Exclusion Restrictions in Dynamic Binary Choice Panel Data Models

Songnian Chen — HKUST

Shakeeb Khan — Boston College

Xun Tang — Rice University

January 26, 2018

[1]

Keywords: Panel Data, Dynamic Binary Choice, Exclusion Restriction

[1] We are grateful to Bo Honore, Arthur Lewbel, Jim Powell, Elie Tamer for helpful feedback. We also thank conference and seminar participants at Cambridge, CEMFI, China ES Meetings, Cornell, CREST, Northwestern, Panel Data Workshop (U Amsterdam, 2015), Pompeu Fabra, SETA (Taiwan), Texas Econometrics Camp (Houston, 2016), Tilburg, UCL, UT Austin, World Congress ES Meetings (Montreal, 2015), and Yale for comments. We thank Yichong Zhang and Michelle Tyler for capable research assistance. The usual disclaimer applies.

1 Introduction

We revisit the use of *exclusion restrictions* in the semiparametric binary choice panel data model with predetermined regressors introduced in Honoré and Lewbel (2002). The identification strategy in Honoré and Lewbel (2002) requires an exclusion restriction (Assumption A.2) that one of the explanatory variables (which we refer to as an "excluded regressor" henceforth), is independent of the individual fixed effect and time-varying idiosyncratic errors, conditional on the other regressors. Their model is presented in a general framework where the explanatory variables are predetermined, e.g., including lagged dependent variables. As explained in Honore and Kyriazidou (2000), such models allow for both "true state dependence" in addition to unobserved heterogeneity.[1] Without such an exclusion restriction, identification and inference in *dynamic* binary choice panel data models (where one of the predetermined explanatory variables is the lagged dependent variable) is complicated and non-standard, as shown in Honore and Kyriazidou (2000) and Hahn (2001). Thus the introduction of exclusion restriction into the model by Honoré and Lewbel (2002) is well motivated.

However, here we show that in a dynamic binary choice panel data model, the exclusion restriction in Honoré and Lewbel (2002) implicitly requires (conditional) *serial independence* of the excluded regressor mentioned above. If such serial independence is violated, then the main identifying condition (Assumption A.2) does not hold in general, and the inverse-density-weighted estimator in Honoré and Lewbel (2002) is generally *inconsistent*. We propose a new identification strategy and estimation method for this semiparametric binary choice panel data model under exclusion restrictions which can accommodate serial dependence of excluded regressors in a dynamic setting. The new estimator converges at the parametric rate to a limiting normal distribution. This rate is faster than the nonparametric rates of existing alternative estimators for the binary choice panel data model, including the static case in Manski (1987) and the dynamic case in Honore and Kyriazidou (2000), both of which do not impose the exclusion restrictions mentioned above.

To develop the intuition for how the exclusion restriction in Assumption A.2 of Honoré and Lewbel (2002) could fail in a dynamic setting, consider the binary choice panel data model:

$$y_{it} = I[v_{it} + x_{it}'\beta_0 + \alpha_i + \epsilon_{it} \geq 0] \tag{1.1}$$

where $i = 1, 2, ..., n$, and $t = 1, 2, ..., T$. Here $I[\cdot]$ is the indicator function that equals one if "$\cdot$" is true and zero otherwise, $v_{it} \in \mathbb{R}$ is the excluded regressor whose coefficient that is normalized to one, x_{it} is a vector of other regressors (possibly predetermined), β_0 is a vector of coefficients, α_i is an individual-specific fixed effect, and the distribution of ϵ_{it} is unknown.

Honoré and Lewbel (2002) estimate this model under an exclusion restriction that $e_{it} \equiv \alpha_i + \epsilon_{it}$ is independent of v_{it}, conditional on x_{it}. In cross-sectional models with no individual-specific fixed effects α_i, such an exclusion restriction has proven useful. Examples include

[1] See also Heckman (1991) for a more detailed discussion on this matter.

Lewbel (2000), Lewbel and Tang (2015) and Chen, Khan, and Tang (2016). However, in a dynamic panel data model, one of the components in x_{it} is the lagged dependent variable y_{it-1}, which itself is a function of α_i, ϵ_{it-1} and v_{it-1}. As a result, serial correlation in v_{it} leads to a complex dependence structure between y_{it-1}, v_{it-1} and $\alpha_i + \epsilon_{it}$, which is generally not compatible with the exclusion restriction (Assumption A.2) in Honoré and Lewbel (2002). This is true even if v_{it} is independent of α_i and ϵ_{it} conditional on the other components in x_{it} in each period.

The rest of the note is organized as follows. Section 2 uses a simplified version of (1.1) to formalize the intuition about why the exclusion restriction in Honoré and Lewbel (2002) does not hold in general when v_{it} is serially correlated. Sections 3 and 4 introduce our new approach for estimating dynamic binary choice panel data models, based on a pairwise comparison approach that allows for serial correlation in the excluded regressor. Section 5 presents simulation evidence. Section 6 concludes. Technical proofs are collected in the appendix.

2 A Simplified Model

To illustrate our main idea, consider a simplified version of the model in (1.1) with two periods following the initial condition y_{i0} ($T = 2$) and only two explanatory variables, which consist of an excluded regressor $v_{it} \in \mathbb{R}$ and the lagged dependent variable y_{it-1}:

$$y_{it} = I[v_{it} + y_{it-1}\gamma_0 + \alpha_i + \epsilon_{it} \geq 0] \text{ for } t = 1, 2, \tag{2.1}$$

where the parameter of interest is γ_0. This specification is subsumed by the original model in (1.1) with $x_{it} \equiv y_{it-1}$. Honoré and Lewbel (2002) maintain the following exclusion restriction (on p.2055) on the model in (1.1) for identification and estimation:

ASSUMPTION A.2: For $t = 1, 2$, $\alpha_i + \epsilon_{it}$ is independent of v_{it} conditional on x_{it} and z_i.

Note that Honoré and Lewbel (2002) states this assumption by conditioning on an instrument z_i, which may overlap with the exogenous variables. Likewise, our argument and results throughout the current paper are also valid conditional on any instruments available. In what follows, we suppress such instruments in the conditioning events to lighten the notation.

For simplicity, suppose that the initial value y_{i0} is degenerate at 0 in the data-generating process. Assumption A.2 in Honoré and Lewbel (2002) requires

$$(\alpha_i + \epsilon_{it}) \perp v_{it} \text{ conditional on } y_{it-1} \text{ for } t = 1, 2. \tag{2.2}$$

We show that, if v_{i1}, v_{i2} are serially correlated, then (2.2) does not hold in general even when $(\alpha_i + \epsilon_{i1}, \alpha_i + \epsilon_{i2})$ is independent of (v_{i1}, v_{i2}).

To simplify notation, we drop the subscript i for all random variables, and let $e_t \equiv -(\alpha + \varepsilon_t)$ for $t = 1, 2$. Assume that (e_1, e_2) is independent of (v_1, v_2); and that the joint distribution of (e_1, e_2) and the joint distribution of (v_1, v_2) are both exchangeable in the index $t \in \{1, 2\}$.[2] Let F (and f) denote the marginal distribution (and density) of v_t; let G (and g) denote the marginal distribution (and density) of e_t. Define $F(s'|s) \equiv \Pr(v_1 \leq s'|v_2 = s)$ and $G(r'|r) \equiv \Pr(e_1 \leq r'|e_2 = r)$. That F, G do not vary with $t = 1, 2$ is a consequence of the exchangeability condition.

We will show that e_2 is not independent of v_2 conditional on $y_1 = 1$ (or equivalently, $v_1 - e_1 \geq 0$) if v_t is serially correlated between $t = 1, 2$. By definition,

$$\begin{aligned}
&\left.\frac{\partial^2 \Pr(e_2 \leq \tilde{r}, v_2 \leq \tilde{s}|y_1 = 1)}{\partial\tilde{r}\partial\tilde{s}}\right|_{\tilde{r}=r,\tilde{s}=s} \\
=\ & \left.\frac{\partial^2}{\partial\tilde{r}\partial\tilde{s}}\left(\frac{\Pr(v_1 - e_1 \geq 0, e_2 \leq \tilde{r}, v_2 \leq \tilde{s})}{\Pr(v_1 - e_1 \geq 0)}\right)\right|_{\tilde{r}=r,\tilde{s}=s} = \frac{\Pr(v_1 - e_1 \geq 0|e_2 = r, v_2 = s)g(r)f(s)}{\Pr(v_1 - e_1 \geq 0)} \\
=\ & \frac{g(r)f(s)\int \Pr(e_1 \leq \tilde{s}|v_1 = \tilde{s}, e_2 = r, v_2 = s)dF(\tilde{s}|s)}{\int G(\tilde{s})dF(\tilde{s})} = \frac{g(r)f(s)\int G(\tilde{s}|r)dF(\tilde{s}|s)}{\int G(\tilde{s})dF(\tilde{s})}.
\end{aligned}$$

where the third and fourth equalities follow from an application of the Law of Total Probability in the numerator and the denominator, and from the independence between (e_1, e_2) and (v_1, v_2). On the other hand,

$$\begin{aligned}
\left.\frac{\partial \Pr(e_2 \leq \tilde{r}|y_1 = 1)}{\partial\tilde{r}}\right|_{\tilde{r}=r} &= \left.\frac{\partial}{\partial\tilde{r}}\left[\frac{\Pr(v_1 - e_1 \geq 0, e_2 \leq \tilde{r})}{\Pr(v_1 - e_1 \geq 0)}\right]\right|_{\tilde{r}=r} = \frac{\Pr(v_1 - e_1 \geq 0|e_2 = r)g(r)}{\Pr(v_1 - e_1 \geq 0)} \\
&= \frac{g(r)\int \Pr(e_1 \leq \tilde{s}|v_1 = \tilde{s}, e_2 = r)dF(\tilde{s})}{\int G(\tilde{s})dF(\tilde{s})} = \frac{g(r)\int G(\tilde{s}|r)dF(\tilde{s})}{\int G(\tilde{s})dF(\tilde{s})}
\end{aligned}$$

and

$$\begin{aligned}
\left.\frac{\partial \Pr(v_2 \leq \tilde{s}|y_1 = 1)}{\partial\tilde{s}}\right|_{\tilde{s}=s} &= \left.\frac{\partial}{\partial\tilde{s}}\left[\frac{\Pr(v_1 - e_1 \geq 0, v_2 \leq \tilde{s})}{\Pr(v_1 - e_1 \geq 0)}\right]\right|_{\tilde{s}=s} = \frac{\Pr(v_1 - e_1 \geq 0|v_2 = s)f(s)}{\Pr(v_1 - e_1 \geq 0)} \\
&= \frac{f(s)\int \Pr(e_1 \leq \tilde{s}|v_1 = \tilde{s}, v_2 = s)dF(\tilde{s}|s)}{\int G(\tilde{s})dF(\tilde{s})} = \frac{f(s)\int G(\tilde{s})dF(\tilde{s}|s)}{\int G(\tilde{s})dF(\tilde{s})},
\end{aligned}$$

where the last two equalities hold because of similar reasons. Thus for all (e, v),

$$\frac{\left(\left.\frac{\partial^2 \Pr(e_2 \leq \tilde{r}, v_2 \leq \tilde{s}|y_1 = 1)}{\partial\tilde{r}\partial\tilde{s}}\right|_{\tilde{r}=r,\tilde{s}=s}\right)}{\left(\left.\frac{\partial \Pr(e_2 \leq \tilde{r}|y_1 = 1)}{\partial\tilde{r}}\right|_{\tilde{r}=r}\right)\left(\left.\frac{\partial \Pr(v_2 \leq \tilde{s}|y_1 = 1)}{\partial\tilde{s}}\right|_{\tilde{s}=s}\right)} = \left(\frac{\int G(\tilde{s}|r)dF(\tilde{s}|s)}{\int G(\tilde{s}|r)dF(\tilde{s})}\right)\left(\frac{\int G(\tilde{s})dF(\tilde{s})}{\int G(\tilde{s})dF(\tilde{s}|s)}\right). \tag{2.3}$$

The right-hand side (r.h.s.) of (2.3) is 1 whenever v_1 and v_2 are serially independent ($F(s'|s) = F(s')$ for all (s', s) on the joint support of (v_1, v_2)). However, if v_1 and v_2 are serially dependent,

[2] Exchangeability is *not* assumed in Honoré and Lewbel (2002), but we introduce it here to demonstrate how Assumption A2 could be violated.

then the right-hand side of (2.3) is not equal to 1 in general. To see this, consider an extreme case where v_t has perfect correlation (v_t is time-invariant with $\Pr(v_1 = v_2) = 1$). Then the right-hand side of (2.3) becomes:

$$\frac{G(s|r)}{\int G(\tilde{s}|r)dF(\tilde{s})} \frac{\int G(\tilde{s})dF(\tilde{s})}{G(s)}.$$

Because $G(.|r)$ varies with r due to the dependence between e_1 and e_2, this expression is in general not equal to 1 for all (s, r) on the support of (e_1, e_2). To sum up, the identifying condition A.2 in Honoré and Lewbel (2002) implicitly requires that the excluded regressors $(v_t)_{t=1,2}$ be serially independent. Otherwise, A.2 does not hold in general, and the estimator in Honoré and Lewbel (2002) is generally inconsistent.

3 Identification with Serially Correlated Regressors

Serial independence of *observed covariates* is hard to justify in a dynamic panel data setting. To address this limitation of the method in Honoré and Lewbel (2002), we introduce an alternative approach that is valid in the presence of serial dependence in the regressors. Consider the simplified version of the dynamic binary choice panel data model with two periods $t = 1, 2$ in (2.1), where the initial value y_{i0} is stochastic and reported in the data. Let $\mathbf{y}_i \equiv (y_{i0}, y_{i1}, y_{i2})$, $\mathbf{v}_i \equiv (v_{i1}, v_{i2})$ and $\boldsymbol{\epsilon}_i \equiv (\epsilon_{i1}, \epsilon_{i2})$. Our method requires the following conditions:

EM1 (*Random Sampling*) For each cross-sectional unit i, the vector $(\mathbf{y}_i, \mathbf{v}_i, \alpha_i, \boldsymbol{\epsilon}_i)$ is independently drawn from the same data-generating process. The vector $(\mathbf{y}_i, \mathbf{v}_i)$ is observed while $(\alpha_i, \boldsymbol{\epsilon}_i)$ is not.

EM2 (*Exclusion Restriction*) $\mathbf{v}_i$ is independent of $(\boldsymbol{\epsilon}_i, \alpha_i, y_{i0})$, and is continuously distributed over a support $\mathcal{V} \subseteq \mathbb{R}^2$.

EM3 (*Exchangeability*) Conditional on y_{i0}, $\mathbf{e}_i \equiv (e_{i1}, e_{i2}) \equiv (-\alpha_i - \epsilon_{i1}, -\alpha_i - \epsilon_{i2})$ is continuously distributed with positive density over $\mathbb{R}^2$, and is exchangeable in $t = 1, 2$.

EM4 (*Overlapping Support*) There exists $v, v' \in \mathcal{V}$ such that either "$v'_1 = v_2$ and $v'_2 + \gamma_0 = v_1$" or "$v'_2 = v_1$ and $v_2 = v'_1 + \gamma_0$".

Unlike Assumption A.2 in Honoré and Lewbel (2002), the exclusion restriction in $EM2$ does not condition on the endogenous lagged dependent variable y_{i1}. The exchangeability in $EM3$ holds, for example, if $\boldsymbol{\epsilon}_i$ is exchangeable in $t = 1, 2$ conditional on (α_i, y_{i0}). We impose no other restriction on the distribution of $(\alpha_i, \boldsymbol{\epsilon}_i)$ given y_{i0} than $EM3$. Condition $EM4$ ensures that the intersection of the marginal support of v_{i1} and v_{i2} is non-empty.

A few remarks about how these conditions are related to the existing literature are in order. First, $EM1$-$EM3$ allow for the serial correlation between regressors in $\mathbf{v}_i$, which as we show above are implicitly ruled out in Assumption A.2 in Honoré and Lewbel (2002). Second, our conditions are non-nested with those in Honore and Kyriazidou (2000), which allows the initial condition y_{i0} to depend on $\mathbf{v}_i$. However, our identification result only requires the data to report two periods $T = 2$ (not including the initial condition) whereas their approach requires $T \geq 3$.

We state our identification theorem for the model in (2.1) with a stochastic initial value y_{i0}.

Theorem 3.1 *Consider the model in (2.1) with $t = 1, 2$. Under Assumptions $EM1, 2, 3, 4$, the coefficient γ_0 is identified.*

Proof of Theorem 3.1. Consider two observations i, j such that $y_{i0} = y_{j0} = 0$ and $v_{j1} = v_{i2} = \bar{v}$ for some $\bar{v}$ and

$$\Pr(y_{i1} = 0, y_{i2} = 1 | \mathbf{v}_i, y_{i0} = 0) = \Pr(y_{j1} = 1, y_{j2} = 0 | \mathbf{v}_j, y_{j0} = 0). \tag{3.1}$$

Such a pair i, j and $\bar{v}$ exist under $EM4$. Under $EM2$, the left-hand side of (3.1) is

$$\Pr(e_{i1} > v_{i1}, e_{i2} \leq v_{i2} | y_{i0} = 0) = \Pr(e_{i1} > v_{i1}, e_{i2} \leq \bar{v} | y_{i0} = 0),$$

and the right-hand side of (3.1) is

$$\Pr(e_{j1} \leq v_{j1}, e_{j2} > v_{j2} + \gamma_0 | y_{j0} = 0) = \Pr(e_{j1} > v_{j2} + \gamma_0, e_{j2} \leq \bar{v} | y_{j0} = 0),$$

where the equality follows from the *exchangeability* of (e_{j1}, e_{j2}) given $y_{j0} = 0$ in $EM3$. It follows from $EM1$ and $EM3$ that (3.1) holds if and only if $v_{i1} = v_{j2} + \gamma_0$. This implies γ_0 is -identified as $\gamma_0 = v_{i1} - v_{j2}$ using any pair i, j such that $y_{i0} = y_{j0} = 0$, $v_{j1} = v_{i2}$ and (3.1) holds.

Likewise, we can look for another pair of cross-sectional units k, l with $y_{k0} = y_{l0} = 1$ and $v_{l2} = v_{k1} = \tilde{v}$ for some $\tilde{v}$ and

$$\Pr(y_{k1} = 0, y_{k2} = 1 | \mathbf{v}_k, y_{k0} = 1) = \Pr(y_{l1} = 1, y_{l2} = 0 | \mathbf{v}_l, y_{l0} = 1) \tag{3.2}$$

By a similar argument, the left-hand side of (3.2) is

$$\Pr(e_{k1} > v_{k1} + \gamma_0, e_{k2} \leq v_{k2} | y_{k0} = 1) = \Pr(e_{k1} > \tilde{v} + \gamma_0, e_{k2} \leq v_{k2} | y_{k0} = 1)$$

and the right-hand side of (3.2) is

$$\Pr(e_{l1} \leq v_{l1} + \gamma_0, e_{l2} > v_{l2} + \gamma_0 | y_{l0} = 1) = \Pr(e_{l1} > \tilde{v} + \gamma_0, e_{l2} \leq v_{l1} + \gamma_0 | y_{l0} = 1).$$

It then follows from $EM1$ and $EM3$ that (3.2) holds if and only if $v_{l1} + \gamma_0 = v_{k2}$. Hence γ_0 is over-identified as $v_{k2} - v_{l1}$ using any pair k, l such that $v_{l2} = v_{k1}$ and (3.2) holds. $\square$

3.1 Estimation

We propose two estimators for γ_0 in (2.1), based on the constructive argument for identification in Theorem 3.1. As we show in Appendix A, both estimators converge at the parametric rate to a limiting normal distribution.

The first estimator has a closed form as follows:

$$\hat{\gamma}_{CF} \equiv \frac{\sum_{j\neq i}\left[\omega_{ij,0}(v_{i1}-v_{j2})+\omega_{ij,1}(v_{i2}-v_{j1})\right]}{\sum_{j\neq i}(\omega_{ij,0}+\omega_{ij,1})}, \tag{3.3}$$

where $\sum_{j\neq i}$ is shorthand notation for the summation over *ordered* pairs $\sum_{i=1}^{N}\sum_{j\in\{1,2,\ldots,N\}\backslash\{i\}}$ and

$$\begin{aligned}
\omega_{ij,0} &\equiv K_h(\hat{p}_{i0}-\hat{q}_{j0}, v_{j1}-v_{i2})(1-y_{i0})(1-y_{j0}),\ \omega_{ij,1}\equiv K_h(\hat{p}_{i1}-\hat{q}_{j1}, v_{j2}-v_{i1})y_{i0}y_{j0},\\
\hat{p}_{i0} &\equiv \frac{\sum_s y_{s2}(1-y_{s1})L_\sigma(\mathbf{v}_s-\mathbf{v}_i)(1-y_{s0})}{\sum_s L_\sigma(\mathbf{v}_s-\mathbf{v}_i)(1-y_{s0})},\ \hat{q}_{j0}\equiv\frac{\sum_s y_{s1}(1-y_{s2})L_\sigma(\mathbf{v}_s-\mathbf{v}_j)(1-y_{s0})}{\sum_s L_\sigma(\mathbf{v}_s-\mathbf{v}_j)(1-y_{s0})},\\
\hat{p}_{i1} &\equiv \frac{\sum_s y_{s2}(1-y_{s1})L_\sigma(\mathbf{v}_s-\mathbf{v}_i)y_{s0}}{\sum_s L_\sigma(\mathbf{v}_s-\mathbf{v}_i)y_{s0}},\ \hat{q}_{j1}\equiv\frac{\sum_s y_{s1}(1-y_{s2})L_\sigma(\mathbf{v}_s-\mathbf{v}_j)y_{s0}}{\sum_s L_\sigma(\mathbf{v}_s-\mathbf{v}_j)y_{s0}},
\end{aligned}$$

with $K_h(.)\equiv\frac{1}{h}K(\frac{\cdot}{h})$ and $L_\sigma(.)\equiv\frac{1}{\sigma}L(\frac{\cdot}{\sigma})$ being shorthand notation for kernel smoothing.

The intuition for the consistency of $\hat{\gamma}_{CF}$ is as follows. First off, under appropriate conditions, the ratio $\left(\sum_{j\neq i}\omega_{ij,0}\right)^{-1}\left[\sum_{j\neq i}\omega_{ij,0}(v_{i1}-v_{j2})\right]$ converges in probability to the expectation of $v_{i1}-v_{j2}$ conditional on $y_{i0}=y_{j0}=0$, $v_{j1}=v_{i2}$ and on the equality in (3.1). By the proof of Theorem 3.1, such a conditional expectation is equal to γ_0. Likewise, $\left(\sum_{j\neq i}\omega_{ij,1}\right)^{-1}\left[\sum_{j\neq i}\omega_{ij,1}(v_{i2}-v_{j1})\right]$ also converges in probability to γ_0. Thus the estimator $\hat{\gamma}_{CF}$ in (3.3) is a weighted average of these two components, each of which is consistent for γ_0. This estimator avoids minimization, but requires multiple kernel smoothing procedures. In Appendix A we show that this estimator is root-n consistent and asymptotically normal (CAN).

The second estimator we propose is a kernel-weighted maximum rank correlation estimator:

$$\hat{\gamma}_{MR} \equiv \max_{\gamma}\frac{1}{n(n-1)}\sum_{j\neq i}\left[\tilde{\omega}_{ij,0}G_{ij,0}(\gamma)+\tilde{\omega}_{ij,1}G_{ij,1}(\gamma)\right], \tag{3.4}$$

where

$$\begin{aligned}
G_{ij,0}(\gamma) &\equiv 1\{d_{i,01}>d_{j,10}\}1\{v_{j2}+\gamma>v_{i1}\}+1\{d_{i,01}<d_{j,10}\}1\{v_{j2}+\gamma<v_{i1}\};\\
G_{ij,1}(\gamma) &\equiv 1\{d_{i,01}>d_{j,10}\}1\{v_{i2}>v_{j1}+\gamma\}+1\{d_{i,01}<d_{j,10}\}1\{v_{i2}<v_{j1}+\gamma\}
\end{aligned}$$

with

$$\begin{aligned}
d_{i,01} &\equiv (1-y_{i1})y_{i2}\ ,\ d_{j,10}\equiv y_{j1}(1-y_{j2}), \qquad (3.5)\\
\tilde{\omega}_{ij,0} &\equiv \tilde{K}_h(v_{j1}-v_{i2})(1-y_{i0})(1-y_{j0})\ ,\ \tilde{\omega}_{ij,1}\equiv\tilde{K}_h(v_{j2}-v_{i1})y_{i0}y_{j0}
\end{aligned}$$

and $\tilde{K}_h(.) \equiv \frac{1}{h}\tilde{K}(\frac{\cdot}{h})$. This estimator is motivated by the maximum rank correlation estimator introduced by Han (1987) for cross-sectional models.

To understand the intuition for the consistency of $\hat{\gamma}_{MR}$, note that under appropriate regularity conditions, the objective function in (3.4) converges in probability to a weighted integral of

$$E[G_{ij,0}(\gamma)|v_{j1} = v_{i2}, y_{i0} = 0, y_{j0} = 0] \text{ and } E[G_{ij,1}(\gamma)|v_{j2} = v_{i1}, y_{i0} = 1, y_{j0} = 1]. \tag{3.6}$$

Both conditional expectations in (3.6) are uniquely maximized at $\gamma = \gamma_0$ under the support condition in $EM4$.[3] The formal argument is analogous to the proof of identification in Han (1987) and omitted for brevity here. Thus the extremum estimator in (3.4) satisfies the identification condition for consistency.

We note that this maximum rank correlation estimator has the advantage of requiring fewer smoothing parameters than the first closed-form estimator. This comes at the expense of higher computational costs as the maximum rank correlation estimator requires optimization of a non-concave objective function. Nonetheless, desirable asymptotic properties such as root-n consistency and asymptotic normality still hold, as we show in Appendix A using an argument similar to Abrevaya, Hausman, and Khan (2010).

4 Model with Multiple Explanatory Regressors

We now discuss the identification and estimation of the model in (1.1) when x_{it} includes other regressors $w_{it} \in \mathbb{R}^L$ as well as the lagged dependent variable y_{it-1}. That is, in the notation of Honoré and Lewbel (2002), $x_{it} \equiv (y_{it-1}, w_{it})$, $\beta_0 \equiv (\gamma_0, \delta_0')'$ and

$$y_{it} = I[v_{it} + \gamma_0 y_{it-1} + w_{it}'\delta_0 + \alpha_i + \epsilon_{it} \geq 0] \text{ for } t = 1, 2. \tag{4.1}$$

Let $\mathbf{w}_i \equiv (w_{i1}, w_{i2})$, $\boldsymbol{\epsilon}_i \equiv (\epsilon_{i1}, \epsilon_{i2})$ and $\mathbf{e}_i \equiv (-\alpha_i - \epsilon_{i1}, -\alpha_i - \epsilon_{i2})$ as before. Let $\mathcal{W} \subset \mathbb{R}^L$denote the support of $\mathbf{w}_i$. We maintain the following conditions:

EEM1 (*Random Sampling*) For each i, $(\mathbf{y}_i, \mathbf{w}_i, \mathbf{v}_i, \alpha_i, \boldsymbol{\epsilon}_i)$ are independently drawn from the same data-generating process. The vector $(\mathbf{y}_i, \mathbf{w}_i, \mathbf{v}_i)$ is reported in data while $(\alpha_i, \boldsymbol{\epsilon}_i)$ is not.

EEM2 (*Exclusion Restriction*) Conditional on $\mathbf{w}_i$, $\mathbf{v}_i$ is independent of $(\mathbf{e}_i, y_{i0})$ and is continuously distributed over a connected support $\mathcal{V} \subseteq \mathbb{R}^2$.

EEM3 (*Exchangeability*) Conditional on $(\mathbf{w}_i, y_{i0})$, $\mathbf{e}_i$ is continuously distributed with positive density over $\mathbb{R}^2$ and is exchangeable in $t = 1, 2$.

[3] Assumption EM4 implies that for all $\gamma \neq \gamma_0$, there is positive probability that $v_{i1} - v_{j2}$ is between γ and γ_0 conditional on $v_{j1} = v_{i2}, y_{i0} = 0, y_{j0} = 0$.

EEM4 (*Support Condition*) There exists $\mathcal{C} \subset \mathbb{R}^L$ such that $\mathcal{C} \otimes \mathcal{C} \subseteq \mathcal{W}$, and the support condition in $EM4$ holds conditional on some $\mathbf{w}_i = (\bar{w}, \bar{w}) \in \mathcal{C} \otimes \mathcal{C}$.

EEM5 (*Rank Condition*) There exists some $\mathcal{W}_0 \subseteq \mathcal{W}$ such that (i) the support of $\{w_{i2} - w_{i1} : \mathbf{w}_i \in \mathcal{W}_0\}$ is not contained in any proper linear subspace, and (ii) conditional on any $\mathbf{w}_i \in \mathcal{W}_0$, there exists $v, \tilde{v} \in \mathcal{V}$ with $v_1 = \tilde{v}_1$, $\tilde{v}_2 = v_2 + \gamma_0$ and $v_1 = \tilde{v}_2 + (w_{i2} - w_{i1})'\delta_0$.

Theorem 4.1 *Consider the model in (4.1) with $t = 1, 2$. Under Assumptions $EEM1, 2, 3, 4, 5$, the coefficients γ_0 and δ_0 are identified.*

Proof of Theorem 4.1. For each i, define $\eta_{it} \equiv -\alpha_i - \epsilon_{it} - w_{i1}'\delta_0$ for $t = 1, 2$ (note that the definition subtracts the *first period* index $w_{i1}'\delta_0$ for *both* $t = 1, 2$). Under $EEM3$, the distribution of $\boldsymbol{\eta}_i \equiv (\eta_{i1}, \eta_{i2})$ is exchangeable in $t = 1, 2$ conditional on $\mathbf{w}_i$. Thus (4.1) can be written as:

$$y_{i1} = I[\eta_{i1} \leq v_{i1} + \gamma_0 y_{i0}] \text{ and } y_{i2} = I[\eta_{i2} \leq v_{i2} + \gamma_0 y_{i1} + \Delta_i],$$

where $\Delta_i \equiv (w_{i2} - w_{i1})'\delta_0$. Consider a vector $\bar{\mathbf{w}}_i$ with $\bar{w}_{i1} = \bar{w}_{i2}$ so that $\Delta_i = 0$. Such a vector exists under the support condition in $EEM4$. Conditional on such a $\bar{\mathbf{w}}_i$, the argument for identifying γ_0 in Theorem 3.1 applies (with the constant index $\bar{w}_{i1}'\delta_0$ absorbed in the fixed effect α_i).

With γ_0 known, we can look for pairs of cross-sectional units i and j such that

$$\mathbf{w}_i = \mathbf{w}_j,\ v_{i1} = v_{j1} \text{ and } v_{j2} = v_{i2} + \gamma_0. \tag{4.2}$$

Such pairs exist due to the condition (ii) in $EEM5$. By construction,

$$\begin{aligned} & \Pr(y_{i1} = 1, y_{i2} = 0 | \mathbf{v}_i, \mathbf{w}_i, y_{i0} = 0) + \Pr(y_{j1} = 0, y_{j2} = 0 | \mathbf{v}_j, \mathbf{w}_j, y_{j0} = 0) \\ = \ & \Pr(\eta_{i1} \leq v_{i1}, \eta_{i2} > v_{i2} + \gamma_0 + \Delta_i | \mathbf{w}_i, y_{i0} = 0) + \Pr(\eta_{j1} > v_{j1}, \eta_{j2} > v_{j2} + \Delta_j | \mathbf{w}_j, y_{j0} = 0) \\ = \ & \Pr(\eta_{j2} > v_{j2} + \Delta_j | \mathbf{w}_j, y_{j0} = 0), \end{aligned}$$

where the first equality follows from $EEM2$ and the second follows from the equalities in (4.2) and $EEM1$. Also by $EEM2$,

$$\Pr(y_{i1} = 0 | \mathbf{v}_i, \mathbf{w}_i, y_{i0} = 0) = \Pr(\eta_{i1} > v_{i1} | \mathbf{w}_i, y_{i0} = 0).$$

It then follows from $EEM3$ that for any pair i and j that satisfy the equalities in (4.2),

$$\begin{aligned} \Pr(y_{i1} &= 1, y_{i2} = 0 | \mathbf{v}_i, \mathbf{w}_i, y_{i0} = 0) + \Pr(y_{j1} = 0, y_{j2} = 0 | \mathbf{v}_j, \mathbf{w}_j, y_{j0} = 0) \\ &= \Pr(y_{i1} = 0 | \mathbf{v}_i, \mathbf{w}_i, y_{i0} = 0) \end{aligned} \tag{4.3}$$

if and only if

$$v_{i1} = v_{j2} + \Delta_j.$$

This identifies $\Delta_j = (w_{j2} - w_{j1})'\delta_0$ for $\mathbf{w}_j$ conditional on the equalities in (4.2). Replicating the same argument conditional on other values of $\mathbf{w}_j$ allows us to identify δ_0 under the rank condition (i) in $EEM5$.[4] $\square$

We now discuss how to translate the identification result in Theorem 4.1 to an estimation procedure, conditioning on the set of time-invariant regressors mentioned in $EEM4$. For ease of illustration, suppose for now that the vector w_{it} consists of discrete components only. Then we can construct a closed-form estimator and a kernel weighted maximum rank correlation estimator for γ_0 similar to those in Section 3.1 conditioning on "$w_{i1} = w_{i2}$", an equality that occurs with positive probability due to the discreteness of w_{it}. Specifically, we need to replace the weights $\omega_{ij,s}, \tilde{\omega}_{ij,s}$ in (3.3) and (3.4) with $\omega_{ij,s}1\{w_{i1} = w_{i2}\}$ and $\tilde{\omega}_{ij,s}1\{w_{i1} = w_{i2}\}$ respectively for $s = 0, 1$. If w_{it} contains a continuous component, the event "$w_{i1} = w_{i2}$" occurs with zero probability and we need to replace $1\{w_{i1} = w_{i2}\}$ with kernel weights $\frac{1}{h}K(\frac{w_{i1}-w_{i2}}{h})$ in order to implement the estimation procedure. Note that in such an estimation procedure we are trimming out all but a shrinking fraction of the cross-sectional population, as opposed to trimming all but a shrinking fraction of pairwise comparisons as before. Consequently, the resulting estimator converges at a nonparametric rate. In comparison, the estimator proposed in Honore and Kyriazidou (2000) does not impose any exclusion restriction, but requires more time periods yet still attains a nonparametric rate.

The non-standard, slower rate of convergence of our estimator described in the preceding paragraph motivates the need to strengthen model assumptions in order to construct root-n CAN estimators. The following two subsections present two cases where root-n CAN estimators are available under strengthened model assumptions.

4.1 Exchangeability in Regressors

Consider the following condition of exchangeability:

EEM3' (*Exchangeability*) The distribution of $\mathbf{e}_i \equiv (e_{i1}, e_{i2})$ conditional on y_{i0} and $\mathbf{w}_i \equiv (w_{i1}, w_{i2})$ is exchangeable in the time index $t = 1, 2$.

EEM4' (*Support Condition*) There exists $(\mathbf{w}, \mathbf{v})$ and $(\widetilde{\mathbf{w}}, \widetilde{\mathbf{v}})$ such that $(w_1, w_2) = (\tilde{w}_2, \tilde{w}_1)$ and either "$\tilde{v}_1 = v_2, v_1 = \tilde{v}_2 + \gamma_0$" or "$\tilde{v}_2 = v_1, v_2 = \tilde{v}_1 + \gamma_0$".

[4] A symmetric argument identifies δ_0 from analogous conditional probabilities under similar support conditions, using other pairs with $y_{i0} = 1$ and $y_{j0} = 1$.

The condition $EEM3'$ strengthens the exchangeability in (e_{i1}, e_{i2}) in $EEM3$ into a stronger notion of exchangeability in (e_{i1}, e_{i2}) *as well as* the explanatory variables (w_{i1}, w_{i2}) that are conditioned on. This notion of exchangeability in $EEM3'$ has been previously used to attain identification results in the econometrics literature. See, for example, Honoré (1992) for censored panel models, Fox (2007) for multinomial choice models, and Altonji and Matzkin (2005) for non-separable models in both cross-sectional and (random-effect) panel data models. In a binary choice random effect model, Altonji and Matzkin (2005) uses the following condition (Assumption 2.3) on the unobserved errors and observed regressors (assuming two time periods for ease of exposition)

$$f(e_{it}|w_{i1}, w_{i2}) = f(e_{it}|w_{i2}, w_{i1}) \text{ for } t = 1, 2,$$

where $f(\cdot|\cdot)$ notes the density of e_{it} conditional on $\mathbf{w}_i$. Hence their Assumption 2.3 means that the value of the conditional density function does not change even if the order of the conditioning variables does. Their Assumption 2.3, along with a conditional i.i.d. (or exchangeability) assumption on ϵ_{it} would imply our Assumption $EEM3'$.

Theorem 4.2 *Consider the model in (4.1) with $t = 1, 2$. Under Assumptions $EEM1, 2, 3', 4', 5$, the coefficients γ_0 and δ_0 are identified .*

Proof of Theorem 4.2. Consider a pair $(\mathbf{w}_i, \mathbf{v}_i)$ and $(\mathbf{w}_j, \mathbf{v}_j)$ such that

$$(w_{i1}, w_{i2}) = (w_{j2}, w_{j1}) \text{ and } v_{j1} = v_{i2} \tag{4.4}$$

and

$$\Pr(y_{i1} = 0, y_{i2} = 1|\mathbf{w}_i, \mathbf{v}_i, y_{i0} = 0) = \Pr(y_{j1} = 1, y_{j2} = 0|\mathbf{w}_j, \mathbf{v}_j, y_{j0} = 0). \tag{4.5}$$

As we show below, such a pair exists under the support condition in $EEM4'$. By construction, the left-hand side of (4.5) is

$$\Pr(e_{i1} > w_{i1}'\beta_0 + v_{i1}, e_{i2} \leq w_{i2}'\beta_0 + v_{i2}|\mathbf{w}_i, y_{i0} = 0),$$

where the equality is due to $EEM2$. Besides, the right-hand side of (4.5) is

$$\begin{aligned} & \Pr(e_{j1} \leq w_{j1}'\beta_0 + v_{j1}, e_{j2} > w_{j2}'\beta_0 + v_{j2} + \gamma_0|\mathbf{W}_j = (w_{j1}, w_{j2}), y_{j0} = 0) \\ = \; & \Pr(e_{j2} \leq w_{j1}'\beta_0 + v_{j1}, e_{j1} > w_{j2}'\beta_0 + v_{j2} + \gamma_0|\mathbf{W}_j = (w_{j2}, w_{j1}), y_{j0} = 0) \\ = \; & \Pr(e_{i2} \leq w_{i2}'\beta_0 + v_{i2}, e_{i1} > w_{i1}'\beta_0 + v_{j2} + \gamma_0|\mathbf{W}_i = (w_{i1}, w_{i2}), y_{i0} = 0), \end{aligned}$$

where the first equality follows from the exchangeability in $EEM3'$ and the second follows from i.i.d. sampling assumption ($EEM1$) and the equalities in (4.4). Thus the equality in (4.5) holds if and only if $v_{i1} = v_{j2} + \gamma_0$. Thus $\gamma_0 = v_{i1} - v_{j2}$ is over-identified using any pair $(\mathbf{w}_i, \mathbf{v}_i)$ and $(\mathbf{w}_j, \mathbf{v}_j)$ that satisfy (4.4) and (4.5).

By a symmetric argument, we can look for pairs of $(\mathbf{w}_i, \mathbf{v}_i)$ and $(\mathbf{w}_j, \mathbf{v}_j)$ such that

$$(w_{i1}, w_{i2}) = (w_{j2}, w_{j1}), v_{j2} = v_{i1}$$

and

$$\Pr(y_{i1} = 0, y_{i2} = 1 | \mathbf{w}_i, \mathbf{v}_i, y_{i0} = 1) = \Pr(y_{j1} = 1, y_{j2} = 0 | \mathbf{w}_j, \mathbf{v}_j, y_{j0} = 1),$$

and show that γ_0 is (over-)identified as $\gamma_0 = v_{i2} - v_{j1}$.

With γ_0 identified, we can replicate the argument in Theorem 4.1 to identify δ_0 through pairwise comparison under the rank condition in $EEM5$. $\square$

Based on the identification strategy in Theorem 4.2, we propose a two-step estimator for γ_0, δ_0. In the first step, use a kernel-weighted maximum rank correlation estimator to estimate γ_0:

$$\hat{\gamma}_{EX} \equiv \max_{\gamma} \frac{1}{n(n-1)} \sum\nolimits_{j \neq i} \mathcal{K}_{ij} \left[\tilde{\omega}_{ij,0} G_{ij,0}(\gamma) + \tilde{\omega}_{ij,1} G_{ij,1}(\gamma) \right]$$

where $d_{i,01}$, $d_{j,10}$, $\tilde{\omega}_{ij,0}$ and $\tilde{\omega}_{ij,1}$ are defined as in (3.5) and $\mathcal{K}_{ij} \equiv \mathcal{K}_{\sigma}(w_{i1} - w_{j2}, w_{i2} - w_{j1})$ is a kernel weight for matching i and j with $(w_{i1}, w_{i2}) = (w_{j2}, w_{j1})$. In the second step, use $\hat{\gamma}_{EX}$ to construct a closed-form estimator for δ_0 by matching pairs i and j that satisfy the equalities in (4.2) and (4.3) simultaneously.

Remark 4.1 *For the model with lagged dependent variables as well as strictly exogenous variables we are able to identify the regression coefficients for three times periods. Furthermore the closed-form estimator we propose converges at the parametric rate with a limiting normal distribution. In comparison, Honore and Kyriazidou (2000) require four periods for identification and achieve a nonparametric rate in estimation. They impose an i.i.d. assumption on ϵ_{it} which is stronger than what we assume here; but they do not impose the exclusion restriction in $EEM2$ nor the exchangeability assumption in $EEM3\prime$. They require that the support of w_{it} to be overlapping over time so that the difference in regressors across adjacent time periods has a positive density in a neighborhood of* 0.

Remark 4.2 *Our new identification and estimation results extend to static binary choice models with fixed effects, where the vector of explanatory variables in a period does not include the lagged dependent variable from the previous period. (See Appendix B for details.) In fact, it can be shown that the same pairwise comparison procedure provides root-n CAN estimators of the regression coefficients under weaker conditions. Specifically, in a static binary choice panel data model with* *the exclusion restriction, we only need observations of the dependent and explanatory variables in two time periods, do not require the exchangeability in* w_{it} *(*$EEM3'$*), and only impose a conditional stationarity restriction on* e_{it}*. The assumed behavior on* e_{it} *would then be identical to that imposed in Manski (1987). The model in Manski (1987) is more general as exclusion was not assumed for his identification result, but the parameters could not be estimated at the parametric rate. Ai and Gan (2010) also study static models (e.g. without lagged dependent variables) and, like ours, their estimator converges at the parametric rate. However, they require* w_{it} *to be independent of* e_{it}*, which we do not impose in the static model considered in Appendix B.*

4.2 Exogenous Initial Condition

Another way to attain root-n CAN estimation is to further require that e_{i1}, v_{i1} and the initial condition y_{i0} be *mutually* independent conditional on w_{i1}. Then a similar two-step estimator based on pairwise comparison across cross-sectional units with different initial conditions can be constructed.

To see this, suppose that e_{i1}, v_{i1} and y_{i0} are mutually independent given w_{i1}, and that e_{i1} is continuously distributed over $\mathbb{R}$ given w_{i1}. First, look for pairs of observations i and j such that

$$w_{i1} = w_{j1} \text{ and } \Pr(y_{i1} = 1|y_{i0} = 0, w_{i1}, v_{i1}) = \Pr(y_{j1} = 1|y_{j0} = 1, w_{j1}, v_{j1}). \tag{4.6}$$

Under the mutual independence condition above, the second equality in (4.6) is equivalent to

$$\Pr(e_{i1} \leq w'_{i1}\delta_0 + v_{i1}|w_{i1}) = \Pr(e_{j1} \leq w'_{j1}\delta_0 + v_{j1} + \gamma_0|w_{j1}).$$

Such pairs exist under mild support conditions. It then follows that γ_0 is over-identified as $\gamma_0 = v_{i1} - v_{j1}$ using any pair of i and j that satisfies both equalities in (4.6). Second, with γ_0 identified, we can identify δ_0 under $EEM1, 2, 3, 5$ by the same argument as in the proof of Theorem 4.1, which uses variables observed in both periods $t = 1, 2$. This line of constructive identification argument lends itself to a two-step estimator that is based on pairwise comparisons, and root-n CAN under appropriate regularity conditions.

5 Simulation Study

In the this section we compare the finite-sample performance of the new estimators we propose with that of existing estimators in dynamic panel data models. The first class of designs we consider have no other strictly exogenous regressors than v_{it}. We randomly generate data from the following equation:

$$y_{it} = I[\alpha_i + v_{it} + \gamma_0 y_{i,t-1} + \epsilon_{it} > 0] \quad t = 1, 2$$

where (v_{i1}, v_{i2}) is bivariate normal with mean $(0, 0)$ and unit standard deviation $(1, 1)$. We report the results for several designs, with the coefficient of correlation between v_{i1} and v_{i2} ranging from 0 to 0.75. The error terms $(\epsilon_{i1}, \epsilon_{i2})$ are independent of $(v_{i1}, v_{i2}, \alpha_i)$ with a bivariate normal distribution with mean $(0, 0)$, standard deviation $(1, 1)$ and a correlation coefficient of 0.5. For the fixed effect α_i, we considered designs where α_i is binary and independent of (v_{i1}, v_{i2}) with $\Pr(\alpha_i = 1) = \Pr(\alpha_i = 0) = 0.5$.

Table 1 reports simulation results for two estimators of $\gamma_0 = 0.5$: the inverse weighting procedure in Honoré and Lewbel (2002) (HL) and our two estimators based on pairwise comparison, i.e., the closed-form estimator (CKT1) and the kernel-weighted maximum rank correlation estimator (CKT2). In practice, each of these three estimators requires some nonparametric estimation procedure and hence smoothing parameters. To focus on these estimators' sensitivity to serial dependence in v_{it} (as opposed to their sensitivity to tuning parameters), we compare the infeasible version of each estimator in our simulation exercises. That is, we use knowledge of the true conditional density for the estimator in Honoré and Lewbel (2002), and the true conditional choice probability for both of our estimators introduced here.

For matching the probabilities in the CKT1 estimator we followed the procedure outlined in Chen, Khan, and Tang (2016), where there is under-smoothing in the choice of bandwidths for the kernel estimation of propensity scores in the preliminary step (relative to the bandwidths used for matching explanatory variables).

We report the mean bias (BIAS) and the root mean square errors (RMSE) of all three estimators for γ_0 (HL, CKT1 and CKT2) in the design above, with the correlation coefficients for (v_{i1}, v_{i2}) ranging between $\rho_v \in \{0, 0.25, 0.5, 0.75\}$. For each sample size $n \in \{200, 400, 800, 1600\}$, we calculate the mean bias and RMSEs using 1601 replications of simulated samples.

The findings from this simulation exercise under dynamic designs are in accordance with our theoretical results. When there is serial correlation in v_{it} ($\rho_v \neq 0$), the mean bias of the HL estimator does not decline monotonically with the sample size, and the RMSE diminishes at a rate much slower than root-n. In contrast, both of our estimators proposed in this paper demonstrate a faster rate of decline in RMSE as the sample size increases, regardless of the level of serial correlation in v_{it}.

TABLE 1. Performance of Estimators for γ_0 in a Simple Model

(Exogenous variable: $\mathbf{v}_i$)

	HL				CKT1				CKT2			
ρ_v	0	1/4	1/2	3/4	0	1/4	1/2	3/4	0	1/4	1/2	3/4
n=200												
BIAS	0.067	0.001	-0.045	-0.123	-0.113	-0.117	-0.130	-0.179	-0.018	-0.006	-0.016	0.008
RMSE	1.572	1.597	1.511	1.501	0.152	0.153	0.161	0.200	0.317	0.318	0.317	0.318
n=400												
BIAS	-0.038	-0.152	-0.149	-0.316	-0.096	-0.099	-0.115	-0.162	-0.001	-0.031	-0.002	-0.004
RMSE	1.196	1.122	1.136	1.216	0.118	0.118	0.129	0.173	0.283	0.288	0.279	0.280
n=800												
BIAS	0.055	-0.097	-0.225	-0.288	-0.086	-0.088	-0.106	-0.147	-0.001	-0.006	0.004	0.009
RMSE	1.027	1.025	1.023	1.216	0.097	0.097	0.114	0.153	0.229	0.232	0.227	0.234
n=1600												
BIAS	0.006	-0.104	-0.214	-0.382	-0.074	-0.079	-0.096	-0.140	-0.003	-0.005	0.003	0.009
RMSE	1.024	0.992	0.857	0.890	0.079	0.084	0.100	0.143	0.176	0.176	0.180	0.173

Next, we study a more general design which includes other explanatory variables w_{it} in addition to v_{it}:

$$y_{it} = I[\alpha_i + v_{it} + w_{it}\delta_0 + \gamma_0 y_{i,t-1} + \epsilon_{it} > 0].$$

In our simulation we let $\delta_0 = 1$ and $\mathbf{w}_i$ be independent of $(\alpha_i, \mathbf{v}_i, \boldsymbol{\epsilon}_i)$. We let (w_{it}, w_{i2}) be serially independent and drawn from a binary distribution $\Pr(w_{it} = 1) = \Pr(w_{it} = 0) = 0.5$ for $t = 1, 2$. The other elements of the model are specified as in the simple design above.

Table 2 and Table 3 report the performance of HL and CKT1 estimators for δ_0 and for γ_0 respectively in small and moderate-sized samples. Table 3 shows in general neither the mean bias or the RMSE of the HL estimator for γ_0 decreases with the sample size in the presence of serial correlation in v_{it}. Table 2 demonstrates similar results for the HL estimator for δ_0. In comparison, CKT1 exhibits a noticeable bias for both δ_0 and γ_0 (especially δ_0) but the RMSE diminishes as the sample size increases.

TABLE 2. Performance of Estimators for δ_0 in a Full Model

(Exogenous variables: $\mathbf{w}_i$ and $\mathbf{v}_i$)

	HL				CKT1			
ρ_v	0	1/4	1/2	3/4	0	1/4	1/2	3/4
200 obs.								
Mean Bias	-0.0546	-0.1286	-0.1522	-0.0655	-0.2412	-0.2541	-0.2651	-0.3244
RMSE	1.0879	0.8127	1.1537	1.2749	0.2475	0.2600	0.2713	0.3323
400 obs.								
Mean Bias	-0.0486	-0.0448	-0.1031	-0.1093	-0.2188	-0.2304	-0.2412	-0.2993
RMSE	1.4033	1.5027	0.7018	1.3616	0.2220	0.2339	0.2447	0.3033
800 obs.								
Mean Bias	0.0324	-0.0057	-0.0869	-0.1413	-0.1990	0.2100	-0.2213	-0.2817
RMSE	1.7034	1.0512	0.8770	0.4258	0.2007	0.2117	0.2231	0.2835
1600 obs.								
Mean Bias	-0.0873	-0.0812	-0.1043	-0.0838	-0.1811	-0.1902	-0.2012	-0.2577
RMSE	0.4085	0.4692	0.8775	0.7066	0.1822	0.1911	0.2011	0.2588

TABLE 3. Performance of Estimators for γ_0 in a Full Model

(Exogenous variables: $\mathbf{w}_i$ and $\mathbf{v}_i$)

	HL				CKT1			
ρ_v	0	1/4	1/2	3/4	0	1/4	1/2	3/4
200 obs.								
Mean Bias	-0.1014	-0.0441	-0.1960	-0.3782	-0.1286	-0.1598	-0.1796	-0.1946
RMSE	7.76673	9.9655	6.1891	4.9221	0.1435	0.1791	0.1897	0.2037
400 obs.								
Mean Bias	-0.0734	-0.1767	-0.0921	-0.2216	-0.1195	-0.1437	-0.1644	-0.1828
RMSE	6.8646	3.2316	3.5076	4.6439	0.1269	0.1493	0.1694	0.1872
800 obs.								
Mean Bias	0.2720	-0.1397	-0.5554	-0.3128	-0.1108	-0.1334	-0.1510	-0.1661
RMSE	6.9441	2.7588	5.4450	2.0464	0.1146	0.1369	0.1539	0.1686
1600 obs.								
Mean Bias	-0.1448	-0.2419	-0.2386	-0.3757	-0.1014	-0.1206	-0.1386	-0.1516
RMSE	2.2697	2.2712	3.4023	2.4554	0.1035	0.1222	0.1401	0.1530

6 Conclusions

We explore the use of exclusion restrictions in dynamic binary choice panel data models introduced in Honoré and Lewbel (2002). Their model was partly motivated by the difficulty in identifying models that allow for both state dependence and unobserved heterogeneity. However, here we show that the exclusion restriction in Honoré and Lewbel (2002) requires (conditional) *serial independence* of the excluded regressor. Thus their inverse-density-weighted estimator in Honoré and Lewbel (2002) is generally *inconsistent* when the excluded regressors are serially correlated in a dynamic panel data model.

We propose a new approach of identification and estimators for semiparametric binary choice panel data model under exclusion restrictions. Our approach accommodates the serial dependence in the excluded regressors, and the new estimators converge at the parametric rate to a limiting normal distribution. This rate is faster than the nonparametric rates of existing alternative estimators for the binary choice panel data model, including the static case in Manski (1987) and the dynamic case in Honore and Kyriazidou (2000).

References

Abrevaya, J., J. Hausman, and S. Khan (2010): "Testing for Causal Effects in a Generalized Regression Model with Endogenous Regressors," *Econometrica*, 48, 2043–2061.

Ai, C., and L. Gan (2010): "An Alternative Root-n Consistent Estimator for Panel Data Binary Choice Model," *Journal of Econometrics*, 157, 93–100.

Altonji, J., and R. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–1102.

Chen, S., S. Khan, and X. Tang (2016): "On the Informational Content of Special Regressors in Heteroskedastic Binary Response Models," *Journal of Econometrics*, 193, 162–182.

Fox, J. (2007): "Semiparametric Estimation of Multinomials Discrete Choice Models using a Subset of Choices," *RAND Journal of Economics*, 38, 1002–1029.

Hahn, J. (2001): "The Information Bound of a Dynamic Panel Logit Model with Fixed Effects," *Econometric Theory*, 17, 913–932.

Han, A. (1987): "The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 303-316.

Heckman, J. (1991): "Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity," *American Economic Review*, pp. 75–79.

Honoré, B. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60(3), 533–65.

Honore, B., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839–874.

Honoré, B., and A. Lewbel (2002): "Semiparametric Binary Choice Panel Data Models without Strictly Exogeneous Regressors," *Econometrica*, 70(5), 2053–2063.

Lewbel, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145–177.

Lewbel, A., and X. Tang (2015): "Identification and Estimation of Games with Incomplete Information using Excluded Regressors," *Journal of Econometrics*, 189, 229–244.

Manski, C. F. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55(2), 357–362.

Sherman, R. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137.

——— (1994): "U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator," *Econometric Theory*, 10, 372–395.

A Regularity Conditions and Asymptotic Theory

We outline the regularity conditions and arguments for deriving the limiting distribution of the closed-form estimator $\hat{\gamma}_{CF}$ in Section A.1 and the rank-based estimator $\hat{\gamma}_{MR}$ in Section A.2. In both cases we focus on the simplified model where the two regressors are the excluded variable v_{it} and the lagged dependent variable y_{it-1}.

A.1 Closed-Form Estimator

Recall the closed-form estimator was expressed as

$$\hat{\gamma}_{CF} \equiv \frac{\sum_{j\neq i}\left[\omega_{ij,0}(v_{i1}-v_{j2})+\omega_{ij,1}(v_{i2}-v_{j1})\right]}{\sum_{j\neq i}\left(\omega_{ij,0}+\omega_{ij,1}\right)} \tag{A.1}$$

where $\sum_{j\neq i}$ denote the summation over $N(N-1)$ ordered pairs and

$$\begin{aligned}
\omega_{ij,0} &\equiv K_h(\hat{p}_{i0}-\hat{q}_{j0}, v_{j1}-v_{i2})(1-y_{i0})(1-y_{j0}) \ , \ \omega_{ij,1} \equiv K_h(\hat{p}_{i1}-\hat{q}_{j1}, v_{j2}-v_{i1})y_{i0}y_{j0}; \\
\hat{p}_{i0} &\equiv \frac{\sum_s y_{s2}(1-y_{s1})L_\sigma(\mathbf{v}_s-\mathbf{v}_i)(1-y_{s0})}{\sum_s L_\sigma(\mathbf{v}_s-\mathbf{v}_i)(1-y_{s0})} \ , \ \hat{q}_{j0} \equiv \frac{\sum_s y_{s1}(1-y_{s2})L_\sigma(\mathbf{v}_s-\mathbf{v}_j)(1-y_{s0})}{\sum_s L_\sigma(\mathbf{v}_s-\mathbf{v}_j)(1-y_{s0})}; \\
\hat{p}_{i1} &\equiv \frac{\sum_s y_{s2}(1-y_{s1})L_\sigma(\mathbf{v}_s-\mathbf{v}_i)y_{s0}}{\sum_s L_\sigma(\mathbf{v}_s-\mathbf{v}_i)y_{s0}} \ , \ \hat{q}_{j1} \equiv \frac{\sum_s y_{s1}(1-y_{s2})L_\sigma(\mathbf{v}_s-\mathbf{v}_j)y_{s0}}{\sum_s L_\sigma(\mathbf{v}_s-\mathbf{v}_j)y_{s0}},
\end{aligned}$$

with $K_h(.) \equiv \frac{1}{h}K(\frac{.}{h})$ and $L_\sigma(.) \equiv \frac{1}{\sigma}L(\frac{.}{\sigma})$.

Our arguments for the limiting distribution theory for the closed form estimator are based on the following conditions:

Assumption A.1 (Non-singularity) The matrix Σ defined in (A.6) is positive and finite.

Assumption A.2 (Kernels for matching) (i) Let $K(\cdot,\cdot) = K_1(\cdot)K_2(\cdot)$ where K_1, K_2 have compact supports, are symmetric around 0, integrate to 1, are twice continuously differentiable and are eighth-order kernels. (ii) $\sup_{t\in\mathbb{R}} h_2^{-1}|K_2(t/h_2)|$, $\sup_{t\in\mathbb{R}} h_1^{-1}|K_1'(t/h_1)|$ and $\sup_{t\in\mathbb{R}} h_1^{-1}|K_1''(t/h_1)|$ are all $O(1)$ as $h_1, h_2 \to 0$.

Assumption A.3 (Bandwidths for matching) $h_1 \propto n^{-\delta_1}$ and $h_2 \propto n^{-\delta_2}$ where $\delta_1 \in (\frac{1}{12}, \frac{1}{9})$ and $2\delta_2 < \frac{2}{3} - \delta_1$.

Assumption A.4 (Smoothness) The functions $\tilde{p}, \tilde{q}$ and the conditional density f_0, g_0, f_1, g_1 defined below are all $M = 6$ times continuously differentiable with bounded derivatives.

Assumption A.5 (Kernel for estimating propensity scores) (i) L has compact support, is symmetric around zero, integrates to one, and is twice continuously differentiable. (ii) L has an m-th order with $m > 12$

Assumption A.6 (Smoothness of population moments) The propensity score $p_0(.)$, $q_0(.)$, $p_1(.)$ and $q_1(.)$ defined below and the density of (v_{i1}, v_{i2}) are continuously differentiable of order m with bounded derivatives, where $m > 12$.

Assumption A.7 (Bandwidth for estimating propensity scores) $\sigma_n \propto n^{-\gamma/3}$, where

$$\frac{3}{m}\left(\frac{1}{3}+\delta_1\right) < \gamma < \frac{1}{3} - 2\delta_1.$$

Assumption A.8 (Finite second moments) The function χ_i defined in (A.13) has finite second moment.

Proposition 1 *Under Assumptions A.1-A.8,*

$$\sqrt{n}\left(\hat{\gamma}_{CF}-\gamma_0\right)\xrightarrow{d}N\left(0,\Sigma^{-2}\mathbb{E}[\chi_i^2]\right)$$

Proof. By construction, we can write

$$\hat{\gamma}_{CF}-\gamma_0=\frac{\sum_{j\neq i}\left[\omega_{ij,0}(v_{i1}-v_{j2}-\gamma_0)+\omega_{ij,1}(v_{i2}-v_{j1}-\gamma_0)\right]}{\sum_{j\neq i}\left(\omega_{ij,0}+\omega_{ij,1}\right)}. \tag{A.2}$$

The proof of the asymptotic distribution of $\hat{\gamma}_{CF}$ requires us to derive a linear representation for the right-hand side of (A.2).

First, we look for the probability limit of $\frac{1}{n(n-1)}\sum_{j\neq i}\omega_{ij,0}$. Under Assumption A.2, we apply a Taylor expansion of $\omega_{ij,0}$ around the actual conditional expectation in the data-generating process $p_{i0}\equiv p_0(v_i)\equiv E\left[y_{i2}(1-y_{i1})|v_i,y_{i0}=0\right]$ and $q_{j0}\equiv q_0(v_j)\equiv E\left[y_{j1}(1-y_{j2})|v_j,y_{j0}=0\right]$. This allows us to write

$$\tfrac{1}{n(n-1)}\sum_{j\neq i}\omega_{ij,0}=\tfrac{1}{n(n-1)}\sum_{j\neq i}\bar{\omega}_{ij,0}+o_p(n^{-1/4}), \tag{A.3}$$

where

$$\bar{\omega}_{ij,0}\equiv K_h(p_{i0}-q_{j0},v_{j1}-v_{i2})(1-y_{i0})(1-y_{j0}). \tag{A.4}$$

That the remainder term in (A.3) is $o_p(n^{-1/4})$ follows from Assumptions A.2, A.3, A.5, A.6, A.7 and an argument used in Lemma D.3 in Chen, Khan, and Tang (2016). Under Assumption A.2, A.3 and A.6, $E[|\bar{\omega}_{ij,0}|^2]=o(n)$. By an application of the Law of Large Numbers for U-statistics (e.g., Lemma 3.1 in Powell, Stock and Stocker (1989)), $\frac{1}{n(n-1)}\sum_{j\neq i}\bar{\omega}_{ij,0}$ converges in probability to the limit of the expectation of $\bar{\omega}_{ij,0}$ as $n\to\infty$. To evaluate this limit, first note that the conditional expectation of $\bar{\omega}_{ij,0}$ given y_{i0},y_{j0} is

$$(1-y_{i0})(1-y_{j0})\int K_{1h}(p-q)K_{2h}(v_{j1}-v_{i2})f_0(p,v_{i2}|y_{i0})g_0(v_{j1},q|y_{j0})dv_{j1}dpdqdv_{i2},$$

where $K_{1h}\equiv\frac{1}{h_1}K_1(\frac{\cdot}{h_1})$, $K_{2h}\equiv\frac{1}{h_2}K_2(\frac{\cdot}{h_2})$, $f_0(.,.|y_{i0})$ denotes the density of (p_{i0},v_{i2}) given y_{i0}, and $g_0(.,.|y_{j0})$ denotes the density of (v_{j1},q_{j0}) given y_{j0}. By changing variables between v_{j1} and $u\equiv(v_{j1}-v_{i2})/h_2$ while fixing (p,q,v_{i2}), we write this expression as:

$$(1-y_{i0})(1-y_{j0})\int K_{1h}(p-q)f_0(p,v_{i2}|y_{i0})\left(\int K_2(u)g_0(v_{i2}+uh_2,q|y_{j0})du\right)dpdqdv_{i2}.$$

Next, change variables between p and $\tilde{u}=(p-q)/h_1$ while fixing (u,q,v_{i2}), we can write this as

$$(1-y_{i0})(1-y_{j0})\int K_1(\tilde{u})K_2(u)f_0(q+\tilde{u}h_1,v_{i2}|y_{i0})g_0(v_{i2}+uh_2,q|y_{j0})dud\tilde{u}dqdv_{i2}.$$

By the Dominated Convergence Theorem, this converges to the following expression as $h_1, h_2 \to 0$:

$$H_0(y_{i0}, y_{j0}) \equiv (1 - y_{i0})(1 - y_{j0}) \int f_0(q, v_{i2}|y_{i0}) g_0(v_{i2}, q|y_{j0}) dq dv_{i2}. \tag{A.5}$$

where we have used the fact $\int K_1(\tilde{u}) d\tilde{u} = \int K_2(u) du = 1$. Thus the probability limit of $\frac{1}{n(n-1)} \sum_{j \neq i} \bar{\omega}_{ij,0}$ is $E[H_0(y_{i0}, y_{j0})]$. It then follows from (A.3) that $\frac{1}{n(n-1)} \sum_{j \neq i} \omega_{ij,0}$ converges in probability to $E[H_0(y_{i0}, y_{j0})]$. Using an analogous argument, we conclude that $\frac{1}{n(n-1)} \sum_{j \neq i} \omega_{ij,1}$ converges in probability to $E[H_1(y_{i0}, y_{j0})]$ where

$$H_1(y_{i0}, y_{j0}) \equiv y_{i0} y_{j0} \int f_1(v_{i1}, q|y_{i0}) g_1(q, v_{i1}|y_{j0}) dq dv_{i1},$$

and $f_1(.,.|y_{i0})$ is the density of (v_{i1}, p_{i1}) given y_{i0}, and $g_1(.,.|y_{j0})$ the density of (q_{j1}, v_{j2}) given y_{j0}, with $p_1(v_i) \equiv E\left[y_{i2}(1 - y_{i1})|v_i, y_{i0} = 1\right] \equiv p_{i1}$ and $q_1(v_j) \equiv E\left[y_{j1}(1 - y_{j2})|v_j, y_{j0} = 1\right] \equiv q_{j1}$. Combining these results, we have shown that

$$\tfrac{1}{n(n-1)} \textstyle\sum_{j \neq i} (\omega_{ij,0} + \omega_{ij,1}) \xrightarrow{p} \Sigma$$

where

$$\Sigma \equiv E[H_0(y_{i0}, y_{j0}) + H_1(y_{i0}, y_{j0})]. \tag{A.6}$$

and is strictly positive and finite under Assumption A.1.

We now turn to the linear representation of the numerator in the right-hand side of (A.2). The first term in the numerator is:

$$\tfrac{1}{n(n-1)} \textstyle\sum_{i \neq j} \omega_{ij,0}(v_{i1} - v_{j2} - \gamma_0). \tag{A.7}$$

By a second-order Taylor expansion around p_{i0} and q_{j0}, we can write this expression as

$$\tfrac{1}{n(n-1)} \textstyle\sum_{i \neq j} \left[\bar{\omega}_{ij,0}(v_{i1} - v_{j2} - \gamma_0) + \bar{\omega}^{(1)}_{ij,0}(v_{i1} - v_{j2} - \gamma_0)(\hat{p}_{i0} - \hat{q}_{j0} - p_{i0} + q_{j0})\right] + R_n \tag{A.8}$$

where $\bar{\omega}_{ij,0}$ is defined in (A.4) and

$$\bar{\omega}^{(1)}_{ij,0} \equiv \frac{1}{h_1^2} K_1' \left(\frac{p_{i0} - q_{j0}}{h_1}\right) \frac{1}{h_2} K_2 \left(\frac{v_{j1} - v_{i2}}{h_2}\right) (1 - y_{i0})(1 - y_{j0}),$$

and R_n is the second-order term in the Taylor expansion. Under Assumptions A.2, A.3, A.5, A.6 and A.7, R_n is $o_p(n^{-1/2})$ by an argument that follows Lemma D.3 in Chen, Khan, and Tang (2016). The lead term $\frac{1}{n(n-1)} \sum_{i \neq j} [\bar{\omega}_{ij,0}(v_{i1} - v_{j2} - \gamma_0)]$ is $o_p(n^{-1/2})$ by our identification result and under the maintained assumptions. It remains to derive a linear representation of the first-order term (i.e., the second term) in the approximation in (A.8). Consider the first additive component in that term:

$$\tfrac{1}{n(n-1)} \textstyle\sum_{i \neq j} \bar{\omega}^{(1)}_{ij,0}(v_{i1} - v_{j2} - \gamma_0)(\hat{p}_{i0} - p_{i0}). \tag{A.9}$$

Let $\hat{m}_{i0}, \hat{f}_{i0}$ denote the numerator and denominator in the definition of $\hat{p}_{i0}$; let $m_{i0} \equiv E[y_{i2}(1 - y_{i1})(1 - y_{i0})|v_i]f(v_i)$ and $f_{i0} \equiv E(1 - y_{i0}|v_i)f(v_i)$ so that $p_{i0} = m_{i0}/f_{i0}$ by construction. Applying a first-order Taylor expansion of (A.9) around (m_{i0}, f_{i0}), we get

$$\tfrac{1}{n(n-1)} \textstyle\sum_{i \neq j} \frac{\bar{\omega}^{(1)}_{ij,0}(v_{i1} - v_{j2} - \gamma_0)}{f_{i0}} \left[\hat{m}_{i0} - m_{i0} - (\hat{f}_{i0} - f_{i0})p_{i0}\right] + \tilde{R}_n \tag{A.10}$$

where $\tilde{R}_n$ is $o_p(n^{-1/2})$ under Assumption A.2, A.3, A.5, A.6 and A.7. Thus we can write the right-hand side of (A.10) as the sum of a third-order U-statistic and some asymptotically negligible term:

$$\tfrac{1}{n(n-1)(n-2)} \textstyle\sum_{i \neq j \neq s} \varphi_n(\xi_i, \xi_j, \xi_s) + o_p(n^{-1/2}) \tag{A.11}$$

where

$$\varphi_n(\xi_i, \xi_j, \xi_s) \equiv \frac{\bar{\omega}^{(1)}_{ij,0}(v_{i1} - v_{j2} - \gamma_0)}{f_{i0}} L_\sigma\left(v_s - v_i\right)\left(1 - y_{s0}\right)\left[y_{s2}(1 - y_{s1}) - p_{i0}\right]$$

where $\xi_i \equiv (y_i, v_i, p_i)$ with $y_i \equiv (y_{i0}, y_{i1}, y_{i2})$, $v_i \equiv (v_{i1}, v_{i2})$ and $p_i \equiv (p_{i0}, p_{i1})$. By Lemma 3.1 in Powell, Stock and Stocker (1989), we can write (A.11) as

$$\theta_n + \frac{1}{n} \textstyle\sum_{i=1}^n \sum_{l=1}^3 [r_n^{(l)}(\xi_i) - \theta_n] + o_p(n^{-1/2})$$

where $r_n^{(l)}(\xi) \equiv E[\varphi_n(\xi_1, \xi_2, \xi_3)|\xi_l = \xi]$ and $\theta_n \equiv E[\varphi_n(\xi_i, \xi_j, \xi_l)]$. Using change of variables and Taylor expansion, as well as the smooth conditions on the kernel $L(.)$, we can show that $\theta_n = E[r_n^{(1)}(\xi_i)] = o(n^{-1/2})$. Furthermore by the Dominated Convergence Theorem, the unconditional variance of $r_n^{(1)}(\xi_i)$ is $o(1)$ under maintained assumptions. It then follows from the Chebyshev's Inequality that

$$\frac{1}{n} \textstyle\sum_{i=1}^n \left[r_n^{(l)}(\xi_i) - \theta_n\right] = o_p(n^{-1/2}) \text{ for } l = 1, 2.$$

By an argument similar to Chen, Khan, and Tang (2016), under Assumptions A2-A8, the third-order U-statistic in (A.11) has the following representation:

$$\frac{1}{n} \sum_{i=1}^n \Gamma_0 \left\{y_{i2}(1 - y_{i1})) - E[y_{i2}(1 - y_{i1})|v_i]\right\} + o_p(n^{-1/2}) \tag{A.12}$$

where Γ_0 is the limit of the following expectation as $n \to \infty$ and $h_1, h_2 \to 0$:

$$E\left[\frac{1}{h_1^2} K_1'\left(\frac{p_{i0} - q_{j0}}{h_1}\right) \frac{1}{h_2} K_{2h}\left(\frac{v_{j1} - v_{i2}}{h_2}\right)(1 - y_{i0})(1 - y_{j0})(v_{i1} - v_{j2} - \gamma_0)\right].$$

That is, $\Gamma_0 = E[(1 - y_{i0})(1 - y_{j0})\mathcal{H}(y_{i0}, y_{j0})]$ with

$$\mathcal{H}(y_{i0}, y_{j0}) \equiv \int G_0(q, q, v_{i2}, v_{i2}, y_{i0}) g_0(v_{i2}, q|y_{j0}) dq dv_{i2},$$

where

$$G_0(p,q,v_{i2},v_{j1}) \equiv -\frac{\partial\{f_0(p,v_{i2}|y_{i0})\,[\tilde{p}_0(p,v_{i2})-\tilde{q}_0(q,v_{j1})-\gamma_0]\}}{\partial p};$$

with $\tilde{p}_0(t,v_{i2}) \equiv \inf\{v_{i1} : p_0(v_{i1},v_{i2}) \leq t\}$; $\tilde{q}_0(t,v_{j1}) \equiv \inf\{v_{j2} : q_0(v_{j1},v_{j2}) \leq t\}$; and f_0 and g_0 denote the joint density of (p_{i0},v_{i2}) and (v_{j1},q_{j0}) conditional on y_{i0} and y_{j0} respectively.

By an analogous argument, the linear representation of

$$\frac{1}{n(n-1)}\sum_{i\neq j}\bar{\omega}^{(1)}_{ij,0}(v_{i1}-v_{j2}-\gamma_0)(\hat{q}_{j0}-q_{i0}).$$

is similar to (A.12), only with $y_{i2}(1-y_{i1})$ replaced by $y_{i1}(1-y_{i2})$. Combining these results, we get the following linear representation of (A.7) as

$$\frac{1}{n}\textstyle\sum_{i=1}^n \Gamma_0\left[y_{i2}-y_{i1}-E\left(y_{i2}-y_{i1}|v_i\right)\right]+o_p(n^{-1/2}).$$

We can use identical arguments to the second part of the numerator:

$$\frac{1}{n(n-1)}\textstyle\sum_{i\neq j}\omega_{ij,1}(v_{i2}-v_{j1}-\gamma_0)$$

and derive the following asymptotic linear representation:

$$\frac{1}{n}\textstyle\sum_{i=1}^n \Gamma_1\left[y_{i2}-y_{i1}-E\left(y_{i2}-y_{i1}|v_i\right)\right]+o_p(n^{-1/2})$$

where $\Gamma_1 = E[y_{i0}y_{j0}\tilde{\mathcal{H}}(y_{i0},y_{j0})]$ with

$$\tilde{\mathcal{H}}(y_{i0},y_{j0}) \equiv \int G_1(q,q,v_{i1},v_{i1},y_{i0})g_1(q,v_{i1}|y_{j0})dq dv_{i1}$$

and

$$G_1(p,q,v_{i1},v_{j2}) \equiv -\frac{\partial\left\{f_1(v_{i1},p|y_{i0})\,[\tilde{p}_1(v_{i1},p)-\tilde{q}_1(v_{j2},q)-\gamma_0]\right\}}{\partial p}$$

and $\tilde{p}_1(v_{i1},t) \equiv \inf\{v_{i2} : p_1(v_{i1},v_{i2}) \geq t\}$ and $\tilde{q}_1(v_{j2},t) \equiv \inf\{v_{j1} : q_1(v_{j1},v_{j2}) \geq t\}$; and $f_1(.,.|y_{i0})$ and $g_1(.,.|y_{j0})$ denote the density of (v_{i1},p_{i1}) and (q_{j1},v_{j2}) conditional on y_{i0} and y_{j0} respectively.

Gathering these results (i.e., the probability limit of the denominator and the linear representation of the numerator), we get the following asymptotic linear representation of our closed-form estimator:

$$\hat{\gamma}_{CF}-\gamma_0 = \Sigma^{-1}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n \chi_i\right)+o_p(n^{-1/2})$$

where

$$\chi_i \equiv (\Gamma_0+\Gamma_1)\left[\Delta y_i - E\left(\Delta y_i|v_i\right)\right], \tag{A.13}$$

with $\Delta y_i \equiv y_{i2}-y_{i1}$.

A.2 Weighted Maximum Rank Correlation Estimator

Recall the weighted maximum rank correlation estimator is

$$\hat{\gamma}_{MR} \equiv \arg\max_{\gamma} \frac{1}{n(n-1)} \sum\nolimits_{j \neq i} \left[\tilde{\omega}_{ij,0} G_{ij,0}(\gamma) + \tilde{\omega}_{ij,1} G_{ij,1}(\gamma)\right], \tag{A.14}$$

where

$$\begin{aligned} G_{n,0}(\gamma) &\equiv 1\{d_{i,01} > d_{j,10}\}1\{v_{j2} + \gamma > v_{i1}\} + 1\{d_{i,01} < d_{j,10}\}1\{v_{j2} + \gamma < v_{i1}\} \\ G_{n,1}(\gamma) &\equiv 1\{d_{i,01} > d_{j,10}\}1\{v_{i2} > v_{j1} + \gamma\} + 1\{d_{i,01} < d_{j,10}\}1\{v_{i2} < v_{j1} + \gamma\} \end{aligned}$$

with

$$\begin{aligned} d_{i,01} &\equiv (1 - y_{i1})y_{i2} \ , \ d_{j,10} \equiv y_{j1}(1 - y_{j2}), \\ \tilde{\omega}_{ij,0} &\equiv \tilde{K}_h(v_{j1} - v_{i2})(1 - y_{i0})(1 - y_{j0}) \ , \ \tilde{\omega}_{ij,1} \equiv \tilde{K}_h(v_{j2} - v_{i1})y_{i0}y_{j0} \end{aligned} \tag{A.15}$$

and $\tilde{K}_h(.) \equiv \frac{1}{h}\tilde{K}(\frac{\cdot}{h})$ being shorthand for kernel smoothing.

Before stating the regularity conditions, we define the following functions:

$$\begin{aligned} \mathcal{G}_r(v_i, v_j, y_{i0}, y_{j0}) &= E[d_{i,01} \geq d_{j,10}]|v_i, v_j, y_{0i}, y_{0j}] \\ \hat{\Upsilon}_n(\gamma) &= \frac{1}{n(n-1)} \sum\nolimits_{j \neq i} \left[\tilde{\omega}_{ij,0} G_{ij,0}(\gamma) + \tilde{\omega}_{ij,1} G_{ij,1}(\gamma)\right] \\ \Upsilon_n(\gamma, y_{0i}, y_{0j}, v_i, v_j) &= E[\hat{\Upsilon}_n(\gamma)|y_{0i}, v_i, y_{0j}, v_j] \\ \bar{\Upsilon}_{1n}(\gamma, y_{0i}, v_i) &= E[\Upsilon_n(\gamma, y_{0i}, v_i, y_{0j}, v_j)|y_{0i}, v_i] \\ \bar{\Upsilon}_{2n}(\gamma, y_{0j}, v_j) &= E[\Upsilon_n(\gamma, y_{0i}, v_i, y_{0j}, v_j)|y_{0j}, v_j] \\ \Upsilon_n(\gamma) &= E[\Upsilon_n(\gamma, y_{0i}, v_i, y_{0j}, v_j)] \\ \Upsilon_0(\gamma) &= \lim_{n \to \infty} \Upsilon_n(\gamma) \end{aligned}$$

Throughout this part of the appendix, we maintain that the true parameter γ_0 lies in the interior of Ξ_0, a compact parameter space (interval) on the real line. Our arguments for the limiting distribution theory for the rank estimator are based on the following conditions:

Assumption MR.1 The constant Σ_0, (defined formally in (A.19)) is positive and finite.

Assumption MR.2 (Kernel for matching) (i) Let $K(\cdot)$ where K has compact support, is symmetric around 0, integrate to 1, is twice continuously differentiable and of order M. (ii) $\sup_{t \in \mathbb{R}} h^{-1}|K(t/h)|$, is $O(1)$ as $h \to 0$.

Assumption MR.3 (Bandwidth for matching) $nh_n^M \to 0$ and $nh_n \to \infty$.

Assumption MR.4 (Smoothness) The functions G_r and the joint density of v_i are all M times continuously differentiable with bounded derivatives.

Assumption MR.5 (Smoothness) The function $\Upsilon_n(\gamma)$ is twice continuously differentiable in γ for all γ in a neighborhood of γ_0 and all n.

Assumption MR.6 (Finite moments) The random variable χ_{1i}, defined in (A.26) has finite second moment.

Proposition 2 *Under Assumptions EM1-EM4 and MR.0-MR.6,*

$$\sqrt{n}\left(\hat{\gamma}_{MR} - \gamma_0\right) \xrightarrow{d} N\left(0, \Sigma_0^{-2}\mathbb{E}[\chi_{1i}^2]\right)$$

where χ_{1i} is a mean 0 random variable formally defined in (A.26).

Proof: We note that the objective function in (A.14) is not smooth in the parameter γ which complicates analysis in the sense that the usual linearization method based on mean value expansions of the sample objective function is not feasible. Nonetheless we can show that the "limiting" objective function, $\Upsilon(\gamma)$ is smooth and work with its quadratic expansion in a neighborhood of $\gamma = \gamma_0$. This approach would be similar to that taken in Sherman (1993), but the presence of the kernel function and bandwidth in our objective function here further complicates things so we make the necessary adjustments used in, e.g. Sherman (1994), Abrevaya, Hausman, and Khan (2010).

We begin by deriving the form of the limiting objective function $\Upsilon_0(\gamma)$. To do so we evaluate the expectation of the term in the double sum in the definition of $\hat{\Upsilon}_n(\gamma)$. Like in the previous proof we will focus on the first "half" as identical arguments can be used for the second half. Taking the expectation we first condition on y_{i0}, y_{j0}, v_i, v_j as before. This gives us the term:

$$(1 - y_{i0})(1 - y_{j0})K_h(v_{j1} - v_{i2})I[v_{j2} + \gamma > v_{i1}]\mathcal{G}_{ra}(v_i, v_j, y_{i0}, y_{j0}) \tag{A.16}$$

where

$$\mathcal{G}_{ra}(v_i, v_j, y_{i0}, y_{j0}) \equiv E[d_{i,01} > d_{j,10}]|v_i, v_j, y_{0i}, y_{0j}] \tag{A.17}$$

we now take the expectation of (A.16) with respect to v_i, v_j, conditional on y_{0i}, y_{0j}. Like before we will change variables $u = (v_{j1} - v_{i2})/h$ yielding an integral of the form for the first half of $G_{ij,0}(\gamma)$:

$$(1-y_{0i})(1-y_{0j})\int K(u)I[v_{j2}+\gamma > v_{i1}]\mathcal{G}_r(v_i, uh+v_{i2}, v_{j2}, y_{i0}, y_{j0})f_1(v_i|y_{0i})f_2(v_{i2}+uh, v_{j2}|y_{0j})dv_i dv_{j2}$$

where above $f_1(\cdot|\cdot)$ denotes the conditional density function of v_i conditional on y_{0i} and $f_2(\cdot|\cdot)$ the conditional density function for v_j given y_{0j}. Taking limits as $h \to 0$ results in a function of y_{0i}, y_{0j}, γ, which we denote here by $S_{1a}(y_{i0}, y_{j0}, \gamma)$. Crucially, given our smoothness assumptions on $K(\cdot)$ and the density of v_i, v_j, $S_{1a}(y_{i0}, y_{j0}, \gamma)$ is a smooth function in γ even though γ is inside an indicator function inside the integral. We can apply identical arguments to the second half of $G_{ij,0}$ now working with the function

$$\mathcal{G}_{rb}(v_i, v_j, y_{i0}, y_{j0}) \equiv E[d_{i,01} < d_{j,10}]|v_i, v_j, y_{0i}, y_{0j}] \tag{A.18}$$

Now when we take limits as $h \to 0$, the resulting function of y_{0i}, y_{0j}, γ will be denoted by $S_{1b}(y_{0i}, y_{0j}, \gamma)$. So we can define

$$\mathcal{S}_1(y_{0i}, y_{0j}, \gamma) \equiv \mathcal{S}_{1a}(y_{0i}, y_{0j}, \gamma) + \mathcal{S}_{1b}(y_{0i}, y_{0j}, \gamma)$$

Also, note we can use identical arguments to express the second "half" of the summand in the objective function, involving $G_{ij,1}(\gamma)$, as $y_{0i}y_{0j}$ times $S_2(y_{i0}, y_{j0}, \gamma)$. So by a LLN for U-processes (see, e.g. Sherman (1994)) we can express:

$$\Upsilon_0(\gamma) = E[(1 - y_{0i})(1 - y_{0j})\mathcal{S}_1(y_{0i}, y_{0j}, \gamma) + y_{0i}y_{0j}\mathcal{S}_2(y_{0i}, y_{0j}, \gamma)]$$

Note that $\Upsilon_0(\gamma)$ is smooth in γ. This permits the following second order expansion of $\Upsilon(\gamma)$ around $\Upsilon(\gamma_0)$:

$$\Upsilon(\gamma) = \Upsilon(\gamma_0) + \Upsilon'(\gamma_0)(\gamma - \gamma_0) + \frac{1}{2}\Upsilon''(\gamma_0)(\gamma - \gamma_0)^2 + o(\gamma - \gamma_0)^2$$

We note that $\Upsilon'(\gamma_0) = 0$ by our point identification result. The second derivative of $\Upsilon(\cdot)$ evaluated at $\gamma = \gamma_0$ relates directly to the Hessian term in our limiting distribution theory:

$$\Sigma_0 = \Upsilon''(\gamma_0) \tag{A.19}$$

To complete our linear representation, we return to (A.14) and work with its Hoeffding decomposition. (see, e.g. Sherman (1994).)

The next term in the decomposition is of the form

$$\frac{1}{n}\sum_{i=1}^{n}(\bar{\Upsilon}_{1n}(\gamma, y_{0i}, v_i) - \Upsilon_n(\gamma)) \tag{A.20}$$

We can handle $\bar{\Upsilon}_{1n}(\gamma, y_{0i}, v_i)$ exactly as we handled $\Upsilon(\gamma)$: changing variables inside the integral with respect to the regressor density. As before this will result in a smooth function of γ which we can again expand around γ_0. Denote the resulting smooth function as $F_1(\gamma, y_{0i}, v_i)$, where

$$\mathcal{F}_1(\gamma, y_{0i}, v_i) = (1 - y_{0i})\xi(v_i, y_{0i}) \tag{A.21}$$

where

$$\begin{aligned}
\xi(v_i, y_{0i}) &= f_1(v_i|y_{0i})E\left[(1-y_{0j})\int I[v_{j2}+\gamma > v_{i1}]\mathcal{G}_{ra}(v_i, v_{i2}, v_{j2}, y_{i0}, y_{j0})f_2(v_{i2}, v_{j2}|y_{0j})dv_{j2}\right] \\
&+ f_1(v_i|y_{0i})E\left[(1-y_{0j})\int I[v_{j2}+\gamma < v_{i1}]\mathcal{G}_{rb}(v_i, v_{i2}, v_{j2}, y_{i0}, y_{j0})f_2(v_{i2}, v_{j2}|y_{0j})dv_{j2}\right]
\end{aligned}$$

so after the expansion the above average in (A.20) can be expressed as:

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{F}_1'(\gamma_0, y_{0i}, v_i)(\gamma - \gamma_0) + r_n \tag{A.22}$$

where the remainder term r_n can shown to be negligible ($o_p(n^{-1})$, uniformly in γ in shrinking neighborhoods of γ_0) using the arguments in Abrevaya, Hausman, and Khan (2010). We can conduct the same exercise for the next term in the decomposition:

$$\frac{1}{n}\sum_{j=1}^{n}\bar{\Upsilon}_{2n}(\gamma, y_{0j}, v_j) - \Upsilon_n(\gamma) \tag{A.23}$$

Using the same arguments we will express this as:

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{F}_2'(\gamma_0, y_{0i}, v_i)(\gamma - \gamma_0) + r_n \tag{A.24}$$

where

$$\mathcal{F}_2(\gamma, y_{0j}, v_j) = (1 - y_{0j})\xi_2(v_j, y_{0j}) \tag{A.25}$$

where

$$\begin{aligned}
\xi_2(v_j, y_{0j}) &= f_2(v_j|y_{0j})E\left[(1-y_{0i})\int I[v_{j2}+\gamma > v_{i1}]\mathcal{G}_{ra}(v_i, v_{i2}, v_{j2}, y_{i0}, y_{j0})f_1(v_{i2}, v_{j2}|y_{0j})dv_{i2}\right] \\
&+ f_2(v_j|y_{0j})E\left[(1-y_{0i})\int I[v_{j2}+\gamma < v_{i1}]\mathcal{G}_{rb}(v_i, v_{i2}, v_{j2}, y_{i0}, y_{j0})f_2(v_{i2}, v_{j2}|y_{0j})dv_{i2}\right]
\end{aligned}$$

Note that $F_1'(\gamma_0, y_{0i}, v_i), F_2'(\gamma_0, y_{0i}, v_i)$ are each mean zero random variables. They came from the linear term in the first "half" of the objective , that involved $G_{ij,0}(\gamma)$. One could use similar arguments for the second "half" of the objective function that involved $G_{ij,1}(\gamma)$. We denote these mean 0 random variables as $F_3'(\gamma_0, y_{0i}, v_i), F_4'(\gamma_0, y_{0i}, v_i)$

Collectively they relate to the influence function in our linear representation in the following way:

$$\chi_{1i} = \mathcal{F}_1'(\gamma_0, y_{0i}, v_i) + \mathcal{F}_2'(\gamma_0, y_{0i}, v_i) + \mathcal{F}_3'(\gamma_0, y_{0i}, v_i) + \mathcal{F}_4'(\gamma_0, y_{0i}, v_i) \tag{A.26}$$

Under the finiteness of the second moment in Assumption MR.6, this implies $\hat{\gamma}_{MR}$ is root-n CAN as stated in Proposition 2.

B Exclusion Restriction in Static Binary Choice Panel Data Models

We can also apply a pairwise approach under the exclusion restriction to estimate static binary choice panel data models which do not include any lagged dependent variable. Consider the following model:

$$y_{it} = 1[w_{it}\beta_0 + v_{it} + \alpha_i + \epsilon_{it} \geq 0] \text{ for } t = 1, 2.$$

where $v_{it} \in \mathbb{R}$ and $w_{it} \in \mathbb{R}^L$ does not include any lagged dependent variable y_{it-1}. Let $y_i \equiv (y_{i1}, y_{i2})$, $w_i \equiv (w_{i1}, w_{i2})$ and $v_i \equiv (v_{i1}, v_{i2})$. Assume the data contains i.i.d. observations of (y_i, w_i, v_i) for $i = 1, 2, ..., N$.

Assumption B.1 $(\epsilon_{i1}, \epsilon_{i2}, \alpha_i)$ are independent of v_i conditional on w_i.

Assumption B.2 The marginal distribution of ϵ_{it} conditional on (α_i, w_i) is continuous with positive density over $\mathbb{R}$, and is the same for $t = 1, 2$.

For $t = 1, 2$, let

$$\pi_t(w_i, v_i, \alpha_i) \equiv E(y_{it}|w_i, v_i, \alpha_i) = \Pr(-\epsilon_{it} \leq w_{it}\beta_0 + v_{it} + \alpha_i | w_i, \alpha_i)$$

where the equality holds because of Assumption B.1. Furthermore, define

$$p_t(w_i, v_i) \equiv E(y_{it}|w_i, v_i) = \int E(y_{it}|w_i, v_i, \alpha_i) dF(\alpha_i|w_i, v_i) = \int \pi_t(w_i, v_i, \alpha_i) dF(\alpha_i|w_i)$$

where the last equality is again due to Assumption B.1. Note that $\pi_t(w_i, v_i, \alpha_i)$ is not identified from the data because the fixed effect α_i is not reported in the data. However, $p_t(w_i, v_i)$ is identified from the data by definition.

Now consider a pair of observations $i, j = 1, 2, ..., N$ such that $w_i = w_j$. Then it can be shown that under Assumptions B.1 and B.2,

$$p_1(w_i, v_i) = p_2(w_j, v_j) \text{ if and only if } w_{i1}\beta_0 + v_{i1} = w_{j2}\beta_0 + v_{j2}. \tag{B.1}$$

To see why (B.1) is true, suppose $w_i = w_j$ and $w_{i1}\beta_0 + v_{i1} = w_{j2}\beta_0 + v_{j2}$. Then under Assumption B.1,

$$\begin{aligned}
\pi_1(w_i, v_i, \alpha) &\equiv \Pr(-\epsilon_{i1} \leq w_{i1}\beta_0 + v_{i1} + \alpha_i | w_i, \alpha_i = \alpha) \\
&= \Pr(-\epsilon_{j2} \leq w_{j2}\beta_0 + v_{j2} + \alpha_i | w_j, \alpha_j = \alpha) \equiv \pi_2(w_j, v_j, \alpha)
\end{aligned}$$

for all α on the support of α_i given w_i (or, equivalently, given w_j, because $w_i = w_j$). The equality above holds because of the fact that observations i and j are independent draws from the same data-generating process, Assumption B.2 as well as that $w_i = w_j$ and $w_{i1}\beta_0 + v_{i1} = w_{j2}\beta_0 + v_{j2}$. Thus

$$p_1(w_i, v_i) = \int \pi_1(w_i, v_i, \alpha) dF(\alpha | w_i) = \int \pi_2(w_j, v_j, \alpha) dF(\alpha | w_j) = p_2(w_j, v_j)$$

because $w_i = w_j$. Next, suppose $w_i = w_j$ and $w_{i1}\beta_0 + v_{i1} > w_{j2}\beta_0 + v_{j2}$. Then $\pi_1(w_i, v_i, \alpha) > \pi_2(w_j, v_j, \alpha)$ for all α, and $p_1(w_i, v_i) > p_2(w_j, v_j)$ using a similar argument. Given this result, the coefficient β_0 is point identified under the following rank condition.

Assumption B.3 The support of $w_{i2} - w_{i1}$ does not lie in any proper linear subspace of $\mathbb{R}^L$.

An implication of Assumption B.3 is that the support of $w_{j2} - w_{i1}$ conditional on $w_i = w_j$ is not contained in any proper linear subspace of $\mathbb{R}^L$. This implies we can recover β_0 by regressing $v_{j2} - v_{i1}$ on $w_{i1} - w_{j2}$ conditional on $w_i = w_j$.

Based on this identification argument, we propose a closed-form estimator for β_0 as follows. Let $K_h(\cdot) \equiv \frac{1}{h} K\left(\frac{\cdot}{h}\right)$, where K is a multivariate (product) kernel function and $h \in \mathbb{R}^{L+2}_{++}$ is a sequence of bandwidth vectors. Define a data-dependent, pairwise weight function:

$$\omega_{ij} = K_h(\hat{p}_{j2} - \hat{p}_{i1}, w_j - w_i),$$

where

$$\hat{p}_{it} = \frac{\sum_{l \neq i} y_{lt} \mathcal{K}_\sigma(v_l - v_i, w_l - w_i)}{\sum_{l \neq i} \mathcal{K}_\sigma(v_l - v_i, w_l - w_i)}$$

is a kernel regression of y_{it} condition on (w_i, v_i), with $\mathcal{K}$ being a multivariate (product) kernel function and $\sigma \in \mathbb{R}^{L+1}$ a sequence of bandwidth vectors. The closed-form estimator of β_0 is

$$\hat{\beta}_{CF} = \left(\sum_{i=1}^N \sum_{j \neq i} \omega_{ij}(w_{j2} - w_{i1})(w_{j2} - w_{i1})'\right)^{-1} \left(\sum_{i=1}^N \sum_{j \neq i} \omega_{ij}(w_{j2} - w_{i1})(v_{i1} - v_{j2})\right).$$

Under suitable kernel regularity conditions, which are standard in the literature and similar to those in Appendix A, this estimator converges at the parametric rate with a limiting normal distribution.

Alternatively, we can define a weighted maximum rank correlation estimator for β_0. To understand how it works, note that conditional on $w_i = w_j$, the ranking between $p_1(w_i, v_i)$ and $p_2(w_j, v_j)$ is identical to the ranking between $w_{i1}\beta_0 + v_{i1}$ and $w_{j2}\beta_0 + v_{j2}$ once conditioning on $w_i = w_j$. Thus a maximum rank correlation estimator can be constructed as follows:

$$\hat{\beta}_{MR} \equiv \max_{\beta} \frac{1}{n(n-1)} \sum\nolimits_i \sum\nolimits_{j \neq i} \tilde{K}_{\tilde{\sigma}}(w_i - w_j) G_{ij}(\beta), \tag{B.2}$$

where

$$G_{ij}(\beta) \equiv 1\{y_{i1} > y_{j2}\}1\{w_{i1}\beta + v_{i1} > w_{j2}\beta + v_{j2}\} + 1\{y_{i1} < y_{j2}\}1\{w_{i1}\beta + v_{i1} < w_{j2}\beta + v_{j2}\}.$$

and $\tilde{K}_{\tilde{\sigma}}(.) \equiv \frac{1}{\tilde{\sigma}}\tilde{K}(\frac{.}{\tilde{\sigma}})$, with $\tilde{K}$ being a product kernel and $\tilde{\sigma} \in \mathbb{R}^L$ a sequence of bandwidth vectors. Note the double sum in (B.2) is over ordered pairs of i and j. This is because $G_{ij}(\beta) \neq G_{ji}(\beta)$ in general. A tradeoff between computational complexity and tuning parameters (similar to the one discussed in the text) exists between $\hat{\beta}_{CF}$ and $\hat{\beta}_{MR}$ proposed above.