Identification and Estimation in a Class of Potential Outcomes Models

Manu Navjeevan
UCLA
mnavjeevan@g.ucla.edu

Rodrigo Pinto
UCLA
rodrig@econ.ucla.edu

Andres Santos
UCLA
andres@econ.ucla.edu

October, 2023

Abstract

This paper develops a class of potential outcomes models characterized by three main features: (i) Unobserved heterogeneity can be represented by a vector of potential outcomes and a "type" describing the manner in which an instrument determines the choice of treatment; (ii) The availability of an instrumental variable that is conditionally independent of unobserved heterogeneity; and (iii) The imposition of convex restrictions on the distribution of unobserved heterogeneity. The proposed class of models encompasses multiple classical and novel research designs, yet possesses a common structure that permits a unifying analysis of identification and estimation. In particular, we establish that these models share a common necessary and sufficient condition for identifying certain causal parameters. Our identification results are constructive in that they yield estimating moment conditions for the parameters of interest. Focusing on a leading special case of our framework, we further show how these estimating moment conditions may be modified to be doubly robust. The corresponding double robust estimators are shown to be asymptotically normally distributed, bootstrap based inference is shown to be asymptotically valid, and the semi-parametric efficiency bound is derived for those parameters that are root-n estimable. We illustrate the usefulness of our results for developing, identifying, and estimating causal models through an empirical evaluation of the role of mental health as a mediating variable in the Moving To Opportunity experiment.

KEYWORDS: Potential outcomes, instrumental variables, mediation, identification, double robustness, Lasso, semiparametric efficiency.

1 Introduction

Potential outcomes models have become the leading framework for identifying and estimating causal effects in applications with heterogeneous treatment responses. Originally developed for randomized experiments (Neyman, 1990) and observational studies (Rubin, 1974), these models have also proven transformative in shaping our understanding of instrumental variable approaches for addressing selection. In this regard, fundamental contributions were made by Imbens and Angrist (1994) and Heckman and Vytlacil (2005) who highlighted the importance to identification of restricting the manner in which an instrument can impact treatment decisions. A subsequent literature has built on their foundational work by developing identifying restrictions for a wide range of empirically relevant settings, including applications involving ordered, unordered, and multiple instruments, as well as the presence of mediating variables.

In this paper, we propose and develop a class of potential outcomes models that unifies and expands upon these identification strategies. The main assumptions imposed by our framework are: (i) Unobserved heterogeneity can be represented by a vector of potential outcomes and a "type" describing the manner in which the instruments determines the treatment decision; (ii) The instrumental variables are conditionally independent of unobserved heterogeneity; and (iii) The distribution of unobserved heterogeneity belongs to a convex set. The third requirement can include, for instance, support restrictions on the unobserved heterogeneity. These encompass, among others, the monotonicity condition of Angrist and Imbens (1995), the partial monotonicity requirement of Mogstad et al. (2021), and the revealed preference based restrictions of Kline and Walters (2016) and Pinto (2021). Additional examples of convex identifying restrictions that go beyond support conditions include the sequential exogeneity requirement of Imai et al. (2010) and the comparative compliers requirement of Mountjoy (2022).

Within the proposed class of models, we study the identification and estimation of parameters that may be expressed as the expectation (or limit of expectations) of identified functions of the unobserved heterogeneity and covariates. These parameters include, for example, local average treatment effects, marginal treatment effects, and conditional expectations of covariates given types as in Abadie (2003). Our main identification result is the characterization of *necessary and sufficient* conditions for such parameters to be identified. In particular, we establish that identification is equivalent to the function whose expectation we aim to identify belonging to the closure of the range of an identified linear map Υ. Intuitively, identification is tantamount to the existence of a sequence of functions $\{\kappa_j\}$ of observable variables such that $\Upsilon(\kappa_j)$ suitably approximates the function of unobserved heterogeneity whose expectation we wish to identify. Critically, the map Υ is the same across all the models in our framework, but the sense in which $\Upsilon(\kappa_j)$ must converge depends on the restrictions being imposed – i.e.

stronger restrictions yield weaker topologies and hence the identification of additional parameters. Our characterization of identification is additionally constructive in that it implies that if the parameter of interest is identified, then it must equal the limit of the expectations of the corresponding approximating functions $\{\kappa_j\}$ of observable variables.

The constructive nature of our identification results further suggests an estimation strategy: Simply estimate the corresponding approximating sequence $\{\kappa_j\}$ and compute its sample average. Establishing asymptotic normality when the nuisance parameters $\{\kappa_j\}$ are estimated via machine learning methods, however, often requires characterizing orthogonal scores that themselves depend on the specific restrictions being imposed. In our estimation analysis, we therefore focus on a leading special case of our framework in which the identifying restrictions being imposed do not depend on the distribution of observable variables.[1] For this class of applications, we derive a double robust moment condition and follow ideas in Smucler et al. (2019), Chernozhukov et al. (2022a), and Chernozhukov et al. (2022c) by employing ℓ_1-regularization to estimate nuisance parameters. We show that the resulting estimators are asymptotically normally distributed and that bootstrap based inference is asymptotically valid even if the estimator converges at a slower than root-n rate. We additionally derive the semiparametric efficiency bound for these parameters and characterize the conditions under which it is finite. As we illustrate in the context of Mogstad et al. (2021), the latter result has important implications for root-n estimability in applications with continuous instruments.

Our results are not only useful in the context of existing models, but can also be instrumental in developing, identifying, and estimating novel causal models. We highlight the utility of our analysis in this regard with an empirical analysis of the Moving to Opportunity (MTO) experiment. Specifically, we evaluate a conjecture by Ludwig et al. (2008) who suggested that improved mental health may play an important role in the causal mechanism through which moving from high to low-poverty neighborhoods impacts economic outcomes. Guided by our necessary and sufficient conditions for identification, we devise a model that enables us to identify and estimate the mediating effects of improved mental health for different subpopulations. Overall, we find evidence in support of the causal channel in which mental health mediates the effect of neighborhood relocation on labor market outcomes.

This paper contributes to a vast literature on potential outcomes models. Our analysis appears to be the first to establish the unifying role that a common linear map Υ plays in determining identification across a variety of models and assumptions. Through this common structure, our analysis delivers *necessary and sufficient* conditions for identification – a result that complements the literature, which has largely focused on *sufficient* conditions for identification.[2] In some applications, our results yield conditions under

[1] As we illustrate in our empirical analysis, estimation under other restrictions is also possible.

[2] Corollary C-1 in Heckman and Pinto (2018), for example, provides necessary and sufficient conditions for linear restrictions implied by the model to deliver identification. Our results in contrast provide

which existing sufficient conditions for identification are in fact necessary – e.g., those in Abadie (2003) and Heckman and Pinto (2018). In other applications, our results yield novel characterizations of what parameters within our framework are identified – e.g., as in Mogstad et al. (2021). Our identification analysis is further related to work providing analytical (Manski, 1990, 2003; Heckman and Vytlacil, 2001) or computational (Balke and Pearl, 1997; Mogstad et al., 2018) bounds for partially identified parameters. By establishing that point identification is determined by a single linear equation, our analysis effectively yields a simple way to derive conditions under which these bounds collapse to a point in the models that fall within our framework.

Our estimation results rely on a double robust moment equation that coincides with that of Tan (2006) and Singh and Sun (2022) for the model of Imbens and Angrist (1994). More generally, however, our results yield the first double robust moment equation for a variety of models and parameters – e.g., under a monotonicity assumption we obtain doubly robust estimators for Heckman and Vytlacil (2005) that do not rely on the propensity score. The semiparametric efficiency bound derived in this paper similarly significantly extends the existing efficiency literature to a variety of models and parameters. Our analysis corrects some approaches in the literature by relying on results by Le Cam and Yang (1988) that enable us to construct the tangent set generated by parametric submodels of the unobserved heterogeneity; see Remark 5.2. This construction further enables us to connect to results in van der Vaart (1991b) and characterize when the efficiency bound is finite. The latter result is, to our knowledge, novel in all the models we consider.

The remainder of the paper is organized as follows. In Section 2 we formally introduce the class of models we study and discuss multiple illustrative examples. Section 3 highlights the empirical implications of our results through an analysis of the MTO experiment. Finally, Sections 4 and 5 contain all theoretical results while Section 6 briefly concludes. All mathematical derivations are included in the Appendix.

2 The Model

We consider applications in which we observe a scalar outcome $Y \in \mathbf{Y} \subseteq \mathbf{R}$, a discrete treatment $T \in \mathbf{T} \equiv \{t_1, \ldots, t_d\}$, an instrument $Z \in \mathbf{Z}$, and covariates $X \in \mathbf{X}$. We model unobserved heterogeneity through a vector of potential outcomes $Y^\star \equiv (Y^\star(t_1), \ldots, Y^\star(t_d))$ and a type $T^\star : \mathbf{Z} \to \{t_1, \ldots, t_d\}$ that describes the manner in which the instrument determines a unit's treatment decision. The observed treatment and outcome are given by the treatment choice induced by the instrument and the potential outcome corresponding to the chosen treatment – i.e. $T = T^\star(Z)$ and $Y = Y^\star(T)$.

necessary and sufficient conditions for identification that reflect all the restrictions of the model.

In order to introduce the assumptions that characterize our model, we first define

$$L^p(Q) \equiv \{f : \|f\|_{Q,p} < \infty\} \qquad \|f\|^p_{Q,p} \equiv \int |f|^p dQ;$$

i.e. $L^p(Q)$ denotes the set of functions that have a finite p^{th} moment under Q. Also recall that a distribution Q is absolutely continuous with respect to (w.r.t.) a distribution Q', denoted $Q \ll Q'$, if Q assigns zero probability to any event to which Q' assigns zero probability. Importantly, whenever Q is absolutely continuous w.r.t. Q' it admits a density w.r.t. Q' that we denote by dQ/dQ'. Finally, we let Q_0 denote the true unknown distribution of $(Y^\star, T^\star, Z, X)$ and P the identified distribution of (Y, T, Z, X).

Given the introduced notation, we impose the following two assumptions:

Assumption 2.1. *(i)* $(Y^\star, T^\star, Z, X) \sim Q_0$ *with* $Y^\star \equiv (Y^\star(t_1), \ldots, Y^\star(t_d)) \in \mathbf{Y}^\star \subseteq \mathbf{R}^d$, $Z \in \mathbf{Z}$, $X \in \mathbf{X}$, *and* $T^\star \in \mathbf{T}^\star$ *with* $\mathbf{T}^*$ *a set of functions from* $\mathbf{Z}$ *to* $\mathbf{T} \equiv \{t_1, \ldots, t_d\}$; *(ii) We observe* $T = T^\star(Z)$, $Y = Y^\star(T)$, Z, *and* X *with* $(Y, T, Z, X) \sim P$.

Assumption 2.2. *(i)* $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ *under* Q_0; *(ii)* $Q_0 \ll \mu$ *for some identified separable probability measure* μ; *(iii)* $dQ_0/d\mu$ *belongs to a set* $\mathcal{Q} \subseteq L^1(\mu)$ *for* $\mathcal{Q}$ *a closed convex subset of Banach Space* $(\mathbf{Q}, \|\cdot\|_{\mathbf{Q}})$ *with* $\|\cdot\|_{\mathbf{Q}}$ *(weakly) stronger than* $\|\cdot\|_{\mu,1}$.

Assumption 2.1 formalizes the data generating process, but has by itself no identifying power. The main conditions powering our identification results are imposed in Assumptions 2.2. In particular, Assumption 2.2(i) requires that Z be exogenous in the sense that it be statistically independent of the unobserved heterogeneity $(Y^\star, T^\star)$ conditional on X. Assumption 2.2(ii) in turn encodes restrictions on the support of the unobserved heterogeneity. For instance, since Q_0 must assign zero probability to any event to which μ assigns zero probability, we may employ μ to rule out certain realizations of T^* – e.g., to rule out "defiers" in Imbens and Angrist (1994). Finally, Assumption 2.2(iii) enables us to accommodate additional convex restrictions on the density of Q_0. These restriction may include both regularity conditions that ensure the parameter of interest is well defined (see Example 2.1 below) as well as more substantive identifying assumptions (see Section 3 below). We note that while not stated explicitly, we may set $\mathcal{Q}$ and $\mathbf{Q}$ to be identified instead of known.

The unknown true distribution Q_0 of $(Y^\star, T^\star, Z, X)$ induces, through Assumption 2.1, the identified distribution P of the observable variables (Y, T, Z, X). Absent restrictive assumptions, Q_0 is not identified in our model because there are alternative distributions Q for $(Y^\star, T^\star, Z, X)$ that induce the distribution P. In what follows, we refer to any such distribution Q as being observationally equivalent to Q_0. While it may not be possible to identify Q_0, it is still possible to restrict it to the identified set

$$\Theta_0 \equiv \{Q : Q \text{ is obs. equiv. to } Q_0,\ (Y^\star, T^\star) \perp\!\!\!\perp Z|X \text{ under } Q,\ Q \ll \mu,\ \frac{dQ}{d\mu} \in \mathcal{Q}\};$$

i.e., the identified set Θ_0 is the set of distributions for $(Y^\star, T^\star, Z, X)$ that induce P and additionally satisfy the requirements imposed on Q_0 in Assumption 2.2.

Our primary goal is to study the identification and estimation of features of the true distribution Q_0. Concretely, we study the identification and estimation of functionals of Q_0 that, for some identified sequence of functions $\{\ell_j\}$, have the structure

$$\lambda_Q \equiv \lim_{j\to\infty} E_Q[\ell_j(Y^\star, T^\star, X)]. \tag{1}$$

Here, the Q subscript is meant to emphasize that the expectation is taken with respect to a distribution Q that may not equal Q_0. For instance, a leading example is to let ℓ_j equal a known function f for all j, in which case identification of λ_{Q_0} is tantamount to the identification of the expectation of $f(Y^\star, T^\star, X)$ under the true distribution Q_0.

2.1 Examples

In order to fix ideas, we next introduce examples that highlight the flexibility of our setup. We will return to some of them throughout the paper to illustrate our results.

Our first examples are based on the most studied models in the literature.

Example 2.1. Following Rosenbaum and Rubin (1983), suppose we observe an outcome $Y \in \mathbf{R}$, a binary treatment $T \in \{0, 1\}$, covariates $X \in \mathbf{X}$, and that potential outcomes $Y^\star \equiv (Y^\star(0), Y^\star(1))$ are independent of T conditional on X. To map this setting into our framework we let $Z = T$, select μ to satisfy the restriction

$$\mu(T^\star(Z) = Z) = 1,$$

and note that Assumption 2.2(i) is then equivalent to the unconfoundedness assumption $(Y^\star(0), Y^\star(1)) \perp\!\!\!\perp T|X$. The classical parameter of interest in the literature is the average treatment effect (ATE), which corresponds to setting $\ell(Y^\star, T^*, X) = Y^\star(1) - Y^\star(0)$ in (1). Ensuring the ATE is well defined requires us to impose that $Y^\star(0)$ and $Y^\star(1)$ have a first moment, which can be accomplished through Assumption 2.2(iii). ■

Example 2.2. Consider a special case of Imbens and Angrist (1994) in which we observe an outcome Y, a binary treatment $T \in \{0, 1\}$, and a binary instrument $Z \in \{0, 1\}$. In this context, $T^\star$ is a random function mapping $\mathbf{Z} \equiv \{0, 1\}$ to $\mathbf{T} \equiv \{0, 1\}$. Following Imbens and Angrist (1994) we may employ Assumption 2.2(ii) to impose that the instrument does not induce individuals out of treatment by setting μ to satisfy

$$\mu(T^\star(1) \geq T^\star(0)) = 1;$$

i.e. μ assigns zero probability to "defiers." Functionals with the structure in (1) include the local average treatment effect (LATE) or, more generally, functionals of the marginal distributions of $Y^\star$ conditional on "compliers" (Imbens and Rubin, 1997; Abadie, 2003). We also note that Assumptions 2.1 and 2.2 can accommodate extensions to ordered discrete treatments (Angrist and Imbens, 1995) or alternatives restrictions on $T^\star$ such as the "extensive margin compliers only" requirement in Rose and Shem-Tov (2021). ■

Example 2.3. Heckman and Vytlacil (1999) study a generalized Roy model in which a unit selects whether to adopt a binary treatment $T \in \{0, 1\}$ according to

$$T = 1\{f(X, Z) \geq \xi\} \tag{2}$$

for f an unknown continuous function, ξ unobservable, and $(Y^\star(0), Y^\star(1), \xi) \perp\!\!\!\perp Z|X$. Assuming that Z is a scalar and $f(X, \cdot)$ is monotonically increasing, we may map this model into our framework by letting $T^\star \equiv 1\{f(X, \cdot) \geq \xi\}$ and setting μ to satisfy[3]

$$\mu(\lim_{z \downarrow z'} T^*(z) = T^*(z') \text{ and } T^\star(z) \geq T^\star(z') \text{ for all } z \geq z') = 1.$$

Common parameters of interest in this literature, such as the policy relevant treatment effect (PRTE) of Heckman and Vytlacil (2005), can be expressed as

$$E[h(Y^\star, \xi, X)] \tag{3}$$

where h is an identified function. Because $T^\star$ is not necessarily an invertible function of ξ, the parameter in (3) may not map into the functionals in (1) that we study. However, under regularity conditions, it is possible to show that a necessary condition for (3) to be identified is that h must depend on ξ only through $T^\star$.[4] Hence, our characterization of identification of (1) also characterizes identification of (3) and our estimation results apply to (3) whenever it is identified. We also note that our framework can accommodate other structural equations models, such as those in Lee and Salanié (2018). ■

Our next three examples illustrate the ability of our framework to accommodate multivalued treatments, vector valued instruments, and mediating variables.

Example 2.4. Kline and Walters (2016) employ the Head Start Impact Study to evaluate the cost-effectiveness of the Head Start Program. In their analysis, $Z \in \{0, 1\}$ denotes whether an individual was offered to attend a Head Start school and the treatment T can take three values: Attend a Head Start School (h), attend other schools (c), or receive home care (n). Here, $T^\star$ maps $\{0, 1\}$ to $\{h, c, n\}$ and we can therefore characterize $T^\star$ as a vector $T^\star = (T^\star(0), T^\star(1))$ taking values in $\{h, c, n\} \times \{h, c, n\}$. The main identification assumption imposed by Kline and Walters (2016) is that receiving

[3]The upper semicontinuity of T^* is a consequence of the continuity of $f(X, \cdot)$. The general case in which Z is not scalar and $f(X, \cdot)$ is not monotonic corresponds to imposing that $T^\star(z) \geq T^\star(z')$ whenever $p(z, X) \geq p(z', X)$ μ-almost surely for $p(Z, X) \equiv P(T = 1|Z, X)$.

[4]Formally, we must have $h(Y^\star, \xi, X) = E[h(Y^\star, \xi, X)|Y^\star, T^\star, X]$ with probability one.

an offer to attend a Head Start school can only (weakly) induce individuals to attend a Head Start school. Formally, they require that $T^\star = (T^\star(0), T^\star(1))$ belong to the set

$$\mathbf{R}^\star \equiv \left\{ \begin{pmatrix} n \\ h \end{pmatrix}, \begin{pmatrix} c \\ h \end{pmatrix}, \begin{pmatrix} n \\ n \end{pmatrix}, \begin{pmatrix} c \\ c \end{pmatrix}, \begin{pmatrix} h \\ h \end{pmatrix} \right\}$$

with probability one, which can be mapped into our framework by setting μ to satisfy $\mu(T^\star \in \mathbf{R}^\star) = 1$. More generally, in applications with a discrete valued instrument Z we may always impose support restrictions on $T^\star$ by demanding that $\mu(T^\star \in \mathbf{R}^\star) = 1$ for some finite set of vectors $\mathbf{R}^\star \equiv \{t_1^\star, \ldots, t_r^\star\}$. Through this observation, our framework can accommodate the unordered monotonicity condition of Heckman and Pinto (2018), the analysis of the Moving to Opportunity experiment by Pinto (2021), and the double threshold crossing model of survey non-response by Dutz et al. (2021). ■

Example 2.5. Mogstad et al. (2021) propose a partial monotonicity condition that can deliver a causal interpretation for the two stage least squares (TSLS) estimand in applications with vector valued instruments. For instance, in an empirical re-examination of Carneiro et al. (2011), the authors consider a setting in which $T \in \{0, 1\}$ indicates whether an individual attended college, Y represents log average hourly wage, and $Z = (C, W)$ where $C \in \{0, 1\}$ indicates whether a college is present in the county of residence at age 14 and W denotes average log earnings in the county of residence at age 17. Mogstad et al. (2021) further suppose that increasing C induces individuals into treatment, while increasing W induces individuals out of treatment. Formally, their requirement may be mapped into our framework by selecting μ to satisfy

$$\mu(T^*(1, w) \geq T^*(0, w) \text{ and } T^*(c, w) \leq T^*(c, w') \text{ for all } c \in \{0, 1\},\ w \geq w') = 1. \quad (4)$$

Under an appropriate choice of sequence $\{\ell_j\}$, parameters such as (1) can then include, for example, analogues to the marginal treatment effect (MTE) of Heckman and Vytlacil (2005). We also note that restrictions analogous to (4) where employed in the empirical study of the returns to two-year colleges by Mountjoy (2022). ■

Example 2.6. Mediation analysis aims to identify how a treatment can affect an outcome through intermediate variables called mediators. Angrist et al. (2022), for instance, argue that engagement in the first year of college is an important mediator through which student grants impact graduation rates. Letting $D \in \{0, 1\}$ indicate whether a student is awarded a grant, $M \in \{e, ne\}$ denote whether she was engaged (e) or not (ne), and $Y \in \{0, 1\}$ indicate whether she graduated within six years, we may map their study into our framework by letting $T = (D, M)$ and $Z = D$. Potential outcomes Y^* are then indexed by $t = (d, m) \in \{0, 1\} \times \{e, ne\}$, while T^* is a function mapping $\mathbf{Z} \equiv \{0, 1\}$ to $\mathbf{T} \equiv \{0, 1\} \times \{e, ne\}$. Assumptions 2.2(i)(ii) can then be employed to impose identifying restrictions such as the sequential ignorability requirement of Imai et al. (2010), while parameters with the structure in (1) include the direct and indirect effects of Pearl (2001) and Robins (2003). Finally, we note our framework can also accommodate IV mediation models, such as those of Imai et al. (2013) and Frölich and Huber (2017). ■

3 Moving to Opportunity

As a preview of our theoretical results, we first illustrate their ability to develop, identify, and estimate a causal model in the context of an empirical analysis of the MTO experiment. MTO was a housing experiment in which households living in high-poverty neighborhoods were offered vouchers that incentivized them to relocate to low-poverty neighborhoods. The experiment targeted disadvantaged families residing in impoverished housing projects from June 1994 to July 1998 (Orr et al., 2003). Approximately 75% of these households relied on welfare support, 92% were female-headed, and only one-third of adult family members had attained a high school diploma.

The MTO literature has found significant impacts on adult mental health, psychological well-being, and risky behavior (Katz et al., 2001; Kling et al., 2005, 2007) as well as on economic outcomes for compliers moving from high to low-poverty neighborhoods (Clampet-Lundquist and Massey, 2008; Pinto, 2021).[5] We revisit MTO to investigate a conjecture by Ludwig et al. (2008), who hypothesize that relocation to low-poverty neighborhoods can improve mental health and empower previously marginalized women to obtain steady employment. Specifically, we employ our theoretical results to obtain the first estimates of the role improved mental health plays as a mediator in the causal channel through which neighborhood relocation affects economic outcomes.

To map this application into our framework, we let $Z \in \{0,1\}$ indicate whether a voucher is offered, Y denote an economic outcome of interest, $D \in \{0,1\}$ indicate whether the household relocated to a low-poverty neighborhood, and $M \in \{0,1\}$ indicate whether the head of household reported having positive mental health.[6] We further set $T = (D, M)$ and let potential outcomes $Y^*(t)$ depend on $t = (d, m)$ to reflect that mental health and relocating neighborhoods can both affect economic outcomes. For our covariates X, we follow the literature in employing experimental site indicators and variables pertaining to household and neighborhood characteristics. We report additional implementation details for this empirical study in Appendix A.4.

3.1 Learning About Types

We begin by studying functionals of the distribution of types T^*, which here describe the heterogeneous manner in which Z affects mental health and the relocation decision.

[5] The evidence on economic impacts from moving from high to medium-poverty neighborhoods is less conclusive, with treatment on the treated estimates often being insignificant (Ludwig et al., 2013).

[6] Specifically, $M = 1$ if the head out household reported feeling calm during the past thirty days.

In particular, for identified functions ℓ, we first estimate expectations with the structure

$$E_{Q_0}[\ell(T^*, X)]. \tag{5}$$

A leading special case of such expectations is the probability that T^* equals a point t^* in its support, which corresponds to setting $\ell(T^*, X) = 1\{T^* = t^*\}$.

By selecting μ in Assumption 2.2(ii) to restrict the support of T^*, our model enables us to restrict how a voucher offer affects relocation decisions and mental health. Under such support restrictions, our identification results imply that (5) is identified if and only if there exists a function κ of (T, Z, X) satisfying the equation

$$\sum_{t\in\mathbf{T}}\sum_{z\in\mathbf{Z}} 1\{t^*(z) = t\}\kappa(t, z, X)P(Z = z|X) = \ell(t^*, X) \tag{6}$$

for every t^* in the support of T^* and all X; see Corollary 4.3. Moreover, any κ satisfying equation (6) can be employed to identify the expectation of $\ell(T^*, X)$ through the equality

$$E_{Q_0}[\ell(T^*, X)] = E_P[\kappa(T, Z, X)]. \tag{7}$$

Guided by this result, we impose three requirements that deliver identification of the distribution of (T^*, X): (i) A voucher offer (weakly) incentivizes households to relocate; (ii) Moving to a low-poverty neighborhood (weakly) improves mental health; and (iii) A voucher offer affects mental health only through the relocation decision. Formally, we impose these restrictions by letting $T^*(z) \equiv (D^*(z), M^*(z))$ with D^* and M^* describing how relocation and mental health respond to a voucher offer, and setting

$$\mu(D^*(1) \geq D^*(0) \text{ and } M^* = F^* \circ D^* \text{ with } F^*(1) \geq F^*(0)) = 1. \tag{8}$$

The imposed restrictions limit the support of T^* to seven possible types. These types, displayed in the first panel of Table 1, are characterized by the possible realizations of D^* and M^* – i.e. whether they are never takers, compliers, or always takers with regards to relocation and mental health status. For instances, types CN, CA, and CC always relocate when offered a voucher. Relocation, however, does not change the mental health status of types CN and CA, but improves the mental health status of type CC.

It is straightforward to verify that, for any function ℓ, equation (6) admits a solution and hence that the expectation of $\ell(T^*, X)$ is identified. In particular, it follows that the probability of each type is identified and that we may apply our asymptotically normal estimator based on (7) to estimate it; see Theorem 5.1. The first panel of Table 1 reports our estimates of the type probabilities. While all types in our model occur with a strictly positive probability, the vast majority of households either do not relocate with a voucher offer (types NN and NA) or only relocate when given a voucher

Table 1: Type Probabilities and Conditional Expectation of Baseline Variables

	Type Definitions and Probabilities						
	NN	NA	CN	CC	CA	AN	AA
$(D^*(0), M^*(0))$	(0,0)	(0,1)	(0,0)	(0,0)	(0,1)	(1,0)	(1,1)
$(D^*(1), M^*(1))$	(0,0)	(0,1)	(1,0)	(1,1)	(1,1)	(1,0)	(1,1)
Probability Point Estimate	0.258	0.253	0.194	0.065	0.203	0.014	0.013
s.e.	(0.014)	(0.014)	(0.013)	(0.023)	(0.022)	(0.004)	(0.003)
	Expected Value of Covariate Conditional on Types						
	NN	NA	CN	CC	CA	AN	AA
Household member has a disability	0.183	0.145	0.156	0.388	0.076	0.091	0.077
	(0.020)	(0.019)	(0.022)	(0.156)	(0.039)	(0.066)	(0.054)
No teens in the household	0.535	0.566	0.656	0.736	0.653	0.602	0.902
	(0.028)	(0.029)	(0.030)	(0.176)	(0.050)	(0.121)	(0.071)
Applied for a Section Eight Voucher	0.394	0.408	0.491	0.277	0.491	0.162	0.399
	(0.027)	(0.029)	(0.033)	(0.182)	(0.052)	(0.088)	(0.125)
Moved 3+ times in past 5 years	0.083	0.080	0.120	0.130	0.079	0.002	0.049
	(0.014)	(0.016)	(0.021)	(0.097)	(0.027)	(0.013)	(0.044)
No friends in the neighborhood	0.355	0.400	0.410	0.699	0.360	0.341	0.659
	(0.026)	(0.029)	(0.033)	(0.204)	(0.051)	(0.115)	(0.118)
Neighborhood is unsafe at night	0.462	0.429	0.534	0.592	0.514	0.471	0.496
	(0.027)	(0.029)	(0.033)	(0.181)	(0.052)	(0.119)	(0.132)
Gangs/Drugs are primary reason to move	0.768	0.741	0.841	0.500	0.848	0.806	0.818
	(0.022)	(0.024)	(0.023)	(0.178)	(0.042)	(0.091)	(0.105)

First panel reports the seven support points of T^* and their respective estimated probabilities. Second panel reports estimates for the expectation of baseline variables conditioned on types. All estimates account for the person-level weight for the adult survey of the interim analyses. Standard errors are displayed in parentheses.

(types CN, CC, CA). In contrast, only 2.6% of households would relocate to low-poverty neighborhoods without a voucher offer (types AN and AA). We also note that only 4.7% of households experience an improvement in mental health upon relocating (type CC).

Since the type probabilities are identified, for any type t^* we may set $\ell(T^*, X) = X1\{T^* = t^*\}/Q_0(T^* = t^*)$, in which case (5) equals the expected value of baseline variables conditional on type. The second panel of Table 1 presents estimates for such identified type characteristics. Interestingly, the observed characteristics of double compliers (type CC) substantially differ from those of other types. The CC households are more likely to include a disabled family member yet are less likely to have teenagers. They seldom apply for Section 8, exhibit higher neighborhood mobility than other types, and are less likely to report having friends in the neighborhood. Although they are slightly more likely to feel unsafe in the neighborhood, they do not cite gang or drug-related issues as the primary reason for seeking to move to relocate.

3.2 Learning About Outcomes

We next turn to estimating treatment effects in our model. A natural starting point is to examine the LATE of Imbens and Angrist (1994), which in our context equals:

$$\text{LATE} \equiv E_{Q_0}[Y^*(1, M^*(1)) - Y^*(0, M^*(0))|T^* \in \{CN, CC, CA\}].$$

The LATE informs us about the treatment effect of relocating to a low-poverty neighborhood for the subgroup of individuals who decide to relocate in response to being offered a voucher. However, the LATE is a weighted average of causal effects across types with different mental health statuses and is, as a result, not suitable for assessing the mediating role of mental health. Specifically, the LATE is a weighted average of:

$$\begin{aligned} \text{CDE}_0 &\equiv E_{Q_0}[Y^*(1,0) - Y^*(0,0)|T^* = CN] \\ \text{CDE}_1 &\equiv E_{Q_0}[Y^*(1,1) - Y^*(0,1)|T^* = CA] \\ \text{CTE} &\equiv E_{Q_0}[Y^*(1,1) - Y^*(0,0)|T^* = CC]; \end{aligned} \tag{9}$$

i.e., the LATE aggregates the "controlled direct effects" of relocating while keeping mental health status constant (CDE_0 and CDE_1) and the "controlled total effect" of simultaneously relocating and improving mental health (CTE).

Because the marginal distribution of T^* is identified, the identification of CDE_0, CDE_1, and CTE reduces to the identification of expectations with the structure

$$E_{Q_0}[\rho(Y^*(t))\ell(T^*, X)] \tag{10}$$

for identified ρ and ℓ. Applying our identification results to this context immediately implies that restriction (8) fails to identify CDE_0, CDE_1, and CTE; see Corollary 4.4. We therefore introduce an "exogeneity of irrelevant mediator choices" (EIMC) assumption: Potential outcomes corresponding to high (resp. low) poverty neighborhood and poor (resp. good) mental health are conditionally independent of what mental health would have been in a low (resp. high) poverty neighborhood. Formally, EIMC requires that

$$\begin{aligned} Y^*(0,0) &\perp M^*(1)|D^*(1) > D^*(0), M^*(0) = 0, X \\ Y^*(1,1) &\perp M^*(0)|D^*(1) > D^*(0), M^*(1) = 1, X \end{aligned}$$

which we note can be imposed in our model through the set $\mathcal{Q}$ in Assumption 2.2(iii).

Our identification results imply that EIMC and restriction (8) secure the identifica-

Table 2: Treatment Effects Estimates

Outcome	CDE_0	CDE_1	CTE	LATE	Implied LATE
Household is Economically Self-Sufficient	0.024 (0.050)	0.048 (0.063)	0.059 (0.078)	0.035 (0.039)	0.039
Sampled Adult is Employed	0.127* (0.066)	-0.003 (0.079)	0.132 (0.122)	0.066 (0.051)	0.070
Sampled Adult not in Labor Force	-0.042 (0.071)	-0.074 (0.072)	-0.412** (0.144)	-0.113** (0.049)	-0.106
Household Total Income	0.498 (2.262)	2.928 (2.191)	2.479 (3.014)	1.811 (1.523)	1.840

Estimates for the treatment effects CDE_0, CDE_1, CTE (as in (9)), and LATE for different economic outcomes. The last column evaluates LATE as a weighted average of CDE_0, CDE_1, and CTE. All estimates account for the person-level weight for the adult survey of the interim analyses. Standard errors are displayed in parentheses.

tion of CDE_0, CDE_1, and CTE; see Theorem 4.3. More generally, our results yield that expectations with the structure in (10) are identified if and only if there is a κ solving

$$E_{Q_0}[\sum_{z\in\mathbf{Z}} 1\{T^*(z)=t\}\kappa(z,X)P(Z=z|X)|V^*(t),X] = E_{Q_0}[\ell(T^*,X)|V^*(t),X], \quad (11)$$

where $V^*(t) = T^*$ if $t \in \{(0,1),(1,0)\}$, $V^*((0,0)) = T^*1\{T^* \notin \{CN,CC\}\}$, and $V^*((1,1)) = T^*1\{T^* \notin \{CA,CC\}\}$.[7] Moreover, any function κ satisfying equation (11) can be employed to identify the expectation of $\ell(T^*,X)$ through the equality

$$E_{Q_0}[\rho(Y^*(t))\ell(T^*,X)] = E_P[\rho(Y)1\{T=t\}\kappa(Z,X)]. \quad (12)$$

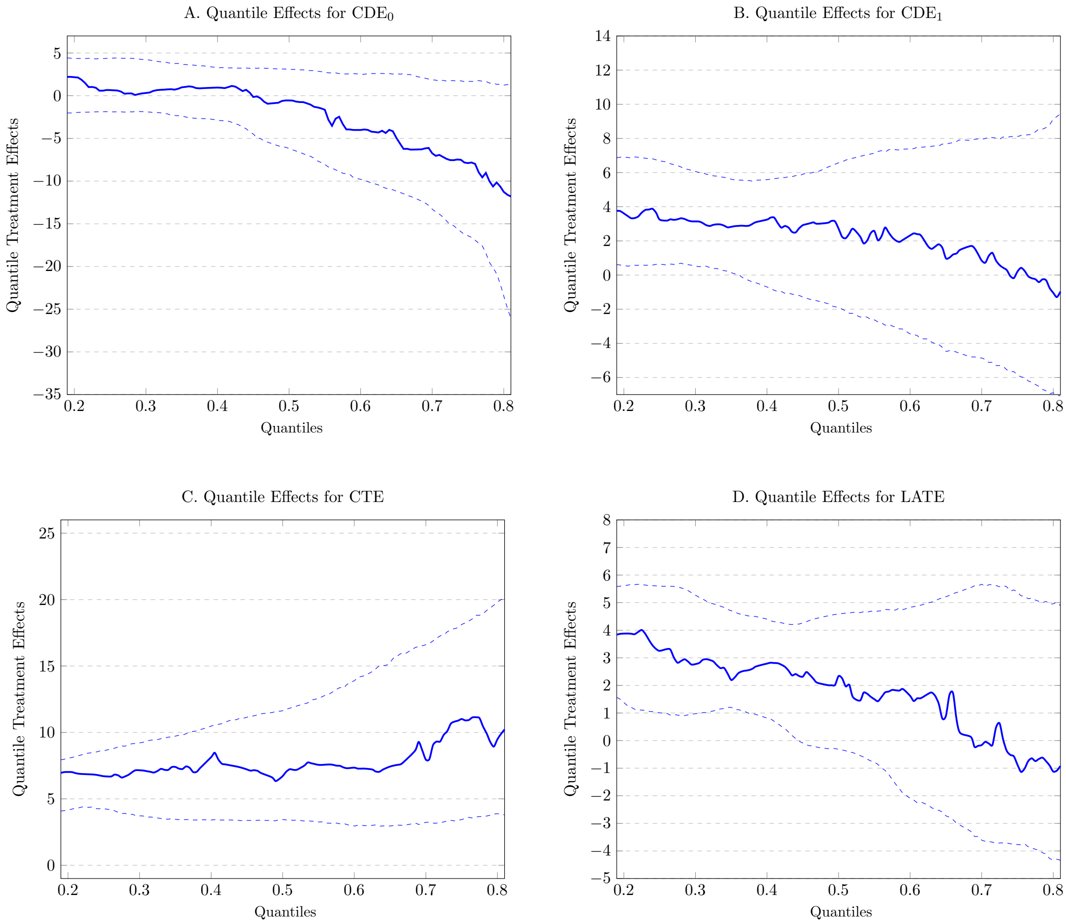
We highlight that the identifying equations in (6) and (11) are both linear, but (11) requires us to "equal" $\ell(T^*,X)$ in a weaker sense than (6). This contrast reflects a deeper observation, established in Theorem 4.1, that identification is driven by a common linear map Υ and a topology that reflects the strength of the identifying assumptions.

Table 2 reports treatment effects estimates based on an orthogonal score of (12); see Appendix A.4 for details. We examine four different outcomes: (i) Household is self-sufficient; [8] (ii) Adult participant is employed; (iii) Adult participant is not in the labor force; and (iv) Household Total Income. LATE estimates suggest that moving from high to low-poverty neighborhoods is associated with improved self-sufficiency, a higher likelihood of being employed, and increased income. The estimates for CDE_0 and CDE_1 indicate that these positive effects from relocation are largely present even if

[7]Here, with some abuse of notation, we understand $T^* \times 1$ to equal T^* and $T^* \times 0$ to equal 0.

[8]Defined as total household income in 2001 being above the poverty line and the household not currently being a recipient of welfare programs, namely, AFDC/TANF, food stamps, SSI, or Medicaid.

Figure 1: Quantile Treatment Effects for Total Household Income



Graph A: Difference in the quantiles of $Y^*(1,0)$ and $Y^*(0,0)$ conditional on type CN. Graph B: Difference in the quantiles of $Y^*(1,1)$ and $Y^*(0,1)$ conditional on type CA. Graph C: Difference in the quantiles of $Y^*(1,1)$ and $Y^*(0,0)$ conditional on type CC. Graph D: Difference in the quantiles of $Y^*(1, M^*(1))$ and $Y^*(0, M^*(0))$ conditional type $\{CA, CN, CC\}$. All estimates consider the person-level weight from the adult survey in the interim analyses. Income is measured in thousands of dollars per year.

mental heath status is unchanged. Parameter CTE encompasses two effects: the impact of moving to a low-poverty neighborhood and the effect of enhanced mental health. In full support of the conjecture by Ludwig et al. (2008), we see that mental health plays an important role in mediating the effects of neighborhood relocation on labor force participation. Table 2 additionally reports the LATE implied by our estimates for CDE_0, CDE_1, CTE, and type probabilities. The implied and estimated LATEs closely align, providing credence to our decomposition of LATE into direct and total effects.

We further investigate treatment impacts across the outcome distribution by computing Quantile Treatment Effects (QTEs) analogues to the average effects estimated in Table 2 – e.g., the QTE for CTE consists of comparing the quantiles of $Y^*(1,1)$ against those of $Y^*(0,0)$ for type CC. Figure 1 reports our QTE estimates for total household income with 95% pointwise confidence regions. Overall, we find the estimates for the QTEs corresponding to relocating while keeping mental health constant (CDE_0 and CDE_1) are decreasing though mostly statistically insignificant. In contrast, we find that the QTEs corresponding to both relocating and improving mental health (CTE) are positive and statistically significant across the quantiles we examine. Reflecting the low proportion of type CC in the population, the LATE QTEs exhibit mixed findings, being decreasing and statistically significant for lower quantiles only.

4 Identification

We next turn to our theoretical results starting, in this section, by developing a characterization of point identification for the functionals that we study.

4.1 Two Key Lemmas

We begin by introducing two lemmas that play a fundamental role in our characterization of identification. The first result is technical in nature, but crucial for our analysis.

Lemma 4.1. *If Assumptions 2.1 and 2.2 hold, then Θ_0 is convex and there is a $\bar{Q} \in \Theta_0$ such that all $Q \in \Theta_0$ are absolutely continuous with respect to $\bar{Q}$.*

In words, Lemma 4.1 establishes the existence of a distribution $\bar{Q}$ that is both in the identified set Θ_0 for the true distribution Q_0 and "larger" than any other distribution in Θ_0. Intuitively, by "larger" we mean that the support of any distribution in the identified set must be contained in the support of the distribution $\bar{Q}$. We note that there may be multiple measures $\bar{Q}$ satisfying the conclusion of Lemma 4.1. However, such measures are equivalent in the sense that they must be mutually absolutely continuous – i.e. they must assign zero probability to the same sets. Hence, whether $\bar{Q}$ assigns probability zero (or one) to a set is a property that is identified from the distribution of the data.

Our second lemma is the cornerstone of our identification analysis. In order to formally state this key result, we first introduce a linear operator Υ that maps functions of (Y, T, Z, X) to functions of (Y^*, T^*, X). Specifically, for any $f \in L^1(P)$ we set

$$\Upsilon(f) \equiv \sum_{t \in \mathbf{T}} E_{P_{Z|X}}[f(Y^\star(t), t, Z, X)1\{T^\star(Z) = t\}],$$

where $P_{Z|X}$ denotes the conditional distribution of Z given X and the notation $E_{P_{Z|X}}$ emphasizes the expectation is taken with respect to Z while $(Y^\star, T^\star, X)$ are kept "fixed." Under our assumptions, it is possible to show that $\Upsilon(f)$ in fact satisfies

$$\Upsilon(f) = E_Q[f(Y, T, Z, X)|Y^*, T^*, X] \tag{13}$$

for any Q in the identified set Θ_0 – i.e. Υ maps functions of observables into functions of unobservables by taking conditional expectations given (Y^*, T^*, X).

The next lemma combines the map Υ with the measure $\bar{Q}$ to obtain a sufficient condition for the expectation of a function ℓ of $(Y^\star, T^\star, X)$ to be identified.

Lemma 4.2. *Let Assumptions 2.1 and 2.2 hold, and ℓ satisfy $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ for some $\kappa \in L^1(P)$. Then it follows that $\ell \in L^1(Q)$ for all $Q \in \Theta_0$ and in addition*

$$E_Q[\ell(Y^\star, T^\star, X)] = E_P[\kappa(Y, T, Z, X)]. \tag{14}$$

The conclusion of Lemma 4.2 is straightforward to obtain after noting that the conditions imposed on κ and the equality in (13) ensure, for any $Q \in \Theta_0$, that

$$E_Q[\kappa(Y, T, Z, X)|Y^*, T^*, X] = \ell(Y^*, T^*, X)$$

from whence result (14) is immediate by the law of iterated expectations. The principal implication of Lemma 4.2 is a recipe for identification and estimation of the expectation of a function ℓ of (Y^*, T^*, X). In particular, Lemma 4.2 suggests estimating the expectation of ℓ by employing sample moments based on an estimator of a function κ solving the equation $\Upsilon(\kappa) = \ell$. In implementing this approach, it is often fruitful to rely on our next corollary, which obtains an alternative representation for κ in terms of the density

$$\pi \equiv \frac{dP_{Z|X}}{d\mu_{Z|X}}.$$

Corollary 4.1. *Let Assumptions 2.1 and 2.2 hold, and suppose $\nu \in L^1(P)$ is such that*

$$\mu(\sum_{t \in \mathbf{T}} E_{\mu_{Z|X}}[\nu(Y^*(t), t, Z, X)1\{T^*(Z) = t\}] = \ell(Y^*, T^*, X)) = 1. \tag{15}$$

If $\mu(\pi(Z, X) > \delta) = 1$ for some $\delta > 0$, then the function $\kappa \equiv \nu/\pi$ satisfies $\kappa \in L^1(P)$ and $\bar{Q}(\Upsilon(\kappa) = f) = 1$, and therefore $E_{Q_0}[\ell(Y^\star, T^\star, X)] = E_P[\kappa(Y, T, Z, X)]$.

Under the requirement that π be bounded away from zero, Corollary 4.1 shows that we may find a function κ solving $\Upsilon(\kappa) = \ell$ by taking the ratio of a function ν satisfying (15) and the density π. This characterization is particularly useful in applications in which the measure μ is known (instead of identified), as is the case in the majority of the examples discussed in Section 2.1. Specifically, if μ is known, then the functions ν satisfying (15) are known in that they may be computed analytically or numerically. In particular, it follows that $\kappa = \nu/\pi$ is known up to the identified density π. We will extensively employ these observations when developing our estimators in Section 5.

Remark 4.1. Revisiting Example 2.1 can be instructive in illustrating the content of Corollary 4.1. In this example, based on Rosenbaum and Rubin (1983), we imposed

$\mu(T^*(Z) = Z) = 1$ and set $\ell(Y^*, T^*, X) = Y^*(1) - Y^*(0)$. In order to identify the ATE, Corollary 4.1 suggests finding a function ν satisfying equation (15). To this end, we select μ to satisfy $\mu(Z = 0|X) = \mu(Z = 1|X) = 1/2$, which implies that

$$\nu(Y, T, Z, X) = 2Y(1\{Z = 1\} - 1\{Z = 0\})$$

solves (15). Moreover, under such a choice of μ, π satisfies $\pi(z, X) = P(Z = z|X)/2$ for $z \in \{0, 1\}$. Therefore, computing $\kappa = \nu/\pi$ and employing Corollary 4.1 yields that

$$E_{Q_0}[Y^*(1) - Y^*(0)] = E_P[\kappa(Y, T, Z, X)] = E_P[\frac{Y1\{Z = 1\}}{P(Z = 1|X)} - \frac{Y1\{Z = 0\}}{P(Z = 0|X)}],$$

which recovers the canonical propensity score reweighing moment for identifying the ATE. Similarly, applying Corollary 4.1 to the model in Imbens and Angrist (1994) recovers the the "κ-weights" identifying equations of Abadie (2003). ■

4.2 Main Result

Lemma 4.2 establishes that a sufficient condition for the identification of the expectation of a function ℓ of $(Y^\star, T^\star, X)$ is the existence of a function κ of (Y, T, Z, X) satisfying $\Upsilon(\kappa) = \ell$ in an appropriate sense. The conclusion of Lemma 4.2 is additionally constructive in that it suggests an estimator for the parameter of interest. However, our analysis so far leaves two important questions unanswered. First: Is it possible to employ a similar approach to identify and estimate the more general class of parameters that interest us? Second: Is such an approach applicable whenever the parameter of interest is identified? In other words, are our sufficient conditions for identification also necessary? We next provide affirmative answers to these questions.

Specifically, we next return to the general class of functionals with the structure

$$\lambda_Q \equiv \lim_{j\to\infty} E_Q[\ell_j(Y^\star, T^\star, X)] \tag{16}$$

and provide necessary and sufficient conditions for the identification of λ_{Q_0}. In particular, we will show that λ_{Q_0} is identified *if and only if* the functional $Q \mapsto \lambda_Q$ is in the "closure" of the set of functions that equal $\Upsilon(\kappa)$ for some κ – notice the distinction with Lemma 4.2, which requires ℓ to exactly equal $\Upsilon(\kappa)$ for some κ. Intuitively, we will establish that for a functional of Q_0 to be identified, it must be the "limit" of functionals of Q_0 whose identification can be shown through Lemma 4.2. Such a characterization of identification is additionally constructive in that it suggests estimating λ_{Q_0} by employing the sample averages of a sequence of functions $\{\kappa_j\}$ for which $\Upsilon(\kappa_j)$ "converges" to the desired functional. Formalizing this discussion, however, first requires to clarify the sense (i.e. topology) in which we mean "converge," "closure," and "limit." We next turn to this task, which requires us to introduce additional assumptions and notation.

Our next assumption introduces the final regularity conditions for our model.

Assumption 4.1. *(i) $\{\ell_j\}_{j=1}^\infty$ is an identified sequence satisfying $\{\ell_j\}_{j=1}^\infty \subset L^1(Q)$ for all $Q \in \Theta_0$; (ii) λ_Q (as in (16)) is well defined and satisfies $|\lambda_Q| < \infty$ for all $Q \in \Theta_0$; (iii) $dQ/d\mu$ belongs to the interior of $\mathcal{Q}$ in $\mathbf{Q}$ for some $Q \in \Theta_0$.*

Assumption 4.1(i) formalizes the requirement that the functions $\{\ell_j\}$ be identified and integrable with respect to every $Q \in \Theta_0$. The latter requirement can be ensured, for example, by imposing suitable regularity conditions through $\mathcal{Q}$ in Assumption 2.2(iii). In turn, Assumption 4.1(ii) formalizes the structure of the parameter of interest by imposing that the limit in (16) exists and is finite for any $Q \in \Theta_0$ – a requirement that can again be ensured through the specification of $\mathcal{Q}$. Finally, Assumption 4.1(iii) will help us establish that our sufficient conditions for identification are also necessary. Intuitively, Assumption 4.1(iii) requires that the restrictions imposed through $\mathcal{Q}$ do not bind at some $Q \in \Theta_0$ and as a result cannot point identify λ_{Q_0}.

As we have informally discussed, the identification of λ_{Q_0} hinges on whether the functional $Q \mapsto \lambda_Q$ is in, an appropriate sense, the closure of the set of functions that equal $\Upsilon(\kappa)$ for some κ. To formally introduce the relevant topology, we first define

$$\langle f, g\rangle_Q \equiv \int fgdQ$$

for any f, g such that $|fg| \in L^1(Q)$ and let Q_V denote the marginal distribution of a random variable V under Q – e.g., Q_{T^*} denotes the marginal distribution of $T^\star$ under Q. It is also useful to note that, for any suitably "smooth" $s \in L^\infty(\bar{Q}_{Y^\star T^\star X})$, the limit

$$\lim_{j\to\infty} \langle s, \ell_j\rangle_Q \tag{17}$$

will often exist for any $Q \in \Theta_0$. For instance, if we let $\mathbf{1}$ be the function that is constant at one and evaluate (17) at $s = \mathbf{1}$, then we recover λ_Q. Given this observation, we set

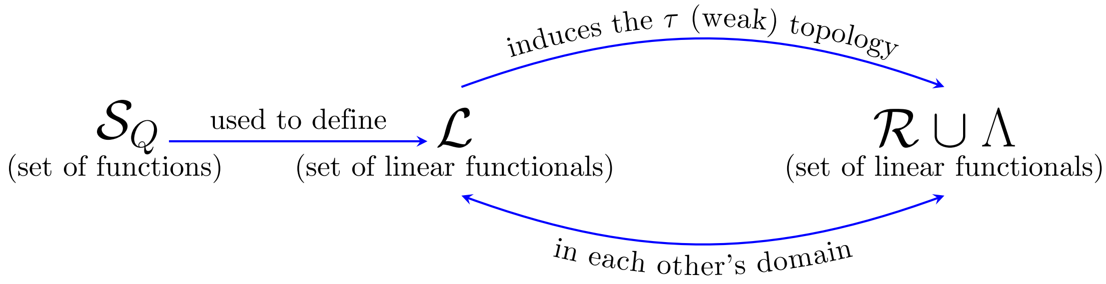
$$\mathcal{S}_Q \equiv \{s \in L^\infty(Q_{Y^*T^*X}) : |\lim_{j\to\infty} \langle s, \ell_j\rangle_Q| < \infty \text{ and } s\frac{dQ}{d\mu} \in \mathbf{Q}\},$$

where we tacitly understand every $s \in \mathcal{S}_Q$ to be such that the limit in (17) exists. Additionally, we let span$\{A\}$ denotes the linear span of a set A, and introduce a vector space $\mathcal{L}$ of linear functionals defined on the space $\bigcap_{Q\in\Theta_0} L^1(Q)$ by setting

$$\mathcal{L} \equiv \text{span}\{L : \bigcap_{Q\in\Theta_0} L^1(Q) \to \mathbf{R} \text{ s.t. } L = \langle\cdot, s\rangle_Q \text{ for some } s \in \mathcal{S},\ Q \in \Theta_0\}.$$

By Lemma 4.2, a function $\Upsilon(\kappa)$ is in the domain of all the linear functionals $L \in \mathcal{L}$ for any $\kappa \in L^1(P)$. Through duality, however, it is more instructive to identify such

Figure 2: Diagram of the definition of the τ topology dictating point identification.



functions with a set of linear functionals on $\mathcal{L}$. We therefore define the set $\mathcal{R}$ by

$$\mathcal{R} \equiv \{L' : \mathcal{L} \to \mathbf{R} \text{ s.t. } L'(L) = L(\Upsilon(\kappa)) \text{ for some } \kappa \in L^1(P)\}.$$

Similarly, $\{\ell_j\}$ generates a linear functional on $\mathcal{L}$, which we denote by Λ and equals

$$\Lambda(L) \equiv \lim_{j\to\infty} L(\ell_j),$$

and we note that any $L \in \mathcal{L}$ can also be viewed as a functional on $\{\mathcal{R} \cup \Lambda\}$ through the relation $L' \mapsto L'(L)$. Given the introduced concepts, we can finally define the topology that dictates identification. Specifically, we let τ denote the weak topology on $\{\mathcal{R} \cup \Lambda\}$ that is generated by the functionals $L \in \mathcal{L}$ – i.e. τ is the weakest topology on $\{\mathcal{R}\cup\Lambda\}$ that makes all $L \in \mathcal{L}$ continuous; see Figure 2 for a diagram summarizing this construction.

The next theorem is our main identification result.

Theorem 4.1. *If Assumptions 2.1, 2.2, and 4.1 hold, then it follows that* λ_{Q_0} *is identified if and only if* Λ *belongs to the the* τ*-closure of* $\mathcal{R}$.

Intuitively, Theorem 4.1 establishes that λ_{Q_0} is identified if and only if there is a sequence $\{L'_j\} \subseteq \mathcal{R}$ converging to Λ in the τ topology.[9] To the best of our knowledge, the characterization of all the functionals that are identified is novel in the context of all the examples in Section 2.1. To gain some insight into why Λ belonging to the τ-closure of $\mathcal{R}$ is a sufficient condition for identification, let $L_{Q_0} \equiv \langle \cdot, \mathbf{1}\rangle_{Q_0}$ and note

$$\lambda_{Q_0} = \lim_{j\to\infty} \langle \ell_j, \mathbf{1}\rangle_{Q_0} = \Lambda(L_{Q_0}). \tag{18}$$

Moreover, since $L'_j \in \mathcal{R}$ implies $L'_j(L) = L(\Upsilon(\kappa_j))$ for some κ_j, it follows from $\{L'_j\}$ converging to Λ in the τ topology that there is a sequence $\{\kappa_j\}$ satisfying

$$\Lambda(L_{Q_0}) = \lim_{j\to\infty} L'_j(L_{Q_0}) = \lim_{j\to\infty} \langle \Upsilon(\kappa_j), \mathbf{1}\rangle_{Q_0} = \lim_{j\to\infty} E_P[\kappa_j(Y, T, Z, X)], \tag{19}$$

where the final equality follows from Lemma 4.2. Importantly, results (18) and (19) not

[9] We discuss sequences for ease of exposition. However, we note that our formal arguments rely on nets because the τ topology may not be first countable and therefore not be metrizable.

only establish the identification of λ_{Q_0}, but also suggest an estimation strategy: Simply employ the sample moments based on estimates of the approximating sequence $\{\kappa_j\}$.

More surprisingly, Theorem 4.1 also establishes that the existence of the desired sequence $\{\kappa_j\}$ is in fact a necessary condition for the identification of λ_{Q_0}.[10] As a result, it is without loss of generality to estimate λ_{Q_0} by employing the discussed estimation strategy that is motivated by results (18) and (19). Moreover, while our preceding discussion suggests that we need only find a sequence $\{\kappa_j\}$ satisfying (19), Theorem 4.1 states that the identification of λ_{Q_0} is in fact only possible if the stronger requirement that $L(\Upsilon(\kappa_j)) \to \Lambda(L)$ for all $L \in \mathcal{L}$ is satisfied. As our next corollary illustrates, the latter observation can be helpful in characterizing the desired sequence $\{\kappa_j\}$.

Corollary 4.2. *Let Assumptions 2.1, 2.2 hold with $\mathcal{Q} = \mathbf{Q} = L^\infty(\mu)$, $\mu \ll \bar{Q}$ with $d\mu/d\bar{Q}$ bounded, and $\lambda_Q \equiv E_Q[\ell(Y^*, T^*, X)]$ for some identified $\ell \in L^1(\mu_{Y^*T^*X})$. Then:*

- *(i) λ_{Q_0} is identified if and only if $\lim_{j\to\infty} \|\ell - \Upsilon(\kappa_j)\|_{\mu,1} = 0$ for some $\{\kappa_j\} \subseteq L^1(P)$. Moreover, any such sequence $\{\kappa_j\}$ satisfies $\lambda_{Q_0} = \lim_{j\to\infty} E_P[\kappa_j(Y, T, Z, X)]$.*
- *(ii) Suppose in addition that $\mu(\pi(Z, X) > \delta) = 1$ for some $\delta > 0$. Then, λ_{Q_0} is identified if and only if there is a sequence $\{\nu_j\} \subseteq L^1(P)$ satisfying*

$$\lim_{j\to\infty} E_\mu[|\ell(Y^*, T^*, X) - \sum_{t\in\mathbf{T}} E_{\mu_{Z|X}}[\nu_j(Y^*(t), t, Z, X)1\{T^*(Z) = t\}]|] = 0. \quad (20)$$

Moreover, for any such $\{\nu_j\}$, $\kappa_j = \nu_j/\pi$ satisfies $\lambda_{Q_0} = \lim_{j\to\infty} E_P[\kappa_j(Y, T, Z, X)]$.

Corollary 4.2 specializes Theorem 4.1 to the case in which the functional of interest is the expectation of a function ℓ of (Y^*, T^*, X) and Assumption 2.2(iii) only imposes that the density of Q_0 be bounded. Within this context, Corollary 4.2(i) shows that λ_{Q_0} is identified if and only if ℓ is the limit of a sequence of functions $\{\Upsilon(\kappa_j)\}$ in the $\|\cdot\|_{\mu,1}$-norm. Paralleling Corollary 4.1, Corollary 4.2(ii) additionally provides conditions under which it is without loss of generality to set $\kappa_j = \nu_j/\pi$ for any sequence $\{\nu_j\}$ satisfying (20). Corollary 4.2(ii) has two important implications for applications in which μ is known and therefore the sequence $\{\nu_j\}$ is known and computable analytically or numerically; see Remark 5.1. First, Corollary 4.2(ii) provides us with a simple characterization of $\{\kappa_j\}$ in terms of the identified density π that we will use in estimation. Second, condition (20) allows us to assess whether the restrictions of our model (as embodied in μ) point identify a functional of interest or not; see our discussion of Example 2.5 below.

4.3 Special Case: Types

Functionals of the joint distribution of $(T^\star, X)$ are often of interest in their own right or as building blocks towards estimating other parameters. In this section, we specialize

[10] Theorem 4.1 only implies the existence of a net, but we again discuss sequences for ease of exposition.

our analysis to such functionals by considering parameters with the structure

$$\lambda_Q \equiv \lim_{j\to\infty} E_Q[\ell_j(T^\star, X)], \tag{21}$$

which we refer to as functionals about "types." While Theorem 4.1 of course continues to apply to this context, the fact that $\{\ell_j\}$ now only depends on $(T^\star, X)$ will allow us to sharpen our identification results. In particular, we will show that λ_{Q_0} is identified if and only if it can be identified from the joint distribution of (T, Z, X) – i.e. from "first stage" information. As a result, in estimating functionals about types we may simplify estimation by only employing the sample for (T, Z, X) (instead of (Y, T, Z, X)).

In order to formally state the conditions for our result, we first define the measure

$$\bar{Q}^{\text{it}} \equiv \bar{Q}_{Y^\star|X}\bar{Q}_{T^\star ZX},$$

which shares the same marginal distributions for (Y^*, X) and (T^*, X) as $\bar{Q}$, but is such that Y^* is independent of T^* conditionally on X. Given this notation we impose:

Assumption 4.2. *(i) $\bar{Q}^{\text{it}} \ll \bar{Q}$; (ii) $(d\bar{Q}^{\text{it}}_{Y^\star T^\star X}/d\bar{Q}_{Y^\star T^\star X})s \in \mathcal{S}_{\bar{Q}}$ for all $s \in L^\infty(\bar{Q}_{T^\star X}) \cap \mathcal{S}_{\bar{Q}}$; (iii) $(dQ_{T^\star X}/d\bar{Q}_{T^\star X})s \in \mathcal{S}_{\bar{Q}}$ for all $Q \in \Theta_0$ and $s \in L^\infty(Q_{T^\star X}) \cap \mathcal{S}_Q$; (iv) For any $Q \in \Theta_0$ and $s \in \mathcal{S}_Q$ we have that $E_Q[s(Y^\star, T^\star, X)|T^\star, X] \in \mathcal{S}_Q$.*

Assumption 4.2(i) essentially requires the support of $Y^\star$ conditional on $(T^\star, X)$ under $\bar{Q}$ to not depend on $T^\star$. We view Assumption 4.2(i) as the key requirement ensuring that the identification of functionals about types can be characterized by the distribution of (T, Z, X). Assumptions 4.2(ii) and 4.2(iii) impose restrictions on the densities $d\bar{Q}^{\text{it}}/d\bar{Q}$ and $dQ_{T^\star X}/d\bar{Q}_{T^\star X}$ (for $Q \in \Theta_0$), while Assumption 4.2(iv) requires that conditional expectations of functions in $\mathcal{S}_Q$ belong to $\mathcal{S}_Q$ as well. Assumptions 4.2(ii)-(iv) can in many applications be verified by appropriately selecting $\mathcal{Q}$ and $\mathbf{Q}$ in Assumption 2.2(iii); see, e.g., Corollary 4.3 and Section 4.3.1 below.

The next theorem is our main result on identification of functionals about types.

Theorem 4.2. *Let Assumptions 2.1, 2.2, 4.1, 4.2 hold, λ_Q be as in (21), and define*

$$\mathcal{R}_T \equiv \{L' : \mathcal{L} \to \mathbf{R} \text{ s.t. } L'(L) = L(\Upsilon(\kappa)) \text{ for some } \kappa \in L^1(P_{TZX})\}.$$

Then, it follows that λ_{Q_0} is identified if and only if Λ belongs to the the τ-closure of $\mathcal{R}_T$.

Theorem 4.2 establishes that functionals about types are identified if and only if they are identified from the distribution of (T, Z, X). Formally, Theorem 4.2 shows that λ_{Q_0} is identified if and only if it belongs to the τ-closure of the subset $\mathcal{R}_T \subseteq \mathcal{R}$ (instead of the τ-closure of $\mathcal{R}$ as in Theorem 4.1). In particular, $\mathcal{R}_T$ is generated by first stage information in that it consists of functionals corresponding to $\Upsilon(\kappa)$ for some κ

depending on (T, Z, X) only. The main implication of this result is that, when estimating functionals about types, we may search for the desired approximating sequence $\{\Upsilon(\kappa_j)\}$ for Λ by considering functions $\{\kappa_j\}$ that depend on (T, Z, X) only.

Theorem 4.2 further yields an analogue to Corollary 4.2. In particular, under parallel conditions to those imposed in Corollary 4.2, it is possible to show that the expectation of a function ℓ of (T^*, X) is identified if and only if ℓ can be approximated by a sequence $\{\Upsilon(\kappa_j)\}$ with κ_j depending only on (T, Z, X). For conciseness, however, we do not formally state such a result. Instead, in our next corollary we highlight the implications of Theorem 4.2 in the empirically salient case of discrete instruments.

Corollary 4.3. *Let Assumption 2.1, 2.2 hold, $\mathcal{Q} = \mathbf{Q} = L^1(\mu)$, $\bar{Q}^{\text{it}} \ll \bar{Q}$ with $d\bar{Q}^{\text{it}}/d\bar{Q}$ bounded, $\ell \in L^1(\bar{Q}_{T^\star X})$ be identified, Z be discrete, $P(Z = z|X) \geq \varepsilon > 0$ a.s. for all $z \in \mathbf{Z}$, and $\bar{Q}(T^\star = t^\star|X) \geq \varepsilon > 0$ a.s. for any $t^\star \in \mathbf{T}^\star$ with $\bar{Q}(T^\star = t^\star) > 0$. Then:*

(i) $\lambda_{Q_0} \equiv E_{Q_0}[\ell(T^\star, X)]$ is identified if and only if $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ for some $\kappa \in L^1(P_{TZX})$. Moreover, any such κ satisfies $\lambda_{Q_0} = E_P[\kappa(T, Z, X)]$.

(ii) Suppose in addition that $\mu(\pi(Z, X) > \delta) = 1$ for some $\delta > 0$ and $\mu \ll \bar{Q}$. Then λ_{Q_0} is identified if and only if there exists a $\nu \in L^1(P_{TZX})$ satisfying

$$\mu(\ell(T^*, X) = \sum_{t \in \mathbf{T}} E_{\mu_{Z|X}}[\nu(t, Z, X)1\{T^*(Z) = t\}]) = 1. \tag{22}$$

Moreover, for any such function ν, $\kappa = \nu/\pi$ satisfies $\lambda_{Q_0} = E_P[\kappa(T, Z, X)]$.

Corollary 4.3(i) specializes our analysis to the case of discrete instruments and no additional identifying assumptions being imposed in Assumption 2.2(ii) – a setting that covers many of the examples in Section 2.1. In this context, Corollary 4.3(i) establishes that the expectation of a function ℓ of (T^*, X) is identified if and only if ℓ equals $\Upsilon(\kappa)$ for some function κ of (T, Z, X). Under its assumptions, Corollary 4.3(i) therefore delivers a converse to Lemma 4.2. In turn, Corollary 4.3(ii) parallels Corollary 4.2 in providing conditions that can be helpful in assessing whether the desired κ exists and estimating it if it does. Such a characterization is particularly useful when μ is known, in which case the validity of (22) for some ν is independent of the distribution of the data.

4.3.1 Examples Revisited

We next revisit Examples 2.4 and 2.5 to illustrate the implications of our results in models with discrete and continuous instruments respectively.

Example 2.4 (cont.) The main features of this example, based on Kline and Walters (2016), are that Z and T^* are discrete and μ imposed $\mu(T^* \in \mathbf{R}^*) = 1$ for some set $\mathbf{R}^* \equiv \{t_1^*, \ldots, t_r^*\}$. Denoting the support of Z by $\mathbf{Z} = \{z_1, \ldots, z_q\}$, we then let

$$\omega_j(t^*) \equiv (1\{t^*(z_j) = t_1\}, \cdots, 1\{t^*(z_j) = t_d\})$$

and note Corollary 4.3(ii) implies the expectation of $\ell(T^*, X)$ is identified if and only if

$$\min_{\{s_j\}_{j=1}^q \subset \mathbf{R}^d} \sum_{i=1}^r (\ell(t_i^*, X) - \sum_{j=1}^q s_j' \omega_j(t_i^*))^2 = 0 \tag{23}$$

with probability one (over X). Moreover, provided condition (23) holds, we can find a function κ of (T, Z, X) whose expectation equals the expectation of $\ell(T^*, X)$ by setting

$$(\kappa(t_1, z_j, X), \ldots, \kappa(t_d, z_j, X)) \equiv \frac{s_j(X)'}{P(Z = z_j|X)}$$

for any $(s_1(X), \ldots, s_q(X))$ minimizing (23). For instance, specializing (23) to Kline and Walters (2016) implies that the distribution of (T^*, X) is identified in that application. More generally, the preceding discussion highlights that identifying and estimating a functional about types reduces to a simple numerical problem when Z is discrete. ■

Example 2.5 (cont.) In this example, based on Mogstad et al. (2021), $Z = (C, W)$ with C binary, W a scalar, and μ imposed that $T^*(c, w)$ be increasing in c and decreasing in w. As in Example 2.3, we also require $T^*(c, w)$ to be lower semicontinuous in w and for simplicity assume that W is continuously distributed with compact support $[\underline{\mathbf{w}}, \overline{\mathbf{w}}]$. Under these restrictions, each T^* can be identified with a unique pair (K_0^*, K_1^*) satisfying

$$T^*(c, w) = \sum_{i=0}^{1} 1\{c = i, K_i^* > w\}$$

and $K_i^* \in [\underline{\mathbf{w}}, \overline{\mathbf{w}}] \cup \{\infty\}$ – note that, for $c = i$, $K_i^* = \underline{\mathbf{w}}$ and $K_i^* = \infty$ corresponds to "never-takers" and "always-takers." We therefore study the identification of the distribution of (K_0^*, K_1^*, X) and note that the restriction that $T^*(c, w)$ be increasing in c is equivalent to imposing $\mu(K_1^* \geq K_0^*) = 1$. It is convenient to let K_i^* be continuously distributed on $(\underline{\mathbf{w}}, \overline{\mathbf{w}}]$ under μ, though we allow μ to possibly assign positive mass to $\{\underline{\mathbf{w}}\}$ and $\{\infty\}$. Under conditions paralleling those in Corollary 4.2(ii), Theorem 4.2 here implies that the expectation of a function ℓ of (K_0^*, K_1^*, X) is identified if and only if

$$\lim_{j \to \infty} E_\mu[|\sum_{i=0}^{1} \int_{K_i^*}^{\bar{\mathbf{w}}} \nu_j(0, i, w, X)dw + \int_{\underline{\mathbf{w}}}^{K_i^*} \nu_j(1, i, w, X)dw - \ell(K_0^*, K_1^*, X)|] = 0 \tag{24}$$

for some sequence $\{\nu_j(T, C, W, X)\}$. Moreover, provided condition (24) holds, the expectation of $\ell(K_0^*, K_1^*, X)$ equals the limit of the expectations of the functions

$$\kappa_j(t, c, w, X) = \frac{\nu_j(t, c, w, X)}{f_{W|CX}(w|c, X)P(C = c|X)}$$

where $f_{W|CX}$ denotes the conditional density of W given (C, X). The characterization of identification obtained in (24) in fact implies that the expectation of a function ℓ of

(K_0^*, K_1^*, X) is identified if and only if ℓ belongs to the $\|\cdot\|_{\mu,1}$-closure of the set

$$\mathcal{T} \equiv \{f : f(K_0^*, K_1^*, X) = g_0(K_0^*, X) + g_1(K_1^*, X) \text{ for some } g_0, g_1\}. \tag{25}$$

For instance, since $1\{K_1^* > a_1,\ K_0^* \le a_0\} = 1\{K_1^* > a_1\} - 1\{K_0^* > a_0\}$ for any $a_0 \ge a_1$ under $\|\cdot\|_{\mu,1}$ due to $\mu(K_1^* \ge K_0^*) = 1$, it follows that the probability of the event $\{K_1^* > a_1,\ K_0^* \le a_0\}$ is identified. Conversely, $1\{K_1^* > a_1,\ K_0^* \le a_0\}$ does not belong to the $\|\cdot\|_{\mu,1}$-closure of $\mathcal{T}$ when $a_1 > a_0$, and hence the probability of the event $\{K_1^* > a_1,\ K_0^* \le a_0\}$ is identified if and only if $a_0 \ge a_1$. ■

4.4 Special Case: Outcomes

We conclude our discussion of identification by specializing our analysis to functionals of the distribution of $(Y^*(t), T^*, X)$ for some $t \in \mathbf{T}$. In particular, we focus on parameters that for some identified function ρ and sequence $\{\ell_j\}$ have the structure

$$\lambda_Q \equiv \lim_{j\to\infty} E_Q[\rho(Y^\star(t))\ell_j(T^\star, X)], \tag{26}$$

which we refer to as functionals about "outcomes." Our primary motivation for studying these functionals is that they include features of the conditional distribution of a potential outcome given types and covariates as a special case.

Intuitively, identification of functionals of the distribution of $(Y^*(t), T^*, X)$ should only be possible from the distribution of observations for which treatment assignment T equals t. As a result, identification will now require us to approximate the sequence $\{\ell_j\}$ by employing only the subset of observations for which T equals t – contrast with Theorem 4.2 which instead employs all treatment values. In order to introduce the assumptions that enable us to formalize this intuition, we first define the measure

$$\bar{Q}^{\text{io}} \equiv \bar{Q}_{Y^\star(t_1)|X} \cdots \bar{Q}_{Y^\star(t_d)|X} \bar{Q}_{T^\star ZX};$$

i.e., for any $t \in \mathbf{T}$, $\bar{Q}^{\text{io}}$ shares the same marginal distributions for $(Y^*(t), X)$ and (T^*, X) as $\bar{Q}$, but is such that all coordinates of Y^* and T^* are mutually independent conditionally on X. We also define a function $\phi_{\bar{Q},\rho}$ of $(Y^*(t), X)$ to be given by[11]

$$\phi_{\bar{Q},\rho}(Y^\star(t), X) \equiv \frac{\rho(Y^\star(t)) - E_{\bar{Q}}[\rho(Y^\star(t))|X]}{\text{Var}_{\bar{Q}}\{\rho(Y^\star(t))|X\}}.$$

Given the introduced notation, we impose the following assumptions.

[11] If $\text{Var}_{\bar{Q}}\{\rho(Y^\star(t))|X\} = 0$, then $\rho(Y^\star(t)) = E_{\bar{Q}}[\rho(Y^\star(t))|X]$ and, setting $0/0 = 0$, we therefore let $\phi_{\bar{Q},\rho}(Y^\star(t), X) = 0$ whenever $\text{Var}_{\bar{Q}}\{\rho(Y^\star(t))|X\} = 0$.

Assumption 4.3. *(i) ρ and $\{\ell_j\}$ are identified, $\rho \in L^\infty(\bar{Q})$, and $\{\ell_j\} \subset L^1(Q)$ for all $Q \in \Theta_0$; (ii) λ_Q (as in (26)) is well defined and satisfies $|\lambda_Q| < \infty$ for all $Q \in \Theta_0$.*

Assumption 4.4. *(i) $E_Q[\rho(Y^\star(t))|T^\star, X] \in \mathcal{S}_Q$ and $E_Q[s(Y^\star, T^\star, X)|T^\star, X] \in \mathcal{S}_Q$ for any $Q \in \Theta_0$ and $s \in \mathcal{S}_Q$; (ii) $\bar{Q}^{\text{io}} \ll \bar{Q}$ and $d\bar{Q}^{\text{io}}/d\bar{Q} \in L^\infty(\bar{Q})$; (iii) $\phi_{\bar{Q},\rho} \in L^\infty(\bar{Q})$, $\text{Var}_{\bar{Q}}\{\rho(Y^\star(t))|X\} > 0$ a.s. under $\bar{Q}$, and $s(d\bar{Q}^{\text{io}}_{Y^\star T^\star X}/d\bar{Q}_{Y^\star T^\star X})\phi_{\bar{Q},\rho} \in \mathcal{S}_{\bar{Q}}$ for all $s \in L^\infty(\bar{Q}_{T^\star X}) \cap \mathcal{S}_{\bar{Q}}$; (iv) $(dQ_{T^\star X}/d\bar{Q}_{T^\star X})s \in \mathcal{S}_{\bar{Q}}$ for all $Q \in \Theta_0$ and $s \in L^\infty(Q_{T^\star X}) \cap \mathcal{S}_Q$.*

Assumption 4.3 ensures that λ_Q is well defined and $\{\ell_j \rho\}$ is integrable under any $Q \in \Theta_0$ – note that Assumption 4.3 essentially imposes that Assumptions 4.1(i)(ii) hold with $\{\ell_j \rho\}$ in place of $\{\ell_j\}$. In turn, Assumption 4.4 is similar in spirit to the conditions we imposed in Assumption 4.2 to establish our results concerning functionals about types. Specifically, we note Assumptions 4.4(i)(iv) imposes restrictions on conditional expectations and densities that parallel those of Assumptions 4.2(ii)-(iv), while Assumption 4.4(ii) imposes a key support requirement that parallels Assumption 4.2(i). Finally, Assumption 4.4(iii) requires that $\text{Var}_{\bar{Q}}\{\rho(Y^\star(t))|X\}$ be positive, which implies the parameter of interest indeed concerns features of the outcomes distribution – e.g. if ρ were constant, then (26) would fall within the framework of Section 4.3.

Our next theorem characterizes the identification of functionals about outcomes.

Theorem 4.3. *Let Assumptions 2.1, 2.2, 4.1(iii), 4.3, 4.4 hold, λ_Q be as in (26), define $L^1(P_{tZX}) \equiv \{f \in L^1(P) : f(T, Z, X) = 1\{T = t\}g(Z, X)$ for some $g \in L^1(P)\}$ and*

$$\mathcal{R}_t \equiv \{L' : \mathcal{L} \to \mathbf{R} \text{ s.t. } L'(L) = L(\Upsilon(\kappa)) \text{ for some } \kappa \in L^1(P_{tZX})\}.$$

Then, it follows that λ_{Q_0} is identified if and only if Λ belongs to the τ-closure of $\mathcal{R}_t$.

Theorem 4.3 establishes that functionals about outcomes are identified if and only if they are identified from the distribution of observations with treatment assignment T equal to t. We emphasize the contrast with Theorem 4.2, which showed identification of functionals about types is equivalent to Λ being in the τ-closure of $\mathcal{R}_T$ (instead of $\mathcal{R}_t$ in Theorem 4.3). In particular, since $\mathcal{R}_t \subseteq \mathcal{R}_T$, it follows that identification of a functional about outcomes for a given sequence $\{\ell_j\}$ implies the identification of the corresponding functional about types. More generally, since Λ depends only on the sequence $\{\ell_j\}$, the identification of λ_{Q_0} for *some* ρ implies that Λ is in the τ-closure of $\mathcal{R}_t$ and therefore that λ_{Q_0} is in fact identified for *all* suitable ρ.

Our next corollary illustrates these implications in the case of discrete instruments.

Corollary 4.4. *Let Assumptions 2.1, 2.2 hold, $\mathcal{Q} = \mathbf{Q} = L^1(\mu)$, $\bar{Q}^{\text{io}} \ll \bar{Q}$ with $d\bar{Q}^{\text{io}}/d\bar{Q}$ bounded, $\ell \in L^1(\bar{Q}_{T^\star X})$ be identified, ρ be bounded, identified, and $\text{Var}_{\bar{Q}}\{\rho(Y^\star(t))|X\} \geq \varepsilon > 0$ a.s.. If Z is discrete, $P(Z = z|X) \geq \varepsilon > 0$ a.s. for all $z \in \mathbf{Z}$, and $\bar{Q}(T^\star = t^\star|X) \geq \varepsilon > 0$ a.s. for any $t^\star \in \mathbf{T}^\star$ with $\bar{Q}(T^\star = t^\star) > 0$, then the following are equivalent:*

(i) $E_{Q_0}[\rho(Y^\star(t))\ell(T^\star, X)]$ *is identified.*

(ii) $E_{Q_0}[f(Y^\star(t))\ell(T^\star, X)]$ *is identified for any bounded* f*.*

(iii) $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ *for some* $\kappa(T, Z, X) = 1\{T = t\}g(Z, X)$ *with* $g \in L^1(P_{ZX})$*, and therefore* $E_{Q_0}[f(Y^*(t))\ell(T^*, X)] = E_P[f(Y)\kappa(T, Z, X)]$ *for any bounded* f*.*

Through the equivalence of (i) and (ii), Corollary 4.4 formalizes that the identification of λ_{Q_0} for some ρ implies the identification of λ_{Q_0} for all ρ. Corollary 4.4 additionally establishes that identification of an expectation about outcomes requires that there be a κ solving $\ell = \Upsilon(\kappa)$. Unlike Corollary 4.3, however, identification of functionals about outcomes further requires κ to only employ observations corresponding to treatment status t – i.e., κ must satisfy $\kappa(T, Z, X) = 1\{T = t\}g(Z, X)$ for some g. Paralleling Corollaries 4.2 and 4.3, it is further possible to show that identification of λ_{Q_0} is also equivalent to the existence of a function $\nu \in L^1(P_{ZX})$ satisfying

$$\ell(T^*, X) = E_{\mu_{Z|X}}[\nu(Z, X)1\{T^*(Z) = t\}] \tag{27}$$

with μ-probability one. Such a result is again particularly helpful when μ is known, in which it is straightforward to asses whether λ_{Q_0} is identified (through (27)) and estimate the desired κ through the relation $\kappa(T, Z, X) = 1\{T = t\}\nu(Z, X)/\pi(Z, X)$.

4.4.1 Examples Revisited

We conclude our discussion on identification by revisiting Examples 2.4 and 2.5.

Example 2.4 (cont.) In this context, Corollary 4.4 implies that the expectation of a function with the structure $\rho(Y^*(t))\ell(T^*, X)$ is identified if and only if

$$\min_{\{s_j\}_{j=1}^q \subset \mathbf{R}} \sum_{i=1}^r (\ell(t_i^*, X) - \sum_{j=1}^q s_j 1\{t_i^*(z_j) = t\})^2 = 0 \tag{28}$$

with probability one (over X). Moreover, provided condition (28) holds, the expectation of $\rho(Y^*(t))\ell(T^*, X)$ is equal to the expectation of $\rho(Y)\kappa(T, Z, X)$ where

$$\kappa(T, z_j, X) = 1\{T = t\}\frac{s_j(X)}{P(Z = z_j|X)}$$

for any $(s_1(X), \ldots, s_q(X))$ minimizing (28). These results highlight that identifying a functional about outcomes reduces to a simple numerical problem when Z is discrete. ■

Example 2.5 (cont.) In this application, under conditions paralleling those in Corollary 4.2, Theorem 4.3 implies that the expectation of a function $\rho(Y^*(0))\ell(K_0^*, K_1^*, X)$

is identified if and only if there exists a sequence $\{\nu_j(C,W,X)\}$ satisfying

$$\lim_{j\to\infty} E_\mu[|\ell(K_0^*,K_1^*,X) - \sum_{c=0}^{1}\int_{K_c^*}^{\bar{\mathbf{w}}} \nu_j(c,w,X)dw|] = 0. \quad (29)$$

Moreover, provided condition (29) holds, the expectation of $\rho(Y^*(0))\ell(K_0^*,K_1^*,X)$ is identified as the limit of the expectations of $\rho(Y)\kappa_j(T,C,W,X)$ with

$$\kappa_j(t,c,w,X) \equiv \frac{1\{t=0\}\nu_j(c,w,X)}{f_{W|CX}(w|c,X)P(C=c|X)}.$$

More generally, our analysis yields that the expectation of $\rho(Y^*(t))\ell(K_0^*,K_1^*,X)$ is identified if and only if ℓ belongs to the $\|\cdot\|_{\mu,1}$-closure of the set $\mathcal{T}_t$, where

$$\mathcal{T}_0 \equiv \{f : f(K_0^*,K_1^*,X) = g_0(K_0^*,X) + g(K_1^*,X) \text{ with } g_0(\infty,X) = g_1(\infty,X) = 0\}$$
$$\mathcal{T}_1 \equiv \{f : f(K_0^*,K_1^*,X) = g_0(K_0^*,X) + g(K_1^*,X) \text{ with } g_0(\underline{\mathbf{w}},X) = g_1(\underline{\mathbf{w}},X) = 0\}.$$

For instance, setting $\ell(K_0^*,K_1^*) = 1\{K_0^*\leq a_0, K_1^* > a_1\}$ with $a_0,a_1 \in [\underline{\mathbf{w}},\bar{\mathbf{w}}]\cup\{\infty\}$ we can conclude that the expectation of $\rho(Y^*(t))1\{K_0^*\leq a_0, K_1^*>a_1\}$ is identified for both $t=0$ and $t=1$ if and only if $a_0 \geq a_1$. In particular, it follows that parameters such as

$$E_{Q_0}[Y^*(1)-Y^*(0)|K_0^*\leq a_0, K_1 = a_1] \text{ and } E_{Q_0}[Y^*(1)-Y^*(0)|K_0^* = a_0, K_1 > a_1] \quad (30)$$

are identified for any points a_0,a_1 satisfying $\underline{\mathbf{w}} \leq a_1 \leq a_0 < \infty$. ■

5 Estimation

In our analysis so far, we have allowed features of our model (e.g., the measure μ and functions $\{\ell_j\}$) to depend on the distribution P of the data. To construct an estimator, however, we need to incorporate additional information on the exact manner in which these features depend on P. For concreteness, in what follows we therefore focus on a leading special case in which μ and $\{\ell_j\}$ are known instead of identified – a setting that encompasses the majority of our examples in Section 2.1.

Our estimation strategy is based on two observations that follow from our identification analysis. First, if the parameter of interest is point identified, then it must equal the limit of expectations of a sequence of unknown functions $\{\kappa_j\}$. Second, provided $\{\ell_j\}$ and μ are known, the functions $\{\kappa_j\}$ can often be set to equal $\kappa_j = \nu_j/\pi$ for known $\{\nu_j\}$ and $\pi = dP_{Z|X}/d\mu_{Z|X}$; see, e.g., Corollaries 4.2, 4.3, and 4.4. These observations enable us to devise double robust identifying moment conditions that readily yield asymptotically normal estimators. We next construct such estimators for functionals about types and about outcomes and characterize their semiparametric efficiency bound.

5.1 Estimation: Types

Recall that functionals about types, as studied in Section 4.3, have the structure

$$\lambda_Q = \lim_{j\to\infty} E_Q[\ell_j(T^\star, X)]. \tag{31}$$

By Theorem 4.2, if λ_{Q_0} is point identified, then it must equal the limit of the expectation of functions $\{\kappa_j\}$ of (T, Z, X). Moreover, in an important class of applications, the functions $\{\kappa_j\}$ satisfy $\kappa_j = \nu_j/\pi$ for some known functions $\{\nu_j\}$ of (T, Z, X).

Our estimator is based on the observation that the structure $\kappa_j = \nu_j/\pi$ with $\pi = dP_{Z|X}/d\mu_{Z|X}$ implies that for any $t \in \mathbf{T}$ and function f we have the equality

$$E_P[\kappa_j(t, Z, X)f(Z, X)] = E_{P_X}[E_{\mu_{Z|X}}[\nu_j(t, Z, X)f(Z, X)]].$$

Therefore, we may equivalently express the expectation of $\kappa_j(T, Z, X)$ as being equal to

$$\begin{aligned} E_P[\kappa_j(T, Z, X)] = \sum_{t\in\mathbf{T}} & E_P[\kappa_j(t, Z, X)(1\{T = t\} - P(T = t|Z, X))] \\ & + \sum_{t\in\mathbf{T}} E_{P_X}[E_{\mu_{Z|X}}[\nu_j(t, Z, X)P(T = t|Z, X)]]. \end{aligned} \tag{32}$$

Crucially, the identifying moment in (32) is double robust in the sense that the equality continues to hold if for any $t \in \mathbf{T}$ we substitute either of the nuisance parameters $\kappa_j(t, Z, X)$ or $P(T = t|Z, X)$ with different functions of (Z, X). This double robustness readily enables estimation through a variety of plug-in machine learning methods. For concreteness, we follow ideas in Smucler et al. (2019), Chernozhukov et al. (2022a), and Chernozhukov et al. (2022c) and employ an ℓ_1-regularized double robust estimator.

Specifically, our estimator is obtained from the following algorithm:

STEP 1. Partition $\{1, \ldots, n\}$ into K subsets $\{I_k\}_{k=1}^K$, select functions $\{b_l\}_{l=1}^p$ of (Z, X) with p potentially larger than n, and let $b(Z, X) \equiv (b_1(Z, X), \ldots, b_p(Z, X))'$. The number of partitions K is fixed with n, and usually set to five or ten. ■

STEP 2. For each treatment value $t \in \mathbf{T}$ and partition k compute the estimators

$$\hat\beta_{t,k} \in \arg\min_{\beta\in\mathbf{R}^p} \sum_{i\in I_k^c} (1\{T_i = t\} - b(Z_i, X_i)'\beta)^2 + \alpha\|\beta\|_1 \tag{33}$$

$$\hat\gamma_{t,k} \in \arg\min_{\gamma\in\mathbf{R}^p} \sum_{i\in I_k^c} \frac{1}{2}(b(Z_i, X_i)'\gamma)^2 - E_{\mu_{Z|X}}[\nu_j(t, Z, X_i)b(Z, X_i)'\gamma] + \alpha\|\gamma\|_1, \tag{34}$$

where $I_k^c = \{1, \ldots, n\} \setminus I_k$. We note that the penalty α need not be the same in both estimation problems, but the set of functions $\{b_l\}_{l=1}^p$ must be the same. The penalty α can be selected in a data-drive way such as, e.g., cross-validation. ■

Step 3. For each k, let $|I_k|$ denote the number of observations in I_k and set $\hat{\lambda}_k$ to equal

$$\hat{\lambda}_k \equiv \frac{1}{|I_k|}\sum_{i\in I_k}\sum_{t\in\mathbf{T}} b(Z_i,X_i)'\hat{\gamma}_{t,k}(1\{T_i=t\}-b(Z_i,X_i)'\hat{\beta}_{t,k})+E_{\mu_{Z|X}}[\nu_j(t,Z,X_i)b(Z,X_i)'\hat{\beta}_{t,k}].$$

Note that in computing the estimator $\hat{\lambda}_k$ we employ estimators $\hat{\gamma}_{t,k}$ and $\hat{\beta}_{t,k}$ that are obtained from data not in partition I_k (see Step 2). ∎

Step 4. The estimator for λ_{Q_0} is given by $\hat{\lambda} \equiv \sum_k \hat{\lambda}_k|I_k|/n$ – i.e. $\hat{\lambda}$ is simply the weighted average of the estimators $\{\hat{\lambda}_k\}_{k=1}^K$ obtained from each partition I_k. ∎

Intuitively, we may view $b(Z,X)'\hat{\beta}_{t,k}$ and $b(Z,X)'\hat{\gamma}_{t,k}$ as estimators for the nuisance parameters $P(T=t|Z,X)$ and $\kappa_j(t,Z,X)$ and $\hat{\lambda}$ as a plug-in estimator based on (32). The sample splitting in Step 1 is important for relaxing our assumptions, though we note $\hat{\lambda}$ will remain asymptotically normal without sample splitting provided we impose sufficiently strong sparsity requirements. We also note that we may substitute $b(Z,X)'\hat{\beta}_{t,k}$ with certain nonlinear estimators, such as logistic regression, and still obtain a double robust estimator for λ_{Q_0} provided $b(Z,X)'\hat{\gamma}_{t,k}$ is modified accordingly as well (Smucler et al., 2019; Chernozhukov et al., 2022a). Alternatively, in Step 3 we may substitute $b(Z,X)'\hat{\beta}_{t,k}$ and $b(Z,X)'\hat{\gamma}_{t,k}$ with any suitably convergent machine learning estimators for the nuisance parameters (Chernozhukov et al., 2018). The resulting estimator for λ_{Q_0}, however, may fail to be double robust in the sense that inference based on it can be invalid if any of the nuisance parameter estimators is inconsistent.

In order to state sufficient conditions for the asymptotic normality of our estimator $\hat{\lambda}$, we first need to introduce some additional notation. To this end, we define

$$\beta_t \in \arg\min_{\beta\in\mathbf{R}^p} E_P[(1\{T=t\}-b(Z,X)'\beta)^2]$$
$$\gamma_t \in \arg\min_{\gamma\in\mathbf{R}^p}\{\frac{1}{2}E_P[(b(Z,X)'\gamma)^2]-E_{P_X}[E_{\mu_{Z|X}}[\nu_j(t,Z,X)b(Z,X)'\gamma]\},$$

which are the estimands for which $\hat{\beta}_{t,k}$ and $\hat{\gamma}_{t,k}$ will be assumed to be consistent for. We additionally denote the estimation error for β_t and γ_t in the prediction norm by

$$r_t^\beta \equiv \max_{1\leq k\leq K}\{E_P[(b(Z,X)'(\hat{\beta}_{t,k}-\beta_t))^2]\}^{1/2} \qquad r_t^\gamma \equiv \max_{1\leq k\leq K}\{E_P[(b(Z,X)'(\hat{\gamma}_{t,k}-\gamma_t))^2]\}^{1/2}.$$

The estimands $b(Z,X)'\beta_t$ and $b(Z,X)'\gamma_t$ are approximations to the nuisance parameters $P(T=t|Z,X)$ and $\kappa_j(t,Z,X)$, and we denote their approximation errors by

$$\delta_t^\beta \equiv \{E_P[(P(T=t|Z,X)-b(Z,X)'\beta_t)^2]\}^{1/2}$$
$$\delta_t^\gamma \equiv \{E_P[(\kappa_j(t,Z,X)-b(Z,X)'\gamma_t)^2]\}^{1/2}.$$

Finally, it will be convenient to denote the influence function of our estimator $\hat{\lambda}$ by

$$\psi(T,Z,X) \equiv \sum_{t\in\mathbf{T}} b(Z,X)'\gamma_t(1\{T=t\} - b(Z,X)'\beta_t) + \sum_{t\in\mathbf{T}} E_{\mu_{Z|X}}[\nu_j(t,Z,X)b(Z,X)'\beta_t] - \lambda_{Q_0}, \quad (35)$$

and to let $\sigma^2 \equiv \mathrm{Var}_P\{\psi(T,Z,X)\}$ denote its variance. While we have suppressed it from the notation, it is important to note that p (the dimension of $b(Z,X)$) and j (as indexing κ_j) can depend on n, and as a result so do all the terms we have defined.

Given the introduced notation, we impose the following assumptions:

Assumption 5.1. *(i) $\{Y_i,T_i,X_i,Z_i\}_{i=1}^n$ is i.i.d.; (ii) There are known $\{\nu_j\} \subseteq L^\infty(P_{TZX})$ such that $\kappa_j \equiv \nu_j/\pi$ satisfies $\Upsilon(\kappa_j) \xrightarrow{\mathcal{T}} \Lambda$; (iii) $\mu_{Z|X} \ll P_{Z|X}$ and $\|1/\pi\|_\infty < \infty$.*

Assumption 5.2. *(i) $\max_t \|b'\beta_t\|_\infty = O(1)$ and $B \equiv \max_t \|b'\gamma_t\|_\infty \vee \|\nu_j\|_\infty < \infty$ satisfies $B\log(n) = o(\sigma\sqrt{n})$; (ii) $r_t^\gamma \vee Br_t^\beta \vee \sqrt{n}r_t^\beta r_t^\gamma = o_P(\sigma)$ for all $t \in \mathbf{T}$; (iii) $\sqrt{n}\delta_t^\beta\delta_t^\gamma = o(\sigma)$ for all $t \in \mathbf{T}$; (iv) $\sqrt{n}|\lambda_{Q_0} - E_P[\kappa_j(T,Z,X)]| = o(\sigma)$; (v) $|I_k| \asymp n$.*

Assumption 5.1(ii) formalizes our conditions on κ_j which, by our identification analysis, is equivalent to the identification of λ_{Q_0} in a variety of applications. Assumption 5.1(iii) imposes that π be bounded away from zero. In turn, Assumption 5.2 states conditions on $\hat{\beta}_{t,k}$ and $\hat{\gamma}_{t,k}$ – we impose high level conditions given the preponderance of results in the literature justifying these assumptions under lower level assumptions. Specifically, Assumption 5.2(ii) demands that $\hat{\beta}_{t,k}$ and $\hat{\gamma}_{t,k}$ be suitably convergent to their respective estimands in the prediction norm. Sufficient conditions for deriving convergence rates for $\hat{\gamma}_{t,k}$ can be found in Chernozhukov et al. (2022c), and for $\hat{\beta}_k$ in Bühlmann and van De Geer (2011) and Bartlett et al. (2012) with and without sparsity assumptions respectively. Assumption 5.2(iii) states our rate requirements on the approximation errors δ_t^γ and δ_t^β. The rate is double robust in that Assumption 5.2(iii) can hold even if one of the estimands is not consistent for its corresponding nuisance parameter. Finally, Assumption 5.2(iv) is automatically satisfied if $\{\kappa_j\}$ does not depend on j (as in Lemma 4.2) and may be viewed as an undersmoothing requirement otherwise.

Remark 5.1. Sufficient conditions for Assumption 5.2(iv) can be analytically derived in certain applications in which κ_j depends on j; see, e.g., our discussion of Example 2.5 below. Alternatively, a numerical bound can be obtained through the inequality

$$|E_{Q_0}[\ell(T^*,X) - \kappa_j(T,Z,X)]| \leq \|\frac{dQ_0}{d\mu}\|_\infty \times E_\mu[|\ell(T^*,X) - \sum_{t\in\mathbf{T}} E_{\mu_{Z|X}}[\nu_j(t,Z,X)1\{T^*(Z)=t\}]|].$$

In particular, if ν_j is computed through, e.g., Corollary 4.2 then we may set it to control the bias in Assumption 2.2(iii) given a sup-norm bound on $dQ_0/d\mu$. ■

Our next result establishes the asymptotic normality of our estimator.

Theorem 5.1. *Let Assumptions 2.1, 2.2, 4.1, 5.1, 5.2 hold, λ_Q and ψ be as defined in (31) and (35), and $\sigma^2 \equiv \mathrm{Var}_P\{\psi(T,Z,X)\}$. Then, there is a $\mathbb{Z} \sim N(0,1)$ satisfying*

$$\frac{\sqrt{n}}{\sigma}(\hat{\lambda} - \lambda_{Q_0}) = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}\psi(T_i, Z_i, X_i) + o_P(1) = \mathbb{Z} + o_P(1). \tag{36}$$

For inference, we will rely on a multiplier bootstrap procedure that approximates the distribution in Theorem 5.1 and further extends to vector valued parameters and their nonlinear functionals. Specifically, for each k we define an estimator for ψ by setting

$$\hat{\psi}_k(T,Z,X) \equiv \sum_{t\in\mathbf{T}} b(Z,X)'\hat{\gamma}_{t,k}(1\{T=t\} - b(Z,X)'\hat{\beta}_{t,k}) + \sum_{t\in\mathbf{T}} E_{\mu_{Z|X}}[\nu_j(t,Z,X)b(Z,X)'\hat{\beta}_{t,k}] - \hat{\lambda}. \tag{37}$$

Our "bootstrapped" estimator $\hat{\lambda}^*$ is then obtained by employing $\hat{\psi}_k$ and an i.i.d. sample $\{W_i\}_{i=1}^n$ of standard normal weights independent of the data to perturb $\hat{\lambda}$ according to

$$\hat{\lambda}^* \equiv \hat{\lambda} + \frac{1}{n}\sum_{k=1}^{K}\sum_{i\in I_k} W_i\hat{\psi}_k(T_i, Z_i, X_i). \tag{38}$$

We employ standard normal weights W to simplify our technical arguments, though under appropriate moment restrictions the proposed bootstrap remains valid provided W satisfies $E[W] = 0$ and $E[W^2] = 1$ – e.g., for W set to be Rademacher weights.

The next result establishes the validity of the proposed bootstrap.

Theorem 5.2. *Let the conditions of Theorem 5.1 hold and $\{W_i\}_{i=1}^n$ be i.i.d. standard normal random variables independent of $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$. Then, there exists a standard normal random variable $\mathbb{Z}^*$ independent of $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$ and satisfying*

$$\frac{\sqrt{n}}{\sigma}(\hat{\lambda}^* - \hat{\lambda}) = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} W_i\psi(T_i, Z_i, X_i) + o_P(1) = \mathbb{Z}^* + o_P(1). \tag{39}$$

Theorems 5.1 and 5.2 justify employing the distribution of $(\hat{\lambda}^* - \hat{\lambda})$ conditional on the data as an approximation to the distribution of $(\hat{\lambda} - \lambda_{Q_0})$. For instance, in order to obtain a two sided confidence region we would: (i) Draw $1 \le b \le B$ samples $\{W_i^{(b)}\}_{i=1}^n$ of the weights independently of the data; (ii) Employ each sample $\{W_i^{(b)}\}_{i=1}^n$ to obtain a bootstrap estimator $\hat{\lambda}^{*(b)}$ through (38); (iii) Compute the $1-\alpha$ quantile $\hat{c}_\alpha$ of $\{|\hat{\lambda}^{*(b)} - \hat{\lambda}|\}_{b=1}^B$; and (iv) Set the two sided confidence region to equal $\hat{\lambda} \pm \hat{c}_\alpha$.

5.1.1 Examples Revisited

We next illustrate our results in the context of Examples 2.4 and 2.5, focusing our discussion on the computation of the terms in our algorithm that are model specific.

Example 2.4 (cont.) Suppose ℓ is a known function of (T^*, X) and recall that we showed the expectation of $\ell(T^*, X)$ is identified if and only if with probability one

$$\min_{\{s_j\}_{j=1}^q \subset \mathbf{R}^d} \sum_{i=1}^r (\ell(t_i^*, X) - \sum_{j=1}^q s_j' \omega_j(t_i^*))^2 = 0, \tag{40}$$

where $\omega_j(t^*) \equiv (1\{t^*(z_j) = t_1\}, \ldots, 1\{t^*(z_j) = t_d\})$. To implement our estimator in this context, let $(s_1(X), \ldots, s_q(X))$ be a minimizer of (40) and $s_{jm}(X)$ denote the m^{th} coordinate of $s_j(X)$. It is then possible to show that ν_j does not depend on j and

$$E_{\mu_{Z|X}}[\nu(t_m, Z, X_i)b(Z, X_i)] = \sum_{j=1}^q s_{jm}(X_i)b(z_j, X_i).$$

This construction yields a double robust estimator of, e.g., $Q_0(T^* \in A)$ for any A for which the probability is identified (i.e. (40) holds with $\ell(t^*, X) = 1\{t^* \in A\}$). ■

Example 2.5 (cont.) We focus on discussing estimators for the expectation of a function ℓ of (K_c^*, X) for some $c \in \{0, 1\}$ – estimators for the expectation of a function of (K_0^*, K_1^*, X) then readily follow from our identification results (see (25)). To this end, suppose $\ell(K_c^*, X)$ is differentiable in K_c^* on $(\underline{\mathbf{w}}, \bar{\mathbf{w}})$ with derivative $\ell'(K_c^*, X)$ and that $\ell(K_c^*, X) = 0$ whenever $K_c^* \in \{\underline{\mathbf{w}}, \bar{\mathbf{w}}, +\infty\}$. It can then be shown that we may set

$$E_{\mu_{Z|X}}[\nu(1, C, W, X)b(C, W, X_i)] = \int_{\underline{\mathbf{w}}}^{\bar{\mathbf{w}}} \ell'(w, X_i)b(c, w, X_i)dw \tag{41}$$

and $\nu(0, C, W, X) = 0$ (see (24)). Expectations of more general functions of (K_c^*, X) can in turn be estimated by approximating them with differentiable functions. For example, the expectation of $\ell(K_c^*) = 1\{a \leq K_c^* \leq b\}$ with $\underline{\mathbf{w}} < a < b < \bar{\mathbf{w}}$ can be approximated by the expectation of $\ell_j(K_c^*) = F((b - K_c^*)/h_j) - F((a - K_c^*)/h_j)$ for some $h_j \downarrow 0$ and F the c.d.f. of a compactly supported mean zero continuous random variable. In this case, we may again set $\nu(0, C, W, X) = 0$ while (41) becomes

$$E_{\mu_{Z|X}}[\nu_j(1, C, W, X)b(C, W, X_i)] = \int_{\underline{\mathbf{w}}}^{\bar{\mathbf{w}}} \frac{1}{h_j}(F'(\frac{a-w}{h_j}) - F'(\frac{b-w}{h_j}))b(c, w, X_i)dw \tag{42}$$

and, under regularity conditions, $B \asymp \sigma^2 \asymp 1/h_j$ and $|\lambda_{Q_0} - E_P[\kappa_j(T, Z, X)]| = O(h^2)$ so that Assumption 5.2 requires us to set $\log^2(n)/(nh_j) = o(1)$ and $nh_j^5 = o(1)$. ■

5.2 Estimation: Outcomes

We next turn to developing an estimator for functionals about outcomes, as studied in Section 4.4. Recall that these functionals are characterized by having the structure

$$\lambda_Q = \lim_{j\to\infty} E_Q[\rho(Y^*(t))\ell_j(T^\star, X)] \tag{43}$$

for some known ρ and $t \in \mathbf{T}$. If λ_{Q_0} is identified, then by Theorem 4.3 it must equal the limit of the expectation of $\{\rho\kappa_j\}$ for some sequence of functions $\{\kappa_j\}$ of (T, Z, X). While the functions $\{\kappa_j\}$ are unknown, in a leading set of applications they satisfy

$$\kappa_j(T, Z, X) = 1\{T = t\}\frac{\nu_j(Z, X)}{\pi(Z, X)} \tag{44}$$

for known functions $\{\nu_j\}$; see, e.g., Corollary 4.4 and subsequent discussion. Due to the similarities between the identifying equations for functionals about types and outcomes, we are able to obtain estimators for functionals about outcomes by slightly modifying our preceding analysis for types. As a result, in what follows we keep exposition brief though note that the discussion and remarks of Section 5.1 apply to this section as well.

Our estimator for functionals about outcomes is obtained though the algorithm:

STEP 1. Partition $\{1, \ldots, n\}$ into K subsets $\{I_k\}_{k=1}^K$, select a set of functions $\{b_l\}_{l=1}^p$ of (Z, X), and let $b(Z, X) \equiv (b_1(Z, X), \ldots, b_p(Z, X))'$. ■

STEP 2. For each partition $1 \le k \le K$ compute the following two estimators

$$\hat\beta_k \in \arg\min_{\beta\in\mathbf{R}^q} \sum_{i\in I_k^c} (\rho(Y_i)1\{T_i = t\} - b(Z_i, X_i)'\beta)^2 + \alpha\|\beta\|_1 \tag{45}$$

$$\hat\gamma_k \in \arg\min_{\gamma\in\mathbf{R}^q} \sum_{i\in I_k^c} \{\frac{1}{2}(b(Z_i, X_i)'\gamma)^2 - E_{\mu_{Z|X}}[\nu_j(Z, X_i)b(Z, X_i)'\gamma]\} + \alpha\|\gamma\|_1, \tag{46}$$

where the set of functions $\{b_l\}_{l=1}^p$ must be the same in both estimation problems. ■

STEP 3. For each partition $1 \le k \le K$ compute the plug-in estimator $\hat\lambda_k$ given by

$$\hat\lambda_k \equiv \frac{1}{|I_k|}\sum_{i\in I_k} b(Z_i, X_i)'\hat\gamma_k(\rho(Y_i)1\{T_i = t\} - b(Z_i, X_i)'\hat\beta_k) + E_{\mu_{Z|X}}[\nu_j(Z, X_i)b(Z, X_i)'\hat\beta_k]$$

where $|I_k|$ denotes number of observations in the partition I_k. ■

STEP 4. Compute $\hat\lambda \equiv \sum_k \hat\lambda_k|I_k|/n$ as the final estimator for λ_{Q_0}. ■

The asymptotic properties of $\hat\lambda$ can unsurprisingly be established under similar conditions to those employed in Section 5.1. Adjusting notation, we now define the estimands

$$\beta \equiv \arg\min_{\beta\in\mathbf{R}^p} E_P[(\rho(Y)1\{T=t\} - b(Z,X)'\beta)^2]$$
$$\gamma \equiv \arg\min_{\gamma\in\mathbf{R}^p} \{\frac{1}{2}E_P[(b(Z,X)'\gamma)^2] - E_{P_X}[E_{\mu_{Z|X}}[\nu_j(Z,X)b(Z,X)'\gamma]]\}$$

and denote the convergence rates for $\hat{\beta}_k$ and $\hat{\gamma}_k$ to β and γ in the prediction norm by

$$r^\beta \equiv \max_{1\le k\le K}\{E_P[(b(Z,X)'(\hat{\beta}_k - \beta))^2]\}^{1/2} \qquad r^\gamma \equiv \max_{1\le k\le K}\{E_P[(b(Z,X)'(\hat{\gamma}_k - \gamma))^2]\}^{1/2}.$$

The functions $b(Z,X)'\beta$ and $b(Z,X)'\gamma$ represent approximations to $E[\rho(Y)1\{T=t\}|Z,X]$ and $\nu_j(Z,X)/\pi(Z,X)$ respectively, and we denote their approximation errors by

$$\delta^\beta \equiv \{E_P[(E_P[\rho(Y)1\{T=t\}|Z,X] - b(Z,X)'\beta)^2]\}^{1/2}$$
$$\delta^\gamma \equiv \{E_P[(\frac{\nu_j(Z,X)}{\pi(Z,X)} - b(Z,X)'\gamma)^2]\}^{1/2}.$$

Finally, we introduce the influence function for our estimator, which here is given by

$$\psi(Y,T,Z,X) \equiv b(Z,X)'\gamma(1\{T=t\}\rho(Y) - b(Z,X)'\beta) + E_{\mu_{Z|X}}[\nu_j(Z,X)b(Z,X)'\beta] - \lambda_{Q_0}, \quad (47)$$

and set $\sigma^2 \equiv \text{Var}_P\{\psi(Y,T,Z,X)\}$. We again note that the introduced parameters are allowed to depend on n, though we suppressed such dependence from the notation.

The following assumptions suffice for estalibshing the asymptotic properties of $\hat{\lambda}$.

Assumption 5.3. *(i) $\{Y_i,T_i,X_i,Z_i\}_{i=1}^n$ is i.i.d.; (ii) There are known $\{\nu_j\} \subseteq L^\infty(P_{ZX})$ such that κ_j given by (44) satisfies $\Upsilon(\kappa_j) \xrightarrow{\mathcal{T}} \Lambda$; (iii) $\mu_{Z|X} \ll P_{Z|X}$ and $\|1/\pi\|_\infty < \infty$.*

Assumption 5.4. *(i) $\|\rho\|_\infty < \infty$, $\|b'\beta\|_\infty = O(1)$, and $B \equiv \|b'\gamma\|_\infty \vee \|\nu_j\|_\infty < \infty$ satisfies $B\log(n) = o(\sigma\sqrt{n})$; (ii) $r^\gamma \vee Br^\beta \vee \sqrt{n}r^\beta r^\gamma = o_P(\sigma)$; (iii) $\sqrt{n}\delta^\beta\delta^\gamma = o(\sigma)$; (iv) $\sqrt{n}|\lambda_{Q_0} - E_P[\rho(Y)\kappa_j(T,Z,X)]| = o(\sigma)$; (v) $|I_k| \asymp n$.*

Assumptions 5.3 and 5.4 are simply adaptations of Assumptions 5.1 and 5.2 to the present estimation problem. The most substantive difference between these sets of assumptions is that Assumption 5.4(i) requires ρ to be bounded – a condition that enables us to establish our results employing convergence rates in the prediction norm. While we impose this requirement for simplicity, we note that it may be relaxed by strengthening the norm under which we require $\hat{\beta}_k$ and $\hat{\gamma}_k$ to converge to β and γ.

The next result establishes the asymptotic normality of our estimator.

Theorem 5.3. *Let Assumptions 2.1, 2.2, 4.1, 5.3, 5.4 hold, λ_Q and ψ be as defined in*

(43) *and* (47)*, and* $\sigma^2 \equiv \mathrm{Var}_P\{\psi(Y,T,Z,X)\}$*. Then, there is a* $\mathbb{Z} \sim N(0,1)$ *satisfying*

$$\frac{\sqrt{n}}{\sigma}(\hat{\lambda} - \lambda_{Q_0}) = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}\psi(Y_i,T_i,Z_i,X_i) + o_P(1) = \mathbb{Z} + o_P(1). \quad (48)$$

For inference we again rely on the multiplier bootstrap. Specifically, for each $1 \leq k \leq K$ in our partition we define an estimator for the influence function by setting

$$\hat{\psi}_k(Y,T,Z,X) \equiv b(Z,X)'\hat{\gamma}_k(1\{T=t\}\rho(Y) - b(Z,X)'\hat{\beta}_k) + E_{\mu_{Z|X}}[\nu_j(Z,X)b(Z,X)'\hat{\beta}_k] - \hat{\lambda}. \quad (49)$$

For $\{W_i\}_{i=1}^n$ an i.i.d. sample of standard normal random variables independent of the data, we then obtain a "bootstrapped" analogue $\hat{\lambda}^*$ to $\hat{\lambda}$ by setting

$$\hat{\lambda}^* \equiv \hat{\lambda} + \frac{1}{n}\sum_{k=1}^{K}\sum_{i\in I_k} W_i\hat{\psi}_k(Y_i,T_i,Z_i,X_i).$$

Our next result establishes the validity of the proposed bootstrap procedure.

Theorem 5.4. *Let the conditions of Theorem 5.3 hold and* $\{W_i\}_{i=1}^n$ *be i.i.d. standard normal random variables independent of* $\{Y_i,T_i,Z_i,X_i\}_{i=1}^n$*. Then, there exists a standard normal random variable* $\mathbb{Z}^*$ *independent of* $\{Y_i,T_i,Z_i,X_i\}_{i=1}^n$ *and satisfying*

$$\frac{\sqrt{n}}{\sigma}(\hat{\lambda}^* - \hat{\lambda}) = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}W_i\psi(Y_i,T_i,Z_i,X_i) + o_P(1) = \mathbb{Z}^* + o_P(1). \quad (50)$$

Theorems 5.3 and 5.4 justify employing the proposed bootstrap to conduct inference on functionals about outcomes. Moreover, together with Theorems 5.1, 5.2, and the Delta method, they also justify employing the bootstrap to conduct inference on parameters such as, e.g., conditional expectations of potential outcomes given types and of types given covariates.[12] Specifically, such parameters have the structure

$$F(\lambda_{Q_01},\ldots,\lambda_{Q_0q})$$

where $F : \mathbf{R}^q \to \mathbf{R}$ is a known differentiable function and each $\lambda_{Q_0j} \in \mathbf{R}$ is a functional about types or outcomes.[13] For instance, to obtain a two sided confidence region we would: (i) Compute estimators $(\hat{\lambda}_1,\ldots,\hat{\lambda}_q)$ for $(\lambda_{Q_01},\ldots,\lambda_{Q_0q})$ using our results for types or outcomes; (ii) Draw B samples $\{W_i^{(b)}\}_{i=1}^n$ of weights independent of the data; (iii) Employ each sample $\{W_i^{(b)}\}_{i=1}^n$ to obtain bootstrap estimators $(\hat{\lambda}_1^{*(b)},\ldots,\hat{\lambda}_q^{*(b)})$ using our results for types or outcomes; (iv) Set $\hat{c}_\alpha$ to equal the $1-\alpha$

[12] See Lemma A.10 in the Appendix for a version of the Delta method suitable for our setting.

[13] E.g., for some event A set $\lambda_{Q_01} = E_{Q_0}[Y^*(t)1\{T^* \in A\}]$, $\lambda_{Q_02} = E_{Q_0}[1\{T^* \in A\}]$, and $F(\lambda_{Q_01},\lambda_{Q_02}) = \lambda_{Q_01}/\lambda_{Q_02}$ to obtain $F(\lambda_{Q_01}/\lambda_{Q_02}) = E_{Q_0}[Y^*(t)|T^* \in A]$.

quantile of $\{|F(\hat{\lambda}_1, \ldots, \hat{\lambda}_q) - F(\hat{\lambda}_1^{*(b)}, \ldots, \hat{\lambda}_q^{*(b)})|\}_{b=1}^B$ across the B samples; and (v) Report $F(\hat{\lambda}_1, \ldots, \hat{\lambda}_q) \pm \hat{c}_\alpha$ as a two sided confidence region. Similarly, our results also allow us to conduct inference on directionally (but not fully) differentiable functionals of $(\lambda_{Q_0 1}, \ldots, \lambda_{Q_0 q})$ by relying on the framework developed in Fang and Santos (2018).

5.2.1 Examples Revisited

Example 2.4 (cont.) We previously established that, for a known function ℓ of (T^*, X), the expectation of $\rho(Y^*(t))\ell(T^*, X)$ is identified if and only if

$$\min_{\{s_j\}_{j=1}^q \subset \mathbf{R}} \sum_{i=1}^r (\ell(t_i^*, X) - \sum_{j=1}^q s_j 1\{t_i^*(z_j) = t\})^2 = 0 \tag{51}$$

with probability one (over X). In order to estimate an identified functional about outcomes (i.e. one for which (51) holds), we may implement our estimator with

$$E_{\mu_{Z|X}}[\nu(Z, X_i) b(Z, X_i)] = \sum_{j=1}^q s_j(X_i) b(z_j, X_i),$$

where $(s_1(X), \ldots, s_q(X))$ is any minimizer of (51). Hence, we may for example conduct inference on $E_{Q_0}[Y^*(t)1\{T^* \in A\}]$ (provided $\ell(t^*, X) = 1\{t^* \in A\}$ satisfies (51)) or, in combination with our results on functionals about types, on $E_{Q_0}[Y^*(t)|T^* \in A]$. ■

Example 2.5 (cont.) When illustrating the implementation of our estimator for functionals about types in this example we employed functions κ_j with the structure $\kappa_j(T, Z, X) = 1\{T = 1\}\nu_j(Z, X)/\pi(Z, X)$. Hence, the same ν_j can be employed to estimate functionals about $Y^*(1)$ – e.g., to estimate $E_{Q_0}[\rho(Y^*(1))1\{a \leq K_c^* \leq b\}]$ we may employ (42). Similarly, to estimate $E_{Q_0}[\rho(Y^*(0))1\{a \leq K_c^* \leq b\}]$ we may set

$$E_{\mu_{Z|X}}[\nu_j(C, W, X_i) b(C, W, X_i)] = \int_{\underline{\mathbf{w}}}^{\bar{\mathbf{w}}} \frac{1}{h_j}(F'(\frac{b - w}{h_j}) - F'(\frac{a - w}{h_j})) b(c, w, X_i) dw,$$

and by combining estimators we may conduct inference on average treatment effects for individuals with $K_c^* \in [a, b]$. More generally, our results enable us to conduct inference on average treatment effects for groups determined by (K_0^*, K_1^*) as in, e.g., (30). ■

5.3 Efficiency Bound

We conclude this section by deriving the semiparametric efficiency bound for the estimation problems studied in Sections 5.1 and 5.2 that required μ to be known (instead of identified). To this end, we first introduce a series of definitions that are standard in the literature on semiparametric efficiency (Bickel et al., 1993).

Definition 5.1. A *path* $\eta \mapsto Q_{\eta,g}$ is a function defined on $[0,1)$ such that $Q_{\eta,g}$ is a probability distribution on $\mathbf{Y}^\star \times \mathbf{T}^\star \times \mathbf{Z} \times \mathbf{X}$ satisfying $Q_{\eta,g} \ll \mu$ for every η and

$$\lim_{\eta \to 0} \int (\frac{1}{\eta}(\frac{dQ_{\eta,g}^{1/2}}{d\mu} - \frac{dQ_{0,g}^{1/2}}{d\mu}) - \frac{1}{2} g \frac{dQ_{0,g}^{1/2}}{d\mu})^2 d\mu = 0. \tag{52}$$

The function $g \in L^2(Q_{0,g})$ is called the *score* of the path $\eta \mapsto Q_{\eta,g}$. ■

Definition 5.2. We say that a path $\eta \mapsto Q_{\eta,g}$ is a *submodel* if: (i) $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under $Q_{\eta,g}$ for all $\eta \in [0,1)$, and (ii) $Q_{0,g} \in \Theta_0$. ■

A *path* is simply a "smooth" one dimensional parametrization of distributions for random variables $(Y^\star, T^\star, Z, X)$. We emphasize that in Definition 5.1 we are relying on the fact that μ is known and is therefore fixed along the path. A *submodel* is a path that in addition: (i) Satisfies the requirements of our model – i.e. $Q_{\eta,g} \ll \mu$ and $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under $Q_{\eta,g}$; and (ii) Induces the distribution P on (Y, T, Z, X) at $\eta = 0$ – i.e. $Q_{0,g}$ is observationally equivalent to Q_0. We note that we do not require that the path satisfy Assumption 2.2(iii). In this regard, our analysis concerns applications in which the conditions encoded in $\mathcal{Q}$ are not informative or we do not want to use such information in estimation. In applications in which $\mathcal{Q}$ encodes regularity conditions, such as in the majority of the examples in Section 2.1, it is often possible to establish that Assumption 4.1(iii) implies that Assumption 2.2(iii) is uninformative, though such arguments rely on the specific choice of $\mathcal{Q}$.

For any η, a distribution $Q_{\eta,g}$ for $(Y^\star, T^\star, Z, X)$ induces a distribution for (Y, T, Z, X) through the relation $(Y, T, Z, X) = (Y^\star(T), T^\star(Z), Z, X)$. As a result, each submodel $\eta \mapsto Q_{\eta,g}$ induces a path $\eta \mapsto P_{\eta,s}$ of probability distributions for (Y, T, Z, X) with a score that we denote by s – i.e. the map $\eta \mapsto P_{\eta,s}$ satisfies smoothness requirements analogous to those imposed in (52) and by construction $P_{0,s} = P$ (Le Cam and Yang, 1988). The resulting set of scores s that can be produced in this manner generate the so-called tangent space for our model, which plays a crucial role in characterizing semiparametric efficiency bounds – see Theorem A.1 in the appendix for a characterization of the tangent space that may be of independent interest.

Remark 5.2. By construction, every path $\eta \mapsto P_{\eta,s}$ we consider satisfies the restrictions of our model. This approach contrasts with, for instance, Frölich (2007) who does not impose that $\eta \mapsto P_{\eta,s}$ be generated by an underlying path $\eta \mapsto Q_{\eta,g}$ satisfying the restrictions of the model. Nonetheless, the efficiency bound of Frölich (2007) is correct because in the model he examines P is just identified in the sense of Chen and Santos (2018). We emphasize, however, than in models in which P is overidentified, neglecting to impose the restrictions of the model can lead to incorrect efficiency bounds. ■

Our first result derives the semiparametric efficiency bound for estimating λ_{Q_0} when

$$\lambda_Q = E_Q[\ell(Y^\star, T^\star, X)] \tag{53}$$

with ℓ a known function. Following Bickel et al. (1993), for any submodel $\eta \mapsto Q_{\eta,g}$ inducing a path $\eta \mapsto P_{\eta,s}$ we define the *information bound* for estimating λ_{Q_0} by

$$I^{-1}(Q_{\cdot,g}) \equiv \{\frac{\partial}{\partial\eta}\lambda_{Q_{\eta,g}}\Big|_{\eta=0}\}^2 \times \{E_P[s^2(Y,T,Z,X)]\}^{-1}. \tag{54}$$

Intuitively, the information bound $I^{-1}(Q_{\cdot,g})$ is the asymptotic variance of the maximum likelihood estimator for λ_{Q_0} in the parametric submodel $\eta \mapsto Q_{\eta,g}$. We note that in order for $I^{-1}(Q_{\cdot,g})$ to be well defined, $\eta \mapsto Q_{\eta,g}$ must be regular in the sense that it induces a path $\eta \mapsto P_{\eta,s}$ whose score has positive variance and hence has positive Fisher information. The semiparametric efficiency bound for estimating λ_{Q_0} is then defined as

$$I^{-1} \equiv \sup_{Q_{\cdot,g}} I^{-1}(Q_{\cdot,g}), \tag{55}$$

where the supremum is taken over all submodels $\eta \mapsto Q_{\eta,g}$ for which $I^{-1}(Q_{\cdot,g})$ is well defined (i.e. the Fisher information of the submodel is positive).

As a final of notation we introduce a map $\mathcal{I}$ mapping functions s of the observables (Y,T,Z,X) to functions $\mathcal{I}(s)$ of (Y^*,T^*,X) by setting

$$\mathcal{I}(s) \equiv \sum_{t\in\mathbf{T}} E_{P_{Z|X}}[s(Y^\star(t),t,Z,X)1\{T^\star(Z)=t\}] - E_P[s(Y,T,Z,X)|X]. \tag{56}$$

The null space of $\mathcal{I}$, defined as $N(\mathcal{I}) \equiv \{s \in L^2(P) : \|\mathcal{I}(s)\|_{\bar{Q},2} = 0\}$, and its orthocomplement $[N(\mathcal{I})]^\perp \equiv \{s \in L^2(P) : \langle s, \tilde{s}\rangle_P = 0 \text{ for all } \tilde{s} \in N(\mathcal{I})\}$, play a crucial role in our next result characterizing the semiparametric efficiency bound for λ_{Q_0}.

Theorem 5.5. *Let Assumptions 2.1 and 2.2 hold, μ be known, $\lambda_Q \equiv E_Q[\ell(Y^*,T^*,X)]$ for some known bounded ℓ, and λ_{Q_0} be identified. Then the following hold:*

- *(i) Suppose $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ for some $\kappa \in L^2(P)$ and let φ denote the projection of κ onto $[N(\mathcal{I})]^\perp$. Then: $I^{-1} = \mathrm{Var}_P\{\varphi(Y,T,Z,X)\} + \mathrm{Var}_P\{E_P[\kappa(Y,T,Z,X)|X]\}$.*
- *(ii) Suppose Assumption 4.1(iii) holds and the projection of ℓ onto the $\|\cdot\|_{\bar{Q},2}$-closure of the range of $\Upsilon : L^2(P) \to L^2(\bar{Q})$ is bounded. If there is no $\kappa \in L^2(P)$ satisfying $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$, then it follows that $I^{-1} = \infty$.*

Theorem 5.5(i) characterizes the semiparametric efficiency bound for estimating λ_{Q_0} when: (i) ℓ is *known*, and (ii) There is a κ such that $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ – i.e., λ_{Q_0} falls within the scope of Lemma 4.2. By Theorem 4.1, we know that λ_{Q_0} may be identified even if there is no κ solving $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$. Subject to an additional regularity condition, however, Theorem 5.5(ii) establishes that such functionals have an infinite semiparametric efficiency bound[14] – a conclusion that is often interpreted as equivalent to the functional not being (regularly) estimable at the root-n rate (Chamberlain, 1986). In summary, we can conclude that: (i) Lemma 4.2 characterizes a set of functionals with a finite semiparametric efficiency bound, and (ii) Theorem 4.1 characterizes all additional functionals that are identified, though such functionals are not root-n estimable.

Theorem 5.5(i) both recovers previously available semiparametric efficiency bounds as special cases (Hahn, 1998; Frölich, 2007) and delivers new semiparametric efficiency bounds for multiple applications (e.g., Heckman and Vytlacil (1999) and Mogstad et al. (2021)). In turn, Theorem 5.5(ii) provides, to our knowledge, the first characterization of when causal parameters are not root-n estimable in these models. For our analysis, an important implication of Theorem 5.5 is its ability to assess whether our proposed estimators are efficient. The next corollary accomplishes this task by providing sufficient conditions for the estimators proposed in Sections 5.1 and 5.2 to be efficient.

Corollary 5.1. *Suppose the conditions of Theorem 5.1 (resp. Theorem 5.3) hold with* $\max_t \delta_t^\beta \vee \delta_t^\gamma = o(1)$ *(resp.* $\delta^\beta \vee \delta^\gamma = o(1)$*) and the conditions of Theorem 5.5(i) hold with a* κ *satisfying Assumption 5.1(ii) (resp. Assumption 5.3(ii)).*

(i) If $s = 0$ *is the only* $s \in L^2(P)$ *satisfying* $\|\Upsilon(s)\|_{\bar{Q},2} = 0$ *and* $E_P[s(Y,T,Z,X)|Z,X] = 0$*, then the estimator of Section 5.1 (resp. Section 5.2) attains the efficiency bound.*

(ii) Let Assumption 4.4(ii) hold, $\delta_t(T) \equiv 1\{T = t\}$ *and suppose, for any* $g \in L^2(P)$ *and* $t \in \mathbf{T}$*,* $\|\Upsilon(g\delta_t)\|_{\bar{Q},2} = 0$ *implies* $\|g\delta_t\|_{\bar{Q},2} = 0$*. If* $s \in L^2(P_{TZX})$ *satisfying* $\|\Upsilon(s)\|_{\bar{Q},2} = 0$ *implies that* $s \in L^2(P_{ZX})$*, then it follows that the estimator of Section 5.1 (resp. Section 5.2) attains the efficiency bound.*

Corollary 5.1(i) provides sufficient conditions for our estimators to be efficient by ensuring P is just identified in the sense of Chen and Santos (2018). In turn, Corollary 5.1(ii) imposes additional restrictions under which verifying whether our estimators are efficient reduces to a more stringent (hence easier to verify) condition than the one obtained in part (i). The requirements of Corollary 5.1 are easily verified in the examples of Section 2.1 to which our semiparametric efficiency analysis applies; see our discusion of Examples 2.4 and 2.5 below. Moreover, we note that Corollary 5.1 further implies that our estimators can be used to efficiently estimate parameters that are differentiable functions of multiple λ_{Q_0} with the structure in (53) (van der Vaart, 1991a). More generally, however, it is important to note that our estimators may fail to be efficient in applications in which P is overidentified in the sense of Chen and Santos (2018).

[14]Intuitively, the additional regularity condition enables us to show that ℓ belonging to the $\|\cdot\|_{\bar{Q},1}$ closure of $\Upsilon(L^1(P))$ implies ℓ also belongs to the $\|\cdot\|_{\bar{Q},2}$ closure of $\Upsilon(L^2(P))$.

5.3.1 Examples Revisited

Example 2.4 (cont.) In this context, Corollary 5.1(ii) can be used to show that the estimators of Section 5.1 and 5.2 are efficient provided that: (i) For all t, the matrix

$$\begin{pmatrix} 1\{t_1^*(z_1)=t\} & \dots & 1\{t_1^*(z_q)=t\} \\ \vdots & \ddots & \vdots \\ 1\{t_r^*(z_1)=t\} & \dots & 1\{t_r^*(z_q)=t\} \end{pmatrix} \tag{57}$$

has rank q; and (ii) Any function f of (T,Z) satisfying the system of equations

$$\sum_{i=1}^{d}\sum_{j=1}^{q} 1\{t_k^*(z_j)=t_i\}f(t_i,z_j)=0 \text{ for all } 1\le k\le r \tag{58}$$

must be such that $f(t,z)=f(t',z)$ for any $t\neq t'$ and any z. The second requirement may be verified analytically or numerically through a linear program. For instance, it is straightforward to analytically verify both requirements in the model of Kline and Walters (2016) and hence that our estimators are efficient in that application. ■

Example 2.5 (cont.) For this application, Corollary 5.1(ii) can be used to show that our estimators are efficient provided the support of K_0^* and K_1^* contains $[\underline{\mathbf{w}},\bar{\mathbf{w}}]$ – i.e. provided a marginal change in W always induces some individuals into treatment. We also note that in Section 5.2 we discussed estimation of parameters such as

$$E_{Q_0}[\rho(Y^*(t))\ell(K_c^*,X)] \tag{59}$$

and found our estimators to be root-n consistent when ℓ is differentiable in K_c^*, but slower than root-n consistent when we set ℓ to equal an indicator function. Theorem 5.5 provides an explanation for this difference, as it implies that (59) has a finite efficiency bound when ℓ is differentiable, and an infinite one when ℓ is an indicator function. ■

6 Conclusion

We proposed and developed a class of potential outcomes models that unifies and extends multiple identification strategies in the literature. By leveraging the rich structure of this class of models, we further derived widely applicable identification and estimation results. We believe that our findings will be valuable to researchers, both in the context of existing models and in the development of novel identification strategies.

Appendix

This Appendix contains the proofs for all the results stated in the paper. Throughout, we employ the notation Q_V to denote the marginal distribution of a random variable V under Q and $Q_{V|W}$ to denote the conditional distribution of V given W under Q. When employing $\bar{Q}^{\mathrm{it}}$ (as in Section 4.3) and $\bar{Q}^{\mathrm{io}}$ (as in Section 4.4), we implicitly assume the conditional distributions $Q_{Y^*|X}$ and $Q_{Y^*(t)|X}$ exist. Finally, distributions Q for (Y^*, T^*, Z, X) are assumed to be defined on a product σ-field generated by $\mathcal{F}_{Y^*} \times \mathcal{F}_{T^*} \times \mathcal{F}_Z \times \mathcal{F}_X$, where $\mathcal{F}_V$ denotes the σ-field on which Q_V is defined.

A.1 Proofs for Section 4

Proof of Lemma 4.1. We first establish the existence of the dominating measure $\bar{Q} \in \Theta_0$. To this end, first note that since μ is separable by Assumption 2.2(ii), Lemma 13.14 in Aliprantis and Border (2006) implies that $L^1(\mu)$ is separable under $\|\cdot\|_{\mu,1}$. Next set $D_0 \equiv \{dQ/d\mu : Q \in \Theta_0\}$ and note that Corollary 3.5 in Aliprantis and Border (2006), $D_0 \subset L^1(\mu)$, and $L^1(\mu)$ being separable imply D_0 is also separable under $\|\cdot\|_{\mu,1}$. Hence, there exists a countable set $\mathcal{D} \equiv \{Q_i\}_{i=1}^{\infty} \subseteq \Theta_0$ such that for any $Q \in \Theta_0$ and $\epsilon > 0$

$$\|\frac{dQ}{d\mu} - \frac{dQ_i}{d\mu}\|_{\mu,1} < \epsilon \tag{A.1}$$

for some $Q_i \in \mathcal{D}$. Next note that by Assumption 2.2(iii), $\mathcal{Q}$ is a closed convex subset of a Banach space $\mathbf{Q}$ with norm $\|\cdot\|_{\mathbf{Q}}$. For any $2 \le n < \infty$ then define

$$\lambda_{in} = \begin{cases} 1 - \sum_{i=2}^{n} \lambda_{in} & \text{if } i = 1 \\ 2^{-i}/\max\{1, \|dQ_i/d\mu\|_{\mathbf{Q}}\} & \text{if } 2 \le i \le n \\ 0 & \text{if } i > n \end{cases} \tag{A.2}$$

and note that $\sum_{i=1}^{n} \lambda_{in} = 1$ and $\lambda_{in} > 0$ for any $1 \le i \le n$ due to $\sum_{i=1}^{\infty} 2^{-i} = 1$. Therefore, since $\mathcal{Q}$ is convex by Assumption 2.2(iii), it follows that

$$f_n \equiv \sum_{i=1}^{n} \lambda_{in} \frac{dQ_i}{d\mu}$$

belongs to $\mathcal{Q}$ for all n. Moreover, for any $n < m$ the triangle inequality yields that

$$\begin{aligned} \|f_n - f_m\|_{\mathbf{Q}} &\le |\lambda_{1n} - \lambda_{1m}| \|\frac{dQ_1}{d\mu}\|_{\mathbf{Q}} + \sum_{i=n+1}^{m} \lambda_{im} \|\frac{dQ_i}{d\mu}\|_{\mathbf{Q}} \\ &\le (\lambda_{1n} - \lambda_{1m}) \|\frac{dQ_1}{d\mu}\|_{\mathbf{Q}} + \sum_{i=n+1}^{\infty} \frac{1}{2^i}, \end{aligned} \tag{A.3}$$

where we employed that $\lambda_{im}\|dQ_i/d\mu\|_{\mathbf{Q}} \le 2^{-i}$ and λ_{1n} is decreasing in n by (A.2).

Furthermore, since $1/2 < \lambda_{1n}$ by (A.2) and $\sum_{i=1}^{\infty} 2^{-i} = 1$, it follows from λ_{1n} being decreasing in n that the sequence $\{\lambda_{1n}\}_{n=1}^{\infty}$ has a limit in $\mathbf{R}$. Hence, by (A.3) we obtain that the sequence $\{f_n\}_{n=1}^{\infty}$ is Cauchy in $\mathbf{Q}$. By completeness of $\mathbf{Q}$, there therefore exists a $q_0 \in \mathbf{Q}$ such that $\|f_n - q_0\|_{\mathbf{Q}} = o(1)$ and since $\mathcal{Q}$ is closed in $\mathbf{Q}$ we obtain that $q_0 \in \mathcal{Q} \subseteq L^1(\mu)$. Finally, we define $\bar{Q}$ satisfying $\bar{Q} \ll \mu$ and $d\bar{Q}/d\mu \in \mathcal{Q}$ by setting

$$\bar{Q}(A) \equiv \int_A q_0 d\mu \tag{A.4}$$

for any measurable set A. Since by Assumption 2.2(iii) we have $\|\cdot\|_{\mu,1} \lesssim \|\cdot\|_{\mathbf{Q}}$ we can also conclude that $\|f_n - q_0\|_{\mu,1} = o(1)$. Therefore, for any bounded g we obtain

$$\lim_{n\to\infty} |\int g d\bar{Q} - \sum_{i=1}^{n} \lambda_{in} \int g dQ_i| \leq \lim_{n\to\infty} \|g\|_{\infty} \int |q_0 - \sum_{i=1}^{n} \lambda_{in} \frac{dQ_i}{d\mu}| d\mu = 0. \tag{A.5}$$

In particular, we note that (A.5) immediately yields $\int d\bar{Q} = 1$ and $0 \leq \bar{Q}(A) \leq 1$ for any measurable set A and hence by (A.4) that $\bar{Q}$ is indeed a probability measure.

We next show that $\bar{Q} \in \Theta_0$. To this end note that (A.5) and $Q_i \in \Theta_0$ implying Q_i induces P yields that for any value $t \in \mathbf{T}$ and (measurable) set V we must have

$$\begin{aligned}\bar{Q}(T^\star(Z) = t, (Y^\star(t), Z, X) \in V) &= \lim_{n\to\infty} \sum_{i=1}^{n} \lambda_{in} Q_i(T^\star(Z) = t, (Y^\star(t), Z, X) \in V) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \lambda_{in} P(T = t, (Y, Z, X) \in V) \\ &= P(T = t, (Y, Z, X) \in V),\end{aligned} \tag{A.6}$$

since $\sum_{i=1}^{n} \lambda_{in} = 1$, which implies $\bar{Q}$ also induces P. Next let f and g be arbitrary bounded functions of $(Y^\star, T^\star, X)$ and (Z, X) respectively and note that

$$\begin{aligned}E_{\bar{Q}}[g(Z,X)f(Y^\star,T^\star,X)] &= \lim_{n\to\infty} \sum_{i=1}^{n} \lambda_{in} E_{Q_i}[g(Z,X) E_{Q_i}[f(Y^\star,T^\star,X)|X]] \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \lambda_{in} E_P[g(Z,X) E_{Q_i}[f(Y^\star,T^\star,X)|X]] \\ &= E_{\bar{Q}}[g(Z,X)(\lim_{n\to\infty} \sum_{i=1}^{n} \lambda_{in} E_{Q_i}[f(Y^\star,T^\star,X)|X])],\end{aligned} \tag{A.7}$$

where the first equality follows from (A.5) and $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q_i and the second equality from Q_i inducing P due to $Q_i \in \Theta_0$. In turn, the third equality in (A.7) follows from the dominated convergence theorem and $\bar{Q}$ inducing P as shown in (A.6). Since

(A.7) holds for any bounded g, Definition 10.1.1 in Bogachev (2007) we obtain

$$E_{\bar{Q}}[f(Y^\star, T^\star, X)|X, Z] = \lim_{n\to\infty} \sum_{i=1}^{n} \lambda_{in} E_{Q_i}[f(Y^\star, T^\star, X)|X]$$
$$= E_{\bar{Q}}[f(Y^\star, T^\star, X)|X], \quad (A.8)$$

where the second equality can be deduced by applying the equalities in (A.7) evaluated at functions g of X only. We have so far shown that result (A.8) holds for any arbitrary bounded f. To extend the result to any $f \in L^1(\bar{Q})$ let $f_M \equiv f1\{|f| \leq M\}$ and note that result (A.8) and Proposition 10.1.7 in Bogachev (2007) imply that

$$E_{\bar{Q}}[f(Y^\star, T^\star, X)|X, Z] = \lim_{M\uparrow\infty} E_{\bar{Q}}[f_M(Y^\star, T^\star, X)|X, Z]$$
$$= \lim_{M\uparrow\infty} E_{\bar{Q}}[f_M(Y^\star, T^\star, X)|X] = E_{\bar{Q}}[f(Y^\star, T^\star, X)|X]. \quad (A.9)$$

Since (A.9) holds for any integrable f, we can conclude that $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under $\bar{Q}$ and therefore, by the preceding results, that $\bar{Q} \in \Theta_0$. Next, fix an arbitrary $Q \in \Theta_0$ and set A with $Q(A) > 0$ and note that there exists a $Q_k \in \mathcal{D}$ satisfying

$$\|\frac{dQ}{d\mu} - \frac{dQ_k}{d\mu}\|_{\mu,1} < \frac{Q(A)}{2}$$

by result (A.1). Hence, the triangle and Jensen's inequalities allow us to conclude that

$$Q_k(A) \geq Q(A) - |\int_A (\frac{dQ}{d\mu} - \frac{dQ_k}{d\mu})d\mu| \geq Q(A) - \|\frac{dQ}{d\mu} - \frac{dQ_k}{d\mu}\|_{\mu,1} > \frac{Q(A)}{2} \quad (A.10)$$

and thus that $Q_k(A) > 0$ as well. Since definition (A.2) implies that the sequence $\{\lambda_{kn}\}_{n=1}^{\infty}$ is bounded away from zero for n sufficiently large, we can combine results (A.5) and (A.10) to obtain that $\bar{Q}(A) \geq \liminf_{n\to\infty} \lambda_{kn} Q_k(A) > 0$. In particular, since A was arbitrary we can conclude that $Q \ll \bar{Q}$ as desired.

In order to establish Θ_0 is convex, let $Q_1, Q_2 \in \Theta_0$, $\gamma \in [0,1]$, and define $Q_\gamma \equiv \gamma Q_1 + (1-\gamma)Q_2$. Then note: (i) $dQ_\gamma/d\mu = \gamma dQ_1/d\mu + (1-\gamma)dQ_2/d\mu \in \mathcal{Q}$ by Assumption 2.2(iii); (ii) Q_γ induces P by the arguments in (A.6) applied with Q_γ in place of $\bar{Q}$; and (iii) $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q_γ by the arguments in (A.7) and (A.8) applied with Q_γ in place of $\bar{Q}$. It follows that $Q_\gamma \in \Theta_0$ and therefore that Θ_0 is convex. ■

Proof of Lemma 4.2. First note that $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ and Lemma 4.1 together imply that $Q(\Upsilon(\kappa) = \ell) = 1$ for all $Q \in \Theta_0$. By Corollary A.1 we can thus conclude

$$f(Y^\star, T^\star, X) = E_Q[\kappa(Y, T, Z, X)|Y^\star, T^\star, X] \quad (A.11)$$

Q-almost surely for any $Q \in \Theta_0$. Result (A.11) and $\kappa \in L^1(P)$ therefore yields that

$f \in L^1(Q)$ for all $Q \in \Theta_0$ as claimed. Hence, for any $Q \in \Theta_0$ we can conclude

$$E_Q[f(Y^\star, T^\star, X)] = E_Q[\kappa(Y, T, Z, X)] = E_P[\kappa(Y, T, Z, X)]$$

where the first equality follows from (A.11) and the law of iterated expectations, and the second equality follows from $Q \in \Theta_0$. ■

Proof of Corollary 4.1. By Lemma A.2, we have $\mu(\Upsilon(\kappa) = \ell) = 1$ and hence also that $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ due to $\bar{Q} \ll \mu$. Thus, the result is immediate from Lemma 4.2. ■

Proof of Theorem 4.1. We first show that λ_{Q_0} being identified implies that Λ belongs to the τ-closure of $\mathcal{R}$. To this end, we let $\mathcal{L}'$ denote the linear span of $\{\mathcal{R} \cup \Lambda\}$ and note that every $L \in \mathcal{L}$ can be identified with a linear functional on $\mathcal{L}'$ through the relation $L' \mapsto L'(L)$. We endow $\mathcal{L}'$ with the weak topology generated by $\mathcal{L}$, which we denote by $\sigma(\mathcal{L}', \mathcal{L})$, and observe $(\mathcal{L}', \sigma(\mathcal{L}', \mathcal{L}))$ is a topological vector space and its topological dual is $\mathcal{L}$; see, e.g., Example 1.3.23 in Bogachev and Smolyanov (2017). Moreover, since $\sigma(\mathcal{L}', \mathcal{L})$ is generated by the family of seminorms $\{|L|\}_{L \in \mathcal{L}}$, Theorem 5.73 in Aliprantis and Border (2006) implies that $(\mathcal{L}', \sigma(\mathcal{L}', \mathcal{L}))$ is additionally locally convex. We also note that by Lemma 2.53 in Aliprantis and Border (2006), the τ topology on $\{\mathcal{R} \cup \Lambda\}$ coincides with the relative topology on $\{\mathcal{R} \cup \Lambda\}$ that is induced by the topology $\sigma(\mathcal{L}', \mathcal{L})$ on $\mathcal{L}'$. Therefore, Λ belongs to the τ-closure of $\mathcal{R}$ (in $\{\mathcal{R} \cup \Lambda\}$) if and only if Λ belongs to the $\sigma(\mathcal{L}', \mathcal{L})$-closure of $\mathcal{R}$ (in $\mathcal{L}'$); see, e.g., Theorem 17.4 in Munkres (2000). Letting $\bar{\mathcal{R}}$ denote the $\sigma(\mathcal{L}', \mathcal{L})$-closure of $\mathcal{R}$ in $\mathcal{L}'$, it then follows that in order to show that Λ belongs to the τ-closure of $\mathcal{R}$ it suffices to establish that $\Lambda \in \bar{\mathcal{R}}$.

We proceed by contradiction and suppose that $\Lambda \notin \bar{\mathcal{R}}$. Since, as argued, $(\mathcal{L}', \sigma(\mathcal{L}', \mathcal{L}))$ is a locally convex topological vector space and $\mathcal{L}$ is its topological dual, Corollary 5.80 in Aliprantis and Border (2006) implies there then is an $L_0 \in \mathcal{L}$ satisfying

$$L_0(\Lambda) \neq 0 \qquad L_0(L') = 0 \text{ for all } L' \in \bar{\mathcal{R}}. \tag{A.12}$$

Moreover, by definition of $\mathcal{L}$ there is a finite collection $\{(s_j, Q_j)\}_{j=1}^J$ with $s_j \in \mathcal{S}_{Q_j}$ and $Q_j \in \Theta_0$ for all $1 \leq j \leq J$ and such that for all $L' \in \mathcal{L}'$ we have $L_0(L') = L'(\sum_{j=1}^J \langle \cdot, s_j \rangle_{Q_j})$. Hence, by definition of $\mathcal{R}$ and Υ we obtain for any $f \in L^1(P)$ that

$$\begin{aligned} 0 &= \sum_{j=1}^J E_{Q_j}[(\sum_{t \in \mathbf{T}} E_{P_{Z|X}}[f(Y^\star(t), t, Z, X) 1\{T^\star(Z) = t\}]) s_j(Y^\star, T^\star, X)] \\ &= \sum_{j=1}^J E_{Q_j}[f(Y, T, Z, X) E_{Q_j}[s_j(Y^\star, T^\star, X) | Y, T, Z, X]] \\ &= E_P[f(Y, T, Z, X)(\sum_{j=1}^J E_{Q_j}[s_j(Y^\star, T^\star, X) | Y, T, Z, X])], \end{aligned} \tag{A.13}$$

where the first equality follows from $L_0(L') = 0$ for all $L' \in \mathcal{R}$, the second from Corollary

A.1 and the law of iterated expectations, and the third from $Q_j \in \Theta_0$ for all $1 \leq j \leq J$. Thus, since (A.13) was shown to hold for any $f \in L^1(P)$ we can conclude that

$$P(\sum_{j=1}^{J} E_{Q_j}[s_j(Y^\star, T^\star, X)|Y, T, Z, X] = 0) = 1.$$

Hence, by Lemma A.3 there are $\tilde{Q}, Q^{\mathrm{a}} \in \Theta_0$ and $\eta > 0$ such that for all (measurable) A

$$\tilde{Q}(A) = Q^{\mathrm{a}}(A) + \eta \sum_{j=1}^{J} E_{Q_j}[s_j(Y^\star, T^\star, X) 1\{(Y^\star, T^\star, Z, X) \in A\}].$$

Letting $\mathbf{1}$ denote the function in $L^\infty(\bar{Q})$ that takes a constant value of one, we then obtain by definition of λ_Q, $L_0(\Lambda) = \Lambda(L_0)$ and the definition of Λ that

$$\lambda_{\tilde{Q}} = \lim_{k\to\infty} \langle \ell_k, \mathbf{1}\rangle_{\tilde{Q}} = \lim_{k\to\infty} \{\langle \ell_k, \mathbf{1}\rangle_{Q^{\mathrm{a}}} + \eta \sum_{j=1}^{J} \langle \ell_k, s_j\rangle_{Q_j}\} = \lambda_{Q^{\mathrm{a}}} + \eta L_0(\Lambda).$$

However, $\eta > 0$ and $L_0(\Lambda) \neq 0$ by (A.12) together imply that $\lambda_{\tilde{Q}} \neq \lambda_{Q^{\mathrm{a}}}$. Thus, since $\tilde{Q}, Q^{\mathrm{a}} \in \Theta_0$ we obtain that λ_Q is not identified reaching a contradiction. We therefore conclude that if λ_Q is identified, then Λ must belong to the τ-closure of $\mathcal{R}$.

For the converse direction, we now suppose that Λ belongs to the τ-closure of $\mathcal{R}$. Since $\mathcal{L}$ is identified (because it only depends on Θ_0), Λ is identified (because $\{\ell_j\}_{j=1}^\infty$ is identified), and $\mathcal{R}$ is identified (because $\Upsilon : L^1(P) \to L^1(\bar{Q})$ is identified), Theorem 2.4 in Aliprantis and Border (2006) implies there is an identified net $\{L'_\alpha\}_{\alpha \in \mathcal{A}}$ with

$$\lim_{\alpha} L'_\alpha(L) = \Lambda(L) \text{ for all } L \in \mathcal{L} \tag{A.14}$$

and $L'_\alpha \in \mathcal{R}$ for all $\alpha \in \mathcal{A}$. Therefore, for any $Q_1, Q_2 \in \Theta_0$ we can then conclude that

$$\lambda_{Q_1} = \Lambda(\langle \cdot, \mathbf{1}\rangle_{Q_1}) = \lim_{\alpha} L'_\alpha(\langle \cdot, \mathbf{1}\rangle_{Q_1}) = \lim_{\alpha} L'_\alpha(\langle \cdot, \mathbf{1}\rangle_{Q_2}) = \Lambda(\langle \cdot, \mathbf{1}\rangle_{Q_2}) = \lambda_{Q_2},$$

where the first and last equalities follow by definition of Λ, the second and fourth equalities by result (A.14), and the third equality by Lemma 4.2 and $L'_\alpha \in \mathcal{R}$. Thus, we conclude that λ_Q is constant in $Q \in \Theta_0$ and is therefore identified. ■

Proof of Corollary 4.2. First note that Assumptions 2.1 and 2.2 were directly imposed. Moreover, Assumption 4.1(i) is satisfied since $\ell \in L^1(\mu)$, $dQ/d\mu \in L^\infty(\mu)$ for all $Q \in \Theta_0$ by Assumption 2.2(iii), and Holder's inequality imply for any $Q \in \Theta_0$ that

$$E_Q[|\ell(Y^\star, T^\star, X)|] = \int |\ell| \frac{dQ}{d\mu} d\mu \leq \|\ell\|_{\mu,1} \|\frac{dQ}{d\mu}\|_{\mu,\infty} < \infty. \tag{A.15}$$

Also note that Assumption 4.1(ii) is immediate since here the sequence $\{\ell_j\}$ is constant, while Assumption 4.1(iii) trivially holds due to $\mathcal{Q} = \mathbf{Q}$. Thus, Theorem 4.1 implies that

λ_Q is identified if and only if Λ belongs to the τ-closure of $\mathcal{R}$.

To establish part (i), note that since $\ell \in L^1(Q)$ for all $Q \in \Theta_0$ and $sd\bar{Q}/d\mu \in L^\infty(\mu)$ for any $s \in L^\infty(\bar{Q}_{Y^\star T^\star X})$ (because $\bar{Q} \ll \mu$ and $d\bar{Q}/d\mu \in L^\infty(\mu)$), it follows that in this application $\mathcal{S}_{\bar{Q}} = L^\infty(\bar{Q}_{Y^\star T^\star X})$. Therefore, Lemma A.4 and the definition of $\mathcal{R}$ imply that there is a sequence $\{\kappa_j\} \subseteq L^1(P)$ satisfying $\|\ell - \Upsilon(\kappa_j)\|_{\bar{Q},1} = o(1)$. Since $\mu \ll \bar{Q}$ and $d\mu/d\bar{Q}$ is bounded, we can conclude $\{\kappa_j\}$ also satisfies $\|\ell - \Upsilon(\kappa_j)\|_{\mu,1} = o(1)$. For the converse, note that if there is a sequence $\{\kappa_j\} \subset L^1(P)$ satisfying $\|\ell - \Upsilon(\kappa_j)\|_{\mu,1} = o(1)$, then $dQ/d\mu \in L^\infty(\mu)$ for all $Q \in \Theta_0$ and Holder's inequality yields

$$\lim_{j\to\infty} \|\ell - \Upsilon(\kappa_j)\|_{Q,1} \leq \|\frac{dQ}{d\mu}\|_{\mu,\infty} \times \lim_{j\to\infty} \|\ell - \Upsilon(\kappa_j)\|_{\mu,1} = 0. \tag{A.16}$$

Therefore, for any $Q_1, Q_2 \in \Theta_0$, Lemma 4.2 and result (A.16) together establish that

$$\lambda_{Q_1} = \lim_{j\to\infty} \int \Upsilon(\kappa_j) dQ_1 = \lim_{j\to\infty} \int \kappa_j dP = \lim_{j\to\infty} \int \Upsilon(\kappa_j) dQ_2 = \lambda_{Q_2},$$

which establishes λ_{Q_0} is identified and $\lambda_{Q_0} = \lim_{j\to\infty} E_P[\kappa_j(Y,T,Z,X)]$. In turn part (ii) of the Corollary follows from part (i) and Lemma A.2. ■

Proof of Theorem 4.2. By Theorem 4.1, λ_{Q_0} is identified if and only if Λ belongs to the τ-closure of $\mathcal{R}$. Since $\mathcal{R}_T \subseteq \mathcal{R}$, it immediately follows that if Λ is in the τ-closure of $\mathcal{R}_T$, then λ_{Q_0} is identified. Thus, to establish the theorem it suffices to show that if Λ is in the τ-closure of $\mathcal{R}$, then it must also belong to the τ-closure of $\mathcal{R}_T$. To this end, note that if Λ belongs to the τ-closure of $\mathcal{R}$, then the definition of $\mathcal{R}$ and Theorem 2.14 in Aliprantis and Border (2006) imply that there exists a net $\{f_\alpha\}_{\alpha\in\mathcal{A}} \subseteq L^1(P)$ satisfying

$$\lim_{\alpha} \langle s, \Upsilon(f_\alpha)\rangle_Q = \Lambda(\langle s, \cdot\rangle_Q) \tag{A.17}$$

for all $s \in \mathcal{S}_Q$ and $Q \in \Theta_0$. Next, set $g_\alpha(t,Z,X) \equiv E_{\bar{Q}_{Y^\star|X}}[f_\alpha(Y^\star(t),t,Z,X)]$ for any $t \in \{t_1,\ldots,t_d\}$. By Jensen's inequality, $\bar{Q} \in \Theta_0$, and the definition of $\bar{Q}^{\text{it}}$ we then obtain

$$\begin{aligned} E_P[|g_\alpha(T,Z,X)|] &= E_{\bar{Q}_{T^\star ZX}}[\sum_{t\in\mathbf{T}} 1\{T^\star(Z)=t\}|E_{\bar{Q}_{Y^\star|X}}[f_\alpha(Y^\star(t),t,Z,X)]|] \\ &\leq E_{\bar{Q}^{\text{it}}}[\sum_{t\in\mathbf{T}} 1\{T^\star(Z)=t\}|f_\alpha(Y^\star(t),t,Z,X)|] \\ &= E_{\bar{Q}^{\text{it}}_{Y^\star T^\star X}}[\sum_{t\in\mathbf{T}} E_{\bar{Q}_{Z|X}}[1\{T^\star(Z)=t\}|f_\alpha(Y^\star(t),t,Z,X)|]], \end{aligned} \tag{A.18}$$

where the final equality follows from Lemma A.1 and $\bar{Q}^{\text{it}}_{ZX} = \bar{Q}_{ZX}$. Letting $\mathbf{1}$ denote the function of $(Y^\star, T^\star, X)$ taking a constant value of 1, note that $\mathbf{1} \in \mathcal{S}_{\bar{Q}}$ and Assumption 4.2(ii) imply $d\bar{Q}^{\text{it}}_{Y^\star T^\star X}/d\bar{Q}_{Y^\star T^\star X} \in \mathcal{S}_{\bar{Q}}$. In particular, since $\mathcal{S}_{\bar{Q}} \subseteq L^\infty(\bar{Q}_{Y^\star T^\star X})$, we may conclude that $d\bar{Q}^{\text{it}}_{Y^\star T^\star X}/d\bar{Q}_{Y^\star T^\star X}$ is bounded, which together with (A.18) yields

$$E_P[|g_\alpha(T,Z,X)|] \lesssim E_{\bar{Q}_{Y^\star T^\star X}}[\sum_{t\in\mathbf{T}} E_{\bar{Q}_{Z|X}}[1\{T^\star(Z)=t\}|f_\alpha(Y^\star(t),t,Z,X)|]]$$
$$= E_P[|f_\alpha(Y,T,Z,X)|], \tag{A.19}$$

where the equality follows from Corollary A.1 and $\bar{Q} \in \Theta_0$ implying $\bar{Q}_{Z|X} = P_{Z|X}$. Thus, since $f_\alpha \in L^1(P)$, result (A.19) implies that $g_\alpha \in L^1(P_{TZX})$.

Next select any $s_0 \in \mathcal{S}_{\bar{Q}} \cap L^\infty(\bar{Q}_{T^\star X})$ and note that the definition of Υ yields that

$$\langle s_0, \Upsilon(g_\alpha)\rangle_{\bar{Q}} = E_{\bar{Q}_{T^\star X}}[E_{P_{Z|X}}[E_{\bar{Q}_{Y^\star|X}}[\sum_{t\in\mathbf{T}} 1\{T^\star(Z)=t\}f_\alpha(Y^\star(t),t,Z,X)s_0(T^\star,X)]]]$$
$$= E_{\bar{Q}^{\text{it}}}[\sum_{t\in\mathbf{T}} 1\{T^\star(Z)=t\}f_\alpha(Y^\star(t),t,Z,X)s_0(T^\star,X)]$$
$$= E_{\bar{Q}^{\text{it}}_{Y^\star T^\star X}}[\sum_{t\in\mathbf{T}} E_{\bar{Q}^{\text{it}}_{Z|X}}[1\{T^\star(Z)=t\}f_\alpha(Y^\star(t),t,Z,X)]s_0(T^\star,X)], \tag{A.20}$$

where in the second equality we employed Lemma A.1, $P_{Z|X} = \bar{Q}_{Z|X}$ due to $\bar{Q} \in \Theta_0$, and the definition of $\bar{Q}^{\text{it}}$, while the final equality follows from Lemma A.1 and $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under $\bar{Q}^{\text{it}}$. Further note that because s_0 and ℓ_j only depend on $(T^\star, X)$ we have

$$\Lambda(\langle\cdot, s_0\rangle_{\bar{Q}}) = \lim_{j\to\infty}\langle \ell_j, s_0\rangle_{\bar{Q}_{T^\star X}} = \lim_{j\to\infty}\langle \ell_j, s_0\rangle_{\bar{Q}^{\text{i}}_{T^\star X}} = \lim_{j\to\infty}\langle \ell_j, s_0\frac{d\bar{Q}^{\text{it}}_{Y^\star T^\star X}}{d\bar{Q}_{Y^\star T^\star X}}\rangle_{\bar{Q}}$$
$$= \lim_\alpha \langle \Upsilon(f_\alpha), s_0\frac{d\bar{Q}^{\text{it}}_{Y^\star T^\star X}}{d\bar{Q}_{Y^\star T^\star X}}\rangle_{\bar{Q}} = \lim_\alpha\langle\Upsilon(f_\alpha), s_0\rangle_{\bar{Q}^{\text{it}}} = \lim_\alpha\langle\Upsilon(g_\alpha), s_0\rangle_{\bar{Q}}, \tag{A.21}$$

where the second equality follows from $\bar{Q}_{T^\star X} = \bar{Q}^{\text{it}}_{T^\star X}$, the fourth from result (A.17), $\bar{Q} \in \Theta_0$, and $s_0(d\bar{Q}^{\text{it}}_{Y^\star T^\star X}/d\bar{Q}_{Y^\star T^\star X}) \in \mathcal{S}_{\bar{Q}}$ by Assumption 4.2(ii), and the sixth from result (A.20), the definition of Υ, and $\bar{Q}^{\text{it}}_{Z|X} = \bar{Q}_{Z|X} = P_{Z|X}$. To conclude, let $\Pi_Q(s)(T^\star, X) \equiv E_Q[s(Y^\star, T^\star, X)|T^\star, X]$ and note that for any $s \in \mathcal{S}_Q$ and $Q \in \Theta_0$ we have

$$\Lambda(\langle\cdot, s\rangle_Q) = \lim_{j\to\infty}\langle \ell_j, \Pi_Q(s)\rangle_Q = \lim_{j\to\infty}\langle \ell_j, \Pi_Q(s)\frac{dQ_{T^\star X}}{d\bar{Q}_{T^\star X}}\rangle_{\bar{Q}}$$
$$= \lim_\alpha\langle\Upsilon(g_\alpha), \Pi_Q(s)\frac{dQ_{T^\star X}}{d\bar{Q}_{T^\star X}}\rangle_{\bar{Q}} = \lim_\alpha\langle\Upsilon(g_\alpha), s\rangle_Q, \tag{A.22}$$

where the first equality follows from ℓ_j depending only on $(T^\star, X)$, the third equality follows from result (A.21) and $\Pi_Q(s)dQ_{T^\star X}/d\bar{Q}_{T^\star X} \in \mathcal{S}_{\bar{Q}}$ by Assumptions 4.2(iii) and 4.2(iv), while the final equality follows from the law of iterated expectations. Thus, result (A.22) and Theorem 2.14 in Aliprantis and Border (2006) implies that Λ belongs to the τ-closure of $\mathcal{R}_T$, which establishes the claim of the theorem. ■

Proof of Corollary 4.3. The fact that existence of a $\kappa \in L^1(P_{TZX})$ satisfying $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ implies λ_{Q_0} is identified follows from Lemma 4.2. To establish the converse, we verify the conditions for Theorem 4.2. To this end, note Assumptions 2.1 and 2.2 were directly assumed, while Assumption 4.1(iii) is immediate from $\mathcal{Q} = \mathbf{Q} = L^1(\mu)$. Define

$$\mathbf{T}_0^\star \equiv \{t^\star \in \mathbf{T}^\star : \bar{Q}(T^\star = t^\star) > 0\},$$

and note that $Q(T^\star \in \mathbf{T}_0^\star) = 1$ for any $Q \in \Theta_0$ due to $Q \ll \bar{Q}$. Moreover, since any $Q \in \Theta_0$ must satisfy $Q_X = \bar{Q}_X = P_X$, it follows by direct calculation that

$$\frac{dQ_{T^\star X}}{d\bar{Q}_{T^\star X}}(T^\star, X) = \sum_{t^\star \in \mathbf{T}_0^\star} 1\{T^\star = t^\star\} \frac{Q(T^\star = t^\star|X)}{\bar{Q}(T^\star = t^\star|X)} \tag{A.23}$$

for any $Q \in \Theta_0$. In particular, since $\bar{Q}(T^\star = t^\star|X) \geq \varepsilon$ a.s. result (A.23) implies that

$$\|\frac{dQ_{T^\star X}}{d\bar{Q}_{T^\star X}}\|_{\bar{Q},\infty} \leq \frac{1}{\varepsilon}. \tag{A.24}$$

Hence, $\ell \in L^1(\bar{Q}_{T^\star X})$ and (A.24) imply $\ell \in L^1(Q)$ for any $Q \in \Theta_0$, verifying Assumptions 4.1(i)(ii). Further note that in this application $\mathcal{S}_Q = L^\infty(Q_{Y^\star T^\star X})$ for any $Q \in \Theta_0$ and therefore Assumptions 4.2(i)(ii) hold (because we assumed $d\bar{Q}^{\text{it}}/d\bar{Q}$ is bounded), Assumption 4.2(iii) follows from (A.24), and Assumption 4.2(iv) holds by Jensen's inequality. Thus, all the conditions of Theorem 4.2 are satisfied and we can conclude that if λ_{Q_0} is identified, then Λ belongs to the τ-closure of $\mathcal{R}_T$. By applying Lemma A.4 we can then conclude that there exists a sequence $\{\kappa_j\} \subseteq L^1(P_{TZX})$ satisfying

$$\lim_{j\to\infty} \|\ell - \Upsilon(\kappa_j)\|_{\bar{Q},1} = 0. \tag{A.25}$$

Next let $\mathbf{S}_0 \equiv \{(t, z) \in \mathbf{T} \times \mathbf{Z} : P(T = t, Z = z) > 0\}$ and note that $\bar{Q} \in \Theta_0$ implying $\bar{Q}$ is observationally equivalent to Q_0 and $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under $\bar{Q}$ yield

$$\begin{aligned} P(T = t, Z = z|X) &= \bar{Q}(T^\star(z) = t, Z = z|X) \\ &= \sum_{t^\star \in \mathbf{T}_0^\star : t^\star(z) = t} \bar{Q}(T^\star = t^\star|X) P(Z = z|X) \geq \varepsilon^2 \end{aligned} \tag{A.26}$$

for any $(t, z) \in \mathbf{S}_0$, and where in the final inequality we employed that $\bar{Q}(T^* = t^*|X) \geq \varepsilon$ for any $t^* \in \mathbf{T}_0^*$ and $P(Z = z|X) \geq \varepsilon$ for any $z \in \mathbf{Z}$ by hypothesis. Hence, for any event E with $P(X \in E) > 0$ and $(t, z) \in \mathbf{S}_0$, Bayes' rule and result (A.26) yield

$$P(X \in E) = \frac{P(X \in E|T = t, Z = z)P(T = t, Z = z)}{P(T = t, Z = z|X \in E)} \leq \frac{P(X \in E|T = t, Z = z)}{\varepsilon^2}.$$

Letting $P_{X|t,z}$ denote the distribution of X conditional on $(T, Z) = (t, z)$ for any $(t, z) \in \mathbf{S}_0$, it therefore follows that $P_X \ll P_{X|t,z}$ and $dP_X/dP_{X|t,z} \leq \varepsilon^{-2}$ almost surely under

$P_{X|t,z}$. In particular, we can conclude for any $(t,z) \in \mathbf{S}_0$ and $1 \leq j < \infty$ that

$$E_{P_X}[|\kappa_j(t,z,X)|] \leq \frac{1}{\varepsilon^2} E_P[|\kappa_j(t,z,X)||T=t,Z=z] \leq \frac{1}{\varepsilon^4} E_P[|\kappa_j(T,Z,X)|] \quad \text{(A.27)}$$

where the final inequality follows by noting (A.26) implies $P(T=t,Z=z) \geq \varepsilon^2$. Thus, $\kappa_j \in L^1(P_{TZX})$ and result (A.27) together imply that $\kappa_j(t,z,\cdot) \in L^1(P_X)$ for any $(t,z) \in \mathbf{S}_0$. Next set $\tilde{\kappa}_j(t,z,X) \equiv \kappa_j(t,z,X)P(Z=z|X)$ and note that

$$\begin{aligned} \|\ell - \Upsilon(\kappa_j)\|_{\bar{Q},1} &= E_{\bar{Q}}[|\ell(T^\star,X) - \sum_{(t,z)\in\mathbf{S}_0} 1\{T^\star(z)=t\}\kappa_j(t,z,X)P(Z=z|X)|] \\ &= E_{\bar{Q}}[|\ell(T^\star,X) - \sum_{(t,z)\in\mathbf{S}_0} 1\{T^\star(z)=t\}\tilde{\kappa}_j(t,z,X)|]. \end{aligned} \quad \text{(A.28)}$$

Since $\tilde{\kappa}_j(t,z,\cdot) \in L^1(P_X)$ for all $(t,z) \in \mathbf{S}_0$, results (A.25), (A.28), and Lemma A.5 imply there are functions $\{f_0(t,z,X)\}_{(t,z)\in\mathbf{S}_0}$ satisfying $f_0(t,z,\cdot) \in L^1(P_X)$ and

$$E_{\bar{Q}}[|\ell(T^\star,X) - \sum_{(t,z)\in\mathbf{S}_0} 1\{T^\star(z)=t\}f_0(t,z,X)|] = 0.$$

Finally, set $\kappa(t,z,X) \equiv f_0(t,z,X)/P(Z=z|X)$ and note that $\kappa \in L^1(P_{TZX})$ because $P(Z=z|X) \geq \varepsilon > 0$ a.s. and $f_0(t,z,\cdot) \in L^1(P_X)$ for any $(t,z) \in \mathbf{S}_0$. We then obtain

$$\begin{aligned} &E_{\bar{Q}}[|\ell(T^\star,X) - \sum_{t\in\mathbf{T}} E_{P_{Z|X}}[1\{T^\star(Z)=t\}\kappa(t,Z,X)]|] \\ &\qquad = E_{\bar{Q}}[|\ell(T^\star,X) - \sum_{(t,z)\in\mathbf{S}_0} 1\{T^\star(z)=t\}f_0(t,z,X)|] = 0, \end{aligned}$$

yielding that identification of λ_Q implies the existence of the desired κ. Hence, we have shown that λ_{Q_0} is identified if and only if $\bar{Q}(\ell = \Upsilon(\kappa)) = 1$ for some $\kappa \in L^1(P_{TZX})$, which establishes part (i) of the corollary. Part (ii) of the corollary is immediate from part (i), Lemma A.2, $\mu \ll \bar{Q}$ by assumption, and $\bar{Q} \ll \mu$ since $\bar{Q} \in \Theta_0$. ■

Proof of Theorem 4.3. We first show that if Λ belongs to the τ-closure of $\mathcal{R}_t$, then λ_{Q_0} is identified. To this end, note that $\mathcal{L}$ is identified (because Θ_0 is identified), Λ is identified (because $\{\ell_j\}$ is identified), and $\mathcal{R}_t$ is identified (because $\Upsilon : L^1(P) \to L^1(\bar{Q})$ is identified). Hence, Theorem 2.14 in Aliprantis and Border (2006) and Λ being in the τ-closure of $\mathcal{R}_t$ imply there is an identified net $\{L'_\alpha\}_{\alpha\in\mathcal{A}}$ satisfying

$$\lim_\alpha L'_\alpha(L) = \Lambda(L) \text{ for all } L \in \mathcal{L}. \quad \text{(A.29)}$$

Next, let $\Pi_Q(s)(T^\star,X) \equiv E_Q[s(Y^\star,T^\star,X)|T^\star,X]$ for any $s \in L^1(Q)$, and note that Assumption 4.3(i) and the law of iterated expectations imply for any $Q \in \Theta_0$ that

$$\lambda_Q = \Lambda(\langle \cdot, \Pi_Q(\rho)\rangle_Q) = \lim_\alpha L'_\alpha(\langle \cdot, \Pi_Q(\rho)\rangle_Q), \quad \text{(A.30)}$$

where the second equality follows from (A.29) and $\Pi_Q(\rho) \in \mathcal{S}_Q$ by Assumption 4.4(i). Also note that, by definition of $\mathcal{R}_t$, there exists a net $\{f_\alpha\}_{\alpha\in\mathcal{A}} \subseteq L^1(P_{tZX})$ satisfying $L'_\alpha(\langle\cdot, s\rangle_Q) = \langle\Upsilon(f_\alpha), s\rangle_Q$ for any $Q \in \Theta_0$ and $s \in \mathcal{S}_Q$. Noting that $f_\alpha(T, Z, X) = g_\alpha(Z, X)1\{T = t\}$ for some function g_α, we then obtain that

$$\begin{aligned} L'_\alpha(\langle\cdot, \Pi_Q(\rho)\rangle_Q) &= E_Q[E_Q[\rho(Y^\star(t))|T^\star, X]E_{P_{Z|X}}[g_\alpha(Z, X)1\{T^\star(Z) = t\}]] \\ &= E_Q[E_{P_{Z|X}}[\rho(Y^\star(t))g_\alpha(Z, X)1\{T^\star(Z) = t\}]] = E_P[\rho(Y)g_\alpha(Z, X)1\{T = t\}], \end{aligned} \quad \text{(A.31)}$$

where the final equality follows from Corollary A.1, the law of iterated expectations, and $Q \in \Theta_0$. Since (A.30) and (A.31) hold for any $Q \in \Theta_0$, it follows that λ_{Q_0} is identified.

We next establish that if λ_{Q_0} is identified, then Λ must belong to the τ-closure of $\mathcal{R}_t$. To this end, we first define the spaces $\mathcal{S}_{Q,\rho}$ and $\mathcal{L}_\rho$ to be given by

$$\mathcal{S}_{Q,\rho} \equiv \{s \in L^\infty(Q_{Y^\star T^\star X}) : |\lim_{j\to\infty}\langle s, \rho\ell_j\rangle_Q| < \infty \text{ and } s\frac{dQ}{d\mu} \in \mathbf{Q}\}$$
$$\mathcal{L}_\rho \equiv \text{span}\{L : \bigcap_{Q\in\Theta_0} L^1(Q) \to \mathbf{R} \text{ s.t. } L = \langle\cdot, s\rangle_Q \text{ for some } s \in \mathcal{S}_{Q,\rho} \text{ and } Q \in \Theta_0\},$$

let $\Lambda_\rho(L) \equiv \lim_{j\to\infty} L(\ell_j\rho)$ for any $L \in \mathcal{L}_\rho$, and τ_ρ denote the weak topology on $\{\mathcal{R}\cup\Lambda_\rho\}$ that is generated by $\mathcal{L}_\rho$ – i.e. $\mathcal{S}_{Q,\rho}$, $\mathcal{L}_\rho$, Λ_ρ, and τ_ρ correspond to our definitions for $\mathcal{L}, \mathcal{S}_Q, \Lambda$, and τ applied with $\{\ell_j\rho\}$ in place of $\{\ell_j\}$. By Theorem 4.1 and λ_{Q_0} being identified, it then follows that Λ_ρ belongs to the τ_ρ-closure of $\mathcal{R}$. Hence, Theorem 2.14 in Aliprantis and Border (2006) implies there is a net $\{L'_\alpha\}_{\alpha\in\mathcal{A}} \subseteq \mathcal{R}$ satisfying

$$\lim_\alpha L'_\alpha(L) = \Lambda_\rho(L) \text{ for all } L \in \mathcal{L}_\rho. \quad \text{(A.32)}$$

Next note that $Y^\star(t)$ being independent of $T^\star$ conditionally on X under $\bar{Q}^{\text{io}}$, the marginal distribution of $(Y^\star(t), X)$ being the same under $\bar{Q}$ and $\bar{Q}^{\text{io}}$, the law of iterated expectations, and Assumptions 4.4(ii)(iii) allow us to conclude that

$$\bar{Q}^{\text{io}}(E_{\bar{Q}^{\text{io}}}[\rho(Y^\star(t))\phi_{\bar{Q},\rho}(Y^\star(t), X)|T^\star, X] = 1) = 1. \quad \text{(A.33)}$$

Fixing an arbitrary $s \in \mathcal{S}_{\bar{Q}}$, then note that the law of iterated expectations, the marginal distribution of $(T^\star, X)$ being the same under $\bar{Q}$ and $\bar{Q}^{\text{io}}$ and result (A.33) yield

$$\begin{aligned} \lim_{j\to\infty}\langle\ell_j, s\rangle_{\bar{Q}} &= \lim_{j\to\infty}\langle\ell_j, \Pi_{\bar{Q}}(s)\rangle_{\bar{Q}} = \lim_{j\to\infty}\langle\ell_j, \Pi_{\bar{Q}}(s)\rangle_{\bar{Q}^{\text{io}}} \\ &= \lim_{j\to\infty}\langle\ell_j\rho, \phi_{\bar{Q},\rho}\Pi_{\bar{Q}}(s)\rangle_{\bar{Q}^{\text{io}}} = \lim_{j\to\infty}\langle\ell_j\rho, \phi_{\bar{Q},\rho}\Pi_{\bar{Q}}(s)\frac{d\bar{Q}^{\text{io}}_{Y^\star T^\star X}}{d\bar{Q}_{Y^\star T^\star X}}\rangle_{\bar{Q}}, \end{aligned} \quad \text{(A.34)}$$

where the final equality follows from Assumption 4.4(ii). Since the limit in (A.34) exists due to $s \in \mathcal{S}_{\bar{Q}}$, Assumptions 4.4(i)(iii) imply $\phi_{\bar{Q},\rho}\Pi_{\bar{Q}}(s)(d\bar{Q}^{\text{io}}_{Y^\star T^\star X}/d\bar{Q}_{Y^\star T^\star X}) \in \mathcal{S}_{\bar{Q},\rho}$. Next note that by definition of $\mathcal{R}$, there exists a net $\{v_\alpha\}_{\alpha\in\mathcal{A}} \subseteq L^1(P)$ such that

$L'_\alpha(\langle\cdot, s\rangle_Q) = \langle\Upsilon(v_\alpha), s\rangle_Q$ for any $Q \in \Theta_0$ and $s \in \mathcal{S}_{Q,\rho}$. In particular, we have

$$\lim_{j\to\infty}\langle\ell_j, s\rangle_{\bar{Q}} = \lim_\alpha\langle\Upsilon(v_\alpha), \phi_{\bar{Q},\rho}\Pi_{\bar{Q}}(s)\frac{d\bar{Q}^{\text{io}}_{Y^\star T^\star X}}{d\bar{Q}_{Y^\star T^\star X}}\rangle_{\bar{Q}} = \lim_\alpha\langle\Upsilon(v_\alpha), \phi_{\bar{Q},\rho}\Pi_{\bar{Q}}(s)\rangle_{\bar{Q}^{\text{io}}} \quad \text{(A.35)}$$

due to (A.32) and (A.34). Next set $f_\alpha(T, Z, X) \equiv 1\{T = t\}g_\alpha(Z, X)$ with g_α given by

$$g_\alpha(Z, X) \equiv E_{\bar{Q}_{Y^\star(t)|X}}[v_\alpha(Y^\star(t), t, Z, X)\phi_{\bar{Q},\rho}(Y^\star(t), X)],$$

and where in the expectation Z and X are kept constant. Also note that $\bar{Q} \in \Theta_0$, Jensen's inequality, and $\phi_{\bar{Q},\rho} \in L^\infty(\bar{Q})$ by Assumption 4.4(iii) imply that

$$\begin{aligned}
E_P[|f_\alpha(T, Z, X)|] &\lesssim E_{\bar{Q}_{T^\star ZX}}[1\{T^\star(Z) = t\}E_{\bar{Q}_{Y^\star(t)|X}}[|v_\alpha(Y^\star(t), t, Z, X)|]\\
&= E_{\bar{Q}^{\text{io}}}[1\{T^\star(Z) = t\}|v_\alpha(Y^\star(t), t, Z, X)|]\\
&\lesssim E_{\bar{Q}}[1\{T^\star(Z) = t\}|v_\alpha(Y^\star(t), t, Z, X)|]\\
&= E_P[|1\{T = t\}v_\alpha(Y, T, Z, X)|], \quad \text{(A.36)}
\end{aligned}$$

where the first equality holds by definition of $\bar{Q}^{\text{io}}$; the second inequality follows from $d\bar{Q}^{\text{io}}/d\bar{Q} \in L^\infty(\bar{Q})$ by Assumption 4.4(ii); and the final equality holds because $\bar{Q} \in \Theta_0$. In particular, since $v_\alpha \in L^1(P)$, result (A.36) implies $f_\alpha \in L^1(P_{tZX})$ and therefore that $\Upsilon(f_\alpha) \in \mathcal{R}_t$. Finally, we observe that the law of iterated expectations, $\bar{Q} \in \Theta_0$, Corollary A.1, and $\bar{Q}^{\text{io}}_{T^\star ZX} = \bar{Q}_{T^\star ZX}$ allow us to conclude for any $s \in \mathcal{S}_{\bar{Q}}$ that

$$\begin{aligned}
&\langle\Upsilon(f_\alpha), s\rangle_{\bar{Q}}\\
&= E_{\bar{Q}^{\text{io}}}[g_\alpha(Z, X)1\{T^\star(Z) = t\}E_{\bar{Q}}[s(Y^\star, T^\star, X)|T^\star, X]]\\
&= E_{\bar{Q}^{\text{io}}}[v_\alpha(Y^\star(t), t, Z, X)1\{T^\star(Z) = t\}\phi_{\bar{Q},\rho}(Y^\star(t), X)E_{\bar{Q}}[s(Y^\star, T^\star, X)|T^\star, X]]\\
&= \sum_{\tilde{t}\in\mathbf{T}} E_{\bar{Q}^{\text{io}}}[v_\alpha(Y^\star(\tilde{t}), \tilde{t}, Z, X)1\{T^\star(Z) = \tilde{t}\}\phi_{\bar{Q},\rho}(Y^\star(t), X)E_{\bar{Q}}[s(Y^\star, T^\star, X)|T^\star, X]]\\
&= \langle\Upsilon(v_\alpha), \phi_{\bar{Q},\rho}\Pi_{\bar{Q}}(s)\rangle_{\bar{Q}^{\text{io}}}, \quad \text{(A.37)}
\end{aligned}$$

where the second equality follow from the definition of $\bar{Q}^{\text{io}}$ and g_α; the third equality from $(Y^\star(\tilde{t}), T^\star, Z)$ being independent of $Y^\star(t)$ conditionally on X under $\bar{Q}^{\text{io}}$ whenever $\tilde{t} \neq t$, $E_{\bar{Q}^{\text{io}}}[\phi_{\bar{Q},\rho}(Y^\star(t), X)|X] = 0$ by definition of $\phi_{\bar{Q},\rho}$ and $\bar{Q}^{\text{io}}_{Y^\star(t)X} = \bar{Q}_{Y^\star(t)X}$; and the final equality holds by Lemma A.1 and $\bar{Q}^{\text{io}}_{ZX} = \bar{Q}_{ZX} = P_{ZX}$. Thus, combining results (A.35) with (A.37) allows us to conclude that for any $s \in \mathcal{S}_{\bar{Q}}$ we have

$$\lim_{j\to\infty}\langle\ell_j, s\rangle_{\bar{Q}} = \lim_\alpha\langle\Upsilon(f_\alpha), s\rangle_{\bar{Q}}. \quad \text{(A.38)}$$

To conclude, note that for any $Q \in \Theta_0$ and $s \in \mathcal{S}_Q$, the law of iterated expectations,

$Q \ll \bar{Q}$, $\Pi_Q(s)dQ_{T^\star X}/d\bar{Q}_{T^\star X} \in \mathcal{S}_{\bar{Q}}$ by Assumptions 4.4(i)(iv), and result (A.38) yield

$$\Lambda(\langle \cdot, s\rangle_Q) = \lim_{j\to\infty} \langle \ell_j, \Pi_Q(s)\rangle_Q = \lim_{j\to\infty} \langle \ell_j, \Pi_Q(s)\frac{dQ_{T^\star X}}{d\bar{Q}_{T^\star X}}\rangle_{\bar{Q}}$$
$$= \lim_{\alpha} \langle \Upsilon(f_\alpha), \Pi_Q(s)\frac{dQ_{T^\star X}}{d\bar{Q}_{T^\star X}}\rangle_{\bar{Q}} = \lim_{\alpha}\langle \Upsilon(f_\alpha), s\rangle_Q.$$

Hence, since $\Upsilon(f_\alpha) \in \mathcal{R}_t$, we conclude that Λ belongs to the τ-closure of $\mathcal{R}_t$. ■

Proof of Corollary 4.4. The proof is similar to that of Corollary 4.3 and we therefore omit some of the details. We first verify that the assumptions of Theorem 4.3 are satisfied. To this end, note that Assumptions 2.1 and 2.2 were directly imposed, while $\mathcal{Q} = \mathbf{Q} = L^1(\mu)$ implies Assumption 4.1(iii) holds. Letting $\mathbf{T}_0^\star \equiv \{t^\star \in \mathbf{T}^\star : \bar{Q}(T^\star = t^\star) > 0\}$, it can then be shown that $\bar{Q}(T^\star = t^\star|X) \geq \varepsilon > 0$ a.s. and $Q_X = P_X$ for any $Q \in \Theta_0$ yield

$$\|\frac{dQ_{T^\star X}}{d\bar{Q}_{T^\star X}}\|_{\bar{Q},\infty} \leq \frac{1}{\varepsilon}. \tag{A.39}$$

In particular, $\ell \in L^1(\bar{Q}_{T^\star X})$ and result (A.39) imply that Assumptions 4.3(i)(ii) also hold. Moreover, since in this application $\mathcal{S}_Q = L^\infty(Q_{Y^\star T^\star X})$ for any $Q \in \Theta_0$, Assumptions 4.4(i) and (iv) hold by Jensen's inequality and result (A.39) respectively. Similarly, we note that Assumption 4.4(ii) was directly imposed, while Assumption 4.4(iii) is satisfied since we assumed $\rho \in L^\infty(\bar{Q})$ and $\mathrm{Var}_{\bar{Q}}\{\rho(Y^\star(t_0))|X\} \geq \varepsilon > 0$ a.s. under $\bar{Q}$. Thus, the conditions of Theorem 4.3 hold.

Next note that if (i) holds, then Theorem 4.3 implies Λ belongs to the τ-closure of $\mathcal{R}_t$. By Lemma A.4, there therefore exists a sequence $\{\kappa_j\} \in L^1(P_{ZXt})$ satisfying

$$\lim_{j\to\infty} \|\ell - \Upsilon(\kappa_j)\|_{\bar{Q},1} = 0. \tag{A.40}$$

Letting $\mathbf{Z}_0 \equiv \{z \in \mathbf{Z} : P(T = t, Z = z) > 0\}$ and noting that $\kappa_j(T,Z,X) = 1\{T = t\}g_j(Z,X)$ for some function g_j by definition of $L^1(P_{tZX})$, it then follows from the same arguments employed in Corollary 4.3 that $g_j(z,\cdot) \in L^1(P_X)$ for any $z \in \mathbf{Z}_0$. Next set $\tilde{g}_j(z,X) \equiv g_j(z,X)P(Z = z|X)$ and observe that by definition of Υ we have

$$\|\ell - \Upsilon(\kappa_j)\|_{\bar{Q},1} = E_{\bar{Q}}[|\ell(T^\star, X) - \sum_{z\in\mathbf{Z}_0} 1\{T^\star(z) = t\}g_j(z,X)P(Z = z|X)|]$$
$$= E_{\bar{Q}}[|\ell(T^\star, X) - \sum_{z\in\mathbf{Z}_0} 1\{T^\star(z) = t\}\tilde{g}_j(z,X)|]. \tag{A.41}$$

Combining results (A.40) and (A.41) with Lemma A.5 then implies that there are functions $\{f_0(z,X)\}_{z\in\mathbf{Z}_0}$ satisfying $f_0(z,\cdot) \in L^1(P_X)$ for all $z \in \mathbf{Z}_0$ and

$$E_{\bar{Q}}[|\ell(T^\star, X) - \sum_{z\in\mathbf{Z}_0} 1\{T^\star(z) = t\}f_0(z,X)|] = 0. \tag{A.42}$$

Hence, setting $\kappa(T, Z, X) \equiv 1\{T = t\} \sum_{z \in \mathbf{Z}_0} 1\{Z = z\} f_0(Z, X)/P(Z = z|X)$ we obtain from $P(Z = z|X) \geq \varepsilon > 0$ a.s. that $\kappa \in L^1(P_{tZX})$ and from (A.42) that $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$. Since $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ implies $\bar{Q}(\Upsilon(f\kappa) = f\ell) = 1$ for any bounded f, Lemma 4.2 allows us to conclude that (i) implies (iii). Thus, because (iii) trivially implies (ii) and (ii) trivially implies (i), the claim of the corollary follows. ■

Lemma A.1. *Let Q be a distribution for $(Y^\star, T^\star, Z, X)$ satisfying $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q. Then it follows that for any $f \in L^1(Q)$ we have:*

$$E_Q[f(Y^\star, T^\star, Z, X)|Y^\star, T^\star, X] = E_{Q_{Z|X}}[f(Y^\star, T^\star, Z, X)].$$

Proof. Let $\mathcal{G}$ denote the σ-field on which Q is defined, which recall we set to equal the σ-field generated by $(\mathcal{G}_{Y^\star} \times \mathcal{G}_{T^\star} \times \mathcal{G}_Z \times \mathcal{G}_X)$ where $\mathcal{G}_V$ denotes the σ-field on which the marginal distribution Q_V is defined. We further define the class of sets

$$\mathcal{A} \equiv \{A \in \mathcal{G} : E_Q[1\{(Y^\star, T^\star, Z, X) \in A\}|Y^\star, T^\star, X] = E_{Q_{Z|X}}[1\{(Y^\star, T^\star, Z, X) \in A\}]\}$$

and note that $\mathbf{Y}^\star \times \mathbf{T}^\star \times \mathbf{Z} \times \mathbf{X} \in \mathcal{A}$. Also observe that if $A_1, A_2 \in \mathcal{A}$ and $A_1 \subseteq A_2$ then

$$\begin{aligned} E_Q[1\{(Y^\star, T^\star, Z, X) \in A_2 \setminus A_1\}|Y^\star, T^\star, X] \\ &= E_Q[1\{(Y^\star, T^\star, Z, X) \in A_2\} - 1\{(Y^\star, T^\star, Z, X) \in A_1\}|Y^\star, T^\star, X] \\ &= E_{Q_{Z|X}}[1\{(Y^\star, T^\star, Z, X) \in A_2\}] - E_{Q_{Z|X}}[1\{(Y^\star, T^\star, Z, X) \in A_1\}] \\ &= E_{Q_{Z|X}}[1\{(Y^\star, T^\star, Z, X) \in A_2 \setminus A_1\}], \qquad \text{(A.43)} \end{aligned}$$

where the first and third equalities follow from $A_1 \subseteq A_2$ while the second equality follows from $A_1, A_2 \in \mathcal{A}$. In particular, result (A.43) implies that $A_2 \setminus A_1 \in \mathcal{A}$. Next, let $\{A_i\}_{i=1}^\infty \subset \mathcal{A}$ be a sequence of pairwise disjoint sets and note that

$$\begin{aligned} E_Q[1\{(Y^\star, T^\star, Z, X) \in \bigcup_{i=1}^\infty A_i\}|Y^\star, T^\star, X] \\ &= \lim_{n\to\infty} E_Q[\sum_{i=1}^n 1\{(Y^\star, T^\star, Z, X) \in A_i\}|Y^\star, T^\star, X] \\ &= \lim_{n\to\infty} E_{Q_{Z|X}}[\sum_{i=1}^n 1\{(Y^\star, T^\star, Z, X) \in A_i\}] \\ &= E_{Q_{Z|X}}[1\{(Y^\star, T^\star, Z, X) \in \bigcup_{i=1}^\infty A_i\}], \qquad \text{(A.44)} \end{aligned}$$

where the first and third equalities follow from Theorem 10.1.5(4) in Bogachev (2007) and $\{A_i\}_{i=1}^\infty$ being disjoint, while the second holds due to $A_i \in \mathcal{A}$ for all i. In particular, (A.44) implies $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$, and we can therefore conclude that $\mathcal{A}$ is a λ-system.

Let $A_{Y^\star} \in \mathcal{G}_{Y^\star}$, $A_{T^\star} \in \mathcal{G}_{T^\star}$, $A_Z \in \mathcal{G}_Z$, and $A_X \in \mathcal{G}_X$ be arbitrary, and then observe

$$\begin{aligned}
&E_Q[1\{(Y^\star, T^\star, Z, X) \in A_{Y^\star} \times A_{T^\star} \times A_Z \times A_X\}|Y^\star, T^\star, X] \\
&\qquad = 1\{Y^\star \in A_{Y^\star}\}1\{T^\star \in A_{T^\star}\}1\{X \in A_X\}E_{Q_{Z|X}}[1\{Z \in A_Z\}] \\
&\qquad = E_{Q_{Z|X}}[1\{(Y^\star, T^\star, Z, X) \in A_{Y^\star} \times A_{T^\star} \times A_Z \times A_X\}] \qquad \text{(A.45)}
\end{aligned}$$

where the first equality follows from $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q and the second by direct manipulation. Result (A.45) implies $A_{Y^\star} \times A_{T^\star} \times A_Z \times A_X \in \mathcal{A}$ and, since $A_{Y^\star}, A_{T^\star}, A_Z, A_X$ were arbitrary, that $(\mathcal{G}_{Y^\star} \times \mathcal{G}_{T^\star} \times \mathcal{G}_Z \times \mathcal{G}_X) \subseteq \mathcal{A}$. Since $(\mathcal{G}_Z \times \mathcal{G}_X \times \mathcal{G}_{T^\star} \times \mathcal{G}_{Y^\star})$ is a π-system and $\mathcal{G}$ equals the σ-field generated by $(\mathcal{G}_{Y^\star} \times \mathcal{G}_{T^\star} \times \mathcal{G}_Z \times \mathcal{G}_X)$, the $\pi - \lambda$ theorem (see, e.g., Theorem 2.38 in Pollard (2002)) then implies $\mathcal{A} = \mathcal{G}$.

To conclude, let $f \in L^1(Q)$ be arbitrary and $\{f_n\}$ be a sequence of simple functions satisfying $|f_n| \leq |f|$ and $f_n(Y^\star, T^\star, Z, X) \to f(Y^\star, T^\star, Z, X)$ on a set with Q-probability one. By Proposition 10.1.7 in Bogachev (2007) we can then conclude that

$$\begin{aligned}
E_Q[f(Y^\star, T^\star, Z, X)|Y^\star, T^\star, X] &= \lim_{n\to\infty} E_Q[f_n(Y^\star, T^\star, Z, X)|Y^\star, T^\star, X] \\
&= \lim_{n\to\infty} E_{Q_{Z|X}}[f_n(Y^\star, T^\star, Z, X)] = E_{Q_{Z|X}}[f(Y^\star, T^\star, Z, X)],
\end{aligned}$$

where the second equality holds due to f_n being a simple function and $\mathcal{A} = \mathcal{G}$. ■

Corollary A.1. *If Assumption 2.1 holds, $Q \in \Theta_0$, and $f \in L^1(P)$, then it follows*

$$E_Q[f(Y, T, Z, X)|Y^\star, T^\star, X] = \sum_{t\in\mathbf{T}} E_{P_{Z|X}}[f(Y^\star(t), t, Z, X)1\{T^\star(Z) = t\}].$$

Proof. The claim is immediate from Lemma A.1 and noting that: (i) $Y = Y^\star(T)$ and $T = T^\star(Z)$ by Assumption 2.1 imply $f(Y, T, Z, X) = \sum_{t\in\mathbf{T}} f(Y^\star(t), t, Z, X)1\{T^\star(Z) = t\}$, and (ii) $Q_{Z|X} = P_{Z|X}$ due to $Q \in \Theta_0$ by hypothesis. ■

Lemma A.2. *Let Assumptions 2.1 and 2.2 hold, and suppose that $\mu(\pi(Z, X) > \delta) = 1$ for some $\delta > 0$. If $\nu \in L^1(P)$, then it follows $\kappa = \nu/\pi$ satisfies $\kappa \in L^1(P)$ and*

$$E_{\mu_{Z|X}}[\nu(Y^*(t), t, Z, X)1\{T^*(Z) = t\}] = E_{P_{Z|X}}[\kappa(Y^*(t), t, Z, X)1\{T^*(Z) = t\}].$$

Proof. First note that since $\mu(\pi(Z, X) > \delta) = 1$ and $\bar{Q} \in \Theta_0$ must satisfy $\bar{Q} \ll \mu$, it follows that $\bar{Q}(\pi(Z, X) > \delta) = 1$. Moreover, $\bar{Q}_{ZX} = P_{ZX}$ due to $\bar{Q} \in \Theta_0$ and therefore $P(\pi(Z, X) > \delta) = 1$ as well. Hence, we can conclude that $1/\pi \in L^\infty(P)$, which together with $\nu \in L^1(P)$ yields that $\kappa = \nu/\pi \in L^1(P)$. Next note that since $\mu(\pi(Z, X) > 0) = 1$, it follows that on a set with μ-probability one we must have

$$\begin{aligned}
&E_{\mu_{Z|X}}[\nu(Y^*(t), t, Z, X)1\{T^\star(Z) = t\}] \\
&\quad = E_{\mu_{Z|X}}[\nu(Y^*(t), t, Z, X)1\{T^\star(Z) = t\}1\{\pi(Z, X) > 0\}] \\
&\quad = E_{P_{Z|X}}[\kappa(Y^*(t), t, Z, X)1\{T^*(Z) = t\}],
\end{aligned}$$

where the final equality follows from the definitions of κ and π. ■

Lemma A.3. *Let Assumptions 2.1, 2.2, and 4.1(iii) hold, and suppose that a finite collection* $\{s_j, Q_j\}_{j=1}^J$ *with* $s_j \in \mathcal{S}_{Q_j}$ *and* $Q_j \in \Theta_0$ *for all* $1 \le j \le J$ *satisfies*

$$P(\sum_{j=1}^J E_{Q_j}[s_j(Y^\star, T^\star, X)|Y, T, Z, X] = 0) = 1. \tag{A.46}$$

Then, there exist a $Q^{\mathrm{a}} \in \Theta_0$ *and a constant* $\eta > 0$ *such that the measure* $\tilde{Q}$ *given by*

$$\tilde{Q}(A) \equiv Q^{\mathrm{a}}(A) + \sum_{j=1}^J \eta E_{Q_j}[s_j(Y^\star, T^\star, X)1\{(Y^\star, T^\star, Z, X) \in A\}]$$

(for any A *in the domain of* $\bar{Q}$*) belongs to the identified set* Θ_0*.*

Proof. First note that by Assumption 4.1(iii) there exists a measure $Q^{\mathrm{i}} \in \Theta_0$ such that $dQ^{\mathrm{i}}/d\mu$ belongs to the interior of $\mathcal{Q}$ in $\mathbf{Q}$. Therefore setting Q^{a} to equal

$$Q^{\mathrm{a}} = \lambda Q^{\mathrm{i}} + \frac{1-\lambda}{J} \sum_{j=1}^J Q_j \tag{A.47}$$

we can conclude that $dQ^{\mathrm{a}}/d\mu$ also belongs to the interior of $\mathcal{Q}$ in $\mathbf{Q}$ provided $\lambda \in (0, 1)$ is chosen sufficiently large, and moreover that $Q^{\mathrm{a}} \in \Theta_0$ due to Θ_0 being convex and $Q_j \in \Theta_0$ for all $1 \le j \le J$. Next note that $\tilde{Q}(A)$ is well defined for any A in the domain of $\bar{Q}$ and that $\tilde{Q}$ is countably additive by the dominated convergence theorem. Moreover, observe that $\|s_j\|_{Q_j,\infty} < \infty$ for all $1 \le j \le J$ since $s_j \in \mathcal{S}_{Q_j} \subseteq L^\infty(Q_j)$. Hence, by the choice of Q^{a} in (A.47) we obtain for any A in the domain of $\bar{Q}$ that

$$\tilde{Q}(A) \ge \sum_{j=1}^J \{\frac{1-\lambda}{J} Q_j(A) - \eta \|s_j\|_{Q_j,\infty} \int_A dQ_j\} = \sum_{j=1}^J Q_j(A)(\frac{1-\lambda}{J} - \eta\|s_j\|_{Q_j,\infty}). \tag{A.48}$$

Thus, result (A.48) implies $\tilde{Q}$ is a positive measure provided we set $\eta > 0$ to satisfy $\eta < (1-\lambda)/(J \max_j \|s\|_{\bar{Q},\infty})$ (which is possible because $\lambda < 1$). Further observe

$$\tilde{Q}(\mathbf{Y}^\star \times \mathbf{T}^\star \times \mathbf{Z} \times \mathbf{X}) = Q^{\mathrm{a}}(\mathbf{Y}^\star \times \mathbf{T}^\star \times \mathbf{Z} \times \mathbf{X}) + \eta \sum_{j=1}^J E_{Q_j}[s_j(Y^\star, T^\star, X)] = 1, \tag{A.49}$$

where the final equality follows from Q^{a} being a probability measure, condition (A.46), the law of iterated expectations, and Q_j being observationally equivalent to Q_0 due to $Q_j \in \Theta_0$ for all $1 \leq j \leq J$. Given the already verified positivity of $\tilde{Q}$ (for η sufficiently small), result (A.49) implies that $\tilde{Q}$ is indeed a probability measure. Also note that since $Q_j \ll \mu$ due to $Q_j \in \Theta_0$ for all $1 \leq j \leq J$, it follows $\tilde{Q} \ll \mu$ and $d\tilde{Q}/d\mu = dQ^{\mathrm{a}}/d\mu + \eta \sum_j s_j dQ_j/d\mu$. Thus, since $dQ^{\mathrm{a}}/d\mu$ belongs to the interior of $\mathcal{Q}$ in $\mathbf{Q}$ and $s_j dQ_j/d\mu \in \mathbf{Q}$ by definition of $\mathcal{S}_{Q_j}$, it follows that $d\tilde{Q}/d\mu \in \mathcal{Q}$ for $\eta > 0$ small.

We next show that $\tilde{Q}$ is observationally equivalent to Q_0. To this end, note that Assumption 2.1(ii) and $Q^{\mathrm{a}} \in \Theta_0$ imply for any $t \in \mathbf{T}$ and (measurable) set V that

$$P(T = t, (Y, Z, X) \in V) = Q^{\mathrm{a}}(T^\star(Z) = t, (Y^\star(t), Z, X) \in V). \tag{A.50}$$

However, for any $t \in \mathbf{T}$ and (measurable) set V, Assumption 2.1(ii) also yields that

$$\begin{aligned}\sum_{j=1}^J E_{Q_j}[s_j(Y^\star, T^\star, X) 1\{T^\star(Z) = t, (Y^\star(t), Z, X) \in V\}] \\ = \sum_{j=1}^J E_{Q_j}[E_{Q_j}[s_j(Y^\star, T^\star, X)|Y, T, Z, X] 1\{T = t, (Y, Z, X) \in V\}] = 0,\end{aligned} \tag{A.51}$$

where the final equality follows from condition (A.46) and Q_j being observationally equivalent to Q_0 due to $Q_j \in \Theta_0$ for all $1 \leq j \leq J$. Hence, (A.50), (A.51), and the definition of $\tilde{Q}$ imply that $\tilde{Q}$ is indeed observationally equivalent to Q_0.

To conclude the proof, it only remains to show that $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under $\tilde{Q}$. To this end, select an $f \in L^1(P_{ZX})$ and let g be any bounded function of $(Y^\star, T^\star, X)$. Then note that since $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q^{a} and all Q_j (due to $Q^{\mathrm{a}}, Q_j \in \Theta_0$) we obtain

$$\begin{aligned}E_{\tilde{Q}}[g(Y^\star, T^\star, X) f(Z, X)] = E_{Q^{\mathrm{a}}}[g(Y^\star, T^\star, X) E_{Q^{\mathrm{a}}}[f(Z, X)|X]] \\ + \eta \sum_{j=1}^J E_{Q_j}[g(Y^\star, T^\star, X) s_j(Y^\star, T^\star, X) E_{Q_j}[f(Z, X)|X]].\end{aligned} \tag{A.52}$$

However, since we showed $\tilde{Q}$ is observationally equivalent to Q_0 and Q^{a} and Q_j are observationally equivalent to Q_0 (due to $Q^{\mathrm{a}}, Q_j \in \Theta_0$) it also follows that $E_{Q^{\mathrm{a}}}[f(Z, X)|X] = E_{Q_j}[f(Z, X)|X] = E_{\tilde{Q}}[f(Z, X)|X]$. Combining this observation with (A.52) then yields

$$E_{\tilde{Q}}[g(Y^\star, T^\star, X) f(Z, X)] = E_{\tilde{Q}}[g(Y^\star, T^\star, X) E_{\tilde{Q}}[f(Z, X)|X]]. \tag{A.53}$$

Since (A.53) holds for any bounded g it follows $E_{\tilde{Q}}[f(Z, X)|Y^\star, T^\star, X] = E_{\tilde{Q}}[f(Z, X)|X]$; see, e.g., Definition 10.1.1 in Bogachev (2007). Thus, since $f \in L^1(P_{ZX})$ was also arbitrary, we conclude $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under $\tilde{Q}$ and therefore that $\tilde{Q} \in \Theta_0$. ■

Lemma A.4. *Let Assumptions 2.1, 2.2 hold,* $\Lambda : \mathcal{L} \to \mathbf{R}$ *satisfy* $\Lambda(\langle \cdot, s \rangle_Q) = \langle \ell, s \rangle_Q$ *for* *some* $\ell \in \bigcap_{Q\in\Theta_0} L^1(Q)$*,* $\mathcal{S}_{\bar{Q}} = L^\infty(\bar{Q}_{Y^\star T^\star X})$*, and* $\mathcal{C} \subseteq \mathcal{R}$ *be convex. If* Λ *belongs to the* τ*-closure of* $\mathcal{C}$*, then there is a sequence* $\{f_n\}_{n=1}^\infty \subseteq L^1(\bar{Q}_{Y^\star T^\star X})$ *such that* $\lim_{n\to\infty} \|\ell - f_n\|_{\bar{Q},1} = 0$ *and* $L_n : \mathcal{L} \to \mathbf{R}$ *given by* $L_n(\langle \cdot, s\rangle_Q) \equiv \langle f_n, s\rangle_Q$ *satisfies* $L_n \in \mathcal{C}$ *for all* n*.*

Proof. First note that for any $L \in \mathcal{C}$, it follows from $\mathcal{C} \subseteq \mathcal{R}$ that there is a $\psi(L) \in \bigcap_{Q\in\Theta_0} L^1(Q)$ such that $L(\langle \cdot, s\rangle_Q) = \langle \psi(L), s\rangle_Q$ for all $s \in \mathcal{S}_Q$ and $Q \in \Theta_0$. Next define $C \equiv \{f \in \bigcap_{Q\in\Theta_0} L^1(Q) : f = \psi(L)$ for some $L \in \mathcal{C}\}$ and note that by Theorem 2.14 in Aliprantis and Border (2006) and Λ belonging to the τ-closure of $\mathcal{C}$, there must exist a net $\{c_\alpha\}_{\alpha\in\mathcal{A}} \subseteq C$ such that for all $s \in \mathcal{S}_{\bar{Q}} = L^\infty(\bar{Q}_{Y^\star T^\star X})$ we have

$$\lim_\alpha \langle s, c_\alpha\rangle_{\bar{Q}} = \langle s, \ell\rangle_{\bar{Q}}. \tag{A.54}$$

A second application of Theorem 2.14 in Aliprantis and Border (2006) and result (A.54) imply $\ell \in L^1(\bar{Q}_{Y^\star T^\star X})$ belongs to the closure of $C \subseteq L^1(\bar{Q}_{Y^\star T^\star X})$ under the weak topology generated by $\mathcal{S}_{\bar{Q}} = L^\infty(\bar{Q}_{Y^\star T^\star X})$. However, since $L^\infty(\bar{Q}_{Y^\star T^\star X})$ is the norm dual of $L^1(\bar{Q}_{Y^\star T^\star X})$ and $C \subseteq L^1(\bar{Q}_{Y^\star T^\star X})$ is convex (by convexity of $\mathcal{C}$), it follows that ℓ belongs to the closure of C under $\|\cdot\|_{\bar{Q},1}$ (see, e.g., Theorem 3.12 in Rudin (1991)). Therefore, by Theorem 2.40(1) in Aliprantis and Border (2006) we can conclude that there is a sequence $\{f_n\} \subseteq C$ satisfying $\|\ell - f_n\|_{\bar{Q},1} = o(1)$ as claimed. ■

Lemma A.5. *Let Assumptions 2.1 and 2.2 hold,* $\mathbf{T}^\star$ *be finite,* $\{C_i\}_{i=1}^r$ *be a finite collection of subsets of* $\mathbf{T}^\star$*, and define* $A : \bigotimes_{i=1}^r L^1(P_X) \to L^1(\bar{Q}_{T^\star X})$ *according to*

$$A(f)(T^\star, X) = \sum_{i=1}^r 1\{T^\star \in C_i\} f_i(X)$$

for any $f = \{f_i\}_{i=1}^r \in \bigotimes_{i=1}^r L^1(P_X)$*. If* $\bar{Q}(T^\star = t^\star|X) \geq \varepsilon > 0$ *a.s. for any* $t^\star \in \mathbf{T}^\star$ *satisfying* $\bar{Q}(T^\star = t^\star) > 0$*, then it follows that the range of* A *is closed.*

Proof. First let $\mathbf{T}_0^\star \equiv \{t^\star \in \mathbf{T}^\star : \bar{Q}(T^\star = t^\star) > 0\}$ denote the support of $T^\star$ under $\bar{Q}$ and enumerate $\mathbf{T}_0^\star$ by $\mathbf{T}_0^\star = \{t_1^\star, \ldots, t_{d^\star}^\star\}$. Further interpret any $\{f_i\}_{i=1}^r \in \bigotimes_{i=1}^r L^1(P_X)$ as a column vector $f(X) \equiv (f_1(X), \ldots, f_r(X))'$ and define a $d^\star \times r$ matrix Ω according to

$$\Omega \equiv \begin{pmatrix} \omega_1' \\ \vdots \\ \omega_{d^\star}' \end{pmatrix} \qquad \omega_j \equiv \begin{pmatrix} 1\{t_j^\star \in C_1\} \\ \vdots \\ 1\{t_j^\star \in C_r\} \end{pmatrix}.$$

Letting $\|a\|_1 \equiv \sum_{i=1}^d |a_i|$ for any $a \equiv (a_1, \ldots, a_d)' \in \mathbf{R}^d$, then note that $\bar{Q}(T^\star = t^\star|X) \geq$

$\varepsilon > 0$ for all $t^\star \in \mathbf{T}_0^\star$ and $\bar{Q}_X = P_X$ due to $\bar{Q} \in \Theta_0$ allow us to conclude

$$E_{\bar{Q}}[|A(f)(T^\star, X)|] = E_{\bar{Q}}[\sum_{j=1}^{d^\star} 1\{T^\star = t_j^\star\}|\sum_{i=1}^{r} 1\{t_j^\star \in C_i\} f_i(X)|]$$
$$= E_{P_X}[\sum_{j=1}^{d^\star} \bar{Q}(T^\star = t_j^\star | X)|\omega_j' f(X)|] \geq \varepsilon \times E_{P_X}[\|\Omega f(X)\|_1]. \quad \text{(A.55)}$$

Let $\Omega^\dagger$ denote the Moore-Penrose pseudoinverse of Ω and note that since $\Omega\Omega^\dagger\Omega = \Omega$ by Proposition 6.11.1(6) in Luenberger (1969) and $\Omega : \text{range}\{\Omega^\dagger\} \to \mathbf{R}^{d^\star}$ is an invertible map (see Chapter 6.11 in Luenberger (1969)), it follows there is an $\eta > 0$ satisfying

$$E_{P_X}[\|\Omega f(X)\|_1] \geq \eta E_{P_X}[\|\Omega^\dagger \Omega f(X)\|_1]. \quad \text{(A.56)}$$

Next suppose there is a sequence $\{f_n\} \in \bigotimes_{i=1}^r L^1(P_X)$ and an $\ell \in L^1(\bar{Q}_{T^\star X})$ such that $\|A(f_n) - \ell\|_{\bar{Q},1} = o(1)$. Combining (A.55) and (A.56) implies the sequence $\{\Omega^\dagger\Omega f_n\}$ is Cauchy in $\bigotimes_{i=1}^r L^1(P_X)$ under the norm $E_{P_X}[\|f(X)\|_1]$. Hence, since $\bigotimes_{i=1}^r L^1(P_X)$ is complete we can conclude that there is an $\tilde{f} \in \bigotimes_{i=1}^r L^1(P_X)$ such that

$$\lim_{n\to\infty} E_{P_X}[\|\Omega^\dagger\Omega f_n(X) - \tilde{f}(X)\|_1] = 0. \quad \text{(A.57)}$$

Therefore, the same manipulations as in result (A.55) and $\Omega = \Omega\Omega^\dagger\Omega$ imply that

$$E_{\bar{Q}}[|\ell(T^\star, X) - A(\tilde{f})(T^\star, X)|] \leq \lim_{n\to\infty} E_{\bar{Q}}[|A(f_n - \tilde{f})(T^\star, X)|]$$
$$\leq \lim_{n\to\infty} E_{P_X}[\|\Omega f_n(X) - \Omega\tilde{f}(X)\|_1] \leq \|\Omega\|_{o,1} E_{P_X}[\|\Omega^\dagger\Omega f_n(X) - \tilde{f}(X)\|_1] = 0,$$

where the final inequality holds for $\|\cdot\|_{o,1}$ the operator norm of $\Omega : \mathbf{R}^r \to \mathbf{R}^{d^\star}$ when both the range and domain are endowed with $\|\cdot\|_1$, while the final equality follows by (A.57). We can thus conclude the range of A is closed, which establishes the lemma. ■

A.2 Proofs for Sections 5.1 and 5.2

Proof of Theorem 5.1. We begin by noting that by definition of $\hat{\lambda}$ and $\hat{\psi}_k$ in (37) we may apply Lemma A.6 with W_i satisfying $P(W_i = 1) = 1$ to obtain that

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}(\psi(T_i, Z_i, X_i) + \lambda_{Q_0}) + o_P(\frac{\sigma}{\sqrt{n}}), \quad \text{(A.58)}$$

which establishes the first equality in (36). Since $\|\psi\|_\infty \lesssim B$ by Assumption 5.2(i) and

$\sigma^2 = \text{Var}_P\{\psi(T,Z,X)\}$ by definition, Theorem 1.1 in Zhai (2018) further yields

$$\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(\psi(T_i,Z_i,X_i) - E_P[\psi(T,Z,X)]) = \mathbb{Z} + O_P(\frac{B\log(n)}{\sigma\sqrt{n}}) \tag{A.59}$$

for $\mathbb{Z}$ a standard normal random variable possibly depending on n. The theorem follows from (A.58), (A.59), Lemma A.7, and $B\log(n) = o(\sigma\sqrt{n})$ by Assumption 5.2(i). ■

Proof of Theorem 5.2. First note that Lemma A.6, $(\hat{\lambda} - \lambda_{Q_0}) = O_P(\sigma/\sqrt{n})$ by Theorem 5.1, and $\sum_i W_i/n = o_P(1)$ due to $E[W] = 0$ together allow us to conclude that

$$\begin{aligned}\frac{\sqrt{n}}{\sigma}(\hat{\lambda}^* - \hat{\lambda}) &= \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}W_i\psi(T_i,Z_i,X_i) + \frac{1}{n}\sum_{i=1}^{n}W_i \times \frac{\sqrt{n}}{\sigma}(\lambda_{Q_0} - \hat{\lambda}) + o_P(1)\\ &= \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}W_i\psi(T_i,Z_i,X_i) + o_P(1).\end{aligned} \tag{A.60}$$

Next note that $\sigma^2 \equiv \text{Var}_P\{\psi(T,Z,X)\}$ by definition, $\|\psi\|_\infty \lesssim B$ and $B\log(n) = o(\sigma\sqrt{n})$ by Assumption 5.2(i), and Bernstein's inequality (see, e.g., Lemma 2.2.9 in van der Vaart and Wellner (1996)), yield that for n sufficiently large

$$P(|\frac{1}{n\sigma}\sum_{i=1}^{n}(\psi(T_i,Z_i,X_i) - E_P[\psi(T,Z,X)])| > \frac{\log(n)}{\sqrt{n}}) \le 2\exp\{-\frac{\log^2(n)}{4}\}. \tag{A.61}$$

In particular, note that result (A.61) together with Lemma A.7 imply that we have

$$|\frac{1}{n}\sum_{i=1}^{n}\{\psi^2(T_i,Z_i,X_i) - (\psi(T_i,Z_i,X_i) - E_P[\psi(T,Z,X)])^2\}| = o_P(\frac{\sigma^2\log(n)}{n}). \tag{A.62}$$

Moreover, Markov's inequality, $\|\psi\|_\infty \lesssim B$, and $\sigma^2 \equiv \text{Var}_P\{\psi(T,Z,X)\}$ further imply

$$\begin{aligned}P(|\frac{1}{n}\sum_{i=1}^{n}\frac{(\psi(T_i,Z_i,X_i) - E_P[\psi(T,Z,X)])^2}{\sigma^2} - 1| > \epsilon)&\\ \le \frac{1}{n\epsilon^2\sigma^4}\text{Var}_P\{(\psi(T,Z,X) - E_P[\psi(T,Z,X)])^2\} &\lesssim \frac{B^2\sigma^2}{n\sigma^4}\end{aligned} \tag{A.63}$$

for any $\epsilon > 0$. Therefore, setting $\hat{\sigma}^2 \equiv \sum_i \psi^2(T_i,Z_i,X_i)/n$, we can conclude from results (A.62) and (A.63) together with $B\log(n) = o(\sigma\sqrt{n})$ by Assumption 5.2(i) that

$$\frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{(\psi(T,Z,X) - E_P[\psi(T,Z,X)])^2}{\sigma^2} + o_P(1) = 1 + o_P(1). \tag{A.64}$$

To conclude, note that $\{W_i\}_{i=1}^n$ being i.i.d. standard normal random variables and

$\{W_i\}_{i=1}^n$ being independent of the sample $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$ imply that the variable

$$\mathbb{Z}^* \equiv \frac{1}{\sqrt{n}\hat{\sigma}} \sum_{i=1}^n W_i \psi(T_i, Z_i, X_i)$$

satisfies $\mathbb{Z}^* \sim N(0,1)$ conditionally on $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$ and hence $\mathbb{Z}^*$ is independent of $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$. The theorem therefore follows from (A.60) and (A.64). ■

Proof of Theorem 5.3. The proof follows by identical arguments as those employed in Theorem 5.1 but relying on Lemmas A.8 and A.9 in place of A.6 and A.7. ■

Proof of Theorem 5.4. The proof follows by identical arguments as those employed in Theorem 5.1 but relying on Theorem 5.3 and Lemmas A.8 and A.9 in place of Theorem 5.1 and Lemmas A.6 and A.7. ■

Lemma A.6. *Let Assumptions 5.1(i)(iii) and 5.2(i)(ii)(v) hold, $\{W_i\}_{i=1}^n$ be an i.i.d. sequence independent of $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$ satisfying $E[W^2] < \infty$, ψ and $\hat{\psi}_k$ be as defined in (35) and (37) respectively, and $\sigma^2 \equiv \text{Var}_P\{\psi(T, Z, X)\}$. Then it follows that*

$$\frac{1}{n}\sum_{k=1}^K \sum_{i \in I_k} W_i\{\hat{\psi}_k(T_i, Z_i, X_i) + \hat{\lambda}_n\} = \frac{1}{n}\sum_{i=1}^n W_i\{\psi(T_i, Z_i, X_i) + \lambda_{Q_0}\} + o_P(\frac{\sigma}{\sqrt{n}}).$$

Proof. Let $\hat{\Delta}_{t,k}^\gamma \equiv (\hat{\gamma}_{t,k} - \gamma_t)$, $\hat{\Delta}_{t,k}^\beta \equiv (\hat{\beta}_{t,k} - \beta_t)$, and note that for $1 \leq k \leq K$ we have

$$\begin{aligned} &\frac{1}{n}\sum_{i \in I_k} W_i\{(\hat{\psi}_k(T_i, Z_i, X_i) + \hat{\lambda}) - (\psi(T_i, Z_i, X_i) + \lambda_{Q_0})\} \\ &\qquad = \frac{1}{n}\sum_{i \in I_k}\sum_{t \in \mathbf{T}} W_i(1\{T_i = t\} - b(Z_i, X_i)'\beta_t) b(Z_i, X_i)'\hat{\Delta}_{t,k}^\gamma \\ &\qquad + \frac{1}{n}\sum_{i \in I_k}\sum_{t \in \mathbf{T}} W_i\{E_{\mu_{Z|X}}[\nu_j(t, Z, X_i) b(Z, X_i)'] - b(Z_i, X_i)'\gamma_t b(Z_i, X_i)'\}\hat{\Delta}_{t,k}^\beta \\ &\qquad - \frac{1}{n}\sum_{i \in I_k}\sum_{t \in \mathbf{T}} W_i(\hat{\Delta}_{t,k}^\gamma)' b(Z_i, X_i) b(Z_i, X_i)'\hat{\Delta}_{t,k}^\beta. \end{aligned} \tag{A.65}$$

Next observe that by definition of β_t, it must satisfy the following first order condition

$$E_P[(1\{T = t\} - b(Z, X)'\beta_t) b(Z, X)] = 0. \tag{A.66}$$

Hence, since $\hat{\gamma}_{t,k}$ is computed using the observations in I_k^c, which are independent of the observations in I_k, and $\{W_i\}_{i=1}^n$ is independent of $\{Y_i, T_i, Z_i, X_i\}_{i=1}^n$, (A.66) yields

$$E[\frac{1}{n}\sum_{i \in I_k}\sum_{t \in \mathbf{T}} W_i(1\{T_i = t\} - b(Z_i, X_i)'\beta_t) b(Z_i, X_i)'\hat{\Delta}_{t,k}^\gamma | \{\hat{\gamma}_{t,k}\}_{t \in \mathbf{T}}] = 0.$$

Therefore, by employing that the observations within I_k are i.i.d., $|I_k| \asymp n$ by Assumption 5.2(v), $\|b'\beta_t\|_\infty$ being bounded by Assumption 5.2(i), and $E[W^2] < \infty$ we obtain

$$\mathrm{Var}\{\frac{1}{n}\sum_{i\in I_k}\sum_{t\in\mathbf{T}} W_i(1\{T_i = t\} - b(Z_i,X_i)'\beta_t)b(Z_i,X_i)'\hat{\Delta}^{\gamma}_{t,k}|\{\hat{\gamma}_{t,k}\}_{t\in\mathbf{T}}\}$$
$$\lesssim \frac{1}{n}\sum_{t\in\mathbf{T}} E_P[(b(Z,X)'\hat{\Delta}^{\gamma}_{t,k})^2]. \quad (A.67)$$

Moreover, by identical arguments but relying on the first order condition for γ_t yields

$$\mathrm{Var}\{\frac{1}{n}\sum_{i\in I_k}\sum_{t\in\mathbf{T}} W_i\{E_{\mu_{Z|X}}[\nu_j(t,Z,X_i)b(Z,X_i)'] - b(Z_i,X_i)'\gamma_t b(Z_i,X_i)'\}\hat{\Delta}^{\beta}_{t,k}|\{\hat{\beta}_{t,k}\}_{t\in\mathbf{T}}\}$$
$$\lesssim \frac{1}{n}\sum_{t\in\mathbf{T}}\{E_P[(E_{\mu_{Z|X}}[\nu_j(t,Z,X)b(Z,X)'\hat{\Delta}^{\beta}_{t,k}])^2] + E_P[(b(Z,X)'\gamma_t)^2(b(Z,X)'\hat{\Delta}^{\beta}_{t,k})^2]$$
$$\lesssim \frac{B^2}{n}\sum_{t\in\mathbf{T}} E_P[(b(Z,X)'\hat{\Delta}^{\beta}_{t,k})^2], \quad (A.68)$$

where in the final inequality we employed Jensen's inequality, Assumptions 5.1(iii) and 5.2(i), and $d\mu_{Z|X}/dP_{Z|X} = 1/\pi$. Finally, note that the Cauchy-Schwarz inequality yields

$$|\frac{1}{n}\sum_{i\in I_k}\sum_{t\in\mathbf{T}} W_i(\hat{\Delta}^{\gamma}_{t,k})'b(Z_i,X_i)b(Z_i,X_i)'\hat{\Delta}^{\beta}_{t,k}|$$
$$\leq \{\frac{1}{n}\sum_{i\in I_k} W_i^2(b(Z_i,X_i)'\hat{\Delta}^{\gamma}_{t,k})^2\}^{1/2} \times \{\frac{1}{n}\sum_{i\in I_k}(b(Z_i,X_i)'\hat{\Delta}^{\beta}_{t,k})^2\}^{1/2}. \quad (A.69)$$

Therefore, combining results (A.65), (A.67), (A.68), and (A.69), $E[W_i^2] < \infty$, $\{W_i\}_{i=1}^n$ being independent of $\{Y_i,T_i,Z_i,X_i\}_{i=1}^n$, and Markov's inequality we obtain that

$$\frac{1}{n}\sum_{i\in I_k} W_i\{(\hat{\psi}_k(T_i,Z_i,X_i) + \hat{\lambda}_n) - (\psi(T_i,Z_i,X_i) + \lambda_{Q_0})\}$$
$$= O_P(\sum_{t\in\mathbf{T}}\frac{r_t^{\gamma} + r_t^{\beta}B}{\sqrt{n}} + r_t^{\beta}r_t^{\gamma}) = o_P(\frac{\sigma}{\sqrt{n}}),$$

where the final equality follows from Assumption 5.2(ii). ■

Lemma A.7. *If Assumptions 5.1(ii), 5.2(iii)(iv) hold, and ψ is as in (35), then*

$$|E_P[\psi(T,Z,X)]| = o(\frac{\sigma}{\sqrt{n}}).$$

Proof. First note that the first order condition implied by the definition of γ_t yields

$$E_P[(b(Z,X)'\gamma_t)(b(Z,X)'\beta_t)] - E_{P_X}[E_{\mu_{Z|X}}[\nu_j(T,Z,X)b(Z,X)'\beta_t]] = 0. \quad (A.70)$$

Hence, by combining the definition of ψ, the first order condition in (A.70), the law of iterated expectations, and $\kappa_j = \nu_j/\pi$ we are able to conclude that

$$\begin{aligned}
&|E_P[\psi(T,Z,X)] - (E_P[\kappa_j(T,Z,X)] - \lambda_{Q_0})| \\
&\quad = |\sum_{t\in\mathbf{T}} E_P[(b(Z,X)'\gamma_t - \kappa_j(t,Z,X))1\{T=t\}]| \\
&\quad = |\sum_{t\in\mathbf{T}} (E_P[(b(Z,X)'\gamma_t - \kappa_j(t,Z,X))(P(T=t|Z,X) - b(Z,X)'\beta_t)]|. \qquad \text{(A.71)}
\end{aligned}$$

Hence, result (A.71), the Cauchy-Schwarz inequality, and Assumptions 5.2(iii)(iv) yield

$$|E_P[\psi(T,Z,X)]| \leq |E_P[\kappa_j(T,Z,X) - \lambda_{Q_0}| + O(\sum_{t\in\mathbf{T}} \delta_t^\beta \delta_t^\gamma) = o(\frac{\sigma}{\sqrt{n}}),$$

which establishes the claim of the lemma. ■

Lemma A.8. *Let Assumptions 5.3(i)(iii) and 5.4(i)(ii)(v) hold, $\{W_i\}_{i=1}^n$ be an i.i.d. sequence independent of $\{Y_i,T_i,Z_i,X_i\}_{i=1}^n$ satisfying $E[W^2]<\infty$, ψ and $\hat{\psi}_k$ be as defined in (47) and (49) respectively, and $\sigma^2 \equiv \mathrm{Var}_P\{\psi(Y,T,Z,X)\}$. Then it follows that*

$$\frac{1}{n}\sum_{k=1}^K \sum_{i\in I_k} W_i\{\hat{\psi}_k(Y_i,T_i,Z_i,X_i) + \hat{\lambda}\} = \frac{1}{n}\sum_{i=1}^n W_i\{\psi(Y_i,T_i,Z_i,X_i) + \lambda_{Q_0}\} + o_P(\frac{\sigma}{\sqrt{n}}).$$

Proof. The proof follows from identical arguments to those in Lemma A.6. ■

Lemma A.9. *If Assumptions 5.3(ii), 5.4(iii)(iv) hold, and ψ is as in (47), then*

$$|E_P[\psi(Y,T,Z,X)]| = o(\frac{\sigma}{\sqrt{n}}).$$

Proof. The proof follows from identical arguments to those in Lemma A.7. ■

Lemma A.10. *Let $V \equiv (Y,T,Z,X)$, $\{V_i\}_{i=1}^n$ be i.i.d., $\{W_i\}_{i=1}^n$ be i.i.d. with $W \sim N(0,1)$ independent of $\{V_i\}_{i=1}^n$, and $\phi : \mathbf{R}^q \to \mathbf{R}^p$ be differentiable at $(\lambda_{Q_01},\ldots,\lambda_{Q_0q})' \equiv \lambda_{Q_0} \in \mathbf{R}^q$ with derivative $\phi'_{\lambda_{Q_0}}$. Suppose $\hat{\lambda} \equiv (\hat{\lambda}_1,\ldots,\hat{\lambda}_q)'$ and $\hat{\lambda}^* \equiv (\hat{\lambda}_1^*,\ldots,\hat{\lambda}_q^*)'$ satisfy*

$$\frac{\sqrt{n}}{\sigma_j}(\hat{\lambda}_j - \lambda_{Q_0j}) = \frac{1}{\sqrt{n}\sigma_j}\sum_{i=1}^n \psi_j(V_i) + o_P(1) \qquad \frac{\sqrt{n}}{\sigma_j}(\hat{\lambda}_j^* - \hat{\lambda}_j) = \frac{1}{\sqrt{n}\sigma_j}\sum_{i=1}^n W_i\psi_j(V_i) + o_P(1)$$

with $\sigma_j^2 \equiv \mathrm{Var}_P\{\psi_j(V)\}$ and let $\bar{\sigma} \equiv \max_{1\leq j\leq q}\sigma_j$. If $B \equiv \max_{1\leq j\leq q}\|\psi_j\|_\infty < \infty$, $E_P[\psi_j(V)] = o(\bar{\sigma}/\sqrt{n})$, $B\log(n) = o(\bar{\sigma}\sqrt{n})$, and $\bar{\sigma} = o(\sqrt{n})$, then it follows

$$\frac{\sqrt{n}}{\bar{\sigma}}(\phi(\hat{\lambda}) - \phi(\lambda_{Q_0})) = \phi'_{\lambda_{Q_0}}(\mathbb{G}) + o_P(1) \qquad \text{(A.72)}$$

$$\frac{\sqrt{n}}{\bar{\sigma}}(\phi(\hat{\lambda}^*) - \phi(\hat{\lambda})) = \phi'_{\lambda_{Q_0}}(\mathbb{G}^*) + o_P(1), \qquad \text{(A.73)}$$

where $\mathbb{G}$ and $\mathbb{G}^*$ have the same distribution and $\mathbb{G}^*$ is independent of $\{V_i\}_{i=1}^n$.

Proof. First set $\psi(V) \equiv (\psi_1(V), \ldots, \psi_q(V))'$ and note that $\sigma_j/\bar{\sigma} \leq 1$ by definition of $\bar{\sigma}$ and our requirements on $(\hat{\lambda}_j - \lambda_{Q_0 j})$ together allow us to conclude that

$$\frac{\sqrt{n}}{\bar{\sigma}}(\hat{\lambda} - \lambda_{Q_0}) = \frac{1}{\sqrt{n}\bar{\sigma}} \sum_{i=1}^{n} \psi(V_i) + o_P(1)$$
$$= \frac{1}{\sqrt{n}\bar{\sigma}} \sum_{i=1}^{n} (\psi(V_i) - E_P[\psi(V)]) + o_P(1) = \mathbb{G} + o_P(1), \quad \text{(A.74)}$$

where the second equality holds due to $E_P[\psi_j(V)] = o(\bar{\sigma}/\sqrt{n})$ by hypothesis, and the final equality holds for some Gaussian $\mathbb{G} \sim N(0, \mathrm{Var}_P\{\psi(V)\}/\bar{\sigma}^2)$ by Theorem 1.1 in Zhai (2018) and $B\log(n)/\bar{\sigma}\sqrt{n} = o(1)$ by hypothesis. In particular, note that since the variance of each coordinate of $\mathbb{G}$ is bounded by one, we must have $\|\mathbb{G}\| = O_P(1)$ and therefore (A.74) implies $\|\hat{\lambda} - \lambda_{Q_0}\| = O_P(\bar{\sigma}/\sqrt{n})$. Hence, $\phi : \mathbf{R}^q \to \mathbf{R}^p$ being differentiable by assumption together with result (A.74) allow us to conclude

$$\frac{\sqrt{n}}{\bar{\sigma}}(\phi(\hat{\lambda}) - \phi(\lambda_{Q_0})) = \frac{\sqrt{n}}{\bar{\sigma}}\phi'_{\lambda_{Q_0}}(\hat{\lambda} - \lambda_{Q_0}) + \frac{\sqrt{n}}{\bar{\sigma}} \times o(\|\hat{\lambda} - \lambda_{Q_0}\|) = \phi'_{\lambda_{Q_0}}(\mathbb{G}) + o_P(1),$$

where in the final result we employed that $\bar{\sigma}/\sqrt{n} = o(1)$ by hypothesis and $\|\hat{\lambda} - \lambda_{Q_0}\| = O_P(\bar{\sigma}/\sqrt{n})$ as already shown. Thus, claim (A.72) holds.

To establish claim (A.73) we first note that since $E_P[\psi_j(V)] = o(\bar{\sigma}/\sqrt{n})$ by hypothesis and $\sqrt{n}\bar{W}_n = O_P(1)$ due to $W \sim N(0,1)$, we can conclude that

$$\frac{1}{\sqrt{n}\bar{\sigma}} \sum_{i=1}^{n} W_i \times \frac{1}{n} \sum_{i=1}^{n} \psi_j(V_i)$$
$$= \frac{1}{\sqrt{n}\bar{\sigma}} \sum_{i=1}^{n} W_i \times \frac{1}{n} \sum_{i=1}^{n} (\psi_j(V_i) - E_P[\psi_j(V)]) + o_P(1) = o_P(1)$$

where the final equality holds due to $\sqrt{n}\bar{W}_n = O_P(1)$, Chebychev's inequality, and $\sigma_j/\bar{\sigma} \leq 1$. Therefore, it follows from our condition on $(\hat{\lambda}^* - \hat{\lambda})$ that we must have

$$\frac{\sqrt{n}}{\bar{\sigma}}(\hat{\lambda}^* - \hat{\lambda}) = \frac{1}{\sqrt{n}\bar{\sigma}} \sum_{i=1}^{n} W_i(\psi(V_i) - \frac{1}{n}\sum_{k=1}^{n} \psi(V_k)) + o_P(1)$$
$$= \mathbb{G}^* + o_P(1), \quad \text{(A.75)}$$

where the final equality holds for some $\mathbb{G}^* \sim N(0, \mathrm{Var}_P\{\psi(V)/\bar{\sigma}\})$ independent of the data by Theorem S.7.1 in Chernozhukov et al. (2022b) – to apply said theorem set, in their notation, $f^{d_n}_{n,P}(V) = (\psi(V) - E_P[\psi(V)])/\bar{\sigma}$ and note that then $C_n = O(1)$ due to $\sigma_j/\bar{\sigma} \leq 1$, $K_n \asymp B/\bar{\sigma}$, $J_{1n} = 0$, and $J_{2n} = O(1)$. Claim (A.73) of the lemma then follows by identical arguments to those employed in showing claim (A.72) but relying on result (A.75) in place of result (A.74). ∎

A.3 Proofs for Section 5.3

Proof of Theorem 5.5. Let $\eta \mapsto Q_{\eta,g}$ be a submodel with $Q_{0,g} = Q \in \Theta_0$ inducing a path $\eta \mapsto P_{\eta,s}$. Then note that by Proposition 4 in Le Cam and Yang (1988) we have

$$s(Y,T,Z,X) = E_Q[g(Y^\star,T^\star,Z,X)|Y,T,Z,X]. \tag{A.76}$$

Next define $T_1(Q) \equiv \{g \in L^2(Q_{Y^\star T^\star X}) : E_Q[g(Y^\star,T^\star,X)|Z,X] = 0\}$ and note Lemma A.12 implies that $g = g_1 + g_2$ for some $g_1 \in T_1(Q)$ and $g_2 \in L_0^2(P_{ZX})$. Further set

$$\Delta(Y,T,Z,X) \equiv \kappa(Y,T,Z,X) - E_P[\kappa(Y,T,Z,X)|Z,X] + E_P[\kappa(Y,T,Z,X)|X]$$

and note that the law of iterated expectations and the definition of $T_1(Q)$ yield that

$$\begin{aligned} E_Q[\kappa(Y,T,Z,X)g_1(Y^\star,T^\star,X)] &= E_Q[\Delta(Y,T,Z,X)g_1(Y^\star,T^\star,X)] \\ E_P[E_P[\kappa(Y,T,Z,X)|X]g_2(Z,X)] &= E_P[\Delta(Y,T,Z,X)g_2(Z,X)]. \end{aligned} \tag{A.77}$$

Next observe that $g = g_1 + g_2$ and Lemma F.1 in Chen and Santos (2018) yield that

$$\frac{\partial}{\partial\eta}\lambda_{Q_{\eta,g}}\Big|_{\eta=0} = E_Q[\ell(Y^\star,T^\star,X)(g_1(Y^\star,T^\star,X) + g_2(Z,X))]. \tag{A.78}$$

Therefore, $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$, $Q \ll \bar{Q}$ for any $Q \in \Theta_0$, Corollary A.1, results (A.77) and (A.78), the law of iterated expectations, and $(Y^\star,T^\star) \perp\!\!\!\perp Z|X$ under Q imply

$$\begin{aligned} \frac{\partial}{\partial\eta}\lambda_{Q_{\eta,g}}\Big|_{\eta=0} &= E_Q[E_Q[\kappa(Y,T,Z,X)|Y^\star,T^\star,X](g_1(Y^\star,T^\star,X) + g_2(Z,X))] \\ &= E_Q[\Delta(Y,T,Z,X)(g_1(Y^\star,T^\star,X) + g_2(Z,X))] \\ &= E_P[\Delta(Y,T,Z,X)s(Y,T,Z,X)], \end{aligned} \tag{A.79}$$

where the final equality follows from Q inducing P due to $Q \in \Theta_0$, the law of iterated expectations, and result (A.76). For any closed linear subspace V of a Hilbert space $\mathbf{H}$ and $f \in \mathbf{H}$ we let $\text{Proj}\{f|V\}$ denote the projection of f onto V (understood to be with respect to the norm $\|\cdot\|_{\mathbf{H}}$ of $\mathbf{H}$). Then note that for $T(P)$ the tangent set and $\bar{T}(P)$ the tangent space (as defined in, e.g., Theorem A.1) result (A.79) yields that

$$\begin{aligned} \sup_{Q_{\cdot,g}} I^{-1}(Q_{\cdot,g}) &= \sup_{0\neq s\in T(P)} \frac{(E_P[\Delta(Y,T,Z,X)s(Y,T,Z,X)])^2}{E_P[s^2(Y,T,Z,X)]} \\ &= \sup_{0\neq s\in \bar{T}(P)} \frac{(E_P[\Delta(Y,T,Z,X)s(Y,T,Z,X)])^2}{E_P[s^2(Y,T,Z,X)]} = \|\text{Proj}\{\Delta|\bar{T}(P)\}\|_{P,2}^2, \end{aligned} \tag{A.80}$$

where the second equality follows by continuity of the objective in s under $\|\cdot\|_{P,2}$.

Moreover, employing that $\bar{T}(P) = [N(\mathcal{I})]^{\perp} \oplus L_0^2(P_{ZX})$ by Theorem A.1 we obtain

$$\begin{aligned}\text{Proj}\{\kappa|\bar{T}(P)\} &= \text{Proj}\{\kappa|[N(\mathcal{I})]^{\perp}\} + \text{Proj}\{\kappa|L_0^2(P_{ZX})\} \\ &= \varphi(Y,T,Z,X) + E_P[\kappa(Y,T,Z,X)|Z,X] - E_P[\kappa(Y,T,Z,X)], \end{aligned} \quad \text{(A.81)}$$

where the second equality follows from the definition of φ and $\text{Proj}\{\kappa|L_0^2(P_{ZX})\} = E_P[\kappa(Y,T,Z,X)|Z,X] - E_P[\kappa(Y,T,Z,X)]$. Since $L_0^2(P_{ZX}) \subseteq \bar{T}(P)$ and the law of iterated expectations implies $E_P[\kappa(Y,T,Z,X)|X] - E_P[\kappa(Y,T,Z,X)|Z,X] \in L_0^2(P_{ZX})$, we thus obtain from result (A.81) and the definition of Δ that

$$\text{Proj}\{\Delta|\bar{T}(P)\} = \varphi(Y,T,Z,X) + E_P[\kappa(Y,T,Z,X)|X] - E_P[\kappa(Y,T,Z,X)]. \quad \text{(A.82)}$$

Part (i) of the theorem therefore follows from results (A.80), (A.82), and $[N(\mathcal{I})]^{\perp}$ and $L_0^2(P_{ZX})$ being orthogonal subspaces of $L_0^2(P)$.

In order to establish part (ii), we first construct a smaller set of submodels that contains the "least favorable" paths. To this end, note that Lemma A.11(ii) implies we may be view $\mathcal{I}$ as a map from $L^2(P)$ to $T_1(\bar{Q})$. Letting $\bar{R}(\mathcal{I},\bar{Q})$ denote the $\|\cdot\|_{\bar{Q},2}$-closure of $R(\mathcal{I},\bar{Q}) \equiv \{g \in L^2(\bar{Q}) : g = \mathcal{I}(f) \text{ for some } f \in L^2(P)\}$, then set

$$\mathcal{P} \equiv \{Q_{\cdot,g} : Q_{\cdot,g} \text{ is a submodel}, Q_{\eta,g} \ll \bar{Q},\ Q_{0,g} = \bar{Q},\ g \in \bar{R}(\mathcal{I},\bar{Q}) \oplus L_0^2(P_{ZX})\};$$

i.e. $\mathcal{P}$ consists of the submodels passing through $\bar{Q}$ whose score belongs to the subspace $\bar{R}(\mathcal{I},\bar{Q}) \oplus L_0^2(P_{ZX})$. Further note that since $R(\mathcal{I},\bar{Q}) \subseteq T_1(\bar{Q})$ and $T_1(\bar{Q})$ is closed under $\|\cdot\|_{\bar{Q},2}$, it follows that $\bar{R}(\mathcal{I},\bar{Q}) \subseteq T_1(\bar{Q})$. Hence, Lemma A.12 implies that

$$T(\bar{Q},\mathcal{P}) \equiv \{g : Q_{\cdot,g} \in \mathcal{P}\} = \bar{R}(\mathcal{I},\bar{Q}) \oplus L_0^2(P_{ZX}). \quad \text{(A.83)}$$

Next, define $\mathcal{I}'_{\bar{Q}}(g) \equiv E_{\bar{Q}}[g(Y^{\star},T^{\star},Z,X)|Y,T,Z,X]$ for any $g \in T(\bar{Q},\mathcal{P})$, and note that by Jensen's inequality we may view $\mathcal{I}'_{\bar{Q}}$ as a map from $T(\bar{Q},\mathcal{P})$ to $L^2(P)$. Similarly set

$$\mathcal{E}(f) \equiv \mathcal{I}(f) + E_P[f(Y,T,Z,X)|Z,X] - E_P[f(Y,T,Z,X)] \quad \text{(A.84)}$$

for any $f \in L^2(P)$, and note that $\mathcal{E}(f) \in T(\bar{Q},\mathcal{P})$ by result (A.83). Moreover, for any $f \in L^2(P)$ and $g \in T(\bar{Q},\mathcal{P})$ we obtain from result (A.83) implying that $g = g_1 + g_2$ for some $g_1 \in \bar{R}(\mathcal{I},\bar{Q})$ and $g_2 \in L_0^2(P_{ZX})$, and the law of iterated expectations that

$$\langle f, \mathcal{I}'_{\bar{Q}}(g)\rangle_P = \langle f, g_1 + g_2\rangle_{\bar{Q}} = \langle \mathcal{E}(f), g\rangle_{\bar{Q}}, \quad \text{(A.85)}$$

where the final equality follows by definition (A.84) and the same arguments employed in (A.77). In particular, result (A.85) implies $\mathcal{E} : L^2(P) \to T(\bar{Q},\mathcal{P})$ is the adjoint of $\mathcal{I}'_{\bar{Q}} : T(\bar{Q},\mathcal{P}) \to L^2(P)$. Thus, since Proposition 4 in Le Cam and Yang (1988) implies that the set of paths $\mathcal{P}$ fit the framework of Section 3 in van der Vaart (1991b), Theorem

4.1 in van der Vaart (1991b) (applied with $A = \mathcal{I}'_{\bar{Q}}$ and $A^* = \mathcal{E}$), $\mathcal{P}$ being a subset of all submodels, and result (A.78) yields that I^{-1} being finite implies

$$\bar{Q}(\mathcal{E}(f_0) = \text{Proj}\{\ell|T(\bar{Q},\mathcal{P})\}) = 1 \text{ for some } f_0 \in L^2(P). \tag{A.86}$$

To conclude the proof, we aim to show that if condition (A.86) holds, then we must have $\|\ell - \Upsilon(\kappa)\|_{\bar{Q},2} = 0$ for some $\kappa \in L^2(P)$. To this end, suppose (A.86) holds and note that since $\mathcal{E}(f) = \mathcal{E}(f+c)$ for any $c \in \mathbf{R}$, we may assume without loss of generality that $f_0 \in L^2_0(P)$. Moreover, result (A.83) and the definition of $\mathcal{E}$ imply

$$\text{Proj}\{\ell|L^2_0(P_{ZX})\} = \text{Proj}\{f_0|L^2_0(P_{ZX})\} \qquad \mathcal{I}(f_0) = \text{Proj}\{\ell|\bar{R}(\mathcal{I},\bar{Q})\}. \tag{A.87}$$

In particular, result (A.87), the law of iterated expectations, $\bar{Q}$ inducing P due to $\bar{Q} \in \Theta_0$, and $f_0 \in L^2_0(P)$ implying $E_P[f_0(Y,T,Z,X)|Z,X] = \text{Proj}\{f_0|L^2_0(P_{ZX})\}$ yield

$$E_{\bar{Q}}[(\ell(Y^\star,T^\star,X) - E_{\bar{Q}}[f_0(Y,T,Z,X)|Y^\star,T^\star,X])E_P[g(Y,T,Z,X)|X]] = 0 \tag{A.88}$$

for any $g \in L^2_0(P)$. Thus, result (A.88) and Corollary A.1 yield, for any $g \in L^2_0(P)$, that

$$\langle \ell - \Upsilon(f_0), \Upsilon(g)\rangle_{\bar{Q}} = \langle \ell - \Upsilon(f_0), \mathcal{I}(g)\rangle_{\bar{Q}} = \langle \ell - \mathcal{I}(f_0), \mathcal{I}(g)\rangle_{\bar{Q}} = 0 \tag{A.89}$$

where the second equality holds by the law of iterated expectations and Corollary A.1, and the final equality by (A.87). Setting $\kappa = f_0 + \lambda_{Q_0}$, then note that $\Upsilon(c) = c$ for any $c \in \mathbf{R}$, the law of iterated expectations, result (A.89), and $\lambda_{Q_0} = E_{\bar{Q}}[\ell(Y^*,T^*,X)]$ due to λ_{Q_0} being identified imply that for any $g \in L^2_0(P)$ and $c \in \mathbf{R}$ we have

$$0 = \langle \ell - \Upsilon(f_0), \Upsilon(g)\rangle_{\bar{Q}} = \langle \ell - \Upsilon(\kappa), \Upsilon(g)\rangle_{\bar{Q}} = \langle \ell - \Upsilon(\kappa), \Upsilon(g+c)\rangle_{\bar{Q}}. \tag{A.90}$$

It follows from (A.90) that $\Upsilon(\kappa)$ equals the projection of ℓ onto the $\|\cdot\|_{\bar{Q},2}$-closure of $\Upsilon(L^2(P))$, and therefore that $\Upsilon(\kappa)$ is bounded by hypothesis. Also note that Assumption 4.1 holds due to ℓ being bounded and Assumption 4.1(iii) holding by hypothesis. Hence, λ_{Q_0} being identified, Theorem 4.1, and Lemma A.4 applied with $\mathcal{C} = \mathcal{R}$ yield

$$0 = \inf_{f \in L^1(P)} \|\ell - \Upsilon(f)\|_{\bar{Q},1} = \sup_{\|f\|_{\bar{Q},\infty} \leq 1} \langle f, \ell\rangle_{\bar{Q}} \text{ s.t. } \langle f, \Upsilon(g)\rangle_{\bar{Q}} = 0 \text{ for all } g \in L^1(P), \tag{A.91}$$

where the second equality follows from Theorem 5.8.1 in Luenberger (1969). Moreover, since we have shown $\|\Upsilon(\kappa)\|_{\bar{Q},\infty} < \infty$ and ℓ is bounded, we next suppose without loss of generality that $\|\ell - \Upsilon(\kappa)\|_{\bar{Q},\infty} > 0$ (because otherwise the theorem trivially follows). By result (A.90), we then obtain that $(\ell - \Upsilon(\kappa))/\|\ell - \Upsilon(\kappa)\|_{\bar{Q},\infty}$ satisfies the constraints in the maximization problem on the right hand side of (A.91), and hence we can conclude

$$0 \geq \langle \ell - \Upsilon(\kappa), \ell\rangle_{\bar{Q}} = \|\ell - \Upsilon(\kappa)\|^2_{\bar{Q},2} + \langle \ell - \Upsilon(\kappa), \Upsilon(\kappa)\rangle_{\bar{Q}} = \|\ell - \Upsilon(\kappa)\|^2_{\bar{Q},2},$$

where the final equality follows $\Upsilon(\kappa)$ equaling the projection of ℓ onto the $\|\cdot\|_{\bar{Q},2}$ closure of $\Upsilon(L^2(P))$. Thus, (A.86) implies $\bar{Q}(\Upsilon(\kappa) = \ell) = 1$ for some $\kappa \in L^2(P)$ and since (A.86) is a necessary condition for I^{-1} to be finite, part (ii) of the theorem follows. ■

Proof of Corollary 5.1. First note that the conditions of part (i) and Lemma A.16 imply $N(\mathcal{I}) = L^2(P_{ZX})$. Hence, since $L^2(P) = L^2(P_{ZX}) \oplus [L^2(P_{ZX})]^\perp$, it follows that the projection of κ onto $[N(\mathcal{I})]^\perp$ equals $\kappa(Y,T,Z,X) - E_P[\kappa(Y,T,Z,X)|Z,X]$. The first claim of the corollary thus follows from Theorem 5.5(i), Lemma A.17, and $E_P[\kappa(Y,T,Z,X)] = \lambda_{Q_0}$ due to $\Upsilon(\kappa) = \ell$ and Lemma 4.2.

Next, let $N(\Upsilon) \equiv \{s \in L^2(P) : \|\Upsilon(s)\|_{\bar{Q},2} = 0\}$ and for any $s \in N(\Upsilon)$ let $\tilde{s}_t$ equal

$$\tilde{s}_t(Z,X) \equiv E_{\bar{Q}_{Y^\star(t)|X}}[s(Y^\star(t),t,Z,X)]. \tag{A.92}$$

Then note that $\bar{Q} \in \Theta_0$, Jensen's inequality, and the definition of $\bar{Q}^{\rm io}$ imply that

$$\begin{aligned} E_P[1\{T=t\}\tilde{s}_t^2(Z,X)] &= E_{\bar{Q}_{T^\star ZX}}[1\{T^\star(Z)=t\}(E_{\bar{Q}_{Y^\star(t)|X}}[s(Y^\star(t),t,Z,X)])^2] \\ &\le E_{\bar{Q}^{\rm io}}[1\{T^\star(Z)=t\}s^2(Y^\star(t),t,Z,X)] \\ &\lesssim E_{\bar{Q}}[1\{T^\star(Z)=t\}s^2(Y^\star(t),t,Z,X)] \\ &\le E_P[s^2(Y,T,Z,X)], \end{aligned} \tag{A.93}$$

where the second inequality follows from $d\bar{Q}^{\rm io}/d\bar{Q}$ being bounded by Assumption 4.4(ii), and the final follows from $\bar{Q} \in \Theta_0$. In particular, since $s \in L^2(P)$, result (A.93) implies $\tilde{s}_t\delta_t \in L^2(P)$ as well. Moreover, $s \in N(\Upsilon)$ and $d\bar{Q}^{\rm io}/d\bar{Q}$ being bounded yield

$$\begin{aligned} 0 = \|\Upsilon(s)\|^2_{\bar{Q},2} \gtrsim \|\Upsilon(s)\|^2_{\bar{Q}^{\rm io},2} &= \|(\sum_{t\in\mathbf{T}}\Upsilon(\delta_t(s-\tilde{s}_t)) + \Upsilon(\sum_{t\in\mathbf{T}}\delta_t\tilde{s}_t)\|^2_{\bar{Q}^{\rm io}} \\ &= \sum_{t\in\mathbf{T}}\|\Upsilon(\delta_t(s-\tilde{s}_t))\|^2_{\bar{Q}^{\rm io},2} + \|\Upsilon(\sum_{t\in\mathbf{T}}(\delta_t\tilde{s}_t)\|^2_{\bar{Q}^{\rm io},2}, \end{aligned} \tag{A.94}$$

where the final equality follows by noting that $\langle\Upsilon(\delta_{t_1}(s-\tilde{s}_{t_1}),\Upsilon(\delta_{t_2}s)\rangle_{\bar{Q}^{\rm io}} = 0$ and $\langle\Upsilon(\delta_{t_1}(s-\tilde{s}_{t_1}),\Upsilon(\delta_{t_2}\tilde{s}_{t_2})\rangle_{\bar{Q}^{\rm io}} = 0$ for any $t_1 \neq t_2$ by definition of $\tilde{s}_t$ and $\bar{Q}^{\rm io}$. Since result (A.94), $\delta_t\tilde{s}_t \in L^2(P)$, and $\bar{Q} \ll \bar{Q}^{\rm io}$ imply $\|\Upsilon(\delta_t(s-\tilde{s}_t)\|_{\bar{Q},2} = 0$, it follows from the hypotheses of part (ii) of the corollary that $\|\delta_t(s-\tilde{s}_t)\|_{\bar{Q},2} = 0$. Hence, we obtain that $s = \sum_{t\in\mathbf{T}}\delta_t\tilde{s}_t$ and since $s \in N(\Upsilon)$ was arbitrary, we can conclude that $N(\Upsilon) \subseteq L^2(P_{TZX})$. However, by hypothesis $N(\Upsilon)\cap L^2(P_{TZX}) \subseteq L^2(P_{ZX})$, and therefore Lemma A.16 implies that $N(\mathcal{I}) = L^2(P_{ZX})$, which together with the same arguments employed in part (i) yields the second claim of the corollary. ■

Theorem A.1. *Let Assumptions 2.1, 2.2 hold, μ be known, and define the tangent set*

$$T(P) \equiv \{s \in L^2(P) : \eta \mapsto P_{\eta,s} \text{ is induced by some submodel } \eta \mapsto Q_{\eta,g}\}.$$

Then, the tangent space satifies $\bar{T}(P) = [N(\mathcal{I})]^\perp \oplus L^2_0(P_{ZX})$, where $\bar{T}(P)$ denotes the

$\|\cdot\|_{P,2}$*-closure of* $T(P)$ *and* $L^2_0(P_{ZX}) \equiv \{f \in L^2(P_{ZX}) : E_P[f(Z,X)] = 0\}$*,*

Proof. First set $T_1(Q) \equiv \{g \in L^2(Q_{Y^\star T^\star X}) : E_Q[g(Y^\star, T^\star, X)|Z,X] = 0\}$ for any $Q \in \Theta_0$ and define a linear map $\mathcal{I}'_Q : L^2(Q) \to L^2(P)$ to be given by

$$\mathcal{I}'_Q(g) \equiv E_Q[g(Y^\star, T^\star, Z, X)|Y, T, Z, X].$$

Next set $N(\mathcal{I}, Q) \equiv \{s \in L^2(P) : \|\mathcal{I}(s)\|_{Q,2} = 0\}$ noting that $N(\mathcal{I}, \bar{Q}) = N(\mathcal{I})$ for $N(\mathcal{I})$ as defined in the main text – for ease of exposition we omitted the dependence on Q from the main text, but we make such dependence explicit in this proof to enhance the clarity of the arguments that follow. Setting $[N(\mathcal{I}, Q)]^\perp \equiv \{s \in L^2(P) : \langle s, s'\rangle_P = 0 \text{ for all } s' \in N(\mathcal{I}, Q)\}$ and letting $\mathrm{cl}\{A\}$ denote the $\|\cdot\|_{P,2}$ closure of any set $A \subseteq L^2(P)$, then observe that Lemma A.11(ii) and Theorem 6.7.3 in Luenberger (1969) imply

$$\mathcal{I}'_Q(T_1(Q)) \subseteq \mathrm{cl}\{\mathcal{I}'_Q(T_1(Q))\} = [N(\mathcal{I}, Q)]^\perp \subseteq [N(\mathcal{I}, \bar{Q})]^\perp, \tag{A.95}$$

where the final set inclusion follows from $Q \ll \bar{Q}$ for any $Q \in \Theta_0$ implying that $N(\mathcal{I}, \bar{Q}) \subseteq N(\mathcal{I}, Q)$. Further note that, by direct calculation, it is possible to verify that $\bar{Q}(\mathcal{I}(s) = 0) = 1$ for any $s \in L^2(P_{ZX})$ and therefore it follows that $[N(\mathcal{I}, \bar{Q})]^\perp$ and $L^2_0(P_{ZX})$ are orthogonal. Hence, if $\{s_n\}$ is a sequence in $[N(\mathcal{I}, \bar{Q})]^\perp + L^2_0(P_{ZX})$, then writing $s_n = s_{1n} + s_{2n}$ for some $\{s_{1n}\} \subset [N(\mathcal{I}, \bar{Q})]^\perp$ and $\{s_{2n}\} \subset L^2_0(P_{ZX})$ we obtain from the orthogonality of $[N(\mathcal{I}, \bar{Q})]^\perp$ and $L^2_0(P_{ZX})$ that $\|s_n\|^2_{P,2} = \|s_{1n}\|^2_{P,2} + \|s_{2n}\|^2_{P,2}$. Therefore, if $\{s_n\}$ is a Cauchy sequence, then so must be $\{s_{1n}\}$ and $\{s_{2n}\}$ and hence, since $[N(\mathcal{I}, \bar{Q})]^\perp$ and $L^2_0(P_{ZX})$ are complete, we can conclude that $\{s_n\}$ has a limit in $[N(\mathcal{I}, \bar{Q})]^\perp + L^2_0(P_{ZX})$. In particular, it follows that $[N(\mathcal{I}, \bar{Q})]^\perp + L^2_0(P_{ZX})$ is closed, which together with result (A.95) and Lemma A.11(i) implies that

$$\bar{T}(P) \subseteq \mathrm{cl}\{[N(\mathcal{I}, \bar{Q})]^\perp + L^2_0(P_{ZX})\} = [N(\mathcal{I}, \bar{Q})]^\perp + L^2_0(P_{ZX}). \tag{A.96}$$

Conversely, note that Lemma A.11(ii) and Theorem 6.7.3 in Luenberger (1969) yield

$$[N(\mathcal{I}, \bar{Q})]^\perp + L^2_0(P_{ZX}) = \mathrm{cl}\{\mathcal{I}'_{\bar{Q}}(T_1(\bar{Q}))\} + L^2_0(P_{ZX}) \subseteq \bar{T}(P), \tag{A.97}$$

where the final set inclusion follows from Lemma A.11(i). The theorem therefore follows from (A.96), (A.97), and the orthogonality of $[N(\mathcal{I}, \bar{Q})]^\perp$ and $L^2_0(P_{ZX})$. ■

Lemma A.11. *Let Assumptions 2.1 and 2.2(i)(ii) hold,* μ *be known, for any* $Q \in \Theta_0$ *let* $T_1(Q) \equiv \{g \in L^2(Q_{Y^\star T^\star X}) : E_Q[g(Y^\star, T^\star, X)|Z,X] = 0\}$*, and for any* $g \in L^2(Q)$ *set*

$$\mathcal{I}'_Q(g) \equiv E_Q[g(Y^\star, T^\star, Z, X)|Y, T, Z, X].$$

Then: (i) $T(P) = \bigcup_{Q \in \Theta_0} \mathcal{I}'_Q(T_1(Q)) + L^2_0(P_{ZX})$ *with* $L^2_0(P_{ZX}) \equiv \{f \in L^2(P_{ZX}) : E_P[f(Z,X)] = 0\}$*; and (ii)* $\mathcal{I}$ *(as in (56)) is the adjoint of* $\mathcal{I}'_Q : T_1(Q) \to L^2(P)$*.*

Proof. For any $Q \in \Theta_0$ let $T(Q) \equiv \{g \in L^2(Q) : \eta \mapsto Q_{\eta,g}$ is a submodel with $Q_{0,g} = Q\}$ and note Lemma A.12, $Q_{ZX} = P_{ZX}$, and the linearity of $\mathcal{I}'_Q : L^2(Q) \to L^2(P)$ imply

$$\mathcal{I}'_Q(T(Q)) = \mathcal{I}'_Q(T_1(Q)) + \mathcal{I}'_Q(L^2_0(P_{ZX})) = \mathcal{I}'_Q(T_1(Q)) + L^2_0(P_{ZX}). \tag{A.98}$$

To establish part (i), then note that Proposition 4 in Le Cam and Yang (1988) implies that any submodel $\eta \mapsto Q_{\eta,g}$ induces a path $\eta \mapsto P_{\eta,s}$ with score $s = \mathcal{I}'_{Q_{0,g}}(g)$. Therefore part (i) of the lemma follows from (A.98) and the definition of $T(P)$.

In order to establish part (ii), first note that for any $s \in L^2(P)$ and $Q \in \Theta_0$ we can conclude from the definition of $\mathcal{I}$ in (56) and Corollary A.1 that

$$\mathcal{I}(s) = E_Q[s(Y,T,Z,X)|Y^\star,T^\star,X] - E_P[s(Y,T,Z,X)|X]. \tag{A.99}$$

Hence, since $(Y^\star,T^\star) \perp\!\!\!\perp X|Z$ under Q and Q induces P due to $Q \in \Theta_0$, result (A.99) and the law of iterated expectations imply that $\mathcal{I}(s) \in T_1(Q)$ for any $s \in L^2(P)$. Next, let $g \in T_1(Q)$ and $s \in L^2(P)$ be arbitrary, and note that the definition of $\mathcal{I}'_Q$, the law of iterated expectations, and $g \in T_1(Q)$ allow us to conclude that

$$\langle \mathcal{I}'_Q(g), s\rangle_P = E_Q[g(Y^\star,T^\star,X)(s(Y,T,Z,X) - E_P[s(Y,T,Z,X)|Z,X]] = \langle g, \mathcal{I}(s)\rangle_Q,$$

where the final equality follows from the law of iterated expectations, $(Y^\star,T^\star) \perp\!\!\!\perp Z|X$ under Q, and (A.99). Hence, $\mathcal{I} : L^2(P) \to T_1(Q)$ is indeed the adjoint of $\mathcal{I}'_Q : T_1(Q) \to L^2(P)$, which establishes part (ii) of the lemma. ■

Lemma A.12. *Let Assumptions 2.1 and 2.2(i)(ii) hold, μ be known, $Q \in \Theta_0$, and set*

$$T_1(Q) \equiv \{g \in L^2(Q_{Y^\star T^\star X}) : E_Q[g(Y^\star,T^\star,X)|Z,X] = 0\}$$
$$T(Q) \equiv \{g \in L^2(Q) : \eta \mapsto Q_{\eta,g} \text{ is a submodel with } Q_{0,g} = Q\}.$$

Then $T(Q) = T_1(Q) + L^2_0(P_{ZX})$, where $L^2_0(P_{ZX}) \equiv \{f \in L^2(P_{ZX}) : E_P[f(Z,X)] = 0\}$. Moreover, the lemma also holds if when defining $T(Q)$ we require $Q_{\eta,g} \ll Q_{0,g}$ for all η.

Proof. Fix $g \in T(Q)$ and note that Lemma A.13 implies that $g = g_1 + g_2$, where

$$g_1(Y^\star,T^\star,X) = E_Q[g(Y^\star,T^\star,Z,X)|Y^\star,T^\star,X] - E_Q[g(Y^\star,T^\star,Z,X)|X]$$
$$g_2(Z,X) = E_Q[g(Y^\star,T^\star,Z,X)|Z,X].$$

Moreover, since $(Y^\star,T^\star) \perp\!\!\!\perp Z|X$ under Q and $Q_{ZX} = P_{ZX}$ because $Q \in \Theta_0$, it follows from the law of iterated expectations and $E_Q[g(Y^\star,T^\star,Z,X)] = 0$ that $g_1 \in T_1(Q)$ and $g_2 \in L^2_0(P_{ZX})$. It thus follows that $T(Q) \subseteq T_1(Q) + L^2_0(P_{ZX})$. In order to establish the reverse inclusion, we rely on a construction from Example 3.2.1 in Bickel et al. (1993).

Specifically, let $g_1 \in T_1(Q)$ and $g_2 \in L_0^2(P_{ZX})$ be arbitrary and set

$$\frac{dQ_\eta}{d\mu} \equiv \frac{dQ}{d\mu}\frac{\Psi(\eta g_1)\Psi(\eta g_2)}{c(\eta)} \qquad c(\eta) \equiv \int \Psi(\eta g_1)\Psi(\eta g_2)dQ \tag{A.100}$$

where $\Psi : \mathbf{R} \to (0, \infty)$ is any continuously differentiable function with $\Psi(0) = \Psi'(0) = 1$ and Ψ, Ψ', and Ψ'/Ψ bounded. Next define $\pi(\eta, X) \equiv E_Q[\Psi(\eta g_1(Y^\star, T^\star, X))|X]$ and note that (A.100), the law of iterated expectations, and $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q yield

$$E_{Q_\eta}[1\{(Z, X) \in A\}] = E_Q[1\{(Z, X) \in A\}\Psi(\eta g_2(Z, X))\frac{\pi(\eta, X)}{c(\eta)}] \tag{A.101}$$

for any measurable A. In particular, (A.101) implies $Q_{\eta,ZX} \ll Q_{ZX}$ and $dQ_{\eta,ZX}/dQ_{ZX} = \Psi(\eta g_2)\pi(\eta, \cdot)/c(\eta)$. Moreover, for any $h \in L^\infty(Q_{\eta,ZX})$ and $f \in L^1(Q_{\eta,Y^\star T^\star X})$, definition (A.100), the law of iterated expectations, $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q, and (A.101) yield

$$\begin{aligned} &E_{Q_\eta}[h(Z, X)f(Y^\star, T^\star, X)] \\ &\quad = E_Q[h(Z, X)\frac{\Psi(\eta g_2(Z, X))}{c(\eta)}E_Q[f(Y^\star, T^\star, X)\Psi(\eta g_1(Y^\star, T^\star, X))|X]] \\ &\quad = E_{Q_\eta}[h(Z, X)\frac{E_Q[f(Y^\star, T^\star, X)\Psi(\eta g_1(Y^\star, T^\star, X))|X]}{\pi(\eta, X)}]. \end{aligned} \tag{A.102}$$

Hence, since (A.102) holds for any bounded h, it follows for any $f \in L^1(Q_{\eta,Y^\star T^\star X})$ that

$$\begin{aligned} E_{Q_\eta}[f(Y^\star, T^\star, X)|Z, X] &= E_{Q_\eta}[f(Y^\star, T^\star, X)|X] \\ &= \frac{E_Q[f(Y^\star, T^\star, X)\Psi(\eta g_1(Y^\star, T^\star, X))|X]}{\pi(\eta, X)}; \end{aligned}$$

see, e.g., Definition 10.1.1 in Bogachev (2007). Therefore, since $f \in L^1(Q_{\eta,Y^\star T^\star X})$ was arbitrary, we can conclude that $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q_η. Finally, note that if $g_1 = g_2 = 0$, then trivially $g_1 + g_2 \in T(Q)$. On the other hand, if either g_1 or g_2 do not equal zero, then Proposition 2.1.1 in Bickel et al. (1993) implies $\eta \mapsto dQ_\eta/d\mu$ is a regular parametric model in a neighborhood of zero. Moreover, by direct calculation

$$\frac{d}{d\eta}\log(\frac{dQ_\eta}{d\mu})\Big|_{\eta=0} = g_1 + g_2$$

due to $\Psi(0) = \Psi'(0) = 1$ and therefore $g_1 + g_2 \in T(Q)$. Thus, we can conclude $T_1(Q) + L_0^2(P_{ZX}) \subseteq T(Q)$, and the claim of the lemma follows. ■

Lemma A.13. *Let Assumptions 2.1 and 2.2(i)(ii) hold, μ be known, $\eta \mapsto Q_{\eta,g}$ be a submodel with $Q_{0,g} = Q \in \Theta_0$, and let $V \equiv (Y^\star, T^\star, Z, X)$. Then, it follows that:*

$$g(V) = E_Q[g(V)|Y^\star, T^\star, X] + E_Q[g(V)|Z, X] - E_Q[g(V)|X].$$

Proof. For notational simplicity we first define the function $\Delta_Q \in L^2(Q)$ to be given by

$$\Delta_Q(V) \equiv g(V) - E_Q[g(V)|Y^\star, T^\star, X] - E_Q[g(V)|Z, X] + E_Q[g(V)|X]. \tag{A.103}$$

Next note that since $Q \in \Theta_0$ we must have $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q and therefore

$$\begin{aligned} E_Q[h(Z,X) - E_Q[h(Z,X)|X]|Y^\star, T^\star, X] &= 0 \\ E_Q[f(Y^\star, T^\star, X) - E_Q[f(Y^\star, T^\star, X)|X]|Z, X] &= 0 \end{aligned} \tag{A.104}$$

for any bounded functions h and f. In particular, definition (A.103), result (A.104), the law of iterated expectations, and Lemma A.14 imply that

$$E_Q[\Delta_Q(V)(h(Z,X) - E_Q[h(Z,X)|X])(f(Y^\star, T^\star, X) - E_Q[f(Y^\star, T^\star, X)|X])] = 0. \tag{A.105}$$

Moreover, the law of iterated expectations and $(Y^\star, T^\star) \perp\!\!\!\perp Z|X$ under Q also yield that

$$E_Q[\Delta_Q(V)|Y^\star, T^\star, X] = E_Q[\Delta_Q(V)|Z, X] = 0. \tag{A.106}$$

Therefore, results (A.105) and (A.106) imply that for any bounded f and h we have

$$E_Q[\Delta_Q(V)h(Z,X)f(Y^\star, T^\star, X)] = 0. \tag{A.107}$$

We next establish the lemma by showing that result (A.107) implies that $\Delta_Q(V) = 0$. To this end, we let $\mathcal{F}$ denote the σ-field generated by $(\mathcal{F}_{Y^\star} \times \mathcal{F}_{T^\star} \times \mathcal{F}_Z \times \mathcal{F}_X)$ which, as in the rest of the literature, we assume equals the σ-field on which Q is defined (here, $\mathcal{F}_U$ denotes the σ-field on which Q_U is defined). We also define the class of sets

$$\mathcal{A} \equiv \{A \in \mathcal{F} : E_Q[1\{(Y^\star, T^\star, Z, X) \in A\}\Delta_Q(V)] = 0\}$$

and note (A.107) implies $\mathbf{Y}^\star \times \mathbf{T}^\star \times \mathbf{X} \times \mathbf{Z} \in \mathcal{A}$. Also, if $A_1, A_2 \in \mathcal{A}$ and $A_1 \subseteq A_2$ then

$$\begin{aligned} &E_Q[\Delta_Q(V)1\{(Y^\star, T^\star, Z, X) \in A_2 \setminus A_1\}] \\ &\qquad = E_Q[\Delta_Q(V)(1\{(Y^\star, T^\star, Z, X) \in A_2\} - 1\{(Y^\star, T^\star, Z, X) \in A_1\})] = 0, \end{aligned}$$

which implies $A_2 \setminus A_1 \in \mathcal{A}$. Similarly, if $\{A_i\}_{i=1}^\infty \subset \mathcal{A}$ is a sequence of pairwise disjoint sets, then the dominated convergence theorem implies that

$$\begin{aligned} E_Q[\Delta_Q(V)1\{(Y^\star, T^\star, Z, X) \in \bigcup_{i=1}^\infty A_i\}] &= \lim_{n\to\infty} E_Q[\Delta_Q(V)1\{(Y^\star, T^\star, Z, X) \in \bigcup_{i=1}^n A_i\}] \\ &= \lim_{n\to\infty} \sum_{i=1}^n E_Q[\Delta_Q(V)1\{(Y^\star, T^\star, Z, X) \in A_i\}] = 0, \end{aligned} \tag{A.108}$$

where the second and third equalities follow from $\{A_i\}_{i=1}^{\infty}$ being disjoint and $A_i \in \mathcal{A}$. Result (A.108) implies $\bigcup_{i=1}^{\infty} \in \mathcal{A}$ and therefore that $\mathcal{A}$ is a λ-system. On the other hand, if $A_{Y^\star} \in \mathcal{F}_{Y^\star}$, $A_{T^\star} \in \mathcal{F}_{T^\star}$, $A_Z \in \mathcal{F}_Z$, and $A_X \in \mathcal{F}_X$, then setting $h(Z,X) = 1\{(Z,X) \in A_Z \times A_X\}$ and $f(Y^\star, T^\star, X) = 1\{(Y^\star, T^\star) \in A_{Y^\star} \times A_{T^\star}\}$ in (A.107) yields

$$\begin{aligned} &E_Q[\Delta_Q(V)1\{(Y^\star, T^\star, Z, X) \in A_{Y^\star} \times A_{T^\star} \times A_Z \times A_X\}] \\ &\qquad = E_Q[\Delta_Q(V)1\{(Y^\star, T^\star) \in A_{Y^\star} \times A_{T^\star}\}1\{(Z,X) \in A_Z \times A_X\}] = 0. \end{aligned}$$

In particular, we obtain that $(\mathcal{F}_{Y^\star} \times \mathcal{F}_{T^\star} \times \mathcal{F}_Z \times \mathcal{F}_X) \subseteq \mathcal{A}$. Hence, since $(\mathcal{F}_{Y^\star} \times \mathcal{F}_{T^\star} \times \mathcal{F}_Z \times \mathcal{F}_X)$ is a π-system and $\mathcal{F}$ is generated by $(\mathcal{F}_{Y^\star} \times \mathcal{F}_{T^\star} \times \mathcal{F}_Z \times \mathcal{F}_X)$, the $\pi - \lambda$ theorem (see, e.g., Theorem 2.38 in Pollard (2002)) yields that $\mathcal{A} = \mathcal{F}$. Thus, we obtain

$$E_Q[|\Delta_Q(V)|] = E_Q[\Delta_Q(V)1\{\Delta_Q(V) \geq 0\}] - E_Q[\Delta_Q(V)1\{\Delta_Q(V) < 0\}] = 0,$$

which establishes the claim of the lemma. ■

Lemma A.14. *Let Assumptions 2.1 and 2.2(i)(ii) hold, μ be known, and $\eta \mapsto Q_{\eta,g}$ be a submodel with $Q_{0,g} = Q \in \Theta_0$. Then, for any $h \in L^\infty(\mu_{ZX})$ and $f \in L^\infty(\mu_{Y^\star T^\star X})$:*

$$E_Q[g(Y^\star, T^\star, Z, X)(h(Z,X) - E_Q[h(Z,X)|X])(f(Y^\star, T^\star, X) - E_Q[f(Y^\star, T^\star, X)|X])] = 0.$$

Proof. In what follows we write E_η in place of $E_{Q_{\eta,g}}$ and E in place of E_Q. Next note that f and h being bounded and Lemma F.1 in Chen and Santos (2018) imply

$$\begin{aligned} &\lim_{\eta \downarrow 0} \frac{1}{\eta}\{E_\eta[h(Z,X)f(Y^\star, T^\star, X)] - E[h(Z,X)f(Y^\star, T^\star, X)]\} \\ &\qquad = E[h(Z,X)f(Y^\star, T^\star, X)g(Y^\star, T^\star, Z, X)]. \quad \text{(A.109)} \end{aligned}$$

Moreover, a second application of Lemma F.1 in Chen and Santos (2018) also yields

$$\begin{aligned} &\lim_{\eta \downarrow 0} \frac{1}{\eta}\{E_\eta[h(Z,X)E_\eta[f(Y^\star, T^\star, X)|X]] - E[h(Z,X)E_\eta[f(Y^\star, T^\star, X)|X]]\} \\ &\quad = \lim_{\eta \downarrow 0} E[h(Z,X)E_\eta[f(Y^\star, T^\star, X)|X]g(Y^\star, T^\star, Z, X)] \\ &\quad = E[h(Z,X)E[f(Y^\star, T^\star, X)|X]g(Y^\star, T^\star, Z, X)] \quad \text{(A.110)} \end{aligned}$$

where the final result follows from the Cauchy-Schwarz inequality and Lemma A.15. Next note that the law of iterated expectations and similar arguments also yield

$$\begin{aligned} &\lim_{\eta \downarrow 0} \frac{1}{\eta}\{E_\eta[E[h(Z,X)|X]f(Y^\star, T^\star, X)] - E[E[h(Z,X)|X]E_\eta[f(Y^\star, T^\star, X)|X]]\} \\ &\quad = \lim_{\eta \downarrow 0} E[E[h(Z,X)|X]E_\eta[f(Y^\star, T^\star, X)|X]g(Y^\star, T^\star, Z, X)] \\ &\quad = E[E[h(Z,X)|X]E[f(Y^\star, T^\star, X)|X]g(Y^\star, T^\star, Z, X)], \quad \text{(A.111)} \end{aligned}$$

while a final application of Lemma F.1 in Chen and Santos (2018) further implies that

$$\lim_{\eta\downarrow 0}\frac{1}{\eta}\{E_\eta[E[h(Z,X)|X]f(Y^\star,T^\star,X)]-E[E[h(Z,X)|X]f(Y^\star,T^\star,X)]\}$$
$$=E[E[h(Z,X)|X]f(Y^\star,T^\star,X)g(Y^\star,T^\star,Z,X)]. \quad (A.112)$$

To conclude, note that since $(Y^\star,T^\star)\perp\!\!\!\perp Z|X$ under $Q_{\eta,g}$ for all $\eta\geq 0$ we must have

$$E_\eta[h(Z,X)E_\eta[f(Y^\star,T^\star,X)|X]]=E_\eta[h(Z,X)f(Y^\star,T^\star,X)] \quad (A.113)$$

for any $\eta\geq 0$. In particular, since $Q_{0,g}=Q$, result (A.113) allows us to conclude that

$$\lim_{\eta\downarrow 0}\frac{1}{\eta}\{E_\eta[h(Z,X)E_\eta[f(Y^\star,T^\star,X)|X]]-E[h(Z,X)E[f(Y^\star,T^\star,X)|X]]\}$$
$$=\lim_{\eta\downarrow 0}\frac{1}{\eta}\{E_\eta[h(Z,X)f(Y^\star,T^\star,X)]-E[h(Z,X)f(Y^\star,T^\star,X)]\}. \quad (A.114)$$

The claim of the lemma therefore follows from combining the equality in (A.114) with results (A.109), (A.110), (A.111) and (A.112). ∎

Lemma A.15. *If μ is known, $\eta\mapsto Q_{\eta,g}$ is a path, and $f\in L^\infty(\mu)$, then it follows that*

$$\lim_{\eta\downarrow 0}E_{Q_{0,g}}[(E_{Q_{\eta,g}}[f(Y^\star,T^\star,Z,X)|X]-E_{Q_{0,g}}[f(Y^\star,T^\star,Z,X)|X])^2]=0.$$

Proof. Set $V\equiv(Y^\star,T^\star,Z,X)$ for notational simplicity and define the sets $A_\eta^+\equiv\{X: E_{Q_{\eta,g}}[f(V)|X]\geq E_{Q_{0,g}}[f(V)|X]\}$ and $A_\eta^-\equiv\{X: E_{Q_{\eta,g}}[f(V)|X]< E_{Q_{0,g}}[f(V)|X]\}$. Then note that since f is bounded by hypothesis we can conclude that

$$E_{Q_{0,g}}[(E_{Q_{\eta,g}}[f(V)|X]-E_{Q_{0,g}}[f(V)|X])^2]\lesssim E_{Q_{0,g}}[|E_{Q_{\eta,g}}[f(V)|X]-E_{Q_{0,g}}[f(V)|X]|]$$
$$=E_{Q_{0,g}}[(1\{X\in A_\eta^+\}-1\{X\in A_\eta^-\})(E_{Q_{\eta,g}}[f(V)|X]-f(V))], \quad (A.115)$$

where the equality follows from the definitions of A_η^+ and A_η^- and the law of iterated expectations. However, by Lemma F.1 in Chen and Santos (2018) we have that

$$\lim_{\eta\downarrow 0}E_{Q_{0,g}}[(1\{X\in A_\eta^+\}-1\{X\in A_\eta^-\})(E_{Q_{\eta,g}}[f(V)|X]-f(V))]$$
$$=\lim_{\eta\downarrow 0}E_{Q_{\eta,g}}[(1\{X\in A_\eta^+\}-1\{X\in A_\eta^-\})(E_{Q_{\eta,g}}[f(V)|X]-f(V))]=0, \quad (A.116)$$

where the final equality follows from the law of iterated expectations. Results (A.115) and (A.116) together establish the claim of the lemma. ∎

Lemma A.16. *Let Assumptions 2.1, 2.2 hold, $N(\Upsilon)\equiv\{s\in L^2(P):\|\Upsilon(s)\|_{\bar{Q},2}=0\}$, and $[L^2(P_{ZX})]^\perp\equiv\{s\in L^2(P):\langle s,\tilde{s}\rangle_P=0$ for all $\tilde{s}\in L^2(P_{ZX})\}$. Then it follows that*

$$N(\mathcal{I})=(N(\Upsilon)\cap[L^2(P_{ZX})]^\perp)\oplus L^2(P_{ZX}).$$

Proof. Let $s_1 \in N(\Upsilon) \cap [L^2(P_{ZX})]^\perp$ and $s_2 \in L^2(P_{ZX})$ be arbitrary and note that

$$\mathcal{I}(s_1 + s_2) = \mathcal{I}(s_1) = -E_P[s_1(Y, T, Z, X)|X]$$
$$= -E_{\bar{Q}}[E_{\bar{Q}}[s_1(Y, T, Z, X)|Y^\star, T^\star, X]|X] = 0$$

where in the first equality we used that $\mathcal{I}(s_2) = 0$ for any $s_2 \in L^2(P_{ZX})$, the second equality follows from $s_1 \in N(\Upsilon)$, the third inequality follows from $\bar{Q} \in \Theta_0$ and the law of iterated expectations, and the final equality from Corollary A.1 and $s_1 \in N(\Upsilon)$. Thus, we must have $(N(\Upsilon) \cap [L^2(P_{ZX})]^\perp) \oplus L^2(P_{ZX}) \subseteq N(\mathcal{I})$. For the reverse inclusion let $s \in N(\mathcal{I})$ be arbitrary and set $s_1 \equiv s - s_2$ with s_2 given by

$$s_2(Z, X) \equiv E_P[s(Y, T, Z, X)|Z, X].$$

Note that $s_2 \in L^2(P_{ZX})$, $s_1 \in [L^2(P_{ZX})]^\perp$, and by the law of iterated expectations

$$\Upsilon(s_2) = \sum_{t \in \mathbf{T}} E_{P_{Z|X}}[s_2(Z, X)1\{T^\star(Z) = t\}] = E_P[s(Y, T, Z, X)|X].$$

Thus, since $s \in N(\mathcal{I})$ we obtain that $\Upsilon(s_1) = \Upsilon(s) - \Upsilon(s_2) = \mathcal{I}(s) = 0$, which implies $s_1 \in N(\Upsilon) \cap [L^2(P_{ZX})]^\perp$. Hence, we conclude $N(\mathcal{I}) \subseteq (N(\Upsilon) \cap [L^2(P_{ZX})]^\perp) \oplus L^2(P_{ZX})$ and the claim of the lemma follows. ■

Lemma A.17. *Suppose that the conditions of Theorem 5.1 (resp. Theorem 5.3) hold with* $\max_t \delta_t^\beta \vee \delta_t^\gamma = o(1)$ *(resp.* $\delta^\beta \vee \delta^\gamma = o(1)$*), the conditions of Theorem 5.5(i) hold with a* κ *satisfying Assumption 5.1(ii) (resp. Assumption 5.3(ii)), and define*

$$\tilde{\psi}(Y, T, Z, X) \equiv \kappa(Y, T, Z, X) - E_P[\kappa(Y, T, Z, X)|Z, X] + E_P[\kappa(Y, T, Z, X)|X] - \lambda_{Q_0}.$$

Then, it follows that the estimator $\hat{\lambda}$ *of Section 5.1 (resp. Section 5.2) satisfies*

$$\sqrt{n}\{\hat{\lambda} - \lambda_{Q_0}\} \xrightarrow{d} N(0, \mathrm{Var}_P\{\tilde{\psi}(Y, T, Z, X)\}).$$

Proof. We only establish the claim concerning Section 5.1, since the claim concerning Section 5.2 follows by identical arguments. First note that by Assumption 5.1(iii), $\kappa \in L^\infty(P_{TZX})$ and hence $\kappa = \nu/\pi$, and the law of iterated expectations yield

$$\tilde{\psi}(Y, T, Z, X) = \sum_{t \in \mathbf{T}} \kappa(t, Z, X)(1\{T = t\} - P(T = t|Z, X))$$
$$+ \sum_{t \in \mathbf{T}} E_{\mu_{Z|X}}[\nu(t, Z, X)P(T = t|Z, X)] - \lambda_{Q_0}. \quad \text{(A.117)}$$

For ψ as in (35), we then obtain from (A.117), $\max_t \|b'\beta_t\|_\infty = O(1)$ by Assumption

5.2(i), $\|\kappa\|_\infty \vee \|\nu\|_\infty < \infty$ by Assumptions 5.1(ii)(iii), and Jensen's inequality that

$$E_P[(\tilde{\psi}(Y,T,Z,X) - \psi(Y,T,Z,X))^2] \lesssim \sum_{t\in\mathbf{T}} E_P[(b(Z,X)'\beta_t - P(T=t|Z,X))^2 + \sum_{t\in\mathbf{T}}(b(Z,X)'\gamma_t - \kappa(t,Z,X))^2] = o(1), \quad (A.118)$$

where the final equality follows from $\max_t \delta_t^\beta \vee \delta_t^\gamma = o(1)$ by hypothesis. Setting $\tilde{\sigma}^2 \equiv \mathrm{Var}_P\{\tilde{\psi}(Y,T,Z,X)\}$ and $\sigma^2 \equiv \mathrm{Var}_P\{\psi(Y,T,Z,X)\}$, it then follows from (A.118) that $\sigma^2 = \tilde{\sigma}^2 + o(1)$ and hence that $\sigma^2 = O(1)$ due to $\|\kappa\|_\infty < \infty$. Thus, $E[\tilde{\psi}(Y,T,Z,X)] = 0$ due to $\Upsilon(\kappa) = \ell$ and Lemma 4.2, Lemma A.7, $\sigma^2 = O(1)$, and (A.118) imply

$$E[(\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(Y_i,T_i,Z_i,X_i) - \tilde{\psi}(Y_i,T_i,Z_i,X_i))^2] = E[(\psi(Y,T,Z,X) - \tilde{\psi}(Y,T,Z,X))^2] + o(1) = o(1). \quad (A.119)$$

Thus, Theorem 5.1, $\sigma^2 = O(1)$, result (A.119), and Markov's inequality yield that

$$\sqrt{n}\{\hat{\lambda} - \lambda_{Q_0}\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{\psi}(Y_i,T_i,Z_i,X_i) + o_P(1),$$

which together with the central limit theorem establishes the claim of the lemma. ■

A.4 Additional Details for Section 3

The MTO experiment offered incentives to households that were socially disadvantaged, encouraging them to relocate from economically deprived areas to more affluent neighborhoods. The experiment was conducted over a period of four years, from June 1994 to July 1998, as documented by Orr et al. (2003). Eligible households were those that belonged to the low-income group and had children under the age of 18, residing in the most impoverished housing projects of five major US cities, namely Baltimore, Boston, Chicago, Los Angeles, and New York. The majority of these households, i.e. 75%, relied on welfare, while only a third had completed high school. African Americans comprised the majority of the sample, constituting 62%, followed by Hispanics at 30%. Female-headed households made up 92% of the participants.

Our dataset comprises 3039 families residing in high-poverty neighborhoods at the onset of the intervention. These families were randomly assigned to either the control group, consisting of 1310 families, or the experimental group, comprising of 1729 families. The experimental group received a rent-subsidizing voucher that incentivized families to relocate from the high-poverty public housing they lived in to low-poverty communities, namely, neighborhoods where less than 10% of households were living below the poverty line according to the 1990 US Census. Families in the control group did not receive any voucher. The Department of Housing and Urban Development (HUD) set the subsidy amount and unit eligibility based on the Applicable Payment Standard (APS). Landlords could not discriminate against a voucher recipient, and leases were automatically renewed. Families that decided to use the experimental voucher were required to live in the low-poverty neighborhood for a year but could move afterward. HUD paid rent directly to the landlord and required that households pay 30% of their monthly adjusted gross income to offset the cost of rent and utilities. A total of 818 out of the 1,729 experimental families agreed to use the voucher to relocate to low-poverty neighborhoods. Experimental families that did not use the voucher and control families were also allowed to move to low-poverty neighborhoods.

We investigate labor market outcomes surveyed at the MTO interim evaluation in 2002 (Orr et al., 2003). For control variables X we follow the literature in employing:

1. Experimental site indicators.
2. Indicator for whether a household member had a disability.
3. Indicator for no teens (ages 13-17) in the household at the onset of the intervention.
4. Indicator for whether the family had previously applied for a Section 8 voucher.
5. Indicator for whether the family had moved more than three times in the five years prior to the onset of the intervention.
6. Indicator for whether respondent reported not having friends in the neighborhood.
7. Indicator for whether respondent was very likely to tell a neighbor if he/she saw a neighbor's child getting into trouble.
8. Indicator for whether a family member had been assaulted during the six months preceding the baseline survey.
9. Assessment of whether the streets near home were very unsafe at night.
10. Baseline respondent's primary or secondary reason for wanting to move was to get away from gangs or drugs.

All our estimates rely on the person-level weights, denoted by $\{\omega_i\}_{i=1}^n$, for the adult survey of the interim analyses, as described in the MTO Interim Impacts Evaluation manual, 2003, Appendix B. We drop any observations with a missing value for any of the outcomes of interest, treatment status, or baseline characteristics.

A.4.1 Additional Details for Section 3.1

All functionals about types fall within the framework of Section 5.1. Moreover, since the instrument $Z \in \{0,1\}$ and treatment $T = (D,M) \in \{0,1\} \times \{0,1\}$ are discrete, the estimation algorithm may be implemented in the manner discussed in Example 2.4. To map this problem into the notation of Example 2.4 simply interpret $T^* \equiv (D^*(0), D^*(1), M^*(0), M^*(1))$ as a vector in $\mathbf{R}^4$ and require that $\mu(T^* \in \mathbf{R}^*) = 1$ where

$$\mathbf{R}^* \equiv \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

Due to the sample size, we do not employ sample splitting – a modification to the algorithm of Section 5.1 that is justified under appropriate sparsity assumptions. We additionally incorporate the weights $\{\omega_i\}_{i=1}^n$ in estimation by proceeding as follows:

Step A.1. Set $b(Z,X) \in \mathbf{R}^p$ to consist of the functions generated by interacting Z and $(1-Z)$ with every coordinate of the baseline covariates X. ■

Step A.2. For each treatment value $t \in \mathbf{T}$ we estimate the following LASSO regression

$$\hat{\beta}_t \in \arg\min_{\beta \in \mathbf{R}^p} \sum_{i=1}^n \omega_i (1\{T_i = t\} - b(Z_i, X_i)'\beta)^2 + \alpha\|\beta\|_1,$$

where the penalty α is chosen through leave-one-out cross validation. We also compute

$$\hat{\gamma}_t \in \arg\min_{\gamma \in \mathbf{R}^p} \sum_{i=1}^n \omega_i \{\frac{1}{2}(b(Z_i, X_i)'\gamma)^2 - E_{\mu_{Z|X}}[\nu(t, Z, X_i) b(Z, X_i)'\gamma]\} + \alpha\|\gamma\|_1,$$

where α is again chosen by leave-one-out cross validation and, since $p < n$, we follow Remark A.0.1 below to compute $\hat{\gamma}_t$ through a LASSO regression. ■

Step A.3. We estimate $\lambda_{Q_0} = E_{Q_0}[\ell(T^*, X)]$ by employing the plug-in estimator

$$\hat{\lambda} \equiv \sum_{i=1}^n \omega_i \{\sum_{t \in \mathbf{T}} b(Z_i, X_i)'\hat{\gamma}_t (1\{T_i = t\} - b(Z_i, X_i)'\hat{\beta}_t) + E_{\mu_{Z|X}}[\nu(t, Z, X_i) b(Z, X_i)'\hat{\beta}_t]\}.$$

Recall $\hat{\lambda}$ is simply a sample analogue to the moment condition in (32). ■

By applying the algorithm with $\ell(T^*, X) = 1\{T^* = t^*\}$ for each possible type t^* we obtain estimates of the type probabilities. The standard error $\hat{\sigma}$ for $\hat{\lambda}$ then satisfies

$$\hat{\sigma}^2 = \sum_{i=1}^n \omega_i^2 (\sum_{t \in \mathbf{T}} b(Z_i, X_i)'\hat{\gamma}_t (1\{T_i = t\} - b(Z_i, X_i)'\hat{\beta}_t) + E_{\mu_{Z|X}}[\nu(t, Z, X_i) b(Z, X_i)'\hat{\beta}_t] - \hat{\lambda})^2$$

To estimate the expectation of a coordinate $X^{(j)}$ of the baseline covariates X conditional on type t^* (as in Table 1), we simply rely on the equality

$$E_{Q_0}[X^{(j)}|T^* = t^*] = \frac{E_{Q_0}[X^{(j)}1\{T^* = t^*\}]}{E_{Q_0}[1\{T^* = t^*\}]}$$

and construct a plug-in estimator by applying the preceding algorithm with $\ell(T^*, X) = X^{(j)}1\{T^* = t^*\}$ and $\ell(T^*, X) = 1\{T^* = t^*\}$. Standard errors for these estimators are obtained via the Delta method.

Remark A.0.1. Whenever the dimension p of $b(Z, X)$ is smaller than n, the estimator $\hat{\gamma}_t$ can be computed through a LASSO regression. Specifically, by setting

$$\tilde{Y}_i \equiv b(Z_i, X_i)'(\sum_{j=1}^{n} \omega_j b(Z_j, X_j) b(Z_j, X_j)')^{-1} \sum_{j=1}^{n} \omega_j E_{\mu_{Z|X}}[\nu(t, Z, X_j) b(Z, X_j)],$$

it is possible to show that $\hat{\gamma}_t$ also equals the solution to the LASSO regression

$$\min_{\gamma \in \mathbf{R}^p} \sum_{i=1}^{n} \omega_i (\tilde{Y}_i - b(Z_i, X_i)'\gamma)^2 + \alpha \|\gamma\|_1.$$

This observation is helpful for computational purposes, because it allows us to rely on readily available LASSO routines to compute the estimator $\hat{\gamma}_t$. ■

A.4.1 Additional Details for Section 3.2

All parameters examined in Section 3.2 depend on expectations with the structure

$$E_{Q_0}[\rho(Y^*(t))\ell(T^*, X)] \tag{A.120}$$

and on type probabilities. Moreover, recall that a necessary and sufficient condition for identification of (A.120) is that there exist a function κ satisfying

$$E_{Q_0}[\sum_{z \in \mathbf{Z}} 1\{T^*(z) = t\}\kappa(z, X) P(Z = z|X)|V^*(t), X] = E_{Q_0}[\ell(T^*, X)|V^*(t), X], \tag{A.121}$$

where $V^*(t) = T^*$ if $t \in \{(0,1), (1,0)\}$, $V^*((0,0)) = T^*1\{T^* \notin \{CN, CC\}\}$, and $V^*((1,1)) = T^*1\{T^* \notin \{CA, CC\}\}$. For certain choices of functions $\ell(T^*, X)$, the identifying equation in (A.121) has the structure assumed in Section 5.2. In particular,

$$E_{Q_0}[\rho(Y^*(t))1\{T^* \in A\}] \text{ with } \begin{cases} t = (0,0) & \text{and } A = \{NN\} \text{ or } \{CN, CC\} \\ t = (0,1) & \text{and } A = \{NA\} \text{ or } \{CA\} \\ t = (1,0) & \text{and } A = \{CN\} \text{ or } \{AN\} \\ t = (1,1) & \text{and } A = \{CC, CA\} \text{ or } \{AA\} \end{cases} \tag{A.122}$$

are identified by $E_P[\rho(Y)1\{T = t\}\kappa(Z, X)]$ with $\kappa(Z, X) = \nu(Z, X)/\pi(Z, X)$ for some known function ν that may be found by proceeding as in our discussion of Example 2.4. For expectations that fall within the scope of (A.122) we therefore employ the algorithm in Section 5.2 but without sample splitting and with the inclusion of person-level weights:

Step B.1. Set $b(Z, X) \in \mathbf{R}^p$ to consist of the functions generated by interacting Z and $(1 - Z)$ with every coordinate of the baseline covariates X. ■

Step B.2. Compute the following two estimators through LASSO regressions

$$\hat{\beta} \in \arg\min_{\beta \in \mathbf{R}^p} \sum_{i=1}^{n} \omega_i(\rho(Y_i)1\{T_i = t\} - b(Z_i, X_i)'\beta)^2 + \alpha\|\beta\|_1$$

$$\hat{\gamma} \in \arg\min_{\gamma \in \mathbf{R}^p} \sum_{i=1}^{n} \omega_i\{\frac{1}{2}(b(Z_i, X_i)'\gamma)^2 - E_{\mu_{Z|X}}[\nu(Z, X_i)b(Z, X_i)'\gamma]\} + \alpha\|\gamma\|_1,$$

where the penalty α is chosen through leave-one-out cross validation and in computing $\hat{\gamma}$ we rely on Remark A.0.1. ■

Step B.3. We estimate $\lambda_{Q_0} = E_{Q_0}[\rho(Y^*(t))\ell(T^*, X)]$ employing the plug-in estimator

$$\hat{\lambda} \equiv \sum_{i=1}^{n} \omega_i\{b(Z_i, X_i)'\hat{\gamma}(\rho(Y_i)1\{T_i = t\} - b(Z_i, X_i)'\hat{\beta}) + E_{\mu_{Z|X}}[\nu(Z, X_i)b(Z, X_i)'\hat{\beta}]\}.$$

Certain parameters that are relevant for our analysis, however, fall outside the scope of Section 5.2 and the preceding algorithm. These parameters have the structure

$$E_{Q_0}[\rho(Y^*(t))1\{T^* = t^*\}] \text{ with } \begin{cases} t = (0,0) & \text{and } t^* = CN \\ t = (1,1) & \text{and } t^* = CA \end{cases}$$

and remain identified by the expectation $E_P[\rho(Y)1\{T = t\}\kappa(Z, X)]$, but the relevant κ no longer satisfies $\kappa(Z, X) = \nu(Z, X)/\pi(Z, X)$ for some known ν. For instance, for identifying $E_{Q_0}[\rho(Y^*(0, 0))1\{T^* = CN\}]$ equation (A.121) implies the relevant κ solves

$$E_{P_{Z|X}}[1\{t^*(Z) = (0, 0)\}\kappa(Z, X)] = 0 \text{ for all } t^* \notin \{CN, CC\}$$

and

$$E_{Q_0}[1\{T^*(Z) = (0, 0)\}\kappa(Z, X)|T^* \in \{CN, CC\}, X] = Q_0(T^* = CN|T^* \in \{CN, CC\}, X).$$

Based on these observations, it is then possible to obtain an orthogonal score for estimating $E_{Q_0}[\rho(Y^*(0,0))1\{T^* = CN\}]$. Specifically, defining the nuisance parameters

$$\begin{aligned} m(Z,X) &\equiv E[\rho(Y)1\{T = (0,0)\}|Z,X] \\ p_{00}(Z,X) &\equiv P(T = (0,0)|Z,X) \\ p_{10}(Z,X) &\equiv P(T = (1,0)|Z,X) \\ \kappa_c(Z,X) &\equiv (\frac{1\{Z=0\}}{P(Z=0|X)} - \frac{1\{Z=1\}}{P(Z=1|X)}) \\ u(X) &\equiv Q_0(T^* \in \{CC, CN\}|X) \\ c(X) &\equiv Q_0(T^* = CN|T^* \in \{CN, CC\}, X), \end{aligned}$$

it is possible to show that the orthogonal score for $E_{Q_0}[\rho(Y^*(0,0))1\{T^* = CN\}]$ equals

$$\begin{aligned} &E[\rho(Y^*(0,0))1\{T^* = CN\}] \\ &\qquad = E[(Y1\{T=(0,0)\} - m(Z,X))c(X)\kappa_c(Z,X)] \\ &\qquad + E[(m(0,X) - m(1,X))c(X)] \\ &\qquad - E[\frac{(m(0,X) - m(1,X))}{u(X)}(1\{T=(1,0)\} - p_{10}(Z,X))\kappa_c(Z,X)] \\ &\qquad - E[\frac{(m(0,X) - m(1,X))}{u(X)}(p_{10}(0,X) - p_{10}(1,X)] \\ &\qquad - E[\frac{(m(0,X) - m(1,X))c(X)}{u(X)}(1\{T=(0,0)\} - p_{00}(Z,X))\kappa_c(Z,X)] \\ &\qquad - E[\frac{(m(0,X) - m(1,X))c(X)}{u(X)}(p_{00}(0,X) - p_{00}(1,X))]. \end{aligned}$$

Given this orthogonal score, we then obtain an estimator by proceding as follows:

STEP C.1. Set $b(Z,X) \in \mathbf{R}^p$ to consist of the functions generated by interacting Z and $(1-Z)$ with every coordinate of the baseline covariates X and compute

$$\begin{aligned} \hat{\beta}_m &\in \arg\min_{\beta \in \mathbf{R}^p} \sum_{i=1}^n \omega_i(\rho(Y_i)1\{T_i = (0,0)\} - b(Z_i,X_i)'\beta)^2 + \alpha\|\beta\|_1 \\ \hat{\beta}_{00} &\in \arg\min_{\beta \in \mathbf{R}^p} \sum_{i=1}^n \omega_i(1\{T_i = (0,0)\} - b(Z_i,X_i)'\beta)^2 + \alpha\|\beta\|_1 \\ \hat{\beta}_{10} &\in \arg\min_{\beta \in \mathbf{R}^p} \sum_{i=1}^n \omega_i(1\{T_i = (1,0)\} - b(Z_i,X_i)'\beta)^2 + \alpha\|\beta\|_1 \end{aligned}$$

through LASSO regression and the penalty α selected by leave-one-out cross validation. Similarly, by relying on Remark A.0.1 we also compute the estimator

$$\hat{\gamma}_\kappa \in \arg\min_{\gamma \in \mathbf{R}^p} \sum_{i=1}^n \omega_i\{\frac{1}{2}(b(Z_i,X_i)'\gamma)^2 - (b(0,X_i) - b(1,X_i))'\gamma\} + \alpha\|\gamma\|_1,$$

through a LASSO regression and select α through leave-one-out cross validation. ■

STEP C.2. Set $f(X) \in \mathbf{R}^q$ to equal X and compute the following penalized estimators

$$\hat{\pi}_{cc} \in \arg\min_{\pi \in \mathbf{R}^q} \sum_{i=1}^{n} \omega_i (1\{T_i \in \{(0,0),(1,0)\}\} b(Z_i, X_i)'\hat{\gamma}_\kappa - f(X_i)'\pi)^2 + \alpha \|\pi\|_1$$

$$\hat{\pi}_{cn} \in \arg\min_{\pi \in \mathbf{R}^q} \sum_{i=1}^{n} \omega_i (1\{T_i = (1,0)\} b(Z_i, X_i)'\hat{\gamma}_\kappa + f(X_i)'\pi)^2 + \alpha \|\pi\|_1,$$

where the penalties α are selected by leave-one-out cross validation. ■

STEP C.3. Using the estimators from Steps 1 and 2 define the following estimators

$$\hat{m}(Z_i, X_i) \equiv b(Z_i, X_i)'\hat{\beta}_m \qquad \hat{p}_{00}(Z_i, X_i) \equiv b(Z_i, X_i)'\hat{\beta}_{00}$$

$$\hat{p}_{10}(Z_i, X_i) \equiv b(Z_i, X_i)'\hat{\beta}_{10} \qquad \hat{\kappa}_c(Z_i, X_i) \equiv b(Z_i, X_i)'\hat{\gamma}_\kappa$$

$$\hat{u}(X_i) \equiv f(X_i)'(\hat{\pi}_{cc} + \hat{\pi}_{cn}) \qquad \hat{c}(X_i) \equiv \frac{f(X_i)'\hat{\pi}_{cn}}{\hat{u}(X_i)}.$$

Employing these estimators we put them all together into the orthogonal score by setting

$$\begin{aligned}
\hat{\psi}(Y_i, T_i, Z_i, X_i) = {} & (\rho(Y_i) 1\{T_i = (0,0)\} - \hat{m}(Z_i, X_i))\hat{c}(X_i)\hat{\kappa}_c(Z_i, X_i) \\
& + (\hat{m}(0, X_i) - \hat{m}(1, X_i))\hat{c}(X_i) \\
& - \frac{(\hat{m}(0, X_i) - \hat{m}(1, X_i))}{\hat{u}(X_i)}(1\{T_i = (1,0)\} - \hat{p}_{10}(Z_i, X_i))\hat{\kappa}_c(Z_i, X_i) \\
& - \frac{(\hat{m}(0, X_i) - \hat{m}(1, X_i))}{\hat{u}(X_i)}(\hat{p}_{10}(0, X_i) - \hat{p}_{10}(1, X_i)) \\
& - \frac{(\hat{m}(0, X_i) - \hat{m}(1, X_i))\hat{c}(X_i)}{\hat{u}(X_i)}(1\{T_i = (0,0)\} - \hat{p}_{00}(Z_i, X_i))\hat{\kappa}_c(Z_i, X_i) \\
& - \frac{(\hat{m}(0, X_i) - \hat{m}(1, X_i))\hat{c}(X_i)}{\hat{u}(X_i)}(\hat{p}_{00}(0, X_i) - \hat{p}_{00}(1, X_i)).
\end{aligned}$$

Our estimator for $E_{Q_0}[Y^*(0,0)1\{T^* = CN\}]$ then equals $\hat{\lambda} = \sum_i \omega_i \hat{\psi}(Y_i, T_i, Z_i, X_i)$. ■

An estimator for $E_{Q_0}[Y^*(1,1)1\{T^* = CA\}]$ can be obtained through similar steps:

STEP D.1. Set $b(Z, X) \in \mathbf{R}^p$ to consist of the functions generated by interacting Z and

$(1-Z)$ with every coordinate of the baseline covariates X and compute

$$\hat{\beta}_m \in \arg\min_{\beta\in\mathbf{R}^p} \sum_{i=1}^n (\rho(Y_i)1\{T_i=(1,1)\} - b(Z_i,X_i)'\beta)^2 + \alpha\|\beta\|_1$$
$$\hat{\beta}_{11} \in \arg\min_{\beta\in\mathbf{R}^p} \sum_{i=1}^n (1\{T_i=(1,1)\} - b(Z_i,X_i)'\beta)^2 + \alpha\|\beta\|_1$$
$$\hat{\beta}_{01} \in \arg\min_{\beta\in\mathbf{R}^p} \sum_{i=1}^n (1\{T_i=(0,1)\} - b(Z_i,X_i)'\beta)^2 + \alpha\|\beta\|_1$$
$$\hat{\gamma}_\kappa \in \arg\min_{\gamma\in\mathbf{R}^p} \sum_{i=1}^n \{\frac{1}{2}(b(Z_i,X_i)'\gamma)^2 - (b(1,X_i)-b(0,X_i))'\gamma\} + \alpha\|\gamma\|_1,$$

with α selected through leave-one-out cross validation. ■

Step D.2. Set $f(X)\in\mathbf{R}^q$ to equal X and compute the following penalized estimators

$$\hat{\pi}_{cc} \in \arg\min_{\pi\in\mathbf{R}^q} \sum_{i=1}^n (1\{T_i=\{(0,0),(1,0)\}\}b(Z_i,X_i)'\hat{\gamma}_\kappa + f(X_i)'\pi)^2 + \alpha\|\pi\|_1$$
$$\hat{\pi}_{ca} \in \arg\min_{\pi\in\mathbf{R}^q} \sum_{i=1}^n (1\{T_i=(0,1)\}b(Z_i,X_i)'\hat{\gamma}_\kappa + f(X_i)'\pi)^2 + \alpha\|\pi\|_1$$

with α selected through leave-one-out cross validation. ■

Step D.3. Using the estimators from Steps 1 and 2 define the following estimators

$$\hat{m}(Z_i,X_i) \equiv b(Z_i,X_i)'\hat{\beta}_m \qquad \hat{p}_{11}(Z_i,X_i) \equiv b(Z_i,X_i)'\hat{\beta}_{11}$$
$$\hat{p}_{01}(Z_i,X_i) \equiv b(Z_i,X_i)'\hat{\beta}_{01} \qquad \hat{\kappa}_c(Z_i,X_i) \equiv b(Z_i,X_i)'\hat{\gamma}_\kappa$$
$$\hat{u}(X_i) \equiv f(X_i)'(\hat{\pi}_{cc}+\hat{\pi}_{ca}) \qquad \hat{c}(X_i) \equiv \frac{f(X_i)'\hat{\pi}_{ca}}{\hat{u}(X_i)}.$$

Employing these estimators we put them all together into the orthogonal score by setting

$$\begin{aligned}
\hat{\psi}(Y_i,T_i,Z_i,X_i) = {} & (\rho(Y_i)1\{T_i=(1,1)\} - \hat{m}(Z_i,X_i))\hat{c}(X_i)\hat{\kappa}_c(Z_i,X_i) \\
& + (\hat{m}(1,X_i)-\hat{m}(0,X_i))\hat{c}(X_i) \\
& - \frac{(\hat{m}(1,X_i)-\hat{m}(0,X_i))}{\hat{u}(X_i)}(1\{T_i=(0,1)\} - \hat{p}_{01}(Z_i,X_i))\hat{\kappa}_c(Z_i,X_i) \\
& - \frac{(\hat{m}(1,X_i)-\hat{m}(0,X_i))}{\hat{u}(X_i)}(\hat{p}_{01}(1,X_i) - \hat{p}_{01}(0,X_i)) \\
& - \frac{(\hat{m}(1,X_i)-\hat{m}(0,X_i))\hat{c}(X_i)}{\hat{u}(X_i)}(1\{T_i=(1,1)\} - \hat{p}_{11}(Z_i,X_i))\hat{\kappa}_c(Z_i,X_i) \\
& - \frac{(\hat{m}(1,X_i)-\hat{m}(0,X_i))\hat{c}(X_i)}{\hat{u}(X_i)}(\hat{p}_{11}(1,X_i) - \hat{p}_{11}(0,X_i)).
\end{aligned}$$

Our estimator for $E_{Q_0}[Y^*(1,1)1\{T^*=CA\}]$ then equals $\hat{\lambda} = \sum_i \omega_i \hat{\psi}(Y_i,T_i,Z_i,X_i)$. ■

All the parameters in Section 3.2 can be computed by employing plug-in estimators based on the preceding algorithms. For instance, to estimate CDE_0 we employ that

$$\text{CDE}_0 = \frac{E_{Q_0}[Y^*(1,0)1\{T^* = CN\}] - E_{Q_0}[Y^*(0,0)1\{T^* = CN\}]}{Q_0(T^* = CN)}$$

and estimate $E_{Q_0}[Y^*(1,0)1\{T^* = CN\}]$ using Steps B.1-B.3, $E_{Q_0}[Y^*(0,0)1\{T^* = CN\}]$ using Steps C.1-C.3, and $Q_0(T^* = CN)$ using Steps A.1-A.3. Similarly, noting

$$\text{CDE}_1 = \frac{E_{Q_0}[Y^*(1,1)1\{T^* = CA\}] - E_{Q_0}[Y^*(0,1)1\{T^* = CA\}]}{Q_0(T^* = CA)}$$

we obtain a plug-in estimator by employing Steps D.1-D.3 to estimate $E_{Q_0}[Y^*(1,1)1\{T^* = CA\}]$, Steps B.1-B.3 to estimate $E_{Q_0}[Y^*(0,1)1\{T^* = CA\}]$, and Steps A.1-A.3 to estimate $Q_0(T^* = CA)$. Finally, to estimate CTE we observe that

$$\text{CTE} = \frac{E_{Q_0}[Y^*(1,1)1\{T^* \in \{CA, CC\}\}] - E_{Q_0}[Y^*(1,1)1\{T^* = CA\})]}{Q_0(T^* = CC)} - \frac{E_{Q_0}[Y^*(0,0)1\{T^* \in \{CN, CC\}\}] - E_{Q_0}[Y^*(0,0)1\{T^* = CN\}]}{Q_0(T^* = CC)},$$

and compute $E_{Q_0}[Y^*(1,1)1\{T^* \in \{CA, CC\}\}]$ and $E_{Q_0}[Y^*(0,0)1\{T^* \in \{CN, CC\}\}]$ employing Steps B.1-B.3, $E_{Q_0}[Y^*(1,1)1\{T^* = CA\}]$ and $E_{Q_0}[Y^*(0,0)1\{T^* = CN\}]$ employing Steps D.1-D.3 and C.1-C.3 respectively, and $Q_0(T^* = CC)$ employing Steps A.1-A.3. All standard errors are obatined through the Delta method.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, **113** 231–263.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis – A Hitchhiker's Guide*. Springer-Verlag, Berlin.

Angrist, J., Autor, D. and Pallais, A. (2022). Marginal effects of merit aid for low-income students. *The Quarterly Journal of Economics*, **137** 1039–1090.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, **90** 431–442.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, **92** 1171–1176.

Bartlett, P. L., Mendelson, S. and Neeman, J. (2012). l 1-regularized linear regression: persistence and oracle inequalities. *Probability theory and related fields*, **154** 193–224.

Bickel, P. J., Klassen, C. A., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York.

Bogachev, V. I. (2007). *Measure Theory: Volume II*. Springer-Verlag, Berlin.

Bogachev, V. I. and Smolyanov, O. G. (2017). *Topological vector spaces and their applications*. Springer.

Bühlmann, P. and van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Carneiro, P., Heckman, J. J. and Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, **101** 2754–2781.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *journal of Econometrics*, **32** 189–218.

Chen, X. and Santos, A. (2018). Overidentification in regular models. *Econometrica*, **86** 1771–1817.

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.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K. and Robins, J. M. (2022a). Locally robust semiparametric estimation. *Econometrica*, **90** 1501–1535.

Chernozhukov, V., Newey, W. K. and Santos, A. (2022b). Constrained conditional moment restriction models. 1509.06311.

Chernozhukov, V., Newey, W. K. and Singh, R. (2022c). Automatic debiased machine learning of causal and structural effects. *Econometrica*, **90** 967–1027.

Clampet-Lundquist, S. and Massey, D. (2008). Neighborhood effects on economic self-sufficiency: A reconsideration of the moving to opportunity experiment. *American Journal of Sociology*, **114** 107–143.

Dutz, D., Huitfeldt, I., Lacouture, S., Mogstad, M., Torgovitsky, A. and van Dijk, W. (2021). Selection in surveys: Using randomized incentives to detect and account for nonresponse bias. Tech. rep., National Bureau of Economic Research.

Fang, Z. and Santos, A. (2018). Inference on directionally differentiable functions. *The Review of Economic Studies*, **86** 377–412.

Frölich, M. (2007). Nonparametric iv estimation of local average treatment effects with covariates. *Journal of Econometrics*, **139** 35–75.

Frölich, M. and Huber, M. (2017). Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 1645–1666.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* 315–331.

Heckman, J. J. and Pinto, R. (2018). Unordered monotonicity. *Econometrica*, **86** 1–35.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, **73** 669–738.

Heckman, J. J. and Vytlacil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences*, **96** 4730–4734.

Heckman, J. J. and Vytlacil, E. J. (2001). *Instrumental variables, selection models, and tight bounds on the average treatment effect*. Springer.

Imai, K., Keele, L. and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, **25** 51–71.

Imai, K., Tingley, D. and Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **176** 5–51.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.

Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, **64** 555–574.

Katz, L. F., Kling, J. R. and Liebman, J. B. (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *The quarterly journal of economics*, **116** 607–654.

Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, **131** 1795–1848.

Kling, J. R., Liebman, J. B. and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, **75** 83–119.

Kling, J. R., Ludwig, J. and Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, **120** 87–130.

Le Cam, L. and Yang, G. L. (1988). On the preservation of local asymptotic normality under information loss. *The Annals of Statistics* 483–520.

Lee, S. and Salanié, B. (2018). Identifying effects of multivalued treatments. *Econometrica*, **86** 1939–1963.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R. and Sanbonmatsu, L. (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American economic review*, **103** 226–231.

Ludwig, J., Liebman, J. B., Kling, J. R., Duncan, G. J., Katz, L. F., Kessler, R. C. and Sanbonmatsu, L. (2008). What can we learn about neighborhood effects from the moving to opportunity experiment? *American Journal of Sociology*, **114** 144–188.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, New York.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *American Economic Review Papers and Proceedings*, **80** 319–323.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

Mogstad, M., Santos, A. and Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, **86** 1589–1619.

Mogstad, M., Torgovitsky, A. and Walters, C. R. (2021). The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review*, **111** 3663–98.

Mountjoy, J. (2022). Community colleges and upward mobility. *American Economic Review*, **112** 2580–2630.

Munkres, J. R. (2000). Topology. *Inc., Upper Saddle River*.

Neyman, J. (1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science* 465–472. Orig. published in 1923.

Orr, L., Feins, J. D., Jacob, R. and Beecroft, E. (2003). *Moving to Opportunity Interim Impacts Evaluation*. U.S. Department of Housing and Urban Development Office of Policy Development & Research, Washington, DC.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers, 411–420.

Pinto, R. (2021). Beyond intention to treat: Using the incentives in moving to opportunity to identify neighborhood effects. *NBER Working Paper*.

Pollard, D. (2002). *A user's guide to measure theoretic probability*, vol. 8. Cambridge University Press.

Robins, J. M. (2003). Semantics of causal dag models and the identification of direct and indirect effects. *Oxford Statistical Science Series* 70–82.

Rose, E. K. and Shem-Tov, Y. (2021). On recoding ordered treatments as binary indicators. URL https://arxiv.org/abs/2111.12258.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70** 41–55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, **66** 688.

Rudin, W. (1991). Functional analysis, mcgrawhill. *Inc, New York*, **45** 46.

Singh, R. and Sun, L. (2022). Double robustness for complier parameters and a semiparametric test for complier characteristics. Tech. rep., Working Paper, MIT.

Smucler, E., Rotnitzky, A. and Robins, J. M. (2019). A unifying approach for doubly-robust ℓ_1 regularized estimation of causal contrasts. *arXiv preprint arXiv:1904.03737*.

Tan, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association*, **101** 1607–1618.

van der Vaart, A. (1991a). Efficiency and hadamard differentiability. *Scandinavian Journal of Statistics* 63–75.

van der Vaart, A. (1991b). On differentiable functionals. *The Annals of Statistics* 178–204.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, New York.

Zhai, A. (2018). A high-dimensional clt in w_2 distance with near optimal convergence rate. *Probability Theory and Related Fields*, **170** 821–845.