# $\sqrt{N}$-Consistent Estimation of a Nonlinear Panel Data Model[†]

Wayne-Roy Gayle[‡]
*University of Virginia*

First Version: February 12, 2007
This Version: January 31, 2009

*Preliminary and Incomplete*

**Abstract**

This paper investigates identification and root-*n* consistent estimation of a class of single index panel data models where: the index function is unspecified; the individual effects may be correlated with all the explanatory variables, and; all the explanatory variables may be predetermined, including lagged dependent variables. The model is extended to allow for a general form of endogeneity and sample selection. For both estimators, we propose kernel based modified backfitting algorithms to estimate the finite and infinite dimensional parameters of interest. The algorithms fully implement all the identification restrictions of the models. We derive consistency and asymptotic normality results for the proposed estimators. Finally, Monte Carlo simulations are performed to study the finite sample properties of both estimators.

**Keywords:** Correlated Random Effects; Single Index; Semiparametric; Panel Data; Predetermined; Lagged Dependent Variables
**JEL classification:** C14, C23, I20, J24

[†]The author is grateful to Xiaohong Chen, Shakeeb Khan, Whitney Newey, Steven Stern and Nese Yildiz for insightful comments and discussions. The author also thanks participants at seminars at the University of Rochester and the University of Virginia. All errors are my own.

[‡]Economics Department, University of Virginia, 2015 Ivy Road, Charlottesville, VA, 22904, phone: (434) 243-4336, e-mail: wg4b@virginia.edu

# 1 Introduction

The objectives of this paper are to: (i) provide new conditions for the identification of a class of nonlinear single-index panel data models with individual specific effects; (ii) derive a new kernel based semiparametric minimum distance estimator for the parameters of the model, and; (iii) develop an algorithm to compute the estimator that fully implements the identification restrictions of the model. The model considered allows for all the explanatory variables to be predetermined, for the individual effect to be correlated with all the observed explanatory variables, and for the index function to be generally unspecified. The estimator is shown to be root-$n$ consistent, and the estimators of the finite dimensional parameters are shown to be asymptotically normal. The basic model under consideration is of the form

$$y_{it} = F_t\left(x_{it}\beta + c_i\right) + u_{it}, \quad i = 1,\cdots,N; \quad t = 1,\cdots,T, \tag{1.1}$$

where $x_{it}$ is a $k$-dimensional vector of explanatory variables which may all be predetermined in that they do not depend on current and future values of $u_{it}$. In particular, this allows for the lagged dependent variable $y_{i,t-1}$ to be included in $x_{it}$. We do not explicitly include the lagged dependent variable because it requires no special treatment in what is to come. The unobserved error term $u_{it}$ is assumed to be mean independent of $c_i$, and current and past values of $x_{it}$. This class of models includes the class on linear single index models given by

$$y_{it} = 1\left\{x_{it}\beta + c_i - \varepsilon_{it} > 0\right\}, \qquad \varepsilon_{it} \sim F_t.$$

The contribution of this paper is therefore discussed generally in the context single-index panel data models and specifically in the context of single-index binary response models.

In the case where $y_{it}$ is continuous, the basic model is extended to jointly allow for sample selection and endogeneity. This extends the results of Das et al. (2003) to the dynamic panel data framework.

The model presented in (1.1) belongs to the class of nonlinear panel data models with correlated individual effects. They belong to a particular subclass that impose restrictions to obtain $\sqrt{N}$-consistent estimators for the finite dimensional parameters. Most of these models can 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 index function. 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 index function. Indeed, Chamberlain (1993) shows that the logit index function is the only one under which $\sqrt{N}$-consistent estimators of $\beta$ can be achieved. One can consider the other end of the spectrum to be the case where no assumption is placed on the index function, but the individual specific effect is defined as a known function of observed regressors.

Models that impose semiparametric restrictions on the index function and the distribution of the individual effects in order to achieve root-$n$ consistency include Chamberlain (1980), Chamberlain (1984), Newey (1994a), Arellano and Carrasco (2003), Chen (1998), and Gayle and Namoro (2005).

Chamberlain (1984) and Newey (1994a) consider a model with only strictly exogenous variables, where the distribution of the error term is normal with unknown mean and variance. Chamberlain (1984) assumes that 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 nonparametric function of the regressors. Chen (1998) extends the model of Newey (1994a) by relaxing parametric assumption on the index function, assuming only that it is a smooth and strictly increasing unknown function. Chen (1998) however assumes that the individual effects are conditionally independent of one continuous regressor given the other regressors. Gayle and Namoro (2005) and Gayle and Viauroux (2007) extend the model of Chen (1998) to include lagged dependent and other predetermined variables.

Moving in a different direction, Arellano and Carrasco (2003) extend the model of Newey (1994a) by allowing all the regressor to be predetermined, as well as allowing all the regressors to be be correlated with the individual effects. The model presented in this paper extends and unifies these two branches by allowing for all the regressors to be predetermined and correlated with the individual effects, as well as assuming only that the index function is a smooth and strictly increasing unknown function. This is achieved by imposing a mild invariance condition on the conditional expectation of the individual effects given one of the regressors. Specifically, for any time period, the marginal relationship between this explanatory variable and the individual effect is invariant over current and past realizations of the random variable, although the marginal relationship may change from period to period. This is a reasonable assumption to make given the nature of the individual effect. One attractive feature of this model is that it is not subject to the time-inconsistency problem faced by the other aforementioned models (excepting that of Arellano and Carrasco (2003)), where the estimator is inconsistent with the arrival of a new wave of observations for each individual.

The paper also presents an estimator for parameters of interest. The estimator is a variation of the generalized additive partial linear model (GAPLM) of Hardle et al. (2004). We also develop a new kernel based modified backfitting algorithm to compute the estimator that fully implements all the identification restrictions of the model. The algorithm makes use of the backfitting estimator proposed in Buja et al. (1989), Mammen et al. (1999), and Mammen et al. (2001). The algorithm is attractive in that the convergence properties are well known, and that it is easy to implement. Furthermore, the algorithm provides a convenient and internally consistent way to impose the identification constraints of the model. We find that the algorithm is well-behaved in that it converges fast and the solution is not sensitive to the choice of starting values.

We perform limited Monte Carlo simulation that illustrate that the estimators perform well in small samples. The simulation results verify $\sqrt{N}$ convergence of the finite dimensional parameters.

## 2 Identification

In this section we prove identification under conditions that allows $c_i$ to be correlated with all the explanatory variables. We first provide conditions for identification of the reduced form version of the basic model 1.1 and then extend these conditions to those sufficient for the identification of

a corresponding reduced form version of the extended model **??**. In both cases, identification in proven in the case where three observations are available for each individual.

## 2.1 Identification of the basic model

We first impose conditions to obtain a reduced form version of the basic model. Specifically, define $x_i^t := (x_{i1}, \cdots, x_{it})$, write $c_i = E[c_i|x_i^t] + v_{it}$ and let $\eta_t(x_i^t) := E[c_i|x_i^t]$. We make the following assumptions.

**Assumption 2.1.**

1. *For $t = 1, \cdots, T$ the index function $F_t$ is strictly increasing on its support.*
2. *For each $t$, the conditional distribution of $v_{it}$ given $x_i^t$ is absolutely continuous and independent of $x_i^t$, with density given by $f_v$.*

Assumption 2.1.1 is standard in the literature of semiparametric identification of single index models. Assumption 2.1.2 is an extension of the assumption Arellano and Carrasco (2003) to an unknown conditional distribution of the individual effects given $x_i^t$.

Given assumption 2.1 we have for all $(i,t)$,

$$\begin{aligned} y_{it} &= F_t(x_{it}\beta + \eta_t(x_i^t) + v_{it}) + u_{it}, \Rightarrow \\ q_{it} := E[y_{it}|x_i^t] &= \Phi_t(x_{it}\beta + \eta_t(x_i^t)), \Rightarrow \\ y_{it} &= \Phi_t(x_{it}\beta + \eta_t(x_i^t)) + e_{it}, \end{aligned} \tag{2.1}$$

where $e_{it} := y_{it} - E[y_{it}|x_i^t]$ with $E[e_{it}|x_i^t] = 0$. Since $F_t$ is strictly increasing, so is $\Phi_t$. Differentiability of $\Phi$ is inherited from $f_v$ and $F_t$. Equation (2.1) provides the starting point of our analysis. This result implies that predictions made from the estimation of equation (2.1) are predictions *after* averaging over the part of the unobserved individual effects that are independent of the observed explanatory variables (ie., the "pure" random effect).

The parameter vector of interest for identification is defined as $\pi := (\Phi, \eta, \beta)$, where $\Phi := (\Phi_1, \cdots, \Phi_T)$, and $\eta := (\eta_1, \cdots, \eta_T)$. We provide the following sufficient conditions for identification of $\pi$.

**Assumption 2.2.**

1. *The random vector $x_{it}$ contains at least one continuous regressor, $x_{it1}$ without loss of generality.*
2. *For $t = 2, \cdots, T$, $\partial\eta_t(x_i^t)/\partial x_{it1} = \partial\eta_t(x_i^t)/\partial x_{i(t-1)1} \neq 0$ with probability one for all $(x_{it1}, x_{i(t-1)1})$.*
3. *For $t = 2, \cdots, T$, $E[(1\ x_{it-1})'(1\ \Delta x_{it})]$ has rank $K+1$.*
4. *$\beta_1 \neq 0$, and $\|\beta\| = 1$.*

5. $E[\eta_t(x_i^t)] = 0, t = 1, \cdots, T$.

Assumption 2.2.1 is essential for the identification of the nonparametric components of the parameter set. This assumption is standard in the literature of semiparametric identification of single index models (see Ichimura (1993) for discussions).

Assumption 2.2.2 is motivated by the Mundlak (1978) specification of fixed effects in linear panel data models, where the individual effect is often specified as a function of time average of the explanatory variables. However, this is not a necessary condition for the assumption to hold. This assumption says that for any time period $t \geq 2$, the marginal relationship between this explanatory variable and the individual effect is invariant over current and (immediate) past realizations of the random variable, although the marginal relationship may change from period to period. Indexing $\eta$ by $t$ is not only theoretically attractive, but has practical significance since the conditioning set $x_i^t$ increases over time. This specification also implies that time-specific effects are not separately identified in this model, because they are absorbed in $\eta_t$.

An example that helps to clarify the restrictions of Assumptions 2.2.2 and 2.1.2 is the following conditional normal distribution. Let the vector $x_{it}$ contain two elements $(x_{it1}, x_{it2})$, with $x_{it1}$ satisfying Assumption 2.2.1 and

$$(c,\ x_{i1}^t)' | x_{i2}^t \sim N\left(\mu(x_{i2}^t),\ \Sigma(t)\right),$$

where $\mu(x_{i2}^t) = (\mu_c(x_{i2}^t), \mu_1(x_{i2}^1), \cdots, \mu_t(x_{i2}^t))'$ and

$$\Sigma(t) = \begin{bmatrix} \sigma_c^2 & \Sigma_{12}(t) \\ \Sigma_{21}(t) & \Sigma_{22}(t) \end{bmatrix}.$$

Then sufficient conditions on $\Sigma(t)$ for Assumptions 2.2.2 and 2.1.2 hold are that: (i) all the elements of $\Sigma_{12}(t)$ are the same, and (ii) $\Sigma_{22}(t)$ is equicorrelated. Under these conditions, the distribution of $c_i$ conditional on $x_i^t$ is given by $c_i = \delta(x_{i2}^t) + \tau_t \sum_{s=1}^t x_{is1} + v_{it}$, where $v_{it} \sim N(0, \sigma_v^2)$ and $v_{it}$ is independent of $x_i^t$. As in Arellano and Carrasco (2003), identification will require the law of iterated expectations (LIE) to hold, so that $E[E[c_i|x_i^t]|x_i^{t-1}] = E[c|x_i^{t-1}]$. In this case however, we require that application of the LIE does not violate Assumptions 2.2.2 and 2.1.2. An additional restriction that ensures this consistency in assumptions is that $x_{i(t+1)2}$ is independent of $x_{i1}^t$ given $x_{i2}^t$, that is, $x_1$ does not Granger cause $x_2$.

This example illustrates that the restrictions imposed on the model by Assumptions 2.2.2 and 2.1.2 are substantial. This example, however, only provides sufficient conditions. Assumption 2.2.2 holds if the pair $(x_{i1}^t)$ enters the conditional density of $c_i$ conditioned on $x_i^t$ only as $\sum_{s=1}^t x_{is1}$, that is, $f_c(c|x_{i1}^t, x_{i2}^t) = f_c(c|\sum_{s=1}^t x_{is1}, x_{i1}^{t-2}, x_{i2}^t)$. Given our mean independence assumption on $u_{it}$, this assumption can be partially tested, since it implies that $E[y_{it}|x_{i1}^t, x_{i2}^t] = E[y_{it}|x_{it1}, \sum_{s=1}^t x_{is1}, x_{i2}^t]$, which is restrictive for $t \geq 3$. The quantities on both sides of the equality can be estimated nonparametrically, which implies that this equality can be tested.

Assumption 2.2.3 is a variation of the usual full rank assumption. Pre-multiplying by $x_{i,t-1}$ instead of $\Delta x_{i,t-1}$ is attractive because it results in only $T = 3$ needed for identification as against

$T = 4$. This assumption implies that the coefficients on observed time-constant random variables are not identified in this model. We also cannot identify the coefficients on random variables, such as age, that change deterministically over time by the same about. Under the assumption that a subset of the explanatory variables are independent of the individual effect, the model may be able to identify these parameters under proper rank conditions. Since the focus of this paper is to allow for general form of correlation between the individual effects and all the explanatory variables, and also to allow for all the explanatory variables to be predetermined, we do not investigate these possibilities.

Assumptions 2.2.4 and 2.2.5 are normalizations used to obtain point identification of the parameters. Assumptions 2.2.4 is only one of a variety of scale normalization of the finite dimensional parameters that can be used (see Gayle and Namoro (2005) for discussion). Assumptions 2.2.6 is analogous to the normalization on the individual effects typically imposed in linear panel data models, that the mean of the individual specific effects is zero.

Denote the true model by $\pi_0 := (\Phi_0, \eta_0, \beta_0)$. Suppose there exists an observationally equivalent model $\pi_1 := (\Phi_1, \eta_1, \beta_1)$ also satisfying assumption 2.2. We show that under assumption 2.2 the parameter set $\theta_0$ is uniquely identified.

**Theorem 2.3.** *Consider the model (2.1) and let Assumption 2.1 and parts 1-3 of Assumption 2.2 hold. Then for constants R and $c > 0$ we have that $\beta_0 = c\beta_1$, $\eta_{t0}(x_i^t) = c\eta_{t1}(x_i^t) + R$, and $\Phi_{t0}(a) = \Phi_{t1}((a-R)/c)$, for $t = 1, \cdots, T$. Furthermore, if parts 4 and 5 of Assumption 2.2 hold, then $c = 1$, and $R = 0$.*

*Proof.* See Appendix A.1 □

Defining $\varphi_t := \Phi_t^{-1}$, equation (2.1) implies that

$$\Delta\varphi_{t0}(q_{it}) = \Delta x_{it}\beta_0 + \Delta\eta_{t0}(x_i^t). \tag{2.2}$$

By the law of iterated expectations, we have that $E[\Delta\eta_{t0}(x_i^t)|x_i^{t-1}] = 0$. This will be the starting point of our estimation strategy.

## 2.2 Identification with sample selection and endogeneity

In this section, we extend the basic model to account for sample selection and endogeneity in a fashion similar to Das et al. (2003). The extended model is given by

$$\begin{aligned} y_{it}^* &= F_t(x_{1it}\beta_1 + z_{2it}\beta_2 + c_i) + u_{it}^*, \quad i = 1, \cdots, N; \quad t = 1, \cdots, T, \\ x_{1it} &= \pi(z_{1it}, z_{2it}) + \varepsilon_{it}, \quad i = 1, \cdots, N; \quad t = 1, \cdots, T, \\ y_{it} &= y_{it}^* d_{it}, \end{aligned} \tag{2.3}$$

where $x_{1it}$, $z_{2it}$, and $z_{1it}$ are $K_1$, $K_2$, and $L_1 (\geq K_1)$ dimensional random vectors. Let $x_{it} = (x_{1it}, z_{2it})$, $z_{it} = (z_{1it}, z_{2it})$, $\beta = (\beta_1', \beta_2')'$, $K = K_1 + K_2$, and $L = L_1 + K_2$, . The vector of functions $\pi$ has

dimension $K_1$. In this extended model, we restrict the dependent variable $y_{it}^*$ to be continuous. We assume that $x_{it}$ and $z_{it}$ are observed for all individuals. As in Das et al. (2003), we assume that the vector $z_{it}$ determines selection. Note that $z_{it}$ may contain lagged values of $d_{it}$ as well as lagged values of $y_{it}$. Define the propensity score $p_{it} = E_t[d_{it}|z_i^t]$. Again, we write $c_i = E[c_i|x_i^t] + v_{it}$ and let $\eta_t(x_i^t) := E[c_i|x_i^t]$. In order to derive the reduced form model, we make the following assumptions:

**Assumption 2.4.**

1. *For $t = 1, \cdots, T$ the index function $F_t$ is strictly increasing on its support.*
2. *For each $t$, the conditional distribution of $v_{it}$ given $x_{it}$, $z_{1it}$ and $d_{it}$ is absolutely continuous and independent of $x_{it}$, $z_{1it}$ and $d_{it}$, with density given by $f_v$.*
3. *$E_t[u_{it}^*|x_i^t, z_i^t, d_{it} = 1] = \lambda_t(\varepsilon_{it}, p_{it})$.*

Assumption 2.4.1 is the same as Assumption 2.1.1. Assumption 2.4.2 strengthens the independence assumption of Assumption 2.1.2. This assumption imposes restrictions on the process by which $d_{it}$ is determined. For example, suppose that $d_{it} = 1\{z_{it}\delta + \mu_i - r_{1it} > 0\}$, and as in the basic model, assume that $\mu_i = E[\mu_i|z_i^t] - r_{2it}$. Let $r_{it} = r_{1it} + r_{2it}$. Then Assumption 2.4.2 implies that $v_{it}$ is independent of $r_{it}$. A necessary condition would therefore be; it is not the case that $c_i = \mu_i$ with probability one. Specifically, all the relationship between $c_i$ and $\mu_i$ is contained in the relationship between $\eta(x_i^t)$ and $E[\mu_i|z_i^t]$. Assumption 2.4.3 is implied by many semiparametric panel limited dependent variables model, including the one introduced here (see Das et al. (2003)). Under these assumptions, we have that

$$\begin{aligned} y_{it}^* &= F_t(x_{it}\beta + \eta_t(x_i^t) + v_{it}) + u_{it}^*, \Rightarrow \\ q_{it} := E[y_{it}|x_i^t, z_i^t, d_{it} = 1] &= \Phi_t(x_{it}\beta + \eta_t(x_i^t)) + \lambda_t(\varepsilon_{it}, p_{it}), \Rightarrow \\ y_{it} &= \Phi_t(x_{it}\beta + \eta_t(x_i^t)) + \lambda_t(\varepsilon_{it}, p_{it}) + e_{it}, \end{aligned} \tag{2.4}$$

where where $e_{it} := y_{it} - E[y_{it}|x_i^t, z_i^t, d_{it} = 1]$ with $E[e_{it}|x_i^t, z_i^t, d_{it} = 1] = 0$. By the same arguments as above, $\Phi_t$ is strictly increasing and differentiable.

The parameters of interest for identification in this selection model are $\pi := (\Phi, \eta, \lambda, \beta)$, where $\lambda := (\lambda_1, \cdots, \lambda_T)$. We impose the following restrictions on the reduced form model (2.4)

**Assumption 2.5.**

1. *The random vector $x_{it}$ contains at least one continuous regressor, $x_{it1}$ without loss of generality.*
2. *For $t = 2, \cdots, T$, $\partial\eta_t(x_i^t)/\partial x_{it1} = \partial\eta_t(x_i^t)/\partial x_{i(t-1)1} \neq 0$ with probability one for all $(x_{it1}, x_{i(t-1)1})$.*
3. *The random vector $z_{it}$ is continuous, the $\lambda_t(\varepsilon_{it}, p_{it})$'s are continuously differentiable, $p_{it}$ and $\pi(z_{it})$ are continuously differentiable in $z_{it}$, and with probability one, rank$[\partial(\pi(z_{it}),\ p(z_i^t))'/\partial z_{1it}] = K_1 + 1$.*
4. *For $t = 2, \cdots, T$, $E[(1\ x_{i(t-1)})'(1\ \Delta x_{it})]$ has rank $K + 1$.*

5. $\beta_1 \neq 0$, *and* $\|\beta\| = 1$.

6. $E[\eta_t(x_i^t)] = 0, t = 1, \cdots, T$.

7. $E[\lambda_t(\varepsilon_{it}, p_{it})] = 0, t = 1, \cdots, T$.

Assumptions 2.5.1 and 2.5.2 are the same Assumptions 2.2.1 and 2.2.2. Assumption 2.5.3 requires that one of the instruments be continuous. This is not necessary for identification, but it makes the exposition clear. The rank condition is a nonparametric generalization of the rank condition for linear models (see Newey et al. (1999)). Parts 4, 5, and 6 of Assumption 2.5 are the same as parts 3, 4, and 5 of Assumption 2.2. Assumption 2.5.7 is a normalization imposed to fix the location of the $\lambda_t$'s. Other normalizations may be used, such as fixing the level of the $\lambda_t$'s for a given $p_{it}$.

Denote the true model by $\pi_0 := (\Phi_0, \eta_0, \lambda_0, \beta_0)$. Suppose there exists an observationally equivalent model $\pi_1 := (\Phi_1, \eta_1, \lambda_1, \beta_1)$ also satisfying Assumption 2.5.

**Theorem 2.6.** *Consider the model (2.4) and let Assumption 2.2 and parts 1-4 of Assumption 2.5 hold. Then for constants* $(R_{1t}, R_{2t})$ *and* $c > 0$ *we have that* $\beta_0 = c\beta_1$, $\Phi_{t0}(a) = \Phi_{t1}((a - R_{2t})/c) - R_{1t}$, $\lambda_{t0}(\varepsilon_{it}, p_{it}) = \lambda_{t1}(\varepsilon_{it}, p_{it}) + R_{1t}$, *and* $\eta_{t0}(x_i^t) = \eta_{t1}(x_i^t) + R_{1t} + R_{2t}$, *for* $t = 1, \cdots, T$. *Furthermore, if parts 5 - 7 of Assumption 2.5 hold, then* $c = 1$, *and* $R_{1t} = R_{2t} = 0$.

*Proof.* See Appendix A.2 □

Inverting and taking first difference of equation 2.4 gives

$$\Delta\varphi_{t0}(q_{it} - \lambda_{t0}(\varepsilon_{it}, p_{it})) = \Delta x_{it}\beta_0 + \Delta\eta_t(x_i^t), \tag{2.5}$$

where again $E[\Delta\eta_{t0}(x_i^t)|x_i^{t-1}] = 0$ by the law of iterated expectations.

# 3 The Estimator

In this section, we propose estimators of the basic and extended models under the restriction that $x_{i1}^t$ enters $\eta_t$ as $\sum_{s=1}^t x_{is1}$. An alternative approach to imposing the restriction would be to impose the constraint directly on $\eta_t$. However, this would require a different formulation since $\eta_t$ is annihilated in the estimator developed in this section. We use the notation $w_{it} := (\sum_{s=1}^t x_{is1}, x_{i2}^t)$ in the basic model and $w_{it} := (\sum_{s=1}^t x_{1is1}, x_{1i2}^t)$ in the extended model.

## 3.1 The basic model

Suppose a sample of $N$ independent realizations $(y_i, x_i\ i = 1, \cdots, N)$ are drawn from the distribution of the $T \times (K+1)$-dimensional random matrix $(y, x)$ with support $\mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y} \subseteq \Re$ and $\mathcal{X} \subseteq$

$\mathfrak{R}^K$. Let $f_x(x)$ be the probability density function of the distribution function defined on $\mathcal{X}$ with respect some dominating measure.

We need to first address implementation of the orthogonality condition $E[\Delta\eta_{t0}(w_{it})|w_{i(t-1)}] = 0$ to obtain an appropriate objective function from equation (2.2). Define $s_{it} = E[\Delta x_{it}|w_{i(t-1)}]$. Then, for $t = 2, \cdots, T$, equation (2.2) implies

$$\begin{aligned} E[\Delta\varphi_t(q_{it})|w_{i(t-1)}] &= s_{it}\beta \Rightarrow \\ \Delta\varphi_t(q_{it}) &= s_{it}\beta + \xi_{it}, \end{aligned} \tag{3.1}$$

where $\xi_{it} := E[\varphi_t(q_{it})|w_{i(t-1)}] - \varphi_t(q_{it})$. Note that since $s_{it}$ is a function of $w_{i(t-1)}$ we have also that $E[\xi_{it}|s_{it}] = 0$. This is simply a conditional expectation derivation of two stage procedure for estimation with instruments.

Because the predicted outcomes $q_{it} := E[y_{it}|x_{it1}, w_{it}]$ has the density of $(x_{it1}, w_{it})$ in the denominator, this density must be bounded away from zero. We therefore impose a fixed trimming condition by defining the compact subset $\chi \subset \mathcal{X}$ where $f_x$ is bounded away from zero on $\chi$. Because $(x_{it1}, w_{it})$ is a transformation of $x_i^t$ with nonzero Jacobian of transformation, the induced density of $(x_{it1}, w_{it})$ is also bounded away from zero on its support (see Mood et al. (1974)). This fixed trimming condition implies that there is a compact subset $\mathcal{K} \subset \mathfrak{R}$ in which all the $q$'s lie. By the same argument, the density of $f_q(q)$ is also bounded away from zero on $\mathcal{K}$. Let $\Lambda^2_{c_2}(\mathcal{K}) := \{f \in \mathcal{C}^2(\mathcal{K}) : \|f\|_{s,2} \leq c_2 < \infty\}$, where $\|\cdot\|_{s,2}$ is the supremum Sobolev norm (see Newey and McFadden (1994)), and $\mathcal{S}_\mathcal{K}$ be a compact subset of $\Lambda^2_{c_2}(\mathcal{K})$, composed of increasing functions.

Assume that $\theta_0 := (\beta_0', \varphi_0)' \in \Theta := \mathcal{B} \times \left(\times_{t=1}^T \mathcal{S}_\mathcal{K}\right)$, where $\mathcal{B} \subseteq \mathfrak{R}^K$ is compact and convex with non-empty interior. Define $s_i := (s_{i2}', \cdots, s_{iT}')$, and the function $\Delta$ as $a := (a_1, \ldots, a_T)' \longmapsto \Delta a := (a_2 - a_1, \ldots, a_T - a_{T-1})'$. Then stacking equation (3.1) in $t$ obtains $\Delta\varphi(q_i) = s_i\beta + \varepsilon_i$. Define $\rho_{it}(\theta) := \rho(q_{it}, s_{it}, \theta) := \Delta\varphi_t(q_{it}) - s_{it}\beta$ and define $\rho_i(\theta)$ accordingly. Finally define the trimming function $\kappa_{it} := 1\{x_{it} \in \chi\}$ and $\kappa_i = \prod_{t=1}^T \kappa_{it}$. Then under the constraints of Assumption 2.2, $\theta_0$ uniquely minimizes

$$Q_0(\theta) := E\left[\kappa\rho(\theta)'\Sigma^{-1}\rho(\theta)\right], \tag{3.2}$$

over $\Theta$, where $\Sigma$ is a (T-1)-dimensional symmetric, positive definite weighting matrix.

In order to derive a feasible empirical analog to 3.2, we first need to estimate $q_{it}$ and $s_{it}$ for all $i$ and $t$. Define the generic function $K_{it}(a) := \sigma^{-d_a}\mathfrak{K}(\sigma^{-1}(a_{it} - a))$, where $d_a$ is the dimension of $a$ and $\mathfrak{K}$ is a generic kernel. Let $\tilde{y}_{it} = (1 \;\; y_{it})$ and $\tilde{x}_{it} = (1 \;\; x_{it})$. Define the kernel estimator of $h_{10}(x_{t1}, w_t) := [f(x_{t1}, w_t) \;\; E[y|x_{t1}, w_t]f(x_{t1}, w_t)]$ as

$$\hat{h}_1(x_{t1}, w_t) := (NT)^{-1}\sum_{j=1}^N\sum_{l=1}^T \tilde{y}_{jl}K_{jl}(x_{t1}, w_t).$$

Define also the kernel estimator of $h_{10}(w_{t-1}) := [f(w_{t-1}) \;\; E[x_{it}|w_{t-1}]f(w_{t-1})]$ as

$$\hat{h}_2(w_{t-1}) := (NT)^{-1} \sum_{j=1}^{N} \sum_{l=1}^{T} \tilde{x}_{jl} K_{jl}(w_{t-1}).$$

With these in hand, let $\hat{q}_t := \hat{h}_{1,2}(x_{t1}, w_{t-1})/\hat{h}_{1,1}(x_{t1}, w_{t-1})$ and $\hat{s}_t := \hat{h}_{2,2}(w_{t-1})/\hat{h}_{2,1}(w_{t-1})$. Define $\hat{\rho}_{it}(\theta) := \rho(\hat{q}_{it}, \hat{s}_{it}, \theta) := \Delta\varphi_t(\hat{q}_{it}) - \hat{s}_{it}\beta$ and define $\hat{\rho}_i(\theta)$ accordingly. Then under the constraints of Assumption 2.2, $\hat{\theta}$ is defined to minimize

$$\hat{Q}(\theta) := \frac{1}{N} \sum_{i=1}^{N} \hat{Q}(x_i; \theta) = \frac{1}{N} \sum_{i=1}^{N} \kappa_i \hat{\rho}_i(\theta)' \hat{\Sigma}^{-1} \hat{\rho}_i(\theta), \tag{3.3}$$

over $\Theta$, where $\hat{\Sigma}$ is a consistent estimator of $\Sigma$.

## 3.2 Estimating the extended model

Suppose now that a sample of $N$ independent realizations $(y_i, x_i, z_i \; i = 1, \cdots, N)$ are drawn from the distribution of the $T \times (K + L + 1)$-dimensional random matrix $(y, x, z)$ with support $\mathcal{Y} \times \mathcal{X} \times \mathcal{Z}$, where $\mathcal{Z} \subseteq \Re^L$. Let $f_{x,z}(x, z)$ be the probability density function of the distribution function defined on $\mathcal{X} \times \mathcal{Z}$ with respect some dominating measure. A similar derivation as in the basic model obtains the following equation that is analogous to equation (3.1)

$$\Delta\varphi_t(q_{it} - \lambda_t(\varepsilon_{it}, p_{it})) = s_{it}\beta + \xi_{it}. \tag{3.4}$$

Let $\theta_0 := (\beta_0', \varphi_0, \lambda_0)' \in \Theta := \mathcal{B} \times \left(\times_{t=1}^{T} \mathcal{S}_{\mathcal{K}}\right) \times \left(\times_{t=1}^{T} \Lambda^2_{c_2}(\mathcal{K})\right)$. Define $\rho_{it}(\theta) := \rho(q_{it}, \varepsilon_{it}, p_{it}, s_{it}, \theta) := \Delta\varphi_t(q_{it} - \lambda_t(\varepsilon_{it}, p_{it})) - s_{it}\beta$ and define $\rho_i(\theta)$ accordingly. Then under the conditions of Assumptions 2.2 and 2.5, $\theta_0$ uniquely minimizes $Q_0(\theta) := E\left[\kappa\rho(\theta)'\Sigma^{-1}\rho(\theta)\right]$. Let $\tilde{x}_{1it} = (1, x_{1it})$ and $\tilde{d}_{it} = (1 \;\; d_{it})$. Define the kernel estimators of $h_{30}(z_t) := [f(z_t) \, E[x_{1it}|z_t] f(z_t)]$ and $h_{40}(z_t) := [f(z_t) \;\; E[d_{it}|z_t] f(z_t)]$ as

$$\hat{h}_3(z_t) := (NT)^{-1} \sum_{j=1}^{N} \sum_{l=1}^{T} \tilde{x}_{1it} K_{jl}(z_t), \text{ and,}$$
$$\hat{h}_4(z_t) := (NT)^{-1} \sum_{j=1}^{N} \sum_{l=1}^{T} \tilde{d}_{jl} K_{jl}(z_t).$$

Let $\hat{\pi}_t(z_t) := \hat{h}_{3,2}/\hat{h}_{3,1}$, $\hat{\varepsilon}_{it} := x_{1it} - \hat{\pi}_t(z_{it})$, and $\hat{p}_t := \hat{h}_{4,2}(z_t)/\hat{h}_{4,1}(z_t)$. Redefine $h_{10}(x_{t1}, w_t)$ to $h_{10}(x_{t1}, w_t, z^t, d_t = 1)$ in the obvious way. Define $\hat{\rho}_{it}$ and $\hat{\rho}_i$ analogous to above. Then $\hat{\theta}$ minimizes

$$\hat{Q}(\theta) := \frac{1}{N} \sum_{i=1}^{N} \hat{Q}(x_i^t, z_i^t; \theta) = \frac{1}{N} \sum_{i=1}^{N} \kappa_i \hat{\rho}_i(\theta)' \hat{\Sigma}^{-1} \hat{\rho}_i(\theta), \tag{3.5}$$

over $\Theta$.

# 4 Computing the estimator

The method presented in this section develops a technique that makes use of the method of alternating projections (Bauschke and Borwein, 1996; Deutsch, 2001) and back fitting algorithm developed in Hastie and Tibshirani (1986), Buja et al. (1989), Mammen et al. (1999) and Mammen et al. (2001). The actual implementation follows closely the two latter references. The advantages of this representation for our purpose are that constraints can be imposed in an internally consistent fashion, and the convergence of properties of these estimators are well understood and obtained under weak conditions. The objective function defined in equation (3.3) is however not the one defined in Mammen et al. (1999) and Mammen et al. (2001). Our first task therefore is to show that this objective function is asymptotically equivalent to those defined in Mammen et al. (2001) under our restrictions on the parameter space and under conditions on the kernels to be used in the estimation of the infinite dimensional parameters. Again, we shall use the generic notation $\omega_{it}(a) := \sigma^{-1}\mathfrak{K}(\sigma^{-1}(a_{it}-a))$ where $\mathfrak{K}$ denotes a generic kernel. It is important to note that this is only to conserve on notation, and does not mean we are using the same kernel to estimate $\hat{q}$, $\hat{p}$, or $\hat{s}$. Furthermore, the kernel notation here makes explicit that we will be working with single dimensional kernels. We make the following assumptions on the kernels used to estimate the $\varphi$'s.

**Assumption 4.1.** *(i) For $a \geq 2$, $\mathfrak{K}(a)$ is differentiable of order $a$, the $a$-th derivatives bounded, $\mathfrak{K}(a)$ is zero outside a bounded set, $\mathfrak{K}(a) \geq 0$, $\int \mathfrak{K}(a)da = 1$, $\int a\mathfrak{K}(a)da = 0$, $\int a^2\mathfrak{K}(a)da < \infty$, and $\int |\mathfrak{K}(a)|^2 da < \infty$. (ii) $\sigma \to 0$ and $N\sigma^2 \to \infty$ as $N \to \infty$.*

These conditions are used to show that the objective function defined in equation (3.3) is equivalent to the those defined in Mammen et al. (2001). Let $m(z_{1i}, z_{2i})$ be a bounded continuous function. The following lemma is essential for the computation of our estimator.

**Lemma 4.2.** *Let $m(z_{1i}, z_{2i})$ twice differentiable in its arguments with bounded second derivatives. Let Assumption 4.1 hold. Then we have that*

$$\frac{1}{N}\sum_{i=1}^{N} m(z_{1i}, z_{2i}) = \int \frac{1}{N}\sum_{i=1}^{N} m(z_1, z_{2i})\omega_i(z_1)dz_1 + o_p(1) \tag{4.1}$$

$$= \int \frac{1}{N}\sum_{i=1}^{N} m(z_{1i}, z_2)\omega_i(z_2)dz_2 + o_p(1) \tag{4.2}$$

$$= \int \frac{1}{N}\sum_{i=1}^{N} m(z_1, z_2)\omega_i(z_1)\omega_i(z_2)dz_1dz_2 + o_p(1) \tag{4.3}$$

*Proof.* See Appendix A.3 □

## 4.1 Computing the estimator for basic model

In order to use Lemma 4.2 in defining the algorithm, require the following assumptions and Lemmas. The following assumption places restrictions on the kernel used to estimate $s_i$. These restric-

tions, while stronger than needed for this section, will be required when evaluating the asymptotic distribution of the finite dimensional parameter estimators.

**Assumption 4.3.** *1. $K(a)$ is differentiable of order $d \geq 2$, the derivatives $d$ are bounded, $K(a)$ is zero outside a bounded set, $\int K(a)da = 1$, there is a positive integer $m$ such that for all $j < m$, $\int K(a)[\otimes_{\ell=1}^{j}]da = 0$. 2. There is a version of $h_{20}(w)$ that is continuously differentiable to order $d$ with bounded derivatives on an open set containing $\mathcal{W}$. 3. There is $p \geq 4$ such that $E[\|\tilde{x}\|^p] < \infty$ and $E[\|\tilde{x}\|^p|w]f_0(w)$ is bounded. 4. The bandwidth $\sigma = \sigma(N)$ satisfies $N^{1-(2/p)}\sigma^L/\ln N \longrightarrow \infty$, $\sqrt{N}\sigma^{2m} \longrightarrow 0$, and $\sqrt{N}\ln N/(N\sigma^{L+2d}) \longrightarrow 0$*

The following assumption on the weighting matrix is easily verified in practice.

**Assumption 4.4.** *The inverse of the weighting matrix, $\hat{\Sigma}^{-1}$, exists and is finite.*

We are now in a position to define the estimator for $\varphi_t$, $t = 1, \cdots, T$ for a given value of $\beta$. For each $i$, let $\varphi(\hat{q}_{i,-t}, q_t)$ be $\varphi(\hat{q}_i)$ with $\hat{q}_{it}$ replaced with $q_t$. Given the assumptions on the parameter space, the second derivative of $\hat{Q}(\theta)$ with respect to $q_t$ is uniformly bounded. Then Lemma 4.2 implies that the objective function (3.3) can be written as

$$\hat{Q}(\theta) := \int \frac{1}{N}\sum_{i=1}^{N} \kappa_i(\Delta\varphi(\hat{q}_{i,-t}, q_t) - \hat{s}_i\beta)'\hat{\Sigma}^{-1}(\Delta\varphi(\hat{q}_{i,-t}, q_t) - \hat{s}_i\beta)\omega_{it}(q_t)dq_t + o_p(1), \tag{4.4}$$

where $\omega_{it}(q_t) = \sigma^{-1}\mathfrak{K}(\sigma^{-1}(\hat{q}_{it} - q_t))$. For fixed $q_t$ the unconstrained estimator $\tilde{\varphi}_t(q_t)$ is defined as follows (see Mammen et al. (2001)):

$$\tilde{\varphi}_t(q_t) = \arg\min_{\varphi_t \in \Lambda^2_{\bar{c}_2}(\mathcal{K})} \frac{1}{N}\sum_{i=1}^{N} \kappa_i(\Delta\varphi(\hat{q}_{i,-t}, q_t) - \hat{s}_i\beta)'\hat{\Sigma}^{-1}(\Delta\varphi(\hat{q}_{i,-t}, q_t) - \hat{s}_i\beta)\omega_{it}(q_t) + o_p(1) \tag{4.5}$$

The solution to (4.4) is characterized by the first-order condition

$$\frac{1}{N}\sum_{i=1}^{N} \kappa_i(\Delta_{\cdot,t})'\hat{\Sigma}^{-1}(\Delta\tilde{\varphi}(\hat{q}_{i,-t}, q_t) - \hat{s}_i\beta)\omega_{it}(q_t) = 0, \tag{4.6}$$

where $\Delta_{\cdot,t}$ denotes the $t$-th column of $\Delta$. Let $a := \Delta'\hat{\Sigma}^{-1}$ and $b := \Delta'\hat{\Sigma}^{-1}\Delta$. Then straightforward calculation show that equation (4.6) implies

$$\tilde{\varphi}_t(q_t) = \sum_{l=2}^{T} \frac{a_{t,l-1}}{b_{tt}} \frac{1}{N}\sum_{i=1}^{N} \kappa_i \frac{\hat{s}_{il}\beta\omega_{it}(q_t)}{\hat{f}_t(q_t)} - \sum_{l\neq t} \frac{b_{tl}}{b_{tt}} \frac{1}{N}\sum_{i=1}^{N} \kappa_i \frac{\tilde{\varphi}_l(\hat{q}_{il})\omega_{it}(q_t)}{\hat{f}_t(q_t)}, \tag{4.7}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \left( \sum_{l=2}^{T} \frac{a_{t,l-1}}{b_{tt}} \hat{s}_{il}\beta - \sum_{l\neq t} \frac{b_{tl}}{b_{tt}} \tilde{\varphi}_l(\hat{q}_{il}) \right) \frac{\kappa_i\omega_{it}(q_t)}{\hat{f}_t(q_t)}, \tag{4.8}$$

where $f_t(q_t) := N_{-1}\sum_{i=1}^{N} \kappa_i\omega_{it}(q_t)$. Note the similarity between the estimator defined in equation (4.7) and equation (12) of Mammen et al. (1999). While the equation (12) of Mammen et al. (1999) requires one- and two-dimensional marginal density estimates, equation (4.8) requires only one-dimensional density estimates. The advantage of this equivalence is that while the estimator defined

here is easier to compute, we can still appeal to the convergence theorems of Mammen et al. (1999) for convergence of the iteration (to be defined). To impose the monotonicity constraint, we further project $\tilde{\varphi}_t$ onto the set $\mathcal{S}_{\mathcal{K}}$. The results of Brunk (1958), and Mammen et al. (2001) imply the following solution to this projection

$$\hat{\varphi}_t(q_t) = \inf_{v \geq q_t} \sup_{u \leq q_t} \frac{\int_{l=u}^{v} \tilde{\varphi}_t(l)\, \hat{f}_t(l) \mathrm{d}l}{\int_{l=u}^{v} \hat{f}_t(l) \mathrm{d}l} \tag{4.9}$$

Given estimates of the $\varphi_t$'s, the estimator for $\beta$ is obtained by solving the objection function (3.3) for $\beta$ to get

$$\tilde{\beta} = \left[ \frac{1}{N} \sum_{i=1}^{N} \kappa_i \hat{s}_i' \hat{\Sigma}^{-1} \hat{s}_i \right]^{-1} \frac{1}{N} \sum_{i=1}^{N} \kappa_i \hat{s}_i' \hat{\Sigma}^{-1} \Delta \hat{\varphi}(\hat{q}_i). \tag{4.10}$$

For arbitrary initial choices $\beta^{[0]}$ and $\varphi_t^{[0]}$, $t = 1, \cdots, T$, the backfitting algorithm therefore works as follows.

*Backfitting Estimator for the Basic Model (BF1)*

Do for $r \geq 1$ until convergence in $\beta$ is reached $\{$
Initialize with $\varphi_t^{[r-1]}(q_t)$ and do for $j \geq 1$ until convergence in $\varphi_t, t = 1, \cdots, T$ is reached $\{$
Do for $t = 1$ to $T\{$

$$\begin{aligned}
\tilde{\varphi}_t^{[j]}(q_t) &= \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{l=2}^{T} \frac{a_{t,l-1}}{b_{tt}} \hat{s}_{il} \beta^{[r-1]} - \sum_{l<t} \frac{b_{tl}}{b_{tt}} \varphi_l^{[j]}(\hat{q}_{il}) - \sum_{l>t} \frac{b_{tl}}{b_{tt}} \varphi_l^{[j-1]}(\hat{q}_{il}) \right) \frac{\kappa_i \omega_{it}(q_t)}{\hat{f}_t(q_t)}, \\
\varphi_t^{[j]}(q_t) &= \inf_{v \geq q_t} \sup_{u \leq q_t} \frac{\int_{l=u}^{v} \tilde{\varphi}_t^{[j]}(l)\, \hat{f}_t(l) \mathrm{d}l}{\int_{l=u}^{v} \hat{f}_t(l) \mathrm{d}l} \\
\varphi_t^{[j]}(q_t) &= \varphi_t^{[j]}(q_t) - \frac{1}{N} \sum_{i=1}^{N} \kappa_i \varphi_t^{[j]}(\hat{q}_{it})
\end{aligned}$$

$\}$
Denote the fixed point values by $\varphi_t^{[r]}, t = 1, \cdots, T$.
$\}$
update $\beta$ by

$$\beta^{[r]} = \left[ \frac{1}{N} \sum_{i=1}^{N} \kappa_i \hat{s}_i' \hat{\Sigma}^{-1} \hat{s}_i \right]^{-1} \frac{1}{N} \sum_{i=1}^{N} \kappa_i \hat{s}_i' \hat{\Sigma}^{-1} \Delta \hat{\varphi}^{[r]}(\hat{q}_i)$$

$\}$
The fixed point solution is denoted as $(\hat{\beta}, \hat{\varphi})$. Notice that we explicitly write the algorithm as a nested loop. The reason is that the projection defined in the inner loop for fixed $\beta$ is well understood to

converge (see Mammen et al. (1999)). The outer loop defines a sequence of alternating projections between two closed convex sets, with one set is finite dimensional. The convergence of this sequence is also well understood (see Cheney and Goldstein (1959) and Gayle and Namoro (2005)).

**Proposition 4.5.** *Under assumption 4.1, the algorithm the algorithm (BF1) converges with probability one.*

## 4.2 Computing the estimator for the extended model

We begin by defining the computation of the $\varphi$'s for fixed $\beta$ and $\lambda_t,\ t=1,\cdots,T$. Let $\hat{\mu}_{it} := \hat{q}_{it} - \lambda(\hat{\varepsilon}_{it},\hat{p}_{it})$ and $\varphi(\hat{\mu}_{i,-t},\mu_t)$ be $\varphi(\hat{\mu}_i)$ with $\hat{\mu}_{it}$ replaced with $\mu_t = q_t - \lambda(\varepsilon_t,p_t)$. Then Lemma 4.2 implies that the objective function (3.5) can be written as equation (4.4), which in turn implies that the unconstrained solution for fixed $\mu_t$ is given by

$$\tilde{\varphi}_t(\mu_t) \quad = \quad \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{l=2}^{T}\frac{a_{t,l-1}}{b_{tt}}\hat{s}_{il}\beta - \sum_{l\neq t}\frac{b_{tl}}{b_{tt}}\tilde{\varphi}_l(\hat{\mu}_{il})\right)\frac{\kappa_i\omega_{it}(\mu_t)}{\hat{f}_t(\mu_t)}. \tag{4.11}$$

The constrained solution is given in equation (4.9).

To compute $\lambda_t,\ t=1,\cdots,T$ at a point, for fixed $\beta$ and $\varphi_t, t=1,\cdots,T$, note that Lemma 4.2 implies that the objective function (3.5) can be written as

$$\begin{aligned}\hat{Q}(\theta) := \int \frac{1}{N}\sum_{i=1}^{N}\kappa_i(\Delta\varphi(\hat{q}_i - \lambda(\hat{\varepsilon}_{i,-t},\varepsilon_t,\hat{p}_{i,-t},p_t)) - \hat{s}_i\beta)'\hat{\Sigma}^{-1}\times \\ (\Delta\varphi(\hat{q}_i - \lambda(\hat{\varepsilon}_{i,-t},\varepsilon_t,\hat{p}_{i,-t},p_t)) - \hat{s}_i\beta)\omega_{it}(\varepsilon_t)\omega_{it}(p_t)d\varepsilon_t dp_t + o_p(1),\end{aligned}$$

where for each $i$, $\lambda(\varepsilon_{i,-t},\varepsilon_t,\hat{p}_{i,-t},p_t)$ is $\lambda(\hat{\varepsilon}_i,\hat{p}_i)$ with $\varepsilon_{it}$ and $p_{it}$ replaced with $\varepsilon_t$ and $p_t$. The first derivative $h(\lambda(\varepsilon_t,p_t))$ and second derivative $H(\lambda(\varepsilon_t,p_t))$ of this objective function with respect to $\lambda_t$ is given by

$$\begin{aligned}h(\lambda_t(\varepsilon_t,p_t)) =& -\frac{1}{N}\sum_{i=1}^{N}\varphi_t'(\hat{q}_{it} - \lambda_t(\varepsilon_t,p_t))(\Delta_{\cdot,t})'\hat{\Sigma}^{-1}\times \\ & (\Delta\varphi(\hat{q}_i - \lambda(\hat{\varepsilon}_{i,-t},\varepsilon_t,\hat{p}_{i,-t},p_t)) - \hat{s}_i\beta)\kappa_i\omega_{it}(\varepsilon_t)\omega_{it}(p_t), \\ =& -\frac{1}{N}\sum_{i=1}^{N}\varphi_t'(\hat{q}_{it} - \lambda_t(\varepsilon_t,p_t))\times \\ & \left\{b_{tt}\varphi_t(\hat{q}_{it} - \lambda(\varepsilon_t,p_t)) + \sum_{l\neq t}b_{tl}\varphi_l(\hat{q}_{il} - \lambda_l(\hat{\varepsilon}_{il},\hat{p}_{il})) - \sum_{l=1}^{T}a_{tl}\hat{s}_{il}\beta\right\}\kappa_i\omega_{it}(\varepsilon_t)\omega_{it}(p_t), \\ H(\lambda_t(\varepsilon_t,p_t)) =& \frac{1}{N}\sum_{i=1}^{N}b_{tt}\left(\varphi_t'(\hat{q}_{it} - \lambda_t(\varepsilon_t,p_t))\right)^2\kappa_i\omega_{it}(\varepsilon_t)\omega_{it}(p_t).\end{aligned}$$

This implies that for given $\lambda_t, t=1,\cdots,T$, one can update $\lambda_t$ by the following single Newton-

Raphson step:

$$\lambda_t^{\text{new}}(\varepsilon_t, p_t) = \lambda_t(\varepsilon_t, p_t) - \frac{h(\lambda_t(\varepsilon_t, p_t))}{H(\lambda_t(\varepsilon_t, p_t))}. \tag{4.12}$$

We are now in a position to state the algorithm for computing the selection model. For arbitrary initial choices $\beta^{[0]}$ and $(\varphi_t^{[0]}, \lambda_t^{[0]})$, $t = 1, \cdots, T$, the backfitting algorithm works as follows.

*Backfitting Estimator for the Extended Model (BF2)*

Do for $r \geq 1$ until convergence in $\beta$ is reached $\{$
Update $\lambda_t$, $t = 1, \cdots, T$ by$\{$

$$\lambda_t^{[r]}(\varepsilon_t, p_t) = \lambda_t^{[r-1]}(\varepsilon_t, p_t) - \frac{h^{[r-1]}(\lambda_t^{[r-1]}(\varepsilon_t, p_t))}{H^{[r-1]}(\lambda_t^{[r-1]}(\varepsilon_t, p_t))}$$

Set $\hat{\mu}_{it}^{[r]} = \hat{q}_{it} - \lambda_t^{[r]}(\hat{\varepsilon}_{it}, \hat{p}_{it})$.
$\}$
Initialize with $\varphi_t^{[r-1]}(\hat{\mu}_{it}^{[r]})$ and do for $j \geq 1$ until convergence in $\varphi_t, t = 1, \cdots, T$ is reached $\{$
Do for $t = 1$ to $T$

$$\begin{aligned}
\tilde{\varphi}_t^{[j]}(\mu_t) &= \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{l=2}^{T}\frac{a_{t,l-1}}{b_{tt}}\hat{s}_{il}\beta^{[r-1]} - \sum_{l<t}\frac{b_{tl}}{b_{tt}}\varphi_l^{[j]}(\hat{\mu}_{il}^{[r]}) - \sum_{l>t}\frac{b_{tl}}{b_{tt}}\varphi_l^{[j-1]}(\hat{\mu}_{il}^{[r]})\right)\frac{\kappa_i\omega_{it}(\mu_t)}{\hat{f}_t(\mu_t)}, \\
\varphi_t^{[j]}(\mu_t) &= \inf_{v\geq\mu_t}\sup_{u\leq\mu_t}\frac{\int_{l=u}^{v}\tilde{\varphi}_t^{[j]}(l)\,\hat{f}_t(l)\mathrm{d}l}{\int_{l=u}^{v}\hat{f}_t(l)\mathrm{d}l} \\
\varphi_t^{[j]}(\mu_t) &= \varphi_t^{[j]}(\mu_t) - \frac{1}{N}\sum_{i=1}^{N}\kappa_i\varphi_t^{[j]}(\hat{\mu}_{it}) \qquad (4.13)
\end{aligned}$$

Denote the fixed point values by $\varphi_t^{[r]}, t = 1, \cdots, T$.
$\}$
update $\beta$ by

$$\beta^{[r]} = \left[\frac{1}{N}\sum_{i=1}^{N}\kappa_i\hat{s}_i'\hat{\Sigma}^{-1}\hat{s}_i\right]^{-1}\frac{1}{N}\sum_{i=1}^{N}\kappa_i\hat{s}_i'\hat{\Sigma}^{-1}\Delta\hat{\varphi}^{[r]}(\hat{\mu}_i^{[r]})$$

$\}$.
The fixed point solution is denoted as $(\hat{\beta}, \hat{\varphi}, \hat{\lambda})$. The algorithm BF2 is exactly BF1 with an added step to update the $\lambda_t$'s. It seems intuitive, therefore, that this algorithm also converges. Indeed based on our experience, the algorithm does converges. However, a direct proof of convergence is beyond the scope of this paper, but does belong to our research agenda.

There is still the unresolved issue of obtaining the estimator $\hat{\Sigma}$ of the weighting matrix $\Sigma$. The

solution is to perform a two-step estimation method by first substituting the $(T-1)$-dimensional identity matrix $I_{T-1}$ for $\hat{\Sigma}$. This provides consistent estimates of the parameters of interest. As will become clear in the following section, a consistent estimator $\hat{\Sigma}$ of $\Sigma$ can then be constructed using these initial consistent estimates. The second step is to substitute this estimate into the objective function and repeat the algorithm. As discussed in Newey and McFadden (1994), the second stage algorithm needs to be executed for a single iteration ($r = j = 1$) if the first stage estimates are used to initialize the second stage algorithm.

# 5 Asymptotic Properties

In this section, we derive the asymptotic properties of the sample selection model. This asymptotic properties of the basic model are presented as a corollaries to the corresponding theorems of the sample selection model.

We assume that $h_{10}$ and $h_{20}$, as well as the kernels used to construct $\hat{h}_1$ $\hat{h}_2$ also satisfy the conditions of Assumption 4.3. We also make the following assumptions on the weighting matrix $\Sigma$ and its estimator $\hat{\Sigma}$.

**Assumption 5.1.** $\|\Sigma\| < \infty$, *and* $\hat{\Sigma} \xrightarrow{p} \Sigma$.

After deriving the asymptotic properties of the estimator under this condition, we will define the efficient choice of $\Sigma$ and a corresponding estimator $\hat{\Sigma}$ that satisfy Assumption 4.4.

## 5.1 Consistency

Define the distance $d$ on $\Theta$ as follows: $d[\theta_1, \theta_2] := \|\beta_1 - \beta_2\|_K + \sum_{t=1}^{T}(\|\varphi_{1t} - \varphi_{2t}\|_{s,2} + \|\lambda_{1t} - \lambda_{2t}\|_{s,2})$, where $\|\cdot\|_K$ is the Euclidean norm on $\Re^K$ and $\|\cdot\|_{s,2}$ is the supremum Sobolev norm of smoothness 2. The supremum for $\varphi$ is with respect to $\mathcal{S}_{\mathcal{K}}$, and the supremum for $\lambda$ is taken over $\Lambda^2_{c_2}(\mathcal{K})$.

**Theorem 5.2.** *Consider the estimator defined as minimizing the objective function defined in equation (3.5). Let the assumptions 2.1, 2.4, 2.5, 4.3, and 4.4 be satisfied. Then* $\hat{\beta} \xrightarrow{p} \beta_0$, *and for* $t = 1, \cdots, T$, $\|\hat{\varphi}_t - \varphi_{t0}\|_{s,2} \xrightarrow{p} 0$, *and* $\|\hat{\lambda}_t - \lambda_{t0}\|_{s,2} \xrightarrow{p} 0$.

*Proof.* See Appendix A.4. □

Consistency of the basic model is therefore presented in the following corollary.

**Corollary 5.3.** *Consider the estimator defined as minimizing the objective function defined in equation (3.3). Let the assumptions 2.1, 2.2, 4.3, and 4.4 hold. Then* $\hat{\beta} \xrightarrow{p} \beta_0$, *and for* $t = 1, \cdots, T$, $\|\hat{\varphi}_t - \varphi_{t0}\|_{s,2} \xrightarrow{p} 0$.

## 5.2 Asymptotic normality

The backfitting algorithm in BF2 work by iteratively solving for $\lambda$ and $\varphi$ given a fixed $\beta$, and then solving for $\beta$. In a neighborhood of the minimum, it is then clear that the algorithm concentrates out $\lambda$ and $\varphi$ by solving for it as a function of $\beta$, and then solving for $\beta$. We shall use the notation $\varphi(q - \lambda(\varepsilon, p; \beta); \beta)$ to make this explicit. Let $B_i(h) := \nabla_\beta \varphi(q_i - \lambda(\varepsilon_i, p_i; \beta); \beta) - s_i$. Also, let $\eta_i := \rho_i + \sum_{k=1}^{K}(\Delta x_{ik} - s_{ik}) + R_{1i}(y_i - q_i) - R_{2i}(d_i - p_i) + R_{3i}\varepsilon_i$ where $R_{1i} := \Delta\varphi_i'(q_i - \lambda(p_i))$, $R_{2i} := R_{1i}\lambda_1(\varepsilon_i, p_i)$, and $R_{3i} := R_{1i}\lambda_2(\varepsilon_i, p_i)$ ($\lambda_1$ and $\lambda_2$ denote the partial derivatives of $\lambda$ with respect to its first and second arguments). Finally let $\Omega := E[\eta_{i0}\eta_{i0}']$, where $\eta_{i0}$ is $\eta_i$ evaluated at the true parameter values. Then we have the following theorem.

**Theorem 5.4.** *Consider the estimator defined as minimizing the objective function defined in equation (3.5). Let the assumptions 2.1, 2.4, 2.5, 4.3, and 4.4 be satisfied. Then* $\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$, *where*

$$V := E\left[\kappa_i B_{i0}(h_{i0})'\Sigma^{-1}B_{i0}(h_{i0})\right] E\left[\kappa_i B_{i0}(h_{i0})'\Sigma^{-1}\Omega\Sigma^{-1}B_{i0}(h_{i0})\right]^{-1} E\left[\kappa_i B_{i0}(h_{i0})'\Sigma^{-1}B_{i0}(h_{i0})\right]$$

*Proof.* See Appendix A.5 □

For the basic model we redefine $B_i(h) := \nabla_\beta \varphi(q_i; \beta)$ and $\varepsilon_i := \rho_i + \sum_{k=1}^{K}(\Delta x_{ik} - s_{ik}) + R_{1i}(y_i - q_i)$.

**Corollary 5.5.** *Consider the estimator defined as minimizing the objective function defined in equation (3.3). Let the assumptions 2.1, 2.2, 4.3, and 4.4 hold. Then* $\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$, *where*

$$V := E\left[\kappa_i B_{i0}(h_{i0})'\Sigma^{-1}B_{i0}(h_{i0})\right] E\left[\kappa_i B_{i0}(h_{i0})'\Sigma^{-1}\Omega\Sigma^{-1}B_{i0}(h_{i0})\right] E\left[\kappa_i B_{i0}(h_{i0})'\Sigma^{-1}B_{i0}(h_{i0})\right]$$

From the preceding, it is clear that the efficient choice of the weighting matrix is given by $\Sigma = \Omega$, reducing the variance matrix to $V = E\left[\kappa_i B_{i0}(h_{i0})'\Sigma^{-1}B_{i0}(h_{i0})\right]$, with the obvious choice for the estimator given by $\hat{\Sigma} = \sum_{i=1}^{N} \kappa_i \hat{\varepsilon}_i \hat{\varepsilon}_i'/N$. The following lemma establishes that these choices satisfy the conditions of Assumption 5.1.

**Lemma 5.6.** *Consider the estimator defined as minimizing the objective function defined in equation (3.5). Let the assumptions 2.1, 2.4, 2.5, and 4.3 be satisfied. The* $\|\Sigma\| < \infty$ *and* $\hat{\Sigma} \xrightarrow{p} \Sigma$.

## 5.3 Consistent asymptotic variance estimation

In order to estimate the asymptotic variance, one also has to obtain an estimate of $B_{i0}(h_{i0})$. We propose the estimator $\hat{B}_i(\hat{h}) := \nabla_\beta \hat{\varphi}(\hat{q}_i - \hat{\lambda}(\hat{p}_i; \hat{\beta}); \hat{\beta})$, where the derivative is taken numerically by perturbing each element of $\hat{\beta}$ and performing a single iteration of the inner loop of the algorithm BF2 (setting $r = j = 1$). The estimator of the asymptotic variance is then given by

$$\hat{V} := \frac{1}{N}\sum_{i=1}^{N} \kappa_i \hat{B}_i(\hat{h})'\hat{\Sigma}^{-1}\hat{B}_i(\hat{h}).$$

**Theorem 5.7.** *Consider the estimator defined as minimizing the objective function defined in equation (3.5). Let the assumptions 2.1, 2.4, 2.5, and 4.3 be satisfied. The* $\hat{V} \xrightarrow{p} V$.

# 6 Monte Carlo study

In this section, we present two limited Monte Carlo exercises to illustrate the performance of the estimators. The first simulation investigates the performance of the estimator for the basic model, where all the explanatory variables are predetermined and correlated with the individual-specific effect. The second investigates the performance of the estimator for the sample selection model, which augments the basic environment.

## 6.1 The basic model

For the first exercise, we consider the following data generating process for $i = 1, \cdots, N$; $t = 1, \cdots, 4$:

$$y_{it} = (\beta_1 x_{it1} + \beta_2 x_{it2} + c_i)^3 + u_{it}.$$

The explanatory variables are generated as follows:

$$x_{itj} = \delta_1 v_{itj} + \delta_2 u_{it-1},\ j = 1, 2,$$

with $v_{itj},\ j = 1, 2$ independently distributed $N(0,1)$ and $u_{it}$ independently distributed $N(0,1)$. We simulate the model for two choices of $\delta$; $(0.9,\ 0.1)$ and $(0.6,\ 0.4)$. The individual-specific effect is generated by $c_i = 0.5(v_{i11} + \cdots + v_{i41})/4 + 0.5(v_{i12} + \cdots + v_{i42})/4 + v_{i4}$, with $v_{i4}$ independently distributed $N(0,1)$. We make the following initializations $u_{i0} = 0$ and set $(\beta_{10},\ \beta_{20}) = (0.6,\ 0.8)$. We perform 100 Monte Carlo replications of the model with three sample sizes N: 100, 200, and 400. The mean bias and root mean square error (RMSE) are calculated for each sample size. 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 EKMD. The results are presented in Table 1.

The results in Table 1 indicate that the estimators perform well in recovering the finite dimensional parameters, even in very small samples. The results also verify $\sqrt{N}$ convergence of the estimator.

### 6.2 The sample selection model

In the sample selection model, we extend the environment of section 6.1 as follows:

$$\begin{aligned} y_{it}^{*} &= (\beta_1 x_{it1} + \beta_2 x_{it2} + c_i)^3 + u_{it}^{*}, \\ y_{it} &= d_{it} y_{it}^{*}, \\ d_{it} &= 1\{-0.1 + 0.2x_{it1} + 0.2x_{it2} + 0.6x_{it3} - \varepsilon_{it} > 0\}. \end{aligned}$$

As we are not interested in recovering the parameters of the selection equation, we choose a relatively straightforward selection rule. Note however, that we can easily accommodate a complicated selection rule, as general as that of the basic model above. $x_{it3}$ is independently distributed $N(0,1)$ and $\varepsilon_{it}$ independently distributed $N(0,0.5)$. These choices result in censoring of approximately 30 percent of the original sample. The other variables are generated as in section 6.1. As above we set $(\beta_{10},\ \beta_{20}) = (0.6,\ 0.8)$ and perform 100 Monte Carlo replications of the model with three sample sizes N: 100, 200, and 400. The results are presented in Table 2.

The results indicate that the estimator performs quite well in relative small samples and with significant censoring. Again the results verify $\sqrt{N}$ convergence of the finite dimensional parameters.

## 7 Conclusion

This paper investigate identification and estimation of a class of single-index panel data models with: (i) All the explanatory variables may be predetermined. (ii) The index function is unspecified. (iii) The individual effects may be correlated with all the explanatory variables. The model is extended to allow for a general form of sample selection. We develop kernel based minimum distance estimators of the finite and infinite dimensional parameters of both models. These estimators extend the minimum distance estimator of Mammen et al. (1999) and Mammen et al. (2001). We develop a new algorithm to compute the estimators that fully implements the restrictions of the models. The algorithm makes use of the backfitting estimator proposed in Buja et al. (1989), Mammen et al. (1999), and Mammen et al. (2001). The algorithm extends that of Gayle and Namoro (2005) for semiparametric panel data models. We find that the algorithm converges fast and is robust to the choice of the starting values. We show that the estimators for the finite dimensional parameters are $\sqrt{N}$-consistent and asymptotically normal. We also show that the estimator of the infinite dimensional parameters are consistent. The paper illustrates the finite sample properties of the estimators by way of small Monte Carlo studies. We find that the estimator perform well in very small samples.

# A LEMMA AND THEOREMS

## A.1 Proof of Theorem 2.3

*Proof.* Since the random variable $q_{it} := E[y_{it}|x_i^t]$ can be consistency estimated, it is known in large samples. This, equation (2.1) gives for each $t$ (dropping the $i$ subscript):

$$\Phi_{t0}(x_{it}\beta_0 + \eta_{t0}(x_i^t)) = \Phi_{t1}(x_{it}\beta_1 + \eta_{t1}(x_i^t)). \tag{A.1}$$

Differentiating with respect to $x_{it1}$ obtains

$$\Phi'_{t0}(x_{it}\beta_0 + \eta_{t0}(x_i^t))(\beta_{01} + \frac{\partial \eta_{t0}(x_i^t)}{\partial x_{it1}}) = \Phi'_{t1}(x_{it}\beta_1 + \eta_{t1}(x_i^t))(\beta_{11} + \frac{\partial \eta_{t1}(x_i^t)}{\partial x_{it1}}). \tag{A.2}$$

Differentiating equation (A.1) with respect to $x_{it-1,1}$ obtains

$$\Phi'_{t0}(x_{it}\beta_0 + \eta_{t0}(x_i^t))\frac{\partial \eta_{t0}(x_i^t)}{\partial x_{it-1,1}} = \Phi'_{t1}(x_{it}\beta_1 + \eta_{t1}(x_i^t))\frac{\partial \eta_{t1}(x_i^t)}{\partial x_{it-1,1}}. \tag{A.3}$$

Assumption 2.2.2, along with equations (A.1) and (A.2) implies that

$$\begin{aligned} \Phi'_{t0}(x_{it}\beta_0 + \eta_{t0}(x_i^t))\beta_{01} &= \Phi'_{t1}(x_{it}\beta_1 + \eta_{t1}(x_i^t))\beta_{11} \Leftrightarrow \\ \Phi'_{t0}(\Phi_{t0}^{-1}(q_{it}))\beta_{01} &= \Phi'_{t1}(\Phi_{t1}^{-1}(q_{it}))\beta_{11} \Leftrightarrow \\ \Phi_{t0}^{-1'}(q_{it}) &= c\Phi_{t1}^{-1'}(q_{it}), \end{aligned} \tag{A.4}$$

where $c = \beta_{01}/\beta_{11}$, which is non-zero and finite by of Assumption 2.2.4. Furthermore, since $\Phi_t > 0$ we have that $c > 0$. Equation (A.4) implies that

$$\Phi_{t0}^{-1}(q_{it}) = c\Phi_{t1}^{-1}(q_{it}) + R_t, \Leftrightarrow \tag{A.5}$$

$$x_{it}\beta_0 + \eta_{t0}(x_i^t) = c(x_{it}\beta_1 + \eta_{t1}(x_i^t)) + R_t, \Leftrightarrow$$

$$x_{it}(\beta_0 - c\beta_1) - R_t = c\eta_{t1}(x_i^t)) - \eta_{t0}(x_i^t) \tag{A.6}$$

where $R_t$ is the integrating constant. Assumptions 2.2.5 and 2.2.6 imply that $R_t = 0,\ t = 1, \cdots, T$. Taking first difference of equation (A.6), defining $\delta_t := ((\beta_0 - c\beta_1)', (R_t - R_{t-1}))'$, pre-multiplying by $(x_{it-1}\ 1)'$ and taking expectations gives

$$E[(x_{it-1}\ 1)'(\Delta x_{it}\ 1)]\delta_t = cE[(x_{it-1}\ 1)'\Delta\eta_{t1}(x_i^t)] - E[(x_{it-1}\ 1)'\Delta\eta_{t0}(x_i^t)]. \tag{A.7}$$

By the law of iterated expectations we have that $E[(x_{it-1}\ 1)'\Delta\eta_t(x_i^t)] = E[(x_{it-1}\ 1)'E[\Delta\eta_t(x_i^t)|x_i^{t-1}]]$ and $E[\Delta\eta_t(x_i^t)|x_i^{t-1}] = E[\eta_t(x_i^t)|x_i^{t-1}] - \eta_t(x_i^{t-1}) = 0$, thus implying $E[(x_{it-1}\ 1)'\Delta\eta_t(x_i^t)] = 0$. This along with equation (A.7) implies that

$$E[(x_{it-1},\ 1)'(\Delta x_{it},\ 1)]\delta_t = 0, \tag{A.8}$$

which, by Assumption 2.2.3 implies that $\beta_0 = c\beta_1$ and $R_t = R_{t-1} = R$. This, along with equations (A.5) and (A.6) imply that $\Phi_{t0}(a) = \Phi_{t1}((a-R)/c)$ and $\eta_{t0}(x_i^t) = \eta_{t1}(x_i^t) + R$.

By Assumption 2.2.4, $1 = \|\beta_0\| = \|\beta_1\|$, which implies that $c = 1$. By Assumption 2.2.5, $E[\eta_{t0}(x_i^t)] = E[\eta_{t1}(x_i^t)] = 0$, which implies that $R = 0$. □

## A.2 Proof of Theorem 2.6

*Proof.* As in A.1 the random variable $q_{it} := E[y_{it}|z_{it}, d_{it} = 1]$ can be consistency estimated, it is known in large samples. This, equations (2.3) gives for each $t$

$$\begin{aligned} &\Phi_{t0}((\pi_t(z_{it}) + \varepsilon_{it})\beta_{10} + z_{2it}\beta_{02} + \eta_{t0}(x_i^t)) + \lambda_{t0}(\varepsilon_{it}, p_{it}) - \\ &\Phi_{t1}((\pi_t(z_{it}) + \varepsilon_{it})\beta_{11} + z_{2it}\beta_{12} + \eta_{t1}(x_i^t)) + \lambda_{t1}(\varepsilon_{it}, p_{it}) = 0. \end{aligned} \tag{A.9}$$

Differentiating equation (A.9) with respect to $z_{it1}$ and $\varepsilon_i t$ obtains

$$\begin{aligned} &\frac{\partial \pi_t}{\partial z_{1it}}' \left[ \left( \Phi_{t0}' \cdot \left( \beta_{10} + \frac{\partial \eta_{t0}}{\partial x_{1it}} \right) \right) - \left( \Phi_{t1}' \cdot \left( \beta_{11} + \frac{\partial \eta_{t1}}{\partial x_{1it}} \right) \right) \right] + \\ &\frac{\partial p_{it}}{\partial z_{1it}}' \left[ \frac{\partial \lambda_{t0}}{\partial p_{it}} - \frac{\partial \lambda_{t1}}{\partial p_{it}} \right] = 0, \end{aligned} \tag{A.10}$$

$$\begin{aligned} &\left[ \left( \Phi_{t0}' \cdot \left( \beta_{10} + \frac{\partial \eta_{t0}}{\partial x_{1it}} \right) \right) - \left( \Phi_{t1}' \cdot \left( \beta_{11} + \frac{\partial \eta_{t1}}{\partial x_{1it}} \right) \right) \right] + \\ &\left[ \frac{\partial \lambda_{t0}}{\partial \varepsilon_{it}} - \frac{\partial \lambda_{t1}}{\partial \varepsilon_{it}} \right] = 0. \end{aligned} \tag{A.11}$$

The full rank condition of Assumption 2.5.3 implies then that

$$\Phi_{t0}' \cdot \left( \beta_{10} + \frac{\partial \eta_{t0}}{\partial x_{1it}} \right) - \Phi_{t1}' \cdot \left( \beta_{11} + \frac{\partial \eta_{t1}}{\partial x_{1it}} \right) = 0, \tag{A.12}$$

$$\frac{\partial \lambda_{t0}}{\partial p_{it}} - \frac{\partial \lambda_{t1}}{\partial p_{it}} = 0, \tag{A.13}$$

$$\frac{\partial \lambda_{t0}}{\partial \varepsilon_{it}} - \frac{\partial \lambda_{t1}}{\partial \varepsilon_{it}} = 0. \tag{A.14}$$

Equations (A.13) and (A.14) imply that $\lambda_{t0}(\varepsilon_{it}, p_{it}) = \lambda_{t1}(\varepsilon_{it}, p_{it}) + R_{1t}$, where $R_{1t}$ is the integrating constant. Differentiating equation (A.9) with respect to $z_{1i(t-1)}$ obtains

$$\frac{\partial \pi_{t-1}}{\partial z_{1i(t-1)}}' \left[ \Phi_{t0}' \cdot \frac{\partial \eta_{t0}}{\partial x_{1i(t-1)}} - \Phi_{t1}' \cdot \frac{\partial \eta_{t1}}{\partial x_{1i(t-1)}} \right] + \frac{\partial p_{it}}{\partial z_{1i(t-1)}}' \left[ \frac{\partial \lambda_{t0}}{\partial p_{it}} - \frac{\partial \lambda_{t1}}{\partial p_{it}} \right] = 0,$$

which implies that

$$\Phi'_{t0} \cdot \frac{\partial \eta_{t0}}{\partial x_{1i(t-1)}} = \Phi'_{t1} \cdot \frac{\partial \eta_{t1}}{\partial x_{1i(t-1)}}. \tag{A.15}$$

Assumption 2.5.2, along with equations (A.12) and (A.15) imply that

$$\Phi'_{t0} \cdot \beta_{110} = \Phi'_{t1} \cdot \beta_{111}. \tag{A.16}$$

Define $\mu_{it} := q_{it} - \lambda_{t0}(\varepsilon_{it}, p_{it})$. Then equation (A.16) can be written as

$$\Phi'_{t0}\left(\Phi_{t0}^{-1}(\mu_{it})\right) \cdot \beta_{110} = \Phi'_{t1}\left(\Phi_{t1}^{-1}(\mu_{it} + R_{1t})\right) \cdot \beta_{111} \Leftrightarrow \tag{A.17}$$

$$\Phi_{t0}^{-1'}(\mu_{it}) = c\Phi_{t1}^{-1'}(\mu_{it} + R_{1t}) \Rightarrow$$

$$\Phi_{t0}^{-1}(\mu_{it}) = c\Phi_{t1}^{-1}(\mu_{it} + R_{1t}) + R_{2t}, \tag{A.18}$$

Where $R_{2t}$ is the integrating constant. The rest of the proof follows exactly Appendix A.1 to obtain $\beta_0 = c\beta_1$, $\Phi_{t0}(a) = \Phi_{t1}((a - R_{2t})/c) - R_{1t}$, $\lambda_{t0}(\varepsilon_{it}, p_{it}) = \lambda_{t1}(\varepsilon_{it}, p_{it}) + R_{1t}$, and $\eta_{t0}(x_i^t) = \eta_{t1}(x_i^t) + R_{1t} + R_{2t}$. The normalizations of Assumptions 2.5.5 - 7 imply then that $c = 1$ and $R_{1t} = R_{2t} = 0$.

□

## A.3 Proof of Lemma 4.2

*Proof.* First note that

$$\begin{aligned}
&\int m(z_1, z_{2i})\omega_i(z_1)dz_1 - m(z_{1i}, z_{2i}) \\
&= \int m(z_{1i} - \sigma v_1, z_{2i})\mathfrak{K}(v_1)dv_1 - m(z_{1i}, z_{2i}) \\
&= \int [m(z_{1i} - \sigma v_1, z_{2i}) - m(z_{1i}, z_{2i})]\mathfrak{K}(v_1)dv_1 \\
&= \int \left[0 - m'_1(z_{1i}, z_{2i})\sigma v_1 + m''_{11}(z_{1i} - \bar{\sigma} v_1, z_{2i})\frac{\sigma^2 v_1^2}{2}\right]\mathfrak{K}(v_1)dv_1 \\
&= \frac{\sigma^2}{2}\int m''_{11}(z_{1i} - \bar{\sigma} v_1, z_{2i})v_1^2\mathfrak{K}(v_1)dv_1, \qquad \text{(A.19)}
\end{aligned}$$

where the first equality comes from the change of variable $v_1 = (z_{1i} - z_1)/\sigma$, the third comes from a the expanding around $\sigma = 0$, and the fourth equality comes from Assumption 4.3. By assumption $|m''_{11}(z_{1i} - \bar{\sigma} v_1, z_{2i})| \leq C < \infty$, thus giving

$$\begin{aligned}
\left|\frac{1}{N}\sum_{i=1}^{N}\left\{\int m(z_1, z_{2i})\omega_i(z_1)dz_1 - m(z_{1i}, z_{2i})\right\}\right| &\leq \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2}{2}\int |m''_{11}(z_{1i} - \bar{\sigma} v_1, z_{2i})|v_1^2\mathfrak{K}(v_1)dv_1 \\
&\leq C\sigma^2\int v^2\mathfrak{K}(v_1)dv_1 = o_p(1). \qquad \text{(A.20)}
\end{aligned}$$

The rest of the proofs follows exactly these arguments. □

## A.4 Proof of Theorem 5.2

*Proof.* Let $h_0 := (h_{10}, h_{20}, h_{30})$, $\hat{h} := (\hat{h}_1, \hat{h}_2, \hat{h}_3)$ and write $\hat{Q}(\theta) = \hat{Q}(\theta, \hat{h})$. Define $Q_N(\theta, h_0) := \frac{1}{N}\sum_{i=1}^{N} \kappa_i(\Delta\varphi_i(q_{i0} - \tau(p_{i0})) - s_{i0}\beta)'\Sigma^{-1}(\Delta\varphi_i(q_{i0} - \tau(p_{i0})) - s_{i0}\beta)$. We make the following claims

$$\sup_{\Theta} |Q_N(\theta, h_0) - Q_0(\theta)| \xrightarrow{p} 0 \tag{A.21}$$

$$\sup_{\Theta} |Q_N(\theta, \hat{h}) - Q_N(\theta, h_0)| \xrightarrow{p} 0 \tag{A.22}$$

*Proof of claim A.21*: Note that for any $\theta \in \Theta$, $\|Q(x, \theta, h_0)\| \leq \kappa_i \|\Delta\varphi(q_i - \lambda(\varepsilon_i, p_i)) - s_i\beta\|^2 \|\Sigma^{-1}\| \leq (\|\varphi(q_i - \lambda(\varepsilon_i, p_i))\| + \|s_i\|)^2 \|\Sigma^{-1}\| \leq C_1 + C_2\|s_i\| + C_3\|s_i\|^2 \leq d(x)$ with $E[d(x)] < \infty$ for some $d(x)$. The third inequality comes from $\varphi_t$ being in the Sobolev ball, and the last inequality comes from Assumption 4.3. This, the compactness of $\Theta$ and Lemma 2.4 of Newey and McFadden (1994) proves claim A.21

*Proof of claim A.22*: For all $\theta \in \Theta$, we have that $|\hat{Q}(\theta, \hat{h}) - Q_N(\theta, h_0)| \leq \sum_i \kappa_i(\hat{\rho}_i'(\theta)\hat{\Sigma}^{-1}\hat{\rho}_i(\theta) - \hat{\rho}_i'(\theta)\Sigma^{-1}\hat{\rho}_i(\theta) + \hat{\rho}_i'(\theta)\Sigma^{-1}\hat{\rho}_i(\theta) - \rho_{i0}'(\theta)\Sigma^{-1}\rho_{i0}(\theta))/N \leq \|\hat{\Sigma}^{-1} - \Sigma^{-1}\| \sum_i \kappa_i \|\hat{\rho}_i(\theta)\|/N + \|\Sigma^{-1}\| \times \sum_i \kappa_i \|\hat{\rho}_i(\theta) - \rho_{i0}(\theta)\|^2/N) + 2\|\Sigma^{-1}\|(\sum_i \kappa_i \|\rho_{i0}(\theta)\|^2/N)^{1/2}(\sum_i \kappa_i \|\hat{\rho}_i(\theta) - \rho_{i0}(\theta)\|^2/N)^{1/2}$. Also, by the mean value theorem and the triangle inequality, we have that $\|\hat{\rho}_i(\theta) - \rho_{i0}(\theta)\|^2 \leq (C_1\|\hat{q}_i - q_{i0}\| + C_2\|\hat{\varepsilon}_i - \varepsilon_{i0}\| + C_3\|\hat{p}_i - p_{i0}\| + C_4\|\hat{s}_i - s_{i0}\|)^2$. Assumptions 4.3, 5.1, and the results of Newey (1994b) therefore implies that $|\hat{Q}(\theta, \hat{h}) - Q_N(\theta, h_0)| \xrightarrow{p} 0$ for all $\theta \in \Theta$. Given that $\Theta$ is compact, the convergence is uniform. This proves claim A.22.

Given that $\hat{\theta}$ minimizes $\hat{\Theta}(\theta, \hat{h})$, we have that

$$\begin{aligned} \hat{Q}(\hat{\theta}, \hat{h}) - Q_0(\theta_0) &\leq \hat{Q}(\theta_0, \hat{h}) - Q_0(\theta_0) \\ &\leq \sup_{\Theta} |Q_N(\theta, \hat{h}) - Q_N(\theta, h_0)| + \sup_{\Theta} |Q_N(\theta, h_0) - Q_0(\theta)| \xrightarrow{p} 0, \end{aligned} \tag{A.23}$$

where convergence comes from claims A.21 and A.22. Also, given that $\theta_0$ minimizes $Q_0(\theta)$ we have that

$$\begin{aligned} 0 &\leq Q_0(\hat{\theta}) - Q_0(\theta_0) \\ &= Q_0(\hat{\theta}) - Q_N(\hat{\theta}, h_0) + Q_N(\hat{\theta}, h_0) - Q_N(\hat{\theta}, \hat{h}) + Q_N(\hat{\theta}, \hat{h}) - Q_0(\theta_0) \\ &\leq \sup_{\Theta} |Q_0(\theta) - Q_N(\theta, h_0)| + \sup_{\Theta} |Q_N(\theta, h_0) - Q_N(\theta, \hat{h})| + |Q_N(\hat{\theta}, \hat{h}) - Q_0(\theta_0)| \xrightarrow{p} 0, \end{aligned} \tag{A.24}$$

by claims A.21 and A.22 and by equation (A.23). The last inequality on the RHS of equation requires that $\hat{\theta} \in \Theta$. The conditions on the kernels, as well as the monotonization of the estimates of the $\varphi$'s ensures that this is the case (see Gayle and Namoro (2005) for discussions). Since the model is identified, for all $\delta > 0$ there exists $\varepsilon > 0$ such that $d[\theta, \theta_0] > \delta \Rightarrow Q_0(\theta) - Q_0(\theta_0) > \varepsilon$, which implies that $\Pr(d[\hat{\theta}, \theta_0] > \delta) \leq \Pr(\{Q_0(\hat{\theta}) - Q_0(\theta_0)\} > \varepsilon)$. The latter converges to zeros in

probability by equation (A.24). □

## A.5 Proof of Theorem 5.4

*Proof.* Given the discussions at the beginning of Section 5.2, Proposition 2 of Newey (1994a) implies that the estimation of $\varphi$ and $\lambda$ can be ignored in calculating the asymptotic distribution of $\hat{\beta}$. We begin by deriving the asymptotic distribution of the unfeasible estimator $\tilde{\beta}$ which substitutes $\Sigma$ for $\hat{\Sigma}$ in the resulting objective function. The first order condition is given by

$$\hat{m}(\tilde{\beta},\hat{h}) := \frac{1}{N}\sum_{i=1}^{N} \kappa_i B_{i0}(\tilde{\beta},\hat{h})'\Sigma^{-1}(\Delta\varphi_0(\hat{q}_i - \lambda_0(\hat{\varepsilon}_i,\hat{p}_i)) - \hat{s}_i\tilde{\beta}) = \frac{1}{N}\sum_{i=1}^{N} \kappa_i B_{i0}(\tilde{\beta},\hat{h})'\Sigma^{-1}\rho_{i0}(\tilde{\beta},\hat{h}) = 0.$$

Expanding $\hat{m}(\tilde{\beta},\hat{h})$ around $\beta_0$ obtains

$$\hat{m}(\tilde{\beta},\hat{h}) = \hat{m}(\beta_0,\hat{h}) + [\hat{M}_1(\bar{\beta},\hat{h}) + \hat{M}_2(\bar{\beta},\hat{h})](\tilde{\beta}-\beta_0), \tag{A.25}$$

where $\bar{\beta}$ is the mean value between $\tilde{\beta}$ and $\beta_0$, $\hat{M}_1(\bar{\beta},\hat{h}) := N^{-1}\sum_{i=1}^{N}\kappa_i B_{i0}(\bar{\beta},\hat{h})'\Sigma^{-1}B_{i0}(\bar{\beta},\hat{h})$, and $\hat{M}_2(\bar{\beta},\hat{h}) := N^{-1}\sum_{i=1}^{N}\kappa_i\left[\rho_{i0}(\bar{\beta},\hat{h})'\Sigma^{-1}\otimes I_K\right]\nabla_\beta B_{i0}(\beta,\hat{h})$. Invertibility implies

(A.26)

$$\sqrt{n}(\tilde{\beta}-\beta_0) = [\hat{M}_1(\bar{\beta},\hat{h}) + \hat{M}_2(\bar{\beta},\hat{h})]^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\kappa_i B_{i0}(\beta_0,\hat{h})'\Sigma^{-1}(\Delta\varphi_0(\hat{q}_i - \lambda_0(\hat{\varepsilon}_i,\hat{p}_i)) - \hat{s}_i\beta_0).$$

Let $g(x_i,z_i,\hat{h}) := \kappa_i B_{i0}(\beta_0,\hat{h})'\Sigma^{-1}(\Delta\varphi_0(\hat{q}_i - \lambda_0(\hat{\varepsilon}_i,\hat{p}_i)) - \hat{s}_i\beta_0)$. Because $\beta_0$ uniquely minimizes $Q_0(\beta,h_0)$, we have that $E[g(x_i,z_i,h_0)] = 0$. Furthermore, $\|g(x_i,z_i,h_0)\| \le \kappa_i\|B_{i0}(h_0)\|\|\Sigma^{-1}\|\|\rho_{i0}(h_0)\| = C\kappa_i\|B_{i0}(h_0)\|$, given $E[\|g(x_i,z_i,h_0)\|^2] < \infty$ by the boundedness conditions in Assumption 4.1. The rest of the proof proceeds by checking the conditions of Theorem 8.11 of Newey and McFadden (1994). Let $\underline{f}(x_i) := \text{diag}(f(x_i^t),\ t = 1,\cdots,T-1)$, and $\underline{f}(w_i) := \text{diag}(f(w_{it}),\quad t = 1,\cdots,T)$. Define $G_{1ik} := \text{diag}((-s_{itk}\quad 1),\quad t = 2,\cdots,T)$, $G_{2i} := \text{diag}((-q_{it}\quad 1),\quad t = 1,\cdots,T)$, and $G_{3i} := \text{diag}((-p_{it}\quad 1),\ t = 1,\cdots,T)$. Define the linear (in $h$) operator

$$D(w_i,h) := \kappa_i B_{i0}'\Sigma^{-1}\left(\underline{f}^{-1}(x_i)\sum_{k=1}^{K} G_{1ik}h_{1ik} + R_{1i}\underline{f}^{-1}(w_i)G_{2i}h_{2i} - R_{2i}\underline{f}^{-1}(w_i)G_{3i}h_{3i}\right).$$

Conditions (i) and (ii) of Theorem 8.11 of Newey and McFadden (1994) are satisfied by noting the appropriate boundedness conditions that gives $\|g(w_i,\hat{h}) - g(w_i,h_0) - D(w_i,\hat{h}-h_0)\| \le c_1(w_i)\|\hat{h}-h_0\|^2$ with $E[c_1(w_i)] < \infty$ and $\|D(w,h)\| \le c_2(w)\|h\|$ with $E[c_2(w)^2] < \infty$. Direct calculations give

$$\int D(w,h)f_0(w)dw = B_0'\Sigma^{-1}\int\kappa\left(\sum_{k=1}^{K} G_{1ik}h_{1ik} + R_{1i}G_{2i}h_{2i} - R_{2i}G_{3i}h_{3i}\right)dw,$$

with $\|h\| < \infty$ on the set $\mathcal{W}$, which is controlled by the trimming function $\kappa$. This satisfies condition (iii) of Theorem 8.11 of Newey and McFadden (1994). Condition (iv) of Newey and McFadden

(1994) is satisfied noting boundedness and almost everywhere continuity of $(G_{1j},\ j=1,\cdots,K)$, $G_2,\ G_3$, $R_1$ and $R_2$ on the set $\mathcal{W}$, assumptions 4.3 and 4.1. See the derivations on pages 2209 and 2210 of Newey and McFadden (1994) for details. Define $v_k(w) := \kappa G_{1k},\ k=1,\cdots,K$ $v_{k+1}(w) := R_1 G_2$ and $v_{k+2}(w) := -R_2 G_{3.}$ Let $\delta_k(w) = v_k(w)\Delta\tilde{x}_k - E[v_k(w)\Delta\tilde{x}_k],\ k=1,\cdots,K$, $\delta_{k+1}(w) := v_{k+1}(w)\tilde{q}$ and $\delta_{k+2}(w) := v_{k+2}(w)\tilde{p}$. Then by Theorem 8.11 of Newey and McFadden (1994) we have that $\sum_{i=1}^{N} g(w_i,\hat{h})/\sqrt{N} \xrightarrow{d} N\left(0, \mathrm{Var}[\kappa g(w_i,h_0) + \kappa B_0'\Sigma^{-1}\sum_k \delta_k(w)]\right)$. Straightforward calculations show that for $k=1,\cdots,K$, $v_k(w)\Delta\tilde{x}_k = \Delta x_k - s_{0k}$ and by the law of iterated expectations $E[v_k(w)\Delta\tilde{x}_k] = 0$, making $\delta_k(w) = \Delta x_k - s_{0k}$. Similar calculations show that $\underline{\delta}_{k+1}(w) = R_1(y_i - q_{i0})$ and $\delta_{k+2} \equiv -R_2(d_i - p_{i0})$. Finally it is straightforward to show that $\|\hat{M}_1(\beta,\hat{h}) - M(w)\| = o_p(1)$ and $\|\hat{M}_1(\beta,\hat{h})\| = o_p(1)$ where $M(w) := E[B_0'\Sigma^{-1}B_0]$ is nonsingular. By Slutsky theorem we have $\sqrt{N}(\tilde{\beta}-\beta_0) \xrightarrow{d} N(0,V)$. From this and under Assumption 5.1 it is straightforward to show that $\sqrt{N}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0,V)$ □

## References


ANDERSON, E. (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society*, Series B, 32, 283–301.

ARELLANO, M. AND R. CARRASCO (2003): "Binary Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics*, 115, 125–157.

BAUSCHKE, H. AND J. BORWEIN (1996): "On Optimal Algorithms for Solving Convex Feasibility Problems," *SIAM Review*, 38, 367–426.

BRUNK, H. D. (1958): "On the Estimation of Parameters Restricted by Inequalities," *Annals of Mathematical Statistics*, 29, 437–454.

BUJA, A., T. HASTIE, AND R. TIBSHIRANI (1989): "Linear Smoothers and Additive Models," *The Annals of Statistics*, 17, 453–555.

CHAMBERLAIN, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economics Studies*, XLVII, 225–238.

——— (1984): *Panel Data*, Elsevier Science Publishers B.V.

——— (1993): "Feedback in Panel Data Models," .

CHEN, S. (1998): "Root-N Consistent Estimation of a Panel Data Sample Selection Model," *MIMEO: The Hong Kong University of Science and Technology*.

CHENEY, W. AND A. GOLDSTEIN (1959): "Proximitymaps for convex sets," *Proceedings of the American Mathematical Society*, 10, 448–450.

DAS, M., W. NEWEY, AND F. VELLA (2003): "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33–58.

DEUTSCH, F. (2001): *Best Approximation in Inner Product Spaces*, Springer.

GAYLE, G. AND C. VIAUROUX (2007): "Root-N Consistent Semiparametric Estimators of a Dynamic Panel Data Sample Selection Model," *Journal of Econometrics*, 141, 179–212.

GAYLE, W.-R. AND S. D. NAMORO (2005): "Semiparametric Estimation of a Nonlinear Panel Data Model with Predetermined Variables and Semiparametric Individual Effects," *MIMEO: University of Pittsburgh*.

HARDLE, W., M. MULLER, S. SPERLICH, AND A. WERWATZ (2004): *Nonparametric and Semiparametric Models*, Springer.

HASTIE, T. AND R. TIBSHIRANI (1986): "Generalized Additive Models," *Statistical Science*, 1, 297–318.

ICHIMURA, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics*, 58, 71–120.

MAMMEN, E., O. LINTON, AND J. NIELSEN (1999): "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm Under Weak Conditions," *The Annals of Statistics*, 27, 1443–1490.

MAMMEN, E., J. MARRON, B. TURLACH, AND M. WAND (2001): "A General Projection Framework for Constrained Smoothing," *Statistical Science*, 16, 232–248.

MOOD, A., F. GRAYBILL, AND D. BOES (1974): *Introduction to the Theory of Statistics*, McGraw Hill.

MUNDLAK, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85.

NEWEY, W. (1994a): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

——— (1994b): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253.

NEWEY, W., J. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565–603.

NEWEY, W. K. AND D. MCFADDEN (1994): *Large Sample Estimation and Hypothesis Testing*, Elsevier Science Publishers.

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

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

| | $\beta_1$ | | $\beta_2$ | |
|---|---|---|---|---|
| | Mean Bias | RMSE | Mean Bias | RMSE |
| | $\delta_1 = 0.9,\ \delta_2 = 0.1$ | | | |
| N=100 | | | | |
| KMD | -0.0352 | 0.3367 | -0.1398 | 0.3906 |
| EKMD | -0.0548 | 0.3383 | -0.0815 | 0.2857 |
| N=200 | | | | |
| KMD | -0.0375 | 0.3097 | -0.0831 | 0.2863 |
| EKMD | -0.0170 | 0.2222 | -0.0677 | 0.1964 |
| N=400 | | | | |
| KMD | -0.0027 | 0.2075 | -0.0481 | 0.1848 |
| EKMD | -0.0023 | 0.0895 | -0.0102 | 0.0752 |
| | $\delta_1 = 0.6,\ \delta_2 = 0.4$ | | | |
| N=100 | | | | |
| KMD | -0.0567 | 0.3463 | -0.1253 | 0.3854 |
| EKMD | -0.0746 | 0.3568 | -0.0907 | 0.3275 |
| N=200 | | | | |
| KMD | -0.1179 | 0.3610 | -0.0514 | 0.3056 |
| EKMD | -0.0712 | 0.2592 | -0.0376 | 0.2807 |
| N=400 | | | | |
| KMD | -0.0148 | 0.1509 | -0.0319 | 0.2146 |
| EKMD | -0.0086 | 0.0736 | -0.0012 | 0.0546 |

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

| | $\beta_1$ | | $\beta_2$ | |
|---|---|---|---|---|
| | Mean Bias | RMSE | Mean Bias | RMSE |
| | $\delta_1 = 0.9,\ \delta_2 = 0.1$ | | | |
| N=100 | | | | |
| KMD | -0.0731 | 0.3802 | -0.1173 | 0.3617 |
| EKMD | -0.1025 | 0.3890 | -0.0831 | 0.3233 |
| N=200 | | | | |
| KMD | -0.0459 | 0.3214 | -0.1010 | 0.3366 |
| EKMD | -0.0039 | 0.2133 | -0.0578 | 0.2273 |
| N=400 | | | | |
| KMD | -0.0072 | 0.2172 | -0.0512 | 0.2084 |
| EKMD | -0.0015 | 0.1010 | -0.0088 | 0.0761 |
| | $\delta_1 = 0.6,\ \delta_2 = 0.4$ | | | |
| N=100 | | | | |
| KMD | -0.1043 | 0.4268 | -0.1131 | 0.3520 |
| EKMD | -0.0598 | 0.3629 | -0.1006 | 0.3178 |
| N=200 | | | | |
| KMD | -0.1285 | 0.3768 | -0.0644 | 0.3394 |
| EKMD | -0.0693 | 0.2723 | -0.0078 | 0.1467 |
| N=400 | | | | |
| KMD | -0.0131 | 0.1677 | -0.0380 | 0.2199 |
| EKMD | -0.0028 | 0.0893 | -0.0061 | 0.0714 |