Fixed Effects Binary Choice Models with Three or More Periods*

Laurent Davezies[†] Xavier D'Haultfœuille[‡] Martin Mugnier[§]

Abstract

We consider fixed effects binary choice models with a fixed number of periods T and without a large support condition on the regressors. If the time-varying unobserved terms are i.i.d. with known distribution F, Chamberlain (2010) shows that the common slope parameter is point-identified if and only if F is logistic. However, he considers in his proof only $T = 2$. We show that actually, the result does not generalize to $T \geq 3$: the common slope parameter and some parameters of the distribution of the shocks can be identified when F belongs to a family including the logit distribution. Identification is based on a conditional moment restriction. We give necessary and sufficient conditions on the covariates for this restriction to identify the parameters. In addition, we show that under mild conditions, the corresponding GMM estimator reaches the semiparametric efficiency bound when $T = 3$.

Keywords: Binary choice model, panel data, point identification, conditional moment restrictions.
JEL Codes: C14, C23, C25.

*We would like to thank Pascal Lavergne for his comments.
[†]CREST, laurent.davezies@ensae.fr
[‡]CREST, xavier.dhaultfoeuille@ensae.fr.
[§]CREST, martin.mugnier@ensae.fr.

1 Introduction

In this paper, we revisit the classical binary choice model with fixed effects. Specifically, let T denote the number of periods and let us suppose to observe, for individual i, $(Y_{it}, X_{it})_{t=1,...,T}$ with

$$Y_{it} = \mathbb{1}\{X_{it}'\beta_0 + \gamma_i - \varepsilon_{it} \geq 0\} \tag{1.1}$$

where $\beta_0 \in \mathbb{R}^K$ is unknown and $\varepsilon_{it} \in \mathbb{R}$ is an idiosyncratic shock. The nonlinear nature of the model and the absence of restriction on the distribution of γ_i conditional on $X_i := (X_{i1}, \dots, X_{iT})$ renders the identification of β_0 difficult. Rasch (1960) shows that if the $(\varepsilon_{it})_{t=1,...,T}$ are i.i.d. with a logitistic distribution, a conditional maximum likelihood can be used to identify and estimate β_0. Chamberlain (2010) establishes a striking converse of Rasch's result: if the $(\varepsilon_{it})_{t=1,...,T}$ are i.i.d. with distribution F and the support of X_i is bounded, β_0 is point identified only if F is logistic. Other papers have circumvented such a negative result by either considering large support regressors (see in particular Manski, 1987; Honore and Lewbel, 2002) or allowing for dependence between the shocks (see Magnac, 2004).

It turns out, however, that Chamberlain (2010) only proves his result for $T = 2$. And in fact, we show that his result does not generalize to $T \geq 3$. Specifically, we consider distributions F satisfying

$$\frac{F(x)}{1-F(x)} = \sum_{k=1}^{\tau} w_k \exp(\lambda_{0k} x) \ \text{ or } \ \frac{1-F(x)}{F(x)} = \sum_{k=1}^{\tau} w_k \exp(-\lambda_{0k} x), \tag{1.2}$$

with $T \geq \tau + 1$, $(w_1, ..., w_\tau) \in (0, \infty) \times [0, \infty)^{\tau-1}$ and $1 = \lambda_{01} < \ldots < \lambda_{0\tau}$. We study the identification of β_0, assuming that $\lambda_0 := (\lambda_{01}, \dots, \lambda_{0\tau})$ is known, but also of $\theta_0 := (\beta_0, \lambda_0)$. In both cases, the weights $w_1, \dots, w_\tau$ remain unknown, thus allowing for much more flexibility on the distribution of ε_{it} than in the logit case. Our main insight is that for any F satisfying (1.2), a conditional moment restriction holds. We then give necessary as well as sufficient conditions for such moment restrictions to identify β_0 or θ_0. The necessary conditions show for instance that with $\tau \geq 2$, point identification of β_0 cannot be achieved with a single, binary. X_{it}. On the other hand, our sufficient conditions imply that at least if γ is constant, θ_0 is identified if conditional on $(X_{j,t'})_{(j,t')\neq(i,t)}$, X_{it} takes at least 2τ values. Note that Johnson (2004) considers the same family with $\tau = 2$ and $T = 3$. However, he does not study the general case and does not show any formal identification result based on the corresponding moment conditions.

Obviously, the conditional moment condition can be used to construct GMM estimators. This means, in particular, that $\sqrt{n}$-consistent estimation is possible beyond the logit case when $T > 2$, overturning again the negative results of Chamberlain (2010) and Magnac (2004). Further, we show that if $T = 3$ and mild additional restrictions hold, the optimal GMM estimator based on our conditional moment conditions reaches the semiparametric efficiency bound of the model. This means that at least when $T = 3$, these moment conditions contain all the information of the model. We also show through simulations that this information is sufficient to form rather precise estimators for usual sample sizes.

The remainder of the paper is organized as follows. Section 2 gives a necessary and sufficient conditions for point identification of β_0 and the λ_{0j}. Section 3 discusses estimation and the semiparametric efficiency bound of the model. Section 4 reports results from a Monte-Carlo study. Section 5 concludes. All the proofs are collected in the appendix.

2 Identification

2.1 The model and moment conditions

We drop the subscript i in the absence of ambiguity and let $Y := (Y_1', \dots, Y_T')'$, $X := (X_1', \dots, X_T')'$ and $X_t := (X_{1,t}, \dots, X_{K,t})'$. For any set $A \subset \mathbb{R}^p$ (for any $p \geq 1$), we let $A^* := A\backslash\{0\}$ and $|A|$ denote the cardinal of A. Hereafter, we maintain the following conditions.

Assumption 1 (Binary choice panel model) *Equation (1.1) holds and:*

1. (X, γ) *and* $(\varepsilon_t)_{1\leq t\leq T}$ *are independent and the* $(\varepsilon_t)_{1\leq t\leq T}$ *are i.i.d. with a known cumulative distribution function (cdf)* F*.*

2. *For all* (k,t)*,* $\mathbb{E}[X_{k,t}^2] < \infty$*.*

3. $\beta_0 \in \mathbb{R}^{K*}$.

The first condition is also considered in Chamberlain (2010). The second condition is a standard moment restriction on the covariates. Finally, we exclude in the third condition the case $\beta_0 = 0$ here. This case can be treated separately, as the following proposition shows.

Proposition 2.1 *Suppose that Assumption 1 holds, F is strictly increasing on $\mathbb{R}$ and there exist $(t,t') \in \{1,\dots,T\}^2$ such that $\mathbb{E}[(X_t - X_{t'})(X_t - X_{t'})']$ is nonsingular. Then $\beta_0 = 0$ if and only if*

$$\mathbb{P}(Y_t = 1, Y_{t'} = 0 | Y_t + Y_{t'} = 1, X_t, X_{t'}) = \frac{1}{2} \quad a.s. \tag{2.1}$$

Condition (2.1) can be tested by a specification test on the nonparametric regression of $D = Y_t(1 - Y_t')$ on $(X_t, X_{t'})$, conditional on the event $Y_t + Y_{t'} = 1$. See, e.g., Bierens (1990) or Hong and White (1995).

Turning to identification on $\mathbb{R}^{K*}$, we first recall the negative result of Chamberlain (2010).

Theorem 2.2 *Suppose that $T = 2$, Assumption 1 holds, F is strictly increasing on $\mathbb{R}$ with bounded, continuous derivative and Supp(X) is compact. If, for all $\beta_0 \in \mathbb{R}^{K*}$, β_0 is identified, then $F(x)/(1 - F(x)) = w \exp(\lambda x)$ for some $(w, \lambda) \in \mathbb{R}^{+*2}$.*

Our results below imply, however, that this negative result does not generalize to $T > 2$. To this end, we consider a family of distribution that includes the logistic distribution and is defined as follows.[1] Hereafter, Λ_τ denotes a subset of $\{(\lambda_1, \dots, \lambda_\tau) \in \mathbb{R}^\tau : 1 = \lambda_1 < \dots < \lambda_\tau\}$.

Assumption 2 ("Generalized" logistic distributions) *There exists a known $\tau \in \{1, \dots, T-1\}$, unknown $w := (w_1, \dots, w_\tau) \in (0,\infty) \times [0,\infty)^{\tau-1}$ and $\lambda_0 := (\lambda_{01}, \dots, \lambda_{0\tau})' \in \Lambda_\tau$ such that:*

$$\begin{aligned} \textit{Either} \quad \frac{F(x)}{1-F(x)} &= \sum_{j=1}^{\tau} w_j \exp(\lambda_{0j} x) \quad \textit{(First type)}, \\ \textit{or} \quad \frac{1-F(x)}{F(x)} &= \sum_{j=1}^{\tau} w_j \exp(-\lambda_{0j} x) \quad \textit{(Second type)}. \end{aligned}$$

[1] Noteworthy, the family of "generalized" logistic distributions we consider differs from those introduced by Balakrishnan and Leung (1988) and Stukel (1988).

We fix $\min\{\lambda_{01}, \ldots, \lambda_{0\tau}\}$ to 1 as the scale of the latent variable $X_{it}'\beta_0 + \gamma_i - \varepsilon_{it}$ is not identified. Also, if F is of the second type, then one can show that the cdf of $-\varepsilon_{it}$ is of the first type. Thus, up to changing (Y_t, X_t) into $(1 - Y_t, -X_t)$, we can assume without loss of generality, as we do afterwards, that F is of the first type. We shall see that $\tau + 1$ periods are sufficient to achieve identification. Hence, we assume, again without loss of generality, that $T = \tau + 1$: if $T > \tau + 1$, we can always focus on $\tau + 1$ periods.

We consider the identification of not only β_0 but also λ_0. We then let $\theta_0 := (\beta_0', \lambda_0')'$ and $\Theta_0 := (\mathbb{R}^{K*}) \times \Lambda_\tau$. We also define, for any $(y, x, \theta) \in \{0,1\}^T \times \text{Supp}(X) \times \Theta_0$,

$$m(y, x; \theta) := \sum_{t=1}^{T} \mathbb{1}\{y_t = 1, y_{t'} = 0 \ \forall t' \neq t\} M_t(x; \theta),$$

where for all $j \in \{1, \ldots, T\}$, $M_j(x; \theta)$ is the $(1, j)$-cofactor of the matrix

$$\begin{pmatrix} 1 & \ldots & 1 \\ \exp(\lambda_1 x_1'\beta) & \ldots & \exp(\lambda_1 x_T'\beta) \\ \vdots & & \vdots \\ \exp(\lambda_\tau x_1'\beta) & \ldots & \exp(\lambda_\tau x_T'\beta) \end{pmatrix}.$$

As we also consider identification of β_0 alone, we also let, with a slight abuse of notation, $m(y, x; \beta) := m(y, x; \, (\beta', \lambda_0')')$. Our first result shows that the conditional moment of $m(Y, X; \theta_0)$ is zero.

Theorem 2.3 *If Assumptions 1-2 hold, we have, almost surely,*

$$\mathbb{E}[m(Y, X; \theta_0)|X] = 0. \tag{2.2}$$

Theorem 2.3 shows there exists a known moment condition which potentially identifies θ_0 in a model more general than the logistic one. It shows that, as the number of periods T increases, there is an increasing class of distributions F for which β_0 (or θ_0) can be point identified. This is consistent with the idea that if $T = \infty$, β_0 is point identified for any F, by using variations in X_t of a single individual. It also complements the results of Chernozhukov et al. (2013) showing that bounds on β_0 for general F shrink quickly as T increases.

Note that the result holds also with $T = \tau + 1 = 2$ (or, more generally, with $T > \tau = 1$). In such a case, the conditional moment condition can be written

$$\mathbb{E}\left[\mathbb{1}\{Y_1 > Y_2\}\exp(X_2'\beta_0) - \mathbb{1}\{Y_2 > Y_1\}\exp(X_1'\beta_0)|X\right] = 0.$$

This conditional moment generate the first-order conditions of the theoretical conditional likelihood, since the latter is equivalent to

$$\mathbb{E}\left[\frac{(X_1 - X_2)}{\exp(X_1'\beta_0) + \exp(X_2'\beta_0)}\left(\mathbb{1}\{Y_1 > Y_2\}\exp(X_2'\beta_0) - \mathbb{1}\{Y_2 > Y_1\}\exp(X_1'\beta_0)\right)\right] = 0.$$

2.2 Necessary and sufficient conditions for identification

The discussion above implies that with $T = \tau + 1 = 2$, β_0 is identified by (2.2) as soon as $\mathbb{E}\left[(X_1 - X_2)(X_1 - X_2)'\right]$ is nonsingular. We now consider sufficient conditions for (2.2) to identify θ_0 (or β_0) more generally, not only with $\tau = 1$. The moment conditions in the general case are highly nonlinear, making it difficult to provide a complete characterization. First, we consider the case where γ is actually constant.[2] For any $(k,t) \in \{1,\dots,K\} \times \{1,\dots,T\}$, we let $X^k := (X_{k,1},\dots,X_{k,T})$, $X^{-k} := (X_{k',t})_{k'\neq k, t=1,\dots,T}$ and $X^k_{-t} = (X_{k,s})_{s\neq t}$.

Proposition 2.4 *Let assume that Assumptions 1-2 are satisfied,* $T = \tau + 1 \geq 2$, $\mathbb{V}(\gamma) = 0$ *and for all* $(k,t) \in \{1,\dots,K\} \times \{1,\dots,T\}$, $|Supp(X^k_t|X^{-k}, X^k_{-t})| \geq 2\tau$. *Then,*

$$\mathbb{E}[m(Y,X;\theta)|X] = 0 \ a.s. \Rightarrow \theta = \theta_0. \tag{2.3}$$

Proposition 2.4 shows that in the absence of fixed effects, the conditional moment condition $\mathbb{E}[m(Y,X;\theta_0)|X] = 0$ is sufficient to identify θ_0 under mild restrictions on the distribution of X.[3] In particular, all components of X may be discrete. The result relies in particular on the fact that for any $\lambda \in \Lambda_\tau$, the family of functions $(v \mapsto \exp(\lambda_j v))_{j=1,\dots,\tau}$ forms a Chebyshev system (see, e.g., Krein and Nudelman, 1977,

[2]Because we consider identification based on (2.2) alone, we suppose this additional restriction to be unknown by the econometrician.

[3]A close inspection of the proof reveals that the support restrictions on X could actually be weakened further, but at the expense of complicating the condition.

Chapter II for the formal definition). This implies that any non-zero "polynomial" $v \mapsto \sum_{j=1}^{T} a_j \exp(\lambda_j v)$ does not vanish more than $T-1$ times.

We now investigate cases where γ is nondegenerate and possibly correlated with X, which is more realistic in practice. For any $(t, \ell, x) \in \{1, \dots, T\} \times \{1, \dots, \tau\} \times \text{Supp}(X)$, let us define

$$a_{t,\ell,x} : v \mapsto \mathbb{E}\left[\frac{\exp(\lambda_{0\ell}\gamma)}{C(\gamma, x; \theta_0, t)\left(1 + \sum_{j=1}^{\tau} w_j \delta_j(x; \theta_0, t) \exp(\lambda_{0j}(\beta_{0k} v + \gamma))\right)}\middle| X = x\right]$$

where $C(\gamma, x; \theta_0, t) := \prod_{t' \neq t}(1 + \sum_{j=1}^{\tau} w_j \exp(\lambda_{0j}(x_{t'}'\beta_0 + \gamma)))$ and $\delta_j(x; \theta_0, t) := \exp(\lambda_{0j} \times x_t'\beta_0)$. We consider the following conditions.[4]

Assumption 3 *1. There exist $(t_0, t_1) \in \{1, ..., T\}^2$ such that $\mathbb{E}[(X_{t_0} - X_{t_1})(X_{t_0} - X_{t_1})']$ is nonsingular.*

2. There exists $k \in \{1, \dots, K\}$ such that $\beta_{0k} \neq 0$ and almost surely, $X^k | X^{-k}$ admits a density with respect to the Lebesgue measure.

Assumption 4 *1. $X^k \perp\!\!\!\perp \gamma | X^{-k}$.*

2. There exists some $t_2 \in \{1, \dots, T\} \backslash \{t_0, t_1\}$ such that, for all $(\beta_k, \lambda) \in (\mathbb{R}^*) \times \Lambda_\tau$, $\{\lambda_1 \beta_k, \dots, \lambda_\tau \beta_k\} \cap \{\lambda_{01}\beta_{0k}, \dots, \lambda_{0\tau}\beta_{0k}\} = \emptyset$ implies that the $\tau(\tau+1)$ functions

$$\{a_{t_2,\ell,x}(v) \exp(\lambda_{0\ell}\beta_{0k} v), a_{t_2,\ell,x}(v) \exp(\lambda_1 \beta_k v), \dots, a_{t_2,\ell,x}(v) \exp(\lambda_\tau \beta_k v)\}_{\ell=1,\dots,\tau}$$

form a free family of functions over $\mathbb{R}$ for almost all $x \in$ Supp(X).

Assumption 4′ *1. $X^k \perp\!\!\!\perp \gamma | X^{-k}$.*

2. There exists some $t_2 \in \{1, \dots, T\} \backslash \{t_0, t_1\}$ such that, for all $\beta_k \in \mathbb{R}^*$, $\{\lambda_{01}\beta_k, \dots, \lambda_{0\tau}\beta_k\} \cap \{\lambda_{01}\beta_{0k}, \dots, \lambda_{0\tau}\beta_{0k}\} = \emptyset$ implies that the $\tau(\tau+1)$ functions

$$\{a_{t_2,\ell,x}(v) \exp(\lambda_{0\ell}\beta_{0k} v), a_{t_2,\ell,x}(v) \exp(\lambda_{01} \beta_k v), \dots, a_{t_2,\ell,x}(v) \exp(\lambda_{0\tau} \beta_k v)\}_{\ell=1,\dots,\tau}$$

form a free family of functions over $\mathbb{R}$ for almost all $x \in$ Supp(X).

[4] Again, Assumptions 3 and 4 (or 3 and 4') are assumed to be unknown by the econometrician.

Assumption 3.1 is necessary to ensure the unique representation of the index difference $(X_{t_0} - X_{t_1})'\beta_0$. Assumption 3.2 imposes that at least one regressor is continuously distributed. Assumption 4 and 4' are very close, with Assumption 4' being a weaker form of Assumption 4 that turns out to be sufficient to identify β_0 only, when λ_0 is supposed to be known. When combined with Assumption 3.2, Assumption 4.1 (or Assumption 4'.1) is similar, but less restrictive, than Assumption R.iii of Magnac and Maurin (2007) or Assumptions A.2-3 in Honore and Lewbel (2002). Importantly, it does not imply any large support restriction. Assumption 4.2 and 4'.2 are high-level conditions that we discuss below.

Proposition 2.5 *Suppose that Assumptions 1-3 hold and $T = \tau + 1$. Then:*

1. *If Assumption 4 holds as well, then* (2.3) *holds.*
2. *If Assumption 4' holds as well,*

$$\mathbb{E}[m(Y, X; \beta)|X] = 0 \ a.s. \Rightarrow \beta = \beta_0. \tag{2.4}$$

The proof relies on two main ingredients. The first is, again, the upper bound on the number of roots of exponential "polynomials". The second is analyticity of the conditional moment as a function of $X_{k,t}$. By a continuation theorem on real analytic functions (see e.g. Corollary 1.2.5 in Krantz and Parks, 2002), this allows us to extend the conditional moment function from any $x \in \text{Supp}(X)$ to any x' such that $x'_{j,t'} = x_{j,t'}$ for all $(j, t') \neq (k, t)$ and $x'_{k,t} \in \mathbb{R}$.

Assumption 4.2 and 4'.2 are high-level and technical. We conjecture that they hold under mild restrictions on the distribution of γ. The following proposition, restricted to $T = 3$ and a binary γ, substantiates this claim.

Proposition 2.6 *Let $T = \tau + 1 = 3$ and $\Lambda_\tau \subset \{(1, \lambda_2) : \lambda_2 > 4\}$. If $|Supp(\gamma|X)| = 2$ almost surely, Assumption 4'.2 is satisfied.*

We now turn to necessary conditions for (2.4) to hold. We consider the following assumption.

Assumption 5 $\mathbb{P}\left(X \in \left\{x \in \mathbb{R}^{KT} : |\{x_1, \ldots, x_T\}| = T\right\}\right) > 0.$

Assumption 5 imposes that there are trajectories of $X = (X_1, \dots, X_T)$ with distinct values at all periods. Since we focus here on $T \geq 3$, this excludes in particular the case where X_t is binary. But contrary to Assumption 3, Assumption 5 does not exclude the case where all covariates are discrete, and can be expected to hold if $|\text{Supp}(X_t)| \geq T$. The following proposition shows that Assumption 5 is actually necessary for the conditional moment condition $\mathbb{E}[m(Y, X; \beta_0)|X] = 0$ to identify β_0.

Proposition 2.7 *Suppose that Assumptions 1-2 are satisfied and $T = \tau + 1 \geq 3$. Then, if* (2.4) *holds, Assumption 5 holds as well.*

3 Estimation

In the following, we assume that λ_0 is known and focus on the estimation of β_0.[5] The conditional moment condition (2.4) can be transformed into unconditional conditions such that standard GMM estimators can easily be constructed. Letting $g(X) \in \mathbb{R}^K$, such estimators $\widehat{\beta}$ satisfy

$$\widehat{\beta} = \arg\min_{\beta \in B} \left(\frac{1}{n} \sum_{i=1}^{n} g(X_i) m(Y_i, X_i; \beta) \right)' \left(\frac{1}{n} \sum_{i=1}^{n} g(X_i) m(Y_i, X_i; \beta) \right), \tag{3.1}$$

where B is a compact subset of $\mathbb{R}^{K*}$. The optimal estimator among this class is obtained by choosing $g^\star(X) := R(X)/\Omega(X)$, with $R(X) = \mathbb{E}[\nabla_\beta m(Y, X; \beta_0)|X]$ and $\Omega(X) = \mathbb{V}[m(Y, X; \beta_0)|X]$ (see Chamberlain, 1987). Given that $R(X)$ and $\Omega(X)$ are unknown, an asymptotically efficient GMM estimator can be obtained in two steps. In a first step, $g(X)$ is chosen arbitrarily and we compute the corresponding estimator $\widehat{\beta}^1$. In a second step, we compute $\widehat{g}^\star(X) = \widehat{R}(X)/\widehat{\Omega}(X)$, where $\widehat{R}(x) = \widehat{\mathbb{E}}[\nabla_\beta m(Y, X; \widehat{\beta}^1)|X]$ and $\widehat{\Omega}(X) = \widehat{\mathbb{V}}[m(Y, X; \widehat{\beta}^1)|X]$ are standard nonparametric estimators (e.g., kernel or series estimators). We then compute the estimator $\widehat{\beta}^\star$ based on $\widehat{g}^\star(X)$. Under regularity conditions displayed in, e.g., Newey (1990), we have

$$\sqrt{n}(\widehat{\beta}^\star - \beta_0) \xrightarrow{d} \mathcal{N}(0, V_0), \tag{3.2}$$

[5] Estimation of θ_0 could be performed in the same way as that of β_0, but it is unclear to us whether the corresponding estimator would reach the semiparametric efficiency bound of θ_0, something we prove below for β_0.

where $V_0 := \mathbb{E}\left[\Omega(X)^{-1}R(X)R(X)'\right]^{-1}$. To obtain this result, two assumptions are worth mentioning. The first is an identifiability condition when using the optimal instruments:

$$\mathbb{E}[g^\star(X)m(Y,X;\beta)] = 0 \Rightarrow \beta = \beta_0.$$

Such a condition may fail to hold, as shown by Dominguez and Lobato (2004). Other estimators relying on the full set of moments can be used to prevent this identification failure (see in particular Dominguez and Lobato, 2004; Hsu and Kuan, 2011; Lavergne and Patilea, 2013). The second condition is that $\mathbb{E}\left[\Omega(X)^{-1}R(X)R(X)'\right]$ exists and is nonsingular. Nonsingularity holds if and only if $\mathbb{E}\left[R(X)R(X)'\right]$ is nonsingular, which is a local identification condition.

We now establish that with $T = \tau + 1 = 3$, the semiparametric efficiency bound actually coincides with the asymptotic variance V_0 of the optimal GMM estimator. The result holds under the following condition.

Assumption 6 *1.* $\mathbb{E}\left[\Omega^{-1}(X)R(X)R(X)'\right]$ *exists and is nonsingular.*

2. $|Supp(\gamma|X)| \geq 10$ *almost surely.*

We already discuss the first condition. The second condition we impose is weaker than that imposed by Chamberlain (2010), namely $\text{Supp}(\gamma|X) = \mathbb{R}$.

Theorem 3.1 *Assume* $T = \tau + 1 = 3$, λ_0 *is known with* $\lambda_{02} \neq 2$ *and Assumptions 1-3 and 6 hold. Then the semiparametric efficiency bound of* β_0, $V^\star(\beta_0)$, *is finite and satisfies* $V^\star(\beta_0) = V_0$.

Intuitively, this result states that all the information content of the model is included in the conditional moment restriction $\mathbb{E}[m(Y,X;\beta_0)|X] = 0$. It complements, for $T = \tau + 1 = 3$, the result of Hahn (1997), which states that the conditional maximum likelihood estimator is the efficient estimator of β_0 if F is logistic. The difference between the two results is that here, (w_1, w_2) is unknown rather than known and equal to $(1, 0)$.

4 Monte-Carlo simulations

We conduct numerical simulations in order to characterize the finite sample performance of $\widehat{\beta}^\star$. We let $T = \tau + 1 = 3$ and consider both $(w_1, w_2) = (0.1, 0.9)$ and $(w_1, w_2) = (0.2, 0.8)$. We fix $\lambda_0 = (1, 5)$ and suppose it is known. Next, we let $K = 1$ and $\beta_0 = 1$, with $X_t \in \{-1, 0, 1\}$ (note that a binary X_t). We first draw X_1 uniformly over $\{-1, 0, 1\}$, then draw X_2 uniformly over $\{-1, 0, 1\}\backslash\{X_1\}$ and finally let X_3 be the remaining element in $\{-1, 0, 1\}\backslash\{X_1, X_2\}$. Note that Assumption 3.2 fails to hold with such a X. But as explained above, this condition is only sufficient, not necessary for identification. We then consider five data generating processes (DGPs) where the r.v. γ is:

i. Constant: $\gamma = 0$.

ii. Discrete and independent of X: $\mathbb{P}(\gamma = -1/2|X) = \mathbb{P}(\gamma = 0|X) = \mathbb{P}(\gamma = 1/2|X) = 1/3$.

iii. Continuous and independent of X: $\gamma|X \sim \mathcal{U}([-1/4, 1/4])$.

iv. Discrete and correlated with X: $\gamma = UZ$ where $(U, Z) \in \{-1/2, 1/2\} \times \{0, 1\}$ and $\mathbb{P}(U = 1/2|X) = 0.5 + X_1X_2/3$, $\mathbb{P}(Z = 1|X, U) = 0.9$.

v. Continuous and correlated with X: $\gamma = UZ$ where $U|X \sim \mathcal{U}([0, 1/2])$ and $Z \in \{-1/2, 1/2\}$, $\mathbb{P}(Z = 1/2|X, U) = 0.5 + X_1X_2/3$.

Hereafter, we consider samples of size $n \in \{500; 1,000; 2,000; 4,000\}$. With the DGPs above, the subsample effectively used in the estimation, namely $\{i \in \{1, \dots, n\} : \sum_{t=1}^{3} Y_{it} = 1\}$, represents on average 47.8% of the initial sample.

To compute the optimal GMM estimator, the usual practice is to estimate $\widehat{g}^\star$ using an inefficient GMM estimator. However, in the current set-up, such estimators are often equal to zero if g is not chosen appropriately. To overcome this finite sample issue, we first use a rough estimator $\widetilde{g}^\star$ of $g^\star$ based on the conditional maximum likelihood estimator $(\widehat{\beta}^1)$ of β_0, assuming a logistic distribution. Then, using $\widetilde{g}^\star$, we obtain an initial GMM estimator $\widetilde{\beta}$, which allows us to compute a second (and consistent) estimator of $\widehat{g}^\star$. Finally, we compute the asymptotically optimal GMM estimator $\widehat{\beta}^\star$ using $\widehat{g}^\star$.

Table 1: Simulation Results for $\widehat{\beta}^{\star}$

		$w_1 = 0.1$		$w_1 = 0.2$	
DGP	n	Bias	RMSE	Bias	RMSE
i.	500	−0.1949	0.2724	−0.2245	0.3119
	1, 000	−0.1692	0.2278	−0.1964	0.2664
	2, 000	−0.0934	0.1722	−0.1137	0.1976
	4, 000	−0.0002	0.0622	−0.0159	0.0831
ii.	500	−0.2784	0.3742	−0.3608	0.4544
	1, 000	−0.2486	0.3311	−0.3118	0.3999
	2, 000	−0.1508	0.2549	−0.1823	0.3018
	4, 000	−0.0403	0.1220	−0.0872	0.1909
iii.	500	−0.2060	0.2892	−0.2467	0.3405
	1, 000	−0.1795	0.2460	−0.2153	0.2910
	2, 000	−0.0991	0.1775	−0.1269	0.2179
	4, 000	−0.0066	0.0777	−0.0229	0.0830
iv.	500	−0.2953	0.3870	−0.3839	0.4719
	1, 000	−0.2596	0.3607	−0.3178	0.4247
	2, 000	−0.1296	0.2409	−0.1546	0.2876
	4, 000	−0.0704	0.1544	−0.1500	0.2685
v.	500	−0.2084	0.2875	−0.2510	0.3418
	1, 000	−0.1785	0.2469	−0.2161	0.2938
	2, 000	−0.0913	0.1734	−0.1176	0.2103
	4, 000	−0.0066	0.0696	−0.0288	0.0931

Notes: $\beta_0 = 1$, $\lambda_0 = (1, 5)$, $w_2 = 1 - w_1$. The optimal instruments are estimated using conditional means and $\widehat{\beta}^1$. The results are based on 10, 000 sample replications.

For each DGP and the two values of w_1, Table 1 reports the estimated bias and root mean square error (RMSE) of $\widehat{\beta}^{\star}$. The estimator $\widehat{\beta}^{\star}$ is precise in the absence of fixed effects. When fixed effects are introduced the bias and RMSE vary with (w, λ_0). Overall, the results suggest that for a given sample size n, the bias and RMSE are lower when $w_2 - w_1$ increases, when $|\text{Supp}(\gamma|X)|$ increases or when γ is uncorrelated with X. The second case is consistent with our conjecture about Assumption 4.

5 Conclusion

This paper addresses the problem of point identification of the common slope parameter in a static panel binary model with exogenous and bounded regressors. We derive necessary and sufficient conditions for global point identification based on a conditional moment restriction when $T \geq 3$ and the unobserved terms belong to a family of generalized logistic distribution that we introduce. Our results generalize those from Chamberlain (2010) and can be used to build a GMM estimator that reaches the semiparametric efficiency bound when $T = 3$. Our paper leaves a few questions unanswered. A first one is whether the family of F considered here is the only one for which point identification can be achieved. Another one is whether the GMM estimator still reaches the semiparametric efficiency bound when $T > 3$. Both questions raise difficult issues and deserve future investigation.

References

Balakrishnan, N. and Leung, M. (1988), 'Order statistics from the type i generalized logistic distribution', *Communications in Statistics-Simulation and Computation* **17**(1), 25–50.

Bierens, H. J. (1990), 'A consistent conditional moment test of functional form', *Econometrica: Journal of the Econometric Society* pp. 1443–1458.

Chamberlain, G. (1987), 'Asymptotic efficiency in estimation with conditional moment restrictions', *Journal of Econometrics* **34**(3), 305–304.

Chamberlain, G. (2010), 'Binary response models for panel data: Identification and information', *Econometrica* **78**(1), 159–168.

Chernozhukov, V., Fernández-Val, I., Hahn, J. and Newey, W. (2013), 'Average and quantile effects in nonseparable panel models', *Econometrica* **81**(2), 535–580.

Davezies, L., D'Haultfœuille, X. and Mugnier, M. (2020), 'Online appendix for fixed effects binary choice models with three or more periods', `https://faculty.crest.fr/xdhaultfoeuille/wp-content/uploads/sites/9/2020/09/online-appendix.pdf`.

Dominguez, M. A. and Lobato, I. N. (2004), 'Consistent estimation of models defined by conditional moment restrictions', *Econometrica* **72**(5), 1601–1615.

Hahn, J. (1997), 'A note on the efficient semiparametric estimation of some exponential panel models', *Econometric Theory* **13**(4), 583–588.

Hong, Y. and White, H. (1995), 'Consistent specification testing via nonparametric series regression', *Econometrica: Journal of the Econometric Society* pp. 1133–1159.

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

Hsu, S.-H. and Kuan, C.-M. (2011), 'Estimation of conditional moment restrictions without assuming parameter identifiability in the implied unconditional moments', *Journal of Econometrics* **165**(1), 87 – 99.

Johnson, E. G. (2004), ‘Identification in discrete choice models with fixed effects’, *Working paper, Bureau of Labor Statistics* .

Krantz, S. and Parks, H. (2002), *A Primer of Real Analytic Functions*, Advanced Texts Series, Birkhäuser Boston.

Krein, M. and Nudelman, A. A. (1977), *The Markov Moment Problem and Extremal Problems*, American Mathematical Society.

Lavergne, P. and Patilea, V. (2013), ‘Smooth minimum distance estimation and testing with conditional estimating equations: uniform in bandwidth theory’, *Journal of Econometrics* **177**(1), 47–59.

Magnac, T. (2004), ‘Binary variables and sufficiency: Generalizing conditional logit’, *Econometrica* **72**(6), 1859–1876.

Magnac, T. and Maurin, E. (2007), ‘Identification and information in monotone binary models’, *Journal of Econometrics* **139**(1), 76–104.

Manski, C. F. (1987), ‘Semiparametric analysis of random effects linear models from binary panel data’, *Econometrica* **5**(2), 357–362.

Newey, W. K. (1990), ‘Efficient instrumental variables estimation of nonlinear models’, *Econometrica* **58**(4), 809–837.

Rasch, G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Copenhagen: Denmarks Paedagogiske Institute.

Stukel, T. A. (1988), ‘Generalized logistic models’, *Journal of the American Statistical Association* **83**(402), 426–431.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

A Proofs of the results

A.1 Proposition 2.1

The sufficient part is obvious. To prove necessity, suppose $\beta_0 \neq 0$. Since $\mathbb{E}[(X_t - X_{t'})(X_t - X_{t'})']$ is non singular, there exist a subset $\mathcal{S}$ of the support of $(X_t, X_{t'})$ such that $\mathbb{P}(\mathcal{S}) > 0$ and for all $(x_t, x_{t'}) \in \mathcal{S}$, $(x_t - x_{t'})'\beta_0$ has constant, non-zero sign. Without loss of generality let us assume $(x_t - x_{t'})'\beta_0 > 0$. Let $G(x) = F(x)/(1-F(x))$. Because G is strictly increasing, we have, for all $g \in \mathbb{R}$,

$$G(x_t'\beta_0 + g) > G(x_{t'}'\beta_0 + g).$$

Equivalently,

$$F(x_t'\beta_0 + g)(1 - F(x_{t'}'\beta_0 + g)) > F(x_{t'}'\beta_0 + g)(1 - F(x_t'\beta_0 + g)).$$

In other words,

$$\mathbb{P}(Y_1 = 1, Y_{t'} = 0 | X_t = x_t, X_{t'} = x_t', \gamma = g) > \mathbb{P}(Y_1 = 0, Y_{t'} = 1 | X_t = x_t, X_{t'} = x_t', \gamma = g),$$

and the result follows by integration over g.

A.2 Theorem 2.3

Let us define

$$A(x, \gamma; \theta) := \begin{pmatrix} \sum_{j=1}^{\tau} w_j \exp(\lambda_j(x_1'\beta + \gamma)) & \dots & \sum_{j=1}^{\tau} w_j \exp(\lambda_j(x_T'\beta + \gamma)) \\ \exp(\lambda_1 x_1'\beta) & \dots & \exp(\lambda_1 x_T'\beta) \\ \vdots & & \vdots \\ \exp(\lambda_\tau x_1'\beta) & \dots & \exp(\lambda_\tau x_T'\beta) \end{pmatrix}.$$

Let $A_i(x, \gamma; \theta)$ denote the ith line of $A(x, \gamma; \theta)$. Then

$$A_1(x, \gamma; \theta) = \sum_{j=1}^{\tau} w_j \exp(\lambda_j \gamma) A_{j+1}(x, \gamma; \theta).$$

It follows that for all $(x, \gamma) \in \text{Supp}(X) \times \mathbb{R}$,

$$\det A(x, \gamma; \theta_0) = 0.$$

By Assumption 2 and since we focus on the first type therein, we have $G(x) := F(x)/(1-F(x)) = \sum_{j=1}^{T-1} w_j \exp(\lambda_{0j} x)$. Now, developping $\det A(x, \gamma; \theta_0)$ with respect to the first row yields, by definition of the function m,

$$\sum_{y \in \{0,1\}^T} m(y, x; \theta_0) \prod_{t: y_t = 1} G(x_t' \beta_0 + \gamma) = 0.$$

Multiplying this equality by $\prod_t (1 - F(x_t'\beta_0 + \gamma))$ we obtain

$$\sum_{y \in \{0,1\}^T} \left[m\left(y, x; \theta_0\right) \prod_{t: y_t = 1} F(x_t'\beta_0 + \gamma) \prod_{t: y_t = 0} (1 - F(x_t'\beta_0 + \gamma)) \right] = 0.$$

This equation is equivalent to $\mathbb{E}\left[m(Y, X; \theta_0) | X, \gamma\right] = 0$ a.s. The result follows by integration over γ.

A.3 Proposition 2.4

Let us suppose that $\theta = (\beta, \lambda) \in \Theta_0$ satisfies

$$\mathbb{E}\left[m(Y, X; \theta) | X\right] = 0, \tag{A.1}$$

and let us show that $\theta = \theta_0$. Since $\gamma = \gamma_0$ almost surely for some γ_0, Equation (A.1) is equivalent to:

$$\sum_{i=1}^{\tau} w_i \exp(\lambda_{0i} \gamma_0) \det\left(A^i(x)\right) = 0, \tag{A.2}$$

for almost all $x \in \text{Supp}(X)$, with

$$A^i(x) := \begin{pmatrix} \exp(\lambda_{0i} x_1'\beta_0) & \dots & \exp(\lambda_{0i} x_T'\beta_0) \\ \exp(\lambda_1 x_1'\beta) & \dots & \exp(\lambda_1 x_T'\beta) \\ \vdots & & \vdots \\ \exp(\lambda_\tau x_1'\beta) & \dots & \exp(\lambda_\tau x_T'\beta) \end{pmatrix}.$$

Let $\mathcal{S}$ denote the subset of $\text{Supp}(X)$ on which (A.2) holds. Further, let $\mathcal{X}(\beta) = \{x \in \mathcal{S} : |\{x_2'\beta, \dots, x_T'\beta\}| = T-1\}$. We first we show that

$$\mathbb{P}(\mathcal{X}(\beta)) > 0. \tag{A.3}$$

This is trivial for $T=2$. Otherwise, note first that there exists k_0 such that $\beta_{0k_0}\neq 0$. Then:

$$\mathcal{X}(\beta)=\left\{x\in\mathcal{S}:\forall t\geq 3, x_{k_0,t}\notin\left\{\frac{x_2'\beta_0-x_{-k_0,3}'\beta_{0-k_0}}{\beta_{0k_0}},\dots,\frac{x_{t-1}'\beta_0-x_{-k_0,t}'\beta_{0-k_0}}{\beta_{0k_0}}\right\}\right\}.$$

The condition $|\text{Supp}(X_{k_0,t}|X_2,\dots,X_{t-1},X_t^{-k_0})|\geq 2\tau$ for all $t\geq 3$ (with $X_t^{-k_0}=(X_{j,t})_{j\neq k_0}$) ensures that almost surely,

$$\text{Supp}(X_{k_0,t}|X_2,\dots,X_{t-1},X_t^{-k_0})\not\subset\left\{\frac{X_2'\beta_0-X_{-k_0,t}'\beta_{0-k_0}}{\beta_{0k_0}},\dots,\frac{X_{t-1}'\beta_0-X_{-k_0,t}'\beta_{0-k_0}}{\beta_{0k_0}}\right\}$$

and thus (A.3) holds.

Now fix $x\in\mathcal{X}(\beta)$ and $k\in\{1,\dots,K\}$. Using again $|\text{Supp}(X_{k,t}|X^{-k},X_{-1}^k)|\geq 2\tau$, there exists $A\subset\mathbb{R}$, $|A|\geq 2\tau$, such that for all $\widetilde{x}$ verifying $\widetilde{x}_{j,t}=x_{j,t}$ for $j\neq k$ or $t\neq 1$ and $\widetilde{x}_{k,1}=x_{k,1}+v$, $v\in A$, we have $\widetilde{x}\in\mathcal{S}$. Applying (A.2) to such $\widetilde{x}$ and developing each determinant with respect to the first column, we obtain that for all $v\in A$,

$$\sum_{j=1}^{\tau}(-1)^j\exp(\lambda_j x_1'\beta)\left(\sum_{i=1}^{\tau}w_i\exp(\lambda_{0i}\gamma_0)\det\left(A^i_{\{j+1,1\}}(x)\right)\right)\exp(\lambda_j\beta_k v)$$
$$+\sum_{i=1}^{\tau}w_i\exp(\lambda_{0i}(x_1'\beta_0+\gamma_0))\det\left(A^i_{\{1,1\}}(x)\right)\exp(\lambda_{0i}\beta_{0k}v)=0,\qquad\text{(A.4)}$$

where $A^i_{j,k}(x)$ denote the sub-matrix of $A^i(x)$ once row j and column k have been removed.

We first assume that $\beta_{0k}\neq 0$. Suppose that there exists i such that for all $j\in\{1,\dots,\tau\}$, $\lambda_j\beta_k\neq\lambda_{0i}\beta_{0k}$. The left hand-side of (A.4) is a polynomial of exponential functions with at most 2τ distinct exponential functions and it is equal to 0 on 2τ distinct points v. Then, by Lemma B.1 and because the coefficient of $\exp(\lambda_{0i}\beta_{0k}v)$ is $w_i\exp(\lambda_{0i}(x_1'\beta_0+\gamma_0))\det\left(A^i_{\{1,1\}}(x)\right)$, we have

$$\det\left(A^i_{\{1,1\}}(x)\right)=0.\qquad\text{(A.5)}$$

Now, because $|\{x_2'\beta,\dots,x_T'\beta\}|=T-1$, the definition of Chebyshev systems implies that $\det\left(A^1_{\{1,1\}}\right)\neq 0$, a contradiction. Hence, for all $i\in\{1,\dots,\tau\}$, there exists $\ell(i)$ such that $\lambda_{\ell(i)}\beta_k=\lambda_{0i}\beta_{0k}$. Because $\lambda_{\ell(i)}$ and λ_{0i} are both positive, the sign of β_k is then equal to the sign of β_{0k}. Let us suppose without loss of generality (since, $\beta_{0k}\neq 0$) that $\beta_{0k}>0$. Then $\lambda_{01}\beta_{0k}<\ldots<\lambda_{0\tau}\beta_{0k}$, implying, since $\beta_k>0$, that

$\lambda_{\ell(1)} < \ldots < \lambda_{\ell(\tau)}$. Hence, $\ell(i) = i$ for all $i \in \{1, \ldots, \tau\}$ and $i = 1$ yields $\beta_k = \beta_{0k}$. In turn, this latter equality implies that $\lambda = \lambda_0$.

We now consider the case $\beta_{0k} = 0$. Let us assume that $\beta_k \neq 0$. The left hand-side of (A.4) is a polynomial of exponential functions with at most $\tau + 1$ distinct exponential functions (since $\lambda_j \beta_k \neq 0$ for all j) and it is equal to 0 on 2τ distinct points v. Then, by Lemma B.1,

$$\sum_{i=1}^{\tau} w_i \exp(\lambda_{0i}(x_1'\beta_0 + \gamma_0)) \det\left(A^i_{\{1,1\}}(x)\right) = 0.$$

Now, notice that $\det\left(A^i_{\{1,1\}}(x)\right) \neq 0$ does not depend on i. As a result,

$$\sum_{i=1}^{\tau} w_i \exp(\lambda_{0i}(x_1'\beta_0 + \gamma_0)) = 0,$$

which is a contradiction. Hence $\beta_k = 0 = \beta_{0k}$. Note that we do not identify λ_0 in this case, but its identification is achieved by the previous paragraph, since there exists k_0 such that $\beta_{0k_0} \neq 0$. This concludes the proof.

A.4 Proposition 2.5

1. Without loss of generality, we assume hereafter that $t_2 = 1$ so that $t_0, t_1 \geq 2$. Let us suppose that $\theta = (\beta, \lambda) \in \Theta_0$ satisfies

$$\mathbb{E}\left[m(Y, X; \theta) | X\right] = 0, \tag{A.6}$$

and let us show that $\theta = \theta_0$. Equation (A.6) is equivalent to

$$\sum_{i=1}^{\tau} w_i \mathbb{E}\left[\frac{\exp(\lambda_{0i}\gamma)}{C(\gamma, x; \theta_0, 1)\left(1 + \sum_{j=1}^{\tau} w_j \exp(\lambda_{0j}(x_1'\beta_0 + \gamma))\right)}\middle| X = x\right]$$
$$\times \det\begin{pmatrix} \exp(\lambda_{0i}x_1'\beta_0) & \ldots & \exp(\lambda_{0i}x_T'\beta_0) \\ \exp(\lambda_1 x_1'\beta) & \ldots & \exp(\lambda_1 x_T'\beta) \\ \vdots & & \vdots \\ \exp(\lambda_\tau x_1'\beta) & \ldots & \exp(\lambda_\tau x_T'\beta) \end{pmatrix} = 0, \tag{A.7}$$

for almost all $x \in \text{Supp}(X)$. Let $\mathcal{S}$ denote the subset of $\text{Supp}(X)$ on which (A.7) holds. Further, let $\mathcal{X}(\beta) = \{x \in \mathcal{S} : |\{x_1'\beta, \ldots, x_T'\beta\}| = T\}$. By Assumption 3,

$\mathbb{P}(\mathcal{X}(\beta)) = 1$. Now, fix $x \in \mathcal{X}(\beta)$. By Assumption 3 again, there exist $\underline{\varepsilon} \leq 0 \leq \overline{\varepsilon}$ with $\max(-\underline{\varepsilon}, \overline{\varepsilon}) > 0$, such that for almost every $\widetilde{x}$ verifying $\widetilde{x}_t = x_t$ for $t > 1$, $\widetilde{x}_{j1} = x_{j1}$ for $j \neq k$, $|\widetilde{x}_{k,1} - x_{k,1}| \in [\underline{\varepsilon}, \overline{\varepsilon}]$, we have $\widetilde{x} \in \text{Supp}(X)$. Applying (A.7) to such $\widetilde{x}$ and using $X^k \perp\!\!\!\perp \gamma | X^{-k}$, we obtain

$$\sum_{i=1}^{\tau} w_i a_{1,i,x}(v) \det\left(A^i(v)\right) = 0, \tag{A.8}$$

for almost every $v \in [\underline{\varepsilon}, \overline{\varepsilon}]$, with

$$A^i(v) = \begin{pmatrix} \exp(\lambda_{0i}(x_1'\beta_0 + \beta_{0k}v)) & \dots & \exp(\lambda_{0i}x_T'\beta_0) \\ \exp(\lambda_1(x_1'\beta + \beta_k v)) & \dots & \exp(\lambda_1 x_T'\beta) \\ \vdots & & \vdots \\ \exp(\lambda_\tau(x_1'\beta + \beta_k v)) & \dots & \exp(\lambda_\tau x_T'\beta) \end{pmatrix}.$$

Let $A^i_{J,K}(v)$ denote the sub-matrix of $A^i(v)$ once the rows and columns with indices in $J \subset \{1, \dots, T\}$ and $K \subset \{1, \dots, T\}$, respectively, have been removed. We simply note $A^i_{J,K}$ when $A^i_{J,K}(v)$ does not depend on v. Then, developping each $A^i(v)$ with respect to the first column, we obtain, for almost every $v \in [\underline{\varepsilon}, \overline{\varepsilon}]$,

$$\sum_{i=1}^{\tau} w_i \Bigg[\det\left(A^i_{\{1\},\{1\}}\right) \exp(\lambda_{0i}(x_1'\beta_0 + \beta_{0k}v)) a_{1,i,x}(v) \\ + \sum_{j=1}^{\tau} (-1)^j \det\left(A^i_{\{j+1\},\{1\}}\right) \exp(\lambda_j(x_1'\beta + \beta_k v)) a_{1,i,x}(v) \Bigg] = 0. \tag{A.9}$$

Now, by Lemma B.2, the left-hand side of (A.9) is real analytic, where we recall that a function $f : I \to \mathbb{R}$ is real analytic if f is equal to its Taylor series at every point of I. Then, by the continuation theorem for real analytic functions (see e.g. Corollary 1.2.5 in Krantz and Parks, 2002), (A.8) holds for all $v \in \mathbb{R}$. Now, fix $i \in \{1, \dots, \tau\}$ and let us assume that there is no $t(i) \in \{1, \dots, \tau\}$ such that $\lambda_{t(i)}\beta_k = \lambda_{0i}\beta_{0k}$. Then, Assumption 4.2 ensures that the functions of v in (A.8) are linearly independent, so that

$$\det\left(A^i_{\{t\},\{1\}}\right) = 0, \quad \forall t \in \{1, \dots, T\}, \tag{A.10}$$

Because $|\{(x_1'\beta, \dots, x_T'\beta\}| = T$, we have, by definition of Chebyshev systems,

$$\det\left(A^i_{\{1\},\{1\}}\right) \neq 0,$$

contradicting equation (A.10). Hence, for all $i \in \{1, \dots, \tau\}$, there exists $t(i)$ such that $\lambda_{t(i)}\beta_k = \lambda_{0i}\beta_{0k}$. Because $\lambda_{t(i)}$ and λ_{0i} are both positive, the sign of β_k is then equal to the sign of β_{0k}. Let us suppose without loss of generality (since, by Assumption 3, $\beta_{0k} \neq 0$) that $\beta_{0k} > 0$. Then $\lambda_{01}\beta_{0k} < \dots < \lambda_{0\tau}\beta_{0k}$, implying, since $\beta_k > 0$, that $\lambda_{t(1)} < \dots < \lambda_{t(\tau)}$. Hence, $t(i) = i$ for all $i \in \{1, \dots, \tau\}$ and $i = 1$ yields $\beta_k = \beta_{0k}$. In turn, this latter equality implies that $\lambda = \lambda_0$.

Now, in (A.8), $\lambda_{0i} = \lambda_i$ for all $i \in \{1, \dots, \tau\}$ and $\beta_{0k} = \beta_k$. With λ replaced by λ_0 and β_k replaced by β_{0k}, (A.9) and Assumption 4.2 still imply that for all $i \in \{1, \dots, \tau\}$,

$$\det\left(A^i_{\{t\},\{1\}}\right) = 0, \quad \forall t \in \{2, \dots, T\}\backslash\{i+1\}. \tag{A.11}$$

Because $|\{x_1'\beta, \dots, x_T'\beta\}| = T$, we have, by definition of Chebyshev systems,

$$\det\left(A^i_{\{1,t\},\{1,n\}}\right) \neq 0, \quad \forall (t,n) \in \{2, \dots, T\}\backslash\{i+1\} \times \{2, \dots, T\}.$$

This, together with (A.11), implies that for $t \in \{2, \dots, T\}\backslash\{i+1\}$ the first row of $A^i_{t,1}$ is a non-trivial linear combination of the other rows. In other words, for all $t \neq i$, there exists a non-zero vector $(w_{t,j})_{j=1,\dots,\tau}$ with $w_{t,t} = 0$ such that for all $s \geq 2$,

$$\exp(\lambda_{0i}x_s'\beta_0) = \sum_{j=1}^{\tau} w_{t,j}\exp(\lambda_{0j}x_s'\beta). \tag{A.12}$$

Let define $P_t(u) = \sum_{j=1}^{\tau} w_{t,j}\exp(\lambda_{0j}u)$ for all $t \in \{1, \dots, \tau\}$. Then, for all $s \geq 2$,

$$P_1(x_s'\beta) = \dots = P_{i-1}(x_s'\beta) = P_{i+1}(x_s'\beta) = \dots = P_\tau(x_s'\beta).$$

Moreover, because $x \in \mathcal{X}(\beta)$, we have $|\{x_2'\beta, \dots, x_T'\beta\}| = \tau$. Then, by Lemma B.1, $P_2 = \dots = P_T$. But this implies that for all $(t,j) \in \{1, \dots, \tau\}\backslash\{i\}$, $w_{t,j} = w_{j,j} = 0$. Therefore, by (A.12) again, there exists strictly positive constants $(c_1, \dots, c_\tau) \in (0, \infty)^\tau$ such that $\exp(\lambda_{0i}x_t'\beta_0) = c_i\exp(\lambda_{0i}x_t'\beta)$ for all $t \geq 2$. In other words, there exists $K \in \mathbb{R}$ such that for all $t \geq 2$,

$$x_t'(\beta_0 - \beta) = K. \tag{A.13}$$

This equality holds in particular for periods t_0 and t_1 in Assumption 3.1. Moreover, because $x \in \mathcal{X}(\beta)$ was arbitrary and $\mathbb{P}(\mathcal{X}(\beta)) = 1$, this implies that almost surely, $(X_{t_0} - X_{t_1})'(\beta_0 - \beta) = 0$. The first part of Assumption 3 implies $\beta = \beta_0$, which ends the proof.

2. We follow the exact same reasoning, except that λ in θ is replaced by λ_0. In particular, we obtain the same equation as (A.9) with λ_0 in place of λ. Then (A.10) holds under Assumption 3'.2 instead of Assumption 3.2. This implies that $\beta_k = \beta_{0k}$. The proof that $\beta_j = \beta_{0j}$ for $j \neq k$ is exactly as above.

A.5 Proposition 2.6

We leave x and the conditioning on $X = x$ implicit here. We also let $C(\gamma) := C(\gamma, x; \theta_0, t_2)$, $\alpha_i := w_i \delta_i(x; \theta_0, t_2)$, $a_i := \lambda_{0i}\beta_k$, $b_i := \lambda_{0i}\beta_{0k}$, $(\gamma_1, \gamma_2) := \text{Supp}(\gamma|X = x)$ and (q_1, q_2) denote the corresponding probabilities. We must prove that for all $\boldsymbol{\mu} = (\mu_{j\ell})_{j=0,1,2,\ell=1,2}$, if for all $v \in \mathbb{R}$,

$$\sum_{j=1}^{2} e^{a_j v} \sum_{p=1}^{2} \frac{q_p}{C(\gamma_p)} \frac{1}{1 + \sum_{i=1}^{2} \alpha_i e^{\lambda_{0i}\gamma_p} e^{b_i v}} \left(\sum_{\ell=1}^{2} \mu_{j\ell} e^{\lambda_{0\ell}\gamma_p} \right)$$
$$+ \sum_{\ell=1}^{2} e^{\lambda_{0\ell}\beta_{0k} v} \sum_{p=1}^{2} \frac{q_p \mu_{0\ell} e^{\lambda_{0\ell}\gamma_\ell}}{C(\gamma_p)} \frac{1}{1 + \sum_{i=1}^{2} \alpha_i e^{\lambda_{0i}\gamma_p} e^{b_i v}} = 0,$$

then $\boldsymbol{\mu} = 0$. Let us define, for $p \in \{1, 2\}$,

$$f_{j,p}(v) = \left| \begin{array}{ll} \frac{e^{a_j v}}{1+\sum_{i=1}^{\tau} \alpha_i e^{\lambda_{0i}\gamma_p} e^{b_i v}} & \text{if } j \in \{1, 2\}, \\ \frac{e^{b_{j-\tau} v}}{1+\sum_{i=1}^{\tau} \alpha_i e^{\lambda_{0i}\gamma_p} e^{b_i v}} & \text{if } j \in \{3, 4\}, \end{array} \right.$$

$$G_{j,p}(\boldsymbol{\mu}) = \left| \begin{array}{ll} \frac{q_p}{C(\gamma_p)} \sum_{\ell=1}^{\tau} \mu_{j\ell} e^{\lambda_{0\ell}\gamma_p} & \text{if } j \in \{1, 2\}, \\ \frac{q_p \mu_{0j-\tau} e^{\lambda_{0j-\tau}\gamma_{j-\tau}}}{C(\gamma_p)} & \text{if } j \in \{3, 4\}. \end{array} \right.$$

Then Assumption 4'.2 can be rewritten as follows:

$$\sum_{j=1}^{4} \sum_{p=1}^{2} G_{j,p}(\boldsymbol{\mu}) f_{j,p}(v) = 0 \quad \forall v \in \mathbb{R} \Rightarrow \boldsymbol{\mu} = 0. \tag{A.14}$$

To prove (A.14), first remark that if $G_{j,p}(\boldsymbol{\mu}) = 0$ for all (j, p), then $\boldsymbol{\mu} = 0$. This is trivial for the $\mu_{0\ell}$. For the $\mu_{j\ell}$, $j \geq 1$, this follows from Lemma B.1. Thus, Assumption 4 holds if the family $(f_{j,p})_{j=1,\dots,4,p=1,2}$ is free, i.e. if for all $\boldsymbol{\nu} = (\nu_{ij})_{i=1,\dots,4,j=1,2}$

$$\sum_{j=1}^{4} \sum_{p=1}^{2} \nu_{jp} f_{j,p}(v) = 0 \quad \forall v \in \mathbb{R} \Rightarrow \boldsymbol{\nu} = 0.$$

Equivalently, we need to show that if for all $v \in \mathbb{R}$,

$$
\begin{aligned}
(\nu_{11}+\nu_{12})e^{a_1 v} + \alpha_1(\nu_{11}e^{\lambda_{01}\gamma_2} + \nu_{12}e^{\lambda_{01}\gamma_1})e^{(a_1+b_1)v} + \alpha_2(\nu_{11}e^{\lambda_{02}\gamma_2} + \nu_{12}e^{\lambda_{02}\gamma_1})e^{(a_1+b_2)v} \\
+(\nu_{21}+\nu_{22})e^{a_2 v} + \alpha_1(\nu_{21}e^{\lambda_{01}\gamma_2} + \nu_{22}e^{\lambda_{01}\gamma_1})e^{(a_2+b_1)v} + \alpha_2(\nu_{21}e^{\lambda_{02}\gamma_2} + \nu_{22}e^{\lambda_{02}\gamma_1})e^{(a_2+b_2)v} \\
+(\nu_{31}+\nu_{32})e^{b_1 v} + \alpha_1(\nu_{31}e^{\lambda_{01}\gamma_2} + \nu_{32}e^{\lambda_{01}\gamma_1})e^{2b_1 v} + \alpha_2(\nu_{31}e^{\lambda_{02}\gamma_2} + \nu_{32}e^{\lambda_{02}\gamma_1})e^{(b_1+b_2)v} \\
+(\nu_{41}+\nu_{42})e^{b_2} + \alpha_1(\nu_{41}e^{\lambda_{01}\gamma_2} + \nu_{42}e^{\lambda_{01}\gamma_1})e^{(b_1+b_2)v} + \alpha_2(\nu_{41}e^{\lambda_{02}\gamma_2} \\
+\nu_{42}e^{\lambda_{02}\gamma_1})e^{2b_2 v} = 0,
\end{aligned}
$$

then $\boldsymbol{\nu} := (\nu_{11}, \nu_{12}, \nu_{21}, \nu_{22}, \nu_{31}, \nu_{32}, \nu_{41}, \nu_{42})' = 0$. The proof of this point, which is long and cumbersome, is detailed in our online Appendix (Davezies et al., 2020).

A.6 Proposition 2.7

Let us suppose that Assumption 5 fails. Without loss of generality, assume that $X_1 = X_2$ almost surely. Let us define $y_1 := (1, 0, ..., 0)$, $y_2 := (0, 1, 0, ..., 0)$ and $f(x; \beta) := \mathbb{E}\left[m(Y, X; \beta, \lambda_0)|X = x\right]$. By definition,

$$
f(X;\beta) = \sum_{y\in\{0,1\}^T} \mathbb{P}(Y = y|X)m(y, X; \beta). \tag{A.15}
$$

Moreover, almost surely,

$$
\begin{aligned}
&\mathbb{P}(Y = y_1|X) \\
=&\int F(X_1'\beta_0+\gamma)(1-F(X_2'\beta_0+\gamma))(1-F(X_3'\beta_0+\gamma))\cdots(1-F(X_T'\beta_0+\gamma))\mathrm{d}F_{\gamma|X}(\gamma) \\
=&\int F(X_2'\beta_0+\gamma)(1-F(X_1'\beta_0+\gamma))(1-F(X_3'\beta_0+\gamma))\cdots(1-F(X_T'\beta_0+\gamma))\mathrm{d}F_{\gamma|X}(\gamma) \\
=&\mathbb{P}(Y = y_2|X).
\end{aligned} \tag{A.16}
$$

Next,

$$
\begin{aligned}
m(y_1, X; \beta) &= \det\begin{pmatrix} \exp(\lambda_{01}X_2'\beta) & \dots & \exp(\lambda_{01}X_T'\beta) \\ \vdots & & \vdots \\ \exp(\lambda_{0T-1}X_2'\beta) & \dots & \exp(\lambda_{0T-1}X_T'\beta)\end{pmatrix} \\
&= \det\begin{pmatrix} \exp(\lambda_{01}X_1'\beta) & \dots & \exp(\lambda_{01}X_T'\beta) \\ \vdots & & \vdots \\ \exp(\lambda_{0T-1}X_1'\beta) & \dots & \exp(\lambda_{0T-1}X_T'\beta)\end{pmatrix} \\
&= -m(y_2, X; \beta).
\end{aligned} \tag{A.17}
$$

Moreover, for all y such that $\sum_t y_t = 1$ and $y \notin \{y_1, y_2\}$, $m(y, X; \beta_0) = 0$ because the cofactor includes two identical columns (since $X_1 = X_2$). Finally, if $\sum_t y_t \neq 1$, we also have $m(y, X; \beta_0) = 0$. In view of (A.15), these last points, combined with (A.16)-(A.17), imply $f(\beta) = 0$. Since β was arbitrary, it means that (2.3) does not identify β_0. The result follows.

A.7 Theorem 3.1

Let us first summarize the proof. We link the current model with a "complete" model where γ is also observed. This model is fully parametric and thus can be analyzed easily. Specifically, we show in a first step that this complete model is differentiable in quadratic mean (see, e.g. van der Vaart, 2000, pp.64-65 for a definition) and has a nonsingular information matrix. In a second step, we establish an abstract expression for the semiparametric efficiency bound. This expression involves in particular the kernel $\mathcal{K}$ of the conditional expectation operator $g \mapsto \mathbb{E}[g(X, Y)|X, \gamma]$. In a third step, we show that

$$\mathcal{K} = \{(x, y) \mapsto q(x)m(x, y; \beta_0), \mathbb{E}[q^2(X)] < \infty\}. \tag{A.18}$$

The fourth step of the proof concludes.

First step: the complete model is differentiable in quadratic mean and has a nonsingular information matrix. Let $p(y|x, g; \beta) = \mathbb{P}(Y = y|X = x, \gamma = g; \beta)$. We check that the conditions of Lemma 7.6 in van der Vaart (2000) hold. Under, Assumptions 1-2, we have

$$p(y|x, g; \beta) = \prod_{t:y_t=1} F(x_{it}'\beta + g) \prod_{t:y_t=0} (1 - F(x_{it}'\beta + g)),$$

where F is C^∞ on $\mathbb{R}$ and takes values in $(0, 1)$. This implies that $\beta \mapsto \ln p(y|x, g; \beta)$ is differentiable. Let $S_\beta = \partial \ln p(Y|X, \gamma; \beta)/\partial\beta$ and let $S_{\beta k}$ denote its k-th component. We prove that $\mathbb{E}[S_{\beta k}^2] < \infty$. First, remark that

$$S_{\beta k} = \sum_{t=1}^{T} \frac{X_{k,t} f(X_t'\beta + \gamma)}{[F(X_t'\beta + \gamma)][1 - F(X_t'\beta + \gamma)]} \left[Y_t - F(X_t'\beta + \gamma)\right].$$

Next, we have

$$
\begin{aligned}
|S_{\beta k}| &\leq \sum_{t=1}^{T} |X_{k,t}| \frac{f(X_t'\beta+\gamma)}{F(X_t'\beta+\gamma)(1-F(X_t'\beta+\gamma))} \\
&= \sum_{t=1}^{T} |X_{k,t}| \frac{\sum_{j=1}^{T-1} w_j \lambda_{0j} e^{\lambda_{0j}(X_t'\beta+\gamma)}}{\sum_{j=1}^{T-1} w_j e^{\lambda_{0j}(X_t'\beta+\gamma)}} \\
&\leq \lambda_{0\tau} \sum_{t=1}^{T} |X_{k,t}|, \qquad \text{(A.19)}
\end{aligned}
$$

where we have used the triangle inequality and $|Y_t - F(X_t'\beta+\gamma)| \leq 1$ to obtain the first inequality. Equation (A.19) and Assumption 1.2 imply that $\mathbb{E}[S_{\beta k}^2] < \infty$. By the dominated convergence theorem and again (A.19), $\beta \mapsto \mathbb{E}[S_\beta S_\beta']$ is continuous. Therefore, the conditions in Lemma 7.6 in van der Vaart (2000) hold, and the complete model is differentiable in quadratic mean. Moreover,

$$
\mathbb{E}[S_\beta S_\beta'] = \mathbb{E}[\mathbb{V}(S_\beta|X,\gamma)] = \sum_{t=1}^{T} \mathbb{E}\left[\left(\frac{f(X_t'\beta+\gamma)}{[F(X_t'\beta+\gamma)][1-F(X_t'\beta+\gamma)]}\right)^2 X_t X_t'\right].
$$

Then, if for some $\lambda \in \mathbb{R}^K$, $\lambda'\mathbb{E}[S_\beta S_\beta']\lambda = 0$, we would have $X_t'\lambda = 0$ almost surely for all $t \in \{1,\dots,T\}$. By Assumption 3.1, this implies $\lambda = 0$. Hence, the information matrix $\mathbb{E}[S_\beta S_\beta']$ is nonsingular.

Second step: $V^\star$ depends on the orthogonal projection of $\mathbb{E}[S_{\beta_0}|X,Y]$ on $\mathcal{K}$. Let $\widetilde{\psi} = (\widetilde{\psi}_1,\dots,\widetilde{\psi}_K)'$ denote the efficient influence function, as defined p.363 of van der Vaart (2000). Then $V^\star = \mathbb{E}[\widetilde{\psi}\widetilde{\psi}']$ and $\mathbb{E}[\widetilde{\psi}] = 0$. Let $\mathscr{S}$ =span(S_{β_0}), $\mathscr{G} = \{q : \mathbb{E}[q^2(X,\gamma)] < \infty, \mathbb{E}[q(X,\gamma)] = 0\}$ and for any closed convex set A and any $h = (h_1,\dots,h_K)'$, let Π_A denote the orthogonal projection on A and $\Pi_A(h) = (\Pi_A(h_1),\dots,\Pi_A(h_K))'$. By Equation (25.29), Lemma 25.34 (since the complete model is differentiable in quadratic mean by the first step) and the same reasoning as in Example 25.36 of van der Vaart (2000), $\widetilde{\psi}$ is the function of (X,Y) of minimal L^2-norm satisfying

$$
\widetilde{\chi} = \Pi_{\mathscr{S}+\mathscr{G}}(\widetilde{\psi}), \qquad \text{(A.20)}
$$

where $\widetilde{\chi}$ is the efficient influence function of the large model. Because this large model is parametric, we have

$$
\widetilde{\chi} = \mathbb{E}[S_{\beta_0}S_{\beta_0}']^{-1}S_{\beta_0}. \qquad \text{(A.21)}
$$

Equation (A.20) implies $\mathbb{E}[(\widetilde{\psi}-\widetilde{\chi})\widetilde{\chi}']=0$. Thus, defining $\ell_{\beta_0}=\mathbb{E}[S_{\beta_0}|Y,X]$, we get

$$\mathbb{E}[\widetilde{\psi}\ell_{\beta_0}']=\mathbb{E}[\widetilde{\psi}S_{\beta_0}']=\mathrm{Id}, \tag{A.22}$$

Moreover, because $\mathbb{E}[S_{\beta_0}|X,\gamma]=0$, $\mathscr{S}$ and $\mathscr{G}$ are orthogonal. Thus, (A.20) is equivalent to $\Pi_{\mathscr{S}}(\widetilde{\chi})=\Pi_{\mathscr{S}}(\widetilde{\psi})$ and $\Pi_{\mathscr{G}}(\widetilde{\chi})=\Pi_{\mathscr{G}}(\widetilde{\psi})$. Moreover, (A.21) implies that $\Pi_{\mathscr{G}}(\widetilde{\chi})=0$. Hence, $\widetilde{\psi}\in\mathcal{K}^K$. Now, because $\Pi_{\mathcal{K}}$ is an orthogonal projector, we have

$$\mathbb{E}[\widetilde{\psi}\Pi_{\mathcal{K}}(\ell_{\beta_0})']=\mathbb{E}[\Pi_{\mathcal{K}}(\widetilde{\psi})\ell_{\beta_0}']=\mathbb{E}[\widetilde{\psi}\ell_{\beta_0}']=\mathrm{Id},$$

where the last equality follows by (A.22). Hence, if $\Pi_{\mathcal{K}}(\ell_{\beta_0})'\lambda=0$ a.s., we would have $\lambda=0$. In other words, $\mathbb{E}[\Pi_{\mathcal{K}}(\ell_{\beta_0})\Pi_{\mathcal{K}}(\ell_{\beta_0})']$ is nonsingular. Now, consider the set,

$$\mathcal{F}=\left\{\mathbb{E}[\Pi_{\mathcal{K}}(\ell_{\beta_0})\Pi_{\mathcal{K}}(\ell_{\beta_0})']^{-1}\Pi_{\mathcal{K}}(\ell_{\beta_0})+v:\ \mathbb{E}[v\Pi_{\mathcal{K}}(\ell_{\beta_0})']=0\right\}.$$

$\mathcal{F}$ is thus the set of vector-valued functions ψ satisfying the equation $\mathbb{E}[\psi\Pi_{\mathcal{K}}(\ell_{\beta_0})]=$Id. Hence, $\widetilde{\psi}$ being the element of $\mathcal{F}$ with minimum L^2-norm, we obtain

$$\widetilde{\psi}=\mathbb{E}[\Pi_{\mathcal{K}}(\ell_{\beta_0})\Pi_{\mathcal{K}}(\ell_{\beta_0})']^{-1}\Pi_{\mathcal{K}}(\ell_{\beta_0}).$$

Finally, because $V^\star=\mathbb{E}[\widetilde{\psi}\widetilde{\psi}']$,

$$V^\star=\mathbb{E}[\Pi_{\mathcal{K}}(\ell_{\beta_0})\Pi_{\mathcal{K}}(\ell_{\beta_0})']^{-1}. \tag{A.23}$$

Third step: (A.18) **holds.** Let $r\in\mathcal{K}$ and let us prove that $r(y,x)=q(x)m(y,x;\beta_0)$ for some q. First, by definition of $\mathcal{K}$, we have, for almost all $(g,x)\in\mathrm{Supp}(\gamma,X)$,

$$\begin{aligned}0=&r((0,0,0),x_0)+r((1,0,0),x_0)G(x_1'\beta_0+g)+r((0,1,0),x_0)G(x_2'\beta_0+g)\\&+r((0,0,1),x_0)G(x_3'\beta_0+g)+r((1,1,0),x_0)G(x_1'\beta_0+g)G(x_2'\beta_0+g)\\&+r((1,0,1),x_0)G(x_1'\beta_0+g)G(x_3'\beta_0+g)+r((0,1,1),x_0)G(x_2'\beta_0+g)G(x_3'\beta_0+g)\\&+r((1,1,1),x_0)G(x_1'\beta_0+g)G(x_2'\beta_0+g)G(x_3'\beta_0+g).\end{aligned} \tag{A.24}$$

Let $a_t:=x_t'\beta_0$ for $t\in\{1,2,3\}$ and, for the sake of conciseness, let us remove the dependence of r on x. Then, using Assumption 2, we obtain, for almost all (g,x),

$$\begin{aligned}0=&A_1e^{0\times g}+A_2e^{g}+A_3e^{\lambda_{02}g}+A_4e^{2g}+A_5e^{2\lambda_{02}g}+A_6e^{(1+\lambda_{02})g}+A_7e^{3g}+A_8e^{(2+\lambda_{02})g}\\&+A_9e^{(1+2\lambda_2)g}+A_{10}e^{3\lambda_{02}g},\end{aligned}$$

where

$$
\begin{aligned}
A_1 =& r(0,0,0),\\
A_2 =& w_1 \left[r(1,0,0)e^{a_1} + r(0,1,0)e^{a_2} + r(0,0,1)e^{a_3}\right],\\
A_3 =& w_2 \left[r(1,0,0)e^{\lambda_{02}a_1} + r(0,1,0)e^{\lambda_{02}a_2} + r(0,0,1)e^{\lambda_{02}a_3}\right],\\
A_4 =& w_1^2 \left[r(1,1,0)e^{(a_1+a_2)} + r(1,0,1)e^{(a_1+a_3)} + r(0,1,1)e^{(a_2+a_3)}\right],\\
A_5 =& w_1 w_2 \Big[r(1,1,0)(e^{a_1+\lambda_{02}a_2} + e^{a_2+\lambda_{02}a_1}) + r(1,0,1)(e^{a_1+\lambda_{02}a_3} + e^{a_3+\lambda_{02}a_1})\\
&+ r(0,1,1)(e^{a_2+\lambda_{02}a_3} + e^{a_3+\lambda_{02}a_2})\Big],\\
A_6 =& w_2^2 \left[r(1,1,0)e^{\lambda_{02}(a_1+a_2)} + r(1,0,1)e^{\lambda_{02}(a_1+a_3)} + r(0,1,1)e^{\lambda_{02}(a_2+a_3)}\right],\\
A_7 =& w_1^3 r(1,1,1)e^{a_1+a_2+a_3},\\
A_8 =& w_1^2 w_2 r(1,1,1)\left[e^{a_1+a_2+\lambda_{02}a_3} + e^{a_1+\lambda_{02}a_2+a_3} + e^{\lambda_{02}a_1+a_2+a_3}\right],\\
A_9 =& w_1 w_2^2 r(1,1,1)\left[e^{a_1+\lambda_{02}(a_2+a_3)} + e^{a_2+\lambda_{02}(a_1+a_3)} + e^{a_3+\lambda_{02}(a_1+a_2)}\right],\\
A_{10} =& w_2^3 r(1,1,1)e^{\lambda_{02}(a_1+a_2+a_3)}.
\end{aligned}
$$

Since $\lambda_{02} = 2$ is excluded by assumption, there are three cases left depending on the number of different exponents in Equation (A.24).

First, we consider $\lambda_{02} \notin \{3/2, 3\}$. By Lemma B.1 and because $|\mathrm{Supp}(\gamma|X)| \geq 10$, we obtain $A_k = 0$ for all $k \in \{1, \dots, 10\}$. $A_1 = A_7 = 0$ imply that $r(0,0,0) = r(1,1,1) = 0$. Next, $A_4 = A_6 = 0$ implies that either $r(1,0,1) = r(1,1,0) = r(0,1,1) = 0$ or

$$
\begin{cases}
r(1,1,0) &= -r(1,0,1)e^{\lambda_{02}(a_3-a_2)} - r(0,1,1)e^{\lambda_{02}(a_3-a_1)},\\
r(1,1,0) &= -r(1,0,1)e^{(a_3-a_2)} - r(0,1,1)e^{(a_3-a_1)}.
\end{cases} \tag{A.25}
$$

Consider the second case. $A_5 = 0$ implies, since $(r(1,0,1), r(1,1,0), r(0,1,1)) \neq (0,0,0)$,

$$
r(1,1,0) = -r(1,0,1)\frac{e^{a_1+\lambda_{02}a_3} + e^{a_3+\lambda_{02}a_1}}{e^{a_1+\lambda_{02}a_2} + e^{a_2+\lambda_{02}a_1}} - r(0,1,1)\frac{e^{a_2+\lambda_{02}a_3} + e^{a_3+\lambda_{02}a_2}}{e^{a_1+\lambda_{02}a_2} + e^{a_2+\lambda_{02}a_1}}.
$$

By assumption, for almost every $x = (x_1, x_2, x_3)$, $a_3 \neq a_2$ and $a_3 \neq a_1$. Then, using the latter display with equation (A.25) yields, since $\lambda_{02} \neq 1$,

$$\begin{aligned} r(1,0,1) &= r(0,1,1)\left[e^{\lambda_{02}(a_3-a_2)} - e^{a_3-a_2}\right]^{-1}\left[e^{a_3-a_1} - e^{\lambda_{02}(a_3-a_1)}\right], \\ r(1,0,1) &= r(0,1,1)\left[e^{\lambda_{02}(a_3-a_2)} - \frac{e^{a_1+\lambda_{02}a_3} + e^{a_3+\lambda_{02}a_1}}{e^{a_1+\lambda_{02}a_2} + e^{a_2+\lambda_{02}a_1}}\right]^{-1} \\ &\quad \times \left[\frac{e^{a_2+\lambda_{02}a_3} + e^{a_3+\lambda_{02}a_2}}{e^{a_1+\lambda_{02}a_2} + e^{a_2+\lambda_{02}a_1}} - e^{\lambda_{02}(a_3-a_1)}\right]. \end{aligned}$$

Since $(r(1,1,0), r(1,0,1), r(0,1,1)) \neq (0,0,0)$, these equalities and (A.25) imply that $r(1,0,1) \neq 0$ and $r(0,1,1) \neq 0$. Then

$$\frac{e^{(1-\lambda_{02})a_2}}{e^{(1-\lambda_{02})a_1}} \frac{e^{a_3+\lambda_{02}a_2+(\lambda_{02}-1)a_1} - e^{\lambda_{02}(a_2+a_3)}}{e^{\lambda_{02}(a_1+a_2)} - e^{(\lambda_{02}-1)a_2+\lambda_{02}a_1+a_3}} = \frac{e^{a_3+\lambda_{02}a_2+(\lambda_{02}-1)a_1} - e^{\lambda_{02}(a_2+a_3)}}{e^{\lambda_{02}(a_1+a_2)} - e^{(\lambda_{02}-1)a_2+\lambda_{02}a_1+a_3}},$$

which is equivalent to $a_1 = a_2$. By assumption, the set of x for which this occurs is of probability zero. In other words, for almost every x,

$$r((1,1,0),x) = r((1,0,1),x) = r((0,1,1),x) = 0.$$

$A_2 = A_3 = 0$ implies that either $r(1,0,0) = r(0,1,0) = r(0,0,1) = 0$ or

$$\begin{cases} r(0,0,1) = -e^{(a_1-a_3)}r(1,0,0) - e^{(a_2-a_3)}r(0,1,0), \\ r(0,0,1) = -e^{\lambda_{02}(a_1-a_3)}r(1,0,0) - e^{\lambda_{02}(a_2-a_3)}r(0,1,0). \end{cases}$$

In the first case, almost surely $r(Y,X) = 0 = 0 \times m(Y,X;\beta_0)$. In the second case, $r(Y,X) = q(X) \times m(Y,X;\beta_0)$ for some $g \in L^2_X$. The result follows.

Now, we turn to $\lambda_{02} = 3/2$. Then, for almost all $(g,x) \in \mathrm{Supp}(\gamma, X)$,

$$0 = A_1e^{0\times g} + A_2e^{g} + A_3e^{\frac{3}{2}g} + A_4e^{2g} + (A_5 + A_7)e^{3g} + A_6e^{\frac{5}{2}g} + A_8e^{\frac{7}{2}g} + A_9e^{4g} + A_{10}e^{\frac{9}{2}g}.$$

By Lemma B.1 and because $|\mathrm{Supp}(\gamma|X)| \geq 9$, we obtain $A_5 + A_7 = 0$ and $A_k = 0$ for all $k \notin \{5,7\}$. $A_1 = A_{10} = 0$ implies that $r(0,0,0) = r(1,1,1) = 0$ which in turn implies that $A_7 = 0$ and thus $A_5 = 0$. Hence, we have $A_k = 0$ for all $k \in \{1,\ldots,10\}$ and the same reasoning as when $\lambda_{02} \notin \{3/2, 3\}$ allows us to obtain the result.

Finally, we consider $\lambda_{02} = 3$. Then, for all (g,x),

$$0 = A_1e^{0\times g} + A_2e^{g} + (A_3 + A_7)e^{3g} + A_4e^{2g} + A_5e^{6g} + A_6e^{4g} + A_7e^{5g} + A_8e^{7g} + A_9e^{9g},$$

By Lemma B.1 and because $|\mathrm{Supp}(\gamma|X)| \geq 9$, we obtain $A_3 + A_7 = 0$ and $A_k = 0$ for all $k \notin \{3,7\}$. $A_1 = A_{10} = 0$ implies that $r(0,0,0) = r(1,1,1) = 0$ which in turn implies that $A_7 = 0$ and thus $A_3 = 0$. Hence, $A_k = 0$ for all $k \in \{1,\ldots,10\}$ and the result follows again as when $\lambda_{02} \notin \{3/2, 3\}$.

Fourth step: conclusion. By Steps 2 and 3, there exists $q_0(X)$ such that $\Pi_{\mathcal{K}}(\ell_{\beta_0}) = q_0(X)m(Y,X;\beta_0)$. Moreover, by definition of the orthogonal projection, $\Pi_{\mathcal{K}}(\ell_{\beta_0}) - \ell_{\beta_0} \in (\mathcal{K}^{\perp})^K$. Hence, again by Step 3, we have, for all $q \in L^2_X$,

$$\mathbb{E}[q_0(X)q(X)m(Y,X;\beta_0)^2] = \mathbb{E}[\ell_{\beta_0}q(X)m(Y,X;\beta_0)].$$

This implies that

$$q_0(X)\Omega(X) = \mathbb{E}[\ell_{\beta_0}m(Y,X;\beta_0)|X].$$

As a result, because $\ell_{\beta_0} = \mathbb{E}[S_{\beta_0}|Y,X]$,

$$\begin{aligned}\Pi_{\mathcal{K}}(\ell_{\beta_0}) =& \Omega^{-1}(X)m(Y,X;\beta_0)\mathbb{E}[\ell_{\beta_0}m(Y,X;\beta_0)|X] \\ =& \Omega^{-1}(X)m(Y,X;\beta_0)\mathbb{E}[S_{\beta_0}m(Y,X;\beta_0)|X].\end{aligned}$$

Then, using (A.23), we obtain

$$V^\star = \mathbb{E}\left[\Omega^{-1}(X)\mathbb{E}[S_{\beta_0}m(Y,X;\beta_0)|X]\mathbb{E}[S_{\beta_0}m(Y,X;\beta_0)|X]'\right]^{-1}.$$

Now, by the end of the proof of Theorem 2.3, we have, for all β,

$$0 = \mathbb{E}_\beta\left[m(Y,X;\beta)|X,\gamma\right].$$

As a result,

$$\begin{aligned}0 &= \nabla_\beta \mathbb{E}_\beta\left[m(Y,X;\beta)|X,\gamma\right] \\ &= \mathbb{E}_\beta\left[\nabla_\beta m(Y,X;\beta)|X,\gamma\right] + \mathbb{E}_\beta\left[m(Y,X;\beta)S_\beta|X,\gamma\right].\end{aligned}$$

Evaluating this equality at β_0 and integrating over γ yields:

$$\mathbb{E}[S_{\beta_0}m(Y,X;\beta_0)|X] = -\mathbb{E}[\nabla_\beta m(Y,X;\beta_0)|X] = -R(X).$$

We conclude that

$$V^\star = \mathbb{E}\left[\Omega^{-1}(X)R(X)R(X)'\right]^{-1} = V_0,$$

which is a well-defined matrix by Assumption 6.1.

B Technical lemmas

The following two lemmas are keys in the proof of Proposition 2.5.

Lemma B.1 *Let $n \geq 1$, $(\alpha_1, \ldots, \alpha_n)$ be n distinct real numbers, $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and $P(x) = \sum_{i=1}^n a_i \exp(\alpha_i x)$. If P has n distinct roots, then $a_1 = \ldots = a_n = 0$.*

Lemma B.2 *For any $(t, \ell) \in \{1, \ldots, T\} \times \{1, \ldots, \tau\}$, $a_{t,\ell,x}$ is real analytic for almost all Supp(X).*

B.1 Proof of Lemma B.1

This follows by induction on n and Rolle's theorem, see e.g. Chapter 2, section 2 of Krein and Nudelman (1977).

B.2 Proof of Lemma B.2

We want to prove that each function $a_{t,\ell,x}$ is real analytic for almost all $x \in \text{Supp}(X)$. Fix $x \in \text{Supp}(X)$, and let $\tilde{w}_j^\gamma := w_j \delta_j(x, \theta_0, t) \exp(\lambda_{0j}\gamma)$, $\tilde{\lambda}_{0j} := \lambda_{0j}\beta_{0k}$. Let us define $f : (v, \gamma) \mapsto 1/\left(1 + \sum_{j=1}^\tau \tilde{w}_j^\gamma \exp(\tilde{\lambda}_{0j} v)\right)$. We have

$$a_{t,\ell,x}(v) = \int \frac{\exp(\lambda_{0\ell}\gamma)}{C(\gamma, x; \theta_0, t)} f(v, \gamma) dF_{\gamma|X=x}(\gamma), \quad \forall v \in \mathbb{R}.$$

We prove the result in three steps. First, we establish a bound on the derivatives of f. Second, we show that $a_{t,\ell,x}$ is C^∞, and we bound its deratives. Finally, we show that $a_{t,\ell,x}$ is real analytic.

First step: for all $k \geq 0$ and all (v, γ),

$$\left|\frac{\partial^k}{\partial v^k} f(v, \gamma)\right| \leq k!(e\lambda_{0\tau}|\beta_{0k}|)^k f(v, \gamma). \tag{B.1}$$

For any infinitely differentiable real function $g : \mathbb{R} \times \text{Supp}(\gamma|X = x) \to \mathbb{R}$, we let $g^{(k)}(v, \gamma) = \partial^k g(v, \gamma)/\partial v^k$ and define $P : (v, \gamma) \mapsto \sum_{j=1}^\tau \tilde{w}_j^\gamma \tilde{\lambda}_{0j} \exp(\tilde{\lambda}_{0j} v)$. First, remark that for any positive integer k,

$$\begin{aligned}\left|P^{(k)}(v,\gamma)\right| =& \left|\sum_{j=1}^{\tau} \tilde{w}_j^{\gamma} \tilde{\lambda}_{0j}^{k+1} \exp(\tilde{\lambda}_{0j} v)\right| \\ \leq& |\tilde{\lambda}_{0\tau}^{k+1}| \left|\sum_{j=1}^{\tau} \tilde{w}_j^{\gamma} \exp(\tilde{\lambda}_{0j} v)\right| \\ \leq& |\tilde{\lambda}_{0\tau}^{k+1}| / f(v,\gamma). \end{aligned} \tag{B.2}$$

Now, we prove (B.1) by induction. The result is trivial for $k = 0$. Suppose that it holds for $j = 0, \ldots, k$, $k \geq 0$. Remark that $f^{(1)} = f \times (fP)$. Then, by applying twice the general Leibniz rule, we obtain

$$\begin{aligned}\left|f^{(k+1)}\right| &= \left|\sum_{j=0}^{k} \binom{k}{j} (f)^{(j)} (fP)^{(k-j)}\right| \\ &\leq \sum_{j=0}^{k} \binom{k}{j} |f^{(j)}||(fP)^{(k-j)}| \\ &\leq f \sum_{j=0}^{k} \binom{k}{j} j!(e\tilde{\lambda}_{0\tau})^j \left|\sum_{i=0}^{k-j} \binom{k-j}{i} f^{(i)} P^{(k-j-i)}\right| \\ &\leq f \sum_{j=0}^{k} \left[\binom{k}{j} j!(e\tilde{\lambda}_{0\tau})^j \sum_{i=0}^{k-j} \binom{k-j}{i} i!(e\tilde{\lambda}_{0\tau})^i f \tilde{\lambda}_{0\tau}^{k-j-i+1} \frac{1}{f}\right] \\ &\leq f \tilde{\lambda}_{0\tau}^{k+1} e^k \sum_{j=0}^{k} \binom{k}{j} j! \left[\sum_{i=0}^{k-j} \binom{k-j}{i} i!\right],\end{aligned}$$

where we used the induction hypothesis to get the second and third inequalities. The last inequality follows from $e^i \leq e^{k-j}, \forall i \leq k-j$. Now, notice that for any $k \in \mathbb{N}^*$, we have

$$\sum_{s=0}^{k} \binom{k}{s} s! = \sum_{s=0}^{k} \frac{k!}{(k-s)!} \leq k!e. \tag{B.3}$$

As a result,

$$\begin{aligned}\left|f^{(k+1)}\right| &\leq f \tilde{\lambda}_{0\tau}^{k+1} e^k \sum_{j=0}^{k} \binom{k}{j} j!(k-j)!e \\ &= f \tilde{\lambda}_{0\tau}^{k+1} e^k \times e \times (k+1) \times k! \\ &= (k+1)! \left(e\tilde{\lambda}_{0\tau}\right)^{k+1} f,\end{aligned}$$

and thus the induction hypothesis holds for $k+1$. This ends the first step.

Second step: $a_{t,\ell,x}$ is C^∞ and for all $k \geq 0$,

$$\sup_{v\in\mathbb{R}} \left| \frac{\partial^k a_{t,\ell,x}(v)}{\partial v^k} \right| \leq C_{t,\ell,x,\theta_0} k!(e\lambda_{0\tau}|\beta_{0k}|)^k, \tag{B.4}$$

for some $C_{t,\ell,x,\theta_0} > 0$.

First, for all $v \in \mathbb{R}$, we have $1/f(v) \geq \tilde{w}_\ell^\gamma \exp(\tilde{\lambda}_{0\ell} v)$ and $C(\gamma, x; \theta_0, t) \geq 1$. Thus,

$$\frac{\exp(\lambda_{0\ell}\gamma)}{C(\gamma, x; \theta_0, t)} f(v, \gamma) \leq \frac{1}{w_\ell \delta_\ell(x, \theta_0, t)}. \tag{B.5}$$

Hence, (B.4) holds for $k = 0$, with $C_{t,\ell,x,\theta_0} = 1/[w_\ell \delta_\ell(x, \theta_0, t)]$. Next, $v \mapsto \exp(\lambda_{0\ell}\gamma) \times f(v, \gamma)/C(\gamma, x; \theta_0, t)$ is C^∞ and by (B.5) and the previous step, we have, for any $k \geq 0$,

$$\begin{aligned} \left| \frac{\partial^k}{\partial v^k} \left(\frac{\exp(\lambda_{0\ell}\gamma)}{C(\gamma, x; \theta_0, t)} f(v, \gamma) \right) \right| &\leq k!(e\lambda_{0\tau}|\beta_{0k}|)^k \frac{\exp(\lambda_{0\ell}\gamma)}{C(\gamma, x; \theta_0, t)} f(v, \gamma) \\ &\leq \frac{k!(e\lambda_{0\tau}|\beta_{0k}|)^k}{w_j \delta_j(x, \theta_0, t)}. \end{aligned}$$

Thus, by the dominated convergence theorem, $a_{t,\ell,x}$ is C^k and we have

$$\begin{aligned} \left| \frac{\partial^k a_{t,\ell,x}(v)}{\partial v^k} \right| &\leq \int \left| \frac{\partial^k}{\partial v^k} \left(\frac{\exp(\lambda_{0\ell}\gamma)}{C(\gamma, x; \theta_0, t)} f(v, \gamma) \right) \right| \mathrm{d}F_{\gamma|X=x}(\gamma) \\ &\leq k!(e\lambda_{0\tau}|\beta_{0k}|)^k \int \frac{\exp(\lambda_{0\ell}\gamma)}{C(\gamma, x; \theta_0, t)} f(v, \gamma) \mathrm{d}F_{\gamma|X=x}(\gamma) \\ &= k!(e\lambda_{0\tau}|\beta_{0k}|)^k a_{t,\ell,x}(v) \\ &\leq C_{t,\ell,x,\theta_0} k!(e\lambda_{0\tau}|\beta_{0k}|)^k. \end{aligned}$$

Third step: $a_{t,\ell,x}$ is real analytic. It suffices to show that there exists $R > 0$ such that for all v, $a_{t,\ell,x}$ coincides with its Taylor expansion at v on $(v - R, v + R)$. Let $R < 1/(2e\lambda_{0\tau}|\beta_{0k}|)$. First, by the second step, we have, for any $v' \in (v - R, v + R)$,

$$\begin{aligned} \left| \frac{1}{k!} (v' - v)^k \frac{\partial^k a_{t,\ell,x}(v)}{\partial v^k} \right| &\leq \frac{1}{k!} R^k \sup_v \left| \frac{\partial^k a_{t,\ell,x}(v)}{\partial v^l} \right| \\ &\leq C_{t,\ell,x,\theta_0} (Re\lambda_{0\tau}|\beta_{0k}|)^k, \end{aligned} \tag{B.6}$$

and the corresponding series converges since $Re\lambda_{0\tau}|\beta_{0k}| < 1$. Thus, the Taylor series of $a_{t,\ell,x}$ converges at v, for any $v' \in (v - R, v + R)$. Finally, by the second step again and Taylor's theorem applied to $a_{t,\ell,x}(v')$, we obtain, for any $K > 0$ and uniformly for $|v - v'| < R$:

$$\begin{aligned}
\left| a_{t,\ell,x}(v') - \sum_{k=0}^{K} \frac{1}{k!} (v' - v)^k \frac{\partial^k a_{t,\ell,x}(v)}{\partial v^k} \right| \leq & \frac{R^{K+1}}{(K+1)!} \sup_{|v'-v|<R} \left| \frac{\partial^{K+1} a_{t,\ell,x}(v')}{\partial v^{K+1}} \right| \\
\leq & C_{t,\ell,x,\theta_0} (Re \lambda_{0\tau} |\beta_{0k}|)^{K+1} \\
\to & \, 0.
\end{aligned}$$

This completes the proof.