ by Restriction 1 and $\mathcal{E}_t \sim U(0, 1)$. Using this result, in turn, yields

$$g(y, u, \varepsilon) \;\leq\; g(y, u, F_{Y_3|Y_2U}(y' \mid y, u)) \;=\; g(y, u, \varepsilon') \;=\; y',$$

where the first inequality follows from $\varepsilon \leq F_{Y_3|Y_2U}(y' \mid y, u)$ by definition of y' as well as the non-decrease of $g(y, u, \cdot)$ by Restriction 1, the next equality follows from the previous result $F_{Y_3|Y_2U}(y' \mid y, u) = \varepsilon'$, and the last equality follows from the definition of ε'. Since y' was chosen as an arbitrary element of $\Lambda(y, u, \varepsilon)$, this shows that $g(y, u, \varepsilon)$ is indeed a lower bound of it. Therefore, $g(y, u, \varepsilon) = \inf\{y' \in g(y, u, (0, 1)) \mid \varepsilon \leq F_{Y_3|Y_2U}(y' \mid y, u)\}$, and g is uniquely determined by $F_{Y_3|Y_2U}$. (Moreover, note that $\inf\{y' \in g(y, u, (0, 1)) \mid \varepsilon \leq F_{Y_3|Y_2U}(y' \mid y, u)\}$ coincides with the definition of the quantile regression $F^{-1}_{Y_3|Y_2U}(\,\cdot \mid y, u)$, hence g is identified by this quantile regression, i.e., $g(y, u, \varepsilon) = F^{-1}_{Y_3|Y_2U}(\varepsilon \mid y, u)$.)

Part (ii) of the lemma can be proved in exactly the same way as in the proof of part (i). In particular, h is identified by the quantile regression: $h(y, u, v) = F^{-1}_{D_2|Y_2U}(v \mid$

$y, u)$. Similarly, part (iii) of the the lemma can be proved in the same way, and ζ is identified by the quantile regression: $\zeta(u, w) = F^{-1}_{Z|U}(w \mid u)$. □

8.2. Lemma 2 (Identification).

PROOF. We will construct six steps for a proof of this lemma. The first step shows that the observed joint distributions uniquely determine $F_{Y_3|Y_2U}$ and $F_{Z|U}$ by a spectral decomposition of a composite linear operator. The second step is auxiliary, and shows that $F_{Y_2Y_1UD_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,1)$ is uniquely determined from the observed joint distributions together with inversion of the operators identified in the first step. The third step again uses spectral decomposition to identify an auxiliary operator with the kernel represented by $F_{Y_1|Y_2UD_2D_1}(\,\cdot\,\mid\,\cdot\,,\,\cdot\,,1,1)$. In the fourth step, solving an integral equation with the adjoint of this auxiliary operator in turn yields another auxiliary operator with the multiplier represented by $F_{Y_2UD_2D_1}(\,\cdot\,,\,\cdot\,,1,1)$. The fifth step uses the three operators identified in Steps 2, 3, and 4 to identify an operator with the kernel represented by $F_{D_2|Y_2U}$ by solving a linear inverse problem. The last step uses results from Steps 1, 2, and 5 to show that the initial joint distribution F_{Y_1U} is uniquely determined from the observed joint distributions. These six steps together prove that the observed joint distributions $F_{Y_2Y_1ZD_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,1)$ and $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1)$ uniquely determine the model $(F_{Y_3|Y_2U}, F_{D_2|Y_2U}, F_{Y_1U}, F_{Z|U})$ as claimed in the lemma.

Given fixed y and z, define the operators $L_{y,z} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_{Y_t})$, $P_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$, $Q_z : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$, $R_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$, $S_y : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$,

$T_y : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$, and $T'_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$ by

$$
\begin{aligned}
(L_{y,z}\xi)(y_3) &= \int f_{Y_3Y_2Y_1ZD_2D_1}(y_3, y, y_1, z, 1, 1) \cdot \xi(y_1) dy_1, \\
(P_y\xi)(y_3) &= \int f_{Y_3|Y_2U}(y_3 \mid y, u) \cdot \xi(u) du, \\
(Q_z\xi)(u) &= f_{Z|U}(z \mid u) \cdot \xi(u), \\
(R_y\xi)(u) &= f_{D_2|Y_2U}(1 \mid y, u) \cdot \xi(u), \\
(S_y\xi)(u) &= \int f_{Y_2Y_1UD_1}(y, y_1, u, 1) \cdot \xi(y_1) dy_1, \\
(T_y\xi)(u) &= \int f_{Y_1|Y_2UD_2D_1}(y_1 \mid y, u, 1, 1) \cdot \xi(y_1) dy_1, \\
(T'_y\xi)(u) &= f_{Y_2UD_2D_1}(y, u, 1, 1) \cdot \xi(u)
\end{aligned}
$$

respectively. We consider $\mathcal{L}^2$ spaces as the normed linear spaces on which these operators are defined, particularly in order to guarantee the existence of its adjoint operator T_y^* to be introduced in Step 4. (Recall that a bounded linear operator between Hilbert spaces admits existence of its adjoint operator.) Identification of the operator leads to that of the associated conditional density (up to null sets), and vice versa. Here, the operators $L_{y,z}$, P_y, S_y, and T_y are integral operators whereas Q_z, R_y, and T'_y are multiplication operators. Note that $L_{y,z}$ is identified from observed joint distribution $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$.

Figure 3.1 illustrates six steps toward identification of $(F_{Y_3|Y_2U}, F_{D_2|Y_2U}, F_{Y_1U}, F_{Z|U})$ from the observed joint distributions $F_{Y_2Y_1ZD_1}(\,\cdot,\,\cdot,\,\cdot,1)$ and $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1)$. The four objects (g, h, F_{Y_1U}, ζ) of interest are enclosed by double lines. The objects that can be observed from data are enclosed by dashed-lines All the other objects are intermediary, and are enclosed by solid lines. Starting out with the observed objects, we show in each step that the intermediary objects are uniquely determined. These uniquely determined intermediary objects in turn show the uniqueness of the four objects (g, h, F_{Y_1U}, ζ) of interest.

Step 1: Uniqueness of $F_{Y_3|Y_2U}$ and $F_{Z|U}$

The kernel $f_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,, y,\,\cdot\,, z, 1, 1)$ of the integral operator $L_{y,z}$ can be rewritten

as

$$
\begin{aligned}
f_{Y_3Y_2Y_1ZD_2D_1}(y_3, y, y_1, z, 1, 1) &= \int f_{Y_3|Y_2Y_1ZUD_2D_1}(y_3 \mid y, y_1, z, u, 1, 1) \\
&\quad \times f_{Z|Y_2Y_1UD_2D_1}(z \mid y, y_1, u, 1, 1) \\
&\quad \times f_{D_2|Y_2Y_1UD_1}(1 \mid y, y_1, u, 1) \cdot f_{Y_2Y_1UD_1}(y, y_1, u, 1)\, du
\end{aligned} \tag{13}
$$

But by Lemma 3 (i), (iv), and (iii), respectively, Restriction 2 implies that

$$
\begin{aligned}
f_{Y_3|Y_2Y_1ZUD_2D_1}(y_3 \mid y, y_1, z, u, 1, 1) &= f_{Y_3|Y_2U}(y_3 \mid y, u), \\
f_{Z|Y_2Y_1UD_2D_1}(z \mid y, y_1, u, 1, 1) &= f_{Z|U}(z \mid u), \\
f_{D_2|Y_2Y_1UD_1}(1 \mid y, y_1, u, 1) &= f_{D_2|Y_2U}(1 \mid y, u).
\end{aligned}
$$

Equation (13) thus can be rewritten as

$$
\begin{aligned}
f_{Y_3Y_2Y_1ZD_2D_1}(y_3, y, y_1, z, 1, 1) &= \int f_{Y_3|Y_2U}(y_3 \mid y, u) \cdot f_{Z|U}(z \mid u) \\
&\quad \times f_{D_2|Y_2U}(1 \mid y, u) \cdot f_{Y_2Y_1UD_1}(y, y_1, u, 1)\, du
\end{aligned}
$$

But this implies that the integral operator $L_{y,z}$ is written as the operator composition

$$L_{y,z} = P_y Q_z R_y S_y.$$

Restriction 3 (i), (ii), (iii), and (iv) imply that the operators P_y, Q_z, R_y, and S_y are invertible, respectively. Hence so is $L_{y,z}$. Using the two values $\{0, 1\}$ of Z, form the product

$$L_{y,1} L_{y,0}^{-1} = P_y Q_{1/0} P_y^{-1}$$

where $Q_{z/z'} := Q_z Q_{z'}^{-1}$ is the multiplication operator with proxy odds defined by

$$(Q_{1/0}\xi)(u) = \frac{f_{Z|U}(1 \mid u)}{f_{Z|U}(0 \mid u)} \xi(u).$$

By Restriction 3 (ii), the operator $L_{y,1} L_{y,0}^{-1}$ is bounded. The expression $L_{y,1} L_{y,0}^{-1} = P_y Q_{1/0} P_y^{-1}$ thus allows unique eigenvalue-eigenfunction decomposition similarly to that of Hu and Schennach (2008).

The distinct proxy odds as in Restriction 3 (ii) guarantee distinct eigenvalues and single dimensionality of the eigenspace associated with each eigenvalue. Within each

of the single-dimensional eigenspace is a unique eigenfunction pinned down by $\mathcal{L}^1$-normalization because of the unity of integrated densities. The eigenvalues $\lambda(u)$ yield the multiplier of the operator $Q_{1/0}$, hence $\lambda(u) = f_{Z|U}(1 \mid u)/f_{Z|U}(0 \mid u)$. This proxy odds in turn identifies $f_{Z|U}(\ \cdot\ \mid u)$ since Z is binary. The corresponding normalized eigenfunctions are the kernels of the integral operator P_y, hence $f_{Y_3|Y_2U}(\ \cdot\ \mid y, u)$. Lastly, Restriction 4 facilitates unique ordering of the eigenfunctions $f_{Y_3|Y_2U}(\ \cdot\ \mid y, u)$ by the distinct concrete values of $u = \lambda(u)$. This is feasible because the eigenvalues $\lambda(u) = f_{Z|U}(1 \mid u)/f_{Z|U}(0 \mid u)$ are invariant from y. That is, eigenfunctions $f_{Y_3|Y_2U}(\ \cdot\ \mid y, u)$ of the operator $L_{y,1}L_{y,0}^{-1}$ across different y can be uniquely ordered in u invariantly from y by the common set of ordered distinct eigenvalues $u = \lambda(u)$.

Therefore, $F_{Y_3|Y_2U}$ and $F_{Z|U}$ are uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_2D_1}(\ \cdot\ ,\ \cdot\ ,\ \cdot\ ,\ \cdot\ , 1, 1)$. Equivalently, the operators P_y and Q_z are uniquely determined for each y and z, respectively.

Step 2: Uniqueness of $F_{Y_2Y_1UD_1}(\ \cdot\ ,\ \cdot\ ,\ \cdot\ , 1)$

By Lemma 3 (ii), Restriction 2 implies $f_{Y_2|Y_1UD_1}(y' \mid y, u, 1) = f_{Y_2|Y_1U}(y' \mid y, u)$. Using this equality, write the density of the observed joint distribution $F_{Y_2Y_1D_1}(\ \cdot\ ,\ \cdot\ , 1)$ as

$$\begin{aligned} f_{Y_2Y_1D_1}(y', y, 1) &= \int f_{Y_2|Y_1UD_1}(y' \mid y, u, 1) f_{Y_1UD_1}(y, u, 1) du \\ &= \int f_{Y_2|Y_1U}(y' \mid y, u) f_{Y_1UD_1}(y, u, 1) du \end{aligned} \tag{14}$$

By Lemma 4 (i), $F_{Y_3|Y_2U}(y' \mid y, u) = F_{Y_2|Y_1U}(y' \mid y, u)$ for all y', y, u. Therefore, we can write the operator P_y as

$$(P_y\xi)(y') = \int f_{Y_3|Y_2U}(y' \mid y, u) \cdot \xi(u) du = \int f_{Y_2|Y_1U}(y' \mid y, u) \cdot \xi(u) du.$$

With this operator notation, it follows from (14) that

$$f_{Y_2Y_1D_1}(\ \cdot\ , y, 1) = P_y f_{Y_1UD_1}(y,\ \cdot\ , 1).$$

By Restriction 3 (i), this operator equation can be solved for $f_{Y_1UD_1}(y,\ \cdot\ , 1)$ as

$$f_{Y_1UD_1}(y,\ \cdot\ , 1) = P_y^{-1} f_{Y_2Y_1D_1}(\ \cdot\ , y, 1) \tag{15}$$

Recall that P_y was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$. The function $f_{Y_2Y_1D_1}(\,\cdot\,,y,1)$ is also uniquely determined by the observed joint distribution $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$ up to null sets. Therefore, (14) shows that $f_{Y_1UD_1}(\,\cdot\,,\,\cdot\,,1)$ is uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$.

Using the solution to the above inverse problem, we can write the kernel of the operator S_y as

$$\begin{aligned} f_{Y_2Y_1UD_1}(y',y,u,1) &= f_{Y_2|Y_1UD_1}(y' \mid y,u,1)\cdot f_{Y_1UD_1}(y,u,1) \\ &= f_{Y_2|Y_1U}(y' \mid y,u)\cdot f_{Y_1UD_1}(y,u,1) \\ &= f_{Y_3|Y_2U}(y' \mid y,u)\cdot f_{Y_1UD_1}(y,u,1) \\ &= f_{Y_3|Y_2U}(y' \mid y,u)\cdot [P_y^{-1} f_{Y_2Y_1D_1}(\,\cdot\,,y,1)](u) \end{aligned}$$

where the second equality follows from Lemma 3 (ii), the third equality follows from Lemma 4 (i), and the forth equality follows from (15). Since $f_{Y_3|Y_2U}$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ and $[P_y^{-1} f_{Y_2Y_1D_1}(\,\cdot\,,y,1)]$ was shown in the previous paragraph to be uniquely determined for each y by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$, it follows that $f_{Y_2Y_1UD_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,1)$ too is uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$. Equivalently, the operator S_y is uniquely determined for each y.

Step 3: Uniqueness of $F_{Y_1|Y_2UD_2D_1}(\,\cdot\,\mid\,\cdot\,,\,\cdot\,,1,1)$

First, note that the kernel of the composite operator $T_y'T_y$ can be written as

$$\begin{aligned} (16)\ f_{Y_2UD_2D_1}(y,u,1,1)\cdot f_{Y_1|Y_2UD_2D_1}(y_1 \mid y,u,1,1) &= f_{Y_2Y_1UD_2D_1}(y,y_1,u,1,1) \\ &= f_{D_2|Y_2Y_1UD_1}(1 \mid y,y_1,u,1)\cdot f_{Y_2Y_1UD_1}(y,y_1,u,1) \\ &= f_{D_2|Y_2U}(1 \mid y,u)\cdot f_{Y_2Y_1UD_1}(y,y_1,u,1) \end{aligned}$$

where the last equality is due to Lemma 3 (iii). But the last expression corresponds to the kernel of the composite operator $R_y S_y$, thus showing that $T_y' T_y = R_y S_y$. But then, $L_{y,z} = P_y Q_z R_y S_y = P_y Q_z T_y' T_y$. Note that the invertibility of R_y and S_y as required by Assumption 3 implies invertibility of T_y' and T_y as well, for otherwise the equivalent composite operator $T_y' T_y = R_y S_y$ would have a nontrivial nullspace.

Using Restriction 3, form the product of operators as in Step 1, but in the opposite order as

$$L_{y,0}^{-1} L_{y,1} = T_y^{-1} Q_{1/0} T_y$$

The disappearance of T_y' is due to commutativity of multiplication operators. By the same logic as in Step 1, this expression together with Restriction 3 (ii) admits unique left eigenvalue-eigenfunction decomposition. Moreover, the point spectrum is exactly the same as the one in Step 1, as is the middle multiplication operator $Q_{1/0}$. This equivalence of the spectrum allows consistent ordering of U with that of Step 1. Left eigenfunctions yield the kernel of T_y pinned down by the normalization of unit integral. This shows that the operator T_y is uniquely determined by the observed joint distribution $F_{Y_3 Y_2 Y_1 Z D_2 D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$.

Step 4: Uniqueness of $F_{Y_2 U D_2 D_1}(\,\cdot\,,\,\cdot\,,1,1)$

Equation (16) implies that

$$\int f_{Y_1|Y_2 U D_2 D_1}(y_1 \mid y, u, 1, 1) \cdot f_{Y_2 U D_2 D_1}(y, u, 1, 1) du = f_{Y_2 Y_1 D_2 D_1}(y, y_1, 1, 1)$$

hence yielding the linear operator equation

$$T_y^* f_{Y_2 U D_2 D_1}(y,\ \cdot, 1, 1) = f_{Y_2 Y_1 D_2 D_1}(y,\ \cdot, 1, 1)$$

where T_y^* denotes the adjoint operator of T_y. Since T_y is invertible, so is its adjoint T_y^*. But then, the multiplier of the multiplication operator T_y' can be given by the unique solution to the above linear operator equation, i.e.,

$$f_{Y_2 U D_2 D_1}(y,\ \cdot, 1, 1) = (T_y^*)^{-1} f_{Y_2 Y_1 D_2 D_1}(y,\ \cdot, 1, 1)$$

T_y hence T_y^* was shown to be uniquely determined by $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ in Step 3, and $f_{Y_2Y_1D_2D_1}(\,\cdot,\,\cdot,1,1)$ is also available from observed data. Therefore, the operator T_y' is uniquely determined by $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$.

Step 5: Uniqueness of $F_{D_2|Y_2U}(1 \mid \,\cdot\,,\,\cdot\,)$

First, the definition of the operators R_y, S_y, T_y, and T_y' and Lemma 3 (iii) yield the operator equality $R_yS_y = T_y'T_y$, where T_y and T_y' have been shown to be uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ in Steps 3 and 4, respectively. Recall that S_y was also shown in Step 2 to be uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$. Restriction 3 (iv) guarantees invertibility of S_y. It follows that the operator inversion $R_y = (R_yS_y)S_y^{-1} = (T_y'T_y)S_y^{-1}$ yields the operator R_y, in turn showing that its multiplier $f_{D_2|Y_2U}(1 \mid y,\,\cdot\,)$ is uniquely determined for each y by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$.

Step 6: Uniqueness of F_{Y_1U}

Recall from Step 2 that $f_{Y_2Y_1UD_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,1)$ is uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$. We can write

$$\begin{aligned} f_{Y_2Y_1UD_1}(y',y,u,1) &= f_{Y_2|Y_1UD_1}(y' \mid y,u,1)f_{D_1|Y_1U}(1 \mid y,u)f_{Y_1U}(y,u) \\ &= f_{Y_2|Y_1U}(y' \mid y,u)f_{D_1|Y_1U}(1 \mid y,u)f_{Y_1U}(y,u) \\ &= f_{Y_3|Y_2U}(y' \mid y,u)f_{D_2|Y_2U}(1 \mid y,u)f_{Y_1U}(y,u), \end{aligned}$$

where the second equality follows from Lemma 3 (ii), and the third equality follows from Lemma 4 (i) and (ii). For a given (y,u), there must exist some y' such that $f_{Y_3|Y_2U}(y' \mid y,u) > 0$ by a property of conditional density functions. Moreover, Restriction 3 (iii) requires that $f_{D_2|Y_2U}(1 \mid y,u) > 0$ for a given y for all u. Therefore, for such a choice of y', we can write

$$f_{Y_1U}(y,u) = \frac{f_{Y_2Y_1UD_1}(y',y,u,1)}{f_{Y_3|Y_2U}(y' \mid y,u)f_{D_2|Y_2U}(1 \mid y,u)}$$

Recall that $f_{Y_3|Y_2U}(\cdot \mid \cdot, \cdot)$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1)$, $f_{Y_2Y_1UD_1}(\cdot, \cdot, \cdot, 1)$ was shown in Step 2 to be uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1)$ and $F_{Y_2Y_1D_1}(\cdot, \cdot, 1)$, and $f_{D_2|Y_2U}(1 \mid \cdot, \cdot)$ was shown in Step 5 to be uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1)$ and $F_{Y_2Y_1D_1}(\cdot, \cdot, 1)$. Therefore, it follows that the initial joint density f_{Y_1U} is uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1)$ and $F_{Y_2Y_1D_1}(\cdot, \cdot, 1)$. □

8.3. Lemma 3 (Independence).

LEMMA 3 (Independence). *The following implications hold:*
(i) Restriction 2 (i) $\Rightarrow \mathcal{E}_3 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2, W) \Rightarrow Y_3 \perp\!\!\!\perp (Y_1, D_1, D_2, Z) \mid (Y_2, U)$.
(ii) Restriction 2 (i) $\Rightarrow \mathcal{E}_2 \perp\!\!\!\perp (U, Y_1, V_1, W) \Rightarrow Y_2 \perp\!\!\!\perp (D_1, Z) \mid (Y_1, U)$.
(iii) Restriction 2 (ii) $\Rightarrow V_2 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1) \Rightarrow D_2 \perp\!\!\!\perp (Y_1, D_1) \mid (Y_2, U)$.
(iv) Restriction 2 (iii) $\Rightarrow W \perp\!\!\!\perp (Y_1, \mathcal{E}_2, V_1, V_2) \Rightarrow Z \perp\!\!\!\perp (Y_2, Y_1, D_2, D_1) \mid U$.

PROOF. In order to prove the lemma, we use the following two properties of conditional independence:
CI.1. $A \perp\!\!\!\perp B$ implies $A \perp\!\!\!\perp B \mid \phi(B)$ for any Borel function ϕ.
CI.2. $A \perp\!\!\!\perp B \mid C$ implies $A \perp\!\!\!\perp \phi(B, C) \mid C$ for any Borel function ϕ.

(i) First, note that Restriction 2 (i) $\mathcal{E}_3 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2, W)$ together with the structural definition $Z = \zeta(U, W)$ implies $\mathcal{E}_3 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2, Z)$. Applying CI.1 to this independence relation $\mathcal{E}_3 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2, Z)$ yields

$$\mathcal{E}_3 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2, Z) \mid (g(Y_1, U, \mathcal{E}_2), U).$$

Since $Y_2 = g(Y_1, U, \mathcal{E}_2)$, it can be rewritten as $\mathcal{E}_3 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2, Z) \mid (Y_2, U)$. Next, applying CI.2 to this conditional independence yields

$$\mathcal{E}_3 \perp\!\!\!\perp (Y_1, h(Y_1, U, V_1), h(Y_2, U, V_2), Z) \mid (Y_2, U).$$

Since $D_t = h(Y_t, U, V_t)$ for each $t \in \{1, 2\}$, it can be rewritten as $\mathcal{E}_3 \perp\!\!\!\perp (Y_1, D_1, D_2, Z) \mid (Y_2, U)$. Lastly, applying CI.2 again to this conditional independence yields

$$g(Y_2, U, \mathcal{E}_3) \perp\!\!\!\perp (Y_1, D_1, D_2, Z) \mid (Y_2, U).$$

Since $Y_3 = g(Y_2, U, \mathcal{E}_3)$, it can be rewritten as $Y_3 \perp\!\!\!\perp (Y_1, D_1, D_2, Z) \mid (Y_2, U)$.

(ii) Note that Restriction 2 (i) $\mathcal{E}_2 \perp\!\!\!\perp (U, Y_1, V_1, W)$ together with the structural definition $Z = \zeta(U, W)$ implies $\mathcal{E}_2 \perp\!\!\!\perp (U, Y_1, V_1, Z)$. Applying CI.1 to this independence relation $\mathcal{E}_2 \perp\!\!\!\perp (U, Y_1, V_1, Z)$ yields

$$\mathcal{E}_2 \perp\!\!\!\perp (U, Y_1, V_1, Z) \mid (Y_1, U).$$

Next, applying CI.2 to this conditional independence yields

$$g(Y_1, U, \mathcal{E}_2) \perp\!\!\!\perp (U, Y_1, V_1, Z) \mid (Y_1, U).$$

Since $Y_2 = g(Y_1, U, \mathcal{E}_2)$, it can be rewritten as $Y_2 \perp\!\!\!\perp (U, Y_1, V_1, Z) \mid (Y_1, U)$. Lastly, applying CI.2 again to this conditional independence yields

$$Y_2 \perp\!\!\!\perp (h(Y_1, U, V_1), Z) \mid (Y_1, U).$$

Since $D_1 = h(Y_1, U, V_1)$, it can be rewritten as $Y_2 \perp\!\!\!\perp (D_1, Z) \mid (Y_1, U)$.

(iii) Applying CI.1 to Restriction 2 (ii) $V_2 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1)$ yields

$$V_2 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1) \mid (g(Y_1, U, \mathcal{E}_2), U).$$

Since $Y_2 = g(Y_1, U, \mathcal{E}_2)$, it can be rewritten as $V_2 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1) \mid (Y_2, U)$. Next, applying CI.2 to this conditional independence yields

$$V_2 \perp\!\!\!\perp (Y_1, h(Y_1, U, V_1)) \mid (Y_2, U).$$

Since $D_1 = h(Y_1, U, V_1)$, it can be rewritten as $V_2 \perp\!\!\!\perp (Y_1, D_1) \mid (Y_2, U)$. Lastly, applying CI.2 to this conditional independence yields

$$h(Y_2, U, V_2) \perp\!\!\!\perp (Y_1, D_1) \mid (Y_2, U).$$

Since $D_2 = h(Y_2, U, V_2)$, it can be rewritten as $D_2 \perp\!\!\!\perp (Y_1, D_1) \mid (Y_2, U)$.

(iv) Note that Restriction 2 (iii) $W \perp\!\!\!\perp (Y_1, \mathcal{E}_2, V_1, V_2)$ together with the structural definition $Z = \zeta(U, W)$ yields $Z \perp\!\!\!\perp (Y_1, \mathcal{E}_2, V_1, V_2) \mid U$. Applying CI.2 to this conditional independence relation $Z \perp\!\!\!\perp (Y_1, \mathcal{E}_2, V_1, V_2) \mid U$ yields

$$Z \perp\!\!\!\perp (Y_1, g(Y_1, U, \mathcal{E}_2), h(Y_1, U, V_1), h(g(Y_1, U, \mathcal{E}_2), U, V_2)) \mid U.$$

Since $D_t = h(Y_t, U, V_t)$ for each $t \in \{1, 2\}$ and $Y_2 = g(Y_1, U, \mathcal{E}_2)$, this conditional independence can be rewritten as $Z \perp\!\!\!\perp (Y_1, Y_2, D_1, D_2) \mid U$. □

8.4. Lemma 4 (Invariant Transition).

LEMMA 4 (Invariant Transition).
(i) Under Restrictions 1 and 2 (i), $F_{Y_3|Y_2U}(y' \mid y, u) = F_{Y_2|Y_1U}(y' \mid y, u)$ *for all* y', y, u.
(ii) Under Restrictions 1 and 2 (ii), $F_{D_2|Y_2U}(d \mid y, u) = F_{D_1|Y_1U}(d \mid y, u)$ *for all* d, y, u.

PROOF. (i) First, note that Restriction 2 (i) $\mathcal{E}_3 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2, W)$ implies $\mathcal{E}_3 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2)$, which in turn implies that $\mathcal{E}_3 \perp\!\!\!\perp (g(Y_1, U, \mathcal{E}_2), U)$, hence $\mathcal{E}_3 \perp\!\!\!\perp (Y_2, U)$. Second, Restriction 2 (i) in particular yields $\mathcal{E}_2 \perp\!\!\!\perp (Y_1, U)$. Using these two independence results, we obtain

$$\begin{aligned} F_{Y_3|Y_2U}(y' \mid y, u) &= \Pr[g(y, u, \mathcal{E}_3) \le y' \mid Y_2 = y, U = u] \\ &= \Pr[g(y, u, \mathcal{E}_3) \le y'] \\ &= \Pr[g(y, u, \mathcal{E}_2) \le y'] \\ &= \Pr[g(y, u, \mathcal{E}_2) \le y' \mid Y_1 = y, U = u] \;=\; F_{Y_2|Y_1U}(y' \mid y, u) \end{aligned}$$

for all y', y, u, where the second equality follows from $\mathcal{E}_3 \perp\!\!\!\perp (Y_2, U)$, the third equality follows from identical distribution of $\mathcal{E}_t$ by Restriction 1, and the forth equality follows from $\mathcal{E}_2 \perp\!\!\!\perp (Y_1, U)$.

(ii) Restriction 2 (ii) $V_2 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_1, \mathcal{E}_2, V_1)$ implies that $V_2 \perp\!\!\!\perp (g(Y_1, U, \mathcal{E}_2), U)$, hence $V_2 \perp\!\!\!\perp (Y_2, U)$. Restriction 2 (ii) also implies $V_1 \perp\!\!\!\perp (Y_1, U)$. Using these two

independence results, we obtain

$$\begin{aligned} F_{D_2|Y_2U}(d \mid y,u) &= \Pr[h(y,u,V_2) \le d \mid Y_2 = y, U = u] \\ &= \Pr[h(y,u,V_2) \le d] \\ &= \Pr[h(y,u,V_1) \le d] \\ &= \Pr[h(y,u,V_1) \le d \mid Y_1 = y, U = u] \;=\; F_{D_1|Y_1U}(d \mid y,u) \end{aligned}$$

for all d, y, u, where the second equality follows from $V_2 \perp\!\!\!\perp (Y_2, U)$, the third equality follows from identical distribution of V_t from Restriction 1, and the forth equality follows from $V_1 \perp\!\!\!\perp (Y_1, U)$. □

9. Appendix: Proofs for Estimation

9.1. Corollary 1 (Constrained Maximum Likelihood).

PROOF. Denote the supports of conditional densities by $I_1 = \{(y_2, y_1, z) \mid f_{Y_2Y_1Z|D_2D_1}(y_2, y_1, z \mid 1) > 0\}$ and $I_2 = \{(y_3, y_2, y_1, z) \mid f_{Y_3Y_2Y_1Z|D_2D_1}(y_3, y_2, y_1, z \mid 1, 1) > 0\}$. The Kullback-Leibler information inequality requires that

$$\int_{I_1} \log\left[\frac{f_{Y_2Y_1Z|D_1}(y_2,y_1,z \mid 1)}{\varphi(y_2,y_1,z)}\right] f_{Y_2Y_1Z|D_1}(y_2,y_1,z \mid 1) d\mu(y_2,y_1,z) \ge 0 \qquad \text{and}$$

$$\int_{I_2} \log\left[\frac{f_{Y_3Y_2Y_1Z|D_2D_1}(y_3,y_2,y_1,z \mid 1,1)}{\psi(y_3,y_2,y_1,z)}\right] f_{Y_3Y_2Y_1Z|D_2D_1}(y_3,y_2,y_1,z \mid 1,1) d\mu(y_3,y_2,y_1,z) \ge 0$$

for all non-negative measurable functions φ and ψ such that $\int \varphi = \int \psi = 1$. These two inequalities hold with equalities if and only if $f_{Y_2Y_1Z|D_1}(\,\cdot,\,\cdot,\,\cdot \mid 1) = \varphi$ and $f_{Y_3Y_2Y_1Z|D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot \mid 1,1) = \psi$, respectively. (Equalities and uniqueness are stated up to the equivalence classes identified by the underlying probability measures.) Let the set of such pairs of functions (φ, ψ) satisfying the above two Kullback-Leibler inequalities be denoted by

$$\Lambda = \left\{(\varphi,\psi) \,\middle|\, \varphi \text{ and } \psi \text{ are non-negative measurable functions with } \int \varphi = \int \psi = 1\right\}.$$

With this notation, the maximization problem

$$\max_{(\varphi,\psi)\in\Lambda} c_1 \mathrm{E}\left[\log \varphi(Y_2, Y_1, Z) \middle| D_1 = 1\right] + c_2 \mathrm{E}\left[\log \psi(Y_3, Y_2, Y_1, Z) \middle| D_2 = D_1 = 1\right] \tag{17}$$

has the unique solution $(\varphi, \psi) = (f_{Y_2Y_1Z|D_1}(\,\cdot,\,\cdot,\,\cdot \mid 1), f_{Y_3Y_2Y_1Z|D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot \mid 1,1))$.

Now, let $F(\,\cdot\,; M)$ denote a distribution function generated by model $M \in \mathcal{F}$. For the true model $M^* := (F^*_{Y_t|Y_{t-1}U}, F^*_{D_t|Y_tU}, F^*_{Y_1U}, F^*_{Z|U})$, we have

$$\begin{aligned} F_{Y_2Y_1ZD_1}(\,\cdot,\,\cdot,\,\cdot,1) &= F_{Y_2Y_1ZD_1}(\,\cdot,\,\cdot,\,\cdot,1;M^*) \\ F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1) &= F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1;M^*) \end{aligned}$$

Moreover, the identification result of Lemma 2 showed that this true model M^* is the unique element in $\mathcal{F}$ that generates the observed parts of the joint distributions $F_{Y_2Y_1ZD_1}(\,\cdot,\,\cdot,\,\cdot,1)$ and $F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1)$, i.e.,

$$\begin{aligned} F_{Y_2Y_1ZD_1}(\,\cdot,\,\cdot,\,\cdot,1) &= F_{Y_2Y_1ZD_1}(\,\cdot,\,\cdot,\,\cdot,1;M) \text{ if and only if } M = M^* \\ F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1) &= F_{Y_3Y_2Y_1ZD_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1;M) \text{ if and only if } M = M^* \end{aligned}$$

But this implies that F^* is the unique model that generates the observable conditional densities $f_{Y_2Y_1Z|D_1}(\,\cdot,\,\cdot,\,\cdot \mid 1)$ and $f_{Y_3Y_2Y_1Z|D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot \mid 1,1)$ among those models $M \in \mathcal{F}$ that are compatible with the observed selection frequencies $f_{D_1}(1)$ and $f_{D_2D_1}(1,1)$, i.e.,

$$\begin{aligned} f_{Y_2Y_1Z|D_1}(\,\cdot,\,\cdot,\,\cdot \mid 1) &= f_{Y_2Y_1Z|D_1}(\,\cdot,\,\cdot,\,\cdot \mid 1;M) \text{ if and only if } M = M^* \\ &\quad \text{given } f_{D_1}(1;M) = f_{D_1}(1), \text{ and} \qquad (18) \\ f_{Y_3Y_2Y_1Z|D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot \mid 1,1) &= f_{Y_3Y_2Y_1Z|D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot \mid 1,1;M) \text{ if and only if } M = M^* \\ &\quad \text{given } f_{D_2D_1}(1,1;M) = f_{D_2D_1}(1,1) \qquad (19) \end{aligned}$$

Since $(\varphi, \psi) = (f_{Y_2Y_1Z|D_1}(\,\cdot,\,\cdot,\,\cdot \mid 1), f_{Y_3Y_2Y_1Z|D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot \mid 1,1))$ is the unique solution to (17), the statements (18) and (19) imply that the true model M^* is the unique solution to

$$\begin{aligned} \max_{M\in\mathcal{F}} \quad & c_1 \mathrm{E}\left[\log f_{Y_2Y_1Z|D_1}(Y_2, Y_1, Z \mid 1; M) \middle| D_1 = 1\right] \\ & + c_2 \mathrm{E}\left[\log f_{Y_3Y_2Y_1Z|D_2D_1}(Y_3, Y_2, Y_1, Z \mid 1,1; M) \middle| D_2 = D_1 = 1\right] \\ \text{s.t.} \quad & f_{D_1}(1;M) = f_{D_1}(1) \quad \text{and} \quad f_{D_2D_1}(1,1;M) = f_{D_2D_1}(1,1) \end{aligned}$$

or equivalently

$$
\begin{aligned}
\max_{M\in\mathcal{F}} \quad & c_1 \mathrm{E}\left[\log f_{Y_2Y_1ZD_1}(Y_2,Y_1,Z,1;M)\middle|\, D_1=1\right] \\
& +c_2 \mathrm{E}\left[\log f_{Y_3Y_2Y_1ZD_2D_1}(Y_3,Y_2,Y_1,Z,1,1;M)\middle|\, D_2=D_1=1\right] \\
\text{s.t.} \quad & f_{D_1}(1;M)=f_{D_1}(1) \quad \text{and} \quad f_{D_2D_1}(1,1;M)=f_{D_2D_1}(1,1)
\end{aligned} \tag{20}
$$

since what have been omitted are constants due to the constraints.

By using Lemmas 3 and 4, we can write the equalities

$$
\begin{aligned}
f_{Y_2Y_1ZD_1}(y_2,y_1,z,1;M) &= \int f_{Y_t|Y_{t-1}U}(y_2\mid y_1,u) f_{D_t|Y_tU}(1\mid y_1,u) f_{Y_1U}(y_1,u) f_{Z|U}(z\mid u) d\mu(u) \\
f_{D_1}(1;M) &= \int f_{D_t|Y_tU}(1\mid y_1,u) f_{Y_1U}(y_1,u) d\mu(y_1,u)
\end{aligned}
$$

$$
\begin{aligned}
f_{Y_3Y_2Y_1ZD_2D_1}(y_3,y_2,y_1,z,1,1;M) &= \int f_{Y_t|Y_{t-1}U}(y_3\mid y_2,u) f_{Y_t|Y_{t-1}U}(y_2\mid y_1,u) f_{D_t|Y_tU}(1\mid y_2,u) \\
&\quad f_{D_t|Y_tU}(1\mid y_1,u) f_{Y_1U}(y_1,u) F_{Z|U}(z\mid u) d\mu(u) \\
f_{D_2D_1}(1,1;M) &= \int f_{Y_t|Y_{t-1}U}(y_2\mid y_1,u) f_{D_t|Y_tU}(1\mid y_2,u) \\
&\quad f_{D_t|Y_tU}(1\mid y_1,u) f_{Y_1U}(y_1,u) d\mu(y_2,y_1,u)
\end{aligned}
$$

for any model $M := (F_{Y_t|Y_{t-1}U}, F_{D_t|Y_tU}, F_{Y_1U}, F_{Z|U}) \in \mathcal{F}$. Substituting these equalities in (20), we conclude that the true model $(F^*_{Y_t|Y_{t-1}U}, F^*_{D_t|Y_tU}, F^*_{Y_1U}, F^*_{Z|U})$ is the unique solution to

$$
\begin{aligned}
\max_{\left(F_{Y_t|Y_{t-1}U}, F_{D_t|Y_tU}, F_{Y_1U}, F_{Z|U}\right)\in\mathcal{F}} \quad & c_1 \mathrm{E}\Big[\log \int f_{Y_t|Y_{t-1}U}(Y_2\mid Y_1,u) f_{D_t|Y_tU}(1\mid Y_1,u) \\
& \qquad f_{Y_1U}(Y_1,u) f_{Z|U}(Z\mid u) d\mu(u)\Big|\, D_1=1\Big] + \\
& c_2 \mathrm{E}\Big[\log \int f_{Y_t|Y_{t-1}U}(Y_3\mid Y_2,u) f_{Y_t|Y_{t-1}U}(Y_2\mid Y_1,u) f_{D_t|Y_tU}(1\mid Y_2,u) f_{D_t|Y_tU}(1\mid Y_1,u) \\
& \qquad f_{Y_1U}(Y_1,u) F_{Z|U}(Z\mid u) d\mu(u)\Big|\, D_2=D_1=1\Big]
\end{aligned}
$$

subject to

$$
\begin{aligned}
&\int f_{D_t|Y_tU}(1\mid y_1,u) f_{Y_1U}(y_1,u) d\mu(y_1,u) = f_{D_1}(1) \qquad \text{and} \\
&\int f_{Y_t|Y_{t-1}U}(y_2\mid y_1,u) f_{D_t|Y_tU}(1\mid y_2,u) f_{D_t|Y_tU}(1\mid y_1,u) f_{Y_1U}(y_1,u) d\mu(y_2,y_1,u) = f_{D_2D_1}(1,1)
\end{aligned}
$$

as claimed. □

9.2. Remark 9 (Unit Lagrange Multipliers). For short-hand notation, we write $f = (f_1, f_2, f_3, f_4) \in \mathcal{F}$ for an element of $\mathcal{F}$, $p_1 := \Pr(D_1 = 1)$, $p_2 := \Pr(D_2 = D_1 = 1)$, and $p := (p_1, p_2)'$. The solution $f^*(\cdot\,; p) \in \mathcal{F}$ has Lagrange multipliers $\lambda^*(p) = (\lambda_1^*(p), \lambda_2^*(p))' \in \Lambda$ such that $(f^*(\cdot\,; p), \lambda^*(p))$ is a saddle point of the Lagrangean functional

$$L(f, \lambda; p) = p_1 L_1(f; p_1) + p_2 L_2(f; p_2) - \lambda_1(L_3(f) - p_1) - \lambda_2(L_4(f) - p_2),$$

where the functionals $L_1, \cdots, L_4$ are defined as

$$\begin{aligned}
L_1(f; p_1) &= \int \left[\log \int f_1(y_2 \mid y_1, u) f_2(1 \mid y_1, u) f_3(y_1, u) f_4(z \mid u) d\mu(u)\right] \\
&\quad \times f_{Y_2 Y_1 Z \mid D_1}(y_2, y_1, z \mid 1) d\mu(y_2, y_1, z) \\
L_2(f; p_2) &= \int \left[\log \int f_1(y_3 \mid y_2, u) f_1(y_2 \mid y_1, u) f_2(1 \mid y_2, u) f_2(1 \mid y_1, u) f_3(y_1, u) f_4(z \mid u) d\mu(u)\right] \\
&\quad \times f_{Y_3 Y_2 Y_1 Z \mid D_2 D_1}(y_3, y_2, y_1, z \mid 1, 1) d\mu(y_3, y_2, y_1, z) \\
L_3(f) &= \int f_2(1 \mid y_1, u) f_3(y_1, u) d\mu(y_1, u) \qquad \text{and} \\
L_4(f) &= \int f_1(y_2 \mid y_1, u) f_2(1 \mid y_2, u) f_2(1 \mid y_1, u) f_3(y_1, u) d\mu(y_2, y_1, u).
\end{aligned}$$

Moreover, $f^*(\cdot\,; p)$ maximizes $L(f, \lambda^*(p); p)$ given λ is restricted to $\lambda^*(p)$. We want to claim that $\lambda^*(p) = (1, 1)'$. The following assumptions are imposed to this end.

Assumption (Regularity for Unit Lagrange Multipliers)**.**
(i) Selection probabilities are positive: $p_1, p_2 > 0$.
(ii) The functionals L_1, L_2, L_3, and L_4 are Fréchet differentiable with respect to f at the solution $f^*(\cdot\,; p)$ for some norm $\|\cdot\|$ on a linear space containing $\mathcal{F}$.
(iii) The solution $(f^*(\cdot\,; p), \lambda^*(p))$ is differentiable with respect to p.
(iv) The solution $f^*(\cdot\,; p)$ is a regular point of the constraint functionals L_3 and L_4.

A sufficient condition for part (ii) of this assumption will be provided later in terms of a concrete normed linear space.

PROOF. Since the Chain Rule holds for a composition of Fréchet differentiable transformations (cf. Luenberger, 1969; pp.176), we have

$$\frac{d}{dp_1}L(f^*(\cdot\,;p),\lambda^*(p);p) = D_{f,\lambda}L(f^*(\cdot\,;p),\lambda^*(p);p)\cdot D_{p_1}(f^*(\cdot\,;p),\lambda^*(p))$$
$$+\frac{\partial}{\partial p_1}L(f^*(\cdot\,;p),\lambda^*(p);p) \;=\; \frac{\partial}{\partial p_1}L(f^*(\cdot\,;p),\lambda^*(p);p)$$

where the second equality follows from the equality constraints and the stationarity of $L(\cdot,\lambda^*(p);p)$ at $f^*(\cdot\,;p)$, which is a regular point of the constraint functionals L_3 and L_4 by assumption.

On one hand, the partial derivative is

$$\frac{\partial}{\partial p_1}L(f^*(\cdot\,;p),\lambda^*(p);p) = \lambda_1^*(p).$$

On the other hand, the complementary slackness yields

$$\frac{d}{dp_1}L(f^*(\cdot\,;p),\lambda^*(p);p) \;=\; \frac{d}{dp_1}[p_1L_1(f^*(\cdot\,;p);p_1)].$$

In order to evaluate the last term, we first note that

$$p_1L_1(f;p_1) \;=\; \int\left[\log\int f_1(y_2\mid y_1,u)f_2(1\mid y_1,u)f_3(y_1,u)f_4(z\mid u)d\mu(u)\right]$$
$$\times\; f_{Y_2Y_1ZD_1}(y_2,y_1,z,1)d\mu(y_2,y_1,z).$$

In view of the proof of Corollary 1, we recall that $f^*(\cdot\,;p)$ maximizes $p_1L_1(\cdot\,;p_1)$, and the solution $f^*(\cdot\,;p)$ satisfies

$$\int f_1^*(y_2\mid y_1,u;p)f_2^*(1\mid y_1,u;p)f_3^*(y_1,u;p)f_4^*(z\mid u;p)d\mu(u) = f_{Y_2Y_1Z|D_1}(y_2,y_1,z\mid 1)\cdot p_1,$$

where the conditional density $f_{Y_2Y_1Z|D_1}(\cdot,\cdot,\cdot\mid 1)$ is invariant from variations in p, and the scale of the integral varies by p_1 which defines the $\mathcal{L}^1$-equivalence class of non-negative functions over which the Kullback-Leibler information inequality is satisfied. Therefore, we have

$$\frac{d}{dp_1}\left[\log\int f_1^*(y_2\mid y_1,u;p)f_2^*(1\mid y_1,u;p)f_3^*(y_1,u;p)f_4^*(z\mid u;p)d\mu(u)\right] = \frac{1}{p_1}.$$

It then follows that

$$\frac{d}{dp_1}L(f^*(\cdot\,;p),\lambda^*(p);p) \;=\; \frac{d}{dp_1}[p_1L_1(f^*(\cdot\,;p);p_1)] = \frac{1}{p_1}\int f_{Y_2Y_1ZD_1}(y_2,y_1,z,1)d\mu(y_2,y_1,z) = 1,$$

showing that $\lambda_1^*(p) = 1$. Similar lines of argument prove $\lambda_2^*(p) = 1$. $\square$

Part (ii) of the above assumption is ambiguous about the definition of underlying topological spaces, as we did not explicitly define the norm. In order to complement for it, here we consider a sufficient condition. Write $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3 \times \mathcal{F}_4$. Define a norm on $\mathcal{F}$ by $\|f\|_s := \|f_1\|_2 + \|f_2\|_2 + \|f_3\|_2 + \|f_4\|_2$, where $\|\cdot\|_2$ denotes the $\mathcal{L}^2$-norm. Also, define the set $B_j(M) = \{f_j \in \mathcal{F}_j \mid \|f_j\|_\infty \leqslant M\}$ for $M \in (0, \infty)$ for each $j = 1, 2, 3, 4$. The following uniform boundedness and integrability together imply part (ii).

Assumption (A Sufficient Condition for Part (ii))**.** There exists $M < \infty$ such that $\mathcal{F}_1 \subset \mathcal{L}^1 \cap B_1(M)$, $\mathcal{F}_2 \subset \mathcal{L}^1 \cap B_2(M)$, $\mathcal{F}_3 \subset \mathcal{L}^1 \cap B_3(M)$, and $\mathcal{F}_4 \subset \mathcal{L}^1 \cap B_4(M)$ hold with the respective Lebesgue measurable spaces.

Note that $\mathcal{F}_1 \subset \mathcal{L}^1 \cap \mathcal{L}^\infty$, $\mathcal{F}_2 \subset \mathcal{L}^1 \cap \mathcal{L}^\infty$, $\mathcal{F}_3 \subset \mathcal{L}^1 \cap \mathcal{L}^\infty$, and $\mathcal{F}_4 \subset \mathcal{L}^1 \cap \mathcal{L}^\infty$ follow from this assumption, since $B(M) \subset \mathcal{L}^\infty$ for each $j = 1, 2, 3, 4$. But then, each of these sets is also square integrable as $\mathcal{L}^1 \cap \mathcal{L}^\infty \subset \mathcal{L}^2$ (cf. Folland, 1999; pp. 185).

To see the Fréchet differentiability of L_1, observe that for any $\eta \in \mathcal{F}$

$$
\begin{aligned}
&\left\| \int (f_1+\eta_1)(f_2+\eta_2)(f_3+\eta_3)(f_4+\eta_4) d\mu(u) - \int f_1 f_2 f_3 f_4 d\mu(u) - DL_1(f;\eta) \right\|_1 \\
\leqslant\ & \|f_1 f_2 \eta_3 \eta_4\|_1 + \|f_1 \eta_2 \eta_3 f_4\|_1 + \|\eta_1 \eta_2 f_3 f_4\|_1 + \|f_1 \eta_2 f_3 \eta_4\|_1 + \|\eta_1 f_2 \eta_3 f_4\|_1 + \|\eta_1 f_2 f_3 \eta_4\|_1 \\
& + \|f_1 \eta_2 \eta_3 \eta_4\|_1 + \|\eta_1 f_2 \eta_3 \eta_4\|_1 + \|\eta_1 \eta_2 f_3 \eta_4\|_1 + \|\eta_1 \eta_2 \eta_3 f_4\|_1 + \|\eta_1 \eta_2 \eta_3 \eta_4\|_1 \\
\leqslant\ & \|f_1\|_\infty \|f_2\|_\infty \|\eta_3\|_2 \|\eta_4\|_2 + \|f_1\|_\infty \|f_4\|_\infty \|\eta_2\|_2 \|\eta_3\|_2 + \|f_3\|_\infty \|f_4\|_\infty \|\eta_1\|_2 \|\eta_2\|_2 \\
& + \|f_1\|_\infty \|f_3\|_\infty \|\eta_2\|_2 \|\eta_4\|_2 + \|f_2\|_\infty \|f_4\|_\infty \|\eta_1\|_2 \|\eta_3\|_2 + \|f_2\|_\infty \|f_3\|_\infty \|\eta_1\|_2 \|\eta_4\|_2 \\
& + \|f_1\|_\infty \|\eta_2\|_\infty \|\eta_3\|_2 \|\eta_4\|_2 + \|\eta_1\|_\infty \|f_2\|_\infty \|\eta_3\|_2 \|\eta_4\|_2 + \|\eta_1\|_\infty \|f_3\|_\infty \|\eta_2\|_2 \|\eta_4\|_2 \\
& + \|\eta_1\|_\infty \|f_4\|_\infty \|\eta_2\|_2 \|\eta_3\|_2 + \|\eta_1\|_\infty \|\eta_2\|_\infty \|\eta_3\|_2 \|\eta_4\|_2 \\
\leqslant\ & (\|f_1\|_\infty \|f_2\|_\infty + \|f_1\|_\infty \|f_4\|_\infty + \|f_3\|_\infty \|f_4\|_\infty + \|f_1\|_\infty \|f_3\|_\infty + \|f_2\|_\infty \|f_4\|_\infty \\
& + \|f_2\|_\infty \|f_3\|_\infty + \|f_1\|_\infty \|\eta_2\|_\infty + \|\eta_1\|_\infty \|f_2\|_\infty + \|\eta_1\|_\infty \|f_3\|_\infty + \|\eta_1\|_\infty \|f_4\|_\infty \\
& + \|\eta_1\|_\infty \|\eta_2\|_\infty) \|\eta\|_s^2 \ \leqslant\ 11 M^2 \|\eta\|_s^2 ,
\end{aligned}
$$

where the $\mathcal{L}^1$-norm in the first line is by integration with respect to (y_2, y_1, z), all the remaining $\mathcal{L}^p$-norms are by integration with respect to (y_2, y_1, z, u), $DL_1(f;\eta) := \int(f_1 f_2 f_3 \eta_4 + f_1 f_2 \eta_3 f_4 + f_1 \eta_2 f_3 f_4 + \eta_1 f_2 f_3 f_4) d\mu(u)$, the first inequality follows from the triangle inequality, the second inequality follows from the Hölder's inequality, the third inequality follows from our definition of the norm on $\mathcal{F}$, and the last inequality follows from our assumption. But then,

$$\lim_{\|\eta\|_s \to 0} \frac{\left\| \int (f_1 + \eta_1)(f_2 + \eta_2)(f_3 + \eta_3)(f_4 + \eta_4) d\mu(u) - \int f_1 f_2 f_3 f_4 d\mu(u) - DL_1(f;\eta) \right\|_1}{\|\eta\|_s} = 0,$$

showing that $DL_1(f;\eta)$ is the Fréchet derivative of the operator $f \mapsto \int f_1 f_2 f_3 f_4 d\mu(u)$. This in turn implies Fréchet differentiability of the functional L_1 at the solution $f^*(\cdot\,; p)$, since the functional $\mathcal{L}^1 \ni \eta \mapsto \int \log \eta dF_{Y_2 Y_1 Z | D_1 = 1}$ is Fréchet differentiable at $\eta = f_{Y_2 Y_1 Z D_1}(\cdot, \cdot, \cdot, 1)$. Similar lines of arguments will show Fréchet differentiability of the other functionals L_2, L_3, and L_4 at $f^*(\cdot\,; p)$.

9.3. Proposition 1 (Consistency of the Nonparametric Estimator). As a setup, we define a normed linear space $(\mathcal{L}, \|\cdot\|)$ containing the model set $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3 \times \mathcal{F}_4$ as follows. We define the uniform norm of f as the essential supremum

$$\|f\|_\infty = \text{ess sup}_x \left| f(x) \right|.$$

Following Newey and Powell (2003) and others, we also define the following version of the uniform norm when characterizing compactness:

$$\|f\|_{R,\infty} = \text{ess sup}_x \left| f(x)(1 + x'x) \right|.$$

Noted that $\|\cdot\|_\infty \leqslant \|\cdot\|_{R,\infty}$ holds. Similarly define the version of the $\mathcal{L}^1$ norm

$$\|f\|_{R,1} = \int \left| f(x) \right| (1 + x'x) dx.$$

Define a norm $\|\cdot\|$ on a linear space containing $\mathcal{F}$ by

$$\|f\| := \|f_1\|_{R,\infty} + \|f_2\|_{R,\infty} + \|f_3\|_{R,\infty} + \|f_4\|_{R,\infty}.$$

We consider $\mathcal{F}$ with the subspace topology of this normed linear space, where Assumption 3 below imposes restrictions on how to appropriately choose such a subset $\mathcal{F}$.

We assume that the data is i.i.d.

ASSUMPTION 1 (Data). The data $\{(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{2i}, D_{1i})\}_{i=1}^n$ is i.i.d.

In order to model the rate at which the complexity of sieve spaces evolve with sample size n, we introduce the notation $N(\cdot\ , \cdot\ , \|\cdot\|)$ for the covering numbers without bracketing. Let $B(f, \varepsilon) = \{f' \in \mathcal{F} \mid \|f - f'\| < \varepsilon\}$ denote the ε-ball around $f \in \mathcal{F}$ with respect to the norm $\|\cdot\|$ defined above. For each $\varepsilon > 0$ and n, let $N(\varepsilon, \mathcal{F}_{k(n)}, \|\cdot\|)$ denote the minimum number of such ε-balls covering $\mathcal{F}_{k(n)}$, i.e., $\min\{|C| \mid \cup_{f \in C} B(f, \varepsilon) \supset \mathcal{F}_{k(n)}\}$. With this notation, we assume the following restriction.

ASSUMPTION 2 (Sieve Spaces).
(i) $\{\mathcal{F}_{k(n)}\}_{n=1}^\infty$ is an increasing sequence, $\mathcal{F}_{k(n)} \subset \mathcal{F}$ for each n, and there exists a sequence $\{\pi_{k(n)} f_0\}_{n=1}^\infty$ such that $\pi_{k(n)} f_0 \in \mathcal{F}_{k(n)}$ for each n.
(ii) $\log N(\varepsilon, \mathcal{F}_{k(n)}, \|\cdot\|) = o(n)$ for all $\varepsilon > 0$.

The next assumption facilitates compactness of the model set and Hölder continuity of the objective functional, both of which are important for nice large sample behavior of the estimator. We assume that the true model f_0 belongs to $\mathcal{F}$ satisfying the following.

ASSUMPTION 3 (Model Set).
(i) $\mathcal{L}^1$ Compactness: Each of $\mathcal{F}_2$ and $\mathcal{F}_3$ is compact with respect to $\|\cdot\|_{R,1}$. Thus, let $M < \infty$ be a number such that $\sup_{f_i \in \mathcal{F}_i} \|f_i\|_{R,1} \leqslant M$ for each $i = 2, 3$.
(ii) $\mathcal{L}^\infty$ Compactness: Each of $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$, and $\mathcal{F}_4$ is compact with respect to $\|\cdot\|_{R,\infty}$. Thus, let $M_\infty < \infty$ be a number such that $\sup_{f_i \in \mathcal{F}_i} \|f_i\|_{R,\infty} \leqslant M_\infty$ for each $i = 1, 2, 3, 4$.

(iii) Uniformly Bounded Density of $\mathcal{E}$: There exists $M_1 < \infty$ such that

$$\sup_{f_1 \in \mathcal{F}_1} \sup_{y_2, y_1} \int |f_1(y_2 \mid y_1, u)| \, du \leqslant M_1.$$

(iv) Uniformly Bounded Density of Y_1: There exists $M_3 < \infty$ such that

$$\sup_{f_3 \in \mathcal{F}_3} \sup_{y_1} \int |f_3(y_1, u)| \, du \leqslant M_3.$$

(v) Bounded Objective:

$$\mathrm{E}\left[\left(\inf_{f \in \mathcal{F}} \int f_1(Y_2 \mid Y_1, u) f_2(1 \mid Y_1, u) f_3(Y_1, u) f_4(Z \mid u) d\mu(u)\right)^{-1}\right] < \infty \quad \text{and}$$

$$\mathrm{E}\left[\left(\inf_{f \in \mathcal{F}} \int f_1(Y_3 \mid Y_2, u) f_1(Y_2 \mid Y_1, u) f_2(1 \mid Y_2, u) f_2(1 \mid Y_1, u) f_3(Y_1, u) f_4(Z \mid u) d\mu(u)\right)^{-1}\right] < \infty$$

Part (i) of this assumption is not redundant, since f_i are not densities, but conditional densities. Despite their appearance, parts (iii) and (iv) of this assumption are not so stringent. Suppose, for example, that the true model f_0 consists of the traditional additively separable dynamic model $Y_t = \alpha Y_{t-1} + U + \mathcal{E}_t$ with a uniformly bounded density of $\mathcal{E}_t$. In this case, the true model f_0 can indeed reside in an $\mathcal{F}$ satisfying the restriction of part (iii) for a suitable choice of M_1. Similarly, the true model f_0 can reside in an $\mathcal{F}$ satisfying the restriction of part (iv) for a suitable choice of M_3, whenever the density of Y_1 is uniformly bounded.

Proof. We show the consistency claim of Proposition 1 by showing that Conditions 3.1, 3.2, 3.4, and 3.5M of Chen (2007) are satisfied by our assumptions (Restrictions 1, 2, 3, and 4 and Assumptions 1, 2, and 3). Restrictions 1, 2, 3, and 4 imply her Condition 3.1 by our identification result yielding Corollary 1 together with Remark 9. Her Condition 3.2 is directly assumed by our Assumption 2 (i). Her Condition 3.4 is implied by our Assumption 3 (ii) applied to the Tychonoff's Theorem. Her Conditions 3.5M (i) and (iii) are directly assumed by our Assumptions 1 and 2 (ii), respectively, provided that we will prove her Condition 3.5M (ii) with $s = 1$. It remains to prove Hölder continuity of $l(y_3, y_2, y_1, z, d_2, d_1; \cdot) : (\mathcal{F}, \|\cdot\|) \to \mathbb{R}$ for each $(y_3, y_2, y_1, z, d_2, d_1)$, which in turn implies her Condition 3.5M (ii).

In order to show Hölder continuity of the functional $l(y_3, y_2, y_1, z, d_2, d_1; \cdot)$, it suffices to prove that of $l_1(y_2, y_1, z; \cdot)$, $l_2(y_3, y_2, y_1, z; \cdot)$, l_3, and l_4. First, consider $l_1(y_2, y_1, z; \cdot)$. For a fixed (y_2, y_1, z), observe

$$
\begin{aligned}
& \left|\exp(l_1(y_2, y_1, z; f)) - \exp(l_1(y_2, y_1, z; \bar{f}))\right| \\
\leqslant\ & \left|\int (f_1 - \bar{f}_1) f_2 f_3 f_4 du\right| + \left|\int \bar{f}_1 (f_2 - \bar{f}_2) f_3 f_4 du\right| \\
& + \left|\int \bar{f}_1 \bar{f}_2 (f_3 - \bar{f}_3) f_4 du\right| + \left|\int \bar{f}_1 \bar{f}_2 \bar{f}_3 (f_4 - \bar{f}_4) du\right| \\
\leqslant\ & \left\|f_1 - \bar{f}_1\right\|_\infty \|f_2\|_\infty \int |f_3|\, du\, \|f_4\|_\infty + \left\|\bar{f}_1\right\|_\infty \left\|f_2 - \bar{f}_2\right\|_\infty \int |f_3|\, du\, \|f_4\|_\infty \\
& + \int \left|\bar{f}_1\right| du \left\|\bar{f}_2\right\|_\infty \left\|f_3 - \bar{f}_3\right\|_\infty \|f_4\|_\infty + \int \left|\bar{f}_1\right| du \left\|\bar{f}_2\right\|_\infty \left\|\bar{f}_3\right\|_\infty \left\|f_4 - \bar{f}_4\right\|_\infty \\
\leqslant\ & 2M_\infty^2 (M_1 + M_3) \left\|f - \bar{f}\right\|,
\end{aligned}
$$

where the first inequality follows from the triangle inequality, the second inequality follows from the Hölder's inequality, and the third inequality uses Assumption 3 (ii), (iii), and (iv), together with the fact that $\|\cdot\|_\infty \leqslant \|\cdot\|_{R,\infty}$. By Assumption 3 (v), there exists a function κ_1 such that $\mathrm{E}[\kappa_1(Y_2, Y_1, Z)] < \infty$ and

$$
\left|l_1(y_2, y_1, z; f) - l_1(y_2, y_1, z; \bar{f})\right| \leqslant 2M_\infty^2 (M_1 + M_3) \left\|f - \bar{f}\right\| \kappa_1(y_2, y_1, z).
$$

This shows Hölder (in particular Lipschitz) continuity of the functional $l_1(y_2, y_1, z; \cdot)$.

By similar calculations using Assumption 3 (ii) and (iii), we obtain

$$
\begin{aligned}
& \left|\exp(l_2(y_3, y_2, y_1, z; f)) - \exp(l_2(y_3, y_2, y_1, z; \bar{f}))\right| \\
\leqslant\ & \left\|f_1 - \bar{f}_1\right\|_\infty \int |f_1|\, du\, \|f_2\|_\infty^2 \|f_3\|_\infty \|f_4\|_\infty + \int \left|\bar{f}_1\right| du \left\|f_1 - \bar{f}_1\right\|_\infty \|f_2\|_\infty^2 \|f_3\|_\infty \|f_4\|_\infty \\
& + \int \left|\bar{f}_1\right| du \left\|\bar{f}_1\right\|_\infty \left\|f_2 - \bar{f}_2\right\|_\infty \|f_2\|_\infty \|f_3\|_\infty \|f_4\|_\infty \\
& + \int \left|\bar{f}_1\right| du \left\|\bar{f}_1\right\|_\infty \left\|\bar{f}_2\right\|_\infty \left\|f_2 - \bar{f}_2\right\|_\infty \|f_3\|_\infty \|f_4\|_\infty \\
& + \int \left|\bar{f}_1\right| du \left\|\bar{f}_1\right\|_\infty \left\|\bar{f}_2\right\|_\infty^2 \left\|f_3 - \bar{f}_3\right\|_\infty \|f_4\|_\infty + \int \left|\bar{f}_1\right| du \left\|\bar{f}_1\right\|_\infty \left\|\bar{f}_2\right\|_\infty^2 \left\|\bar{f}_3\right\|_\infty \left\|f_4 - \bar{f}_4\right\|_\infty \\
\leqslant\ & 6M_\infty^4 M_1 \left\|f - \bar{f}\right\|.
\end{aligned}
$$

By Assumption 3 (v), there exists a function κ_2 such that $\mathrm{E}[\kappa_2(Y_3, Y_2, Y_1, Z)] < \infty$ and

$$\left|l_2(y_3, y_2, y_1, z; f) - l_2(y_3, y_2, y_1, z; \bar{f})\right| \leqslant 6M_\infty^4 M_1 \left\|f - \bar{f}\right\| \kappa_2(y_3, y_2, y_1, z).$$

This shows Lipschitz continuity of the functional $l_2(y_3, y_2, y_1, z; \cdot)$.

Next, using Assumption 3 (i) yields

$$\begin{aligned}\left|l_3(f) - l_3(\bar{f})\right| &\leqslant \left|\int (f_2 - \bar{f}_2) f_3\right| + \left|\int \bar{f}_2 (f_3 - \bar{f}_3)\right| \\ &\leqslant \left\|f_2 - \bar{f}_2\right\|_\infty \|f_3\|_1 + \left\|\bar{f}_2\right\|_1 \left\|f_3 - \bar{f}_3\right\|_\infty \leqslant 2M \left\|f - \bar{f}\right\|,\end{aligned}$$

Similarly, using Assumption (i) and (ii) yields

$$\begin{aligned}\left|l_4(f) - l_4(\bar{f})\right| &\leqslant \left\|f_1 - \bar{f}_1\right\|_\infty \|f_2\|_\infty^2 \|f_3\|_1 + \left\|\bar{f}_1\right\|_\infty \left\|f_2 - \bar{f}_2\right\|_\infty \|f_2\|_\infty \|f_3\|_1 \\ &\quad + \left|\bar{f}_1\right|_\infty \left\|\bar{f}_2\right\|_\infty \left\|f_2 - \bar{f}_2\right\|_\infty \|f_3\|_1 + \left|\bar{f}_1\right|_\infty \left\|\bar{f}_2\right\|_\infty \|f_2\|_1 \left\|f_3 - \bar{f}_3\right\|_\infty \\ &\leqslant 4MM_\infty^2 \left\|f - \bar{f}\right\|.\end{aligned}$$

It follows that l_3 and l_4 are also Lipschitz continuous. These in particular implies Hölder continuity of the functionals $l_1(y_2, y_1, z; \cdot)$, $l_2(y_3, y_2, y_1, z; \cdot)$, l_3, and l_4, hence $l(y_3, y_2, y_1, z, d_2, d_1; \cdot)$. Therefore, Chen's Condition 3.5M (ii) is satisfied with $s = 1$ by our assumptions. □

9.4. Semiparametric Estimation. Section 5.2 proposed an estimator which treats the quadruple $(f_{Y_t|Y_{t-1}U}, f_{D_t|Y_tU}, f_{Y_1U}, f_{Z|U})$ of the density functions nonparametrically. In practice, it may be more useful to specify one or more of these densities semi-parametrically. For example, the dynamic model g is conventionally specified by

$$g(y, u, \varepsilon) = \alpha y + u + \varepsilon.$$

By denoting the nonparametric density functions of $\mathcal{E}_t$ by $f_\mathcal{E}$, we can represent the density $f_{Y_t|Y_{t-1}U}$ by $f_{Y_t|Y_{t-1}U}(y' \mid y, u) = f_\mathcal{E}(y' - \alpha y - u)$. Consequently, a model is represented by $(\alpha, f_\mathcal{E}, f_{D_t|Y_tU}, f_{Y_1U}, f_{Z|U})$. For ease of writing, let this model be denoted by $\theta = (\alpha, \tilde{f}_1, f_2, f_3, f_4)$. Accordingly, write a set of such models by $\Theta = \mathcal{A} \times \tilde{F}_1 \times F_2 \times F_3 \times F_4$

Under these notations, Corollary 1 and Remark 9 characterize a sieve semiparametric estimator $\hat{\theta}$ of θ_0 as the solution to

$$\max_{\theta \in \Theta_{k(n)}} \frac{1}{n} \sum_{i=1}^{n} l(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}; \theta)$$

for some sieve space $\Theta_{k(n)} = \mathcal{A}_{k(n)} \times \tilde{\mathcal{F}}_{1,k_1(n)} \times \mathcal{F}_{2,k_2(n)} \times \mathcal{F}_{3,k_3(n)} \times \mathcal{F}_{4,k_4(n)} \subset \Theta$, where

$$\begin{aligned}
l(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}; \theta) &:= \mathbb{1}\{D_{i1} = 1\} \cdot l_1(Y_{i2}, Y_{i1}, Z_i; \theta) \\
&\quad + \mathbb{1}\{D_{i2} = D_{i1} = 1\} \cdot l_2(Y_{i3}, Y_{i2}, Y_{i1}, Z_i; \theta) - l_3(\theta) - l_4(\theta), \\
l_1(Y_{i2}, Y_{i1}, Z_i; \theta) &:= \log \int \tilde{f}_1(Y_{i2} - \alpha Y_{i1} - u) f_2(1 \mid Y_{i1}, u) f_3(Y_{i1}, u) f_4(Z_i \mid u) d\mu(u), \\
l_2(Y_{i3}, Y_{i2}, Y_{i1}, Z_i; \theta) &:= \log \int \tilde{f}_1(Y_{i3} - \alpha Y_{i2} - u) f_1(Y_{i2} - \alpha Y_{i1} - u) \\
&\quad \times f_2(1 \mid Y_{i2}, u) f_2(1 \mid Y_{i1}, u) f_3(Y_{i1}, u) f_4(Z_i \mid u) d\mu(u), \\
l_3(\theta) &:= \int f_2(1 \mid y_1, u) f_3(y_1, u) d\mu(y_1, u), \qquad \text{and} \\
l_4(\theta) &:= \int \tilde{f}_1(y_2 - \alpha y_1 - u) f_2(1 \mid y_2, u) f_2(1 \mid y_1, u) f_3(y_1, u) d\mu(y_2, y_1, u).
\end{aligned}$$

The asymptotic distribution of $\hat{\alpha}$ can be derived by following the method of Ai and Chen (2003), which was also used in Blundell, Chen, Kristensen (2007) and Hu and Schennach (2008). First, I introduce auxiliary notations. Define the path-wise derivative

$$l'_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; \theta - \theta_0) = \lim_{r \to 0} \frac{l(y_3, y_2, y_1, z, d_2, d_1; \theta(\theta_0, r)) - l(y_3, y_2, y_1, z, d_2, d_1; \theta_0)}{r}$$

where $\theta(\theta_0, \cdot) : \mathbb{R} \to \Theta$ denotes a path such that $\theta(\theta_0, 0) = \theta_0$ and $\theta(\theta_0, 1) = \theta$. Similarly define the path-wise derivative with respect to each component of $(\tilde{f}_1, f_2, f_3, f_4)$ by

$$\begin{aligned}
\frac{d}{d\tilde{f}_1} l_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; \tilde{f}_1 - \tilde{f}_{10}) &= \lim_{r \to 0} \frac{l(y_3, y_2, y_1, z, d_2, d_1; \tilde{f}_1(\theta_0, r)) - l(y_3, y_2, y_1, z, d_2, d_1; \theta_0)}{r} \\
\frac{d}{df_2} l_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; f_2 - f_{20}) &= \lim_{r \to 0} \frac{l(y_3, y_2, y_1, z, d_2, d_1; f_2(\theta_0, r)) - l(y_3, y_2, y_1, z, d_2, d_1; \theta_0)}{r} \\
\frac{d}{df_3} l_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; f_3 - f_{30}) &= \lim_{r \to 0} \frac{l(y_3, y_2, y_1, z, d_2, d_1; f_3(\theta_0, r)) - l(y_3, y_2, y_1, z, d_2, d_1; \theta_0)}{r} \\
\frac{d}{df_4} l_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; f_4 - f_{40}) &= \lim_{r \to 0} \frac{l(y_3, y_2, y_1, z, d_2, d_1; f_4(\theta_0, r)) - l(y_3, y_2, y_1, z, d_2, d_1; \theta_0)}{r}
\end{aligned}$$

where $\tilde{f}_1(\theta_0, \cdot) : \mathbb{R} \to \tilde{F}_1$, $f_2(\theta_0, \cdot) : \mathbb{R} \to F_2$, $f_3(\theta_0, \cdot) : \mathbb{R} \to F_3$, and $f_4(\theta_0, \cdot) : \mathbb{R} \to F_4$ denote paths as before.

Recenter the set of parameters by $\Omega = \Theta - \theta_0$ so that

$$\langle v_1, v_2 \rangle = \mathrm{E}\left[l'_{\theta_0}(Y_3, Y_2, Y_1, Z, D_2, D_1; v_1) l'_{\theta_0}(Y_3, Y_2, Y_1, Z, D_2, D_1; v_2) \right]$$

defines an inner product on Ω. Furthermore, by taking the closure $\overline{\Omega}$, we obtain a complete space $\overline{\Omega}$ with respect to the topology induced by $\langle \cdot, \cdot \rangle$, hence a Hilbert space $(\overline{\Omega}, \langle \cdot, \cdot \rangle)$. It can be written as $\overline{\Omega} = \mathbb{R} \times \overline{W}$ where $W = \tilde{F}_1 \times F_2 \times F_3 \times F_4 - (\tilde{f}_{10}, f_{20}, f_{30}, f_{40})$. Given these notations, define

$$\begin{aligned} w^* := (\tilde{f}_1^*, f_2^*, f_3^*, f_4^*) = \arg\min_{w \in \overline{W}} \mathrm{E}\Bigg[\Bigg(& \frac{d}{d\alpha} l(Y_3, Y_2, Y_1, Z, D_2, D_1; \theta_0) - \frac{d}{d\tilde{f}_1} l_{\theta_0}(Y_3, Y_2, Y_1, Z, D_2, D_1; \tilde{f}_1) \\ & - \frac{d}{df_2} l_{\theta_0}(Y_3, Y_2, Y_1, Z, D_2, D_1; f_2) - \frac{d}{df_3} l_{\theta_0}(Y_3, Y_2, Y_1, Z, D_2, D_1; f_3) \\ & - \frac{d}{df_4} l_{\theta_0}(Y_3, Y_2, Y_1, Z, D_2, D_1; f_4) \Bigg)^2 \Bigg]. \end{aligned}$$

Given this w^*, next define

$$\begin{aligned} \Phi_{w^*}(y_3, y_2, y_1, z, d_2, d_1) := & \frac{d}{d\alpha} l(y_3, y_2, y_1, z, d_2, d_1; \theta_0) - \frac{d}{d\tilde{f}_1} l_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; \tilde{f}_1) \\ & - \frac{d}{df_2} l_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; f_2) - \frac{d}{df_3} l_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; f_3) \\ & - \frac{d}{df_4} l_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; f_4). \end{aligned}$$

The next assumption sets a moment condition.

Assumption 4 (Bounded Second Moment). $\sigma := \mathrm{E}\left[\Phi_{w^*}(Y_3, Y_2, Y_1, Z, D_2, D_1)^2 \right] < \infty$

The mapping $\theta - \theta_0 \overset{s}{\mapsto} \alpha - \alpha_0$ is a linear functional on $\overline{\Omega}$. Since $(\overline{\Omega}, \langle \cdot, \cdot \rangle)$ is a Hilbert space, the Riesz Representation Theorem guarantees the existence of $v^* \in \overline{\Omega}$ such that $s(\theta - \theta_0) = \langle v^*, \theta - \theta_0 \rangle$ for all $\theta \in \overline{\Theta}$ under Assumption 4. Moreover, this representing vector has the explicit formula $v^* = (\sigma^{-1}, -\sigma^{-1} w^*)$. Using Corollary 1 of Shen (1997) yields asymptotic distribution of $\sqrt{N}(\alpha - \alpha_0) = \sqrt{N} \left\langle v^*, \hat{\theta} - \theta_0 \right\rangle$, which is $N(0, \sigma^{-1})$.

In order to invoke Shen's corollary, a couple of additional notations need to be introduced. The remainder of the linear approximation is

$$\begin{aligned} r(y_3, y_2, y_1, z, d_2, d_1; \theta - \theta_0) \quad := \quad & l(y_3, y_2, y_1, z, d_2, d_1; \theta) - l(y_3, y_2, y_1, z, d_2, d_1; \theta_0) \\ & - l'_{\theta_0}(y_3, y_2, y_1, z, d_2, d_1; \theta - \theta_0) \end{aligned}$$

A divergence measure is defined by

$$K(\theta_0, \theta) := \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}\left[l(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}; \theta_0) - l(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}; \theta) \right].$$

Denote the empirical process induced by g by

$$\nu_n(g) := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(g(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}) - \mathrm{E} g(Y_{i3}, Y_{i2}, Y_{i1}, Z_i, D_{i2}, D_{i1}) \right)$$

For a perturbation ϵ_n such that $\epsilon_n = o(n^{-1/2})$, let $\theta^*(\theta, \epsilon_n) = (1 - \epsilon_n)\theta + \epsilon_n(u^* + \theta_0)$ where $u^* = \pm v^*$. Lastly, P_n denote the projection $\Theta \to \Theta_n$. The following high-level assumptions of Shen (1997) guarantees asymptotic normality of $\sqrt{N} \left\langle v^*, \hat{\theta} - \theta_0 \right\rangle$, or equivalently of $\sqrt{N}(\alpha - \alpha_0)$.

Assumption 5 (Regularity). (i) $\sup_{\{\theta \in \Theta_n | \|\theta - \theta_0\| \leqslant \delta_0\}} n^{-1/2} \nu_n(r(Y_3, Y_2, Y_1, Z, D_2, D_1; \theta - \theta_0) - r(Y_3, Y_2, Y_1, Z, D_2, D_1; P_n(\theta^*(\theta, \epsilon_n)) - \theta_0)) = O_p(\epsilon_n^2)$. (ii) $\sup_{\{\theta \in \Theta_n | 0 < \|\theta - \theta_0\| \leqslant \delta_n\}} [K(\theta_0, P_n(\theta^*(\theta, \epsilon_n)) - K(\theta_0, \theta)] - \frac{1}{2} \left[\|\theta^*(\theta, \epsilon_n) - \theta_0\|^2 - \|\theta - \theta_0\|^2 \right] = O(\epsilon_n^2)$. (iii) $\sup_{\{\theta \in \Theta_n | 0 < \|\theta - \theta_0\| \leqslant \delta_n\}} \|\theta^*(\theta, \epsilon_n) - P_n(\theta^*(\theta, \epsilon_n))\| = O(\delta_n^{-1} \epsilon_n^2)$. (iv) $\sup_{\{\theta \in \Theta_n | \|\theta - \theta_0\| \leqslant \delta_n\}} n^{-1/2} \nu_n(l'_{\theta_0}(\cdots ; \theta^*(\theta, \epsilon_n) - P_n(\theta^*(\theta, \epsilon_n)))) = O_p(\epsilon_n^2)$. (v) $\sup_{\{\theta \in \Theta_n | \|\theta - \theta_0\| \leqslant \delta_n\}} n^{-1/2} \nu_n (l'_{\theta_0}(\cdots ; \theta - \theta_0)) = O_p(\epsilon_n)$.

Proposition 2 (Asymptotic Distribution of a Semiparametric Estimator). *Suppose that Restrictions 1, 2, 3, and 4 and Assumptions 4 and 5 hold. Then,* $\sqrt{N}(\alpha - \alpha_0) \xrightarrow{d} N(0, \sigma^{-1})$.

10. Appendix: Special Cases and Generalizations of the Baseline Model

10.1. A Variety of Missing Observations. While the baseline model considered in the paper induces a permanent dropout from data by a hazard selection, variants of the model can be encompassed as special cases under which the main

identification remains to hold. Specifically, we consider the following Classes 1 and 2 as special models of Class 3.

Class 1 (Nonseparable Dynamic Panel Data Model).

$$\begin{cases} Y_t = g(Y_{t-1}, U, \mathcal{E}_t) & t = 2, \cdots, T \qquad \text{(State Dynamics)} \\ F_{Y_1 U} & \text{(Initial joint distribution of } (Y_1, U)) \\ Z = \zeta(U, W) & \text{(Optional: nonclassical proxy of } U) \end{cases}$$

□

Class 2 (Nonseparable Dynamic Panel Data Model with Missing Observations).

$$\begin{cases} Y_t = g(Y_{t-1}, U, \mathcal{E}_t) & t = 2, \cdots, T \qquad \text{(State Dynamics)} \\ D_t = h(Y_t, U, V_t) & t = 1, \cdots, T-1 \qquad \text{(Selection)} \\ F_{Y_1 U} & \text{(Initial joint distribution of } (Y_1, U)) \\ Z = \zeta(U, W) & \text{(Optional: nonclassical proxy of } U) \end{cases}$$

where Y_t is censored by the binary indicator D_t of sample selection as follows:

$$\begin{cases} Y_t \text{ is observed} & \text{if } D_{t-1} = 1 \text{ or } t = 1. \\ Y_t \text{ is unobserved} & \text{if } D_{t-1} = 0 \text{ and } t > 1. \end{cases}$$

□

A representative example of this instantaneous selection is the Roy model such as $h(y, u, v) = \mathbb{1}\{\mathrm{E}[\pi(g(y, u, \mathcal{E}_{t+1}), u)] \geqslant c(u, v)\}$ where π measures payoffs and c measures costs. The following is the baseline model considered in the paper.

Class 3 (Nonseparable Dynamic Panel Data Model with Hazards).

$$\begin{cases} Y_t = g(Y_{t-1}, U, \mathcal{E}_t) & t = 2, \cdots, T \qquad \text{(State Dynamics)} \\ D_t = h(Y_t, U, V_t) & t = 1, \cdots, T-1 \qquad \text{(Hazard Model)} \\ F_{Y_1 U} & \text{(Initial joint distribution of } (Y_1, U)) \\ Z = \zeta(U, W) & \text{(Optional: nonclassical proxy of } U) \end{cases}$$

where $D_t = 0$ induces a hazard of permanent dropout in the following manner:

$$\begin{cases} Y_1 \text{ is observed,} & \\ Y_2 \text{ is observed} & \text{if } D_1 = 1, \\ Y_3 \text{ is observed} & \text{if } D_1 = D_2 = 1. \end{cases}$$

□

The present appendix section proves that identification of Class 3 implies identification of Classes 1 and 2. The observable parts of the joint distributions in each of the three classes include (but are not limited to) the following:

Class 1: Observe $F_{Y_3Y_2Y_1ZD_2D_1}$, $F_{Y_2Y_1ZD_1}$, $F_{Y_3Y_1ZD_2}$, and $F_{Y_3Y_2ZD_2}$

Class 2: Observe $F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\cdot,\cdot,\cdot,1,1)$, $F_{Y_2Y_1ZD_1}(\cdot,\cdot,\cdot,1)$, and $F_{Y_3Y_1ZD_2}(\cdot,\cdot,\cdot,1)$

Class 3: Observe $F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\cdot,\cdot,\cdot,1,1)$ and $F_{Y_2Y_1ZD_1}(\cdot,\cdot,\cdot,1)$

Since the selection variable D_t in Class 1 is not defined, we assume without loss of generality that it is degenerate at $D_t = 1$ in Class 1.

The problem of identification under each class can be characterized by the well-definition of the following maps:

Class 1: $(F_{Y_3Y_2Y_1ZD_2D_1}, F_{Y_2Y_1ZD_1}, F_{Y_3Y_1ZD_2}, F_{Y_3Y_2ZD_2}) \overset{\iota_1}{\mapsto} (g, F_{Y_1U}, \zeta)$

Class 2: $(F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\cdot,\cdot,\cdot,1,1), F_{Y_2Y_1ZD_1}(\cdot,\cdot,\cdot,1), F_{Y_3Y_1ZD_2}(\cdot,\cdot,\cdot,1)) \overset{\iota_2}{\mapsto} (g, h, F_{Y_1U}, \zeta)$

Class 3: $(F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\cdot,\cdot,\cdot,1,1), F_{Y_2Y_1ZD_1}(\cdot,\cdot,\cdot,1) \overset{\iota_3}{\mapsto} (g, h, F_{Y_1U}, \zeta)$

The main identification result of this paper was to show well-definition of the map ι_3. Therefore, in order to argue that identification of Class 3 implies identification of Classes 1 and 2, it suffices to claim that the well-definition of ι_3 implies well-definition of the maps ι_1 and ι_2.

First, note that the trivial projections

$$(F_{Y_3Y_2Y_1ZD_2D_1}, F_{Y_2Y_1ZD_1}, F_{Y_3Y_1ZD_2}, F_{Y_3Y_2ZD_2}) \overset{\pi_1}{\mapsto} (F_{Y_3Y_2Y_1ZD_2D_1}, F_{Y_2Y_1ZD_1}, F_{Y_3Y_1ZD_2}) \quad \text{and}$$

$$(F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\cdot,\cdot,\cdot,1,1), F_{Y_2Y_1ZD_1}(\cdot,\cdot,\cdot,1), F_{Y_3Y_1ZD_2}(\cdot,\cdot,\cdot,1))$$

$$\overset{\pi_2}{\mapsto} (F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\cdot,\cdot,\cdot,1,1), F_{Y_2Y_1ZD_1}(\cdot,\cdot,\cdot,1))$$

are well-defined. Second, by the construction of degenerate random variable $D_t = 1$ in Class 1, the map

$$\begin{aligned}&(F_{Y_3Y_2Y_1ZD_2D_1}, F_{Y_2Y_1ZD_1}, F_{Y_3Y_1ZD_2}, F_{Y_3Y_2ZD_2})\\ &\quad \overset{\kappa_1}{\mapsto} (F_{Y_3Y_2Y_1ZD_2D_1}(\cdot,\cdot,\cdot,\cdot,1,1), F_{Y_2Y_1ZD_1}(\cdot,\cdot,\cdot,1), F_{Y_3Y_1ZD_2}(\cdot,\cdot,\cdot,1))\end{aligned}$$

is well-defined in Class 1. Third, the trivial projection

$$(g, h, F_{Y_1U}, \zeta) \overset{\rho}{\mapsto} (g, F_{Y_1U}, \zeta)$$

is well-defined.

Now, notice that

$$\begin{aligned}\iota_1 &= \rho \circ \iota_3 \circ \kappa_1 \circ \pi_1 \quad \text{in Class 1, and}\\ \iota_2 &= \iota_3 \circ \pi_2 \qquad\qquad\quad \text{in Class 2.}\end{aligned}$$

Therefore, the well-definition of ι_3 implies well-definition of ι_1 and ι_2 in particular. Therefore, identification of Class 3 implies identification of Classes 1 and 2.

10.2. Identification without a Nonclassical Proxy Variable. The main result of this paper assumed use of a nonclassical proxy variable Z. However, this use was mentioned to be optional, and one can substitute a slightly longer panel $T = 6$ for use of a proxy variable. In this section we show how the model (g, h, F_{Y_1U}) can be identified from the joint distribution $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\cdot,\cdot,\cdot,\cdot,\cdot,\cdot,1,1,1,1,1)$ that follows from $T = 6$ time periods of unbalanced panel data without additional information Z.

Restriction 5 (Independence).

(i) Exogeneity of $\mathcal{E}_t$: $\quad \mathcal{E}_t \perp\!\!\!\perp (U, Y_1, \{\mathcal{E}_s\}_{s<t}, \{V_s\}_{s<t}, W)$ for all $t \geq 2$.

(ii) Exogeneity of V_t: $\quad V_t \perp\!\!\!\perp (U, Y_1, \{\mathcal{E}_s\}_{s\leqslant t}, \{V_s\}_{s<t})$ for all $t \geq 1$.

For simplicity of notation, we compress the nondegenerate random variable Y_3 into a binary random variable $Z := \eta(Y_3)$ with a known transformation η such that part (iii) of the following rank condition is satisfied. As the notation suggests, this Z serves as a substitute for a nonclassical proxy variable.

RESTRICTION 6 (Rank Conditions). The following conditions hold for every $y \in \mathcal{Y}$:

(i) Heterogeneous Dynamics: the integral operator $P_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$ defined by $P_y\xi(y') = \int f_{Y_3|Y_2U}(y' \mid y, u) \cdot \xi(u)du$ is bounded and invertible.

(ii) There exist y_4 and y_2 satisfying the following conditions:

Nondegeneracy: $f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 1, 1, 1, 1 \mid y_2, u)$ is bounded away from 0 and 1 for all u.

Relevance: $\frac{f_{Y_4ZD_5D_4D_3|Y_2U}(y_4,1,1,1,1|y_2,u)}{f_{Y_4ZD_5D_4D_3|Y_2U}(y_4,0,1,1,1|y_2,u)} \neq \frac{f_{Y_4ZD_5D_4D_3|Y_2U}(y_4,1,1,1,1|y_2,u')}{f_{Y_4ZD_5D_4D_3|Y_2U}(y_4,0,1,1,1|y_2,u')}$ whenever $u \neq u'$.

(iii) No Extinction: $f_{D_2|Y_2U}(1 \mid y, u) > 0$ for all $u \in \mathcal{U}$.

(iv) Initial Heterogeneity: the integral operator $S_y : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$ defined by $S_y\xi(u) = \int f_{Y_2Y_1UD_1U}(y, y', u, 1) \cdot \xi(y')dy'$ is bounded and invertible.

LEMMA 5 (Independence). *The following implications hold:*

(i) Restriction 5 (i) $\Rightarrow \mathcal{E}_6 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5, V_1, V_2, V_3, V_4, V_5)$

$\Rightarrow Y_6 \perp\!\!\!\perp (Y_1, Y_2, Y_3, Y_4, D_1, D_2, D_3, D_4, D_5) \mid (Y_5, U)$.

(ii) Restriction 5 (i) & (ii) $\Rightarrow (\mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5, V_3, V_4, V_5) \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2)$

$\Rightarrow (Y_3, Y_4, Y_5, D_3, D_4, D_5) \perp\!\!\!\perp (Y_1, D_1, D_2) \mid (Y_2, U)$.

(iii) Restriction 5 (i) $\Rightarrow \mathcal{E}_2 \perp\!\!\!\perp (U, Y_1, V_1) \Rightarrow Y_2 \perp\!\!\!\perp D_1 \mid (Y_1, U)$.

(iv) Restriction 5 (ii) $\Rightarrow V_2 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1) \Rightarrow D_2 \perp\!\!\!\perp (Y_1, D_1) \mid (Y_2, U)$.

PROOF. In order to prove the lemma, we use the following two properties of conditional independence:

CI.1. $A \perp\!\!\!\perp B$ implies $A \perp\!\!\!\perp B \mid \phi(B)$ for any Borel function ϕ.

CI.2. $A \perp\!\!\!\perp B \mid C$ implies $A \perp\!\!\!\perp \phi(B, C) \mid C$ for any Borel function ϕ.

(i) First, applying CI.1 to $\mathcal{E}_6 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5, V_1, V_2, V_3, V_4, V_5)$ and using the definition of g yield

$$\mathcal{E}_6 \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5, V_1, V_2, V_3, V_4, V_5) \mid (Y_5, U).$$

Next, applying CI.2 to this conditional independence and using the definitions of g and h yield

$$\mathcal{E}_6 \perp\!\!\!\perp (Y_1, Y_2, Y_3, Y_4, D_1, D_2, D_3, D_4, D_5, Z) \mid (Y_5, U).$$

Applying CI.2 again to this conditional independence and using the definition of g yield

$$Y_6 \perp\!\!\!\perp (Y_1, Y_2, Y_3, Y_4, D_1, D_2, D_3, D_4, D_5, Z) \mid (Y_5, U).$$

(ii) First, applying CI.1 to $(\mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5, V_3, V_4, V_5) \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2)$ and using the definition of g yield

$$(\mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5, V_3, V_4, V_5) \perp\!\!\!\perp (U, Y_1, \mathcal{E}_2, V_1, V_2) \mid (Y_2, U)$$

Next, applying CI.2 to this conditional independence and using the definitions of g and h yield

$$(\mathcal{E}_3, \mathcal{E}_4, \mathcal{E}_5, V_3, V_4, V_5) \perp\!\!\!\perp (Y_1, D_1, D_2) \mid (Y_2, U)$$

Applying CI.2 again to this conditional independence and using the definition of g yield

$$(Y_3, Y_4, Y_5, D_3, D_4, D_5) \perp\!\!\!\perp (Y_1, D_1, D_2) \mid (Y_2, U)$$

(iii) The proof is the same as that of Lemma 3 (ii).

(iv) The proof is the same as that of Lemma 3 (iii). □

LEMMA 6 (Invariant Transition).

(i) Under Restrictions 1 and 5 (i), $F_{Y_t|Y_{t-1}U}(y' \mid y, u) = F_{Y_{t'}|Y_{t'-1}U}(y' \mid y, u)$ *for all* y', y, u, t, t'.

(ii) Under Restrictions 1 and 5 (ii), $F_{D_2|Y_2U}(d \mid y, u) = F_{D_1|Y_1U}(d \mid y, u)$ *for all* d, y, u.

This lemma can be proved similarly to Lemma 4.

LEMMA 7 (Identification). *Under Restrictions 1, 4, 5, & 6,* $(F_{Y_3|Y_2U}, F_{D_2|Y_2U}, F_{Y_1U})$ *is uniquely determined by* $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, 1, 1, 1, 1, 1)$ *and* $F_{Y_2Y_1D_1}(\cdot, \cdot, 1)$.

PROOF. Given fixed (y_5, y_4, z, y_2), define the operators $L_{y_5,y_4,z,y_2} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_{Y_t})$, $P_{y_5} : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$, $Q_{y_5,y_4,z,y_2} : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$, $R_{y_2} : \mathcal{L}^2(F_U) \to$

$\mathcal{L}^2(F_U)$, $S_{y_2} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$, $T_{y_2} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$, and $T'_{y_2} : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$ by

$$\begin{aligned}
(L_{y_5,y_4,z,y_2}\xi)(y_6) &= \int f_{Y_6Y_5Y_4ZY_2Y_1D_5D_4D_3D_2D_1}(y_6, y_5, y_4, z, y_2, y_1, 1, 1, 1, 1, 1) \cdot \xi(y_1)dy_1, \\
(P_{y_5}\xi)(y_3) &= \int f_{Y_6|Y_5U}(y_6 \mid y_5, u) \cdot \xi(u)du, \\
(Q_{y_5,y_4,z,y_2}\xi)(u) &= f_{Y_5Y_4ZD_5D_4D_3|Y_2U}(y_5, y_4, z, 1, 1, 1 \mid y_2, u) \cdot \xi(u), \\
(R_{y_2}\xi)(u) &= f_{D_2|Y_2U}(1 \mid y_2, u) \cdot \xi(u), \\
(S_{y_2}\xi)(u) &= \int f_{Y_2Y_1UD_1}(y_2, y_1, u, 1) \cdot \xi(y_1)dy_1, \\
(T_{y_2}\xi)(u) &= \int f_{Y_1|Y_2UD_2D_1}(y_1 \mid y_2, u, 1, 1) \cdot \xi(y_1)dy_1, \\
(T'_{y_2}\xi)(u) &= f_{Y_2UD_2D_1}(y_2, u, 1, 1) \cdot \xi(u)
\end{aligned}$$

respectively.

Step 1: Uniqueness of $F_{Y_6|Y_5U}$ and $F_{Y_5Y_4ZD_5D_4D_3|Y_2U}(\ \cdot,\ \cdot,\ \cdot, 1, 1, 1 \mid\ \cdot,\ \cdot)$

The kernel $f_{Y_6Y_5Y_4ZY_2Y_1D_5D_4D_3D_2D_1}(\ \cdot, y_5, y_4, z, y_2,\ \cdot, 1, 1, 1, 1, 1)$ of the integral operator L_{y_5,y_4,z,y_2} can be rewritten as

$$\begin{aligned}
& f_{Y_6Y_5Y_4ZY_2Y_1D_5D_4D_3D_2D_1}(y_6, y_5, y_4, z, y_2, y_1, 1, 1, 1, 1, 1) \\
= & \int f_{Y_6|Y_5Y_4ZY_2Y_1UD_5D_4D_3D_2D_1}(y_6 \mid y_5, y_4, z, y_2, y_1, u, 1, 1, 1, 1, 1) \\
& \times f_{Y_5Y_4ZD_5D_4D_3|Y_2Y_1UD_2D_1}(y_5, y_4, z, 1, 1, 1 \mid y_2, y_1, u, 1, 1) \\
& \times f_{D_2|Y_2Y_1UD_1}(1 \mid y_2, y_1, u, 1) \cdot f_{Y_2Y_1UD_1}(y_2, y_1, u, 1)\ du
\end{aligned} \tag{21}$$

But by Lemma 5 (i), (ii), and (iv), respectively, Restriction 5 implies that

$$f_{Y_6|Y_5Y_4ZY_2Y_1UD_5D_4D_3D_2D_1}(y_6 \mid y_5, y_4, z, y_2, y_1, u, 1, 1, 1, 1, 1) = f_{Y_6|Y_5U}(y_5 \mid y_5, u),$$

$$f_{Y_5Y_4ZD_5D_4D_3|Y_2Y_1UD_2D_1}(y_5, y_4, z, 1, 1, 1 \mid y_2, y_1, u, 1, 1) = f_{Y_5Y_4ZD_5D_4D_3|Y_2U}(y_5, y_4, z, 1, 1, 1 \mid y_2, u),$$

$$f_{D_2|Y_2Y_1UD_1}(1 \mid y_2, y_1, u, 1) = f_{D_2|Y_2U}(1 \mid y_2, u).$$

Equation (21) thus can be rewritten as

$$\begin{aligned} & f_{Y_6Y_5Y_4ZY_2Y_1D_5D_4D_3D_2D_1}(y_6, y_5, y_4, z, y_2, y_1, 1, 1, 1, 1, 1) \\ = & \int f_{Y_6|Y_5U}(y_6 \mid y_5, u) \cdot f_{Y_5Y_4ZD_5D_4D_3|Y_2U}(y_5, y_4, z, 1, 1, 1 \mid y_2, u) \\ & \times f_{D_2|Y_2U}(1 \mid y_2, u) \cdot f_{Y_2Y_1UD_1}(y_2, y_1, u, 1)\ du \end{aligned}$$

But this implies that the integral operator L_{y_4,y_3,y_2} is written as the operator composition

$$L_{y_5,y_4,z,y_2} = P_{y_5} Q_{y_5,y_4,z,y_2} R_{y_2} S_{y_2}.$$

Restriction 6 (i), (ii), (iii), and (iv) imply that the operators P_{y_5}, Q_{y_5,y_4,z,y_2}, R_{y_2}, and S_{y_2} are invertible, respectively. Hence so is L_{y_5,y_4,z,y_2}. Using the two values $\{0, 1\}$ of Z, form the product

$$L_{y_5,y_4,1,y_2} L^{-1}_{y_5,y_4,0,y_2} = P_{y_5} Q_{y_4,1/0,y_2} P^{-1}_{y_5}$$

where $Q_{y_4,1/0,y_2} = Q_{y_5,y_4,1,y_2} Q^{-1}_{y_5,y_4,0,y_2}$ is the multiplication operator with proxy odds defined by

$$\begin{aligned} (Q_{y_4,1/0,y_2}\xi)(u) &= \frac{f_{Y_5Y_4ZD_5D_4D_3|Y_2U}(y_5, y_4, 1, 1, 1, 1 \mid y_2, u)}{f_{Y_5Y_4ZD_5D_4D_3|Y_2U}(y_5, y_4, 0, 1, 1, 1 \mid y_2, u)} \xi(u) \\ &= \frac{f_{Y_5|Y_4U}(y_5 \mid y_4, u) \cdot f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 1, 1, 1, 1 \mid y_2, u)}{f_{Y_5|Y_4U}(y_5 \mid y_4, u) \cdot f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 0, 1, 1, 1 \mid y_2, u)} \xi(u) \\ &= \frac{f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 1, 1, 1, 1 \mid y_2, u)}{f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 0, 1, 1, 1 \mid y_2, u)} \xi(u). \end{aligned}$$

Note the invariance of this operator in y_5, hence the notation. By Restriction 6 (ii), the operator $L_{y_5,y_4,1,y_2} L^{-1}_{y_5,y_4,0,y_2}$ is bounded. The expression $L_{y_5,y_4,1,y_2} L^{-1}_{y_5,y_4,0,y_2} = P_{y_5} Q_{y_4,1/0,y_2} P^{-1}_{y_5}$ thus allows unique eigenvalue-eigenfunction decomposition as in the proof of Lemma 2.

The distinct proxy odds as in Restriction 6 (ii) guarantee distinct eigenvalues and single dimensionality of the eigenspace associated with each eigenvalue. Within each of the single-dimensional eigenspace is a unique eigenfunction pinned down by $\mathcal{L}^1$-normalization because of the unity of integrated densities. The eigenvalues $\lambda(u)$ yield the multiplier of the operator $Q_{y_4,1/0,y_2}$, hence $\lambda_{y_4,y_2}(u) = f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 1, 1, 1, 1 \mid$

$y_2, u)/f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 0, 1, 1, 1 \mid y_2, u)$. This proxy odds in turn identifies the function $f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, \cdot, 1, 1, 1 \mid y_2, u)$ since Z is binary. The corresponding normalized eigenfunctions are the kernels of the integral operator P_{y_5}, hence $f_{Y_6|Y_5U}(\cdot \mid y_5, u)$. Lastly, Restriction 4 facilitates unique ordering of the eigenfunctions $f_{Y_6|Y_5U}(\cdot \mid y_5, u)$ by the distinct concrete values of $u = \lambda_{y_4,y_2}(u)$. This is feasible because the eigenvalues $\lambda_{y_4,y_2}(u) = f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 1, 1, 1, 1 \mid y_2, u)/f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, 0, 1, 1, 1 \mid y_2, u)$ are invariant from y_5. That is, eigenfunctions $f_{Y_6|Y_5U}(\cdot \mid y_5, u)$ of the operator $L_{y_5,y_4,1,y_2}L^{-1}_{y_5,y_4,0,y_2}$ across different y_5 can be uniquely ordered in u invariantly from y_5 by the common set of ordered distinct eigenvalues $u = \lambda_{y_4,y_2}(u)$.

Therefore, $F_{Y_6|Y_5U}$ and $F_{Y_4ZD_5D_4D_3|Y_2U}(y_4, \cdot, 1, 1, 1 \mid y_2, u)$ are uniquely determined by the joint distribution $F_{Y_6Y_5Y_4ZY_2Y_1D_5D_4D_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, 1, 1, 1, 1, 1)$, which in turn is uniquely determined by the $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, 1, 1, 1, 1, 1)$. The multiplier of the operator Q_{y_5,y_4,z,y_2} is of the form

$$\begin{aligned} f_{Y_5Y_4ZD_5D_4D_3|Y_2U}(y_5, y_4, z, 1, 1, 1 \mid y_2, u) &= f_{Y_5|Y_4U}(y_5 \mid y_4, u) \cdot f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, z, 1, 1, 1 \mid y_2, u) \\ &= f_{Y_6|Y_5U}(y_5 \mid y_4, u) \cdot f_{Y_4ZD_5D_4D_3|Y_2U}(y_4, z, 1, 1, 1 \mid y_2, u) \end{aligned}$$

by Lemma 6 (i), where the right-hand side object has been identified. Consequently, the operators P_{y_5} and Q_{y_5,y_4,z,y_2} are uniquely determined for each combination of y_5, y_4, z, y_2.

Step 2: Uniqueness of $F_{Y_2Y_1UD_1}(\cdot, \cdot, \cdot, 1)$

By Lemma 5 (iii), Restriction 5 implies $f_{Y_2|Y_1ZUD_1}(y' \mid y, z, u, 1) = f_{Y_2|Y_1U}(y' \mid y, u)$. Using this equality, write the density of the observed $F_{Y_2Y_1ZD_1}(\cdot, \cdot, \cdot, 1)$ as

$$\begin{aligned} f_{Y_2Y_1D_1}(y_2, y_1, 1) &= \int f_{Y_2|Y_1UD_1}(y_2 \mid y_1, u, 1) f_{Y_1UD_1}(y_1, u, 1) du \\ &= \int f_{Y_2|Y_1U}(y_2 \mid y_1, u) f_{Y_1UD_1}(y_1, u, 1) du \end{aligned} \tag{22}$$

By Lemma 4 (i), $F_{Y_6|Y_5U}(y' \mid y, u) = F_{Y_2|Y_1U}(y' \mid y, u)$ for all y', y, u. Therefore, we can write the operator P_y as

$$(P_{y_1}\xi)(y_2) = \int f_{Y_6|Y_5U}(y_2 \mid y_1, u) \cdot \xi(u) du = \int f_{Y_2|Y_1U}(y_2 \mid y_1, u) \cdot \xi(u) du.$$

With this operator notation, it follows from (22) that

$$f_{Y_2Y_1D_1}(\,\cdot\,, y_1, 1) = P_{y_1} f_{Y_1UD_1}(y_1,\ \cdot\,, 1).$$

By Restriction 6 (i) and (ii), this operator equation can be solved for $f_{Y_1UD_1}(y,\ \cdot\,, 1)$ as

$$f_{Y_1UD_1}(y_1,\ \cdot\,, 1) = P_{y_1}^{-1} f_{Y_2Y_1D_1}(\,\cdot\,, y_1, 1) \tag{23}$$

Recall that P_y was shown in Step 1 to be uniquely determined by the observed $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot, 1, 1, 1, 1, 1)$. The function $f_{Y_2Y_1D_1}(\,\cdot\,, y, 1)$ is also uniquely determined by the observed joint distribution $F_{Y_2Y_1D_1}(\,\cdot\,,\ \cdot\,, 1)$ up to null sets. Therefore, (22) shows that $f_{Y_1UD_1}(\,\cdot\,,\ \cdot\,, 1)$ is uniquely determined by the pair of the observed joint distributions $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot, 1, 1, 1, 1, 1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\ \cdot\,, 1)$.

Using the solution to the above inverse problem, we can write the kernel of the operator S_{y_2} as

$$\begin{aligned}
f_{Y_2Y_1UD_1}(y_2, y_1, u, 1) &= f_{Y_2|Y_1UD_1}(y_2 \mid y_1, u, 1) \cdot f_{Y_1UD_1}(y_1, u, 1) \\
&= f_{Y_2|Y_1U}(y_2 \mid y_1, u) \cdot f_{Y_1UD_1}(y_1, u, 1) \\
&= f_{Y_6|Y_5U}(y_2 \mid y_1, u) \cdot f_{Y_1UD_1}(y_1, u, 1) \\
&= f_{Y_6|Y_5U}(y_2 \mid y_1, u) \cdot [P_{y_1}^{-1} f_{Y_2Y_1D_1}(\,\cdot\,, y_1, 1)](u)
\end{aligned}$$

where the second equality follows from Lemma 5 (iii), the third equality follows from Lemma 4 (i), and the forth equality follows from (23). Since $f_{Y_6|Y_5U}$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot, 1, 1, 1, 1, 1)$ and $[P_{y_1}^{-1} f_{Y_2Y_1D_1}(\,\cdot\,, y_1, 1)]$ was shown in the previous paragraph to be uniquely determined for each y_1 by the observed joint distributions $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot, 1, 1, 1, 1, 1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\ \cdot\,, 1)$, it follows that $f_{Y_2Y_1UD_1}(\,\cdot\,,\ \cdot\,,\ \cdot\,, 1)$ too is uniquely determined by the observed joint distributions $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot,\ \cdot, 1, 1, 1, 1, 1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\ \cdot\,, 1)$. Equivalently, the operator S_{y_2} is uniquely determined for each y_2.

Step 3: Uniqueness of $F_{Y_1|Y_2UD_2D_1}(\cdot \mid \cdot, \cdot, 1, 1)$

This step is the same as Step 3 in the proof of Lemma 2, except that $L_{y,z}$ and $Q_{1/0}$ are replaced by L_{y_5,y_4,z,y_2} and $Q_{y_4,1/0,y_2}$, respectively, which were defined in Step 1 of this proof. $F_{Y_1|Y_2UD_2D_1}(\cdot \mid \cdot, \cdot, 1, 1)$ or the operator T_y is uniquely determined by the observed joint distribution $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, 1, 1, 1, 1, 1)$.

Step 4: Uniqueness of $F_{Y_2UD_2D_1}(\cdot, \cdot, 1, 1)$

This step is the same as Step 4 in the proof of Lemma 2. $F_{Y_2UD_2D_1}(\cdot, \cdot, 1, 1)$ or the auxiliary operator T'_y is uniquely determined by the pair of the observed joint distributions $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, 1, 1, 1, 1, 1)$ and $F_{Y_2Y_1D_2D_1}(\cdot, \cdot, 1, 1)$.

Step 5: Uniqueness of $F_{D_2|Y_2U}(1 \mid \cdot, \cdot)$

This step is the same as Step 5 in the proof of Lemma 2. $F_{D_2|Y_2U}(1 \mid \cdot, \cdot)$ or the auxiliary operator T'_y is uniquely determined by the pair of the observed joint distributions $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, 1, 1, 1, 1, 1)$ and $F_{Y_2Y_1D_2D_1}(\cdot, \cdot, 1, 1)$.

Step 6: Uniqueness of F_{Y_1U}

Recall from Step 2 that $f_{Y_2Y_1UD_1}(\cdot, \cdot, \cdot, 1)$ is uniquely determined by the observed joint distributions $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, 1, 1, 1, 1, 1)$ and $F_{Y_2Y_1D_1}(\cdot, \cdot, 1)$. We can write

$$\begin{aligned} f_{Y_2Y_1UD_1}(y_2, y_1, u, 1) &= f_{Y_2|Y_1UD_1}(y_2 \mid y_1, u, 1) f_{D_1|Y_1U}(1 \mid y_1, u) f_{Y_1U}(y_1, u) \\ &= f_{Y_2|Y_1U}(y_2 \mid y_1, u) f_{D_1|Y_1U}(1 \mid y_1, u) f_{Y_1U}(y_1, u) \\ &= f_{Y_6|Y_5U}(y_2 \mid y_1, u) f_{D_2|Y_2U}(1 \mid y_1, u) f_{Y_1U}(y_1, u), \end{aligned}$$

where the second equality follows from Lemma 5 (iii), and the third equality follows from Lemma 4 (i) and (ii). For a given (y_1, u), there must exist some y_2 such that $f_{Y_6|Y_5U}(y_2 \mid y_1, u) > 0$ by a property of conditional density functions. Moreover, Restriction 6 (iii) requires that $f_{D_2|Y_2U}(1 \mid y_1, u) > 0$ for a given y_1 for all u. Therefore,

for such a choice of y_2, we can write

$$f_{Y_1U}(y_1,u)=\frac{f_{Y_2Y_1UD_1}(y_2,y_1,u,1)}{f_{Y_6|Y_5U}(y_2\mid y_1,u)f_{D_2|Y_2U}(1\mid y_1,u)}$$

Recall that $f_{Y_6|Y_5U}(\,\cdot\,\mid\,\cdot\,,\,\cdot\,)$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1,1,1,1)$, $f_{Y_2Y_1UD_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,1)$ was shown in Step 2 to be uniquely determined by the pair of the observed joint distributions $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1,1,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$, and $f_{D_2|Y_2U}(1\mid\,\cdot\,,\,\cdot\,)$ was shown in Step 5 to be uniquely determined by the observed joint distributions $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1,1,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$. Therefore, it follows that the initial joint density f_{Y_1U} is uniquely determined by the observed $F_{Y_6Y_5Y_4Y_3Y_2Y_1D_5D_4D_3D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot,\,\cdot,\,\cdot,1,1,1,1,1)$ and $F_{Y_2Y_1D_1}(\,\cdot\,,\,\cdot\,,1)$. □

We next discuss an identification-preserving criterion analogously to Corollary 1. Let $\mathcal{F}$ denote the set of all the admissible model representations

$$\mathcal{F}=\{(F_{Y_t|Y_{t-1}U},F_{D_t|Y_tU},F_{Y_1U},F_{Z|U})\mid(g,h,F_{Y_1U},\zeta)\text{ satisfies Restrictions 1, 4, 5, and 6}\}.$$

A natural consequence of the main identification result of Lemma 7 is that the true model $(F^*_{Y_t|Y_{t-1}U},F^*_{D_t|Y_tU},F^*_{Y_1U},F^*_{Z|U})$ is the unique maximizer of the following criterion.

Corollary 2 (Constrained Maximum Likelihood). *If the quadruple for the true model $(F^*_{Y_t|Y_{t-1}U},F^*_{D_t|Y_tU},F^*_{Y_1U},F^*_{Z|U})$ is an element of $\mathcal{F}$, then it is the unique solution to*

$$\max_{\left(F_{Y_t|Y_{t-1}U},F_{D_t|Y_tU},F_{Y_1U},F_{Z|U}\right)\in\mathcal{F}}\ c_1E\left[\log\int f_{Y_t|Y_{t-1}U}(Y_2\mid Y_1,u)f_{D_t|Y_tU}(1\mid Y_1,u)f_{Y_1U}(Y_1,u)d\mu(u)\middle|D_1=1\right]+$$
$$c_2E\left[\log\int\prod_{s=1}^{5}f_{Y_t|Y_{t-1}U}(Y_{s+1}\mid Y_s,u)f_{D_t|Y_tU}(1\mid Y_s,u)f_{Y_1U}(Y_1,u)d\mu(u)\middle|D_5=\cdots=D_1=1\right]$$

for any $c_1,c_2>0$ subject to

$$\int f_{D_t|Y_tU}(1\mid y_1,u)f_{Y_1U}(y_1,u)d\mu(y_1,u)=f_{D_1}(1)\qquad\textit{and}$$
$$\int\prod_{s=2}^{5}f_{Y_t|Y_{t-1}U}(y_s\mid y_{s-1},u)f_{D_t|Y_tU}(1\mid y_s,u)f_{D_t|Y_tU}(1\mid y_1,u)f_{Y_1U}(y_1,u)d\mu(y_2,y_1,u)$$
$$=f_{D_5D_4D_3D_2D_1}(1,1,1,1,1).$$

10.3. Models with Higher-Order Lags. The model discussed in this paper can be extended to the following model

$$\begin{cases} Y_t = g(Y_{t-1}, \cdots, Y_{t-\tau}, U, \mathcal{E}_t) & \text{for } t = \tau+1, \cdots, T \\ D_t = h(Y_t, \cdots, Y_{t-\tau+1}, U, V_t) & \text{for } t = \tau, \cdots, T-1 \\ F_{Y_\tau \cdots Y_1 U D_{\tau-1} \cdots D_1}(\cdots, \cdot, (1)) \\ Z = \zeta(U, W) \end{cases}$$

where g is a τ-th order Markov process with heterogeneity U, and the attrition model depends on the past as well as the current state. In this set up, we can observe the parts, $F_{Y_{\tau+2} \cdots Y_1 Z D_{\tau+1} \cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1} \cdots Y_1 Z D_\tau \cdots D_1}(\cdots, \cdot, (1))$, of the joint distributions if $T = \tau + 2$. I claim that $T = \tau + 2$ suffices for identification. In other words, it can be shown that $(g, h, F_{Y_\tau \cdots Y_1 U D_{\tau-1} \cdots D_1}(\cdots, \cdot, (1)), \zeta)$ is uniquely determined by $F_{Y_{\tau+2} \cdots Y_1 Z D_{\tau+1} \cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1} \cdots Y_1 Z D_\tau \cdots D_1}(\cdots, \cdot, (1))$ up to equivalence classes. To this end, we replace Restrictions 2 and 3 by the following restrictions.

Restriction 7 (Independence).

(i) Exogeneity of $\mathcal{E}_t$: $\mathcal{E}_t \perp\!\!\!\perp (U, \{Y_s\}_{s=1}^{\tau}, \{D_s\}_{s=1}^{\tau}, \{\mathcal{E}_s\}_{s<t}, \{V_s\}_{s<t}, W)$ for all $t \geq \tau + 1$.

(ii) Exogeneity of V_t: $V_t \perp\!\!\!\perp (U, \{Y_s\}_{s=1}^{\tau-1}, \{D_s\}_{s=1}^{\tau-1}, \{\mathcal{E}_s\}_{s \leqslant t}, \{V_s\}_{s<t})$ for all $t \geq \tau$.

(iii) Exogeneity of W: $W \perp\!\!\!\perp (\{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, \{\mathcal{E}_t\}_t, \{V_t\}_t)$.

Restriction 8 (Rank Conditions). The following conditions hold for every $(y) \in \mathcal{Y}^\tau$:

(i) Heterogeneous Dynamics: the integral operator $P_{(y)} : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$ defined by $P_{(y)}\xi(y') = \int f_{Y_{\tau+2}|Y_{\tau+1} \cdots Y_2 U}(y' \mid (y), u) \cdot \xi(u) du$ is bounded and invertible.

(ii) Nondegenerate Proxy Model: $f_{Z|U}(1 \mid u)$ is bounded away from 0 and 1 for all u.

Relevant Proxy: $f_{Z|U}(1 \mid u) \neq f_{Z|U}(1 \mid u')$ whenever $u \neq u'$.

(iii) No Extinction: $f_{D_{\tau+1}|Y_{\tau+1} \cdots Y_2 U}(1 \mid (y), u) > 0$ for all $u \in \mathcal{U}$.

(iv) Initial Heterogeneity: the integral operator $S_{(y)} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$ defined by $S_{(y)}\xi(u) = \int f_{Y_{\tau+1} \cdots Y_2 Y_1 U D_\tau \cdots D_1}((y), y', u, (1)) \cdot \xi(y') dy'$ is bounded and invertible.

Lemma 8 (Independence). *The following implications hold:*

(i) Restriction 7 (i) $\Rightarrow Y_{\tau+2} \perp\!\!\!\perp (Y_1, \{D_t\}_{t=1}^{\tau+1}, Z) \mid (\{Y_t\}_{t=2}^{\tau+1}, U)$.

(ii) Restriction 7 (i) $\Rightarrow Y_{\tau+1} \perp\!\!\!\perp (\{D_t\}_{t=1}^{\tau}, Z) \mid (\{Y_t\}_{t=1}^{\tau}, U)$.

(iii) Restriction 7 (ii) $\Rightarrow D_{\tau+1} \perp\!\!\!\perp (Y_1, \{D_t\}_{t=1}^{\tau}) \mid (\{Y_t\}_{t=2}^{\tau+1}, U)$.

(iv) Restriction 7 (iii) $\Rightarrow Z \perp\!\!\!\perp (\{Y_t\}_{t=1}^{\tau+1}, \{D_t\}_{t=1}^{\tau+1}) \mid U$.

PROOF. As in the proof of Lemma 3, we use the following two properties of conditional independence:

CI.1. $A \perp\!\!\!\perp B$ implies $A \perp\!\!\!\perp B \mid \phi(B)$ for any Borel function ϕ.

CI.2. $A \perp\!\!\!\perp B \mid C$ implies $A \perp\!\!\!\perp \phi(B, C) \mid C$ for any Borel function ϕ.

(i) First, note that Restriction 2 (i) $\mathcal{E}_{\tau+2} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, \mathcal{E}_{\tau+1}, V_{\tau+1}, W)$ together with the structural definition $Z = \zeta(U, W)$ implies the independence restriction $\mathcal{E}_{\tau+2} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, \mathcal{E}_{\tau+1}, V_{\tau+1}, V_\tau, Z)$. Applying CI.1 to this independence relation $\mathcal{E}_{\tau+2} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, \mathcal{E}_{\tau+1}, V_{\tau+1}, Z)$ yields

$$\mathcal{E}_{\tau+2} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, \mathcal{E}_{\tau+1}, V_{\tau+1}, Z) \mid (g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1}), \{Y_t\}_{t=2}^{\tau}, U).$$

Since $Y_{\tau+1} = g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1})$, this conditional independence relation can be rewritten as $\mathcal{E}_{\tau+2} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, \mathcal{E}_{\tau+1}, V_{\tau+1}, Z) \mid (\{Y_t\}_{t=2}^{\tau+1}, U)$. Next, applying CI.2 to this conditional independence yields

$$\mathcal{E}_{\tau+2} \perp\!\!\!\perp (Y_1, h(Y_{\tau+1}, \cdots, Y_2, U, V_{\tau+1}), \{D_t\}_{t=1}^{\tau}, Z) \mid (\{Y_t\}_{t=2}^{\tau+1}, U).$$

Since $D_{\tau+1} = h(Y_{\tau+1}, \cdots, Y_2, U, V_{\tau+1}))$, this conditional independence can be rewritten as $\mathcal{E}_{\tau+2} \perp\!\!\!\perp (Y_1, \{D_t\}_{t=1}^{\tau+1}, Z) \mid (\{Y_t\}_{t=2}^{\tau+1}, U)$. Lastly, applying CI.2 again to this conditional independence yields

$$g(Y_{\tau+1}, \cdots, Y_2, U, \mathcal{E}_{\tau+2}) \perp\!\!\!\perp (Y_1, \{D_t\}_{t=1}^{\tau+1}, Z) \mid (\{Y_t\}_{t=2}^{\tau+1}, U).$$

Since $Y_{\tau+2} = g(Y_{\tau+1}, \cdots, Y_2, U, \mathcal{E}_{\tau+2})$, this conditional independence relation can be rewritten as $Y_{\tau+2} \perp\!\!\!\perp (Y_1, \{D_t\}_{t=1}^{\tau+1}, Z) \mid (\{Y_t\}_{t=2}^{\tau+1}, U)$.

(ii) Note that Restriction 2 (i) $\mathcal{E}_{\tau+1} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, W)$ together with the structural definition $Z = \zeta(U, W)$ implies $\mathcal{E}_{\tau+1} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, Z)$. Applying CI.1 to this independence relation yields

$$\mathcal{E}_{\tau+1} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, Z) \mid (\{Y_t\}_{t=1}^{\tau}, U).$$

Next, applying CI.2 to this conditional independence yields

$$g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1}) \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, Z) \mid (\{Y_t\}_{t=1}^{\tau}, U).$$

Since $Y_{\tau+1} = g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1})$, this conditional independence relation can be rewritten as $Y_{\tau+1} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, Z) \mid (D_1, U)$. Lastly, applying CI.2 again to this conditional independence yields $Y_{\tau+1} \perp\!\!\!\perp (\{D_t\}_{t=1}^{\tau}, Z) \mid (\{Y_t\}_{t=1}^{\tau}, U)$.

(iii) Applying CI.1 to Restriction 2 (ii) $V_{\tau+1} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau-1}, \mathcal{E}_{\tau+1}, V_\tau)$ yields

$$V_{\tau+1} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau-1}, \mathcal{E}_{\tau+1}, V_\tau) \mid (g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1}), \{Y_t\}_{t=2}^{\tau}, U).$$

Since $Y_{\tau+1} = g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1})$ by construction, it can be rewritten as $V_{\tau+1} \perp\!\!\!\perp (U, \{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau-1}, \mathcal{E}_{\tau+1}, V_\tau) \mid (\{Y_t\}_{t=2}^{\tau+1}, U)$. Next, applying CI.2 to this conditional independence yields

$$V_{\tau+1} \perp\!\!\!\perp (Y_1, h(Y_\tau, \cdots, Y_1, U, V_\tau), \{D_t\}_{t=1}^{\tau-1}) \mid (\{Y_t\}_{t=2}^{\tau+1}, U).$$

Since $D_\tau = h(Y_\tau, \cdots, Y_1, U, V_\tau)$, it can be rewritten as $V_{\tau+1} \perp\!\!\!\perp (Y_1, \{D_t\}_{t=1}^{\tau}) \mid (\{Y_t\}_{t=2}^{\tau+1}, U)$. Lastly, applying CI.2 to this conditional independence yields

$$h(Y_{\tau+1}, \cdots, Y_2, U, V_2) \perp\!\!\!\perp (Y_1, \{D_t\}_{t=1}^{\tau}) \mid (\{Y_t\}_{t=2}^{\tau+1}, U).$$

Since $D_{\tau+1} = h(Y_{\tau+1}, \cdots, Y_2, U, V_{\tau+1})$, it can be rewritten as $D_{\tau+1} \perp\!\!\!\perp (Y_1, \{D_t\}_{t=1}^{\tau}) \mid (\{Y_t\}_{t=2}^{\tau+1}, U)$.

(iv) Note that Restriction 2 (iii) $W \perp\!\!\!\perp (\{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, \mathcal{E}_{\tau+1}, V_{\tau+1})$ together with the structural definition $Z = \zeta(U, W)$ yields $Z \perp\!\!\!\perp (\{Y_t\}_{t=1}^{\tau}, \{D_t\}_{t=1}^{\tau}, \mathcal{E}_{\tau+1}, V_{\tau+1}) \mid U$. Applying CI.2 to this conditional independence relation yields

$$Z \perp\!\!\!\perp (g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1}), \{Y_t\}_{t=1}^{\tau}, h(g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1}), U, V_{\tau+1}), \{D_t\}_{t=1}^{\tau}) \mid U.$$

Since $Y_{\tau+1} = g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1})$ and $D_{\tau+1} = h(Y_{\tau+1}, U, V_{\tau+1})$, this conditional independence can be rewritten as $Z \perp\!\!\!\perp (\{Y_t\}_{t=1}^{\tau+1}, \{D_t\}_{t=1}^{\tau+1}) \mid U$. □

LEMMA 9 (Invariant Transition).

(i) Under Restrictions 1 and 7 (i), $F_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(y' \mid (y), u) = F_{Y_{\tau+1}|Y_\tau \cdots Y_1 U}(y' \mid (y), u)$ *for all* $y', (y), u$.

(ii) Under Restrictions 1 and 7 (ii), $F_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2 U}(d \mid (y), u) = F_{D_1|Y_\tau \cdots Y_1 U}(d \mid (y), u)$ *for all* $d, (y), u$.

PROOF. (i) First, note that Restriction 7 (i) implies $\mathcal{E}_{\tau+2} \perp\!\!\!\perp (U, Y_\tau, \cdots, Y_1, \mathcal{E}_{\tau+1})$, which in turn implies that $\mathcal{E}_{\tau+2} \perp\!\!\!\perp (g(Y_\tau, \cdots, Y_1, U, \mathcal{E}_{\tau+1}), Y_\tau, \cdots, Y_2, U)$, hence

$\mathcal{E}_{\tau+2} \perp\!\!\!\perp (Y_{\tau+1}, \cdots, Y_2, U)$. Second, Restriction 7 (i) in particular yields $\mathcal{E}_{\tau+1} \perp\!\!\!\perp (Y_\tau, \cdots, Y_1, U)$. Using these two independence results, we obtain

$$\begin{aligned}
F_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(y' \mid (y), u) &= \Pr[g((y), u, \mathcal{E}_{\tau+2}) \leq y' \mid (Y_{\tau+1}, \cdots, Y_2) = (y), U = u] \\
&= \Pr[g((y), u, \mathcal{E}_{\tau+2}) \leq y'] \\
&= \Pr[g((y), u, \mathcal{E}_{\tau+1}) \leq y'] \\
&= \Pr[g((y), u, \mathcal{E}_{\tau+1}) \leq y' \mid (Y_\tau, \cdots, Y_1) = (y), U = u] \\
&= F_{Y_{\tau+1}|Y_\tau \cdots Y_1 U}(y' \mid (y), u)
\end{aligned}$$

for all $y', (y), u$, where the second equality follows from $\mathcal{E}_{\tau+2} \perp\!\!\!\perp (Y_{\tau+1}, \cdots, Y_2, U)$, the third equality follows from identical distribution of $\mathcal{E}_t$ by Restriction 1, and the forth equality follows from $\mathcal{E}_{\tau+1} \perp\!\!\!\perp (Y_\tau, \cdots, Y_1, U)$.

(ii) Restriction 7 (ii) implies that $V_{\tau+1} \perp\!\!\!\perp (g(Y_{\tau+1}, \cdots, Y_1, U, \mathcal{E}_{\tau+1}), Y_\tau, \cdots, Y_1, U)$, hence $V_{\tau+1} \perp\!\!\!\perp (Y_{\tau+1}, \cdots, Y_2, U)$. Restriction 7 (ii) also implies $V_\tau \perp\!\!\!\perp (Y_\tau, \cdots, Y_1, U)$. Using these two independence results, we obtain

$$\begin{aligned}
F_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2 U}(d \mid (y), u) &= \Pr[h((y), u, V_{\tau+1}) \leq d \mid (Y_{\tau+1}, \cdots, Y_2) = (y), U = u] \\
&= \Pr[h((y), u, V_{\tau+1}) \leq d] \\
&= \Pr[h((y), u, V_\tau) \leq d] \\
&= \Pr[h((y), u, V_\tau) \leq d \mid (Y_\tau, \cdots, Y_1) = (y), U = u] \\
&= F_{D_1|Y_\tau \cdots Y_1 U}(d \mid (y), u)
\end{aligned}$$

for all $d, (y), u$, where the second equality follows from $V_{\tau+1} \perp\!\!\!\perp (Y_{\tau+1}, \cdots, Y_2, U)$, the third equality follows from identical distribution of V_t from Restriction 1, and the forth equality follows from $V_\tau \perp\!\!\!\perp (Y_\tau, \cdots, Y_1, U)$. □

LEMMA 10 (Identification). *Under Restrictions 1, 4, 7, and 8, the quadruple* $(F_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}, F_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2 U}, F_{Y_\tau \cdots Y_1 U D_{\tau-1}\cdots D_1}(\cdots, \cdot, (1)), F_{Z|U})$ *is uniquely determined by* $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$ *and* $F_{Y_{\tau+1}\cdots Y_1 Z D_\tau \cdots D_1}(\cdots, \cdot, (1))$.

PROOF. Given fixed (y) and z, define the operators $L_{(y),z} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_{Y_t})$, $P_{(y)} : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$, $Q_z : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$, $R_{(y)} : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$, $S_{(y)} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$, $T_{(y)} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_U)$, and $T'_{(y)} : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$ by

$$\begin{aligned}
(L_{(y),z}\xi)(y_{\tau+2}) &= \int f_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(y_{\tau+2},(y),y_1,z,(1)) \cdot \xi(y_1)dy_1, \\
(P_{(y)}\xi)(y_{\tau+2}) &= \int f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(y_{\tau+2} \mid (y),u) \cdot \xi(u)du, \\
(Q_z\xi)(u) &= f_{Z|U}(z \mid u) \cdot \xi(u), \\
(R_{(y)}\xi)(u) &= f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2 U}(1 \mid (y),u) \cdot \xi(u), \\
(S_{(y)}\xi)(u) &= \int f_{Y_{\tau+1}\cdots Y_2 Y_1 U D_\tau \cdots D_1}((y),y_1,u,(1)) \cdot \xi(y_1)dy_1, \\
(T_{(y)}\xi)(u) &= \int f_{Y_1|Y_{\tau+1}\cdots Y_2 U D_{\tau+1} D_1}(y_1 \mid (y),u,(1)) \cdot \xi(y_1)dy_1, \\
(T'_{(y)}\xi)(u) &= f_{Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}((y),u,(1)) \cdot \xi(u)
\end{aligned}$$

respectively. The operators $L_{(y),z}$, $P_{(y)}$, $S_{(y)}$, and $T_{(y)}$ are integral operators whereas Q_z, $R_{(y)}$, and $T'_{(y)}$ are multiplication operators. Note that $L_{(y),z}$ is identified from observed joint distribution $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$.

Step 1: Uniqueness of $F_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}$ and $F_{Z|U}$

The kernel $f_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdot, (y), \cdot, z, (1))$ of the integral operator $L_{(y),z}$ can be rewritten as

$$\begin{aligned}
f_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(y_{\tau+2},(y),y_1,z,(1)) = \int & f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_1 Z U D_{\tau+1}\cdots D_1}(y_{\tau+2} \mid (y),y_1,z,u,(1)) \\
& \times f_{Z|Y_{\tau+1}\cdots Y_1 U D_{\tau+1}\cdots D_1}(z \mid (y),y_1,u,(1)) \\
& \times f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}(1 \mid (y),y_1,u,(1)) \\
& \times f_{Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}((y),y_1,u,(1))\, du
\end{aligned} \tag{24}$$

But by Lemma 8 (i), (iv), and (iii), respectively, Restriction 7 implies that

$$\begin{aligned}
f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_1 Z U D_{\tau+1}\cdots D_1}(y_{\tau+2} \mid (y),y_1,z,u,(1)) &= f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(y_{\tau+2} \mid (y),u), \\
f_{Z|Y_{\tau+1}\cdots Y_1 U D_{\tau+1}\cdots D_1}(z \mid (y),y_1,u,(1)) &= f_{Z|U}(z \mid u), \\
f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}(1 \mid (y),y_1,u,(1)) &= f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2 U}(1 \mid (y),u).
\end{aligned}$$

Equation (24) thus can be rewritten as

$$\begin{aligned} f_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(y_{\tau+2},(y),y_1,z,(1)) \;&=\; \int f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(y_{\tau+2}\mid (y),u)\cdot f_{Z|U}(z\mid u) \\ &\quad \times f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2 U}(1\mid (y),u) \\ &\quad \times f_{Y_{\tau+1}\cdots Y_1 U D_\tau\cdots D_1}((y),y_1,u,(1))\; du \end{aligned}$$

But this implies that the integral operator $L_{y,z}$ is written as the operator composition

$$L_{(y),z} = P_{(y)} Q_z R_{(y)} S_{(y)}.$$

Restriction 8 (i), (ii), (iii), and (iv) imply that the operators $P_{(y)}$, Q_z, $R_{(y)}$, and $S_{(y)}$ are invertible, respectively. Hence so is $L_{(y),z}$. Using the two values $\{0,1\}$ of Z, form the product

$$L_{(y),1} L_{y,0}^{-1} = P_{(y)} Q_{1/0} P_{(y)}^{-1}$$

where $Q_{z/z'} := Q_z Q_{z'}^{-1}$. By Restriction 8 (ii), the operator $L_{(y),1} L_{(y),0}^{-1}$ is bounded. The expression $L_{(y),1} L_{(y),0}^{-1} = P_{(y)} Q_{1/0} P_{(y)}^{-1}$ thus allows unique eigenvalue-eigenfunction decomposition.

The distinct proxy odds as in Restriction 8 (ii) guarantee distinct eigenvalues and single dimensionality of the eigenspace associated with each eigenvalue. Within each of the single-dimensional eigenspace is a unique eigenfunction pinned down by $\mathcal{L}^1$-normalization because of the unity of integrated densities. The eigenvalues $\lambda(u)$ yield the multiplier of the operator $Q_{1/0}$, hence $\lambda(u) = f_{Z|U}(1 \mid u)/f_{Z|U}(0 \mid u)$. This proxy odds in turn identifies $f_{Z|U}(\;\cdot\;\mid u)$ since Z is binary. The corresponding normalized eigenfunctions are the kernels of the integral operator $P_{(y)}$, hence $f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(\;\cdot\;\mid (y),u)$. Lastly, Restriction 4 facilitates unique ordering of the eigenfunctions $f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(\;\cdot\;\mid (y),u)$ by the distinct concrete values of $u = \lambda(u)$. This is feasible because the eigenvalues $\lambda(u) = f_{Z|U}(1 \mid u)/f_{Z|U}(0 \mid u)$ are invariant from (y). That is, eigenfunctions $f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(\;\cdot\;\mid (y),u)$ of the operator $L_{(y),1} L_{(y),0}^{-1}$ across different (y) can be uniquely ordered in u invariantly from (y) by the common set of ordered distinct eigenvalues $u = \lambda(u)$.

Therefore, $F_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2U}$ and $F_{Z|U}$ are uniquely determined by the observed joint distribution $F_{Y_{\tau+2}\cdots Y_1ZD_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$. Equivalently, the operators $P_{(y)}$ and Q_z are uniquely determined for each (y) and z, respectively.

Step 2: Uniqueness of $F_{Y_{\tau+1}\cdots Y_1UD_\tau\cdots D_1}(\cdots, \cdot, (1))$

By Lemma 8 (ii), Restriction 7 implies $f_{Y_{\tau+1}|Y_\tau\cdots Y_1UD_\tau\cdots D_1}(y' \mid (y), u, (1)) = f_{Y_{\tau+1}|Y_\tau\cdots Y_1U}(y' \mid (y), u)$. Using this equality, write the density of the observed joint distribution $F_{Y_{\tau+1}\cdots Y_1D_\tau\cdots D_1}(\cdots, (1))$ as

$$\begin{aligned} f_{Y_{\tau+1}\cdots Y_1D_\tau\cdots D_1}(y', (y), (1)) &= \int f_{Y_{\tau+1}|Y_\tau\cdots Y_1UD_\tau\cdots D_1}(y' \mid (y), u, (1)) \\ &\quad \times f_{Y_\tau\cdots Y_1UD_\tau\cdots D_1}((y), u, (1))du \\ &= \int f_{Y_{\tau+1}|Y_\tau\cdots Y_1U}(y' \mid (y), u) \\ &\quad \times f_{Y_\tau\cdots Y_1UD_\tau\cdots D_1}((y), u, (1))du \end{aligned} \tag{25}$$

By Lemma 9 (i), $F_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2U}(y' \mid (y), u) = F_{Y_{\tau+1}|Y_\tau\cdots Y_1U}(y' \mid (y), u)$ for all $y', (y), u$. Therefore, we can write the operator $P_{(y)}$ as

$$(P_{(y)}\xi)(y') = \int f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2U}(y' \mid (y), u) \cdot \xi(u)du = \int f_{Y_{\tau+1}|Y_\tau\cdots Y_1U}(y' \mid (y), u) \cdot \xi(u)du.$$

With this operator notation, it follows from (25) that

$$f_{Y_{\tau+1}\cdots Y_1D_\tau\cdots D_1}(\cdot, (y), (1)) = P_{(y)}f_{Y_\tau\cdots Y_1UD_\tau\cdots D_1}((y), \cdot, (1)).$$

By Restriction 8 (i) and (ii), this operator equation can be solved for the function $f_{Y_\tau\cdots Y_1UD_\tau\cdots D_1}((y), \cdot, (1))$ as

$$f_{Y_\tau\cdots Y_1UD_\tau\cdots D_1}((y), \cdot, (1)) = P_{(y)}^{-1}f_{Y_{\tau+1}\cdots Y_1D_\tau\cdots D_1}(\cdot, (y), (1)) \tag{26}$$

Recall that $P_{(y)}$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_{\tau+2}\cdots Y_1ZD_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$. The function $f_{Y_{\tau+1}\cdots Y_1D_\tau\cdots D_1}(\cdot, (y), (1))$ is also uniquely determined by the observed joint distribution $f_{Y_{\tau+1}\cdots Y_1D_\tau\cdots D_1}(\cdots, (1))$. Therefore, (25) shows that $f_{Y_\tau\cdots Y_1UD_\tau\cdots D_1}(\cdots, \cdot, (1))$ is uniquely determined by the observed joint distributions $F_{Y_{\tau+2}\cdots Y_1ZD_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1}\cdots Y_1D_\tau\cdots D_1}(\cdots, (1))$.

Using the solution to the above inverse problem, we can write the kernel of the operator $S_{(y)}$ as

$$\begin{aligned} f_{Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}(y',(y),u,(1)) &= f_{Y_{\tau+1}|Y_\tau \cdots Y_1 U D_\tau \cdots D_1}(y' \mid (y),u,(1)) \cdot f_{Y_\tau \cdots Y_1 U D_\tau \cdots D_1}((y),u,(1)) \\ &= f_{Y_{\tau+1}|Y_\tau \cdots Y_1 U}(y' \mid (y),u) \cdot f_{Y_\tau \cdots Y_1 U D_\tau \cdots D_1}((y),u,(1)) \\ &= f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(y' \mid (y),u) \cdot f_{Y_\tau \cdots Y_1 U D_\tau \cdots D_1}((y),u,(1)) \\ &= f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(y' \mid (y),u) \\ &\quad \times [P_{(y)}^{-1} f_{Y_{\tau+1}\cdots Y_1 Z D_\tau \cdots D_1}(\,\cdot\,,(y),z,(1))](u) \end{aligned}$$

where the second equality follows from Lemma 8 (ii), the third equality follows from Lemma 9 (i), and the forth equality follows from (26). Since $f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots,\,\cdot\,,(1))$ and $[P_{(y)}^{-1} f_{Y_{\tau+1}\cdots Y_1 Z D_\tau \cdots D_1}(\,\cdot\,,(y),z,(1))]$ was shown in the previous paragraph to be uniquely determined for each y by the observed joint distributions $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots,\,\cdot\,,(1))$ and $f_{Y_{\tau+1}\cdots Y_1 D_\tau \cdots D_1}(\cdots,(1))$, it follows that $f_{Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}(\cdots,\,\cdot\,,(1))$ too is uniquely determined by the observed joint distributions $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots,\,\cdot\,,(1))$ and $f_{Y_{\tau+1}\cdots Y_1 D_\tau \cdots D_1}(\cdots,(1))$. Equivalently, the operator $S_{(y)}$ is uniquely determined for each (y).

Step 3: Uniqueness of $F_{Y_1|Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}(\,\cdot\mid\cdots,\,\cdot\,,(1))$

First, note that the kernel of the composite operator $T'_{(y)}T_{(y)}$ can be written as

$$\begin{aligned} & f_{Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}((y),u,(1)) \cdot f_{Y_1|Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}(y_1 \mid (y),u,(1)) \\ = \; & f_{Y_{\tau+1}\cdots Y_1 U D_{\tau+1}\cdots D_1}((y),y_1,u,(1)) \\ = \; & f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}(1 \mid (y),y_1,u,(1)) \cdot f_{Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}((y),y_1,u,(1)) \\ = \; & f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2 U}(1 \mid (y),u) \cdot f_{Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}((y),y_1,u,(1)) \qquad (27) \end{aligned}$$

where the last equality is due to Lemma 8 (iii). But the last expression corresponds to the kernel of the composite operator $R_{(y)}S_{(y)}$, thus showing that $T'_{(y)}T_{(y)} = R_{(y)}S_{(y)}$. But then, $L_{(y),z} = P_{(y)}Q_z R_{(y)}S_{(y)} = P_{(y)}Q_z T'_{(y)}T_{(y)}$. Note that the invertibility of $R_{(y)}$ and $S_{(y)}$ as required by Assumption 8 implies invertibility of $T'_{(y)}$ and $T_{(y)}$ as

well, for otherwise the equivalent composite operator $T'_{(y)}T_{(y)} = R_{(y)}S_{(y)}$ would have a nontrivial nullspace.

Using Restriction 8, form the product of operators as

$$L^{-1}_{(y),0}L_{(y),1} = T^{-1}_{(y)}Q_{1/0}T_{(y)}$$

The disappearance of $T'_{(y)}$ is due to commutativity of multiplication operators. By the same logic as in Step 1, this expression together with Restriction 8 (ii) admits unique left eigenvalue-eigenfunction decomposition. Moreover, the point spectrum is exactly the same as the one in Step 1, as is the middle multiplication operator $Q_{1/0}$. This equivalence of the spectrum allows consistent ordering of U with that of Step 1. Left eigenfunctions yield the kernel of $T_{(y)}$ pinned down by the normalization of unit integral. This shows that the operator $T_{(y)}$ is uniquely determined by the observed joint distribution $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$.

Step 4: Uniqueness of $F_{Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$

Equation (27) implies that

$$\int f_{Y_1|Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}(y_1 \mid (y), u, (1)) \cdot f_{Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}((y), u, (1))du$$
$$= f_{Y_{\tau+1}\cdots Y_1 D_{\tau+1}\cdots D_1}((y), y_1, (1))$$

hence yielding the linear operator equation

$$T^*_{(y)} f_{Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}((y), \cdot, (1)) = f_{Y_{\tau+1}\cdots Y_1 D_{\tau+1}\cdots D_1}((y), \cdot, (1))$$

where $T^*_{(y)}$ denotes the adjoint operator of $T_{(y)}$. Since $T_{(y)}$ is invertible, so is its adjoint operator $T^*_{(y)}$. But then, the multiplier of the multiplication operator $T'_{(y)}$ can be given by the unique solution to the above linear operator equation, i.e.,

$$f_{Y_{\tau+1}\cdots Y_2 U D_{\tau+1}\cdots D_1}((y), \cdot, (1)) = (T^*_{(y)})^{-1} f_{Y_{\tau+1}\cdots Y_1 D_{\tau+1}\cdots D_1}((y), \cdot, (1))$$

$T_{(y)}$ hence $T^*_{(y)}$ was shown to be uniquely determined by $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$ in Step 3, and $f_{Y_{\tau+1}\cdots Y_1 D_{\tau+1}\cdots D_1}(\cdots, (1))$ is also available from observed data. Therefore, the operator $T'_{(y)}$ is uniquely determined by $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$.

Step 5: Uniqueness of $F_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2U}(1 \mid \cdots, \cdot)$

First, the definition of the operators $R_{(y)}$, $S_{(y)}$, $T_{(y)}$, and $T'_{(y)}$ and Lemma 8 (iii) yield the operator equality $R_{(y)}S_{(y)} = T'_{(y)}T_{(y)}$, where $T_{(y)}$ and $T'_{(y)}$ have been shown to be uniquely determined by the observed joint distribution $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$ in Steps 3 and 4, respectively. Recall that $S_{(y)}$ was also shown in Step 2 to be uniquely determined by the observed joint distributions $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1}\cdots Y_1 D_\tau \cdots D_1}(\cdots, (1))$. Restriction 8 (iv) guarantees invertibility of $S_{(y)}$. It follows that the operator inversion $R_{(y)} = (R_{(y)}S_{(y)})S_{(y)}^{-1} = (T'_{(y)}T_{(y)})S_{(y)}^{-1}$ yields the operator $R_{(y)}$, in turn showing that its multiplier $f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2U}(1 \mid (y), \cdot)$ is uniquely determined for each (y) by the observed joint distributions $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1}\cdots Y_1 D_\tau \cdots D_1}(\cdots, (1))$.

Step 6: Uniqueness of $F_{Y_\tau \cdots Y_1 U D_{\tau-1}\cdots D_1}(\cdots, \cdot, (1))$

Recall from Step 2 that $f_{Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}(\cdots, \cdot, (1))$ is uniquely determined by the observed joint distributions $F_{Y_{\tau+2}\cdots Y_1 Z D_{\tau+1}\cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1}\cdots Y_1 D_\tau \cdots D_1}(\cdots, (1))$. We can write

$$
\begin{aligned}
f_{Y_{\tau+1}\cdots Y_1 U D_\tau \cdots D_1}(y', (y), u, (1)) &= f_{Y_{\tau+1}|Y_\tau \cdots Y_1 U D_\tau \cdots D_1}(y' \mid (y), u, (1)) \\
&\quad \times f_{D_\tau|Y_\tau \cdots Y_1 U D_{\tau-1}\cdots D_1}(1 \mid (y), u, (1)) \\
&\quad \times f_{Y_\tau \cdots Y_1 U D_{\tau-1}\cdots D_1}((y), u, (1)) \\
&= f_{Y_{\tau+1}|Y_\tau \cdots Y_1 U}(y' \mid (y), u) \cdot f_{D_\tau|Y_\tau \cdots Y_1 U}(1 \mid (y), u) \\
&\quad \times f_{Y_\tau \cdots Y_1 U D_{\tau-1}\cdots D_1}((y), u, (1)) \\
&= f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2 U}(y' \mid (y), u) \cdot f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2 U}(1 \mid (y), u) \\
&\quad \times f_{Y_\tau \cdots Y_1 U D_{\tau-1}\cdots D_1}((y), u, (1)),
\end{aligned}
$$

where the second equality follows from Lemma 8 (ii), and the third equality follows from Lemma 9 (i) and (ii). For a given $((y), u)$, there must exist some y' such that $f_{Y_{\tau+2}|Y_{\tau+1}\cdots Y_2U}(y' \mid (y), u) > 0$ by a property of conditional density functions. Moreover, Restriction 8 (iii) requires that $f_{D_{\tau+1}|Y_{\tau+1}\cdots Y_2U}(1 \mid (y), u) > 0$ for a given

(y) for all u. Therefore, for such a choice of y', we can write

$$f_{Y_\tau \cdots Y_1 U D_{\tau-1} \cdots D_1}((y), u, (1)) = \frac{f_{Y_{\tau+1} \cdots Y_1 U D_\tau \cdots D_1}(y', (y), u, (1))}{f_{Y_{\tau+2}|Y_{\tau+1} \cdots Y_2 U}(y' \mid (y), u) \cdot f_{D_{\tau+1}|Y_{\tau+1} \cdots Y_2 U}(1 \mid (y), u)}$$

$f_{Y_{\tau+2}|Y_{\tau+1} \cdots Y_2 U}(y' \mid (y), u)$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_{\tau+2} \cdots Y_1 Z D_{\tau+1} \cdots D_1}(\cdots, \cdot, (1))$, $f_{Y_{\tau+1} \cdots Y_1 U D_\tau \cdots D_1}(y', (y), u, (1))$ was shown in Step 2 to be uniquely determined by the observed joint distributions $F_{Y_{\tau+2} \cdots Y_1 Z D_{\tau+1} \cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1} \cdots Y_1 D_\tau \cdots D_1}(\cdots, (1))$, and $f_{D_{\tau+1}|Y_{\tau+1} \cdots Y_2 U}(1 \mid (y), u)$ was shown in Step 5 to be uniquely determined by the observed joint distributions $F_{Y_{\tau+2} \cdots Y_1 Z D_{\tau+1} \cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1} \cdots Y_1 D_\tau \cdots D_1}(\cdots, (1))$. Therefore, it follows that the joint density $f_{Y_\tau \cdots Y_1 U D_{\tau-1} \cdots D_1}(\cdots, \cdot, (1))$ is uniquely determined by the observed joint distributions $F_{Y_{\tau+2} \cdots Y_1 Z D_{\tau+1} \cdots D_1}(\cdots, \cdot, (1))$ and $f_{Y_{\tau+1} \cdots Y_1 D_\tau \cdots D_1}(\cdots, (1))$. □

10.4. Models with Time-Specific Effects. The baseline model (1) that we considered in this paper assumes that the dynamic model g is time-invariant. It is often more realistic to allow this model to have time-specific effects. Consider the following variant of the model (1).

$$\begin{cases} Y_t = g_t(Y_{t-1}, U, \mathcal{E}_t) & t = 2, \cdots, T \qquad \text{(State Dynamics)} \\ D_t = h(Y_t, U, V_t) & t = 1, \cdots, T-1 \qquad \text{(Hazard Model)} \\ F_{Y_1 U} & \text{(Initial joint distribution of } (Y_1, U)) \\ Z = \zeta(U, W) & \text{(Optional: nonclassical proxy of } U) \end{cases}$$

The differences from (1) are the t subscripts under g.

The objective is to identify the model $(\{g_t\}_{t=2}^T, h, F_{Y_1 U}, \zeta)$. The main obstacle is that the invariant transition of Lemma 4 (i) is no longer useful. As a result, Steps 2 and 6 in the proof of Lemma 2 break down. In order to remedy this hole, we need to observe data of an additional time period prior to the start of the data, i.e., $t = 0$. For brevity, we show this result for the case of $T = 3$.

LEMMA 11 (Identification). *Suppose that Restrictions 1, 2, 3, and 4 hold conditionally on* $Pr(D_0 = 1)$. *Then the model* $(\{F_{Y_t|Y_{t-1}U}\}_{t=2}^3, F_{D_t|Y_t U}, F_{Y_1 U|D_0=1}, F_{Z|U})$

is uniquely determined by the observed joint distributions $F_{Y_1Y_0ZD_0}(\cdot, \cdot, \cdot, 1)$, $F_{Y_2Y_1Y_0ZD_1D_0}(\cdot, \cdot, \cdot, \cdot, 1, 1)$, *and* $F_{Y_3Y_2Y_1Y_0ZD_2D_1D_0}(\cdot, \cdot, \cdot, \cdot, \cdot, 1, 1, 1)$.

PROOF. Many parts of the proof Lemma 2 remains available. However, under the current model with time-specific transition, the operator P_y is time-specific. Therefore, we use two operators $P_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$ and $P'_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$ for each y defined as

$$\begin{aligned}(P_y\xi)(y_2) &= \int f_{Y_2|Y_1U}(y_2 \mid y, u) \cdot \xi(u)du, \\ (P'_y\xi)(y_3) &= \int f_{Y_3|Y_2U}(y_3 \mid y, u) \cdot \xi(u)du,\end{aligned}$$

Accordingly, we employ the two observable operators $L_{y,z} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_{Y_t})$ and $L'_{y,z} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_{Y_t})$ for each (y, z) defined as

$$\begin{aligned}(L_{y,z}\xi)(y_2) &= \int f_{Y_2Y_1Y_0ZD_1D_0}(y_2, y, y_0, z, 1, 1) \cdot \xi(y_0)dy_0, \\ (L'_{y,z}\xi)(y_3) &= \int f_{Y_3Y_2Y_1ZD_2D_1D_0}(y_3, y, y_1, z, 1, 1, 1) \cdot \xi(y_1)dy_1.\end{aligned}$$

All the other operators directly carry over from the proof of Lemma 2 as:

$$\begin{aligned}(Q_z\xi)(u) &= f_{Z|U}(z \mid u) \cdot \xi(u), \\ (R_y\xi)(u) &= f_{D_2|Y_2U}(1 \mid y, u) \cdot \xi(u), \\ (S_y\xi)(u) &= \int f_{Y_2Y_1UD_1D_0}(y, y_1, u, 1, 1) \cdot \xi(y_1)dy_1, \\ (T_y\xi)(u) &= \int f_{Y_1|Y_2UD_2D_1D_0}(y_1 \mid y, u, 1, 1, 1) \cdot \xi(y_1)dy_1, \\ (T'_y\xi)(u) &= f_{Y_2UD_2D_1D_0}(y, u, 1, 1, 1) \cdot \xi(u)\end{aligned}$$

except that the additional argument $D_0 = 1$ is attached to the kernels of S_y and T_y and the multiplier of T'_y.

The first task is to identify the kernels of these two integral operators. Following Step 1 of the proof of Lemma 2 by using the observed operator $L'_{y,z}$ shows that P'_y and Q_z are identified. Equivalently, $F_{Y_3|Y_2U}$ and $F_{Z|U}$ are identified. Similarly, following

Step 1 by using the observed operator $L_{y,z}$ shows that P_y and Q_z are identified. Equivalently, $F_{Y_2|Y_1U}$ is identified as well.

Next, follow Step 2 of the proof of Lemma 2, except that we use our current definition of P_y instead of P'_y. It follows that

$$f_{Y_2Y_1UD_1D_0}(y', y, u, 1, 1) \;=\; f_{Y_2|Y_1U}(y' \mid y, u) \cdot [P_y^{-1} f_{Y_2Y_1D_1D_0}(\,\cdot\,, y, 1, 1)](u)$$

where $f_{Y_2|Y_1U}$ was identified as the kernel of P_y in the previous step, P_y was identified in the previous step, and $f_{Y_2Y_1D_1D_0}(\,\cdot,\ \cdot, 1, 1)$ is observable from data. This shows that the operator S_y is identified for each y.

Steps 3–5 analogously follow from the proof of Lemma 2 except that the current definitions of $L_{y,z}$, R_y, S_y, T_y, and T'_y are used. These steps show that R_y in particular are identified for each y.

Lastly, extending the argument of Step 6 in the proof of Lemma 2 yields

$$f_{Y_1U|D_0}(y, u \mid 1) = \frac{f_{Y_2Y_1UD_1D_0}(y', y, u, 1, 1)}{f_{Y_2|Y_1U}(y' \mid y, u) f_{D_2|Y_2U}(1 \mid y, u) f_{D_0}(1)}$$

where $f_{Y_2Y_1UD_1D_0}(\,\cdot,\ \cdot,\ \cdot, 1, 1)$ was identified in the second step, $f_{Y_2|Y_1U}$ was identified in the first step, $f_{D_2|Y_2U}$ was identified in the previous step, and $f_{D_0}(1)$ is observable from data. It follows that $F_{Y_1U|D_0=1}$ is identified. □

10.5. Censoring by Contemporaneous D_t instead of Lagged D_t. For the main identification result discussed, we assumed that lagged selection indicator D_t induces censored observation of Y_t as follows:

$$\begin{aligned} &\text{observe } Y_1, \\ &\text{observe } Y_2 \quad \text{if } D_1 = 1, \\ &\text{observe } Y_3 \quad \text{if } D_1 = D_2 = 1. \end{aligned}$$

In many application, contemporaneous D_t instead of lagged D_t may induce censored observation of Y_t as follows:

$$\begin{aligned} &\text{observe } Y_1, \quad \text{if } D_1 = 1 \\ &\text{observe } Y_2 \quad \text{if } D_1 = D_2 = 1, \\ &\text{observe } Y_3 \quad \text{if } D_1 = D_2 = D_3 = 1. \end{aligned}$$

where the model follows a slight modification of (1):

$$\begin{cases} Y_t = g(Y_{t-1}, U, \mathcal{E}_t) & t = 2, \cdots, T \qquad \text{(State Dynamics)} \\ D_t = h(Y_t, U, V_t) & t = 1, \cdots, T \qquad \text{(Hazard Model)} \\ F_{Y_1 U} & \text{(Initial joint distribution of } (Y_1, U)) \\ Z = \zeta(U, W) & \text{(Optional: nonclassical proxy of } U) \end{cases}$$

(The difference from the baseline model (1) is that the hazard model is defined for all $t = 1, \cdots, T$.) In this model, the problem of identification is to show the well-definition of

$$(F_{Y_2 Y_1 Z D_2 D_1}(\cdot,\ \cdot,\ \cdot, 1, 1),\ F_{Y_3 Y_2 Y_1 Z D_3 D_2 D_1}(\cdot,\ \cdot,\ \cdot,\ \cdot, 1, 1, 1)) \ \mapsto\ (g, h, F_{Y_1 U | D_1 = 1}, \zeta).$$

First, consider the following auxiliary lemma, which can be proved similarly to Lemma 3.

Lemma 12 (Independence). *The following implications hold:*
(i) Restriction 2 (i) $\Rightarrow Y_3 \perp\!\!\!\perp (Y_1, D_1, D_2, D_3, Z) \mid (Y_2, U)$.
(ii) Restriction 2 (i) $\Rightarrow Y_2 \perp\!\!\!\perp (D_1, D_2, Z) \mid (Y_1, U)$.
(iii) Restriction 2 (ii) $\Rightarrow D_3 \perp\!\!\!\perp Y_2 \mid (Y_3, U)$.
(iv) Restriction 2 (iii) $\Rightarrow Z \perp\!\!\!\perp (Y_2, Y_1, D_3, D_2, D_1) \mid U$.

Some of the rank conditions of Restriction 3 are replaced as follows.

Restriction 9 (Rank Conditions). The following conditions hold for every $y \in \mathcal{Y}$:
(i) Heterogeneous Dynamics: the integral operator $P_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$ defined by $P_y \xi(y') = \int f_{Y_3 | Y_2 U}(y' \mid y, u) \cdot \xi(u) du$ is bounded and invertible.
(ii) Nondegenerate Proxy Model: there exists $\delta > 0$ such that $\delta \leqslant f_{Z|U}(1 \mid u) \leqslant 1 - \delta$ for all u.

Relevant Proxy: $f_{Z|U}(1 \mid u) \neq f_{Z|U}(1 \mid u')$ whenever $u \neq u'$.
(iii) No Extinction: $f_{D_2 | Y_2 U}(1 \mid y, u) > 0$ for all $u \in \mathcal{U}$.
(iv) Initial Heterogeneity: the two integral operators $\tilde{L}_y : \mathcal{L}^2(Y_t) \to \mathcal{L}^2(U)$, and

$S_y : \mathcal{L}^2(U) \to \mathcal{L}^2(Y_t)$ respectively defined by $\tilde{L}_y\xi(u) = \int f_{Y_2Y_1UD_3D_2D_1}(y, y_1, u, 1, 1, 1) \cdot \xi(y_1)dy_1$ and $S_y\xi(y_1) = \int f_{Y_2Y_1UD_2D_1}(y, y_1, u, 1, 1) \cdot \xi(u)du$ are bounded and invertible.

LEMMA 13 (Identification). *Under Restrictions 1, 2, 4, and 9, the quadruple* $(F_{Y_3|Y_2U}, F_{D_3|Y_3U}, F_{Y_1U|D_1=1}, F_{Z|U})$ *is uniquely determined by* $F_{Y_2Y_1ZD_2D_1}(\cdot, \cdot, \cdot, 1, 1)$ *and* $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1, 1)$.

PROOF. Given fixed y and z, define the operators $L_{y,z} : \mathcal{L}^2(F_{Y_t}) \to \mathcal{L}^2(F_{Y_t})$, $P_y : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_{Y_t})$, $Q_z : \mathcal{L}^2(F_U) \to \mathcal{L}^2(F_U)$, $\tilde{L}_y : \mathcal{L}^2(Y_t) \to \mathcal{L}^2(U)$, and $S_y : \mathcal{L}^2(U) \to \mathcal{L}^2(Y_t)$ by

$$\begin{aligned}
(L_{y,z}\xi)(y_3) &= \int f_{Y_3Y_2Y_1ZD_3D_2D_1}(y_3, y, y_1, z, 1, 1, 1) \cdot \xi(y_1)dy_1, \\
(P_y\xi)(y_3) &= \int f_{Y_3|Y_2U}(y_3 \mid y, u) \cdot \xi(u)du, \\
(Q_z\xi)(u) &= f_{Z|U}(z \mid u) \cdot \xi(u), \\
(\tilde{L}_y\xi)(u) &= \int f_{Y_2Y_1UD_3D_2D_1}(y, y_1, u, 1, 1, 1) \cdot \xi(y_1)dy_1, \\
(S_y\xi)(y_1) &= \int f_{Y_2Y_1UD_2D_1}(y, y_1, u, 1, 1) \cdot \xi(u)du
\end{aligned}$$

respectively. Similarly to the proof of Lemma 2, the operator $L_{y,z}$ is identified from observed joint distribution $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1, 1)$.

Step 1: Uniqueness of $F_{Y_3|Y_2U}$ and $F_{Z|U}$

The kernel $f_{Y_3Y_2Y_1ZD_3D_2D_1}(\cdot, y, \cdot, z, 1, 1, 1)$ of the integral operator $L_{y,z}$ can be rewritten as

$$\begin{aligned}
f_{Y_3Y_2Y_1ZD_3D_2D_1}(y_3, y, y_1, z, 1, 1, 1) &= \int f_{Y_3|Y_2Y_1ZUD_3D_2D_1}(y_3 \mid y, y_1, z, u, 1, 1, 1) \\
&\quad \times f_{Z|Y_2Y_1UD_3D_2D_1}(z \mid y, y_1, u, 1, 1, 1) \\
&\quad \times f_{Y_2Y_1UD_3D_2D_1}(y, y_1, u, 1, 1, 1)\ du
\end{aligned} \tag{28}$$

But by Lemma 12 (i) and (iv) respectively, Restriction 2 implies that

$$\begin{aligned}
f_{Y_3|Y_2Y_1ZUD_3D_2D_1}(y_3 \mid y, y_1, z, u, 1, 1, 1) &= f_{Y_3|Y_2U}(y_3 \mid y, u), \\
f_{Z|Y_2Y_1UD_3D_2D_1}(z \mid y, y_1, u, 1, 1, 1) &= f_{Z|U}(z \mid u).
\end{aligned}$$

Equation (28) thus can be rewritten as

$$f_{Y_3Y_2Y_1ZD_3D_2D_1}(y_3, y, y_1, z, 1, 1, 1) = \int f_{Y_3|Y_2U}(y_3 \mid y, u) \cdot f_{Z|U}(z \mid u) \times f_{Y_2Y_1UD_3D_2D_1}(y, y_1, u, 1, 1, 1)\, du$$

But this implies that the integral operator $L_{y,z}$ is written as the operator composition

$$L_{y,z} = P_y Q_z \tilde{L}_y$$

Restriction 9 (i), (ii), and (iv) imply that the operators P_y, Q_z, and $\tilde{L}_y$ are invertible, respectively. Hence so is $L_{y,z}$. Using the two values $\{0, 1\}$ of Z, form the product

$$L_{y,1} L_{y,0}^{-1} = P_y Q_{1/0} P_y^{-1}$$

where $Q_{z/z'} := Q_z Q_{z'}^{-1}$ is the multiplication operator with proxy odds defined by

$$(Q_{1/0}\xi)(u) = \frac{f_{Z|U}(1 \mid u)}{f_{Z|U}(0 \mid u)} \xi(u).$$

The rest of Step 1 is analogous to that of the proof of Lemma 2. Therefore, $F_{Y_3|Y_2U}$ and $F_{Z|U}$ are uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot,\,\cdot,\,\cdot,\,\cdot, 1, 1, 1)$. Equivalently, the operators P_y and Q_z are uniquely determined for each y and z, respectively.

Step 2: Uniqueness of $F_{Y_2Y_1UD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,, 1, 1)$

By Lemma 12 (ii), Restriction 2 implies $f_{Y_2|Y_1UD_2D_1}(y' \mid y, u, 1, 1) = f_{Y_2|Y_1U}(y' \mid y, u)$. Using this equality, write the density of the observed joint distribution $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\,\cdot\,, 1, 1)$ as

$$\begin{aligned} f_{Y_2Y_1D_2D_1}(y', y, 1, 1) &= \int f_{Y_2|Y_1UD_2D_1}(y' \mid y, u, 1, 1) f_{Y_1UD_2D_1}(y, u, 1, 1) du \\ &= \int f_{Y_2|Y_1U}(y' \mid y, u) f_{Y_1UD_2D_1}(y, u, 1, 1) du \end{aligned} \tag{29}$$

By Lemma 4 (i), $F_{Y_3|Y_2U}(y' \mid y, u) = F_{Y_2|Y_1U}(y' \mid y, u)$ for all y', y, u. Therefore, we can write the operator P_y as

$$(P_y\xi)(y') = \int f_{Y_3|Y_2U}(y' \mid y, u) \cdot \xi(u) du = \int f_{Y_2|Y_1U}(y' \mid y, u) \cdot \xi(u) du.$$

With this operator notation, it follows from (29) that

$$f_{Y_2Y_1D_2D_1}(\,\cdot\,, y, 1, 1) = P_y f_{Y_1UD_2D_1}(y,\ \cdot\,, 1, 1).$$

By Restriction 9 (i), this operator equation can be solved for $f_{Y_1UD_2D_1}(y,\ \cdot\,, 1, 1)$ as

$$f_{Y_1UD_2D_1}(y,\ \cdot\,, 1, 1) = P_y^{-1} f_{Y_2Y_1D_2D_1}(\,\cdot\,, y, 1, 1) \tag{30}$$

Recall that P_y was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\ \cdot\,,\ \cdot\,,\ \cdot\,, 1, 1, 1)$. The function $f_{Y_2Y_1D_2D_1}(\,\cdot\,, y, 1, 1)$ is also uniquely determined by the observed joint distribution $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\ \cdot\,, 1, 1)$ up to null sets. Therefore, (29) shows that $f_{Y_1UD_2D_1}(\,\cdot\,,\ \cdot\,, 1, 1)$ is uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\ \cdot\,,\ \cdot\,,\ \cdot\,, 1, 1, 1)$ and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\ \cdot\,, 1, 1)$.

Using the solution to the above inverse problem, we can write the kernel of the operator S_y as

$$\begin{aligned}
f_{Y_2Y_1UD_2D_1}(y', y, u, 1, 1) &= f_{Y_2|Y_1UD_2D_1}(y' \mid y, u, 1, 1) \cdot f_{Y_1UD_2D_1}(y, u, 1, 1) \\
&= f_{Y_2|Y_1U}(y' \mid y, u) \cdot f_{Y_1UD_2D_1}(y, u, 1, 1) \\
&= f_{Y_3|Y_2U}(y' \mid y, u) \cdot f_{Y_1UD_2D_1}(y, u, 1, 1) \\
&= f_{Y_3|Y_2U}(y' \mid y, u) \cdot [P_y^{-1} f_{Y_2Y_1D_2D_1}(\,\cdot\,, y, 1, 1)](u)
\end{aligned}$$

where the second equality follows from Lemma 12 (ii), the third equality follows from Lemma 4 (i), and the forth equality follows from (30). Since $f_{Y_3|Y_2U}$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\ \cdot\,,\ \cdot\,,\ \cdot\,, 1, 1, 1)$ and $[P_y^{-1} f_{Y_2Y_1D_2D_1}(\,\cdot\,, y, 1, 1)]$ was shown in the previous paragraph to be uniquely determined for each y by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\ \cdot\,,\ \cdot\,,\ \cdot\,, 1, 1, 1)$ and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\ \cdot\,, 1, 1)$, it follows that $f_{Y_2Y_1UD_2D_1}(\,\cdot\,,\ \cdot\,,\ \cdot\,, 1, 1)$ too is uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\ \cdot\,,\ \cdot\,,\ \cdot\,, 1, 1, 1)$ and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\ \cdot\,, 1, 1)$. Equivalently, the operator S_y is identified for each y.

Step 3: Uniqueness of $F_{Y_3D_3|Y_3U}(\,\cdot\,, 1 \mid\ \cdot\,,\ \cdot\,)$

The density of the observed joint distribution $F_{Y_3Y_2Y_1D_3D_2D_1}(y_3, y_2, y_1, 1, 1, 1)$ can be decomposed as

$$\begin{aligned} f_{Y_3Y_2Y_1D_3D_2D_1}(y_3, y_2, \cdot, 1, 1, 1) &= \int f_{Y_3D_3|Y_2Y_1UD_2D_1}(y_3, 1 \mid y_2, y_1, u, 1, 1) \\ &\qquad \times f_{Y_2Y_1UD_2D_1}(y_2, y_1, u, 1, 1)du \\ &= \int f_{Y_3D_3|Y_2U}(y_3, 1 \mid y_2, u) \cdot f_{Y_2Y_1UD_2D_1}(y_2, y_1, u, 1, 1)du \\ &= S_{y_2} \cdot f_{Y_3D_3|Y_2U}(y_3, 1 \mid y_2, \cdot) \end{aligned}$$

for each y_3 and y_2, where the second equality follows from Lemma 12 (i) and (iii). By Restriction 9 (iv), S_{y_2} is invertible, and we can rewrite the above equality as

$$f_{Y_3D_3|Y_2U}(y_3, 1 \mid y_2, \cdot) = S_{y_2}^{-1} f_{Y_3Y_2Y_1D_3D_2D_1}(y_3, y_2, \cdot, 1, 1, 1).$$

Recall that S_{y_2} was shown to be uniquely determined in Step 2 by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1, 1)$ and $F_{Y_2Y_1D_2D_1}(\cdot, \cdot, 1, 1)$. Therefore, $F_{Y_3D_3|Y_2U}(\cdot, 1 \mid \cdot, \cdot)$ is identified by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1, 1)$ and $F_{Y_2Y_1D_2D_1}(\cdot, \cdot, 1, 1)$.

Step 4: Uniqueness of $F_{D_3|Y_3U}(1 \mid \cdot, \cdot)$

The density of the observed joint distribution $F_{Y_3D_3|Y_2U}(y_3, 1 \mid y_2, u)$ can be decomposed as

$$\begin{aligned} f_{Y_3D_3|Y_2U}(y_3, 1 \mid y_2, u) &= f_{D_3|Y_3Y_2U}(1 \mid y_3, y_2, u) \cdot f_{Y_3|Y_2U}(y_3 \mid y_2, u) \\ &= f_{D_3|Y_3U}(1 \mid y_3, u) \cdot f_{Y_3|Y_2U}(y_3 \mid y_2, u) \end{aligned}$$

where the second equality follows from Lemma 12 (iii). For each pair (y_3, u) in the support, there exists y_2 such that $f_{Y_3|Y_2U}(y_3 \mid y_2, u) > 0$. For such y_2, rewrite the above equation as

$$f_{D_3|Y_3U}(1 \mid y_3, u) = \frac{f_{Y_3D_3|Y_2U}(y_3, 1 \mid y_2, u)}{f_{Y_3|Y_2U}(y_3 \mid y_2, u)}.$$

Recall that Step 1 showed that $F_{Y_3|Y_2U}$ is uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1, 1)$, and Step 3 showed that $F_{Y_3D_3|Y_2U}(\cdot, 1 \mid \cdot, \cdot)$ is identified by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\cdot, \cdot, \cdot, \cdot, 1, 1, 1)$

and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\,\cdot\,,1,1)$. Therefore, $F_{D_3|Y_3U}(1 \mid \,\cdot\,,\,\cdot\,)$ is identified by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1,1)$ and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\,\cdot\,,1,1)$.

Step 5: Uniqueness of $F_{Y_1U|D_1=1}$

Recall from Step 2 that $f_{Y_2Y_1UD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ is uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1,1)$ and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\,\cdot\,,1,1)$. We can write

$$\begin{aligned} f_{Y_2Y_1UD_2D_1}(y',y,u,1,1) &= f_{D_2|Y_2Y_1UD_1}(1 \mid y',y,u,1) f_{Y_2|Y_1UD_1}(y' \mid y,u,1) f_{Y_1UD_1}(y,u,1) \\ &= f_{D_2|Y_2U}(1 \mid y',u) f_{Y_2|Y_1U}(y' \mid y,u) f_{Y_1UD_1}(y,u,1) \\ &= f_{D_3|Y_3U}(1 \mid y',u) f_{Y_3|Y_2U}(y' \mid y,u) f_{Y_1UD_1}(y,u,1) \end{aligned}$$

where the second equality follows from Lemma 12 (ii), and the third equality follows from Lemma 4 (i) and (ii). For a given (y,u), there must exist some y' such that $f_{Y_3|Y_2U}(y' \mid y,u) > 0$ by a property of conditional density functions. Moreover, Restriction 9 (iii) requires that $f_{D_3|Y_3U}(1 \mid y',u) > 0$ for a given y' for all u. Therefore, for such a choice of y', we can write

$$f_{Y_1UD_1}(y,u,1) = \frac{f_{Y_2Y_1UD_2D_1}(y',y,u,1,1)}{f_{Y_3|Y_2U}(y' \mid y,u) f_{D_3|Y_3U}(1 \mid y',u)}$$

Recall that $f_{Y_3|Y_2U}(\,\cdot\mid\,\cdot\,,\,\cdot\,)$ was shown in Step 1 to be uniquely determined by the observed joint distribution $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1,1)$, $f_{Y_2Y_1UD_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1)$ was shown in Step 2 to be uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1,1)$ and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\,\cdot\,,1,1)$, and $f_{D_3|Y_3U}(1 \mid \,\cdot\,,\,\cdot\,)$ was shown in Step 4 to be uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1,1)$ and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\,\cdot\,,1,1)$. Therefore, it follows that the initial joint density $f_{Y_1U|D_1=1}$ is uniquely determined by the observed joint distributions $F_{Y_3Y_2Y_1ZD_3D_2D_1}(\,\cdot\,,\,\cdot\,,\,\cdot\,,\,\cdot\,,1,1,1)$ and $F_{Y_2Y_1D_2D_1}(\,\cdot\,,\,\cdot\,,1,1)$. □

CHAPTER 2

Structural Partial Effects

1. Introduction

Economists are often interested in the structural partial effect β of models of the form

$$Y = \alpha + \beta X + \mathcal{E}, \qquad X \not\perp\!\!\!\perp \mathcal{E}.$$

In this constant-coefficient affine structure, the local instrumental variable (LIV) defined by $\frac{d}{dz}\mathrm{E}[Y \mid Z = z]/\frac{d}{dz}\mathrm{E}[X \mid Z = z]$ using any point of an instrumental variable $Z = z$ identifies this structural parameter β even if the first stage is nonparametric and nonseparable. It boils down to the two-stage least squares when the first stage is a constant-coefficient affine model.

More generally, economists are interested in the structural partial effect $\beta(X, \mathcal{E}) := \frac{\partial}{\partial x} g(X, \mathcal{E})$ of nonparametric and nonseparable structural models of the form

$$Y = g(X, \mathcal{E}), \qquad X \not\perp\!\!\!\perp \mathcal{E}.$$

This paper shows that the LIV continues to identify the structural partial effect even in this nonparametric framework under certain first-stage restrictions. Moreover, we generalize the LIV identification methods to accommodate a more general class of first-stage models.

An identification concept of nonseparable models under exogeneity and monotonicity is discussed by Matzkin (2003). Hoderlein and Mammen (2007) discuss what can be identified without monotonicity. Chesher (2003) identifies structural partial effects of nonseparable models under endogeneity. In the meanwhile, control variable approaches are proposed as ways to turn endogeneity into conditional exogeneity (Altonji and Matzkin, 2005; Imbens and Newey, 2009). Identification of quantile structural functions under endogeneity is studied by Chernozhukov and Hansen (2005),

Chernozhukov, Imbens and Newey (2007), Torgovitzky (2011), and D'Haultfoeuille and Février (2011). Our work is perhaps most closely related to Chesher (2003) who specifically identifies structural partial effects for well-defined subpopulations of economic agents.

Because the statistical parameter (two-stage least squares) coincides with the structural partial effect in the classical affine regression models, it is worth starting out with this classical idea. We explore possible directions in which this classical idea can be extend to nonparametric and nonseparable models. Heckman and Vytlacil (1999, 2005, 2007) demonstrate that the localized version of the IV estimator (LIV) identifies structural causal effects under nonparametric binary treatment models. We show that the same statistical object can be used to identify average structural partial effects for a class of nonseparable and nonparametric models in Section 2. We further extend this idea in Section 3 for identification of the general marginal treatment effect (MTE) projected on subpopulations characterized by all observed variables, i.e., $\mathrm{E}[\beta(X,\mathcal{E}) \mid YXZ]$. The identifying statistical parameter of a special case of the MTE corresponds to well-known formula previously proposed in the literature (Chesher, 2003; Imbens and Newey, 2009) - see Section 4.2.

2. The Local Instrumental Variable Estimator

Recall the classical two-stage constant-coefficient affine structures of the form

$$\begin{cases} Y = \alpha + \beta X + \mathcal{E} \\ X = \gamma + \delta Z + U \end{cases} \qquad \text{where } \mathrm{E}[(\mathcal{E},U) \mid Z] = (0,0). \tag{31}$$

The structural parameter β is identified by the ratio of the two reduced-form mean regressions

$$\frac{\frac{d}{dz}\mathrm{E}[Y \mid Z=z]}{\frac{d}{dz}\mathrm{E}[X \mid Z=z]}. \tag{32}$$

A sample analog of this fraction is nothing but the two-stage least-squares estimator of β.[1]

[1] Recall that the two-stage least squares is $\beta_{2SLS} = e_2'\mathrm{E}[Z'X]^{-1}\mathrm{E}[Z'Y] = \mathrm{Cov}(Z,Y)/\mathrm{Cov}(Z,X) = [\mathrm{Cov}(Z,Y)/\mathrm{Var}(Z)]/[\mathrm{Cov}(Z,X)/\mathrm{Var}(Z)] = \frac{d}{dz}\mathrm{E}[Y \mid Z=z]/\frac{d}{dz}\mathrm{E}[X \mid Z=$

This fraction (32) remains to identify a variety of causal effects in another important class of models. Heckman and Vytlacil (1999, 2005, 2007) consider a nonparametric framework of binary treatment model of the form

$$\begin{cases} Y = g(X, \mathcal{E}) \\ X = 1\{h(Z) \geqslant U\} \end{cases} \qquad \text{where } Z \perp\!\!\!\perp (\mathcal{E}, U). \tag{33}$$

Despite the apparent discrepancy between the models (31) and (33), they show that the same fraction (32) still identifies the marginal treatment effect $\mathrm{E}[g(1,\mathcal{E}) - g(0,\mathcal{E}) \mid U]$ under (33), which in turn can be used to recover various treatment parameters. Heckman and Vytlacil call (32) the local instrumental variable (LIV) estimator.[2]

Given that the LIV identifies causal effects in both the parametric continuous treatment model (31) and the nonparametric binary treatment model (33), our natural question is: to how much extent can we nonparametrically generalize (31) while keeping the LIV capable of identifying causal effects? This question is practically important because the LIV formula (32) is a natural generalization of the conventional two-stage least squares that certainly work for the classical model (31), whereas the true model may be more general than (31).

In order to answer this question, consider a class of nonparametric and nonseparable two-stage structures of the form

$$\begin{cases} Y = g(X, \mathcal{E}) \\ X = h(Z, U) \end{cases} \qquad \text{where } Z \perp\!\!\!\perp (\mathcal{E}, U), \tag{34}$$

We define the partial effect by $\beta(x, \varepsilon) := \frac{\partial}{\partial x} g(x, \varepsilon)$. It is straightforward to see that the local average partial effect $\mathrm{E}[\beta(X, \mathcal{E}) \mid Z = z]$ can be identified by the LIV formula (32) under nonparametric regression models. We provide an exact condition under

$z]$. The numerator $\frac{d}{dz}\mathrm{E}[Y \mid Z = z]$ identifies the reduced-form composite parameter $\beta\delta$ and the denominator $\frac{d}{dz}\mathrm{E}[X \mid Z = z]$ identifies the reduced-form first-stage parameter δ under the constant-coefficient affine model (31).

[2] They define the LIV as the derivative $d\mathrm{E}[Y \mid P = p]/dp$ where $P = \mathrm{E}[X \mid Z]$, but it is equivalent to (32).

which this LIV identification result remains to hold under the nonseparable model (34).

ASSUMPTION 6 (Local Rank Condition for LIV). $\frac{d}{dz}\mathrm{E}[X \mid Z = z] \neq 0$.

ASSUMPTION 7 (Stochastically Separable First Stage).
$\mathrm{Cov}\left(\beta(h(z,U),\mathcal{E}),\ \frac{\partial}{\partial z}h(z,U)\right) = 0$.

Assumption 6 is a generalization of the conventional rank condition $\delta \neq 0$ under the classical model (31). In the case of endogeneity, Assumption 7 generally amounts to separable first stage model $h(z,u) = \mu(z) + u$, unless the second stage model takes the specific form of a constant-coefficient affine model $g(x,\varepsilon) = \alpha + \beta x + \varepsilon$ as in (31). The following theorem states that this assumption is essential for the LIV formula (32) to identify the local average partial effect $\mathrm{E}[\beta(X,\mathcal{E}) \mid Z = z]$.

THEOREM 2 (LIV: Necessary and Sufficient Condition). *Suppose that Assumption 6 is satisfied for the model (34).*[3] *Then,*

$$\mathbb{E}\left[\beta(X,A)|Z=z\right] = \frac{\frac{d}{dz}E[Y \mid Z=z]}{\frac{d}{dz}E[X \mid Z=z]}$$

holds if and only if Assumption 7 is true.

If we allow the second stage structural function g to take more arbitrary forms than the simple affine model (31), this necessary and sufficient condition generally amounts separable first stage, $h(z,u) = \mu(z) + u$, whose representative example is of course the nonparametric mean regression model. Therefore, given the result of Theorem 2, we hereafter discuss the LIV within the framework of the following assumption.

ASSUMPTION 8 (Separable First Stage). $h(z,u) = \mu(z) + u$.

Recall that the two-stage least squares identifying the structural parameter β under the classical model (31) can be interpreted as the coefficient in the regression of

[3] In addition, we also assume the following regularity conditions: g and h are continuously differentiable with respect to their first arguments; and $\beta(h(z,\cdot),\cdot)$ is dominated in absolute value by an $L^1(F_{\mathcal{E}U})$ function.

Y on the predicted value of X, i.e., the partial effect of δZ on Y. This interpretation carries over to the nonparametric model (34) under Assumption 8. That is, the structural partial effect $\mathrm{E}[\beta(X,\mathcal{E}) \mid Z = z]$ can be interpreted as and can be identified by the partial effect of $\mu(Z)$ on Y.

THEOREM 3 (Mean and Quantile LIV). *Suppose that Assumptions 6 and 8 are satisfied for the model (34).*[4] *With the notation* $P := \mu(Z)$*, the following equalities hold.*

$$(i) \qquad E[\beta(X,\mathcal{E}) \mid Z = z] \;=\; \frac{d}{dp} E[Y \mid P = p]\Big|_{p=\mu(z)} \qquad \textit{and}$$

$$(ii) \qquad E[\beta(X,\mathcal{E}) \mid Y = Q_{Y|P}(\tau \mid \mu(z)),\ Z = z] \;=\; \frac{\partial}{\partial p} Q_{Y|P}(\tau \mid p)\Big|_{p=\mu(z)}$$

where $Q_{Y|P}(\tau \mid p) := \inf\{y \mid F_{Y|P}(y \mid p) \geqslant \tau\}$ *denotes the* τ*-th quantile regression of* Y *on* P*.*

Combining the identifying equality of Theorem 2 with Theorem 3 (i), we have

$$\mathbb{E}\left[\beta(X,A)|Z=z\right] = \underbrace{\frac{\frac{d}{dz}\mathrm{E}[Y \mid Z=z]}{\frac{d}{dz}\mathrm{E}[X \mid Z=z]}}_{(a)} = \underbrace{\frac{d}{dp}\mathrm{E}[Y \mid P=p]\Big|_{p=\mu(z)}}_{(b)},$$

which confirms that the conventional property extends to the current nonparametric setting, i.e., the structural partial effect can be identified by both (a) the ratio of reduced-form partial effects and (b) the partial effect of Y on the predicted value P of X.

This result even extends to the quantile counterpart. Quantile regressions are only statistical objects, and usually do not have structural interpretations particularly under endogeneity. Theorem 3 (ii), however, shows that the slope of the quantile regression $Q_{Y|P}$ on the right-hand side *does* identify an average of the structural partial effects $\beta(X,A)$ on the left-hand side. Notice that the identifying quantile regression is $Q_{Y|P}$, which is in general different from $Q_{Y|X}$.

[4] In addition, we also assume the following regularity conditions: g is continuously differentiable with respect to x; $F_{Y|P}$ is continuously differentiable with respect to y; $Q_{Y|P}$ is continuously differentiable with respect to p; $f_{Y|P}$ is continuous in p; and $\beta(h(z,\cdot),\cdot)$ is dominated in absolute value by an $L^1(F_{\mathcal{E}U})$ function.

We make a remark on the analogous expressions between parts (i) and (ii) of Theorem 3. The mean regression $\mathrm{E}[Y \mid P]$ is used to identify $\mathrm{E}[\beta(X,\mathcal{E}) \mid Z]$ in part (i), whereas the quantile regression $Q_{Y|P}$ is used to identify $\mathrm{E}[\beta(X,\mathcal{E}) \mid Y,Z]$ in part (ii). This parallel is intuitively straightforward, but is not too simple to be explained concisely with logical precision, because quantile regressions do not directly transform into moments in general. See the proof in the appendix to find out different approaches used to prove the seemingly similar formulas in (i) and (ii).

3. Marginal Treatment Effects

The previous section studied identifiability of the local averages of the structural partial effects $\mathrm{E}[\beta(X,\mathcal{E}) \mid Z]$ and $\mathrm{E}[\beta(X,\mathcal{E}) \mid Y,Z]$ by the LIV. However, the LIV is nothing but one particular statistical expression, and need not be considered as the sole identifying device. In this section, we ask if we can extend the idea of the LIV to identify causal effects, $\mathrm{E}[\beta(X,\mathcal{E}) \mid X,Z]$ and $\mathrm{E}[\beta(X,\mathcal{E}) \mid Y,X,Z]$ on finer subpopulations characterized by the additional conditioning variable X. Using the same idea as Theorem 3 (i) under Assumptions 6 and 8 yields

$$\begin{aligned}\mathrm{E}[\beta(X,\mathcal{E}) \mid X=\mu(z)+u, Z=z] &= \frac{d}{dp}\mathrm{E}[Y \mid X=p+u, P=p]\Big|_{p=\mu(z)} \\ &= \frac{\partial}{\partial x}\mathrm{E}[Y \mid X=\mu(z)+u, Z=z] + \frac{\frac{\partial}{\partial z}\mathrm{E}[Y \mid X=\mu(z)+u, Z=z]}{\mu'(\mu(z))} \qquad (35)\end{aligned}$$

The identifying statistical object on the right-hand side is no longer the LIV, but the structural partial effect $\mathrm{E}[\beta(X,\mathcal{E}) \mid X,Z]$ of interest is indeed identified. This section presents generalization of this heuristic result by replacing Assumptions 6 and 8 by the following assumptions.

Assumption 9 (Invertibility). $h(z,\cdot\,)$ is invertible at each z, and $v(z,\cdot\,)$ denotes the inverse.

Assumption 10 (Local Rank Condition for MTE). $\frac{\partial}{\partial z}v(z,x) \neq 0$.

THEOREM 4 (Mean and Quantile MTE). *Suppose that Assumptions 9 and 10 are satisfied for the model (34).*[5] *With the notation* $\rho(z,x) := \left[\frac{\partial \upsilon(z,x)}{\partial x}\right] / \left[\frac{\partial \upsilon(z,x)}{\partial z}\right]$, *the following equalities hold.*

$$(i) \quad E[\beta(X,\mathcal{E}) \mid X=x, Z=z] \;=\; \frac{\partial}{\partial x} E[Y \mid X=x, Z=z] - \rho(z,x) \cdot \frac{\partial}{\partial z} E[Y \mid X=x, Z=z] \qquad \textit{and}$$

$$(ii) \quad E[\beta(X,\mathcal{E}) \mid Y = Q_{Y|XZ}(\tau \mid x,z), X=x, Z=z] \;=\; \frac{\partial}{\partial x} Q_{Y|XZ}(\tau \mid x,z) - \rho(z,x) \cdot \frac{\partial}{\partial z} Q_{Y|XZ}(\tau \mid x,z)$$

REMARK 11. While the model (34) entails the strong instrument independence $Z \perp\!\!\!\perp (\mathcal{E}, U)$, only a weaker form of instrument independence $Z \perp\!\!\!\perp \mathcal{E} \mid U$ is required for Theorem 4. In other words, the instrument Z may be correlated with the first-stage unobservable U.

Observe the parallel between parts (i) and (ii) of Theorem 4, which is similar to that of Theorem 3. The mean regression $\mathrm{E}[Y \mid X, Z]$ is used to identify $\mathrm{E}[\beta(X,\mathcal{E}) \mid X, Z]$ in part (i), whereas the quantile regression $Q_{Y|X,Z}$ is used to identify $\mathrm{E}[\beta(X,\mathcal{E}) \mid Y, X, Z]$ in part (ii). Again, this parallel is not too simple to be explained concisely because differentiating quantile regressions do not directly transform into moments of derivatives.

Theorem 4 (i) proposes identification of $\mathrm{E}[\beta(X,\mathcal{E}) \mid X, Z]$, but the conditioning variable X is by itself an endogenous outcome of more primitive variables (Z, U). In order to give precise economic interpretation to this causal effect, it would be useful to identify causal effects conditional on a set of primitive variables, e.g., $\mathrm{E}[\beta(x,\mathcal{E}) \mid Z, U]$ instead of $\mathrm{E}[\beta(X,\mathcal{E}) \mid X, Z]$. Under Assumption 9, they in fact coincide to each other:

$$(36) \qquad \mathrm{E}[\beta(x,\mathcal{E}) \mid \underbrace{Z=z, U=\upsilon(z,x)}_{\text{Primitive Condition}}] \;=\; \mathrm{E}[\beta(X,\mathcal{E}) \mid \underbrace{X=x,\ Z=z}_{\text{Endogenous Condition}}].$$

[5] In addition, we also assume the following regularity conditions: g is continuously differentiable with respect to x; $F_{Y|XZ}$ is continuously differentiable with respect to y; $Q_{Y|XZ}$ is continuously differentiable with respect to (x,z); $f_{Y|XZ}$ is continuously differentiable with respect to (x,z); and $\beta(x,\cdot)$ is dominated in absolute value by an $L^1(F_{\mathcal{E}|XZ})$ function.

This shows that the formula in Theorem 4 (i) identifies the causal effects $\mathrm{E}[\beta(x,\mathcal{E}) \mid Z,U]$, which is close in spirit to Heckman and Vytlacil's (2005) marginal treatment effect (MTE) proposed in the context of the binary treatment model (33). We therefore refer to the causal effects indentified in this section as the MTE.

Because Theorem 4 requires Assumption 9, identification of the MTE presumes our knowledge of the first-stage structure h. In many economic applications, structural construction of economic models provide explicit formula for the first-stage function h. The following example illustrates the case in point.

EXAMPLE 2. Suppose that $Y = g(X,\mathcal{E})$ models the demand for a single good Y as a function of income X and preferences $\mathcal{E}$ as in the study of Engel curves. Income X can be taken to be within-period total expenditure, which is justified under preference restrictions (Lewbel, 1999). The first-stage endogenous choice $X = h(Z,U)$ is modeled as a result of optimization behaviors.

Suppose that the dynamic consumption decision of economic agent at time t is given by

$$\max_{\{X_{t+\tau}\}_{\tau=0}^{\bar{T}-t}} \mathrm{E}_t\left[\sum_{\tau=0}^{\bar{T}-t}\beta^\tau u(X_{t+\tau};\theta)\right] \quad \text{s.t.} \quad M_{t+1} = (M_t - X_t)R + Z_{t+1}$$

where $u(\cdot;\theta)$ is the CARA utility function with parameter θ, Z_t is the consumer's idiosyncratic labor income, M_t denotes assets, and R is the interest factor which is fixed and deterministic for simplicity. If the growth of Z_t is stochastic with Gaussian iid innovation with variance σ^2, then the first-stage function for individuals with no initial assets is given by

$$X_t = Z_t - \frac{\sigma^2\theta}{2\beta},$$

where individuals have heterogeneous structural parameters (β,θ,σ^2). Applying Theorem 4 (i) together with (36) yields identification of the local average maginal Engel coefficient

$$\mathrm{E}[\beta(x,\mathcal{E}) \mid Z = z,\ \sigma^2\theta/\beta = 2(z-x)] = \frac{\partial}{\partial x}\mathrm{E}[Y \mid X = x, Z = z] + \frac{\partial}{\partial z}\mathrm{E}[Y \mid X = x, Z = z]$$

The object identified in this equation has a clear economic interpretation: the average of heterogeneous marginal Engel coefficients at $X = x$ among the subpopulation of consumers earning z units of income with volatility σ^2, having preference parameters (β, θ) satisfying the relation $\sigma^2\theta/\beta = 2(z - x)$.

Note that, in this example, the first-stage structure is trivially identified to be $X = Z + U$ where $U := -0.5\sigma^2\theta/\beta$. This trivial identification of the first-stage holds even without any statistical or mean independence conditions between Z and U. In other words, instrument independence for the first-stage unobservables need not hold for the purpose of identifying the structural causal effects (see Remark 11).

4. Two Special Cases of the MTE

Example 2 illustrated a clear economic interpretation of the identified causal effects (MTE) when the first stage is structurally constructed. Many applications, however, lack such structural motivations. In the absence of structural models, economists have often substituted statistical objects such as mean regressions and quantile regressions. In this section, we demonstrate that the MTE is also compatible with such statistical devices, though its economic interpretation becomes less clear than in the case of Example 2. Sections 4.1 and 4.2 propose the special cases of the MTE when the first stage is abstractly summarized by a mean regression and a quantile regression, respectively.

4.1. Case 1: When First Stage Is a Mean Regression. Suppose that the first-stage model is a nonparametric mean regression of the form

$$h(z, u) = \mu(z) + u \qquad \text{where } \mathrm{E}[U \mid Z] = 0.$$

In this case, Assumption 9 is trivially satisfied. Furthermore, the traditional local rank condition $\mu'(z) \neq 0$ satisfies Assumption 10, and Theorem 4 can therefore be used. Note that Remark 11 following Theorem 4 states that statistical independence between the instrument Z and the first-stage unobservable U is not required. Therefore, heteroscedasticity in the first stage $\mathrm{E}[U^2 \mid Z] \neq 0$ is admissible in particular.

Applying the theorem to this special case entails $\rho(z,x) = -1/[\mu'(z)]$ and yields the following result.

COROLLARY 3 (MTE When First Stage is a Mean Regression). *Suppose that the model is given by*

$$\begin{cases} Y = g(X,\mathcal{E}) & \text{where } Z \perp\!\!\!\perp \mathcal{E} \mid U \\ X = \mu(Z) + U & \text{where } E[U \mid Z] = 0 \text{ and } \mu'(z) \neq 0 \end{cases}$$

Then, the following identifying equalities hold.

$$(i) \quad E[\beta(X,\mathcal{E}) \mid X = x, Z = z] = \frac{\partial}{\partial x} E[Y \mid X = x, Z = z] + \frac{\frac{\partial}{\partial z} E[Y \mid X = x, Z = z]}{\mu'(z)} \qquad \text{and}$$

$$(ii) \quad E[\beta(X,\mathcal{E}) \mid Y = Q_{Y|XZ}(\tau \mid x,z), X = x, Z = z] = \frac{\partial}{\partial x} Q_{Y|XZ}(\tau \mid x,z) + \frac{\frac{\partial}{\partial z} Q_{Y|XZ}(\tau \mid x,z)}{\mu'(z)}$$

Not surprisingly, Corollary 3 (i) corresponds to (35), by which the MTE was motivated as a natural extension of the LIV under separable first stage models.

4.2. Case 2: When First Stage Is a Quantile Regression. Another important special case is when the first-stage model is represented by a nonparametric quantile regression of the form

$$h(z,u) = Q_{X|Z}(u \mid z) \qquad \text{where } Z \perp\!\!\!\perp U.$$

In this case, Assumption 9 is satisfied if the $X \mid Z = z$ is continuously distributed. Furthermore, the traditional local rank condition $\frac{\partial}{\partial z} F_{X|Z}(x \mid z) \neq 0$ satisfies Assumption 10. Applying Theorem 4 to this special case entails $\rho(z,x) = -\left[\frac{\partial}{\partial z} Q_{X|Z}(u \mid z)\right]^{-1}$ where $u = F_{X|Z}(x \mid z)$, and therefore yields the following result.

COROLLARY 4 (MTE When First Stage is a Quantile Regression). *Suppose that the model is given by*

$$\begin{cases} Y = g(X,\mathcal{E}) \\ X = Q_{X|Z}(U \mid Z) \end{cases} \qquad \text{where } Z \perp\!\!\!\perp (\mathcal{E}, U)$$

Then, the following identifying equalities hold.

$$(i) \quad E[\beta(X,\mathcal{E}) \mid X=x, Z=z] \;=\; \frac{\partial}{\partial x} E[Y \mid X=x, Z=z] + \frac{\frac{\partial}{\partial z} E[Y \mid X=x, Z=z]}{\frac{\partial}{\partial z} Q_{X|Z}(u \mid z)} \qquad \textit{and}$$

$$(ii) \quad E[\beta(X,\mathcal{E}) \mid Y=Q_{Y|XZ}(\tau \mid x,z), X=x, Z=z] \;=\; \frac{\partial}{\partial x} Q_{Y|XZ}(\tau \mid x,z) + \frac{\frac{\partial}{\partial z} Q_{Y|XZ}(\tau \mid x,z)}{\frac{\partial}{\partial z} Q_{X|Z}(u \mid z)}$$

where $u := F_{X|Z}(x \mid z)$.

The right-hand side of Corollary 4 (ii) turns out to be the same as the identifying equality

$$\frac{\partial}{\partial x} Q_{Y|XU}(y \mid x,u) \;=\; \frac{\partial}{\partial x} Q_{Y|XZ}(\tau \mid x,z) + \frac{\frac{\partial}{\partial z} Q_{Y|XZ}(\tau \mid x,z)}{\frac{\partial}{\partial z} Q_{X|Z}(u \mid z)} \tag{37}$$

derived by Imbens and Newey (2009; Theorem 2). The left-hand side expressions, however, are different. Equation (37) identifies the statistical object $\frac{\partial}{\partial x} Q_{Y|XU}(y \mid x,u)$, whereas Corollary 4 (ii) identifies a local average of the structural object $\beta(X,\mathcal{E})$. Quantile partial effects do not generally represent structural partial effects unless rank invariance is assumed. Our result therefore adds a structural interpretation to (37), and parallels Chesher (2003), who originally derived this formula to identify non-averaged structural partial effect.

The identifying equalities in Corollary 4 (i) and (ii) remain to hold whenever the first-stage function h is monotone with respect to the unobservables u, because a quantile regression can be used to represent a monotone first stage. The monotonicity, however, is also the exact limit up to which they remain to hold. In other words, the following assumption is a necessary and sufficient condition for these identifying equalities.

Assumption 11 (Monotonicity). There exist functions $\bar{h} : \mathbb{R}^2 \to \mathbb{R}$ and $\iota : \mathbb{R}^M \to \mathbb{R}$ such that $h(z,u) = \bar{h}(z, \iota(u))$ and $\bar{h}$ is strictly monotone in its second argument.

Proposition 3 (Necessary and Sufficient Condition). *Suppose that* $F_{X|Z}(\cdot \mid z)$ *is strictly increasing in the model (34).*[6] *Then the identifying equalities in Corollary 4*

[6] In addition, we also assume the following regularity conditions: g is continuously differentiable with respect to x; h is continuously differentiable; $F_{X|Z}$ is absolutely continuous; $Q_{X|Z}$ is continuously differentiable with respect to z; and $f_{U|XZ}$ is continuously differentiable with respect to (x,z).

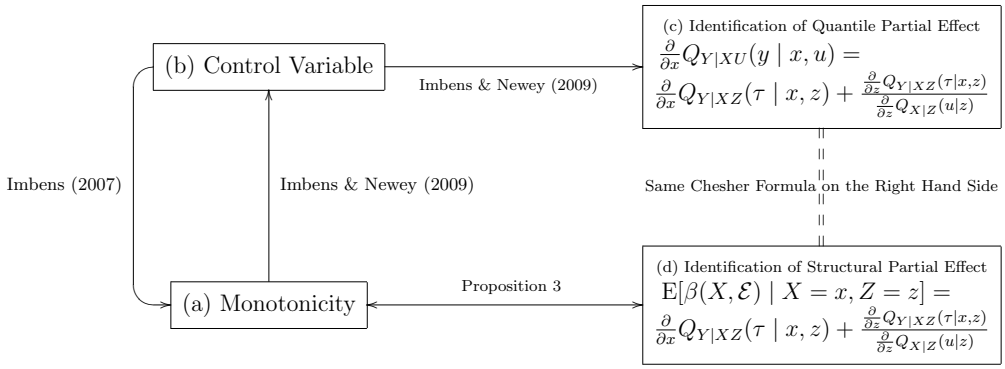


FIGURE 2.1. The role of monotonicity in related identification results.

(i) and (ii) hold for all structural models $(g, F_{\mathcal{E}|U})$ *if and only if Assumption 11 holds for the first-stage function* h*.*

Imbens and Newey (2009) show that the monotonicity is sufficient for a control variable,[7] which in turn is sufficient for the identifying equality (37). We show that the monotonicity is necessary and sufficient for the identifying equality in Corollary 4 (ii). These relations are summarized in the logical diagram in Figure 2.1.

5. Nonlinear Heterogeneous Effects of Smoking

Adverse effects of smoking during pregnancy on infant birth weights have been extensively studied in the health economic literature (e.g., Rosenzweig and Schultz (1983); Evans and Ringel (1999); Lien and Evans (2005)). Most papers, including those in the medical literature, have suggested that the effect of smoking (as binary variable) on infant birth weights ranges from -200 grams to -400 grams. Given that the average number of cigarettes smoked by smoking pregnant women is 12 between years 1989 and 1999, average effects of one cigarette on infant birth weight thus ranges from -17 grams to -33 grams. The goal of our analysis is to provide a much more detailed assessment of the effect of smoking, in particular we consider the

[7] Imbens (2007) discusses its necessity. This is followed up by Kasy (2011).

heterogeneous marginal effects of a single cigarette as opposed to these coarse average effects.

Specifically, we analyze the effects of the number of cigarettes on infant birth weight, extending an older idea of Evans and Ringel (1999). We allow for arbitrary nonlinear, endogenous and heterogeneous effects of smoking, and want to obtain averages of causal marginal effects for various subpopulations defined by treatment intensity, as well as other variables that proxy for unobserved heterogeneity as detailed below. Evans and Ringel use cigarette excise tax rate as source of exogenous variation to mitigate confounding factors in identifying the effects of smoking. We follow this idea; in our framework tax rates hence play the role of Z, while number of cigarettes per day and infant birth weight are X and Y, respectively. The causal model is then given by

$$\begin{cases} Y = g(X, S, \mathcal{E}) \\ X = h(Z, S, U) \end{cases}$$

where $\mathcal{E}$ captures other unobserved factors related to the lifestyle of the mother that impact the child's birth weight. Other observed characteristics of the mother, denoted S, are also controlled for, including maternal age, alcohol intake, number of prenatal visits, and number of live births experienced. We use a cross section of the natality data from the Natality Vital Statistics System of the National Center for Health Statistics. The main variables in the data are summarized in Table 2.1. From this data set we extract a random sample of size 100,000 from the time period between 1989 to 1999.

The structural features of interest are the averages marginal effect of a cigarette, $\beta(X, S, \mathcal{E})$, using subpopulation defined by combinations of Y, X and Z. Such causal effects are identified in Theorems 2–4 and Corollaries 3 and 4. Note that all these identifying equalities are proposed with derivatives of nonparametric mean and/or quantile regressions, whose estimation and large sample theories are very standard in the econometric literature (see Fan and Gijbels, 1996). We do not elaborate on these standard statistical results in this paper. Estimates of the structural partial effects

Variable	Mean	Std. Dev.	Description
Birth Weight	3330	606	Infant birth weight measured in grams
Cigarette	1.75	5.51	Number of cigarettes smoked per day
Tax	30.4	15.5	Excise tax rate on cigarettes in percentage
Age	26.7	6.0	Maternal age
Drinks	0.04	0.75	Number of times of drinking per week
Visits	11.3	4.1	Number of prenatal care visits
Births	1.97	1.00	Number of live births experienced

TABLE 2.1. Descriptive statistics of the data.

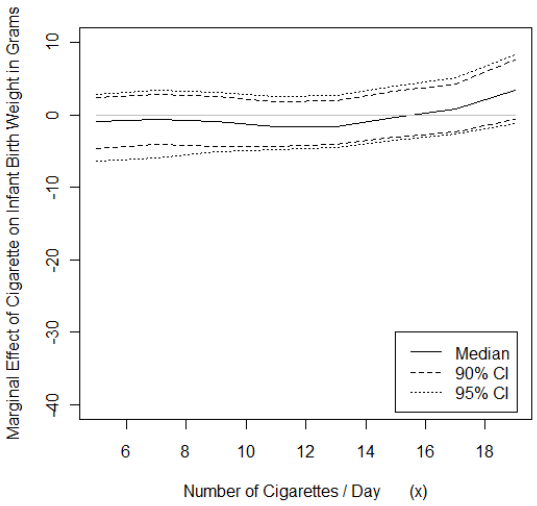


FIGURE 2.2. Confidence intervals of $\mathbb{E}[\beta(X,S,A) \mid Y = 2500, X = x, Z = 0.30, S = \bar{s}]$.

$\beta(X,S,\mathcal{E})$ projected on (Y,X,Z) are plotted in Figures 2.2–2.7). Due to the point mass of the distribution of X at $X = 0$ which conflicts the assumption of absolute continuity, our analysis focuses on the domain outside of this locality. With this framework in place, we make the following observations:

1. Comparing the graphs with lower Z (e.g., Figure 2.2) and higher Z (e.g., Figure 2.6), we observe ceteris paribus a great deal of heterogeneity in overall effects. In

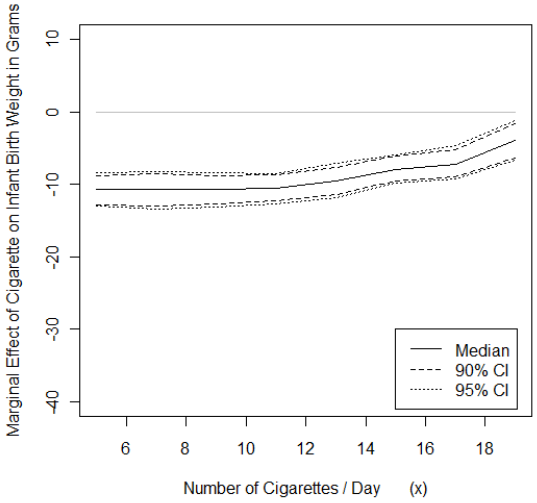


FIGURE 2.3. Confidence intervals of $\mathbb{E}[\beta(X, S, A) \mid Y = 3000, X = x, Z = 0.30, S = \bar{s}]$.

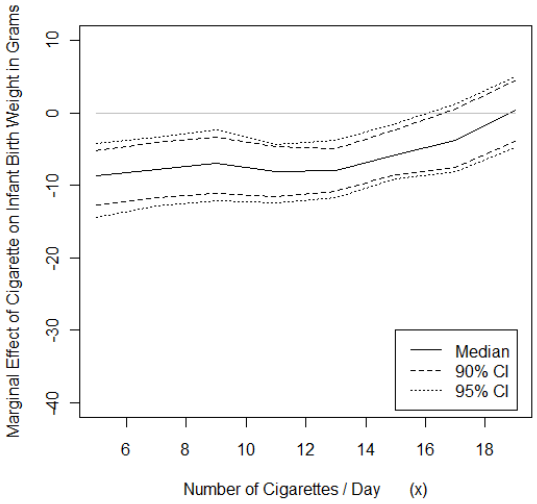


FIGURE 2.4. Confidence intervals of $\mathbb{E}[\beta(X, S, A) \mid Y = 2500, X = x, Z = 0.40, S = \bar{s}]$.

particular, the marginal effects under higher tax rates are relatively larger in magnitude. In other words, pregnant women who still choose to smoke despite facing higher tax rates exhibit larger marginal effects of smoking on infant birth weights. We will discuss this phenomenon in more detail below.

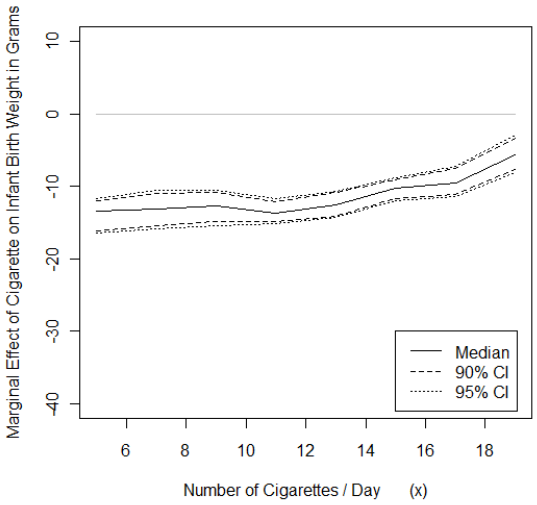


FIGURE 2.5. Confidence intervals of $\mathbb{E}[\beta(X, S, A) \mid Y = 3000, X = x, Z = 0.40, S = \bar{s}]$.

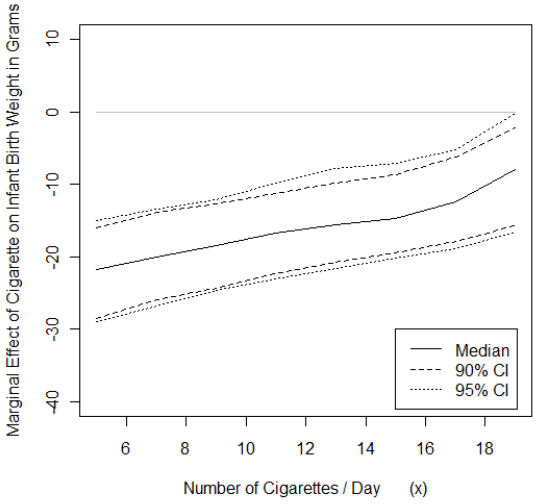


FIGURE 2.6. Confidence intervals of $\mathbb{E}[\beta(X, S, A) \mid Y = 2500, X = x, Z = 0.50, S = \bar{s}]$.

2. Comparing the marginal effects across X, we observe a common tendency for marginal effects to diminish towards $x = 20$ (e.g., Figures 2.3–2.7). That is, the negatively sloped structural function g will eventually flatten on average as x increases. This phenomenon reflects the reduction in harm of an additional cigarette as the number of cigarettes increases, i.e., diminishing marginal effects. It is imperative to

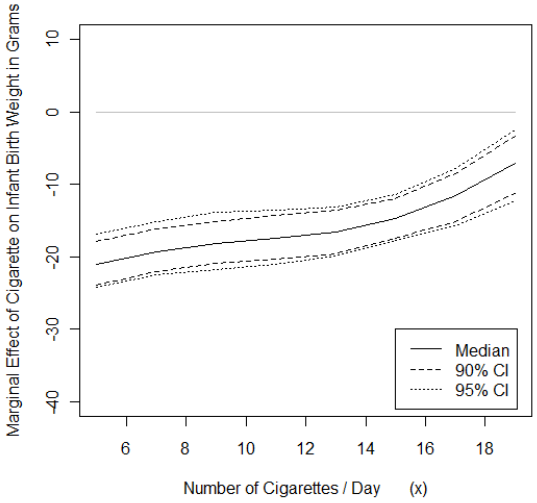


FIGURE 2.7. Confidence intervals of $\mathbb{E}[\beta(X,S,A) \mid Y = 3000, X = x, Z = 0.50, S = \bar{s}]$.

keep in mind, however, that a woman who smoked 20 cigarettes a day has already inflicted a large cumulative effect on her child.

3. Comparing the graphs with different values of Y (e.g., Figures 2.2 and 2.3), we observe some differences in marginal effects across quantiles of Y, especially at lower tax rates $z = 30$. Marginal effects of smoking on birth weights tend to be smaller for lower quantiles of Y. This makes sense as it is more difficult to reduce a birth weight that is already low by the same absolute value (though a similar percentage reduction seems conceivable). These quantile differences are milder at higher tax rates $z \geqslant 40$. However, the differences in Y are not pronounced in this application, and as a consequence it may be justified to focus on the difference across (x, z) by integrating out Y. These effects are illustrated in Figures 2.8–2.10, and they reinforce nicely the observations made in the first two points above.

It is instructive to examine the first point in more detail and provide likely causal explanations. As the graphs indicate, the magnitude of partial effects tends to be negatively related to Z for each fixed value of X. Suppose now that $z' > z$. The subpopulation who smokes x cigarettes when the taxes are z' is then characterized by a higher preference for smoking than the subpopulation that smokes x cigarettes

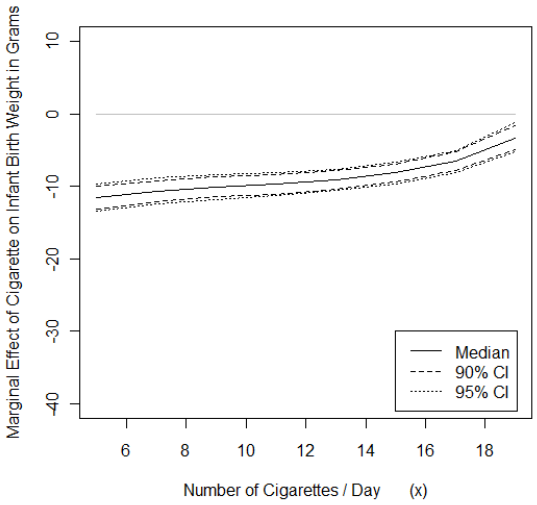


FIGURE 2.8. Confidence intervals of $\mathbb{E}[\beta(X, S, A) \mid X = x, Z = 0.30, S = \bar{s}]$.

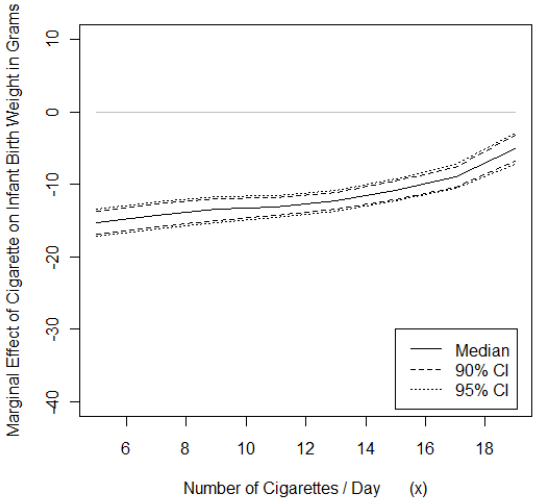


FIGURE 2.9. Confidence intervals of $\mathbb{E}[\beta(X, S, A) \mid X = x, Z = 0.40, S = \bar{s}]$.

at the lower price (tax) z. What causes endogeneity is now precisely the correlation between this preference for smoking and other factors in $\mathcal{E}$, in particular adverse ones, say, a preference for an unhealthy lifestyle, and/or a partner who also smokes. The graphical results imply that the magnitude of partial effects tends to be positively related to higher taxes in excess of the effect already incurred through X, suggesting

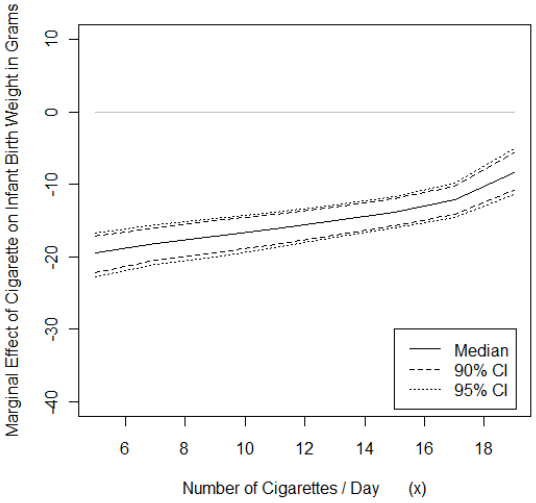


FIGURE 2.10. Confidence intervals of $\mathbb{E}[\beta(X, S, A) \mid X = x, Z = 0.50, S = \bar{s}]$.

this revealed preference for a negative lifestyle as explanation. Moreover, it implies that the magnitude of partial effects tends to be positively related with unhealthy factors in $\mathcal{E}$, other things fixed. Lastly, this implies the cross partial sign

$$0 < \frac{\partial}{\partial \varepsilon} \left| \frac{\partial}{\partial x} g(x, \varepsilon) \right| = -\frac{\partial^2}{\partial \varepsilon \partial x} g(x, \varepsilon).$$

Therefore, smoking X and other unhealthy behavioral inputs $\mathcal{E}$ are likely to be complementary negative inputs in the birth weight "production" function g. So, based on our results, policy should not just discourage smoking, but also the negative and unhealthy life style associated with it that exacerbates its effect.

6. Summary

Economists are generally interested in estimating structural objects such as the structural partial effects $\beta(X, \mathcal{E})$. When $\mathcal{E}$ is multi-dimensional and/or the structural function g is not invertible with respect to $\mathcal{E}$, quantile partial effects generally do not represent this structural partial effects of interest. This paper explored possibilities of identifying local means of the structural partial effects. The main results are summarized in Table 2.2.

We first considered the LIV as a natural extension of the classical two-stage least squares. This method identifies a useful policy parameter, but at the cost of separable first stage assumption, which is necessary as well as sufficient (Theorem 2). The classical idea of "regressing" Y on the predicted values of X remains to work in identifying $\mathrm{E}[\beta(X, \mathcal{E}) \mid Z]$, even if g is nonparametric and nonseparable (Theorem 3 (i)). Moreover, this idea extends to identification of $\mathrm{E}[\beta(X, \mathcal{E}) \mid YZ]$ simply by replacing the mean regression by the corresponding quantile regression (Theorem 3 (ii)).

We next considered identifying the structural partial effects controlling for the endogenous choice variable X. Equation (35) heuristically demonstrated that such a calsal effect can be identified as a result of a slight extension of the LIV concept, given that we have a prior knowledge about how the first-stage model looks like. This heuristic finding was generalized in Theorem 4 (i), showing identification of $\mathrm{E}[\beta(X, \mathcal{E}) \mid XZ]$. Replacing the mean regressions by the corresponding quantile regressions allowed identification of $\mathrm{E}[\beta(X, \mathcal{E}) \mid YXZ]$ (Theorem 4 (ii)). This parallell between parts (i) and (ii) of Theorem 4 resembles that of Theorem 3. These structural partial effects (MTE) have clear economic interpretations when the first-stage model is structurally constructed, as demonstrated in Example 2.

In the absence of structural motivations in the first stage, we can still substitute statistical models such as the mean regression or the quantile regression to represent a first-stage model (Corollaries 3 and 4). The identifying formula in the special case of using quantile regressions to represent the first-stage model (Corollary 4 (ii)) coincides with the well-known formula previously proposed by Chesher (2003) and Imbens and Newey (2009). The identified objects, however, are different from each other.

With the abilities of the identified objects to describe heterogeneous structural partial effects, we studied causal effects of smoking on infant birth weights. Smoking is significantly malignant with diminishing marginal effects. Moreover, these marginal effects tend to be greater for those mothers smoking under higher cigarette excise tax

		Identified Structural Object	Identifying Statistical Object	First-Stage Structural/Statistical Model
LIV	Theorem 2	$\mathrm{E}[\beta(X,\mathcal{E})\|Z]$	Local Two-Stage Least Squares $\frac{d}{dz}\mathrm{E}[Y \mid Z]/\frac{d}{dz}\mathrm{E}[X \mid Z]$	Stochastically Separable Structure $\mathrm{Cov}(\beta(h(z,U),\mathcal{E}),\frac{\partial}{\partial z}h(z,U))=0$
	Theorem 3	(i) $\mathrm{E}[\beta(X,\mathcal{E})\|Z]$	$\frac{d}{dp}\mathrm{E}[Y\|P]$ where $P:=\mu(Z)$	Separable Structure
		(ii) $\mathrm{E}[\beta(X,\mathcal{E})\|YZ]$	$\frac{\partial}{\partial p}Q(Y\|P)$ where $P:=\mu(Z)$	$X=\mu(Z)+U, \quad Z \perp\!\!\!\perp U$
MTE	Theorem 4	(i) $\mathrm{E}[\beta(X,\mathcal{E})\|XZ]$	$\frac{\partial}{\partial x}\mathrm{E}[Y\|XZ]-\frac{\partial}{\partial z}\mathrm{E}[Y\|XZ]\cdot\rho(Z,X)$	General Identifiable Structure (see Example 2)
		(ii) $\mathrm{E}[\beta(X,\mathcal{E})\|YXZ]$	$\frac{\partial}{\partial x}Q(Y\|XZ)-\frac{\partial}{\partial z}Q(Y\|XZ)\cdot\rho(Z,X)$	$X=h(Z,U)$, Z and U may be correlated
	Corollary 3	(i) $\mathrm{E}[\beta(X,\mathcal{E})\|XZ]$	$\frac{\partial}{\partial x}\mathrm{E}[Y\|XZ]+\frac{\partial}{\partial z}\mathrm{E}[Y\|XZ]/\frac{d}{dz}\mathrm{E}[X\|Z]$	Mean Regression
		(ii) $\mathrm{E}[\beta(X,\mathcal{E})\|YXZ]$	$\frac{\partial}{\partial x}Q(Y\|XZ)+\frac{\partial}{\partial z}Q(Y\|XZ)/\frac{d}{dz}\mathrm{E}[X\|Z]$	$X=\mu(Z)+U, \quad \mathrm{E}[U\|Z]=0$
	Corollary 4	(i) $\mathrm{E}[\beta(X,\mathcal{E})\|XZ]$	$\frac{\partial}{\partial x}\mathrm{E}[Y\|XZ]+\frac{\partial}{\partial z}\mathrm{E}[Y\|XZ]/\frac{\partial}{\partial z}Q(X\|Z)$	Quantile Regression
		(ii) $\mathrm{E}[\beta(X,\mathcal{E})\|YXZ]$	$\frac{\partial}{\partial x}Q(Y\|XZ)+\frac{\partial}{\partial z}Q(Y\|XZ)/\frac{d}{dz}Q(X\|Z)$	$X=Q_{X\|Z}(U\|Z), \quad Z \perp\!\!\!\perp U$

TABLE 2.2. Summary of identified structural parameters and the respective first-stage models.

rate. Our inspection of this result implies that smoking and its associated unhealthy life style have complementary negative effects on infant birth weights.

7. Mathematical Appendix

7.1. Proof of Theorem 2.

PROOF. Using the definition (34) of the structural model, we have

$$\mathbb{E}\left[Y|Z=z\right] \;=\; \int\int g(h(z,u),\varepsilon) f_{\mathcal{E}U|Z}(\varepsilon,u \mid z)d\varepsilon du \;=\; \int\int g(h(z,u),\varepsilon) f_{\mathcal{E}U}(\varepsilon,u)d\varepsilon du$$

where the last equality is due to the instrument independence in (34). Taking derivatives on the both sides produces

$$\begin{aligned}
\frac{d}{dz}\mathbb{E}\left[Y|Z=z\right] &= \int\int \beta(h(z,u),\varepsilon)\left[\frac{\partial}{\partial z}h(z,u)\right] f_{\mathcal{E}U}(\varepsilon,u)d\varepsilon du \\
&= \int\int \beta(h(z,u),\varepsilon)\left[\frac{\partial}{\partial z}h(z,u)\right] f_{\mathcal{E}U|Z}(\varepsilon,u \mid z)d\varepsilon du \\
&= \mathrm{E}\left[\beta(X,\mathcal{E})\cdot\frac{\partial}{\partial z}h(Z,U)\middle| Z=z\right] \\
&= \mathrm{E}\left[\beta(X,\mathcal{E})|Z=z\right]\cdot\mathrm{E}\left[\frac{\partial}{\partial z}h(Z,U)\middle| Z=z\right] \\
&\quad +\mathrm{Cov}\left(\beta(X,\mathcal{E}),\ \frac{\partial}{\partial z}h(Z,U)\middle| Z=z\right)
\end{aligned}$$

where the first equality is due to the differentiability of g and h with respect their first arguments as well as the L^1 dominance of the integrand, and the second equality

is again due to the instrument independence in (34). The instrument independence also yields

$$\begin{aligned} \mathrm{E}\left[\left.\frac{\partial}{\partial z}h(Z,U)\right| Z=z\right] &= \frac{d}{dz}\mathrm{E}\left[X\right| Z=z] \qquad \text{and} \\ \mathrm{Cov}\left(\beta(X,\mathcal{E}),\ \left.\frac{\partial}{\partial z}h(Z,U)\right| Z=z\right) &= \mathrm{Cov}\left(\beta(h(z,U),\mathcal{E}),\ \frac{\partial}{\partial z}h(z,U)\right). \end{aligned}$$

Substituting these equalities and rearranging terms under Assumption 6, we obtain

$$\mathrm{E}\left[\beta(X,\mathcal{E})|Z=z\right] = \frac{\frac{d}{dz}\mathbb{E}\left[Y|Z=z\right] - \mathrm{Cov}\left(\beta(h(z,U),\mathcal{E}),\ \frac{\partial}{\partial z}h(z,U)\right)}{\frac{d}{dz}\mathrm{E}\left[X\right| Z=z]}$$

Therefore, the desired equality holds if and only if Assumption 7 is true. □

7.2. Proof of Theorem 3.

Proof. (i) Using Assumption 8, we can write

$$\begin{aligned} \frac{d}{dp}\mathrm{E}[Y \mid P=p]\Big|_{p=\mu(z)} &= \frac{d}{dp}\int\int g(p+u,\varepsilon) f_{\mathcal{E}U|P}(\varepsilon,u \mid p)\Big|_{p=\mu(z)} \\ &= \frac{d}{dp}\int\int g(p+u,\varepsilon) f_{\mathcal{E}U}(\varepsilon,u)\Big|_{p=\mu(z)} \\ &= \int\int \beta(\mu(z)+u,\varepsilon) f_{\mathcal{E}U}(\varepsilon,u) \\ &= \int\int \beta(\mu(z)+u,\varepsilon) f_{\mathcal{E}U|Z}(\varepsilon,u \mid z) \ = \ \mathrm{E}[\beta(X,\mathcal{E}) \mid Z=z], \end{aligned}$$

where the second and fourth equalities are due to the instrumental independence in the model (34), and the third equality is due to the differentiability of g with respect to x as well as the L^1-domination of β.

(ii) We derive the following three auxiliary equations. First,

$$\begin{aligned} & \Pr[g(X,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p+\delta) \mid P=p+\delta] - \Pr[g(X,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p) \mid P=p+\delta] \\ = \ & F_{Y|P}(Q_{Y|P}(\tau \mid p+\delta) \mid p+\delta) - F_{Y|P}(Q_{Y|P}(\tau \mid p) \mid p+\delta) \\ (38)= \ & \delta\left[\frac{\partial}{\partial p}Q_{Y|P}(\tau \mid p)\right] f_{Y|P}(Q_{Y|P}(\tau \mid p) \mid p+\delta) + o(\delta) \end{aligned}$$

holds under the differentiability of $F_{Y|P}$ and $Q_{Y|P}$ with respect to y and p, respectively. Second,

$$\begin{aligned}
& \Pr[g(X,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p) \mid P = p+\delta] - \Pr[g(X+\delta,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p) \mid P = p] \\
= \;& \Pr[g(p+\delta+U,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p) \mid P = p+\delta] - \Pr[g(p+\delta+U,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p) \mid P = p] \\
(39) = \;& \Pr[g(p+\delta+U,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p)] - \Pr[g(p+\delta+U,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p)] \;=\; 0,
\end{aligned}$$

where the second equality is due to the instrumental independence in the model (34). Third, using the short-hand notation $B = \beta(X,\mathcal{E})$, we have

$$\begin{aligned}
& \Pr[g(X+\delta,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p) \mid P = p] - \Pr[g(X,\mathcal{E}) \leqslant Q_{Y|P}(\tau \mid p) \mid P = p] \\
= \;& \Pr[Q_{Y|P}(\tau \mid p) < Y \leqslant Q_{Y|P}(\tau \mid p) - (g(X+\delta,\mathcal{E}) - Y) \mid P = p] \\
& -\Pr[Q_{Y|P}(\tau \mid p) - (g(X+\delta,p) - Y) < Y \leqslant Q_{Y|P}(\tau \mid p) \mid P = p] \\
= \;& \Pr[Q_{Y|P}(\tau \mid p) \leqslant Y \leqslant Q_{Y|P}(\tau \mid p) - \delta B \mid P = p] \\
& -\Pr[Q_{Y|P}(\tau \mid p) - \delta B \leqslant Y \leqslant Q_{Y|P}(\tau \mid p) \mid P = p] + o(\delta) \\
= \;& \int_{Q_{Y|P}(\tau|p)}^{\infty} \int_{-\infty}^{-\delta^{-1}[y - Q_{Y|P}(\tau|p)]} f_{YB|P}(y,b \mid p) db dy \\
& - \int_{-\infty}^{Q_{Y|P}(\tau|p)} \int_{-\delta^{-1}[y-Q_{Y|P}(\tau|p)]}^{\infty} f_{YB|P}(y,b \mid p) db dy + o(\delta) \\
= \;& -\delta \int_{-\infty}^{0} b f_{YB|P}(Q_{Y|P}(\tau \mid p), b \mid p) db - \delta \int_{0}^{\infty} b f_{YB|P}(Q_{Y|P}(\tau \mid p), b \mid p) db + o(\delta) \\
(40) = \;& -\delta \mathbb{E}[B \mid Y = Q_{Y|P}(\tau \mid p), P = p] \cdot f_{Y|P}(Q_{Y|P}(\tau \mid p) \mid p) + o(\delta),
\end{aligned}$$

where the second equality is due to the differentiability of g and $F_{Y|P}$ with respect to x and y, respectively, and the fourth equality is due to change of variables and integration by parts. Add these three equations (38), (39) and (40) together to get

$$\begin{aligned}
0 \;=\; & \delta \left[\frac{\partial}{\partial p} Q_{Y|P}(\tau \mid p) \right] \cdot f_{Y|P}(Q_{Y|P}(\tau \mid p) \mid p+\delta) \\
& - \delta \mathbb{E}[B \mid Y = Q_{Y|P}(\tau \mid p), P = p] \cdot f_{Y|P}(Q_{Y|P}(\tau \mid p) \mid p) + o(\delta).
\end{aligned}$$

Under the condition that $f_{Y|P}$ is continuous in p, letting $\delta \to 0$ yields

$$\mathrm{E}[\beta(X,\mathcal{E}) \mid Y = Q_{Y|P}(\tau \mid p),\; P = p] \;=\; \frac{\partial}{\partial p} Q_{Y|P}(\tau \mid p).$$

Since $Z = z$ and $P = \mu(z)$ are the same events under Assumption 6, setting $p = \mu(z)$ yields the result. □

7.3. Proof of Theorem 4.

Proof. (i) We derive the following two auxiliary equations. First,

$$\begin{aligned}\frac{\partial}{\partial x}\mathbb{E}[Y \mid X = x, Z = z] &= \partial_x \int g(x,\varepsilon) f_{\mathcal{E}|XZ}(\varepsilon \mid x, z) d\varepsilon \;=\; \frac{\partial}{\partial x}\int g(x,\varepsilon) f_{\mathcal{E}|UZ}(\varepsilon \mid \upsilon(z,x), z) d\varepsilon \\ &= \mathbb{E}[\beta(X,\mathcal{E}) \mid U = \upsilon(z,x), Z = z] \\ &+ \frac{\partial}{\partial x}\upsilon(z,x) \cdot \mathbb{E}\left[g(X,\mathcal{E}) \frac{\partial}{\partial u} \log f_{\mathcal{E}|UZ}(\mathcal{E} \mid U, Z) \middle| U = \upsilon(z,x), Z = z \right],\end{aligned}$$

where the second equality is due to Assumption 9 and the third equality is due to the differentiability of g with respect to x as well as the L^1 dominance of the integrand. Second, a similar calculation yields

$$\frac{\partial}{\partial z}\mathbb{E}[Y \mid X = x, Z = z] = \frac{\partial}{\partial z}\upsilon(z,x) \cdot \mathbb{E}\left[g(X,\mathcal{E}) \frac{\partial}{\partial u} \log f_{\mathcal{E}|UZ}(\mathcal{E} \mid U, Z) \middle| U = \upsilon(z,x), Z = z \right],$$

where the instrumental independence in model (34) was used to vanish $\frac{\partial}{\partial z} f_{\mathcal{E}|UZ}$. Combining the above two equations and rearranging by Assumption 10 yield the desired result.

(ii) Assumptions 9 and 10 provide the parameterized curve $h \mapsto (h, \delta_z(h)) =: (\delta_x, \delta_z)$ that solves the implicit function equation $\upsilon(z + \delta_z, x + \delta_x) - \upsilon(z,x) = 0$ of a smooth submanifold in a neighborhood of $h = 0$. Furthermore, $\delta_z(0) = 0$ and $(\delta_x, \delta_z) \to 0$ as $h \to 0$. By these properties, we have

$$\frac{\delta_z}{\delta_x} = -\frac{\frac{\partial}{\partial x}\upsilon(z,x)}{\frac{\partial}{\partial z}\upsilon(z,x)} + o(1) \quad \text{as } h \to 0. \tag{41}$$

Next, we derive the following four auxiliary equations. First,

$$\begin{aligned}&\Pr[g(x+\delta_x, \mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x + \delta_x, z + \delta_z) \mid X = x + \delta_x, Z = z + \delta_z] \\ &-\Pr[g(x+\delta_x, \mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z + \delta_z) \mid X = x + \delta_x, Z = z + \delta_z] \\ =\; &F_{Y|XZ}(Q_{Y|XZ}(\tau \mid x + \delta_x, z + \delta_z) \mid x + \delta_x, z + \delta_z) \\ &-F_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z + \delta_z) \mid x + \delta_x, z + \delta_z) \\ =\; &\delta_x \frac{\partial}{\partial x} Q_{Y|XZ}(\tau \mid x, z) f_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z) \mid x + \delta_x, z + \delta_z) + o(\delta_x),\end{aligned} \tag{42}$$

where the last equality is due to the differentiability of $Q_{Y|XZ}$ and $F_{Y|XZ}$ with respect to x and y, respectively. Second, similar lines of calculations yield

$$\begin{aligned}
& \Pr[g(x+\delta_x,\mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z+\delta_z) \mid X = x+\delta_x, Z = z+\delta_z] \\
& -\Pr[g(x+\delta_x,\mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid X = x+\delta_x, Z = z+\delta_z] \\
= \; & F_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z+\delta_z) \mid x+\delta_x, z+\delta_z) \\
& -F_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z) \mid x+\delta_x, z+\delta_z) \\
= \; & \delta_z \frac{\partial}{\partial z} Q_{Y|XZ}(\tau \mid x, z) f_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z) \mid x+\delta_x, z+\delta_z) + o(\delta_z), \qquad (43)
\end{aligned}$$

under the differentiability of $Q_{Y|XZ}$ with respect to z. Third,

$$\begin{aligned}
& \Pr[g(x+\delta_x,\mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid X = x+\delta_x, Z = z+\delta_z] \\
& -\Pr[g(x+\delta_x,\mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid X = x, Z = z] \\
= \; & \Pr[g(x+\delta_x,\mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid Z = z+\delta_z, U = \upsilon(z+\delta_z, x+\delta_x)] \\
& -\Pr[g(x+\delta_x,\mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid Z = z, U = \upsilon(z, x)] \\
= \; & \Pr[g(x+\delta_x,\mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid Z = z+\delta_z, U = \upsilon(z, x)] \\
& -\Pr[g(x+\delta_x,\mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid Z = z, U = \upsilon(z, x)] \; = \; 0, \qquad (44)
\end{aligned}$$

where the first equality is due to Assumption 9, the second equality is due to the definition of (δ_x, δ_z), and the last equality is due to the instrument independence in

the model (34). Fourth, with the short-hand notation $B := \beta(X, \mathcal{E})$, we have

$$
\begin{aligned}
& \Pr[g(x+\delta_x, \mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid X = x, Z = z] \\
& -\Pr[g(x, \mathcal{E}) \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid X = x, Z = z] \\
= \; & \Pr[Q_{Y|XZ}(\tau \mid x, z) < Y \leqslant Q_{Y|XZ}(\tau \mid x, z) - (g(x+\delta_x, \mathcal{E}) - Y) \mid X = x, Z = z] \\
& -\Pr[Q_{Y|XZ}(\tau \mid x, z) - (g(x+\delta_x, z) - Y) < Y \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid X = x, Z = z] \\
= \; & \Pr[Q_{Y|XZ}(\tau \mid x, z) \leqslant Y \leqslant Q_{Y|XZ}(\tau \mid x, z) - \delta_x B \mid X = x, Z = z] \\
& -\Pr[Q_{Y|XZ}(\tau \mid x, z) - \delta_x B \leqslant Y \leqslant Q_{Y|XZ}(\tau \mid x, z) \mid X = x, Z = z] + o(\delta_x) \\
= \; & \int_{Q_{Y|XZ}(\tau|x,z)}^{\infty} \int_{-\infty}^{-\delta_x^{-1}[y - Q_{Y|XZ}(\tau|x,z)]} f_{YB|XZ}(y, b \mid x, z) db dy \\
& - \int_{-\infty}^{Q_{Y|XZ}(\tau|x,z)} \int_{-\delta_x^{-1}[y - Q_{Y|XZ}(\tau|x,z)]}^{\infty} f_{YB|XZ}(y, b \mid x, z) db dy + o(\delta_x) \\
= \; & -\delta_x \int_{-\infty}^{0} b f_{YB|XZ}(Q_{Y|XZ}(\tau \mid x, z), b \mid x, z) db \\
& -\delta_x \int_{0}^{\infty} b f_{YB|XZ}(Q_{Y|XZ}(\tau \mid x, z), b \mid x, z) db + o(\delta_x) \\
(45) \; = \; & -\delta_x \mathbb{E}[B \mid Y = Q_{Y|XZ}(\tau \mid x, z), X = x, Z = z] f_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z) \mid x, z) + o(\delta_x),
\end{aligned}
$$

where the second equality is due to the differentiability of g and $F_{Y|XZ}$ with respect to x and y, respectively, and the fourth equality is due to change of variables and integration by parts. Add the above four equations (42), (43), (44) and (45) together to get

$$
\begin{aligned}
0 \; = \; & \delta_x \frac{\partial}{\partial x} Q_{Y|XZ}(\tau \mid x, z) f_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z) \mid x + \delta_x, z + \delta_z) \\
& + \delta_z \frac{\partial}{\partial z} Q_{Y|XZ}(\tau \mid x, z) f_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z) \mid x + \delta_x, z + \delta_z) \\
& - \delta_x \mathbb{E}[B \mid Y = Q_{Y|XZ}(\tau \mid x, z), X = x, Z = z] f_{Y|XZ}(Q_{Y|XZ}(\tau \mid x, z) \mid x, z) \\
& + o(\delta_x) + o(\delta_z).
\end{aligned}
$$

The desired result follows from this equation together with Equation (41), Assumption 9, and the differentiability of $f_{Y|XZ}$ with respect to the conditioning variables (x, z). □

7.4. Proof of Proposition 3. We prove the following four auxiliary lemmas to prove Proposition 3.

LEMMA 14. *Suppose that $F_{X|Z}(\cdot \mid z)$ is strictly increasing in the model (34). Then, Assumption 11 holds if and only if $F_{X|Z}(h(z,u) \mid z) = F_{X|Z}(h(z',u) \mid z')$ holds for all z, z' and u.*

PROOF. ($\Rightarrow$) Suppose that $h(z,u) = \bar{h}(z,\iota(u))$ holds for all (z,u), where $\bar{h}$ is strictly increasing in its second argument. Then, $F_{X|Z}(h(z,u) \mid z) = F_{X|Z}(\bar{h}(z,\iota(u)) \mid z) = \Pr(\bar{h}(z,\iota(U)) \leqslant \bar{h}(z,\iota(u)) \mid Z = z) \overset{(1)}{=} \Pr(\iota(U) \leqslant \iota(u) \mid Z = z) \overset{(2)}{=} \Pr(\iota(U) \leqslant \iota(u) \mid Z = z') \overset{(1)}{=} \Pr(\bar{h}(z',\iota(U)) \leqslant \bar{h}(z',\iota(u)) \mid Z = z') = F_{X|Z}(\bar{h}(z',\iota(u)) \mid z') = F_{X|Z}(h(z',u) \mid z')$, where equalities (1) are due to the premise that $\bar{h}$ is strictly increasing in its second argument, and equality (2) is due to the instrument independence in (34).

($\Leftarrow$) Let z^0 be an arbitrarily chosen element of the support of Z. Define $\iota : \mathbb{R}^M \to \mathbb{R}$ by $\iota(u) = F_{X|Z}(h(z^0,u) \mid z^0)$. For each $\eta \in (0,1)$, there exists $\xi(\eta)$ on the support of $X \mid (Z = z^0)$ such that $\eta = F_{X|Z}(\xi(\eta) \mid z^0)$. But then, for this $\xi(\eta)$ as an element of the support of $X \mid (Z = z^0)$, there exists $\upsilon(\eta)$ on the support of U (not necessarily unique) such that $\xi(\eta) = h(z^0,\upsilon(\eta))$, i.e., $\eta = F_{X|Z}(h(z^0,\upsilon(\eta)) \mid z^0)$.

As such an element $\upsilon(\eta)$ is not generally unique for a given η, the map $(z,\eta) \overset{\bar{h}}{\mapsto} h(z,\upsilon(\eta))$ need not be well-defined. However, we will show that such a map $\bar{h}$ is indeed well-defined if $F_{X|Z}(h(z,u) \mid z) = F_{X|Z}(h(z',u) \mid z')$ holds for all z, z' and u. To show its well-definition, consider $\upsilon(\eta)$ and $\tilde{\upsilon}(\eta)$ such that $\xi(\eta) = h(z^0,\upsilon(\eta)) = h(z^0,\tilde{\upsilon}(\eta))$. Note that $F_{X|Z}(h(z,\upsilon(\eta)) \mid z) \overset{(*)}{=} F_{X|Z}(h(z^0,\upsilon(\eta)) \mid z^0) = F_{X|Z}(\xi(\eta) \mid z^0) = F_{X|Z}(h(z^0,\tilde{\upsilon}(\eta)) \mid z^0) \overset{(*)}{=} F_{X|Z}(h(z,\tilde{\upsilon}(\eta)) \mid z)$ holds where equalities $(*)$ are due to $F_{X|Z}(h(z,u) \mid z) = F_{X|Z}(h(z',u) \mid z')$. This implies that $h(z,\upsilon(\eta)) = h(z,\tilde{\upsilon}(\eta))$ by the assumption that $F_{X|Z}(\cdot \mid z)$ is strictly increasing. Therefore, the mapping $\bar{h} : \mathbb{R}^2 \to \mathbb{R}$ defined by $(z,\eta) \overset{\bar{h}}{\mapsto} h(z,\upsilon(\eta))$ is indeed well-defined.

Having proven the well-definition of $\bar{h}$, we next show that $h(z,u) = \bar{h}(z,\iota(u))$ holds for all (z,u). Observe $F_{X|Z}(\bar{h}(z,\iota(u)) \mid z) \overset{(1)}{=} F_{X|Z}(h(z,\upsilon(\iota(u))) \mid z) \overset{(2)}{=} F_{X|Z}(h(z^0,\upsilon(\iota(u))) \mid z^0) \overset{(3)}{=} \iota(u) \overset{(4)}{=} F_{X|Z}(h(z^0,u) \mid z^0) \overset{(2)}{=} F_{X|Z}(h(z,u) \mid z)$, where equality (1) is due to the definition of the well-defined map $\bar{h}$, equalities (2) are due to $F_{X|Z}(h(z,u) \mid z) = F_{X|Z}(h(z',u) \mid z')$, equality (3) is due to the definition of υ,

and equality (4) is due to the definition of ι. This implies $\bar{h}(z, \iota(u)) = h(z, u)$ for each (z, u) by the assumption that $F_{X|Z}(\cdot \mid z)$ is strictly increasing.

It remains to show that $\bar{h}$ constructed in this way is strictly monotone in its second argument. To see this, let $0 < \eta_1 < \eta_2 < 1$. Then, $F_{X|Z}(\bar{h}(z, \eta_1) \mid z) \overset{(1)}{=} F_{X|Z}(h(z, \upsilon(\eta_1)) \mid z) \overset{(2)}{=} F_{X|Z}(h(z^0, \upsilon(\eta_1)) \mid z^0) \overset{(3)}{=} \eta_1 < \eta_2 \overset{(3)}{=} F_{X|Z}(h(z^0, \upsilon(\eta_2)) \mid z^0) \overset{(2)}{=} F_{X|Z}(h(z, \upsilon(\eta_2)) \mid z) \overset{(1)}{=} F_{X|Z}(\bar{h}(z, \eta_2) \mid z)$, where equalities (1) are due to the definition of the well-defined map $\bar{h}$, equalities (2) are due to $F_{X|Z}(h(z, u) \mid z) = F_{X|Z}(h(z', u) \mid z')$, and equalities (3) are due to the definition of υ. It follows from this inequality that $\bar{h}(z, \eta_1) < \bar{h}(z, \eta_2)$ by the assumption that $F_{X|Z}(\cdot \mid z)$ is strictly increasing. Therefore $\bar{h}$ is strictly increasing in its second argument. □

LEMMA 15. *Suppose that $F_{X|Z}(\cdot \mid z)$ is strictly increasing.*[8] *If Assumption 11 holds and $\frac{\partial}{\partial z} Q_{X|Z}(v \mid z) \neq 0$, then for the choice of $c := [\frac{\partial}{\partial z} Q_{X|Z}(v \mid z)]^{-1}$,*

$$\frac{\partial}{\partial x} \log f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) + c \frac{\partial}{\partial z} \log f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) = 0$$

holds for all u on the support of $F_{U|XZ}(\ \cdot\ \mid Q_{X|Z}(v \mid z), z)$.

PROOF. Let $V := F_{X|Z}(h(Z, U) \mid Z)$. Then, Assumption 11 and Lemma 14 imply that V does not depend on Z, thus $V = \nu(U)$ for some function ν. Since $(U, V) = (U, \nu(U))$, we have $(U, V) \perp\!\!\!\perp Z$ by the instrument independence in the model (34). Using this independence restriction yields $F_{UZ|V} = \frac{F_{UV|Z}}{F_V} F_Z = \frac{F_{UV}}{F_V} F_Z = F_{U|V} F_Z = F_{U|V} F_{Z|V}$, showing that $U \perp\!\!\!\perp Z \mid V$.

Now, note that the map $(v, z) \mapsto (Q_{X|Z}(v \mid z), z)$ is well-defined and injective, owing to the well-definition and injectivity of the map $v \mapsto Q_{X|Z}(v \mid z)$ by the absolute continuity of $F_{X|Z}$ (note that the absolute continuity of $F_{X|Z}$ in particular implies that there is no singular part in its Lebesgue-Radon-Nikodym decomposition, thus the quantile is strictly increasing in v). Therefore, we have $f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) = f_{U|VZ}(u \mid v, z)$ for all z and v in their respective domains. Finally, use the

[8] In addition, we also assume the following regularity conditions: $F_{X|Z}$ is absolutely continuous; $Q_{X|Z}$ is continuously differentiable with respect to z; and $f_{U|XZ}$ is continuously differentiable with respect to (x, z)

independence condition $U \perp\!\!\!\perp Z \mid V$ obtained in the last paragraph to conclude that $f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) = f_{U|V}(u \mid v)$, which is constant in z.

Since $f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z)$ is constant in z, we have

$$\begin{aligned} 0 &= \frac{d}{dz} f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) \\ &= \frac{\partial}{\partial z} Q_{X|Z}(v \mid z) \cdot \frac{\partial}{\partial x} f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) + \frac{\partial}{\partial z} f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) \end{aligned}$$

under the differentiability of $Q_{X|Z}$ with respect to z and the differentiability of $f_{U|XZ}$ with respect to (x, z). Divide this equation by $[\frac{\partial}{\partial z} Q_{X|Z}(v \mid z)] \cdot f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z)$ to prove the lemma. □

LEMMA 16. *Suppose that $F_{X|Z}(\cdot \mid z)$ is strictly increasing.*[9] *If Assumption 11 does not hold, then there exists a set $\bar{\mathcal{U}} \subset \mathcal{U}$ of positive measure such that*

$$\frac{d}{dz} f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) \neq 0. \tag{46}$$

holds for some $z \in \mathcal{Z}$ and for all $u \in \bar{\mathcal{U}}$.

PROOF. Let $V := F_{X|Z}(h(Z, U) \mid Z)$. Write $H(z, u) := F_{X|Z}(h(z, u) \mid z)$. The differentiability of h and $F_{X|Z}$ imply $H \in C^1(\mathbb{R}^{M+1}, \mathbb{R})$ where M is the dimension of U. As Assumption 11 does not hold, we have $\nabla_z H(\bar{z}, \bar{u}) \neq 0$ for some $(\bar{z}, \bar{u}) \in \mathcal{Z} \times \mathcal{U}$. Let j be a coordinate of U in h satisfying $dh(z, u)/du_j \neq 0$ at $(\bar{z}, \bar{u})$. Then, we have $\nabla_{u_j} H(\bar{z}, \bar{u}) \neq 0$. Thus we have sufficient conditions to invoke the Implicit Function Theorem to obtain a continuous function $\lambda : \mathcal{Z} \supset \mathcal{B}_\delta(\bar{z}) \to \mathcal{U}$ defined in a neighborhood of $\bar{z}$ such that $H(z, \lambda(z)) = H(\bar{z}, \bar{u}) =: \bar{v}$. It follows that a continuum of the level set of $V = \bar{v}$ exist in a neighborhood of $z = \bar{z}$. But this level set does not contain arbitrarily close horizontal or vertical displacements $(z \pm \delta, u)$ and $(z, u \pm \delta e_j)$, due to $\nabla_z H(\bar{z}, \bar{u}) \neq 0$ and $\nabla_{u_j} H(\bar{z}, \bar{u}) \neq 0$. These two facts (i.e., existence of a continuum of the level set and no containment of arbitrarily close horizontal or vertical displacements) imply that $\text{supp}[(Z, U_j) \mid V = \bar{v}] \neq \text{supp}[Z \mid V = \bar{v}] \times \text{supp}[U_j \mid V =$

[9] In addition, we also assume the following regularity conditions: h is continuously differentiable; $F_{X|Z}$ is continuously differentiable; $Q_{X|Z}$ is continuously differentiable with respect to z; and $f_{U|XZ}$ is continuously differentiable with respect to (x, z).

$\bar{v}]$, i.e., the support of $(Z, U_j) \mid (V = \bar{v})$ is not rectangular. Since a rectangular support of the joint distribution is a necessary condition for independence, this implies that $Z \perp\!\!\!\perp U_j \mid (V = \bar{v})$ does not hold, which in turn implies that $Z \perp\!\!\!\perp U \mid (V = \bar{v})$ does not hold.

Keeping the last result in mind, we now want to prove that (46) holds for some $z \in \mathcal{Z}$ and for all $u \in \bar{\mathcal{U}}$ with $\bar{\mathcal{U}}$ a set of positive measure. But by the assumptions of continuous differentiability of $f_{U|XZ}$ and $Q_{X|Z}$, it suffices to show that (46) holds for some $(z, u) \in \mathcal{Z} \times \mathcal{U}$, since the continuity of the derivatives then yields the corresponding result throughout a neighborhood of such u. Suppose, by way of contradiction, that $\frac{d}{dz} f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) = 0$ holds for all $(z, u) \in \mathcal{Z} \times \mathcal{U}$. As in the proof of Lemma 15, $f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) = f_{U|VZ}(u \mid v, z)$. Hence, $\frac{d}{dz} f_{U|VZ}(u \mid v, z) = 0$ holds for all $(z, u) \in \mathcal{Z} \times \mathcal{U}$, showing that $Z \perp\!\!\!\perp U \mid V$. This is a contradiction with the conclusion of the previous paragraph. □

LEMMA 17. *(i) Suppose that the set of structural models satisfies (34).*[10] *Then,*

$$\mathbb{E}\left[\beta(X, \mathcal{E}) | X = x, Z = z\right] = \frac{\partial}{\partial x}\mathbb{E}\left[Y | X = x, Z = z\right] + c\frac{\partial}{\partial z}\mathbb{E}\left[Y | X = x, Z = z\right] - B(c, x, z)$$

holds for any $c \in \mathbb{R}$, where the bias term is

$$B(c, x, z) = \mathbb{E}\left[Y \left\{ \frac{\partial}{\partial x} \log f_{U|XZ}(U|x, z) + c\frac{\partial}{\partial z} \log f_{U|XZ}(U|x, z) \right\} | X = x, Z = z \right].$$

(ii) If in addition $v = F_{X|Z}(x \mid z)$ and $\frac{\partial}{\partial z} Q_{X|Z}(v \mid z) \neq 0$, then Assumption 11 is sufficient to make $B([\frac{\partial}{\partial z} Q_{X|Z}(v \mid z)]^{-1}, Q_{X|Z}(v \mid z), z) = 0$ for all structural models $(g, F_{\mathcal{E}|U}) \in \mathcal{G} \times \mathcal{F}$.

(iii) Assumption 11 is also necessary to make $B([\frac{\partial}{\partial z} Q_{X|Z}(v \mid z)]^{-1}, Q_{X|Z}(v \mid z), z) = 0$ for all structural models $(g, F_{\mathcal{E}|U}) \in \mathcal{G} \times \mathcal{F}$.

[10] In addition, we also assume the following regularity assumptions: g is continuously differentiable with respect to x; and $f_{U|XZ}$ is continuously differentiable with respect to (x, z).

PROOF. We write the observable mean regression $\mathbb{E}[Y|X=x,Z=z]$ as

$$\begin{aligned}\mathbb{E}[Y|X=x,Z=z] &= \int\int g(x,\varepsilon)f_{\mathcal{E}|UXZ}(\varepsilon \mid u,x,z)d\varepsilon f_{U|XZ}(u \mid x,z)du \\ &= \int\int g(x,\varepsilon)f_{\mathcal{E}|U}(\varepsilon \mid u)d\varepsilon f_{U|XZ}(u \mid x,z)du,\end{aligned}$$

where the last equality is due to the instrumental independence in (34). Take derivatives to obtain

$$\begin{aligned}&\frac{\partial}{\partial x}\mathbb{E}[Y|X=x,Z=z] \\ =\ &\int\int \beta(x,\varepsilon)f_{\mathcal{E}|U}(\varepsilon \mid u)d\varepsilon f_{U|XZ}(u \mid x,z)du + \int\int g(x,\varepsilon)f_{\mathcal{E}|U}(\varepsilon \mid u)d\varepsilon \frac{\partial}{\partial x}f_{U|XZ}(u \mid x,z)du \\ =\ &\mathbb{E}[\beta(x,\mathcal{E})|X=x,Z=z] + \mathbb{E}\left[Y\frac{\partial}{\partial x}\log f_{U|XZ}(U \mid x,z)|X=x,Z=z\right],\end{aligned}$$

where the first equality is due to the differentiability of g and $f_{U|XZ}$ with respect to x as well as the L^1 dominance of the integrand. By similar calculations, we have

$$c\frac{\partial}{\partial z}\mathbb{E}[Y|X=x,Z=z] = \mathbb{E}\left[Yc\frac{\partial}{\partial z}\log f_{U|XZ}(U \mid x,z)|X=x,Z=z\right].$$

Combining these two inequalities yields

$$\mathbb{E}[\beta(X,\mathcal{E})|X=x,Z=z] = \frac{\partial}{\partial x}\mathbb{E}[Y|X=x,Z=z] + c\frac{\partial}{\partial z}\mathbb{E}[Y|X=x,Z=z] - B(c,x,z),$$

where the bias term is

$$B(c,x,z) = \mathbb{E}\left[Y\left\{c\frac{\partial}{\partial z}\log f_{U|XZ}(U|x,z) + \frac{\partial}{\partial x}\log f_{U|XZ}(U|x,z)\right\}|X=x,Z=z\right].$$

This proves part (i) of the theorem. Apply Lemma 15 to prove part (ii).

Lastly, we prove part (iii) of the theorem by applying Lemma 16. We prove the contrapositive statement, that if Assumption11 does not hold then there exists a structural model $(g,F_{\mathcal{E}|U}) \in \mathcal{G}\times\mathcal{F}$ such that $B([\nabla_z Q_{X|Z}(v \mid z)]^{-1}, Q_{X|Z}(v \mid z), z) \neq 0$. By Lemma 16, there exists a set $\bar{\mathcal{U}} \subset \mathcal{U}$ of positive measure such that

$$\frac{d}{dz}f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) \neq 0.$$

holds for some $z \in \mathcal{Z}$ and for all $u \in \bar{\mathcal{U}}$. There exists a subset $\tilde{\mathcal{U}}$ of $\bar{\mathcal{U}}$ with positive measure on which the sign is positive or negative throughout. Without loss of generality, assume that the above expression is positive on $\tilde{\mathcal{U}}$. Pick a structure

$(g, F_{\mathcal{E}|U}) \in \mathcal{G} \times \mathcal{F}$ such that $\int g(x,\varepsilon) f_{\mathcal{E}|U}(\varepsilon \mid u) d\varepsilon$ is positive on $\tilde{\mathcal{U}}$ and zero outside $\tilde{\mathcal{U}}$ for some x.[11] With this choice of $(g, F_{\mathcal{E}|U})$, we have

$$
\begin{aligned}
& \frac{\partial}{\partial z} Q_{X|Z}(v \mid z) B([\frac{\partial}{\partial z} Q_{X|Z}(v \mid z)]^{-1}, Q_{X|Z}(v \mid z), z) \\
= \; & \mathbb{E}\left[Y \left\{ \frac{\partial}{\partial z} \log f_{U|XZ}(U \mid x, z) + \frac{\partial}{\partial z} Q_{X|Z}(v \mid z) \cdot \frac{\partial}{\partial x} \log f_{U|XZ}(U \mid x, z) \right\} \mid X = x, Z = z \right] \\
= \; & \int \left[\int g(x,\varepsilon) f_{\mathcal{E}|U}(\varepsilon \mid u) d\varepsilon \right] \left[\frac{d}{dz} f_{U|XZ}(u \mid Q_{X|Z}(v \mid z), z) \right] du \; > \; 0.
\end{aligned}
$$

This shows that $B([\nabla_z Q_{X|Z}(v \mid z)]^{-1}, Q_{X|Z}(v \mid z), z) \neq 0$. □

Proposition 3 concerning the identifying equality of Corollary 4 (i) follows from this lemma. The conclusion concerning the identifying equality of Corollary 4 (ii) follows from similar lines of argument. □

[11] There exist many such structures $(g, F_{\mathcal{E}|U})$. As one example of a way to construct such a structure, consider an orthonormal basis $\mathcal{B}$ of the $L^2(m)$ space. Form a function ϕ and an indexed family $\left\{ f_{\mathcal{E}|U}(\,\cdot \mid u) \right\}_{u \in \mathcal{U}}$ of density functions by linear combinations of $\mathcal{B}$ so that $\left\langle \phi, f_{\mathcal{E}|U}(\,\cdot \mid u) \right\rangle > 0$ for all $u \in \tilde{\mathcal{U}}$ and $\left\langle \phi, f_{\mathcal{E}|U}(\,\cdot \mid u) \right\rangle = 0$ for all $u \in \mathcal{U} \backslash \tilde{\mathcal{U}}$ by applying the Parseval-Bessel equality. Then, the required property is satisfied by any pair $(g, F_{\mathcal{E}|U})$ with any function g such that $g(x, \cdot\,) = \phi$ for some x. Note g is required to be continuously differentiable only with respect to x, and thus will not be violated by this construction of g.

CHAPTER 3

Nonparametric Model Tests with Discrete Instruments

1. Introduction

The following is a list of four common micro-econometric models:

$$\begin{aligned}
&\text{(i)} && Y_i = \beta S_i + \varepsilon(A_i) && \text{– Constant coefficient model}\\
&\text{(ii)} && Y_i = \phi(S_i) + \varepsilon(A_i) && \text{– Nonlinear separable model}\\
&\text{(iii)} && Y_i = \beta(A_i) S_i + \varepsilon(A_i) && \text{– Random coefficient model}\\
&\text{(iv)} && Y_i = \phi(S_i, A_i) && \text{– Nonlinear nonseparable model}
\end{aligned}$$

where Y_i, S_i, and A_i denote observed outcome, observed endogenous choice, and unobserved heterogeneity, respectively. Finding out the correct model reveals the underlying economic structure. For example, if we find that model (i) or (ii) is correct, then we can deduce that variations in first-stage choice come from preferences or costs, rather than from heterogeneity in marginal returns. Moreover, finding out the correct model allows us to choose the correct specification on which to conduct statistical inferences.[1] Therefore, there are both economic and econometric reasons why one may be interested in distinguishing these four types of models. This paper proposes a practically feasible method of testing to this end.

In the ideal setting where a continuous instrument induces smooth first-stage effects to construct a continuous control variable, the existing methods of nonparametric inference would achieve this objective. However, this ideal setting is not often the case, as one can see in the survey by Angrist and Krueger (2001) and Card (2001).

[1] Various identification and estimation theories have been proposed. Model (ii) is studied by Blundell and Powell (2003), Florens (2003), Newey and Powell (2003), Hall and Horowitz (2005), Darolles et al. (2011), and Horowitz (2011). Model (iii) is studied by Garen (1984) and Heckman and Vytlacil (1998), and the associated IV quantile regression has been studied by Ma and Koenker (2006), Blundell and Powell (2007), Lee (2007), and Jun (2009). Model (iv) is studied by Chesher (2003, 2005), Altonji and Matzkin (2005), Chernozhukov and Hansen (2005), Chernozhukov et al. (2007), Horowitz and Lee (2007), Hoderlein and Mammen (2009), Imbens and Newey (2009), and Torgovitsky (2011).

In most empirical data covered in this survey, instruments exhibit only coarse and discrete variations. An important example of discrete instruments is the quarter of birth used by Angrist and Krueger (1991). With discrete instruments, structural features are only partially identified under the nonlinear nonseparable model (iv) (Chesher , 2005).[2] In this light, I show how partially identified parameters can be used to distinguish the model types (i), (ii), and (iii) against (iv).

Another practical limitation in common empirical data is the locality of instrumental effects, which prohibits identification of global shape of structural functions. Again, an example is Angrist and Krueger in which the quarter-of-birth instrument affects schooling choices only near the 9th to the 10th grades (for most cohorts and states). While the locality often results in the smaller power of tests, we will empirically show that models (i) and (ii) are rejected for wage outcome as a function of years of education. Likewise, it will be empirically shown that models (i) and (iii) are rejected for infant birth weight as a function of smoking intensity.

The objective of this paper is related to the literature on heterogeneity testing. For example, Chernozhukov and Hansen (2006) propose a test of heterogeneous quantile regression parameters under endogeneity. This idea is related to distinguishing the model (i) from model (iii). The objective of this paper is also related to the literature on nonparametric testing. For example, Horowitz and Lee (2009) propose tests of parametric quantile regressions against nonparametric family of quantile regressions under endogeneity. This idea can be used to distinguish the model (iii) from model (iv).

The value added by this paper to this existing literature is twofold. First, we develop a single device that can be used to distinguish the four types of models, (i), (ii), (iii) and (iv). Second, more importantly, our method accounts for the aforementioned difficulty associated with discrete instruments. This practically important issue has

[2] Also related is Jun, Pinkse, and Xu (2011).

not been explicitly addressed in the existing methods of heterogeneity or specification testing under endogeneity.

The paper is organized as follows. We first discuss how to use partially identified parameters to distinguish the four models in Section 2. Estimates of these partially identified parameters are used to construct test statistics in Section 3. The tests are applied to two empirical problems in Section 4. Section 5 summarizes the paper. The appendix contains mathematical notes.

2. Bounds as Means of Specification Testing

How can we distinguish the four model types? Let $\frac{\partial \phi}{\partial s}$ denote the partial effect of the structural function ϕ with respect to s. Under each model from (i) to (iii), this partial effect reduces to

$$\begin{aligned}
&\tfrac{\partial \phi}{\partial s}\big|_{(s,a)} = \beta && \text{under (i) the constant coefficient model,}\\
&\tfrac{\partial \phi}{\partial s}\big|_{(s,a)} = \phi'(s) && \text{under (ii) the nonlinear separable model, and}\\
&\tfrac{\partial \phi}{\partial s}\big|_{(s,a)} = \beta(a) && \text{under (iii) the linear random coefficient model.}
\end{aligned}$$

They are a constant in (i), a function of only s in (ii), and a function of only a in (iii). Models (i) and (iii) entail s-invariance of the partial effects, whereas models (i) and (ii) entail a-invariance of the partial effects.

The next step is to translate these discriminatory characteristics into empirically testable restrictions. As noted in the introductory section, empirical data often exhibit only local instrumental effects near a certain point of s. Under this common limitation, the s-invariance and the a-invariance of the partial effects at a given s are characterized by

$$\begin{aligned}
&\mathcal{H}_0^S : \frac{\partial^2}{\partial s^2}\phi(s,a) \;=\; 0 \text{ for all } a \in \text{supp}(A) \text{ at a given } s, \qquad \text{and}\\
&\mathcal{H}_0^A : \frac{\partial^2}{\partial s \partial a}\phi(s,a) = 0 \text{ for all } a \in \text{supp}(A) \text{ at a given } s,
\end{aligned}$$

respectively, where $\text{supp}(\cdot)$ denotes the support of the random variable. Rejection of $\mathcal{H}_0^S$ will result in falsification of model types (i) and (iii). Likewise, rejection of $\mathcal{H}_0^A$ will result in falsification of model types (i) and (ii).

Specification tests therefore require some knowledge of the partial effects, which is not observable to us. We use as an alternative device the average partial effect, denoted by

$$APE(s,z) := \mathrm{E}\left[\frac{\partial}{\partial s}\phi(S,A)\middle| S=s, Z=z\right].$$

This $APE(s,z)$ contains aggregate information about $\frac{\partial\phi}{\partial s}$ over some set of population near the locality of s.[3] The instrument Z as an additional conditioning variable plays a key role in recovering the control variable (cf. Imbens and Newey (2009)), which in turn reveals variations of $\frac{\partial\phi}{\partial s}$ in a. Consequently, sensitivity of $APE(s,z)$ in z implies whether $\mathcal{H}_0^A$ is true or not.

Point identification of the APE would allow us to test the hypotheses $\mathcal{H}_0^S$ and $\mathcal{H}_0^A$ easily. However, as noted in the introductory section, discreteness instruments are obstacles for recovering smooth counterfactuals. Consequently, the APE is generally at best partially identified, and our testing criteria will be based on the bounds of the APE, and their intersections. The following subsection discusses relationships between $APE(s,z)$ and its bounds, which will establish empirically testable restrictions under the null hypotheses $\mathcal{H}_0^S$ and $\mathcal{H}_0^A$.

2.1. Bounds of the APE and Their Implications for the Hypotheses. In the literature on nonseparable models, the first-stage restrictions are often used for identification of the second-stage features. Consider the following nonseparable first-stage:

$$S = \psi(Z,V),$$

where Z is an instrumental variable which may be discrete, and V denotes the reduced-form unobserved heterogeneity. We impose the following standard restrictions:

[3] Even if ϕ is differentiable everywhere, this object need not exist in general due to the Borel-Kolmogorov paradox. But we assume it does exist, a sufficient condition for which is provided in Appendix Section 6.

Assumption 12 (Restrictions).

(IM) First-Stage Monotonicity: $\psi(z,v) \leqslant \psi(z,v') \iff \psi(z',v) \leqslant \psi(z',v')$ for all $z, z' \in \text{supp}(Z)$ and for all $v, v' \in \text{supp}(V)$.

(IV) Instrument Independence: $(A,V) \perp\!\!\!\perp Z$.

(AC) Absolute Continuity: the distribution of $S \mid_{Z=z}$ is absolutely continuous with a convex support for every $z \in \text{supp}(Z)$.

(SI) Strong Instrument: $\psi(z,v)$ is strictly increasing in z.

(IM) states that the first-stage choice is monotone in some index of unobserved attributes, an assumption useful for constructing control variables (e.g., Imbens and Newey (2009)). (IV) states the standard instrument independence. (AC) requires that S is continuously distributed, which is an admissible abstraction of real data if S has many support points such as years of schooling or number of cigarettes. On the other hand, we maintain the assumption that Z is discrete since instruments often have much fewer support points. This (AC) guarantees that the distribution of $S \mid_{Z=z}$ has a probability density function denoted by $f_{S|Z}(\,\cdot \mid z)$, and $F_{S|Z}(\,\cdot \mid z)$ is invertible on its support. Under this restriction, we denote the quantile regression by $Q_{S|Z} := F^{-1}_{S|Z}$. (SI) requires nontrivial first-stage effects. Alternatively, $\psi(z,v)$ can be strictly decreasing in z, in which case we only need to relabel z into negative values. In practice, we are generally restricted to the locality of s at which (SI) is satisfied, e.g., $s \approx$ 9-10 for Angrist and Krueger (1991).

As a device of characterizing model types, we use the notion of the local shapes of the structural function ϕ. To formalize the sense of locality, consider the following finite sets

$$\begin{aligned} \mathcal{Z}^+_{s,z} &:= \{z' \in \text{supp}(Z) \mid z' > z, s \in \text{supp}(S \mid Z = z')\} \quad \text{and} \\ \mathcal{Z}^-_{s,z} &:= \{z' \in \text{supp}(Z) \mid z' < z, s \in \text{supp}(S \mid Z = z')\} \end{aligned}$$

for a given $(s,z) \in \text{supp}(S,Z)$. $\mathcal{Z}^+_{s,z}$ consists of the "right instruments," whereas $\mathcal{Z}^-_{s,z}$ consists of the "left instruments." The right (respectively, left) instruments counterfactually induce higher (respectively, lower) treatment values under (SI). Define

$z^+_{s,z} = \min \mathcal{Z}^+_{s,z}$ and $s^-_{s,z} = \max \mathcal{Z}^-_{s,z}$, the right and left instruments closest to z, respectively. Define the smallest local interval

$$I(s,z) := \big(\min\{Q_{S|Z}(F_{S|Z}(s \mid z) \mid z^+_{s,z}), Q_{S|Z}(F_{S|Z}(s \mid z) \mid z^-_{s,z})\}, \\ \max\{Q_{S|Z}(F_{S|Z}(s \mid z) \mid z^+_{s,z}), Q_{S|Z}(F_{S|Z}(s \mid z) \mid z^-_{s,z})\}\big) .$$

bounded by the closest counterfactual levels of treatments induced by the instruments.

Definition 1 (Local Shape). Suppose that $\phi \in C^2$. The following three local shapes are defined:

(DR) Local Concavity: $\left.\frac{\partial^2 \phi}{\partial s^2}\right|_{(s',a)} < 0$ for all $s' \in I(s,z)$ and $a \in \text{supp}(A)$.

(CR) Local Linearity: $\left.\frac{\partial^2 \phi}{\partial s^2}\right|_{(s',a)} = 0$ for all $s' \in I(s,z)$ and $a \in \text{supp}(A)$.

(IR) Local Convexity: $\left.\frac{\partial^2 \phi}{\partial s^2}\right|_{(s',a)} > 0$ for all $s' \in I(s,z)$ and $a \in \text{supp}(A)$.

These local concavity, linearity, and convexity determine the order relation between the APE and its upper and lower bounds. To discuss its intuition, let B^R and B^L denote the difference quotients of ϕ that can be formed by counterfactual variations in treatment S induced by the right and left instruments, respectively. Under the local concavity of ϕ, we will see $B^R < APE < B^L$. The inequality will be reversed under the local convexity of ϕ. Moreover, we expect to see the equality $B^R = APE = B^L$ under the local linearity of ϕ. The following theorem formalizes these informal arguments.

Theorem 5 (Bounds of the APE). *Suppose that Assumption 12 holds. Then, the (in-)equalities*

$$\begin{aligned} B(s; z^+_{s,z}, z) &< APE(s,z) < B(s; s^-_{s,z}, z) && \textit{under (DR)} \\ B(s; z^+_{s,z}, z) &= APE(s,z) = B(s; s^-_{s,z}, z) && \textit{under (CR)} \\ B(s; z^+_{s,z}, z) &> APE(s,z) > B(s; s^-_{s,z}, z) && \textit{under (IR)} \end{aligned}$$

hold for $z^+_{s,z} = \min \mathcal{Z}^+_{s,z}$ *and* $s^-_{s,z} = \max \mathcal{Z}^-_{s,z}$*, where* B *is defined with* $\theta_z(s) := F_{S|Z}(s \mid z)$ *by*

$$B(s; z', z) \quad := \quad \frac{E\left[Y \left| S = Q_{S|Z}\left(\theta_z(s) \mid z'\right), Z = z'\right.\right] - E\left[Y \left| S = Q_{S|Z}\left(\theta_z(s) \mid z\right), Z = z\right.\right]}{Q_{S|Z}\left(\theta_z(s) \mid z'\right) - Q_{S|Z}\left(\theta_z(s) \mid z\right)}.$$

REMARK 12. When $\mathcal{Z}^+_{s,z}$ (respectively $\mathcal{Z}^-_{s,z}$) is an empty set, there is no lower bound (respectively no upper bound) under Definition 1 (DR). The opposite is the case under Definition 1 (IR).

REMARK 13. As the support of Z becomes richer to constitute a cluster point at z, these bounds asymptotically converge to the true APE under some regularity assumptions.

REMARK 14. Notice that formula of the bound $B(s; z', z)$ resembles that of the LATE (Imbens and Angrist, 1994). It takes the form of the "heterogeneous reduced-form effects" divided by the "heterogeneous first-stage effects." This parallels the LATE which takes the form of the "average reduced-form effects" divided by the "average first-stage effects."

Implication for Hypothesis Testing: The theorem suggests that comparing the order relation between $B(s; z', z)$ and $B(s; z'', z)$, which can be identified from observed data, reveals the local shape of the structural function ϕ. If data indicates $B(s; z', z) < B(s; z'', z)$ or $B(s; z', z) > B(s; z'', z)$, then we are in a position to reject (CR), and hence to reject $\mathcal{H}^S_0$. That is, variations of B in its *second* argument (i.e., z' or z'') are used to test the hypothesis $\mathcal{H}^S_0$.

On the other hand, variations of B in its *third* argument (i.e., z) can be used to test $\mathcal{H}^A_0$. Note that under the null hypothesis $\mathcal{H}^A_0$, the partial effect is invariant in variations in a. Varying z while fixing s at a given locality causes variations in the control variable v which in turn affects a under endogeneity. Thus, $APE(s, z)$ should be insensitive to variations in z under $\mathcal{H}^A_0$. This in turn implies non-emptiness of the intersection bounds $\bigcap_{z \in \mathcal{Z}}[B(s; z^+_{s,z}, z), B(s; z^-_{s,z}, z)]$ under $\mathcal{H}^A_0$. Contrapositively, emptiness of these intersection bounds leads to rejection of $\mathcal{H}^A_0$.

Example – Model (i): For an instance of the linear constant coefficient model

$$\begin{cases} Y = \beta S + A \\ S = \pi Z + V \end{cases} \qquad \text{with } (A, V) \sim N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right),$$

some calculation yields $Q_{S|Z}(\theta \mid z) = \pi z + Q_V(\theta)$ and $\mathrm{E}[Y \mid S = s, Z = z] = (1+\beta)s - \pi z$. Substituting these expressions in the bound formula shows

$$B(s; z', z) = \beta \qquad \text{for all } (s, z', z).$$

The bounds $B(s; z', z)$ universally point-identify the constant partial derivative, which is β. As the upper and lower bounds coincide at β, we fail to reject $\mathcal{H}_0^S$. Moreover, by non-emptiness of the intersection bounds $\bigcap_{z \in \mathcal{Z}} [B(s; z^+_{s,z}, z), B(s; s^-_{s,z}, z)] = \{\beta\}$, we fail to reject $\mathcal{H}_0^A$ too. □

Example – Model (ii): For an instance of the nonlinear regression model

$$\begin{cases} Y = \beta\sqrt{S} + A \\ S = \pi Z + V \end{cases} \qquad \text{with } (A, V) \sim N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right),$$

some calculation yields $Q_{S|Z}(\theta \mid z) = \pi z + Q_V(\theta)$ and $\mathrm{E}[Y \mid S = s, Z = z] = \beta\sqrt{s} + s - \pi z$. Substituting these expressions in the bound formula shows

$$B(s; z', z) = \beta \frac{\sqrt{\pi(z' - z) + s} - \sqrt{s}}{\pi(z' - z)} \qquad \text{for all } (s, z', z).$$

Therefore, taking $(s; z', z) = (\pi; 2, 1)$ for example yields

$$\underbrace{B(\pi; 2, 1)}_{= \frac{\beta}{\sqrt{\pi}}(\sqrt{2}-1)} < \underbrace{APE(\pi, 1)}_{= \frac{\beta}{\sqrt{\pi}} \frac{1}{2}} < \underbrace{B\left(\pi; \frac{1}{2}, 1\right)}_{= \frac{\beta}{\sqrt{\pi}}(2-\sqrt{2})} \qquad \text{if } \beta > 0.$$

As we have a strict inequality between upper and lower bounds, we reject $\mathcal{H}_0^S$. On the other hand, since the set enclosed by the intersection bounds containing the z-invariant APE is non-empty, we fail to reject $\mathcal{H}_0^A$. □

Example – Model (iii): For an instance of the linear random coefficient model

$$\begin{cases} Y = (\beta A)S + A \\ S = \pi Z + V \end{cases} \qquad \text{with } (A, V) \sim N\left(0, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right),$$

some calculation yields $Q_{S|Z}(\theta \mid z) = \pi z + Q_V(\theta)$ and $\mathrm{E}[Y \mid S = s, Z = z] = (\beta s + 1)(s - \pi z)$. Substituting these expressions in the bound formula shows

$$B(s; z', z) = \beta \cdot (s - \pi z) \qquad \text{for all } (s, z', z).$$

On the other hand, the true APE takes the following form:

$$APE(s, z) = \beta \mathrm{E}[A \mid S = s, Z = z] = \beta \mathrm{E}[A \mid V = s - \pi z] = \beta \cdot (s - \pi z) \qquad \text{for all } (s, z).$$

Therefore, the bounds $B(s; z', z)$ point-identify the heterogeneous $APE(s, z)$ for every (s, z). As the upper and lower bounds coincide at $\beta(s - \pi z)$, we fail to reject $\mathcal{H}_0^S$. On the other hand, given the empty intersection bounds $\bigcap_{z \in \mathcal{Z}} [B(s; z_{s,z}^+, z), B(s; s_{s,z}^-, z)] = \bigcap_{z \in \mathcal{Z}} \{\beta \cdot (s - \pi z)\} = \phi$ under $\pi \neq 0$, we reject $\mathcal{H}_0^A$. □

These three examples reconfirm the relevance of the empirically testable hypotheses for our main goal of specification testing. Indeed, we correctly fail to reject $\mathcal{H}_0^S$ under models (i) and (iii). Similarly, we correctly fail to reject $\mathcal{H}_0^A$ under models (i) and (ii).

2.2. Numerical Illustrations. This subsection graphically illustrates the implications of Theorem 5 for specification tests. Consider the following family of data-generating models as a numerical example:

$$Y = \phi(S, A) := \frac{5^{p_s}}{1 - p_s} (S - 5)^{1 - p_s} e^{p_a A} + (1 - p_a) A,$$

where p_s and p_a are parameters that generate heterogeneity across S and A, respectively. These parameters reduce the model into the four types in the following manner:

$$\begin{aligned}
p_s = 0 \quad p_a = 0 \quad &\Longrightarrow \quad \text{(i)} \quad \text{Linear constant coefficient model} \\
p_s \neq 0 \quad p_a = 0 \quad &\Longrightarrow \quad \text{(ii)} \quad \text{Nonlinear separable model} \\
p_s = 0 \quad p_a \neq 0 \quad &\Longrightarrow \quad \text{(iii)} \quad \text{Linear random coefficient model} \\
p_s \neq 0 \quad p_a \neq 0 \quad &\Longrightarrow \quad \text{(iv)} \quad \text{Nonlinear nonseparable model}
\end{aligned}$$

In order to keep the true APE analytically tractable, we assume the linear first stage $S = 10 + Z + V$ where Z is uniform on $\{-1, 0, 1\}$ and (A, V) follows a joint normal distribution with positive covariance.

Figure 3.1 depicts the true $APE(10,z)$ together with its bounds $B(10;z+1,z)$ and $B(10;z-1,z)$. Notice that the bound inequalities are strict under the nonlinear models (ii) and (iv), whereas the upper and lower bounds exactly coincide under the linear models (i) and (iii). This result agrees with the (in-)equalities in Theorem 5, and these characteristics can be used in an attempt to reject $\mathcal{H}_0^S$, or as a way to distinguish types (ii) and (iv) from types (i) and (iii).

Also notice that the true $APE(10,z)$ is non-constant in z under the heterogeneous models (iii) and (iv), whereas the true $APE(10,z)$ is constant across z under the homogeneous models (i) and (ii). The non-constancy of $APE(10,z)$ across z under the heterogeneous models (iii) and (iv) causes empty intersection bounds across z, and this characteristic can be used in an attempt to reject $\mathcal{H}_0^A$, or as a way to distinguish types (iii) and (iv) from types (i) and (ii).

3. The Test Statistics

The previous section analyzed the role that bounds play in distinguishing the four types of the models under discrete instruments. A sample analog of the bound in Theorem 5 is

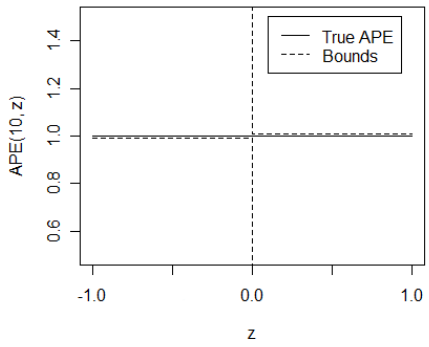
$$\widehat{B}(s;z',z) \quad = \quad \left(\mathbb{E}_{n,h_n}\left[Y\,|S=\widehat{q}(s,z;z'),Z=z'\right]-\mathbb{E}_{n,h_n}\left[Y\,|S=s,Z=z\right]\right)/\left(\widehat{q}(s,z;z')-s\right),$$

where $\widehat{q}(s,z;z')$ is an r_n-consistent estimator of $Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right)$, and $\mathbb{E}_{n,h_n}$ denotes the Nadaraya-Watson estimator

$$\mathbb{E}_{n,h_n}\left[Y\,|S=s,Z=z\right]:=\left(\sum_{i=1}^{n}K\left(\frac{S_i-s}{h_n}\right)\mathbb{1}(Z_i=z)Y_i\right)/\left(\sum_{i=1}^{n}K\left(\frac{S_i-s}{h_n}\right)\mathbb{1}(Z_i=z)\right).$$

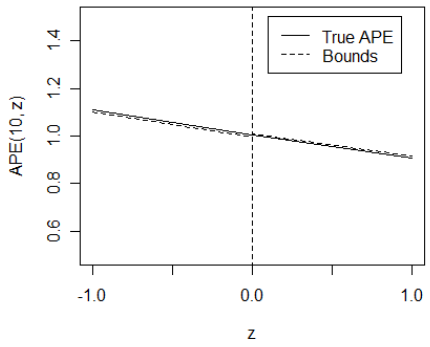
Because the econometric and statistical literature has suggested numerous estimators of conditional distribution and quantile regressions, we will not elaborate on large sample properties of $\widehat{q}(s,z;z')$. It suffices to assume $r_n=\sqrt{n}$, which is in general true when Z is discrete and is compatible with the following assumption.

Assumption 13 (Large Sample). (i) $\widehat{q}(s,z;z')-Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right)=O_p(r_n^{-1})$. (ii) $h_n\to 0$, $r_n\to\infty$, $h_n^3r_n^2\to\infty$, $nh_h\to\infty$, and $nh_nr_n^{-2}\to 0$ as $n\to\infty$. (iii) K is symmetric and Lipschitz-continuous with $\int K=1$, $\text{supp}(K)\subset(-1,1)$, and $\|K\|_{2+\delta}<\infty$ for some $\delta>0$. (iv) (A_i,V_i,Z_i) is i.i.d. (v) With $\varepsilon:=Y-\text{E}[Y\mid S,Z]$, the



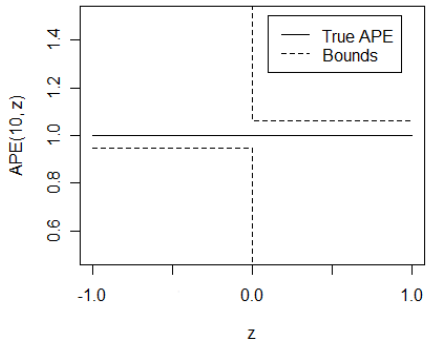
(i) Linear Constant Coefficient

$(p_s = 0.0,\ p_a = 0.0)$



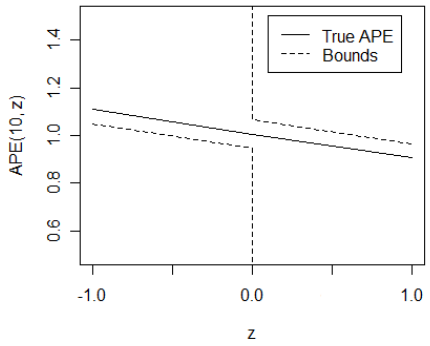
(iii) Linear Random Coefficient

$(p_s = 0.0,\ p_a = 0.1)$



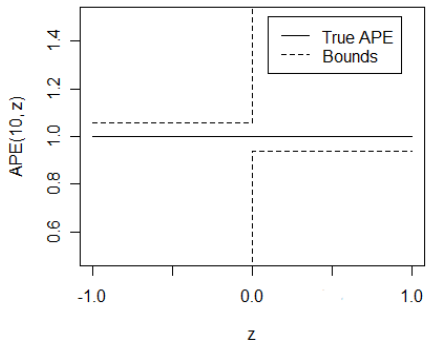
(ii) Nonlinear Separable

$(p_s = 0.5,\ p_a = 0.0)$



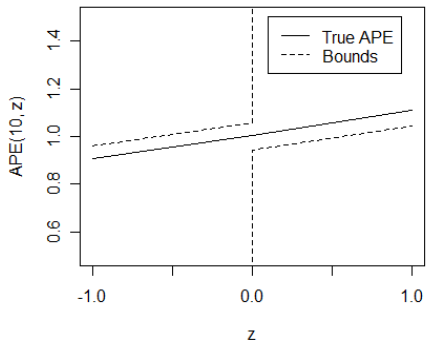
(iv) Nonlinear Nonseparable

$(p_s = 0.5,\ p_a = 0.1)$



(ii) Nonlinear Separable

$(p_s = -0.5,\ p_a = 0.0)$



(iv) Nonlinear Nonseparable

$(p_s = -0.5,\ p_a = -0.1)$

FIGURE 3.1. Relationships between the true $APE(10, z)$ and their bounds.

skedastic function $\sigma^2(s,z) := \mathrm{E}[\varepsilon^2 \mid S = s, Z = z]$ is twice continuously differentiable with bounded second derivatives, and $\mathrm{E}[|\varepsilon|^{2+\delta} \mid S = s, Z = z] < \infty$ for some $\delta > 0$. (vi) f_{SZ} is twice continuously differentiable with respect to s with bounded second derivatives. (vii) $\mathrm{E}[Y \mid S = s, Z = z]$ is twice continuously differentiable with respect to s with a bounded second derivative.

Under this set of assumptions, asymptotic distributions of the bound estimators $\widehat{B}(s; z', z)$ are obtained in Lemmas 22–24 in the appendix section. These asymptotic distributions will be used in turn to derive the asymptotic behavior of the test statistics to be presented in Sections 3.1 and 3.2.

Recall the two practical limitations discussed in the introductory section, discrete instruments and local instrumental effects. We fix short-hand notations in this light. Because we focus on discrete instrumental variation, let $\mathrm{supp}(Z) = \{z_1, \cdots, z_K\}$ with $z_1 < \cdots < z_K$. Moreover, because we focus on local instrumental effects, let $s \in \mathrm{supp}(S)$ be fixed hereafter.[4]

3.1. Test of the Hypothesis $\mathcal{H}_0^S$. For a test of model types (i) and (iii), we consider the null hypothesis

$$\mathcal{H}_0^S : \frac{\partial^2}{\partial s^2}\phi(s,a) \;=\; 0 \text{ for all } a \in \mathrm{supp}(A),$$

against the list of two alternatives

$$\begin{aligned}
\mathcal{H}_{-1}^S : \frac{\partial^2}{\partial s^2}\phi(s,a) \;&<\; 0 \text{ for some } a \text{ over a non-null set, and} \\
\mathcal{H}_{+1}^S : \frac{\partial^2}{\partial s^2}\phi(s,a) \;&>\; 0 \text{ for some } a \text{ over a non-null set.}
\end{aligned}$$

By Theorem 5, the null hypothesis $\mathcal{H}_0^S$ implies the empirically testable hypothesis

$$\mathcal{H}_0^{S'} : B(s; z_{k+1}, z_k) \;=\; B(s; z_{k-1}, z_k) \text{ for each } k = 2, \cdots, K-1.$$

[4] In case of non-local instrumental effects, one can extend our test statistics by summing / integrating over s to gain power. Summing over a finite set of s is a straightforward extension of our results. A drawback with integrating over continuous s is that the stochastic process based on our nonparametric estimation does not converge weakly to a tight process (hence non-Donsker). One way to overcome this difficulty is to directly obtain the extreme value distribution of the limit process, e.g., Chernozhukov, Lee, and Rosen (2009) Section 3.5.

Contrapositively, rejection of this $\mathcal{H}_0^{S'}$ concludes rejection of the original $\mathcal{H}_0^S$.

Figure 3.2 depicts relative orderings of the bounds $B(s;\ \cdot\ ,\ \cdot\ )$ under each case of these null and alternative hypotheses. A natural approach is to use a measure of discrepancy between upper and lower bounds to form a test statistic. As such, consider the simple test statistic

$$\widehat{T}_{k,n}^S \ := \ \frac{\sqrt{nh_n}\left(\widehat{B}(s; z_{k+1}, z_k) - \widehat{B}(s; z_{k-1}, z_k)\right)}{\sqrt{\Gamma_{k,k-1} + \Gamma_{k,k+1} - 2\Gamma_{k,k-1,k+1}}}$$

for any $k \in \{2, \cdots, K-1\}$, where an $(2K-2)\times(2K-2)$ matrix Γ is given in Section 7 in the appendix. Large negative values of $\widehat{T}_n^S$ reject $H_0^{S'}$ in favor of H_{-1}^S. Large positive values of $\widehat{T}_n^S$ reject $H_0^{S'}$ in favor of H_{+1}^S. This test statistic asymptotically follows the standard normal distribution.

PROPOSITION 4 (Test of $\mathcal{H}_0^S$). *Suppose that Assumptions 12 and 13 hold. Then,* $\widehat{T}_{k,n}^S \xrightarrow{d} Z \sim N(0,1)$ *under* $\mathcal{H}_0^S$.

3.2. Test of the Hypothesis $\mathcal{H}_0^A$. For a test of model types (i) and (ii), we consider the null hypothesis

$$\mathcal{H}_0^A : \frac{\partial}{\partial s}\phi(s,a) = \beta_s \text{ for some constant } \beta_s \text{ for } [P_{A|S=s}]\text{-a.s. } a$$

A test statistic will be constructed through the following logic. Given discrete z, $\mathcal{H}_0^A$ implies

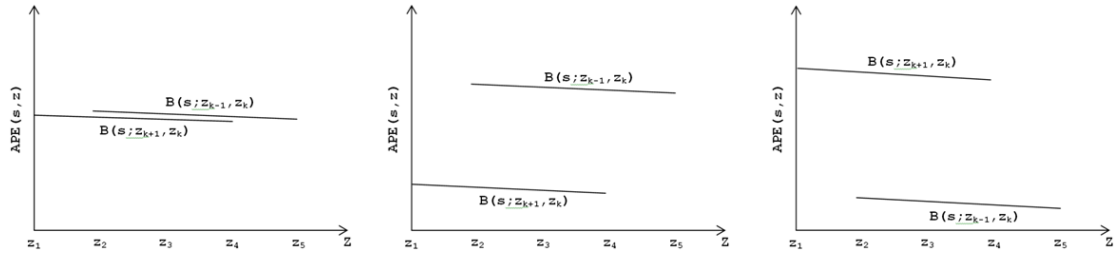
$$\mathcal{H}_0^{A'} : APE(s,z) = \beta_s \text{ for some constant } \beta_s \text{ for all } z \in \text{supp}(Z \mid S = s).$$

Suppose that ϕ exhibits (DR) or (CR), i.e., (IR) is ruled out. Theorem 5 under $\mathcal{H}_0^{A'}$ yields

$$\begin{aligned} B(s; z_{k+1}, z_k) &\leqslant \beta_s \quad \text{for all} \quad k = 1, \cdots, K-1 \text{ and} \\ \beta_s &\leqslant B(s; z_{k-1}, z_k) \quad \text{for all} \quad k = 2, \cdots, K. \end{aligned}$$

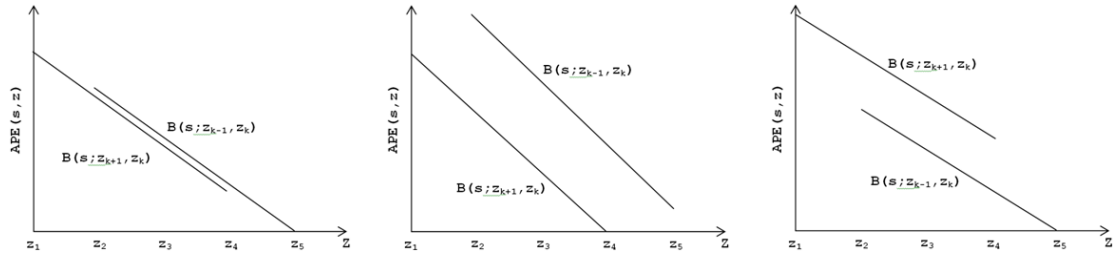
Satisfaction of these inequalities implies the following empirically testable hypothesis:

$$\mathcal{H}_0^{A''} : B(s; z_{k+1}, z_k) \leqslant B(s; z_{k'-1}, z_{k'}) \text{ for all } k, k'$$



Fail to reject $\mathcal{H}_0^S$ | Reject $\mathcal{H}_0^S$ in favor of $\mathcal{H}_{-1}^S$ | Reject $\mathcal{H}_0^S$ in favor of $\mathcal{H}_{+1}^S$

Fail to reject $\mathcal{H}_0^A$ | Fail to reject $\mathcal{H}_0^A$ | Fail to reject $\mathcal{H}_0^A$



Fail to reject $\mathcal{H}_0^S$ | Reject $\mathcal{H}_0^S$ in favor of $\mathcal{H}_{-1}^S$ | Reject $\mathcal{H}_0^S$ in favor of $\mathcal{H}_{+1}^S$

Reject $\mathcal{H}_0^A$ | Reject $\mathcal{H}_0^A$ | Reject $\mathcal{H}_0^A$

FIGURE 3.2. Graphical characterizations of specification tests.

Rejection of $\mathcal{H}_0^{A''}$ implies rejection of $\mathcal{H}_0^{A'}$, which in turn implies rejection of $\mathcal{H}_0^A$. Figure 3.2 depicts the various cases in which the last null hypothesis is rejected or fails to be rejected.

The last testable form of the hypothesis $\mathcal{H}_0^{A''}$ requires that no lower bound exceeds any uppper bound. Consider a test statistic as a measure of the largest deviation from this requirement

$$\widehat{T}_n^A(W) \quad := \quad \sqrt{nh_n} \max_{k,k'} \left\{ W_{k,k'} \left(\widehat{B}(s; z_{k+1}, z_k) - \widehat{B}(s; z_{k'-1}, z_{k'}) \right) \right\}$$

where W is a weighting matrix. Consider the random variable $T^A(W) := \max_{k,k'} \{W_{k,k'} (L_k - U_{k'})\}$, where $(L_1, U_2, L_2, U_3, L_3, \cdots, U_K)$ is a $2(K-1)$-tuple random vector following the normal law $N(0, \Gamma)$ with Γ given in Section 7 in the appendix.

PROPOSITION 5 (Test of $\mathcal{H}_0^A$). *Suppose that Assumptions 12 and 13 hold. Then,*

$$\lim_{n\to\infty} \sup_{H \in \mathcal{H}_0^A} Pr \left(\widehat{T}_n^A \geqslant F_{T^A}^{-1}(1-\alpha) \middle| H \right) \quad \leqslant \quad \alpha$$

holds for all $\alpha \in (0,1)$ *under (DR) or (CR).*

REMARK 15. Under (IR) or (CR), we can develop a similar test statistic by switching the roles of $\widehat{B}(s; z_{k+1}, z_k)$ and $\widehat{B}(s; z_{k'-1}, z_{k'})$.

The size of this test is in general conservative. The exact size α is achieved under the case of point-identification for every $z \in \text{supp}(Z)$, that is, when Assumption 1 (CR) holds. Tendency of under-rejection becomes more likely as partially identified regions become wider.

3.3. Monte Carlo Evidences. Consider the family of data-generating models introduced in Section 2.2:

$$Y = \frac{5^{p_s}}{1-p_s}(S-5)^{1-p_s}e^{p_a A} + (1-p_a)A$$

Recall that p_s and p_a are parameters that induce heterogeneity in the dimensions of S and A, respectively. The null hypothesis $\mathcal{H}_0^S$ is true when $p_s = 0$. Similarly, the null hypothesis $\mathcal{H}_0^A$ is true when $p_a = 0$. We expect that the power of the test of $\mathcal{H}_0^S$ increases as p_s deviates away from zero, and that the power of the test of $\mathcal{H}_0^A$ increases as p_a deviates away from zero.

Figure 3.3 draws MC-simulated power curves of the 95% level tests across different values of p_s and p_a for various sample sizes of 1,000, 2,000, 5,000, and 10,000 observations. The size is indeed correct under both tests with about 5% rejection probability at $p_s = 0$ and $p_a = 0$ for the tests of $\mathcal{H}_0^s$ and $\mathcal{H}_0^a$, respectively. The power approaches one as the sample size increases when $p_s \neq 0$ and $p_a \neq 0$ under the tests of $\mathcal{H}_0^s$ and $\mathcal{H}_0^a$, respectively. These simulation results evidence the power as well as the unbiasedness of the proposed tests. The next section further evidences their power with empirical data.

4. Testing with Empirical Data

In this section, we apply the proposed methods to two economic problems, both of which have been extensively studied in the empirical literature. The first application

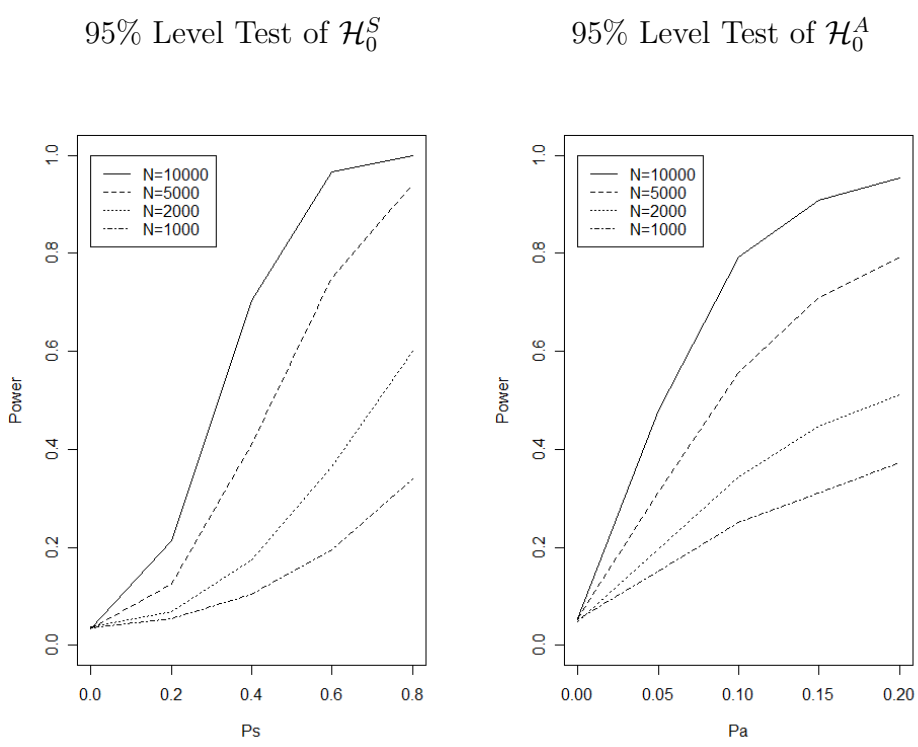


FIGURE 3.3. Simulated power curves of the specification tests.

analyzes wage returns to years of schooling (Section 4.1). We will reject $\mathcal{H}_0^A$, hence precluding (i) the linear constant coefficient model and (ii) the nonlinear separable model. The second application analyzes infant birth weight as an outcome of smoking intensity (Section 4.2). We will reject $\mathcal{H}_0^S$, hence precluding (i) the linear constant coefficient model and (iii) the linear random coefficient model.

4.1. Returns to Schooling. Empirical assessment of the marginal returns to schooling has long been of interest in labor economics. An important scene in the literature was the emergence of natural experiments by instrumental variables (IV) as sources of exogenous variations in endogenous choice, e.g., Angrist and Krueger (1991). Interpretations of the IV estimator has been discussed in the econometric literature, e.g., Angrist and Imbens (1995); (1997). For instance, (1997) shows that 2SLS identifies the average partial effect under a special case of structural type (iii). In this way, the knowledge of the true model type may allow a sensible structural interpretation of the common statistical parameters.

We attempt to test the structural types (i), (ii), and/or (iii) using the data of Angrist and Krueger (1991). They used the quarter of birth as an exogenous source of variation in years of schooling in order to study partial effects of education on wage outcomes. The data consists of three decades of birth cohorts with log wage outcome (Y), endogenous choice of years of schooling (S), and several attribute characteristics including quarter of birth (Z) and state of birth as well as other standard covariates.

Before starting the tests, we note that the instrumental effects should be limited to the locality of the 9th to 10th years of schooling, where compulsory education laws are supposed to induce difference by quarter of birth, although there are some variations across time and states. Lleras-Muney (2002) describes the details of this policy. Recall that this practical limitation was the reason for our setting of the null hypotheses $\mathcal{H}_0^S$ and $\mathcal{H}_0^A$ at fixed s instead of global s. It is necessary to focus on this locality where the instrument indeed matters, as characterized by the testable restriction

$$\text{Local First-Stage Effects} := Q_{S|Z}\left(\theta_z(s) \mid z'\right) - Q_{S|Z}\left(\theta_z(s) \mid z\right) \neq 0.$$

In order to confirm this restriction of Assumption 12 (SI), we estimate the heterogeneous first-stage effects across (s, z) using smooth quantile regression estimation[5] for the subsample of individuals born in Arkansas, Kentucky, or Tennessee, the three states associated with strongest first-stage effects (Hoogerheide et al. , 2007). Figure 3.4 shows that the first-stage effects in these three states are nonzero at $s = 9$, whereas the instrument may be irrelevant at the college level ($s = 12$ and $s = 16$). These differential first-stage effects coincide with the pattern implied by compulsory education policies. We drop the first quarter because the transition between the first and second quarters induces no difference even at $s = 9$. Therefore, we use the instrumental variations among $\text{supp}(Z) = \{2, 3, 4\}$.

[5] This follows from Horowitz (1998) replacing the indicator function with a smooth function in the objective function of quantile estimators. We do so for our treatment of the years S as a continuous variable.

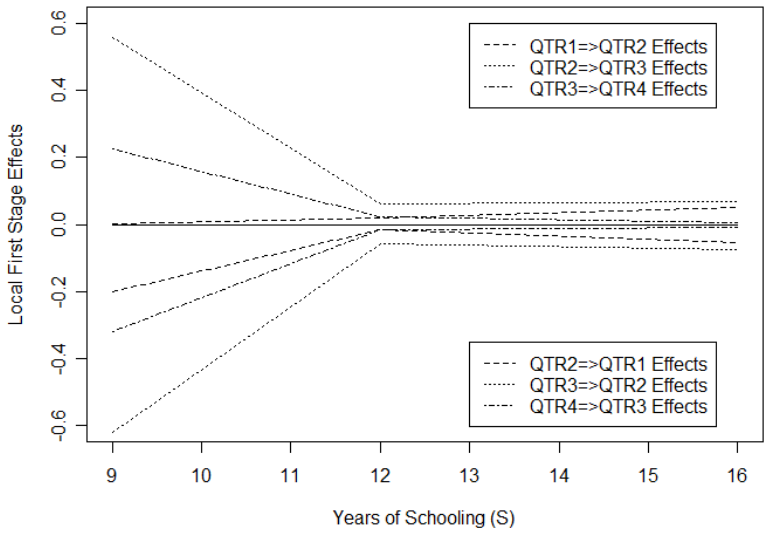


FIGURE 3.4. Heterogeneous first-stage effects across years of schooling and quarters of birth. Sample: place of birth from Arkansas, Kentucky, or Tennessee for all birth years.

The set estimates and confidence intervals for $APE(9.5, z)$ for $z \in \{2, 3, 4\}$ are depicted in Figure 3.5. Comparing Figure 3.5 with Figure 3.1, we can conjecture that the closer model in a statistical sense is (iii) the linear random coefficient model. Moreover, comparing Figure 3.5 with Figure 3.2, we can see that the bottom left picture in Figure 3.2 best resembles Figure 3.5 in a statistical sense. According to Figure 3.5, the upper and lower bounds of $APE(9.5, 3)$ do not seem to significantly differ from each other, hence we are not likely to reject $\mathcal{H}_0^S$. On the other hand, the figure shows that $APE(9.5, z)$ tends to decrease in z, which implies that we will probably reject $\mathcal{H}_0^A$.

Let us formalize these visual analyses as follows. First, consider the hypothesis $\mathcal{H}_0^S : \frac{\partial^2}{\partial s^2}\phi(9.5, a) = 0$ for all $a \in \text{supp}(A)$. Recall that large negative (respectively, positive) values of the test statistic $\widehat{T}_n^S$ reject the null hypothesis $\mathcal{H}_0^S$ of constant returns in favor of the alternative hypothesis $\mathcal{H}_{-1}^S$ of decreasing returns (respectively, $\mathcal{H}_{+1}^S$ of increasing returns). Table 3.1 shows that the test statistic is negative, but not significantly so. Hence, we fail to reject $\mathcal{H}_0^S$. Second, consider a test of the hypothesis

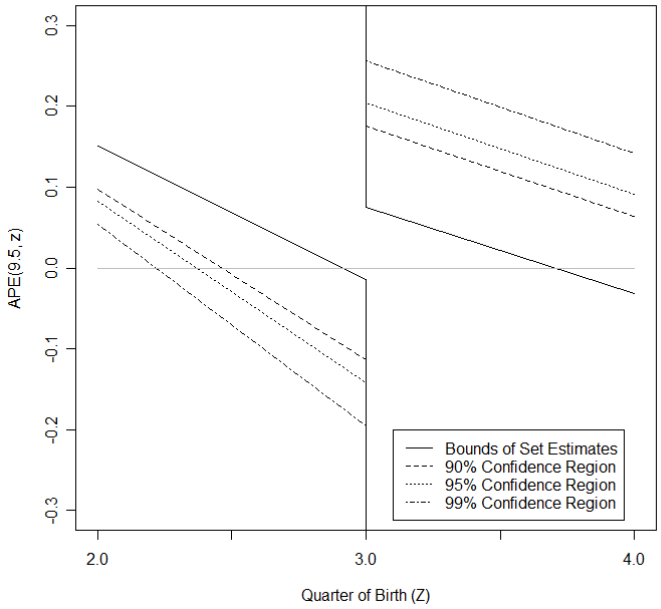


FIGURE 3.5. Bounds and confidence regions of $APE(9.5, z)$ for $z = 2, 3, 4$. Sample: place of birth in Arkansas, Kentucky, or Tennessee.

Null Hypothesis	Alternative Hypothesis	Test Statistic	p-value
$\mathcal{H}_0^S : \frac{\partial \phi}{\partial s}$ is constant across S	$\mathcal{H}_{-1}^S : \phi$ is concave in S $\mathcal{H}_{+1}^S : \phi$ is convex in S	$\widehat{T}_n^S = -0.90$	0.367
$\mathcal{H}_0^A : \frac{\partial \phi}{\partial s}$ is constant across A	$\mathcal{H}_1^A : \frac{\partial \phi}{\partial s}$ varies with A	$\widehat{T}_n^A = 106.38$	0.018^{**}

TABLE 3.1. Results of specification tests for Angrist & Krueger (1991) data.

$\mathcal{H}_0^A : \frac{\partial}{\partial s}\phi(9.5, a) = \beta_s$ for some constant β_s for all $a \in \text{supp}(A)$. Table 3.1 shows that the test statistic is significantly large at the level of 5%. We therefore reject $\mathcal{H}_0^A$.

These results imply that the wage production function may be linear in years of schooling. However, the constant returns to education are likely to be heterogeneous, conceivably across unobserved abilities. The structural specifications of (i) the constant coefficient model and (ii) the nonlinear separable model are rejected, whereas (iii) the linear random coefficient model survived our attempt at rejection.

4.2. Smoking and Infant Birth Weights. The effects of smoking by pregnant women on infant birth weights have been studied by an extensive body of both the health economic literature (e.g., Rosenzweig and Schultz, 1983; Evans and Ringel, 1999; and Lien and Evans, 2005) and the medical literature (e.g., Lightwood, Phibbs, and Glantz, 1999). In this section, we analyze the structural type of the birth-weight production function that takes cigarettes as a negative production factor. Cigarette excise tax rates are used to instrument for variations in cigarette consumption, following the approach of Evans and Ringel (1999). From natality data of the National Vital Statistics System of the National Center for Health Statistics, we extract a random sample of 100,000 observations from 1989 to 1999. Since three categories of instrumental variations suffice for our purpose, we categorize the tax rate into three groups, high for 50–100%, medium for 25–50%, and low for 0–25%, which are labeled as z=1, 2, and 3, respectively.

Cigarette excise tax rates have nearly continuous variations. This instrumental variable, unlike the example of Section 4.1 or many other empirical data, is rich enough to allow nonparametric inferences without partial identification. We nevertheless consider this application for the sake of demonstrating the power of our test for $\mathcal{H}_0^S$.

In order to confirm the restriction of Assumption 12 (SI), we estimate the heterogeneous first-stage effects across (s, z) using smooth quantile regression estimation. Figure 3.6 suggests that the first-stage effects are nonzero for $2 \leqslant s \leqslant 8$. Both positive and negative variations are large around $s = 5$. Therefore, we focus on a neighborhood of $s = 5$.

The set estimates and confidence intervals for $APE(5, z)$ are depicted in Figure 3.7. We can conjecture the true structural type by comparing Figure 3.7 with Figure 3.1 or 3.2. In comparison with Figure 3.1, we see that the graph representing (ii) the nonlinear separable model most closely resembles Figure 3.7. The upper and lower bounds of $APE(5, 2)$ seem to differ significantly from each other, hence we will probably reject $\mathcal{H}_0^S$. On the other hand, the figure does not imply a tendency of either increase or decrease for the true $APE(5, z)$ in z, thus we are not likely to reject $\mathcal{H}_0^A$.

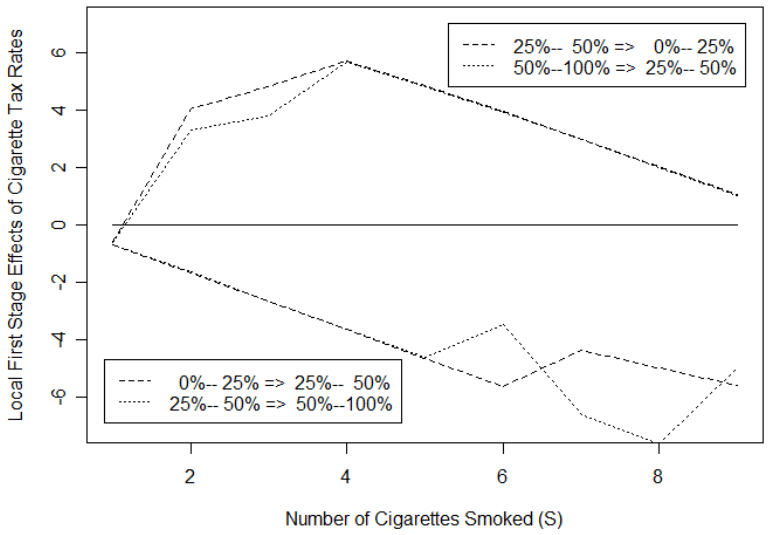


FIGURE 3.6. Heterogeneous first-stage effects across number of cigarettes smoked and cigarette tax rate.

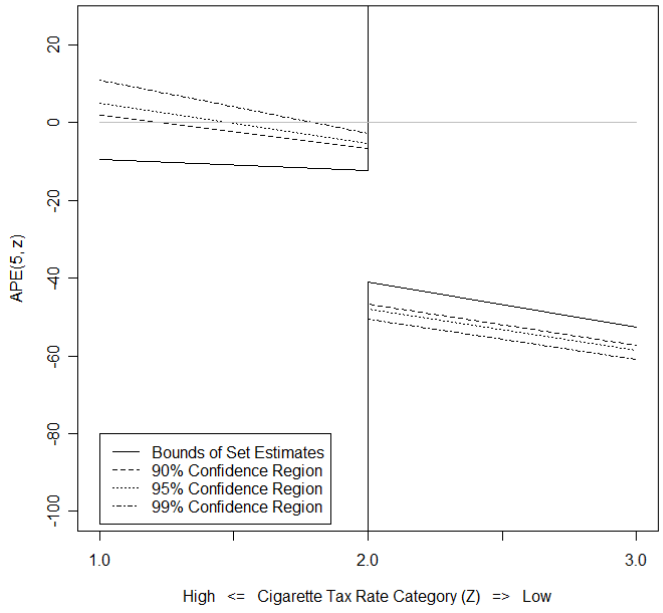


FIGURE 3.7. Bounds and confidence regions of $APE(5, z)$ for $z = 1, 2, 3$ with $h = 3$.

Let us formalize these visual analyses. First, consider a test of hypothesis $\mathcal{H}_0^S : \frac{\partial^2}{\partial s^2}\phi(5, a) = 0$ for all $a \in \text{supp}(A)$. Table 3.1 shows that the test statistic is

$s = 4$

Null Hypothesis	Alternative Hypothesis	Test Statistic	p-value
$\mathcal{H}_0^S : \frac{\partial \phi}{\partial s}$ is constant across S	$\mathcal{H}_{-1}^S : \phi$ is concave in S $\mathcal{H}_{+1}^S : \phi$ is convex in S	$\widehat{T}_n^S = 5.86$	0.000^{***}
$\mathcal{H}_0^A : \frac{\partial \phi}{\partial s}$ is constant across A	$\mathcal{H}_1^A : \frac{\partial \phi}{\partial s}$ varies with A	$\widehat{T}_n^A = 0.00$	0.625

$s = 5$

Null Hypothesis	Alternative Hypothesis	Test Statistic	p-value
$\mathcal{H}_0^S : \frac{\partial \phi}{\partial s}$ is constant across S	$\mathcal{H}_{-1}^S : \phi$ is concave in S $\mathcal{H}_{+1}^S : \phi$ is convex in S	$\widehat{T}_n^S = 4.07$	0.000^{***}
$\mathcal{H}_0^A : \frac{\partial \phi}{\partial s}$ is constant across A	$\mathcal{H}_1^A : \frac{\partial \phi}{\partial s}$ varies with A	$\widehat{T}_n^A = 0.00$	0.618

$s = 6$

Null Hypothesis	Alternative Hypothesis	Test Statistic	p-value
$\mathcal{H}_0^S : \frac{\partial \phi}{\partial s}$ is constant across S	$\mathcal{H}_{-1}^S : \phi$ is concave in S $\mathcal{H}_{+1}^S : \phi$ is convex in S	$\widehat{T}_n^S = 2.06$	0.040^{**}
$\mathcal{H}_0^A : \frac{\partial \phi}{\partial s}$ is constant across A	$\mathcal{H}_1^A : \frac{\partial \phi}{\partial s}$ varies with A	$\widehat{T}_n^A = 0.00$	0.612

TABLE 3.2. Results of specification tests for smoking and infant birth weights.

significantly positive. Hence, we reject $\mathcal{H}_0^S$ in favor of $\mathcal{H}_{+1}^S$, that is, convexity. Second, consider a test of the hypothesis $\mathcal{H}_0^A : \frac{\partial}{\partial s}\phi(5, a) = \beta_s$ for some constant β_s for all $a \in \text{supp}(A)$. Table 3.2 shows that the test statistic is not significant. We therefore fail to reject $\mathcal{H}_0^A$.

These results imply that the birth-weight production function is convex in the number of cigarettes, and the shapes of these functions are perhaps homogeneous

across individuals. In other words, the negative marginal effects of smoking diminish in number of cigarettes, but may not vary across unobserved physiological characteristics of mothers. The structural specifications of (i) the constant coefficient model and (iii) the linear random coefficient model are rejected, whereas (ii) the nonlinear separable model survived our attempt at rejection.

5. Conclusion

This paper proposed methods of specification testing that are effective even when instruments exhibit only discrete variations as in the case of many empirical data. To reflect this limitation, we developed an idea to use partially identified parameters to construct test statistics. The tests are designed to distinguish (i) the linear constant coefficient model, (ii) the nonlinear separable model, and (iii) the linear random coefficient model, against the alternative of (iv) the nonlinear nonseparable model. We showed the empirical relevance of the method. Specifications (i) and (ii) are rejected for log wages as a function of years of education. Specifications (i) and (iii) are rejected for infant birth weights as a function of smoking intensity.

6. Appendix: Well-Defined Conditional Expectations

Define $\Psi := \frac{\partial}{\partial s}\phi(S, A)$. Assume that there exists a function $h \in \mathcal{L}^1(F_{SZ})$ such that

$$\int_F h(s, z) dF_{SZ}(s, z) = \int_{\mathbb{R}\times F} \psi dF_{\Psi SZ}(\psi, s, z)$$

holds for every Borel set $F \in \mathcal{B}(\mathbb{R}^2)$. Under this assumption, we can define the conditional expectation

$$\mathrm{E}\left[\frac{\partial}{\partial s}\phi(S, A) \middle| (S, Z) = \ \cdot \ \right] := h,$$

which do not suffer from the Borel-Kolmogorov paradox. We similarly assume that other conditional expectations and conditional distributions used through this paper are well-defined.

7. Appendix: Covariance Matrices Γ and $\tilde{\Gamma}$

The covariance matrix Γ is given by

$$\begin{pmatrix} \Gamma_{12} & & & & & & \\ & \Gamma_{21} & \Gamma_{213} & & & O & \\ & \Gamma_{213} & \Gamma_{23} & & & & \\ & & & \Gamma_{32} & \Gamma_{324} & & \\ & & & \Gamma_{324} & \Gamma_{34} & & \\ & O & & & & \ddots & \\ & & & & & & \Gamma_{K,K-1} \end{pmatrix}$$

where

$$\begin{aligned} \Gamma_{k,k'} &:= \|K\|_2^2 \frac{\sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{k'}\right), z_{k'})}{f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{k'}\right), z_{k'}) \left[Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{k'}\right) - s\right]^2} \\ &\quad + \|K\|_2^2 \frac{\sigma^2(s, z_k)}{f_{SZ}(s, z_k) \left[Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{k'}\right) - s\right]^2} \\ \Gamma_{k,k',k''} &:= \|K\|_2^2 \frac{\sigma^2(s, z_k)}{f_{SZ}(s, z_k) \left[Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{k''}\right) - s\right] \left[Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{k'}\right) - s\right]} \end{aligned}$$

The covariance matrix $\tilde{\Gamma}$ is the middle $2(K-2) \times 2(K-2)$ elements of Γ.

8. Appendix: Auxiliary Lemmas

8.1. Lemma 18.

LEMMA 18. *Suppose that Assumption 12 holds. Then*

$$f_{V|SZ}\left(v \left| Q_{S|Z}(\theta \mid z'), z'\right.\right) = f_{V|SZ}\left(v \left| Q_{S|Z}(\theta \mid z), z\right.\right)$$

holds for all $\theta \in (0,1)$, for all $v \in supp(V)$, and for all $z, z' \in supp(Z)$.

PROOF. With this specification, we have

$$\begin{aligned} F_{S|Z}(\psi(z,v) \mid z) &= Pr(\psi(z,V) \leqslant \psi(z,v) \mid Z = z) \overset{(1)}{=} Pr(\psi(z,V) \leqslant \psi(z,v) \mid Z = z') \\ &\overset{(2)}{=} Pr(\psi(z',V) \leqslant \psi(z',v) \mid Z = z') = F_{S|Z}(\psi(z',v) \mid z') \end{aligned}$$

for all $(a,v) \in \text{supp}(A,V)$ and $z, z' \in \text{supp}(Z)$, where step (1) is due to (IV), and step (2) is due to (IM). But then, the random variable defined by $\Theta := F_{S|Z}(\psi(z,V) \mid z)$ does not depend on z, hence $\Theta = h(V)$ for some function h. Since $(V,\Theta) =$

$(V, h(A,V))$, (IV) implies $(V,\Theta) \perp\!\!\!\perp Z$. But then, $f_{V|\Theta} = \frac{f_{V\Theta|Z}}{f_\Theta} f_Z = \frac{f_{V\Theta}}{f_\Theta} f_Z = f_{V|\Theta} f_Z = f_{V|\Theta} f_{Z|\Theta}$, showing that $V \perp\!\!\!\perp Z \mid \Theta$.

By (AC), $F_{S|Z}(\cdot \mid z)$ is invertible on its support, and the map $\theta \mapsto F^{-1}_{S|Z}(\theta \mid z)$ is injective. But then, so is the map $(\theta, z) \mapsto \left(F^{-1}_{S|Z}(\theta \mid z), z\right)$. This implies that $f_{V|SZ}\left(v \left| F^{-1}_{S|Z}(\theta \mid z), z\right.\right) = f_{V|\Theta Z}(v \mid \theta, z)$ holds for all a v, θ, and z. Now apply $V \perp\!\!\!\perp Z \mid \Theta$ to this equality to get $f_{V|SZ}\left(v \left| F^{-1}_{S|Z}(\theta \mid z), z\right.\right) = f_{V|\Theta Z}(v \mid \theta, z) = f_{V|\Theta}(v \mid \theta)$. Therefore, it follows that

$$f_{V|SZ}\left(v \left| F^{-1}_{S|Z}(\theta \mid z'), z'\right.\right) - f_{V|SZ}\left(v \left| F^{-1}_{S|Z}(\theta \mid z), z\right.\right) = f_{V|\Theta}(v \mid \theta) - f_{V|\Theta}(v \mid \theta) = 0.$$

□

8.2. Lemma 19.

Lemma 19. *Suppose that (IV) holds. Then,*

$$E[Y \mid S = s, Z = z] = \int \phi(s,a) f_{A|V}(a \mid v) f_{V|SZ}(v \mid s, z) d(a, v)$$

holds for all (a, s, v, z) for which the conditional distributions are well-defined.

Proof. First, note that $f_{A|VSZ}(a \mid v, s, z) = f_{A|VSZ}(a \mid v, \psi(z, v), z) = f_{A|VZ}(a \mid v, z)$ holds on the support of $f_{V|SZ}$. Using this fact, we obtain

$$\mathrm{E}[Y \mid S = s, Z = z] \;=\; \int\int \phi(s,a) f_{A|VZ}(a \mid v, z) f_{V|SZ}(v \mid s, z) da dv$$

Next, using (IV) reduces $f_{A|VZ}$ into $f_{A|V}$, hence proving the lemma. □

8.3. Lemma 20.

Lemma 20. *Suppose that Assumption 12 holds. If $z, z' \in supp(Z)$ and $\theta \in (0,1)$, then*

$$E[\phi(s', A) - \phi(s, A) \mid S = s, Z = z] \;=\; E[Y \mid S = s', Z = z'] - E[Y \mid S = s, Z = z] \text{ and}$$
$$E[\phi(s', A) - \phi(s, A) \mid S = s', Z = z'] \;=\; E[Y \mid S = s', Z = z'] - E[Y \mid S = s, Z = z]$$

hold, where $s = Q_{S|Z}(\theta \mid z)$ and $s' = Q_{S|Z}(\theta \mid z')$.

PROOF. First, note that the convex support condition in (AC) guarantees the invertibility of $F_{S|Z}(\cdot \mid z)$, hence $F^{-1}_{S|Z}(\cdot \mid z)$ is well-defined on $(0,1)$. For notational simplicity, write

$$\Lambda(z', z, \theta) := \int\int \phi(F^{-1}_{S|Z}(\theta \mid z'), a) f_{A|V}(a \mid v) f_{V|SZ}(v \mid F^{-1}_{S|Z}(\theta \mid z), z) da dv.$$

Then, Lemma 19 states

$$\mathrm{E}[Y \mid S = F^{-1}_{S|Z}(\theta \mid z'), Z = z'] = \Lambda(z', z', \theta) \quad \text{and} \quad \mathrm{E}[Y \mid S = F^{-1}_{S|Z}(\theta \mid z), Z = z] = \Lambda(z, z, \theta).$$

On the other hand, Lemma 18 implies $\Lambda(z', z', \theta) - \Lambda(z', z, \theta) = 0$ and $\Lambda(z, z', \theta) - \Lambda(z, z, \theta) = 0$. Combining these two results yields

$$\begin{aligned} & \mathrm{E}[Y \mid S = F^{-1}_{S|Z}(\theta \mid z'), Z = z'] - \mathrm{E}[Y \mid S = F^{-1}_{S|Z}(\theta \mid z), Z = z] \\ = \; & \underbrace{\Lambda(z', z', \theta) - \Lambda(z', z, \theta)}_{0} + \underbrace{\Lambda(z', z, \theta) - \Lambda(z, z, \theta)}_{(*)} \end{aligned}$$

where the $(*)$ part is

$$\begin{aligned} (*) &= \int\int \left[\phi(F^{-1}_{S|Z}(\theta \mid z'), a) - \phi(F^{-1}_{S|Z}(\theta \mid z), a)\right] f_{A|V}(a \mid v) f_{V|SZ}(v \mid F^{-1}_{S|Z}(\theta \mid z), z) da dv \\ &= \mathrm{E}\left[\phi(F^{-1}_{S|Z}(\theta \mid z'), A) - \phi(F^{-1}_{S|Z}(\theta \mid z), A) \middle| S = F^{-1}_{S|Z}(\theta \mid z), Z = z\right] \end{aligned}$$

Now, substitute $s' = F^{-1}_{S|Z}(\theta \mid z')$ and $s = F^{-1}_{S|Z}(\theta \mid z)$ to obtain

$$\mathrm{E}[Y \mid S = s', Z = z'] - \mathrm{E}[Y \mid S = s, Z = z] = \mathrm{E}[\phi(s', A) - \phi(s, A) \mid S = s, Z = z].$$

Similarly, we write

$$\begin{aligned} & \mathrm{E}[Y \mid S = F^{-1}_{S|Z}(\theta \mid z'), Z = z'] - \mathrm{E}[Y \mid S = F^{-1}_{S|Z}(\theta \mid z), Z = z] \\ = \; & \underbrace{\Lambda(z', z', \theta) - \Lambda(z, z', \theta)}_{(**)} + \underbrace{\Lambda(z, z', \theta) - \Lambda(z, z, \theta)}_{0} \end{aligned}$$

where the $(**)$ part is

$$\begin{aligned} (**) &= \int\int \left[\phi(F^{-1}_{S|Z}(\theta \mid z'), a) - \phi(F^{-1}_{S|Z}(\theta \mid z), a)\right] f_{A|V}(a \mid v) f_{V|SZ}(v \mid F^{-1}_{S|Z}(\theta \mid z'), z') da dv \\ &= \mathrm{E}\left[\phi(F^{-1}_{S|Z}(\theta \mid z'), A) - \phi(F^{-1}_{S|Z}(\theta \mid z), A) \middle| S = F^{-1}_{S|Z}(\theta \mid z'), Z = z'\right] \end{aligned}$$

Now, substitute $s' = F^{-1}_{S|Z}(\theta \mid z')$ and $s = F^{-1}_{S|Z}(\theta \mid z)$ to obtain

$$\mathrm{E}[Y \mid S = s', Z = z'] - \mathrm{E}[Y \mid S = s, Z = z] = \mathrm{E}[\phi(s', A) - \phi(s, A) \mid S = s', Z = z'].$$

□

8.4. Lemma 21.

LEMMA 21 (Monotonicity). *(i) Suppose that (IV) and (SI) hold. Then, for a fixed* $\theta \in (0,1)$, $Q_{S|Z}(\theta \mid z)$ *is increasing in* z. *(ii) If in addition (AC) holds, then* $\theta \in (0,1)$, $Q_{S|Z}(\theta \mid z)$ *is strictly increasing in* z.

PROOF. Let $z' > z$. Since ψ is increasing by (SI), $Pr(\psi(z,V) \leqslant s) \geqslant Pr(\psi(z',V) \leqslant s)$ for all s. But since $Pr(\psi(z,V) \leqslant s) = Pr(\psi(Z,V) \leqslant s \mid Z = z) = F_{S|Z}(s \mid z)$ and similarly $Pr(\psi(z',V) \leqslant s) = F_{S|Z}(s \mid z')$ by (IV), the above inequality reduces to $F_{S|Z}(s \mid z) \geqslant F_{S|Z}(s \mid z')$ for all s. This inequality in turn yields the set relation $\left\{s \left| F_{S|Z}(s \mid z) \geqslant \theta \right.\right\} \supset \left\{s \left| F_{S|Z}(s \mid z') \geqslant \theta \right.\right\}$. But then,

$$F^{-1}_{S|Z}(\theta \mid z) = \inf\left\{s \left| F_{S|Z}(s \mid z) \geqslant \theta \right.\right\} \leqslant \inf\left\{s \left| F_{S|Z}(s \mid z') \geqslant \theta \right.\right\} = F^{-1}_{S|Z}(\theta \mid z')$$

which proves part (i).

To prove part (ii), assume (AC) in addition. Note that (AC) implies the absolute continuity of the distribution of $\psi(z,V)$ with a convex support for each $z \in \text{supp}(Z)$. But then, $z < z'$ and (SI) yield $Pr(\psi(z,V) \leqslant s) > Pr(\psi(z',V) \leqslant s)$, hence $\theta := F_{S|Z}(s \mid z) > F_{S|Z}(s \mid z')$ by the same argument as in part (i). Since (AC) implies $\frac{\partial}{\partial\theta} F^{-1}_{S|Z}(\theta \mid z) > 0$, we obtain

$$F^{-1}_{S|Z}(\theta \mid z') = F^{-1}_{S|Z}(F_{S|Z}(F^{-1}_{S|Z}(\theta \mid z') \mid z) \mid z) > F^{-1}_{S|Z}(F_{S|Z}(F^{-1}_{S|Z}(\theta \mid z') \mid z') \mid z) = F^{-1}_{S|Z}(\theta \mid z),$$

which proves part (ii). □

8.5. Lemma 22.

LEMMA 22 (Asymptotic Distribution). *Suppose that Assumption 13 holds. If* $z' \neq z$ *and* $Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right) \neq s$, *then* $\sqrt{nh_n}\left(\widehat{B}(s;z',z) - B(s;z',z)\right) \xrightarrow{d} \xi \sim N(0,V)$, *where*

$$V := \|K\|_2^2 \frac{f_{SZ}(s,z)\sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right), z') + f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right), z')\sigma^2(s,z)}{f_{SZ}(s,z) f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right), z')\left[Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right) - s\right]^2}$$

PROOF. By Assumption 13 (i), $\widehat{q}(s,z;z') \xrightarrow{p} Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)$ is granted. Therefore, it suffices to show convergence of the numerator of $\widehat{B}(s;z',z)$ to that of $B(s;z'z)$ in law. I use some variants of the standard asymptotic theories of kernel-based estimators (e.g., Pagan and Ullah, 1999).

First, the Lyapunov's CLT with Assumption 13 (ii)–(vi) yields

$$\begin{aligned}\sqrt{nh_n}\left(\mathbb{E}_{n,h}[Y \mid S=s, Z=z] - \mathrm{E}[Y \mid S=s, Z=z]\right) &= \frac{\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n K\left(\frac{S_i-s}{h_n}\right)\mathbb{1}\left(Z_i=z\right)\varepsilon_i}{\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{S_i-s}{h_n}\right)\mathbb{1}\left(Z_i=z\right)} \\ &= \frac{\frac{1}{\sqrt{n(z)h_n}}\sum_{i=1}^n K\left(\frac{S_i-s}{h_n}\right)\mathbb{1}\left(Z_i=z\right)\varepsilon_i}{\sqrt{\frac{n(z)}{n}}\frac{1}{n(z)h_n}\sum_{i=1}^n K\left(\frac{S_i-s}{h_n}\right)\mathbb{1}\left(Z_i=z\right)} \xrightarrow{d} \frac{\xi_1}{\sqrt{f_Z(z)}f_{S|Z}(s \mid z)}\end{aligned}$$

where $n(z) := \sum_{i=1}^n \mathbb{1}\left(Z_i=z\right)$ and $\xi_1 \sim N\left(0, f_{S|Z}(s \mid z)\left\|K\right\|_2^2\sigma^2(s,z)\right)$. Similarly, we have

$$\begin{aligned}&\sqrt{nh_n}\left(\mathbb{E}_{n,h}[Y \mid S=\widehat{q}(s,z;z'), Z=z'] - \mathrm{E}[Y \mid S=Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right), Z=z']\right) \\ &= \frac{\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n K\left(\frac{S_i-\widehat{q}(s,z;z')}{h_n}\right)\mathbb{1}\left(Z_i=z'\right)\varepsilon_i}{\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{S_i-\widehat{q}(s,z;z')}{h_n}\right)\mathbb{1}\left(Z_i=z'\right)} + \\ &\underbrace{\sqrt{nh_n}\left(\mathrm{E}[Y \mid S=\widehat{q}(s,z;z'), Z=z'] - \mathrm{E}[Y \mid S=Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right), Z=z']\right)}_{=o_p(1)}\end{aligned}$$

First, we show that the last term is $o_p(1)$. To see this, note that by the mean value expansion together with Assumption 13 (i), (ii), and (vii), we have

$$\begin{aligned}&\sqrt{nh_n}\left(\mathrm{E}[Y \mid S=\widehat{q}(s,z;z'), Z=z'] - \mathrm{E}[Y \mid S=Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right), Z=z']\right) \\ &\qquad = n^{1/2}h_n^{1/2}O_p(r_n^{-1}) = O_p((nh_nr_n^{-2})^{1/2}) = o_p(1).\end{aligned}$$

To study the asymptotic behavior of

$$\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n K\left(\frac{S_i-\widehat{q}(s,z;z')}{h_n}\right)\mathbb{1}\left(Z_i=z'\right)\varepsilon_i$$

rewrite it as

$$\underbrace{\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n} K\left(\frac{S_i - Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)}{h_n}\right) \mathbb{1}\left(Z_i = z'\right)\varepsilon_i}_{\xrightarrow{d}\sqrt{f_Z(z')}\xi_2}$$

$$+\underbrace{\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left[K\left(\frac{S_i - \widehat{q}(s,z;z')}{h_n}\right) - K\left(\frac{S_i - Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)}{h_n}\right)\right]\mathbb{1}\left(Z_i = z'\right)\varepsilon_i}_{o_p(1)}$$

Again, the Lyapunov's CLT with Assumption 13 (ii)–(vi) yields convergence of the first term as

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n} K\left(\frac{S_i - Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)}{h_n}\right)\mathbb{1}\left(Z_i = z'\right)\varepsilon_i \\
= \;& \sqrt{\frac{n(z')}{n}}\frac{1}{\sqrt{n(z')h_n}}\sum_{i=1}^{n} K\left(\frac{S_i - Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)}{h_n}\right)\mathbb{1}\left(Z_i = z'\right)\varepsilon_i \xrightarrow{d} \sqrt{f_Z(z')}\xi_2
\end{aligned}$$

where $\xi_2 \sim N\left(0, f_{S|Z}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right) \mid z')\left\|K\right\|_2^2 \sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right), z')\right)$. On the other hand,

$$\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left[K\left(\frac{S_i - \widehat{q}(s,z;z')}{h_n}\right) - K\left(\frac{S_i - Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)}{h_n}\right)\right]\mathbb{1}\left(Z_i = z'\right)\varepsilon_i$$

is $o_p(1)$ To see this, let M denote the Lipschitz constant of K as granted by Assumption 13 (iii), and note that

$$\begin{aligned}
&\left|K\left(\frac{S_i - \widehat{q}(s,z;z')}{h_n}\right) - K\left(\frac{S_i - Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)}{h_n}\right)\right| \\
&\qquad \leqslant M\left|\frac{\widehat{q}(s,z;z') - Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)}{h_n}\right| = O_p(h_n^{-1}r_n^{-1})
\end{aligned}$$

by Assumption 13 (i). Note that this expression is independent of the subscript i. Another application of the Lyapunov's CLT with Assumption 13 (ii)–(vi) yields

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{1}\left(Z_i = z'\right)\varepsilon_i = \sqrt{\frac{n(z')}{n}}\frac{1}{\sqrt{n(z')}}\sum_{i=1}^{n}\mathbb{1}\left(Z_i = z'\right)\varepsilon_i = O_p(1),$$

Hence, it follows that

$$\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left[K\left(\frac{S_i-\widehat{q}(s,z;z')}{h_n}\right)-K\left(\frac{S_i-Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right)}{h_n}\right)\right]\mathbb{1}\left(Z_i=z'\right)\varepsilon_i$$
$$=h_n^{-1/2}O_p(h_n^{-1}r_n^{-1})O_p(1)=O_p\left((h_n^3r_n^2)^{-1/2}\right)=o_p(1)$$

by Assumption 13 (ii), as desired. A similar decomposition method will show

$$\frac{1}{nh_n}\sum_{i=1}^{n}K\left(\frac{S_i-\widehat{q}(s,z;z')}{h_n}\right)\mathbb{1}\left(Z_i=z'\right)$$
$$=\frac{n(z')}{n}\frac{1}{n(z')h_n}\sum_{i=1}^{n}K\left(\frac{S_i-\widehat{q}(s,z;z')}{h_n}\right)\mathbb{1}\left(Z_i=z'\right)\xrightarrow{p}f_Z(z')f_{S|Z}(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right)\mid z')$$
$$=f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),z')$$

Putting all these pieces together, we obtain

$$\sqrt{nh_n}\left(\mathbb{E}_{n,h}[Y\mid S=\widehat{q}(s,z;z'),Z=z']-\mathrm{E}[Y\mid S=Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),Z=z']\right)$$
$$\xrightarrow{d}\frac{\sqrt{f_Z(z')}\xi_2}{f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),z')}$$

where $\xi_2\sim N\left(0,f_{S|Z}(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right)\mid z')\left\|K\right\|_2^2\sigma^2(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),z')\right)$. Since $z\neq z'$, $Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right)\neq s$, $h_n\to 0$, and $\mathrm{supp}(K)\subset(-1,1)$ by assumption,

$$\sqrt{nh_n}\left(\mathbb{E}_{n,h}[Y\mid S=s,Z=z]-\mathrm{E}[Y\mid S=s,Z=z]\right)$$

and

$$\sqrt{nh_n}\left(\mathbb{E}_{n,h}[Y\mid S=\widehat{q}(s,z;z'),Z=z']-\mathrm{E}[Y\mid S=Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),Z=z']\right)$$

are asymptotically independent. Using this fact, we see that

$$\begin{aligned}\widehat{\xi}\;:=\;&\sqrt{nh_n}((\mathbb{E}_{n,h}[Y\mid S=\widehat{q}(s,z;z'),Z=z']-\mathbb{E}_{n,h}[Y\mid S=s,Z=z])\\&-(\mathrm{E}[Y\mid S=Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),Z=z']-\mathrm{E}[Y\mid S=s,Z=z]))\\\xrightarrow{d}\xi:=&\frac{\sqrt{f_Z(z')}\xi_2}{f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),z')}-\frac{\sqrt{f_Z(z)}\xi_1}{f_{SZ}(s,z)}\end{aligned}$$

where

$$\xi\sim N\left(0,\frac{\left\|K\right\|_2^2\sigma^2(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),z')}{f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right),z')}+\frac{\left\|K\right\|_2^2\sigma^2(s,z)}{f_{SZ}(s,z)}\right)$$

Lastly, noting Assumption 13 (i) and (ii) yields

$$\begin{aligned}
& \sqrt{nh_n}\left(\widehat{B}(s;z',z) - B(s;z',z)\right) \\
= \quad & \frac{\widehat{\xi}}{\widehat{q}(s,z;z') - s} + \frac{\sqrt{nh_n}\left(\widehat{q}(s,z;z') - Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right)\right)}{\left(\widehat{q}(s,z;z') - s\right)\left(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right) - s\right)} \\
= \quad & \frac{\widehat{\xi}}{\widehat{q}(s,z;z') - s} + \frac{O_p((nh_n r_n^{-2})^{1/2})}{\left(\widehat{q}(s,z;z') - s\right)\left(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right) - s\right)} \\
\xrightarrow{d} \quad & \frac{\xi}{Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right) - s} \sim N(0, V)
\end{aligned}$$

where

$$V = \frac{\frac{\|K\|_2^2 \sigma^2(Q_{S|Z}\left(F_{S|Z}(s|z)|z'\right),z')}{f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s|z)|z'\right),z')} + \frac{\|K\|_2^2 \sigma^2(s,z)}{f_{SZ}(s,z)}}{\left[Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z'\right) - s\right]^2}$$

□

8.6. Lemma 23.

Lemma 23 (Asymptotic Joint Distribution). *Suppose that Assumption 13 holds. If* $z_* < z < z^*$ *and* $Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z_*\right) < s < Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z^*\right)$*, then*

$$\sqrt{nh_n}\begin{pmatrix} \widehat{B}(s;z^*,z) - B(s;z^*,z) \\ \widehat{B}(s;z_*,z) - B(s;z_*,z) \end{pmatrix} \xrightarrow{d} b \sim N\left(0, \begin{pmatrix} \Sigma_{11}(s,z) & \Sigma_{12}(s,z) \\ \Sigma_{12}(s,z) & \Sigma_{22}(s,z) \end{pmatrix}\right)$$

where

$$\begin{aligned}
\Sigma_{11}(s,z) &:= \|K\|_2^2 \frac{f_{SZ}(s,z)\sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z^*\right), z^*) + f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z^*\right), z^*)\sigma^2(s,z)}{f_{SZ}(s,z) f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z^*\right), z^*)\left[Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z^*\right) - s\right]^2} \\
\Sigma_{12}(s,z) &:= \|K\|_2^2 \frac{\sigma^2(s,z)}{f_{SZ}(s,z)\left[Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z^*\right) - s\right]\left[Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z_*\right) - s\right]} \\
\Sigma_{22}(s,z) &:= \|K\|_2^2 \frac{f_{SZ}(s,z)\sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z_*\right), z_*) + f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z_*\right), z_*)\sigma^2(s,z)}{f_{SZ}(s,z) f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z_*\right), z_*)\left[Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z_*\right) - s\right]^2}
\end{aligned}$$

Proof. By using a similar argument to the proof of Lemma 22 (i), we obtain

$$\sqrt{nh_n}\begin{pmatrix} \widehat{B}(s;z^*,z) - B(s;z^*,z) \\ \widehat{B}(s;z_*,z) - B(s;z_*,z) \end{pmatrix} = \widehat{Q}(s,z;z^*,z_*)^{-1}\begin{pmatrix} \widehat{\xi}^* \\ \widehat{\xi}_* \end{pmatrix} + O_p((nh_n r_n^{-2})^{1/2})$$

where

$$\begin{aligned}\widehat{Q}(s,z;z^*,z_*) &:= diag\left(\widehat{q}(s,z;z^*)-s,\widehat{q}(s,z;z_*)-s\right)\\ &\xrightarrow{p} diag\left(Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z^*\right)-s,Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z_*\right)-s\right)\end{aligned}$$

is nonsingular with probability approaching one under the assumption that $Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z_*\right) < s < Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z^*\right)$, and

$$\begin{aligned}\widehat{\xi}^* &:= \sqrt{nh_n}((\mathbb{E}_{n,h}[Y\mid S=\widehat{q}(s,z;z^*),Z=z^*]-\mathbb{E}_{n,h}[Y\mid S=s,Z=z])\\ &\quad -(\mathrm{E}[Y\mid S=Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z^*\right),Z=z^*]-\mathrm{E}[Y\mid S=s,Z=z]))\\ \widehat{\xi}_* &:= \sqrt{nh_n}((\mathbb{E}_{n,h}[Y\mid S=\widehat{q}(s,z;z_*),Z=z_*]-\mathbb{E}_{n,h}[Y\mid S=s,Z=z])\\ &\quad -(\mathrm{E}[Y\mid S=Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z_*\right),Z=z_*]-\mathrm{E}[Y\mid S=s,Z=z]))\end{aligned}$$

Since $z\neq z'$, $Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z'\right)\neq s$, $h_n\to 0$, and $\mathrm{supp}(K)\subset(-1,1)$ by assumption,

$$\sqrt{nh_n}\left(\mathbb{E}_{n,h}[Y\mid S=s,Z=z]-\mathrm{E}[Y\mid S=s,Z=z]\right),$$

$$\sqrt{nh_n}\left(\mathbb{E}_{n,h}[Y\mid S=\widehat{q}(s,z;z^*),Z=z^*]-\mathrm{E}[Y\mid S=Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z^*\right),Z=z^*]\right),$$

and

$$\sqrt{nh_n}\left(\mathbb{E}_{n,h}[Y\mid S=\widehat{q}(s,z;z_*),Z=z_*]-\mathrm{E}[Y\mid S=Q_{S|Z}\left(F_{S|Z}(s\mid z)\mid z_*\right),Z=z_*]\right)$$

are asymptotically independent. Therefore, by a similar argument to the proof of Lemma 22 (i), the Lyapunov's CLT together with Assumption 13 (ii)-(vi) yields

$$\begin{pmatrix}\widehat{\xi}^*\\ \widehat{\xi}_*\end{pmatrix}\xrightarrow{p}\begin{pmatrix}\xi^*\\ \xi_*\end{pmatrix}\sim N\left(0,\begin{pmatrix}\Lambda^1_{11}(s,z) & \Lambda^1_{12}(s,z)\\ \Lambda^1_{12}(s,z) & \Lambda^1_{22}(s,z)\end{pmatrix}\right)$$

where

$$\begin{aligned}
\Lambda^1_{11}(s,z) &:= \frac{\|K\|_2^2\,\sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z^*\right), z^*)}{f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z^*\right), z^*)} + \frac{\|K\|_2^2\,\sigma^2(s,z)}{f_{SZ}(s,z)} \\
\Lambda^1_{12}(s,z) &:= \frac{\|K\|_2^2\,\sigma^2(s,z)}{f_{SZ}(s,z)} \\
\Lambda^1_{22}(s,z) &:= \frac{\|K\|_2^2\,\sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z_*\right), z_*)}{f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z) \mid z_*\right), z_*)} + \frac{\|K\|_2^2\,\sigma^2(s,z)}{f_{SZ}(s,z)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sqrt{nh_n}\begin{pmatrix} \widehat{B}(s;z^*,z) - B(s;z^*,z) \\ \widehat{B}(s;z_*,z) - B(s;z_*,z) \end{pmatrix} &\xrightarrow{d} \operatorname{plim}\left(\widehat{Q}(s,z;z^*,z_*)\right)^{-1}\begin{pmatrix} \xi^* \\ \xi_* \end{pmatrix} \\
&\sim N\left(0, \begin{pmatrix} \Sigma_{11}(s,z) & \Sigma_{12}(s,z) \\ \Sigma_{12}(s,z) & \Sigma_{22}(s,z) \end{pmatrix}\right)
\end{aligned}$$

□

8.7. Lemma 24.

LEMMA 24 (Asymptotic Joint Distribution). *Suppose that Assumption 13 holds for s and all $z \in supp(Z)$, and assume that $Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right) \neq Q_{S|Z}\left(F_{S|Z}(s \mid z_{k'}) \mid z_j\right)$ for $k \neq k'$, $k \neq i$, and $k' \neq j$. If $supp(Z) = \{z_1, \ldots, z_K\}$ with $z_1 < \cdots < z_K$ and $s \in supp(S)$, then*

$$\begin{aligned}
\sqrt{nh_n} \cdot vec\begin{pmatrix} \widehat{B}(s;z_2,z_1) - B(s;z_2,z_1) & \cdots & \widehat{B}(s;z_1,z_K) - B(s;z_1,z_K) \\ \vdots & \ddots & \vdots \\ \widehat{B}(s;z_K,z_1) - B(s;z_K,z_1) & \cdots & \widehat{B}(s;z_{K-1},z_K) - B(s;z_{K-1},z_K) \end{pmatrix}_{(K-1)\times(K-1)} \\
\xrightarrow{d} b \sim N\left(0, \begin{pmatrix} \Sigma(z_1) & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & \Sigma(z_K) \end{pmatrix}_{(K-1)^2\times(K-1)^2}\right)
\end{aligned}$$

where for each $k = 1, \cdots, K$, the (i,i)–element of $\Sigma(z_k)$ is

$$\|K\|_2^2\,\frac{f_{SZ}(s,z_k)\sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right), z_i) + f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right), z_i)\sigma^2(s,z_k)}{f_{SZ}(s,z_k) f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right), z_i)\left[Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right) - s\right]^2}$$

for $i = 1, \cdots, k-1$, the (i,i)–element of $\Sigma(z_k)$ is

$$\|K\|_2^2\,\frac{f_{SZ}(s,z_k)\sigma^2(Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{i+1}\right), z_{i+1}) + f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{i+1}\right), z_{i+1})\sigma^2(s,z_k)}{f_{SZ}(s,z_k) f_{SZ}(Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{i+1}\right), z_{i+1})\left[Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_{i+1}\right) - s\right]^2}$$

for $i = k+1, \cdots, K-1$, the (i,j)–element of $\Sigma(z_k)$ is

$$\|K\|_2^2 \frac{\sigma^2(s, z_k)}{f_{SZ}(s, z_k)\left[Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right) - s\right]\left[Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_j\right) - s\right]}$$

for $i \neq j$ with $i, j < k$ and similarly for other cases.

PROOF. The upper $(K-1) \times K$ block of the left-hand-side matrix consists of elements of the form

$$\sqrt{nh_n}\left(\widehat{B}(s; z_i, z_k) - B(s; z_i, z_k)\right) = \frac{\widehat{\xi}_{ik}}{\widehat{q}(s, z_k; z_i) - s} + O_p((nh_n r_n^{-2})^{1/2})$$

where, as in the proof of Lemma 23,

$$\widehat{q}(s, z_k; z_i) - s \xrightarrow{p} Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right) - s \neq 0$$

and

$$\begin{aligned}\widehat{\xi}_{ik} \quad := \quad & \sqrt{nh_n}((\mathbb{E}_{n,h}[Y \mid S = \widehat{q}(s, z_k; z_i), Z = z_i] - \mathbb{E}_{n,h}[Y \mid S = s, Z = z_k]) \\ & -(\mathrm{E}[Y \mid S = Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right), Z = z_i] - \mathrm{E}[Y \mid S = s, Z = z_k]))\end{aligned}$$

for which Assumption 13 (ii)–(vi) facilitated a sufficient condition for the Lyapunov's CLT to be invoked. It remains to investigate in the elements of the variance-covariance matrix, but this follows from the same argument as the proof of Lemma 23.

Lastly, we show that all the elements off the block diagonal are zero. To this end, it suffices to observe asymptotic independence between $\widehat{\xi}_{ik}$ and $\widehat{\xi}_{jk'}$ for $k \neq k'$. But this asymptotic independence clearly holds by noting that the data is i.i.d., $h_n \to 0$ as $n \to \infty$, the assumption of the lemma that $Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right) \neq Q_{S|Z}\left(F_{S|Z}(s \mid z_{k'}) \mid z_j\right)$, and that $k \neq k'$ implies $z_k \neq z_{k'}$, $s \neq Q_{S|Z}\left(F_{S|Z}(s \mid z_k) \mid z_i\right)$ for all $i \neq k$, and $s \neq Q_{S|Z}\left(F_{S|Z}(s \mid z_{k'}) \mid z_j\right)$ for all $j \neq k'$. □

9. Appendix: Proofs of the Theorem and the Propositions

9.1. Proof of Theorem 5.

PROOF. We prove the statement for the case of Assumption 1 (DR). Other cases can be similarly proved by simply replacing inequalities. Let $z' := z_{s,z}^+$ and $s' =$

$F_{S|Z}^{-1}\left(F_{S|Z}(s \mid z) \mid z'\right)$. Since $s = F_{S|Z}^{-1}\left(F_{S|Z}(s \mid z) \mid z\right)$, Lemma 21 implies $s' > s$. Then (DR) or (CR) implies

$$\frac{\phi(s', a) - \phi(s, a)}{s' - s} < \frac{\partial}{\partial S}\phi(s, a)$$

for all a. But then,

$$\begin{aligned} \frac{\mathrm{E}\left[\phi(s', A) - \phi(s, A) \mid S = s, Z = z\right]}{s' - s} &= \int \frac{\phi(s', a) - \phi(s, a)}{s' - s} f_{A|SZ}(a \mid s, z)da \\ < \int \frac{\partial}{\partial S}\phi(s, a) f_{A|SZ}(a \mid s, z)da &= \mathrm{E}\left[\left.\frac{\partial}{\partial S}\phi(s, A)\right| S = s, Z = z\right]. \end{aligned}$$

Moreover, Lemma 20 yields

$$\mathrm{E}\left[\phi(s', A) - \phi(s, A) \mid S = s, Z = z\right] = \mathrm{E}[Y \mid S = s', Z = z'] - \mathrm{E}[Y \mid S = s, Z = z]$$

since $s = F_{S|Z}^{-1}(\theta \mid z)$ and $s' = F_{S|Z}^{-1}(\theta \mid z')$ where $\theta = F_{S|Z}(s \mid z)$. Substituting this equality into the last inequality yields

$$\frac{\mathrm{E}[Y \mid S = s', Z = z'] - \mathrm{E}[Y \mid S = s, Z = z]}{s' - s} < \mathrm{E}\left[\left.\frac{\partial}{\partial S}\phi(s, A)\right| S = s, Z = z\right].$$

Next, let $z'' := z_{s,z}^{-}$ and $s'' = F_{S|Z}^{-1}\left(F_{S|Z}(s \mid z) \mid z''\right)$. Since $s = F_{S|Z}^{-1}\left(F_{S|Z}(s \mid z) \mid z\right)$, Lemma 21 implies $s'' < s$. Using (DR) or (CR) and a similar argument to the previous paragraph, we obtain

$$\mathrm{E}\left[\left.\frac{\partial}{\partial S}\phi(s, A)\right| S = s, Z = z\right] < \frac{\mathrm{E}[Y \mid S = F_{S|Z}^{-1}\left(F_{S|Z}(s \mid z) \mid z''\right), Z = z''] - \mathrm{E}[Y \mid S = s, Z = z]}{F_{S|Z}^{-1}\left(F_{S|Z}(s \mid z) \mid z''\right) - s}.$$

□

9.2. Proof of Proposition 4.

Proof. Under $\mathcal{H}_0^S$, which implies $\mathcal{H}_0^{S'}$, we have

$$\begin{aligned} \widehat{T}_n^S &:= \sqrt{nh_n}\frac{\widehat{B}(s; z_{k+1}, z_k) - \widehat{B}(s; z_{k-1}, z_k)}{\sqrt{\Gamma_{k,k-1} + \Gamma_{k,k+1} - 2\Gamma_{k,k-1,k+1}}} \\ &:= \sqrt{nh_n}\frac{\left(\widehat{B}(s; z_{k+1}, z_k) - B(s; z_{k+1}, z_k)\right) - \left(\widehat{B}(s; z_{k-1}, z_k) - B(s; z_{k-1}, z_k)\right)}{\sqrt{\Gamma_{k,k-1} + \Gamma_{k,k+1} - 2\Gamma_{k,k-1,k+1}}}. \end{aligned}$$

We have from Lemma 24 that

$$\sqrt{nh_n}\begin{pmatrix} \widehat{B}(s;z_1,z_2)-B(s;z_1,z_2) \\ \widehat{B}(s;z_3,z_2)-B(s;z_3,z_2) \\ \vdots \\ \widehat{B}(s;z_{K-2},z_{K-1})-B(s;z_{K-2},z_{K-1}) \\ \widehat{B}(s;z_K,z_{K-1})-B(s;z_K,z_{K-1}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} U_2 \\ L_2 \\ \vdots \\ U_{K-1} \\ L_{K-1} \end{pmatrix} \sim N(0,\tilde{\Gamma}).$$

Hence, the continuous mapping theorem yields

$$\sqrt{nh_n}\frac{\left(\widehat{B}(s;z_{k+1},z_k)-B(s;z_{k+1},z_k)\right)-\left(\widehat{B}(s;z_{k-1},z_k)-B(s;z_{k-1},z_k)\right)}{\sqrt{\Gamma_{k,k-1}+\Gamma_{k,k+1}-2\Gamma_{k,k-1,k+1}}}, \\ \xrightarrow{d} Z_k \sim N(0,1).$$

□

9.3. Proof of Proposition 5.

PROOF. First, note that for each realization of $\widehat{B}(s;\ \cdot\ ,\cdot\ )$, we have

$$\begin{aligned} \widehat{T}_n^A &= \sqrt{nh_n}\max_{k,k'}\left\{W_{k,k'}\left(\widehat{B}(s;z_{k+1},z_k)-\widehat{B}(s;z_{k'-1},z_{k'})\right)\right\} \\ &\leqslant \sqrt{nh_n}\max_{k,k'}\left\{W_{k,k'}\left(\widehat{B}(s;z_{k+1},z_k)-B(s;z_{k+1},z_k)\right)-W_{k,k'}\left(\widehat{B}(s;z_{k'-1},z_{k'})\right.\right. \\ &\quad \left.\left.-B(s;z_{k'-1},z_{k'})\right)\right\} \end{aligned}$$

under $\mathcal{H}_0^A$, since $\mathcal{H}_0^A$ implies $\mathcal{H}_0^{A''}$, which in turn requires $B(s;z_{k+1},z_k) \leqslant B(s;z_{k'-1},z_{k'})$ for each k,k'. Therefore, by denoting by T_n^* the random variable taking the form of the right hand side, we obtain $\widehat{T}_n^A \preceq T_n^*$, where $\preceq$ denotes the first-order stochastic dominance relation. Under the special state of $\mathcal{H}_0^{A''}$ in which $B(s;z_{k+1},z_k) = B(s;z_{k'-1},z_{k'})$ for each k,k', the above inequality holds with equality, hence $\widehat{T}_n^A = T_n^*$ under this least favorable state of $\mathcal{H}_0^A$.

Second, note that by Lemma 24, we have

$$\sqrt{nh_n}\begin{pmatrix} \widehat{B}(s;z_2,z_1)-B(s;z_2,z_1) \\ \widehat{B}(s;z_1,z_2)-B(s;z_1,z_2) \\ \widehat{B}(s;z_2,z_1)-B(s;z_2,z_1) \\ \vdots \\ \widehat{B}(s;z_{K-1},z_K)-B(s;z_{K-1},z_K) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} L_1 \\ U_2 \\ L_2 \\ \vdots \\ U_K \end{pmatrix} \sim N(0,\Gamma).$$

Therefore, by the Continuous Mapping Theorem, we have $T_n^* \xrightarrow{d} T^A$.

Since $\widehat{T}_n^A \preceq T_n^*$ under $\mathcal{H}_0^A$, we have

$$Pr\left(\widehat{T}_n^A \geqslant F_{T^A}^{-1}(1-\alpha) \mid H\right) \leqslant Pr\left(T_n^* \geqslant F_{T^A}^{-1}(1-\alpha) \mid H\right)$$

for any $\alpha \in (0,1)$ under $H \in \mathcal{H}_0^A$. Taking the supremum over $\mathcal{H}_0^A$ yields

$$\sup_{H\in\mathcal{H}_0^A} Pr\left(\widehat{T}_n \geqslant F_{T^A}^{-1}(1-\alpha) \mid H\right) \leqslant \sup_{H\in\mathcal{H}_0^A} Pr\left(T_n^* \geqslant F_{T^A}^{-1}(1-\alpha) \mid H\right)$$

Lastly, it follows from $T_n^* \xrightarrow{d} T$ that

$$\begin{aligned} \lim_{n\to\infty} \sup_{H\in\mathcal{H}_0^A} Pr\left(\widehat{T}_n \geqslant F_{T^A}^{-1}(1-\alpha) \mid H\right) &\leqslant \lim_{n\to\infty} \sup_{H\in\mathcal{H}_0^A} Pr\left(T_n^* \geqslant F_{T^A}^{-1}(1-\alpha) \mid H\right) \\ &= Pr\left(T \geqslant F_{T^A}^{-1}(1-\alpha)\right) = \alpha \end{aligned}$$

□

Bibliography

Adams, Peter, Michael D. Hurd, Daniel McFadden, Angela Merrill, and Tiago Ribeiro (2003) "Healthy, Wealthy and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status," *Journal of Econometrics,* Vol. 112 (1), pp. 3–56.

Aguirregabiria, Victor (2010) "Another Look at the Identification of Dynamic Discrete Decision Processes: An Application to Retirement Behavior," *Journal of Business and Economic Statistics,* Vol. 28 (2), pp. 201–218.

Aguirregabiria, Victor, and Pedro Mira (2007) "Sequential Estimation of Dynamic Discrete Games," *Econometrica,* Vol. 75 (1), pp. 1–53.

Ahn, Seung C. and Peter Schmidt (1995) "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics,* Vol. 68 (1), pp. 5–27.

Ai, Chunrong, and Xiaohong Chen (2003) "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica,* Vol. 71 (6), pp. 1795–1843.

Almond, Douglas (2006) "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population," *Journal of Political Economy,* Vol. 114 (4), pp. 672–712.

Almond, Douglas and Bhashkar Mazumder (2005) "The 1918 Influenza Pandemic and Subsequent Health Outcomes: An Analysis of SIPP Data," *American Economic Review,* Vol. 95 (2), Papers and Proceedings, pp. 258–262.

Altonji, Joseph G. and Rosa L. Matzkin (2005) "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, Vol. 73 (4), pp. 1053-1102.

Altonji, Joseph G., Hidehiko Ichimura, and Taisuke Otsu (2011) "Estimating Derivatives in Nonseparable Models with Limited Dependent Variables," *Econometrica*,

forthcoming.

Anderson, T.W. and Cheng Hsiao (1982) "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics,* Vol. 18 (1), pp. 47–82.

Andrews, Donald W.K. (2011) "Examples of L^2-Complete and Boundedly-Complete Distributions," Cowles Foundation Discussion Paper No. 1801.

Angrist, Joshua D. and Guido W. Imbens (1995) "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 431–442.

Angrist, Joshua D. and Alan B. Krueger (1991) "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, Vol. 106, No. 4, pp. 979–1014.

Angrist, Joshua D. and Alan B. Krueger (2001) "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 69–85.

Arcidiacono, Peter and Robert A. Miller (2011) "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, forthcoming.

Arellano, Manuel and Stephen Bond (1991) "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies,* Vol. 58 (2), pp. 277–297.

Arellano, Manuel and Stéphane Bonhomme (2009) "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," Cemmap Working Paper No. 22/09.

Arellano, Manuel and Olympia Bover (1995) "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics,* Vol. 68 (1), pp. 29–51.

Arellano, Manuel and Jinyong Hahn (2005) "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," Invited Lecture, Econometric Society World Congress, London, August 2005.

Bai, Jushan (2009) "Panel data models with interactive fixed effects," *Econometrica,* Vol 77 (4), pp. 1229–1279.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007) "Estimating Dynamic Models of Imperfect Competition," *Econometrica,* Vol. 75 (5), pp. 1331-1370.

Baltagi, Badi H. (1985) "Pooling Cross-Sections with Unequal Time-Series Lengths," *Economics Letters,* Vol. 18 (2), pp. 133–136.

Baltagi, Badi H. and Young-Jae Chang (1994) "Incomplete Panels: A Comparative Study of Alternative Estimators for the Unbalanced One-Way Error Component Regression Model," *Journal of Econometrics,* Vol. 62 (2), pp. 67–89.

Belzil, Christian and Jörgen Hansen (2002) "Unobserved Ability and the Return to Schooling," *Econometrica,* Vol. 70 (5), pp. 2075–2091.

Bester, Alan C. and Christian Hansen (2009) "A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects," *Journal of Business and Economic Statistics,* Vol. 27 (2), pp. 131-148.

Bhattacharya, Debopam (2008) "Inference in panel data models under attrition caused by unobservables," *Journal of Econometrics,* Vol. 144 (2), pp. 430–446.

Blundell, Richard and Stephen Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics,* Vol. 87 (1), pp. 115–143.

Blundell, Richard, Xiaohong Chen, and Dennis Kristensen (2007) "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica,* Vol. 75 (6), pp. 1613–1669.

Blundell, Richard., and James L. Powell (2003) "Endogeneity in Nonparametric and Semiparametric Regression Models," in Advances in Economics and Econometrics: Theory and Applications, Vol. 2, ed. by M. Dewatripont, L.-P. Hansen, and S. J. Turnovsky. Cambridge, U.K.: Cambridge University Press, pp. 312-357.

Blundell, Richard., and James L. Powell (2007) "Censored Regression Quantiles with Endogenous Regressors," *Journal of Econometrics*, Vol. 141, No. 1, pp. 65–83.

Bonhomme, Stéphane (2010) "Functional Differencing," Working Paper, New York University.

Bound, John, Todd Stinebrickner, and Timothy Waidmann (2010) "Health, Economic Resources and the Work Decisions of Older Men," *Journal of Econometrics,* Vol. 156, No. 1, pp. 106–129.

Brown, Stephen J., William Goetzmann, Roger G. Ibbotson, and Stephen A. Ross (1992) "Survivorship Bias in Performance Studies," *Review of Financial Studies,* Vol. 5 (4), pp. 553-580

Cameron, Stephen V. and James J. Heckman (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy,* Vol. 106 (2), pp. 262–333.

Card, David (2001) "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, Vol. 69, No. 5, pp. 1127–1160.

Carrasco, Marine, Jean-Pierre Florens, and Eric Renault (2007) "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in J.J. Heckman and E.E. Leamer (ed.) *Handbook of Econometrics*, Vol. 6, Ch. 77.

Case, Anne, Angela Fertig, and Christina Paxson (2005) "The Lasting Impact of Childhood Health and Circumstance," *Journal of Health Economics,* Vol. 24 (2), pp. 365–389.

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

Chen, Xiaohong (2007) "Large Sample Sieve Estimation of Semi-Nonparametric Models," in J.J. Heckman and E.E. Leamer (ed.) *Handbook of Econometrics*, Vol. 6, Ch. 76.

Chernozhukov, Victor and Christian Hansen (2005) "An IV Model of Quantile Treatment Effects," *Econometrica*, Vol. 73 (1), pp. 245–261.

Chernozhukov, Victor and Christian Hansen (2006) "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics*, Vol. 132, No. 2, pp. 491–525.

Chernozhukov, Victor, Guido W. Imbens, and Whitney K. Newey (2007) "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, Vol. 139 (1), pp. 4–14.

Chernozhukov, Victor, Sokbae Lee, and Adam M. Rosen (2009) "Intersection Bounds: Estimation and Inference," Working Paper

Chernozhukov, Victor, Iván Fernández-Val, Jinyong Hahn, and Whitney K. Newey (2009) "Identification and Estimation of Marginal Effects in Nonlinear Panel Models," CeMMAP Working Papers CWP05/09.

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

Chesher, Andrew (2003) "Identification in Nonseparable Models," *Econometrica*, Vol. 71 (5), pp. 1405–1441.

Chesher, Andrew (2005) "Nonparametric Identification under Discrete Variation," *Econometrica*, Vol. 73 (5), pp. 1525–1550.

Contoyannis, Paul, Andrew M. Jones, and Nigel Rice (2004) "The dynamics of health in the British Household Panel Survey," *Journal of Applied Econometrics,* Vol. 19 (4), pp. 473–503.

Crawford, Gregory S. and Matthew Shum (2005) "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica,* Vol. 73 (4), pp. 1137–1173.

Cunha, Flavio, and James J. Heckman (2008) "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources,* Vol. 43 (4), pp. 738–782.

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach (2010) "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica,* Vol. 78 (3), pp. 883-931.

Currie, Janet and Enrico Moretti (2003) "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence From College Openings," *Quarterly Journal of Economics,* Vol. 118 (4), pp. 1495–1532.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney (2006) "The Determinants of Mortality," *Journal of Economic Perspectives,* Vol. 20 (3), pp. 97–120.

Darolles, Serge, Yanqin Fan, Jean-Pierre Florens, and Eric Renault (2011) "Nonparametric Instrumental Regression," *Econometrica,* Vol. 79 (5), pp. 1541-1565.

Das, Mitali (2004) "Simple Estimators for Nonparametric Panel Data Models with Sample Attrition," *Journal of Econometrics,* Vol. 120 (1), pp. 159–180.

Deaton, Angus S. and Christina Paxson (2001) "Mortality, Education, Income, and Inequality among American Cohorts," in *Themes in the Economics of Aging,* Wise D.A. eds. University of Chicago Press, pp. 129–170.

D'Haultfoeuille, X. and P Février (2011) "Identification of Nonseparable Models with Endogeneity and Discrete Instruments," Working Paper.

Eckstein, Zvi and Kenneth I. Wolpin (1999) "Why Youths Drop out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica,* Vol. 67 (6), pp. 1295-1339.

Elbers, Chris and Geert Ridder (1982) "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies,* Vol. 49 (3), pp. 403–409.

Evans, William N. and Jeanne Ringel (1999) "Can Higher Cigarette Taxes Improve Birth Outcomes?" *Journal of Public Economics*, Vol. 72, No. 1, pp. 135–54.

Evdokimov, Kirill (2009) "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Yale University.

Fan, Jianqing and Irène Gijbels (1996) "Local Polynomial Modelling and Its Applications" Chapman & Hall.

Florens, Jean-Pierre (2003) "Inverse Problems and Structural Econometrics: The Example of Instrumental Variables," in Advances in Economics and Econometrics: Theory and Applications, Dewatripont, Vol. 2, ed. by L.-P. Hansen and S. J.

Turnovsky. Cambridge, U.K.: Cambridge University Press, pp. 284-311.

Florens, Jean-Pierre, James J. Heckman, Costas Meghir, and Edward J. Vytlacil (2008) "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, Vol. 76 (5), pp. 1191-1206.

Folland, Gerald, B. (1999) "Real Analysis: Modern Techniques and Their Applications," Wiley.

French, Eric (2005) "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies,* Vol. 72 (2), pp. 395-427.

French, Eric and John B. Jones (2011) "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica,* Vol. 79 (3), pp. 693-732.

Garen, John (1984) "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, Vol. 52 (5), pp. 1199–1218.

Graham, Bryan S. and James Powell (2008) "Identification and Estimation of 'Irregular' Correlated Random Coefficient Models," NBER Working Paper 14469.

Hahn, Jinyong (1999) "How Informative is the Initial Condition in the Dynamic Panel Model with Fixed Effects?" *Journal of Econometrics,* Vol. 93 (2), pp. 309–326.

Hall, Peter and Joel L. Horowitz (2005) "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, Vol. 33 (6), pp. 2904-2929.

Halliday, Timothy J. (2008) "Heterogeneity, State Dependence and Health," *Econometrics Journal,* Vol. 11 (3), pp. 499–516.

Hausman, Jerry A. and William E. Taylor (1981) "Panel Data and Unobservable Individual Effects," *Econometrica,* Vol. 49 (6), pp. 1377–1398.

Hausman, Jerry A. and David A. Wise (1979) "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica,* Vol. 47 (2), pp. 455–473.

Heckman, James J. (1981a) "Heterogeneity and State Dependence," in *Studies in Labor Markets,* Rosen S., eds. University of Chicago Press. pp. 91–139.

Heckman, James J. (1981b) "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data with Econometric Applications,* Manski C.F., McFadden D., eds. MIT Press. pp. 179–195.

Heckman, James J. (1991) "Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity," *American Economic Review,* Vol. 81 (2), Papers and Proceedings, pp. 75–79.

Heckman, James J. (2000) "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," *Quarterly Journal of Economics,* Vol. 115, pp. 45-97.

Heckman, James J. (2007) "The Economics, Technology, and Neuroscience of Human Capability Formation," *Proceedings of the National Academy of Sciences,* Vol. 104 (33), pp. 13250–13255.

Heckman, James J. and Salvador Navarro (2007) "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics,* Vol. 136 (2), pp. 341–396.

Heckman, James J. and Burton Singer (1984) "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica,* Vol. 52 (2), pp. 271–320.

Heckman, James J. and Edward Vytlacil (1998) "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling," *Journal of Human Resources,* Vol. 33 (4), pp. 974–987.

Heckman, James J. and Edward Vytlacil (1999) "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Science*, USA, Vol. 96, pp. 4730-4734

Heckman, James J. and Edward Vytlacil (2005) "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, Vol. 73, No. 3, pp. 669-738

Heckman, James J. and Edward J. Vytlacil (2007) "Econometric Evaluations of Social Programs, Part 1: Causal Models, Structural Models and Econometric Policy Evaluation," in J.J. Heckman and E.E. Leamer (ed.) *Handbook of Econometrics*, Vol. 6, Ch. 70.

Hellerstein, Judith K. and Guido W. Imbens (1999) "Imposing Moment Restrictions from Auxiliary Data by Weighting," *Review of Economics and Statistics,* Vol. 81 (1), pp. 1–14.

Henry, Marc, Yuichi Kitamura, and Bernard Salanié (2010) "Identifying Finite Mixtures in Econometric Models," Cowles Foundation Discussion Paper No. 1767.

Hirano, Keisuke, Guido W. Imbens, Geert Ridder, and Donald B. Rubin (2001) "Combining Panel Data Sets with Attrition and Refreshment Samples," *Econometrica,* Vol. 69 (6), pp. 1645–1659.

Hoderlein, Stefan and Enno Mammen (2007) "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," *Econometrica*, Vol. 75 (5), pp. 1513-1518

Hoderlein, Stefan and Enno Mammen (2009) "Identification and Estimation of Local Average Derivatives in Non-Separable Models without Monotonicity," *Econometrics Journal,* Vol. 12, No. 1, pp. 1-25.

Hoderlein, Stefan and Halbert White (2009) "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," CeMMAP Working Papers CWP33/09.

Honoré, Bo E. (1990) "Simple Estimation of a Duration Model with Unobserved Heterogeneity," *Econometrica,* Vol. 58 (2), pp. 453–473.

Honoré, Bo E. and Elie Tamer (2006) "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica,* Vol. 74 (3), pp. 611-629.

Hoogerheide, Lennart, Frank Kleibergen, and Herman K. van Dijk (2007) "Natural Conjugate Priors for the Instrumental Variables Regression Model Applied to the AngristKrueger Data," *Journal of Econometrics*, Vol. 138 (1), pp. 63–103.

Horowitz, Joel L. (1998) "Bootstrap Methods for Median Regression Models," *Econometrica*, Vol. 66, No. 6, pp. 1327–1351.

Horowitz, Joel L. (1999) "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity," *Econometrica,* Vol. 67 (5), pp. 1001-1028.

Horowitz, Joel L. (2011) "Applied Nonparametric Instrumental Variables Estimation," *Econometrica,* Vol. 79, No. 2, pp. 347–394.

Horowitz, Joel L. and Sokbae Lee (2007) "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, Vol. 75 (4), pp. 1191-1208.

Horowitz, Joel L. and Sokbae Lee (2009) "Testing a Parametric Quantile-Regression Model with an Endogenous Explanatory Variable Against a Nonparametric Alternative" *Journal of Econometrics*, Vol. 152 (2), pp. 141–152.

Hotz, Joseph V. and Robert A. Miller (1993) "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies,* Vol. 60 (3), pp. 497–529.

Hsiao, Cheng (1975) "Some Estimation Methods for a Random Coefficient Model," *Econometrica*, Vol. 43, No. 2, pp. 305–325.

Hsiao, Cheng (2003) "Analysis of Panel Data." Cambridge University Press.

Hsiao, Cheng and Hashem M. Pesaran (2004) "Random Coefficient Panel Data Models," IZA Discussion Papers 1236.

Hu, Yingyao (2008) "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics,* Vol. 144 (1), pp. 27–61.

Hu, Yingyao and Susanne M. Schennach (2008) "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica,* Vol. 76 (1), pp. 195–216.

Hu, Yingyao and Matthew Shum (2010) "Nonparametric Identification of Dynamic Models with Unobserved State Variables," Johns Hopkins University Working Paper #543.

Imbens, Guido W. and Joshua D. Angrist (1994) "Identification and Estimation of Local Average Treatment Effects " *Econometrica*, Vol. 62 (2), pp. 467–475.

Imbens, Guido W. (2007) "Nonadditive Models with Endogenous Regressors," in *Advances in Economics and Econometrics: Theory and Applications,* R. Blundell, W. Newey, T. Persson, eds. Cambridge University Press.

Imbens, Guido W. and Whitney K. Newey (2009) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, Vol. 77 (5), pp 1481–1512.

Jun, Sung Jae (2009) "Local Structural Quantile Effects in a Model with a Nonseparable Control Variable," *Journal of Econometrics*, Vol. 151 (1), pp 82–97.

Jun, Sung Jae, Joris Pinkse, and Haiqing Xu (2011) "Tighter Bounds in Triangular Systems," *Journal of Econometrics,* Forthcoming.

Karlstrom, Anders, Marten Palme, and Ingemar Svensson (2004) "A Dynamic Programming Approach to Model the Retirement Behaviour of Blue-Collar Workers in Sweden," *Journal of Applied Econometrics,* Vol. 19 (6), pp. 795-807.

Kasahara, Hiroyuki, and Katsumi Shimotsu (2009) "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices, *Econometrica,* Vol. 77 (1), pp. 135–175.

Kasy, M. (2011) "Identification in Triangular Systems Using Control Functions," *Econometric Theory,* Vol. 27, No. 3, pp. 1–9.

Khan, Shakeeb, Maria Ponomareva, and Elie Tamer (2011) "Inference on Panel Data Models with Endogenous Censoring," Duke University Working Paper 11-07.

Kyriazidou, Ekaterini (1997) "Estimation of a Panel Data Sample Selection Model," *Econometrica,* Vol. 65 (6), pp. 1335–1364.

Lancaster, Tony (1979) "Econometric Methods for the Duration of Unemployment," *Econometrica,* Vol. 47 (4), pp. 939–956.

Lee, Sokbae (2007) "Endogeneity in Quantile Regression Models: A Control Function Approach," *Journal of Econometrics*, Vol. 141 (2), pp. 1131–1158.

Lehman, E.L. (1974) "Nonparametrics: Statistical Methods Based on Ranks." Holden-Day, San Francisco.

Lewbel, A. (1999); "Consumer Demand Systems and Household Expenditure", in Pesaran, H. and M. Wickens (Eds.), Handbook of Applied Econometrics, Blackwell Handbooks in economics.

Lewbel, Arthur (2007) "Estimation of Average Treatment Effects with Misclassification," *Econometrica,* Vol. 75 (2), pp. 537-551.

Lien, Diana S. and William N. Evans (2005) "Estimating the Impact of Large Cigarette Tax Hikes: The Case of Maternal Smoking and Infant Birth Weight." *Journal of Human Resources*, Vol. 40, No. 2, pp. 373–392.

Lightwood, James M., Ciaran S. Phibbs, and Stanton A. Glantz (1999) "Short-term Health and Economic Benefits of Smoking Cessation: Low Birth Weight," *Pediatrics,* Vol. 104, No. 6, pp. 1312–1320.

Lleras-Muney, Adriana (2002) "Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939," *Journal of Law and Economics,* Vol. 45, No. 2, pp. 401–435.

Lleras-Muney, Adriana (2005) "The Relationship Between Education and Adult Mortality in the United States," *Review of Economic Studies,* Vol. 72 (1), pp. 189-221.

Luenberger, David, G. (1969) "Optimization by Vector Space Methods," Wiley-Interscience.

Ma, Lingjie and Roger Koenker (2006) "Quantile Regression Methods for Recursive Structural Equation Models," *Journal of Econometrics*, Vol. 134, pp. 471506.

Maccini, Sharon, and Dean Yang (2009) "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *American Economic Review,* Vol. 99 (3), pp. 1006–1026.

Magnac, Thierry and David Thesmar (2002) "Identifying Dynamic Discrete Decision Process," *Econometrica,* Vol. 70 (2), pp. 801–816.

Mahajan, Aprajit (2006) "Identification and Estimation of Regression Models with Misclassification," *Econometrica,* Vol. 74 (3), pp. 631-665.

Marschak, Jacob (1953) "Economic Measurements for Policy and Prediction," in: Hood, W., Koopmans, T. (Eds.), *Studies in Econometric Method*. Wiley, New York, pp. 1-26.

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

Matzkin, Rosa, L. (2007) "Nonparametric Identification," in J.J. Heckman and E.E. Leamer (ed.) *Handbook of Econometrics*, Vol. 6, Ch. 73.

Moffitt, Robert, John Fitzgerald, and Peter Gottschalk (1999) "Sample Attrition in Panel Data: The Role of Selection on Observables," *Annales d'Économie et de Statistique* No. 55/56, pp. 129–152

Newey, Whitney K. and James L. Powell (2003) "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, Vol. 71 (5), pp. 1565–1578.

Pagan, Adrian and Aman Ullah (1999) "Nonparametric Econometrics," Cambridge University Press.

Pakes, Ariel, Michael Ostrovsky, and Steven Berry (2007) "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)," *The RAND Journal of Economics,* Vol. 38 (2), pp. 373–399.

Pearl, Judea (2000) "Causality." Cambridge University Press.

Pesaran, Hashem M. and Ron Smith (1995) "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics,* Vol. 68 (1), pp. 79-113.

Pesendorfer, Martin, and Philipp Schmidt-Dengler (2008) "Asymptotic Least Squares Estimator for Dynamic Games," *Review of Economic Studies,* Vol. 75 (3), pp. 901–928.

Ridder, Geert (1990) "The Non-Parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies,* Vol. 57 (2), pp. 167–181.

Ridder, Geert (1992) "An Empirical Evaluation of Some Models for Non-Random Attrition in Panel Data," *Structural Change and Economic Dynamics,* Vol. 3 (2), pp. 337–355.

Ridder, Geert and Tiemen M. Woutersen (2003) "The Singularity of the Information Matrix of the Mixed Proportional Hazard Model," *Econometrica,* Vol. 71 (5), pp. 1579–1589.

Rosenzweig, Mark R. and T. Paul Schultz (1983) "Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight," *Journal of Political Economy*, Vol. 91, No. 5, pp. 723–746 .

Ruhm, Christopher J. (2000) "Are Recessions Good for Your Health?" *Quarterly Journal of Economics,* Vol. 115 (2), pp. 617–650.

Rust, John (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica,* Vol. 55 (5), pp. 999–1033.

Rust, John (1994) "Structural Estimation of Markov Decision Processes," in R. Engle and D. McFadden (ed.) *Handbook of Econometrics*, Vol. 4, Ch. 51.

Rust, John and Christopher Phelan (1997) "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets," *Econometrica,* Vol. 65 (4), pp. 781–831.

Schennach, Susanne M. (2007) "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica,* Vol. 75 (1), pp. 201-239.

Schennach Susanne M., Suyong Song, and Halbert White (2011) "Identification and Estimation of Nonseparable Models with Measurement Errors," unpublished manuscript.

Schultz, Paul T. (2002) "Wage Gains Associated with Height as a Form of Health Human Capital," *American Economic Review,* Vol. 92 (2), pp. 349–353.

Shen, Xiaotong (1997) "On Methods of Sieves and Penalization," *Annals of statistics,* Vol. 25 (6), pp. 2555–2591.

Shiu, Ji-Liang, and Yingyao Hu (2011) "Identification and Estimation of Nonlinear Dynamic Panel Data Models with Unobserved Covariates," Johns Hopkins University Working Paper #557.

Snyder, Stephen E., and William N. Evans (2006) "The Effect of Income on Mortality: Evidence from the Social Security Notch," *Review of Economics and Statistics,* Vol.

88 (3), pp. 482–495.

Sullivan, Daniel, and Till von Wachter (2009a) "Average Earnings and Long-Term Mortality: Evidence from Administrative Data," *American Economic Review,* Vol. 99 (2), pp. 133-138.

Sullivan, Daniel, and Till von Wachter (2009b) "Job Displacement and Mortality: An Analysis Using Administrative Data," *Quarterly Journal of Economics,* Vol. 124 (3), pp. 1265–1306.

Stock, James H. and David A. Wise (1990) "Pensions, the Option Value of Work, and Retirement," *Econometrica,* Vol. 58 (5), pp. 1151–1180.

Torgovitsky, Alexander (2011) "Identification and Estimation of Nonparametric Quantile Regressions with Endogeneity," Working Paper.

Wooldridge, Jeffrey, M. (1997) "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model," *Economics Letters,* Vol. 56 (2), pp. 129–133.

Wooldridge, Jeffrey M. (2001) "Econometric Analysis of Cross Section and Panel Data," The MIT Press.

Wooldridge, Jeffrey M. (2002) "Inverse Probability Weighted M-Estimators for Sample Selection, Attrition, and Stratification," *Portuguese Economic Journal,* Vol. 1 (2), pp. 117–139.

Wooldridge, Jeffrey M. (2005) "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics,* Vol. 20 (1), pp. 39–54.