Nonlinear Panel Data Models with Lagged Dependent Variables[1]

Xiaohong Chen[2] and Edward Vytlacil[3]

First Draft, June 2004

Abstract

……

---

[1]We would like to thank Jim Heckman, Bo Honore, Katerina Kyriazidou, Jim Powell and Nese Yildiz for helpful comments. Chen acknowledges financial support from the National Science Foundation and C.V. Starr Center at NYU. Vytlacil acknowledges support from ...

[2]Department of Economics, New York University, 269 Mercer Street, New York, NY 10003; E-mail: xiaohong.chen@nyu.edu

[3]Department of Economics, Stanford University.

1 Introduction

......

The rest of the paper is organized as follows. In Section 2 we first present the binary choice dynamic panel data model with general interaction, lagged dependent variables and arbitrarily correlated unobservables. We then consider the identification for the model with two periods. We also study the identification for the model with periods greater than two. In Section 3 we present two semiparametric efficient estimation procedures for binary choice dynamic panel data model with two periods. In Section 4 we study semiparametric efficient estimation for models with three (or higher) periods. Here to resolve the problem of curve of dimensionality, we impose some parametric dependence structure among the unobservables. Section 5 contains a small monte carlo study and Section 6 concludes.

2 Binary Choice Dynamic Panel Data Model

Binary-choice panel data model:

$$Y_1 = \mathbf{1}[X_{0,1} + X_{1,1}\beta_1 - U_1 \geq 0]$$

$$Y_2 = \mathbf{1}[X_{0,2} + X_{1,2}\beta_2 + Y_1(X_2\delta_2 + \gamma_2) - U_2 \geq 0]$$

$$\vdots$$

$$Y_T = \mathbf{1}[X_{0,T} + X_{1,T}\beta_T + \sum_{j=1}^{T-1} Y_j(X_T\delta_T^j + \gamma_T^j) + \sum_{j=1}^{T-1}\sum_{k=j+1}^{T-1} Y_jY_k(X_T\delta_T^{jk} + \gamma_T^{jk}) - U_T \geq 0]$$

where $X_t = (X_{0,t}, X_{1,t})$ with $X_{0,t}$ a scalar random variable and $X_{1,t}$ a $1 \times d_t$ row random vector. We will also partition $\delta_t = (\delta_{0t}, \delta_{1t}')'$ where δ_{0t} is a scalar.

Maintained Assumptions:

(A1) $(U_1, ..., U_T) \perp\!\!\!\perp (X_1, ..., X_T)$

(A2) Distribution of $(U_1, ..., U_T)$ has continuous and strictly positive density w.r.t. Lebesgue measure on $\Re^T$.

(A3) For $t = 1, ..., T$, $X_t \in R^{d_t+1}$, X_t does not contain an intercept, and X_t does not lie in a proper linear subspace of R^{d_t+1} with probability one.

(A4) For $t = 1, ..., T$, X_t has a density with respect to Lebesgue measure on $\Re^{d_t+1}$.

We will also require an exclusion restrictions: A continuous variable enters each index that does not appear in any other index. We will state this restriction formally for each special case of the model. Without loss of generality, let this regressor be $X_{0,t}$, the first element of X_t.

The model is written imposing that the coefficient on $X_{0,t}$ is one in each period. This normalization is an innocuous scale normalization given our assumption that $X_{0,t}$ has a nonzero coefficient. Assumption (A3) includes the restriction that X_t does not contain an intercept, which provides us with an innocuous location normalization. Assumption (A4) is not strictly required for our approach and can be considerably relaxed, but is imposed in order to simplify the analysis. We will require one continuous regressor in each period that does not enter the index for other periods, but the assumption that all regressors are continuous is only imposed for ease of exposition.

(ADD: emphasize comparison to Honore and Kyriazidou – (A1) is stronger, need independence, but do not require permanent-transitory structure but can instead allow very general forms of dependence in the U_t structure, can allow marginal distribution of U_t to vary with t, can allow β_t to vary with t, can allow for interactions . . . etc.. Also, add comparison to Cameron and Heckman, cannot allow single spell duration model but will not require identification at infinity).

2.1 Identification for Two Period Case

2.1.1 Simplest case: two period, no interactions

For ease of exposition, first consider the special case:

- 2 Period Model
- All Regressors Continuous
- No interactions between Y_1 and X_2 in the Y_2 equation.

We consider this case first in order to show the central identification argument in a simple context. We will then show how our analysis extends to more complicated models. The assumption of no interactions between Y_1 and X_2 in the Y_2 equation is the assumption that the effect of the lagged dependent variable on the current choice does not depend on regressors. For example (include example) This assumption is frequently imposed both in practice and in econometric theory. Thus, consider:

$$\begin{aligned} Y_1 &= \mathbf{1}[X_{0,1} + X_{1,1}\beta_1 - U_1 \geq 0] \\ Y_2 &= \mathbf{1}[X_{0,2} + X_{1,2}\beta_2 + Y_1\gamma_2 - U_2 \geq 0]. \end{aligned} \tag{1}$$

First consider identification of β_1 and β_2. Identification of these parameters will follow by exploiting standard analysis for average derivatives. From (A1) and (A2) we have $E(Y_1 \mid X_1) = F_{U_1}(X_{0,1} + X_{1,1}\beta_1)$ is continuously differentiable in X_1 with derivative given by

$$\begin{aligned} \frac{\partial}{\partial X_{0,1}} E(Y_1 \mid X_1) &= f_{U_1}(X_{0,1} + X_{1,1}\beta_1) \\ \frac{\partial}{\partial X_{1,1}} E(Y_1 \mid X_1) &= f_{U_1}(X_{0,1} + X_{1,1}\beta_1)\beta_1. \end{aligned}$$

Given assumption (A3) and (A4), we can then identify β_1 by

$$\beta_1 = \left[E\left(\frac{\partial}{\partial X_{0,1}} E(Y_1 \mid X_1) \right) \right]^{-1} E\left(\frac{\partial}{\partial X_{1,1}} E(Y_1 \mid X_1) \right).$$

Given identification of β_1, consider $E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2)$. From (A1) and (A2), we have that $E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2) = \Pr[U_2 \leq X_{0,2} + X_{1,2}\beta_2 \mid U_1 > X_{0,1} + X_{1,1}\beta_1]$ is differentiable in X_2 with derivative given by

$$\frac{\partial}{\partial X_2} E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2) = g(X_{0,2} + X_{1,2}\beta_2 \mid X_{0,1} + X_{1,1}\beta_1) \begin{bmatrix} 1 \\ \beta_2 \end{bmatrix}$$

where $g(\cdot \mid t)$ denotes the density of U_2 conditional on $U_1 > t$. Thus, we can identify β_2 by

$$\beta_2 = \left[E\left(\frac{\partial}{\partial X_{0,2}} E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2) \right) \right]^{-1} E\left(\frac{\partial}{\partial X_{1,2}} E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2$$

Given identification of β_1 and β_2, now consider identification of γ_2.

Using (A1) and (A2), we have that $E\big(Y_1Y_2 \big| X_{0,1} + X_{1,1}\beta_1 = t_1, X_{0,2} + X_{1,2}\beta_2 = t_2\big)$, $E\big(Y_1 \mid X_{0,1} + X_{1,1}\beta_1 = t_1\big)$, $E\big((1 - Y_1)Y_2 \big| X_{0,1} + X_{1,1}\beta_1 = t_1, X_{0,2} + X_{1,2}\beta_2 = t_2\big)$, and $E\big((1 - Y_1) \mid X_{0,1} + X_{1,1}\beta_1 = t_1\big)$ are differentiable functions of t_1, t_2. Define:

$$h_1(t_1, t_2) \equiv \frac{\frac{\partial}{\partial t_1} E\big(Y_1Y_2 \big| X_{0,1} + X_{1,1}\beta_1 = t_1, X_{0,2} + X_{1,2}\beta_2 = t_2\big)}{\frac{\partial}{\partial t_1} E\big(Y_1 \mid X_{0,1} + X_{1,1}\beta_1 = t_1\big)}$$

$$h_0(t_1, t_2) \equiv \frac{\frac{\partial}{\partial t_1} E\big((1 - Y_1)Y_2 \big| X_{0,1} + X_{1,1}\beta_1 = t_1, X_{0,2} + X_{1,2}\beta_2 = t_2\big)}{\frac{\partial}{\partial t_1} E\big((1 - Y_1) \mid X_{0,1} + X_{1,1}\beta_1 = t_1\big)}$$

Given identification of β_1 and β_2, we have identification of the relevant conditional expectations a.e. with respect to the joint distribution of $(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2)$. Given our assumption that X_t itself has a density with respect to Lebesgue measure, we have that identification of the conditional expectations a.e. implies identification of the derivatives of the conditional expectations a.e., and thus implies identification of h_1 and h_0 a.e. with respect to the joint distribution of $(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2)$.

Using (A1) and (A2), one can easily show that h_1 and h_0 are given by

$$h_1(t_1,t_2) = \Pr[U_2 \le t_2+\gamma_2 \mid U_1 = t_1]$$

$$h_0(t_1,t_2) = \Pr[U_2 \le t_2 \mid U_1 = t_1].$$

(include graph suggested by Honore to provide intuition for result?) From assumption (A2), we have $\Pr[U_2 \in B \mid U_1 = t_1] > 0$ for any open interval B, and thus

$$h_1(t_1,t_2) - h_0(t_1,t_2^*) = 0$$

if and only if

$$t_2 - \gamma_2 = t_2^*.$$

In other words, γ_2 is the unique solution to

$$h_1(t_1,t_2-\gamma_2) - h_0(t_1,t_2) = 0.$$

(include graph? show to cdfs shifted, similar to graph in Klein and Sherman? include footnote comparing to Klein and Sherman?) Since we identify h_1 and h_0 a.e. with respect to the joint distribution of $(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2)$, we have that γ_2 is identified if the following support condition holds:

(A5.1) $\text{Supp}(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2+\gamma_2)\bigcap \text{Supp}(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2) \neq \emptyset.$

(Will add discussion of condition).

Given identification $(\beta_1,\beta_2,\gamma_2)$, consider identification of the distribution of (U_1,U_2). From

$$E(Y_1Y_2 \mid X_1,X_2) = F_{U_1,U_2}(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2+\gamma_2),$$

we identify F_{U_1,U_2} a.e. with respect to the distribution of $(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2+\gamma_2)$. From

$$E((1-Y_1)Y_2 \mid X_1,X_2) = F_{-U_1,U_2}(-X_{0,1}-X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2),$$

we identify F_{-U_1,U_2} a.e. with respect to the distribution of $(-X_{0,1}-X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2)$. We thus identify $f_{U_1,U_2}(t_1,t_2)$ for $(t_1,t_2) \in$Supp$(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2+\gamma_2) \bigcup \text{Supp}(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2)$.

We now state this result formally.

Theorem 1 *Assume the model given by equation (1). Assume (A1)-(A4) and (A5.1). Then:*

1. $(\beta_1, \beta_2, \gamma_2)$ is identified,

2. $F_{U_1,U_2}(t_1, t_2)$ is identified for $(t_1, t_2) \in \text{Supp}(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2 + \gamma_2)$,

3. $F_{-U_1,U_2}(t_1, t_2)$ is identified for $(t_1, t_2) \in \text{Supp}(-X_{0,1} - X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2)$,

4. $f_{U_1,U_2}(t_1, t_2)$ is identified for $(t_1, t_2) \in \text{Supp}(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2+\gamma_2) \bigcup \text{Supp}(X_{0,1}+X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2)$.

2.1.2 Two periods, allowing interactions

We now relax that assumption of no X_2, Y_1 interactions in the Y_2 equation. For ease of exposition, we maintain the assumption of a two period model and that all regressors are continuous:

- 2 Period Model
- All Regressors Continuous

Thus, consider:

$$\begin{aligned} Y_1 &= \mathbf{1}[X_{0,1} + X_{1,1}\beta_1 - U_1 \geq 0] \\ Y_2 &= \mathbf{1}[X_{0,2} + X_{1,2}\beta_2 + Y_1(X_2\delta_2 + \gamma_2) - U_2 \geq 0]. \end{aligned} \tag{2}$$

First consider identification of β_1 and β_2. We identify β_1 exactly as before for the case of no interactions. Recall that we partition δ_2 by $\delta_2 = [\delta_{02}, \delta_{12}']'$ where δ_{02} is the first element of δ_2. Now consider identification of β_2 and of $\beta_2 + \delta_{12}$ up to scale given identification of β_1. From (A1) and (A2) we have that $E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2) = \Pr[U_2 \leq X_{0,2} + X_{1,2}\beta_2 \mid U_1 > X_{0,1} + X_{1,1}\beta_1]$ and $E(Y_2 \mid Y_1 = 1, X_{0,1} + X_{1,1}\beta_1, X_2) = \Pr[U_2 \leq X_{0,2}(1 + \delta_{02}) + X_{1,2}(\beta_2 + \delta_{12}) + \gamma_2 \mid U_1 \leq X_{0,1} + X_{1,1}\beta_1]$ are differentiable in X_2 with derivatives given by

$$\begin{aligned} \frac{\partial}{\partial X_{0,2}} E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2) &= g_0(X_{0,2} + X_{1,2}\beta_2 \mid X_{0,1} + X_{1,1}\beta_1) \\ \frac{\partial}{\partial X_{1,2}} E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2) &= g_0(X_{0,2} + X_{1,2}\beta_2 \mid X_{0,1} + X_{1,1}\beta_1)\beta_2 \end{aligned}$$

and

$$\begin{aligned} &\frac{\partial}{\partial X_2} E(Y_2 \mid Y_1 = 1, X_{0,1} + X_{1,1}\beta_1, X_2) \\ &\qquad = g_1(X_{0,2}(1 + \delta_{02}) + X_{1,2}(\beta_2 + \delta_{12}) + \gamma_2 \mid X_{0,1} + X_{1,1}\beta_1)(1 + \delta_{02}) \end{aligned}$$

$$\frac{\partial}{\partial X_2} E(Y_2 \mid Y_1 = 1, X_{0,1} + X_{1,1}\beta_1, X_2)$$
$$= g_1(X_{0,2}(1+\delta_{02}) + X_{1,2}(\beta_2 + \delta_{12}) + \gamma_2 \mid X_{0,1} + X_{1,1}\beta_1)(\beta_2 + \delta_{12})$$

where $g_0(\cdot \mid t)$ denotes the density of U_2 conditional on $U_1 > t$ and $g_1(\cdot \mid t)$ denotes the density of U_2 conditional on $U_1 \leq t$. Thus, we can identify

$$\beta_2 = \left[E\left(\frac{\partial}{\partial X_{0,2}} E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2) \right) \right]^{-1} E\left(\frac{\partial}{\partial X_{1,2}} E(Y_2 \mid Y_1 = 0, X_{0,1} + X_{1,1}\beta_1, X_2$$

and

$$(1/(1+\delta_{02})(\beta_2 + \delta_{12})$$
$$= \left[E\left(\frac{\partial}{\partial X_{0,2}} E(Y_2 \mid Y_1 = 1, X_{0,1} + X_{1,1}\beta_1, X_2) \right) \right]^{-1} E\left(\frac{\partial}{\partial X_{1,2}} E(Y_2 \mid Y_1 = 1, X_{0,1} + X_{1,1}\beta_1, X_2) \right)$$

Let $\tilde{\beta}_2 = (1/(1+\delta_{02})(\beta_2 + \delta_{12})$.

Given identification of β_1, β_2, and $\tilde{\beta}_2 = (1/(1+\delta_{02})(\beta_2 + \delta_{12})$, consider identification of δ_2 and γ_2. Note that identification of β_2 and of $(1/(1+\delta_{02})(\beta_2 + \delta_{12})$ does not immediately imply identification of δ_{12} is one has not identified the scale term $(1+\delta_{02})$. If we could identify δ_{02}, we could use knowledge of β_2 and $\tilde{\beta}_2 = (1/(1+\delta_{02})(\beta_2 + \delta_{12})$ to identify δ_{12}. We now turn to identification of δ_{02} and γ_2.

Redefine:

$$h_1(t_1, t_2) \equiv \frac{\frac{\partial}{\partial t_1} E\big(Y_1 Y_2 \big| X_{0,1} + X_{1,1}\beta_1 = t_1, X_2 = t_2\big)}{\frac{\partial}{\partial t_1} E\big(Y_1 \mid X_{0,1} + X_{1,1}\beta_1 = t_1\big)}$$
$$h_0(t_1, t_2) \equiv \frac{\frac{\partial}{\partial t_1} E\big((1-Y_1) Y_2 \big| X_{0,1} + X_{1,1}\beta_1 = t_1, X_2 = t_2\big)}{\frac{\partial}{\partial t_1} E\big((1-Y_1) \mid X_{0,1} + X_{1,1}\beta_1 = t_1\big)}$$

One can easily show:

$$h_1(t_1, t_2) = \Pr[U_2 \leq t_{02}(1+\delta_{02}) + t_{12}(\beta_2 + \delta_{12}) + \gamma_2 \mid U_1 = t_1]$$

$$h_0(t_1, t_2) = \Pr[U_2 \leq t_{02} + t_{12}\beta_2 \mid U_1 = t_1].$$

where we are partitioning t_2 as $t_2 = [t_{02}, t_{12}']'$ with t_{02} a scalar. From assumption (A2), we have $\Pr[U_2 \in B \mid U_1 = t_1] > 0$ for any open interval B, and thus

$$h_1(t_1, t_2) - h_0(t_1, t_2^*) = 0$$

if and only if

$$t_{02}^* + t_{12}^*\beta_2 = \gamma_2 + t_{02}(1+\delta_{02}) + t_{12}(\beta_2 + \delta_{12}).$$

Substituting in our identified quantity $\tilde{\beta}_2 = (1/(1+\delta_{02}))(\beta_2 + \delta_{12})$, we have

$$h_1(t_1, t_2) - h_0(t_1, t_2^*) = 0$$

if and only if

$$t_{02}^* + t_{12}^* \beta_2 = \gamma_2 + (1+\delta_{02})(t_{02} + t_{12}\tilde{\beta}_2).$$

This last equality gives us the identified function $t_{02}^* + t_{12}^*\beta_2$ as a linear function of the identified function $t_{02} + t_{12}\tilde{\beta}_2$ with slope the unknown coefficient $(1+\delta_{02})$ and intercept the unknown coefficient γ_2. Thus, if we can find points (t_1, t_2, t_2^*), $(\tilde{t}_1, \tilde{t}_2, \tilde{t}_2^*)$, such that $t_2 \neq \tilde{t}_2$ and $h_1(t_1, t_2) - h_0(t_1, t_2^*) = 0$ and $h_1(\tilde{t}_1, \tilde{t}_2) - h_0(\tilde{t}_1, \tilde{t}_2^*) = 0$, than we can solve the resulting two linear equations for the intercept γ_2 and the slope $(1+\delta_{02})$. (include graph?) Thus, we identify γ_2 and δ_{02} if the following support condition holds:

(A5.2)

$$\#\{\text{Supp}(X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}) + X_{1,2}(\beta_2+\delta_{12}) + \gamma_2) \bigcap \text{Supp}(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2)\} \geq 2$$

where recall that for any set A we have defined $\#A$ to denote the cardinality of the set A. (Will add discussion of condition) Given identification of γ_2 and δ_{02} and having previously identified β_1, β_2 and $(1/(1+\delta_{02}))(\beta_2+\delta_{12})$, we now have identification of $(\beta_1, \beta_2, \delta_2, \gamma_2)$.

Now consider identification of the joint distribution of (U_1, U_2). From

$$E(Y_1 Y_2 \mid X_1, X_2) = F_{U_1,U_2}(X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}) + X_{1,2}(\beta_2+\delta_{12}) + \gamma_2),$$

we identify F_{U_1,U_2} a.e. with respect to the distribution of $(X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}) + X_{1,2}(\beta_2+\delta_{12}) + \gamma_2)$. From

$$E((1-Y_1)Y_2 \mid X_1, X_2) = F_{-U_1,U_2}(-X_{0,1} - X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2),$$

we identify F_{-U_1,U_2} a.e. with respect to the distribution of $(-X_{0,1} - X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2)$. We thus identify $f_{U_1,U_2}(t_1, t_2)$ for $(t_1, t_2) \in$Supp$(X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}) + X_{1,2}(\beta_2+\delta_{12}) + \gamma_2)$ $\bigcup$ Supp$(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2)$.

We now state this result formally.

Theorem 2.1 *Assume the model given by equation (2). Assume (A1)-(A4) and (A5.2), and impose scale normalization (N). Than:*

1. $(\beta_1, \beta_2, \delta_2, \gamma_2)$ *is identified,*

2. *$F_{U_1,U_2}(t_1,t_2)$ is identified for $(t_1,t_2) \in$ Supp$(X_{0,1}+X_{1,1}\beta_1, X_{0,2}(1+\delta_{02})+X_{1,2}(\beta_2+\delta_{12})+\gamma_2)$,*

3. *$F_{-U_1,U_2}(t_1,t_2)$ is identified for $(t_1,t_2) \in$ Supp$(-X_{0,1}-X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2)$,*

4. *$f_{U_1,U_2}(t_1,t_2)$ is identified for $(t_1,t_2) \in$Supp$(X_{0,1}+X_{1,1}\beta_1, X_{0,2}(1+\delta_{02})+X_{1,2}(\beta_2+\delta_{12})+\gamma_2) \bigcup$ Supp$(X_{0,1}+X_{1,1}\beta_1, X_{0,2}+X_{1,2}\beta_2)$.*

2.2 Identification with three period case

Now consider the 3 period version of the model, Allowing Interactions, Allowing Multiple Lags:

- 3 Period Model
- All Regressors Continuous
- Y_1, Y_2, Y_1Y_2 allowed to enter equation for Y_3.

Thus, consider:

$$\begin{aligned} Y_1 &= \mathbf{1}[X_{0,1}+X_{1,1}\beta_1-U_1\geq 0] \\ Y_2 &= \mathbf{1}[X_{0,2}+X_{1,2}\beta_2+Y_1(X_2\delta_2^1+\gamma_2^1)-U_2\geq 0] \\ Y_3 &= \mathbf{1}[X_{0,3}+X_{1,3}\beta_3+Y_2(X_3\delta_3^2+\gamma_3^2)+Y_1(X_3\delta_3^1+\gamma_3^1)+Y_1Y_2(X_3\delta_3^{12}+\gamma_3^{12})-U_3\geq 0] \end{aligned} \quad (3)$$

Assume (A1)-(A4) and (A5.2) so that $(\beta_1,\beta_2,\delta_2^1,\gamma_2^1)$ is identified. We will now consider identification of the remaining parameters. Partition δ_3^1, δ_3^2 and δ_3^{12} by $\delta_3^j=[\delta_{03}^j,(\delta_{12}^j)']'$ where δ_{03}^j is the first element of δ_3^j and $j=1,2,12$. From (A1) and (A2) we have that

$$\begin{aligned} &E(Y_3 \mid Y_1=0,Y_2=0,X_{0,1}+X_{1,1}\beta_1,X_{0,2}+X_{1,2}\beta_2,X_3) \\ &\qquad = \Pr[U_3\leq X_{0,3}+X_{1,3}\beta_3 \mid U_1>X_{0,1}+X_{1,1}\beta_1, U_2>X_{0,2}+X_{1,2}\beta_2] \end{aligned}$$

$$\begin{aligned} &E(Y_3 \mid Y_1=0,Y_2=1,X_{0,1}+X_{1,1}\beta_1,X_{0,2}+X_{1,2}\beta_2,X_3) \\ &= \Pr[U_3\leq X_{0,3}(1+\delta_{03}^2)+X_{1,3}(\beta_3+\delta_{13}^2)+\gamma_3^2 \mid U_1>X_{0,1}+X_{1,1}\beta_1, U_2\leq X_{0,2}+X_{1,2}\beta_2] \end{aligned}$$

$$\begin{aligned} &E(Y_3 \mid Y_1=1,Y_2=0,X_{0,1}+X_{1,1}\beta_1,X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1),X_3) \\ &= \Pr[U_3\leq X_{0,3}(1+\delta_{03}^1)+X_{1,3}(\beta_3+\delta_{13}^1)+\gamma_3^1 \mid U_1\leq X_{0,1}+X_{1,1}\beta_1, U_2>X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1)+ \end{aligned}$$

$$E(Y_3 \mid Y_1 = 1, Y_2 = 1, X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1 + \delta^1_{02}) + X_{1,2}(\beta_2 + \delta^1_{12}), X_3)$$
$$= \Pr[U_3 \le X_{0,3}(1 + \delta^1_{03} + \delta^2_{03} + \delta^{12}_{03}) + X_{1,3}(\beta_3 + \delta^1_{13} + \delta^2_{13} + \beta^{23}_{13}) + (\gamma^1_3 + \gamma^2_3 + \gamma^{12}_3)$$
$$\mid U_1 \le X_{0,1} + X_{1,1}\beta_1, U_2 > X_{0,2}(1 + \delta^1_{02}) + X_{1,2}(\beta_2 + \delta^1_{12}) + \gamma_2]$$

are differentiable in X_3 with derivatives given by

$$\frac{\partial}{\partial X_3} E(Y_3 \mid Y_1 = 0, Y_2 = 0, X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_3)$$
$$= g_{00}(X_{0,3} + X_{1,3}\beta_3 \mid X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2) \begin{bmatrix} 1 \\ \beta_3 \end{bmatrix}$$

$$\frac{\partial}{\partial X_3} E(Y_3 \mid Y_1 = 0, Y_2 = 1, X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_3)$$
$$= g_{01}(X_{0,3}(1 + \delta^2_{03}) + X_{1,3}(\beta_3 + \delta^2_{13}) + \gamma^2_3 \mid X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2) \begin{bmatrix} 1 + \delta^2_{03} \\ \beta_3 + \delta^2_{13} \end{bmatrix}$$

$$\frac{\partial}{\partial X_3} E(Y_3 \mid Y_1 = 1, Y_2 = 0, X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1 + \delta^1_{02}) + X_{1,2}(\beta_2 + \delta^1_{12}), X_3)$$
$$= g_{10}(X_{0,3}(1+\delta^1_{03}) + X_{1,3}(\beta_3+\delta^1_{13}) + \gamma^1_3) \mid X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1+\delta^1_{02}) + X_{1,2}(\beta_2+\delta^1_{12}) + \gamma_2) \begin{bmatrix} 1 + \delta^1_{03} \\ \beta_3 + \delta^1_{13} \end{bmatrix}$$

$$\frac{\partial}{\partial X_3} E(Y_3 \mid Y_1 = 1, Y_2 = 1, X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1 + \delta^1_{02}) + X_{1,2}(\beta_2 + \delta^1_{12}), X_3)$$
$$= g_{11}(X_{0,3}(1 + \delta^1_{03} + \delta^2_{03} + \delta^{12}_{03}) + X_{1,3}(\beta_3 + \delta^1_{13} + \delta^2_{13} + \delta^{12}_{13}) + (\gamma^1_3 + \gamma^2_3 + \gamma^{12}_3)$$
$$\mid X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1 + \delta^1_{02}) + X_{1,2}(\beta_2 + \delta^1_{12}) + \gamma_2) \begin{bmatrix} 1 + \delta^1_{03} + \delta^2_{03} + \delta^{12}_{03} \\ \beta_3 + \delta^1_{13} + \delta^2_{13} + \delta^{12}_{13} \end{bmatrix},$$

where $g_{00}(\cdot \mid t_1, t_2)$ denotes the density of U_3 conditional on $U_1 > t_1$, $U_2 > t_2$; $g_{01}(\cdot \mid t_1, t_2)$ denotes the density of U_3 conditional on $U_1 > t_1$, $U_2 \le t_2$; $g_{10}(\cdot \mid t_1, t_2)$ denotes the density of U_3 conditional on $U_1 \le t_1$, $U_2 > t_2$; and $g_{11}(\cdot \mid t_1, t_2)$ denotes the density of U_3 conditional on $U_1 \le t_1$, $U_2 \le t_2$. Thus, we can identify

$$\beta_3 = \left[E\left(\frac{\partial}{\partial X_{0,3}} E(Y_3 \mid Y_1 = 0, Y_2 = 0, X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_3) \right) \right]^{-1}$$
$$\times E\left(\frac{\partial}{\partial X_{1,3}} E(Y_3 \mid Y_1 = 0, Y_2 = 0, X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_3) \right)$$

$$(\beta_3+\delta^2_{13})/(1+\delta^2_{03}) = \left[E\left(\frac{\partial}{\partial X_{0,3}} E(Y_3 \mid Y_1 = 0, Y_2 = 1, X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_3) \right) \right]^{-1}$$
$$\times E\left(\frac{\partial}{\partial X_{1,3}} E(Y_3 \mid Y_1 = 0, Y_2 = 1, X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_3) \right)$$

$$(\beta_3+\delta_{13}^1)/(1+\delta_{03}^1)$$
$$= \left[\frac{\partial}{\partial X_{0,3}} E(Y_3 \mid Y_1=1, Y_2=0, X_{0,1}+X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1), X_3)\right]^{-1}$$
$$\times E\left(\frac{\partial}{\partial X_{1,3}} E(Y_3 \mid Y_1=1, Y_2=0, X_{0,1}+X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1), X_3)\right)$$

$$(\beta_3+\delta_{13}^1+\delta_{13}^2+\delta_{13}^{12})/(1+\delta_{03}^1+\delta_{03}^2+\delta_{03}^{12})$$
$$= \left[\frac{\partial}{\partial X_{0,3}} E(Y_3 \mid Y_1=1, Y_2=1, X_{0,1}+X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1), X_3)\right]^{-1}$$
$$\times E\left(\frac{\partial}{\partial X_{1,3}} E(Y_3 \mid Y_1=1, Y_2=1, X_{0,1}+X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1), X_3)\right)$$

Let
$$\tilde\beta_3^1 = (\beta_3+\delta_{13}^1)/(1+\delta_{03}^1),$$
$$\tilde\beta_3^2 = (\beta_3+\delta_{13}^2)/(1+\delta_{03}^2),$$
and let
$$\tilde\beta_3^{12} = (\beta_3+\delta_{13}^1+\delta_{13}^2+\delta_{13}^{12})/(1+\delta_{03}^1+\delta_{03}^2+\delta_{03}^{12}).$$
Given identification of $\beta_3, \tilde\beta_3^1$, $\tilde\beta_3^2$, and $\tilde\beta_3^{12}$, we now turn to identification of the δ_{03}^1, δ_{03}^2, and δ_{03}^{12}.

Define:
$$h_{11}(t_1,t_2,t_3) \equiv \frac{\frac{\partial}{\partial t_1 \partial t_2} E\big(Y_1Y_2Y_3 \big| X_{0,1}+X_{1,1}\beta_1=t_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1)=t_2-\gamma_2^1, X_3=t_3\big)}{\frac{\partial}{\partial t_1 \partial t_2} E\big(Y_1Y_2 \mid X_{0,1}+X_{1,1}\beta_1=t_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1)=t_2\big)}$$

$$h_{10}(t_1,t_2,t_3)$$
$$\equiv \frac{\frac{\partial}{\partial t_1 \partial t_2} E\big(Y_1(1-Y_2)Y_3 \big| X_{0,1}+X_{1,1}\beta_1=t_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1)=t_2-\gamma_2^1, X_3=t_3\big)}{\frac{\partial}{\partial t_2} E\big(Y_1(1-Y_2) \mid X_{0,1}+X_{1,1}\beta_1=t_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1)=t_2\big)}$$
$$h_{01}(t_1,t_2,t_3) \equiv \frac{\frac{\partial}{\partial t_1 \partial t_2} E\big((1-Y_1)Y_2Y_3 \big| X_{0,1}+X_{1,1}\beta_1=t_1, X_{0,2}+X_{1,2}\beta_2=t_2, X_3=t_3\big)}{\frac{\partial}{\partial t_1 \partial t_2} E\big(Y_1Y_2 \mid X_{0,1}+X_{1,1}\beta_1=t_1, X_{0,2}+X_{1,2}\beta_2=t_2\big)}$$
$$h_{00}(t_1,t_2,t_3) \equiv \frac{\frac{\partial}{\partial t_1 \partial t_2} E\big((1-Y_1)(1-Y_2)Y_3 \big| X_{0,1}+X_{1,1}\beta_1=t_1, X_{0,2}+X_{1,2}\beta_2=t_2, X_3=t_3\big)}{\frac{\partial}{\partial t_1 \partial t_2} E\big(Y_1Y_2 \mid X_{0,1}+X_{1,1}\beta_1=t_1, X_{0,2}+X_{1,2}\beta_2=t_2\big)}$$

One can easily show:
$$\begin{aligned}
h_{11}(t_1,t_2,t_3) &= \Pr[U_3 \le t_{03}(1+\delta_{03}^1+\delta_{03}^2+\delta_{03}^{12})+t_{13}(\beta_3+\delta_{13}^1+\delta_{13}^2+\delta_{13}^{12}) \\
&\qquad + (\gamma_3^1+\gamma_3^2+\gamma_3^{12}) \mid U_1=t_1, U_2=t_2] \\
h_{01}(t_1,t_2,t_3) &= \Pr[U_3 \le t_{03}(1+\delta_{03}^2)+t_{13}(\beta_3+\delta_{13}^2)+\gamma_3^2 \mid U_1=t_1, U_2=t_2] \\
h_{10}(t_1,t_2,t_3) &= \Pr[U_3 \le t_{03}(1+\delta_{03}^1)+t_{13}(\beta_3+\delta_{13}^1)+\gamma_3^1 \mid U_1=t_1, U_2=t_2] \\
h_{00}(t_1,t_2,t_3) &= \Pr[U_3 \le t_{03}+t_{13}\beta_3 \mid U_1=t_1, U_2=t_2].
\end{aligned}$$

From assumption (A2), we have $\Pr[U_3 \in B \mid U_1 = t_1, U_2 = t_2] > 0$ for any open interval B, and thus

$$h_{11}(t_1,t_2,t_3) - h_{00}(t_1,t_2,t_3^*) = 0 \text{ if and only if } t_{03}^* + t_{13}^*\beta_3 = (\gamma_3^1 + \gamma_3^2 + \gamma_3^{12}) + t_{03}(1 + \delta_{03}^1 + \delta_{03}^2 + \delta_{03}^{12}) + t_{13}(\beta_3 + \delta_{13}^1 + \delta_{13}^2 + \delta_{13}^{12})$$

$$h_{01}(t_1,t_2,t_3) - h_{00}(t_1,t_2,t_3^*) = 0 \text{ if and only if } t_{03}^* + t_{13}^*\beta_3 = \gamma_3^2 + t_{03}(1 + \delta_{03}^2) + t_{13}(\beta_3 + \delta_{13}^2)$$

$$h_{10}(t_1,t_2,t_3) - h_{00}(t_1,t_2,t_3^*) = 0 \text{ if and only if } t_{03}^* + t_{13}^*\beta_3 = \gamma_3^1 + t_{03}(1 + \delta_{03}^1) + t_{13}(\beta_3 + \delta_{13}^1)$$

Plugging in for the identified quantities $\tilde{\beta}_3^1, \tilde{\beta}_3^2, \tilde{\beta}_3^{12}$, we obtain

$$h_{11}(t_1,t_2,t_3) - h_{00}(t_1,t_2,t_3^*) = 0 \text{ if and only if } t_{03}^* + t_{13}^*\beta_3 = (\gamma_3^1 + \gamma_3^2 + \gamma_3^{12}) + (1 + \delta_{03}^1 + \delta_{03}^2 + \delta_{03}^{12})(t_{03} + t_{13}\tilde{\beta}_3^{12})$$

$$h_{01}(t_1,t_2,t_3) - h_{00}(t_1,t_2,t_3^*) = 0 \text{ if and only if } t_{03}^* + t_{13}^*\beta_3 = \gamma_3^2 + (1 + \delta_{03}^2)(t_{03} + t_{13}\tilde{\beta}_3^2)$$

$$h_{10}(t_1,t_2,t_3) - h_{00}(t_1,t_2,t_3^*) = 0 \text{ if and only if } t_{03}^* + t_{13}^*\beta_3 = \gamma_3^1 + (1 + \delta_{03}^1)(t_{03} + t_{13}\tilde{\beta}_3^1)$$

These equalities define a system of three linear equations. Thus, if we can find points (t_1,t_2,t_3,t_3^*), $(\tilde{t}_1,\tilde{t}_2,\tilde{t}_3,\tilde{t}_3^*)$, that both satisfy the equalities $h_{11}(t_1,t_2,t_3) - h_{00}(t_1,t_2,t_3^*) = 0$ and $h_{11}(\tilde{t}_1,\tilde{t}_2,\tilde{t}_3) - h_{00}(\tilde{t}_1,\tilde{t}_2,\tilde{t}_3^*) = 0$, than we can solve the resulting linear equation for the intercept $(\gamma_3^1 + \gamma_3^2 + \gamma_3^{12})$ and the slope $(1 + \delta_{03}^1 + \delta_{03}^2 + \delta_{03}^{12})$. Likewise, given the appropriate pairs of points, we can solve the the second and third linear equations to obtain γ_3^1, γ_3^2, $(1+\delta_{03}^1)$ and $(1+\delta_{03}^2)$. If we can solve each linear equation for the slopes and intercepts, we can then solve for each of the original coefficients. Define assumption (A5.4a)-(A5.4c) by:

(A5.4a)
$$\#\Big\{\mathrm{Supp}\big(X_{0,1} + X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}^1) + X_{1,2}(\beta_2 + \delta_{12}^1) + \gamma_2^1, X_{0,3}(1 + \delta_{03}^1 + \delta_{03}^2 + \delta_{03}^{12}) + X_{1,3}(\beta_3 + \delta_{13}^1 + \delta_{13}^2 + \beta_{13}^{23}) + (\gamma_3^1 + \gamma_3^2 + \gamma_3^{12})\big) \bigcap \mathrm{Supp}\big(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_{0,3} + X_{1,3}\beta_3\big)\Big\} \geq 2,$$

(A5.4b)
$$\#\Big\{\mathrm{Supp}\big(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_{0,3}(1 + \delta_{03}^2) + X_{1,3}(\beta_3 + \delta_{13}^2) + \gamma_3^2\big) \bigcap \mathrm{Supp}\big(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_{0,3} + X_{1,3}\beta_3\big)\Big\} \geq 2,$$

(A5.4c)
$$\#\Big\{\mathrm{Supp}\big(X_{0,1}+X_{1,1}\beta_1, X_{0,2}(1+\delta_{02}^1)+X_{1,2}(\beta_2+\delta_{12}^1)+\gamma_2^1, X_{0,3}(1+\delta_{03}^1)+X_{1,3}(\beta_3+\delta_{13}^1)+\gamma_3^1\big) \bigcap \mathrm{Supp}\big(X_{0,1} + X_{1,1}\beta_1, X_{0,2} + X_{1,2}\beta_2, X_{0,3} + X_{1,3}\beta_3\big)\Big\} \geq 2.$$

(discuss: is stronger than we need, do not need to compare each outcome to $(0,0)$ outcome). We now state this result formally.

Theorem 2.2 *Assume the model given by equation (3). Assume (A1)-(A4), (A5.2), and (A5.4a)-(A5.4c). Then:*

1. $(\beta_1, \beta_2, \delta_2^1, \gamma_2^1, \beta_3, \delta_3^1, \delta_3^2, \delta_3^{12}, \gamma_3^1, \gamma_3^2, \gamma_3^{12})$ is identified,
2. $F_{U_1,U_2,U_3}(t_1,t_2,t_3)$ is identified for $(t_1,t_2,t_3) \in$
3. $f_{U_1,U_2,U_3}(t_1,t_2,t_3)$ is identified for $(t_1,t_2,t_3) \in$

3 Semiparametric Efficient Estimation in Two Period Case

We first consider the semiparametric efficient estimation when $T = 2$. Let $\{Z_i = (X_{1i}, X_{2i}, Y_{1i}, Y_{2i})'\}_{i=1}^n$ denote an i.i.d. drawn from the same distribution of $Z = (X_1, X_2, Y_1, Y_2)'$, which satisfies the specification in the previous section. Due to scale normalization for the slope coefficients in each period, we assume $X_t = (X_{0,t}, X_{1,t}) \in \Re^1 \times \Re^{d_t}$, $\beta_t \in \Re^{d_t}$ for $t = 1, 2$, while $\delta_2 \in \Re^{d_2+1}$ and $\gamma_2 \in \Re^1$. To simplify notation, we let $\theta = (\beta_1', \beta_2', \gamma_2, \delta_2')' \in \Theta$, a compact set in $\Re^{d_\theta}$ with $d_\theta = d_1 + 2d_2 + 2$. Let $f(u_1, u_2)$ be a candidate joint density function of (U_1, U_2), $f_1(u_1) = \int_{-\infty}^{\infty} f(u_1, v)dv$ be the corresponding candidate marginal density function and $F_1(u_1)$ the cdf of U_1, similarly $f_2(u_2)$ the candidate marginal density and $F_2(u_2)$ the cdf of U_2; $F_{2|1}(u_1, u_2) = \Pr\{U_2 \le u_2 | U_1 = u_1\} = \int_{-\infty}^{u_2} f(u_1, v)dv / f_1(u_1)$ be the corresponding candidate conditional cdf of U_2 given $U_1 = u_1$, and $f_{2|1}(u_1, u_2) = f(u_1, u_2)/f_1(u_1)$ be the corresponding candidate conditional density of U_2 given $U_1 = u_1$. Throughout this section we also denote $\alpha_o = (\theta_o, f_o) = (\beta_{o1}', \beta_{o2}', \gamma_{o2}, \delta_{o2}', f_o)$ as the true unknown parameters of interests and $\alpha = (\theta, f)$ as any candidate parameter values in the parameter space $\Theta \times \mathcal{F}$, where $\mathcal{F}$ denote the space of probability density functions $f : \Re^2 \to [0, \infty)$ that are Hölder continuous with exponent $r > 1$. Let $\mathcal{F}_n$ denote a sieve space for $\mathcal{F}$.

In this section we shall present two estimation procedures: optimally weighted sieve MD and sieve ML. By applying the general theory of Ai and Chen (2003), the optimally weighted sieve MD estimator of θ_o will be root-n consistent and semiparametrically efficient. By applying the general theory of Shen (1997), the sieve ML procedure will lead to semiparametric efficient estimation of any smooth functionals of the α_o, in particular the sieve ML estimators of θ_o is efficient.

3.1 Optimally weighted sieve MD estimation

We note that the model with $T = 2$ is equivalent to the following set of conditional moment restrictions:

$$
\begin{aligned}
E\left[Y_1Y_2 - \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} \{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}} f_o(u_1,u_2)du_2\}du_1 \mid X_1, X_2\right] &= 0, \\
E\left[(1-Y_1)Y_2 - \int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty} \{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}} f_o(u_1,u_2)du_2\}du_1 \mid X_1, X_2\right] &= 0, \\
E\left[Y_1(1-Y_2) - \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} \{\int_{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}}^{\infty} f_o(u_1,u_2)du_2\}du_1 \mid X_1, X_2\right] &= 0 \quad (4)
\end{aligned}
$$

We note that the following forth conditional moment restriction is redundant given the above three:

$$
E\left[(1-Y_1)(1-Y_2) - \int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty} \{\int_{X_{0,2}+X_{1,2}\beta_{o2}}^{\infty} f_o(u_1,u_2)du_2\}du_1 \mid X_1, X_2\right] = 0.
$$

Let $\rho(Z,\alpha) = (\rho_1(Z,\alpha), \rho_2(Z,\alpha), \rho_3(Z,\alpha))'$ be a $3 \times 1-$vector valued function with

$$
\begin{aligned}
\rho_1(Z,\alpha) &= Y_1Y_2 - \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_1} \{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_2+\gamma_2+X_2\delta_2} f(u_1,u_2)du_2\}du_1, \\
\rho_2(Z,\alpha) &= (1-Y_1)Y_2 - \int_{X_{0,1}+X_{1,1}\beta_1}^{\infty} \{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_2} f(u_1,u_2)du_2\}du_1, \\
\rho_3(Z,\alpha) &= Y_1(1-Y_2) - \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_1} \{\int_{X_{0,2}+X_{1,2}\beta_2+\gamma_2+X_2\delta_2}^{\infty} f(u_1,u_2)du_2\}du_1.
\end{aligned}
$$

Since $\rho(Z,\alpha) - \rho(Z,\alpha_o)$ does not depend on the endogenous variables Y_1, Y_2, we can apply the three-step optimally weighted sieve generalized least squares (GLS) or the one-step continuously updated sieve GLS to estimate θ_o efficiently. In the following we let $\Sigma(X_1, X_2, \theta, f) = Var\{\rho(Z,\theta,f)|X_1,X_2\}$ and

$$
\begin{aligned}
\Sigma_o(X_1,X_2) &= E\{\rho(Z,\theta_o,f_o)\rho(Z,\theta_o,f_o)'|X_1,X_2\} \\
&= \begin{bmatrix} E\{\rho_1(Z,\alpha_o)^2|X\} & E\{\rho_1(Z,\alpha_o)\rho_2(Z,\alpha_o)|X\} & E\{\rho_1(Z,\alpha_o)\rho_3(Z,\alpha_o)|X\} \\ E\{\rho_1(Z,\alpha_o)\rho_2(Z,\alpha_o)|X\} & E\{\rho_2(Z,\alpha_o)^2|X\} & E\{\rho_2(Z,\alpha_o)\rho_3(Z,\alpha_o)|X\} \\ E\{\rho_1(Z,\alpha_o)\rho_3(Z,\alpha_o)|X\} & E\{\rho_2(Z,\alpha_o)\rho_3(Z,\alpha_o)|X\} & E\{\rho_3(Z,\alpha_o)^2|X\} \end{bmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
E\{\rho_1(Z,\alpha_o)^2|X\} &= E\{Y_1Y_2|X\} \times [1 - E\{Y_1Y_2|X\}], \\
E\{\rho_2(Z,\alpha_o)^2|X\} &= E\{(1-Y_1)Y_2|X\} \times [1 - E\{(1-Y_1)Y_2|X\}], \\
E\{\rho_3(Z,\alpha_o)^2|X\} &= E\{Y_1(1-Y_2)|X\} \times [1 - E\{Y_1(1-Y_2)|X\}],
\end{aligned}
$$

and

$$
\begin{aligned}
E\{\rho_1(Z,\alpha_o)\rho_2(Z,\alpha_o)|X\} &= -E\{Y_1Y_2|X\} \times E\{(1-Y_1)Y_2|X\}, \\
E\{\rho_1(Z,\alpha_o)\rho_3(Z,\alpha_o)|X\} &= -E\{Y_1Y_2|X\} \times E\{Y_1(1-Y_2)|X\}, \\
E\{\rho_2(Z,\alpha_o)\rho_3(Z,\alpha_o)|X\} &= -E\{(1-Y_1)Y_2|X\} \times E\{Y_1(1-Y_2)|X\}.
\end{aligned}
$$

The *three-step optimally weighted sieve GLS* estimator $\widehat{\alpha}_{ogls} = (\widehat{\theta}_{ogls}, \widehat{f}_{ogls})$ can be obtained as follows:

Step 1: $(\widehat{\theta}_{gls}, \widehat{f}_{gls}) = \arg\min_{\theta\in\Theta, f\in\mathcal{F}_n} \frac{1}{n}\sum_{i=1}^{n} \rho(Z_i,\theta,f)'\rho(Z_i,\theta,f)$, where $\mathcal{F}_n$ is a sieve space for $\mathcal{F}$;

Step 2: compute a consistent estimator $\widehat{\Sigma}_o(X_1,X_2)$ for $\Sigma_o(X_1,X_2)$. We could either get a nonparametric sieve LS regression estimator of $\Sigma_o(X_1,X_2)$, or we could obtain a plug-in consistent estimator by using the $\Sigma_o(X_1,X_2)$ expression, the conditional moment restrictions (4) and the Step 1 estimates $(\widehat{\theta}_{gls}, \widehat{f}_{gls})$;

Step 3: compute the optimally weighted sieve GLS estimator $\widehat{\alpha}_{ogls} = (\widehat{\theta}_{ogls}, \widehat{f}_{ogls})$ as

$$
(\widehat{\theta}_{ogls}, \widehat{f}_{ogls}) = \arg \min_{\theta\in\Theta, f\in\mathcal{F}_n : ||\alpha-\alpha_o|| = o(n^{-1/4})} \frac{1}{n}\sum_{i=1}^{n} \rho(Z_i,\theta,f)'\{\widehat{\Sigma}_o(X_{1i},X_{2i})\}^{-1}\rho(Z_i,\theta,f). \tag{5}
$$

Let $\widehat{\Sigma}(X_1,X_2,\theta,f)$ be a consistent nonparametric estimator of $\Sigma(X_1,X_2,\theta,f)$. Then the *one-step continuously updated sieve GLS* estimator $\widehat{\alpha}_{cgls} = (\widehat{\theta}_{cgls}, \widehat{f}_{cgls})$ is given by:

$$
(\widehat{\theta}_{cgls}, \widehat{f}_{cgls}) = \arg \min_{\theta\in\Theta, f\in\mathcal{F}_n} \frac{1}{n}\sum_{i=1}^{n} \rho(Z_i,\theta,f)'\{\widehat{\Sigma}(X_{1i},X_{2i},\theta,f)\}^{-1}\rho(Z_i,\theta,f).
$$

By applying the general theory of Ai and Chen (2003), we can show that both $\widehat{\theta}_{ogls}$ and $\widehat{\theta}_{cgls}$ are $\sqrt{n}-$consistent, asymptotically normal and semiparametrically efficient for θ_o. However since the one-step continuously updated sieve GLS is computationally intensive, in this paper we only present sufficient conditions to ensure the semiparametric efficiency of $\widehat{\theta}_{ogls}$.

3.1.1 Fisher-like norm

For any direction $v = (v_\theta, v_f)$ with $v_\theta = (v'_{\beta_1}, v'_{\beta_2}, v_{\gamma_2}, v'_{\delta_2})'$ being a $d_\theta \times 1$-vector, v_{β_1} a $d_1 \times 1$-vector, v_{β_2} a $d_2 \times 1$-vector, v_{γ_2} a scalar, v_{δ_2} a $(d_2+1)\times 1$-vector, and v_f a scalar valued function in $\mathcal{F} - \{f_o\}$, we define the Fisher-like norm of v as

$$
||v||^2 = E\left(\left\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'}[v_\theta] + \frac{\partial\rho(Z,\alpha_o)}{\partial f}[v_f]\right\}' \{\Sigma_o(X_1,X_2)\}^{-1} \left\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'}[v_\theta] + \frac{\partial\rho(Z,\alpha_o)}{\partial f}[v_f]\right\}\right),
$$

where

$$\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'}[v_\theta] = \frac{\partial\rho(Z,\alpha_o)}{\partial\beta_1'}v_{\beta_1} + \frac{\partial\rho(Z,\alpha_o)}{\partial\beta_2'}v_{\beta_2} + \frac{\partial\rho(Z,\alpha_o)}{\partial\gamma_2}v_{\gamma_2} + \frac{\partial\rho(Z,\alpha_o)}{\partial\delta_2'}v_{\delta_2}$$

is a $3\times 1-$vector, and

$$\begin{aligned}
\frac{\partial\rho(Z,\alpha_o)}{\partial\beta_1'} &= \left(\frac{\partial\rho_1(Z,\alpha_o)}{\partial\beta_1'}, \frac{\partial\rho_2(Z,\alpha_o)}{\partial\beta_1'}, \frac{\partial\rho_3(Z,\alpha_o)}{\partial\beta_1'}\right)' \text{ is } 3\times d_1\text{-matrix},\\
\frac{\partial\rho_1(Z,\alpha_o)}{\partial\beta_1'} &= -X_{1,1}\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}} f_o(X_{0,1}+X_{1,1}\beta_{o1},u_2)du_2,\\
\frac{\partial\rho_2(Z,\alpha_o)}{\partial\beta_1'} &= X_{1,1}\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}} f_o(X_{0,1}+X_{1,1}\beta_{o1},u_2)du_2,\\
\frac{\partial\rho_3(Z,\alpha_o)}{\partial\beta_1'} &= -X_{1,1}\int_{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}}^{\infty} f_o(X_{0,1}+X_{1,1}\beta_{o1},u_2)du_2,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\rho(Z,\alpha_o)}{\partial\beta_2'} &= \left(\frac{\partial\rho_1(Z,\alpha_o)}{\partial\beta_2'}, \frac{\partial\rho_2(Z,\alpha_o)}{\partial\beta_2'}, \frac{\partial\rho_3(Z,\alpha_o)}{\partial\beta_2'}\right)' \text{ is } 3\times d_2\text{-matrix},\\
\frac{\partial\rho_1(Z,\alpha_o)}{\partial\beta_2'} &= -X_{1,2}\int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1,X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1 = X_{1,2}\frac{\partial\rho_1(Z,\alpha_o)}{\partial\gamma_2},\\
\frac{\partial\rho_2(Z,\alpha_o)}{\partial\beta_2'} &= -X_{1,2}\int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty} f_o(u_1,X_{0,2}+X_{1,2}\beta_{o2})du_1,\\
\frac{\partial\rho_3(Z,\alpha_o)}{\partial\beta_2'} &= -\frac{\partial\rho_1(Z,\alpha_o)}{\partial\beta_2'} = X_{1,2}\frac{\partial\rho_3(Z,\alpha_o)}{\partial\gamma_2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\rho(Z,\alpha_o)}{\partial\gamma_2} &= \left(\frac{\partial\rho_1(Z,\alpha_o)}{\partial\gamma_2}, \frac{\partial\rho_2(Z,\alpha_o)}{\partial\gamma_2}, \frac{\partial\rho_3(Z,\alpha_o)}{\partial\gamma_2}\right)' \text{ is } 3\times 1\text{-matrix},\\
\frac{\partial\rho_1(Z,\alpha_o)}{\partial\gamma_2} &= -\int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1,X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1,\\
\frac{\partial\rho_2(Z,\alpha_o)}{\partial\gamma_2} &= 0,\\
\frac{\partial\rho_3(Z,\alpha_o)}{\partial\gamma_2} &= -\frac{\partial\rho_1(Z,\alpha_o)}{\partial\gamma_2},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\rho(Z,\alpha_o)}{\partial\delta_2'} &= \left(\frac{\partial\rho_1(Z,\alpha_o)}{\partial\delta_2'}, \frac{\partial\rho_2(Z,\alpha_o)}{\partial\delta_2'}, \frac{\partial\rho_3(Z,\alpha_o)}{\partial\delta_2'}\right)' \text{ is } 3\times (d_2+1)\text{-matrix},\\
\frac{\partial\rho_1(Z,\alpha_o)}{\partial\delta_2'} &= -X_2\int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1,X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1 = X_2\frac{\partial\rho_1(Z,\alpha_o)}{\partial\gamma_2},\\
\frac{\partial\rho_2(Z,\alpha_o)}{\partial\delta_2'} &= 0,\\
\frac{\partial\rho_3(Z,\alpha_o)}{\partial\delta_2'} &= -\frac{\partial\rho_1(Z,\alpha_o)}{\partial\delta_2'} = X_2\frac{\partial\rho_3(Z,\alpha_o)}{\partial\gamma_2},
\end{aligned}$$

$$\frac{\partial\rho(Z,\alpha_o)}{\partial f}[v_f] = \left(\frac{\partial\rho_1(Z,\alpha_o)}{\partial f}[v_f], \frac{\partial\rho_2(Z,\alpha_o)}{\partial f}[v_f], \frac{\partial\rho_3(Z,\alpha_o)}{\partial f}[v_f]\right)' \text{ is } 3\times 1\text{-matrix},$$

$$\frac{\partial\rho_1(Z,\alpha_o)}{\partial f}[v_f] = -\int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}}\{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}} v_f(u_1,u_2)du_2\}du_1,$$

$$\frac{\partial\rho_2(Z,\alpha_o)}{\partial f}[v_f] = -\int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty}\{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}} v_f(u_1,u_2)du_2\}du_1,$$

$$\frac{\partial\rho_3(Z,\alpha_o)}{\partial f}[v_f] = -\int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}}\{\int_{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}}^{\infty} v_f(u_1,u_2)du_2\}du_1.$$

3.1.2 Convergence rate

Theorem 2 *Assume all the conditions for identification in the case $T=2$ are satisfies. Suppose further that the joint density of (U_1,U_2) is Hölder continuous with exponent greater than one. Then $||\widehat{\alpha}_{gls}-\alpha_o|| = o_p(n^{-1/4})$.*

3.1.3 Root-n normality and efficiency

For any $v=(v_\theta,v_f)$, let $v_f=-w\times v_\theta$ with $w=(w^1,...,w^{d_\theta})$ being a $1\times d_\theta$ -vector valued function. Then

$$||v||^2 = v_\theta' E\left(\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w]\}'\{\Sigma_o(X_1,X_2)\}^{-1}\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w]\}\right)v_\theta.$$

Let θ_j denote the $j-th$ element of the $d_\theta\times 1-$vector θ. We denote $w_o=(w_o^1,...,w_o^{d_\theta})$ be the solution to

$$\inf_{w^j} E\left(\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta_j} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w^j]\}'\{\Sigma_o(X_1,X_2)\}^{-1}\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta_j} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w^j]\}\right).$$

Denote $\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w_o]$ as the $3\times d_\theta-$matrix with the $j-th$ column given by $\frac{\partial\rho(Z,\alpha_o)}{\partial\theta_j} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w_o^j]$ for $j=1,...,d_\theta$. Finally we denote

$$V_o = E\left[\left\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w_o]\right\}'\{\Sigma_o(X_1,X_2)\}^{-1}\left\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w_o]\right\}'\right].$$

Then we can apply the general theory of Ai and Chen (2003) to establish

Theorem 3 *Assume all the conditions for identification in the case $T=2$ are satisfies. Suppose further that the joint density of (U_1,U_2) is holder continuous with exponent greater than one. Then the optimally weighted sieve GLS estimator $\widehat{\theta}_{ogls}$ given in (5) is root-n consistent, asymptotically normal and semiparametrically efficient, with asymptotic variance V_o^{-1}.*

3.2 Sieve ML Estimation

The log-likelihood evalued at $i-th$ observation is:

$$
\begin{aligned}
&\ell(Z_i,\theta,f)\\
=\ & Y_{1i}Y_{2i}\log\left\{\int_{-\infty}^{X_{0,1i}+X_{1,1i}\beta_1}\{\int_{-\infty}^{X_{0,2i}+X_{1,2i}\beta_2+\gamma_2+X_{2i}\delta_2} f(u_1,u_2)du_2\}du_1\right\}\\
&+(1-Y_{1i})Y_{2i}\log\left\{\int_{X_{0,1i}+X_{1,1i}\beta_1}^{\infty}\{\int_{-\infty}^{X_{0,2i}+X_{1,2i}\beta_2} f(u_1,u_2)du_2\}du_1\right\}\\
&+Y_{1i}(1-Y_{2i})\log\left\{\int_{-\infty}^{X_{0,1i}+X_{1,1i}\beta_1}\{\int_{X_{0,2i}+X_{1,2i}\beta_2+\gamma_2+X_{2i}\delta_2}^{\infty} f(u_1,u_2)du_2\}du_1\right\}\\
&+(1-Y_{1i})(1-Y_{2i})\log\left\{\int_{X_{0,1i}+X_{1,1i}\beta_1}^{\infty}\{\int_{X_{0,2i}+X_{1,2i}\beta_2}^{\infty} f(u_1,u_2)du_2\}du_1\right\}.
\end{aligned}
$$

Then the sieve ML estimator $\widehat{\alpha}_{ml}=(\widehat{\theta}_{ml},\widehat{f}_{ml})$ is the solution to:

$$
(\widehat{\theta}_{ml},\widehat{f}_{ml})=\arg\max_{\theta\in\Theta,f\in\mathcal{F}_n}\frac{1}{n}\sum_{i=1}^{n}\ell(Z_i,\theta,f),
$$

where $\mathcal{F}_n$ is a sieve space for $\mathcal{F}$.

3.2.1 Fisher norm

For any direction $v=(v_\theta,v_f)$ with $v_\theta=(v_{\beta_1},v_{\beta_2},v_{\gamma_2},v_{\delta_2})$, we define the Fisher norm as

$$
||v||^2=E\left(\left\{\frac{\partial\ell(Z,\alpha_o)}{\partial\theta'}[v_\theta]+\frac{\partial\ell(Z,\alpha_o)}{\partial f}[v_f]\right\}^2\right),
$$

where

$$
\frac{\partial\ell(Z,\alpha_o)}{\partial\theta'}[v_\theta]=\frac{\partial\ell(Z,\alpha_o)}{\partial\beta_1'}v_{\beta_1}+\frac{\partial\ell(Z,\alpha_o)}{\partial\beta_2'}v_{\beta_2}+\frac{\partial\ell(Z,\alpha_o)}{\partial\gamma_2}v_{\gamma_2}+\frac{\partial\ell(Z,\alpha_o)}{\partial\delta_2'}v_{\delta_2}
$$

and

$$
\begin{aligned}
&\frac{\partial\ell(Z,\alpha_o)}{\partial\beta_1'}\\
=\ &\frac{Y_1Y_2}{E[Y_1Y_2|X_1,X_2]}X_{1,1}\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}} f_o(X_{0,1}+X_{1,1}\beta_{o1},u_2)du_2\\
&-\frac{(1-Y_1)Y_2}{E[(1-Y_1)Y_2|X_1,X_2]}X_{1,1}\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}} f_o(X_{0,1}+X_{1,1}\beta_{o1},u_2)du_2\\
&+\frac{Y_1(1-Y_2)}{E[Y_1(1-Y_2)|X_1,X_2]}X_{1,1}\int_{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}}^{\infty} f_o(X_{0,1}+X_{1,1}\beta_{o1},u_2)du_2\\
&-\frac{(1-Y_1)(1-Y_2)}{E[(1-Y_1)(1-Y_2)|X_1,X_2]}X_{1,1}\int_{X_{0,2}+X_{1,2}\beta_{o2}}^{\infty} f_o(X_{0,1}+X_{1,1}\beta_{o1},u_2)du_2,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \ell(Z,\alpha_o)}{\partial \beta_2'} \\
= \; & \frac{Y_1Y_2}{E[Y_1Y_2|X_1,X_2]} X_{1,2} \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1, X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1 \\
&+\frac{(1-Y_1)Y_2}{E[(1-Y_1)Y_2|X_1,X_2]} X_{1,2} \int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty} f_o(u_1, X_{0,2}+X_{1,2}\beta_{o2})du_1 \\
&-\frac{Y_1(1-Y_2)}{E[Y_1(1-Y_2)|X_1,X_2]} X_{1,2} \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1, X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1 \\
&-\frac{(1-Y_1)(1-Y_2)}{E[(1-Y_1)(1-Y_2)|X_1,X_2]} X_{1,2} \int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty} f_o(u_1, X_{0,2}+X_{1,2}\beta_{o2})du_1,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \ell(Z,\alpha_o)}{\partial \gamma_2} \\
= \; & \frac{Y_1Y_2}{E[Y_1Y_2|X_1,X_2]} \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1, X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1 \\
&-\frac{Y_1(1-Y_2)}{E[Y_1(1-Y_2)|X_1,X_2]} \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1, X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1,
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \ell(Z,\alpha_o)}{\partial \delta_2'} \\
= \; & \frac{Y_1Y_2}{E[Y_1Y_2|X_1,X_2]} X_2 \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1, X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1 \\
&-\frac{Y_1(1-Y_2)}{E[Y_1(1-Y_2)|X_1,X_2]} X_2 \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} f_o(u_1, X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})du_1,
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\partial \ell(Z,\alpha_o)}{\partial f}[v_f] \\
= \; & \frac{Y_1Y_2}{E[Y_1Y_2|X_1,X_2]} \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} \{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}} v_f(u_1,u_2)du_2\}du_1 \\
&+\frac{(1-Y_1)Y_2}{E[(1-Y_1)Y_2|X_1,X_2]} \int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty} \{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}} v_f(u_1,u_2)du_2\}du_1 \\
&+\frac{Y_1(1-Y_2)}{E[Y_1(1-Y_2)|X_1,X_2]} \int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}} \{\int_{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}}^{\infty} v_f(u_1,u_2)du_2\}du_1 \\
&+\frac{(1-Y_1)(1-Y_2)}{E[(1-Y_1)(1-Y_2)|X_1,X_2]} \int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty} \{\int_{X_{0,2}+X_{1,2}\beta_{o2}}^{\infty} v_f(u_1,u_2)du_2\}du_1.
\end{aligned}
$$

3.2.2 Convergence rate

Theorem 4 *Assume all the conditions for identification in the case* $T = 2$ *are satisfies. Suppose further that the joint density of* (U_1, U_2) *is Hölder continuous with exponent greater than one. Then* $||\widehat{\alpha}_{ml} - \alpha_o|| = o_p(n^{-1/4})$.

3.2.3 Root-n normality and efficiency

For any $v = (v_\theta, v_f)$, let $v_f = -w \times v_\theta$ with $w = (w^1, ..., w^{d_\theta})$ being a $1 \times d_\theta$ -vector valued function. Let θ_j denote the $j-th$ element of the $d_\theta \times 1-$vector θ. We denote $w_* = (w_*^1, ..., w_*^{d_\theta})$ be the solution to

$$\inf_{w^j} E\left(\left\{\frac{\partial \ell(Z,\alpha_o)}{\partial \theta_j} - \frac{\partial \ell(Z,\alpha_o)}{\partial f}[w^j]\right\}^2\right).$$

Denote $\frac{\partial \ell(Z,\alpha_o)}{\partial \theta'} - \frac{\partial \ell(Z,\alpha_o)}{\partial f}[w_*]$ as the $1 \times d_\theta-$matrix with the $j-th$ column given by $\frac{\partial \ell(Z,\alpha_o)}{\partial \theta_j} - \frac{\partial \ell(Z,\alpha_o)}{\partial f}[w_*^j]$ for $j = 1, ..., d_\theta$. Finally we denote

$$I_* = E\left[\left\{\frac{\partial \ell(Z,\alpha_o)}{\partial \theta'} - \frac{\partial \ell(Z,\alpha_o)}{\partial f}[w_*]\right\}'\left\{\frac{\partial \ell(Z,\alpha_o)}{\partial \theta'} - \frac{\partial \ell(Z,\alpha_o)}{\partial f}[w_*]\right\}\right].$$

Then we can apply the general theory on sieve MLE in Shen (1997) to establish

Theorem 5 *Assume all the conditions for identification in the case* $T = 2$ *are satisfies. Suppose further that the joint density of* (U_1, U_2) *is holder continuous with exponent greater than one. Then the sieve ML estimator* $\widehat{\theta}_{ml}$ *is root-n consistent, asymptotically normal and semiparametrically efficient, with asymptotic variance* I_*^{-1}.

Remark: we can verify that actually $V_o = I_*$ hence the optimally weighted sieve GLS is as efficient as the sieve ML estimator.

3.3 Efficiency Gain by Imposing Parametric Copula

Denote f_{oj} as the true unknown marginal density and F_{oj} as the true unknown marginal cdf of U_j for $j = 1, 2$. By Sklar's (1959) theorem, there is a unique copula density function $c_o(e_1, e_2)$ such that $f_o(u_1, u_2) \equiv c_o(F_{o1}(u_1), F_{o2}(u_2)) \times f_{o1}(u_1)f_{o2}(u_2)$. In this subsection we impose a parametric structure on the copula function, which is equivalently to specify the joint density of (U_1, U_2) semiparametrically as

$$f_o(u_1, u_2) \equiv c(F_{o1}(u_1), F_{o2}(u_2); \lambda_o) \times f_{o1}(u_1)f_{o2}(u_2)$$

where the copula density $c()$ is of known functional form up to an unknown finite-dimensional parameter λ_o, but the marginal densities f_{oj} of U_j is unspecified for $j = 1, 2$.

We note that the model with $T = 2$ is now equivalent to the following set of conditional moment restrictions:

$$
\begin{aligned}
E\left[Y_1Y_2 - \int_0^{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}\{\int_0^{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})} c(e_1,e_2;\lambda_o)de_2\}de_1 \mid X_1,X_2\right] &= 0,\\
E\left[(1-Y_1)Y_2 - \int_{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}^{1}\{\int_0^{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2})} c(e_1,e_2;\lambda_o)de_2\}de_1 \mid X_1,X_2\right] &= 0,\\
E\left[Y_1(1-Y_2) - \int_0^{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}\{\int_{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})}^{1} c(e_1,e_2;\lambda_o)de_2\}de_1 \mid X_1,X_2\right] &= 0. \quad (6)
\end{aligned}
$$

In this subsection we let $\alpha = (\theta, f)$ with $\theta = (\beta_1', \beta_2', \gamma_2, \delta_2')' \in \Theta$ as before except that now $f \in \mathcal{F}^c$ with

$$
\mathcal{F}^c = \left\{ \begin{array}{l} f(u_1,u_2) = c(F_1(u_1), F_2(u_2); \lambda) \times f_1(u_1)f_2(u_2) : \lambda \in \Lambda, \\ f_1, f_2 \text{ are univariate densities that are Lip. continuous} \end{array} \right\}.
$$

Let $\rho(Z,\alpha) = (\rho_1(Z,\alpha), \rho_2(Z,\alpha), \rho_3(Z,\alpha))'$ be a $3 \times 1-$vector valued function with

$$
\begin{aligned}
\rho_1(Z,\alpha) &= Y_1Y_2 - \int_0^{F_1(X_{0,1}+X_{1,1}\beta_1)}\{\int_0^{F_2(X_{0,2}+X_{1,2}\beta_2+\gamma_2+X_2\delta_2)} c(e_1,e_2;\lambda)de_2\}de_1,\\
\rho_2(Z,\alpha) &= (1-Y_1)Y_2 - \int_{F_1(X_{0,1}+X_{1,1}\beta_1)}^{1}\{\int_0^{F_2(X_{0,2}+X_{1,2}\beta_2)} c(e_1,e_2;\lambda)de_2\}de_1,\\
\rho_3(Z,\alpha) &= Y_1(1-Y_2) - \int_0^{F_1(X_{0,1}+X_{1,1}\beta_1)}\{\int_{F_2(X_{0,2}+X_{1,2}\beta_2+\gamma_2+X_2\delta_2)}^{1} c(e_1,e_2;\lambda)de_2\}de_1.
\end{aligned}
$$

The three-step optimally weighted sieve GLS estimator $\widehat{\alpha}_{ogls} = (\widehat{\theta}_{ogls}, \widehat{f}_{ogls})$ can be obtained as follows:

Step 1: $(\widehat{\theta}_{gls}, \widehat{f}_{gls}) = \arg\min_{\theta\in\Theta, f\in\mathcal{F}_n^c} \frac{1}{n}\sum_{i=1}^n \rho(Z_i,\theta,f)'\rho(Z_i,\theta,f)$, where $\mathcal{F}_n^c$ is a sieve space for $\mathcal{F}^c$;

Step 2: compute a consistent estimator $\widehat{\Sigma}_o(X_1,X_2)$ for $\Sigma_o(X_1,X_2)$; We could again obtain a plug-in consistent estimator by using the $\Sigma_o(X_1,X_2)$ expression, the conditional moment restrictions (6) and the Step 1 estimates $(\widehat{\theta}_{gls}, \widehat{f}_{gls})$;

Step 3: compute the optimally weighted sieve GLS estimator $\widehat{\alpha}_{ogls} = (\widehat{\theta}_{ogls}, \widehat{f}_{ogls})$ as

$$
(\widehat{\theta}_{ogls}, \widehat{f}_{ogls}) = \arg \min_{\theta\in\Theta, f\in\mathcal{F}_n^c : ||\alpha-\alpha_o|| = o(n^{-1/4})} \frac{1}{n}\sum_{i=1}^n \rho(Z_i,\theta,f)'\{\widehat{\Sigma}_o(X_{1i},X_{2i})\}^{-1}\rho(Z_i,\theta,f). \quad (7)
$$

For any direction $v = (v_\theta, v_f)$ with $v_\theta = (v_{\beta_1}', v_{\beta_2}', v_{\gamma_2}, v_{\delta_2}')'$ being a $d_\theta \times 1$-vector, v_{β_1} a $d_1 \times 1$-vector, v_{β_2} a $d_2 \times 1$-vector, v_{γ_2} a scalar, v_{δ_2} a $(d_2+1)\times 1$-vector, and v_f a scalar valued function in $\mathcal{F}^c - \{f_o\}$, we define the Fisher-like norm of v as

$$
||v||^2 = E\left(\left\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'}[v_\theta] + \frac{\partial\rho(Z,\alpha_o)}{\partial f}[v_f]\right\}'\{\Sigma_o(X_1,X_2)\}^{-1}\left\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'}[v_\theta] + \frac{\partial\rho(Z,\alpha_o)}{\partial f}[v_f]\right\}\right),
$$

where

$$\frac{\partial \rho(Z,\alpha_o)}{\partial \theta'}[v_\theta] = \frac{\partial \rho(Z,\alpha_o)}{\partial \beta_1'}v_{\beta_1} + \frac{\partial \rho(Z,\alpha_o)}{\partial \beta_2'}v_{\beta_2} + \frac{\partial \rho(Z,\alpha_o)}{\partial \gamma_2}v_{\gamma_2} + \frac{\partial \rho(Z,\alpha_o)}{\partial \delta_2'}v_{\delta_2}$$

is a $3\times 1-$vector, and

$$\begin{aligned}
\frac{\partial \rho(Z,\alpha_o)}{\partial \beta_1'} &= \left(\frac{\partial \rho_1(Z,\alpha_o)}{\partial \beta_1'}, \frac{\partial \rho_2(Z,\alpha_o)}{\partial \beta_1'}, \frac{\partial \rho_3(Z,\alpha_o)}{\partial \beta_1'}\right)' \text{ is } 3\times d_1\text{-matrix,}\\
\frac{\partial \rho_1(Z,\alpha_o)}{\partial \beta_1'} &= -X_{1,1}f_{o1}(X_{0,1}+X_{1,1}\beta_{o1})\int_0^{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})} c(F_{o1}(X_{0,1}+X_{1,1}\beta_{o1}), e_2; \lambda_o)\\
\frac{\partial \rho_2(Z,\alpha_o)}{\partial \beta_1'} &= X_{1,1}f_{o1}(X_{0,1}+X_{1,1}\beta_{o1})\int_0^{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2})} c(F_{o1}(X_{0,1}+X_{1,1}\beta_{o1}), e_2; \lambda_o)de_2,\\
\frac{\partial \rho_3(Z,\alpha_o)}{\partial \beta_1'} &= -X_{1,1}f_{o1}(X_{0,1}+X_{1,1}\beta_{o1})\int_{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})}^1 c(F_{o1}(X_{0,1}+X_{1,1}\beta_{o1}), e_2; \lambda_o)d
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \rho(Z,\alpha_o)}{\partial \beta_2'} &= \left(\frac{\partial \rho_1(Z,\alpha_o)}{\partial \beta_2'}, \frac{\partial \rho_2(Z,\alpha_o)}{\partial \beta_2'}, \frac{\partial \rho_3(Z,\alpha_o)}{\partial \beta_2'}\right)' \text{ is } 3\times d_2\text{-matrix,}\\
\frac{\partial \rho_1(Z,\alpha_o)}{\partial \beta_2'} &= X_{1,2}\frac{\partial \rho_1(Z,\alpha_o)}{\partial \gamma_2},\\
\frac{\partial \rho_2(Z,\alpha_o)}{\partial \beta_2'} &= -X_{1,2}f_{o2}(X_{0,2}+X_{1,2}\beta_{o2})\int_{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}^1 c(e_1, F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}))de_1,\\
\frac{\partial \rho_3(Z,\alpha_o)}{\partial \beta_2'} &= X_{1,2}\frac{\partial \rho_3(Z,\alpha_o)}{\partial \gamma_2} = -\frac{\partial \rho_1(Z,\alpha_o)}{\partial \beta_2'},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \rho(Z,\alpha_o)}{\partial \gamma_2} &= \left(\frac{\partial \rho_1(Z,\alpha_o)}{\partial \gamma_2}, \frac{\partial \rho_2(Z,\alpha_o)}{\partial \gamma_2}, \frac{\partial \rho_3(Z,\alpha_o)}{\partial \gamma_2}\right)' \text{ is } 3\times 1\text{-matrix,}\\
\frac{\partial \rho_1(Z,\alpha_o)}{\partial \gamma_2} &= -f_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})\times\\
&\quad \int_0^{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})} c(e_1, F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}); \lambda_o)de_1,\\
\frac{\partial \rho_2(Z,\alpha_o)}{\partial \gamma_2} &= 0,\\
\frac{\partial \rho_3(Z,\alpha_o)}{\partial \gamma_2} &= -\frac{\partial \rho_1(Z,\alpha_o)}{\partial \gamma_2},
\end{aligned}$$

$$\frac{\partial\rho(Z,\alpha_o)}{\partial\delta_2'} = \left(\frac{\partial\rho_1(Z,\alpha_o)}{\partial\delta_2'}, \frac{\partial\rho_2(Z,\alpha_o)}{\partial\delta_2'}, \frac{\partial\rho_3(Z,\alpha_o)}{\partial\delta_2'}\right)' \text{ is } 3\times(d_2+1)\text{-matrix},$$

$$\frac{\partial\rho_1(Z,\alpha_o)}{\partial\delta_2'} = X_2\frac{\partial\rho_1(Z,\alpha_o)}{\partial\gamma_2},$$

$$\frac{\partial\rho_2(Z,\alpha_o)}{\partial\delta_2'} = 0,$$

$$\frac{\partial\rho_3(Z,\alpha_o)}{\partial\delta_2'} = X_2\frac{\partial\rho_3(Z,\alpha_o)}{\partial\gamma_2} = -\frac{\partial\rho_1(Z,\alpha_o)}{\partial\delta_2'},$$

$$\frac{\partial\rho(Z,\alpha_o)}{\partial f}[v_f] = \left(\frac{\partial\rho_1(Z,\alpha_o)}{\partial f}[v_f], \frac{\partial\rho_2(Z,\alpha_o)}{\partial f}[v_f], \frac{\partial\rho_3(Z,\alpha_o)}{\partial f}[v_f]\right)' \text{ is } 3\times 1\text{-matrix},$$

$$\frac{\partial\rho_1(Z,\alpha_o)}{\partial f}[v_f] = -\int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}}\{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}} v_f(u_1,u_2)du_2\}du_1,$$

$$\frac{\partial\rho_2(Z,\alpha_o)}{\partial f}[v_f] = -\int_{X_{0,1}+X_{1,1}\beta_{o1}}^{\infty}\{\int_{-\infty}^{X_{0,2}+X_{1,2}\beta_{o2}} v_f(u_1,u_2)du_2\}du_1,$$

$$\frac{\partial\rho_3(Z,\alpha_o)}{\partial f}[v_f] = -\int_{-\infty}^{X_{0,1}+X_{1,1}\beta_{o1}}\{\int_{X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2}}^{\infty} v_f(u_1,u_2)du_2\}du_1.$$

For any $v=(v_\theta,v_f)$, let $v_f=-w\times v_\theta$ with $w=(w^1,...,w^{d_\theta})$ being a $1\times d_\theta$ -vector valued function. Then

$$||v||^2 = v_\theta' E\left(\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w]\}'\{\Sigma_o(X_1,X_2)\}^{-1}\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w]\}\right)v_\theta.$$

Let θ_j denote the $j-th$ element of the $d_\theta\times 1-$vector θ. We denote $w_o=(w_o^1,...,w_o^{d_\theta})$ be the solution to

$$\inf_{w^j} E\left(\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta_j} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w^j]\}'\{\Sigma_o(X_1,X_2)\}^{-1}\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta_j} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w^j]\}\right).$$

Denote $\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w_o]$ as the $3\times d_\theta-$matrix with the $j-th$ column given by $\frac{\partial\rho(Z,\alpha_o)}{\partial\theta_j} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w_o^j]$ for $j=1,...,d_\theta$. Finally we denote

$$V_{oc} = E\left[\left\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w_o]\right\}'\{\Sigma_o(X_1,X_2)\}^{-1}\left\{\frac{\partial\rho(Z,\alpha_o)}{\partial\theta'} - \frac{\partial\rho(Z,\alpha_o)}{\partial f}[w_o]\right\}'\right].$$

Then we can apply the general theory of Ai and Chen (2003) to establish

Theorem 6 *Assume all the conditions for identification in the case $T=2$ are satisfies. Suppose further that the joint density of (U_1,U_2) satisfies $f_o(u_1,u_2)\equiv c(F_{o1}(u_1),F_{o2}(u_2);\lambda_o)f_{o1}(u_1)f_{o2}($ and the marginal densities $f_{o1}(u_1)$, $f_{o2}(u_2)$ are holder continuous with exponent greater than 1/2. Then the optimally weighted sieve GLS estimator $\widehat{\theta}_{ogls}$ given in (7) is root-n consistent, asymptotically normal and semiparametrically efficient, with asymptotic variance V_{oc}^{-1}.*

Remark: Notice that $\mathcal{F}^c \subset \mathcal{F}$, we have $V_{oc} \geq V_o$. Hence $V_{oc}^{-1} \leq V_o^{-1}$, that is, by imposing parametric copula dependence structure among (U_1, U_2) will lead to a more efficient estimation of $\theta_o = (\beta'_{o1}, \beta'_{o2}, \gamma_{o2}, \delta'_{o2})'$.

4 Semiparametric Efficient Estimation for Greater than Two Periods

When $T \geq 3$, although we could still identify the joint density $f_{U_1,...,U_T}$ of $(U_1, ..., U_T)$, due to the well-known "curse of dimensionality", we shall impose some parametric dependence structure among $U_1, ..., U_T$ while still leave the marginal densities $f_{U_t}, t = 1, ..., T$ unspecified. By Sklar's (1959) theorem, we have

$$f_{U_1,...,U_T}(u_1, ..., u_T) \equiv c(F_{U_1}(U_1), ..., F_{U_T}(U_T)) \prod_{t=1}^{T} f_{U_t}(u_t)$$

where $c(F_{U_1}(U_1), ..., F_{U_T}(U_T))$ is the copula function associated with $(U_1, ..., U_T)$. In the following we denote F_{oj} as the true unknown marginal cdf of U_j for $j = 1, ..., T$. To avoid the curse of dimensionality, we impose a parametric structure on the copula function, which is equivalently to specify the joint density of $(U_1, ..., U_T)$ semiparametrically as

$$f_{U_1,...,U_T}(u_1, ..., u_T) = c(F_{o1}(U_1), ..., F_{oT}(U_T); \lambda_o) \prod_{t=1}^{T} f_{ot}(u_t) \tag{8}$$

where the copula density $c()$ is of known functional form up to the finite-dimensional unknown parameter λ_o and the marginal densities f_{ot} of U_t is unspecified for $t = 1, ..., T$.

In this paper, $c(F_{o1}(U_1), ..., F_{oT}(U_T); \lambda_o)$ can be any parametric copula density function[4] such as the Normal (or Gaussian) copula, the Student's t-copula, the Frank copula, the Gumble copula, and the Clayton copula, see Joe (1997) and Nelsen (1999) for examples and properties of copulas. Here we present three copula examples:

Example 1 (Normal copula): Let Φ denote the scalar standard normal distribution, and $\Phi_{\Sigma,T}$ the $T-$dimensional normal distribution with correlation matrix Σ. Then the $T-$dimensional normal copula with correlation matrix $\lambda = \Sigma$ is

$$C\left(\mathbf{e};\Sigma\right) = \Phi_{\Sigma,T}(\Phi^{-1}(e_1), ..., \Phi^{-1}(e_T)),$$

whose copula density is

$$c\left(\mathbf{e};\Sigma\right) = \frac{1}{\sqrt{\det(\Sigma)}} \exp\left\{ -\frac{\left(\Phi^{-1}(e_1), ..., \Phi^{-1}(e_T)\right)' \left(\Sigma^{-1} - I_T\right) \left(\Phi^{-1}(e_1), ..., \Phi^{-1}(e_T)\right)}{2} \right\}.$$

[4]See Chen and Fan (2003) for model selection among multiple possibly misspecified parametric copula models.

Normal copula with $\Sigma \neq 0$ generates joint symmetric dependence, but there is no tail dependence (i.e., there is no joint extreme events).

Example 2 (Student's t-copula): Let $\mathcal{T}_\nu$ be the scalar standard Student's t distribution with $\nu > 2$ degrees of freedom, and $\mathcal{T}_{\Sigma,\nu}$ be the $T-$dimensional Student's t distribution with $\nu > 2$ degrees of freedom and a shape matrix Σ. Then the $T-$dimensional Student's t-copula with parameters $\lambda = (\Sigma, \nu)$ is

$$C(\mathbf{e};\Sigma,\nu) = \mathcal{T}_{\Sigma,\nu}(\mathcal{T}_\nu^{-1}(u_1), ..., \mathcal{T}_\nu^{-1}(u_T)).$$

The Student's t copula density is:

$$\begin{aligned} c(\mathbf{e};\Sigma,\nu) &= \frac{\Gamma(\frac{\nu+T}{2})[\Gamma(\frac{\nu}{2})]^{T-1}}{\sqrt{\det(\Sigma)}[\Gamma(\frac{\nu+1}{2})]^T}\left(1+\frac{\mathbf{x}'\Sigma^{-1}\mathbf{x}}{\nu}\right)^{-\frac{\nu+T}{2}}\prod_{i=1}^{T}\left(1+\frac{x_i^2}{\nu}\right)^{\frac{\nu+1}{2}}, \\ \text{where } \mathbf{x} &= (x_1, ..., x_T)', \quad x_i = \mathcal{T}_\nu^{-1}(u_i). \end{aligned}$$

The Student's t copula with $\Sigma \neq 0$ can generate joint symmetric tail dependence, hence allow for *joint* fat tails (i.e., an increased probability of joint extreme events).

Example 3 (Clayton copula): the $T-$dimensional Clayton copula with parameter $\lambda > 0$ is:

$$C(\mathbf{e};\lambda) = [e_1^{-\lambda} + \ldots + e_T^{-\lambda} - T + 1]^{-1/\lambda}, \quad \text{where } \lambda > 0. \tag{9}$$

The copula density of the Clayton copula is given by

$$c(\mathbf{e};\lambda) = \{\Pi_{j=1}^T[1+(j-1)\lambda]\}\{\Pi_{j=1}^T u_j^{-(\lambda+1)}\}[\sum_{j=1}^T u_j^{-\lambda} - T + 1]^{-(\lambda^{-1}+T)}, \quad \text{where } \lambda > 0.$$

Unlike the Gaussian and Student's t copulas, the Clayton copula can generate asymmetric dependence. In particular, the Clayton copula has lower tail dependence, but no upper tail dependence.

In the following we only consider the estimation of the model when $T = 3$, since the general case $T > 3$ is notationally very tidious. (Recall that we have $2^T - 1$ number of conditional moment restrictions for a model with T periods.) Let $X = (X_1', X_2', X_3')'$, then we have the following 7 conditional moment restrictions implied by the model with $T = 3$ and imposing the parametric copula structure (8).

$$\begin{aligned} & E[Y_1Y_2Y_3|X] \\ = & \int_0^{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}\int_0^{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})} \times \\ & \int_0^{F_{o3}(X_{0,3}+X_{1,3}\beta_{o3}+\gamma_{o3}^2+X_3\delta_{o3}^2+\gamma_{o3}^1+X_3\delta_{o3}^1+\gamma_{o3}^{12}+X_3\delta_{o3}^{12})} c(e_1,e_2,e_3;\lambda_o)de_3de_2de_1 \end{aligned}$$

$$E[(1-Y_1)Y_2Y_3|X] = \int_{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}^{1}\int_0^{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2})}\int_0^{F_{o3}(X_{0,3}+X_{1,3}\beta_{o3}+\gamma_{o3}^2+X_3\delta_{o3}^2)} c(e_1,e_2,e_3;\lambda_o)de_3de_2de_1$$

$$E[Y_1(1-Y_2)Y_3|X] = \int_0^{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}\int_{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})}^{1}\int_0^{F_{o3}(X_{0,3}+X_{1,3}\beta_{o3}+\gamma_{o3}^1+X_3\delta_{o3}^1)} c(e_1,e_2,e_3;\lambda_o)de_3de_2de_1$$

$$E[(1-Y_1)(1-Y_2)Y_3|X] = \int_{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}^{1}\int_{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2})}^{1}\int_0^{F_{o3}(X_{0,3}+X_{1,3}\beta_{o3})} c(e_1,e_2,e_3;\lambda_o)de_3de_2de_1$$

$$E[Y_1Y_2(1-Y_3)|X] = \int_0^{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}\int_0^{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})} \times \int_{F_{o3}(X_{0,3}+X_{1,3}\beta_{o3}+\gamma_{o3}^2+X_3\delta_{o3}^2+\gamma_{o3}^1+X_3\delta_{o3}^1+\gamma_{o3}^{12}+X_3\delta_{o3}^{12})}^{1} c(e_1,e_2,e_3;\lambda_o)de_3de_2de_1$$

$$E[(1-Y_1)Y_2(1-Y_3)|X] = \int_{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}^{1}\int_0^{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2})}\int_{F_{o3}(X_{0,3}+X_{1,3}\beta_{o3}+\gamma_{o3}^2+X_3\delta_{o3}^2)}^{1} c(e_1,e_2,e_3;\lambda_o)de_3de_2de_1$$

$$E[Y_1(1-Y_2)(1-Y_3)|X] = \int_0^{F_{o1}(X_{0,1}+X_{1,1}\beta_{o1})}\int_{F_{o2}(X_{0,2}+X_{1,2}\beta_{o2}+\gamma_{o2}+X_2\delta_{o2})}^{1}\int_{F_{o3}(X_{0,3}+X_{1,3}\beta_{o3}+\gamma_{o3}^1+X_3\delta_{o3}^1)}^{1} c(e_1,e_2,e_3;\lambda_o)de_3de_2de_1$$

In the following we denote $\alpha = (\theta, f)$ with $\theta' = \left(\beta_1', \beta_2', \gamma_2, \delta_2', \beta_3', \gamma_3^2, \delta_3^{2\prime}, \gamma_3^1, \delta_3^{1\prime}, \gamma_3^{12}, \delta_3^{12\prime}\right) \in \Theta$ and $f = (F_1, F_2, F_3, \lambda) \in \mathcal{F}^c$,

$$\mathcal{F}^c = \left\{ \begin{array}{c} f(u_1,u_2,u_3) = c(F_1(u_1), F_2(u_2), F_3(u_3); \lambda) f_1(u_1) f_2(u_2) f_3(u_3) : \lambda \in \Lambda, \\ f_1, f_2, f_3 \text{ are univariate densities that are Lip. continuous.} \end{array} \right\}.$$

Also denote $\alpha_o = (\theta_o, f_o)$. Let $\rho(Z,\alpha) = (\rho_1(Z,\alpha), ..., \rho_7(Z,\alpha))'$ be a $7 \times 1-$vector valued function with

$$\rho_1(Z,\alpha) = Y_1Y_2Y_3 - \int_0^{F_1(X_{0,1}+X_{1,1}\beta_1)}\int_0^{F_2(X_{0,2}+X_{1,2}\beta_2+\gamma_2+X_2\delta_2)} \times \int_0^{F_3(X_{0,3}+X_{1,3}\beta_3+\gamma_3^2+X_3\delta_3^2+\gamma_3^1+X_3\delta_3^1+\gamma_3^{12}+X_3\delta_3^{12})} c(e_1,e_2,e_3;\lambda)de_3de_2de_1$$

$$\rho_2(Z,\alpha) = (1-Y_1)Y_2Y_3 - \int_{F_1(X_{0,1}+X_{1,1}\beta_1)}^{1} \int_0^{F_2(X_{0,2}+X_{1,2}\beta_2)} \times \int_0^{F_3(X_{0,3}+X_{1,3}\beta_3+\gamma_3^2+X_3\delta_3^2)} c(e_1,e_2,e_3;\lambda)de_3de_2de_1$$

$$\rho_3(Z,\alpha) = Y_1(1-Y_2)Y_3 - \int_0^{F_1(X_{0,1}+X_{1,1}\beta_1)} \int_{F_2(X_{0,2}+X_{1,2}\beta_2+\gamma_2+X_2\delta_2)}^{1} \times \int_0^{F_3(X_{0,3}+X_{1,3}\beta_3+\gamma_3^1+X_3\delta_3^1)} c(e_1,e_2,e_3;\lambda)de_3de_2de_1$$

$$\rho_4(Z,\alpha) = (1-Y_1)(1-Y_2)Y_3 - \int_{F_1(X_{0,1}+X_{1,1}\beta_1)}^{1} \int_{F_2(X_{0,2}+X_{1,2}\beta_2)}^{1} \times \int_0^{F_3(X_{0,3}+X_{1,3}\beta_3)} c(e_1,e_2,e_3;\lambda)de_3de_2de_1$$

$$\rho_5(Z,\alpha) = Y_1Y_2(1-Y_3) - \int_0^{F_1(X_{0,1}+X_{1,1}\beta_1)} \int_0^{F_2(X_{0,2}+X_{1,2}\beta_2+\gamma_2+X_2\delta_2)} \times \int_{F_3(X_{0,3}+X_{1,3}\beta_3+\gamma_3^2+X_3\delta_3^2+\gamma_3^1+X_3\delta_3^1+\gamma_3^{12}+X_3\delta_3^{12})}^{1} c(e_1,e_2,e_3;\lambda)de_3de_2de_1$$

$$\rho_6(Z,\alpha) = (1-Y_1)Y_2(1-Y_3) - \int_{F_1(X_{0,1}+X_{1,1}\beta_1)}^{1} \int_0^{F_2(X_{0,2}+X_{1,2}\beta_2)} \times \int_{F_3(X_{0,3}+X_{1,3}\beta_3+\gamma_3^2+X_3\delta_3^2)}^{1} c(e_1,e_2,e_3;\lambda)de_3de_2de_1$$

$$\rho_7(Z,\alpha) = Y_1(1-Y_2)(1-Y_3) - \int_0^{F_1(X_{0,1}+X_{1,1}\beta_1)} \int_{F_2(X_{0,2}+X_{1,2}\beta_2+\gamma_2+X_2\delta_2)}^{1} \times \int_{F_3(X_{0,3}+X_{1,3}\beta_3+\gamma_3^1+X_3\delta_3^1)}^{1} c(e_1,e_2,e_3;\lambda)de_3de_2de_1$$

The three-step optimally weighted sieve GLS estimator $\widehat{\alpha}_{ogls} = (\widehat{\theta}_{ogls}, \widehat{f}_{ogls})$ can be obtained as follows:

Step 1: $(\widehat{\theta}_{gls}, \widehat{f}_{gls}) = \arg\min_{\theta\in\Theta, f\in\mathcal{F}_n^c} \frac{1}{n}\sum_{i=1}^n \rho(Z_i,\theta,f)'\rho(Z_i,\theta,f)$, where $\mathcal{F}_n^c$ is a sieve space for $\mathcal{F}^c$;

Step 2: compute a consistent estimator $\widehat{\Sigma}_o(X)$ for $\Sigma_o(X) = Var\{\rho(Z_i,\theta_o,f_o)|X\}$; We could again obtain a plug-in consistent estimator by using the $\Sigma_o(X_1)$ expression, the above 7 conditional moment restrictions and the Step 1 estimates $(\widehat{\theta}_{gls}, \widehat{f}_{gls})$;

Step 3: compute the optimally weighted sieve GLS estimator $\widehat{\alpha}_{ogls} = (\widehat{\theta}_{ogls}, \widehat{f}_{ogls})$ as

$$(\widehat{\theta}_{ogls}, \widehat{f}_{ogls}) = \arg \min_{\theta \in \Theta, f \in \mathcal{F}_n^c : ||\alpha - \alpha_o|| = o(n^{-1/4})} \frac{1}{n} \sum_{i=1}^{n} \rho(Z_i, \theta, f)' \{\widehat{\Sigma}_o(X_{1i})\}^{-1} \rho(Z_i, \theta, f).$$

Finally we denote

$$V_{oc} = E\left[\left\{\frac{\partial \rho(Z, \alpha_o)}{\partial \theta'} - \frac{\partial \rho(Z, \alpha_o)}{\partial f}[w_o]\right\}' \{\Sigma_o(X)\}^{-1} \left\{\frac{\partial \rho(Z, \alpha_o)}{\partial \theta'} - \frac{\partial \rho(Z, \alpha_o)}{\partial f}[w_o]\right\}'\right].$$

Then we can apply the general theory of Ai and Chen (2003) to establish

Theorem 7 *Assume all the conditions for identification in the case $T = 3$ are satisfies. Suppose further that the joint density of (U_1, U_2, U_3) satisfies*

$$f_o(u_1, u_2, u_3) \equiv c(F_{o1}(u_1), F_{o2}(u_2), F_{o3}(u_3); \lambda_o) f_{o1}(u_1) f_{o2}(u_2) f_{o3}(u_3),$$

and the marginal densities $f_{oj}(u_j)$, $j = 1, 2, 3$ are holder continuous with exponent greater than 1/2. Then the optimally weighted sieve GLS estimator $\widehat{\theta}_{ogls}$ is root-n consistent, asymptotically normal and semiparametrically efficient, with asymptotic variance V_{oc}^{-1}.