Identification and $\sqrt{N}$-Consistent Estimation of a Nonlinear Panel Data Model with Correlated Unobserved Effects

Wayne-Roy Gayle†

University of Virginia

August 7, 2012

Abstract

This paper investigates identification and root-*n*–consistent estimation of a class of single-index panel data models in which the link function is unknown, the unobserved individual effects may be correlated with all the explanatory variables, and all the explanatory variables may be predetermined, including lagged dependent variables. We propose two sets of sufficient conditions, one in which link function is assumed to be strictly increasing, and the other in which it is not. We propose simple kernel-based estimators for the models, and derive consistency and asymptotic normality results for the proposed estimators. Finally, we present results of two Monte-Carlo studies of the estimators.

Keywords: Correlated Random Effects; Single Index; Semiparametric; Panel Data; Predetermined; Lagged Dependent Variables

JEL classification: C14, C23, I20, J24

†Economics Department, University of Virginia, Monroe Hall, Charlottesville, VA, 22903, phone: (434) 243-4336, fax: 434-982-2904, e-mail: `wg4b@virginia.edu`

1 Introduction

This paper provides new conditions for identification of a class of dynamic nonlinear single-index panel data models with unobserved individual-specific effects, and derives a new kernel-based estimator for the parameters of the model that fully implements the identification restrictions of the model. The model allows for all the explanatory variables to be predetermined, for the unobserved effect to be correlated with all the observed explanatory variables, and for the link function to be generally unspecified. All models considered in this paper are restrictions on the following model:

$$y_{it} = H_t\left(w_{it}B + c_i, \varepsilon_{it}\right), \quad t = 1, \cdots, T, \tag{1.1}$$

where w_{it} is a $(1 \times K+1)$-dimensional vector of explanatory variables, B is a $K+1$ dimensional vector of parameters, c_i is an unobserved individual-specific effect, and ε_{it} is an individual-time–specific unobservable. The unobserved error term ε_{it} is assumed to be independent of current and past values of w_{it}.[1] Three specific examples of models encompassed by equation (1.1) are

$$y_{it} = 1\{w_{it}B + c_i - \varepsilon_{it} \geq 0\}, \tag{1.2}$$

$$y_{it} = H(G_t(w_{it}B + c_i) + \varepsilon_{it}), \text{ and} \tag{1.3}$$

$$y_{it} = H(G_t(w_{it}B + c_i)\varepsilon_{it}), \tag{1.4}$$

where $1\{A\}$ is the indicator function equal to one if A is true, and zero otherwise. The unknown functions H, and G_t are not assumed to be strictly monotone on their domains. Under the restrictions this paper proposes, these models produce the same reduced form of interest.

The model presented in (1.2) is the widely investigated linear index panel data model, and merits some discussion. The approach taken here falls into the subclass of identification strategies that impose restrictions on the link function, as well as on the joint distribution of the observable

[1] This restriction can be relaxed to allow for contemporaneous correlation between ε_{it} and w_{it} using the methods of Newey and Powell [2003] and Blundell and Powell [2004].

and unobservable covariates, to obtain identification and $\sqrt{n}$-consistent estimation of the finite-dimensional parameters. Consequently, these models may be characterized by the tradeoff between the restrictions placed on the correlation between the regressors and the individual effects on one hand, and the restrictions placed on the link function on the other hand. At one end of this spectrum is the conditional logit model of Rasch [1960] and Anderson [1970], which makes no assumptions about the correlation between the regressors and the individual effects, but assumes a logit link function. Indeed, Chamberlain [2010] shows the logit error structure is the only one under which one can achieve $\sqrt{n}$-consistent estimators of β without restricting the correlation between the regressors and the individual effects. The other end of the spectrum is the case in which no assumption is placed on the link function, and the individual-specific effect is defined as a known function of observed regressors.

We provide two sets of sufficient conditions for identification. The first is in the spirit of Chamberlain [1980], Chamberlain [1984], Manski [1987], Horowitz [1992], Newey [1994a], Arellano and Carrasco [2003], Chen [1998], Honoré and Lewbel [2002], Ai and Gan [2010], and Gayle and Namoro [2012], in that we assume the link function is strictly increasing. This shape constraint, along with restrictions on the distribution of the individual effects, achieves root-n consistency of the finite-dimensional parameters. The second set of conditions for identification is in the spirit of Ichimura [1993], in that we do not restrict the link function to be strictly increasing.

Chamberlain [1984] and Newey [1994a] consider a model with only strictly exogenous variables and for which the distribution of the error term is normal with unknown mean and variance. Chamberlain [1984] assumes the individual effect is a parametric function of the regressors, and Newey [1994a] relaxes this assumption by imposing that the individual effect is an unknown function of the regressors. Chen [1998] extends the model of Newey [1994a] by relaxing the parametric assumption on the link function, assuming only an unknown, smooth, and strictly increasing function. Chen [1998], however, assumes the individual effects are conditionally independent of one continuous regressor (the “special regressor”) given the other regressors, and the coefficient on the "special regressor" is known.

Moving in a different direction, Arellano and Carrasco [2003] extend the model of Newey [1994a] by allowing all the regressors to be predetermined, as well as allowing all the regressors to be correlated with the individual effects. However, Arellano and Carrasco [2003] maintain the assumption of a known and strictly increasing link function.

The model presented in this paper extends these two branches by allowing for all the regressors to be predetermined and correlated with the individual effects, as well as by relaxing the strict monotonicity restriction on the link function. One attractive feature of this model is that it is not subject to the time-inconsistency problem the aforementioned models (excepting that of Arellano and Carrasco [2003]) face, in that the estimator is inconsistent with the arrival of a new wave of observations for each individual.

Other important advancements in identification and estimation of equation (1.2) include Manski [1987], Horowitz [1992], Honoré and Kyriazidou [2000], Honoré and Lewbel [2002], Honore and Tamer [2006], and Ai and Gan [2010]. Compared to Manski [1987], Horowitz [1992], and Honoré and Kyriazidou [2000], our estimator imposes stronger restrictions on the joint distribution of the observable regressors and individual effects, but allows for heteroskedasticity over time, and is $\sqrt{n}$-consistent. Also, unlike our estimator, those of Manski [1987] and Horowitz [1992] apply only to the static panel data framework. Honoré and Kyriazidou [2000], and Honore and Tamer [2006] allow for the inclusion of lagged dependent variables, but assume the other explanatory variables are strictly exogenous.

Honoré and Lewbel [2002] and Ai and Gan [2010] propose estimators of equation (1.2) that achieve $\sqrt{n}$ consistency of the finite-dimensional parameters. Compared to Manski [1987] and Horowitz [1992], these estimators impose stronger restrictions on the joint distribution of the observed regressors and the individual effects. In particular, Honoré and Lewbel [2002] assumes the existence of a "special regressor" that (i) is continuous with prior known sign of its coefficient, (ii) is independent of the individual effects, conditional on the other explanatory variables and instruments, and (iii) has bounded support that is large enough to contain the rest of the linear index, including the unobservable variables. Ai and Gan [2010] argue Honoré and Lewbel's third assumption is the most restrictive, and propose an estimator that relaxes this assumption. Ai and Gan [2010] impose stronger conditions than Honoré and Lewbel [2002] on the error term, resulting in a more restricted form of heteroskedasticity, but accommodates logit and probit models. Identification of equation (1.2) in our framework maintains the first key assumption of Honoré and Lewbel [2002]. Our assumption, which is analogous to Honoré and Lewbel's second one, is stronger in that we assume we can decompose the individual effect into a conditional expectations term and a term that is independent of the observed regressors. Our assumption is closer to those imposed by the studies in the previous two paragraphs. We also assume the individual effects are independent of our "special regressor" given the other observable regressors and a "dimension-reducing" function of the time vector of "special regressor." Altonji and Matzkin [2005] also impose this dimension-reducing restriction to identify their local average response. Our third assumption is closer to that of Ai and Gan [2010] than Honoré and Lewbel [2002] in that the support condition is not a function of the unknown error term and individual effects. However, unlike Ai and Gan's model, ours allows for time-specific heteroskedasticity. The identification conditions of this paper are therefore somewhat stronger than those made by Honoré and Lewbel [2002], and Ai and Gan [2010], in the static, binary-choice fixed effects panel data model. However, this paper is more concerned with identification and estimation of nonlinear, dynamic panel data models of the form given in equation (1.1).

Equations (1.3) and (1.4) are panel data, generalized linear models, with unknown, cascading, and potentially non-monotonic link functions. These equations encompass a large class of panel data models. The identification and estimation strategies proposed in this paper also apply to models of forms presented in equations (1.3) and (1.4). We discuss the application of these models to an extension of the classical Mincer wage equation at the end of section 2.

This paper presents kernel-based estimators for parameters of interest. The estimation strategy proceeds in two stages. The first stage concerns the problem of estimating the parameters of a generalized partial linear model (GPLM) with an unknown link function. This estimator is of interest in its own right because, to the best of our knowledge, no kernel-based estimator of a GPLM with an unknown and non-invertible link function exists.[2] Furthermore, kernel estimation of GPLM, as proposed by Linton [2000], requires the investigator to have initial consistent estimates of the nuisance parameters. These initial consistent estimates are generally obtained by the marginal integration technique (e.g., Linton and Hardle [1996]). The estimator proposed in this paper does not rely on the availability of initial consistent estimates.

Given estimation of the key infinite-dimensional parameters in the first stage, the second stage of the estimator reduces to a dynamic linear panel data model that employs the GMM-based method of Arellano and Bond [1991] and Arellano and Bover [1995] to identify and estimate the finite-dimensional parameters.

The rest of this paper is structured as follows. Section 2 outlines the identification and estimation strategy applied to the single, linear index, static panel data, discrete choice framework, and explains the extension to dynamic panel data models. Section 3 formalizes the sufficient conditions for identification of features of the model with and without the assumption of an invertible link function. Section 4 presents a kernel-based estimator of the parameters of interest that fully implements the identification restrictions presented in section 3. In section 5, consistency and uniform rates of convergence results are derived for the parameters, and asymptotic normality is proven for the estimators of the finite-dimensional parameters. Section 6 investigates the small sample properties of the estimator applied to equations (1.2) and (1.3). Section 7 concludes. The appendix contains the proofs of all lemmas and theorems.

[2] See Hardle et al. [2004] for discussions on estimating GPLMs.

2 Model

To delineate the identification restrictions of this paper, consider the simple, static panel data, binary choice model:

$$y_{it} = 1\{w_{it}B + c_i - \varepsilon_{it} \geq 0\}, \quad t = 1, \cdots, T. \tag{2.1}$$

Partition $w_{it} = (v_{it}, x_{it})$, where v_{it} is unidimensional, and assume the coefficient on v_{it} in equation (2.1) is one, so that

$$y_{it} = 1\{v_{it} + x_{it}\beta + c_i - \varepsilon_{it} \geq 0\}. \tag{2.2}$$

As discussed in the introduction, a large body of literature investigates identifying features of equation (2.2). These works can largely be distinguished by the tradeoff between restrictions on the correlation between c_i and w_{it} on one hand, and restrictions imposed on the distribution of ε_{it} on the other. See Arellano and Honoré [2001] for discussions of these various alternative restrictions. The identification strategy of this paper follows Chamberlain [1980], and subsequent authors, by imposing restrictions on the conditional distribution of c_i given $w_i := (w_{i1}, \cdots, w_{iT})$. Chamberlain [1980] assumes $c_i = \eta w_i + \varepsilon_i^c$, where $\varepsilon_{it} - \varepsilon_i^c$ is normally distributed and independent of w_i. Newey [1994a] assumes $c_i = \eta(w_i) + \varepsilon_i^c$, where η is an unknown function, but assumes the distribution of $\varepsilon_{it} - \varepsilon_i^c$ is known up to scale. Chen [1998] and Gayle and Namoro [2012] extend the model of Newey [1994a] by allowing the distribution of $\varepsilon_{it} - \varepsilon_i^c$ to be unknown.

Assume $c_i = \eta(w_i) + \varepsilon_i^c$, and $\varepsilon_{it} - \varepsilon_i^c$ is independent of w_i with mean zero and cdf, $\Phi(\cdot)$, that is absolutely continuous with respect to a Lebesgue measure. Then equation (2.2) obtains

$$E[y_{it}|v_i, x_i] = \Phi(v_{it} + x_{it}\beta + \eta(v_i, x_i)). \tag{2.3}$$

The papers cited in the previous paragraph rely critically on two key properties: the link function

Φ is invertible, and the invariance of $\eta(v_i, x_i)$ over time generates intertemporal (within) variation of w_{it} given $\eta(w_i)$.[3] The model in equation (2.1) may be specified so the resulting link function, Φ, in equation (2.3) is not strictly monotonic (e.g., equation (1.3)). The results of this paper provide a way to identify the parameters of equation (2.3) without assuming invertibility of the link function. However, to illustrate the basic approach, we will continue to motivate our model assuming that the link function is invertible.

The first restriction is $\eta(v_i, x_i) = E[c_i|v_i, x_i]$, and depends on v_i through a function $R(v_i)$, where the range of R is less than T and depends on the restriction imposed so that the density of $f_v(v_{it}|R(v_i), x_i) > 0$ for all $v_{it} \in \Re$, where $f_v(\cdot|R(v_i), x_i)$ is the density of v_{it} conditioned on w_i. The purpose of this restriction is to generate intra-temporal variation of v_{it} given $\eta(w_i)$. As in Chamberlain [1980], this restriction is most suitably formulated in Bayesian terms, where $\eta(v_i, x_i)$ is interpreted as the posterior mean of c_i given (v_i, x_i). For example, suppose the following equation generates v_{it}:

$$v_{it} = \rho_t(x_{it}) + c_i + \varepsilon_{it}, \quad t = 1, \cdots, T, \tag{2.4}$$

where ε_{it}, conditional on c_i and x_{it}, is i.i.d. $N(0, \sigma^2)$. Suppose c_i is i.i.d $N(0, \sigma_c^2)$. Then the distribution of c_i conditional on (v_i, x_i) is given by

$$c_i|(v_i, x_i) \sim N(\eta(v_i, x_i), \sigma_T^2),$$

where

$$\eta(v_i, x_i) = E[c_i|v_i, x_i] = \frac{\sigma_c^2}{\sigma^2 + T\sigma_c^2} \sum_{s=1}^{T} (v_{is} + \rho_s(x_{is})) = E[c|R(v_i), x_i] = \eta(R(v_i), x_i),$$

and $R(v_i) = \sum_{s=1}^{T} v_{is}$. Also,

$$\sigma_T^2 = \frac{\sigma^2 \sigma_c^2}{\sigma^2 + T\sigma_c^2}.$$

[3] Suppose the alternative decomposition is imposed as $c_i = \eta(w_{it}) + \varepsilon_{it}^c$, where $\varepsilon_{it}^c \perp w_{it}$. Then even if Φ is assumed to be known and strictly increasing on its support, the inversion method of Newey [1994a] does not identify β, because differencing over time does not eliminate $\eta(w_{it})$.

Thus we have that $c_i = \eta(R(v_i), x_i) + \varepsilon_i^c$, where $\varepsilon_i^c | v_i, x_i \sim N(0, \sigma_T^2)$.

The above example satisfies all the restrictions on the distribution of c_i given (v_i, x_i) set out so far, with $R(v_i) = \sum_{s=1}^T v_{is}$. Note we do not assume the investigator knows ρ_t, σ^2, and σ_c^2. Note further that this example allows for general forms of correlation between c_i and x_i. As in Chamberlain [1980], the model for v_{it} can be extended to a stationary autoregressive framework. More generally, one may assume $c_i = \rho(R(v_i), x_i) + \varepsilon_i^c$, where ε_i^c is independent of (v_i, x_i) with an unknown distribution function, and the conditional density $f_v(v_{it}|R(v_i), x_i) > 0$, $v_{it} \in \Re$.

Define $z_i = (R(v_i), x_i)$ and $\mu_t(z_i) = x_{it}\beta + \eta(z_i)$. Then equation (2.3) becomes

$$E[y_{it}|v_i, x_i] = E[y_{it}|v_{it}, z_i] = \Phi(v_{it} + \mu_t(z_i)). \tag{2.5}$$

To understand the second restriction, note that allowing Φ and μ_t to be unknown necessitates location normalization. To see why, observe that for any constant b, the alternative model $(\tilde{\Phi}, \tilde{\mu}_t)$, where $\tilde{\Phi}(a) = \Phi(a+b)$ and $\tilde{\mu}_t(z) = \mu_t(z) - b$, generates the same conditional expectation on the LHS of equation (2.5) as (Φ, μ_t). In other words, these two alternative models are observationally equivalent with respect to the conditional expectation. Therefore, without loss of generality, impose the normalization condition that for a given value of z_i, $\check{z}$, and a known constant, b, $\mu_t(\check{z}) = b$.

The strategy for identification is to exploit the intra-temporal variation of v_{it} given z_i and the index restriction of equation (2.5) to recover Φ and $\mu_t, t = 1, \cdots, T$, then to exploit the within variation of x_{it} given $\eta(z_i)$ to recover β. First note

$$E[y_{it}|\tilde{v} = v_{it}, \tilde{z} = \check{z}] = \Phi(v_{it} + \mu_t(\check{z})) = \Phi(v_{it} + b)$$

traces out Φ by varying v_{it} on its support. In the example given above, this device can be implemented for T as small as 2, where v_{it} can be varied holding $v_{i1} + v_{i2}$ fixed. To recover $\mu_t(z)$, we make use of the normalization and the index restriction of the model. Specifically, note that for any

$\overline{v_t}$ and fixed $\overline{z}$, we have the following equalities:

$$\begin{aligned}
E[y_{it}|\tilde{v} = \overline{v_t}, \tilde{z} = \overline{z}] &= \Phi(\overline{v_t} + \mu_t(\overline{z})) \\
&= \Phi([\overline{v_t} + \mu_t(\overline{z}) - b] + b) \\
&= E[y_{it}|\tilde{v} = \overline{v_t} + \mu_t(\overline{z_t}) - b, \tilde{z} = \check{z}] \\
&= E[y_{it}|\tilde{v} = \check{v}_t, \tilde{z} = \check{z}], \qquad (2.6)
\end{aligned}$$

where $\check{v}_t = \check{v}(\overline{v_t}, \overline{z}) = \overline{v_t} + \mu_t(\overline{z}) - b$. Now, for any $\overline{z}$, $E[y_{it}|\tilde{v} = v_{it} + \mu_t(\overline{z}) - b, \tilde{z} = \check{z}]$ is identified because it is a function of the observable random variables. Equation (2.6) holds for all v so that $\mu_t(z)$ can be solved as the minimizer of

$$\begin{aligned}
Q(z, \mu_t) &= E\left[\left(E[y_{it}|\tilde{v}, \tilde{z} = z] - E[y_{it}|\tilde{v} + \mu_t - b, \tilde{z} = \check{z}]\right)^2\right], \\
&= \int \left(E[y_{it}|\tilde{v}, \tilde{z} = z] - E[y_{it}|\tilde{v} + \mu_t - b, \tilde{z} = \check{z}]\right)^2 f_v(\tilde{v}) d\tilde{v}, \qquad (2.7)
\end{aligned}$$

where f_v is the marginal density of v_{it}. For any z, the minimizer of equation (2.7) is unique if the link function Φ is strictly monotone. If Φ is not strictly monotone then, similar to Ichimura [1993], a sufficient condition for uniqueness of μ_t in equation (2.7) is that Φ is not periodic on its domain.

The method we propose to identify Φ and μ_t requires the distribution of v_{it} conditional on z_i to be absolutely continuous with large support. In this sense, v_{it} is similar to the special regressor proposed in Lewbel [2000], Honoré and Lewbel [2002], and Ai and Gan [2010]. The restrictions this paper impose on the conditional distribution of c_i given (v_i, z_i) are somewhat stronger than those the previous referenced papers propose. However, their results rely on the binary choice framework. The restrictions this paper propose are sufficient to identify the parameters in a larger class of continuous and limited-dependent variable models, allow for the link function not to be strictly monotone, and allow for all the regressors to be predetermined.

Given μ_t, identification of β comes from recognizing

$$\mu_t(z_i) = x_{it}\beta + \eta(z_i), \quad t = 1, \cdots, T,$$

is a static linear panel data model, in which $\mu_t(z_i)$ is the dependent variable. Because $\mu_t(z_i)$ is identified under the restrictions set out, we can treat it as known. The within or first-difference approaches are both valid. Using the first-difference approach, we have the closed-form expression

$$\beta = (E[\Delta x_i' \Delta x_i])^{-1} E[\Delta x_i' \Delta \mu(z_i)],$$

where $\Delta x_i = (\Delta x_{i2}', \cdots, \Delta x_{iT}')'$ and $\Delta\mu(z_i) = (\Delta\mu_2(z_i), \cdots, \Delta\mu_T(z_i))'$.

This method of identification extends to the case in which some or all of the explanatory variables are pre-determined. It is well understood that if x_{it} contains lagged values of y_{it}, or if (v_{it}, x_{it}) depends on lagged values of y_{it}, then $E[c_i|v_i, x_i]$ is necessarily correlated with the error term in each period. As in Arellano and Carrasco [2003], we impose the restriction $c_i = E[c_i|v_i^t, x_i^t] + \varepsilon_{it}^c$ with $\varepsilon_{it}^c \perp (v_i^t, x_i^t)$, where $v_i^t = (v_{i1}, \cdots, v_{it})$ and $x_i^t = (x_{i1}, \cdots, x_{it})$. Arellano and Carrasco [2003] show this decomposition, along with the law of iterated expectations, $E[E[c_i|v_i^t, x_i^t]|v_i^{t-1}, x_i^{t-1}] = E[c_i|v_i^{t-1}, x_i^{t-1}]$, is sufficient to identify the finite-dimensional parameters if the link functions $\Phi_t, t = 1, \cdots, T$ are strictly monotonic and known. However, if the link functions are unknown and not assumed to be strictly monotonic, additional restrictions are necessary for identification. As in the static panel data model, we impose the restriction that $E[c_i|v_i^t, x_i^t] = E[c_i|R_t(v_i^t), x_i^t]$, where the R_t is chosen so that v_{it} can be varied continuously holding $\eta_t(R_t(v_i^t), x_i^t)$ fixed. In the example given above,

$$c_i|(v_i^t, x_i^t) \sim N(\eta(v_i^t, x_i^t), \sigma_t^2),$$

where

$$\eta_t(v_i^t, x_i^t) = E[c_i|v_i^t, x_i^t] = \frac{\sigma_c^2}{\sigma^2 + t\sigma_c^2} \sum_{s=1}^{t} (v_s + \rho_s(x_s)) = E[c_i|R_t(v_i^t), x_i^t] = \eta_t(R_t(v_i^t), x_i^t),$$

and $R_t(v_i^t) = \sum_{s=1}^t v_s$. Also,

$$\sigma_t^2 = \frac{\sigma^2 \sigma_c^2}{\sigma^2 + t\sigma_c^2}.$$

Thus, we have $c_i = \eta_t(R_t(v_i^t), x_i^t) + \varepsilon_{it}^c$, where $\varepsilon_{it}^c | v_i^t, x_i^t \sim N(0, \sigma_t^2)$. Furthermore, by the law of iterated expectations, $E[E[c_i|R_t(v_i^t), x_i^t]|x_i^{t-1}] = E[E[c_i|R_{t-1}(v_i^{t-1}), x_i^{t-1}]|x_i^{t-1}]$.

More generally, one may assume $c_i = \rho_t(R_t(v_i^t), x_i^t) + \varepsilon_{it}^c$, where ε_{it}^c is independent of (v_i^t, x_i^t) with an unknown time-specific distribution function. This more general assumption is closely related to Arellano and Carrasco [2003], who assume $(c_i + \varepsilon_{it})|v_i^t, x_i^t \sim N(E[c_i|v_i^t, x_i^t], \sigma_t^2)$. Compared to Arellano and Carrasco [2003], our model relaxes the normality assumption and strengthens the conditional mean assumption.

These simple examples satisfy all the restrictions to identify the model in the case in which some or all of the explanatory variables are predetermined. Define $z_{it} = (R(v_i^t), x_i^t)$ and $\mu_t(z_{it}) = x_{it}\beta + \eta_t(z_{it})$. Then equation (2.3) becomes

$$E[y_{it}|v_i^t, x_i^t] = E[y_{it}|v_{it}, z_{it}] = \Phi(v_{it} + \mu_t(z_{it})). \tag{2.8}$$

The normalization $\mu_t(\check{z}_{it}) = b_t$ identifies Φ_t and μ_t as in the static framework. Unlike in the static panel data case, we allow b_t to vary over t because doing so may be more appropriate given the dimension of z_{it} is necessarily increasing in t without additional restrictions. Imposing the law of iterated expectations identifies β, because

$$\begin{aligned} E[x_i^{t-1}\Delta\eta_t(z_{it})] &= E[x_i^{t-1}(E[c_i|z_{it}] - E[c_i|z_{it-1}])] \\ &= E[x_i^{t-1}(E[c_i|R_t(v_i^t), x_i^t] - E[c_i|R_t(v_i^{t-1}), x_i^{t-1}])] \\ &= E[x_i^{t-1}E[E[c_i|R_t(v_i^t), x_i^t] - E[c_i|R_{t-1}(v_i^{t-1}), x_i^{t-1}]|x^{t-1}]] = 0. \end{aligned} \tag{2.9}$$

Therefore, identification of β reduces to identification in the following linear, dynamic panel data framework:

$$\mu_t(z_{it}) = x_{it}\beta + \eta_t(z_{it}), \quad t = 2, \cdots, T,$$

where $\mu_t(z_{it})$ is the dependent variable, $\eta_t(z_{it})$ is the error term, and x_i^{t-1} serve as the period t instruments in the first-difference transformed model. Specifically, for $T \geq 3$, let X_i be the $(T-1) \times KT(T-1)$ matrix of the form $X_i =$diag$[(x_{i1}, \cdots, x_{is}),\ s = 1, \cdots, T-1]$, $\Delta x_i = (\Delta x'_{i3}, \cdots, \Delta x'_{iT})'$, and $\Delta\mu(z_i) = (\Delta\mu_3(z_{i3}), \cdots, \Delta\mu_T(z_{iT}))'$. Let Ω be any $KT(T-1) \times KT(T-1)$ symmetric, positive-definite weighting matrix, including the identity matrix. Then the results of Arellano and Bond [1991] obtain the following closed-form expression for β:

$$\beta = (C'\Omega C)^{-1} C'\Omega E[X_i' \Delta\mu(z_i)],$$

where $C = E[X_i' \Delta x_i]$.

In sum, the strategy outlined in this section to recover the finite-dimensional parameters is to first recover the time-specific “reduced-form” link functions. Using these link functions, the second step recovers the individual-time–specific index functions. Given the index functions, the third step is to recover the finite-dimensional parameters by applying GMM for linear panel data models to the index functions.

Notice the proposed method does not require constructing the joint distribution of sequences of the dependent variable, given the explanatory variables and individual-specific effects. In the non-linear dynamic panel data context, such a strategy typically requires specification of the initial conditions, that is, the distribution of y_{i1} given (x_{i1}, c_i), $f(y_{i1}|x_{i1}, c_i)$. Our identification strategy does not require specification of $f(y_{i1}|x_{i1}, c_i)$. Two important exceptions, where identification involves evaluating the joint distribution of sequences of the dependent variable without specifying $f(y_{i1}|x_{i1}, c_i)$, are Honoré and Kyriazidou [2000], and Honore and Tamer [2006]. Honoré and Kyriazidou [2000] propose methods for identifying and estimating fixed-effect discrete choice models with strictly exogenous covariates, as well as lags of the dependent variable. Their methods make no assumptions on the distribution of $f(y_{i1}|x_i,c_i)$. However, their methods rule out time-specific effects and require at least four time periods for identification. Honore and Tamer [2006] relax these restrictions of Honoré and Kyriazidou [2000] and derive tight bounds for the set of finite-dimensional parameters. Their results suggest the identified set for the finite-dimensional parameters is small in many cases. In contrast to Honore and Tamer [2006], this paper imposes restrictions on the conditional distribution of c_i, but leaves the distribution of the individual-time–specific unobservables unspecified.

The estimator proposed in this paper also applies to models with continuous outcome variables of the form given in equations (1.3), and (1.4), where the functions H and G_t are unknown and possibly nonlinear. An example of a potential application of equation (1.3) is an extension of the classical Mincer wage regression model. In this example, G_t are time-specific link functions, which capture aggregate changes in the effect of the individual-time–specific index, $w_{it}B+c_i$, on wages. It is natural to assume the derivatives of G_t are non-negative over their support. H is an unknown function that captures potential nonlinear effects of the index $G_t(\cdot)+\varepsilon_{it}$ on wages. It is natural to assume the derivative of H is also non-negative over its support. These monotonicity restrictions on G_t and H are not necessary to identify the finite-dimensional parameters, B. Notice that if H is the exponential function and G_t are identity functions, the model reduces to the classical log-linear wage regression. This extension of the classical log-linear wage regression is useful, at least as a robustness check, because economic theory generally does not imply the parametric forms of H and G_t.

3 Identification

This section formalizes the results obtained in the previous section. Recall the general model of interest is given by

$$y_{it}=H_t\left(w_{it}B+c_i,\varepsilon_{it}\right),\quad t=1,\cdots,T, \tag{3.1}$$

where w_{it} is a $K+1$-dimensional row vector of explanatory variables, which may all be predetermined, c_i is an unobserved individual specific-effect, and ε_{it} is the individual-time–specific unobserved effect.

Partition $w_{it} = (v_{it}, x_{it})$, where v_{it} is unidimensional and x_{it} is $(1 \times K)$-dimensional to have $y_{it} = H_t(\alpha v_{it} + x_{it}\beta + c_i, \varepsilon_{it})$ with $\alpha \in \mathcal{A} \subset \Re$, and $\beta \in \mathcal{B} \subset \Re^K$. The random vector x_{it} may contain discrete and continuous regressors. Recall $w_i^t = (w_{i1}, \cdots, w_{it})$.

Assumption 3.1. *$c_i = E[c_i|w_i^t] + \varepsilon_{it}^c$, and $(\varepsilon_{it}, \varepsilon_{it}^c) \perp w_i^t$.*

Assumption 3.1 is essentially the assumption Arellano and Carrasco [2003] make, extended to allow for an unknown distribution for ε_{it}^c. Although Assumption 3.1 is strong, we know from Chamberlain [2010] that strong restrictions on the joint distribution of (c_i, w_i^t) are necessary to derive $\sqrt{n}$-consistent estimators of the finite-dimensional parameters without knowledge of the distribution of ε_{it}. Assumption 3.1 is one such restriction, which allows for estimation for a large class of dynamic nonlinear panel data models. Note this assumption allows for correlation between ε_{it} and ε_{is}^c, as well as between ε_{it}^c and ε_{is}^c for all (t,s). It also allows for correlation between ε_{it}^c and w_{is} for $s > t$, which necessarily occurs if w_{it} contains lagged values of the dependent variable. The previous section gave two examples in which Assumption 3.1 holds. In other contexts, this assumption can be viewed as a reduced-form restriction much in the spirit of the control-function approach of Blundell and Powell [2004].

Under Assumption 3.1, equation (3.1) obtains

$$
\begin{aligned}
E[y_{it}|w_i^t] &= \int H_t\left(w_{it}B + E[c_i|w_i^t] + \varepsilon_{it}^c, \varepsilon_{it}\right) f_{\varepsilon_t^c,\varepsilon_t}(\varepsilon_{it}^c, \varepsilon_{it}) d\varepsilon_{it}^c d\varepsilon_{it}, \\
&= \Phi_t\left(\alpha v_{it} + x_{it}\beta + E[c_i|w_i^t]\right) \\
&= \Phi_t\left(\alpha v_{it} + x_{it}\beta + \eta_t(w_i^t)\right), \qquad (3.2)
\end{aligned}
$$

where $f_{\varepsilon_t^c,\varepsilon_t}$ is the joint density of $(\varepsilon_t^c, \varepsilon_t)$, and $\Phi_t(a) = \int H_t(a + \varepsilon_{it}^c, \varepsilon_{it}^c) f_{\varepsilon_t^c,\varepsilon_t}(\varepsilon_{it}^c, \varepsilon_{it}) d\varepsilon_{it}^c d\varepsilon_{it}$.

We provide two sets of sufficient conditions for identification of the model defined in equation (3.2). The distinction between these two sets of conditions is that in the first case, the reduced-form link functions Φ_t are assumed to be strictly increasing, in which case, the sign of the coefficient on v_{it} is identified. In the second case, the link functions are not assumed to be strictly increasing. For both approaches to identification, we make the following common assumptions.

Assumption 3.2.

1. *Let* $z_{it} = (R_t(v_i^t), x_i^t)$, *where* $R_t : \Re^t \mapsto \Re^{L_t}$, $L_t < t$ *is known. For* $t = 2, \cdots, T$, $E[c_i|v_i^t, x_i^t] = E[c_i|z_{it}]$, *and the conditional distribution of* v_{it} *given* z_{it} *is absolutely continuous with respect to a Lebegue measure with non-degenerate Radon-Nikodym conditional density* $f_v(v_{it}|z_{it})$, *with support equal to* $\Re$, *and* L_t *depends on the restriction adopted.*

2. *Let* $\mu_t(z_{it}) = x_{it}\beta + \eta_t(z_{it})$. *For* $t = 2, \cdots, T$, *for a given value of* $\check{z}_t$, *and a given constant,* b_t, $\mu_t(\check{z}_t) = b_t$.

3. *For* $t = 2, \cdots, T$, $E[(x_i^{t-1})'\Delta x_{it}]$ *has rank* K.

The purpose of Assumption 3.2.1 is to be able to vary v_{it} holding the conditional mean of c_i given w_i^t constant. This condition is key in identifying the link functions, as well as the functions $\mu_t(z_{it})$. Except for the restriction that $R_t(v_i^t) = v_i^{t-1}$, Assumption 3.2.1 generally requires $T \geq 3$. Given Assumption 3.1, the first part of Assumption 3.2.1 implies the following restriction on the conditional density of c_i: $f_c(c_i|v_i^t, x_i^t) = f_c(c_i|R_t(v_i^t), x_i^t)$, where R_t is such that $f_v(v_{it}|R_t(v_i^t), x_i^t)$ has support $\Re$. This condition is similar to the restrictions imposed in Altonji and Matzkin [2005]. The restriction on $f_v(\cdot|w_i^t)$ is similar to the conditional independence assumption made in Honoré and Lewbel [2002] and Ai and Gan [2010], where $R_t(v_i^t)$ serves as instruments for v_{it}. Indeed, one may include instruments in z_{it} to help pick up any dependence between v_{it} and c_i. We opt for existence of the dimension-reducing restriction in Assumption3.2.1, instead of assuming the existence of excluded instruments, because the traditional motivation of panel data models is that they may be able to account for permanent unobserved heterogeneity without the need for external instruments. Note this dimension-reducing condition is not too hard to satisfy, because the investigator is typically able to choose from a variety of candidate functions, and Altonji and Matzkin [2005] show the selection can be based on a test that compares $E[y_{it}|R_t(v_i^t), x_i^t]$ with $E[y_{it}|v_i^t, x_i^t]$ over the candidates for R_t. The restrictions Assumption 3.2.1 imposes are different from those Altonji and Matzkin [2005] impose; they are neither weaker nor stronger. However, the restrictions this paper imposes allow for the accommodation of relatively general dynamics in the panel data model and do not require monotonicity restrictions. Some examples of R_t for which Assumption 3.2.1 holds include $R_t(v_i^t) = \sum_{s=1}^t v_{is}$, $R_t(v_i^t) = (v_{it} + v_{i(t-1)}, v_i^{t-2})$, $R(v_i^t) = (v_{it} - v_{i(t-1)}, v_i^{t-2})$, and $R(v_i^t) = (|v_{it} - v_{i(t-1)}|, v_i^{t-2})$. See section 2.2 of Altonji and Matzkin [2005] for a detailed discussion of these and other choices of R_t. In the first example, $L_t = 1$, and in the other three, $L_t = t-1$. Note that in all four examples, $L_t < t$. This is condition is necessary for Assumption 3.2.1 to hold. As the previous section discusses, any restriction R_t such that v_{it} can be varied continuously holding $R_t(v_i^t)$ constant suffices. In the second example, this restriction requires that for any v_t and constant, a, in the support of $v_t + v_{t-1}$, the subpopulation with v_{t-1} so that $v_t + v_{t-1} = a$ has positive measure. Again, this restriction is testable.

Similar to Honoré and Lewbel [2002] and Ai and Gan [2010], Assumption 3.2.1 imposes a large support condition on the conditional distribution of v_{it} given z_{it}. The estimators of Honoré and Lewbel [2002] and Ai and Gan [2010] do not require Assumption 3.1. Therefore, their identification conditions are weaker than those presented in this paper. However, the estimators Honoré and Lewbel [2002] and Ai and Gan [2010] propose are valid only for linear index discrete choice models. The method this paper presents applies to a larger class of limited dependent variables models, as well as to continuous outcome models.

As discussed in the previous section, Φ_t and μ_t are only identified up to a location constant. Assumption 3.2.2 fixes the location of both. In a previous version of this paper, the choice of normalization was that $\mu_t(\mathbf{0}) = 0$, so that $\check{z}_t = \mathbf{0}$ and $b_t = 0$. We also suggested that if $\mathbf{0} \notin \mathcal{Z}_t$, redefine the coordinates by a translation so that in the new coordinates, $\mathbf{0} \in \mathcal{Z}_t$, where $\mathcal{Z}_t$ is the support of z_{it}. Although this choice of restriction is without loss of generality, it gives the impression that the "zero condition" is essential for identification. Therefore, we drop this condition in the current version of the paper to highlight this condition is a normalization one; any value of z_{it} and any constant in the range of μ_t suffice. In practice, one would choose $\check{z}_t$ to be in the interior of the support of z_{it} and, b_t to be in the interior of the support of v_{it}.

Given appropriate normalization on α (depending on whether Φ_t is assumed to be strictly increasing), Assumptions 3.1, 3.2.1, and 3.2.2 are sufficient for identifying Φ_t and μ_t. Given identification of μ_t, Assumption 3.2.3, along with the law of iterated expectations, is sufficient for identifying β from the linear, dynamic panel data model

$$\mu_t(z_{it}) = x_{it}\beta + \eta_t(z_{it}), \quad t = 2, \cdots, T,$$

where $\mu_t(z_{it})$ is the dependent variable, and $\eta_t(z_{it}) = E[c_i|z_{it}]$ is the error term with the condition that

$$\begin{aligned}
E[x_i^{t-1}(E[c_i|z_{it}] - E[c_i|z_{it-1}])] &= E[x_i^{t-1}(E[c_i|R_t(v_i^t), x_i^t] - E[c_i|R_t(v_i^{t-1}), x_i^{t-1}])] \\
&= E[x_i^{t-1}E[E[c_i|R_t(v_i^t), x_i^t] - E[c_i|R_{t-1}(v_i^{t-1}), x_i^{t-1}]|x^{t-1}]] \\
&= 0.
\end{aligned}$$

The parameter vector of interest for identification is defined as $\theta := (\Phi, \eta, B)$, where $\Phi := (\Phi_2, \cdots, \Phi_T)$, and $\eta := (\eta_2, \cdots, \eta_T)$, and $B = (\alpha, \beta')$, where α is the coefficient on v_{it} and β is the coefficient on x_{it}. We provide two alternative sets of sufficient conditions for identification of θ. The first is similar to Newey [1994a], Arellano and Carrasco [2003], and Gayle and Namoro [2012] in that it requires the link function to be strictly increasing.

Assumption 3.3.

1. *For $t = 1, \cdots, T$ the link function Φ_t is strictly increasing on its domain.*

2. $|\alpha| = 1$.

Assumption 3.3.1 allows us to invert the link functions and extract the finite-dimensional parameters using the law of iterated expectations. Assumption 3.3.2 is a normalization used to obtain point identification of the parameters. It is only one of a variety of scale normalizations of the finite-dimensional parameters that can be implemented. For example, Manski [1987] and Gayle and Namoro [2012] impose the condition $\|B\| = 1$. We choose Assumption 3.3.2 because it is simpler to impose in the estimation strategy to come. Denote the true model by $\theta_0 := (\Phi_0, \eta_0, B_0)$.

Theorem 3.4. *Consider model (3.2) and let Assumptions 3.1, 3.2, and 3.3 hold. Then θ_0 is uniquely identified.*

Proof. See Appendix A.1 □

The intuition behind Theorem 3.4 is that under the restrictions imposed by R_t, v_{it} can be varied holding z_{it} fixed, particularly at $z_{it} = \check{z}_{it}$. This independent variation assumption identifies the sign of α and traces out the link function, which can then be inverted to recover the finite dimensional parameters via a first-difference mechanism and the law of iterated expectations.

Along with Assumptions 3.1 and 3.2, Assumption 3.3 suffices in many applications. However, the condition that the link function is strictly increasing may be too restrictive for some applications. The next assumption provides an alternative set of sufficient conditions that relaxes this restriction. Essentially, these assumptions extend the results of Ichimura [1993] to the panel data setting.

Assumption 3.5.

1. *For $t = 1, \cdots, T$, Φ_t is not periodic on its domain.*

2. $\alpha = 1$.

As in Ichimura [1993], Assumption 3.5.1 places restrictions on the potential oscillation of the link functions Φ_t. In models (1.3) and (1.4), a sufficient condition for Assumption 3.5.1 to hold is that $G_t, t = 1, \cdots, T$ are not periodic on their domains. Assumption 3.5.2 is without loss of generality, because we no longer assume the link functions are strictly increasing.

Theorem 3.6. *Consider model (3.2) and let Assumptions 3.1, 3.2, and 3.5 hold. Then θ_0 is uniquely identified.*

Proof. See Appendix A.2 $\square$

The intuition behind the identification result in Theorem 3.6 begins, as above, by tracing out the index function by varying v_{it}, holding $z_{it} = \check{z}_t$. However, because the link function is not strictly monotone, the inversion device is no longer available. Instead, the non-periodicity of the link function is employed to recover $\mu_{0t}(z_t) = x_t\beta_0 + \eta_{0t}(z_t)$ for any fixed z_t. This operation is done for all z_{it}. The rest of the proof then becomes identical to Theorem 3.4.

Remember the identification strategy presented in this section applies to any model of the form in equation (3.1) that satisfies the restrictions of Assumptions 3.1 and 3.2, along with 3.3 or 3.5.

Finally, we can allow for the individual effect to be time varying, denoted by c_{it}. Time-varying individual effects may be appropriate in a variety of applications. Indeed, in the managerial incentive problem of Holmstrom [1999], where the unobserved individual effect is the manager's ability, ability must be time-varying to prevent it from becoming fully known by both parties over time. As another example, consider a wage-determination model, in which $c_{it} = \lambda_t' c_i$, c_i is a vector of unobserved individual-specific productivity characteristics, and λ_t is the corresponding vector of time-specific market prices of the productivity characteristics. Bai [2009] investigates this interactive-effects specification in a dynamic linear panel data model. In the case in which the individual effect is time varying, the identification strategy of this paper goes through with the addition of one condition:

$$E[E[c_{it}|z_{it}]|x_i^t] = E[E[c_{it-1}|z_{it-1}]|x_i^{t-1}].$$

This condition is satisfied if the sequence $\{\eta_t(z_{it}) = E[c_{it}|z_{it}]\}_{t=2}^T$ is a martingale with respect to $\{x_i^t\}_{t=1}^T$. For the motivating example presented in the previous section, this condition is satisfied if

$c_{it} = c_i + r_{it}$, with $r_{it} \perp w_i^t$ and r_{it} is i.i.d. $N(0, \sigma_r^2)$.

4 The Estimator

The previous section provided two alternative sets of sufficient conditions for identification of the parameters of the model given in equation (3.2). This section presents estimators of these parameters. In what follows, we present the estimator for the model under Assumptions 3.1, 3.2, and 3.5, hereafter called "model 2." In section 4.5, we discuss the estimator for the model under Assumptions 3.1, 3.2, and 3.3 (hereafter "model 1").

The approach to estimating the β_0 proceeds in two mains steps. Recall $\mu_t(z_{it}) = x_{it}\beta + \eta_t(z_{it})$. The strategy is to first estimate (Φ_{0it}, μ_{0it}), $t = 3, \cdots, T$, which amounts to estimating the parameters of a generalized partial linear model (GPLM) with unknown link functions. This stage of the estimator is of interest in its own right, because no known kernel-based estimator for a GPLM with unknown, non-monotonic link function that is not restricted to be strictly monotone exists. Horowitz [2001] and Gayle and Namoro [2012] develop kernel-based estimators for the case in which the link functions are assumed to be strictly monotone. The second-stage estimates β_0 using the estimates of μ_{0it}.

Suppose the investigator has data: $((y_i, w_i)\ i = 1, \cdots, n)$. Define the generic function

$$K_\sigma(a - a_i) := \sigma^{-d_a} \prod_{k=1}^{d_a} \mathfrak{K}(\sigma^{-1}(a_k - a_{ik})),$$

where d_a is the dimension of a, $\mathfrak{K}$ is a kernel, and σ is a bandwidth that converges to 0 and $n \to \infty$.

4.1 Estimating Φ_{0t}

One approach to estimating Φ_{0t} is to obtain initial consistent estimates of μ_{0t} and use them in the estimation of Φ_{0t}. The corresponding infeasible estimator for Φ_{0t} assumes μ_{0t} is known (e.g., Linton and Hardle [1996] and Linton [2000]). However, under the identification conditions, an estimate of Φ_{0t} can be obtained without the use of initial consistent estimates. Recall that under Assumption 3.2.2 (that $\mu_t(\check{z}_t) = b_t$), we have that $E[y_{it}|\tilde{v} = v, \tilde{z}_t = \check{z}_t] = \Phi_{0t}(v + b_t)$, or equivalently,

$$E[y_{it}|\tilde{v} = v - b_t, \tilde{z}_t = \check{z}_t] = \Phi_{0t}(v). \tag{4.1}$$

Thus, by assumption 3.2.1, Φ_{0t} can be recovered by varying v over its support, $\mathcal{V}$, holding z_{it} fixed at $\check{z}_t$. By analogy, an estimator for Φ_{0t} is given by

$$\hat{\Phi}_t(v) = \frac{\sum_{i=1}^n y_{it} K_\sigma(v - b_t - v_{it}) K_\sigma(\check{z}_t - z_{it})}{\sum_{i=1}^n K_\sigma(v - b_t - v_{it}) K_\sigma(\check{z}_t - z_{it})}. \tag{4.2}$$

4.2 Estimating μ_{0t}

Let $\tau(w) := 1\{w \in \mathcal{W}_1\}$, where $\mathcal{W}_1$ is a subset of $\mathcal{W}$, the support of w. $\mathcal{W}_1$ is chosen so that the density $f(w)$ is bounded away from zero on $\mathcal{W}_1$. See Newey and McFadden [1994] and Newey [1994b] for discussion on this fixed trimming condition. It can be shown that the resulting density of the transformed random variable (v, z_t) is also bounded away from zero of the corresponding support $\mathcal{V}_1 \times Z_{t1}$ (see Mood et al. [1974] for details). Now, from equation (2.7), we have for fixed $z_t \in \mathcal{Z}_{t1}$, $\mu_{0t}(z_t) = x_t\beta_0 + \eta_0(z_t)$ minimizes

$$\begin{aligned}
Q(z_t, \mu_t) &= E\left[\tau(\tilde{v}, z_t)\left(E[y_{it}|\tilde{v}, z_t] - E[y_{it}|\tilde{v} + \mu_t - b, \check{z}_t]\right)^2\right], \\
&= E\left[\tau(\tilde{v}, z_t)\left(\Phi_{0t}(\tilde{v} + \mu_{0t}(z_t)) - \Phi_{0t}(\tilde{v} + \mu_t)\right)^2\right], \\
&= E\left[\tau(\tilde{v}, z_t)\left(q_t(\tilde{v}, z_t) - \Phi_{0t}(\tilde{v} + \mu_t)\right)^2\right], \qquad (4.3)
\end{aligned}$$

where $q_t(v, z_t) = E[y_{it}|v, z_t]$. For fixed z_t of z_{it}, an estimator for $\mu_{0t}(z_t)$, denoted by $\hat{\mu}_t(z_t)$, is defined as the minimizer of

$$\begin{aligned}\hat{Q}(z_t, \mu_t) &= \frac{1}{n}\sum_{i=1}^{n}\tau(v_{it}, z_t)\left(\hat{E}[y_{jt}|v_{it}, z_t] - \hat{E}[y_{jt}|v_{it} + \mu_t - b_t, \check{z}_t]\right)^2, \\ &= \frac{1}{n}\sum_{i=1}^{n}\tau(v_{it}, z_t)\left(\hat{q}_t(v_{it}, z_t) - \hat{\Phi}_t(v_{it} + \mu_t)\right)^2, \end{aligned} \tag{4.4}$$

where

$$\hat{q}_t(v_{it}, z_t) = \hat{E}[y_{jt}|v_{it}, z_t] = \frac{\sum_{j=1}^{n} y_{jt} K_\sigma(v_{it} - v_{jt}) K_\sigma(z_t - z_{jt})}{\sum_{j=1}^{n} K_\sigma(v_{it} - v_{jt}) K_\sigma(z_t - z_{jt})}.$$

This problem is an unconstrained, unidimensional, nonlinear optimization problem that is easy to understand and compute. Any statistical package that solves nonlinear optimization problems (e.g., Matlab) should suffice.

4.3 Estimating β_0

As discussed in section 2, given knowledge of μ_{0t}, the problem of recovering β_0 takes on the following closed form:

$$\beta_0 = (C'\Omega C)^{-1} C'\Omega E[X_i'\tau_i \Delta\mu_0(z_i)], \tag{4.5}$$

where $C = E[X_i'\tau_i\Delta x_i]$, $X_i = \text{diag}[(x_{i1}, \cdots, x_{is}),\ s = 1, \cdots, T-1]$, $\Delta x_i = (\Delta x_{i3}', \cdots, \Delta x_{iT}')'$, $\Delta\mu_0(z_i) = (\Delta\mu_{03}(z_{i3}), \cdots, \Delta\mu_{0T}(z_{iT}))'$, and Ω is any $KT(T-1) \times KT(T-1)$ symmetric, positive-definite weighting matrix. To define an analogous estimator for β_0, let $\tau_i = \text{diag}(\tau(w_{it})\tau(w_{i(t+1)}), t = 2, \cdots, T-1)$, and let $\hat{\Omega}$ be a consistent estimator for Ω. Then

$$\hat{\beta} = (\hat{C}'\hat{\Omega}\hat{C})^{-1}\hat{C}'\hat{\Omega}\frac{1}{n}\sum_{i=1}^{n} X_i'\tau_i\Delta\hat{\mu}(z_i), \tag{4.6}$$

where $\hat{C} = \frac{1}{n}\sum_{i=1}^{n} X_i'\tau_i\Delta x_i$, and $\Delta\hat{\mu}(z_i) = (\Delta\hat{\mu}_3(z_{i3}), \cdots, \Delta\hat{\mu}_T(z_{iT}))'$.

4.4 Estimating η_{0t}

Given the estimators $\hat{\mu}_t$ and $\hat{\beta}$, an estimator for η_{0t} is given by $\hat{\eta}_t(z_{it}) := \hat{\mu}_t(z_{it}) - x_{it}\hat{\beta}$. A kernel smoother can be used to estimate η_{0t} for z_t off the data grid.

4.5 Estimating model 1

Recall that model 1 is defined under Assumptions 3.1, 3.2, and 3.3, which differs from model 2 in that the link function is assumed to be strictly increasing and the sign of α, the coefficient on v_{it}, is unknown. Under these restrictions, we can obtain estimators of the parameters by implementing the methods developed in Hardle et al. [2004] or Gayle and Namoro [2012] with minor modifications. These modifications are necessary because of the differences in the normalizing restrictions. However, the estimator for the parameters of model 1 requires only modest modifications of the estimator developed so far in this paper.

The first task is to estimate the sign of α, which is the same task faced by Honoré and Lewbel [2002], who propose a super-consistent estimator for the sign of α. The same estimator is appropriate in the framework of this paper.

The second issue is that of imposing the monotonicity restriction on Φ_0 in estimation. The results of Brunk [1958] and Mammen et al. [2001] suggest the monotonicity constraint can be imposed as follows:

$$\Phi_t^*(v) = \min_{u \geq v} \max_{l \leq v} \frac{1}{u - l + 1} \sum_{\tilde{v}=l}^{u} \hat{\Phi}_t(\tilde{v}).$$

5 Asymptotic Properties

In this section, we derive the asymptotic properties of the proposed estimators. The results in this section draw mainly from results in Severini and Wong [1992], Newey and McFadden [1994], and Newey [1994b] on the asymptotic properties of finite-dimensional estimators with infinite-dimensional "plug-in" nuisance parameters. This section focuses on the asymptotic properties of model 2, with only brief discussions at the end about adjustments necessary for the results to go through for model 1.

Let $\Lambda_j = \{g \in \mathcal{C}^j : \|g\|_{s,2} < \infty\}$, where $\|\cdot\|_{s,j}$ is the sobolev supremum norm of smoothness at least j. The following assumptions are necessary in what follows:

Assumption 5.1.

1. *For each* $t = 1, \cdots, T$, Φ_{0t} *is in the interior of* Λ_2 *and* μ_{0t} *are in the interior of the set* Λ_0.
2. β_0 *is in the interior of* $\mathcal{B}$, *a compact subset of* $\Re^K$.

Let $F_{y,w}$ be the joint density of $[(y_{it}, w_{it}),\ t = 1, \cdots, T]$.

Assumption 5.2. *The sample of n realizations is drawn from* $F_{y,w}$. *For each* $i = 1, \cdots, n$, $[(y_{it}, w_{it}), t = 1, \cdots, T]$ *is observed.*

Define $\gamma_2 := (\gamma_{21}, \gamma_{22})'$, where $\gamma_{22}(w) = E[q|w]f(w)$ and $\gamma_{21} = f(w)$.

Assumption 5.3.

1. $\mathfrak{K}(a)$ *is differentiable of order* $d \geq 2$, *the derivatives of order d are bounded,* $\mathfrak{K}(a)$ *is zero outside a bounded set,* $\int \mathfrak{K}(a)da = 1$, *and a positive integer m exists such that for all* $j < m$, $\int a^j \mathfrak{K}(a)da = 0$.
2. *A version of* $\gamma_2(w)$ *exists that is continuously differentiable to order d with bounded derivatives on an open set containing* $\mathcal{W}$.

3. $p \geq 4$ *exists such that* $E[||y||^p] \leq \infty$ *and* $E[||y||^p|w]f(w)$ *is bounded.*

Assumption 5.4. $\hat{\Omega} \xrightarrow{p} \Omega$

5.1 Consistency and rates of convergence

Theorem 5.5. *Suppose (i) Assumptions 3.1 and 3.2 hold, (ii) Assumption 3.5 holds, and (iii) Assumptions 5.1 - 5.4 hold. Let* $\sigma^m \to 0$, *and* $(\ln(n))^{1/2}(n\sigma^{K+2d})^{-1/2} \to 0$. *Then for* $t = 2, \cdots, T$,

$$\left\|\hat{\Phi}_t - \Phi_{0t}\right\|_{s,2} \xrightarrow{p} 0,$$

$$\|\hat{\mu}_t - \mu_{0t}\|_{s,0} \xrightarrow{p} 0,$$

and $\hat{\beta} \xrightarrow{p} \beta_0$.

Proof. See Appendix A.3 □

Theorem 5.6. *Suppose (i) Assumptions 3.1 and 3.2 hold, (ii) Assumption 3.5 holds, and (iii) Assumptions 5.1 - 5.4 hold. Let* $\sqrt{n}\sigma^{2m} \to 0$, *and* $\sqrt{n}\ln(n)/n\sigma^{K+2d} \to 0$. *Then for* $t = 1, \cdots, T$,

$$\left\|\hat{\Phi}_t - \Phi_{0t}\right\|_{s,2} = o_p(n^{-1/4}),$$

and

$$\|\hat{\mu}_t - \mu_{0t}\|_{s,0} = o_p(n^{-1/4}).$$

Proof. See Appendix A.4 □

Notice the weaker convergence results for $\hat{\mu}_t$ as compared to $\hat{\Phi}_t$. These results can be strengthened at the cost of stronger regularity conditions, particulary on the smoothness of Φ_{0t} (e.g., Lemma 5 of Severini and Wong [1992]). However, these conditions are sufficient to obtain the required rate of convergence of $\hat{\beta}$.

5.2 Asymptotic normality

Theorem 5.7. *Suppose (i) Assumptions 3.1 and 3.2 hold, (ii) Assumption 3.5 holds, and (iii) Assumptions 5.1 - 5.4 hold. Let* $\sqrt{n}\sigma^{2m} \to 0$, *and* $\sqrt{n}\ln(n)/n\sigma^{K+2d} \to 0$. *Then* $\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0,V)$, *where* $V := (C'\Omega C)^{-1}C'\Omega\Sigma\Omega C(C'\Omega C)^{-1}$, $\Sigma = E[X_i'\tau_i\Delta(\eta_{0i}+\delta_i e_i)(\Delta(\eta_{0i}+\delta_i e_i))'\tau_i X_i]$, $e_i = y_i - q_i$, *and* $\delta_i = diag(\delta(w_{it}), t = 2, \cdots, T)$, *where*

$$\delta(w_{it}) = \frac{f(v_{it})f(z_{it})}{Q^{(2)}(z_{it})f(w_{it})},$$

and $Q^{(2)}(z_{it})$ *is the second derivative of* $Q(z_{it},\mu_t)$ *with respect to* μ_t *evaluated at* μ_{0t}.

Proof. See Appendix A.5 ☐

From the preceding, the efficient choice of the weighting matrix is clearly given by $\Omega = \Sigma^{-1}$, reducing the variance matrix to $V = [C\Sigma^{-1}C']^{-1}$. In either case, we obtain a consistent estimator for V by replacing the expectations, population densities, and objective functions with their respective sample counterparts. We can apply the results of Newey and McFadden [1994] and Newey [1994b] to prove consistency of the resulting estimator.

6 Monte Carlo Study

In this section, we present two Monte Carlo exercises to illustrate the performance of the estimators. In both exercises, v_{it} given y_{it-1} is i.i.d. $N(\alpha y_{it-1}, \sigma_v^2)$, c_i is i.i.d. $N(0,\sigma_c^2)$, and the covariance between v_{it} and c_i is σ_{vc}. Hence,

$$v_{it} = \alpha y_{it-1} + \frac{\sigma_{vc}}{\sigma_c^2}c_i + \varepsilon_{it}^v, \tag{6.1}$$

where ε_{it}^{v} i.i.d. $N(0,\sigma_v^2-\sigma_{vc}^2/\sigma_c^2)$, and

$$c_i|v_i^t,y_i^{t-1}\sim N\left(\frac{\sigma_{vc}}{\sigma_v^2}\sum_{s=1}^{t}(v_{is}-\alpha y_{is-1}),\sigma_c^2-t\frac{\sigma_{vc}^2}{\sigma_v^2}\right). \tag{6.2}$$

Thus, $R_t(v^t)=\sum_{s=1}^{t}v_{is}$, $z_{it}=(R_t(v_i^t),y_i^{t-1})$, and $\eta_t(R_t(v_i^t),y_i^{t-1})=\frac{\sigma_{vc}}{\sigma_v^2}\sum_{s=1}^{t}(v_{is}-\alpha y_{is-1})$. In both simulation exercises, $\sigma_v^2=3$, $\sigma_c^2=0.5$, $\sigma_{vc}=0.1$, and $\alpha=0.2$. We take $R_t(v^t)=\sum_{s=1}^{t}v_{is}$ and perform 100 Monte Carlo replications of the model with $T=4$, and three sample sizes N: 200, 400, and 800, where we estimate (β,η,Φ) for each replication. The mean bias (MB), mean absolute bias (MAB), and the root mean square error (RMSE) of the simulated values of β are calculated for each sample size. In both exercises, we set $\check{z}_t=\mathbf{0}$, and $b_t=0$. The results of the simulation exercises are robust to the normalization choice. As a general guide, one should set $\check{z}_t$ to be in the interior of support of z_{it}, and b_t to be in the interior of the support of v_{it}.

6.1 Design 1

For the first exercise, we consider the following data-generating process for the discrete outcome dependent variable:

$$y_{it}=1\{v_{it}+\beta y_{it-1}+c_i-\varepsilon_{it}\geq 0\}, \tag{6.3}$$

where ε_{it} is i.i.d. $N(0,1)$, $\beta=-0.5$. The process is initialized as follows: $y_{i0}=1\{\varepsilon_{i0}\leq 0\}$. Because parameters are estimated using minimum distance, and not maximum likelihood, we do not need to specify the initial conditions for y_{i0} in estimation. Therefore, allowing for y_{i0} to depend on c_i does not pose a problem. Given the above specification for the distribution of (v_{it},c_i), equation (6.3) is equivalently given by

$$y_{it}=1\{v_{it}+\beta y_{it-1}+\eta_t(z_{it})-u_{it}\geq 0\},\quad t=1,\cdots,4,$$

where u_{it} is independent of (v_{it}, z_{it}) and distributed $N(0, 1+\sigma_c^2 - t\frac{\sigma_{vc}^2}{\sigma_v^2})$.

For comparison, we also simulate the model assuming u_{it} is known. The resulting estimator is then similar to that proposed in Arellano and Carrasco [2003]. The first-stage estimator where $\hat{\Sigma}$ is set to be the identity matrix is denoted by KMD, and the second-stage estimator is denoted by OKMD. The corresponding first- and second-stage estimators for the model in which the distribution of u_{it} is assumed to be known are denoted by *AC* and *OAC*, respectively. The results are presented in Table 1.

The simulation results show the proposed estimator works well in small samples. As expected, the estimator in which the link function is known outperforms the proposed estimator in terms of mean squared error. However, our estimator does better in terms of mean bias. To understand why, notice the conditional expectations are functions of two continuous random variables, v_{it} and $\sum_{s=1}^{t} v_{is}$. To obtain $\sqrt{n}$ convergence of β, high-order kernels are therefore necessary. We use the third-order Epanechnikov kernel in this exercise. The result is that some of the estimated probabilities, $\hat{q}_{it}$, are outside the unit interval, whereas the inverse link function is not defined outside the unit interval. These estimated probabilities are therefore trimmed away, resulting in the higher mean bias.

6.2 Design 2

For the second exercise, consider the following data-generating process for the binary outcome dependent variable:

$$y_{it} = 1\{G(v_{it} + \beta y_{it-1} + c_i) - \varepsilon_{it} \geq 0\},$$

where $G(a) = 0.25a + SIN(a)$, ε_{it} i.i.d. $N(0,1)$, $\beta = -0.5$, and $y_{i0} = 1\{\varepsilon_{i0} \leq 0\}$. Unlike in design 1, the resulting index function is not strictly monotonic. Indeed, it exhibits significant oscillation, possessing six turning points on the interval $[-10, 10]$. However, this index function is not periodic on its support. Table 2 presents the simulation results in the same format as Table 1.

Table 1: Small sample properties of the estimator of model design 1.

n=200			
	MB	MAB	RMSE
KMD	0.0890	0.6610	1.0210
OKMD	0.0906	0.4865	0.7598
AC	-0.4234	0.6181	0.7522
OAC	-0.4176	0.5330	0.6641
n=400			
	MB	MAB	RMSE
KMD	0.0820	0.4597	0.5852
OKMD	0.0894	0.3547	0.4617
AC	-0.1987	0.3119	0.3894
OAC	-0.2432	0.2931	0.3648
n=800			
	MB	MAB	RMSE
KMD	-0.0391	0.3783	0.5162
OKMD	0.0190	0.2810	0.3803
AC	-0.1312	0.1878	0.2300
OAC	-0.1553	0.1959	0.2355

Table 2: Small sample properties of the estimator of model design 2.

n=200			
	MB	MAB	RMSE
KMD	0.0596	0.9309	1.3410
OKMD	0.0649	0.6964	1.1100
n=400			
	MB	MAB	RMSE
KMD	0.1639	0.7275	1.0590
OKMD	0.1129	0.5671	0.8318
n=800			
	MB	MAB	RMSE
KMD	0.0340	0.4759	0.6436
OKMD	0.0991	0.3824	0.5228

7 Conclusion

This paper provides new identification conditions for a class of single-index panel data models in which (i) all the explanatory variables may be predetermined, (ii) the link function is unspecified, and (iii) the individual effects may be correlated with all the explanatory variables. We propose two alternative sets of sufficient conditions to identify the parameters of interest. The first requires the link function to be strictly increasing. The second does not assume the link function is strictly increasing, but imposes restrictions on its potential oscillation. We propose kernel-based estimators for the two models. We also show the estimators for β are $\sqrt{n}$-consistent and asymptotically normal with a simple form for the covariance matrix.

Individually, all of the identification assumptions are not new, and have been used in other works on nonlinear panel data models. However, the combination of the assumptions this paper proposes provides a powerful source of identification of the key parameters of interest to empirical researchers. Note that the structural link function is not identified. What we do identify is the integral of the structural link function with respect to the distribution of the "pure" random effects. As a result, one should interpret partial effects and predictions made from the estimated model as the partial effects and predictions after integrating over the distribution of the pure random effects.

The first-stage estimator for the link function and the nonparametric function μ_t proposed in this paper is interesting in its own right because, to the best or our knowledge, no known kernel estimator of a GPLM with an unknown link function that is not restricted to be strictly increasing exists. This estimator generalizes the semiparametric least squares estimator of Ichimura [1993] to panel data and partial linear contexts.

Acknowledgements

The author is grateful to Shakeeb Khan, Nathan Larson, Whitney Newey, Steven Stern, Nese Yildiz, and three anonymous referees for insightful comments and discussions. The author also thanks participants at the conference on Panel Data (Bonn, Germany) and in seminars at the University of Rochester and the University of Virginia. All errors are my own.

A LEMMAS AND THEOREMS

A.1 Proof of Theorem 3.4

Proof. Suppose another parameter vector $\theta_1 = (\Phi_1, \eta_1, B_1)$ exist that satisfies Assumptions 3.1, 3.2, and 3.3 such that

$$E[y_{it}|w_i^t] = \Phi_{t1}(\alpha_1 v_{it} + x_{it}\beta_1 + \eta_{t1}(z_{it})), \quad t = 1, \cdots, T.$$

Because $\check{z}_t \in \mathcal{Z}_t$ and

$$\mu_{t0}(\check{z}_t) = \check{x}_{it}\beta_0 + \eta_{t0}(\check{z}_{it}) = b_t = \check{x}_{it}\beta_1 + \eta_{t1}(\check{z}_{it}) = \mu_{t1}(\check{z}_t),$$

taking expectation of y_{it} conditioned on $(\tilde{v} = v_{it}, \tilde{z}_t = \check{z}_t)$ gives $q(v_{it}, \check{z}_t) = E[y_{it}|\tilde{v} = v_{it}, \tilde{z}_t = \check{z}_t]$, which is known in the population. This result, and equation (3.2) give for each t

$$\Phi_{t0}(v_{it}\alpha_0 + b_t) = \Phi_{t1}(v_{it}\alpha_1 + b_t). \tag{A.1}$$

Differentiating with respect to v_{it} obtains

$$\Phi'_{t0}(v_{it}\alpha_0 + b_t) \cdot \alpha_0 = \Phi'_{t1}(v_{it}\alpha_1 + b_t) \cdot \alpha_1. \tag{A.2}$$

Because $|\alpha_0| = |\alpha_1| = 1$ and the index function is strictly increasing, taking the absolute value of equation (A.2) obtains

$$\begin{aligned} \Phi'_{t0}(v_{it}\alpha_0 + b_t)|\alpha_0| &= \Phi'_{t1}(v_{it}\alpha_1 + b_t)|\alpha_1| \Rightarrow \\ \Phi'_{t0}(v_{it}\alpha_0 + b_t) &= \Phi'_{t1}(v_{it}\alpha_1 + b_t) \Rightarrow \\ \alpha_0 &= \alpha_1. \end{aligned} \tag{A.3}$$

Substituting equation (A.3) into equation (A.1) gives

$$\Phi_{t0}(v_{it}\alpha_0 + b_t) = \Phi_{t1}(v_{it}\alpha_0 + b_t).$$

This equality holds for all v_{it} by assumption 3.2.1 so that $\Phi_{t0}(\cdot) = \Phi_{t1}(\cdot)$. Therefore, taking expectations of y_{it} conditional on w_i^t obtains

$$\Phi_{t0}(v_{it}\alpha_0 + x_{it}\beta_0 + \eta_{t0}(z_{it})) = \Phi_{t1}(v_{it}\alpha_1 + x_{it}\beta_1 + \eta_{t1}(z_{it})) = \Phi_{t0}(v_{it}\alpha_0 + x_{it}\beta_1 + \eta_{t1}(z_{it})),$$

which implies

$$x_{it}\beta_0 + \eta_{t0}(z_{it}) = x_{it}\beta_1 + \eta_{t1}(z_{it}), \tag{A.4}$$

because Φ_{t0} is strictly increasing. Taking first difference of equation (A.4) and stacking obtains the $T-1$ system of equations

$$\Delta x_i\beta_0 + \Delta\eta_0(z_i) = \Delta x_i\beta_1 + \Delta\eta_1(z_i). \tag{A.5}$$

By the law of iterated expectations, for $t \geq 2$,

$$\begin{aligned}
E[x_i^{t-1}\Delta\eta_0(z_{it})] &= E[x_i^{t-1}(E[c_i|z_{it}] - E[c_i|z_{it-1}])] \\
&= E[x_i^{t-1}(E[c_i|R_t(v_i^t), x_i^t] - E[c_i|R_t(v_i^{t-1}), x_i^{t-1}])] \\
&= E[x_i^{t-1}E[E[c_i|R_t(v_i^t), x_i^t] - E[c_i|R_{t-1}(v_i^{t-1}), x_i^{t-1}]|x^{t-1}]] \\
&= 0.
\end{aligned} \tag{A.6}$$

By the same argument, $E[x_i^{t-1}\Delta\eta_1(z_{it})] = 0$. Let X_i be the $(T-1)\times KT(T-1)$ matrix of the form $X_i =$diag$[(x_{i1},\cdots,x_{is}),\ s = 1,\cdots,T-1]$, let Ω be any $KT(T-1)\times KT(T-1)$ symmetric, positive-definite weighting matrix, and let $C = E[X_i'\Delta x_i]$. Premultiplying equation (A.5) by $C\Omega X_i'$ and taking expectations obtains

$$[C'\Omega C]\beta_0 + C'\Omega E[X_i'\Delta\eta_0(z_i)] = [C'\Omega C]\beta_1 + C\Omega E[X_i'\Delta\eta_1(z_i)]. \tag{A.7}$$

Equations (A.6) and (A.7), along with Assumption 3.2.3, obtain $\beta_0 = \beta_1$. This equality, and equation (A.4) imply $\eta_{t0}(\cdot) = \eta_{t1}(\cdot),\ t = 2, \cdots, T$.

Note that given identification of $\mu_0(z_i) = x_i\beta_0 + \eta(z_i)$, equation (A.6) and Assumption 3.2.3 obtain the following closed-form expression for β_0:

$$\beta_0 = (C'\Omega C)^{-1}C'\Omega E[X_i \Delta\mu(z_i)].$$

□

A.2 Proof of Theorem 3.6

Proof. Because $\check{z}_t \in \mathcal{Z}_t$, $\eta_{t0}(\check{z}_t) = \eta_{t1}(\check{z}_t) = b_t$, and $\alpha_0 = \alpha_1 = 1$, taking expectation of y_{it} conditioned on $(\tilde{v} = v_{it}, \tilde{z}_t = \check{z}_t)$ gives

$$\Phi_{t0}(v_{it} + b_t) = \Phi_{t1}(v_{it} + b_t).$$

This equality holds for all v_{it} so that by Assumption 3.2.1, $\Phi_{t0}(\cdot) = \Phi t1(\cdot)$. Taking expectation of y_{it} conditioned on $(\tilde{v} = v_{it}, \tilde{z} = \overline{z_t})$ gives

$$\Phi_{t0}(v_{it} + \overline{x_t}\beta_0 + \eta_{t0}(\overline{z_t})) \quad = \quad \Phi_{t0}(v_{it} + \overline{x_t}\beta_1 + \eta_{t1}(\overline{z_t})),$$

which again holds for all v_{it}. Assumptions 3.2.1 and 3.5.1 imply $\overline{x_t}\beta_0 + \eta_{t0}(\overline{z_t}) = \overline{x_t}\beta_1 + \eta_{t1}(\overline{z_t})$, which holds for any $\overline{z_t}$ so that $x_{it}\beta_0 + \eta_{t0}(z_{it}) = x_{it}\beta_1 + \eta_{t1}(z_{it})$. The rest of the proof is similar to A.1. □

A.3 Proof of Theorem 5.5

Proof. Under the bandwidth conditions of the theorem, lemma A.2 obtains

$$\sup_{z_t}\sup_{\mu_t}|\hat{Q}(z_t,\mu_t)-Q(z_t,\mu_t)|=o_p(1). \tag{A.8}$$

Because $\hat{\mu}_t$ is the minimizer of $\hat{Q}(z_t,\mu_t)$, which is nonnegative, and $Q(z_t,\mu_{0t})=0$ identically in z_t, $0\leq|\hat{Q}(z_t,\hat{\mu}_t)-Q(z_t,\mu_{0t})|\leq|\hat{Q}(z_t,\mu_{0t})-Q(z_t,\mu_{0t})|\leq|\hat{Q}(z_t,\mu_{0t})-Q(z_t,\mu_{0t})|$ for any z_t, so that

$$0\leq\sup_{z_t}|\hat{Q}(z_t,\hat{\mu}_t)-Q(z_t,\mu_{0t})|\leq\sup_{z_t}|\hat{Q}(z_t,\mu_{0t})-Q(z_t,\mu_{0t})|,$$

obtaining

$$0\leq\sup_{z_t}|\hat{Q}(z_t,\hat{\mu}_t)-Q(z_t,\mu_{0t})|\leq\sup_{z_t}\sup_{\mu_t}|\hat{Q}(z_t,\mu_t)-Q(z_t,\mu_t)|=o_p(1) \tag{A.9}$$

by equation (A.8). By the triangular inequality,

$$|Q(z_t,\hat{\mu}_t)-Q(z_t,\mu_{0t})|\leq|\hat{Q}(z_t,\hat{\mu}_t)-Q(z_t,\mu_t)|+|\hat{Q}(z_t,\hat{\mu}_t)-Q(z_t,\mu_{0t})|.$$

Therefore, equations (A.8) and (A.9) imply

$$\begin{aligned}0\leq\sup_{z_t}|Q(z_t,\hat{\mu}_t)-Q(z_t,\mu_{0t})|&\leq\sup_{z_t}\sup_{\mu_t}|\hat{Q}(z_t,\mu_t)-Q(z_t,\mu_t)|\\&\quad+\sup_{z_t}|\hat{Q}(z_t,\hat{\mu}_t)-Q(z_t,\mu_{0t})|=o_p(1).\end{aligned} \tag{A.10}$$

Given the model is identified, for all $\delta>0$, $\varepsilon>0$ exists such that $|\hat{\mu}_t(z_t)-\mu_{0t}(z_t)|>\delta\Rightarrow|Q(z_t,\hat{\mu}_t)-Q(z_t,\mu_{0t})|>\varepsilon$ for any z_t. Therefore, by equation (A.10),

$$P\left\{\sup_{z_t}|\hat{\mu}_t(z_t)-\mu_{0t}(z_t)|>\delta\right\}\leq P\left\{\sup_{z_t}|Q(z_t,\hat{\mu}_t)-Q(z_t,\mu_t)|>\varepsilon\right\}=o_p(1).$$

Thus $\sup_{z_t} |\hat{\mu}_t(z_t) - \mu_{0t}(z_t)| = o_p(1)$. Finally, given Assumption 5.4 and the above result, it is straightforward to show $\hat{\beta} \xrightarrow{p} \beta_0$. $\square$

To prove the rest of the theorems of the text, additional notations are necessary. For fixed z_t and for $r = 0, 1, 2$, let

$$\begin{aligned}
Q^{(r)}(v_{it}, z_t, \mu_t) &= \partial^r (q(v_{it}, z_t) - \Phi_{0t}(v_{it} + \mu_t))^2 / \partial \mu_t^r, \\
\hat{Q}^{(r)}(v_{it}, z_t, \mu_t) &= \partial^r (\hat{q}(v_{it}, z_t) - \hat{\Phi}_t(v_{it} + \mu_t))^2 / \partial \mu_t^r, \\
Q^{(r)}(z_t, \mu_t) &= E[\tau(v_{it}, z_t) Q^{(r)}(v_{it}, z_t, \mu_t)], \\
\tilde{Q}^{(r)}(z_t, \mu_t) &= \sum_i \tau(v_{it}, z_t) Q^{(r)}(v_{it}, z_t, \mu_t)/n, \text{ and} \\
\hat{Q}^{(r)}(z_t, \mu_t) &= \sum_i \tau(v_{it}, z_t) \hat{Q}^{(r)}(v_{it}, z_t, \mu_t)/n.
\end{aligned}$$

We will also use $\phi_{0t}(v) = \partial \Phi_{0t}(v)/\partial v$, and $\hat{\phi}_t(v) = \partial \hat{\Phi}_t(v)/\partial v$.

A.4 Proof of Theorem 5.6

Proof. The bandwidth conditions of the theorem along with the results of lemma A.1 imply the required result of $\hat{\Phi}_t$. Because $\hat{Q}^{(1)}(z_t, \hat{\mu}_t) = Q^{(1)}(z_t, \mu_{0t}) = 0$, by the mean value expansion,

$$0 = \hat{Q}^{(1)}(z_t, \hat{\mu}_t) - Q^{(1)}(z_t, \hat{\mu}_t) + Q^{(2)}(z_t, \overline{\mu}_t)(\hat{\mu}_t - \mu_{0t}),$$

where $\overline{\mu}_t$ is between μ_{0t} and $\hat{\mu}_t$. Thus

$$\hat{\mu}_t(z_t) - \mu_{0t}(z_t) = -\left[Q^{(2)}(z_t, \overline{\mu}_t)\right]^{-1} [\hat{Q}^{(1)}(z_t, \hat{\mu}_t) - Q^{(1)}(z_t, \hat{\mu}_t)].$$

By identification of μ_{0t} and consistency of $\hat{\mu}_t$, for n large enough, for $\delta > 0$, we have $\inf_{z_t} |Q^{(2)}(z_t, \overline{\mu}_t)| \leq \inf_{z_t} \sup_{|\mu_t - \mu_{0t}| \leq \delta} |Q^{(2)}(z_t, \mu_t)| > 0$. Thus we have

$$\begin{aligned}
\sup_{z_t} |\hat{\mu}_t(z_t) - \mu_{0t}(z_t)| &\leq \sup_{z_t} \left\| \left[Q^{(2)}(z_t, \overline{\mu}_t) \right]^{-1} \right\| \sup_{z_t} \|\hat{Q}^{(1)}(z_t, \hat{\mu}_t) - Q^{(1)}(z_t, \hat{\mu}_t)\| \\
&\leq \left[\inf_{z_t} \|Q^{(2)}(z_t, \overline{\mu}_t)\| \right]^{-1} \sup_{\mu_t} \sup_{z_t} \|\hat{Q}^{(1)}(z_t, \mu_t) - Q^{(1)}(z_t, \mu_t)\| \\
&= O_p((\ln(n)/n\sigma^{d_w+2r})^{1/2} + \sigma^m + n^{-1/2}),
\end{aligned}$$

where the last equality comes from lemma A.2. Applying the bandwidth conditions of the theorem completes the proof. □

A.5 Proof of Theorem 5.7

Proof. Recall $\mu_0(z_i) = x_t \beta_0 + \eta(z_i)$ so that

$$\begin{aligned}
\hat{\beta} &= (\hat{C}'\hat{\Omega}\hat{C})^{-1}\hat{C}'\hat{\Omega}\frac{1}{n}\sum_{i=1}^{n} X_i'\tau_i \Delta\hat{\mu}(z_i) \\
&= \beta_0 + (\hat{C}'\hat{\Omega}\hat{C})^{-1}\hat{C}'\hat{\Omega}\frac{1}{n}\sum_{i=1}^{n} X_i'\tau_i \Delta\eta_0(z_i) + (\hat{C}'\hat{\Omega}\hat{C})^{-1}\hat{C}'\hat{\Omega}\frac{1}{n}\sum_{i=1}^{n} X_i'\tau_i \Delta(\hat{\mu}(z_i) - \mu_0(z_i)).
\end{aligned}$$

Thus,

$$\sqrt{n}(\hat{\beta} - \beta_0) = (\hat{C}'\hat{\Omega}\hat{C})^{-1}\hat{C}'\hat{\Omega}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'\tau_i \Delta\eta_0(z_i) + (\hat{C}'\hat{\Omega}\hat{C})^{-1}\hat{C}'\hat{\Omega}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'\tau_i \Delta(\hat{\mu}(z_i) - \mu_0(z_i)).$$

To compute the limiting distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$, consider the scaled sample average $\sum_i a(z_{it})(\hat{\mu}_t(z_{it}) - \mu_{0t}(z_{it}))/\sqrt{n}$ and recall that $\hat{Q}^{(1)}(z_t, \hat{\mu}_t) = 0$. Under the bandwidth conditions of the theorem, and using lemmas A.1 and A.2, expanding the latter term gives

$$
\begin{aligned}
0 \quad &= \tilde{Q}^{(1)}(z_t,\mu_{0t}) \\
&+\tilde{Q}^{(2)}(z_t,\mu_{0t})(\hat{\mu}_t(z_t)-\mu_{0t}(z_t)) \\
&+\tfrac{1}{n}\textstyle\sum_i \tau(v_{it},z_t)\phi_{0t}(v_{it}+\mu_{0t}(z_t))[\hat{\Phi}_t(v_{it}+\mu_{0t}(z_t))-\Phi_{0t}(v_{it}+\mu_{0t}(z_t))] \\
&+\tfrac{1}{n}\textstyle\sum_i (q(v_{it},z_t))-\Phi_{0t}(v_{it}+\mu_{0t}(z_t))[\hat{\phi}_t(v_{it}+\mu_{0t}(z_t))-\phi_{0t}(v_{it}+\mu_{0t}(z_t))] \\
&+\tfrac{1}{n}\textstyle\sum_i \tau(v_{it},z_t)\phi_{0t}(v_{it}+\mu_{0t}(z_t))[\hat{q}(v_{it},z_t)-q(v_{it},z_t)]+o_p(1/\sqrt{n}).
\end{aligned}
$$

Because $q(v_{it},z_t)=\Phi_{0t}(v_{it}+\mu_{0t}(z_t))$ for all $(v_{it},z_t)\in\mathcal{W}_1$, $\tilde{Q}^{(1)}(z_t,\mu_{0t})=0$. Therefore, the first and fourth terms of the above equation are identically zero. Therefore, by lemma A.2, we have

$$
\begin{aligned}
\hat{\mu}_t(z_t)-\mu_{0t}(z_t) \quad &= \tfrac{1}{n}\textstyle\sum_i \tau(v_{it},z_t)\phi_{0t}(v_{it}+\mu_{0t}(z_t))[\hat{q}(v_{it},z_t)-\hat{\Phi}_t(v_{it}+\mu_{0t}(z_t))]/Q^{(2)}(z_t,\mu_{0t})+o_p(1/\sqrt{n}) \\
&= \tfrac{1}{n}\textstyle\sum_i Q^{(2)}(z_t)^{-1}G(v_{it},z_t)[\gamma_{2t}(v_{it},z_t)-\gamma_{1t}(v_{it},z_t)]+o_p(1/\sqrt{n}),
\end{aligned}
$$

where $G(v_{it},z_t)=\tau(v_{it},z_t)f(v_{it},z_t)^{-1}\phi_{0t}(v_{it}+\mu_{0t}(z_t))[-\Phi_{0t}(v_{it}+\mu_{0t}(z_t)),1]$,
$\hat{\gamma}_1(v_{it},z_t)=\sum_j(1,\hat{q}(w_{jt}))'K_\sigma(v_{it}+\mu_{0t}(z_t)-b_t-v_{jt})K_\sigma(\check{z}_t-z_{jt})$,
and $\hat{\gamma}_2(v_{it},z_t)=\sum_j(1,y_{jt})'K_\sigma(v_{it}-v_{jt})K_\sigma(z_t-z_{jt})$. Define $b(z_{it})=a(z_{it})Q^{(2)}(z_{it})^{-1}$. Then

$$
\frac{1}{n}\sum_i a(z_{it})(\hat{\mu}_t(z_{it})-\mu_{0t}(z_{it}))=\frac{1}{n}\sum_i b(z_{it})\frac{1}{n}\sum_j G(v_{jt},z_{it})[\gamma_{2t}(v_{jt},z_{it})-\gamma_{1t}(v_{jt},z_{it})]+o_p(1/\sqrt{n}).
$$

Now the first term

$$
\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})\gamma_{2t}(v_{jt},z_{it}) \\
&= \tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})\sum_l(1,y_{lt})'K_\sigma(v_{jt}-v_{lt})K_\sigma(z_{it}-z_{lt}) \\
&= \tfrac{1}{n}\textstyle\sum_l(1,y_{lt})\tfrac{1}{n}\sum_i b(z_{it})K_\sigma(z_{it}-z_{lt})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})'K_\sigma(v_{jt}-v_{lt}) \\
&= \tfrac{1}{n}\textstyle\sum_l(1,y_{lt})\tfrac{1}{n}\sum_i b(z_{it})K_\sigma(z_{it}-z_{lt})\hat{h}_l(z_{it}) \\
&= \tfrac{1}{n}\textstyle\sum_l(1,y_{lt})\tfrac{1}{n}\sum_i D_l(z_{it})\hat{h}_l(z_{it}).
\end{aligned}
$$

Define $h_{0l}(z_{it}) := f(v_{lt})G(v_{lt}, z_{it})'$, $\overline{h}_l(z_{it}) = E[\hat{h}_l(z_{it})]$, $m_l(v_{jt}, z_{it}) := b(z_{it})K_\sigma(z_{it} - z_{lt})G(v_{jt}, z_{it})'K_\sigma(v_{jt} - v_{lt})$, and consider

$$\frac{1}{n}\sum_i b(z_{it})K_\sigma(z_{it} - z_{lt})(\hat{h}(v_{lt}, z_{it}) - h_0(v_{lt}, z_{it})).$$

Under Assumptions 5.1 and 5.2, we have that $E[\|m_l(v_{it}, z_{it})\|^2] = O(\sigma^{-2d_w})$ and $E[\|m_l(v_{jt}, z_{it})\|^2] = O(\sigma^{-2d_w})$ so that by lemma 8.4 of Newey and McFadden [1994],

$$\sqrt{n}\left\{\frac{1}{n}\sum_i D_l(z_{it})(\hat{h}_l(z_{it}) - \overline{h}_l(z_{it})) - E[D_l(z_{it})(\hat{h}_l(z_{it}) - \overline{h}_l(z_{it}))]\right\} = o_p(1).$$

Also, similar to results of Newey and McFadden [1994], we can show $\|\sqrt{n}E[D_l(z_{it})(\overline{h}_l(z_{it}) - h_{0l}(z_{it}))]\| = O(\sqrt{n}\sigma^m)$ and $E\left[\|D_l(z_{it})(\overline{h}_l(z_{it}) - h_{0l}(z_{it}))\|^2\right] = O(1/(\sqrt{n}\sigma^{d_w}))$. Therefore, under the bandwidth conditions of the theorem,

$$\sqrt{n}\left\{\frac{1}{n}\sum_i D_l(z_{it})(\hat{h}_l(z_{it}) - h_{0l}z_{it})) - E[D_l(z_{it})(\hat{h}_l(z_{it}) - h_{0l}(z_{it}))]\right\} = o_p(1).$$

Furthermore, similar calculations show $\|\sqrt{n}E[E[D_l(z_{it})(\hat{h}_l(z_{it}) - h_{0l}(z_{it}))]]\| = \sqrt{n}\|E[D_l(z_{it})(\overline{h}_l(z_{it}) - h_{0l}(z_{it}))]\| = o_p(1)$ and $\sqrt{n}E[\|E[D_l(z_{it})(\hat{h}_l(z_{it}) - h_{0l}(z_{it}))]\|^2] = o_p(1)$ so that by the Tchebychev inequality, $\sqrt{n}E[D_l(z_{it})(\overline{h}_l(z_{it}) - h_{0l}(z_{it}))] = o_p(1)$. Therefore,

$$\sum_i b(z_{it})K_\sigma(z_{it} - z_{lt})\frac{1}{n}\sum_j G(v_{jt}, z_{it})'K_\sigma(v_{jt} - v_{lt}) = \sum_i f(v_{lt})b(z_{it})G(v_{lt}, z_{it})'K_\sigma(z_{it} - z_{lt}) + o_p(1/\sqrt{n}).$$

Because $E[\|y\|^4] < \infty$,

$$\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_l(1, y_{lt})\tfrac{1}{n}\sum_i b(z_{it})K_\sigma(z_{it} - z_{lt})\tfrac{1}{n}\sum_j G(v_{jt}, z_{it})'K_\sigma(v_{jt} - v_{lt}) \\
&= \tfrac{1}{n}\textstyle\sum_l(1, y_{lt})\tfrac{1}{n}\sum_i f(v_{lt})b(z_{it})G(v_{lt}, z_{it})'K_\sigma(z_{it} - z_{lt}) + o_p(1/\sqrt{n}) \\
&= \tfrac{1}{n}\textstyle\sum_l(1, y_{lt})f(v_{lt})f(z_{lt})b(z_{lt})G(v_{lt}, z_{lt})' + o_p(1/\sqrt{n}) \\
&= \tfrac{1}{n}\textstyle\sum_i a(z_{it})(Q^{(2)}(z_{it})f(w_{it}))^{-1}f(v_{lt})f(z_{lt})(y_{it} - \Phi_{0t}(v_{it} + \mu_{0t}(z_{it})) + o_p(1/\sqrt{n}),
\end{aligned}$$

where the second equality is achieved by repeating the procedure above.

Consider now

$$\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})\gamma_{1t}(v_{jt},z_{it})\\
&= \tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})\tfrac{1}{n}\sum_l (1,\hat{q}_{lt})' K_\sigma(v_{jt}+\mu_{0t}(z_{it})-b_t-v_{lt})K_\sigma(\check{z}_t-z_{lt})\\
&= \tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})\tfrac{1}{n}\sum_l (1,q_{lt})' K_\sigma(v_{jt}+\mu_{0t}(z_{it})-b_t-v_{lt})K_\sigma(\check{z}_t-z_{lt})\\
&\quad + \tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})\tfrac{1}{n}\sum_l (0,(\hat{q}_{lt}-q_{lt}))' K_\sigma(v_{jt}+\mu_{0t}(z_{it})-b_t-v_{lt})K_\sigma(\check{z}_t-z_{lt}).
\end{aligned}$$

Also, define

$$m_i(v_{jt},z_{lt}) \quad = \tfrac{1}{n}\textstyle\sum_j G(v_{jt},z_{it})\tfrac{1}{n}\sum_l (\hat{q}_{lt}-q_{lt})' K_\sigma(v_{jt}+\mu_{0t}(z_{it})-b_t-v_{lt})K_\sigma(\check{z}_t-z_{lt}).$$

Calculations similar to those above obtain $E[m_i(v_{jt},z_{jt})^2] = O(n^{-1/2}\sigma^{-2d_w})$ and $E[m_i(v_{jt},z_{lt})^2] = O(n^{-1/2}\sigma^{-2d_w})$. Also

$$\sqrt{n}\|E[G(v_{jt},z_{it})E[(\hat{q}_{lt}-q_{lt})' K_\sigma(v_{jt}+\mu_{0t}(z_{it})-b_t-v_{lt})K_\sigma(\check{z}_t-z_{lt})]\| = O(\sqrt{n}\sigma^m),$$

so that under the bandwidth conditions of the theorem,

$$\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})\gamma_{1t}(v_{jt},z_{it})\\
&= \tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})\tfrac{1}{n}\sum_l (1,q_{lt})' K_\sigma(v_{jt}+\mu_{0t}(z_{it})-b_t-v_{lt})K_\sigma(\check{z}_t-z_{lt}) + o(1/\sqrt{n})\\
&= \tfrac{1}{n}\textstyle\sum_l (1,q_{lt})K_\sigma(\check{z}_t-z_{lt})\tfrac{1}{n}\sum_i b(z_{it})\tfrac{1}{n}\sum_j G(v_{jt},z_{it})' K_\sigma(v_{jt}+\mu_{0t}(z_{it})-b_t-v_{lt}) + o(1/\sqrt{n})\\
&= \tfrac{1}{n}\textstyle\sum_l (1,q_{lt})K_\sigma(\check{z}_t-z_{lt})\tfrac{1}{n}\sum_i b(z_{it})f(v_{lt}-\mu_0(z_{it}))G(v_{lt}-\mu_{0t}(z_{it}),z_{it})' + o(1/\sqrt{n})\\
&= \tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_l f(v_{lt}-\mu_0(z_{it}))G(v_{lt}-\mu_{0t}(z_{it}),z_{it})(1,q_{lt})' K_\sigma(\check{z}_t-z_{lt}) + o(1/\sqrt{n})\\
&= \tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_l f(v_{lt}-\mu_0(z_{it}))f(v_{lt}-\mu_0(z_{it}),z_{it})^{-1}[q(w_{lt})-q(v_{lt}-\mu_0(z_{it}),z_{it})]K_\sigma(\check{z}_t-z_{lt}) + o(1/\sqrt{n})\\
&= \tfrac{1}{n}\textstyle\sum_i b(z_{it})\tfrac{1}{n}\sum_l f(v_{lt}-\mu_0(z_{it}))f(v_{lt}-\mu_0(z_{it}),z_{it})^{-1}[q(w_{lt})-\Phi_{0t}(v_{lt})]K_\sigma(\check{z}_t-z_{lt}) + o(1/\sqrt{n})\\
&= o(1/\sqrt{n}),
\end{aligned}$$

where the third and seventh equalities are derived as above. Therefore,

$$\begin{aligned}\frac{1}{n}\sum_i a(z_{it})(\hat{\mu}_t(z_{it})-\mu_{0t}(z_{it})) &= \frac{1}{n}\sum_i a(z_{it})\frac{f(v_{it})f(z_{it})}{Q^{(2)}(z_{it})f(w_{it})}(y_{it}-\Phi_{0t}(v_{it}+\mu_{0t}(z_{it}))+o_p(1/\sqrt{n}),\\ &= \frac{1}{n}\sum_i a(z_{it})\delta(w_{it})e_{it}+o_p(1/\sqrt{n}).\end{aligned}$$

Let δ_i :=diag$(\delta(w_{it}), t = 2, \cdots, T)'$. Then putting all this together, we have that

$$\sqrt{n}(\hat{\beta}-\beta_0) = (\hat{C}'\hat{\Omega}\hat{C})^{-1}\hat{C}'\hat{\Omega}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i'\tau_i\Delta(\eta_{0i}+\delta_i e_i)+o_p(1).$$

By Assumption 5.4 and the WLLN, $\hat{C}\xrightarrow{p}C$, and by assumption, $\hat{\Omega}\xrightarrow{p}\Omega$. Therefore, by the Lindberg-Levy CLT and the Slutsky theorem we have that $\sqrt{n}(\hat{\beta}-\beta_0)\xrightarrow{d}N(0,V)$, where

$$V = (C'\Omega C)^{-1}C'\Omega\Sigma\Omega C(C'\Omega C)^{-1},$$

and $\Sigma = E[X_i'\tau_i\Delta(\eta_{0i}+\delta_i e_i)(\Delta(\eta_{0i}+\delta_i e_i))'\tau_i X_i]$.

□

A.6 Auxiliary Lemmas

Lemma A.1. *Suppose (i) Assumptions 3.1 and 3.2 hold, (ii) Assumption 3.5 holds, and (iii) Assumptions 5.1 - 5.3 hold. Then* $\|\hat{\Phi}_t-\Phi_{0t}\|_{s,r} = O_p((\ln(n)/n\sigma^{d_w+2r})^{1/2}+\sigma^m)$.

Proof. By noting that $\hat{\Phi}_t(v)$ in equation (4.2) is a standard kernel estimator of a conditional expectation, the result immediately follows by lemma B.3 of Newey [1994b]. □

Lemma A.2. *Suppose (i) Assumptions 3.1 and 3.2 hold, (ii) Assumption 3.5 holds, and (iii) As-*

sumptions 5.1 - 5.3 hold. Then

$$\sup_{z_t}\sup_{\mu_t}\|\hat{Q}^{(r)}(z_t,\mu_t)-Q^{(r)}(z_t,\mu_t)\| = O_p((\ln(n)/n\sigma^{d_w+2r})^{1/2}+\sigma^m+n^{-1/2}).$$

Proof. Because $\mathcal{W}$ and $\Lambda_{s,2}$ are compact, and $q(w)$ is bounded, standard results show

$$\sup_{z_t}\sup_{\mu_t}\|\tilde{Q}^{(r)}(z_t,\mu_t)-Q^{(r)}(z_t,\mu_t)\| = o_p(n^{-1/2}).$$

Now, $\hat{Q}(v_{it},z_t,\mu_t) = ((q(v_{it},z_t)-\Phi_{0t}(v_{it}+\mu_t))+(\hat{q}(v_{it},z_t)-q(v_{it},z_t))-(\hat{\Phi}_t(v_{it}+\mu_t)-\Phi_{0t}(v_{it}+\mu_t)))^2$, so expanding the brackets gives $\hat{Q}(v_{it},z_t,\mu_t) = Q(v_{it},z_t,\mu_t)+\sum_{l=1}^{5} T_l(v_{it},z_t,\mu_t)$, where

$$\begin{aligned}
T_1(v_{it},z_t,\mu_t) &= (\hat{q}(v_{it},z_t)-q(v_{it},z_t))^2,\\
T_2(v_{it},z_t,\mu_t) &= (\hat{\Phi}_t(v_{it}+\mu_t)-\Phi_{0t}(v_{it}+\mu_t))^2,\\
T_3(v_{it},z_t,\mu_t) &= (q(v_{it},z_t)-\Phi_{0t}(v_{it}+\mu_t))(\hat{q}(v_{it},z_t)-q(v_{it},z_t)),\\
T_4(v_{it},z_t,\mu_t) &= -(q(v_{it},z_t)-\Phi_{0t}(v_{it}+\mu_t))(\hat{\Phi}_t(v_{it}+\mu_t)-\Phi_{0t}(v_{it}+\mu_t)),\\
T_5(v_{it},z_t,\mu_t) &= -(\hat{q}(v_{it},z_t)-q(v_{it},z_t))(\hat{\Phi}_t(v_{it}+\mu_t)-\Phi_{0t}(v_{it}+\mu_t)).
\end{aligned}$$

Therefore, using the results in lemma A.1 obtains

$$\begin{aligned}
\sup_{z_t}\sup_{\mu_t}\|\hat{Q}^{(r)}(z_t,\mu_t)-\tilde{Q}^{(r)}(z_t,\mu_t)\| &\le \sum_{l=1}^{5} n^{-1}\sum_i \sup_{z_t}\sup_{\mu_t}\left\|\tau(v_{it},z_t)\frac{\partial^r}{\partial\mu_t^r}T_l(v_{it},z_t,\mu_t)\right\|\\
&= O_p((\ln(n)/n\sigma^{d_w+2r})^{1/2}+\sigma^m). \qquad \text{(A.11)}
\end{aligned}$$

Thus

$$\sup_{z_t}\sup_{\mu_t}\|\hat{Q}^{(r)}(z_t,\mu_t)-Q^{(r)}(z_t,\mu_t)\| = O_p((\ln(n)/n\sigma^{d_w+2r})^{1/2}+\sigma^m+n^{-1/2}).$$

□

References

Chunrong Ai and Li Gan. An alternative root-n consistent estimator for panel data binary choice models. *Journal of Econometrics*, 157:93–100, 2010.

J.G. Altonji and R.L. Matzkin. Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73(4):1053–1102, 2005.

E. Anderson. Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society*, Series B, 32:283–301, 1970.

M. Arellano and S.R. Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies*, 58:277–297, 1991.

M. Arellano and O. Bover. Another look at the instrumental variable estimation of error-comonents models. *Journal of Econometrics*, 68:29–51, 1995.

M. Arellano and R. Carrasco. Binary choice panel data models with predetermined variables. *Journal of Econometrics*, 115:125–157, 2003.

M. Arellano and B. Honoré. *Panel Data Models: More Recent Developments*. Elsevier Science Publishers B.V., 2001.

Jushan Bai. Likelihood approach to small t dynamic panel data models with interactive effects. *MIMEO: Columbia University*, 2009.

R. Blundell and J.L. Powell. Endogeneity in semiparametric binary response models. *Review of Economic Studies*, 71:665–679, 2004.

H. D. Brunk. On the estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics*, 29:437–454, 1958.

G. Chamberlain. Analysis of covariance with qualitative data. *Review of Economics Studies*, XLVII: 225–238, 1980.

G. Chamberlain. *Panel Data*. Elsevier Science Publishers B.V., 1984.

G. Chamberlain. Binary response models for panel data: Identifcation and information. *Econometrica*, 78:159168, 2010.

S. Chen. Root-n consistent estimation of a panel data sample selection model. *MIMEO: The Hong Kong University of Science and Technology*, 1998.

Wayne-Roy Gayle and Soiliou D. Namoro. Estimation of a nonlinear panel data model with semiparametric individual effects. *MIMEO: University of Virginia*, 2012.

W Hardle, M Muller, S Sperlich, and A Werwatz. *Nonparametric and Semiparametric Models*. Springer, 2004.

Bengt Holmstrom. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66:169–182, 1999.

B.E Honoré and E. Kyriazidou. Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874, 2000.

B.E. Honoré and A. Lewbel. Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica*, 70(5):2053–2063, 2002.

B.E. Honore and E. Tamer. Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74(3):661–629, 2006.

J. Horowitz. A smoothed maximum score estimator for the binary response model. *Econometrica*, 60(3):505–531, 1992.

J. Horowitz. Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica*, 69(2):499–513, 2001.

H. Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single index models. *Journal of Econometrics*, 58:71–120, 1993.

A. Lewbel. Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables. *Journal of Econometrics*, 97:145–177, 2000.

O.B. Linton. Efficient estimation of generalized additive nonparametric regression models. *Econometric Theory*, 16:502523, 2000.

O.B. Linton and W. Hardle. Estimating additive regression models with known links. *Biometrica*, 83:529–540, 1996.

E. Mammen, J.S. Marron, B.A. Turlach, and M.P. Wand. a general projection framework for constrained smoothing. *Statistical Science*, 16(3):232–248, 2001.

C. Manski. Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55(2):357–362, 1987.

A. Mood, F Graybill, and D Boes. *Introduction to the Theory of Statistics*. McGraw Hill, 1974.

Whitney K. Newey. Asymptotic variance of semiparametric estimators. *Econometrica*, 62(6):1349–1382, 1994a.

Whitney K. Newey. Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10(2):233–253, 1994b.

Whitney K. Newey and Daniel McFadden. *Large Sample Estimation and Hypothesis Testing*. Elsevier Science Publishers, 1994.

Whitney K. Newey and James L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

G. Rasch. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Denmark Pædagogiske Institut., 1960.

T A. Severini and W H. Wong. Profile likelihood and conditionally parametric models. *The Annals of Statistics*, 20(4):1768–1802, 1992.