Guarantees on Correct Conclusions with Incorrect Likelihoods*

Thomas Wiemann†

June 29, 2025

Abstract

This note studies robustness properties of (non)linear control function estimands such as (mixed) Logistic or Poisson pseudo maximum likelihood estimands. I show that under misspecification, commonly-applied estimands are not informative about the sign of the true partial effects. For example, (mixed) logistic regression estimands potentially imply positive partial effects even if all true partial effects are negative. I provide sufficient conditions to admit valid conclusions about the sign of partial effects. For a large class of estimands, including common pseudo maximum likelihood estimands based on natural exponential family distributions, nonparametrically conditioning on the control function is sufficient for sign preservation.

JEL codes: C14, C18, C21, C25, C51.

Keywords: Weakly causal, Logistic regression, Poisson regression, pseudo maximum likelihood, natural exponential family distributions.

---

*I thank Stéphane Bonhomme, Pradeep Chintagunta, Giovanni Compiani, Max Farrell, Christian Hansen, Samuel Higbee, Ali Hortaçsu, Sanjog Misra, and Alexander Torgovitsky for valuable comments and suggestions, along with participants at the University of Chicago Econometrics advising group and the Booth Quantitative Marketing brown bag. All remaining errors are my own.

†University of Chicago, wiemann@uchicago.edu.

1 Introduction

Empirical analyses frequently place structural assumptions on unknown relationships between observed and unobserved variables to ease estimation. Many such convenience assumptions are seldom motivated by domain knowledge or economic theory. This note studies the robustness of conclusions about relationships between observed variables when these convenience assumptions are incorrect. This is motivated by the fact that commonly considered convenience assumptions can result in incorrect conclusions about even the *direction* of association between outcomes and variables of interest.

As a simple example, consider consumers choosing between a single good ($j = 1$) and an outside option ($j = 0$), and suppose the researcher is interested in how a choice feature D_1 impacts these choices. Letting $W_j \equiv (D_j, V_j^\top)^\top$ denote the observed features of choice j, it is convenient to model the consumer's choice as

$$Y = \underset{j \in \{0,1\}}{\text{argmax}} \; W_j^\top \theta + U_j, \tag{1}$$

where $W_j \perp\!\!\!\perp U_j \overset{iid}{\sim} \text{T1EV}$. When θ is a fixed parameter and the outside option is normalized to $W_0 = 0$, (1) implies familiar Logit conditional choice probabilities (CCPs) $m_\theta(W) \equiv \exp(W_j^\top \theta)/(1 + \exp(W_j^\top \theta))$ that allow for straightforward estimation. If the modeling assumptions are correct so that the true CCPs are indeed given by m_θ, then standard maximum likelihood theory implies that the Logit maximum likelihood estimator (MLE) is consistent for θ. As a consequence, the researcher's conclusions about the partial effects of D_1 on Y based on the Logit MLE are consistent as well. But what if the modeling assumptions are incorrect?

Even if the linear index structure of (1) is correct, it is possible to find a distribution for the latent utility shocks U (mean-zero and independent of W) such that the Logit *pseudo* MLE implies partial effects of D on Y that are of the opposite sign as the true partial effects. Example 1 provides a simple numerical example of such a distribution. Hence, even

with infinite data and no omitted variables, the researcher may incorrectly judge whether D increases or decreases choice probabilities.

Example 1. *Consider* $(Y, D, V^\top, U)$ *with distribution* P *where*

$$\begin{aligned}
&\Pr(D=0, V_2=0, V_3=0) = 0.01, && \Pr(D=1, V_2=0, V_3=0) = 0.36,\\
&\Pr(D=0, V_2=1, V_3=0) = 0.01, && \Pr(D=1, V_2=1, V_3=0) = 0.18,\\
&\Pr(D=0, V_2=0, V_3=1) = 0.01, && \Pr(D=1, V_2=0, V_3=1) = 0.18,\\
&\Pr(D=0, V_2=1, V_3=1) = 0.24, && \Pr(D=1, V_2=1, V_3=1) = 0.01,
\end{aligned}$$

and $U \perp\!\!\!\perp (D, V)$ *with marginal distribution*

$$\Pr(U=-9.5) = 0.2, \quad \Pr(U=1) = 0.05, \quad \Pr(U=2) = 0.7, \quad \Pr(U=9) = 0.05.$$

Finally, the binary outcome is determined via

$$Y = \mathbb{1}\{-5D + 5V_2 + 5V_3 - 2 + U \geq 0\}.$$

Clearly, partial effects with respect to D *are negative in this example. However, the Logit pseudo MLE implies a slope-coefficient for* D *that converges to approximately* $+2$ *so that model-implied partial effects with respect to* D *are positive.*

After defining the framework formally in Section 2, I show in Section 3 that Example 1 is not a special case. For a large class of estimands, including (mixed) Logit, Normal, or Poisson pseudo maximum likelihood estimands (PMLE), it is generally not possible to infer the sign of the true partial effects from the sign of the model-implied partial effects. The result highlights potentially grave negative consequences of convenience assumptions and motivates a need for estimands that are more robust to misspecification.

Section 4 defines two robustness criteria: Weak and strong sign preservation. Heuristically, a *non*-sign preserving estimand can imply partial effects that are of the opposite sign as all true partial effects. A sign preserving estimand guarantees against such drastically

incorrect conclusions. While sign preservation is not sufficient to judge the quality of quantitative conclusions about partial effects or counterfactuals, it is difficult to motivate the use of estimands that may lead to even wrong qualitative conclusions. Indeed, robustness properties of linear estimands analogous to sign preservation — including the "weak causality" property of Blandhol et al. (2022) — are often viewed as minimal robustness properties (see also Bugni et al., 2023; Sävje, 2024; Leung, 2024).

Section 5 then presents simple sufficient conditions for sign preservation. The presented conditions are satisfied by a large class of commonly used estimands, including versions of Gaussian, Poisson, and Logit PMLEs. Indeed, PMLEs based on the class of natural exponential family distributions are shown to have appealing robustness properties under misspecification. This has potentially surprising implications for discrete choice estimation: The Logit PMLEs discussed in this note are sign preserving and thus guaranteed not to result in drastically wrong qualitative conclusions about the direction of partial effects. This guarantee holds even if the true partial effects are generated by a mixed Logit model *regardless of the true mixing distribution*. In contrast, there are no known results that guarantee similarly robust conclusions on the sign of partial effects based on a mixed Logit PMLE if, for example, the distribution of random coefficients that the researcher considers differs from the true mixing distribution.[1]

Finally, Section 6 relates sign preservation to causal inference using control functions. In particular, I consider the causal model $Y = g(D, U)$ where g is a (potentially unknown) structural function and U are all other (at least partially unobserved) determinants of Y other than D. Following Imbens and Newey (2009), a control function is then simply any (observed or identified) random vector V such that $D \perp\!\!\!\perp U|V$. Examples of control functions

[1]This result does not contradict the influential universal approximation properties of mixed Logit models of McFadden and Train (2000) and further developed in Lu and Saito (2022) and Chang et al. (2022), since these previously-established results rely crucially on availability of a nonparametric mixing distribution. In contrast, I focus on parametric mixing distributions as conventionally applied in practice.

V used in practice include observed controls or residuals from a first stage regression of D on instruments.[2] In this control function setup, sign preservation of an estimand is shown to be equivalent to correct conclusions about the sign of conditional causal effects of D on Y.

Throughout the main text, I focus on a simple setting with a scalar-valued outcome Y, a scalar-valued variable of interest D, and a vector of covariates V. Their joint distribution is P in a class of possible distributions $\mathcal{P}$. For all $P \in \mathcal{P}$, I assume that Y, D, and V have finite second moments, that the support of Y given by $\mathcal{Y} \subset [\underline{y}, \overline{y}]$ with either $\underline{y}$ or $\overline{y}$ potentially infinite, the support of D given by $\mathcal{D} \equiv \{d_1, \ldots, d_J\}$ with $J \in \mathbb{N}$,[3] and the support of V denoted by $\mathcal{V}$ are fixed, and that the support of (D, V) is $\mathcal{D} \times \mathcal{V}$. These assumptions greatly simplify the exposition while still allowing for discussion of the main insights. In Appendix A, I extend the result of the main text to vector-valued Y and D to also accommodate analysis of, for example, multinomial Logit PMLEs.

Literature. This note draws from and contributes to several strands of literature. A large literature in econometrics studies statistical properties of estimators for so-called pseudo-true parameters that maximize a population-level objective function (e.g., White, 1982). Standard maximum likelihood theory shows, for example, that under mild conditions, the MLE converges to the parameter that minimizes the Kullback-Leibler distance between the true likelihood and the likelihood the researcher considers. As highlighted by Andrews et al. (2024), however, these "pseudo-true" parameters are not in general policy relevant if the

[2]In demand estimation, control function approaches are often viewed as alternatives to the structural IV estimands proposed by Berry et al. (1995). While placing strong assumptions on the source of endogeneity (Blundell and Matzkin, 2014), control function approaches have several advantages. For example, control function estimands are more readily applicable in settings with relatively few consumers per market than the estimator of Berry et al. (1995) — see, in particular, Petrin and Train (2010) and Kim and Petrin (2019) for control function approaches in demand estimation. See also Wooldridge (2015) for a general overview of control function approaches.

[3]See also Remark 3 for heuristic extension to settings with continuous D.

policy depends on the *value* of the true parameter.[4] In contrast, the analysis of this note suggests that a large class of pseudo-true parameters is policy relevant if the policy depends on the *sign* of the true parameter.

In the context of discrete-choice estimation, recent literature has highlighted the worry that conclusions based on pseudo-true parameters are driven by particular distributional assumptions such as T1EV or normally distributed random coefficients as is common in mixed Logit estimation (Compiani, 2022; Tebaldi et al., 2023). In response, Compiani (2022) and Tebaldi et al. (2023) propose nonparametric demand estimators to avoid possible model-misspecification. While guaranteeing quantitatively accurate conclusions with infinite data, the fully nonparametric approaches pose substantially more difficult estimation problems than commonly considered in applied research. This note takes an alternative approach, characterizing the extent to which conclusions based on commonly used estimands are robust to incorrect distributional assumptions.[5] In this regard, the results of this note complement recent work of Andrews et al. (2023), who study the robustness properties of structural IV estimands such as those suggested by Berry et al. (1995). The proposed sign preservation property of this note is a substantially stronger robustness property than the sharp-zero consistency property of Andrews et al. (2023). In particular, in contrast to sign preservation, sharp-zero consistency cannot guarantee that strictly positive model-implied partial effects were not generated by a distribution with strictly negative true partial effects. The additional strength of the proposed sign preservation property comes at the cost of analyzing a more

[4]Important exceptions are discussed in Gourieroux et al. (1984), who show that for linear exponential family PMLEs, correct specification of the conditional expectation function (rather than the full distribution) suffices for correct inference on partial effects. This note extends these results by analyzing robustness also under misspecification of the conditional expectation function.

[5]This also relates to the work of Ruud (1983) and Li and Duan (1989) who provide sufficient conditions for maximum likelihood estimands to be correct up to an unknown scalar. In contrast to their work that imposes linear index assumptions on the outcome model, I focus primarily on settings without distributional assumptions on the data generating process.

restricted class of estimands: While I show that mixed Logit estimands can suffer from sign reversal under misspecification, this note does not provide (non-trivial) sufficient conditions for sign preservation of mixed Logit estimands with non-degenerate mixing distributions. Instead, sufficient conditions focus on estimands corresponding to models with coefficients that are allowed to vary in observed covariates.[6]

The approach of characterizing robustness of economic conclusions based on pseudo-true parameters draws heavily from the literature on positively-weighted causal effects in linear regression (Yitzhaki, 1996; Angrist, 1998; Angrist and Krueger, 1999; Angrist and Pischke, 2009; Słoczyński, 2022), linear two stage least squares (e.g., Angrist et al., 2000; Blandhol et al., 2022; Borusyak and Hull, 2024), and difference-in-differences (e.g., de Chaisemartin and Xavier d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; Baker et al., 2022; Borusyak et al., 2024). In contrast to the estimands studied previously, however, the estimands I consider do not generally ensure that model-implied partial effects are within the convex hull of the true partial effects.[7] Instead, I define a weaker version of robustness that is based only on the property that *a* convex combination of model-implied partial effects is equal to a convex combination of the true partial effects.

The analysis of pseudo-true estimands in misspecified nonlinear models is also connected to recent work by Silva and Winkelmann (2024) who highlight sufficient conditions for Poisson PMLEs to capture average marginal effects. The authors show that for jointly normally distributed covariates, the model-implied average partial effects corresponding to the Poisson linear index PMLE equal the true average partial effects. While the results in this note also

[6]See, for example, Dubé and Misra (2023) who apply a Logit model with coefficients given by nonparametric functions of consumer characteristics in their welfare analysis of personalized pricing.

[7]In particular, in nonlinear models where the model-implied partial effect varies with covariates (e.g., Logit), it is not generally possible to guarantee that that model-implied partial effects are within the convex hull of the true partial effects. This follows immediately from considering a setting where the true partial effects are constant in the covariates.

suggest that this equality of model-implied and true partial effects holds more generally for a large class of estimands under assumptions of (conditional) normality, I focus primarily on settings without restrictions on the joint distribution of covariates. This seems particularly important as assumptions on the joint distribution of covariates — such as linearity of the conditional expectation of D given covariates V — are seldom easier to motivate in practice than assumptions on the joint distribution of the outcome and covariates that initially motivated the robustness analysis.

Finally, the results in this note provide a new motivation for recently proposed partially and locally linear index estimators. In particular, I show that partially and locally linear index natural exponential family estimators such as those proposed by Liu et al. (2021), and special cases of Athey et al. (2019) and Farrell et al. (2021) target sign preserving estimands without parametric distributional assumptions on the joint distribution of the data. This is in contrast to fully linear index natural exponential family estimands (such as the Poisson PMLE considered by Silva and Winkelmann (2024)) that are only known to be sign preserving when the conditional expectation of D given covariates V is linear in V.

2 Setup

I consider the following setting: Given samples from a distribution $P \in \mathcal{P}$, a researcher constructs an estimator for a parameter $\theta^*(P) \in \Theta$. Equipped with their estimate of $\theta^*(P)$, the researcher then makes conclusions about the association between observed variables of interest in P. A large literature in econometrics focuses on differentiating estimators by their statistical properties (e.g., efficiency) while keeping the parameter of interest θ^* fixed. In contrast, this note abstracts away from any statistical questions and focuses entirely on characterizing the population-value of the estimator — i.e., the estimand $\theta^* : \mathcal{P} \to \Theta$ — and whether it is informative about the relationships of interest.

To fix ideas, I assume that the researcher forms conclusions about associations using a *model* m_θ, indexed by $\theta \in \Theta$, for the conditional expectation of Y given a variable of interest D and other covariates V. Notably, the researcher's model may be misspecified.

Definition 1. *The* ***researcher's model*** *of Y given D and V is a class of functions $m_\theta \in L_{2,\mathcal{P}} : \mathcal{D} \times \mathcal{V} \to \mathcal{Y}$, indexed by a finite dimensional parameter $\theta \in \Theta$.*[8]

Examples 2-5 give examples of commonly considered models. Example 2 references the mixed Logit model with Normal random coefficients as used frequently in demand estimation. Examples 3-5 discuss special cases of exponential family distributions, most notably the Normal, Poisson, and Logit models, with parametric, semiparametric, and nonparametric linear index functions. I refer to these models as linear, partially linear, and locally linear index natural exponential family models, respectively.

Example 2 (Linear Index Mixed Logit)**.** *Let $w \equiv (d, v^\top)^\top$, and define*

$$s_\theta(w, \psi) \equiv \frac{\exp\{w^\top(\mu + \Sigma^{\frac{1}{2}}\psi)\}}{1 + \exp\{w^\top(\mu + \Sigma^{\frac{1}{2}}\psi)\}}, \quad m_\theta(w) \equiv \int s_\theta(w, \psi) dF(\psi)$$

where $\theta = (\mu, \Sigma)$, $F(\psi)$ denotes a standard multivariate Normal distribution function, and $m_\theta(w)$ is the model-implied mean given $W = w$. The mixed Logit (pseudo) likelihood is

$$L(y, w; \theta) \equiv m_\theta(w)^y \left(1 - m_\theta(w)\right)^{1-y}.$$

Example 3 (Linear Index Natural Exponential Family)**.** *Consider the conditional (pseudo) likelihood of y given d and v defined by*

$$L(y, d, v; \theta) \propto \exp\left((d\alpha + v^\top\beta)y - A(d\alpha + v^\top\beta)\right),$$

where $\theta \equiv (\alpha, \beta)$ and A is a known function. Then, by properties of natural exponential family distributions, the model-implied mean given d and v is $m_\theta(d, v) = A'(d\alpha + v^\top\beta)$. Key

[8]Notation: $L_{2,\mathcal{P}}$ denotes the class of functions with finite second moments $\forall P \in \mathcal{P}$.

examples for commonly considered models are the Normal, Poisson, and Logit models where

$$\begin{aligned}
\textit{Normal with known } \sigma\textit{:} \quad A(d\alpha + v^\top\beta) &= \frac{1}{2}(d\alpha + v^\top\beta)^2, \\
A'(d\alpha + v^\top\beta) &= d\alpha + v^\top\beta, \\
\textit{Poisson:} \quad A(d\alpha + v^\top\beta) &= \exp\{d\alpha + v^\top\beta\}, \\
A'(d\alpha + v^\top\beta) &= \exp\{d\alpha + v^\top\beta\}, \\
\textit{Logit:} \quad A(d\alpha + v^\top\beta) &= \log(1 + \exp(d\alpha + v^\top\beta)), \\
A'(d\alpha + v^\top\beta) &= \frac{\exp(d\alpha + v^\top\beta)}{1 + \exp(d\alpha + v^\top\beta)}.
\end{aligned}$$

Example 4 (Partially Linear Index Natural Exponential Family)**.** *Consider the conditional (pseudo) likelihood of y given d and v defined by*

$$L(y, d, v; \theta) \propto \exp\left((d\alpha + b(v))y - A(d\alpha + b(v))\right),$$

where $\theta \equiv (\alpha, b)$ and $b : \mathcal{V} \to \mathbb{R}$ is unknown. The model-implied mean given d and v is $m_\theta(d, v) = A'(d\alpha + b(v))$.

Example 5 (Locally Linear Index Natural Exponential Family)**.** *Consider the conditional (pseudo) likelihood of y given d and v defined by*

$$L(y, d, v; \theta) \propto \exp\left((da(v) + b(v))y - A(da(v) + b(v))\right),$$

where $\theta \equiv (a, b)$ and $a : \mathcal{V} \to \mathbb{R}$ and $b : \mathcal{V} \to \mathbb{R}$ are unknown. The model-implied mean given d and v is $m_\theta(d, v) = A'(a(v) + b(v))$.

3 Sign Reversal

This section shows that a large class of estimands is generally uninformative about the sign of partial effects. For this purpose, I define subsets of $\mathcal{P}$ in which every conditional partial effect of D on Y given V is strictly positive or negative.

Definition 2. *D is said to have **strict P-positive association** with Y given V if*

$$\mathrm{E}_P[Y|D=d',V] \overset{a.s.}{>} \mathrm{E}_P[Y|D=d,V], \quad \forall d' \geq d \in \mathcal{D}.$$

Strict P-negative association is defined analogously. For distributions $P \in \mathcal{P}$, let $\mathcal{P}_{++} \subset \mathcal{P}$ denote the subset of distributions under which D has strict P-positive association with Y given X. Analogously, let $\mathcal{P}_{--} \subset \mathcal{P}$ denote the subset of distributions under which D has strict P-negative association with Y given V.

Uninformativeness of an estimand about the sign of partial effects can then be described by *sign reversal* (Definition 3). Sign reversal implies that for every value of the estimand, there exist at least one distribution P_{++} with strictly positive partial effects and at least one distribution P_{--} with strictly negative partial effect, both of which imply the same estimand value. Of course, given any particular setting, only either can be true, yet the researcher cannot differentiate between these distributions P_{++} and P_{--} when equipped with only the sign-reversing estimand.

Definition 3. *An estimand $\theta^* : \mathcal{P} \to \Theta$ is said to be **sign reversing** for the expected change in Y due to D given V if*

$$\forall \theta \in \theta^*(\mathcal{P}), \quad \exists P_{--} \in \mathcal{P}_{--}, P_{++} \in \mathcal{P}_{++} : \quad \theta = \theta^*(P_{--}) = \theta^*(P_{++}).$$

Sign reversal is a very undesirable property. Still, in the absence of functional form assumptions, a large class of commonly used estimands is sign reversing. These estimands are characterized by Assumptions 1-3. Assumption 1 characterizes the class of distributions $\mathcal{P}$ under consideration. For ease of exposition, I focus on the simplest setting with a single covariate throughout this section.

Assumption 1. *$\mathcal{P}$ is the set of joint distributions of the random vector (Y, D, V) where $\mathcal{Y} \subset [\underline{y}, \overline{y}] \subset \mathbb{R}$, $\mathcal{D} = \{d_1, \ldots, d_J\} \subset \mathbb{R}$ with $0 \leq d_1 < \ldots < d_J$, and $\mathcal{V} = \{v_1, \ldots, v_K\} \subset \mathbb{R}$. Further, $\forall P \in \mathcal{P}$, the support of (D, V) is $\mathcal{D} \times \mathcal{V}$.*

Assumption 2 defines the estimand through moment conditions. Examples 6 and 7 show that (pseudo) maximum likelihood estimands for the mixed Logit model and commonly considered linear index natural exponential family models satisfy Assumption 2.[9]

Assumption 2. *Let $Q \in \mathbb{N}$. The estimand $\theta^* : \mathcal{P} \to \Theta$ is defined as the solution to*

$$\begin{aligned} \mathrm{E}_P[(Y - m_{\theta^*(P)}(D,V)) f^q_{\theta^*(P)}(D,V)] &= 0, \\ \mathrm{E}_P[D(Y - m_{\theta^*(P)}(D,V)) h^q_{\theta^*(P)}(D,V)] &= 0, \\ \mathrm{E}_P[V(Y - m_{\theta^*(P)}(D,V)) l^q_{\theta^*(P)}(D,V)] &= 0, \end{aligned} \tag{2}$$

$\forall q \in [Q]$, where $(f^q_\theta)_{q\in[Q]}$, $(h^q_\theta)_{q\in[Q]}$, and $(l^q_\theta)_{q\in[Q]}$ are functions in $L_{2,\mathcal{P}}$ indexed by θ. Further, $m_\theta(\mathcal{D},\mathcal{V}) \subset \mathrm{int}(\mathcal{Y}), \forall \theta \in \theta^*(\mathcal{P})$.

Example 6 (Mixed Logit Pseudo Maximum Likelihood)**.** *Suppose $W \equiv (D, V, 1)^\top$ satisfies Assumption 1. Consider the pseudo maximum likelihood estimand that maximizes the likelihood of Example 2, or equivalently,*

$$\theta^*(P) \equiv \operatorname*{argmax}_{\theta\in\Theta} \mathrm{E}_P[Y \log m_\theta(w) + (1-Y)\log(1 - m_\theta(w))].$$

Suppose for simplicity that $\Sigma^{1/2} = \mathrm{diag}(\sigma_D, \sigma_V, \sigma_1)$ corresponding to the random coefficient-shocks $\psi = (\psi_D, \psi_V, \psi_1)$. Hence, $\theta = (\mu, \sigma_D, \sigma_V, \sigma_1)$. The corresponding first order conditions with respect to μ are

$$\mathrm{E}_P\left[W(Y - m_\theta(W)) \frac{\int s_\theta(W,\psi)(1 - s_\theta(W,\psi)) dF(\psi)}{m_\theta(W,\psi)(1 - m_\theta(W,\psi))}\right] = 0.$$

[9] PMLEs as discussed here can be understood as limiting objects of frequentist pseudo maximum likelihood estimators as in White (1982), or as limiting Bayes estimators under a L_2-loss whenever the Bernstein-von Mises theorem applies.

The first order conditions with respect to $(\sigma_D, \sigma_V, \sigma_1)$ *are*

$$\begin{aligned}
\mathrm{E}_P\left[D(Y - m_\theta(W))\frac{\int \psi_D s_\theta(W,\psi)(1 - s_\theta(W,\psi))dF(\psi)}{m_\theta(W,\psi)(1 - m_\theta(W,\psi))}\right] &= 0 \\
\mathrm{E}_P\left[V(Y - m_\theta(W))\frac{\int \psi_V s_\theta(W,\psi)(1 - s_\theta(W,\psi))dF(\psi)}{m_\theta(W,\psi)(1 - m_\theta(W,\psi))}\right] &= 0 \\
\mathrm{E}_P\left[(Y - m_\theta(W))\frac{\int \psi_1 s_\theta(W,\psi)(1 - s_\theta(W,\psi))dF(\psi)}{m_\theta(W,\psi)(1 - m_\theta(W,\psi))}\right] &= 0.
\end{aligned}$$

It follows that the mixed Logit pseudo maximum likelihood estimand satisfies (2) *with* $Q = 2$.

Example 7 (Linear Index Natural Exponential Family Pseudo Maximum Likelihood)**.** *Suppose* $W \equiv (D, V, 1)^\top$ *satisfies Assumption 1. Consider the pseudo maximum likelihood estimand that maximizes the likelihood of Example 3, or equivalently,*

$$\theta^*(P) \equiv \operatorname*{argmax}_{\theta \in \Theta} \mathrm{E}_P[(W^\top\theta)Y - A(W^\top\theta)].$$

The first order conditions with respect to θ *are*

$$\mathrm{E}_P[W(Y - A'(W^\top\theta))] = 0.$$

By properties of natural exponential families, we have $A'(W^\top\theta) = m_\theta(W)$*. Hence, the linear index natural exponential family PMLE satisfies* (2) *with* $Q = 1$.

Finally, Assumption 3 places a rank condition on the functions $(f_\theta^q, h_\theta^q, l_\theta^q)_{q\in[Q]}$ referenced in Assumption 2.[10] The assumption ensures that these functions do not place all weight on fewer than two unique support points of D and fewer than three unique support points of V. Note also that Assumption 3 is directly satisfied for constant $(f_\theta^q, h_\theta^q, l_\theta^q)_{q\in[Q]}$ when $K \geq 3$.[11]

[10]Notation: For a positive integer $S \in \mathbb{N}^*$, let $[S] \equiv \mathbb{N}^*_{\leq S}$ denote the set of integers from 1 to S.

[11]For example, suppose that $Q = 1$ and $f_\theta, h_\theta, l_\theta$ are all equal to 1 as in Example 7. Then Assumption 3 is equivalent to non-singularity of

$$\begin{bmatrix} 1 & 1 & 1 \\ d_1 & d_J & d_J \\ v_1 & v_2 & v_3 \end{bmatrix}.$$

Assumption 3. $\forall P \in \mathcal{P}$, $\exists (q', q'', q''') \in [Q]^3$, $(d', d'') \in [J]^2$, *and* $(v', v'', v''') \in [K]^3$ *s.t.*

$$\begin{bmatrix} f^{q'}_{\theta^*(P)}(d', v') & f^{q'}_{\theta^*(P)}(d'', v'') & f^{q'}_{\theta^*(P)}(d'', v''') \\ d' h^{q''}_{\theta^*(P)}(d', v') & d'' h^{q''}_{\theta^*(P)}(d'', v'') & d'' h^{q''}_{\theta^*(P)}(d'', v''') \\ v' l^{q'''}_{\theta^*(P)}(d', v') & v'' l^{q'''}_{\theta^*(P)}(d'', v'') & v''' l^{q'''}_{\theta^*(P)}(d'', v''') \end{bmatrix}$$

is non-singular.

Proposition 1 shows that assumptions 1-3 characterize a class of estimands that are sign reversing. The result highlights potential problems with even basic qualitative conclusions of commonly used estimands under misspecification, and thus motivates the search for more robust estimands that I turn to next.

Proposition 1. *Let assumptions 1-3 hold. Then θ^* is sign reversing.*

Proof. See Appendix B.1. □

Remark 1. *Note that the Normal linear index PMLE defined is Example 7 is equivalent to a linear regression estimand. For this special case, Proposition 1 is similar to results in Blandhol et al. (2022) that highlight potential sign reversal of linear regression and linear two-stage least squares estimands without additional parametric distributional assumptions (cf. Proposition 7 in Blandhol et al., 2022). Proposition 1 complements these existing results for linear models with analyses of estimands for nonlinear models, including (mixed) Logit and Poisson estimands.*

4 Sign Preservation

This section defines sign preservation, a robustness property of estimands that guarantees qualitatively correct conclusions about the direction of the partial effects. I define two versions of sign preservation (weak and strong) to allow for better differentiation of partially and locally linear index model estimands in subsequent sections.

This follows immediately from Assumption 1 when $K \geq 3$.

Weak sign preservation is defined in Definition 5. A weakly sign preserving estimand θ^* associated with a researcher's model m_θ is guaranteed to result in model-implied partial effects that are of the same sign as the true partial effects *if* all true partial effects have the same direction (as defined in Definition 4).

Definition 4. *D is said to have P-**positive association** with Y given V if*

$$\mathrm{E}_P[Y|D=d',V] \overset{a.s.}{\geq} \mathrm{E}_P[Y|D=d,V], \quad \forall d' \geq d \in \mathcal{D}.$$

P-negative association is defined analogously. For distributions $P \in \mathcal{P}$, let $\mathcal{P}_+ \subset \mathcal{P}$ denote the subset of distributions under which D has P-positive association with Y given V. Analogously, let $\mathcal{P}_- \subset \mathcal{P}$ denote the subset of distributions under which D has P-negative association with Y given V.

Definition 5. *An estimand $\theta^* : \mathcal{P} \to \Theta$ corresponding to a model $m_\theta, \theta \in \Theta$, is said to be **sign preserving** for the expected change in Y due to D given V if,*

$$m_{\theta^*(P)}(d',V) \overset{a.s.}{\geq} m_{\theta^*(P)}(d,V), \quad \forall P \in \mathcal{P}_+,\ d' \geq d \in \mathcal{D},$$
$$m_{\theta^*(P)}(d',V) \overset{a.s.}{\leq} m_{\theta^*(P)}(d,V), \quad \forall P \in \mathcal{P}_-,\ d' \geq d \in \mathcal{D}.$$

To allow for correct conclusions about the sign of partial effects in settings where the true partial effects are not all of the same sign, I also define *strong sign preservation* in Definition 7. In contrast to weak sign preservation, strong sign preservation ensures that the model-implied partial effects are of the same sign as the true partial effects *for every value of V*. Both weak and strong sign preservation require monotone effects of D on Y given V.

Definition 6. *D is said to have P-**monotone association** with Y given V if, $\forall v \in \mathcal{V}$,*

$$\{\mathrm{E}_P[Y|D=d',V=v] \geq \mathrm{E}_P[Y|D=d,V=v], \quad \forall d' \geq d \in \mathcal{D}\}$$
$$\textit{or} \quad \{\mathrm{E}_P[Y|D=d',V=v] \leq \mathrm{E}_P[Y|D=d,V=v], \quad \forall d' \geq d \in \mathcal{D}\}.$$

For distributions $P \in \mathcal{P}$, let $\mathcal{P}_{\pm} \subset \mathcal{P}$ denote the subset of distributions under which D has P-monotone association with Y given V. Further, for $P \in \mathcal{P}_{\pm}$, let $\mathcal{V}_+$ and $\mathcal{V}_-$ denote values of V with positive and negative partial effects, respectively.

Definition 7. *An estimand $\theta^* : \mathcal{P} \to \Theta$ corresponding to a model $m_\theta, \theta \in \Theta$, is said to be* ***strongly sign preserving*** *for the expected change in Y due to D given V if, $\forall P \in \mathcal{P}_{\pm}$,*

$$m_{\theta^*(P)}(d', v) \geq m_{\theta^*(P)}(d', v), \quad \forall d' \geq d \in \mathcal{D}, v \in \mathcal{V}_+,$$
$$m_{\theta^*(P)}(d', v) \leq m_{\theta^*(P)}(d', v), \quad \forall d' \geq d \in \mathcal{D}, v \in \mathcal{V}_-.$$

5 Sufficient Conditions for Sign Preservation

This section provides sufficient conditions for estimands to satisfy weak and strong sign preservation. In addition to providing high-level sufficient conditions, I also provide sufficient conditions for linear, partially linear, and locally linear index natural exponential family pseudo maximum likelihood estimands (hereafter: linear, partially linear, locally linear index NEF-PMLEs, respectively) corresponding to the models discussed in Examples 3-5.

5.1 Sufficient Conditions for Weak Sign Preservation

A key intermediate result leveraged to characterize sign preserving estimands expresses a convex combination of model-implied partial effects as a convex combination of true partial effects.[12] This result is formally stated in Lemma 1 and is based on a single simple moment condition stated in Assumption 4.

Assumption 4. *The estimand $\theta^* : \mathcal{P} \to \Theta$ is unique and satisfies*

$$\mathrm{E}_P(D - \mathrm{E}_P[D|V])(Y - m_{\theta^*(P)}(D, V)) = 0, \quad \forall P \in \mathcal{P}. \tag{3}$$

[12]The coefficients $\omega_P(j, v)$ of Lemma 1 are weakly positive and integrate to a constant, but do not necessarily integrate to one. It is possible to normalize the coefficients by dividing with $\sum_{j=2}^{J} \mathrm{E}_P[\omega_P(j, V)]$.

Lemma 1. *Suppose that Assumption 4 holds, then* $\forall P \in \mathcal{P}$,

$$\begin{aligned} & \sum_{j=2}^{J} \mathrm{E}_P \left[\left(\mathrm{E}_P[Y|D = d_j, V] - \mathrm{E}_P[Y|D = d_{j-1}, V] \right) \omega_P(j, V) \right] \\ = \quad & \sum_{j=2}^{J} \mathrm{E}_P \left[\left(m_\theta(d_j, V) - m_\theta(d_{j-1}, V) \right) \omega_P(j, V) \right], \end{aligned} \tag{4}$$

where $\omega_P(j, V) \equiv \left(\mathrm{E}_P[D|D \geq d_j, V] - \mathrm{E}_P[D|D < d_j, V] \right) \Pr(D \geq d_j|V) \Pr(D < d_j|V)$.

Proof. See Appendix B.2. □

Remark 2. *Lemma 1 is closely connected to the literature on the weakly causal interpretation of linear regression, two-stage least squares, and difference-in-difference estimands (e.g., Yitzhaki, 1996; Blandhol et al., 2022; de Chaisemartin and Xavier d'Haultfoeuille, 2020, and other citations in the introduction). Unlike this literature on linear estimands, however, Lemma 1 does not express a single parameter as a convex combination of partial effects. While such a result is immediately implied for the versions of Normal PMLEs considered here when plugging in the expression in Example 3 for* m_θ*, it is not possible to obtain similar expressions in the more general class of NEF-PMLEs.*

Remark 3. *While I focus on discrete* D *throughout the main text, it is straightforward to extend the analysis to continuously distributed* D *under regularity conditions as in Silva and Winkelmann (2024) who analyze Poisson PMLEs. In particular,* (4) *would instead correspond to*

$$\mathrm{E}_P \left[\int_{\mathcal{D}} \frac{\partial}{\partial d} \mathrm{E}_P[Y|D = d, V] \big|_{d=j} \omega_P(j, V) dj \right] = \mathrm{E}_P \left[\int_{\mathcal{D}} \frac{\partial}{\partial d} m_\theta(d, V) \big|_{d=j} \omega_P(j, V) dj \right], \tag{5}$$

where $\omega_P(j, V)$ *is as in Lemma 1. Normalizing both sides of the equation by* $\mathrm{E}_P[\int_{\mathcal{D}} \omega_P(j, V) dj]$*, we obtain the result that a convex combination of model-implied partial derivatives equals a convex combination of the true partial effects. Note further that for normally distributed* $D|V$*, the coefficients* $\omega_P(\cdot, V)$ *are proportional to the conditional density of* $D|V$ *(Stoker, 1986; Wooldridge, 2010, Section 15.6).* (5) *thus implies that for normally distributed* $D|V$*,*

the model-implied average partial derivative is equal to the true average partial derivative for all unique estimands satisfying (3)*. As shown below, this includes linear index NEF-PMLEs under linearity of* $\mathrm{E}_P[D|V]$*, as well as all partially and locally linear index NEF-PMLEs.*

To help characterize robustness properties of estimands, it is also useful to highlight an important feature of the models highlighted in Examples 3-5:

Definition 8. *A class of scalar-valued functions* m_θ *on* $\mathcal{D} \times \mathcal{V}$ *indexed by* $\theta \in \Theta$ *is said to be* ***monotone*** *in* D *given* V *if,* $\forall \theta \in \Theta, P \in \mathcal{P}$,

$$\left\{ m_\theta(d', V) \overset{a.s.}{\geq} m_\theta(d, V),\ \forall d' \geq d \in \mathcal{D} \right\} \quad or \quad \left\{ m_\theta(d', V) \overset{a.s.}{\leq} m_\theta(d, V),\ \forall d' \geq d \in \mathcal{D} \right\}.$$

As shown by Proposition 2, simple sufficient conditions for an estimand to be weakly sign preserving are then given by assumptions 4 and 5.

Assumption 5. *The researcher's model* m_θ *is monotone in* D *given* V*.*

Proposition 2. *If assumptions 4-5 hold, then* θ^* *is weakly sign preserving for the expected change in* Y *due to* D *given* V*.*

Proof. See Appendix B.3. □

Proposition 2 shows that any estimand corresponding to a monotone model and that can be shown to satisfy the moment condition (3) is weakly sign preserving. Corollaries 1 and 2 directly link these assumptions to commonly used estimands. First, Example 7 shows that all identified linear index NEF-PMLEs satisfy Assumption 6. As stated in Corollary 1, this assumption allows for a sufficiency result in the special case that the conditional expectation of D given V is linear in V.

Assumption 6. *The estimand* $\theta^* : \mathcal{P} \to \Theta$ *is unique and satisfies*

$$\mathrm{E}_P D(Y - m_{\theta^*(P)}(D, V)) = 0, \quad \mathrm{E}_P V(Y - m_{\theta^*(P)}(D, V)) = 0, \qquad \forall P \in \mathcal{P}.$$

Corollary 1. *Suppose that assumptions 5-6 hold. If in addition* $\mathrm{E}_P[D|V]$ *is linear in* V*,* $\forall P \in \mathcal{P}$*, then* θ^* *is weakly sign preserving for the expected change in* Y *due to* D *given* V*.*

Proof. See Appendix B.4.1. □

Remark 4. *Given its equivalence to the linear regression estimand, the well-known robustness results of Yitzhaki (1996) and Angrist and Krueger (1999) directly imply sign preservation under linearity of* $\mathrm{E}_P[D|V]$ *for the Normal linear index PMLE. Corollary 1 extends these results to all other linear index NEF-PMLEs.*

A key caveat of Corollary 1 is that sign preservation of the estimand was only shown under linearity of $\mathrm{E}_P[D|V]$. In the absence of explicit functional form assumptions (that may be equally difficult to motivate in practice as functional form assumptions for $\mathrm{E}_P[Y|D,V]$), linearity is guaranteed only in the special cases when D is randomly assigned or when V is fully saturated. To also accommodate settings with complex V (e.g., a continuous covariate), Assumption 7 provides an alternative sufficient condition that guarantees weak sign preservation of estimands associated with monotone models also in the absence of any parametric restrictions on the joint distribution of D and V.

Assumption 7. *The estimand* $\theta^* : \mathcal{P} \to \Theta$ *is unique and satisfies*

$$\mathrm{E}_P D(Y - m_{\theta^*(P)}(D,V)) = 0, \quad \mathrm{E}_P[Y - m_{\theta^*(P)}(D,V)|V] \overset{a.s.}{=} 0, \qquad \forall P \in \mathcal{P}.$$

Examples 8 and 9 show that any identified partially or locally linear index NEF-PMLE satisfies Assumption 7. Corollary 2 then formally states that weak sign preservation is guaranteed for these estimands.

Example 8 (Partially Linear Index NEF-PMLE)**.** *Consider the pseudo maximum likelihood estimand that maximizes the likelihood of Example 4, or equivalently,*

$$\theta^*(P) \equiv \underset{\theta \in \Theta}{\operatorname{argmax}}\, \mathrm{E}_P[(D\alpha + b(V))Y - A(D\alpha + b(V))].$$

The first order condition with respect to α is

$$\mathrm{E}_P[D(Y - A'(D\alpha + b(V)))] = 0.$$

The first order conditions with respect to b are

$$\mathrm{E}_P[Y - A'(D\alpha + b(V))|V] \overset{a.s.}{=} 0.$$

By properties of natural exponential families, we have $A'(D\alpha + b(V)) = m_\theta(D, V)$. Hence, all identified partially linear index NEF-PMLEs satisfy Assumption 7.

Example 9 (Locally Linear Index NEF-PMLE)**.** *Consider the pseudo maximum likelihood estimand that maximizes the likelihood of Example 5, or equivalently,*

$$\theta^*(P) \equiv \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathrm{E}_P[(Da(V) + b(V))Y - A(Da(V) + b(V))].$$

The first order condition with respect to a are

$$\mathrm{E}_P[D(Y - A'(Da(V) + b(V)))|V] \overset{a.s.}{=} 0, \tag{6}$$

so that taking expectations over V we have $\mathrm{E}_P[D(Y - A'(Da(V) + b(V)))] = 0$. The first order conditions with respect to b are

$$\mathrm{E}_P[Y - A'(Da(V) + b(V))|V] \overset{a.s.}{=} 0.$$

By properties of natural exponential families, we have $A'(Da(V)+b(V)) = m_\theta(D, V)$. Hence, all identified locally linear index NEF-PMLEs satisfy Assumption 7.

Corollary 2. *Suppose that assumptions 5 and 7 hold. Then θ^* is weakly sign preserving for the expected change in Y due to D given V.*

Proof. See Appendix B.4.2. □

Remark 5. *Given the equivalence between linear regression estimations with saturated V and partially linear regression estimands (e.g., Robinson, 1988; Chernozhukov et al., 2018),*

Corollary 2 follows straightforwardly from results in Angrist and Krueger (1999) for the Normal partially linear index PMLE. The presented results here further generalize sign preservation to all partially and locally linear index NEF-PMLEs, including estimands targeted by the partially linear Logit estimators provided by Liu et al. (2021) or the locally linear Logit estimators applied in Farrell et al. (2021) and Dubé and Misra (2023).

5.2 Sufficient Conditions for Strong Sign Preservation

Example 9 and Corollary 2 showed that all identified locally linear index NEF-PMLEs are weakly sign preserving. Example 9 however also showed that these estimands satisfy the stronger moment conditions (6). Corollary 3 shows that these moment conditions are sufficient conditions for strong sign preservation. Hence, among the estimands considered in this note, locally linear index NEF-PMLEs allow for the strongest conclusions about the direction of partial effects under misspecification.

Assumption 8. *The estimand $\theta^* : \mathcal{P} \to \Theta$ is unique and satisfies*

$$\mathrm{E}_P[D(Y - m_{\theta^*(P)}(D,V))|V] \overset{a.s.}{=} 0, \quad \mathrm{E}_P[Y - m_{\theta^*(P)}(D,V)|V] \overset{a.s.}{=} 0, \qquad \forall P \in \mathcal{P}.$$

Corollary 3. *Suppose that assumptions 5 and 8 hold. Then θ^* is strongly sign preserving for the expected change in Y due to D given V.*

Proof. See Appendix B.5. □

6 Implications for Control Function Approaches

Strong and weak sign preservation as discussed thus far describe an estimands ability to correctly characterize features of the *reduced form* (i.e., descriptive) relationship between Y and D. In most policy settings analyzed in economics, however, researchers employing estimands such as those listed in Examples 3-5 are also interested in analyzing features

of *causal* effects of D on Y. This section shows that the previous analyses generalizes straightforwardly to the study of causal effects in a control function setting.

To fix ideas, consider a standard all-causes model that relates the outcome of interest Y to a scalar-valued variable of interest D and all other (partially unobserved) determinants of U (see, e.g., Heckman and Vytlacil, 2007). Following Imbens and Newey (2009), a control function is any (observed or identified) V such that D is as-good-as randomly assigned given V. The setting is stated in Assumption 9, where g is a (unknown) structural response function.

Assumption 9. *$Y = g(D, U)$ where $D \perp\!\!\!\perp U|V$.*

Lemma 2 states that under Assumption 9, the descriptive properties captured by sign preserving estimands are equivalent to properties of the causal effects of D on Y given V. It thus follows that for control function estimands, sign preservation captures robustness to correct inference on the direction of causal effects under misspecification.

Lemma 2. *Let Assumption 9 hold. Define $CATE_P^{d',d}(V) \equiv \mathrm{E}_P[g(d', U) - g(d, U)|V]$.*

(a) $CATE_P^{d',d}(V) \overset{a.s.}{\geq} 0$, $\forall d' \geq d \in \mathcal{D}$, iff D has P-positive association with Y given V.

(b) $CATE_P^{d',d}(V) \overset{a.s.}{\leq} 0$, $\forall d' \geq d \in \mathcal{D}$, iff D has P-negative association with Y given V.

(c) $\forall v \in \mathcal{V}$, $\left\{CATE_P^{d',d}(v) \geq 0, \forall d' \geq d \in \mathcal{D}\right\}$ or $\left\{CATE_P^{d',d}(v) \leq 0, \forall d' \geq d \in \mathcal{D}\right\}$, iff D has P-monotone association with Y given V.

Proof. See Appendix B.6. □

7 Conclusion

This note shows that under misspecification, a large class of estimands, including (mixed) Logit and Poisson PMLEs, are generally not informative about the direction of the true partial effects. To ensure robustness to misspecification of conclusions about the sign of the

true partial effects, I consider a minimal robustness property: Sign preservation. Intuitively, this property simply ensures that the model-implied partial effects are not all of the opposite sign as the true partial effects.

The note then develops simple sufficient conditions that guarantee sign preservation of an estimand. The results shows partially and locally linear index natural exponential family PMLEs are highly robust to misspecification of the outcome model without any parametric assumptions on the joint distribution of covariates. This suggests, in particular, that the recently proposed estimators of Liu et al. (2021) and special cases of the local maximum likelihood estimators of Athey et al. (2019) and Farrell et al. (2021) are highly attractive in many applied settings.

References

Andrews, I., Barahona, N., Gentzkow, M., Rambachan, A., and Shapiro, J. M. (2023). Causal interpretation of structural IV estimands. NBER working paper.

Andrews, I., Barnhard, H., and Carlson, J. (2024). True and pseudo-true parameters. Working paper.

Angrist, J. D. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica*, 66(2):249–288.

Angrist, J. D., Graddy, K., and Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies*, 67(3):499–527.

Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In Ashenfelter, O. C. and Card, D., editors, *Handbook of Labor Economics*, volume 3, chapter 23, pages 1277–1366. Elsevier, Amsterdam.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2):1148–1178.

Baker, A. C., Larcker, D. F., and Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Blandhol, C., Bonney, J., Mogstad, M., and Torgovitsky, A. (2022). When is TSLS actually LATE? NBER working paper.

Blundell, R. and Matzkin, R. L. (2014). Control functions in nonseparable simultaneous equations models. *Quantitative Economics*, 5(2):271–295.

Borusyak, K. and Hull, P. (2024). Negative weights are no concern in design-based specifications. *AEA Papers & Proceedings*, 114(1):597–600.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*, forthcoming.

Bugni, F. A., Canay, I. A., and McBride, S. (2023). Decomposition and interpretation of treatment effects in settings with delayed outcomes. *arXiv preprint arXiv:2302.11505*.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Chang, H., Narita, Y., and Saito, K. (2022). Approximating choice data by discrete choice models. *arXiv preprint arXiv:2205.01882*.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Compiani, G. (2022). Market counterfactuals and the specification of multiproduct demand: A nonparametric approach. *Quantitative Economics*, 13(2):545–591.

de Chaisemartin, C. and Xavier d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Dubé, J.-P. and Misra, S. (2023). Personalized pricing and consumer welfare. *Journal of Political Economy*, 131(1):131–189.

Farrell, M. H., Liang, T., and Misra, S. (2021). Deep learning for individual heterogeneity: An automatic inference framework. *arXiv preprint arXiv:2010.14694*.

Goldsmith-Pinkham, P., Hull, P., and Kolesár, M. (2024). Contamination bias in linear regressions. *arXiv preprint arXiv:2106.05024*.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In Heckman, J. J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6, chapter 70, pages 4779–4874. Elsevier, Amsterdam.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512.

Kim, K. and Petrin, A. (2019). Control function corrections for unobserved factors in differentiated product models. *Quantitative Marketing and Economics*, forthcoming.

Leung, M. P. (2024). Causal interpretation of estimands defined by exposure mappings. *arXiv preprint arXiv:2403.08183*.

Li, K.-C. and Duan, N. (1989). Regression analysis under link violation. *Annals of Statistics*, 17(3):1009–1052.

Liu, M., Zhang, Y., and Zhou, D. (2021). Double/debiased machine learning for logistic partially linear model. *The Econometrics Journal*, 24(3):559–588.

Lu, J. and Saito, K. (2022). Mixed logit and pure characteristics models. Working paper.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, 47(1):3–13.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954.

Ruud, P. A. (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. *Econometrica*, 51(1):225–228.

Sävje, F. (2024). Rejoinder: Causal inference with misspecified exposure mappings: separating definitions and assumptions. *Biometrika*, 111(1):25–29.

Silva, J. S. and Winkelmann, R. (2024). Misspecified exponential regressions: Estimation, interpretation, and average marginal effects. *Review of Economics and Statistics*, forthcoming.

Słoczyński, T. (2022). Interpreting ols estimands when treatment effects are heterogeneous: Smaller groups get larger weights. *Review of Economics and Statistics*, 104(3):501–509.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica*, 54(6):1461–1481.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Tebaldi, P., Torgovitsky, A., and Yang, H. (2023). Nonparametric estimates of demand in the california health insurance exchange. *Econometrica*, 91(1):107–146.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445.

Yitzhaki, S. (1996). On using linear regressions in welfare economics. *Journal of Business & Economic Statistics*, 14(4):478–486.

A Sign Preservation in Multivariate-Outcome Models

This section extends the analysis of sign preserving estimands to multivariate outcomes $Y = (Y_1, \ldots, Y_S)^\top$ and multivariate variables of interest $D = (D_1, \ldots, D_S)^\top$, for some $S \in \mathbb{N}$. Analogous the analysis of the main text, I assume that the support of Y is $\mathcal{Y} \subset [\underline{y}, \overline{y}]^S$, the support of D is $\mathcal{D} \equiv \{d_1, \ldots, d_J\}^S$, and the support of the covariates V is $\mathcal{V}$. Further, the support of (D, V) is $\mathcal{D} \times \mathcal{V}$.

The results have implications, in particular, for multinomial Logit estimands where Y denotes the choice among S alternatives (and one outside option) and D denotes the corresponding choice features. See, in particular, Examples 10-11. Throughout, for a random vector $D = (D_1, \ldots, D_S)^\top$, let D_s and D_{-s} denote the sth element of D and D without its sth element, respectively. Further, the researcher's model $m_\theta : \mathcal{D} \times \mathcal{V} \to \mathcal{Y}$ is a S-dimensional vector with elements $m_{s,\theta} \in L_{2,\mathcal{P}} : \mathcal{D}_s \times \mathcal{D}_{-s} \times \mathcal{V} \to \mathcal{Y}$.

Example 10 (Multivariate Linear Index Natural Exponential Family)**.** *For $\theta = (\alpha, \beta)$, define the S-dimensional vector $\eta_\theta(d, v) \equiv (d_s\alpha + v^\top\beta)_{s=1}^S$. Consider the conditional (pseudo) likelihood of y given d and v defined by*

$$L(y, d, v; \theta) \propto \exp\left(\eta_\theta(d, v)^\top y - A(\eta_\theta(d, v))\right),$$

where A is a known scalar-valued function. Then, by properties of natural exponential families, the model-implied mean of Y_s given $D = d$ and $V = v$ is $m_{s,\theta}(d, v) = A^s(\eta_\theta(d, v))$, where A^s denotes the partial derivative of A with respect to its sth argument. A key example is the multinomial Logit model where

$$A(\eta_\theta(d, v)) = \log(1 + \sum_{s=1}^{S} \exp(d_s\alpha + v^\top\beta)), \quad A^s(\eta_\theta(d, v)) = \frac{\exp(d_s\alpha + v^\top\beta)}{1 + \sum_{l=1}^{S} \exp(d_l\alpha + v^\top\beta)}.$$

Example 11 (Multivariate Partially Linear Index Natural Exponential Family)**.** *For $\theta = (\alpha, b_1, \ldots, b_S)$, with $b_s : \mathcal{V} \to \mathbb{R}, \forall s \in [S]$, define the S-dimensional vector $\eta_\theta(d, v) \equiv (d_s\alpha +$*

$b_s(v))_{s=1}^S$. *Consider the conditional (pseudo) likelihood of* y *given* d *and* v *defined by*

$$L(y, d, v; \theta) \propto \exp\left(\eta_\theta(d, v)^\top y - A(\eta_\theta(d, v))\right),$$

where A *is a known scalar-valued function. The model-implied mean of* Y_s *given* $D = d$ *and* $V = v$ *is* $m_{s,\theta}(d, v) = A^s(\eta_\theta(d, v))$. *An example is the partially linear multinomial Logit model where*

$$A(\eta_\theta(d, v)) = \log(1 + \sum_{s=1}^{S} \exp(d_s\alpha + b_s(v))), \quad A^s(\eta_\theta(d, v)) = \frac{\exp(d_s\alpha + b_s(v))}{1 + \sum_{l=1}^{S} \exp(d_l\alpha + b_l(v))}.$$

Section A.1 defines and discusses weak *diagonal* sign preservation. Section A.2 then states sufficient conditions analogous to those in Section 5.1 of the main text. I focus on extending the results for weak sign preservation for brevity.

A.1 Weak Diagonal Sign Preservation

Definition 9 defines *diagonal* positive (or negative) association to allow for characterization subsets of the distributions $\mathcal{P}$ with weakly positive (or negative) partial effects. Diagonal here refers to the sth element of Y being positively (or negatively) associated with the sth element of D. In the discrete choice setting, for example, Y_s would correspond to the choice of the sth good and D_s could be a marketing variable of the sth good. Positive diagonal association would then say that the choice probability of good sth is weakly increasing in the marketing variable of the sth good *holding all other marketing variables* D_{-s} *and covariates* V *fixed*. In analogy to Definition 5, Definition 10 then formulates *diagonal* sign preservation.

Definition 9. D *is said to have* P***-positive diagonal association*** *with* Y *given given* V *if, for every* $s \in [S]$, D_s *has* P*-positive association with* Y_s *given* D_{-s} *and* V. P*-negative diagonal association is defined analogously. For distributions* $P \in \mathcal{P}$, *let* $\mathcal{P}_+^S \subset \mathcal{P}$ *denote the subset of distributions under which* D *has* P*-positive diagonal association with* Y *given* V. *Analogously, let* $\mathcal{P}_-^S \subset \mathcal{P}$ *denote the subset of distributions under which* D *has* P*-negative diagonal association with* Y *given* V.

Definition 10. *An estimand $\theta^* : \mathcal{P} \to \Theta$ is said to be* ***diagonal sign preserving*** *for the expected change in Y due to D given V if*

$$m_{s,\theta^*(P)}(d', D_{-s}, X) \overset{a.s.}{\geq} m_{s,\theta^*(P)}(d, D_{-s}, X), \quad \forall P \in \mathcal{P}_+^S,\ d' \geq d \in \mathcal{D}_s,\ s \in [S],$$
$$m_{s,\theta^*(P)}(d', D_{-s}, X) \overset{a.s.}{\leq} m_{s,\theta^*(P)}(d, D_{-s}, X), \quad \forall P \in \mathcal{P}_-^S,\ d' \geq d \in \mathcal{D}_s,\ s \in [S].$$

A.2 Sufficient Conditions for Weak Diagonal Sign Preservation

This section states sufficient conditions for weak diagonal sign preservation. As in the main text, I begin with a high-level condition (Assumption 10) that implies that the model-implied diagonal partial effects are a positively weighted average of the true partial effects. This is formally stated in Lemma 3.

Assumption 10. *The estimand $\theta^* : \mathcal{P} \to \Theta$ is unique and satisfies*

$$\sum_{s=1}^{S} \mathrm{E}_P(D_s - \mathrm{E}_P[D_s|D_{-s}, V])(Y_s - m_{s,\theta^*(P)}(D, V)) = 0, \quad \forall P \in \mathcal{P}. \tag{7}$$

Lemma 3. *Suppose that Assumption 10 holds, then $\forall P \in \mathcal{P}$,*

$$\begin{aligned} &\sum_{s=1}^{S}\sum_{j=2}^{J} \mathrm{E}_P\left[\left(\mathrm{E}_P[Y_s|D_s = d_j, D_{-s}, V] - \mathrm{E}_P[Y_s|D_s = d_{j-1}, D_{-s}, V]\right)\omega_{s,P}(j, D_{-s}, V)\right] \\ = &\sum_{s=1}^{S}\sum_{j=2}^{J} \mathrm{E}_P\left[\left(m_{s,\theta}(d_j, D_{-s}, V) - m_{s,\theta}(d_{j-1}, D_{-s}, V)\right)\omega_{s,P}(j, D_{-s}, V)\right], \end{aligned} \tag{8}$$

where

$$\begin{aligned} \omega_{s,P}(j, D_{-s}, V) \equiv &\left(\mathrm{E}_P[D_s|D_s \geq d_j, D_{-s}, V] - \mathrm{E}_P[D_s|D_s < d_j, D_{-s}, V]\right) \\ &\times \Pr(D_s \geq d_j|D_{-s}, V)\Pr(D_s < d_j|D_{-s}, V). \end{aligned}$$

Proof. Fix an arbitrary $P \in \mathcal{P}$.

Note that for $\mathcal{D}_s = \{d_1, \ldots, d_J\}$, we can write

$$\mathrm{E}_P[Y_s|D, V] = \mathrm{E}_P[Y_s|D_s = d_1, D_{-s}, V] + \sum_{j=2}^{D_s} \Delta_{s,P}^j(D_{-s}, V),$$

where

$$\Delta^j_{s,P}(D_{-s}, V) \equiv \mathrm{E}_P[Y_s|D_s = d_j, D_{-s}, V] - \mathrm{E}_P[Y_s|D_s = d_{j-1}, D_{-s}, V].$$

Similarly, $\forall \theta \in \Theta$,

$$m_{s,\theta}(D, V) = m_{s,\theta}(d_1, D_{-s}, V) + \sum_{j=2}^{D} \Delta^j_{s,\theta}(D_{-s}, V),$$

where

$$\Delta^j_{s,\theta}(D, V) \equiv m_{s,\theta}(d_j, D_{-s}, V) - m_{s,\theta}(d_{j-1}, D_{-s}, V).$$

Let $\theta = \theta^*(P)$. By Assumption 10, we then have

$$\begin{aligned}
0 &= \sum_{s=1}^{S} \mathrm{E}_P(D_s - \mathrm{E}_P[D_s|D_{-s}, V])(Y_s - m_{s,\theta^*(P)}(D, V)) \\
&= \sum_{s=1}^{S} \mathrm{E}_P(D_s - \mathrm{E}_P[D_s|D_{-s}, V]) \left(\mathrm{E}_P[Y_s|D_s = d_1, D_{-s}, V] - m_{s,\theta^*(P)}(d_1, D_{-s}, V)\right) \\
&\quad + \sum_{s=1}^{S} \mathrm{E}_P(D_s - \mathrm{E}_P[D_s|D_{-s}, V]) \sum_{j=2}^{D_s} \left(\Delta^j_{s,P}(D_{-s}, V) - \Delta^j_{s,\theta^*(P)}(D_{-s}, V)\right) \\
&\overset{[1]}{=} \sum_{s=1}^{S} \mathrm{E}_P\mathrm{E}_P[(D_s - \mathrm{E}_P[D_s|D_{-s}, V])|D_{-s}, V] \left(\mathrm{E}_P[Y_s|D_s = d_1, D_{-s}, V] - m_{s,\theta^*(P)}(d_1, D_{-s}, V)\right) \\
&\quad + \sum_{s=1}^{S} \sum_{j=2}^{J} \mathrm{E}_P(D_s - \mathrm{E}_P[D_s|D_{-s}, V]) \mathbb{1}\{d_j \leq D_s\} \left(\Delta^j_{s,P}(D_{-s}, V) - \Delta^j_{s,\theta^*(P)}(D_{-s}, V)\right) \\
&\overset{[2]}{=} \sum_{s=1}^{S} \sum_{j=2}^{J} \mathrm{E}_P\mathrm{E}_P[(D_s - \mathrm{E}_P[D_s|D_{-s}, V])\mathbb{1}\{d_j \leq D_s\}|D_{-s}, V] \left(\Delta^j_{s,P}(D_{-s}, V) - \Delta^j_{s,\theta^*(P)}(D_{-s}, V)\right),
\end{aligned}$$

where [1] and [2] follow from the law of iterated expectations. As a consequence,

$$\sum_{s=1}^{S} \sum_{j=2}^{J} \mathrm{E}_P \Delta^j_{s,P}(D_{-s}, V) \omega_{s,P}(j, D_{-s}, V) = \sum_{s=1}^{S} \sum_{j=2}^{J} \mathrm{E}_P \Delta^j_{s,\theta^*(P)}(D_{-s}, V) \omega_{s,P}(j, D_{-s}, V),$$

where, $\forall j \in [J] \setminus 1$,

$$\begin{aligned}\omega_{s,P}(j, D_{-s}, V) &\equiv \mathrm{E}_P[(D_s - \mathrm{E}_P[D_s|D_{-s}, V])\mathbb{1}\{d_j \leq D_s\}|D_{-s}, V] \\ &= (\mathrm{E}_P[D_s|D_s \geq d_j, D_{-s}, V] - \mathrm{E}_P[D_s|D_s < d_j, D_{-s}, V]) \\ &\quad \times \Pr(D_s \geq d_j|D_{-s}, V) \Pr(D_s < d_j|D_{-s}, V) \overset{a.s.}{\geq} 0.\end{aligned}$$

Since the choice of $P \in \mathcal{P}$ was arbitrary, this concludes the proof. □

As shown by Proposition 3, Assumption 10 is sufficient under additional restrictions on the researcher's model as stated in Assumption 11. Note, in particular, that commonly used multinomial Logit models as in Examples 10 and 11 satisfy Assumption 11.

Definition 11. *A class of scalar-valued functions m_θ on $\mathcal{D} \times \mathcal{V}$ indexed by $\theta \in \Theta$ is said to be **diagonally monotone** in D given V if, $\forall \theta \in \Theta, P \in \mathcal{P}, s \in [S]$,*

$$\begin{aligned}&\left\{m_{s,\theta}(d', D_{-s}, V) \overset{a.s.}{\geq} m_{s,\theta}(d, D_{-s}, V),\ \forall d' \geq d \in \mathcal{D}_s\right\} \\ \text{or} \quad &\left\{m_{s,\theta}(d', D_{-s}, V) \overset{a.s.}{\leq} m_{s,\theta}(d, D_{-s}, V),\ \forall d' \geq d \in \mathcal{D}_s\right\}.\end{aligned}$$

Assumption 11. *The researcher's model m_θ is diagonally monotone in D given V.*

Proposition 3. *If assumptions 10 and 11 hold, then θ^* is weakly diagonal sign preserving for the expected change in Y due to D given V.*

Proof. Note that Lemma 3 applies by Assumption 10. By Assumption 11, $m_{\theta(P)}$ is diagonally monotone in D given V, so that the right hand side of (8) is either a positively weighted sum of weakly negative terms or a positively weighted sum of weakly positive terms. Hence, the sign of the right hand side (weakly) determines the sign of all terms in the sum.

Now, if $P \in \mathcal{P}_+^S$, then Lemma 3 implies

$$0 \leq \sum_{j=2}^{J} \mathrm{E}_P\left[(m_{s,\theta}(d_j, D_{-s}, V) - m_{s,\theta}(d_{j-1}, D_{-s}, V))\, \omega_{s,P}(j, D_{-s}, V)\right], \quad \forall s \in [S].$$

Hence, since $\omega_{s,P}(j, D_{-s}, V) \overset{a.s.}{\geq} 0$, diagonal monotonicity implies

$$m_{s,\theta^*(P)}(d', D_{-s}, V) \overset{a.s.}{\geq} m_{s,\theta^*(P)}(d', D_{-s}, V), \quad \forall d' \geq d \in \mathcal{D}_s, s \in [S].$$

Analogous arguments imply that if $P \in \mathcal{P}_-^S$, then

$$m_{s,\theta^*(P)}(d', D_{-s}, V) \overset{a.s.}{\leq} m_{s,\theta^*(P)}(d', D_{-s}, V), \quad \forall d' \geq d \in \mathcal{D}_s, s \in [S].$$

Since the choice of $P \in \mathcal{P}$ was arbitrary, this concludes the proof. □

To show that the multinomial Logit linear index PMLE can also satisfy Assumption 10, I state a simple set of alternative moment conditions in (12). See Example 12. Corollary 5 then confirms diagonal sign preservation of the corresponding estimand under identification and *three* additional conditions: 1) The feature D_s is mean-independent of the other features D_{-s} given V, 2) the conditional expectation of D_s given V is linear, and 3) $\mathrm{E}_P[D_s|V]$ is the same for all choices $s \in [S]$. All of these conditions are strong. In industrial organization and quantitative marketing applications where multinomial Logit estimands are applied to study, for example, own-price elasticities, it is not commonly assumed that the vector of prices is mutually mean-independent. Linearity of $E[D_s|V]$ is guaranteed only for saturated V or randomly assigned D. The assumption of equality of $E[D_s|V]$, perhaps, seems most difficult to motivate, however, since it does not even permit differences in the unconditional mean of D_s for different goods $s \in [S]$.

Assumption 12. *The estimand $\theta^* : \mathcal{P} \to \Theta$ is unique and satisfies*

$$\sum_{s=1}^{S} \mathrm{E}_P D_s(Y_s - m_{s,\theta^*(P)}(D, V)) = 0, \quad \sum_{s=1}^{S} \mathrm{E}_P V(Y_s - m_{s,\theta^*(P)}(D, V)) = 0, \qquad \forall P \in \mathcal{P}. \quad (9)$$

Example 12 (Multivariate Linear Index NEF-PMLE)**.** *Consider the pseudo maximum likelihood estimand that maximizes the likelihood of Example 10, or equivalently,*

$$\theta^*(P) \equiv \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathrm{E}_P[\eta_\theta(D, V)^\top Y - A(\eta_\theta(D, V))],$$

where $\theta = (\alpha, \beta)$. The first order conditions with respect to α is

$$\sum_{s=1} \mathrm{E}_P[D_s(Y_s - A^s(\eta_\theta(D, V)))] = 0,$$

where A^s denotes the partial derivative of A with respect to its sth argument, and similarly for the first order conditions with respect to β, we have

$$\sum_{s=1} \mathrm{E}_P[V(Y_s - A^s(\eta_\theta(D, V)))] = 0.$$

By properties of natural exponential families, we have $A^s(\eta_\theta(D, V)) = m_{s,\theta}(D, V)$. Hence, the multivariate linear index NEF-PMLE satisfies (9)*.*

Corollary 4. *Suppose that assumptions 11-12 hold. If in addition $\forall P \in \mathcal{P}$, $\exists \gamma_P \in \mathbb{R}^{|\mathcal{V}|}$: $\mathrm{E}_P[D_s|D_{-s}, V] = V^\top \gamma_P, \forall s \in [S]$, then θ^* is weakly diagonal sign preserving for the expected change in Y due to D given V.*

Proof. Fix an arbitrary $P \in \mathcal{P}$.

By Assumption 12, we have $\forall \gamma \in \mathbb{R}^{|\mathcal{V}|}$ that

$$\sum_{s=1} \mathrm{E}_P V^\top \gamma(Y_s - m_{s,\theta^*(P)}(D, V)) = 0.$$

Take γ_P satisfying $\mathrm{E}_P[D_s|D_{-s}, V] = V^\top \gamma_P, \forall s \in [S]$, which was assumed to exist in the statement of the corollary. Then, by Assumption 12, we have

$$\begin{aligned}
0 &= \sum_{s=1} \mathrm{E}_P D_s(Y_s - m_{s,\theta^*(P)}(D, V)) - \mathrm{E}_P V^\top \gamma_P(Y_s - m_{s,\theta^*(P)}(D, V)) \\
&= \sum_{s=1} \mathrm{E}_P(D_s - V^\top \gamma_P)(Y_s - m_{s,\theta^*(P)}(D, V)) \\
&= \sum_{s=1} \mathrm{E}_P(D_s - \mathrm{E}_P[D_s|D_{-s}, V])(Y_s - m_{s,\theta^*(P)}(D, V)).
\end{aligned}$$

Since the choice of $P \in \mathcal{P}$ was arbitrary, this implies that Assumption 10 is satisfied. Applying Proposition 3 completes the proof. □

To avoid two of the three strong restrictions on $\mathcal{P}$ that suffice for diagonal sign preservation of the multinomial Logit linear PMLE of Example 10, I provide alternative sufficient conditions in Assumption 13. As shown in Example 13, these weaker conditions are satisfied by the multinomial Logit *partially linear index* PMLE. Corollary 5 then shows that diagonal sign preservation can be guaranteed without parametric restrictions on $E[D_s|V]$. In contrast to the results of Corollary 2 of the main text, however, mean-independence of D_s and D_{-s} given V is still used in the statement of Corollary 5. Exploring alternative multinomial Logit estimands that can be guaranteed sign preserving robustness properties without these restrictions on the joint distribution of (D, V) might thus be an interesting area for future research.[13]

Assumption 13. *The estimand $\theta^* : \mathcal{P} \to \Theta$ is unique and satisfies*

$$\sum_{s=1}^{S} \mathrm{E}_P D_s(Y_s - m_{s,\theta^*(P)}(D,V)) = 0, \quad \mathrm{E}_P[Y_s - m_{s,\theta^*(P)}(D,V)|V] \overset{a.s.}{=} 0, \qquad \forall P \in \mathcal{P}. \quad (10)$$

Example 13 (Multivariate Partially Linear Index NEF-PMLE)**.** *Consider the pseudo maximum likelihood estimand that maximizes the likelihood of Example 11, or equivalently,*

$$\theta^*(P) \equiv \underset{\theta\in\Theta}{\operatorname{argmax}}\, \mathrm{E}_P[\eta_\theta(D,V)^\top Y - A(\eta_\theta(D,V))],$$

where $\theta = (\alpha, b_1, \ldots, b_S)$, with $b_s : \mathcal{V} \to \mathbb{R}, \forall s \in [S]$. The first order conditions with respect to α is

$$\sum_{s=1} \mathrm{E}_P[D_s(Y_s - A^s(\eta_\theta(D,V)))] = 0.$$

For every $s \in [S]$, the first order conditions with respect to b_s are

$$\mathrm{E}_P[Y_s - A^s(\eta_\theta(D,V))|V] \overset{a.s.}{=} 0.$$

[13]Issues arising from mutually dependent D are closely connected to contamination bias arising in linear regression (see, e.g., Goldsmith-Pinkham et al., 2024).

By properties of natural exponential families, we have $A^s(\eta_\theta(D,V)) = m_{s,\theta}(D,V)$. *Hence, the multivariate linear index NEF-PMLE satisfies* (10).

Corollary 5. *Suppose that assumptions 11-13 hold. If in addition* $\mathrm{E}_P[D_s|D_{-s},V] \overset{a.s.}{=} \mathrm{E}_P[D_s|V], \forall s \in [S]$, *then* θ^* *is weakly diagonal sign preserving for the expected change in* Y *due to* D *given* V.

Proof. Fix an arbitrary $P \in \mathcal{P}$.

By Assumption 13, we have $\forall (h_s)_{s=1}^S$ where $h_s : \mathcal{V} \to \mathbb{R}$ that

$$\sum_{s=1} \mathrm{E}_P h_s(V)(Y_s - m_{s,\theta^*(P)}(D,V)) = 0.$$

Take $(h_{s,P})_{s=1}^S$ where $h_{s,P}(V) \equiv \mathrm{E}_P[D_s|V], \forall s \in [S]$. Then, by Assumption 13, we have

$$\begin{aligned}
0 &= \sum_{s=1} \mathrm{E}_P D_s(Y_s - m_{s,\theta^*(P)}(D,V)) - \sum_{s=1} \mathrm{E}_P h_{s,P}(V)(Y_s - m_{s,\theta^*(P)}(D,V)) \\
&= \sum_{s=1} \mathrm{E}_P(D_s - h_{s,P}(V))(Y_s - m_{s,\theta^*(P)}(D,V)) \\
&= \sum_{s=1} \mathrm{E}_P(D_s - \mathrm{E}_P[D_s|V])(Y_s - m_{s,\theta^*(P)}(D,V)) \\
&= \sum_{s=1} \mathrm{E}_P(D_s - \mathrm{E}_P[D_s|D_{-s},V])(Y_s - m_{s,\theta^*(P)}(D,V)),
\end{aligned}$$

where the final equality follows from $\mathrm{E}_P[D_s|D_{-s},V] \overset{a.s.}{=} \mathrm{E}_P[D_s|V], \forall s \in [S]$ which was assumed to exist in the statement of the corollary.

Since the choice of $P \in \mathcal{P}$ was arbitrary, this implies that Assumption 10 is satisfied. Applying Proposition 3 completes the proof. □

B Proofs of the Main Results

B.1 Proof of Proposition 1

Take any $\theta \in \theta^*(\mathcal{P})$. Let $\mathcal{P}_\theta \equiv \{P \in \mathcal{P} : \theta^*(P) = \theta\}$. The goal is to show that $\mathcal{P}_\theta \bigcap \mathcal{P}_{++} \neq \emptyset$ *and* $\mathcal{P}_\theta \bigcap \mathcal{P}_{--} \neq \emptyset$. The proof constructs a distribution P with strictly *negative*

partial effects with respect to D given V such that (a) $\theta^*(P) = \theta$, (b) $\mathrm{E}_P[Y|D,V] \in \mathrm{int}(\mathcal{Y}) = (\underline{y}, \overline{y})$, (c) the support of (D,V) is $\mathcal{D} \times \mathcal{V}$. Given this P, it follows that $P \in \mathcal{P}_\theta \bigcap \mathcal{P}_{--} \neq \emptyset$. A distribution with strictly positive partial effects satisfying (a)-(c) can be found using analogous arguments. I omit the full derivation here to avoid repetition.

B.1.1 Setup

Let $A_k^j \equiv m_\theta(d_j, v_k), \forall (j,k) \in [J] \times [K]$ and note that $A_k^j \in (\underline{y}, \overline{y})$ by Assumption 2. Similarly, let $f_k^{j,q} = f_\theta^q(d_j, v_k)$, $h_k^{j,q} = h_\theta^q(d_j, v_k)$, and $l_k^{j,q} = l_\theta^q(d_j, v_k)$. Let (d', d'') be those elements of $\mathcal{D}$ that satisfy Assumption 3, and let (J', J'') be the corresponding indices, respectively. Further, let (v', v'', v''') be those elements of $\mathcal{V}$ that satisfy Assumption 3, and let (K', K'', K''') be the corresponding indices, respectively.

To define a distribution $P \in \mathcal{P}_{--}$, I choose values for the corresponding conditional expectation functions $\mathrm{E}_P[Y|D,V]$ and the corresponding joint mass function $\Pr_P(D = d, V = v)$. I define P as a perturbation of a distribution $\tilde{P} \in \mathcal{P}$ (not necessarily with strictly negative/positive partial effects) where $\tilde{g}_k^j \equiv \mathrm{E}_{\tilde{P}}[Y|D = d_j, V = v_k] = A_k^j$ and where $p_k^j \equiv \Pr_{\tilde{P}}(D = d_j, V = v_k)$ are arbitrary strictly positive probabilities. By Assumption 2, $\tilde{P}$ satisfies (a)-(b), and by Assumption 1, $\tilde{P}$ satisfies (c).

B.1.2 Construction of P with strictly negative partial effects

The perturbations for $\tilde{P}$ are defined by shifting the values of the conditional expectation function. In particular, defining $g_k^j \equiv \mathrm{E}_P[Y|D = d_k, V = v_k]$, consider

$$g_k^j = \begin{cases} \tilde{g}_k^j + b_k^j - r' + \gamma j \epsilon = A_k^{J'} - r' + \gamma j \epsilon & \text{if } k = K' \\ \tilde{g}_k^j + c_k^j - r'' + \gamma j \epsilon = A_k^{J''} - r'' + \gamma j \epsilon & \text{if } k = K'' \\ \tilde{g}_k^j + c_k^j - r''' + \gamma j \epsilon = A_k^{J''} - r''' + \gamma j \epsilon & \text{if } k = K''' \\ \tilde{g}_k^j + c_k^j + \gamma j \epsilon = A_k^{J''} + \gamma j \epsilon & \text{if } k \in [K] \setminus \{K', K'', K'''\}, \end{cases} \tag{11}$$

where $(r', r'', r''', \gamma, \epsilon)$ are numbers that I choose, and $c_k^j \equiv A_k^{J''} - A_k^j, b_k^j \equiv A_k^{J'} - A_k^j, \forall (j,k) \in [J] \times [K]$, are given (since θ is fixed). Note for $\gamma = -1$, we have

$$g_k^{j+1} - g_k^j = -\epsilon, \quad \forall j \geq 2, k \in [K].$$

Hence, for arbitrary choice of $(r', r'', r''') \in \mathbb{R}^3$ and arbitrary positive $\epsilon > 0$, a joint distribution with conditional expectation values g_k^j would correspond to strictly negative partial effects when $\gamma = -1$.[14] For the remainder of the proof, I let $\gamma = -1$.

B.1.3 Construction of P satisfying $\theta^*(P) = \theta$

In order for P to satisfy $\theta^*(P) = \theta$, P must satisfy the moment conditions given by Assumption 2. Note that $\tilde{P} \in \mathcal{P}$ constructed previously trivially satisfies the moment conditions — that is, $\forall q \in [Q]$,

$$0 = \sum_{jk} f_k^{j,q} (\tilde{g}_k^j - A_k^j) p_k^j, \quad 0 = \sum_{jk} d_j h_k^{j,q} (\tilde{g}_k^j - A_k^j) p_k^j, \quad 0 = \sum_{jk} x_k l_k^{j,q} (\tilde{g}_k^j - A_k^j) p_k^j, \tag{12}$$

Solving (11) for $\tilde{g}_k^j$ results in

$$\tilde{g}_k^j = \begin{cases} g_k^j - b_k^j + r' + j\epsilon & \text{if } k = K' \\ g_k^j - c_k^j + r'' + j\epsilon & \text{if } k = K'' \\ g_k^j - c_k^j + r''' + j\epsilon & \text{if } k = K''' \\ g_k^j - c_k^j + j\epsilon & \text{if } k \in [K] \setminus \{K', K'', K'''\}. \end{cases} \tag{13}$$

[14] Similarly, the distribution would have strictly positive partial effects when $\gamma = 1$.

Substituting into (12) we have

$$
\begin{aligned}
0 &= \sum_{jk} f_k^{j,q}(g_k^j - A_k^j)p_k^j \\
&\quad + \sum_j f_{K'}^{j,q}(b_{K'}^j + r' + j\epsilon])p_{K'}^j + \sum_j f_{K''}^{j,q}(-c_{K''}^j + r'' + j\epsilon])p_{K''}^j \\
&\quad + \sum_j f_{K'''}^{j,q}(-c_{K'''}^j + r''' + j\epsilon])p_{K'''}^j + \sum_{j,k\notin\{K',K'',K'''\}} q_k^{j,q}(-c_k^j + j\epsilon)p_k^j \\
0 &= \sum_{jk} d_j h_k^{j,q}(g_k^j - A_k^j)p_k^j \\
&\quad + \sum_j d_j h_{K'}^{j,q}(-b_{K'}^j + r' + j\epsilon])p_{K'}^j + \sum_j d_j h_{K''}^{j,q}(-c_{K''}^j + r'' + j\epsilon])p_{K''}^j \\
0 &= \sum_{jk} v_k l_k^{j,q}(g_k^j - A_k^j)p_k^j \\
&\quad + v_{K'}\sum_j l_{K'}^{j,q}(-b_{K'}^j + r' + j\epsilon])p_{K'}^j + v_{K''}\sum_j l_{K''}^{j,q}(-c_{K''}^j + r'' + j\epsilon])p_{K''}^j \\
&\quad + v_{K'''}\sum_j l_{K'''}^{j,q}(-c_{K'''}^j + r''' + j\epsilon])p_{K'''}^j + \sum_{j,k\notin\{K',K'',K'''\}} v_k l_k^{j,q}(-c_k^j + j\epsilon)p_k^j,
\end{aligned}
\tag{14}
$$

$\forall q \in [Q]$. Clearly, $\theta^*(P) = \theta$ if, $\forall q \in [Q]$,

$$
\begin{aligned}
0 &= \sum_j f_{K'}^{j,q}(b_{K'}^j + r' + j\epsilon])p_{K'}^j + \sum_j f_{K''}^{j,q}(-c_{K''}^j + r'' + j\epsilon])p_{K''}^j \\
&\quad + \sum_j f_{K'''}^{j,q}(-c_{K'''}^j + r''' + j\epsilon])p_{K'''}^j + \sum_{j,k\notin\{K',K'',K'''\}} q_k^{j,q}(-c_k^j + j\epsilon)p_k^j \\
0 &= \sum_j d_j h_{K'}^{j,q}(-b_{K'}^j + r' + j\epsilon])p_{K'}^j + \sum_j d_j h_{K''}^{j,q}(-c_{K''}^j + r'' + j\epsilon])p_{K''}^j \\
0 &= v_{K'}\sum_j l_{K'}^{j,q}(-b_{K'}^j + r' + j\epsilon])p_{K'}^j + v_{K''}\sum_j l_{K''}^{j,q}(-c_{K''}^j + r'' + j\epsilon])p_{K''}^j \\
&\quad + v_{K'''}\sum_j l_{K'''}^{j,q}(-c_{K'''}^j + r''' + j\epsilon])p_{K'''}^j + \sum_{j,k\notin\{K',K'',K'''\}} v_k l_k^{j,q}(-c_k^j + j\epsilon)p_k^j.
\end{aligned}
\tag{15}
$$

B.1.4 Construction of P such that $P \in \mathcal{P}$

As illustrated above, arbitrary choices of (r', r'', r'''), $\epsilon > 0$, and probabilities $(p_k^j)_{j,k}$ such that (15) holds imply that P has strictly negative partial effects and that $\theta^*(P) = \theta$. It remains to show that such a P exists that also satisfies (b) $\mathrm{E}_P[Y|D,V] \in \mathrm{int}(\mathcal{Y}) = (\underline{y}, \overline{y})$, and (c) the

support of (D, V) is $\mathcal{D} \times \mathcal{V}$. I do so by first writing the solution of (15) in $r \equiv (r', r'', r''')^\top$ as a function of ϵ and the probabilities $(p_k^j)_{j,k}$. Then, I show that this solution in r is arbitrarily small as ϵ and $(p_k^j)_{j,k}$ are taken to be arbitrarily small (but strictly positive) values. (c) then follows directly from strictly positive $(p_k^j)_{j,k}$, and (b) follows from the fact that ϵ and the solution in r can be taken to be arbitrarily small so that combining with (11) and the fact that $A_k^j \in (\underline{y}, \overline{y})$ by Assumption 2 delivers the desired result.

For some $\delta > 0$ (that I choose), set $p_k^j = \delta, \forall (j,k) \in ([J] \times [K]) \backslash \{(J', K'), (J'', K''), (J'', K''')\}$. Also fix $p_{K'}^{J'} = \frac{1}{3}$, $p_{K''}^{J''} = \frac{1}{3}$, and so $p_{K'''}^{J'} = \frac{1}{3} - \delta(JK - 3)$, for $\delta > 0$ sufficiently small such that $\sum_{jk} p_k^j = 1$.

Rewriting (15) in matrix notation, we have

$$-\Delta_{\delta,\epsilon} = M_\delta r, \tag{16}$$

where $\Delta_{\delta,\epsilon} \equiv \begin{bmatrix} \Delta_{\delta,\epsilon}^{1,1} & \Delta_{\delta,\epsilon}^{1,2} & \Delta_{\delta,\epsilon}^{1,3} & \dots & \Delta_{\delta,\epsilon}^{q,1} & \Delta_{\delta,\epsilon}^{q,2} & \Delta_{\delta,\epsilon}^{q,3} & \dots & \Delta_{\delta,\epsilon}^{Q,1} & \Delta_{\delta,\epsilon}^{Q,2} & \Delta_{\delta,\epsilon}^{Q,3} \end{bmatrix}^\top$ with

$$\begin{aligned}
\Delta_{\delta,\epsilon}^{q,1} &\equiv \epsilon J' f_{K'}^{J',q} p_{K'}^{J'} + \epsilon J'' f_{K''}^{J'',q} p_{K''}^{J''} + \epsilon J'' f_{K'''}^{J'',q} p_{K'''}^{J''} \\
&\quad + \sum_{j \neq J'} f_{K'}^{j,q} (-b_{K'}^j + j\epsilon) p_{K'}^j + \sum_{j \neq J''} f_{K''}^{j,q} (-c_{K''}^j + j\epsilon) \\
&\quad + \sum_{j \neq J''} f_{K'''}^{j,q} (-c_{K'''}^j + j\epsilon) p_{K'''}^j + \sum_{j,k \notin \{K', K'', K'''\}} f_k^{j,q} (-c_k^j + j\epsilon) p_k^j, \\
&= \epsilon \left(\frac{J' f_{K'}^{J',q}}{3} + \frac{J'' f_{K''}^{J'',q}}{3} + \frac{J'' f_{K'''}^{J'',q}}{3} - \delta J'' f_{K'''}^{J'',q} (JK - 3) \right) \\
&\quad + \delta \Bigg(\sum_{j \neq J'} f_{K'}^{j,q} (-b_{K'}^j + j\epsilon) + \sum_{j \neq J''} f_{K''}^{j,q} (-c_{K''}^j + j\epsilon) \\
&\qquad + \sum_{j \neq J''} f_{K'''}^{j,q} (-c_{K'''}^j + j\epsilon) + \sum_{j,k \notin \{K', K'', K'''\}} f_k^{j,q} (-c_k^j + j\epsilon) \Bigg),
\end{aligned}$$

$$
\begin{aligned}
\Delta^{q,2}_{\delta,\epsilon} \equiv\ & \epsilon d_{K'} J' h^{J',q}_{K'} p^{J'}_{K'} + \epsilon d_{J''} J'' h^{J'',q}_{K''} p^{J''}_{K''} + \epsilon d_{J''} J'' h^{J'',q}_{K'''} p^{J''}_{K'''} \\
& + \sum_{j \neq J'} d_j h^{j,q}_{K'} (-b^j_{K'} + j\epsilon) p^j_{K'} + \sum_{j \neq J''} d_j h^{j,q}_{K''} (-c^j_{K''} + j\epsilon) \\
& + \sum_{j \neq J''} d_j h^{j,q}_{K'''} (-c^j_{K'''} + j\epsilon) p^j_{K'''} + \sum_{j,k \notin \{K',K'',K'''\}} d_j h^{j,q}_k (-c^j_k + j\epsilon) p^j_k, \\
=\ & \epsilon \Big(\frac{d_{J'} J' h^{J',q}_{K'}}{3} + \frac{d_{J''} J'' h^{J'',q}_{K''}}{3} + \frac{d_{J''} J'' h^{J'',q}_{K'''}}{3} - \delta J'' d_{J''} h^{J'',q}_{K'''} (JK - 3) \Big) \\
& + \delta \Bigg(\sum_{j \neq J'} d_j h^{j,q}_{K'} (-b^j_{K'} + j\epsilon) + \sum_{j \neq J''} d_j h^{j,q}_{K''} (-c^j_{K''} + j\epsilon) \\
& \qquad + \sum_{j \neq J''} d_j h^{j,q}_{K'''} (-c^j_{K'''} + j\epsilon) + \sum_{j,k \notin \{K',K'',K'''\}} d_j h^{j,q}_k (-c^j_k + j\epsilon) \Bigg), \\
\Delta^{q,3}_{\delta,\epsilon} \equiv\ & \epsilon v_{K'} J' l^{J',q}_{K'} p^{J'}_{K'} + \epsilon v_{K''} J'' l^{J'',q}_{K''} p^{J''}_{K''} + \epsilon v_{K'''} J'' l^{J'',q}_{K'''} p^{J''}_{K'''} \\
& + v_{K'} \sum_{j \neq J'} l^{j,q}_{K'} (-b^j_{K'} + j\epsilon) p^j_{K'} + v_{K''} \sum_{j \neq J''} l^{j,q}_{K''} (-c^j_{K''} + j\epsilon) \\
& + v_{K'''} \sum_{j \neq J''} l^{j,q}_{K'''} (-c^j_{K'''} + j\epsilon) p^j_{K'''} + \sum_{j,k \notin \{K',K'',K'''\}} v_k l^{j,q}_k (-c^j_k + j\epsilon) p^j_k, \\
=\ & \epsilon \Big(\frac{v_{K'} J' l^{J',q}_{K'}}{3} + \frac{v_{K''} J'' l^{J''}_{K''}}{3} + \frac{v_{K'''} J'' l^{J'',q}_{K'''}}{3} - \delta J'' v_{K'''} l^{J'',q}_{K'''} (JK - 3) \Big) \\
& + \delta \Bigg(v_{K'} \sum_{j \neq J'} l^{j,q}_{K'} (-b^j_{K'} + j\epsilon) + v_{K''} \sum_{j \neq J''} l^{j,q}_{K''} (-c^j_{K''} + j\epsilon) \\
& \qquad + v_{K'''} \sum_{j \neq J''} l^{j,q}_{K'''} (-c^j_{K'''} + j\epsilon) + \sum_{j,k \notin \{K',K'',K'''\}} v_k l^{j,q}_k (-c^j_k + j\epsilon) \Bigg),
\end{aligned}
$$

where I used that $b^{J'}_{K'} = 0, c^{J''}_{K''} = c^{J''}_{K''} = 0$. Further, M_δ is an $3Q \times 3$ matrix

$$
M_\delta \equiv \begin{bmatrix} M^{1,1}_\delta & M^{1,2}_\delta & M^{1,3}_\delta & \dots & M^{q,1}_\delta & M^{q,2}_\delta & M^{q,3}_\delta & \dots & M^{Q,1}_\delta & M^{Q,2}_\delta & M^{Q,3}_\delta \end{bmatrix}^\top
$$

with

$$
M^{q,1}_\delta \equiv \begin{bmatrix} f^{J',q}_{K'} p^{J'}_{K'} + \sum_{j \neq J'} f^{j,q}_{K'} p^j_{K'} \\ f^{J'',q}_{K''} p^{J''}_{K''} + \sum_{j \neq J''} f^{j,q}_{K''} p^j_{K''} \\ f^{J'',q}_{K'''} p^{J''}_{K'''} + \sum_{j \neq J''} f^{j,q}_{K'''} p^j_{K'''} \end{bmatrix} = \begin{bmatrix} \frac{f^{J',q}_{K'}}{3} + \delta \sum_{j \neq J'} f^{j,q}_{K'} \\ \frac{f^{J'',q}_{K''}}{3} + \delta \sum_{j \neq J''} f^{j,q}_{K''} \\ \frac{f^{J'',q}_{K'''}}{3} + \delta (\sum_{j \neq J''} f^{j,q}_{K'''} + f^{J'',q}_{K'''} (3 - JK)) \end{bmatrix},
$$

$$M_\delta^{q,2} \equiv \begin{bmatrix} d_{J'} h_{K'}^{j,q} p_{K'}^{J'} + \sum_{j \neq J'} d_j h_{K'}^{j,q} p_{K'}^{j} \\ d_{J''} h_{K''}^{J'',q} p_{K''}^{J''} + \sum_{j \neq J''} d_j h_{K''}^{j,q} p_{K''}^{j} \\ d_{J''} h_{K'''}^{J'',q} p_{K'''}^{J''} + \sum_{j \neq J''} d_j h_{K'''}^{j,q} p_{K'''}^{j} \end{bmatrix} = \begin{bmatrix} \frac{d_{J'} h_{K'}^{j,q}}{3} + \delta \sum_{j \neq J'} d_j h_{K'}^{j,q} \\ \frac{d_{J''} h_{K''}^{J'',q}}{3} + \delta \sum_{j \neq J''} d_j h_{K''}^{j,q} \\ \frac{d_{J''} h_{K'''}^{J'',q}}{3} + \delta (\sum_{j \neq J''} d_j h_{K'''}^{j,q} + d_{J''} h_{K'''}^{J'',q} (3 - JK)) \end{bmatrix},$$

$$M_\delta^{q,3} \equiv \begin{bmatrix} v_{K'} (l_{K'}^{J',q} p_{K'}^{J'} + \sum_{j \neq J'} l_{K'}^{j,q} p_{K'}^{j}) \\ v_{K''} (l_{K''}^{J'',q} p_{K''}^{J''} + \sum_{j \neq J''} l_{K''}^{j,q} p_{K''}^{j}) \\ v_{K'''} (l_{K'''}^{J'',q} p_{K'''}^{J''} + \sum_{j \neq J''} l_{K'''}^{j,q} p_{K'''}^{j}) \end{bmatrix} = \begin{bmatrix} \frac{v_{K'} l_{K'}^{J',q}}{3} + \delta v_{K'} \sum_{j \neq J'} l_{K'}^{j,q} \\ \frac{v_{K''} l_{K''}^{J'',q}}{3} + \delta v_{K''} \sum_{j \neq J''} l_{K''}^{j,q} \\ \frac{v_{K'''} l_{K'''}^{J'',q}}{3} + \delta v_{K'''} (\sum_{j \neq J''} l_{K'''}^{j,q} + l_{K'''}^{J'',q} (3 - JK)) \end{bmatrix}.$$

Now note that

$$\lim_{\delta \to +0} M_\delta \equiv \overline{M} = \begin{bmatrix} \overline{M}^{1,1} & \overline{M}^{1,2} & \overline{M}^{1,3} & \dots & \overline{M}^{q,1} & \overline{M}^{q,2} & \overline{M}^{q,3} & \dots & \overline{M}^{Q,1} & \overline{M}^{Q,2} & \overline{M}^{Q,3} \end{bmatrix}^\top$$

where

$$\overline{M}^{q,1} \equiv \frac{1}{3} \begin{bmatrix} f_{K'}^{J',q} \\ f_{K''}^{J'',q} \\ f_{K'''}^{J'',q} \end{bmatrix}, \quad \overline{M}^{q,2} \equiv \frac{1}{3} \begin{bmatrix} d_{J'} h_{K'}^{J',q} \\ d_{J''} h_{K''}^{J'',q} \\ d_{J''} h_{K'''}^{J'',q} \end{bmatrix}, \quad \overline{M}^{q,3} \equiv \frac{1}{3} \begin{bmatrix} v_{K'} l_{K'}^{J',q} \\ v_{K''} l_{K''}^{J'',q} \\ v_{K'''} l_{K'''}^{J'',q} \end{bmatrix}.$$

By Assumption 3, it thus follows that $\overline{M}$ has full column-rank so that $\lim_{\delta \to +0} M_\delta^\top M_\delta = \overline{M}^\top \overline{M}$ is non-singular. Hence, by continuity of $\|\cdot\|$ and continuity of the inverse, $\lim_{\delta \to +0} \|(M_\delta^\top M_\delta)^{-1}\| = O(1)$.

Further, note that $\lim_{(\epsilon,\delta) \to +(0,0)} M_\delta^\top \Delta_{\delta,\epsilon} = 0$. Combining, it then follows that

$$\begin{aligned} \lim_{(\epsilon,\delta) \to +(0,0)} |(M_\delta^\top M_\delta)^{-1} M_\delta^\top \Delta_{\delta,\epsilon}| \leq & \lim_{(\epsilon,\delta) \to +(0,0)} \|(M_\delta^\top M_\delta)^{-1}\| \|M_\delta^\top \Delta_{\delta,\epsilon}\|_2 \\ = & O(1) \left\| \lim_{(\epsilon,\delta) \to +(0,0)} M_\delta^\top \Delta_{\delta,\epsilon} \right\|_2 = 0, \end{aligned}$$

where the first inequality follows from Cauchy-Schwarz.

It then follows from (16) that $r = (r', r'', r''')$ can be made arbitrarily close to 0 by choosing $(\epsilon, \delta) > 0$ sufficiently small as desired. This concludes the proof.

B.2 Proof of Lemma 1

Fix an arbitrary $P \in \mathcal{P}$. Note that for $\mathcal{D} = \{d_1, \ldots, d_J\}$, we can write

$$\mathrm{E}_P[Y|D,V] = \mathrm{E}_P[Y|D = d_1, V] + \sum_{j=2}^{D} \Delta_P^j(V),$$

where $\Delta_P^j(V) \equiv \mathrm{E}_P[Y|D = d_j, V] - \mathrm{E}_P[Y|D = d_{j-1}, V]$. Similarly, $\forall \theta \in \Theta$,

$$m_\theta(D,V) = m_\theta(d_1,V) + \sum_{j=2}^{D} \Delta_\theta^j(V),$$

where $\Delta_\theta^j(V) \equiv m_\theta(d_j,V) - m_\theta(d_{j-1},V)$.

Let $\theta = \theta^*(P)$. By Assumption 4, we then have

$$\begin{aligned}
0 &= \mathrm{E}_P(D - \mathrm{E}_P[D|V])(Y - m_{\theta^*(P)}(D,V)) \\
&= \mathrm{E}_P(D - \mathrm{E}_P[D|V]) \left(\mathrm{E}_P[Y|D = d_1, V] - m_{\theta^*(P)}(d_1,V)\right) \\
&\quad + \mathrm{E}_P(D - \mathrm{E}_P[D|V]) \sum_{j=2}^{D} \left(\Delta_P^j(V) - \Delta_{\theta^*(P)}^j(V)\right) \\
&\overset{[1]}{=} \mathrm{E}_P\mathrm{E}_P[(D - \mathrm{E}_P[D|V])|V] \left(\mathrm{E}_P[Y|D = d_1, V] - m_{\theta^*(P)}(d_1,V)\right) \\
&\quad + \sum_{j=2}^{J} \mathrm{E}_P(D - \mathrm{E}_P[D|V]) \mathbb{1}\{d_j \le D\} \left(\Delta_P^j(V) - \Delta_{\theta^*(P)}^j(V)\right) \\
&\overset{[2]}{=} \sum_{j=2}^{J} \mathrm{E}_P\mathrm{E}_P[(D - \mathrm{E}_P[D|V]) \mathbb{1}\{d_j \le D\}|V] \left(\Delta_P^j(V) - \Delta_{\theta^*(P)}^j(V)\right),
\end{aligned}$$

where [1] and [2] follow from the law of iterated expectations. As a consequence,

$$\sum_{j=2}^{J} \mathrm{E}_P \Delta_P^j(V) \omega_P(j,V) = \sum_{j=2}^{J} \mathrm{E}_P \Delta_{\theta^*(P)}^j(V) \omega_P(j,V),$$

where, $\forall j \in [J] \setminus 1$,

$$\begin{aligned}
\omega_P(j,V) &\equiv \mathrm{E}_P[(D - \mathrm{E}_P[D|V]) \mathbb{1}\{d_j \le D\}|V] \\
&= (\mathrm{E}_P[D|D \ge d_j, V] - \mathrm{E}_P[D|D < d_j, V]) \Pr(D \ge d_j|V) \Pr(D < d_j|V) \overset{a.s.}{\ge} 0.
\end{aligned}$$

Since the choice of $P \in \mathcal{P}$ was arbitrary, this concludes the proof.

B.3 Proof of Proposition 2

Fix an arbitrary $P \in \mathcal{P}$. Note that Lemma 1 applies by Assumption 4. By Assumption 5, $m_{\theta(P)}$ is monotone in D given V, so that the right hand side of (4) is either a positively weighted sum of weakly negative terms or a positively weighted sum of weakly positive terms. Hence, the sign of the right hand side (weakly) determines the sign of all terms in the sum.

Now, if $P \in \mathcal{P}_+$, then Lemma 1 implies

$$0 \leq \sum_{j=2}^{J} \mathrm{E}_P \left[\left(m_\theta(d_j, V) - m_\theta(d_{j-1}, V) \right) \omega_P(j, V) \right].$$

Hence, because $\omega_P(j, V) \overset{a.s.}{\geq} 0$, monotonicity implies

$$m_{\theta^*(P)}(d', V) \overset{a.s.}{\geq} m_{\theta^*(P)}(d', V), \quad \forall d' \geq d \in \mathcal{D}.$$

Analogous arguments imply that if $P \in \mathcal{P}_-$, then

$$m_{\theta^*(P)}(d', V) \overset{a.s.}{\leq} m_{\theta^*(P)}(d', V), \quad \forall d' \geq d \in \mathcal{D}.$$

Since the choice of $P \in \mathcal{P}$ was arbitrary, this concludes the proof.

B.4 Proof of Corollaries 1-3

B.4.1 Proof of Corollary 1

Fix an arbitrary $P \in \mathcal{P}$. By Assumption 6, we have $\forall \gamma \in \mathbb{R}^{|\mathcal{V}|}$ that $\mathrm{E}_P V^\top \gamma (Y - m_{\theta^*(P)}(D, V)) = 0$. Take γ_P satisfying $\mathrm{E}_P[D|V] \overset{a.s.}{=} V^\top \gamma_P$ which exists by linearity of $\mathrm{E}_P[D|V]$. Then, by Assumption 6, we have

$$\begin{aligned}
0 &= \mathrm{E}_P D(Y - m_{\theta^*(P)}(D, V)) - \mathrm{E}_P V^\top \gamma_P (Y - m_{\theta^*(P)}(D, V)) \\
&= \mathrm{E}_P (D - V^\top \gamma_P)(Y - m_{\theta^*(P)}(D, V)) \\
&= \mathrm{E}_P (D - \mathrm{E}_P[D|V])(Y - m_{\theta^*(P)}(D, V)).
\end{aligned}$$

Since the choice of $P \in \mathcal{P}$ was arbitrary, this implies that Assumption 4 is satisfied. Applying Proposition 2 completes the proof.

B.4.2 Proof of Corollary 2

The proof uses similar arguments as the previous proof and is included here for completeness.

Fix an arbitrary $P \in \mathcal{P}$. By Assumption 7, we have $\mathrm{E}_P h(V)(Y - m_{\theta^*(P)}(D,V)) = 0$ for all L_2-integrable $h : \mathcal{V} \to \mathbb{R}$. Take $h_P(V) \equiv \mathrm{E}_P[D|V]$. Then, by Assumption 7, we have

$$\begin{aligned} 0 &= \mathrm{E}_P D(Y - m_{\theta^*(P)}(D,V)) - \mathrm{E}_P h_P(V)(Y - m_{\theta^*(P)}(D,V)) \\ &= \mathrm{E}_P(D - h_P(V))(Y - m_{\theta^*(P)}(D,V)) \\ &= \mathrm{E}_P(D - \mathrm{E}_P[D|V])(Y - m_{\theta^*(P)}(D,V)). \end{aligned}$$

Since the choice of $P \in \mathcal{P}$ was arbitrary, this implies that Assumption 4 is satisfied. Applying Proposition 2 completes the proof.

B.5 Proof of Corollary 3

The proof uses similar arguments as the proof of Proposition 2 and Corollary 2 and is included here for completeness.

By Assumption 8, we have $\mathrm{E}_P[h(V)(Y - m_{\theta^*(P)}(D,V))|V] \overset{a.s.}{=} 0$, for all L_2-integrable $h : \mathcal{V} \to \mathbb{R}$ and $P \in \mathcal{P}$. Take $h_P(V) \equiv \mathrm{E}_P[D|V]$, then by Assumption 8, we have

$$\begin{aligned} 0 &= \mathrm{E}_P[D(Y - m_{\theta^*(P)}(D,V))|V] - \mathrm{E}_P[h_P(V)(Y - m_{\theta^*(P)}(D,V))|V] \\ &= \mathrm{E}_P[(D - \mathrm{E}_P[D|V])(Y - m_{\theta^*(P)}(D,V))|V], \quad \forall P \in \mathcal{P}. \end{aligned} \tag{17}$$

I continue with proving a simple intermediate lemma analogous to Lemma 1.

Lemma 4. *Suppose that the estimand $\theta^* : \mathcal{P} \to \Theta$ is unique and satisfies* (17)*, $\forall P \in \mathcal{P}$. Then, $\forall P \in \mathcal{P}, v \in \mathcal{V}$,*

$$\begin{aligned} & \sum_{j=2}^{J} \left(\mathrm{E}_P[Y|D = d_j, V = v] - \mathrm{E}_P[Y|D = d_{j-1}, V = v] \right) \omega_P(j, v) \\ = \quad & \sum_{j=2}^{J} \left(m_\theta(d_j, v) - m_\theta(d_{j-1}, v) \right) \omega_P(j, v), \end{aligned} \tag{18}$$

where $\omega_P(j, V)$ is defined as in Lemma 1.

Proof. Fix an arbitrary $P \in \mathcal{P}$. Also fix an arbitrary $v \in \mathcal{V}$. Now let P_v denote the joint distribution of $(Y, D)|V = v$. Replacing P with P_v in the proof of Lemma 1, as well as replacing references to Assumption 4 with references to (17), then gives the desired result. □

To finish the proof of Corollary 3, fix an arbitrary $P \in \mathcal{P}_\pm, v \in \mathcal{V}$. Note that Lemma 4 applies. By Assumption 5, $m_{\theta^*(P)}$ is monotone in D given V, so that the right hand side of (18) is either a positively weighted sum of weakly negative terms or a positively weighted sum of weakly positive terms. Hence, the sign of the right hand side (weakly) determines the sign of all terms in the sum.

Now, if $v \in \mathcal{V}_+$, then Lemma 4 implies

$$0 \leq \sum_{j=2}^{J} \left(m_\theta(d_j, v) - m_\theta(d_{j-1}, v) \right) \omega_P(j, v).$$

Hence, because $\omega_P(j, v) \geq 0$, monotonicity implies

$$m_{\theta^*(P)}(d', v) \geq m_{\theta^*(P)}(d', v), \quad \forall d' \geq d \in \mathcal{D}.$$

Analogous arguments imply that if $v \in \mathcal{V}_-$, then

$$m_{\theta^*(P)}(d', v) \leq m_{\theta^*(P)}(d', v), \quad \forall d' \geq d \in \mathcal{D}.$$

Since the choice of $P \in \mathcal{P}_\pm$ and $v \in \mathcal{V}$ was arbitrary, this concludes the proof.

B.6 Proof of Lemma 2

Note that, $\forall d', d \in \mathcal{D}$,

$$\begin{aligned}
\text{CATE}_P^{d',d}(V) &= \text{E}_P[g(d', U)|V] - \text{E}_P[g(d, U)|V] \\
&= \text{E}_P[g(D, U)|D = d', V] - \text{E}_P[g(D, U)|D = d, V] \\
&= \text{E}_P[Y|D = d', V] - \text{E}_P[Y|D = d, V],
\end{aligned}$$

where the equalities follow from Assumption 9. Substituting the final expression for $\text{CATE}_P^{d',d}(V)$ in (a)-(c) gives the desired results.