Many Alternatives, Random Incentives: Sharp Bounds and Efficiency in Treatment Selection Models*

Atom Vayalinkal†

JOB MARKET PAPER

November 4, 2026

Frequently updated; click here for latest version.

In this paper, I present identification and estimation results for models with essential heterogeneity, multivalued treatments and (as-good-as) randomly assigned multivalued instruments. I consider a broad class of models that restrict how instruments influence selection into treatment while remaining agnostic about outcomes, including *all* models that simply rule out certain ways of responding to the instrument(s). I characterize the testable implications of such models as convex moment inequalities and, for a broad and flexible class of target parameters, express sharp bounds as solutions to finite-dimensional convex programs. These results follow from a convex representation of such models' empirical content, with key support functions computable exactly from finitely many linear programs even when outcomes are continuous. This enables statistical inference using existing methods, without discretizing outcomes. I also derive a semiparametric efficiency bound for estimation of the bounds and propose an estimator that asymptotically attains the efficiency bound in suitably regular cases, while mitigating the inward bias associated with analog estimators in irregular cases. Simulations show that this approach can be substantially more informative than discretization-based approaches, often at much lower computational cost.

*A first version of this paper was presented in June 2023 at two talks at the University of Chicago, under the title "Many Treatments, Some Discrete Instruments: Choice Models and the Identification of Causal Effects": to the Chicago Experiments Group, June 22, 2023, hosted by John List, and at the Center for the Economics of Human Development, June 27, 2023, hosted by Jim Heckman. A second version was presented, under the title "Sharp Identification Regions in General Selection Models with (Un)ordered Treatments and Discrete Instruments", at the Chicago Causal Inference Student Conference at the University of Chicago, March 22, 2024.

†Ph.D. Candidate, Dept. of Economics, University of Toronto (Contact: atom.vayalinkal@mail.utoronto.ca).
Acknowledgments: I am indebted to my supervisors Jiaying Gu and Ismael Mourifié for their consistent support, encouragement and guidance, without which this project would not have been possible, and I am grateful to Kory Kroft and Yuanyuan Wan for their advice and support. I thank Saminul Haque, Matthew Koster, David Ledvinka and Jessica Liu for technical suggestions that significantly improved the paper, and Nanxi Fu, Jacob Kelly, Oliver Scott-Pankratz, Olivia Robins, Fardin Syed and Amalrose Vayalinkal for feedback on early versions. This paper benefitted from discussions with Rawnak Amin, Yuehao Bai, Salim Benhachmi, Alison Cane, JoonHwan Cho, Michael Cuna, Faith Fatchen, Dalia Ghanem, Leonard Goff, Sid Kankanala, Quinlan Lee, Ariel Listo, James Macek, Anthony McCanny, Talha Naeem, David Novgorodsky, Rodrigo Pinto, Lina Ramirez, Thomas Russell, Adrian Schröder, Azeem Shaikh and Gregory Sun. I thank the participants in various seminars and brownbags at the University of Toronto, the University of Chicago, Washington University in St. Louis, and Université du Québec à Montréal (Canadian Economic Association 2025 Meeting); I also thank attendees to poster sessions at the Centre Interuniversitaire de Recherche en Économie Quantitative and at the Canadian Econometrics Study Group (2024 and 2026 Meetings). I gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada. I am grateful to the Chicago Experiments Group at the University of Chicago and to the Department of Economics at Washington University in St. Louis for supporting me hospitably as a visitor while I worked on this project in Summer 2022 and Summer 2023, and Winter 2024, respectively.

1. Introduction

In many settings of interest to economists, individuals self-select into one among many available alternative treatments or programs using idiosyncratic information that is related to their potential outcomes, but not observed by the economist. In some cases, outcomes may only be observed for individuals that select into certain treatments (e.g., wages are only observed for the employed). Evaluating the effects of treatments in such settings is a common problem in economics. One widely used solution is to rely on instrumental variables (IV) methods, which exploit natural, policy-induced, or experimental variation that is "as-good-as" randomly assigned and can affect individuals' realized treatment but not their potential outcomes. What this variation reveals about treatment effects, however, depends on how individuals respond to the instrument. When there are many alternatives, for example, changes in the instrument may induce different individuals to switch between different treatments, and when individuals respond heterogeneously to treatment, the resulting changes in realized outcomes depends both on which switches occur and on who makes them. Therefore, using IV to evaluate the effects of treatments requires the economist to impose a model, referred to here as a *treatment selection model*, for how changes in the instrument can influence selection into treatment. In this paper, I present a new general framework for the partial identification of a broad and flexible class of target parameters, including counterfactual choice probabilities, mean treatment effects, and ratios of treatment effects, in such models, and for sharply testing the compatibility of such models with the observed data.

I present new results for settings where causal relationships can be described by:

$$\text{Treatment Selection Equation: } D = g_D(Z, X, \boldsymbol{V}) \,, \tag{1.1}$$

$$\text{Outcome Equation: } Y = g_Y(D, X, \boldsymbol{V}, \boldsymbol{\varepsilon}_Y) \,, \tag{1.2}$$

where g_D and g_Y are (generally unknown) structural functions, $\boldsymbol{V}$ and $\boldsymbol{\varepsilon}_Y$ are (potentially infinite dimensional) independent unobservables, X is a vector of baseline covariates, Y is a (potentially continuous) outcome, D is a categorical treatment, and Z is a categorical instrument randomly assigned (potentially conditional on X). I show that, under these basic requirements, this model is formally equivalent to the usual potential outcomes IV model. By treatment selection model (TSM), I mean models that impose restrictions on the treatment selection equation, like shape restrictions on g_D or restrictions on the distribution of $\boldsymbol{V}$.

Identification in this setting is hindered by the presence of the unobservable confounder $\boldsymbol{V}$, which can generate selection bias. As in Heckman and Pinto (2018), this paper's approach to the selection bias problem uses the instrument Z to partition the support of $\boldsymbol{V}$ into cells that induce the same vector of counterfactual treatment assignments, or *response-type*.[1]

[1] In order to accommodate a richer class of parameters, the notion of response-types used in this paper is a suitably "enriched" version of the usual notion of response-type (or "principal stratum"), as considered in Balke and Pearl (1994), Imbens and Angrist (1994), Heckman and Pinto (2018), Frangakis and Rubin (2002), Gu and Russell (2022), Lee and Salanié (2023), and many other papers; the response-types considered in this paper does not, however, involve outcomes directly, unlike the "generalized principal strata" considered by Bai, Huang, Moon, Santos, Shaikh, and Vytlacil (2026).

I show that assumptions on g_D and/or $\boldsymbol{V}$ have identifying power *only* through restrictions on the relative frequencies of different response-types (including, in particular, which response-types are eliminated by the model); although some results are developed in full generality, the strongest results apply to models that restrict these frequencies ("response-type probabilities") to lie in a convex set, which includes all models that simply eliminate some types. Heckman and Pinto (2018) showed that the problem of identifying response-type conditional counterfactuals can be represented as a finite mixture model where the mixture components are the conditional distributions generated by response-types. I establish the sharpness of this representation for treatment selection models and attack this mixture problem directly, leveraging a geometric characterization of such mixture models to develop new results on the partial identification of components of overlapping mixtures, including a characterization of bounds on mean features of such components in terms of convex moment inequalities. This approach also allows the sharp testable implications of such models to be cast in terms of a particularly simple class of such moment inequalities.

The proposed approach sharply incorporates the full identifying power of randomly assigned instruments, and can tractably accommodate continuous outcomes without discretization. The unified approach of this paper allows researchers to flexibly choose the target parameter and selection model most relevant for their policy question and setting, without requiring new derivations, even when such a model/parameter has not been previously considered in the methodological literature. Since D and Z are only required to be finitely-supported, the results in this paper apply to settings with multiple dimensions of treatment (e.g., mediation analysis) and/or unobserved heterogeneity. In particular, the framework of this paper applies to settings with "multiple" instruments or treatments just as easily as it does to settings with "multivalued" instruments or treatments: such distinctions are incorporated via restrictions on the selection equation. Moreover, by allowing Z (or some components of it) to be included as components of D, the results also apply to models where conventional IV exclusion restrictions may be violated (fully or partially, as specified by the TSM) . For several important cases of models and parameters, I also show that the sharp bounds admit analytic expressions, which can help provide intuition for the source and structure of identifying power in such models. I emphasize, however, that such expressions are not needed to implement the approach proposed in this paper, which remains highly tractable even when such expressions are not available.

The identification results in this paper are sharp in the sense that they exhaust all of the information in the model and in the available data. Importantly, these results are sharp under the empirically and theoretically common assumption of random assignment ("full" independence) of the instrument, as opposed to the weaker assumption of mean independence considered in some of the related literature on partial identification in IV models (as in, e.g., Manski (1990); Mogstad, Santos, and Torgovitsky (2018)). This distinction can be important, since the stronger full independence assumption can lead to tighter bounds on target parameters, as discussed below. Moreover, as noted by Mogstad and Torgovitsky (2018), it is

not common to encounter practical applications where there are reasons to believe that mean independence holds but full independence fails to hold.

To the best of my knowledge, this paper is the first to obtain these sharp identification results at this level of generality; in particular, I believe this paper is the first to derive sharp bounds, under general support restrictions and for general principal strata conditional mean treatment effects, in potential outcome models with non-binary instruments and continuous outcomes. For several special cases, other analytic and computational approaches have been proposed in the literature (Cheng and Small, 2006; Kolesár, Olea, and Roth, 2026). These cases only cover a very small subset of the models that fall into the general class of treatment selection models considered here, however, and rely on special features of their particular settings that are not generally available for the broad class considered here.

Much of the literature on IV models with multiple treatments and/or instruments have focused on point identification under relatively restrictive assumptions on treatment selection: Angrist and Imbens (1995) assume "monotonicity" of the instrument and extend Imbens and Angrist (1994)'s results to a setting with multiple levels of (ordered) treatments, Heckman, Urzua, and Vytlacil (2008) considers multiple treatments within the Marginal Treatment Effect (MTE) framework of Heckman and Vytlacil (1999, 2001, 2007a,b) but rely on continuous monotonic instruments, Heckman and Pinto (2018) consider point identification in the same setting as this paper and propose the "Unordered Monotonicity" condition which allows several mean counterfactuals and treatment effects to be point-identified, Mountjoy (2022) provides point identification for some parameters of interest using continuous instruments and the less restrictive "Unordered Partial Monotonicity" assumption, which builds on the "Partial Monotonicity" assumption introduced by Mogstad, Torgovitsky, and Walters (2021). Lee and Salanié (2018) consider a broad class of treatment selection models based on combinations of threshold crossing rules and develop point identification results that rely on continuous instruments. When only finitely supported instruments are available, the partial identification framework of this paper can sharply incorporate any of these assumptions on treatment selection.[2]

The convex moment characterization of the model specification test and parameter inclusion tests provided here enable statistical inference using existing moment inequality methods that do not require discretization of continuous outcomes, such as Chernozhukov, Lee, and Rosen (2013, "CLR") and Andrews and Shi (2013), under the regularity conditions required by these methods. I also propose a simple adaptation of the procedure of CLR that, in this setting, asymptotically controls size uniformly over any class of data distributions with instrument assignment probabilities uniformly bounded away from zero, and is consistent

[2]In order to directly and sharply apply the identification results in this paper for a given TSM, the researcher needs to know the *a priori* restrictions placed by the TSM on the relative frequencies of response-types. In the leading case where the TSM simply *eliminates* some response-types (e.g., Ordered/Unordered/Vector/Partial Monotonicity, and the models considered in Lee and Salanié (2018) when the thresholds are known or identified.), this information simply consists of knowing what types are ruled out by the model. More generally, although this information is always "known" from the model, it may be non-trivial to derive for complicated selection models (e.g., parametric models).

against every fixed alternative. In particular, when the instrument assignment probabilities are known, the class contains *all* data distributions P with those assignment probabilities.

In a highly complementary paper, Bai et al. (2026, BHMSSV) propose a general computational approach for inference in settings with multivalued treatments, instruments, and outcomes. Their approach, which does not focus on treatment selection models specifically, can be used for inference in our setting when the outcome is finite and the treatment selection model simply eliminates some response-types. BHMSSV sets up their identification problem as one "large" linear program and use the inference approach recently proposed by Fang, Santos, Shaikh, and Torgovitsky (2023). In contrast, the approach here leverages the unique structure afforded by treatment selection models to instead work with a continuum of "small" linear programs for each treatment level. The papers are complementary in the finite outcome TSM setting, because the results in this paper can be used to implement the approach of BHMSSV with a much smaller coefficient matrix, when the model of interest is a TSM, while still obtaining a statistically identical procedure.[3] This can substantially reduce the computational burden of BHMSSV's approach when the model of interest is a TSM, while avoiding the sample-splitting required by the moment-based inference approach proposed in this paper. For continuous outcomes, however, where their approach relies on discretizing the outcome into finitely many bins and leveraging a high-dimensional result that allows them to increase the number of bins with the sample size (subject to a rate requirement) while still controlling size, their approach can be computationally demanding (even after the reduction enabled by the results in this paper) and potentially very conservative. In simulations, I find that the sample-splitting and moment inequality based approach described above outperforms the discretization-based approach of BHMSSV on both informativeness and runtime, as shown in Figure 1. The cause of tighter confidence intervals is entirely structural (i.e. due to binning), and cannot be improved by using a more powerful inference procedure: in fig. 1, the mean 95% CI obtained by the approach proposed here, with a sample size of 12,000, is narrower than the *population level* discretized identified set, with the outcome discretized to 50 bins, over 90% of the time.

Almost all of the related sharp partial identification results in IV settings with multiple treatments assume that the instrument satisfies an exogeneity condition (random assignment being the strongest such assumption) but do not consider further restrictions on selection into treatment. For example, Manski (2003) assumes that the instrument and outcomes satisfy a marginal statistical independence condition, and provides a sharp characterization of potential outcome distributions for settings with many treatments and a finitely supported instrument. Similarly, Beresteanu, Molchanov, and Molinari (2012) assume random assignment of the instrument, and use random set theory to develop sharp identification results for settings with multiple treatments and multivalued (potentially continuous) instruments, extending the results of Manski (2003), Balke and Pearl (1997), and Kitagawa (2021); the

[3] This reduction may not be numerically identical because the solutions return by linear programming solvers can be sensitive to the exact way in which the problem is specified when there are multiple optimal solutions.

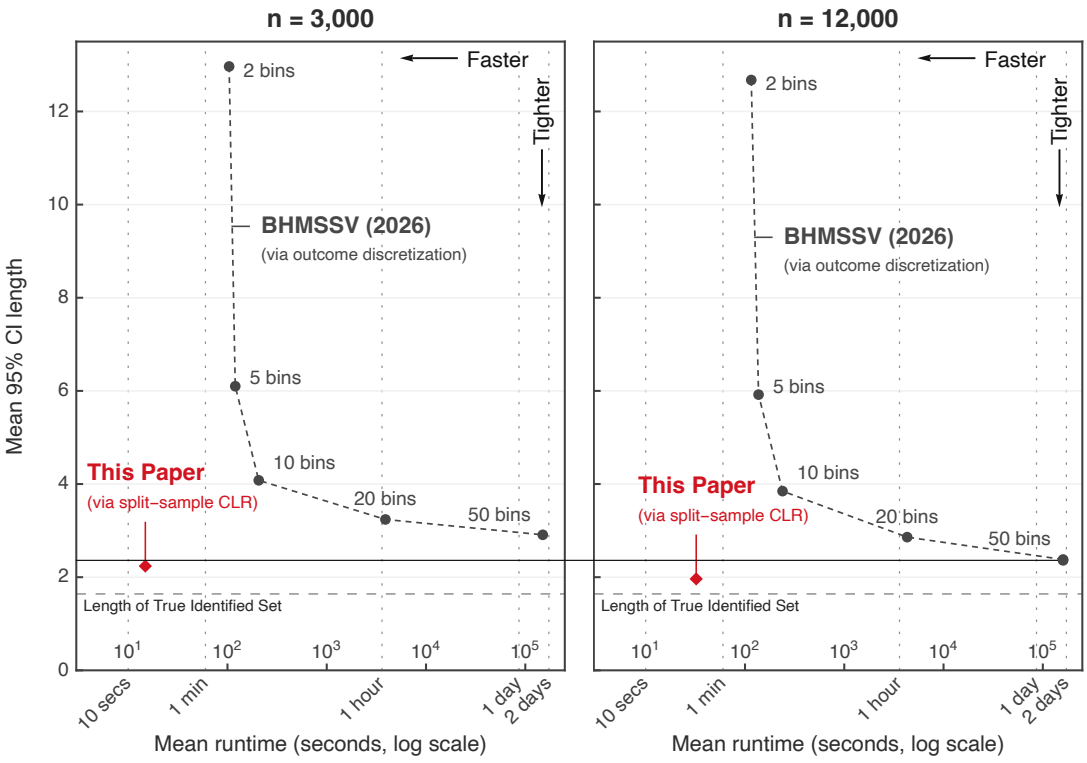


FIG 1. *The DGP used in this comparison and the runtime/width numbers reported for BHMSSV are from Fig. 1 in BHMSSV (2026). The runtime and width for the split-sample CLR procedure of this paper are based on 1000 replications run on a 2021 M1 Macbook Pro.*

notion of response-types is not present in their analysis, however, and they do not consider monotonicity or any other assumptions on treatment selection.[4] An exception is Cheng and Small (2006), who consider a setting with binary outcomes, three treatments and three instrument values, and provide sharp bounds on response-type conditional treatment effects for particular forms of one-sided noncompliance.

This paper is also related to the literature on specification testing in IV models, for which a more general graph-theoretic approach was recently proposed by Kaido and Ponomarev (2025). The moment characterization obtained in this paper is specific to TSMs, and is different from the one obtained by Kaido and Ponomarev (2025); in particular, the conditional moments obtained here are of moment functions of the treatment and the outcome and conditions on the instrument, not conditional expectations with respect to the outcome, like in Kaido and Ponomarev (2025) (and many other papers like Kitagawa (2015); Mourifié and Wan (2017); Bai, Huang, and Tabord-Meehan (2025)). The resulting statistics are similar, both in spirit and mathematical provenance, to the inclusion tests derived in Beresteanu,

[4] Although this paper's approach to identification does not involve random sets, the *Minkowski integral* operation from convex geometry (a generalization of the Minkowski sum), central to the approach proposed here, is related to the notion of *Aumann expectation* from random set theory, used in Beresteanu et al. (2012) and many other papers.

Molchanov, and Molinari (2011) for incomplete models with convex moment predictions. The moment functions derived here depend on the data because they are "profiled" in the sense that they are solutions to an optimization problem that depends on the data. The feasible set for these problems does not depend on the data, however, and so, it is straightforward to obtain a "deprofiled" characterization that only involves moments with known functions, which may be of interest for applying certain off-the-shelf inference methods, and avoiding the sample-splitting of the inference approach discussed above.

This paper extends the results of Kitagawa (2015) and Mourifié and Wan (2017) on sharp testable implications of LATE assumptions to treatment selection models beyond monotonicity and to the general multivalued treatment and instrument setting, and also provides an alternative characterization of their tests. As special cases, this paper also provides a sharp test for random assignment, for which a test was proposed by Kédagni and Mourifié (2020), and also for Unordered Montonicity, for which a test was proposed by Sun (2023). Neither of these papers show sharpness under full instrument independence, however, unlike this paper.[5] The framework of this paper can also be used to perform sensitivity analysis in the spirit of Noack (2021), but I do not pursue this formally in this draft. This paper is also related to the literature on the identification of joint distribution of potential outcomes, as investigated by Heckman and Smith (1995, 1998), Heckman, Smith, and Clements (1997), and more recently by Mourifié, Henry, and Méango (2020), Jun and Lee (2023) and Kaji and Cao (2023); the approach of this paper is fully nonparametric and in a relatively more general setting. Finally, this paper is related to the literature on nonparametric identification of mixture components, as in Horowitz and Manski (1995, 1998, 2000), Cross and Manski (2002), and Henry, Kitamura, and Salanié (2014), and to the literature on nonparametric sample selection, as in Lee (2009). For several special cases, the analytic results derived in this paper share features with the analytic results of Horowitz and Manski (1995), Cross and Manski (2002), Lee (2009), Kitagawa (2021), Kaji and Cao (2023), and Marx (2024).

Notation. I use ":=" to indicate that an equality holds by definition. For any set K, $|K|$ denotes its cardinality. For any number $n \in \mathbb{N}$, $[n] := \{1, \ldots, n\}$. For any two sets A and B, A^B denotes the set of all functions from B to A; when B is finite, we will usually denote $\boldsymbol{x}(b)$ by $\boldsymbol{x}_{(b)}$, for any given $\boldsymbol{x} \in A^B$ and $b \in B$. For any $n \in \mathbb{N}$, $A^n := A^{[n]}$. Throughout, any vector of the form $(x_1, \ldots, x_K)$ will be understood to be the column vector $[x_1 \cdots x_K]^\intercal$. For any pair of matrices (or vectors) A, B with the same number of columns, (A, B) is the matrix M obtained by stacking A and B vertically; i.e. $M_{(i,\cdot)} = A_{(i,\cdot)}$ for $1 \leq i \leq |\operatorname{rows}(A)|$ and $M_{(i,\cdot)} = B_{(i-|\operatorname{rows}(A)|,\cdot)}$ for $|\operatorname{rows}(A)| \leq i \leq |\operatorname{rows}(A)| + |\operatorname{rows}(B)|$. For any vector $\boldsymbol{v}$, $\{\boldsymbol{v}\}_+$ denotes the nonnegative part of $\boldsymbol{v}$, i.e. $\left(\{\boldsymbol{v}\}_+\right)_{(i)} := \max\{\boldsymbol{v}_{(i)}, 0\}$; similarly, $\{\boldsymbol{v}\}_-$ denotes the non-positive part of $\boldsymbol{v}$, i.e. $\left([\{\boldsymbol{v}\}_-\right)_{(i)} := \min\{\boldsymbol{v}_{(i)}, 0\}$. For any event A, denote by $\mathbb{1}\{A\}$ the indicator function takes value 1 if A is true, and 0 otherwise. For any random variable X, $\operatorname{Law}(X)$ denotes its distribution, and $\operatorname{supp}(X)$ denotes its support, defined

[5] In fact, the results in this paper suggest that the conjectured sharpness (under random assignment/"full independence" of Z) of the approach in Kédagni and Mourifié (2020) does not hold.

here as the smallest closed set that contains X with probability 1. All optimization variables without an explicitly specified domain should be understood to have their domain be the set of real-valued vectors of conformable dimension. For a measure μ on a measurable space $\mathcal{Y}$, and any subspace $\mathcal{X}$ of a finite-dimensional Euclidean space, $L^1(\mu;\mathcal{X})$ is the space of μ-integrable functions $\mathcal{Y} \to \mathcal{X}$, and $L^2(\mu;\mathcal{X})$ is the space of such functions that are also square-integrable; when the second argument is omitted, $\mathcal{X}$ is understood to be $\mathbb{R}$.

2. Analytic Framework

We will work with a complete and sufficiently rich probability space, $(\Omega, \Sigma, \mathbb{P})$, on which all random variables in this paper are defined; following the usual convention, I identify the sample space Ω with the population of individuals. For each individual $\omega \in \Omega$, we observe $(Y(\omega), D(\omega), Z(\omega), X(\omega))$, where Y is their (potentially continuous and/or vector-valued) outcome(s), D represents the treatment(s) they received, Z represents their assigned instrument value(s), and X denotes a vector of their baseline covariates.

Denote by P the distribution of the observed data (Y, D, Z, X). In Sections 2 to 4, since our focus is on identification, I will generally treat P as known; estimation and inference based on finite samples is addressed in Section 5 and Appendix C, respectively. I condition on X throughout the remainder, and therefore suppress it in the notation. I assume that D and Z are categorical variables, and only assume that Y satisfies some mild regularity conditions. Formally, I assume throughout that P satisfies the following condition.

Condition D (Setting)**.** *The random variables D ("treatment") and Z ("instrument") take on values in finite non-empty sets $\mathcal{D}$ and $\mathcal{Z}$ respectively, with $|\mathcal{D}| \geq 2$ and*

$$P(Z = z) > 0, \ \forall z \in \mathcal{Z} \ .$$

The random variable Y ("outcome") takes on values in a closed Borel set $\mathcal{Y} \subseteq \mathbb{R}^k$, equipped with the Borel σ-algebra $\mathscr{B}_{\mathcal{Y}}$ and a dominating σ-finite measure μ.

Note that this condition allows the distribution of Y to be discrete, continuous, or mixed.

Following Heckman and Pinto (2018), I formalize the basic instrumental variables (IV) setting using the following condition on P, which we will use to formally define the notion of treatment selection models used in this paper.

Condition M (Structural IV Model)**.** *P admits a representation of the following form:*

$$\textit{Treatment Selection Equation:} \qquad D = g_D(Z, \boldsymbol{V}) \ , \tag{2.1}$$

$$\textit{Outcome Equation:} \qquad Y = g_Y(D, \boldsymbol{V}, \varepsilon_Y) \ , \tag{2.2}$$

$$\textit{Independence Condition:} \qquad Z, \boldsymbol{V}, \textit{ and } \varepsilon_Y \textit{ are mutually independent.} \tag{2.3}$$

for random variables $\boldsymbol{V}$ and ε_Y taking values in measurable spaces $\boldsymbol{\mathcal{V}}$ and $\boldsymbol{\mathcal{E}}_Y$, respectively, and measurable functions $g_D : \mathcal{Z} \times \boldsymbol{\mathcal{V}} \to \mathcal{D}$ and $g_Y : \mathcal{D} \times \boldsymbol{\mathcal{V}} \times \boldsymbol{\mathcal{E}}_Y \to \mathcal{Y}$ such that, for all $d \in \mathcal{D}$, $g_Y(d, \boldsymbol{V}, \varepsilon_Y)$ has a μ-dominated distribution.

This model makes explicit the (potentially infinite-dimensional) latent variables $\boldsymbol{V}$ and ε_Y and the structural functions g_D and g_Y, but is equivalent to the usual IV potential outcomes model.[6] The independence condition is equivalent to Z being randomly assigned.

Under Condition M, we can define potential choices and outcomes as follows. For each $d \in \mathcal{D}, z \in \mathcal{Z}$, and $\omega \in \Omega$, define

$$D_z(\omega) := g_D(z, \boldsymbol{V}(\omega)) \text{ and } Y_d(\omega) := g_Y(d, \boldsymbol{V}(\omega), \varepsilon_Y(\omega)) . \tag{2.4}$$

Now, under Condition M, the observed choices and outcomes are generated by (2.1)-(2.2), which implies the usual formulation via (2.4),

$$D = \sum_{z \in \mathcal{Z}} D_z \mathbb{1}\{Z = z\} \text{ and } Y = \sum_{d \in \mathcal{D}} Y_d \mathbb{1}\{D = d\} . \tag{2.5}$$

In this paper, we will focus on Treatment Selection Models (TSMs), defined as follows.

Definition 2.1 (Treatment Selection Model)**.** *A **treatment selection model** (TSM) is a restriction on the treatment selection mechanism*

$$D = g_D(Z, \boldsymbol{V})$$

in Condition M. Precisely, on a specified support $\boldsymbol{\mathcal{V}}$, a TSM $\mathbf{S}$ is specified as a nonempty collection of admissible pairs $(F_{\boldsymbol{V}}, g_D)$, where $F_{\boldsymbol{V}}$ is the distribution of the latent variable $\boldsymbol{V}$ and g_D is the treatment selection rule.

A TSM $\mathbf{S}$ may restrict the distribution of the latent confounder $\boldsymbol{V}$, the shape of the structural function g_D, or both. Note that $\boldsymbol{V}$ generally has no representation-invariant meaning under Condition M alone, but can become economically interpretable once a TSM $\mathbf{S}$ is imposed. In the remainder, I assume that the analyst has specified a treatment selection model (TSM) $\mathbf{S}$. Along with $\mathbf{S}$, I also assume that the analyst has specified a latent subpopulation $\boldsymbol{\mathcal{V}}^\star \subseteq \boldsymbol{\mathcal{V}}$ of interest, the interpretation of which depends on $\mathbf{S}$.

A leading class of $\boldsymbol{\mathcal{V}}^\star$ is principal-strata, sets of the form $\boldsymbol{\mathcal{V}}^\star = \{(D_z : z \in \mathcal{Z}) \in A^\star\}$, which stratify the population based on the set of counterfactual choices they would make when faced with different values of Z. Generally, however, the subpopulation of interest need not be determined by principal stratum. For example, a TSM may introduce latent costs or preferences that can distinguish between individuals with identical treatment choices over the observed instrument support.

2.1. **Response-Types.** Define the random vector $\boldsymbol{D}$ indicating the principal-stratum that an individual belongs to,

$$\boldsymbol{D} = \boldsymbol{D}(\boldsymbol{V}) := (g_D(z, \boldsymbol{V}) : z \in \mathcal{Z}) = (D_z : z \in \mathcal{Z}) ,$$

[6] I provide a formal statement and proof of this claim in Appendix D.

and then define the random vector $\boldsymbol{T}$

$$\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{V}) := (\boldsymbol{D}, \mathbb{1}\{\boldsymbol{V} \in \mathcal{V}^\star\}) \ ,$$

which denotes an individual's *response-type*. If membership in $\boldsymbol{\mathcal{V}}^*$ is determined by $\boldsymbol{D}$, then we may simply take $\boldsymbol{T} = \boldsymbol{D}$. Otherwise $\boldsymbol{T}$ can be viewed as an "enriched" stratification that further distinguishes individuals in the target population within each principal-stratum.

2.2. **Parameters of Interest.** We will focus on target parameters that can be expressed in one of the following forms, for some measurable $\boldsymbol{\mathcal{V}}^* \subseteq \boldsymbol{\mathcal{V}}$.

(i) Response-type probabilities: $\mathbb{P}(\boldsymbol{V} \in \boldsymbol{\mathcal{V}}^*)$. For example, $\mathbb{P}(\boldsymbol{T} = \boldsymbol{\tau}) = \mathbb{P}\left(\boldsymbol{V} \in \boldsymbol{T}^{-1}(\boldsymbol{\tau})\right)$.
(ii) (Un)conditional mean outcomes: for known $\alpha_d \in \mathbb{R}$ and integrable $\kappa_d : \mathcal{Y} \to \mathbb{R}$,

$$\mathbb{E}\left[\sum_{d \in \mathcal{D}} \alpha_d \kappa_d(Y_d) \,\middle|\, \boldsymbol{V} \in \boldsymbol{\mathcal{V}}^*\right] \ .$$

For example, $\mathbb{E}[Y_d - Y_{d'} \,|\, \boldsymbol{T} = \boldsymbol{\tau}]$ or $\mathbb{E}[Y_d - Y_{d'}]$.
(iii) Ratios of (un)conditional mean outcomes: for known $\alpha_d, \beta_d \in \mathbb{R}$ (with $\beta_d \neq 0$ for at least one $d \in \mathcal{D}$) and integrable $\kappa_d : \mathcal{Y} \to \mathbb{R}$,

$$\frac{\mathbb{E}\left[\sum_{d \in \mathcal{D}} \alpha_d \kappa_d(Y_d) \,\middle|\, \boldsymbol{V} \in \boldsymbol{\mathcal{V}}^*\right]}{\mathbb{E}\left[\sum_{d \in \mathcal{D}} \beta_d \kappa_d(Y_d) \,\middle|\, \boldsymbol{V} \in \boldsymbol{\mathcal{V}}^*\right]} \ .$$

For example, $\frac{\mathbb{E}[Y_{(0,1)} - Y_{(0,0)} \,|\, \boldsymbol{T} = \boldsymbol{\tau}]}{\mathbb{E}[Y_{(1,1)} - Y_{(0,0)} \,|\, \boldsymbol{T} = \boldsymbol{\tau}]}$ or $\frac{\mathbb{E}[Y_{(0,1)} - Y_{(0,0)}]}{\mathbb{E}[Y_{(1,1)} - Y_{(0,0)}]}$.

2.3. **Parent Distributions.** Note that our parameters of interest are now features of the (unknown) joint distribution, denoted Q, of

$$((Y_d : d \in \mathcal{D}), \boldsymbol{T}, Z) \ ,$$

and we shall refer to such distributions as parent distributions. Therefore, we can use Q as our primitive object moving forward. Let $\mathcal{M}$ denote the set of possible parent distributions; precisely, define $\mathcal{M}$ as the set of probability laws of $((Y_d : d \in \mathcal{D}), \boldsymbol{T}, Z)$ over the specified supports, with $Q_{Y_d} \ll \mu$ for all $d \in \mathcal{D}$ and $Z \perp ((Y_d : d \in \mathcal{D}), \boldsymbol{T})$. For each $Q \in \mathcal{M}$, denote by P^Q the distribution of (Y, D, Z) induced by Q via (2.5). Now, the identified set for Q under Condition M is given by

$$\mathcal{Q} = \mathcal{Q}(P) := \left\{Q \in \mathcal{M} \,\middle|\, P^Q = P\right\} \ .$$

We will say that a parent distribution Q *satisfies* $\mathbf{S}$ if and only if there exists a structural representation satisfying Condition M, with $(F_{\boldsymbol{V}}, g_D) \in \mathbf{S}$, that induces Q as the joint distribution of $\left((Y_d)_{d \in \mathcal{D}}, \boldsymbol{T}, Z\right)$. The identified set for Q under $\mathbf{S}$, denoted $\mathcal{Q}_{\mathbf{S}} = \mathcal{Q}_{\mathbf{S}}(P)$, is defined as the set of parent distributions that (i) generate the observed data and (ii) satisfy $\mathbf{S}$. Precisely,

$$\mathcal{Q}_{\mathbf{S}} := \{Q \in \mathcal{Q} \,|\, Q \text{ satisfies } \mathbf{S}\} \ .$$

2.4. **Identifying Power of Treatment Selection Models.** Denote the set of response-types possible under $\mathbf{S}$, by

$$\boldsymbol{\mathcal{T}} := \operatorname{supp}(\boldsymbol{T}) = \{\boldsymbol{T}(\boldsymbol{v}) \mid \boldsymbol{v} \in \boldsymbol{\mathcal{V}}\} \ .$$

For each response-type $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$, denote by $\boldsymbol{\tau}(z)$ the treatment taken by those of response-type $\boldsymbol{\tau}$ when $Z = z$. Also define, for each treatment d, the subset $\boldsymbol{\mathcal{T}}_d \subseteq \boldsymbol{\mathcal{T}}$ of types that are *not* "d-Never-Takers", i.e. $\boldsymbol{\mathcal{T}}_d := \{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \,|\, \exists z \in \mathcal{Z} : \boldsymbol{\tau}(z) = d\}$.

$\boldsymbol{\mathcal{T}}$ is determined by the assumptions imposed on g_D and $\boldsymbol{V}$ by the TSM $\mathbf{S}$. If $\mathbf{S}$ places no restrictions on treatment selection, then $\boldsymbol{\mathcal{T}} = \mathcal{D}^{\mathcal{Z}} \times \{0,1\}$. More generally, $\mathbf{S}$ can restrict the distribution of $\boldsymbol{T}$ in other ways, beyond just restricting its support. The distribution of $\boldsymbol{T}$ is captured by the (unknown) vector of response-type probabilities (i.e. the relative frequencies of response-types), which we denote by $\boldsymbol{\pi}$,

$$\boldsymbol{\pi} := \left(\mathbb{P}(\boldsymbol{T} = \boldsymbol{\tau}) : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}\right) .$$

For $Q \in \mathcal{M}$, denote by $\boldsymbol{\pi}^Q$ the $\boldsymbol{\pi}$ induced by Q, i.e. $\boldsymbol{\pi}^Q := (Q(\boldsymbol{T} = \boldsymbol{\tau}) : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}})$. The identified set for $\boldsymbol{\pi}$ under $\mathbf{S}$, denoted $\boldsymbol{\Pi}$, is defined as

$$\boldsymbol{\Pi} := \left\{\boldsymbol{\pi}^Q \,\middle|\, Q \in \mathcal{Q}_{\mathbf{S}}\right\} \ .$$

The next result shows that treatment selection models have identifying power for features of parent distributions only through restrictions on the distribution of response-types. Denote by $\mathcal{R}_{\boldsymbol{\pi}} = \mathcal{R}_{\boldsymbol{\pi}}(\mathbf{S})$ the *a priori* restrictions imposed on $\boldsymbol{\pi}$ by $\mathbf{S}$, defined as

$$\begin{aligned}\mathcal{R}_{\boldsymbol{\pi}} &:= \left\{ \left(\mathbb{P}_{F_{\boldsymbol{V}}}\left[\left(\left(g_D(z, \boldsymbol{V})\right)_{z \in \mathcal{Z}}, \mathbb{1}\{\boldsymbol{V} \in \boldsymbol{\mathcal{V}}^\star\}\right) = \boldsymbol{\tau}\right] : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}\right) \,\middle|\, (F_{\boldsymbol{V}}, g_D) \in \mathbf{S} \right\} \\ &= \left\{\boldsymbol{\pi}^Q \,\middle|\, Q \in \mathcal{M} \text{ and } Q \text{ satisfies } \mathbf{S}\right\} \ .\end{aligned}$$

Note that $\mathcal{R}_{\boldsymbol{\pi}}$ is feature of $\mathbf{S}$ and does not depend on the observed data distribution P, since the definition does not impose $Q \in \mathcal{Q}$. Moreover, $\mathcal{R}_{\boldsymbol{\pi}}$ is always a finite-dimensional bounded set. We now have the following result.

Theorem 2.1. *Suppose P satisfies Condition D. Then, for any treatment selection model $\mathbf{S}$,*

$$\mathcal{Q}_{\mathbf{S}} = \{Q \in \mathcal{Q} \mid \boldsymbol{\pi}_Q \in \mathcal{R}_{\boldsymbol{\pi}}(\mathbf{S})\} \ .$$

Proof. See Appendix A.1.1. ■

In particular, we now have that $\boldsymbol{\Pi} = \left\{\boldsymbol{\pi}^Q \,\middle|\, Q \in \mathcal{Q}\right\} \cap \mathcal{R}_{\boldsymbol{\pi}}$, where the first set depends on P, but not $\mathbf{S}$, and the second set depends on $\mathbf{S}$, but not P.

2.5. **Requirements on S.** The reduction in Theorem 2.1 applies to arbitrary TSMs, but operationalizing this result generally requires additional structure on $\mathcal{R}_{\boldsymbol{\pi}}$. I now introduce convexity and polyhedrality conditions, which are satisfied by many familiar restrictions on treatment selection.

I first introduce the notion of support functions of convex sets, which will be used extensively in the remainder.

Definition 2.2. *For any non-empty closed convex set $C \subseteq \mathbb{R}^n$, let $h_C : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be its **support function**, defined as the convex function such that, for every $\boldsymbol{c} \in \mathbb{R}^n$,*

$$h_C(\boldsymbol{c}) := \sup_{\boldsymbol{x} \in C} \boldsymbol{c}^\intercal \boldsymbol{x} \ .$$

By convention, the support function of $\emptyset$ is taken to be constant at $-\infty$.

We will see later that when $\mathcal{R}_{\boldsymbol{\pi}}$ is closed (therefore compact) and convex, the identified set for $\boldsymbol{\pi}$ under $\mathbf{S}$ is a compact convex set, leading to the following condition.

Condition C. *$\mathcal{R}_{\boldsymbol{\pi}}(\mathbf{S})$ is nonempty, closed, and convex, with support function $h_{\mathbf{S}} := h_{\mathcal{R}_{\boldsymbol{\pi}}(\mathbf{S})}$.*

When $\mathbf{S}$ only satisfies Condition C, I will generally assume that the analyst knows $h_{\mathbf{S}}$, the support function of $\mathcal{R}_{\boldsymbol{\pi}}$. Many results will be most useful when $\mathbf{S}$ satisfies the following stronger condition, which is satisfied by the selection models most frequently considered in the literature, including the leading class of models that impose only support restrictions on $\boldsymbol{T}$ (i.e. models that simply eliminate some response-types).

Condition CP. *$\mathcal{R}_{\boldsymbol{\pi}}(\mathbf{S})$ is a nonempty closed convex polytope,[7] with*

$$\mathcal{R}_{\boldsymbol{\pi}}(\mathbf{S}) = \left\{ \boldsymbol{\pi} \in \mathbb{R}^{\mathcal{T}} \,\middle|\, \mathbf{A}_{\mathbf{S}} \boldsymbol{\pi} \leq \mathbf{a}_{\mathbf{S}} \right\} \ ,$$

for known matrix $\mathbf{A}_{\mathbf{S}}$ and vector $\boldsymbol{a}_{\mathbf{S}}$. In this case, $h_{\mathbf{S}}(\boldsymbol{c}) := h_{\mathcal{R}_{\boldsymbol{\pi}}}(\boldsymbol{c}) = \inf_{\boldsymbol{\lambda} \geq 0 : \mathbf{A}_{\mathbf{S}}^\intercal \boldsymbol{\lambda} = \boldsymbol{c}} \boldsymbol{\lambda}^\intercal \mathbf{a}_{\mathbf{S}}$.

Any selection model that simply eliminates response-types (including, e.g., Ordered/Unordered/Vector/Partial Monotonicity) automatically satisfies this condition. More generally, Condition CP is satisfied by any selection model $\mathbf{S}$ such that the (*a priori*) restrictions on the type distribution imposed by $\mathbf{S}$ can be characterized by a set of linear (in)equalities involving response-type probabilities.[8] The main utility of Condition CP is tractable computation of the support function of $\boldsymbol{\Pi}$; we will see later that $\boldsymbol{\Pi}$ itself is a convex polytope whenever $\mathbf{S}$ satisfies Condition CP.

3. Sharp Testable Implications of Treatment Selection Models

Notation. The following notation is used throughout the remainder. Fix an (arbitrary) ordering on $\mathcal{D}$, $\mathcal{Z}$, and $\mathcal{D}^{\mathcal{Z}}$, and assume throughout that all relevant tuples indexed by any subset of these sets follow this ordering. We will generally identify $\mathcal{D}$ with the set $\left\{d_1, \ldots, d_{|\mathcal{D}|}\right\}$.

[7] We will say that a set $K \subseteq \mathbb{R}^n$ is a ***convex polytope*** if and only if there exists, for some $m \in \mathbb{N}$, a matrix $A \in \mathbb{R}^{m \times n}$ and vector $a \in \mathbb{R}^m$ such that $K = \left\{x \in \mathbb{R}^n \,\middle|\, Ax \leq a\right\}$. Note that this definition allows K to be unbounded.

[8] In this case, $\mathcal{R}_{\boldsymbol{\pi}}$ is simply the intersection of the standard $\mathcal{D}^{\mathcal{Z}}$-simplex with the polytope defined by the non-negative solutions to such linear (in)equalities.

For each treatment $d \in \mathcal{D}$, define the response-matrix $\mathbf{R}_d \in \{0,1\}^{|\mathcal{Z}|\times|\boldsymbol{\mathcal{T}}|} \setminus \{\mathbf{0}\}$ by

$$\mathbf{R}_d[z, \boldsymbol{\tau}] := \mathbb{1}\{\boldsymbol{\tau}(z) = d\} \ ,$$

so that each column corresponds to a type $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$ and indicates the instrument values under which individuals of that type take treatment d.

For $d \in \mathcal{D}, z \in \mathcal{Z}$, define the observed subdistribution $P_{d|z}$ as follows, for Borel $A \subseteq \mathcal{Y}$,

$$P_{d|z}(A) := P(Y \in A, D = d \,|\, Z = z) \ ,$$

and denote by $\boldsymbol{P}_d$ the tuple formed by stacking over $z \in \mathcal{Z}$,

$$\boldsymbol{P}_d^P(A) := \left(P_{d|z}(A) : z \in \mathcal{Z}\right) \ .$$

Similarly, for parent distribution $Q \in \mathcal{M}$ and $d \in \mathcal{D}, \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$, define the subdistribution $Q_{d,\boldsymbol{\tau}}$ as follows, for Borel $A \subseteq \mathcal{Y}$,

$$Q_{d,\boldsymbol{\tau}}(A) := Q(Y_d \in A, \boldsymbol{T} = \boldsymbol{\tau}) \ ,$$

and stack these over $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$ to get

$$\boldsymbol{Q}_d^Q := (Q_{d,\boldsymbol{\tau}}(A) : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}) \ .$$

3.1. **Identified Set for Parent Distributions.**

Lemma 3.1 (Sharp Characterization of $\mathcal{Q}_{\mathbf{S}}$)**.** *Suppose P satisfies Condition D and let $Q \in \mathcal{M}$ with $Q_Z = P_Z$ be given. Then, $Q \in \mathcal{Q}_{\mathbf{S}}$ if and only if, for Borel subsets $A \subseteq \mathcal{Y}$,*

$$\begin{aligned} &\mathbf{R}_d \boldsymbol{Q}_d^Q(A) = \boldsymbol{P}_d^P(A) \quad \forall d \in \mathcal{D} \ , \\ &\boldsymbol{\pi}^Q \in \mathcal{R}_{\boldsymbol{\pi}} \ . \end{aligned} \tag{$\star$}$$

Proof. See Appendix A.1.2. ■

Note that ($\star$) only restricts the marginal distribution of $(Y_d, \boldsymbol{T})$ under Q and, in particular, does not involve the joint distribution of potential outcomes. Indeed, any restrictions imposed by $\mathbf{S}$ on the dependence structure of $(Y_d)_{d\in\mathcal{D}}$ must be enforced through the restriction $\boldsymbol{\pi} \in \mathcal{R}_{\boldsymbol{\pi}}$, which restricts the distribution of $\boldsymbol{T}$ and plays the role of a "mutual compatibility" requirement. As formalized in the next subsection, we will therefore be able to evaluate the compatibility of the observed data with $\mathbf{S}$ by first studying the possible values of treatment-specific outcome distributions within each response-type, and then examining what (if anything) remains possible once we additionally require that these treatment-specific outcome distributions can each be coupled with a *common* $\boldsymbol{T}$ distribution that satisfies the second requirement.

In the remainder, it will prove convenient to generally represent the relevant outcome distributions by their densities with respect to the common dominating measure μ. Throughout

the remainder, denote by $p^P_{d|z}$ the sub-density of $P_{d,z}$, i.e. $p^P_{d|z} \geq 0$ and, for any $A \in \mathscr{B}_\mathcal{Y}$,

$$P_{d|z}(A) = \int_A p^P_{d|z}(y)\, d\mu(y) \ .$$

We will usually suppress the dependence on P and just write $p_{d|z}$. Note that $p_{d|z}$ is not generally a density because $\int_\mathcal{Y} p_{d|z}(y)\, d\mu(y) = P(D = d \,|\, Z = z)$, which can be strictly less than 1. When $P(D = d \,|\, Z = z) > 0$, we have $p_{d|z}(y) = P(D = d \,|\, Z = z)\, f_{Y|D=d,Z=z}(y)$ where $f_{Y|D=d,Z=z}$ is the density of the conditional distribution of Y given $\{D = d, Z = z\}$. Define $\boldsymbol{p}_d : \mathcal{Y} \to \mathbb{R}^\mathcal{Z}_{\geq 0}$ as the vector-valued function obtained by stacking $p_{d|z}$ over $z \in \mathcal{Z}$,

$$\boldsymbol{p}_d(y) = \boldsymbol{p}^P_d(y) := \left(p^P_{d|z}(y) : z \in \mathcal{Z}\right) \ .$$

For each parent distribution $Q \in \mathcal{M}$ and $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$, define $q_{d|\boldsymbol{\tau}} = q^Q_{d|\boldsymbol{\tau}}$ as the sub-density of the sub-measure $Q_{d,\boldsymbol{\tau}}$, defined as $Q_{d,\boldsymbol{\tau}}(A) := Q(Y_d \in A, \boldsymbol{T} = \boldsymbol{\tau})$ for all $A \in \mathscr{B}_\mathcal{Y}$. Define $\boldsymbol{q}_d : \mathcal{Y} \to \mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\geq 0}$ as the vector-valued function obtained by stacking $q_{d|\boldsymbol{\tau}}$ over $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$,

$$\boldsymbol{q}_d(y) = \boldsymbol{q}^Q_d(y) := \left(q^Q_{d|\boldsymbol{\tau}}(y) : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}\right) \ .$$

Note that $\int_\mathcal{Y} \boldsymbol{q}^Q_d(y)\, d\mu(y) = \boldsymbol{\pi}^Q$.

Throughout, all pointwise statements involving $y \in \mathcal{Y}$ should be interpreted as holding μ-almost-everywhere. When I say "all" or "each" $y \in \mathcal{Y}$, I mean "μ-almost-all", and when I say that a property holds for "some $y \in \mathcal{Y}$", or that "there exists $y \in \mathcal{Y}$" with some property, I mean that the set of $y \in \mathcal{Y}$ with the property has positive μ measure. Similarly, all results and (in)equalities involving measurable functions over $\mathcal{Y}$ should be understood as holding only up to a zero μ-measure subset of $\mathcal{Y}$.

Now, we can rewrite ($\star$) in terms of densities to get that $\mathbf{R}_d \boldsymbol{q}_d(y) = \boldsymbol{p}_d(y)$, for all treatments d and outcome values y. However, once $\boldsymbol{\pi} \in \mathcal{R}_{\boldsymbol{\pi}}$ is imposed, it suffices to impose the relaxed requirement $\mathbf{R}_d \boldsymbol{q}_d(y) \leq \boldsymbol{p}_d(y)$. To see this, first note that $\boldsymbol{\pi} \in \mathcal{R}_{\boldsymbol{\pi}}$ implies that $\sum_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}} \boldsymbol{\pi}_{\boldsymbol{\tau}} = 1$, and so, we must have

$$\sum_{d \in \mathcal{D}} (\mathbf{R}_d \boldsymbol{\pi})_{(z)} = \sum_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}} \boldsymbol{\pi}_{\boldsymbol{\tau}} = 1 = \sum_{d \in \mathcal{D}} P(D = d \,|\, Z = z) = \sum_{d \in \mathcal{D}} \int_\mathcal{Y} p_{d|z}(y)\, d\mu(y) \ ,$$

which implies

$$\sum_{d \in \mathcal{D}} \int_\mathcal{Y} \mathbf{R}_d \boldsymbol{q}_d(y) - \boldsymbol{p}_d(y)\, d\mu(y) = \sum_{d \in \mathcal{D}} \left(\mathbf{R}_d \boldsymbol{\pi} - \int_\mathcal{Y} \boldsymbol{p}_d(y)\, d\mu(y) \right) = 0 \ ,$$

and so, $\mathbf{R}_d \boldsymbol{q}_d(y) - \boldsymbol{p}_d(y) \leq 0$ for all $d \in \mathcal{D}, y \in \mathcal{Y}$ must hold with equality in order to be compatible with $\boldsymbol{\pi} \in \mathcal{R}_{\boldsymbol{\pi}}$. Therefore, we have now obtained the following characterization of the identified set for Q under $\mathbf{S}$,

$$\mathcal{Q}_\mathbf{S}(P) = \left\{ Q \in \mathcal{M} \;\middle|\; \begin{array}{l} Q_Z = P_Z \ , \ \boldsymbol{\pi}^Q \in \mathcal{R}_{\boldsymbol{\pi}} \ , \\ \mathbf{R}_d \boldsymbol{q}^Q_d(y) \leq \boldsymbol{p}^P_d(y) \ \forall d \in \mathcal{D}, y \in \mathcal{Y} \ . \end{array} \right\} \ . \tag{3.1}$$

This characterization is useful because $\mathbf{R}_d \boldsymbol{q}_d(y) \leq \boldsymbol{p}_d(y)$ always admits a nonnegative solution (i.e. $\boldsymbol{q}_d(y) = \mathbf{0}$), whereas the unrelaxed system $\mathbf{R}_d \boldsymbol{q}_d(y) = \boldsymbol{p}_d(y)$ only admits a solution when $\boldsymbol{p}_d(y) \in \operatorname{cone}(\mathbf{R}_d)$. In the next subsection, we exploit this property, along with $\mathbf{R}_d$ being invariant across y, to reduce the continuum of treatment-specific restrictions imposed by (3.1) to a finite-dimensional feasibility problem.

Remark 3.1 (Sample Selection)**.** For any $d \in \mathcal{D}$, taking $A = \mathcal{Y}$ in ($\star$), we get

$$\mathbf{R}_d \boldsymbol{\pi} = \left(P(D = d \mid Z = z) : z \in \mathcal{Z} \right), \tag{3.2}$$

which is the propensity score equality. If Y is observed for all treatment states, as we generally assume in this paper, specifically requiring (3.2) is redundant given ($\star$). In some settings, however, (Y, D, Z) may only be observed conditional on $D \in \mathcal{D}^\dagger$, where $\mathcal{D}^\dagger \subsetneq \mathcal{D}$, with only (D, Z) being observed when $D \in \mathcal{D} \setminus \mathcal{D}^\dagger$ (i.e. there is sample selection). In such cases, Lemma 3.1 continues to hold if, for each $d \in \mathcal{D} \setminus \mathcal{D}^\dagger$, the requirement that ($\star$) must hold for each $A \subseteq \mathcal{Y}$ is replaced with the (much simpler) requirement that (3.2) must hold. With this modification in place, the approach of this paper extends to settings with sample selection. For an example of such an application, see Kroft, Mourifié, and Vayalinkal (2025). ♦

3.2. **Identified Set for Response-Type Probabilities.** For $d \in \mathcal{D}$ and $y \in \mathcal{Y}$, let $\operatorname{Fibre}_d(\boldsymbol{p}_d(y))$ denote the set of non-negative solutions to $\mathbf{R}_d \boldsymbol{q}_d(y) \leq \boldsymbol{p}_d(y)$,

$$\operatorname{Fibre}_d(\boldsymbol{p}_d(y)) := \left\{ \boldsymbol{x} \in \mathbb{R}^{\mathcal{T}}_{\geq 0} \;\middle|\; \mathbf{R}_d \boldsymbol{x} \leq \boldsymbol{p}_d(y) \right\},$$

which is a closed convex polytope. Defining $\operatorname{Fibre}_d(\boldsymbol{p}_d(y))$ in terms of the relaxed $\mathbf{R}_d \boldsymbol{q}_d(y) \leq \boldsymbol{p}_d(y)$ system, instead of $\mathbf{R}_d \boldsymbol{q}_d(y) = \boldsymbol{p}_d(y)$, ensures that it is always nonempty (since $\boldsymbol{p}_d \geq 0$ implies that $\boldsymbol{x} = 0$ is a solution). , but is unbounded whenever the model does not eliminate all "d-never-taker" types (i.e. individuals that don't take treatment d when faced with any $z \in \mathcal{Z}$).

The support function of $\operatorname{Fibre}_d(\boldsymbol{p}_d(y))$, which we will also view as a function of $\boldsymbol{p}_d(y)$, will play a central role in our remaining analysis. In the remainder, denote this support function by the shorthand $h_d(\,\cdot\,; \boldsymbol{p}_d(y))$. Precisely, for every $\boldsymbol{c} \in \mathbb{R}^{\mathcal{T}}$, define

$$\begin{aligned} h_d(\boldsymbol{c}; \boldsymbol{p}_d(y)) &:= h_{\operatorname{Fibre}_d(\boldsymbol{p}_d(y))}(\boldsymbol{c}) \\ &= \sup_{\boldsymbol{x} \geq 0 : \mathbf{R}_d \boldsymbol{x} \leq \boldsymbol{p}_d(y)} \boldsymbol{c}^\intercal \boldsymbol{x}, \end{aligned}$$

and note that the right hand side above is a linear program.

Note that $h_d(\boldsymbol{c}; \boldsymbol{p}_d(y)) = \infty$ iff there exists $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$ such that $\boldsymbol{\tau}^{-1}(d) = \emptyset$ (i.e. $\boldsymbol{\tau}$ is a "d-Never-Taker") and $\boldsymbol{c}_{(\boldsymbol{\tau})} > 0$. In other words, for each $y \in \mathcal{Y}$, $\boldsymbol{c} \mapsto h_d(\boldsymbol{c}; \boldsymbol{p}_d(y))$ is a bounded function over directions that only put positive weight on $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d$ (i.e. the non-"d-Never-Taker" types). For later use, note that, by duality of linear programs, we now have that for any

$\boldsymbol{c} \in \mathbb{R}^{\mathcal{T}}$ with $\boldsymbol{c}_{(\boldsymbol{\tau})} \leq 0 \; \forall \boldsymbol{\tau} \in \mathcal{T} \setminus \mathcal{T}_d$,

$$h_d\left(\boldsymbol{c}; \boldsymbol{p}_d\left(y\right)\right) = \min_{\boldsymbol{\ell} \geq 0 : \mathbf{R}_d^{\intercal} \boldsymbol{\ell} \geq \boldsymbol{c}} \boldsymbol{\ell}^{\intercal} \boldsymbol{p}_d\left(y\right) \; . \tag{3.3}$$

Now, since every $Q \in \mathcal{Q}_{\mathbf{S}}$ must satisfy $\mathbf{R}_d \boldsymbol{q}_d^Q\left(y\right) \leq \boldsymbol{p}_d\left(y\right)$ for all $y \in \mathcal{Y}$ and $\int_{\mathcal{Y}} \boldsymbol{q}_d\left(y\right) \, d\mu(y) = \boldsymbol{\pi}$, we can define a set $\boldsymbol{\Pi}_d$ of preliminary "candidates" for $\boldsymbol{\pi}$, as follows,

$$\boldsymbol{\Pi}_d := \left\{ \int_{\mathcal{Y}} \boldsymbol{x}(y) \, d\mu\left(y\right) \;\middle|\; \begin{array}{l} \text{measurable } \boldsymbol{x} : \mathcal{Y} \to \mathbb{R}_{\geq 0}^{\mathcal{T}} \text{ s.t.} \\ \boldsymbol{x}\left(y\right) \in \mathrm{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right) \; \forall y \in \mathcal{Y} \; . \end{array} \right\} \; .$$

Since this argument holds for every $d \in \mathcal{D}$, we must have that $\boldsymbol{\pi} \in \bigcap_{d \in \mathcal{D}} \boldsymbol{\Pi}_d$. Note also that this intersection is always nonempty, since each $\boldsymbol{\Pi}_d$ contains the $\mathbf{0}$ vector. Once intersected with $\mathcal{R}_{\boldsymbol{\pi}}$, however, such frivolous candidates are removed and any additional restrictions required by $\mathbf{S}$ are imposed, yielding the following result that shows this is sharp.

Theorem 3.1. *Suppose P satisfies Condition D, and let the TSM $\mathbf{S}$ be arbitrary. Then,*

$$\boldsymbol{\Pi} = \left(\bigcap_{d \in \mathcal{D}} \boldsymbol{\Pi}_d \right) \cap \mathcal{R}_{\boldsymbol{\pi}} \; . \tag{3.4}$$

In particular, $\mathcal{Q}_{\mathbf{S}} \neq \emptyset$ if and only if $\left(\bigcap_{d \in \mathcal{D}} \boldsymbol{\Pi}_d\right) \cap \mathcal{R}_{\boldsymbol{\pi}} \neq \emptyset$.

Further, if $\mathbf{S}$ satisfies Condition C, $\boldsymbol{\Pi}$ is a compact convex set, and if $\mathbf{S}$ satisfies Condition CP, $\boldsymbol{\Pi}$ is a compact convex polytope.

Proof. See Appendix A.1.5 ■

As suggested by our discussion in the previous subsection, this result formalizes the idea that agreement of treatment-specific allocations on a common $\mathbf{S}$-compatible $\boldsymbol{\pi}$, is sufficient to obtain a compatible parent distribution Q with $\boldsymbol{\pi}^Q = \boldsymbol{\pi}$. The proof of this result further makes it clear that, as might be expected, we can take this compatible Q to also have potential outcome distributions that agree with aforementioned treatment-specific allocations, and we will exploit this fact in the next section to obtain bounds on mean outcomes and treatment effects.

Although Equation (3.4) reduces compatibility to the non-emptiness of finitely many finite-dimensional sets, these sets are still defined in terms of infinite dimensional objects satisfying a continuum of restrictions. The key feature that enables us to operationalize this characterization is that the pointwise linear restrictions $\mathbf{R}_d \boldsymbol{x} \leq \boldsymbol{p}_d\left(y\right)$ all share the same coefficient matrix $\mathbf{R}_d$ across the continuum of y-values, with only the observed sub-density vector $\boldsymbol{p}_d$ varying with y.

To illustrate why this feature is so powerful, it is useful to first consider the following equality version of Fibre_d,

$$\mathrm{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)^{=} := \left\{ \boldsymbol{x} \in \mathbb{R}_{\geq 0}^{\mathcal{T}} \;\middle|\; \mathbf{R}_d \boldsymbol{x} = \boldsymbol{p}_d\left(y\right) \right\} \; .$$

We can view each $\mathrm{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)^{=}$ as the feasible region of a linear program, with $\mathrm{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)^{=} \neq \emptyset$ if and only if $\boldsymbol{p}_d\left(y\right) \in \mathrm{cone}\,\mathbf{R}_d$, where $\mathrm{cone}\,\mathbf{R}_d$ is the set, a convex cone, generated by the nonnegative linear combinations of the columns of $\mathbf{R}_d$. To make this concrete, let $\boldsymbol{c} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ be a vector of weights on response-types, and consider the linear program (LP)

$$\max\left\{\boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} \,\middle|\, \boldsymbol{x} \in \mathrm{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)^{=}\right\} = \max\left\{\boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} \,\middle|\, \mathbf{R}_d\boldsymbol{x} = \boldsymbol{p}_d\left(y\right),\ \boldsymbol{x} \geq 0\right\}\ . \tag{3.5}$$

For simplicity, first assume that $\boldsymbol{p}_d\left(y\right) \in \mathrm{cone}\,\mathbf{R}_d\ \forall y \in \mathcal{Y}$ so that the LP above is feasible. In our setting, since $\mathbf{R}_d$ is "fixed", while $\boldsymbol{p}_d$ is "moving" (with y), a natural question to ask is what happens to the LP as we move over $y \in \mathcal{Y}$, causing $\boldsymbol{p}_d$ to move over $\mathrm{cone}\,\mathbf{R}_d$? As it turns out, the answer is that $\mathrm{cone}\,\mathbf{R}_d$ admits a finite partition into "chambers" with the following property: for any fixed vector of weights[9] $\boldsymbol{c}$, the optimal value is given by the same linear function of $\boldsymbol{p}_d\left(y\right)$ as long as $\boldsymbol{p}_d\left(y\right)$ stays in the same chamber. This partition depends only on $\mathbf{R}_d$ and works simultaneously for every such $\boldsymbol{c}$, although the linear function depends on $\boldsymbol{c}$.

This linearity allows us to interchange integration and optimization within each chamber. Precisely, if $B \subseteq \mathcal{Y}$ is a region such that $\left\{\boldsymbol{p}_d\left(y\right) : y \in B\right\}$ is a subset of a single chamber, then

$$\int_B \max_{\substack{\boldsymbol{x}\geq 0 \\ \mathbf{R}_d\boldsymbol{x}=\boldsymbol{p}_d(y)}} \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x}\, d\mu(y) = \max_{\substack{\boldsymbol{v}\geq 0 \\ \mathbf{R}_d\boldsymbol{v}=\int_B \boldsymbol{p}_d(y)\, d\mu(y)}} \boldsymbol{c}^{\mathsf{T}}\boldsymbol{v}.$$

We can use nonnegative slack variables to convert our inequality fibres $\mathrm{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)$ into equality fibres of the form just discussed. In particular, $\boldsymbol{\Pi}_d$ can be decomposed into a finite Minkowski sum, where each component of this sum is formed by "integrating" polyhedrons with the same "shape", in the sense that linear programs defined over such polyhedrons produce the same optimal "basis" for every objective function.[10] The same structure underlies Algorithm 3.1, which extends solutions across outcome values as needed, without first enumerating the collection of possible chambers.

Therefore, in order to operationalize this result, we will work with the support function of $\boldsymbol{\Pi}_d$, which can be evaluated through integrals of the form we just discussed. In order to state the most powerful version of that result, I first state the following result, which shows that the dualization (3.3) allows integrals of $h_d\left(\boldsymbol{c}; \boldsymbol{p}_d\left(y\right)\right)$ to be cast as moments.

Lemma 3.2. *Suppose P satisfies Condition D. Then, for $d \in \mathcal{D}$ and integrable $\boldsymbol{c} : \mathcal{Y} \to \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ with $\boldsymbol{c}_{(\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d)} \leq \mathbf{0}$, there exists integrable $\boldsymbol{\ell}_d^{\star}\left(\boldsymbol{c}\left(\cdot\right); \boldsymbol{p}_d^P\left(\cdot\right)\right) : \mathcal{Y} \to \mathbb{R}^{\mathcal{Z}}_{\geq 0}$ with*

$$\boldsymbol{\ell}_d^{\star}\left(\boldsymbol{c}\left(y\right); \boldsymbol{p}_d\left(y\right)\right) \in \operatorname*{argmin}_{\substack{\boldsymbol{\ell}\geq 0 : \mathbf{R}_d^{\mathsf{T}}\boldsymbol{\ell} \geq \boldsymbol{c}(y), \\ \|\boldsymbol{\ell}\|_\infty \leq \|\boldsymbol{c}(y)\|_\infty}} \boldsymbol{\ell}^{\mathsf{T}}\boldsymbol{p}_d\left(y\right)\ ,$$

[9] We need $\boldsymbol{c}$ to be such that the maximum is finite; i.e. no positive weight on d-never-takers.

[10] Formally, these polyhedrons are considered "normally equivalent". See the proof of Lemma A.3 for a more careful discussion of this idea.

and

$$\int_{\mathcal{Y}} h_d\left(\boldsymbol{c}\left(y\right);\boldsymbol{p}_d\left(y\right)\right)d\mu\left(y\right) = \int_{\mathcal{Y}} \boldsymbol{\ell}_d^{\star}\left(\boldsymbol{c};\boldsymbol{p}_d\left(y\right)\right)^{\intercal}\boldsymbol{p}_d^P\left(y\right)\,d\mu\left(y\right)$$
$$= \sum_{z\in\mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_d^P\left(\boldsymbol{c}\left(Y\right);W\right)\middle|\, Z=z\right]\ .$$

where $\boldsymbol{m}_d^P\left(\ \cdot\ ;y,d',z\right) := \boldsymbol{\ell}_{d,(z)}^{\star}\left(\ \cdot\ ;\boldsymbol{p}_d^P\left(y\right)\right)\mathbb{1}\left\{d'=d\right\}$.

Proof. See Appendix A.1.3. ■

In this section and the next, given $\boldsymbol{c}\in\mathbb{R}^{\boldsymbol{\mathcal{T}}}$ with $\boldsymbol{c}_{(\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d)}\leq\mathbf{0}$, choose an arbitrary measurable $\boldsymbol{\ell}_d^{\star}\left(\boldsymbol{c};\boldsymbol{p}_d\left(y\right)\right)$ as in Lemma 3.2, and let $\boldsymbol{m}_d^P\left(\boldsymbol{c};y,d',z\right) := \boldsymbol{\ell}_{d,(z)}^{\star}\left(\boldsymbol{c};\boldsymbol{p}_d^P\left(y\right)\right)\mathbb{1}\left\{d'=d\right\}$.

Throughout, we will use the following shorthand, for any matrix A, with rows $\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k\}$, define the vectors $h_d\left(A;\boldsymbol{p}_d\left(y\right)\right)$ and $\boldsymbol{m}_d\left(A;W\right)$ as

$$h_d\left(A;\boldsymbol{p}_d\left(y\right)\right) := \left(h_d\left(\boldsymbol{a}_j;\boldsymbol{p}_d\left(y\right)\right)\right)_{j=1}^k \text{ and } \boldsymbol{m}_d\left(A;W\right) := \left(\boldsymbol{m}_d\left(\boldsymbol{a}_j;W\right)\right)_{j=1}^k\ .$$

We now have the following result which states that the support function of $\boldsymbol{\Pi}_d$ can be obtained by integrating the pointwise support function $h_d\left(\boldsymbol{c};\boldsymbol{p}_d\left(y\right)\right)$, and therefore can be expressed in terms of the moments we have just defined.

Lemma 3.3. *Suppose P satisfies Condition D. Then, for $d\in\mathcal{D}$ and $\boldsymbol{c}\in\mathbb{R}^{\boldsymbol{\mathcal{T}}}$ with $\boldsymbol{c}_{(\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d)}\leq\mathbf{0}$,*

$$h_{\boldsymbol{\Pi}_d}\left(\boldsymbol{c}\right) = \int_{\mathcal{Y}} h_d\left(\boldsymbol{c};\boldsymbol{p}_d\left(y\right)\right)d\mu\left(y\right) = \sum_{z\in\mathcal{Z}}\mathbb{E}\left[\boldsymbol{m}_d\left(\boldsymbol{c};W\right)\middle|\, Z=z\right]\ ,$$

Further, there exists a matrix $\mathbf{A}_d\in\mathbb{R}^{k_d\times\boldsymbol{\mathcal{T}}}$, depending only on $\mathbf{R}_d$ (not on P), with rows $\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_{k_d}\}$ such that $\boldsymbol{a}_{j,(\boldsymbol{\tau})}=0\ \forall\boldsymbol{\tau}\in\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d, j\in\{1,\ldots,k_d\}$ and

$$\boldsymbol{\Pi}_d = \left\{\boldsymbol{x}\geq 0\middle|\,\mathbf{A}_d\boldsymbol{x}\leq\int_{\mathcal{Y}} h_d\left(\mathbf{A}_d;\boldsymbol{p}_d\left(y\right)\right)d\mu\left(y\right)\right\}$$
$$= \left\{\boldsymbol{x}\geq 0\middle|\,\mathbf{A}_d\boldsymbol{x}\leq\sum_{z\in\mathcal{Z}}\mathbb{E}\left[\boldsymbol{m}_d\left(\mathbf{A}_d;W\right)\middle|\, Z=z\right]\right\}\ ,$$

Proof. See Appendix A.1.4. ■

With this lemma and Theorem 3.1 in hand, we immediately have the following result, which characterizes the sharp testable implications of TSMs that satisfy Condition CP, including the leading case where $\mathbf{S}$ consists solely of support restrictions on $\boldsymbol{T}$.

Theorem 3.2 (Sharp Testable Implications under Condition CP)**.** *Suppose P satisfies Condition D, and $\mathbf{S}$ satisfies Condition CP. Then, $\boldsymbol{\Pi}$ is a convex polytope, with*

$$\boldsymbol{\Pi} = \left\{\boldsymbol{x}\geq 0\middle|\,\begin{bmatrix}\mathbf{A}_{\mathbf{S}}\\ \mathbf{A}_{\boldsymbol{\Pi}}\end{bmatrix}\boldsymbol{x}\leq\begin{bmatrix}\mathbf{a}_{\mathbf{S}}\\ \sum_{z\in\mathcal{Z}}\mathbb{E}\left[\boldsymbol{m}_{\boldsymbol{\Pi}}\left(W\right)\middle|\, Z=z\right]\end{bmatrix}\right\}\ ,$$

where $\mathbf{A}_{\mathbf{\Pi}} := \begin{bmatrix} \mathbf{A}_{d_1} \\ \vdots \\ \mathbf{A}_{d_{|\mathcal{D}|}} \end{bmatrix}$ *and* $\boldsymbol{m}_{\mathbf{\Pi}}(W) := \begin{bmatrix} \boldsymbol{m}_{d_1}(\mathbf{A}_{d_1}; W) \\ \vdots \\ \boldsymbol{m}_{d_{|\mathcal{D}|}}(\mathbf{A}_{d_{|\mathcal{D}|}}; W) \end{bmatrix}$.

Further, for any $\boldsymbol{c} \in \mathbb{R}^{\mathcal{T}}$, $\Lambda(\boldsymbol{c}) := \left\{(\boldsymbol{\lambda}_{\mathbf{S}}, \boldsymbol{\lambda}_{\mathbf{\Pi}}) \geq 0 \,\middle|\, \mathbf{A}_{\mathbf{S}}^{\intercal}\boldsymbol{\lambda}_{\mathbf{S}} + \mathbf{A}_{\mathbf{\Pi}}^{\intercal}\boldsymbol{\lambda}_{\mathbf{\Pi}} \geq \boldsymbol{c}\right\} \neq \emptyset$ *and*

$$h_{\mathbf{\Pi}}(\boldsymbol{c}) = \inf_{(\boldsymbol{\lambda}_{\mathbf{S}}, \boldsymbol{\lambda}_{\mathbf{\Pi}}) \in \Lambda(\boldsymbol{c})} \left\{ \boldsymbol{\lambda}_{\mathbf{S}}^{\intercal}\mathbf{a}_{\mathbf{S}} + \sum_{z \in \mathcal{Z}} \mathbb{E}\left[\boldsymbol{\lambda}_{\mathbf{\Pi}}^{\intercal}\boldsymbol{m}_{\mathbf{\Pi}}(W) \,\middle|\, Z = z\right] \right\},$$

which is a linear program.[11] *In particular,*

$$\mathcal{Q}_{\mathbf{S}} \neq \emptyset \iff \boldsymbol{\lambda}_{\mathbf{S}}^{\intercal}\mathbf{a}_{\mathbf{S}} + \sum_{z \in \mathcal{Z}} \mathbb{E}\left[\boldsymbol{\lambda}_{\mathbf{\Pi}}^{\intercal}\boldsymbol{m}_{\mathbf{\Pi}}(W) \,\middle|\, Z = z\right] \geq 0 \quad \forall (\boldsymbol{\lambda}_{\mathbf{S}}, \boldsymbol{\lambda}_{\mathbf{\Pi}}) \in \Lambda(\mathbf{0}) \ .$$

Proof. See Appendix A.1.6. ■

In order to operationalize Theorem 3.1 for TSMs that only satisfy the weaker condition Condition C, we will need the following general characterization of the support function of convex intersections. More importantly, this result is of practical relevance, since the matrix $\mathbf{A}_d$ is not generally easy to compute, whereas integrals of h_d can be easily computed via Algorithm 3.1.

Lemma 3.4 (Infimal Convolution Representation of Support Functions of Intersections)**.** *Let* $K_1, \ldots, K_N$ *be nonempty closed convex subsets of* $\mathbb{R}^k$. *First, if* $\bigcap_{i=1}^N K_i$ *is compact and nonempty, then, for all* $\boldsymbol{c} \in \mathbb{R}^k$,

$$h_{\bigcap_{i=1}^N K_i}(\boldsymbol{c}) = \inf_{\{\boldsymbol{c}_i\}_{i=1}^N : \sum_{i=1}^N \boldsymbol{c}_i = \boldsymbol{c}} \left\{ \sum_{i=1}^N h_{K_i}(\boldsymbol{c}_i) \right\},$$

and, in particular, the right hand side equals 0 *when* $\boldsymbol{c} = \mathbf{0}$.

Second, if either (i) at least one K_i *is compact or (ii) every* K_i *is polyhedral, then*

$$\bigcap_{i=1}^N K_i \neq \emptyset \iff \sum_{i=1}^N h_{K_i}(\boldsymbol{c}_i) \geq 0 \quad \forall \boldsymbol{c}_1, \ldots, \boldsymbol{c}_N \in \mathbb{R}^k : \sum_{i=1}^N \boldsymbol{c}_i = \mathbf{0} \ .$$

Proof. The first claim follows from a suitable application of the Fenchel-Moreau theorem. The second claim follows from strong separation of $\prod_{i=1}^N K_i$ and the diagonal in the first case, and from Farkas' lemma in the second case. See Appendix A.1.7. ■

Define $\mathcal{C}_d := \left\{\boldsymbol{c} \in \mathbb{R}^{\mathcal{T}} \,\middle|\, \boldsymbol{c}_{(\mathcal{T} \setminus \mathcal{T}_d)} = \mathbf{0}\right\}$, let $\mathcal{C} := \prod_{d \in \mathcal{D}} \mathcal{C}_d$, and, for $\boldsymbol{u} \in \mathcal{C}$, use the shorthand $\boldsymbol{u} := (\boldsymbol{u}_d)_{d \in \mathcal{D}}$ to split it into its $\mathcal{C}_d$ components.

[11] Recalling that $h_{\emptyset} := -\infty$, this expression for $h_{\mathbf{\Pi}}$ holds even when $\mathbf{\Pi} = \emptyset$, but is finite if and only if $\mathbf{\Pi} \neq \emptyset$.

Theorem 3.3 (Sharp Testable Implications under Condition C)**.** *Suppose P satisfies Condition D, and $\mathbf{S}$ satisfies Condition C. Then, for any direction $\boldsymbol{c} \in \mathbb{R}^{\mathcal{T}}$,*

$$h_{\mathbf{\Pi}}(\boldsymbol{c}) = \inf_{\boldsymbol{u} \in \mathcal{C}} \left\{ \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} h_d(\boldsymbol{u}_d; \boldsymbol{p}_d(y))\, d\mu(y) + h_{\mathcal{R}_\pi}\left(\boldsymbol{c} - \sum_{d \in \mathcal{D}} \boldsymbol{u}_d \right) \right\}$$
$$= \inf_{\boldsymbol{u} \in \mathcal{C}} \left\{ \sum_{d \in \mathcal{D}} \sum_{z \in \mathcal{Z}} \mathbb{E}\left[\boldsymbol{m}_d(\boldsymbol{u}_d; W) \mid Z = z \right] + h_{\mathcal{R}_\pi}\left(\boldsymbol{c} - \sum_{d \in \mathcal{D}} \boldsymbol{u}_d \right) \right\},$$

which is a convex program. In particular, $\mathbf{S}$ has the following sharp testable implications:

$$\mathcal{Q}_{\mathbf{S}} \neq \emptyset \iff \sum_{d \in \mathcal{D}} \sum_{z \in \mathcal{Z}} \mathbb{E}\left[\boldsymbol{m}_d(\boldsymbol{u}_d; W) \mid Z = z \right] + h_{\mathcal{R}_\pi}\left(-\sum_{d \in \mathcal{D}} \boldsymbol{u}_d \right) \geq 0 \quad \forall \boldsymbol{u} \in \mathcal{C}\,.$$

Proof. See Appendix A.1.8. ■

The second display above says that $\mathcal{Q}_{\mathbf{S}} = \emptyset$ if and only if there exists $(\boldsymbol{u}_d)_{d \in \mathcal{D}}$ such that

$$\sum_{d \in \mathcal{D}} \sum_{z \in \mathcal{Z}} \mathbb{E}\left[\boldsymbol{m}_d(\boldsymbol{u}_d; W) \mid Z = z \right] < -h_{\mathcal{R}_\pi}\left(-\sum_{d \in \mathcal{D}} \boldsymbol{u}_d \right) = \inf_{\boldsymbol{\pi} \in \mathcal{R}_\pi} \left(\sum_{d \in \mathcal{D}} \boldsymbol{u}_d \right)^{\intercal} \boldsymbol{\pi}\,,$$

which is equivalent, by Lemma 3.4, to the existence of a $\boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}$ such that

$$\sup_{\boldsymbol{\pi} \in \bigcap_{d \in \mathcal{D}} \mathbf{\Pi}_d} \boldsymbol{v}^{\intercal} \boldsymbol{\pi} < \inf_{\boldsymbol{\pi} \in \mathcal{R}_\pi} \boldsymbol{v}^{\intercal} \boldsymbol{\pi}$$

where we use that $\sup_{\boldsymbol{\pi} \in \bigcap_{d \in \mathcal{D}} \mathbf{\Pi}_d} \boldsymbol{v}^{\intercal} \boldsymbol{\pi} = \inf_{(\boldsymbol{v}_d)_d : \sum_d \boldsymbol{v}_d = \boldsymbol{v}} \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} h_d(\boldsymbol{v}_d; \boldsymbol{p}_d(y)) d\mu(y)$.

3.3. **Computation of $\int h_d$.** In order to represent $\mathbf{R}_d \boldsymbol{q}_d \leq \boldsymbol{p}_d$ as an equality system, we will use the augmented matrix $\mathbf{R} := [\mathbf{R}_d \;\; \mathbf{I}_{\mathcal{Z}}]$, where $\mathbf{I}_{\mathcal{Z}}$ is the conformable identity matrix and represents (non-negative) slack variables.

For any suitable subset B of columns of $\mathbf{R}$, denote by $\mathbf{R}_B$ the submatrix of $\mathbf{R}$ consisting of those columns, and denote by $\mathbf{R}_{\neg B}$ the submatrix consisting of only those columns *not* in B. Since columns in $\mathbf{R}_d$ have a one-to-one correspondence with response-types, we can view B as a subset of response-types/slacks. For any vector $\boldsymbol{c} \in \mathbb{R}^{\mathcal{T}}$ we will also use, in an abuse of notation, $\boldsymbol{c}$ to represent the version of $\boldsymbol{c}$ mapped into $\mathbb{R}^{\mathcal{T}} \times \mathbb{R}^{\mathcal{Z}}$ by setting the additional coordinates (slacks) to zero and define $\boldsymbol{c}_B$ as the subvector consisting of just those components that correspond to response-types/slacks in B, and define $\boldsymbol{c}_{\neg B}$ as the subvector consisting of components corresponding to response-types/slacks not in B.

Given the output of Algorithm 3.1, we can construct the value function as

$$h_d^{\text{alg}}(\boldsymbol{c}(\cdot); \boldsymbol{p}_d(\cdot)) : y \mapsto \sum_{\tilde{y} \in \mathcal{G}} \mathbb{1}\{y \in O(\tilde{y})\} \boldsymbol{c}_{B(\tilde{y})}(y)^{\intercal} \mathbf{R}_{B(\tilde{y})}^{-1} \boldsymbol{p}_d(y)\,,$$

Algorithm 3.1 (Computation of h_d)**.**

Input: Model and Data: matrix $\mathbf{R} := [\mathbf{R}_d \;\; \mathbf{I}_{\mathcal{Z}}]$ and function $\boldsymbol{p}_d : \mathcal{Y} \to \mathbb{R}^{\mathcal{Z}}_{\geq 0}$.
Parameters: function $\boldsymbol{c} : \mathcal{Y} \to \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ with $\boldsymbol{c}_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} \leq \mathbf{0}$.
Oracle: $\mathrm{LP}_d(\boldsymbol{c}, \boldsymbol{b})$ returns a primal-dual optimal basis for $\max_{(\boldsymbol{x},\boldsymbol{s}) \geq 0 : \mathbf{R}(\boldsymbol{x},\boldsymbol{s}) = \boldsymbol{b}} \boldsymbol{c}^{\intercal} \boldsymbol{x}$.

Output: Solved Points: $\mathcal{G} \subseteq \mathcal{Y}$,
Optimal Basis Map: $B : y \mapsto B(y)$ for $y \in \mathcal{G}$,
Support Sets: $O : y \mapsto O(y)$ for $y \in \mathcal{G}$.

1: Set $\mathcal{G} := \emptyset$ and $O(\mathcal{G}) := \emptyset$
2: **while** $O(\mathcal{G}) \neq \mathcal{Y}$ **do** ▷ Find optimal solution for all $y \in \mathcal{Y}$
3: Choose $\tilde{y} \in \mathcal{Y} \setminus O(\mathcal{G})$ arbitrarily
4: Set $B(\tilde{y}) := \mathrm{LP}_d(\boldsymbol{c}(\tilde{y}), \boldsymbol{p}_d(\tilde{y}))$ ▷ Query LP solver
5: Set $O(\tilde{y})$ to the subset of $\mathcal{Y} \setminus O(\mathcal{G})$ where $B(\tilde{y})$ remains feasible and optimal:

$$O(\tilde{y}) := \left\{ y \in \mathcal{Y} \setminus O(\mathcal{G}) \;\middle|\; \begin{array}{l} \mathbf{R}_{B(\tilde{y})}^{-1} \boldsymbol{p}_d(y) \geq 0 \,, \\ \boldsymbol{c}_{B(\tilde{y})}(y)^{\intercal} \mathbf{R}_{B(\tilde{y})}^{-1} \mathbf{R}_{\neg B(\tilde{y})} - \boldsymbol{c}_{\neg B(\tilde{y})}(y)^{\intercal} \geq 0 \,. \end{array} \right\}$$

6: Update $\mathcal{G} \leftarrow \mathcal{G} \sqcup \{\tilde{y}\}$
7: Update $O(\mathcal{G}) \leftarrow \bigsqcup_{y \in \mathcal{G}} O(y)$
8: **end while**
9: **return** $\left(\mathcal{G}, \{(B(y), O(y))\}_{y \in \mathcal{G}}\right)$

---

the dual solution as

$$\boldsymbol{\ell}_d^{\mathrm{alg}}(\boldsymbol{c}(\cdot); \boldsymbol{p}_d(\cdot)) : y \mapsto \sum_{\tilde{y} \in \mathcal{G}} \mathbb{1}\{y \in O(\tilde{y})\} \mathbf{R}_{B(\tilde{y})}^{-\intercal} \boldsymbol{c}_{B(\tilde{y})}(y) \;\; .$$

For the primal solution, first define, for each $B \in \{B(y)\}_{y \in \mathcal{G}}$, the matrix $\boldsymbol{J}_B$ such that, for any $\boldsymbol{x} \in \mathbb{R}^{\mathcal{Z}}$, $\boldsymbol{J}_B \boldsymbol{x} = \tilde{\boldsymbol{x}} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ such that $\tilde{\boldsymbol{x}}_{(B \cap \boldsymbol{\mathcal{T}})} = \boldsymbol{x}$ and $\tilde{\boldsymbol{x}}_{(\boldsymbol{\mathcal{T}} \setminus B)} = \mathbf{0}$.[12] Then, the primal solution is given by

$$\boldsymbol{q}_d^{\mathrm{alg}}(\boldsymbol{c}(\cdot); \boldsymbol{p}_d(\cdot)) : y \mapsto \sum_{\tilde{y} \in \mathcal{G}} \mathbb{1}\{y \in O(\tilde{y})\} \boldsymbol{J}_{B(\tilde{y})} \mathbf{R}_{B(\tilde{y})}^{-1} \boldsymbol{p}_d(y) \;.$$

The following result shows that these constructions exactly recover their corresponding targets, and that Algorithm 3.1 always terminates in finitely many steps.

[12] Precisely, let $\boldsymbol{J}_B$ consist of the columns indexed by B of $[\boldsymbol{I}_{|\boldsymbol{\mathcal{T}}|} \; \mathbf{0}_{|\boldsymbol{\mathcal{T}}| \times |\mathcal{Z}|}]$, in the same order as in $\mathbf{R}_B$.

Theorem 3.4. *Suppose P satisfies Condition D and let $\mathbf{S}$ be any TSM. Given $d \in \mathcal{D}$ and P-integrable $\boldsymbol{c} : \mathcal{Y} \to \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ with $\boldsymbol{c}_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} \leq \mathbf{0}$, we have that Algorithm 3.1 satisfies the following properties, for any oracle solver $\mathrm{LP}_d(\boldsymbol{c}, \boldsymbol{b})$ that returns a primal-dual optimal basis for the specified LP,*

(i) *Algorithm 3.1 terminates in finitely many steps,*
(ii) *After termination, the resulting $\left(\mathcal{G}, \{(B(y), O(y))\}_{y \in \mathcal{G}}\right)$ satisfies the following :*
   a) *(Integrability) $y \mapsto h_d^{\mathrm{alg}}(\boldsymbol{c}(y); \boldsymbol{p}_d(y))$ is integrable.*
   b) *(Optimality) For all $y \in \mathcal{Y}$,*

$$\begin{aligned} h_d(\boldsymbol{c}(y); \boldsymbol{p}_d(y)) &= h_d^{\mathrm{alg}}(\boldsymbol{c}(y); \boldsymbol{p}_d(y)) \\ &= \boldsymbol{c}(y)^{\intercal} \boldsymbol{q}_d^{\mathrm{alg}}(\boldsymbol{c}(y); \boldsymbol{p}_d(y)) \\ &= \boldsymbol{\ell}_d^{\mathrm{alg}}(\boldsymbol{c}(y); \boldsymbol{p}_d(y))^{\intercal} \boldsymbol{p}_d(y) \ . \end{aligned}$$

   c) *(Feasibility) For all $y \in \mathcal{Y}$,*

$$\mathbf{R}_d \boldsymbol{q}_d^{\mathrm{alg}}(\boldsymbol{c}(y); \boldsymbol{p}_d(y)) \leq \boldsymbol{p}_d(y) \ ,$$

   and

$$\begin{aligned} \boldsymbol{\ell}_d^{\mathrm{alg}}(\boldsymbol{c}(y); \boldsymbol{p}_d(y)) &\geq \mathbf{0} \ , \\ \mathbf{R}_d^{\intercal} \boldsymbol{\ell}_d^{\mathrm{alg}}(\boldsymbol{c}(y); \boldsymbol{p}_d(y)) &\geq \boldsymbol{c}(y) \ , \\ \| \boldsymbol{\ell}_d^{\mathrm{alg}}(\boldsymbol{c}(y); \boldsymbol{p}_d(y)) \|_\infty &\leq \| \boldsymbol{c}(y) \|_\infty \ . \end{aligned}$$

Proof. See Appendix A.1.9 ■

4. Sharp Bounds on Counterfactual Means and Treatment Effects

For any P-integrable $\kappa : \mathcal{Y} \to \mathbb{R}$ and $d \in \mathcal{D}$, denote by $\boldsymbol{\eta}_{d;(\kappa)}$ the vector of response-type-probability-weighted response-type conditional $\kappa(Y_d)$ means,

$$\begin{aligned} \boldsymbol{\eta}_{d,(\kappa)} = \boldsymbol{\eta}_{d,(\kappa)}^{Q} &:= \left(\mathbb{E}\left[\kappa(Y_d) \mathbb{1}\{\boldsymbol{T} = \boldsymbol{\tau}\}\right] : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \right) \\ &= \left(\mathbb{P}(\boldsymbol{T} = \boldsymbol{\tau}) \mathbb{E}\left[\kappa(Y_d) \mid \boldsymbol{T} = \boldsymbol{\tau}\right] : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \right) \ . \end{aligned}$$

To simplify the exposition in the main text, I focus on the special case where θ is a conditional mean. Precisely, in the remainder of this section, I fix $d \in \mathcal{D}$ and a nonempty subset of response-types $A \subseteq \boldsymbol{\mathcal{T}}$, let $\kappa := \kappa_d$ and $\boldsymbol{\eta} := \boldsymbol{\eta}_{d;(\kappa)}$, and assume θ has the form

$$\theta = \frac{\boldsymbol{c}_A^{\intercal} \boldsymbol{\eta}}{\boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}} = \mathbb{E}\left[\kappa_d(Y_d) \,|\, \boldsymbol{T} \in A\right] \ , \tag{4.1}$$

where $\boldsymbol{c}_A \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ indicates which response-types are in A, i.e. $\boldsymbol{c}_{A;(A)} = \mathbf{1}$ and $\boldsymbol{c}_{A;(\boldsymbol{\mathcal{T}} \setminus A)} = \mathbf{0}$; I also assume that A contains no d-Never-Takers and so, $\boldsymbol{c}_{A;(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} = \mathbf{0}$. I also assume that

$\mathcal{K} := \kappa(\mathcal{Y})$ is compact, and define

$$\underline{\kappa} := \min \mathcal{K}\,, \quad \overline{\kappa} := \max \mathcal{K} \quad \text{and} \quad \Theta_0 := [\underline{\kappa}, \overline{\kappa}]\,.$$

All results in this section are stated and proved in full generality in Appendix A.3.

Since θ^Q is only well-defined when $\boldsymbol{c}^\intercal \boldsymbol{\pi}^Q > 0$, I take the identified set to be the a priori range whenever zero conditioning mass is compatible with the model and the data. Precisely, I define the identified set for θ under $\mathbf{S}$, denoted $\Theta_{\mathbf{S}}$, as

$$\Theta_{\mathbf{S}} = \Theta_{\mathbf{S}}(P) := \begin{cases} \emptyset & \text{iff } \mathcal{Q}_{\mathbf{S}}(P) = \emptyset\,, \\ \Theta_0 & \text{if } \inf_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \boldsymbol{c}^\intercal \boldsymbol{\pi} = 0\,, \\ \left\{ \dfrac{\boldsymbol{c}_A^\intercal \boldsymbol{\eta}^Q}{\boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q} \,\middle|\, Q \in \mathcal{Q}_{\mathbf{S}}(P) \right\} & \text{otherwise}\,. \end{cases}$$

4.1. **"Direct" Approach.** In order to proceed, we start by considering the joint identified set for $(\boldsymbol{\eta}, \boldsymbol{\pi})$, $\left\{ \left(\boldsymbol{\eta}^Q, \boldsymbol{\pi}^Q\right) \middle| Q \in \mathcal{Q}_{\mathbf{S}} \right\}$. In the previous section, we obtained $\boldsymbol{\Pi}_d$, the set of "candidates" for $\boldsymbol{\pi}$, by integrating functions that lie in Fibre_d over $\mathcal{Y}$. Noting that $\int_{\mathcal{Y}} \kappa(y)\, \boldsymbol{q}_{d^\star}(y)\, d\mu(y) = \boldsymbol{\eta}$, we can consider a set $\Upsilon_{d^\star}$ of "candidates" for $(\boldsymbol{\eta}, \boldsymbol{\pi})$. Precisely, for $d \in \mathcal{D}$, define

$$\Upsilon_d := \left\{ \left(\int_{\mathcal{Y}} \kappa_d(y)\, \boldsymbol{x}(y)\, d\mu(y), \int_{\mathcal{Y}} \boldsymbol{x}(y)\, d\mu(y) \right) \middle| \begin{array}{l} \text{measurable } \boldsymbol{x} : \mathcal{Y} \to \mathbb{R}^{\mathcal{T}}_{\geq 0} \text{ s.t.} \\ \boldsymbol{x}(y) \in \mathrm{Fibre}_d(\boldsymbol{p}_d(y)) \ \forall y \in \mathcal{Y} \end{array} \right\}.$$

By Theorem 3.1, we now have

$$\left\{ \left(\boldsymbol{\eta}^Q, \boldsymbol{\pi}^Q\right) \middle| Q \in \mathcal{Q}_{\mathbf{S}} \right\} = \Upsilon_{d^\star} \cap \left(\mathbb{R}^{\mathcal{T}} \times \boldsymbol{\Pi}\right)\,, \tag{4.2}$$

which, together with Lemma 3.4, immediately suggests that the sharp upper bound on the *numerator* $\boldsymbol{c}_A^\intercal \boldsymbol{\eta} = \mathbb{P}(\boldsymbol{T} \in A)\, \mathbb{E}[\kappa(Y_{d^\star}) \,|\, \boldsymbol{T} \in A]$ in (4.1) can be obtained by evaluating the support function of $\Upsilon_{d^\star} \cap \left(\mathbb{R}^{\mathcal{T}} \times \boldsymbol{\Pi}\right)$ in the direction $(\boldsymbol{c}, \mathbf{0})$. Precisely, for any $Q \in \mathcal{Q}_{\mathbf{S}}$, we must have the following upper bound and, by (4.2), it is sharp under $\mathbf{S}$:

$$\begin{aligned} Q(\boldsymbol{T} \in A)\, \mathbb{E}_Q[\kappa(Y_{d^\star}) \,|\, \boldsymbol{T} \in A] &\leq \inf_{\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}} \left\{ h_{\Upsilon_{d^\star}}(\boldsymbol{c}_A - \boldsymbol{u}, -\boldsymbol{v}) + h_{\mathbb{R}^{\mathcal{T}} \times \boldsymbol{\Pi}}(\boldsymbol{u}, \boldsymbol{v}) \right\} \\ &= \inf_{\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}} \left\{ h_{\Upsilon_{d^\star}}(\boldsymbol{c}_A - \boldsymbol{u}, -\boldsymbol{v}) + h_{\mathbb{R}^{\mathcal{T}}}(\boldsymbol{u}) + h_{\boldsymbol{\Pi}}(\boldsymbol{v}) \right\} \\ &= \inf_{\boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}} \left\{ h_{\Upsilon_{d^\star}}(\boldsymbol{c}_A, -\boldsymbol{v}) + h_{\boldsymbol{\Pi}}(\boldsymbol{v}) \right\}\,, \end{aligned} \tag{4.3}$$

where the second line follows from the general fact that $h_{K_1 \times K_2}(\boldsymbol{a}_1, \boldsymbol{a}_2) = h_{K_1}(\boldsymbol{a}_1) + h_{K_2}(\boldsymbol{a}_2)$ for nonempty convex K_1, K_2,[13] and the third from $h_{\mathbb{R}^{\mathcal{T}}}(\boldsymbol{u}) = \infty$ for all $\boldsymbol{u} \neq 0$.

Importantly, like we saw with $h_{\boldsymbol{\Pi}_d}$ in the previous section, the support function $h_{\Upsilon_{d^\star}}$ of $\Upsilon_{d^\star}$ also possesses a pointwise fibre integral representation. To see this, first note that, for

[13] For $K_1, K_2 \neq \emptyset$, $\sup_{\boldsymbol{x}_1 \in K_1, \boldsymbol{x}_2 \in K_2} \boldsymbol{c}_1^\intercal \boldsymbol{x}_1 + \boldsymbol{c}_2^\intercal \boldsymbol{x}_2 = \sup_{\boldsymbol{x}_1 \in K_1} \boldsymbol{c}_1^\intercal \boldsymbol{x}_1 + \sup_{\boldsymbol{x}_2 \in K_2} \boldsymbol{c}_2^\intercal \boldsymbol{x}_2$.

$(\boldsymbol{\eta}, \boldsymbol{\pi}) \in \Upsilon_{d^\star}$ and $(\boldsymbol{c}_\eta, \boldsymbol{c}_\pi) \in \mathbb{R}^{\mathcal{T}} \times \mathbb{R}^{\mathcal{T}}$, there exists measurable $\boldsymbol{x} \in \mathrm{Fibre}_{d^\star}(\boldsymbol{p}_{d^\star})$ with

$$(\boldsymbol{c}_\eta, \boldsymbol{c}_\pi)^\intercal (\boldsymbol{\eta}, \boldsymbol{\pi}) = \int_{\mathcal{Y}} \boldsymbol{c}_\eta^\intercal \kappa(y)\, \boldsymbol{x}(y) + \boldsymbol{c}_\pi^\intercal \boldsymbol{x}(y)\, d\mu(y) = \int_{\mathcal{Y}} (\kappa(y)\, \boldsymbol{c}_\eta + \boldsymbol{c}_\pi)^\intercal \boldsymbol{x}(y)\, d\mu(y)\ ,$$

which suggests the following extension of Lemma 3.3.

Lemma 4.1. *Suppose P satisfies Condition D. Then, for $d \in \mathcal{D}$ and $(\boldsymbol{c}_\eta, \boldsymbol{c}_\pi) \in \mathbb{R}^{\mathcal{T}} \times \mathbb{R}^{\mathcal{T}}$ with $\boldsymbol{c}_{\pi;(\mathcal{T}\setminus\mathcal{T}_d)} \leq \mathbf{0}$ and $\kappa(y)\, \boldsymbol{c}_{\eta;(\mathcal{T}\setminus\mathcal{T}_d)} \leq \mathbf{0}$ $\forall y \in \mathcal{Y}$,*

$$h_{\Upsilon_d}(\boldsymbol{c}_\eta, \boldsymbol{c}_\pi) = \int_{\mathcal{Y}} h_d(\kappa(y)\, \boldsymbol{c}_\eta + \boldsymbol{c}_\pi; \boldsymbol{p}_d(y))\, d\mu(y)$$
$$= \sum_{z \in \mathcal{Z}} \mathbb{E}\left[\boldsymbol{m}_d(\kappa(Y)\, \boldsymbol{c}_\eta + \boldsymbol{c}_\pi; Y, D, Z) \,|\, Z = z\right]\ .$$

Proof. See Appendix A.2.1. ■

Applying Lemma 4.1 now provides a tractable moment representation of the sharp upper bound (4.3) on the numerator in (4.1), which is a convex program.

Conditional means, like our target parameter $\frac{\boldsymbol{c}_A^\intercal \boldsymbol{\eta}}{\boldsymbol{c}_A^\intercal \boldsymbol{\pi}}$, are not affine functions of $(\boldsymbol{\eta}, \boldsymbol{\pi})$, however, and so, bounds on such parameters cannot be directly obtained in this way. One way to proceed is to start from (4.3) and add the constraint $\boldsymbol{c}_A^\intercal \boldsymbol{\pi} = p$ by intersecting the hyperplane $\left\{\boldsymbol{x} \,\middle|\, \boldsymbol{c}_A^\intercal \boldsymbol{x} = p\right\}$ with $\Upsilon_{d^\star} \cap \left(\mathbb{R}^{\mathcal{T}} \times \boldsymbol{\Pi}\right)$, applying Lemma 3.4 as before, and simplifying to get

$$\max_{Q \in \mathcal{Q}_{\mathbf{S}}: \boldsymbol{c}^\intercal \boldsymbol{\pi}^Q = p} \boldsymbol{c}_A^\intercal \boldsymbol{\eta} = \inf_{\lambda \in \mathbb{R}, \boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}} \left\{h_{\Upsilon_{d^\star}}(\boldsymbol{c}_A, -\boldsymbol{v}) + h_{\boldsymbol{\Pi}}(\boldsymbol{v} - \lambda \boldsymbol{c}_A) + \lambda p\right\}\ ,$$

which then gives the following sharp upper bound on θ (whenever $\underline{p} > 0$),

$$\mathbb{E}_Q\left[\kappa(Y_{d^\star}) \,|\, \boldsymbol{T} \in A\right] \leq \max_{p \in [\underline{p}, \overline{p}]} \frac{\inf_{\lambda \in \mathbb{R}, \boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}} \left(h_{\Upsilon_{d^\star}}(\boldsymbol{c}_A, -\boldsymbol{v}) + h_{\boldsymbol{\Pi}}(\boldsymbol{v} - \lambda \boldsymbol{c}_A) + \lambda p\right)}{p}\ ,$$

where $\underline{p} := \inf_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} \boldsymbol{c}_A^\intercal \boldsymbol{\pi}$ and $\overline{p} := \sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} \boldsymbol{c}_A^\intercal \boldsymbol{\pi}$. This is not a jointly convex program,[14] but this characterization of the bound is nevertheless useful, especially for obtaining analytic bounds. For our main results, however, we focus on an alternate "root"-based characterization that allows us to get generally computationally tractable representations for the sharp bounds and inclusion test, as developed in the following subsection, and used extensively in Section 5 to develop estimators for these bounds.

4.2. **Characterization Based on "Extreme Residuals".** We start by observing that $\boldsymbol{c}_A^\intercal \boldsymbol{\eta}^Q$ is zero whenever $\boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q = Q(\boldsymbol{T} \in A)$ is zero, since $\boldsymbol{c}_A^\intercal \boldsymbol{\eta}^Q = \left(\boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q\right) \mathbb{E}_Q\left[\kappa(Y_{d^\star}) \,|\, \boldsymbol{T} \in A\right]$.

[14] Note, however, that the inner problem is convex, with value function concave in p, making the outer maximization a concave(-linear) fractional program, which can be cast as a convex program via the so-called Charnes-Cooper transformation (see Schaible (1995) for a review, or Boyd and Vandenberghe (2004, Exercise 4.7)).

Therefore, we have that

$$\Theta_{\mathbf{S}} = \left\{ \gamma \in \Theta_0 \,\middle|\, \exists Q \in \mathcal{Q}_{\mathbf{S}} : \boldsymbol{c}_A^{\intercal} \boldsymbol{\eta}^Q - \gamma \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q = 0 \right\} ,$$

where we note that, for Q such that $\boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q = 0$, all $\gamma \in \Theta_0$ satisfy $\boldsymbol{c}_A^{\intercal} \boldsymbol{\eta}^Q - \gamma \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q = 0$, as required. Now, we have

$$\gamma \in \Theta_{\mathbf{S}} \iff \exists Q \in \mathcal{Q}_{\mathbf{S}} : \boldsymbol{c}_A^{\intercal} \boldsymbol{\eta}^Q - \gamma \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q = 0 \tag{4.4a}$$

$$\iff \inf_{Q \in \mathcal{Q}_{\mathbf{S}}} \left\{ \boldsymbol{c}_A^{\intercal} \boldsymbol{\eta}^Q - \gamma \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q \right\} \leq 0 \leq \sup_{Q \in \mathcal{Q}_{\mathbf{S}}} \left\{ \boldsymbol{c}_A^{\intercal} \boldsymbol{\eta}^Q - \gamma \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q \right\} , \tag{4.4b}$$

and by (4.2), the last implication (4.4b) is equivalent, under Condition C, to

$$-h_{\Upsilon_{d^\star} \cap (\mathbb{R}^{\mathcal{T}} \times \mathbf{\Pi})} \left(-\boldsymbol{c}_A, \gamma \boldsymbol{c}_A \right) \leq 0 \leq h_{\Upsilon_{d^\star} \cap (\mathbb{R}^{\mathcal{T}} \times \mathbf{\Pi})} \left(\boldsymbol{c}_A, -\gamma \boldsymbol{c} \right) .$$

In the nonlinear optimization literature, the quantity $\boldsymbol{c}^{\intercal} \boldsymbol{\eta}^Q - \gamma \boldsymbol{c}^{\intercal} \boldsymbol{\pi}^Q$ is sometimes referred to as the "(nonlinear) residual" associated with the candidate γ (e.g., Kelley (2003)). I therefore refer to the infimum and supremum in (4.4b) as the lower and upper extreme residuals, and denote them by $\underline{\Psi}(\gamma; P)$ and $\overline{\Psi}(\gamma; P)$, respectively. Applying Lemma 3.4 now gives the following residual-based characterization of the identified set for θ.

Lemma 4.2 (Residual-based Representation of $\Theta_{\mathbf{S}}$)**.** *Suppose P satisfies Condition D and $\mathbf{S}$ satisfies Condition C. Then,*

$$\Theta_{\mathbf{S}}(P) = \left\{ \gamma \in \Theta_0 \,\middle|\, \underline{\Psi}(\gamma; P) \leq 0 \leq \overline{\Psi}(\gamma; P) \right\} .$$

Moreover, $\overline{\Psi}(\cdot\,; P) : \Theta_0 \to \mathbb{R}$ is convex and

$$\overline{\Psi}(\gamma; P) = \max_{\substack{(\boldsymbol{x}_\eta, \boldsymbol{x}_\pi) \in \Upsilon_{d^\star}: \\ \boldsymbol{x}_\pi \in \mathbf{\Pi}}} \left\{ \boldsymbol{c}_A^{\intercal} \boldsymbol{x}_\eta - \gamma \boldsymbol{c}_A^{\intercal} \boldsymbol{x}_\pi \right\} \tag{4.5a}$$

$$= \inf_{\boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}} \left\{ \int_{\mathcal{Y}} h_{d^\star} \left(\left(\kappa(y) - \gamma \right) \boldsymbol{c}_A - \boldsymbol{v}; \boldsymbol{p}_{d^\star}(y) \right) d\mu(y) + h_{\mathbf{\Pi}}(\boldsymbol{v}) \right\} . \tag{4.5b}$$

Similarly, $\underline{\Psi}(\cdot\,; P) : \Theta_0 \to \mathbb{R}$ is concave and

$$\underline{\Psi}(\gamma; P) = \min_{\substack{(\boldsymbol{x}_\eta, \boldsymbol{x}_\pi) \in \Upsilon_{d^\star}: \\ \boldsymbol{x}_\pi \in \mathbf{\Pi}}} \left\{ \boldsymbol{c}_A^{\intercal} \boldsymbol{x}_\eta - \gamma \boldsymbol{c}_A^{\intercal} \boldsymbol{x}_\pi \right\} \tag{4.6a}$$

$$= - \inf_{\boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}} \left\{ \int_{\mathcal{Y}} h_{d^\star} \left(\left(\gamma - \kappa(y) \right) \boldsymbol{c}_A - \boldsymbol{v}; \boldsymbol{p}_{d^\star}(y) \right) d\mu(y) + h_{\mathbf{\Pi}}(\boldsymbol{v}) \right\} . \tag{4.6b}$$

Proof. See Appendix A.2.2. ■

In Lemma 4.2, (4.5a) and (4.6a) are "primal" representations of the extreme residual, whereas (4.5b) and (4.6b) are (Lagrangian) "dual" representations which can also be expressed in terms of moments, by applying Lemmas 3.3 and 4.1. As we will see in the next section, both representations will play a role in our proposed estimator for the sharp bounds on θ. If P is compatible with $\mathbf{S}$, i.e. $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$, then (4.5a) and (4.6a) can be written as a

maximum and minimum, respectively, but this is not necessarily true for (4.5b) and (4.6b), where, in general, the optimal values need not be attained by any optimizer.

4.2.1. *Estimation Equation for Sharp Bounds.* The characterization of $\Theta_{\mathbf{S}}$ given in Lemma 4.2 immediately suggests using the dual representations (4.5b) and (4.6b) to obtain a moment-inequality based test of candidate inclusion, and we will indeed formalize this idea in a computationally convenient form at the end of this section. First, however, we have the following result, based on (4.4a), that represents the sharp bounds on θ as the *unique root* of the corresponding extreme residual, and also expresses each bound as the solution to a convex program.

Theorem 4.1 (Sharp Bounds on Conditional Means)**.** *Suppose P satisfies Condition D and* $\mathbf{S}$ *satisfies Condition C, with $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. If $\min_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q > 0$, then the following bounds on θ are sharp under* $\mathbf{S}$*,*

$$\sup\left\{\gamma \in \Theta_0 \,|\, -\underline{\Psi}(\gamma; P) \leq 0\right\} =: \underline{\theta}(P) \;\leq\; \theta \;\leq\; \bar{\theta}(P) := \inf\left\{\gamma \in \Theta_0 \,\middle|\, \overline{\Psi}(\gamma; P) \leq 0\right\},$$

and $\underline{\Psi}(\,\cdot\,; P)$ and $\overline{\Psi}(\,\cdot\,; P)$ are strictly decreasing, with unique roots $\underline{\theta}(P)$ and $\bar{\theta}(P)$, respectively.

Proof. See Appendix A.2.3. ■

This representation of $\bar{\theta}$ and $\underline{\theta}$ as unique roots suggests a natural connection to the general theory on Z-estimators defined as roots of random data-dependent maps,[15] which will prove useful in the next section.

Remark 4.1 (Computationally Convenient Representation of $\bar{\theta}$ and $\underline{\theta}$)**.** Since $\overline{\Psi}(\,\cdot\,; P)$ and $-\underline{\Psi}(\,\cdot\,; P)$ are convex, the bounds $\underline{\theta}$ and $\bar{\theta}$ given in Theorem 4.1 are defined by convex programs. Computing $\overline{\Psi}(\,\cdot\,; P)$, e.g., requires (convex) optimization and so, it may seem that computing the bounds requires solving nested optimization problems. This is not the case, however, since the apparent nesting can be eliminated by jointly optimizing over the inner- and outer-problems in a unified convex program. Precisely, whenever $\min_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q > 0$, the sharp upper bound $\bar{\theta}$ (e.g.; lower bound is analogous) can be computed using the following "lifted" convex program, under the convention that an empty feasible set implies $\bar{\theta} = \bar{\kappa}$,[16]

$$\inf_{\substack{\gamma \in [\underline{\kappa}, \bar{\kappa}], \\ \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}}} \quad \gamma \quad \text{such that} \quad \int_{\mathcal{Y}} h_{d^\star}\left(\left(\kappa(y) - \gamma\right) \boldsymbol{c}_A - \boldsymbol{v}; \boldsymbol{p}_{d^\star}(y)\right) d\mu(y) + h_{\boldsymbol{\Pi}}(\boldsymbol{v}) \leq 0\,,$$

[15] See, e.g., Van Der Vaart and Wellner (1996, Chapter 3.3) or Kosorok (2008, Chapter 13).

[16] Under Condition C, this lifted program has an empty feasible set if and only if $\bar{\theta} = \bar{\kappa}$ and the infimum defining $\overline{\Psi}(\bar{\kappa}; P)$ is not attained; under Condition CP and $\min_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \boldsymbol{c}^{\intercal} \boldsymbol{\pi}^Q > 0$, the infimum is attained. For formal statements and proofs of the claims in this remark, see Appendix A.2.5.

which has the following moment-inequality representation,

$$\inf_{\substack{\gamma\in[\underline{\kappa},\overline{\kappa}],\boldsymbol{v}\in\mathbb{R}^{\boldsymbol{\mathcal{T}}},\\ \boldsymbol{u}\in\mathcal{C}}} \gamma \;\text{ s.t. }\; \sum_{z\in\mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}^P_{d^\star}\left(\left(\kappa\left(Y\right)-\gamma\right)\boldsymbol{c}_A-\boldsymbol{v}\,;Y,D,Z\right)\Big|\,Z=z\right] + \sum_{d\in\mathcal{D}}\sum_{z\in\mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}^P_d\left(\boldsymbol{u}_d\,;W\right)\Big|\,Z=z\right] + h_{\mathcal{R}_\pi}\left(\boldsymbol{v}-\sum_{d\in\mathcal{D}}\boldsymbol{u}_d\right)\leq 0\,.$$

Under Condition CP, this simplifies to

$$\inf_{\substack{\gamma\in[\underline{\kappa},\overline{\kappa}],\\ \boldsymbol{u}\in\mathcal{C},\\ \boldsymbol{\lambda}\geq 0}} \gamma \quad \text{s.t.} \quad \sum_{z\in\mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}^P_\theta\left(\gamma,\boldsymbol{u},\boldsymbol{\lambda};W\right)\Big|\,Z=z\right]\leq 0\,, \tag{4.7}$$

where

$$\begin{aligned}\boldsymbol{m}^P_\theta\left(\gamma,\boldsymbol{u},\boldsymbol{\lambda};W\right) := \boldsymbol{m}^P_{d^\star}&\left(\left(\kappa(Y)-\gamma\right)\boldsymbol{c}_A-\mathbf{A}_{\mathbf{S}}^{\intercal}\boldsymbol{\lambda}-\sum_{d\in\mathcal{D}}\boldsymbol{u}_d;W\right)\\ &+\sum_{d\in\mathcal{D}}\boldsymbol{m}^P_d\left(\boldsymbol{u}_d;W\right)+\frac{1}{|\mathcal{Z}|}\boldsymbol{\lambda}^{\intercal}\mathbf{a}_{\mathbf{S}}.\end{aligned} \tag{4.8}$$

and we recall that $\mathcal{C}:=\prod_{d\in\mathcal{D}}\mathcal{C}_d$, with $\mathcal{C}_d:=\left\{\boldsymbol{u}\in\mathbb{R}^{\boldsymbol{\mathcal{T}}}\,\middle|\,\boldsymbol{u}_{(\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d)}=\mathbf{0}\right\}$, and the shorthand $\boldsymbol{u}:=(\boldsymbol{u}_d)_{d\in\mathcal{D}}$ is used to split $\boldsymbol{u}$ into its $\mathcal{C}_d$ components. Note, however, that the moment functions above depend on P, which presents additional challenges for estimation and inference, as discussed in the next two sections.

Finally, I note that, under Condition CP,

$$\overline{\Psi}\left(\gamma;P\right)=\inf_{\substack{\boldsymbol{u}\in\mathcal{C},\\ \boldsymbol{\lambda}\geq 0}}\sum_{z\in\mathcal{Z}}\mathbb{E}_P\left[\boldsymbol{m}^P_\theta\left(\gamma,\boldsymbol{u},\boldsymbol{\lambda};W\right)\Big|\,Z=z\right]\,,$$

and the optimal solution to (4.7) above, whenever it exists, is a root of $\overline{\Psi}\left(\,\cdot\,;P\right)$. ♦

4.2.2. *Inclusion Test.* From Lemma 4.2, we immediately have that, for any $\gamma\in\Theta_0$,

$$\begin{aligned}\gamma\in\Theta_{\mathbf{S}} \iff 0 &\leq \min\left\{\overline{\Psi}\left(\gamma;P\right),-\underline{\Psi}\left(\gamma;P\right)\right\}\\ &=\inf_{\substack{\boldsymbol{v}\in\mathbb{R}^{\boldsymbol{\mathcal{T}}}\\ t\in\{-1,1\}}}\left\{\int_{\mathcal{Y}} h_{d^\star}\left(t\left(\kappa\left(y\right)-\gamma\right)\boldsymbol{c}_A-\boldsymbol{v};\boldsymbol{p}_{d^\star}\left(y\right)\right)d\mu\left(y\right)+h_{\boldsymbol{\Pi}}\left(\boldsymbol{v}\right)\right\}\,.\end{aligned}$$

Convexifying this problem and using positive homogeneity (in $(t,\boldsymbol{v})$) we have the following characterization of the inclusion test as a convex program.

Theorem 4.2 (Inclusion Test for Conditional Means)**.** *Suppose P satisfies Condition D and $\mathbf{S}$ satisfies Condition C. Then, for any norm $\|\cdot\|$ on $\mathbb{R}^{\boldsymbol{\mathcal{T}}}\times\mathbb{R}$,*

$$\Theta_{\mathbf{S}}=\left\{\gamma\in\Theta_0\,\middle|\,\inf_{\substack{\boldsymbol{v}\in\mathbb{R}^{\boldsymbol{\mathcal{T}}},t\in\mathbb{R},\\ \|(\boldsymbol{v},t)\|\leq 1}}\left(\int_{\mathcal{Y}} h_{d^\star}\left(t\left(\kappa\left(y\right)-\gamma\right)\boldsymbol{c}_A-\boldsymbol{v};\boldsymbol{p}_{d^\star}\left(y\right)\right)d\mu\left(y\right)+h_{\boldsymbol{\Pi}}\left(\boldsymbol{v}\right)\right)\geq 0\right\}\,.$$

Proof. See Appendix A.2.4. ■

4.3. **Point Identification.** For completeness, I also revisit sufficient conditions for point identification, due to Heckman and Pinto (2018).

Proposition 4.3 (Heckman and Pinto (2018))**.** *Suppose P satisfies Condition D. Then, for any treatment selection model $\mathbf{S}$ with $\mathcal{Q}_{\mathbf{S}} \neq \emptyset$, the following statements hold:*

(i) *For any $d \in \mathcal{D}$, if $\boldsymbol{c} \in \text{row-span}\left(\mathbf{R}_{d,\mathcal{T}_{\mathbf{S}}}\right)$, then $\boldsymbol{c}^{\intercal}\boldsymbol{q}_d$ is point-identified under $\mathbf{S}$*

(ii) *If $\boldsymbol{c} \in \text{row-span}\left(\begin{bmatrix} \mathbf{R}_{1,\mathcal{T}_{\mathbf{S}}} \\ \vdots \\ \mathbf{R}_{|\mathcal{D}|,\mathcal{T}_{\mathbf{S}}} \end{bmatrix}\right)$, then $\boldsymbol{c}^{\intercal}\boldsymbol{\pi}$ is point-identified under $\mathbf{S}$*

For a proof, see Theorem T-2 in Heckman and Pinto (2018), which is a mathematically equivalent result, except stated in terms of Moore-Penrose inverses of $\mathbf{R}_d$. Note that Proposition 4.3(i) is also sufficient for the point identification of mean outcomes, since $\boldsymbol{c} \in \text{row-span}\left(\mathbf{R}_{d,\mathcal{T}_{\mathbf{S}}}\right)$ implies that $\left\{ \frac{\boldsymbol{c}^{\intercal}\boldsymbol{\eta}}{\boldsymbol{c}^{\intercal}\boldsymbol{\pi}} \;\middle|\; (\boldsymbol{\eta},\boldsymbol{\pi}) \in \Upsilon_d \right\}$ is a singleton. Similarly, for point identification of treatment effects, $\boldsymbol{c} \in \text{row-span}\left(\mathbf{R}_{d,\mathcal{T}_{\mathbf{S}}}\right) \cap \text{row-span}\left(\mathbf{R}_{d',\mathcal{T}_{\mathbf{S}}}\right)$ is sufficient for $\left\{ \frac{\boldsymbol{c}^{\intercal}\boldsymbol{\eta}_d - \boldsymbol{c}^{\intercal}\boldsymbol{\eta}_{d'}}{\boldsymbol{c}^{\intercal}\boldsymbol{\pi}} \;\middle|\; (\boldsymbol{\eta}_d,\boldsymbol{\pi}) \in \Upsilon_d \,,\; (\boldsymbol{\eta}_{d'},\boldsymbol{\pi}) \in \Upsilon_{d'} \,. \right\}$ to collapse to a point.

These conditions are not, however, *necessary* for point-identification of any of these parameters. In the absence of these conditions, point identification can still be obtained if some "knife-edge" conditions hold, as discussed in the following remark.[17]

Remark 4.2 ("Knife-edge" Conditions for Point Identification)**.** Let $\mathbf{S}$ be a treatment selection model such that $\mathcal{Q}_{\mathbf{S}} \neq \emptyset$. If $\boldsymbol{\Pi} = \mathcal{R}_{\boldsymbol{\pi}} \cap \bigcap_{d\in\mathcal{D}} \boldsymbol{\Pi}_d$ is a singleton, say $\{\boldsymbol{o}\}$, then obviously $\boldsymbol{\pi} = \boldsymbol{o}$ is point-identified, but the (weighted) density $\boldsymbol{q}_d$ may not be point-identified for any d. If $\boldsymbol{o}$ is a *vertex* of $\boldsymbol{\Pi}_d$, however, then $\boldsymbol{q}_d$ is point-identified! I sketch the argument below. A similar argument shows that $\boldsymbol{c}^{\intercal}\boldsymbol{q}_d$ is point-identified if $\mathcal{R}_{\boldsymbol{\pi}} \cap \bigcap_{d\in\mathcal{D}} \boldsymbol{\Pi}_d$ is a (non-empty) subset of the extreme face of the polytope $\boldsymbol{\Pi}_d$ with respect to the direction $\boldsymbol{c}$.[18]

To see this, first note that any vertex of $\boldsymbol{\Pi}_d$ can be obtained by integrating along the corresponding vertex of $\text{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)$. In fact, it is possible to show that the extreme face of the polytope $\boldsymbol{\Pi}_d$, with respect to a direction $\boldsymbol{c}$, is the following set

$$\left\{ \int_{\mathcal{Y}} \boldsymbol{g}(y)\, d\mu\left(y\right) \;\middle|\; \begin{array}{l} \boldsymbol{g} : \mathcal{Y} \to \mathbb{R}^{\mathcal{T}}_{\geq 0} \text{ is measurable,} \\ \boldsymbol{g}\left(y\right) \in \left(\text{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)\right)^{\boldsymbol{c}} \forall y \in \mathcal{Y} \end{array} \right\}, \tag{4.9}$$

where $\left(\text{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)\right)^{\boldsymbol{c}}$ is defined as the (maximal) extreme face of the polytope $\text{Fibre}_d\left(\boldsymbol{p}_d\left(y\right)\right)$ with respect to direction $\boldsymbol{c}$.[19] Since $\boldsymbol{o}$ is a vertex of $\boldsymbol{\Pi}_d$, there exists a direction $\boldsymbol{c}_{\boldsymbol{o}}$ such that $\text{argmax}_{\boldsymbol{x}\in\boldsymbol{\Pi}_d}\left\{\boldsymbol{c}_{\boldsymbol{o}}^{\intercal}\boldsymbol{x}\right\} = \{\boldsymbol{o}\}$. It is straightforward to show that (4.9) is

[17] See also Navjeevan, Pinto, and Santos (2023) and Goff (2024) for a detailed discussion of related point-identification results. In particular, Theorem 2 in Goff (2024) shows that the conditions in Proposition 4.3 are also necessary to ensure that $\mathbf{S}$ delivers point-identification of such parameters under *any* distribution of the observables that is consistent with $\mathbf{S}$. This is in contrast to Remark 4.2 here, which suggests that certain parameters are point-identified under $\mathbf{S}$ (i.e. their identified set, under $\mathbf{S}$, is a singleton) whenever the distribution of observables satisfies certain conditions, even when $\mathbf{S}$ fails to satisfy the corresponding condition in Proposition 4.3.

[18] In other words, if $\mathcal{R}_{\boldsymbol{\pi}} \cap \bigcap_{d\in\mathcal{D}} \boldsymbol{\Pi}_d \subseteq \text{argmax}_{\boldsymbol{x}\in\boldsymbol{\Pi}_d}\, \boldsymbol{c}^{\intercal}\boldsymbol{x}$.

[19] This claim generalizes Proposition 1.2 in Billera and Sturmfels (1992), and follows from Lemma A.2.

a singleton if and only if $(\text{Fibre}_d(\boldsymbol{p}_d(y)))^{\boldsymbol{c}_o}$ is a singleton, say $\{\boldsymbol{b}_d^{\boldsymbol{c}_o}(y)\}$, for every $y \in \mathcal{Y}$, which immediately gives the result. Indeed, when this is the case, $\boldsymbol{q}_d(y) = \boldsymbol{b}_d^{\boldsymbol{c}_o}(y)$ for every $y \in \mathcal{Y}$, which also implies the identification of the corresponding mean counterfactuals. ♦

5. Estimating the Sharp Bounds

So far, we have treated the distribution of the observed data P as known. In this section, I assume that we only have access to a finite sample

$$\{(Y_i, D_i, Z_i)\}_{i=1}^n = \{W_i\}_{i=1}^n \overset{i.i.d.}{\sim} P$$

of independently and identically distributed draws from an unknown distribution P satisfying Condition D. As usual, we define the empirical distribution $\hat{P}_n$ by

$$\hat{P}_n(Y \leq y, D = d, Z = z) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{Y_i \leq y, D_i = d, Z_i = z\} \ .$$

Throughout this section, I assume that $\mathbf{S}$ satisfies Condition CP, and leave the development of estimation methods for TSMs that only satisfy Condition C to future work. As in the previous section, we will continue to take the target parameter θ to be the conditional mean (4.1) and maintain the assumptions on κ and $\boldsymbol{c}_A$. Throughout this section, I also assume that P is compatible with the TSM $\mathbf{S}$ and that the denominator $Q(\boldsymbol{T} \in A)$ in (4.1) is strictly positive for all parent distributions Q that are consistent with the model and the data. Precisely, I maintain the following assumption.

Assumption 5.1. *The observed data P and the treatment selection model $\mathbf{S}$ are such that*

$$\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset \ \text{ and } \ \min_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} Q(\boldsymbol{T} \in A) > 0 \ .$$

For simplicity, I only discuss estimation of $\bar{\theta}$, the upper endpoint of the identified set for θ; estimation of $\underline{\theta}$ is analogous.

Under these conditions, we have the following result that says we can restrict the (outer) optimization in the definition of $\overline{\Psi}(\gamma; P)$ to a uniformly bounded set of directions.

Lemma 5.1. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, Assumption 5.1 holds, and $\gamma \in \Theta_0$. Then, there exists $B_\theta < \infty$, depending only on $\mathbf{S}$, $\overline{\kappa} - \underline{\kappa} < \infty$ and $\boldsymbol{c}_A$, such that, for all such P and γ,*

$$\overline{\Psi}(\gamma; P) = \min_{\substack{\boldsymbol{u} \in \mathcal{C}, \boldsymbol{\lambda} \geq 0 \\ \|(\boldsymbol{u}, \boldsymbol{\lambda})\|_\infty \leq B_\theta}} \sum_{z \in \mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_\theta^P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}; W) \,\middle|\, Z = z\right] \ . \tag{5.1}$$

Proof. See Appendix B.1.1. ■

In the remainder, we will treat the moment functions $\boldsymbol{m}_\theta$ as uniformly bounded, which is justified by this lemma, together with Lemma 3.2. So far, we have defined the moment functions $\boldsymbol{m}_\theta$ via an arbitrary choice of inner optimizers $\ell_d^\star$, as given by Lemma 3.2.

This is convenient for identification, since our results there only rely on quantities like $\sum_{z\in\mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_\theta^P(\gamma,\boldsymbol{u},\boldsymbol{\lambda};W)\,\middle|\,Z=z\right]$ which are invariant to the choice of $\ell_d^\star$. The individual moment functions, however, are not themselves invariant and so, the choice can be consequential for the estimation and efficiency results below. For computational convenience, we will continue to *estimate* the moment functions $\boldsymbol{m}$ using arbitrary measurable selections of optimal inner dual multipliers,[20] but we will now be more particular about the inner dual multipliers used in the definition of their *population* counterparts, which will allow us to avoid imposing needlessly strong restrictions on P.

To this end, define $\mathfrak{M}_P^\star$ as the set of moments optimal for (5.1), including all possible optimal selections of both the outer *and inner* problems, subject only to bounds that hold over all $\mathbf{S}$-compatible P. Precisely, with $\mathcal{W}:=\mathcal{Y}\times\mathcal{D}\times\mathcal{Z}$, define

$$\mathfrak{M}_P^\star := \left\{ \begin{array}{l|l} \text{measurable} & \sum_{z\in\mathcal{Z}}\mathbb{E}_P\left[m\left(W\right)|Z=z\right]=0\ ,\ \text{and} \\ m:\mathcal{W}\to\mathbb{R} & m\left(y,d,z\right)=\mathbb{1}\left\{d=d^\star\right\}\ell_{0,(z)}\left(y\right)+\ell_{d,(z)}\left(y\right)+\frac{\boldsymbol{\lambda}^\intercal\mathbf{a}_\mathbf{S}}{|\mathcal{Z}|}\ : \\ & \text{measurable } \ell_0,\ell_{d_1},\ldots,\ell_{d_{|\mathcal{D}|}}:\mathcal{Y}\to\mathbb{R}^{\mathcal{Z}}_{\geq 0}:\forall y\in\mathcal{Y}, \\ & \left\|\ell_0\left(y\right)\right\|_\infty\leq\left\|\boldsymbol{c}_0(y)\right\|_\infty \text{ and } \mathbf{R}_{d^\star}^\intercal\ell_0\left(y\right)\geq\boldsymbol{c}_0\left(y\right), \\ & \left\|\ell_d\left(y\right)\right\|_\infty\leq\left\|\boldsymbol{u}_d\right\|_\infty \text{ and } \mathbf{R}_d^\intercal\ell_d\left(y\right)\geq\boldsymbol{u}_d\ \forall d\in\mathcal{D}, \\ & \boldsymbol{u}\in\mathcal{C},\ \boldsymbol{\lambda}\geq 0,\ \left\|(\boldsymbol{u},\boldsymbol{\lambda})\right\|_\infty\leq B_\theta, \\ & \text{and } \boldsymbol{c}_0\left(y\right)=\left(\kappa\left(y\right)-\bar{\theta}\left(P\right)\right)\boldsymbol{c}_A-\mathbf{A}_\mathbf{S}^\intercal\boldsymbol{\lambda}-\sum_{d\in\mathcal{D}}\boldsymbol{u}_d\ . \end{array} \right\}.$$

In the remainder, when we require that there exists a population optimizer satisfying some requirement, we will simply say that there exists $m\in\mathfrak{M}_P^\star$ meeting that requirement.

5.1. **A Model-Compatible "Plug-in" Estimator: Consistency and Limitations.** Given the characterization of $\bar{\theta}\left(P\right)$ provided in Theorem 4.1, a natural approach is to obtain estimates $\check{\boldsymbol{p}}_{d;n}$ of the observed sub-densities $\boldsymbol{p}_d$ and solve the convex program (4.7), with $\check{\boldsymbol{p}}_{d;n}$ in place of the unknown true $\boldsymbol{p}_d^P$. In our setting, however, this familiar "plug-in" approach faces two immediate complications.

First, when Y is not finitely-supported, suitably estimating $\boldsymbol{p}_d$ generally requires $\check{\boldsymbol{p}}_{d;n}$ to be obtained via smoothing, or some other form of regularization, which can lead to bias. Second, even when $\mathcal{Q}_\mathbf{S}\left(P\right)\neq\emptyset$, the estimated observed-data distribution $\check{P}_n$ induced by these $\check{\boldsymbol{p}}_{d,n}$ may not be compatible with $\mathbf{S}$ in finite samples, leading to empty identified sets. When P lies on the "boundary" of the set of $\mathbf{S}$-compatible observed-data distributions, $\mathcal{Q}_\mathbf{S}\left(\check{P}_n\right)$ may remain empty with nontrivial probability even in large samples. A natural solution to this second problem is to project $\left(\check{\boldsymbol{p}}_{d,n}\right)_{d\in\mathcal{D}}$ onto the set of "$\mathbf{S}$-compatible sub-densities" to obtain a compatible estimate $\left(\tilde{\boldsymbol{p}}_{d,n}\right)_{d\in\mathcal{D}}$ that allows us to construct a suitable observed-data distribution $\tilde{P}_n$, with $\boldsymbol{p}_d^{\tilde{P}_n}=\tilde{\boldsymbol{p}}_{d,n}$ for each d, such that $\mathcal{Q}_\mathbf{S}(\tilde{P}_n)\neq\emptyset$. With such a $\tilde{P}_n$ in hand,

[20] This measurability requirement is naturally satisfied by the estimated moment functions computed using the basis-extension approach of Algorithm 3.1, while allowing the algorithm itself to be implemented using arbitrary solutions provided by any linear programming solver.

we can now obtain a *compatible plug-in* $\tilde{\bar{\theta}}_n$ by solving (4.7) with $\tilde{\boldsymbol{p}}_{d,n}$ in place of $\boldsymbol{p}_d^P$, i.e.

$$\tilde{\bar{\theta}}_n := \bar{\theta}(\tilde{P}_n) \,. \tag{5.2}$$

To formalize and operationalize this construction, I now make our requirements on $\tilde{P}_n$ concrete and provide a feasible procedure for computing (5.2). First, note that P is fully determined by the sub-density vectors $\left(\boldsymbol{p}_d^P\right)_{d\in\mathcal{D}}$, which determine the conditional distributions $(Y,D)\,|Z=z$ for $z\in\mathcal{Z}$, and the marginal distribution P_Z of Z. Next, define the set of $\mathbf{S}$-compatible sub-densities as

$$\mathscr{C}_{\mathbf{S}} := \left\{ (\boldsymbol{b}_d)_{d\in\mathcal{D}} \in \prod_{d\in\mathcal{D}} L^1\left(\mu, \mathbb{R}_{\geq 0}\right)^{\mathcal{Z}} \;\middle|\; \begin{array}{l} \exists \boldsymbol{\pi}\in\mathcal{R}_{\boldsymbol{\pi}}, (\boldsymbol{x}_d)_{d\in\mathcal{D}} \in \prod\limits_{d\in\mathcal{D}} L^1\left(\mu,\mathbb{R}_{\geq 0}\right)^{\mathcal{T}}: \\ \int_{\mathcal{Y}} \boldsymbol{x}_d \, d\mu = \boldsymbol{\pi} \;\forall d, \;\; \sum\limits_{d\in\mathcal{D}} \int_{\mathcal{Y}} \boldsymbol{b}_d \, d\mu = \mathbf{1}, \\ \mathbf{R}_d \boldsymbol{x}_d(y) \leq \boldsymbol{b}_d(y) \;\; \forall d, y. \end{array} \right\} \,, \tag{5.3}$$

and note that, as derived in the discussion immediately following Lemma 3.1, the requirements imposed in the definition of $\mathscr{C}_{\mathbf{S}}$ ensure that (i) $(\boldsymbol{b}_d)_{d\in\mathcal{D}}$ is a stacked vector of sub-densities, and (ii) by (3.1), any P satisfying Condition D with $\left(\boldsymbol{p}_d^P\right)_{d\in\mathcal{D}} \in \mathscr{C}_{\mathbf{S}}$ must also have $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. In this sense, $\mathscr{C}_{\mathbf{S}}$ provides a suitable formalization of the notion of the set of "$\mathbf{S}$-compatible sub-densities", as required. Note, however, that projection onto $\mathscr{C}_{\mathbf{S}}$ is generally an infinite-dimensional operation when Y is continuous, depending on the density estimator used to construct $\check{P}_n$. See Remark 5.1 below for a convenient construction that allows compatibility to be imposed using a single linear program.

Finally, for any observed data law P', with $P'_{d|z} \ll \mu$ for all $d\in\mathcal{D}, z\in\mathcal{Z}$, define

$$\Delta\left(P', P\right) := \sum_{d\in\mathcal{D}} \sum_{z\in\mathcal{Z}} \int_{\mathcal{Y}} \left| p^{P'}_{d|z}(y) - p^{P}_{d|z}(y) \right| \, d\mu(y) \,,$$

which is the distance between the conditional sub-density vectors induced by P' and P, under the vector-valued L^1 metric induced by the ℓ^1 norm, and will serve, for our purposes, as an appropriate metric by which $\tilde{P}_n$ must approximate P. As an aside, I also note that

$$\Delta\left(P', P\right) = 2 \sum_{z\in\mathcal{Z}} \mathrm{TV}\left(P'_{(Y,D)|Z=z}, P_{(Y,D)|Z=z}\right) \,,$$

where TV denotes the total-variation distance.

In the remainder, we will assume that $\tilde{P}_n$ satisfies the following requirements.

Assumption 5.2. *For every n, $\tilde{P}_n$ is a distribution of (Y,D,Z) that satisfies:*

(i) $\left(\boldsymbol{p}_d^{\tilde{P}_n}\right)_{d\in\mathcal{D}} =: \left(\tilde{\boldsymbol{p}}_{d;n}\right)_{d\in\mathcal{D}} \in \mathscr{C}_{\mathbf{S}}$,

(ii) $\tilde{P}_{Z;n} = \hat{P}_{Z;n}$,

(iii) $\Delta\left(\tilde{P}_n, P\right) \xrightarrow{P} 0$.

As mentioned above, the first requirement ensures that $\mathcal{Q}_{\mathbf{S}}(\tilde{P}) \neq \emptyset$ and the other two requirements ensure that $\tilde{P}_n$ can still serve as a suitable approximation for P. We now have the following result, which shows that, under these conditions, $\tilde{\bar{\theta}}_n$ is consistent.

Theorem 5.1. *Suppose P satisfies Condition D and $\mathbf{S}$ satisfies Condition CP. Then, under Assumptions 5.1 and 5.2,*

$$\tilde{\bar{\theta}}_n \xrightarrow{P} \bar{\theta}(P) \ .$$

Proof. See Appendix B.1.2. ■

Remark 5.1 (Computation of $\tilde{\bar{\theta}}$)**.** The theoretical results in this section permit $\check{P}_n$ to be any preliminary estimator satisfying the specified rate, μ-measurability, and other conditions, as needed for each result. In practice, however, with continuous Y, projection onto $\mathscr{C}_{\mathbf{S}}$, the set of $\mathbf{S}$-compatible sub-densities, may not be tractable for all estimators.

One convenient way to render this step tractable is to construct the estimate $\check{P}_n$ using conditional sub-densities estimated using products of triangular hat functions on a common grid. In this case, we can impose linear coefficient restrictions that ensure the entire fitted sub-density vector lies in $\mathscr{C}_{\mathbf{S}}$. Under Condition CP, computing the minimum (in ℓ^1) coefficient adjustment required to satisfy these restrictions is a linear program. I provide the explicit construction in Appendix E, where the convergence rate of this estimator is established, and other computational details are discussed. ♦

5.1.1. *Limitations of the Plug-In Estimator.* In the remainder, suppress the dependence of $\tilde{P}$ and $\hat{P}$ on n.

Our construction of $\tilde{P}$ above enables us to overcome the challenges described earlier and define a coherent and operationalizable notion of "plug-in" estimation that is suitable for our setting. $\tilde{P}$ can be a biased estimator of P, however, and may converge at a rate slower that root-n. Moreover, $\bar{\theta}(P)$ is not generally first-order insensitive to perturbations in P and so, poor estimation of P can translate to poor estimation of $\bar{\theta}(P)$.

Finally, a further complication arises from the fact that $\tilde{\bar{\theta}}_n$ is defined as the root of a decreasing function that is itself the infimum of an estimated function, which can lead $\tilde{\bar{\theta}}_n$ to be biased downward in finite samples. To see this, first note that (5.2) implies, by definition, that $\tilde{\bar{\theta}}_n$ is the unique[21] value of γ that satisfies

$$0 = \overline{\Psi}\left(\gamma; \tilde{P}\right) \ .$$

[21] Under the assumption that $\min_{Q \in \mathcal{Q}_{\mathbf{S}}(\tilde{P})} \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}^Q > 0$ which holds w.p.a.1 under Assumptions 5.1 and 5.2, but can fail in finite-samples.

Moreover, as noted in Remark 4.1, under Condition CP and Assumption 5.1, we have that

$$\overline{\Psi}\left(\gamma;\tilde{P}\right) = \min_{\substack{\boldsymbol{u}\in\mathcal{C},\boldsymbol{\lambda}\geq 0\\ \|(\boldsymbol{u},\boldsymbol{\lambda})\|_\infty\leq B_\theta}} \sum_{z\in\mathcal{Z}} \mathbb{E}_{\tilde{P}}\left[\boldsymbol{m}_\theta^{\tilde{P}}\left(\gamma,\boldsymbol{u},\boldsymbol{\lambda};W\right)\middle| Z=z\right] , \tag{5.4}$$

which is the minimum of an estimated function and so, can itself suffer from potentially severe downward bias in finite samples because sample optimizers can be selected based on realized sampling error, as has been noted in the related literature.[22]. Finally, by Theorem 4.1, we have that, under these conditions, $\overline{\Psi}\left(\gamma;P\right)$ is a strictly decreasing function of γ, which implies that

$$\overline{\Psi}\left(\bar{\theta}\left(P\right);P\right) = 0 = \overline{\Psi}\left(\tilde{\bar{\theta}}_n;\tilde{P}\right) < \overline{\Psi}\left(\tilde{\bar{\theta}}_n;P\right) \implies \bar{\theta}\left(P\right) > \tilde{\bar{\theta}}_n ,$$

thereby transmitting the downward bias of $\overline{\Psi}\left(\gamma;\tilde{P}\right)$ to its root $\tilde{\bar{\theta}}_n$.

In the next subsection, I propose a bias-corrected estimator motivated by these concerns.

5.2. **A Bias-Corrected Estimator.** We begin this section by noting that $\tilde{\bar{\theta}}$ can generally be obtained by solving (4.7) (with $\tilde{P}_n$ in place of P) by convex optimization methods, using Algorithm 4.1 to evaluate the pointwise support functions and recover associated inner dual multipliers, which also yields the moment function $\boldsymbol{m}_\theta^{\tilde{P}}$ and optimizers $\tilde{\boldsymbol{u}} := \boldsymbol{u}(\tilde{\bar{\theta}},\tilde{P}), \tilde{\boldsymbol{\lambda}} := \boldsymbol{\lambda}(\tilde{\bar{\theta}},\tilde{P})$ associated with the solution, such that

$$\bar{\Psi}\left(\tilde{\bar{\theta}};\tilde{P}\right) = \sum_{z\in\mathcal{Z}} \mathbb{E}_{\tilde{P}}\left[\boldsymbol{m}_\theta^{\tilde{P}}\left(\tilde{\bar{\theta}},\tilde{\boldsymbol{u}},\tilde{\boldsymbol{\lambda}};W\right)\middle| Z=z\right] = 0 .$$

This (4.7) based approach relies on the dual representation (4.5b) of $\bar{\Psi}\left(\gamma;P\right)$ but, as noted earlier, $\bar{\Psi}\left(\gamma;P\right)$ also has the primal representation (4.5a). Indeed, it is also possible to obtain, from the outputs of Algorithm 3.1, a primal solution $(\tilde{\boldsymbol{\eta}},\tilde{\boldsymbol{\pi}})$ such that

$$\bar{\Psi}\left(\tilde{\bar{\theta}}_n;\tilde{P}\right) = \max_{\substack{(\boldsymbol{x}_\eta,\boldsymbol{x}_\pi)\in\Upsilon_d(\tilde{P}):\\ \boldsymbol{x}_\pi\in\boldsymbol{\Pi}(\tilde{P})}} \left\{\boldsymbol{c}_A^\intercal\boldsymbol{x}_\eta - \gamma\boldsymbol{c}_A^\intercal\boldsymbol{x}_\pi\right\} = \boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\eta}} - \tilde{\bar{\theta}}_n\boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\pi}} = 0 .$$

Since the feasible set for the primal problem does not depend on γ, we also have that

$$\bar{\Psi}\left(\gamma;\tilde{P}\right) \geq \boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\eta}} - \gamma\boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\pi}} = -\boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\pi}}\left(\gamma - \tilde{\bar{\theta}}_n\right) ,$$

where the last equality uses $\boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\eta}} = \left(\boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\pi}}\right)\tilde{\bar{\theta}}_n$. The inequality above holds with equality when $\gamma = \tilde{\bar{\theta}}_n$ and so, $-\boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\pi}} \in \partial_\gamma\bar{\Psi}\left(\tilde{\bar{\theta}}_n;\tilde{P}\right)$, where ∂_γ denotes the subdifferential. Note that, in general, neither the primal nor dual solutions are necessarily unique. When the optimal conditioning probability $\boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\pi}}$ is unique, $\bar{\Psi}\left(\gamma;\tilde{P}\right)$ is differentiable with respect to γ at its root $\tilde{\bar{\theta}}_n$, and $-\boldsymbol{c}_A^\intercal\tilde{\boldsymbol{\pi}}$ is its γ-derivative.

[22] For example, see the discussions in Manski and Pepper (2000, 2009); Haile and Tamer (2003); Kreider and Pepper (2007); Chernozhukov et al. (2013). In particular, the approach proposed by Chernozhukov et al. (2013) suggests an alternative solution to this problem, based on inverting the inclusion test proposed in Theorem 4.2 to obtain a half-median-unbiased estimator.

5.2.1. *The Correction.* In order to motivate a suitable correction for $\tilde{\bar{\theta}}_n$, first suppose that P and $\tilde{P}$ are such that $\bar{\Psi}(\gamma; P)$ admits a first-order expansion around $\left(\tilde{\bar{\theta}}_n, \tilde{P}\right)$, and that a suitable envelope theorem applies. Then, we can linearize the estimation equation for $\bar{\theta}(P)$ around $\left(\tilde{\bar{\theta}}_n, \tilde{P}\right)$ as follows,

$$\begin{aligned} 0 &= \bar{\Psi}\left(\bar{\theta}(P); P\right) \\ &\approx \bar{\Psi}\left(\tilde{\bar{\theta}}_n; \tilde{P}\right) + \sum_{z \in \mathcal{Z}} \left(P_{Z=z} - \tilde{P}_{Z=z}\right) \boldsymbol{m}_\theta^{\tilde{P}}\left(\tilde{\bar{\theta}}_n, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\lambda}}_n; \cdot\right) - \left(\boldsymbol{c}_A^\intercal \tilde{\boldsymbol{\pi}}\right)\left(\bar{\theta}(P) - \tilde{\bar{\theta}}_n\right), \end{aligned}$$

where, for integrable functions f,

$$\left(P_{Z=z} - \tilde{P}_{Z=z}\right) f(\cdot) := \mathbb{E}_P\left[f(W) \,|\, Z = z\right] - \mathbb{E}_{\tilde{P}}\left[f(W) \,|\, Z = z\right].$$

Solving for $\bar{\theta}(P)$ and using $\bar{\Psi}\left(\tilde{\bar{\theta}}_n; \tilde{P}\right) = \sum_z \mathbb{E}_{\tilde{P}}\left[\boldsymbol{m}_\theta^{\tilde{P}}\left(\tilde{\bar{\theta}}_n, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\lambda}}; W\right) \,\middle|\, Z = z\right] = 0$ yields

$$\bar{\theta}(P) \approx \tilde{\bar{\theta}}_n + \frac{1}{\boldsymbol{c}_A^\intercal \tilde{\boldsymbol{\pi}}} \sum_{z \in \mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_\theta^{\tilde{P}}\left(\tilde{\bar{\theta}}_n, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\lambda}}; W\right) \,\middle|\, Z = z\right]. \tag{5.5}$$

Although the linearization used to obtain (5.5) relies on regularity conditions that will not generally be imposed in the remainder, the resulting correction also possesses several properties that are desirable for one-sided control, based on primal-dual inequalities that do not rely on differentiability. To see this, first note that the feasible set over which the infimum in (5.4) is taken does not depend on γ or P, we have that the $\tilde{P}$-optimal moment remains feasible at the true P, yielding the following weak-duality inequality

$$\sum_{z \in \mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_\theta^{\tilde{P}}\left(\tilde{\bar{\theta}}_n, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\lambda}}; W\right) \,\middle|\, Z = z\right] \geq \overline{\Psi}\left(\tilde{\bar{\theta}}; P\right).$$

Next, consider a parent distribution $Q \in \mathcal{Q}_{\mathbf{S}}(P)$ such that $\theta^Q = \bar{\theta}(P)$, in which case we have $\boldsymbol{c}_A^\intercal \boldsymbol{\eta}_{d^\star}^Q = \bar{\theta}(P)\, \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q$, which yields

$$\overline{\Psi}\left(\tilde{\bar{\theta}}; P\right) \geq \boldsymbol{c}_A^\intercal \boldsymbol{\eta}_{d^\star}^Q - \tilde{\bar{\theta}} \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q = \left(\bar{\theta}(P) - \tilde{\bar{\theta}}\right) \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q,$$

where the first inequality follows from the primal representation (4.5a) of $\overline{\Psi}(\gamma; P)$. Together, the inequalities in the last two displays imply that, for every such endpoint-optimal parent,

$$\sum_{z \in \mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_\theta^{\tilde{P}}\left(\tilde{\bar{\theta}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\lambda}}; W\right) \,\middle|\, Z = z\right] \geq \left(\bar{\theta}(P) - \tilde{\bar{\theta}}\right) \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q, \tag{5.6}$$

without any differentiability requirements. Now, substituting this inequality into the right hand side of (5.5), we get, for any such parent distribution Q,

$$\begin{aligned} &\tilde{\bar{\theta}}_n + \frac{1}{\boldsymbol{c}_A^\intercal \tilde{\boldsymbol{\pi}}} \sum_{z \in \mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_\theta^{\tilde{P}}\left(\tilde{\bar{\theta}}_n, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\lambda}}; W\right) \,\middle|\, Z = z\right] \\ &\qquad \geq \tilde{\bar{\theta}}_n + \frac{\bar{\theta}(P) - \tilde{\bar{\theta}}_n}{\boldsymbol{c}_A^\intercal \tilde{\boldsymbol{\pi}}} \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q = \bar{\theta}(P) + \frac{\left(\tilde{\bar{\theta}}_n - \bar{\theta}(P)\right)\left(\boldsymbol{c}_A^\intercal \tilde{\boldsymbol{\pi}} - \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q\right)}{\boldsymbol{c}_A^\intercal \tilde{\boldsymbol{\pi}}} \end{aligned}$$

$$\geq \bar{\theta}(P) - \frac{\left|\tilde{\bar{\theta}}_n - \bar{\theta}(P)\right| \left|\boldsymbol{c}_A^{\mathsf{T}}\tilde{\boldsymbol{\pi}} - \boldsymbol{c}_A^{\mathsf{T}}\boldsymbol{\pi}^Q\right|}{\boldsymbol{c}_A^{\mathsf{T}}\tilde{\boldsymbol{\pi}}}, \quad (5.7)$$

where the equality in the penultimate line follows from adding and subtracting $\bar{\theta}(P)$, then simplifying.

5.2.2. *The Estimator.* Since the expectation in (5.5) is taken with respect to the true P, which is unknown, we need to evaluate the moment term in (5.5) at a suitable estimated distribution, in order to operationalize the correction suggested by our derivation above. Therefore, we will estimate the correction term by evaluating the expectation in its numerator on held-out data.

To formally define the bias-corrected estimator, I introduce the following notation. Randomly partition the sample into K folds $I_1, \ldots, I_K \subseteq [n]$, for fixed $K \geq 2$, and define $n^{(k)} := |I_k|$ and $n_z^{(k)} := |\{i \in I_k \,|\, Z_i = z\}|$. For each k, define $\hat{P}^{(k)}$ and $\tilde{P}^{(k)}$ as the versions of $\hat{P}$ and $\tilde{P}$, respectively, defined using the subsample $\{(Y_i, D_i, Z_i)\}_{i \in I_k}$, define $\hat{P}^{(-k)}$ and $\tilde{P}^{(-k)}$ as the analogous versions defined using $\{(Y_i, D_i, Z_i)\}_{i \in [n] \setminus I_k}$, and extend the remaining notation accordingly. Finally, define the bias-corrected estimator $\hat{\tilde{\theta}}_n^{(\mathrm{CF})}$ as follows,

$$\hat{\tilde{\theta}}_n^{(\mathrm{CF})} := \sum_{k=1}^{K} \frac{n^{(k)}}{n} \left(\tilde{\bar{\theta}}_n^{(-k)} + \frac{1}{\boldsymbol{c}_A^{\mathsf{T}}\tilde{\boldsymbol{\pi}}^{(-k)}} \sum_{z \in \mathcal{Z}} \frac{1}{n_z^{(k)}} \sum_{\substack{i \in I_k: \\ Z_i = z}} \boldsymbol{m}_\theta^{\tilde{P}^{(-k)}} \left(\tilde{\bar{\theta}}_n^{(-k)}, \tilde{\boldsymbol{u}}^{(-k)}, \tilde{\boldsymbol{\lambda}}^{(-k)}; W_i \right) \right).$$

In the remainder, I assume that the folds are selected independently of the data and, for each $k \in \{1, \ldots, K\}$, $|I_k| > 0$ and $\frac{n^{(k)}}{n} \to c_k$ for some $c_k \in (0, 1)$. Throughout, in all results involving $\hat{\tilde{\theta}}_n^{(\mathrm{CF})}$, conditions imposed on $\tilde{P}$, including Assumption 5.2, will be understood to hold for each $\tilde{P}^{(-k)}$.

5.2.3. *Consistency and Root-n Control of Inward Error.* First, we have that $\hat{\tilde{\theta}}_n^{(\mathrm{CF})}$ is consistent under essentially the same conditions needed for consistency of $\tilde{\bar{\theta}}_n$.

Theorem 5.2. *Suppose P satisfies Condition D and* **S** *satisfies Condition CP. Then, under Assumptions 5.1 and 5.2,*

$$\hat{\tilde{\theta}}_n^{(\mathrm{CF})} \xrightarrow{P} \bar{\theta}(P).$$

Proof. See Appendix B.1.3. ■

In order to control inward error, we need the following technical regularity condition that permits continuous, discrete (finitely-supported), and mixed distributions.

Assumption 5.3. *Under P, the conditional distribution of $\kappa(Y)$ given $D = d^\star$ has an essentially bounded density with respect to the Lebesgue measure, with the exception of finitely many atoms.*

Informally, Assumption 5.3 restricts how much probability can concentrate on small intervals of $\kappa(Y)$.[23] Recall that $\{x\}_- := \min\{x, 0\}$. We now have the following result formalizing the sense in which the inward error of $\hat{\bar{\theta}}_n^{(\mathrm{CF})}$ is controlled at root-n scale. Importantly, note that this result imposes no differentiability requirements on the extreme residual $\overline{\Psi}$ or on the endpoint functional $\bar{\theta}(P)$ itself and so, holds even in highly irregular cases.

Theorem 5.3. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and*

$$\Delta\left(\tilde{P}, P\right) = O_P\left(n^{-1/3}\right) . \tag{5.8}$$

Then, under Assumptions 5.1 to 5.3,

$$\sqrt{n}\left\{\hat{\bar{\theta}}_n^{(\mathrm{CF})} - \bar{\theta}(P)\right\}_- = O_P(1) .$$

Proof. See Appendix B.1.4. ■

The rate requirement (5.8) is fairly mild and does not require the sub-densities to be estimated at root-n rate. In particular, it does not necessarily impose any higher order smoothness requirements on the outcome (sub-)densities: when Y is a continuous scalar, the estimator discussed in Remark 5.1, with bandwidth sequence $h_n \asymp n^{-1/3}$, achieves (5.8) whenever Y has bounded Lipschitz sub-densities (see Appendix E). When the conditioning probability $\mathbb{P}(\boldsymbol{T} \in A)$ is point-identified under $\mathbf{S}$,[24] the conclusion of Theorem 5.3 continues to hold even without Assumption 5.3, and with (5.8) weakened to $\Delta\left(\tilde{P}_n, P\right) = O_P\left(n^{-1/4}\right)$.

5.2.4. *Root-n Consistency and Gaussianity.* For each $k \in \{1, \ldots, K\}$, $d \in \mathcal{D}$ and $z \in \mathcal{Z}$, define the signed product error

$$R_{k,d,z}\left(\tilde{P}, P; m\right) := \int_{\mathcal{Y}} \left[p^{P}_{d|z}(y) - p^{\tilde{P}^{(-k)}}_{d|z}(y)\right] \times \left[\boldsymbol{m}_\theta^{\tilde{P}^{(-k)}}\left(\tilde{\bar{\theta}}_n^{(-k)}, \tilde{\boldsymbol{u}}^{(-k)}, \tilde{\boldsymbol{\lambda}}^{(-k)}; y, d, z\right) - m(y, d, z)\right] d\mu(y) .$$

For root-n consistency of $\hat{\bar{\theta}}_n^{(\mathrm{CF})}$, we need the following additional assumption, on the control of the outward error of the sum of these signed errors.

Assumption 5.4 (Approximation of Optimizer Sequence)**.** *For $k \in \{1, \ldots, K\}$, there exists a measurable sequence $m^{(k)}_{n;P} \in \mathfrak{M}^\star_P$ such that*

$$\left\{\sum_{d\in\mathcal{D}}\sum_{z\in\mathcal{Z}} R_{k,d,z}\left(\tilde{P}, P; m^{(k)}_{n;P}\right)\right\}_+ = O_P\left(n^{-1/2}\right) .$$

Under (5.8), which requires that $\tilde{P}$ suitably approximates P at rate $n^{-1/3}$, a simple sufficient condition for Assumption 5.4 is that each fold's estimated optimal moment function

[23] The main role of this assumption is to rule out cases where $\kappa(Y)$ has a density that diverges near a point (e.g., a Beta distribution with a shape parameter less than 1, which has a "bathtub" shaped density).

[24] In the strong ("generic") sense of the row-span condition required by Proposition 4.3(ii).

$\boldsymbol{m}_\theta^{\tilde{P}}\left(\tilde{\bar{\theta}}_n, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{\lambda}}; \cdot, \cdot, \cdot\right)$ approaches a population optimal moment function at rate $n^{-1/6}$, uniformly over (y, d, z); uniformity is not necessary, however, since poor approximations of the optimal moment on "low probability" regions of $\mathcal{Y}$ (e.g., due to changes in optimal bases across $\tilde{P}$ and P) may contribute only negligibly to the integrated product error $R_{k,d,z}$.

Theorem 5.4. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and (5.8) holds. Then, under Assumptions 5.1 to 5.4,*

$$\sqrt{n}\left(\hat{\bar{\theta}}_n^{(\mathrm{CF})} - \bar{\theta}(P)\right) = O_P(1) \ .$$

Proof. See Appendix B.1.5. ■

For asymptotic Gaussianity, we need significantly stronger stability and remainder control conditions, which replace Assumptions 5.3 and 5.4, as follows.

Assumption 5.5. *There exists $p_A^\dagger(P) \in (0, 1]$ such that, for all $Q \in \mathcal{Q}_{\mathbf{S}}(P)$ with $\theta^Q = \bar{\theta}(P)$,*

$$\boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q = p_A^\dagger(P) \ .$$

Assumption 5.6 (Common Score of Optimal Moments)**.** *For $m_1, m_2 \in \mathfrak{M}_P^\star$ and $z \in \mathcal{Z}$,*

$$m_1(Y, D, z) - \mathbb{E}_P[m_1(W) \mid Z = z] = m_2(Y, D, z) - \mathbb{E}_P[m_2(W) \mid Z = z] \ ,$$

$P_{Y,D|Z=z}$-almost surely.

Assumption 5.7. *For $k \in \{1, \ldots, K\}$, there exists a measurable sequence $m_{n;P}^{(k)} \in \mathfrak{M}_P^\star$ such that $\Delta\left(\tilde{P}, P\right)\left|\boldsymbol{c}_A^\intercal \tilde{\boldsymbol{\pi}}^{(-k)} - \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q\right| = o_P\left(n^{-1/2}\right)$ for some $Q \in \mathcal{Q}_{\mathbf{S}}(P)$ with $\theta^Q = \bar{\theta}(P)$, and*

$$\left\{\sum_{d \in \mathcal{D}} \sum_{z \in \mathcal{Z}} R_{k,d,z}\left(\tilde{P}, P; m_{n;P}^{(k)}\right)\right\}_+ = o_P\left(n^{-1/2}\right) \ .$$

The product rate requirement is easily satisfied when $\mathcal{Y}$ is finite since $\Delta\left(\tilde{P}, P\right) = O_P\left(n^{-1/2}\right)$ also implies a $o_P(1)$ denominator error. More generally, the $n^{1/3}$ rate of (5.8) together with $o_P(n^{-1/6})$ denominator error can deliver the required product rate.

Under Assumptions 5.5 and 5.6, define (with $P_Z(z) := P(Z = z)$) the conditionally centred score ϕ_P as

$$\phi_P(W) := \frac{m(W) - \mathbb{E}_P[m(W) \mid Z]}{P_Z(Z)\, p_A^\dagger(P)} \ , \tag{5.9}$$

for any $m \in \mathfrak{M}_P^\star$, with $\phi_P(W)$ invariant to choice of m under Assumption 5.6.

Theorem 5.5. *Suppose P satisfies Condition D and $\mathbf{S}$ satisfies Condition CP. Then, under Assumptions 5.1, 5.2 and 5.5 to 5.7,*

$$\sqrt{n}\left(\hat{\bar{\theta}}_n^{(\mathrm{CF})} - \bar{\theta}\left(P\right)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_P\left(W_i\right) + o_P(1) \rightsquigarrow \mathcal{N}\left(0, \mathbb{E}_P\left[\phi_P^2\left(W\right)\right]\right) .$$

Proof. See Appendix B.1.6. ■

5.3. **Efficient Estimation.**

5.3.1. *Preliminaries and Statistical Model.* In the remainder, I will generally assume that P_Z is entirely unknown and estimated from the data; where relevant, I will provide the minor modifications required to adapt the following results to the case where the assignment probabilities are known *a priori* (as in, e.g., a randomized controlled trial).

Let ν be the product measure on $\mathcal{Y}\times\mathcal{D}\times\mathcal{Z}$ produced by μ on $\mathcal{Y}$ and the counting measure on $\mathcal{D}\times\mathcal{Z}$. Define the observable law (statistical) model

$$\mathfrak{P}_{\mathbf{S}} := \left\{ P \ll \nu \;\middle|\; P_Z\left(z\right) > 0 \;\forall z \in \mathcal{Z},\; \mathcal{Q}_{\mathbf{S}}\left(P\right) \neq \emptyset,\; \min_{Q\in\mathcal{Q}_{\mathbf{S}}(P)} \boldsymbol{c}_A^{\intercal}\boldsymbol{\pi}^Q > 0 \right\} .$$

If assignment probabilities are known, then take $\mathfrak{P}_{\mathbf{S}}$ to be the submodel that only includes P with P_Z equal to the known probabilities.

In order to state the efficiency results precisely, I now formally define several notions from the semiparametric estimation literature that are relevant to our efficiency theory. For a detailed discussion of these notions, see Van Der Vaart (1998, Section 25.3) and Bickel, Klaassen, Ritov, and Wellner (1998, Chs. 3 and 5.2). Denote the set of mean zero square-integrable functions by $L_0^2\left(P\right) := \left\{ s \in L^2\left(P\right) \,\middle|\, \mathbb{E}_P\left[s\left(W\right)\right] = 0 \right\}$.

Definition 5.1 (Quadratic-mean Differentiable Paths and Tangent Space)**.** *A two-sided path through P in $\mathfrak{P}_{\mathbf{S}}$ is a map $t \mapsto P_t$ such that $P_0 = P$ and, for some $\epsilon > 0$, $P_t \in \mathfrak{P}_{\mathbf{S}}$ for all $t \in (-\epsilon, \epsilon)$. Denoting by $r_t := \frac{dP_t}{d\nu}$ the Radon-Nikodym derivative of P_t with respect to ν, we will say that such a path is **quadratic mean differentiable** (QMD) at zero with score $s \in L_0^2\left(P\right)$ iff, as $t \to 0$,*

$$\left\| \frac{\sqrt{r_t} - \sqrt{r_0}}{t} - \frac{1}{2} s \sqrt{r_0} \right\|_{L^2(\nu)} \to 0 .$$

*Finally, define the tangent set of $\mathfrak{P}_{\mathbf{S}}$ at P as the set consisting of scores of all such QMD paths through P, and define the **tangent space** $\dot{\mathfrak{P}}_{\mathbf{S}}^P$ of $\mathfrak{P}_{\mathbf{S}}$ at P as the closed linear span in $L^2\left(P\right)$ of this tangent set.*

With this definition in hand, we can now define the notion of parameters being "pathwise differentiable" with respect to the data distribution P, relative to our model $\mathfrak{P}_{\mathbf{S}}$, as follows.

Definition 5.2 (Pathwise Differentiability and Efficient Influence Function)**.** *For any Banach space $\mathbb{B}$, we will say that a parameter $\vartheta : \mathfrak{P}_{\mathbf{S}} \to \mathbb{B}$ is **pathwise differentiable** at P with*

respect to $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$ *iff there exists a continuous linear map* $\dot{\vartheta}_P : \dot{\mathfrak{P}}_{\mathbf{S}}^{P} \to \mathbb{B}$*, called the pathwise derivative of* ϑ *at* P*, such that every two-sided QMD path* $(P_t)_t$ *through* P *in* $\mathfrak{P}_{\mathbf{S}}$*, with score* $s \in L_0^2(P)$*, satisfies, as* $t \to 0$*,*

$$\left\| \frac{1}{t} \left(\vartheta(P_t) - \vartheta(P) \right) - \dot{\vartheta}_P(s) \right\|_{\mathbb{B}} \to 0 \, .$$

When $\mathbb{B} = \mathbb{R}$*, there exists a (*P*-almost-surely-)unique* $g \in \dot{\mathfrak{P}}_{\mathbf{S}}^{P}$ *such that*

$$\dot{\vartheta}_P(s) = \mathbb{E}_P\left[g(W) \, s(W) \right] \ \forall s \in \dot{\mathfrak{P}}_{\mathbf{S}}^{P} \, , \tag{5.10}$$

and we refer to g *as the* ***efficient influence function*** *(EIF)* $\phi^*_{P;\vartheta}$ *of* ϑ *at* P*.*

In the above definition, the existence and uniqueness of such a g follows from the Riesz representation theorem, and given a scalar parameter ϑ that is pathwise differentiable at P, with influence function g, its EIF $\phi^*_{P;\vartheta}$ is the orthogonal projection of g onto $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$.

We are now ready to formally define the notion of asymptotic efficiency that we will work with in the remainder.

Definition 5.3 (Asymptotic Linearity, Regularity, and Efficiency)**.** *An estimator* T_n *for parameter* $\vartheta : \mathfrak{P}_{\mathbf{S}} \to \mathbb{R}$ *is said to be* ***asymptotically linear*** *at* P*, with influence function* $g \in L_0^2(P)$*, iff* $\sqrt{n}\left(T_n - \vartheta(P)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(W_i) + o_P(1)$*.*

T_n *is said to be* ***regular*** *at* P *in* $\mathfrak{P}_{\mathbf{S}}$ *iff there exists a distribution* L *such that, for every two-sided path* $(P_t)_t$*, through* P *in* $\mathfrak{P}_{\mathbf{S}}$*, that is differentiable in quadratic mean at zero, and every fixed* $h \in \mathbb{R}$*, under* $P_{h/\sqrt{n}}$*,*

$$\sqrt{n}\left(T_n - \vartheta\left(P_{h/\sqrt{n}} \right) \right) \rightsquigarrow \mathscr{L} \, . \tag{5.11}$$

If, at P*,* ϑ *is pathwise differentiable and* T_n *is regular, then* T_n *is* ***asymptotically efficient*** *at* P *iff* (5.11) *holds with*

$$\mathscr{L} = \mathcal{N}\left(0, \mathbb{E}_P\left[\phi^*_{P;\vartheta}(W)^2 \right] \right) \, .$$

These definitions provide the foundation needed to formally assess the efficiency of $\hat{\bar{\theta}}_n^{(\mathrm{CF})}$. In the remainder of the main text, we will write $\phi^*_P := \phi^*_{P;\bar{\theta}}$. First, note that, for any regular asymptotically linear estimator of $\bar{\theta}$, with influence function g, we have

$$\mathbb{E}_P\left[g(W)^2 \right] = \mathbb{E}_P\left[\phi^*_P(W)^2 \right] + \mathbb{E}_P\left[\left(g(W) - \phi^*_P(W) \right)^2 \right] \, ,$$

since $\mathbb{E}_P\left[\phi^*_P(W) \left(g(W) - \phi^*_P(W) \right) \right] = 0$ as ϕ^*_P is the orthogonal projection of $g = \phi^*_P + (g - \phi^*_P)$ onto $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$. Therefore, any such estimator is efficient if and only if its influence function is P-almost-surely equal to ϕ^*_P. In the remaining subsections, I characterize ϕ^*_P and provide additional conditions under which (i) $\hat{\bar{\theta}}_n^{(\mathrm{CF})}$ is regular, with influence function equal to the function ϕ_P introduced earlier, and (ii) ϕ_P coincides with ϕ^*_P.

5.3.2. *A Semiparametric Efficiency Bound for Sharp Bounds on Conditional Means.* In order to characterize efficiency at P in $\mathfrak{P}_{\mathbf{S}}$, we first need to obtain a suitable characterization of the tangent space $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$ of $\mathfrak{P}_{\mathbf{S}}$ at P, the closed linear span in $L^2(P)$ of the set of scores generated by two-sided QMD paths through P in $\mathfrak{P}_{\mathbf{S}}$. First, recall $\mathscr{C}_{\mathbf{S}}$, the set of $\mathbf{S}$-compatible sub-densities defined in (5.3) above, and note that for any path $t \mapsto P_t$, we have

$$P_t \in \mathfrak{P}_{\mathbf{S}} \iff P_{t;Z} > 0, \ \min_{Q \in \mathcal{Q}_{\mathbf{S}}(P_t)} \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q > 0 \ \text{ and } \ (\boldsymbol{p}_d^{P_t})_{d \in \mathcal{D}} \in \mathscr{C}_{\mathbf{S}}.$$

Since we already impose the first two requirements at $P = P_0 \in \mathfrak{P}_{\mathbf{S}}$, and since $h_{\mathbf{\Pi}}$ as given in Theorem 3.2 is a continuous function of $\left(\boldsymbol{p}_d^P\right)_{d \in \mathcal{D}}$, we must have that any QMD path satisfies $P_{t;Z} > 0$ and $\min_{Q \in \mathcal{Q}_{\mathbf{S}}(P_t)} \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q = -h_{\mathbf{\Pi}(P_t)}(-\boldsymbol{c}_A) > 0$ for sufficiently small t. Therefore, the final requirement, that the conditional sub-densities associated with such a path must lie in $\mathscr{C}_{\mathbf{S}}$ is the substantive requirement. This suggests examining the conditions under which a QMD path P_t has conditional sub-densities that lie in $\mathscr{C}_{\mathbf{S}}$. Note that for any such path, we have $r_t(y,d,z) = P_{t;Z}(z)\, p_{d|z}^{P_t}(y)$, with $r_0(y,d,z) := P_Z(z)\, p_{d|z}^{P}(y)$. We now have the following lemma that derives some useful pathwise derivatives.

Lemma 5.2. *Suppose $P \in \mathfrak{P}_{\mathbf{S}}$. Then, $p_{d|z}^P : \mathfrak{P}_{\mathbf{S}} \to L^1(\mu)$, $P_Z : \mathfrak{P}_{\mathbf{S}} \to \mathbb{R}^{\mathcal{Z}}$ and $r_0 : \mathfrak{P}_{\mathbf{S}} \to L^1(\nu)$ are pathwise differentiable. Along any two-sided QMD path $(P_t)_t$ through P in $\mathfrak{P}_{\mathbf{S}}$, with score $s \in L_0^2(P)$,*[25]

$$\begin{aligned}
\dot{p}_{d|z}^P(s)(y) &= p_{d|z}^P(y)\left(s(y,d,z) - \mathbb{E}_P[s(W) \mid Z = z]\right), \\
\dot{P}_Z(s)(z) &= P_Z(z)\, \mathbb{E}_P[s(W) \mid Z = z], \\
\dot{r}_0(s)(y,d,z) &= r_0(y,d,z)\, s(y,d,z),
\end{aligned}$$

and, with $h_s(y,d,z) := s(y,d,z) - \mathbb{E}_P[s(W) \mid Z = z]$, as $t \to 0$,

$$\sum_{d \in \mathcal{D}, z \in \mathcal{Z}} P_Z(z) \left\| \frac{\sqrt{p_{d|z}^{P_t}} - \sqrt{p_{d|z}^{P}}}{t} - \frac{1}{2} h_s(\cdot,d,z) \sqrt{p_{d|z}^{P}} \right\|_{L^2(\mu)}^2 \to 0. \tag{5.12}$$

Proof. See Appendix B.1.7 ■

Given $s \in L_0^2(P)$, Lemma 5.2 suggests decomposing s orthogonally, with h_s as defined there, into $s(y,d,z) = h_s(y,d,z) + \mathbb{E}_P[s(W) \mid Z = z]$, where $z \mapsto \mathbb{E}_P[s(W) \mid Z = z]$ reflects changes in the assignment probabilities and h_s reflects changes in the sub-densities $\boldsymbol{p}_d$. Therefore, h_s is the component of s that is constrained by the requirement that $(\boldsymbol{p}_d^{P_t})_{d \in \mathcal{D}} \in \mathscr{C}_{\mathbf{S}}$. For $s \in L_0^2(P)$, define $\dot{\boldsymbol{p}}_d^P(s) := \left(p_{d|z}^P h_s(\cdot,d,z)\right)_{z \in \mathcal{Z}}$ and $\dot{\boldsymbol{p}}^P(s) := \left(\dot{\boldsymbol{p}}_d^P(s)\right)_{d \in \mathcal{D}}$.

[25] Note that the lemma, as written, does not depend on whether or not assignment probabilities are known. However, if the assignment probabilities are known, then s will necessarily satisfy $\mathbb{E}_P[s(W) \mid Z = z] = 0$ for all $z \in \mathcal{Z}$, and $\dot{P}_Z$ will be zero, as required.

In order to characterize $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$, we first define a suitable "tangent space" $\dot{\mathscr{C}}_{\mathbf{S}}^{P}$ of $\mathscr{C}_{\mathbf{S}}$ at P. Precisely, noting that $\dot{p}_{d|z}(s)(y) = p_{d|z}^{P}(y)\, h_s(y,d,z) = 0$ whenever $p_{d|z}^{P}(y) = 0$, define the ambient space of sub-density perturbations at P, denoted $\mathbb{H}_P$, as

$$\mathbb{H}_P := \left\{ (\boldsymbol{b}_d)_{d\in\mathcal{D}} = \left(\left(b_{d|z}\right)_{z\in\mathcal{Z}} \right)_{d\in\mathcal{D}} \middle| \begin{array}{l} b_{d|z} \in \{\mathcal{Y} \to \mathbb{R}\} \text{ measurable}, \\ b_{d|z}(y) = 0 \ \forall y \text{ with } p_{d|z}^{P}(y) = 0, \\ \displaystyle\sum_{d\in\mathcal{D}, z\in\mathcal{Z}} P_Z(z) \int_{\{p_{d|z}^{P}(y)>0\}} \frac{b_{d|z}(y)^2}{p_{d|z}(y)}\, d\mu(y) < \infty . \end{array} \right\}$$

and endow it with the norm $\|\cdot\|_P$, defined, for $\boldsymbol{b} = (\boldsymbol{b}_d)_{d\in\mathcal{D}}$, as

$$\|\boldsymbol{b}\|_P := \left(\sum_{d\in\mathcal{D}, z\in\mathcal{Z}} P_Z(z) \int_{\{y \,|\, p_{d|z}^{P}(y)>0\}} \frac{b_{d|z}(y)^2}{p_{d|z}(y)}\, d\mu(y) \right)^{1/2} .$$

Then, for every $\boldsymbol{b} = (\boldsymbol{b}_d)_d$ in $\mathbb{H}_P$, there exists a P-almost-surely-unique $h \in L^2(P)$ such that $b_{d|z}(y) = p_{d|z}(y)\, h(y,d,z)$,[26] with

$$\|\boldsymbol{b}\|_P = \sqrt{\mathbb{E}\left[h(W)^2\right]} = \|h\|_{L^2(P)} .$$

Conversely, given any $h \in L^2(P)$, we have that $\left(\left(p_{d|z}(\cdot) h(\cdot,d,z)\right)_{d\in\mathcal{D}} \right)_{z\in\mathcal{Z}} \in \mathbb{H}_P$, with its norm unchanged. Therefore, the mapping $h \mapsto \left(\left(p_{d|z}(\cdot) h(\cdot,d,z)\right)_{d\in\mathcal{D}} \right)_{z\in\mathcal{Z}}$ defines a linear bijective isometry between $L^2(P)$ and $\mathbb{H}_P$, making $\mathbb{H}_P$ a Hilbert space. By construction, we have that for any $s \in L_0^2(P)$ such that there exists a two-way QMD path $(P_t)_t$ through P in $\mathfrak{P}_{\mathbf{S}}$ with score s, the pathwise derivative $\dot{\boldsymbol{p}}^{P}(s)$ is in $\mathbb{H}_P$. We now define $\dot{\mathscr{C}}_{\mathbf{S}}^{P} \subseteq \mathbb{H}_P$ as the closed linear span in $\mathbb{H}_P$ of the set of $\dot{\boldsymbol{p}}^{P}(s)$ generated by scores s of such paths. Note that $\dot{\boldsymbol{p}}^{P}(s)$ is a vector of pathwise derivatives in $L^1(\mu)$, as given in Lemma 5.2, whereas the closure in the definition of $\dot{\mathscr{C}}_{\mathbf{S}}^{P}$ is in the norm corresponding to $L^2(P)$ scores. This yields the following representation of $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$, which will prove useful.

Define $\mathcal{L}_P := \left\{ s \in L_0^2(P) \,\middle|\, \mathbb{E}_P[s(W) \,|\, Z = z] = 0 \ \forall z \in \mathcal{Z} \right\}$ if the assignment probabilities are known, and set $\mathcal{L}_P := L_0^2(P)$ otherwise.

Lemma 5.3. *Suppose* $\mathbf{S}$ *satisfies Condition CP and* $P \in \mathfrak{P}_{\mathbf{S}}$. *Then,*

$$\dot{\mathfrak{P}}_{\mathbf{S}}^{P} = \left\{ s \in \mathcal{L}_P \,\middle|\, \dot{\boldsymbol{p}}^{P}(s) \in \dot{\mathscr{C}}_{\mathbf{S}}^{P} \right\} .$$

In particular, $s \in \dot{\mathfrak{P}}_{\mathbf{S}}^{P}$ *if and only if there exists* $a_s : \mathcal{Z} \to \mathbb{R}$ *and* $h_s \in L_0^2(P)$ *such that*

$$\sum_{z\in\mathcal{Z}} P_Z(z)\, a_s(z) = 0, \ \mathbb{E}_P[h_s(W) \,|\, Z] = 0 \ \textit{a.s.}, \ \left(\left(p_{d|z}^{P}(\cdot)\, h_s(\cdot,d,z) \right)_{d\in\mathcal{D}} \right)_{z\in\mathcal{Z}} \in \dot{\mathscr{C}}_{\mathbf{S}}^{P} ,$$

and $s(y,d,z) = a_s(z) + h_s(y,d,z)$, *with* $a_s(z) = 0$ *for every* $z \in \mathcal{Z}$ *when* P_Z *is known.*

[26] Simply take $h(y,d,z) := b_{d|z}(y)/p_{d|z}(y)$ whenever $p_{d|z}(y) > 0$ and 0 otherwise. Any other solution to $b_{d|z}(y) = p_{d|z}(y)\, h(y,d,z)$ must agree with this one wherever $p_{d|z}(y) > 0$, hence P-almost-surely.

Proof. See Appendix B.1.8. ■

This result already explicitly characterizes the assignment component of the scores that belong to $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$, as the space of all functions $a : \mathcal{Z} \to \mathbb{R}$ with $\mathbb{E}_P[a(Z)] = 0$. Informally, with estimated P_Z,[27] the set of allowable first-order perturbations $\dot{P}_Z$ of the assignment probabilities at P under $\mathfrak{P}_{\mathbf{S}}$ is the set of all $\dot{P}_Z$ satisfying $\sum_{z\in\mathcal{Z}} \dot{P}_Z(z) = 0$. Therefore, we now focus on $\dot{\mathscr{C}}_{\mathbf{S}}^{P}$.

In order to provide a more concrete characterization of $\dot{\mathscr{C}}_{\mathbf{S}}^{P}$, we will work with the characterization of $\mathbf{\Pi}_d$ given in Lemma 3.3. Let $\mathbf{A}_d$ be as given by Lemma 3.3, with rows $\{\boldsymbol{a}_{d,1}, \ldots, \boldsymbol{a}_{d,k_d}\}$, and define $J_d(P)$ as the subset of rows of $\mathbf{A}_d$ that bind at P, i.e.

$$J_d^P := \left\{ j \,\middle|\, \boldsymbol{a}_{d,j}^{\intercal}\boldsymbol{\pi} = \int_{\mathcal{Y}} h_d(\boldsymbol{a}_{d,j}; \boldsymbol{p}_d(y))\, d\mu(y)\ \forall \boldsymbol{\pi} \in \mathbf{\Pi}(P) \right\}.$$

Next, let $\mathbf{A}_{d,J_d^P}$ be the submatrix of $\mathbf{A}_d$ consisting of the rows in J_d^P, and define, for $y \in \mathcal{Y}$, the pointwise face $\mathscr{F}_d^P(y)$ as

$$\mathscr{F}_d^P(y) := \left\{ \boldsymbol{x} \geq \mathbf{0} \,\middle|\, \begin{array}{l} \mathbf{R}_d\boldsymbol{x} = \boldsymbol{p}_d^P(y),\ \mathbf{A}_{d,J_d^P}\boldsymbol{x} = h_d\left(\mathbf{A}_{d,J_d^P}; \boldsymbol{p}_d(y)\right), \\ \boldsymbol{x}_{(\boldsymbol{\tau})} = 0\ \forall \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \text{ with } \max\limits_{\boldsymbol{\pi}\in\mathbf{\Pi}(P)} \boldsymbol{\pi}_{(\boldsymbol{\tau})} = 0. \end{array} \right\}.$$

The main utility of defining $\mathscr{F}_d^P(y)$ is that, informally, any allowable first-order perturbation $\dot{\boldsymbol{p}}_d^P$ of $\boldsymbol{p}_d^P$ must be supported by a "latent perturbation" $\dot{\boldsymbol{q}}_d = \left(\dot{q}_{d|\boldsymbol{\tau}}\right)_{\boldsymbol{\tau}\in\boldsymbol{\mathcal{T}}}$ that satisfies $\dot{q}_{d|\boldsymbol{\tau}}(y) = 0$ whenever $\sup_{\boldsymbol{x}\in\mathscr{F}_d^P(y)} \boldsymbol{x}_{(\boldsymbol{\tau})} = 0$. This must hold for all $d \in \mathcal{D}$ and these latent perturbations must be mutually compatible with each other and the restrictions on $\boldsymbol{\pi}$ that are binding at P, in the sense formalized in the next result.

Define the set of rows of $\mathbf{A}_{\mathbf{S}}$, the coefficient matrix in the definition of $\mathcal{R}_{\boldsymbol{\pi}}$ under Condition CP, corresponding to inequalities that hold with equality throughout $\mathbf{\Pi}(P)$,

$$J_{\mathbf{S}}^P := \left\{ j \,\middle|\, \mathbf{A}_{\mathbf{S},(j,\cdot)}^{\intercal}\boldsymbol{\pi} = \mathbf{a}_{\mathbf{S},(j)}\ \forall \boldsymbol{\pi} \in \mathbf{\Pi}(P) \right\},$$

and denote by $\mathbf{A}_{\mathbf{S},J_{\mathbf{S}}^P}$ the submatrix of $\mathbf{A}_{\mathbf{S}}$ consisting of the rows in $J_{\mathbf{S}}^P$. We now have the following explicit characterization of $\dot{\mathscr{C}}_{\mathbf{S}}^{P}$, and therefore of $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$.

Theorem 5.6. *Suppose* $\mathbf{S}$ *satisfies Condition CP and* $P \in \mathfrak{P}_{\mathbf{S}}$. *Then,*

$$\dot{\mathscr{C}}_{\mathbf{S}}^{P} = \left\{ (\mathbf{R}_d\dot{\boldsymbol{q}}_d)_{d\in\mathcal{D}} \in \mathbb{H}_P \,\middle|\, \begin{array}{l} (\dot{\boldsymbol{q}}_d)_{d\in\mathcal{D}} \in \prod\limits_{d\in\mathcal{D}} L^1(\mu)^{\boldsymbol{\mathcal{T}}},\ \dot{q}_{d|\boldsymbol{\tau}}(y) = 0\ \forall y : \sup\limits_{\boldsymbol{x}\in\mathscr{F}_d^P(y)} \boldsymbol{x}_{(\boldsymbol{\tau})} = 0, \\ \exists \dot{\boldsymbol{\pi}} \in \ker\left(\mathbf{A}_{\mathbf{S},J_{\mathbf{S}}^P}\right) : \forall d \in \mathcal{D},\ \int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d(y)\, d\mu(y) = \dot{\boldsymbol{\pi}}. \end{array} \right\}.$$

Proof. See Appendix B.1.9 ■

With this result in hand, we are now well-equipped to derive the efficient influence function. First, define $0 < \underline{p}_A^{\star}(P) \leq \overline{p}_A^{\star}(P) \leq 1$ as the endpoints of the interval of conditioning

[27] When assignment probabilities are known, then we must have $\dot{P}_Z = 0$.

probabilities that can support the sharp upper bound, i.e.

$$\begin{aligned}\left[\underline{p}_A^\star(P), \overline{p}_A^\star(P)\right] &:= \left\{\boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q \,\middle|\, Q \in \mathcal{Q}_{\mathbf{S}}(P) : \theta^Q = \bar{\theta}(P)\right\} \\ &= \left\{\boldsymbol{c}_A^\intercal \boldsymbol{\pi} \,\middle|\, (\boldsymbol{\eta}, \boldsymbol{\pi}) \in \Upsilon_d d^\star(P), \boldsymbol{\pi} \in \boldsymbol{\Pi}(P), \boldsymbol{c}_A^\intercal \boldsymbol{\eta} = \bar{\theta}(P)\, \boldsymbol{c}_A^\intercal \boldsymbol{\pi}\right\} .\end{aligned}$$

Next, for $\dot{\boldsymbol{p}} \in \dot{\mathscr{C}}_{\mathbf{S}}^P$, define

$$\delta_P(\dot{\boldsymbol{p}}) := \inf_{m \in \mathfrak{M}_P^\star} \sum_{d \in \mathcal{D}, z \in \mathcal{Z}} \int_{\mathcal{Y}} m(y,d,z)\, \dot{p}_{d|z}(y)\, d\mu(y) ,$$

which can be understood as the smallest possible change in the optimal moments in direction $\dot{\boldsymbol{p}}$; note that δ_P need not be linear, but is always finite, continuous and concave.

In order for an efficient influence function to exist, we need $\bar{\theta}(P)$ to be pathwise differentiable. In general, however, $\bar{\theta}(P)$ may only be directionally pathwise differentiable. The following lemma provides a characterization of the pathwise differentiability of $\bar{\theta}(P)$.

Lemma 5.4. *Suppose* $\mathbf{S}$ *satisfies Condition CP and* $P \in \mathfrak{P}_{\mathbf{S}}$*. Then, for every two-sided QMD path* $(P_t)_t$ *through* P *in* $\mathfrak{P}_{\mathbf{S}}$ *with score* s*,*

$$\lim_{t \downarrow 0} \frac{\bar{\theta}(P_t) - \bar{\theta}(P)}{t} = \max_{p \in \left[\underline{p}_A^\star(P), \overline{p}_A^\star(P)\right]} \frac{\delta_P\left(\dot{\boldsymbol{p}}^P(s)\right)}{p} = \begin{cases} \dfrac{\delta_P\left(\dot{\boldsymbol{p}}^P(s)\right)}{\underline{p}_A^\star(P)} \text{ if } \delta_P\left(\dot{\boldsymbol{p}}^P(s)\right) \geq 0 , \\ \dfrac{\delta_P\left(\dot{\boldsymbol{p}}^P(s)\right)}{\overline{p}_A^\star(P)} \text{ if } \delta_P\left(\dot{\boldsymbol{p}}^P(s)\right) < 0 . \end{cases}$$

Further, $\bar{\theta}(P)$ *is pathwise differentiable at* P *with respect to* $\dot{\mathfrak{P}}_{\mathbf{S}}^P$ *if and only if*

$$\delta_P\left(\dot{\boldsymbol{p}}^P(s)\right) = \begin{cases} \langle \dot{\boldsymbol{p}}^{P,\star}, \dot{\boldsymbol{p}}^P(s) \rangle_P & \text{if } \langle \dot{\boldsymbol{p}}^{P,\star}, \dot{\boldsymbol{p}}^P(s) \rangle_P \geq 0 , \\ \dfrac{\overline{p}_A^\star(P)}{\underline{p}_A^\star(P)} \langle \dot{\boldsymbol{p}}^{P,\star}, \dot{\boldsymbol{p}}^P(s) \rangle_P & \text{if } \langle \dot{\boldsymbol{p}}^{P,\star}, \dot{\boldsymbol{p}}^P(s) \rangle_P < 0 , \end{cases} \tag{5.13}$$

where $\dot{\boldsymbol{p}}^{P,\star}$ *is the unique solution to the strongly concave optimization problem*

$$\max_{\dot{\boldsymbol{p}} \in \dot{\mathscr{C}}_{\mathbf{S}}^P} \left\{ 2\delta_P(\dot{\boldsymbol{p}}) - \|\dot{\boldsymbol{p}}\|_P^2 \right\} . \tag{5.14}$$

Proof. See Appendix B.1.10 ■

We now have the following result.

Theorem 5.7 (Efficient Influence Function and Semiparametric Efficiency Bound)**.** *Suppose* $\mathbf{S}$ *satisfies Condition CP,* $P \in \mathfrak{P}_{\mathbf{S}}$ *and the condition* (5.13) *holds, so that* $\bar{\theta}(P)$ *is pathwise differentiable at* P *with respect to* $\dot{\mathfrak{P}}_{\mathbf{S}}$*. Then, with*

$$\dot{\boldsymbol{p}}^{P,\star} = \left(\left(\dot{p}_{d|z}^{P,\star}\right)_{z \in \mathcal{Z}}\right)_{d \in \mathcal{D}} = \operatorname*{argmax}_{\dot{\boldsymbol{p}} \in \dot{\mathscr{C}}_{\mathbf{S}}^P} \left\{ 2\delta_P(\dot{\boldsymbol{p}}) - \|\dot{\boldsymbol{p}}\|_P^2 \right\}$$

defined as in Lemma 5.4, the efficient influence function ϕ_P^* *of* $\bar{\theta}$ *at* P *is*

$$\phi_P^*(y,d,z) = \begin{cases} \dfrac{\dot{p}_{d|z}^{P,\star}}{\underline{p}_A^\star(P)\, p_{d|z}^P(y)} & \text{if } p_{d|z}^P(y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, whenever $\bar{\theta}$ *is pathwise differentiable, the semiparametric efficiency bound for estimating* $\bar{\theta}(P)$ *in the model* $\mathfrak{P}_{\mathbf{S}}$ *is*

$$\mathbb{E}_P\left[\phi_P^*(W)^2\right] = \frac{\|\dot{\boldsymbol{p}}^{P,\star}\|_P^2}{\underline{p}_A^\star(P)^2}\,.$$

Further, $\mathbb{E}_P[\phi_P^*(W) \mid Z = z] = 0$ *for all* $z \in \mathcal{Z}$.

Proof. See Appendix B.1.11 ■

Corollary 5.1. *Suppose* $\mathbf{S}$ *satisfies Condition CP,* $P \in \mathfrak{P}_{\mathbf{S}}$ *and Assumption 5.5 holds, so that* $\underline{p}_A^\star(P) = \overline{p}_A^\star(P) = p_A^\dagger(P)$. *Then, the following statements are equivalent:*

(i) $\bar{\theta}$ *is pathwise differentiable at* P *with respect to* $\dot{\mathfrak{P}}_{\mathbf{S}}$.
(ii) δ_P *is linear on* $\dot{\mathscr{C}}_{\mathbf{S}}^P$.
(iii) *For any* $m_1, m_2 \in \mathfrak{M}_P^\star$ *and* $\dot{\boldsymbol{p}} \in \dot{\mathscr{C}}_{\mathbf{S}}^P$,

$$\sum_{d\in\mathcal{D}, z\in\mathcal{Z}} \int_{\mathcal{Y}} (m_1(y,d,z) - m_2(y,d,z))\, \dot{p}_{d|z}(y)\, d\mu(y) = 0\,.$$

Suppose, in addition, that these equivalent statements hold. Then, for ***any*** $m \in \mathfrak{M}_P^*$, *the strongly concave problem (5.14), with unique solution* $\dot{\boldsymbol{p}}^{P,\star}$, *is equivalent to*

$$\max_{\dot{\boldsymbol{p}}\in\dot{\mathscr{C}}_{\mathbf{S}}^P} \left\{ 2 \sum_{d\in\mathcal{D}, z\in\mathcal{Z}} \int_{\mathcal{Y}} m(y,d,z)\, \dot{p}_{d|z}(y)\, d\mu(y) - \|\dot{\boldsymbol{p}}\|_P^2 \right\}\,,$$

and the EIF ϕ_P^* *and semiparametric efficiency bound in Theorem 5.7 reduce to*

$$\phi_P^*(y,d,z) = \begin{cases} \dfrac{\dot{p}_{d|z}^{P,\star}}{p_A^\dagger(P)\, p_{d|z}^P(y)} & \text{if } p_{d|z}^P(y) > 0, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \mathbb{E}_P\left[\phi_P^*(W)^2\right] = \frac{\|\dot{\boldsymbol{p}}^{P,\star}\|_P^2}{p_A^\dagger(P)^2}\,.$$

Proof. See Appendix B.1.12 ■

5.3.3. *Efficiency of Bias-corrected Estimator.* ,

For $d \in \mathcal{D}$ and $Q \in \mathcal{Q}_{\mathbf{S}}$, denote by $\mathbf{R}_d(y;Q)$ the matrix obtained from $\mathbf{R}_d$ by retaining only the columns that correspond to response-types $\boldsymbol{\tau}$ with $q_{d|\boldsymbol{\tau}}^Q(y) > 0$.

Assumption 5.8 (Weak Interiority)**.** *There exists* $Q^\dagger \in \mathcal{Q}_{\mathbf{S}}(P)$ *such that*

(i) $\boldsymbol{\pi}^{Q^\dagger} \in \text{relint}(\mathcal{R}_{\boldsymbol{\pi}})$,

(ii) *For every* $d \in \mathcal{D}$, $\int_{\mathcal{Y}} \min_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d} q^{Q^\dagger}_{d|\boldsymbol{\tau}}(y)\, d\mu(y) > 0$,

(iii) *For every* $d \in \mathcal{D}$ *and* $y \in \mathcal{Y}$,

$$\operatorname{col}\left(\mathbf{R}_d\left(y; Q^\dagger\right)\right) = \operatorname{col}\left(\mathbf{R}_d\right) \cap \left\{\boldsymbol{v} \in \mathbb{R}^{\mathcal{Z}} \,\middle|\, \boldsymbol{v}_{(z)} = 0 \forall z \in \mathcal{Z} : \boldsymbol{p}^P_{d|z}(y) = 0\right\} .$$

Denote by $E_{\mathbf{S}}$ the set of rows of $\mathbf{A}_{\mathbf{S}}$ corresponding to inequalities that hold with equality throughout $\mathcal{R}_{\boldsymbol{\pi}}$, i.e.

$$E_{\mathbf{S}} := \left\{j \,\middle|\, \mathbf{A}_{\mathbf{S};(j,\cdot)}\boldsymbol{\pi} = \mathbf{a}_{\mathbf{S};(j,\cdot)}\ \forall \boldsymbol{\pi} : \mathbf{A}_{\mathbf{S}}\boldsymbol{\pi} \leq \mathbf{a}_{\mathbf{S}}\right\} ;$$

$E_{\mathbf{S}}$ can be found using a linear programs over the known polytope $\mathcal{R}_{\boldsymbol{\pi}}$ (see Appendix E). Denote by $\mathbf{A}_{\mathbf{S},E}$ the submatrix of $\mathbf{A}_{\mathbf{S}}$ consisting of only the rows in $E_{\mathbf{S}}$, and note that $\operatorname{span}\left(\mathcal{R}_{\boldsymbol{\pi}} - \mathcal{R}_{\boldsymbol{\pi}}\right) = \ker\left(\mathbf{A}_{\mathbf{S},E}\right)$. Informally, we can now characterize the set of two-sided perturbations of a $\boldsymbol{\pi} \in \operatorname{relint}\left(\mathcal{R}_{\boldsymbol{\pi}}\right)$ that can support an allowed perturbation $\dot{\boldsymbol{p}}$ in terms of the globally-binding restrictions of $\mathbf{S}$. Precisely, given $\boldsymbol{\pi}^\dagger \in \operatorname{relint}\left(\mathcal{R}_{\boldsymbol{\pi}}\right)$, we have

$$\boldsymbol{\pi}^{Q^\dagger} + t\dot{\boldsymbol{\pi}} \in \mathcal{R}_{\boldsymbol{\pi}} \text{ for all sufficiently small } |t| \iff \mathbf{A}_{\mathbf{S},E}\dot{\boldsymbol{\pi}} = \mathbf{0} .$$

The relative interiority of $\boldsymbol{\pi}^{Q^\dagger}$ ensures that the remaining inequalities implied by $\mathbf{S}$ are locally slack and so impose no additional restrictions on $\dot{\boldsymbol{\pi}}$.

Lemma 5.5. *Suppose* $\mathbf{S}$ *satisfies Condition CP,* $P \in \mathfrak{P}_{\mathbf{S}}$*, and Assumption 5.8 holds. Then,*

$$\dot{\mathscr{C}}^P_{\mathbf{S}} = \left\{ \dot{\boldsymbol{p}} \in \mathbb{H}_P \,\middle|\, \begin{array}{l} \dot{\boldsymbol{p}}_d(y) \in \text{col-span}\left(\mathbf{R}_d\right)\ \forall d \in \mathcal{D}, \\ \exists \dot{\boldsymbol{\pi}} \in \ker\left(\mathbf{A}_{\mathbf{S},E}\right) : \forall d \in \mathcal{D},\ \int_{\mathcal{Y}} \dot{\boldsymbol{p}}_d(y)\ d\mu(y) = \mathbf{R}_d\dot{\boldsymbol{\pi}} . \end{array} \right\} .$$

Proof. See Appendix B.1.13 ■

Theorem 5.8 (Regularity and Efficiency of $\hat{\bar{\theta}}^{(\mathrm{CF})}_n$)**.** *Suppose the conditions of Theorem 5.5 hold. Then,* $\hat{\bar{\theta}}^{(\mathrm{CF})}_n$ *is regular at* P*, with influence function* ϕ_P*, as defined in (5.9).*

If Assumption 5.8 also holds, then $\hat{\bar{\theta}}^{(\mathrm{CF})}_n$ *is asymptotically efficient if and only if*

$$\left(p_{d|z}\phi_P(y,d,z)\right)_{z\in\mathcal{Z}} \in \text{col-span}\left(\mathbf{R}_d\right)\ \forall d \in \mathcal{D} \text{ and } y \in \mathcal{Y} , \tag{5.15}$$

and there exists $\dot{\boldsymbol{\pi}} \in \ker\left(\mathbf{A}_{\mathbf{S},E}\right)$ *such that*

$$\left(\mathbb{E}_P\left[\mathbb{1}\left\{D = d\right\}\phi_P(W) \,\middle|\, Z = z\right]\right)_z = \mathbf{R}_d\dot{\boldsymbol{\pi}}\ \forall d \in \mathcal{D} . \tag{5.16}$$

In such cases, $\phi^*_P = \phi_P$ *and*

$$\mathbb{E}_P\left[\phi^*_P(W)^2\right] = \frac{1}{p^\dagger_A(P)^2}\sum_{z\in\mathcal{Z}} \frac{\operatorname{Var}\left(m(W) \,\middle|\, Z = z\right)}{P_Z(z)} ,$$

for any $m \in \mathfrak{M}^\star_P$.

Further, (5.15) *and* (5.16) *hold whenever* $\mathbf{S}$ *satisfies*

$$\operatorname{rank}\begin{bmatrix}\mathbf{A}_{\mathbf{S},E}\\ \mathbf{R}_{d_1}\\ \vdots\\ \mathbf{R}_{d_{|\mathcal{D}|}}\end{bmatrix} - \operatorname{rank}\left(\mathbf{A}_{\mathbf{S},E}\right) = |\mathcal{Z}|\left(|\mathcal{D}|-1\right)\ , \tag{5.17}$$

and, in the case where $\mathbf{S}$ *simply restricts the support of* $\boldsymbol{T}$*,* (5.17) *reduces to*

$$\operatorname{rank}\begin{bmatrix}\mathbf{R}_{d_1}\\ \vdots\\ \mathbf{R}_{d_{|\mathcal{D}|}}\end{bmatrix} = 1 + |\mathcal{Z}|\left(|\mathcal{D}|-1\right)\ .$$

Proof. See Appendix B.1.14 ■

6. Conclusion

This paper presents new specification testing and partial identification results for instrumental variable models with multiple levels (or types) of treatment and discrete-valued, randomly assigned, instruments. The proposed framework allows for the partial identification of a broad class of causal parameters under many different kinds of assumptions about the treatment selection mechanism. Other than random assignment of the instrument, I do not require any additional restrictions on treatment selection, but allow the analyst to flexibly model selection into treatment. Using a novel geometric characterization of IV mixture models, this paper provides a characterization of the empirical content and testable implications of such models, and a a new characterization of the identified set for counterfactual choices, outcome means, and treatment effects under such models.

I use the aforementioned convex representation to derive a semiparametric efficiency bound for the estimation of sharp bounds on conditional mean parameters under such models, and propose a cross-fitted, bias-corrected estimator that attains the bound in suitably regular cases, while controlling inward error at the root-n rate in irregular cases. The convex moment characterization of the model specification test and parameter inclusion tests provided here enable statistical inference using existing moment inequality methods that do not require discretization of continuous outcomes, such as Chernozhukov et al. (2013, "CLR") and Andrews and Shi (2013). I also propose an adaptation of the procedure of CLR that, in this setting, asymptotically controls size uniformly over any class of data distributions with instrument assignment probabilities uniformly bounded away from zero, and is consistent against every fixed alternative. In particular, when the instrument assignment probabilities are known, the class contains *all* data distributions P with those assignment probabilities.

References

Andrews, D. W. K. and X. Shi (2013). Inference Based on Conditional Moment Inequalities. *Econometrica 81*(2), 609–666.

Angrist, J. D. and G. W. Imbens (1995, June). Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association 90*(430), 431–442.

Aumann, R. J. (1965, August). Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications 12*(1), 1–12.

Bai, Y., S. Huang, S. Moon, A. Santos, A. M. Shaikh, and E. J. Vytlacil (2026, September). Inference for Treatment Effects Conditional on Generalized Principal Strata using Instrumental Variables.

Bai, Y., S. Huang, and M. Tabord-Meehan (2025, February). Sharp Testable Implications of Encouragement Designs.

Balke, A. and J. Pearl (1994, January). Counterfactual Probabilities: Computational Methods, Bounds and Applications. In R. L. de Mantaras and D. Poole (Eds.), *Uncertainty Proceedings 1994*, pp. 46–54. San Francisco (CA): Morgan Kaufmann.

Balke, A. and J. Pearl (1997). Bounds on Treatment Effects From Studies With Imperfect Compliance. *Journal of the American Statistical Association 92*(439), 1171–1176.

Beresteanu, A., I. Molchanov, and F. Molinari (2011). Sharp Identification Regions in Models with Convex Moment Predictions. *Econometrica 79*(6), 1785–1821.

Beresteanu, A., I. Molchanov, and F. Molinari (2012, January). Partial identification using random set theory. *Journal of Econometrics 166*(1), 17–32.

Beresteanu, A. and F. Molinari (2008). Asymptotic Properties for a Class of Partially Identified Models. *Econometrica 76*(4), 763–814.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer.

Billera, L. J. and B. Sturmfels (1992). Fiber Polytopes. *Annals of Mathematics 135*(3), 527–549.

Boyd, S. P. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge, UK ; New York: Cambridge University Press.

Cheng, J. and D. S. Small (2006, November). Bounds on Causal Effects in Three-Arm Trials With Non-Compliance. *Journal of the Royal Statistical Society Series B: Statistical Methodology 68*(5), 815–836.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection Bounds: Estimation and Inference. *Econometrica 81*(2), 667–737.

Cross, P. J. and C. F. Manski (2002). Regressions, Short and Long. *Econometrica 70*(1), 357–368.

De Loera, J. A., J. Rambau, and F. Santos (2010). *Triangulations*, Volume 25 of *Algorithms and Computation in Mathematics*. Berlin, Heidelberg: Springer.

Fang, Z., A. Santos, A. M. Shaikh, and A. Torgovitsky (2023). Inference for Large-Scale Linear Systems With Known Coefficients. *Econometrica 91*(1), 299–327.

Frangakis, C. E. and D. B. Rubin (2002, March). Principal Stratification in Causal Inference. *Biometrics 58*(1), 21–29.

Goff, L. (2024). When does IV identification not restrict outcomes?

Gu, J. and T. M. Russell (2022, June). Partial identification in nonseparable binary response models with endogenous regressors. *Journal of Econometrics*.

Haile, P. A. and E. Tamer (2003, February). Inference with an Incomplete Model of English Auctions. *Journal of Political Economy 111*(1), 1–51.

Heckman, J. J. and R. Pinto (2018). Unordered Monotonicity. *Econometrica 86*(1), 1–35.

Heckman, J. J., J. Smith, and N. Clements (1997). Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *The Review of Economic Studies 64*(4), 487–535.

Heckman, J. J. and J. A. Smith (1995). Assessing the Case for Social Experiments. *The Journal of Economic Perspectives 9*(2), 85–110.

Heckman, J. J. and J. A. Smith (1998, May). Evaluating the Welfare State.

Heckman, J. J., S. Urzua, and E. Vytlacil (2008). Instrumental Variables in Models with Multiple Outcomes: The General Unordered Case. *Annales d'Économie et de Statistique* (91/92), 151–174.

Heckman, J. J. and E. Vytlacil (2001). Policy-Relevant Treatment Effects. *American Economic Review 91*(2), 107–111.

Heckman, J. J. and E. J. Vytlacil (1999, April). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences of the United States of America 96*(8), 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2007a, January). Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, pp. 4779–4874. Elsevier.

Heckman, J. J. and E. J. Vytlacil (2007b, January). Chapter 71 Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, pp. 4875–5143. Elsevier.

Henry, M., Y. Kitamura, and B. Salanié (2014). Partial identification of finite mixtures in econometric models. *Quantitative Economics 5*(1), 123–144.

Horowitz, J. L. and C. F. Manski (1995). Identification and Robustness with Contaminated and Corrupted Data. *Econometrica 63*(2), 281–302.

Horowitz, J. L. and C. F. Manski (1998, May). Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations. *Journal of Econometrics 84*(1), 37–58.

Horowitz, J. L. and C. F. Manski (2000). Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data. *Journal of the American Statistical Association 95*(449), 77–84.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and Estimation of Local Average Treatment Effects. *Econometrica : journal of the Econometric Society 62*(2), 467.

Jun, S. J. and S. Lee (2023, February). Identifying the Effect of Persuasion. *Journal of Political Economy*.

Kaido, H. and K. Ponomarev (2025, December). Testing Exclusion and Shape Restrictions in Potential Outcomes Models.

Kaji, T. and J. Cao (2023, June). Assessing Heterogeneity of Treatment Effects. http://arxiv.org/abs/2306.15048.

Kédagni, D. and I. Mourifié (2020, September). Generalized instrumental inequalities: Testing the instrumental variable independence assumption. *Biometrika 107*(3), 661–675.

Kelley, C. T. (2003). *Solving Nonlinear Equations with Newton's Method*. Fundamentals of Algorithms. Philadelphia: Society for Industrial and Applied Mathematics.

Kitagawa, T. (2015). A Test for Instrument Validity. *Econometrica 83*(5), 2043–2063.

Kitagawa, T. (2021, December). The identification region of the potential outcome distributions under instrument independence. *Journal of Econometrics 225*(2), 231–253.

Kolesár, M., J. L. M. Olea, and J. Roth (2026, March). Evaluating Counterfactual Policies Using Instruments.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics Ser. [New York, NY: Springer.

Kreider, B. and J. V. Pepper (2007, June). Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors. *Journal of the American Statistical Association 102*(478), 432–441.

Kroft, K., I. Mourifié, and A. Vayalinkal (2025, February). Horowitz-Manski-Lee Bounds with Multilayered Sample Selection.

Lee, D. S. (2009, July). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies 76*(3), 1071–1102.

Lee, S. and B. Salanié (2018). Identifying Effects of Multivalued Treatments. *Econometrica 86*(6), 1939–1963.

Lee, S. and B. Salanié (2023, November). Treatment Effects with Targeting Instruments.

Manski, C. F. (1990). Nonparametric Bounds on Treatment Effects. *The American Economic Review 80*(2), 319–323.

Manski, C. F. (2003, May). *Partial Identification of Probability Distributions*. Springer Science & Business Media.

Manski, C. F. and J. V. Pepper (2000). Monotone Instrumental Variables: With an Application to the Returns to Schooling. *Econometrica 68*(4), 997–1010.

Manski, C. F. and J. V. Pepper (2009). More on monotone instrumental variables. *The Econometrics Journal 12*(S1), S200–S216.

Marx, P. (2024, January). Sharp bounds in the latent index selection model. *Journal of Econometrics 238*(2), 105561.

Matousek, J. and B. Gärtner (2007, July). *Understanding and Using Linear Programming*. Springer Science & Business Media.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica 86*(5), 1589–1619.

Mogstad, M. and A. Torgovitsky (2018). Identification and Extrapolation of Causal Effects with Instrumental Variables. *Annual Review of Economics 10*(1), 577–613.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2021, November). The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables. *American Economic Review 111*(11), 3663–3698.

Mountjoy, J. (2022, August). Community Colleges and Upward Mobility. *American Economic Review 112*(8), 2580–2630.

Mourifié, I., M. Henry, and R. Méango (2020, August). Sharp Bounds and Testability of a Roy Model of STEM Major Choices. *Journal of Political Economy 128*(8), 3220–3283.

Mourifié, I. and Y. Wan (2017, May). Testing Local Average Treatment Effect Assumptions. *The Review of Economics and Statistics 99*(2), 305–313.

Navjeevan, M., R. Pinto, and A. Santos (2023, October). Identification and Estimation in a Class of Potential Outcomes Models.

Noack, C. (2021, June). Sensitivity of LATE Estimates to Violations of the Monotonicity Assumption. https://arxiv.org/abs/2106.06421v1.

Rockafellar, R. T. (2015). *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton: Princeton University Press.

Rockafellar, R. T. and R. J. B. Wets (1998). *Variational Analysis*, Volume 317 of *Grundlehren Der Mathematischen Wissenschaften*. Berlin, Heidelberg: Springer.

Rudin, W. (1987). *Real and Complex Analysis* (3rd ed ed.). New York: McGraw-Hill.

Schaible, S. (1995). Fractional Programming. In R. Horst and P. M. Pardalos (Eds.), *Handbook of Global Optimization*, pp. 495–608. Boston, MA: Springer US.

Sun, Z. (2023, October). Instrument Validity for Heterogeneous Causal Effects. http://arxiv.org/abs/2009.01995.

Van Der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK ; New York, NY, USA: Cambridge University Press.

Van Der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York, NY: Springer.

Appendices

Unless otherwise specified, $y, d, z, \boldsymbol{\tau}$ range over $\mathcal{Y}, \mathcal{D}, \mathcal{Z}, \boldsymbol{\mathcal{T}}$, respectively. These ranges are implicit whenever the domains of quantification, summation or integration are omitted. All optimization variables without an explicitly specified domain should be understood to have their domain be the set of real-valued vectors of conformable dimension.

Throughout the appendices, θ is one of the three types of parameters considered in Section 2.2. Fix a representation of θ using measurable functions $\kappa_d, \varkappa_d : \mathcal{Y} \to \mathbb{R}$ and fixed vectors $\boldsymbol{c}_d, \boldsymbol{a}_d, \boldsymbol{c}_\pi, \boldsymbol{a}_\pi \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$. We can now represent all three types of θ in the following form,

$$\theta^Q = \frac{\mathsf{N}^Q}{\mathsf{D}^Q}. \tag{$\dagger$}$$

where

$$\mathsf{N}^Q := \boldsymbol{c}_\pi^\intercal \boldsymbol{\pi}^Q + \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} \kappa_d(y) \boldsymbol{c}_d^\intercal \boldsymbol{q}_d^Q(y)\, d\mu(y)\ ,$$

$$\mathsf{D}^Q := \boldsymbol{a}_\pi^\intercal \boldsymbol{\pi}^Q + \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} \varkappa_d(y) \boldsymbol{a}_d^\intercal \boldsymbol{q}_d^Q(y)\, d\mu(y)\ .$$

The following condition rules out θ being an effects ratio.

Condition T1. *θ is a parameter of the form considered in Section 2.2, with the affine ratio representation given in (†), with $\boldsymbol{a}_d = \mathbf{0}$ for every $d \in \mathcal{D}$.*

The following condition allows θ to be an effects ratio, but does not allow κ_d and $\varkappa_d$ to differ if the d-relevant coefficients put weight on d-Never-Taker types.

Condition T2. *θ is a parameter of the form considered in Section 2.2, with the affine ratio representation given in (†), such that, for each $d \in \mathcal{D}$, at least one of the following holds:*

(i) $\kappa_d = \varkappa_d$;
(ii) $\boldsymbol{a}_{d;(\mathcal{T}\setminus\mathcal{T}_d)} = \mathbf{0}$.

The following condition requires that θ be defined using bounded loadings.

Condition T3. *θ is a parameter of the form considered in Section 2.2, with the affine ratio representation given in (†), with*

$$\max_{d\in\mathcal{D}} \operatorname*{ess\,sup}_{y\in\mathcal{Y}} \left(|\kappa_d(y)| \|\boldsymbol{c}_d\|_\infty + |\varkappa_d(y)| \|\boldsymbol{a}_d\|_\infty\right) < \infty,$$

where the essential supremum is with respect to μ.

Appendix A: Main Identification Results

In the remainder, we make extensive use of Aumann's (1965) definition of integrals of set-valued functions. Since we apply them to maps derived from polyhedral correspondences and rely heavily on generalizations of results developed in Billera and Sturmfels (1992), I follow their terminology and refer to these integrals as "Minkowski integrals".

Definition A.1 (Minkowski Integral)**.** *For a correspondence $F : \mathcal{Y} \rightrightarrows \mathbb{R}^k$ and measurable $A \subseteq \mathcal{Y}$, define the **Minkowski integral** of F as*

$$\int_A F(y)\, d\mu(y) := \left\{ \int_A \boldsymbol{x}\, d\mu \,\middle|\, \boldsymbol{x} \in L^1(\mu;\mathbb{R}^k),\ \boldsymbol{x}(y) \in F(y)\ \forall y \in A \right\}.$$

The definition above is more general than the one in Billera and Sturmfels (1992), since we need this notion to apply to a broader class of maps than the integrably bounded "polytope bundles" considered in Billera and Sturmfels (1992).[28]

[28] The formulation here applies to deterministic correspondences on $\mathcal{Y}$. For a random set on a probability space, the same construction gives its Aumann expectation; see, Definition 5 in Beresteanu and Molinari (2008).

A.1. **Proofs of Identification Results in Sections 2 and 3**

A.1.1. *Proof of Theorem 2.1.*

Proof. "$\subseteq$" follows from the definition of $\mathcal{R}_{\boldsymbol{\pi}}$ and so it suffices to show "$\supseteq$".

To this end, let $Q \in \mathcal{Q}(P)$ with $\boldsymbol{\pi}^Q \in \mathcal{R}_{\boldsymbol{\pi}}$ be given. Then, there exists $(F_{\boldsymbol{V}}, g_D) \in \mathbf{S}$ that induces $\boldsymbol{\pi}^Q$. Define mutually independent random variables $(\boldsymbol{e}(\boldsymbol{\tau}) : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}})$, $\boldsymbol{V} \sim F_{\boldsymbol{V}}$, $Z \sim P_Z$, where, for each $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$, $\boldsymbol{e}(\boldsymbol{\tau})$ is defined as a random vector with distribution $Q_{(Y_d)_{d\in\mathcal{D}}|\boldsymbol{T}=\boldsymbol{\tau}}$. Define $\boldsymbol{\varepsilon}_Y := (\boldsymbol{e}(\boldsymbol{\tau}))_{\boldsymbol{\tau}\in\boldsymbol{\mathcal{T}}}$ and $g_Y(d, \boldsymbol{v}, \boldsymbol{e}') := \boldsymbol{e}(\boldsymbol{T}(\boldsymbol{v}))_{(d)}$. When $\boldsymbol{\pi}^Q_{(\boldsymbol{\tau})} = 0$, take the components of $\boldsymbol{e}(\boldsymbol{\tau})$ to be mutually independent with common law P_Y, which does not affect the observed data. Finally, define $D = g_D(Z, \boldsymbol{V})$ and $Y = g_Y(D, \boldsymbol{V}, \boldsymbol{\varepsilon}_Y)$. By construction $(Z, \boldsymbol{V}, \boldsymbol{\varepsilon}_Y; g_D, g_Y)$ satisfies the independence condition, generates Q and satisfies $\mathbf{S}$, as required. ■

A.1.2. *Proof of Lemma 3.1.*

Proof. First, note that the random assignment of Z implies that for any $Q \in \mathcal{M}$ with $Q_Z = P_Z$, for any $d \in \mathcal{D}, z \in \mathcal{Z}$ and Borel $B \subseteq \mathcal{Y}$, we have

$$P_Q(Y \in B, D = d, Z = z) = Q(Y_d \in B, D_z = d)\, Q(Z = z) \ , \tag{A.1}$$

and since $Q_Z = P_Z$, we have[29]

$$\begin{aligned} P_Q(Y \in B, D = d | Z = z) &= Q(Y_d \in B, D_z = d) \\ &= \sum_{\boldsymbol{\tau}\in\boldsymbol{\mathcal{T}}:\boldsymbol{\tau}(z)=d} Q(Y_d \in B, \boldsymbol{T} = \boldsymbol{\tau}) = \left(\mathbf{R}_d \boldsymbol{Q}^Q_d(B)\right)_{(z)} \ . \end{aligned}$$

The "$\Longrightarrow$" direction now follows immediately by noting that $Q \in \mathcal{Q}_{\mathbf{S}}$ implies (i) $\boldsymbol{\pi}^Q \in \mathcal{R}_{\boldsymbol{\pi}}$ by Theorem 2.1 and (ii) $P_Q = P$ and then stacking both sides over $z \in \mathcal{Z}$. The other direction follows by noting that the right hand side of (A.1) determines P_Q fully and so, if Q satisfies it with P in place of P_Q, then $P_Q = P$ and so, $Q \in \mathcal{Q}$, and then applying Theorem 2.1. ■

A.1.3. *Proof of Lemma 3.2.*

Proof. By Lemma A.7, the map $y \mapsto \boldsymbol{\ell}^\star_d\left(\boldsymbol{c}(y); \boldsymbol{p}^P_d(y)\right)$ is measurable and, for all $y \in \mathcal{Y}$,

$$\boldsymbol{\ell}^\star_d\left(\boldsymbol{c}(y); \boldsymbol{p}^P_d(y)\right) \in \underset{\substack{\boldsymbol{\ell}\in\mathbb{R}^{\mathcal{Z}}_{\geq 0} \\ \mathbf{R}_d^\intercal \boldsymbol{\ell} \geq \boldsymbol{c}(y) \\ \|\boldsymbol{\ell}\|_\infty \leq \|\boldsymbol{c}(y)\|_\infty}}{\operatorname{argmin}} \boldsymbol{\ell}^\intercal \boldsymbol{p}^P_d(y) \ .$$

The norm bound gives

$$\mathbb{E}_P\left[\left\|\boldsymbol{\ell}^\star_d\left(\boldsymbol{c}(Y); \boldsymbol{p}^P_d(Y)\right)\right\|_\infty\right] \leq \mathbb{E}_P\left[\|\boldsymbol{c}(Y)\|_\infty\right] < \infty \ ,$$

[29] For both implications, we are free to assume WLOG that $\boldsymbol{\pi}^Q \in \mathcal{R}_{\boldsymbol{\pi}}$, since this is necessary by Theorem 2.1 and is imposed in the converse. This avoids irritants related to the appropriate index for $\boldsymbol{\tau}$-indexed objects etc., and allows us to take the sum over $\boldsymbol{\mathcal{T}}$.

and so this map is P-integrable. Let $p_{\min} := \min_{z\in\mathcal{Z}} P_Z(z) > 0$. By the definition of $\boldsymbol{m}_d^P$ and the pointwise value equality in Lemma A.7,

$$\begin{aligned}\sum_{z\in\mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_d^P\left(\boldsymbol{c}(Y);W\right)\,\middle|\,Z=z\right] &= \int_{\mathcal{Y}} \boldsymbol{\ell}_d^\star\left(\boldsymbol{c}(y);\boldsymbol{p}_d^P(y)\right)^{\mathsf{T}}\boldsymbol{p}_d^P(y)\,d\mu(y)\\ &= \int_{\mathcal{Y}} h_d\left(\boldsymbol{c}(y);\boldsymbol{p}_d^P(y)\right)\,d\mu(y)\\ &\le \int_{\mathcal{Y}} \|\boldsymbol{c}(y)\|_\infty \sum_{z\in\mathcal{Z}} p_{d|z}^P(y)\,d\mu(y)\\ &\le \frac{1}{p_{\min}}\mathbb{E}_P\left[\|\boldsymbol{c}(Y)\|_\infty\right] < \infty\,. \end{aligned}$$

■

A.1.4. *Proof of Lemma 3.3.*

Proof. Fix $d\in\mathcal{D}$ and a coefficient $\boldsymbol{c}$ satisfying the hypothesis. We first show that maximizing separately at each y gives the support function of $\boldsymbol{\Pi}_d$.

Lemma A.2 supplies an integrable pointwise maximizer $\boldsymbol{x}^\star$, and $\boldsymbol{\pi}^\star := \int_{\mathcal{Y}} \boldsymbol{x}^\star\,d\mu$ belongs to $\boldsymbol{\Pi}_d$. For any $\boldsymbol{\pi}\in\boldsymbol{\Pi}_d$, choose an integrable selection $\boldsymbol{x}(y)\in\mathrm{Fibre}_d(\boldsymbol{p}_d^P(y))$ with $\int_{\mathcal{Y}}\boldsymbol{x}\,d\mu = \boldsymbol{\pi}$. Then

$$\begin{aligned}\boldsymbol{c}^{\mathsf{T}}\boldsymbol{\pi} &= \int_{\mathcal{Y}} \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x}(y)\,d\mu(y)\\ &\le \int_{\mathcal{Y}} h_d\left(\boldsymbol{c};\boldsymbol{p}_d^P(y)\right)\,d\mu(y)\\ &= \int_{\mathcal{Y}} \boldsymbol{c}^{\mathsf{T}}\boldsymbol{x}^\star(y)\,d\mu(y) = \boldsymbol{c}^{\mathsf{T}}\boldsymbol{\pi}^\star\,.\end{aligned}$$

Thus the integrated pointwise maximum is attained in $\boldsymbol{\Pi}_d$, proving the first equality. Lemma 3.2 gives the moment representation:

$$\begin{aligned}h_{\boldsymbol{\Pi}_d}(\boldsymbol{c}) &= \int_{\mathcal{Y}} h_d\left(\boldsymbol{c};\boldsymbol{p}_d^P(y)\right)\,d\mu(y)\\ &= \sum_{z\in\mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_d^P\left(\boldsymbol{c};W\right)\,\middle|\,Z=z\right]\,.\end{aligned}$$

Next, we show why finitely many support inequalities suffice, with directions that depend only on $\mathbf{R}_d$. Let $\mathcal{B} := \mathcal{B}_d(\boldsymbol{p}_d^P)$ be the partition from Corollary A.1, and define

$$\boldsymbol{r}_B := \int_B \boldsymbol{p}_d^P\,d\mu, \qquad C_B := \left(\mathrm{Fibre}_d(\boldsymbol{r}_B)\right)_{(\boldsymbol{\tau}_d)}, \qquad K := \sum_{B\in\mathcal{B}} C_B\,.$$

By that corollary, K is a nonempty compact convex polytope and

$$\boldsymbol{\Pi}_d = \left\{\boldsymbol{x}\ge\mathbf{0}\,\middle|\,\boldsymbol{x}_{(\boldsymbol{\tau}_d)}\in K\right\}\,.$$

Thus it suffices to describe K; the remaining coordinates of $\boldsymbol{x}$ are unrestricted beyond nonnegativity.

Take $\mathscr{N}_d$, $\boldsymbol{a}_{d,j}$, and J_Γ from Lemma A.4. For each $\Gamma \in \mathscr{N}_d$, that lemma supplies $\boldsymbol{v}_{\Gamma,\boldsymbol{r}_B} \in C_B$ maximizing $\boldsymbol{a}^\intercal \boldsymbol{x}$ over C_B simultaneously for all $\boldsymbol{a} \in \Gamma$. Therefore, with

$$\boldsymbol{v}_\Gamma := \sum_{B \in \mathcal{B}} \boldsymbol{v}_{\Gamma,\boldsymbol{r}_B} \in K \ ,$$

the definition of a Minkowski sum gives

$$h_K(\boldsymbol{a}) = \sum_{B \in \mathcal{B}} h_{C_B}(\boldsymbol{a}) = \boldsymbol{a}^\intercal \boldsymbol{v}_\Gamma \qquad \forall \boldsymbol{a} \in \Gamma \ .$$

In particular, h_K is linear on each cone Γ.

Every $\boldsymbol{u} \in K$ satisfies $\boldsymbol{a}_{d,j}^\intercal \boldsymbol{u} \le h_K(\boldsymbol{a}_{d,j})$ for all $j \in [k_d]$. Conversely, suppose $\boldsymbol{u} \in \mathbb{R}^{\boldsymbol{\mathcal{T}_d}} \setminus K$. The sets K and $\{\boldsymbol{u}\}$ are disjoint, nonempty, compact and convex subsets of $\mathbb{R}^{\boldsymbol{\mathcal{T}_d}}$, and so the strict hyperplane separation theorem, applied to these two sets, supplies $\boldsymbol{a}$ with

$$\boldsymbol{a}^\intercal \boldsymbol{u} > \max_{\boldsymbol{v} \in K} \boldsymbol{a}^\intercal \boldsymbol{v} = h_K(\boldsymbol{a}) \ .$$

Completeness of $\mathscr{N}_d$ places $\boldsymbol{a}$ in some $\Gamma \in \mathscr{N}_d$, where $\boldsymbol{a} = \sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_{d,j}$ for coefficients $\alpha_j \ge 0$. If all the proposed inequalities held at $\boldsymbol{u}$, linearity on Γ would give

$$\begin{aligned} \boldsymbol{a}^\intercal \boldsymbol{u} &= \sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_{d,j}^\intercal \boldsymbol{u} \\ &\le \sum_{j \in J_\Gamma} \alpha_j h_K(\boldsymbol{a}_{d,j}) = h_K(\boldsymbol{a}) \ , \end{aligned}$$

contradicting the strict separation. Therefore,

$$K = \left\{ \boldsymbol{u} \in \mathbb{R}^{\boldsymbol{\mathcal{T}_d}} \,\middle|\, \boldsymbol{a}_{d,j}^\intercal \boldsymbol{u} \le h_K(\boldsymbol{a}_{d,j}) \ \forall j \in [k_d] \right\} \ .$$

To express these inequalities in the full coordinates, define

$$\mathbf{A}_{d;(j,\boldsymbol{\mathcal{T}_d})} := \boldsymbol{a}_{d,j}^\intercal, \qquad \mathbf{A}_{d;(j,\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}_d})} := \mathbf{0}^\intercal \qquad \forall j \in [k_d] \ .$$

This matrix depends only on $\mathbf{R}_d$. Moreover,

$$h_K(\boldsymbol{a}_{d,j}) = h_{\boldsymbol{\Pi}_d}\left(\mathbf{A}_{d;(j,\cdot)}^\intercal\right) \ ,$$

and the first part of the proof expresses each of these support values as the required integral and moment. Substituting them into the finite inequalities for K gives

$$\begin{aligned} \boldsymbol{\Pi}_d &= \left\{ \boldsymbol{x} \ge \mathbf{0} \,\middle|\, \mathbf{A}_d \boldsymbol{x} \le \int_{\mathcal{Y}} h_d\left(\mathbf{A}_d; \boldsymbol{p}_d^P(y)\right) \, d\mu(y) \right\} \\ &= \left\{ \boldsymbol{x} \ge \mathbf{0} \,\middle|\, \mathbf{A}_d \boldsymbol{x} \le \sum_{z \in \mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_d^P(\mathbf{A}_d; W) \,\middle|\, Z = z \right] \right\} \ . \end{aligned}$$

■

A.1.5. *Proof of Theorem 3.1.*

Proof. For "$\subseteq$", let $Q \in \mathcal{Q}_{\mathbf{S}}$. By Lemma 3.1,

$$\boldsymbol{\pi}^Q \in \mathcal{R}_{\boldsymbol{\pi}}, \qquad \mathbf{R}_d \boldsymbol{q}_d^Q(y) = \boldsymbol{p}_d^P(y), \qquad \int_{\mathcal{Y}} \boldsymbol{q}_d^Q \, d\mu = \boldsymbol{\pi}^Q \quad \forall d \in \mathcal{D} \, .$$

Therefore, $\boldsymbol{q}_d^Q(y) \in \mathrm{Fibre}_d(\boldsymbol{p}_d^P(y)) \ \forall y \in \mathcal{Y}$ and is integrable with $\int_{\mathcal{Y}} \boldsymbol{q}_d^Q \, d\mu = \boldsymbol{\pi}^Q$. Therefore, $\boldsymbol{\pi}^Q \in \boldsymbol{\Pi}_d$ for every $d \in \mathcal{D}$, which gives $\boldsymbol{\Pi} \subseteq \mathcal{R}_{\boldsymbol{\pi}} \cap \bigcap_{d \in \mathcal{D}} \boldsymbol{\Pi}_d$.

For "$\supseteq$", fix $\boldsymbol{\pi} \in \mathcal{R}_{\boldsymbol{\pi}} \cap \bigcap_{d \in \mathcal{D}} \boldsymbol{\Pi}_d$. For each $d \in \mathcal{D}$ there exists, by definition of $\boldsymbol{\Pi}_d$, $\boldsymbol{q}_d = (q_{d|\boldsymbol{\tau}})_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}} \in L^1(\mu; \mathbb{R}_{\geq 0}^{\boldsymbol{\mathcal{T}}})$ with

$$\int_{\mathcal{Y}} \boldsymbol{q}_d \, d\mu = \boldsymbol{\pi}, \qquad \mathbf{R}_d \boldsymbol{q}_d(y) \leq \boldsymbol{p}_d^P(y) \, .$$

We will now construct a parent law that preserves these marginal sub-densities.

First, by the argument already given in the main text, $\mathbf{R}_d \boldsymbol{q}_d(y) \leq \boldsymbol{p}_d^P(y)$ must hold with equality at this $\boldsymbol{q}_d$: for each $z \in \mathcal{Z}$,

$$\begin{aligned} \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} \left[p_{d|z}^P(y) - (\mathbf{R}_d \boldsymbol{q}_d(y))_{(z)} \right] d\mu(y) &= 1 - \sum_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}} \boldsymbol{\pi}_{(\boldsymbol{\tau})} \sum_{d \in \mathcal{D}} \mathbb{1}\{\boldsymbol{\tau}(z) = d\} \\ &= 1 - \mathbf{1}^{\intercal} \boldsymbol{\pi} = 0 \, . \end{aligned}$$

Every integrand is nonnegative, and so $\mathbf{R}_d \boldsymbol{q}_d(y) = \boldsymbol{p}_d^P(y)$ for every $d \in \mathcal{D}$.

Next, whenever $\boldsymbol{\pi}_{(\boldsymbol{\tau})} > 0$, the function $q_{d|\boldsymbol{\tau}} / \boldsymbol{\pi}_{(\boldsymbol{\tau})}$ is a probability density because $\int_{\mathcal{Y}} q_{d|\boldsymbol{\tau}} \, d\mu = \boldsymbol{\pi}_{(\boldsymbol{\tau})}$. Take $\boldsymbol{T}$ with probabilities $\boldsymbol{\pi}$ and, conditional on $\boldsymbol{T} = \boldsymbol{\tau}$, take the Y_d's to be mutually independent with these respective densities. Take $Z \sim P_Z$ independently of $((Y_d)_{d \in \mathcal{D}}, \boldsymbol{T})$, and denote the resulting joint law by

$$Q := \mathrm{Law}\left((Y_d)_{d \in \mathcal{D}}, \boldsymbol{T}, Z\right) \, . \tag{A.2}$$

This construction gives $Q \in \mathcal{M}$, $Q_Z = P_Z$, $\boldsymbol{\pi}^Q = \boldsymbol{\pi}$, and

$$Q\left(Y_d \in B, \boldsymbol{T} = \boldsymbol{\tau}\right) = \int_B q_{d|\boldsymbol{\tau}} \, d\mu \quad \forall d \in \mathcal{D}, \ \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}, \ B \in \mathscr{B}_{\mathcal{Y}} \, .$$

In particular,

$$\begin{aligned} \mathbf{R}_d \boldsymbol{Q}_d^Q(B) &= \int_B \mathbf{R}_d \boldsymbol{q}_d \, d\mu \\ &= \int_B \boldsymbol{p}_d^P \, d\mu = \boldsymbol{P}_d^P(B) \, . \end{aligned}$$

Together with $\boldsymbol{\pi}^Q \in \mathcal{R}_{\boldsymbol{\pi}}$ and $Q_Z = P_Z$, this implies $Q \in \mathcal{Q}_{\mathbf{S}}$, by Lemma 3.1, which also gives $\boldsymbol{\Pi} = \left\{ \boldsymbol{\pi}^Q \,\middle|\, Q \in \mathcal{Q}_{\mathbf{S}} \right\} \neq \emptyset$.

Finally, Lemma 3.3 makes every $\boldsymbol{\Pi}_d$ a closed convex polyhedron. Under Condition C, their intersection with $\mathcal{R}_{\boldsymbol{\pi}}$ is closed and convex, and it is bounded because

$$\mathcal{R}_{\boldsymbol{\pi}} \subseteq \left\{ \boldsymbol{\pi} \geq \mathbf{0} \,\middle|\, \mathbf{1}^{\intercal} \boldsymbol{\pi} = 1 \right\} \, ,$$

and therefore compact. Under Condition CP, $\mathcal{R}_{\boldsymbol{\pi}}$ also has a finite linear-inequality representation, and so the same intersection is a compact convex polytope. ■

A.1.6. *Proof of Theorem 3.2.*

Proof. Let

$$\boldsymbol{b}_{\Pi}(P) := \sum_{z \in \mathcal{Z}} \mathbb{E}_P \left[\boldsymbol{m}_{\boldsymbol{\Pi}}(W) \mid Z = z \right].$$

By Theorem 3.1 and lemma 3.3,

$$\boldsymbol{\Pi}(P) = \left\{ \boldsymbol{\pi} \geq \mathbf{0} \,|\, \mathbf{A}_{\mathbf{S}} \boldsymbol{\pi} \leq \mathbf{a}_{\mathbf{S}},\ \mathbf{A}_{\boldsymbol{\Pi}} \boldsymbol{\pi} \leq \boldsymbol{b}_{\Pi}(P) \right\}.$$

This is a convex polytope, and it is compact since it is contained in $\mathcal{R}_{\boldsymbol{\pi}}$.

Fix $\boldsymbol{c} \in \mathbb{R}^{\mathcal{T}}$. The linear program maximizing $\boldsymbol{c}^{\intercal} \boldsymbol{\pi}$ over the nonempty compact set $\mathcal{R}_{\boldsymbol{\pi}}$ has a dual solution $\boldsymbol{\lambda}^0_{\mathbf{S}} \geq \mathbf{0}$ satisfying $\mathbf{A}_{\mathbf{S}}^{\intercal} \boldsymbol{\lambda}^0_{\mathbf{S}} \geq \boldsymbol{c}$ and so, $(\boldsymbol{\lambda}^0_{\mathbf{S}}, \mathbf{0}) \in \Lambda(\boldsymbol{c})$.

If $\boldsymbol{\Pi}(P) \neq \emptyset$, strong duality gives

$$h_{\boldsymbol{\Pi}(P)}(\boldsymbol{c}) = \min_{(\boldsymbol{\lambda}_{\mathbf{S}}, \boldsymbol{\lambda}_{\boldsymbol{\Pi}}) \in \Lambda(\boldsymbol{c})} \left\{ \boldsymbol{\lambda}_{\mathbf{S}}^{\intercal} \mathbf{a}_{\mathbf{S}} + \boldsymbol{\lambda}_{\boldsymbol{\Pi}}^{\intercal} \boldsymbol{b}_{\Pi}(P) \right\},$$

$$\boldsymbol{\lambda}_{\boldsymbol{\Pi}}^{\intercal} \boldsymbol{b}_{\Pi}(P) = \sum_{z \in \mathcal{Z}} \mathbb{E}_P \left[\boldsymbol{\lambda}_{\boldsymbol{\Pi}}^{\intercal} \boldsymbol{m}_{\boldsymbol{\Pi}}(W) \mid Z = z \right].$$

If $\boldsymbol{\Pi}(P) = \emptyset$, Farkas' lemma, applied to the displayed system defining $\boldsymbol{\Pi}(P)$, gives $(\boldsymbol{\eta}_{\mathbf{S}}, \boldsymbol{\eta}_{\boldsymbol{\Pi}}) \geq \mathbf{0}$ with

$$\mathbf{A}_{\mathbf{S}}^{\intercal} \boldsymbol{\eta}_{\mathbf{S}} + \mathbf{A}_{\boldsymbol{\Pi}}^{\intercal} \boldsymbol{\eta}_{\boldsymbol{\Pi}} \geq \mathbf{0},$$

$$\beta := \boldsymbol{\eta}_{\mathbf{S}}^{\intercal} \mathbf{a}_{\mathbf{S}} + \boldsymbol{\eta}_{\boldsymbol{\Pi}}^{\intercal} \boldsymbol{b}_{\Pi}(P) < 0.$$

For every $t \geq 0$, $(\boldsymbol{\lambda}^0_{\mathbf{S}} + t\boldsymbol{\eta}_{\mathbf{S}}, t\boldsymbol{\eta}_{\boldsymbol{\Pi}}) \in \Lambda(\boldsymbol{c})$, with objective

$$(\boldsymbol{\lambda}^0_{\mathbf{S}} + t\boldsymbol{\eta}_{\mathbf{S}})^{\intercal} \mathbf{a}_{\mathbf{S}} + t\boldsymbol{\eta}_{\boldsymbol{\Pi}}^{\intercal} \boldsymbol{b}_{\Pi}(P) = (\boldsymbol{\lambda}^0_{\mathbf{S}})^{\intercal} \mathbf{a}_{\mathbf{S}} + t\beta \to -\infty.$$

The dual infimum therefore equals $-\infty = h_{\emptyset}(\boldsymbol{c})$ in this case as well.

Finally, at $\boldsymbol{c} = \mathbf{0}$, the dual infimum is zero if $\boldsymbol{\Pi}(P) \neq \emptyset$ and $-\infty$ otherwise. Since $(\mathbf{0}, \mathbf{0}) \in \Lambda(\mathbf{0})$, this infimum is zero exactly when every feasible dual objective is nonnegative. By Theorem 3.1, $\boldsymbol{\Pi}(P) \neq \emptyset$ is equivalent to $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$, giving the final claim. ■

A.1.7. *Proof of Lemma 3.4.*

Proof. Let $K := \bigcap_{i=1}^{N} K_i$. Suppose first that K is nonempty and compact, and define

$$f(\boldsymbol{c}) := \inf_{\substack{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N \in \mathbb{R}^k \\ \sum_{i=1}^{N} \boldsymbol{c}_i = \boldsymbol{c}}} \sum_{i=1}^{N} h_{K_i}(\boldsymbol{c}_i).$$

The function f is convex. For any $\boldsymbol{x}_0 \in K$,

$$f(\boldsymbol{c}) \geq \boldsymbol{c}^{\intercal} \boldsymbol{x}_0, \qquad f(\mathbf{0}) = 0,$$

and so f is proper.

For $C \subseteq \mathbb{R}^k$, let δ_C equal zero on C and ∞ elsewhere. For $g : \mathbb{R}^k \to \mathbb{R} \cup \{\infty\}$, let $g^*(\boldsymbol{x}) := \sup_{\boldsymbol{u} \in \mathbb{R}^k} \left\{ \boldsymbol{x}^\intercal \boldsymbol{u} - g(\boldsymbol{u}) \right\}$ for its convex conjugate and $g^{**} := (g^*)^*$ for its biconjugate.[30] Since each K_i is nonempty, closed, and convex, $h^*_{K_i} = \delta_{K_i}$, and

$$\begin{aligned} f^*(\boldsymbol{x}) &= \sup_{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N \in \mathbb{R}^k} \sum_{i=1}^N \left(\boldsymbol{c}_i^\intercal \boldsymbol{x} - h_{K_i}(\boldsymbol{c}_i) \right) \\ &= \sum_{i=1}^N \delta_{K_i}(\boldsymbol{x}) = \delta_K(\boldsymbol{x}). \end{aligned}$$

The Fenchel-Moreau theorem (Rockafellar, 2015, Theorem 12.2) therefore gives $\operatorname{cl} f = f^{**} = h_K$. Compactness of K implies that h_K is finite throughout $\mathbb{R}^k$. By Theorem 7.4 and Corollary 7.4.1 in Rockafellar (2015), f and $\operatorname{cl} f$ have the same relative interior of their domains and agree there. Therefore,

$$\begin{aligned} \operatorname{ri}(\operatorname{dom} f) = \operatorname{ri}(\operatorname{dom} \operatorname{cl} f) &= \mathbb{R}^k, \\ f(\boldsymbol{c}) = (\operatorname{cl} f)(\boldsymbol{c}) = h_K(\boldsymbol{c}) \qquad & \forall \boldsymbol{c} \in \mathbb{R}^k, \end{aligned}$$

which proves the first claim.

For the second claim, if $\boldsymbol{x} \in K$ and $\sum_{i=1}^N \boldsymbol{c}_i = \boldsymbol{0}$, then

$$\sum_{i=1}^N h_{K_i}(\boldsymbol{c}_i) \geq \sum_{i=1}^N \boldsymbol{c}_i^\intercal \boldsymbol{x} = 0.$$

For the converse, suppose $K = \emptyset$.

If one K_i is compact, relabel the sets so that K_1 is compact. Define

$$S := \prod_{i=1}^N K_i, \qquad \Delta := \left\{ (\boldsymbol{y}, \ldots, \boldsymbol{y}) \,\middle|\, \boldsymbol{y} \in \mathbb{R}^k \right\}.$$

These sets have positive distance. Otherwise, there would be $\boldsymbol{x}_{i,n} \in K_i$ and $\boldsymbol{y}_n \in \mathbb{R}^k$ satisfying

$$\max_{i \in [N]} \|\boldsymbol{x}_{i,n} - \boldsymbol{y}_n\|_2 \leq \frac{1}{n}.$$

Compactness supplies a subsequence n_j with $\boldsymbol{x}_{1,n_j} \to \boldsymbol{x} \in K_1$, and

$$\begin{aligned} \|\boldsymbol{x}_{i,n_j} - \boldsymbol{x}\|_2 &\leq \|\boldsymbol{x}_{i,n_j} - \boldsymbol{y}_{n_j}\|_2 + \|\boldsymbol{y}_{n_j} - \boldsymbol{x}_{1,n_j}\|_2 + \|\boldsymbol{x}_{1,n_j} - \boldsymbol{x}\|_2 \\ &\leq \frac{2}{n_j} + \|\boldsymbol{x}_{1,n_j} - \boldsymbol{x}\|_2 \to 0 \qquad \forall i \in [N]. \end{aligned}$$

Closedness of each K_i would then give $\boldsymbol{x} \in K$, a contradiction.

[30] See, e.g., Rockafellar (2015, Section 12, p. 104) for more details.

The hyperplane separation theorem, applied to the nonempty closed convex sets S and Δ at positive distance, gives $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N \in \mathbb{R}^k$ such that

$$\begin{aligned}\sum_{i=1}^{N} h_{K_i}(\boldsymbol{c}_i) &= \sup_{\boldsymbol{x}_i \in K_i,\ i \in [N]} \sum_{i=1}^{N} \boldsymbol{c}_i^{\intercal} \boldsymbol{x}_i \\ &< \inf_{\boldsymbol{y} \in \mathbb{R}^k} \left(\sum_{i=1}^{N} \boldsymbol{c}_i \right)^{\intercal} \boldsymbol{y}.\end{aligned}$$

Since S is nonempty, the left-hand side exceeds $-\infty$. The right-hand side is $-\infty$ unless $\sum_{i=1}^{N} \boldsymbol{c}_i = \boldsymbol{0}$, and so this sum must be zero. The right-hand side is then zero, giving the required strict inequality.

If instead every K_i is polyhedral, we can write

$$K_i = \left\{ \boldsymbol{x} \in \mathbb{R}^k \,\middle|\, \boldsymbol{A}_i \boldsymbol{x} \leq \boldsymbol{b}_i \right\},$$

where $\boldsymbol{A}_i \in \mathbb{R}^{m_i \times k}$ and $\boldsymbol{b}_i \in \mathbb{R}^{m_i}$. Since the combined system is infeasible, Farkas' lemma gives $\boldsymbol{\lambda}_i \in \mathbb{R}^{m_i}_{\geq 0}$ satisfying

$$\sum_{i=1}^{N} \boldsymbol{A}_i^{\intercal} \boldsymbol{\lambda}_i = \boldsymbol{0}, \qquad \sum_{i=1}^{N} \boldsymbol{b}_i^{\intercal} \boldsymbol{\lambda}_i < 0.$$

Set $\boldsymbol{c}_i := \boldsymbol{A}_i^{\intercal} \boldsymbol{\lambda}_i$. For every $\boldsymbol{x} \in K_i$,

$$\boldsymbol{c}_i^{\intercal} \boldsymbol{x} = \boldsymbol{\lambda}_i^{\intercal} \boldsymbol{A}_i \boldsymbol{x} \leq \boldsymbol{\lambda}_i^{\intercal} \boldsymbol{b}_i,$$

and so

$$\sum_{i=1}^{N} \boldsymbol{c}_i = \boldsymbol{0}, \qquad \sum_{i=1}^{N} h_{K_i}(\boldsymbol{c}_i) \leq \sum_{i=1}^{N} \boldsymbol{\lambda}_i^{\intercal} \boldsymbol{b}_i < 0.$$

This proves the converse. ■

A.1.8. *Proof of Theorem 3.3.*

Proof. By Theorem 3.1,

$$\boldsymbol{\Pi}(P) = \mathcal{R}_{\boldsymbol{\pi}} \cap \bigcap_{d \in \mathcal{D}} \boldsymbol{\Pi}_d(P).$$

By assumption, $\mathcal{R}_{\boldsymbol{\pi}}$ is a nonempty compact convex set. By Corollary A.1, each $\boldsymbol{\Pi}_d(P)$ is a nonempty closed convex polyhedron, with d-Never-Taker coordinates ranging unrestricted over $\mathbb{R}^{\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d}_{\geq 0}$. For $\boldsymbol{c} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ and $\boldsymbol{u} \in \prod_{d \in \mathcal{D}} \mathbb{R}^{\boldsymbol{\mathcal{T}}}$, define

$$F_{\boldsymbol{c}}(\boldsymbol{u}) := \sum_{d \in \mathcal{D}} h_{\boldsymbol{\Pi}_d(P)}(\boldsymbol{u}_d) + h_{\mathcal{R}_{\boldsymbol{\pi}}} \left(\boldsymbol{c} - \sum_{d \in \mathcal{D}} \boldsymbol{u}_d \right).$$

If $\mathbf{\Pi}(P) \neq \emptyset$, its compactness and Lemma 3.4 give

$$h_{\mathbf{\Pi}(P)}(\boldsymbol{c}) = \inf_{\boldsymbol{u} \in \prod_{d \in \mathcal{D}} \mathbb{R}^{\mathcal{T}}} F_{\boldsymbol{c}}(\boldsymbol{u}).$$

If $\mathbf{\Pi}(P) = \emptyset$, the second claim of Lemma 3.4 gives $\boldsymbol{a}_0, (\boldsymbol{a}_d)_{d \in \mathcal{D}} \in \mathbb{R}^{\mathcal{T}}$ with

$$\boldsymbol{a}_0 + \sum_{d \in \mathcal{D}} \boldsymbol{a}_d = \mathbf{0}\ .$$

Let $\eta := h_{\mathcal{R}_{\boldsymbol{\pi}}}(\boldsymbol{a}_0) + \sum_{d \in \mathcal{D}} h_{\mathbf{\Pi}_d(P)}(\boldsymbol{a}_d) < 0$. By subadditivity and positive homogeneity of support functions, for every $t > 0$, we have

$$F_{\boldsymbol{c}}\left((t\boldsymbol{a}_d)_{d \in \mathcal{D}}\right) \leq h_{\mathcal{R}_{\boldsymbol{\pi}}}(\boldsymbol{c}) + t\eta \to -\infty.$$

Thus the same infimum identity holds when $\mathbf{\Pi}(P) = \emptyset$, with $h_{\emptyset}(\boldsymbol{c}) = -\infty$.

It remains to restrict the infimum to $\mathcal{C}$. If $F_{\boldsymbol{c}}(\boldsymbol{u}) < \infty$, the unrestricted nonnegative coordinates of $\mathbf{\Pi}_d(P)$ imply $\boldsymbol{u}_{d;(\mathcal{T} \setminus \mathcal{T}_d)}$ for all $d \in \mathcal{D}$. Define $\widetilde{\boldsymbol{u}}_d$ by $\widetilde{\boldsymbol{u}}_{d;(\mathcal{T}_d)} := \boldsymbol{u}_{d;(\mathcal{T}_d)}$ and $\widetilde{\boldsymbol{u}}_{d;(\mathcal{T} \setminus \mathcal{T}_d)} := \mathbf{0}$. Then $\widetilde{\boldsymbol{u}} \in \mathcal{C}$, $\widetilde{\boldsymbol{u}}_d \geq \boldsymbol{u}_d$, and

$$h_{\mathbf{\Pi}_d(P)}(\widetilde{\boldsymbol{u}}_d) = h_{\mathbf{\Pi}_d(P)}(\boldsymbol{u}_d),$$
$$h_{\mathcal{R}_{\boldsymbol{\pi}}}\left(\boldsymbol{c} - \sum_d \widetilde{\boldsymbol{u}}_d\right) \leq h_{\mathcal{R}_{\boldsymbol{\pi}}}\left(\boldsymbol{c} - \sum_d \boldsymbol{u}_d\right),$$

where the inequality uses $\mathcal{R}_{\boldsymbol{\pi}} \subseteq \mathbb{R}^{\mathcal{T}}_{\geq 0}$. So, $F_{\boldsymbol{c}}(\widetilde{\boldsymbol{u}}) \leq F_{\boldsymbol{c}}(\boldsymbol{u})$, and restricting the infimum to $\mathcal{C}$ does not change its value.

For $\boldsymbol{u} \in \mathcal{C}$, Lemmas 3.2 and A.2 give

$$\begin{aligned} h_{\mathbf{\Pi}_d(P)}(\boldsymbol{u}_d) &= \int_{\mathcal{Y}} h_d\left(\boldsymbol{u}_d; \boldsymbol{p}_d^P(y)\right) d\mu(y) \\ &= \sum_{z \in \mathcal{Z}} \mathbb{E}_P\left[\boldsymbol{m}_d^P(\boldsymbol{u}_d; W) \mid Z = z\right]. \end{aligned}$$

Convexity follows because support functions are convex and their arguments are affine. The final claim now follows immediately by noting that when $\boldsymbol{c} = \mathbf{0}$, the infimum is 0 if $\mathbf{\Pi}(P) \neq \emptyset$ and $-\infty$ otherwise. ■

A.1.9. *Proof of Theorem 3.4.*

Proof. First, note that $(\mathbf{0}, \boldsymbol{p}_d(y))$ and $\|\boldsymbol{c}(y)\|_\infty \mathbf{1}$ are primal and dual feasible, respectively, so every pointwise LP $\max_{(\boldsymbol{x},\boldsymbol{s}) \geq 0 : \mathbf{R}(\boldsymbol{x},\boldsymbol{s}) = \boldsymbol{p}_d(y)} \boldsymbol{c}(y)^{\mathsf{T}} \boldsymbol{x}$ is always feasible; note that the value of this LP is exactly $h_d(\boldsymbol{c}(y); \boldsymbol{p}_d(y))$. Next, since $\boldsymbol{c}_{(\mathcal{T} \setminus \mathcal{T}_d)} \leq \mathbf{0}$ and since every slack column and every column of $\mathbf{R}_d$ indexed by $\mathcal{T}_d$ contains a 1, we have that

$$0 \leq h_d(\boldsymbol{c}(y); \boldsymbol{p}_d(y)) \leq \|\boldsymbol{c}(y)\|_\infty \mathbf{1}^{\mathsf{T}} \boldsymbol{p}_d(y) < \infty\ ,$$

and so each pointwise LP has a primal-dual optimal basis. The finite termination claim now follows immediately by noting that there are only finitely many possible bases, and that each

iteration of the while loop covers every remaining point where the basis found that iteration is primal-dual feasible, ensuring no basis recurs. This also implies that the regions $\{O(y)\}_{y\in\mathcal{G}}$ forms a finite partition of $\mathcal{Y}$ and since the regions are defined by linear inequalities involving the measurable functions $\boldsymbol{c}$ and $\boldsymbol{p}_d$, this partition is a measurable partition.

On each region, their defining inequalities give primal and dual feasibility, with common objective $\boldsymbol{c}_B(y)^\intercal \mathbf{R}_B^{-1}\boldsymbol{p}_d(y)$, which immediately implies that B is primal-dual optimal there and so, (ii)b)-c) holds, except for $\|\boldsymbol{\ell}_d^{\mathrm{alg}}(\boldsymbol{c}(y);\boldsymbol{p}_d(y))\|_\infty \le \|\boldsymbol{c}(y)\|_\infty$, [31] for which we now obtain a direct argument, as follows. For any basis B that is primal-dual optimal at y, let $\boldsymbol{\ell} := \mathbf{R}_B^{-\intercal}\boldsymbol{c}_B(y)$, and note that

$$\mathbf{R}_B^\intercal \boldsymbol{\ell} = \mathbf{R}_B^\intercal \mathbf{R}_B^{-\intercal}\boldsymbol{c}_B(y) = \boldsymbol{c}_B(y)\,,$$

and since $\mathbf{R}_B$ is necessarily invertible, it does not have any zero rows, and so, given $z\in\mathcal{Z}$ it must have a column, say column j, with a 1 in its z-th row, so that, since $\mathbf{R}_B$ is a binary matrix,

$$0 \le \boldsymbol{\ell}_{(z)} \le \mathbf{R}_{B;(\cdot,j)}^\intercal \boldsymbol{\ell} = \boldsymbol{c}_{B;(j)}(y) \le \|\boldsymbol{c}(y)\|_\infty\,.$$

The result now follows immediately by noting that the constructed functions are all measurable, by the argument above, and, since $\boldsymbol{c}$ is P-integrable and $P_Z > 0$,

$$0 \le \int_{\mathcal{Y}} h_d^{\mathrm{alg}}(\boldsymbol{c}(y);\boldsymbol{p}_d(y))\,d\mu(y) \le \frac{\mathbb{E}_P[\|\boldsymbol{c}(Y)\|_\infty]}{\min_{z\in\mathcal{Z}} P_Z(z)} < \infty\,. \qquad \blacksquare$$

A.2. **Proofs of Identification Results in Section 4**

A.2.1. *Proof of Lemma 4.1.*

Proof. Define

$$\mathcal{E}_d(y) := \left\{ (\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathbb{R}^{\boldsymbol{\mathcal{T}}}\times\mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\ge 0} \;\middle|\; \begin{array}{l} \boldsymbol{x}_1 = \kappa(y)\boldsymbol{x}_2, \\ \mathbf{R}_d\boldsymbol{x}_2 \le \boldsymbol{p}_d^P(y) \end{array} \right\}\,.$$

By the same argument as in the proof of Lemma 3.3, applied to the coefficient $y\mapsto \kappa(y)\boldsymbol{c}_\eta + \boldsymbol{c}_\pi$, we have

$$h_{\Upsilon_d}(\boldsymbol{c}_\eta,\boldsymbol{c}_\pi) = \int_{\mathcal{Y}} h_{\mathcal{E}_d(y)}(\boldsymbol{c}_\eta,\boldsymbol{c}_\pi)\,d\mu(y)\,.$$

The first equality now follows by noting that

$$\begin{aligned} h_{\mathcal{E}_d(y)}(\boldsymbol{c}_\eta,\boldsymbol{c}_\pi) &= \sup_{(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathcal{E}_d(y)} \left\{\boldsymbol{c}_\eta^\intercal\boldsymbol{x}_1 + \boldsymbol{c}_\pi^\intercal\boldsymbol{x}_2\right\} \\ &= \sup_{(\boldsymbol{x}_1,\boldsymbol{x}_2)\in\mathcal{E}_d(y)} \left\{\boldsymbol{c}_\eta^\intercal\kappa(y)\boldsymbol{x}_2 + \boldsymbol{c}_\pi^\intercal\boldsymbol{x}_2\right\} \\ &= \sup_{\boldsymbol{x}\in\mathrm{Fibre}_d(\boldsymbol{p}_d^P(y))} (\kappa(y)\boldsymbol{c}_\eta + \boldsymbol{c}_\pi)^\intercal\,\boldsymbol{x} \end{aligned}$$

[31] Since this constraint is not part of the ordinary dual problem associated with the pointwise LPs.

$$= h_d\left(\kappa(y)\boldsymbol{c}_\eta + \boldsymbol{c}_\pi; \boldsymbol{p}_d^P(y)\right) \ .$$

The moment representation follows from Lemma 3.2 with $\boldsymbol{c}(y) = \kappa(y)\boldsymbol{c}_\eta + \boldsymbol{c}_\pi$. ■

A.2.2. *Proof of Lemma 4.2.*

Proof. Suppose first that $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. Let K_P be the nonempty compact convex set in Lemma A.10. For every $\gamma \in \mathbb{R}$, its image under $(\mathsf{N}, a) \mapsto \mathsf{N} - \gamma a$ is the compact interval

$$\{\mathsf{N} - \gamma a \,|\, (\mathsf{N}, a) \in K_P\} = \left[\underline{\Psi}(\gamma; P), \overline{\Psi}(\gamma; P)\right].$$

If $\min_{(\mathsf{N},a)\in K_P} a > 0$, this gives

$$\begin{aligned}\gamma \in \Theta_{\mathbf{S}}(P) &\iff \exists\, (\mathsf{N}, a) \in K_P : \mathsf{N} = \gamma a \\ &\iff \underline{\Psi}(\gamma; P) \leq 0 \leq \overline{\Psi}(\gamma; P) \qquad \forall \gamma \in \Theta_0.\end{aligned}$$

If instead the minimum mass is zero, compactness supplies $(\mathsf{N}, 0) \in K_P$. The bound $|\mathsf{N}| \leq \|\kappa\|_\infty a$ implies $\mathsf{N} = 0$, and so every $\gamma \in \Theta_0$ passes the residual test, as required by the definition $\Theta_{\mathbf{S}}(P) = \Theta_0$ in this case.

By (4.2), the extrema over K_P are exactly the primal extrema in (**??**). For the support representations, Lemma A.11, applied first to κ at γ and then to $-\kappa$ at $-\gamma$, gives

$$\overline{\Psi}(\gamma; P) = \inf_{\substack{\boldsymbol{v}\in\mathbb{R}^{\boldsymbol{\tau}} \\ \boldsymbol{v}_{(\boldsymbol{\tau}\setminus\boldsymbol{\tau}_{d^\star})}\geq \mathbf{0}}} \left\{ \int_{\mathcal{Y}} h_{d^\star}\left((\kappa(y) - \gamma)\boldsymbol{c}_A - \boldsymbol{v}; \boldsymbol{p}_{d^\star}^P(y)\right) d\mu(y) + h_{\boldsymbol{\Pi}(P)}(\boldsymbol{v}) \right\},$$

$$-\underline{\Psi}(\gamma; P) = \inf_{\substack{\boldsymbol{v}\in\mathbb{R}^{\boldsymbol{\tau}} \\ \boldsymbol{v}_{(\boldsymbol{\tau}\setminus\boldsymbol{\tau}_{d^\star})}\geq \mathbf{0}}} \left\{ \int_{\mathcal{Y}} h_{d^\star}\left((\gamma - \kappa(y))\boldsymbol{c}_A - \boldsymbol{v}; \boldsymbol{p}_{d^\star}^P(y)\right) d\mu(y) + h_{\boldsymbol{\Pi}(P)}(\boldsymbol{v}) \right\}.$$

The restriction on $\boldsymbol{v}$ is exactly the condition that the inactive coefficients in the fibre objective are nonpositive; all remaining coefficients are bounded, and so both fibre integrals are finite by Lemma A.2.

Finally, $\overline{\Psi}$ is a supremum of affine functions of γ, and $\underline{\Psi}$ is an infimum of those functions, proving convexity and concavity, respectively. Both are finite by compactness of K_P.

If $\mathcal{Q}_{\mathbf{S}}(P) = \emptyset$, then $\boldsymbol{\Pi}(P) = \emptyset$ by Theorem 3.1. The residual definitions give $\underline{\Psi}(\gamma; P) = \infty$ and $\overline{\Psi}(\gamma; P) = -\infty$, and so no $\gamma \in \Theta_0$ passes the test. On the displayed domain of $\boldsymbol{v}$, the fibre integrals are finite and $h_{\boldsymbol{\Pi}(P)}(\boldsymbol{v}) = -\infty$, giving the same extended values in the support formulas. ■

A.2.3. *Proof of Theorem 4.1.*

Proof. Let K_P be the compact convex set in Lemma A.10, and set

$$a_* := \min_{(\mathsf{N},a)\in K_P} a > 0, \qquad \theta_- := \min_{(\mathsf{N},a)\in K_P} \frac{\mathsf{N}}{a}, \qquad \theta_+ := \max_{(\mathsf{N},a)\in K_P} \frac{\mathsf{N}}{a}.$$

Both ratio extrema are attained by Lemma A.10 and belong to Θ_0.

For every $(\mathsf{N}, a) \in K_P$ and $\gamma' > \gamma$, $a_* \leq a \leq 1$ gives

$$\begin{aligned}(\mathsf{N} - \gamma a) - (\gamma' - \gamma) &\leq \mathsf{N} - \gamma' a \\ &\leq (\mathsf{N} - \gamma a) - a_*(\gamma' - \gamma).\end{aligned}$$

Taking either minima or maxima over K_P gives

$$\Psi(\gamma) - (\gamma' - \gamma) \leq \Psi(\gamma') \leq \Psi(\gamma) - a_*(\gamma' - \gamma)$$

for each $\Psi \in \{\underline{\Psi}(\,\cdot\,; P), \overline{\Psi}(\,\cdot\,; P)\}$. Both functions are therefore continuous and strictly decreasing, implying that their roots, if any, are unique. Therefore, it suffices to exhibit a root for each.

For every $(\mathsf{N}, a) \in K_P$,

$$\mathsf{N} - \theta_- a \geq 0, \qquad\qquad \mathsf{N} - \theta_+ a \leq 0,$$

with equality at a minimizing and maximizing ratio pair, respectively. Therefore, we have $\underline{\Psi}(\theta_-; P) = \overline{\Psi}(\theta_+; P) = 0$, where the roots are unique, and we now have

$$\begin{aligned}\{\gamma \in \Theta_0 \,|\, -\underline{\Psi}(\gamma; P) \leq 0\} &= [\underline{\kappa}, \theta_-], \\ \left\{\gamma \in \Theta_0 \,\middle|\, \overline{\Psi}(\gamma; P) \leq 0\right\} &= [\theta_+, \overline{\kappa}].\end{aligned}$$

By Lemma 4.2, $\Theta_{\mathbf{S}}(P) = [\theta_-, \theta_+]$, and so the bounds are sharp; in this case, both are also attained. ■

A.2.4. *Proof of Theorem 4.2.*

Proof. Note that, in this case, $\tilde{\Theta}_{\mathbf{S}} = \Theta_{\mathbf{S}}$, where $\tilde{\Theta}_{\mathbf{S}}$ is as defined in (A.3) above. The result now follows immediately from Theorem A.1 by setting $\boldsymbol{c}_{d^\star} = \boldsymbol{a}_\pi = \boldsymbol{c}_A$, setting $\boldsymbol{a}_{d^\star} = \boldsymbol{c}_\pi$ to zero along with all other $\boldsymbol{a}_d$ and $\boldsymbol{c}_d$, using $A \subseteq \boldsymbol{\mathcal{T}}_{d^\star}$ to eliminate $\boldsymbol{\zeta}_\gamma$, and then noticing that non-zero choices of $\boldsymbol{v}_d$, for $d \neq d^\star$, can only ever increase the objective (since $\mathbf{0}$ is always feasible for the maximization problem defining h_d). ■

A.2.5. *Proofs of Claims in Remark 4.1.*

Proof. Let G_P, H_P, and, under Condition CP, $\mathscr{J}_P$ and $\mathcal{D}_\theta$ be as defined in Lemma A.11. That lemma gives

$$\begin{aligned}\overline{\Psi}(\gamma; P) &= \inf_{\boldsymbol{v} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}} G_P(\gamma, \boldsymbol{v}) \\ &= \inf_{\substack{\boldsymbol{v} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}} \\ \boldsymbol{u} \in \mathcal{C}}} H_P(\gamma, \boldsymbol{v}, \boldsymbol{u}),\end{aligned}$$

and establishes joint convexity of both objective functions and the value $\bar{\theta}(P)$ of both lifted programs.

For every choice with $G_P(\gamma, \boldsymbol{v}) < \infty$, Lemma 3.2 gives

$$H_P(\gamma, \boldsymbol{v}, \boldsymbol{u}) = \sum_{z \in \mathcal{Z}} \mathbb{E}_P \left[\boldsymbol{m}_{d^\star}^P \left((\kappa(Y) - \gamma)\boldsymbol{c}_A - \boldsymbol{v}; W \right) + \sum_{d \in \mathcal{D}} \boldsymbol{m}_d^P(\boldsymbol{u}_d; W) \,\middle|\, Z = z \right] + h_{\mathcal{R}_{\boldsymbol{\pi}}} \left(\boldsymbol{v} - \sum_{d \in \mathcal{D}} \boldsymbol{u}_d \right),$$

which is the stated moment representation. Under Condition C, the two lifted programs agree in infimum value; feasibility at $\gamma = \bar{\theta}(P)$ is equivalent to attainment of the respective infimum defining $\overline{\Psi}(\bar{\theta}(P); P) = 0$.

Under Condition CP, Lemmas 3.2 and A.11 give

$$\mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}) = \sum_{z \in \mathcal{Z}} \mathbb{E}_P \left[\boldsymbol{m}_\theta^P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}; W) \mid Z = z \right],$$

$$\overline{\Psi}(\gamma; P) = \inf_{(\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{D}_\theta} \mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}).$$

Here $\mathcal{D}_\theta$ specifies exactly the choices for which the target fibre integral is finite. By Lemma A.12, there exists $(\boldsymbol{u}^\star, \boldsymbol{\lambda}^\star) \in \mathcal{D}_\theta$ such that

$$\mathscr{J}_P\left(\bar{\theta}(P), \boldsymbol{u}^\star, \boldsymbol{\lambda}^\star\right) = \overline{\Psi}(\bar{\theta}(P); P) = 0.$$

Thus (4.7) attains $\bar{\theta}(P)$, and every optimal solution has its γ coordinate equal to the unique root of $\overline{\Psi}(\,\cdot\,; P)$.

Finally, since $(\kappa(y) - \overline{\kappa})\boldsymbol{c}_A \le \mathbf{0}$ and each fibre contains $\mathbf{0}$,

$$G_P(\overline{\kappa}, \mathbf{0}) = H_P(\overline{\kappa}, \mathbf{0}, \mathbf{0}) = 0,$$

$$\mathscr{J}_P(\overline{\kappa}, \mathbf{0}, \mathbf{0}) = 0 \qquad \text{under Condition CP.}$$

All the displayed programs therefore have nonempty feasible sets, under the conditions imposed throughout the remark. Replacing κ by $-\kappa$, which yields a parameter with sharp upper bound $-\underline{\theta}(P)$, proves the analogous claims for the lower-bound. ■

A.3. **Identification Results for General Parameters**

In this section, let $\theta = \frac{\mathsf{N}^Q}{\mathsf{D}^Q}$ be represented as in (†). Let Θ_0 be the set of *a priori* possible values for θ, which we will treat as known and assume satisfies $\Theta_0 \supseteq \overline{\{\mathsf{N}^Q/\mathsf{D}^Q \,|\, Q \in \mathcal{M}, \boldsymbol{\pi}^Q \in \mathcal{R}_{\boldsymbol{\pi}}\}}$. Define $\Theta_{\mathbf{S}}$, the identified set for θ under $\mathbf{S}$, as

$$\Theta_{\mathbf{S}}(P) := \begin{cases} \emptyset & \text{iff } \mathcal{Q}_{\mathbf{S}}(P) = \emptyset, \\ \Theta_0 & \text{if } \inf_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \mathsf{D}^Q \le 0 \le \sup_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \mathsf{D}^Q, \\ \left\{ \mathsf{N}^Q/\mathsf{D}^Q \;\middle|\; Q \in \mathcal{Q}_{\mathbf{S}}(P) \right\} & \text{otherwise}. \end{cases}$$

For effects ratios, we will require the following condition, which assumes the zero denominator case away.

Assumption A.1. $0 \notin \overline{\{\mathsf{D}^Q \,|\, Q \in \mathcal{Q}_{\mathbf{S}}(P)\}}$

A.3.1. *Inclusion Test for General Parameters.* Define[32]

$$\tilde{\Theta}_{\mathbf{S}}(P) := \left\{ \gamma \in \Theta_0 \,\middle|\, \inf_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \left(\mathsf{N}^Q - \gamma \mathsf{D}^Q\right) \leq 0 \leq \sup_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \left(\mathsf{N}^Q - \gamma \mathsf{D}^Q\right) \right\} . \tag{A.3}$$

For all three types of target parameters, under the requirements of Theorem A.1 below, we have $\Theta_{\mathbf{S}}(P) \subseteq \tilde{\Theta}_{\mathbf{S}}(P) \subseteq \overline{\Theta_{\mathbf{S}}(P)}$ with $\tilde{\Theta}_{\mathbf{S}}(P) = \overline{\Theta_{\mathbf{S}}(P)}$ for target parameter types (i) and (ii); for effects ratios, $\tilde{\Theta}_{\mathbf{S}}(P) = \overline{\Theta_{\mathbf{S}}(P)}$ whenever the $\varkappa_d$s are bounded and the requirements of Theorem A.1 holds.

Recall that $\mathcal{C} := \prod_{d \in \mathcal{D}} \mathcal{C}_d$, with $\mathcal{C}_d := \left\{ \boldsymbol{u} \in \mathbb{R}^{\mathcal{T}} \,\middle|\, \boldsymbol{u}_{(\mathcal{T} \setminus \mathcal{T}_d)} = \mathbf{0} \right\}$, and the shorthand $\boldsymbol{v} := (\boldsymbol{v}_d)_{d \in \mathcal{D}}$ is used to split $\boldsymbol{v}$ into its $\mathcal{C}_d$ components. For any $\boldsymbol{b} \in \mathbb{R}^{\mathcal{T}}$, denote by $\boldsymbol{b}_{\restriction d}$ the vector generated from $\boldsymbol{b}$ by setting its d-Never-Taker coordinates to 0. Define

$$F_P(\boldsymbol{v}, t; \gamma) := \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} h_d\left(t\left(\kappa_d(y)\, \boldsymbol{c}_d - \gamma \varkappa_d(y)\, \boldsymbol{a}_d\right)_{\restriction d} - \boldsymbol{v}_d; \boldsymbol{p}_d(y)\right) d\mu(y) + h_{\mathbf{\Pi}}\left(\sum_{d \in \mathcal{D}} \boldsymbol{v}_d + t\left(\boldsymbol{c}_\pi - \gamma \boldsymbol{a}_\pi\right) + \boldsymbol{\zeta}_\gamma(t)\right) , \tag{A.4}$$

where $\boldsymbol{\zeta}_\gamma$ is the following "Never-Taker" adjustment term,

$$\boldsymbol{\zeta}_\gamma(t) := \left(\sum_{d \in \mathcal{D}: \boldsymbol{\tau} \notin \boldsymbol{\mathcal{T}}_d} \left(\overline{\kappa}_d \left\{ t\left(\boldsymbol{c}_{d;(\boldsymbol{\tau})} - \gamma \boldsymbol{a}_{d;(\boldsymbol{\tau})}\right) \right\}_+ + \underline{\kappa}_d \left\{ t\left(\boldsymbol{c}_{d;(\boldsymbol{\tau})} - \gamma \boldsymbol{a}_{d;(\boldsymbol{\tau})}\right) \right\}_- \right) : \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \right) ,$$

with $\overline{\kappa}_d := \operatorname{ess\,sup}_\mu \kappa_d$ and $\underline{\kappa}_d := \operatorname{ess\,inf}_\mu \kappa_d$, and where we extend the domain of $h_{\mathbf{\Pi}}$ to include ∞, with the convention $0 \cdot \infty = 0$.[33] Note that the $\boldsymbol{\zeta}_\gamma(t) = \mathbf{0}$ if $\boldsymbol{c}_{d;(\mathcal{T} \setminus \mathcal{T}_d)} = \boldsymbol{a}_{d;(\mathcal{T} \setminus \mathcal{T}_d)} = \mathbf{0}$ for all $d \in \mathcal{D}$; more generally, $\boldsymbol{\zeta}_\gamma(t)$ satisfies

$$\boldsymbol{\zeta}_\gamma(t) = \begin{cases} t\boldsymbol{\zeta}_\gamma(1) & \text{if } t \geq 0 , \\ -t\boldsymbol{\zeta}_\gamma(-1) & \text{if } t < 0 . \end{cases}$$

and so, the test below can be implemented as the minimum of a pair of convex programs, one with $t \geq 0$ and another with $t \leq 0$.

We now have the following inclusion test for $\tilde{\Theta}_{\mathbf{S}}$.

Theorem A.1. *Suppose P satisfies Condition D and $\mathbf{S}$ satisfies Condition C. Further suppose that either (i) θ satisfies Condition T1, or (ii) Assumption A.1 holds and θ satisfies Condition T2. Then, for any norm $\|\cdot\|$ on $\left(\prod_{d \in \mathcal{D}} \mathbb{R}^{\mathcal{T}}\right) \times \mathbb{R}$,*

$$\tilde{\Theta}_{\mathbf{S}}(P) = \left\{ \gamma \in \Theta_0 \,\middle|\, \inf_{\substack{\boldsymbol{v} \in \mathcal{C}, t \in \mathbb{R} \\ \|(\boldsymbol{v}, t)\| \leq 1}} F_P(\boldsymbol{v}, t; \gamma) \geq 0 \right\} .$$

[32] Using our usual convention that infimum (resp. supremum) over empty set is ∞ (resp. $-\infty$).

[33] Note that this is unambiguous here since $\boldsymbol{\zeta}_\gamma \in \bar{\mathbb{R}} \setminus \{-\infty\}$

Proof. Fix $\gamma, t \in \mathbb{R}$, and define

$$\begin{aligned} \boldsymbol{g}_d(y) &:= t\left(\kappa_d(y)\boldsymbol{c}_d - \gamma \varkappa_d(y)\boldsymbol{a}_d\right)_{\restriction d}, \\ \boldsymbol{b} &:= t(\boldsymbol{c}_\pi - \gamma \boldsymbol{a}_\pi) + \boldsymbol{\zeta}_\gamma(t). \end{aligned}$$

Each $\boldsymbol{g}_d$ is P-integrable and takes values in $\mathcal{C}_d$. Writing $r_d := \sum_z p^P_{d|z}$, Lemma A.2 gives, for every $\boldsymbol{v}_d \in \mathcal{C}_d$,

$$\begin{aligned} 0 &\le \int_{\mathcal{Y}} h_d\left(\boldsymbol{g}_d(y) - \boldsymbol{v}_d; \boldsymbol{p}^P_d(y)\right) d\mu(y) \\ &\le \int_{\mathcal{Y}} \left(\|\boldsymbol{g}_d(y)\|_\infty + \|\boldsymbol{v}_d\|_\infty\right) r_d(y)\, d\mu(y) \\ &\le \frac{\mathbb{E}_P\|\boldsymbol{g}_d(Y)\|_\infty + \|\boldsymbol{v}_d\|_\infty}{\min_z P_Z(z)} < \infty. \end{aligned}$$

If $\mathcal{Q}_{\mathbf{S}}(P) = \emptyset$, then $\boldsymbol{\Pi}(P) = \emptyset$, and so $F_P(\boldsymbol{v}, t; \gamma) = -\infty$ for every $\boldsymbol{v} \in \mathcal{C}$. Both sets in the statement are therefore empty. Assume henceforth that $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$.

First consider $d \in \mathcal{D}$ and $\boldsymbol{\tau} \notin \boldsymbol{\mathcal{T}}_d$. Under either Condition T1 or Condition T2,

$$\begin{aligned} f(y) &:= t\left(\kappa_d(y)\boldsymbol{c}_{d;(\boldsymbol{\tau})} - \gamma \varkappa_d(y)\boldsymbol{a}_{d;(\boldsymbol{\tau})}\right) \\ &= t\left(\boldsymbol{c}_{d;(\boldsymbol{\tau})} - \gamma \boldsymbol{a}_{d;(\boldsymbol{\tau})}\right)\kappa_d(y). \end{aligned}$$

Let ρ_0 be the density supplied by Lemma A.1 for the functions $(\kappa_d, \varkappa_d)_{d\in\mathcal{D}}$. For any $a < \operatorname{ess\,sup}_\mu f$, the density

$$\rho_a(y) := \frac{\rho_0(y)\mathbb{1}\{f(y) > a\}}{\int_{\mathcal{Y}} \rho_0(u)\mathbb{1}\{f(u) > a\}\, d\mu(u)}$$

has finite κ_d- and $\varkappa_d$-moments and satisfies $\int_{\mathcal{Y}} f\rho_a\, d\mu > a$. Since a is arbitrary, the supremum of $\int_{\mathcal{Y}} fq\, d\mu$ over nonnegative densities of mass $\boldsymbol{\pi}_{(\boldsymbol{\tau})}$ is

$$\boldsymbol{\pi}_{(\boldsymbol{\tau})} \operatorname*{ess\,sup}_\mu f = \boldsymbol{\pi}_{(\boldsymbol{\tau})}\left[\overline{\kappa}_d\left\{t\left(\boldsymbol{c}_{d;(\boldsymbol{\tau})} - \gamma\boldsymbol{a}_{d;(\boldsymbol{\tau})}\right)\right\}_+ + \underline{\kappa}_d\left\{t\left(\boldsymbol{c}_{d;(\boldsymbol{\tau})} - \gamma\boldsymbol{a}_{d;(\boldsymbol{\tau})}\right)\right\}_-\right].$$

This also holds when $\boldsymbol{\pi}_{(\boldsymbol{\tau})} = 0$, using $0 \cdot \infty = 0$.

These conditional outcome distributions can be chosen separately for every $(d, \boldsymbol{\tau})$ with $\boldsymbol{\tau} \notin \boldsymbol{\mathcal{T}}_d$. They leave $\boldsymbol{\pi}$ and every $\mathbf{R}_d\boldsymbol{q}_d$ unchanged. Draw the potential outcomes independently conditional on $\boldsymbol{T}$, and draw Z independently with law P_Z; Lemma 3.1 then gives a parent in $\mathcal{Q}_{\mathbf{S}}(P)$. Summing the preceding suprema therefore gives

$$\begin{aligned} &\sup_{Q\in\mathcal{Q}_{\mathbf{S}}(P)} t\left(\mathsf{N}^Q - \gamma\mathsf{D}^Q\right) \\ &= \sup_{Q\in\mathcal{Q}_{\mathbf{S}}(P)} \left\{\sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} \boldsymbol{g}_d(y)^\intercal \boldsymbol{q}^Q_d(y)\, d\mu(y) + \boldsymbol{b}^\intercal\boldsymbol{\pi}^Q\right\}. \end{aligned}$$

Here finite suprema are approached by the densities just constructed; if one contribution is infinite and has positive mass, it can be made arbitrarily large while all other contributions remain finite.

If $\boldsymbol{b}_{(\boldsymbol{\tau})} = \infty$ and $\boldsymbol{\pi}_{(\boldsymbol{\tau})} > 0$ for some $\boldsymbol{\pi} \in \mathbf{\Pi}(P)$, both the last supremum and $F_P(\boldsymbol{v}, t; \gamma)$ are infinite for every $\boldsymbol{v} \in \mathcal{C}$. Otherwise, every coordinate on which $\boldsymbol{b}$ is infinite has zero mass under every $\boldsymbol{\pi} \in \mathbf{\Pi}(P)$. Replace those components of $\boldsymbol{b}$ by zero to obtain $\boldsymbol{b}^0 \in \mathbb{R}^{\mathcal{T}}$. Then, for every $\boldsymbol{v} \in \mathcal{C}$,

$$\boldsymbol{b}^\intercal \boldsymbol{\pi} = (\boldsymbol{b}^0)^\intercal \boldsymbol{\pi} \qquad \forall \boldsymbol{\pi} \in \mathbf{\Pi}(P),$$

$$h_{\mathbf{\Pi}(P)}\left(\boldsymbol{b} + \sum_d \boldsymbol{v}_d\right) = h_{\mathbf{\Pi}(P)}\left(\boldsymbol{b}^0 + \sum_d \boldsymbol{v}_d\right).$$

Apply Lemma A.9 with $D_0 = \mathcal{D}$, $\Pi_0 = \mathbf{\Pi}(P)$, loadings $(\boldsymbol{g}_d)_{d \in \mathcal{D}}$, and coefficient $\boldsymbol{b}^0$. In either case, we obtain

$$\sup_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} t\left(\mathsf{N}^Q - \gamma \mathsf{D}^Q\right) = \inf_{\boldsymbol{v} \in \mathcal{C}} F_P(\boldsymbol{v}, t; \gamma). \tag{A.5}$$

At $t = 1$ and $t = -1$, this identity and (A.3) give

$$\gamma \in \tilde{\Theta}_{\mathbf{S}}(P) \iff F_P(\boldsymbol{v}, t; \gamma) \geq 0 \quad \forall \boldsymbol{v} \in \mathcal{C},\ t \in \{-1, 1\}.$$

At $t = 0$, the same identity gives $F_P(\boldsymbol{v}, 0; \gamma) \geq 0$ for every $\boldsymbol{v} \in \mathcal{C}$. Moreover, positive homogeneity allows us to restrict to the unit ball, as required. ■

A.3.2. *Sharp Bounds for General Parameters.* Note that, whenever $\mathbf{S}$ satisfies Condition C, Assumption A.1 implies that either $\inf_{Q \in \mathcal{Q}_{\mathbf{S}}} \mathsf{D}^Q > 0$ or $\sup_{Q \in \mathcal{Q}_{\mathbf{S}}} \mathsf{D}^Q < 0$. Define

$$s_P := \begin{cases} 1 & \text{if } \inf_{Q \in \mathcal{Q}_{\mathbf{S}}} \mathsf{D}^Q > 0\,, \\ -1 & \text{otherwise}\,, \end{cases}$$

and then define the generalized extreme residuals as

$$\overline{\Psi}(\gamma; P) := \sup_{Q \in \mathcal{Q}_{\mathbf{S}}} s_P\left(\mathsf{N}^Q - \gamma \mathsf{D}^Q\right)\,,$$

$$\underline{\Psi}(\gamma; P) := \inf_{Q \in \mathcal{Q}_{\mathbf{S}}} s_P\left(\mathsf{N}^Q - \gamma \mathsf{D}^Q\right)\,.$$

Theorem A.2. *Suppose P satisfies Condition D and $\mathbf{S}$ satisfies Condition C. Further suppose that θ satisfies Condition T1 or Condition T2 (or both). Then, $\overline{\Psi}(\gamma; P) : \Theta_0 \to \mathbb{R} \sqcup \{-\infty, \infty\}$ is convex, $\underline{\Psi}(\gamma; P) : \Theta_0 \to \mathbb{R} \sqcup \{-\infty, \infty\}$ is concave, and*

$$\overline{\Psi}(\gamma; P) = \inf_{\boldsymbol{v} \in \mathcal{C}} F_P(\boldsymbol{v}, s_P; \gamma)\,,$$

$$\underline{\Psi}(\gamma; P) = -\inf_{\boldsymbol{v} \in \mathcal{C}} F_P(\boldsymbol{v}, -s_P; \gamma)\,.$$

If also $\mathcal{Q}_{\mathbf{S}} \neq \emptyset$ and Assumption A.1 holds, then the following bounds on θ are sharp under $\mathbf{S}$,

$$\underline{\theta}(P) := \sup\{\gamma \in \Theta_0 \,|\, -\underline{\Psi}(\gamma; P) \leq 0\} \;\leq\; \theta \;\leq\; \inf\left\{\gamma \in \Theta_0 \,\middle|\, \overline{\Psi}(\gamma; P) \leq 0\right\} =: \bar{\theta}(P)\,,$$

and, if these bounds are finite and $\sup_{Q \in \mathcal{Q}_{\mathbf{S}}} s_P \mathsf{D}^Q < \infty$, then $\underline{\Psi}(\cdot\,; P)$ and $\overline{\Psi}(\cdot\,; P)$ are strictly decreasing, with unique roots $\underline{\theta}(P)$ and $\bar{\theta}(P)$, respectively.

Proof. Applying (A.5) at $t = s_P$ and $t = -s_P$ gives

$$\overline{\Psi}(\gamma; P) = \inf_{\boldsymbol{v} \in \mathcal{C}} F_P(\boldsymbol{v}, s_P; \gamma),$$
$$\underline{\Psi}(\gamma; P) = -\inf_{\boldsymbol{v} \in \mathcal{C}} F_P(\boldsymbol{v}, -s_P; \gamma).$$

For each Q, the function $\gamma \mapsto s_P(\mathsf{N}^Q - \gamma \mathsf{D}^Q)$ is affine, and so its supremum is convex and its infimum is concave.

Now suppose $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$ and Assumption A.1 holds. By Lemma 3.1 and convexity of $\mathcal{R}_{\boldsymbol{\pi}}$, mixtures of compatible parents remain compatible: their independent Z-marginal is always P_Z, their response equations are preserved, and their type vectors remain in $\mathcal{R}_{\boldsymbol{\pi}}$. In particular,

$$\mathsf{D}^{(1-\lambda)Q_0 + \lambda Q_1} = (1-\lambda)\mathsf{D}^{Q_0} + \lambda \mathsf{D}^{Q_1} \qquad \forall \lambda \in [0,1].$$

The denominator image is therefore an interval. Since its closure excludes zero,

$$a_* := \inf_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} s_P \mathsf{D}^Q > 0.$$

Define

$$\theta_- := \inf_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \frac{\mathsf{N}^Q}{\mathsf{D}^Q}, \qquad\qquad \theta_+ := \sup_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \frac{\mathsf{N}^Q}{\mathsf{D}^Q}.$$

For every $Q \in \mathcal{Q}_{\mathbf{S}}(P)$,

$$s_P(\mathsf{N}^Q - \gamma \mathsf{D}^Q) = s_P \mathsf{D}^Q(\theta^Q - \gamma),$$

and so, for every $\gamma \in \mathbb{R}$,

$$-\underline{\Psi}(\gamma; P) \leq 0 \iff \gamma \leq \theta_-,$$
$$\overline{\Psi}(\gamma; P) \leq 0 \iff \gamma \geq \theta_+.$$

Every finite endpoint θ_- or θ_+ belongs to Θ_0 by its defining requirement. If $\theta_- = -\infty$ or $\theta_+ = \infty$, the corresponding set in the bound formula is empty. Thus the stated formulas give $\underline{\theta}(P) = \theta_-$ and $\bar{\theta}(P) = \theta_+$, including infinite bounds.

For $Q_0, Q_1 \in \mathcal{Q}_{\mathbf{S}}(P)$, the ratio

$$\frac{(1-\lambda)\mathsf{N}^{Q_0} + \lambda \mathsf{N}^{Q_1}}{(1-\lambda)\mathsf{D}^{Q_0} + \lambda \mathsf{D}^{Q_1}}$$

is continuous in $\lambda \in [0,1]$ and takes every value between θ^{Q_0} and θ^{Q_1}. Every value strictly between θ_- and θ_+ is therefore attained, and the bounds are sharp.

Finally, suppose $\theta_-, \theta_+ \in \mathbb{R}$ and

$$b_* := \sup_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} s_P \mathsf{D}^Q < \infty.$$

For every $\gamma \in \mathbb{R}$ and $Q \in \mathcal{Q}_{\mathbf{S}}(P)$,

$$\left|\mathsf{N}^Q - \gamma \mathsf{D}^Q\right| \leq b_* \left(\max\{|\theta_-|, |\theta_+|\} + |\gamma|\right),$$

and so both residuals are finite. For $\gamma' > \gamma$,

$$\begin{aligned} s_P(\mathsf{N}^Q - \gamma \mathsf{D}^Q) - b_*(\gamma' - \gamma) &\leq s_P(\mathsf{N}^Q - \gamma' \mathsf{D}^Q) \\ &\leq s_P(\mathsf{N}^Q - \gamma \mathsf{D}^Q) - a_*(\gamma' - \gamma). \end{aligned}$$

Taking suprema or infima over Q gives

$$\Psi(\gamma) - b_*(\gamma' - \gamma) \leq \Psi(\gamma') \leq \Psi(\gamma) - a_*(\gamma' - \gamma)$$

for each $\Psi \in \{\underline{\Psi}(\,\cdot\,; P), \overline{\Psi}(\,\cdot\,; P)\}$. Both functions are therefore continuous and strictly decreasing, implying their roots, if any, are unique. Therefore, it suffices to exhibit a root for each.

Choose sequences $(Q_n^-)_{n\in\mathbb{N}}, (Q_n^+)_{n\in\mathbb{N}} \subseteq \mathcal{Q}_{\mathbf{S}}(P)$ such that $\theta^{Q_n^-} \to \theta_-$ and $\theta^{Q_n^+} \to \theta_+$, so that we now have

$$\left|s_P\left(\mathsf{N}^{Q_n^\pm} - \theta_\pm \mathsf{D}^{Q_n^\pm}\right)\right| \leq b_* \left|\theta^{Q_n^\pm} - \theta_\pm\right| \to 0.$$

All residuals at θ_- are nonnegative, and all residuals at θ_+ are nonpositive, and so, $\underline{\Psi}(\theta_-; P) = \overline{\Psi}(\theta_+; P) = 0$, as required. ■

A.4. **Auxiliary Identification Lemmas**

Lemma A.1. *Let measurable $f_1, \ldots, f_N : \mathcal{Y} \to \mathbb{R}$ be given. Then, there exists measurable $\rho_0 : \mathcal{Y} \to \mathbb{R}$ with $\rho_0(y) > 0\ \forall y \in \mathcal{Y}$, $\int_{\mathcal{Y}} \rho_0 \, d\mu = 1$ and $\int_{\mathcal{Y}} \left(\sum_{k=1}^N |f_k|\right) \rho_0 \, d\mu < \infty$.*

Proof. Since μ is σ-finite, Lemma 6.9 in Rudin (1987) gives the existence of $w \in L^1(\mu)$ with $0 < w(y) < 1\ \forall y \in \mathcal{Y}$; result follows immediately by defining $\rho_0 := \frac{w}{c(1+\sum_{k=1}^N |f_k|)}$, where

$$0 < c := \int_{\mathcal{Y}} \frac{w(y)}{1 + \sum_{k=1}^N |f_k(y)|} \, d\mu < \infty\,. \qquad ■$$

Lemma A.2. *Suppose P satisfies Condition D and fix $d \in \mathcal{D}$. Then, for any P-integrable $\boldsymbol{c} : \mathcal{Y} \to \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ with $\boldsymbol{c}_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} \leq \mathbf{0}$, there exists $\boldsymbol{x}^\star \in L^1(\mu; \mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\geq 0})$ with $\boldsymbol{x}^\star_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} = \mathbf{0}$ such that, for all $y \in \mathcal{Y}$,*

$$\boldsymbol{x}^\star(y) \in \operatorname*{argmax}_{\boldsymbol{v} \in \mathrm{Fibre}_d(\boldsymbol{p}_d^P(y))} \boldsymbol{c}(y)^\intercal \boldsymbol{v}\,,$$

and

$$0 \le \|\boldsymbol{x}^\star(y)\|_1 \le \sum_{z\in\mathcal{Z}} \left(\mathbf{R}_d \boldsymbol{x}^\star(y)\right)_z \le \sum_{z\in\mathcal{Z}} p^P_{d|z}(y).$$

Moreover,

$$\max_{\substack{\boldsymbol{x}\in L^1(\mu;\mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\ge 0}) \\ \boldsymbol{x}(y)\in \mathrm{Fibre}_d(\boldsymbol{p}^P_d(y)) \\ \boldsymbol{x}_{\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d}(y)=\mathbf{0}}} \int_{\mathcal{Y}} \boldsymbol{c}(y)^\intercal \boldsymbol{x}(y)\, d\mu(y) = \int_{\mathcal{Y}} h_d\left(\boldsymbol{c}(y); \boldsymbol{p}^P_d(y)\right)\, d\mu(y) < \infty.$$

Further, for $\boldsymbol{c}' : \mathcal{Y} \to \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ that is also P-integrable and satisfies $\boldsymbol{c}'_{(\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d)} \le \mathbf{0}$,

$$\begin{aligned}
&\left| \int_{\mathcal{Y}} h_d\left(\boldsymbol{c}(y); \boldsymbol{p}^P_d(y)\right)\, d\mu(y) - \int_{\mathcal{Y}} h_d\left(\boldsymbol{c}'(y); \boldsymbol{p}^P_d(y)\right)\, d\mu(y) \right| \\
&\le \int_{\mathcal{Y}} \left(\sum_{z\in\mathcal{Z}} p^P_{d|z}(y) \right) \left\| (\boldsymbol{c}(y) - \boldsymbol{c}'(y))_{\boldsymbol{\mathcal{T}}_d} \right\|_\infty d\mu(y).
\end{aligned}$$

Proof. For $y \in \mathcal{Y}$, define the nonempty compact set (closed, bounded, and contains $\mathbf{0}$)

$$\mathrm{Fibre}^0_d\left(\boldsymbol{p}^P_d(y)\right) := \left\{ \boldsymbol{x} \in \mathrm{Fibre}_d\left(\boldsymbol{p}^P_d(y)\right) \,\middle|\, \boldsymbol{x}_{(\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d)} = \mathbf{0} \right\}.$$

Every column of $\mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}_d)}$ contains at least one 1 and so, for every $\boldsymbol{x} \in \mathrm{Fibre}^0_d\left(\boldsymbol{p}^P_d(y)\right)$,

$$\|\boldsymbol{x}\|_1 = \sum_{\boldsymbol{\tau}\in\boldsymbol{\mathcal{T}}_d} x_{(\boldsymbol{\tau})} \le \sum_z \left(\mathbf{R}_d \boldsymbol{x}\right)_{(z)} \le \sum_z p^P_{d|z}(y).$$

Define

$$f(y, \boldsymbol{x}) := \begin{cases} -\boldsymbol{c}(y)^\intercal \boldsymbol{x} & \text{if } \boldsymbol{x} \in \mathrm{Fibre}^0_d\left(\boldsymbol{p}^P_d(y)\right), \\ \infty & \text{otherwise,} \end{cases}$$

and note that the following set

$$\mathrm{epi}\, f(y,\cdot) := \left\{ (\boldsymbol{x}, a) : \boldsymbol{x} \in \mathrm{Fibre}^0_d\left(\boldsymbol{p}^P_d(y)\right),\ -\boldsymbol{c}(y)^\intercal \boldsymbol{x} \le a \right\}$$

is always a closed and measurable set, since it is defined by linear constraints with measurable coefficients, by Theorem 14.36 in Rockafellar and Wets (1998). This, in turn, implies, by Theorem 14.37 in Rockafellar and Wets (1998), that there exists a measurable $\boldsymbol{x}^\star$ with

$$\boldsymbol{x}^\star(y) \in \operatorname*{argmax}_{\boldsymbol{x}\in\mathrm{Fibre}^0_d(\boldsymbol{p}^P_d(y))} \boldsymbol{c}(y)^\intercal \boldsymbol{x}.$$

Since $\mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d)}$ has only zero columns and $\boldsymbol{c}_{(\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d)} \le \mathbf{0}$, we have

$$\boldsymbol{c}(y)^\intercal \boldsymbol{x}^\star(y) = h_d\left(\boldsymbol{c}(y); \boldsymbol{p}^P_d(y)\right).$$

The first claim now follows since, for every measurable $\boldsymbol{x}$ with $\boldsymbol{x}(y) \in \mathrm{Fibre}^0_d\left(\boldsymbol{p}^P_d(y)\right)$,

$$\int_{\mathcal{Y}} \|\boldsymbol{x}(y)\|_1\, d\mu(y) \le \sum_z \int_{\mathcal{Y}} p^P_{d|z}(y)\, d\mu(y) = \sum_z P(D = d \mid Z = z) \le |\mathcal{Z}|.$$

Next, since $P_Z > \mathbf{0}$, we have

$$\begin{aligned}\int_{\mathcal{Y}} |\boldsymbol{c}(y)^\intercal \boldsymbol{x}(y)|\, d\mu(y) &\le \int_{\mathcal{Y}} \|\boldsymbol{c}(y)\|_\infty \left(\sum_z p^P_{d|z}(y)\right) d\mu(y) \\ &= \mathbb{E}_P\left[\frac{\mathbb{1}\{D=d\}\|\boldsymbol{c}(Y)\|_\infty}{P_Z(Z)}\right] \\ &\le \frac{1}{\min_z P_Z(z)} \mathbb{E}_P\left[\|\boldsymbol{c}(Y)\|_\infty\right] < \infty\,,\end{aligned}$$

which gives the second claim since the right hand side doesn't depend on $\boldsymbol{x}$, and we already have $\boldsymbol{c}(y)^\intercal \boldsymbol{x}^\star(y) = h_d\left(\boldsymbol{c}(y); \boldsymbol{p}^P_d(y)\right)$.

Finally, note that

$$\begin{aligned}h_d\left(\boldsymbol{c}(y); \boldsymbol{p}^P_d(y)\right) - h_d\left(\boldsymbol{c}'(y); \boldsymbol{p}^P_d(y)\right) &\le (\boldsymbol{c}(y) - \boldsymbol{c}'(y))^\intercal \boldsymbol{x}^\star(y) \\ &\le \left(\sum_z p^P_{d|z}(y)\right) \left\|(\boldsymbol{c}(y) - \boldsymbol{c}'(y))_{(\boldsymbol{\tau}_d)}\right\|_\infty\,,\end{aligned}$$

and the result follows by noting that, since switching the roles of $\boldsymbol{c}$ and $\boldsymbol{c}'$ gives the reverse inequality, the above continues to hold if we take the absolute value on the left hand side, so we can integrate the absolute value of the left hand side and apply the integral triangle inequality to get the result. ■

Lemma A.3. *Let $\boldsymbol{H} \in \{0,1\}^{m\times n}$ be an arbitrary binary matrix with no zero columns. Then, there exists $N_{\boldsymbol{H}} < \infty$, depending only on $\boldsymbol{H}$, such that, for every $\boldsymbol{b} \in L^1(\mu; \mathbb{R}^m)$ with $\boldsymbol{b}(y) \in \operatorname{cone} \boldsymbol{H}$, there is a measurable partition $\mathcal{B}$ of $\mathcal{Y}$, with $|\mathcal{B}| \le N_{\boldsymbol{H}}$, such that, for every $B \in \mathcal{B}$ and every measurable $A \subseteq B$,*

$$\left\{\boldsymbol{u} \in \mathbb{R}^n_{\ge 0} \,\middle|\, \boldsymbol{H}\boldsymbol{u} = \int_A \boldsymbol{b}\, d\mu\right\} = \left\{\boldsymbol{u} \in \mathbb{R}^n_{\ge 0} \,\middle|\, \begin{array}{l} \exists\, \boldsymbol{x}_A \in L^1(\mu; \mathbb{R}^n_{\ge 0}) \text{ such that} \\ \displaystyle\int_A \boldsymbol{x}_A\, d\mu = \boldsymbol{u}, \\ \boldsymbol{H}\boldsymbol{x}_A(y) = \boldsymbol{b}(y)\ \forall y \in A, \\ \boldsymbol{x}_{A;(j)}(y) > 0 \iff \boldsymbol{u}_{(j)} > 0\ \forall y \in A, j \in [n] \end{array}\right\}.$$

Proof. We will first partition the possible values of $\boldsymbol{b}(y)$ so that the same linear solution formulas work throughout each cell. Their linearity will then let us construct a pointwise solution with the prescribed integral.

Since $\boldsymbol{H}$ has no zero columns, the "fibre" $\mathcal{H}(\boldsymbol{r}) := \{\boldsymbol{x} \ge \mathbf{0} \mid \boldsymbol{H}\boldsymbol{x} = \boldsymbol{r}\}$ is compact for any $r \in \operatorname{cone} \boldsymbol{H}$ and so, is a polytope. For each $\boldsymbol{r} \in \operatorname{cone} \boldsymbol{H}$, we can view the fibre $\mathcal{H}(\boldsymbol{r})$ as the feasible region of a linear program. Associated with every such fibre is a "normal fan",[34] which is a collection of cones such that the relative interiors of the cones in the normal fan partitions the space of objective functions (i.e. the space of linear functionals on $\mathcal{H}(\boldsymbol{r})$) into functions that are maximized at the same face of $\mathcal{H}(\boldsymbol{r})$. Now, by Theorem 5.4.7 in De Loera

[34] See Definition 2.1.8 in De Loera, Rambau, and Santos (2010).

et al. (2010),[35] there is a finite partition $\mathscr{C}_{\boldsymbol{H}}$ of $\operatorname{cone} \boldsymbol{H}$ into relatively open polyhedral cones ("chambers") such that, within each chamber C, the fibres $\{\mathcal{H}(\boldsymbol{r})\}_{\boldsymbol{r}\in C}$ are normally equivalent, i.e. they have the same normal fan. In particular, given $\boldsymbol{r}_1, \boldsymbol{r}_2 \in C$ and any objective function $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{x}$, the set of optimal bases[36] is the same at $\mathcal{H}(\boldsymbol{r}_1)$ and $\mathcal{H}(\boldsymbol{r}_2)$. In fact, the set of bases that are feasible at $\mathcal{H}(\boldsymbol{r})$ is invariant over $\boldsymbol{r} \in C$. Let $\mathscr{J}_C$ denote this common set of feasible bases:

$$\mathscr{J}_C := \left\{ J \subseteq [n] \;\middle|\; \begin{array}{l} |J| = \operatorname{rank} \boldsymbol{H}_{(\cdot,J)} = \operatorname{rank} \boldsymbol{H}, \\ C \subseteq \operatorname{cone} \boldsymbol{H}_{(\cdot,J)} \end{array} \right\}.$$

For every $J \in \mathscr{J}_C$, the basic solution $\boldsymbol{H}_J^{-1}\boldsymbol{r}$ depends linearly on $\boldsymbol{r}$ and is nonnegative for every $\boldsymbol{r} \in C$. By the basic feasible solution characterization of optimal solutions to linear programs (see, e.g. Theorem 4.4.1 Matousek and Gärtner (2007)), every linear objective attains its maximum over $\mathcal{H}(\boldsymbol{r})$ at one of these solutions, which are the vertices of $\mathcal{H}(\boldsymbol{r})$ and so,

$$\mathcal{H}(\boldsymbol{r}) = \operatorname{conv}\left\{ \boldsymbol{H}_J^{-1}\boldsymbol{r} \,\middle|\, J \in \mathscr{J}_C \right\} \qquad \forall \boldsymbol{r} \in C.$$

If needed, set $\boldsymbol{b}(y) = \boldsymbol{0}$ on the measurable null set

$$N := \{ y \in \mathcal{Y} \,|\, \boldsymbol{b}(y) \notin \operatorname{cone} \boldsymbol{H} \}.$$

This changes none of the integrals or almost-everywhere assertions in the claim, and ensures that $\boldsymbol{b}(y) \in \operatorname{cone} \boldsymbol{H}$ for every $y \in \mathcal{Y}$. Let $\mathcal{B}$ consist of the nonempty sets

$$B_C := \{ y \in \mathcal{Y} \,|\, \boldsymbol{b}(y) \in C \}, \qquad C \in \mathscr{C}_{\boldsymbol{H}}.$$

Since the chambers are Borel sets, $\mathcal{B}$ is a measurable partition of $\mathcal{Y}$, with

$$|\mathcal{B}| \le N_{\boldsymbol{H}} := |\mathscr{C}_{\boldsymbol{H}}|.$$

Fix $C \in \mathscr{C}_{\boldsymbol{H}}$, measurable $A \subseteq B_C$, and $\boldsymbol{u} \ge \boldsymbol{0}$ satisfying

$$\boldsymbol{H}\boldsymbol{u} = \boldsymbol{b}_A := \int_A \boldsymbol{b}\, d\mu.$$

If $\mu(A) = 0$, then

$$0 \le \|\boldsymbol{u}\|_1 \le \boldsymbol{1}^{\mathsf{T}}\boldsymbol{H}\boldsymbol{u} = 0,$$

[35] Also see Section 1.2 and the discussion following Corollary 9.5.7 in De Loera et al. (2010).

[36] A feasible basis at $\boldsymbol{r}$ is a subset $B \subseteq [n]$ such that the submatrix $\boldsymbol{H}_B$ is invertible (assume, for simplicity, that $\boldsymbol{H}$ has full row rank, otherwise this remark applies to any submatrix of $\boldsymbol{H}$ consisting of linearly independent rows), and $\boldsymbol{H}_B^{-1}\boldsymbol{r} \ge 0$, where $\boldsymbol{H}_B^{-1}\boldsymbol{r}$ is defined, in an abuse of notation, as the vector $\boldsymbol{x}$ such that $\boldsymbol{x}_{[n]\setminus B} = \boldsymbol{0}$ and $\boldsymbol{x}_B = \boldsymbol{H}_B^{-1}\boldsymbol{r}$. For any objective function $\boldsymbol{c} \in \mathbb{R}^n$ and $\boldsymbol{r} \in \operatorname{cone} \boldsymbol{H}$, there exists a basis B such that $\boldsymbol{H}_B^{-1}\boldsymbol{r} \in \mathcal{H}(\boldsymbol{r})$ is an optimal solution to the linear programming problem $\max\{\boldsymbol{c}^{\mathsf{T}}\boldsymbol{x} \,|\, \boldsymbol{H}\boldsymbol{x} = \boldsymbol{r}, \boldsymbol{x} \ge 0\}$, in which case it is called an optimal basic solution and we call B an optimal basis (which then also implies that $\boldsymbol{c}^{\mathsf{T}}\boldsymbol{H}_B^{-1}\boldsymbol{r}$ is the optimal value for this LP).

and we can take $\boldsymbol{x}_A = \mathbf{0}$. Suppose now that $\mu(A) > 0$. Since C is a relatively open polyhedral cone, there exist matrices $\mathbf{E}_C, \mathbf{G}_C$ with m columns such that

$$C = \{\boldsymbol{r} \in \mathbb{R}^m \,|\, \mathbf{E}_C \boldsymbol{r} = \mathbf{0}, \quad \mathbf{G}_C \boldsymbol{r} > \mathbf{0}\}.$$

Since $\boldsymbol{b}(y) \in C$ for every $y \in A$, integration gives

$$\mathbf{E}_C \boldsymbol{b}_A = \int_A \mathbf{E}_C \boldsymbol{b}\, d\mu = \mathbf{0},$$
$$\mathbf{G}_C \boldsymbol{b}_A = \int_A \mathbf{G}_C \boldsymbol{b}\, d\mu > \mathbf{0},$$

where strict positivity follows from $\mu(A) > 0$, and so $\boldsymbol{b}_A \in C$.

The convex-hull representation of $\mathcal{H}(\boldsymbol{b}_A)$ now gives coefficients $\alpha_J \geq 0$, $J \in \mathscr{J}_C$, satisfying

$$\sum_{J \in \mathscr{J}_C} \alpha_J = 1, \qquad \boldsymbol{u} = \sum_{J \in \mathscr{J}_C} \alpha_J \boldsymbol{H}_J^{-1} \boldsymbol{b}_A.$$

Use these same coefficients at every $y \in A$ to define

$$\boldsymbol{x}_A(y) := \begin{cases} \displaystyle\sum_{J \in \mathscr{J}_C} \alpha_J \boldsymbol{H}_J^{-1} \boldsymbol{b}(y) & \text{if } y \in A, \\ \mathbf{0} & \text{if } y \in \mathcal{Y} \setminus A. \end{cases}$$

This map is measurable and nonnegative, and

$$\boldsymbol{H} \boldsymbol{x}_A(y) = \sum_{J \in \mathscr{J}_C} \alpha_J \boldsymbol{b}(y) = \boldsymbol{b}(y) \qquad \forall y \in A.$$

Since every column of $\boldsymbol{H}$ contains a 1,

$$\begin{aligned} \int_{\mathcal{Y}} \|\boldsymbol{x}_A(y)\|_1 \, d\mu(y) &\leq \int_A \mathbf{1}^{\mathsf{T}} \boldsymbol{H} \boldsymbol{x}_A(y) \, d\mu(y) \\ &= \int_A \|\boldsymbol{b}(y)\|_1 \, d\mu(y) < \infty, \end{aligned}$$

and linearity gives

$$\int_A \boldsymbol{x}_A \, d\mu = \sum_{J \in \mathscr{J}_C} \alpha_J \boldsymbol{H}_J^{-1} \left(\int_A \boldsymbol{b}\, d\mu \right) = \boldsymbol{u}.$$

Finally, for each $j \in [n]$, the linear function

$$\ell_j(\boldsymbol{r}) := \left(\sum_{J \in \mathscr{J}_C} \alpha_J \boldsymbol{H}_J^{-1} \boldsymbol{r} \right)_{(j)}$$

is nonnegative on C, and so is either identically zero or strictly positive throughout C. Indeed, if $\ell_j(\boldsymbol{r}_0) = 0 < \ell_j(\boldsymbol{r}_1)$ for some $\boldsymbol{r}_0, \boldsymbol{r}_1 \in C$, relative openness gives $\boldsymbol{r}_0 - \epsilon \boldsymbol{r}_1 \in C$ for

sufficiently small $\epsilon > 0$, but

$$\ell_j(\boldsymbol{r}_0 - \epsilon \boldsymbol{r}_1) = -\epsilon \ell_j(\boldsymbol{r}_1) < 0,$$

a contradiction. Since $\boldsymbol{b}(y), \boldsymbol{b}_A \in C$, the $\subseteq$ direction now follows from

$$\boldsymbol{x}_{A;(j)}(y) > 0 \quad \Longleftrightarrow \quad \ell_j(\boldsymbol{b}_A) > 0 \quad \Longleftrightarrow \quad \boldsymbol{u}_{(j)} > 0 \qquad \forall y \in A,\ j \in [n].$$

For the reverse inclusion, any $\boldsymbol{x}_A$ satisfying the conditions on the right gives

$$\boldsymbol{H}\boldsymbol{u} = \boldsymbol{H} \int_A \boldsymbol{x}_A \, d\mu = \int_A \boldsymbol{H}\boldsymbol{x}_A \, d\mu = \int_A \boldsymbol{b} \, d\mu,$$

as required. ■

Corollary A.1. *Fix $d \in \mathcal{D}$ and $\boldsymbol{b} \in L^1(\mu; \mathbb{R}^{\mathcal{Z}}_{\geq 0})$. Let $\mathcal{B}_d(\boldsymbol{b})$ be a partition supplied by Lemma A.3 for*

$$\boldsymbol{H}_d := \begin{bmatrix} \mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}_d)} & \boldsymbol{I}_{|\mathcal{Z}|} \end{bmatrix}.$$

Then, $\sum_{B \in \mathcal{B}_d(\boldsymbol{b})} \left(\mathrm{Fibre}_d\left(\int_B \boldsymbol{b}\, d\mu\right)\right)_{(\boldsymbol{\mathcal{T}}_d)}$ is a nonempty polytope, and

$$\int_{\mathcal{Y}} \mathrm{Fibre}_d(\boldsymbol{b}(y))\, d\mu(y) = \left\{ \boldsymbol{u} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\geq 0} \,\middle|\, \boldsymbol{u}_{(\boldsymbol{\mathcal{T}}_d)} \in \sum_{B \in \mathcal{B}_d(\boldsymbol{b})} \left(\mathrm{Fibre}_d\left(\int_B \boldsymbol{b}\, d\mu \right) \right)_{(\boldsymbol{\mathcal{T}}_d)} \right\}.$$

where t

Proof. The matrix $\boldsymbol{H}_d$ has no zero columns, and its identity block ensures that $\boldsymbol{b}(y) \in \mathrm{cone}\, \boldsymbol{H}_d$. For every $B \in \mathcal{B}_d(\boldsymbol{b})$,

$$\int_B \left(\mathrm{Fibre}_d(\boldsymbol{b}(y))\right)_{(\boldsymbol{\mathcal{T}}_d)} \, d\mu(y) = \left(\mathrm{Fibre}_d\left(\int_B \boldsymbol{b}\, d\mu \right) \right)_{(\boldsymbol{\mathcal{T}}_d)}.$$

To prove $\subseteq$, take $\boldsymbol{v} \in \int_B \left(\mathrm{Fibre}_d(\boldsymbol{b}(y))\right)_{(\boldsymbol{\mathcal{T}}_d)} \, d\mu(y)$. By the definition of the Minkowski integral, there exists $\boldsymbol{x} \in L^1(\mu; \mathbb{R}^{\boldsymbol{\mathcal{T}}_d}_{\geq 0})$ such that $\mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}_d)}\boldsymbol{x}(y) \leq \boldsymbol{b}(y)\ \forall y \in B$, with

$$\boldsymbol{v} = \int_B \boldsymbol{x}\, d\mu\,,$$

and integrating both sides of the inequalities satisfied by this $\boldsymbol{x}$ gives $\boldsymbol{v} \geq \mathbf{0}$ and

$$\mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}_d)}\boldsymbol{v} = \int_B \mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}_d)}\boldsymbol{x}\, d\mu \leq \int_B \boldsymbol{b}\, d\mu,$$

and so $\boldsymbol{v} \in \left(\mathrm{Fibre}_d\left(\int_B \boldsymbol{b}\, d\mu\right)\right)_{(\boldsymbol{\mathcal{T}}_d)}$.

Conversely, take $\boldsymbol{v} \in \left(\mathrm{Fibre}_d\left(\int_B \boldsymbol{b}\, d\mu\right)\right)_{(\boldsymbol{\mathcal{T}}_d)}$. The vector

$$\begin{bmatrix} \boldsymbol{v} \\ \int_B \boldsymbol{b}\, d\mu - \mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}_d)}\boldsymbol{v} \end{bmatrix}$$

is nonnegative and has image $\int_B \boldsymbol{b}\, d\mu$ under $\boldsymbol{H}_d$. By Lemma A.3, there exists $\boldsymbol{x}_B \in L^1(\mu; \mathbb{R}^{\mathcal{T}_d}_{\geq 0})$ satisfying

$$\mathbf{R}_{d;(\cdot,\mathcal{T}_d)} \boldsymbol{x}_B(y) \leq \boldsymbol{b}(y) \qquad \forall y \in B,$$
$$\int_B \boldsymbol{x}_B \, d\mu = \boldsymbol{v},$$

and so $\boldsymbol{v}$ belongs to the Minkowski integral on the left.

Restricting selections to the cells, and combining them as $\sum_{B \in \mathcal{B}_d(\boldsymbol{b})} \mathbb{1}_B \boldsymbol{x}_B$, we get

$$\int_{\mathcal{Y}} (\mathrm{Fibre}_d(\boldsymbol{b}(y)))_{(\mathcal{T}_d)} \, d\mu(y) = \sum_{B \in \mathcal{B}_d(\boldsymbol{b})} \left(\mathrm{Fibre}_d \left(\int_B \boldsymbol{b}\, d\mu \right) \right)_{(\mathcal{T}_d)}.$$

The columns of $\mathbf{R}_d$ outside $\mathcal{T}_d$ are zero. Thus, any selection in this last Minkowski integral can be extended to a selection of $\mathrm{Fibre}_d(\boldsymbol{b}(y))$ with any prescribed nonnegative integral $\boldsymbol{u}_{(\mathcal{T} \setminus \mathcal{T}_d)}$ in the remaining coordinates, by setting $\boldsymbol{x}_{(\mathcal{T} \setminus \mathcal{T}_d)}(y) = \rho_0(y) \boldsymbol{u}_{(\mathcal{T} \setminus \mathcal{T}_d)}$, where ρ_0 is the probability density from Lemma A.1. Conversely, every integrable selection of the full fibre restricts to an integrable selection of its $\mathcal{T}_d$ coordinates, and so the claimed equality follows.

Finally, each summand is a polyhedron containing zero, and every $\boldsymbol{v}$ in the summand indexed by B satisfies

$$\|\boldsymbol{v}\|_1 \leq \mathbf{1}^\intercal \mathbf{R}_{d;(\cdot,\mathcal{T}_d)} \boldsymbol{v} \leq \int_B \|\boldsymbol{b}(y)\|_1 \, d\mu(y) < \infty.$$

Each summand is therefore a nonempty polytope, and so is their finite Minkowski sum. ■

Lemma A.4. *Fix $d \in \mathcal{D}$. There exists a finite complete polyhedral fan $\mathscr{N}_d$ in $\mathbb{R}^{\mathcal{T}_d}$, consisting of pointed cones*[37] *and depending only on $\mathbf{R}_d$, such that, for every $\Gamma \in \mathscr{N}_d$ and $\boldsymbol{r} \in \mathbb{R}^{\mathcal{Z}}_{\geq 0}$, there exists $\boldsymbol{v}_{\Gamma,\boldsymbol{r}} \in (\mathrm{Fibre}_d(\boldsymbol{r}))_{(\mathcal{T}_d)}$ satisfying*

$$\max_{\boldsymbol{x} \in (\mathrm{Fibre}_d(\boldsymbol{r}))_{(\mathcal{T}_d)}} \boldsymbol{a}^\intercal \boldsymbol{x} = \boldsymbol{a}^\intercal \boldsymbol{v}_{\Gamma,\boldsymbol{r}} \qquad \forall \boldsymbol{a} \in \Gamma.$$

Moreover, there exist a nonnegative integer k_d and vectors $\boldsymbol{a}_{d,j} \in \mathbb{R}^{\mathcal{T}_d} \setminus \{\mathbf{0}\}$, $j \in [k_d]$, depending only on $\mathbf{R}_d$, such that every $\Gamma \in \mathscr{N}_d$ admits $J_\Gamma \subseteq [k_d]$ with

$$\Gamma = \left\{ \sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_{d,j} \;\middle|\; \alpha_j \geq 0 \; \forall j \in J_\Gamma \right\}.$$

Proof. For every $\boldsymbol{r} \in \mathbb{R}^{\mathcal{Z}}_{\geq 0}$, the set $(\mathrm{Fibre}_d(\boldsymbol{r}))_{(\mathcal{T}_d)}$ contains $\mathbf{0}$, and

$$\|\boldsymbol{x}\|_1 \leq \mathbf{1}^\intercal \mathbf{R}_{d;(\cdot,\mathcal{T}_d)} \boldsymbol{x} \leq \|\boldsymbol{r}\|_1 \qquad \forall \boldsymbol{x} \in (\mathrm{Fibre}_d(\boldsymbol{r}))_{(\mathcal{T}_d)},$$

[37] A cone is said to be "pointed" if it does not contain any line (i.e. no full line passing through origin).

and so $(\mathrm{Fibre}_d(\boldsymbol{r}))_{(\boldsymbol{\mathcal{T}}_d)}$ is a nonempty polytope. Denote its normal fan by $\mathscr{N}_{d,\boldsymbol{r}}$, and its normal cone at a vertex $\boldsymbol{v}$ by $N_{d,\boldsymbol{r}}(\boldsymbol{v})$. By Farkas' lemma,

$$N_{d,\boldsymbol{r}}(\boldsymbol{v}) = \operatorname{cone}\left(\left\{\mathbf{R}^{\mathsf{T}}_{d;(z,\boldsymbol{\mathcal{T}}_d)} \,\middle|\, z \in \mathcal{Z},\ \mathbf{R}_{d;(z,\boldsymbol{\mathcal{T}}_d)}\boldsymbol{v} = \boldsymbol{r}_{(z)}\right\} \cup \left\{-\boldsymbol{e}_{\boldsymbol{\tau}} \,\middle|\, \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d,\ \boldsymbol{v}_{(\boldsymbol{\tau})} = 0\right\}\right),$$

where $\boldsymbol{e}_{\boldsymbol{\tau};(\boldsymbol{\tau}')} = \mathbb{1}\{\boldsymbol{\tau} = \boldsymbol{\tau}'\}$ for $\boldsymbol{\tau}, \boldsymbol{\tau}' \in \boldsymbol{\mathcal{T}}_d$. Every $N_{d,\boldsymbol{r}}(\boldsymbol{v})$ is therefore generated by a subset of the fixed finite set

$$\mathcal{V}_d := \left\{\mathbf{R}^{\mathsf{T}}_{d;(z,\boldsymbol{\mathcal{T}}_d)} \,\middle|\, z \in \mathcal{Z}\right\} \cup \left\{-\boldsymbol{e}_{\boldsymbol{\tau}} \,\middle|\, \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d\right\}.$$

Since $\mathscr{N}_{d,\boldsymbol{r}}$ consists of the vertex normal cones $N_{d,\boldsymbol{r}}(\boldsymbol{v})$ and their faces, the family

$$\mathfrak{N}_d := \left\{\mathscr{N}_{d,\boldsymbol{r}} \,\middle|\, \boldsymbol{r} \in \mathbb{R}^{\mathcal{Z}}_{\geq 0}\right\}$$

is finite.

Define the coordinate orthant fan by

$$\mathscr{O}_d := \left\{\prod_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d} O_{\boldsymbol{\tau}} \,\middle|\, O_{\boldsymbol{\tau}} \in \{\{0\}, \mathbb{R}_{\geq 0}, \mathbb{R}_{\leq 0}\}\ \forall \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d\right\}.$$

Let $\mathscr{N}_d$ be the common refinement of the fans in $\mathfrak{N}_d \cup \{\mathscr{O}_d\}$; explicitly,

$$\mathscr{N}_d := \left\{O \cap \bigcap_{\mathscr{F} \in \mathfrak{N}_d} \Gamma_{\mathscr{F}} \,\middle|\, O \in \mathscr{O}_d,\ \Gamma_{\mathscr{F}} \in \mathscr{F}\ \forall \mathscr{F} \in \mathfrak{N}_d\right\}.$$

The fan $\mathscr{N}_d$ is finite and complete because every fan in the finite family $\mathfrak{N}_d \cup \{\mathscr{O}_d\}$ is finite and complete. Every $\Gamma \in \mathscr{N}_d$ is contained in some $O \in \mathscr{O}_d$, and so

$$\Gamma \cap (-\Gamma) \subseteq O \cap (-O) = \{\mathbf{0}\},$$

which establishes pointedness. The construction of $\mathscr{N}_d$ depends only on $\mathbf{R}_d$.

For every $\Gamma \in \mathscr{N}_d$ and $\boldsymbol{r} \in \mathbb{R}^{\mathcal{Z}}_{\geq 0}$, refinement of $\mathscr{N}_{d,\boldsymbol{r}}$ gives a vertex $\boldsymbol{v}_{\Gamma,\boldsymbol{r}}$ of $(\mathrm{Fibre}_d(\boldsymbol{r}))_{(\boldsymbol{\mathcal{T}}_d)}$ such that

$$\Gamma \subseteq N_{d,\boldsymbol{r}}(\boldsymbol{v}_{\Gamma,\boldsymbol{r}}).$$

By the definition of $N_{d,\boldsymbol{r}}(\boldsymbol{v}_{\Gamma,\boldsymbol{r}})$,

$$\boldsymbol{a}^{\mathsf{T}}(\boldsymbol{x} - \boldsymbol{v}_{\Gamma,\boldsymbol{r}}) \leq 0 \qquad \forall \boldsymbol{a} \in \Gamma,\ \boldsymbol{x} \in (\mathrm{Fibre}_d(\boldsymbol{r}))_{(\boldsymbol{\mathcal{T}}_d)},$$

and so

$$\max_{\boldsymbol{x} \in (\mathrm{Fibre}_d(\boldsymbol{r}))_{(\boldsymbol{\mathcal{T}}_d)}} \boldsymbol{a}^{\mathsf{T}}\boldsymbol{x} = \boldsymbol{a}^{\mathsf{T}}\boldsymbol{v}_{\Gamma,\boldsymbol{r}} \qquad \forall \boldsymbol{a} \in \Gamma.$$

For each distinct extreme ray of the cones in $\mathscr{N}_d$, choose a nonzero vector on the ray, enumerate the resulting finite list as $\boldsymbol{a}_{d,1}, \ldots, \boldsymbol{a}_{d,k_d}$, and define

$$J_\Gamma := \{j \in [k_d] \,|\, \mathbb{R}_{\geq 0}\boldsymbol{a}_{d,j} \text{ is an extreme ray of } \Gamma\} \qquad \forall \Gamma \in \mathscr{N}_d.$$

Every pointed polyhedral cone $\Gamma \in \mathscr{N}_d$ is generated by its extreme rays, and so

$$\Gamma = \left\{ \sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_{d,j} \,\middle|\, \alpha_j \geq 0 \; \forall j \in J_\Gamma \right\}. \qquad \blacksquare$$

Lemma A.5. *Let $K \subseteq \mathbb{R}^n$ be nonempty, compact, and convex, and let $\mathscr{N}$ be a finite complete fan of pointed polyhedral cones in $\mathbb{R}^n$. Choose one nonzero vector on each distinct extreme ray of the cones in $\mathscr{N}$, and enumerate these vectors as $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$. For every $\Gamma \in \mathscr{N}$, define*

$$J_\Gamma := \{j \in [k] \,|\, \mathbb{R}_{\geq 0}\boldsymbol{a}_j \text{ is an extreme ray of } \Gamma\}.$$

Then the following statements hold.

(i) *Suppose that, for every $\Gamma \in \mathscr{N}$,*

$$h_K\left(\sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_j\right) = \sum_{j \in J_\Gamma} \alpha_j h_K(\boldsymbol{a}_j) \qquad \forall (\alpha_j)_{j \in J_\Gamma} \in \mathbb{R}^{J_\Gamma}_{\geq 0}.$$

Then

$$K = \left\{\boldsymbol{x} \in \mathbb{R}^n \,\middle|\, \boldsymbol{a}_j^\intercal \boldsymbol{x} \leq h_K(\boldsymbol{a}_j) \; \forall j \in [k]\right\}.$$

(ii) *Let $C : \mathcal{Y} \rightrightarrows \mathbb{R}^n$ be a correspondence with nonempty compact convex values. Suppose that, for every $\boldsymbol{a} \in \mathbb{R}^n$, the function $y \mapsto h_{C(y)}(\boldsymbol{a})$ belongs to $L^1(\mu)$, and*

$$h_K(\boldsymbol{a}) = \int_{\mathcal{Y}} h_{C(y)}(\boldsymbol{a})\, d\mu(y).$$

Suppose also that

$$h_{C(y)}\left(\sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_j\right) = \sum_{j \in J_\Gamma} \alpha_j h_{C(y)}(\boldsymbol{a}_j)$$

for every $y \in \mathcal{Y}$, $\Gamma \in \mathscr{N}$, and $(\alpha_j)_{j \in J_\Gamma} \in \mathbb{R}^{J_\Gamma}_{\geq 0}$, under the standing convention for pointwise assertions in y. Then

$$K = \left\{\boldsymbol{x} \in \mathbb{R}^n \,\middle|\, \boldsymbol{a}_j^\intercal \boldsymbol{x} \leq \int_{\mathcal{Y}} h_{C(y)}(\boldsymbol{a}_j)\, d\mu(y) \; \forall j \in [k]\right\}.$$

Proof. Every $\boldsymbol{x} \in K$ satisfies $\boldsymbol{a}_j^\intercal \boldsymbol{x} \leq h_K(\boldsymbol{a}_j)$ for all $j \in [k]$, by the definition of the support function h_K.

Conversely, suppose $\boldsymbol{x} \in \mathbb{R}^n$ satisfies these inequalities. For any $\boldsymbol{a} \in \mathbb{R}^n$, completeness of $\mathscr{N}$ gives $\Gamma \in \mathscr{N}$ with $\boldsymbol{a} \in \Gamma$. Since Γ is generated by its extreme rays, there exist $\alpha_j \geq 0$,

$j \in J_\Gamma$, such that $\boldsymbol{a} = \sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_j$. The hypothesis in part 1 gives

$$\begin{aligned} \boldsymbol{a}^\intercal \boldsymbol{x} &= \sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_j^\intercal \boldsymbol{x} \\ &\leq \sum_{j \in J_\Gamma} \alpha_j h_K(\boldsymbol{a}_j) = h_K(\boldsymbol{a}). \end{aligned}$$

Thus $\boldsymbol{a}^\intercal \boldsymbol{x} \leq h_K(\boldsymbol{a})$ for every $\boldsymbol{a} \in \mathbb{R}^n$, and so hyperplane separation gives $\boldsymbol{x} \in K$.

Under the conditions of the second claim, for every $\Gamma \in \mathscr{N}$ and every $(\alpha_j)_{j \in J_\Gamma} \in \mathbb{R}^{J_\Gamma}_{\geq 0}$,

$$\begin{aligned} h_K\left(\sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_j\right) &= \int_{\mathcal{Y}} h_{C(y)}\left(\sum_{j \in J_\Gamma} \alpha_j \boldsymbol{a}_j\right) d\mu(y) \\ &= \int_{\mathcal{Y}} \sum_{j \in J_\Gamma} \alpha_j h_{C(y)}(\boldsymbol{a}_j)\, d\mu(y) \\ &= \sum_{j \in J_\Gamma} \alpha_j h_K(\boldsymbol{a}_j). \end{aligned}$$

The first claim therefore applies, and the second claim follows immediately by substituting $h_K(\boldsymbol{a}_j) = \int_{\mathcal{Y}} h_{C(y)}(\boldsymbol{a}_j)\, d\mu(y)$. ∎

Corollary A.2. *Suppose P satisfies Condition D, and fix $d \in \mathcal{D}$. For every $\boldsymbol{a} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}_d}$, define its zero extension $\widetilde{\boldsymbol{a}} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ by*

$$\widetilde{\boldsymbol{a}}_{(\boldsymbol{\mathcal{T}}_d)} = \boldsymbol{a}, \qquad \widetilde{\boldsymbol{a}}_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} = \boldsymbol{0}.$$

Let $\boldsymbol{a}_{d,1}, \ldots, \boldsymbol{a}_{d,k_d}$ be the vectors supplied by Lemma A.4, with zero extensions $\widetilde{\boldsymbol{a}}_{d,1}, \ldots, \widetilde{\boldsymbol{a}}_{d,k_d}$. Then

$$\boldsymbol{\Pi}_d = \left\{ \boldsymbol{u} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\geq 0} \,\middle|\, \widetilde{\boldsymbol{a}}_{d,j}^\intercal \boldsymbol{u} \leq \int_{\mathcal{Y}} h_d\left(\widetilde{\boldsymbol{a}}_{d,j}; \boldsymbol{p}_d^P(y)\right)\, d\mu(y)\ \forall j \in [k_d] \right\}.$$

In particular, $\boldsymbol{\Pi}_d$ is a nonempty closed convex polyhedron.

Proof. Define

$$K := (\boldsymbol{\Pi}_d)_{(\boldsymbol{\mathcal{T}}_d)}, \qquad C(y) := \left(\mathrm{Fibre}_d(\boldsymbol{p}_d^P(y))\right)_{(\boldsymbol{\mathcal{T}}_d)}.$$

By Corollary A.1, K is a nonempty polytope, and

$$\boldsymbol{\Pi}_d = \left\{ \boldsymbol{u} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\geq 0} \,\middle|\, \boldsymbol{u}_{(\boldsymbol{\mathcal{T}}_d)} \in K \right\}.$$

For every $\boldsymbol{a} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}_d}$, Lemma A.2 applied to the constant coefficient $\widetilde{\boldsymbol{a}}$ gives

$$\begin{aligned} 0 \leq h_K(\boldsymbol{a}) &= \int_{\mathcal{Y}} h_{C(y)}(\boldsymbol{a})\, d\mu(y) \\ &= \int_{\mathcal{Y}} h_d\left(\widetilde{\boldsymbol{a}}; \boldsymbol{p}_d^P(y)\right)\, d\mu(y) \end{aligned}$$

$$\leq \|\boldsymbol{a}\|_\infty \sum_{z\in\mathcal{Z}} P(D=d \mid Z=z) < \infty.$$

In particular, $y \mapsto h_{C(y)}(\boldsymbol{a})$ belongs to $L^1(\mu)$.

For every $\Gamma \in \mathscr{N}_d$, Lemma A.4 supplies $\boldsymbol{v}_{\Gamma,\boldsymbol{p}_d^P(y)} \in C(y)$ such that, for all nonnegative coefficients $(\alpha_j)_{j\in J_\Gamma}$,

$$\begin{aligned} h_{C(y)}\left(\sum_{j\in J_\Gamma} \alpha_j \boldsymbol{a}_{d,j}\right) &= \left(\sum_{j\in J_\Gamma} \alpha_j \boldsymbol{a}_{d,j}\right)^\mathsf{T} \boldsymbol{v}_{\Gamma,\boldsymbol{p}_d^P(y)} \\ &= \sum_{j\in J_\Gamma} \alpha_j h_{C(y)}(\boldsymbol{a}_{d,j}). \end{aligned}$$

Lemma A.5(ii), with $n = |\boldsymbol{\mathcal{T}}_d|$, $\mathscr{N} = \mathscr{N}_d$, and generators $\boldsymbol{a}_j = \boldsymbol{a}_{d,j}$, therefore gives

$$K = \left\{ \boldsymbol{v} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}_d} \,\middle|\, \boldsymbol{a}_{d,j}^\mathsf{T}\boldsymbol{v} \leq \int_{\mathcal{Y}} h_d\left(\widetilde{\boldsymbol{a}}_{d,j}; \boldsymbol{p}_d^P(y)\right)\, d\mu(y)\ \forall j \in [k_d] \right\}.$$

Substituting this representation of K into the displayed representation of $\boldsymbol{\Pi}_d$, and using

$$\widetilde{\boldsymbol{a}}_{d,j}^\mathsf{T}\boldsymbol{u} = \boldsymbol{a}_{d,j}^\mathsf{T}\boldsymbol{u}_{(\boldsymbol{\mathcal{T}}_d)},$$

proves the claimed equality. The resulting representation of $\boldsymbol{\Pi}_d$ is a finite intersection of closed halfspaces containing $\mathbf{0}$, and so $\boldsymbol{\Pi}_d$ is a nonempty closed convex polyhedron. ■

Lemma A.6. *Let $n_x \in \mathbb{N}$, let $g \in L^1(\mu;\mathbb{R}_{\geq 0})$, and let measurable functions $\boldsymbol{x}_n : \mathcal{Y} \to \mathbb{R}^{n_x}$, $n \in \mathbb{N}$, satisfy*

$$\|\boldsymbol{x}_n(y)\|_\infty \leq g(y) \qquad \forall n \in \mathbb{N},\ y \in \mathcal{Y}.$$

Then there exist integers $1 \leq n_1 < n_2 < \cdots$ and $\boldsymbol{x} \in L^1(\mu;\mathbb{R}^{n_x})$, with $\|\boldsymbol{x}(y)\|_\infty \leq g(y)$, such that

$$\left\| \int_{\mathcal{Y}} w(y)\big(\boldsymbol{x}_{n_k}(y) - \boldsymbol{x}(y)\big)\, d\mu(y) \right\|_1 \to 0 \qquad (k \to \infty)$$

for every measurable $w : \mathcal{Y} \to \mathbb{R}$ satisfying

$$\int_{\mathcal{Y}} |w(y)| g(y)\, d\mu(y) < \infty.$$

Moreover, for any $m \in \mathbb{N}$, matrix $\boldsymbol{G} \in \mathbb{R}^{m\times n_x}$, and measurable $\boldsymbol{b} : \mathcal{Y} \to \mathbb{R}^m$,

$$\boldsymbol{G}\boldsymbol{x}_n(y) \leq \boldsymbol{b}(y) \quad \forall n \in \mathbb{N},\ y \in \mathcal{Y} \qquad \Longrightarrow \qquad \boldsymbol{G}\boldsymbol{x}(y) \leq \boldsymbol{b}(y) \quad \forall y \in \mathcal{Y}.$$

Proof. Define the finite measure

$$\nu(A) := \int_A g\, d\mu \qquad \forall A \in \mathscr{B}_{\mathcal{Y}},$$

and, for every $n \in \mathbb{N}$, define

$$\boldsymbol{u}_n(y) := \begin{cases} \dfrac{\boldsymbol{x}_n(y)}{g(y)} & \text{if } g(y) > 0, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Then $\|\boldsymbol{u}_n(y)\|_\infty \le 1$. Since $\mathscr{B}_\mathcal{Y}$ is countably generated and ν is finite, $L^1(\nu;\mathbb{R}^{n_x})$ is separable. Banach-Alaoglu and weak-* metrizability of the dual unit ball therefore give integers $1 \le n_1 < n_2 < \cdots$ and a measurable $\boldsymbol{u}$, with $\|\boldsymbol{u}(y)\|_\infty \le 1$ for ν-almost every y, such that

$$\left\| \int_\mathcal{Y} f(y)\big(\boldsymbol{u}_{n_k}(y) - \boldsymbol{u}(y)\big)\, d\nu(y) \right\|_1 \to 0 \qquad \forall f \in L^1(\nu).$$

Set $\boldsymbol{x}(y) := g(y)\boldsymbol{u}(y)$. Then

$$\|\boldsymbol{x}(y)\|_\infty \le g(y),$$
$$\int_\mathcal{Y} \|\boldsymbol{x}(y)\|_1\, d\mu(y) \le n_x \int_\mathcal{Y} g\, d\mu < \infty.$$

For every measurable w with $\int_\mathcal{Y} |w| g\, d\mu < \infty$, we have $w \in L^1(\nu)$, and so

$$\left\| \int_\mathcal{Y} w(\boldsymbol{x}_{n_k} - \boldsymbol{x})\, d\mu \right\|_1 = \left\| \int_\mathcal{Y} w(\boldsymbol{u}_{n_k} - \boldsymbol{u})\, d\nu \right\|_1 \to 0.$$

Now suppose $\boldsymbol{G}\boldsymbol{x}_n \le \boldsymbol{b}$ for every n. For each $i \in [m]$, define

$$A_i := \left\{ y \in \mathcal{Y} \,\middle|\, \boldsymbol{G}_{(i,\cdot)}\boldsymbol{x}(y) > \boldsymbol{b}_{(i)}(y) \right\}.$$

For every $y \in A_i$,

$$-\|\boldsymbol{G}_{(i,\cdot)}\|_1 g(y) \le \boldsymbol{G}_{(i,\cdot)}\boldsymbol{x}_1(y) \le \boldsymbol{b}_{(i)}(y),$$
$$\boldsymbol{b}_{(i)}(y) < \boldsymbol{G}_{(i,\cdot)}\boldsymbol{x}(y) \le \|\boldsymbol{G}_{(i,\cdot)}\|_1 g(y),$$

and so $\mathbb{1}_{A_i}\boldsymbol{b}_{(i)} \in L^1(\mu)$. Applying the established moment convergence with $w = \mathbb{1}_{A_i}$ gives

$$\begin{aligned} 0 \le & \int_{A_i} \big(\boldsymbol{G}_{(i,\cdot)}\boldsymbol{x} - \boldsymbol{b}_{(i)}\big)\, d\mu \\ &= \lim_{k\to\infty} \int_{A_i} \big(\boldsymbol{G}_{(i,\cdot)}\boldsymbol{x}_{n_k} - \boldsymbol{b}_{(i)}\big)\, d\mu \\ &\le 0. \end{aligned}$$

Since $\boldsymbol{G}_{(i,\cdot)}\boldsymbol{x}(y) - \boldsymbol{b}_{(i)}(y) > 0$ on A_i, this implies $\mu(A_i) = 0$. Applying this argument for every $i \in [m]$ gives $\boldsymbol{G}\boldsymbol{x} \le \boldsymbol{b}$, as required. ■

Corollary A.3. *Let $n_x, m_G, m \in \mathbb{N}$, $g \in L^1(\mu;\mathbb{R}_{\ge 0})$, and $\boldsymbol{G} \in \mathbb{R}^{m_G \times n_x}$. Let $\boldsymbol{b} : \mathcal{Y} \to \mathbb{R}^{m_G}$ and $\boldsymbol{W} : \mathcal{Y} \to \mathbb{R}^{m \times n_x}$ be measurable, with*

$$\int_\mathcal{Y} |\boldsymbol{W}_{(i,j)}(y)| g(y)\, d\mu(y) < \infty \qquad \forall i \in [m],\ j \in [n_x].$$

Define

$$\mathcal{X} := \left\{ \boldsymbol{x} \in L^1(\mu; \mathbb{R}^{n_x}) \,\middle|\, \|\boldsymbol{x}(y)\|_\infty \le g(y),\ \boldsymbol{G}\boldsymbol{x}(y) \le \boldsymbol{b}(y)\ \forall y \in \mathcal{Y} \right\}.$$

Then

$$\mathcal{J} := \left\{ \int_{\mathcal{Y}} \boldsymbol{W}(y)\boldsymbol{x}(y)\, d\mu(y) \,\middle|\, \boldsymbol{x} \in \mathcal{X} \right\}$$

is compact and convex (potentially empty).

Proof. For every $\boldsymbol{x} \in \mathcal{X}$,

$$\int_{\mathcal{Y}} \|\boldsymbol{W}(y)\boldsymbol{x}(y)\|_1\, d\mu(y) \le \sum_{i=1}^{m}\sum_{j=1}^{n_x} \int_{\mathcal{Y}} |\boldsymbol{W}_{(i,j)}(y)| g(y)\, d\mu(y) < \infty,$$

and so $\mathcal{J} \subseteq \mathbb{R}^m$ is well defined. If $\mathcal{X} = \emptyset$, then $\mathcal{J} = \emptyset$ is compact and convex.

Otherwise, take any sequence $(\boldsymbol{z}_n)_{n\in\mathbb{N}} \subseteq \mathcal{J}$, and choose $\boldsymbol{x}_n \in \mathcal{X}$ with

$$\boldsymbol{z}_n = \int_{\mathcal{Y}} \boldsymbol{W}(y)\boldsymbol{x}_n(y)\, d\mu(y) \qquad \forall n \in \mathbb{N}.$$

By Lemma A.6, there exist integers $1 \le n_1 < n_2 < \cdots$ and $\boldsymbol{x} \in \mathcal{X}$ such that, taking $w(y) = \boldsymbol{W}_{(i,j)}(y)$,

$$\int_{\mathcal{Y}} \boldsymbol{W}_{(i,j)}(y)\big(\boldsymbol{x}_{n_k;(j)}(y) - \boldsymbol{x}_{(j)}(y)\big)\, d\mu(y) \to 0 \qquad \forall i \in [m],\ j \in [n_x].$$

Therefore,

$$\begin{aligned}\left\| \boldsymbol{z}_{n_k} - \int_{\mathcal{Y}} \boldsymbol{W}(y)\boldsymbol{x}(y)\, d\mu(y) \right\|_1 &\le \sum_{i=1}^{m}\sum_{j=1}^{n_x} \left| \int_{\mathcal{Y}} \boldsymbol{W}_{(i,j)}(y)\big(\boldsymbol{x}_{n_k;(j)}(y) - \boldsymbol{x}_{(j)}(y)\big)\, d\mu(y) \right| \\ &\to 0.\end{aligned}$$

Since $\boldsymbol{x} \in \mathcal{X}$, the limit $\int_{\mathcal{Y}} \boldsymbol{W}\boldsymbol{x}\, d\mu$ belongs to $\mathcal{J}$. Thus $\mathcal{J}$ is sequentially compact in $\mathbb{R}^m$, and hence compact.

For $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ and $t \in [0,1]$, set $\boldsymbol{x}_t := t\boldsymbol{x} + (1-t)\boldsymbol{x}'$. Then $\boldsymbol{x}_t \in L^1(\mu; \mathbb{R}^{n_x})$, and

$$\begin{aligned}\|\boldsymbol{x}_t(y)\|_\infty &\le t\|\boldsymbol{x}(y)\|_\infty + (1-t)\|\boldsymbol{x}'(y)\|_\infty \le g(y), \\ \boldsymbol{G}\boldsymbol{x}_t(y) &= t\boldsymbol{G}\boldsymbol{x}(y) + (1-t)\boldsymbol{G}\boldsymbol{x}'(y) \le \boldsymbol{b}(y),\end{aligned}$$

and so $\boldsymbol{x}_t \in \mathcal{X}$. Therefore,

$$t\int_{\mathcal{Y}} \boldsymbol{W}\boldsymbol{x}\, d\mu + (1-t)\int_{\mathcal{Y}} \boldsymbol{W}\boldsymbol{x}'\, d\mu = \int_{\mathcal{Y}} \boldsymbol{W}\boldsymbol{x}_t\, d\mu \in \mathcal{J},$$

which proves convexity of $\mathcal{J}$. ∎

Lemma A.7. *Fix $d \in \mathcal{D}$. There exist a finite nonempty index set $\mathscr{J}_d$, matrices $\boldsymbol{B}_{d,J} \in \mathbb{R}^{\mathcal{Z} \times \boldsymbol{\mathcal{T}}}$, $J \in \mathscr{J}_d$, depending only on $\mathbf{R}_d$, and a jointly Borel function*

$$\boldsymbol{\ell}^\star_d : \left\{ \boldsymbol{c} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}} \,\middle|\, \boldsymbol{c}_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} \le \mathbf{0} \right\} \times \mathbb{R}^{\mathcal{Z}}_{\ge 0} \to \mathbb{R}^{\mathcal{Z}}_{\ge 0}$$

such that, for every $(\boldsymbol{c}, \boldsymbol{b})$ in its domain,

$$\begin{aligned} \mathbf{R}_d^\intercal \boldsymbol{\ell}^\star_d(\boldsymbol{c}; \boldsymbol{b}) &\ge \boldsymbol{c}, \\ \boldsymbol{b}^\intercal \boldsymbol{\ell}^\star_d(\boldsymbol{c}; \boldsymbol{b}) &= h_d(\boldsymbol{c}; \boldsymbol{b}), \\ \|\boldsymbol{\ell}^\star_d(\boldsymbol{c}; \boldsymbol{b})\|_\infty &\le \|\boldsymbol{c}\|_\infty, \\ \boldsymbol{\ell}^\star_d(\boldsymbol{c}; \boldsymbol{b}) &\in \left\{ \boldsymbol{B}_{d,J} \boldsymbol{c} \,|\, J \in \mathscr{J}_d \right\}. \end{aligned}$$

Proof. Define

$$\boldsymbol{G}_d := \begin{bmatrix} \boldsymbol{I}_{|\mathcal{Z}|} \\ \mathbf{R}_d^\intercal \end{bmatrix}, \qquad \boldsymbol{E}_d := \begin{bmatrix} \mathbf{0}_{|\mathcal{Z}| \times |\boldsymbol{\mathcal{T}}|} \\ \boldsymbol{I}_{|\boldsymbol{\mathcal{T}}|} \end{bmatrix},$$

$$\mathscr{J}_d := \left\{ J \subseteq [|\mathcal{Z}| + |\boldsymbol{\mathcal{T}}|] \,\middle|\, |J| = |\mathcal{Z}|,\ \det \boldsymbol{G}_{d;(J,\cdot)} \neq 0 \right\}.$$

For every $J \in \mathscr{J}_d$, set

$$\boldsymbol{B}_{d,J} := \boldsymbol{G}^{-1}_{d;(J,\cdot)} \boldsymbol{E}_{d;(J,\cdot)}. \tag{A.6}$$

Note that $\mathscr{J}_d$ is finite and $[|\mathcal{Z}|] \in \mathscr{J}_d$.

Fix $(\boldsymbol{c}, \boldsymbol{b})$ in the specified domain of $\boldsymbol{\ell}^\star_d$. The vector $\|\boldsymbol{c}\|_\infty \mathbf{1}_{\mathcal{Z}}$ is dual feasible, since every column of $\mathbf{R}_d$ indexed by $\boldsymbol{\mathcal{T}}_d$ contains a one, and the other columns are zero with corresponding coefficients nonpositive. Since $\boldsymbol{b} \ge \mathbf{0}$, the dual objective $\boldsymbol{b}^\intercal \boldsymbol{\ell}$ is bounded below by zero. The dual admits an optimal vertex $\boldsymbol{\ell}$ of value $h_d(\boldsymbol{c}; \boldsymbol{b})$, and some $J \in \mathscr{J}_d$ satisfies

$$\boldsymbol{G}_{d;(J,\cdot)} \boldsymbol{\ell} = \boldsymbol{E}_{d;(J,\cdot)} \boldsymbol{c}, \qquad \boldsymbol{\ell} = \boldsymbol{B}_{d,J} \boldsymbol{c}\,.$$

Consider any $J \in \mathscr{J}_d$ for which $\boldsymbol{\ell} := \boldsymbol{B}_{d,J} \boldsymbol{c}$ is dual feasible. By construction,

$$\boldsymbol{G}_{d;(J,\cdot)} \boldsymbol{\ell} = \boldsymbol{E}_{d;(J,\cdot)} \boldsymbol{c}\,.$$

If $\boldsymbol{\ell}_{(z)} > 0$, the identity row for coordinate z cannot belong to this system, since its equation would be $\boldsymbol{\ell}_{(z)} = 0$. Since $\boldsymbol{G}_{d;(J,\cdot)}$ is invertible, its z-column is nonzero, and so some row $\mathbf{R}^\intercal_{d;(\cdot,\boldsymbol{\tau})}$ in this system has $\mathbf{R}_{d;(z,\boldsymbol{\tau})} = 1$. Its equation and the nonnegativity of $\boldsymbol{\ell}$ give

$$\begin{aligned} 0 < \boldsymbol{\ell}_{(z)} &\le \sum_{z' \in \mathcal{Z}} \mathbf{R}_{d;(z',\boldsymbol{\tau})} \boldsymbol{\ell}_{(z')} \\ &= \boldsymbol{c}_{(\boldsymbol{\tau})} \le \|\boldsymbol{c}\|_\infty\,, \end{aligned}$$

and so every dual-feasible $\boldsymbol{B}_{d,J} \boldsymbol{c}$ satisfies $\|\boldsymbol{B}_{d,J} \boldsymbol{c}\|_\infty \le \|\boldsymbol{c}\|_\infty$.

Fix an enumeration $\mathscr{J}_d = \{J_1, \dots, J_{N_d}\}$, and define

$$v_j(\boldsymbol{c}, \boldsymbol{b}) := \begin{cases} \boldsymbol{b}^{\mathsf{T}} \boldsymbol{B}_{d,J_j} \boldsymbol{c} & \text{if } \boldsymbol{G}_d \boldsymbol{B}_{d,J_j} \boldsymbol{c} \geq \boldsymbol{E}_d \boldsymbol{c}, \\ \infty & \text{otherwise.} \end{cases}$$

Each v_j is Borel, and their minimum is finite on the stated domain. Set

$$j^\star(\boldsymbol{c}, \boldsymbol{b}) := \min \operatorname*{argmin}_{j \in [N_d]} v_j(\boldsymbol{c}, \boldsymbol{b}),$$

$$\boldsymbol{\ell}_d^\star(\boldsymbol{c}; \boldsymbol{b}) := \boldsymbol{B}_{d,J_{j^\star(\boldsymbol{c},\boldsymbol{b})}} \boldsymbol{c}.$$

The sets

$$\{j^\star = j\} = \{v_j < \infty\} \cap \bigcap_{k<j} \{v_j < v_k\} \cap \bigcap_{k>j} \{v_j \leq v_k\}$$

are Borel, and so $\boldsymbol{\ell}_d^\star$ is jointly Borel. Its feasibility, optimality, and norm bound follow from the construction. ∎

Lemma A.8. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition C, and $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. Fix a nonempty $D_0 \subseteq \mathcal{D}$ and P-integrable functions $\boldsymbol{g}_d : \mathcal{Y} \to \mathcal{C}_d$, $d \in D_0$. Define*

$$K_d := \left\{ \left(\int_{\mathcal{Y}} \boldsymbol{x}_{(\boldsymbol{\mathcal{T}}_d)} \, d\mu, \int_{\mathcal{Y}} \boldsymbol{g}_d^{\mathsf{T}} \boldsymbol{x} \, d\mu \right) \middle| \begin{array}{l} \boldsymbol{x} \in L^1(\mu; \mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\geq 0}), \\ \boldsymbol{x}(y) \in \operatorname{Fibre}_d\left(\boldsymbol{p}_d^P(y)\right), \\ \boldsymbol{x}_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)}(y) = \mathbf{0} \end{array} \right\}.$$

Each K_d is nonempty, compact, and convex. Moreover,

$$\begin{aligned} &\left\{ (\boldsymbol{\pi}, (\boldsymbol{u}_d, t_d)_{d \in D_0}) \middle| \begin{array}{l} \boldsymbol{\pi} \in \boldsymbol{\Pi}(P), \\ (\boldsymbol{u}_d, t_d) \in K_d, \quad \boldsymbol{u}_d = \boldsymbol{\pi}_{(\boldsymbol{\mathcal{T}}_d)} \quad \forall d \in D_0 \end{array} \right\} \\ &= \left\{ \left(\boldsymbol{\pi}^Q, \left(\boldsymbol{\pi}^Q_{(\boldsymbol{\mathcal{T}}_d)}, \int_{\mathcal{Y}} \boldsymbol{g}_d^{\mathsf{T}} \boldsymbol{q}_d^Q \, d\mu \right)_{d \in D_0} \right) \middle| Q \in \mathcal{Q}_{\mathbf{S}}(P) \right\}, \end{aligned} \tag{A.7}$$

and this common set is nonempty, compact, and convex.

Proof. Let $r_d := \sum_{z \in \mathcal{Z}} p^P_{d|z}$ and $p_{\min} := \min_{z \in \mathcal{Z}} P_Z(z) > 0$. Every function in the definition of K_d satisfies

$$\|\boldsymbol{x}(y)\|_\infty \leq \|\boldsymbol{x}(y)\|_1 \leq \mathbf{1}^{\mathsf{T}} \mathbf{R}_d \boldsymbol{x}(y) \leq r_d(y),$$

and

$$\int_{\mathcal{Y}} (1 + \|\boldsymbol{g}_d(y)\|_\infty) \, r_d(y) \, d\mu(y) \leq |\mathcal{Z}| + \frac{1}{p_{\min}} \mathbb{E}_P \|\boldsymbol{g}_d(Y)\|_\infty < \infty.$$

Let $n_d := |\boldsymbol{\mathcal{T}}_d|$. Apply Corollary A.3 to the active coordinates, with

$$\boldsymbol{G} := \begin{bmatrix} -\boldsymbol{I}_{n_d} \\ \mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}_d)} \end{bmatrix}, \qquad \boldsymbol{b}(y) := \begin{bmatrix} \mathbf{0} \\ \boldsymbol{p}_d^P(y) \end{bmatrix},$$

$$\boldsymbol{W}(y) := \begin{bmatrix} \boldsymbol{I}_{n_d} \\ \boldsymbol{g}_{d;(\boldsymbol{\mathcal{T}}_d)}(y)^\intercal \end{bmatrix},$$

and envelope r_d. Its image is K_d, which contains $(\mathbf{0}, 0)$. Apply Corollary A.3 By Theorem 3.1, $\boldsymbol{\Pi}(P)$ is compact and convex, and so the set on the left of (A.7) is compact and convex.

For $Q \in \mathcal{Q}_{\mathbf{S}}(P)$, define

$$\boldsymbol{x}_{d;(\boldsymbol{\tau})}(y) := q^Q_{d|\boldsymbol{\tau}}(y)\mathbb{1}\{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d\} \qquad \forall d \in D_0.$$

By Lemma 3.1, these functions satisfy the defining constraints of K_d, and, since $\boldsymbol{g}_d(y) \in \mathcal{C}_d$,

$$\left(\int_{\mathcal{Y}} \boldsymbol{x}_{d;(\boldsymbol{\mathcal{T}}_d)}\, d\mu, \int_{\mathcal{Y}} \boldsymbol{g}_d^\intercal \boldsymbol{x}_d\, d\mu\right) = \left(\boldsymbol{\pi}^Q_{(\boldsymbol{\mathcal{T}}_d)}, \int_{\mathcal{Y}} \boldsymbol{g}_d^\intercal \boldsymbol{q}_d^Q\, d\mu\right).$$

This proves inclusion from right to left in (A.7) and nonemptiness.

Conversely, take a point on the left of (A.7) and choose its defining functions $\boldsymbol{x}_d$. Choose $Q^0 \in \mathcal{Q}_{\mathbf{S}}(P)$ with $\boldsymbol{\pi}^{Q^0} = \boldsymbol{\pi}$, and let ρ_0 be the probability density from Lemma A.1. For $d \in D_0$, define

$$q_{d|\boldsymbol{\tau}}(y) := \begin{cases} \boldsymbol{x}_{d;(\boldsymbol{\tau})}(y) & \text{if } \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d, \\ \boldsymbol{\pi}_{(\boldsymbol{\tau})}\rho_0(y) & \text{otherwise.} \end{cases}$$

For $d \notin D_0$, take $\boldsymbol{q}_d = \boldsymbol{q}_d^{Q^0}$. Then

$$\int_{\mathcal{Y}} \boldsymbol{q}_d\, d\mu = \boldsymbol{\pi}, \qquad \mathbf{R}_d \boldsymbol{q}_d \le \boldsymbol{p}_d^P \qquad \forall d \in \mathcal{D},$$
$$\int_{\mathcal{Y}} \boldsymbol{g}_d^\intercal \boldsymbol{q}_d\, d\mu = t_d \qquad \forall d \in D_0.$$

The construction in the proof of Theorem 3.1, applied to this collection, supplies $Q \in \mathcal{Q}_{\mathbf{S}}(P)$ with $\boldsymbol{\pi}^Q = \boldsymbol{\pi}$ and $\boldsymbol{q}_d^Q = \boldsymbol{q}_d$ for every $d \in \mathcal{D}$, proving the reverse inclusion. ■

Lemma A.9. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition C, and $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. Fix a nonempty $D_0 \subseteq \mathcal{D}$, P-integrable functions $\boldsymbol{g}_d : \mathcal{Y} \to \mathcal{C}_d$ for $d \in D_0$, $\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$, and a nonempty compact convex set $\Pi_0 \subseteq \boldsymbol{\Pi}(P)$. Then*

$$\max_{\substack{Q \in \mathcal{Q}_{\mathbf{S}}(P) \\ \boldsymbol{\pi}^Q \in \Pi_0}} \left\{ \sum_{d \in D_0} \int_{\mathcal{Y}} \boldsymbol{g}_d(y)^\intercal \boldsymbol{q}_d^Q(y)\, d\mu(y) + \boldsymbol{b}^\intercal \boldsymbol{\pi}^Q \right\}$$
$$= \inf_{(\boldsymbol{v}_d)_{d \in D_0} \in \prod_{d \in D_0} \mathcal{C}_d} \left\{ \sum_{d \in D_0} \int_{\mathcal{Y}} h_d\left(\boldsymbol{g}_d(y) - \boldsymbol{v}_d; \boldsymbol{p}_d^P(y)\right)\, d\mu(y) + h_{\Pi_0}\left(\boldsymbol{b} + \sum_{d \in D_0} \boldsymbol{v}_d\right) \right\}.$$

In particular, the maximum on the left is attained.

Proof. Let K_d be the compact convex set in Lemma A.8, for the function $\boldsymbol{g}_d$, and define

$$X := \Pi_0 \times \prod_{d \in D_0} K_d,$$
$$L := \left\{ (\boldsymbol{\pi}, (\boldsymbol{w}_d, t_d)_{d \in D_0}) \,\middle|\, \boldsymbol{w}_d = \boldsymbol{\pi}_{(\boldsymbol{\mathcal{T}}_d)} \; \forall d \in D_0 \right\}.$$

Here L is a linear subspace of $\mathbb{R}^{\boldsymbol{\mathcal{T}}} \times \prod_{d \in D_0} (\mathbb{R}^{\boldsymbol{\mathcal{T}}_d} \times \mathbb{R})$. By Lemma A.8, $X \cap L$ is nonempty and compact, and the maximum in the statement equals

$$\max_{(\boldsymbol{\pi}, (\boldsymbol{w}_d, t_d)_d) \in X \cap L} \left\{ \boldsymbol{b}^\intercal \boldsymbol{\pi} + \sum_{d \in D_0} t_d \right\}.$$

Moreover,

$$L^\perp = \left\{ \left(\sum_{d \in D_0} \boldsymbol{v}_d, \left(-\boldsymbol{v}_{d;(\boldsymbol{\mathcal{T}}_d)}, 0 \right)_{d \in D_0} \right) \,\middle|\, \boldsymbol{v}_d \in \mathcal{C}_d \; \forall d \in D_0 \right\}.$$

Since h_L is zero on $L^\perp$ and infinite elsewhere, Lemma 3.4, applied to X and L, gives

$$\max_{(\boldsymbol{\pi}, (\boldsymbol{w}_d, t_d)_d) \in X \cap L} \left\{ \boldsymbol{b}^\intercal \boldsymbol{\pi} + \sum_{d \in D_0} t_d \right\}$$
$$= \inf_{\boldsymbol{v}_d \in \mathcal{C}_d, \; d \in D_0} \left\{ h_{\Pi_0} \left(\boldsymbol{b} + \sum_{d \in D_0} \boldsymbol{v}_d \right) + \sum_{d \in D_0} \max_{(\boldsymbol{w}, t) \in K_d} \left(t - \boldsymbol{v}_{d;(\boldsymbol{\mathcal{T}}_d)}^\intercal \boldsymbol{w} \right) \right\}.$$

By the definition of K_d and Lemma A.2,

$$\max_{(\boldsymbol{w}, t) \in K_d} \left(t - \boldsymbol{v}_{d;(\boldsymbol{\mathcal{T}}_d)}^\intercal \boldsymbol{w} \right) = \int_{\mathcal{Y}} h_d \left(\boldsymbol{g}_d(y) - \boldsymbol{v}_d; \boldsymbol{p}_d^P(y) \right) \, d\mu(y) \,,$$

and the result follows immediately, with attainment following from $X \cap L$ being compact. ■

Lemma A.10. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition C, and $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. Fix $d^\star \in \mathcal{D}$, nonempty $A \subseteq \boldsymbol{\mathcal{T}}_{d^\star}$, and P-integrable $\kappa : \mathcal{Y} \to \mathbb{R}$. Define*

$$\mathsf{N}^Q := \int_{\mathcal{Y}} \kappa(y) \boldsymbol{c}_A^\intercal \boldsymbol{q}_{d^\star}^Q(y) \, d\mu(y), \qquad \mathsf{D}^Q := \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q,$$
$$K_P := \left\{ (\mathsf{N}^Q, \mathsf{D}^Q) \,\middle|\, Q \in \mathcal{Q}_{\mathbf{S}}(P) \right\},$$
$$\overline{\Psi}(\gamma; P) := \sup_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \left\{ \mathsf{N}^Q - \gamma \mathsf{D}^Q \right\} \qquad \forall \gamma \in \mathbb{R}.$$

Then K_P is nonempty, compact, and convex, and

$$\overline{\Psi}(\gamma; P) = \max_{(\mathsf{N}, a) \in K_P} \{ \mathsf{N} - \gamma a \} \qquad \forall \gamma \in \mathbb{R}.$$

For every $\gamma \in \mathbb{R}$, the set

$$\mathscr{I}_P(\gamma) := \left\{ a \in \mathbb{R} \;\middle|\; \begin{array}{l} \exists\, \mathsf{N} \in \mathbb{R} : (\mathsf{N}, a) \in K_P, \\ \mathsf{N} - \gamma a = \overline{\Psi}(\gamma; P) \end{array} \right\}$$

is nonempty and compact, and

$$\lim_{t \downarrow 0} \frac{\overline{\Psi}(\gamma + t; P) - \overline{\Psi}(\gamma; P)}{t} = -\min \mathscr{I}_P(\gamma),$$
$$\lim_{t \uparrow 0} \frac{\overline{\Psi}(\gamma + t; P) - \overline{\Psi}(\gamma; P)}{t} = -\max \mathscr{I}_P(\gamma).$$

If $a_*(P) := \min_{Q \in \mathcal{Q}_{\mathsf{S}}(P)} \mathsf{D}^Q > 0$, both bounds on $\theta^Q = \mathsf{N}^Q/\mathsf{D}^Q$ are attained, and

$$\overline{\Psi}(\bar{\theta}(P); P) = 0,$$
$$\mathscr{I}_P(\bar{\theta}(P)) = \left\{ \mathsf{D}^Q \,\middle|\, Q \in \mathcal{Q}_{\mathsf{S}}(P),\ \theta^Q = \bar{\theta}(P) \right\}.$$

If this set is $\{p_A^\dagger(P)\}$, then

$$\partial_\gamma \overline{\Psi}(\gamma; P)\big|_{\gamma = \bar{\theta}(P)} = -p_A^\dagger(P).$$

Proof. Apply Lemma A.8 with $D_0 = \{d^\star\}$ and $\boldsymbol{g}_{d^\star} = \kappa \boldsymbol{c}_A$. The continuous linear map

$$\left(\boldsymbol{\pi}, (\boldsymbol{\pi}_{(\boldsymbol{\tau}_{d^\star})}, t)\right) \mapsto \left(t, \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}\right)$$

maps the compact convex set in (A.7) onto K_P. For every γ, the set

$$K_P(\gamma) := \operatorname*{argmax}_{(\mathsf{N}, a) \in K_P} \{\mathsf{N} - \gamma a\}$$

is nonempty and compact, and so is its projection $\mathscr{I}_P(\gamma)$ onto the second coordinate.

The function $(\gamma, \mathsf{N}, a) \mapsto \mathsf{N} - \gamma a$ has continuous derivative $-a$ with respect to γ. Danskin's theorem therefore applies to its maximum over the nonempty compact set K_P and gives

$$\lim_{t \downarrow 0} \frac{\overline{\Psi}(\gamma + t; P) - \overline{\Psi}(\gamma; P)}{t} = \max_{(\mathsf{N}, a) \in K_P(\gamma)} (-a) = -\min \mathscr{I}_P(\gamma),$$
$$\lim_{t \uparrow 0} \frac{\overline{\Psi}(\gamma + t; P) - \overline{\Psi}(\gamma; P)}{t} = -\lim_{u \downarrow 0} \frac{\overline{\Psi}(\gamma - u; P) - \overline{\Psi}(\gamma; P)}{u}$$
$$= -\max_{(\mathsf{N}, a) \in K_P(\gamma)} a = -\max \mathscr{I}_P(\gamma).$$

If $a_*(P) > 0$, the function $(\mathsf{N}, a) \mapsto \mathsf{N}/a$ is continuous on K_P, and so attains both extrema. Moreover, for every $(\mathsf{N}, a) \in K_P$,

$$\mathsf{N} - \bar{\theta}(P) a = a \left(\frac{\mathsf{N}}{a} - \bar{\theta}(P) \right) \leq 0,$$

with equality exactly at the upper-endpoint maximizers. This proves the assertions at $\bar{\theta}(P)$. When their masses coincide, the two one-sided derivatives coincide as well. ∎

Lemma A.11. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition C, and $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. Fix $d^\star \in \mathcal{D}$, nonempty $A \subseteq \mathcal{T}_{d^\star}$, and a P-integrable $\kappa : \mathcal{Y} \to \mathbb{R}$. For $\gamma \in \mathbb{R}$, $\boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}$, and $\boldsymbol{u} \in \mathcal{C}$, define*

$$G_P(\gamma, \boldsymbol{v}) := \int_{\mathcal{Y}} h_{d^\star}\left((\kappa(y) - \gamma)\boldsymbol{c}_A - \boldsymbol{v}; \boldsymbol{p}^P_{d^\star}(y)\right) d\mu(y) + h_{\boldsymbol{\Pi}(P)}(\boldsymbol{v}),$$

$$H_P(\gamma, \boldsymbol{v}, \boldsymbol{u}) := \int_{\mathcal{Y}} h_{d^\star}\left((\kappa(y) - \gamma)\boldsymbol{c}_A - \boldsymbol{v}; \boldsymbol{p}^P_{d^\star}(y)\right) d\mu(y) + \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} h_d\left(\boldsymbol{u}_d; \boldsymbol{p}^P_d(y)\right) d\mu(y) + h_{\mathcal{R}_{\boldsymbol{\pi}}}\left(\boldsymbol{v} - \sum_{d \in \mathcal{D}} \boldsymbol{u}_d\right).$$

These functions are jointly convex, and

$$\overline{\Psi}(\gamma; P) = \inf_{\boldsymbol{v} \in \mathbb{R}^{\mathcal{T}}} G_P(\gamma, \boldsymbol{v}) = \inf_{\substack{\boldsymbol{v} \in \mathbb{R}^{\mathcal{T}} \\ \boldsymbol{u} \in \mathcal{C}}} H_P(\gamma, \boldsymbol{v}, \boldsymbol{u}) \qquad \forall \gamma \in \mathbb{R}. \tag{A.8}$$

Under Condition CP, let $m_{\mathbf{S}}$ be the number of rows of $\mathbf{A}_{\mathbf{S}}$ and define

$$\mathcal{D}_\theta := \left\{ (\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{C} \times \mathbb{R}^{m_{\mathbf{S}}}_{\geq 0} \,\middle|\, \left(\mathbf{A}_{\mathbf{S}}^\intercal \boldsymbol{\lambda} + \sum_{d \in \mathcal{D}} \boldsymbol{u}_d\right)_{(\mathcal{T} \setminus \mathcal{T}_{d^\star})} \geq \mathbf{0} \right\},$$

$$\boldsymbol{c}_0(y; \gamma, \boldsymbol{u}, \boldsymbol{\lambda}) := (\kappa(y) - \gamma)\boldsymbol{c}_A - \mathbf{A}_{\mathbf{S}}^\intercal \boldsymbol{\lambda} - \sum_{d \in \mathcal{D}} \boldsymbol{u}_d,$$

$$\mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}) := \int_{\mathcal{Y}} h_{d^\star}\left(\boldsymbol{c}_0(y; \gamma, \boldsymbol{u}, \boldsymbol{\lambda}); \boldsymbol{p}^P_{d^\star}(y)\right) d\mu(y) + \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} h_d\left(\boldsymbol{u}_d; \boldsymbol{p}^P_d(y)\right) d\mu(y) + \boldsymbol{\lambda}^\intercal \mathbf{a}_{\mathbf{S}}.$$

The function $\mathscr{J}_P$ is finite and jointly convex on $\mathbb{R} \times \mathcal{D}_\theta$, and

$$\overline{\Psi}(\gamma; P) = \inf_{(\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{D}_\theta} \mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}) \qquad \forall \gamma \in \mathbb{R}. \tag{A.9}$$

Suppose additionally that

$$a_*(P) := \min_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q > 0.$$

Then

$$\begin{aligned} \bar{\theta}(P) &= \inf\left\{\gamma \in \mathbb{R} \,\middle|\, \exists\, \boldsymbol{v} \in \mathbb{R}^{\mathcal{T}} : G_P(\gamma, \boldsymbol{v}) \leq 0\right\} \\ &= \inf\left\{\gamma \in \mathbb{R} \,\middle|\, \exists\, (\boldsymbol{v}, \boldsymbol{u}) \in \mathbb{R}^{\mathcal{T}} \times \mathcal{C} : H_P(\gamma, \boldsymbol{v}, \boldsymbol{u}) \leq 0\right\}, \end{aligned}$$

and, under Condition CP,

$$\bar{\theta}(P) = \inf\left\{\gamma \in \mathbb{R} \,\middle|\, \exists\, (\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{D}_\theta : \mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}) \leq 0\right\}.$$

In each program, $\gamma = \bar{\theta}(P)$ is feasible if and only if the respective infimum in (A.8) or (A.9) is attained at zero.

If $\underline{\kappa} \le \kappa \le \overline{\kappa}$ for finite $\underline{\kappa}, \overline{\kappa}$, each program may instead restrict γ to $\Theta_0 = [\underline{\kappa}, \overline{\kappa}]$, using the convention that its value is $\overline{\kappa}$ when its feasible set is empty.

Proof. Apply Lemma A.9 with $D_0 = \{d^\star\}$, $\boldsymbol{g}_{d^\star} = (\kappa - \gamma)\boldsymbol{c}_A$, $\boldsymbol{b} = \mathbf{0}$, and $\Pi_0 = \mathbf{\Pi}(P)$. This gives

$$\overline{\Psi}(\gamma; P) = \inf_{\boldsymbol{v} \in \mathcal{C}_{d^\star}} G_P(\gamma, \boldsymbol{v}).$$

If $G_P(\gamma, \boldsymbol{v}) < \infty$, then $\boldsymbol{v}_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_{d^\star})} \ge \mathbf{0}$. Replacing these coordinates by zero leaves the fibre support unchanged and cannot increase $h_{\mathbf{\Pi}(P)}(\boldsymbol{v})$, since $\mathbf{\Pi}(P) \subseteq \mathbb{R}^{\boldsymbol{\mathcal{T}}}_{\ge 0}$. Thus the infimum is unchanged when $\boldsymbol{v}$ ranges over $\mathbb{R}^{\boldsymbol{\mathcal{T}}}$.

By Theorem 3.3,

$$G_P(\gamma, \boldsymbol{v}) = \inf_{\boldsymbol{u} \in \mathcal{C}} H_P(\gamma, \boldsymbol{v}, \boldsymbol{u}),$$

which proves (A.8). Under Condition CP, strong duality gives

$$h_{\mathcal{R}_{\boldsymbol{\pi}}}(\boldsymbol{w}) = \min_{\substack{\boldsymbol{\lambda} \ge \mathbf{0} \\ \mathbf{A}_{\mathbf{S}}^{\mathsf{T}} \boldsymbol{\lambda} = \boldsymbol{w}}} \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{a}_{\mathbf{S}} \qquad \forall \boldsymbol{w} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}.$$

Substitute this representation into H_P and eliminate $\boldsymbol{v}$ using

$$\boldsymbol{v} = \mathbf{A}_{\mathbf{S}}^{\mathsf{T}} \boldsymbol{\lambda} + \sum_{d \in \mathcal{D}} \boldsymbol{u}_d.$$

The target fibre support is finite exactly when $(\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{D}_\theta$. This proves (A.9). Joint convexity follows from convexity of support functions and the affine dependence of their arguments on the optimization variables. Finiteness of $\mathscr{J}_P$ on its stated domain follows from Lemma A.2.

Now suppose $a_*(P) > 0$. By Lemma A.10, there is $Q^\star \in \mathcal{Q}_{\mathbf{S}}(P)$ with $\theta^{Q^\star} = \bar{\theta}(P)$, and $\overline{\Psi}(\bar{\theta}(P); P) = 0$. For $\gamma < \bar{\theta}(P)$,

$$\begin{aligned} \overline{\Psi}(\gamma; P) &\ge \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}^{Q^\star} \left(\bar{\theta}(P) - \gamma \right) \\ &\ge a_*(P) \left(\bar{\theta}(P) - \gamma \right) > 0. \end{aligned}$$

For $\gamma > \bar{\theta}(P)$,

$$\begin{aligned} \overline{\Psi}(\gamma; P) &= \max_{Q \in \mathcal{Q}_{\mathbf{S}}(P)} \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}^{Q} \left(\theta^{Q} - \gamma \right) \\ &\le a_*(P) \left(\bar{\theta}(P) - \gamma \right) < 0. \end{aligned}$$

Hence none of the three programs is feasible below $\bar{\theta}(P)$, and each is feasible at every $\gamma > \bar{\theta}(P)$, by (A.8) and (A.9). At $\gamma = \bar{\theta}(P)$, feasibility is exactly attainment of the respective infimum at zero.

Finally, when κ takes values in Θ_0, $\bar{\theta}(P) \in \Theta_0$. If $\bar{\theta}(P) < \overline{\kappa}$, the preceding argument supplies feasible values arbitrarily close to $\bar{\theta}(P)$ within Θ_0. If $\bar{\theta}(P) = \overline{\kappa}$, the restricted feasible set either contains this endpoint or is empty, and our convention gives the same value. ■

Lemma A.12. *Maintain the assumptions and notation of Lemma A.11, with Condition CP, and suppose*

$$\underline{\kappa} \le \kappa(y) \le \overline{\kappa} \qquad \forall y \in \mathcal{Y},$$

for finite $\underline{\kappa}, \overline{\kappa}$. There exists $B_\theta < \infty$, depending only on the fixed model matrices, $\overline{\kappa} - \underline{\kappa}$, and $\boldsymbol{c}_A$, such that, for every such P and every $\gamma \in [\underline{\kappa} - 1, \overline{\kappa} + 1]$, the set

$$\mathcal{O}_\theta(P, \gamma) := \left\{ (\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{D}_\theta \;\middle|\; \begin{array}{l} \|(\boldsymbol{u}, \boldsymbol{\lambda})\|_\infty \le B_\theta, \\ \mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}) = \overline{\Psi}(\gamma; P) \end{array} \right\}$$

is nonempty. One member of $\mathcal{O}_\theta(P, \gamma)$ can be chosen with $\boldsymbol{u}_{d^\star} = \boldsymbol{0}$.

If $a_*(P) > 0$, the program in Lemma A.11 involving $\mathscr{J}_P$ attains $\bar{\theta}(P)$ with outer variables satisfying this bound.

Proof. Fix P and $\gamma \in [\underline{\kappa} - 1, \overline{\kappa} + 1]$. Let $\boldsymbol{x} := (\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathbb{R}^{p_\theta}$, where $p_\theta := |\mathcal{D}||\boldsymbol{\mathcal{T}}| + m_{\mathbf{S}}$. Enumerate $\mathcal{D} = \{d_1, \dots, d_J\}$ and set $d_0 := d^\star$.

We first identify finitely many linear inequalities that determine a cone of the fan in Lemma A.4 for every coefficient appearing in $\mathscr{J}_P$. Define the fixed vectors $\boldsymbol{a}_i$ and linear maps $\boldsymbol{B}_i$ by

$$\boldsymbol{a}_0 := \boldsymbol{c}_A, \qquad \boldsymbol{B}_0 \boldsymbol{x} := -\mathbf{A}_{\mathbf{S}}^{\intercal} \boldsymbol{\lambda} - \sum_{d \in \mathcal{D}} \boldsymbol{u}_d,$$

$$\boldsymbol{a}_i := \boldsymbol{0}, \qquad \boldsymbol{B}_i \boldsymbol{x} := \boldsymbol{u}_{d_i} \qquad \forall i \in [J],$$

and let

$$\boldsymbol{c}_i(y; \boldsymbol{x}) := \boldsymbol{a}_i(\kappa(y) - \gamma) + \boldsymbol{B}_i \boldsymbol{x}.$$

Then

$$\mathscr{J}_P(\gamma, \boldsymbol{x}) = \sum_{i=0}^{J} \int_{\mathcal{Y}} h_{d_i}\left(\boldsymbol{c}_i(y; \boldsymbol{x}); \boldsymbol{p}_{d_i}^P(y)\right) d\mu(y) + \boldsymbol{\lambda}^{\intercal} \mathbf{a}_{\mathbf{S}}.$$

For $\boldsymbol{x} \in \mathcal{D}_\theta$, every $\boldsymbol{c}_i(y; \boldsymbol{x})$ has nonpositive coordinates outside $\boldsymbol{\mathcal{T}}_{d_i}$, and so its fibre support depends only on its $\boldsymbol{\mathcal{T}}_{d_i}$ coordinates.

For each $i \in \{0, \dots, J\}$ and $\Gamma \in \mathscr{N}_{d_i}$, choose a finite representation

$$\Gamma = \left\{ \boldsymbol{a} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}_{d_i}} \;\middle|\; \boldsymbol{F}_{i,\Gamma} \boldsymbol{a} \ge \boldsymbol{0} \right\},$$

where $\boldsymbol{F}_{i,\Gamma} \in \mathbb{R}^{m_{i,\Gamma} \times \boldsymbol{\mathcal{T}}_{d_i}}$. Let

$$\mathcal{J} := \left\{ (i, \Gamma, r) \,|\, i \in \{0, \dots, J\},\ \Gamma \in \mathscr{N}_{d_i},\ r \in [m_{i,\Gamma}] \right\}.$$

For $j = (i, \Gamma, r) \in \mathcal{J}$, define

$$\boldsymbol{v}_j^{\intercal} := \boldsymbol{F}_{i,\Gamma;(r,\cdot)} \boldsymbol{B}_{i;(\boldsymbol{\mathcal{T}}_{d_i}, \cdot)},$$

$$b_j := \boldsymbol{F}_{i,\Gamma;(r,\cdot)} \boldsymbol{a}_{i;(\boldsymbol{\mathcal{T}}_{d_i})}.$$

The corresponding inequality for $\boldsymbol{c}_{i;(\boldsymbol{\tau}_{d_i})}(y;\boldsymbol{x}) \in \Gamma$ is

$$\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x} + b_j(\kappa(y) - \gamma) \geq 0.$$

Set

$$M := \max\left\{1,\ (\overline{\kappa} - \underline{\kappa} + 1)\max_{j\in\mathcal{J}} |b_j|\right\}.$$

Then $|b_j(\kappa(y) - \gamma)| \leq M$ for every $j \in \mathcal{J}$ and $y \in \mathcal{Y}$.

Fix any $\boldsymbol{x}_0 \in \mathcal{D}_\theta$, and let $K(\boldsymbol{x}_0)$ be the intersection of $\mathcal{D}_\theta$ with the following constraint for each $j \in \mathcal{J}$:

$$\begin{cases} \boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x} = \boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x}_0 & \text{if } |\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x}_0| \leq M, \\ \boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x} \geq M & \text{if } \boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x}_0 > M, \\ \boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x} \leq -M & \text{if } \boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x}_0 < -M. \end{cases}$$

This polyhedron contains $\boldsymbol{x}_0$. Its constraint matrices range over a fixed finite collection, and every right-hand side has absolute value at most M.

If $\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x}_0 + b_j(\kappa(y) - \gamma) \geq 0$, then, for every $\boldsymbol{x} \in K(\boldsymbol{x}_0)$,

$$\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{x} + b_j(\kappa(y) - \gamma) \geq 0.$$

Indeed, the expression is unchanged in the first case, is at least $M - |b_j(\kappa(y) - \gamma)| \geq 0$ in the second, and cannot be nonnegative at $\boldsymbol{x}_0$ in the third.

For each i and y, choose $\Gamma \in \mathscr{N}_{d_i}$ containing $\boldsymbol{c}_{i;(\boldsymbol{\tau}_{d_i})}(y;\boldsymbol{x}_0)$. The preceding inequalities imply

$$\boldsymbol{c}_{i;(\boldsymbol{\tau}_{d_i})}(y;\boldsymbol{x}) \in \Gamma \qquad \forall \boldsymbol{x} \in K(\boldsymbol{x}_0).$$

By Lemma A.4, the fibre support is linear on Γ. Hence, for $\boldsymbol{x}, \boldsymbol{x}' \in K(\boldsymbol{x}_0)$ and $t \in [0,1]$,

$$\begin{aligned} &h_{d_i}\left(\boldsymbol{c}_i(y; t\boldsymbol{x} + (1-t)\boldsymbol{x}'); \boldsymbol{p}_{d_i}^P(y)\right) \\ &= t\, h_{d_i}\left(\boldsymbol{c}_i(y;\boldsymbol{x}); \boldsymbol{p}_{d_i}^P(y)\right) + (1-t)\, h_{d_i}\left(\boldsymbol{c}_i(y;\boldsymbol{x}'); \boldsymbol{p}_{d_i}^P(y)\right). \end{aligned}$$

Integrating and adding $\boldsymbol{\lambda}^{\mathsf{T}}\mathbf{a}_{\mathbf{S}}$ shows that $\mathscr{J}_P(\gamma,\cdot)$ is affine on $K(\boldsymbol{x}_0)$. It is finite there by Lemma A.2, and bounded below by $\overline{\Psi}(\gamma;P) \in \mathbb{R}$, by Lemma A.11.

We can therefore minimize an affine function over $K(\boldsymbol{x}_0)$. To bound an optimizer independently of $\boldsymbol{x}_0$, convert each possible constraint system for $K(\boldsymbol{x}_0)$ into standard form by writing $\boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^-$, adding nonnegative slack variables, and deleting redundant equations. Let $\mathscr{A}$ be the resulting finite collection of full-row-rank constraint matrices. Every resulting right-hand side $\boldsymbol{b}$ satisfies $\|\boldsymbol{b}\|_\infty \leq M$. Define

$$C_{\mathrm{bas}} := \max\left(\{1\} \cup \left\{ \|\boldsymbol{B}^{-1}\|_{\infty\to\infty} \;\middle|\; \begin{array}{l}\boldsymbol{B} \text{ is an invertible square column} \\ \text{submatrix of some } \boldsymbol{A} \in \mathscr{A}\end{array} \right\}\right),$$

$$B_\theta := 2MC_{\mathrm{bas}}.$$

An optimal basic feasible solution $\boldsymbol{w}^\star$ of the standard-form program satisfies

$$\boldsymbol{w}^\star_{(B)} = \boldsymbol{B}^{-1}\boldsymbol{b}, \qquad \boldsymbol{w}^\star_{(B^c)} = \mathbf{0},$$
$$\|\boldsymbol{x}^\star\|_\infty \le 2\|\boldsymbol{w}^\star\|_\infty \le 2MC_{\mathrm{bas}} = B_\theta,$$

where B indexes its basic columns and $\boldsymbol{x}^\star = (\boldsymbol{x}^+)^\star - (\boldsymbol{x}^-)^\star$. In particular,

$$\mathscr{J}_P(\gamma, \boldsymbol{x}^\star) \le \mathscr{J}_P(\gamma, \boldsymbol{x}_0).$$

The matrices defining $\mathscr{A}$ depend only on the fixed model matrices and the fans in Lemma A.4; the right-hand-side bound M additionally depends only on $\overline{\kappa} - \underline{\kappa}$ and $\boldsymbol{c}_A$. Thus B_θ has the asserted dependence.

Since $\boldsymbol{x}_0$ was arbitrary, the infimum over $\mathcal{D}_\theta$ equals the infimum over

$$\{\boldsymbol{x} \in \mathcal{D}_\theta \,|\, \|\boldsymbol{x}\|_\infty \le B_\theta\},$$

which is nonempty and compact. For $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{D}_\theta$, Lemma A.2 gives

$$\begin{aligned}&\left|\mathscr{J}_P(\gamma, \boldsymbol{x}) - \mathscr{J}_P(\gamma, \boldsymbol{x}')\right| \\ &\le \left[\sum_{i=0}^{J} \|\boldsymbol{B}_i\|_{\infty\to\infty} \int_{\mathcal{Y}} \|\boldsymbol{p}^P_{d_i}(y)\|_1 \, d\mu(y) + \|\mathbf{a}_{\mathbf{S}}\|_1\right] \|\boldsymbol{x} - \boldsymbol{x}'\|_\infty.\end{aligned}$$

The coefficient is finite, and so $\mathscr{J}_P(\gamma, \cdot)$ is continuous. Its minimum on the compact set is attained, proving $\mathcal{O}_\theta(P, \gamma) \neq \emptyset$.

Finally, take a bounded optimizer and replace $\boldsymbol{u}_{d^\star}$ by $\mathbf{0}$. Its target coefficient changes from $\boldsymbol{c}_0(y)$ to $\boldsymbol{c}_0(y) + \boldsymbol{u}_{d^\star}$, and

$$h_{d^\star}\left(\boldsymbol{c}_0(y) + \boldsymbol{u}_{d^\star}; \boldsymbol{p}^P_{d^\star}(y)\right) \le h_{d^\star}\left(\boldsymbol{c}_0(y); \boldsymbol{p}^P_{d^\star}(y)\right) + h_{d^\star}\left(\boldsymbol{u}_{d^\star}; \boldsymbol{p}^P_{d^\star}(y)\right).$$

Thus the replacement cannot increase $\mathscr{J}_P$. It preserves $\mathcal{D}_\theta$, since $\boldsymbol{u}_{d^\star}$ is zero outside $\mathcal{T}_{d^\star}$, and cannot increase the outer norm. The modified tuple is therefore also in $\mathcal{O}_\theta(P, \gamma)$.

When $a_*(P) > 0$, Lemma A.10 gives $\overline{\Psi}(\bar\theta(P); P) = 0$, and the result at $\gamma = \bar\theta(P)$ proves endpoint attainment. ∎

Corollary A.4. *Maintain the assumptions and notation of Lemma A.12. Fix $\gamma \in [\underline{\kappa} - 1, \overline{\kappa} + 1]$ and $(\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{O}_\theta(P, \gamma)$, and let $\boldsymbol{c}_0(y) := \boldsymbol{c}_0(y; \gamma, \boldsymbol{u}, \boldsymbol{\lambda})$. There exist measurable $\boldsymbol{\ell}_0, \boldsymbol{\ell}_d : \mathcal{Y} \to \mathbb{R}^{\mathcal{Z}}_{\ge 0}$, $d \in \mathcal{D}$, satisfying*

$$\begin{aligned}\mathbf{R}^\intercal_{d^\star}\boldsymbol{\ell}_0(y) &\ge \boldsymbol{c}_0(y), & \|\boldsymbol{\ell}_0(y)\|_\infty &\le \|\boldsymbol{c}_0(y)\|_\infty, \\ \boldsymbol{p}^P_{d^\star}(y)^\intercal\boldsymbol{\ell}_0(y) &= h_{d^\star}\left(\boldsymbol{c}_0(y); \boldsymbol{p}^P_{d^\star}(y)\right), & & \\ \mathbf{R}^\intercal_d\boldsymbol{\ell}_d(y) &\ge \boldsymbol{u}_d, & \|\boldsymbol{\ell}_d(y)\|_\infty &\le \|\boldsymbol{u}_d\|_\infty, \\ \boldsymbol{p}^P_d(y)^\intercal\boldsymbol{\ell}_d(y) &= h_d\left(\boldsymbol{u}_d; \boldsymbol{p}^P_d(y)\right) \qquad \forall d \in \mathcal{D}. & &\end{aligned}$$

For every such choice, define

$$m(y,d,z) := \mathbb{1}\{d = d^\star\}\boldsymbol{\ell}_{0;(z)}(y) + \boldsymbol{\ell}_{d;(z)}(y) + \frac{\boldsymbol{\lambda}^\intercal \mathbf{a}_\mathbf{S}}{|\mathcal{Z}|}.$$

There exists $C_m < \infty$, independent of P, γ, and these choices, such that

$$\|m\|_\infty \le C_m,$$
$$\sum_{z\in\mathcal{Z}} \mathbb{E}_P[m(W) \mid Z = z] = \overline{\Psi}(\gamma; P).$$

The same bound holds for every $m \in \mathfrak{M}^\star_P$. If $a_*(P) > 0$, then $\mathfrak{M}^\star_P \neq \emptyset$.

Proof. The measurable multipliers and their pointwise properties follow from Lemma 3.2. Set

$$C_0 := (\overline{\kappa} - \underline{\kappa} + 1)\|\boldsymbol{c}_A\|_\infty + B_\theta \|\mathbf{A}_\mathbf{S}^\intercal\|_{\infty\to\infty} + |\mathcal{D}|B_\theta,$$
$$C_m := C_0 + B_\theta + \frac{B_\theta \|\mathbf{a}_\mathbf{S}\|_1}{|\mathcal{Z}|}.$$

Since $\|(\boldsymbol{u},\boldsymbol{\lambda})\|_\infty \le B_\theta$,

$$\|\boldsymbol{c}_0(y)\|_\infty \le C_0,$$
$$|m(y,d,z)| \le \|\boldsymbol{\ell}_0(y)\|_\infty + \|\boldsymbol{\ell}_d(y)\|_\infty + \frac{\|\boldsymbol{\lambda}\|_\infty \|\mathbf{a}_\mathbf{S}\|_1}{|\mathcal{Z}|} \le C_m.$$

Integrating the pointwise optimality equalities gives

$$\begin{aligned}\sum_{z\in\mathcal{Z}} \mathbb{E}_P[m(W) \mid Z = z] &= \int_\mathcal{Y} \boldsymbol{p}^P_{d^\star}(y)^\intercal \boldsymbol{\ell}_0(y)\, d\mu(y) \\ &\quad + \sum_{d\in\mathcal{D}} \int_\mathcal{Y} \boldsymbol{p}^P_d(y)^\intercal \boldsymbol{\ell}_d(y)\, d\mu(y) + \boldsymbol{\lambda}^\intercal \mathbf{a}_\mathbf{S} \\ &= \mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}) = \overline{\Psi}(\gamma; P).\end{aligned}$$

If $a_*(P) > 0$, take $\gamma = \bar{\theta}(P)$. By Lemmas A.10 and A.12, the required outer tuple exists and the last display equals zero. The resulting m satisfies every condition in the definition of $\mathfrak{M}^\star_P$. ∎

Appendix B: Estimation Results and Proofs

Throughout, we will use shorthands of the form $\boldsymbol{q} := (\boldsymbol{q}_d)_{d\in\mathcal{D}}$. For any P satisfying Condition D, define

$$\mathcal{F}^P_\mathbf{S} := \left\{ (\boldsymbol{\pi}, \boldsymbol{q}) \;\middle|\; \begin{array}{l} \int \boldsymbol{q}_d\, d\mu = \boldsymbol{\pi} \in \mathcal{R}_{\boldsymbol{\pi}}\ ;\ \boldsymbol{q}_d \in L^1\left(\mu; \mathbb{R}^{\mathcal{T}}_{\ge 0}\right)\ \forall d\ ; \\ \mathbf{R}_d \boldsymbol{q}_d(y) = \boldsymbol{p}_d(y)\ \ \forall d, y\ . \end{array} \right\},$$

and for any P also satisfying Assumption 5.1, define

$$\mathcal{A}_\theta(P) := \left\{ (\boldsymbol{\pi}, \boldsymbol{q}) \in \mathcal{F}_{\mathbf{S}}^P \,\middle|\, \int_{\mathcal{Y}} (\kappa - \bar{\theta}(P)) \boldsymbol{c}_A^\intercal \boldsymbol{q}_{d^\star} \, d\mu = 0 \right\} .$$

For $(\boldsymbol{\pi}, \boldsymbol{q}) \in \mathcal{F}_{\mathbf{S}}^P, (\boldsymbol{\pi}', \boldsymbol{q}') \in \mathcal{F}_{\mathbf{S}}^{P'}$, define $\|(\boldsymbol{\pi}, \boldsymbol{q}) - (\boldsymbol{\pi}', \boldsymbol{q}')\|_{\mathcal{F},1} := \|\boldsymbol{\pi} - \boldsymbol{\pi}'\|_1 + \sum_d \int \|\boldsymbol{q}_d - \boldsymbol{q}_d'\|_1 \, d\mu$. Define the vector $\boldsymbol{q}_d^a$ and the matrices E_d^a and E_d^0 by

$$\begin{aligned}
(\boldsymbol{q}_d^a)_{(\boldsymbol{\tau})} &:= q_{d|\boldsymbol{\tau}} \mathbb{1}\{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d\} \ \forall \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} , \\
(E_d^a)_{(\boldsymbol{\tau}, \boldsymbol{\tau}')} &:= \mathbb{1}\{\boldsymbol{\tau} = \boldsymbol{\tau}'\} \ \forall (\boldsymbol{\tau}, \boldsymbol{\tau}') \in \boldsymbol{\mathcal{T}}_d \times \boldsymbol{\mathcal{T}} , \\
(E_d^0)_{(\boldsymbol{\tau}, \boldsymbol{\tau}')} &:= \mathbb{1}\{\boldsymbol{\tau} = \boldsymbol{\tau}'\} \ \forall (\boldsymbol{\tau}, \boldsymbol{\tau}') \in (\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d) \times \boldsymbol{\mathcal{T}} ,
\end{aligned}$$

so that $\boldsymbol{q}_d = (E_d^a)^\intercal E_d^a \boldsymbol{q}_d + (E_d^0)^\intercal E_d^0 \boldsymbol{q}_d = \boldsymbol{q}_d^a + (E_d^0)^\intercal E_d^0 \boldsymbol{q}_d$.

B.1. **Proofs of Estimation and Efficiency Results in Section 5**

B.1.1. *Proof of Lemma 5.1.*

Proof. Apply Lemma A.12 with $\gamma \in \Theta_0$. For every $(\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{D}_\theta$, the pointwise optimality equalities in Lemma 3.2 give

$$\sum_{z \in \mathcal{Z}} \mathbb{E}_P \left[\boldsymbol{m}_\theta^P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}; W) \,\middle|\, Z = z \right] = \mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}).$$

Substituting this equality into the bounded minimum supplied by Lemma A.12 gives (5.1). ■

B.1.2. *Proof of Theorem 5.1.*

Proof. Since $\mathcal{Z}$ is finite and $P_Z(z) > 0$ for every $z \in \mathcal{Z}$, $\min_{z \in \mathcal{Z}} \hat{P}_{Z;n}(z) > 0$ w.p.a.1 and so, we condition on this event in the remainder of the proof.

Apply Corollary B.1 to the conditional-mean parameter in (4.1), noting that its hypotheses hold since κ is bounded, and by Assumption 5.1, and so there exist constants $\delta, a, C > 0$ such that, for every $P' \in \mathcal{U}_\delta(P)$,

$$\begin{aligned}
\min_{Q \in \mathcal{Q}_{\mathbf{S}}(P')} \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q &\geq a, \\
\left| \bar{\theta}(P') - \bar{\theta}(P) \right| &\leq C \Delta(P', P).
\end{aligned}$$

In particular, the conditioning probability remains bounded away from zero throughout this neighbourhood, and the error in the upper bound is controlled by the error in the conditional sub-densities. By Assumption 5.2, $\tilde{P}_n$ satisfies Condition D, $\mathcal{Q}_{\mathbf{S}}(\tilde{P}_n) \neq \emptyset$, and $\Delta(\tilde{P}_n, P) \xrightarrow{P} 0$. Therefore $\tilde{P}_n \in \mathcal{U}_\delta(P)$ with probability tending to one, and so the displayed inequality and (5.2) give

$$\tilde{\bar{\theta}}_n - \bar{\theta}(P) = \bar{\theta}(\tilde{P}_n) - \bar{\theta}(P) = o_P(1) . \qquad ■$$

B.1.3. *Proof of Theorem 5.2.*

Proof. First, by Assumption 5.1 and Corollary B.1, there exist $\delta, a > 0$ and $L < \infty$ such that, for every law R satisfying Condition D, $\mathcal{Q}_{\mathbf{S}}(R) \neq \emptyset$, and $\Delta(R, P) < \delta$,

$$\begin{aligned}\min_{Q \in \mathcal{Q}_{\mathbf{S}}(R)} \boldsymbol{c}_A^{\intercal} \boldsymbol{\pi}^Q \geq & a, \\ |\bar{\theta}(R) - \bar{\theta}(P)| \leq & L\Delta(R, P) \ ,\end{aligned}$$

and note that Corollary A.4 supplies $M < \infty$ such that $\|m\|_\infty \leq M$ for every $m \in \mathfrak{M}_R^\star$ and every such R.

Next, for each fold, take

$$\begin{aligned}a_{n,k} &:= \boldsymbol{c}_A^{\intercal} \tilde{\boldsymbol{\pi}}^{(-k)}, \\ m_{n,k}(W) &:= \boldsymbol{m}_\theta^{\tilde{P}^{(-k)}} \left(\tilde{\bar{\theta}}_n^{(-k)}, \tilde{\boldsymbol{u}}^{(-k)}, \tilde{\boldsymbol{\lambda}}^{(-k)}; W \right). \end{aligned} \tag{B.1}$$

Now, $m_{n,k} \in \mathfrak{M}^\star_{\tilde{P}_n^{(-k)}}$ and the primal solution defining $\tilde{\boldsymbol{\pi}}^{(-k)}$ attains $\bar{\theta}(\tilde{P}_n^{(-k)})$. Define

$$\begin{aligned}\Delta_{n,k} :=& \Delta(\tilde{P}_n^{(-k)}, P), \\ b_{n,k} :=& \bar{\theta}(\tilde{P}_n^{(-k)}) - \bar{\theta}(P) + \frac{1}{a_{n,k}} \sum_z \mathbb{E}_P[m_{n,k}(W) \mid Z = z], \\ \phi_{n,k}(W) :=& \frac{m_{n,k}(W) - \mathbb{E}_P[m_{n,k}(W) \mid Z]}{P_Z(Z) a_{n,k}},\end{aligned}$$

where $P_Z(z) := P(Z = z)$ and the expectations involving functions estimated on the training folds are understood as taken conditional on the training folds, i.e. holding training-estimated quantities fixed.

By Assumption 5.2 and positivity of all $P_Z(z)$, $\max_k \Delta_{n,k} = o_P(1)$, and every $\tilde{P}_n^{(-k)}$ satisfies the conditions imposed on R above w.p.a.1. Since $\sum_z \mathbb{E}_{\tilde{P}_n^{(-k)}}[m_{n,k}(W) \mid Z = z] = 0$, w.p.a.1,

$$\begin{aligned}\left| \sum_z \mathbb{E}_P[m_{n,k}(W) \mid Z = z] \right| \leq & M \Delta_{n,k}, \\ \max_k |b_{n,k}| \leq & \left(L + \frac{M}{a} \right) \max_k \Delta_{n,k} = o_P(1) \ ,\end{aligned}$$

and note that we now have $a_{n,k} \geq a$ and $\|m_{n,k}\|_\infty \leq M$ for every k, w.p.a.1. Therefore, we can apply Lemma B.8 with $\eta_{n,k} = \bar{\theta}(\tilde{P}_n^{(-k)})$ and $\vartheta = \bar{\theta}(P)$, to get

$$\hat{\bar{\theta}}_n^{(\mathrm{CF})} - \bar{\theta}(P) = \sum_{k=1}^K \frac{n^{(k)}}{n} b_{n,k} + O_P(n^{-1/2}) \ ,$$

and the result follows immediately, since $\left| \sum_{k=1}^K \frac{n^{(k)}}{n} b_{n,k} \right| \leq \max_k |b_{n,k}| = o_P(1)$. ■

B.1.4. *Proof of Theorem 5.3.*

Proof. Use $a_{n,k}, \Delta_{n,k}, b_{n,k}, a, L$ from the proof of Theorem 5.2, and define

$$\mathcal{I}_\theta(P) := \left\{ \boldsymbol{c}_A^\intercal \boldsymbol{\pi}^Q : Q \in \mathcal{Q}_{\mathbf{S}}(P),\ \theta^Q = \bar{\theta}(P) \right\}.$$

By Lemma A.10, this is a nonempty compact interval. Applying the lower bound in Lemma B.7 with $R = \tilde{P}_n^{(-k)}$ and a population conditioning probability nearest to $a_{n,k}$ gives, w.p.a.1,

$$b_{n,k} \geq -\frac{L}{a} \Delta_{n,k} \operatorname{dist}(a_{n,k}, \mathcal{I}_\theta(P)),$$

where $\operatorname{dist}(x, I) := \inf_{r \in I} |x - r|$. By Lemma B.6 and the assumed preliminary rate,

$$\begin{aligned} \max_k |\{b_{n,k}\}_-| &\leq \frac{L}{a} \max_k \Delta_{n,k} \operatorname{dist}(a_{n,k}, \mathcal{I}_\theta(P)) \\ &\leq \frac{LC_P}{a} \max_k \Delta_{n,k}^{3/2} = O_P(n^{-1/2}). \end{aligned}$$

Finally, the expansion and score bound in Lemma B.8, applied as in the consistency proof, give

$$\hat{\bar{\theta}}_n^{(\mathrm{CF})} - \bar{\theta}(P) = \sum_{k=1}^{K} \frac{n^{(k)}}{n} b_{n,k} + O_P(n^{-1/2}).$$

Since these weights are nonnegative and sum to one, the preceding bound implies

$$\sqrt{n}\{\hat{\bar{\theta}}_n^{(\mathrm{CF})} - \bar{\theta}(P)\}_- = O_P(1). \qquad \blacksquare$$

B.1.5. *Proof of Theorem 5.4.*

Proof. Let $b_{n,k}$ and $a > 0$ be as in the proof of Theorem 5.2, and note that the proof of Theorem 5.3 establishes

$$\max_k |\{b_{n,k}\}_-| = O_P(n^{-1/2}).$$

For the moments $m_{n;P}^{(k)} \in \mathfrak{M}_P^\star$ supplied by Assumption 5.4, define

$$e_{n,k} := \sum_{d,z} R_{k,d,z}(\tilde{P}, P; m_{n;P}^{(k)}).$$

By Lemma B.7, w.p.a.1,

$$b_{n,k} \leq \frac{e_{n,k}}{a_{n,k}} \leq \frac{\{e_{n,k}\}_+}{a},$$

and we have $\max_k \{e_{n,k}\}_+ = O_P(n^{-1/2})$ by assumption. Together with the lower bound just obtained, this implies $\max_k |b_{n,k}| = O_P(n^{-1/2})$, and applying Lemma B.8 immediately gives the result. $\blacksquare$

B.1.6. *Proof of Theorem 5.5.*

Proof. Use the fold quantities in (B.1), and use $\Delta_{n,k}, b_{n,k}, \phi_{n,k}, a, L, C_m$ from the proof of Theorem 5.2. By Assumption 5.5, every $Q \in \mathcal{Q}_{\mathbf{S}}(P)$ with $\theta^Q = \bar{\theta}(P)$ has conditioning probability $p_A^\dagger(P)$, as defined there, and so, by Assumption 5.7, we have

$$\max_k \Delta_{n,k} |a_{n,k} - p_A^\dagger(P)| = o_P(n^{-1/2}).$$

For the moments $m_{n;P}^{(k)} \in \mathfrak{M}_P^\star$ supplied by Assumption 5.7, define

$$e_{n,k} := \sum_{d,z} R_{k,d,z}(\tilde{P}, P; m_{n;P}^{(k)}) \ ,$$

and note that the assumption gives $\max_k \{e_{n,k}\}_+ = o_P(n^{-1/2})$ here. By Lemma B.7, w.p.a.1,

$$-\frac{L}{a} \Delta_{n,k} |a_{n,k} - p_A^\dagger(P)| \leq b_{n,k} \leq \frac{\{e_{n,k}\}_+}{a} \ ,$$

and so, $\max_k |b_{n,k}| = o_P(n^{-1/2})$.

It remains to verify convergence of the centred scores. Fix $m_* \in \mathfrak{M}_P^\star$ and set $p_{\min} := \min_z P_Z(z) > 0$. By Lemmas B.9 and B.10,

$$a_{n,k} \xrightarrow{P} p_A^\dagger(P),$$

$$\| m_{n,k} - \mathbb{E}_P[m_{n,k}(W) \mid Z] - m_* + \mathbb{E}_P[m_*(W) \mid Z] \|_{L^2(P)} \xrightarrow{P} 0.$$

The denominators $a_{n,k}$ are bounded below by the constant $a > 0$ from Corollary B.1, w.p.a.1. Using $\|m_*\|_\infty \leq C_m$ gives

$$\begin{aligned} \|\phi_{n,k} - \phi_P\|_{L^2(P)} \leq \frac{1}{a p_{\min}} \| m_{n,k} - \mathbb{E}_P[m_{n,k}(W) \mid Z] - m_* + \mathbb{E}_P[m_*(W) \mid Z] \|_{L^2(P)} \\ + \frac{2C_m}{a p_{\min} p_A^\dagger(P)} |a_{n,k} - p_A^\dagger(P)| = o_P(1). \end{aligned}$$

Set $\xi_{n,k} := \phi_{n,k} - \phi_P$, where ϕ_P is defined in (5.9), and note that these functions satisfy the training measurability condition in Lemma B.8 by construction, and

$$\mathbb{E}_P[\xi_{n,k}(W)] = 0, \qquad \max_k \|\xi_{n,k}\|_{L^2(P)} = o_P(1).$$

The last assertion of Lemma B.8, applied to $\xi_{n,k}$, gives

$$\frac{1}{\sqrt{n}} \sum_{k=1}^K \sum_{i \in I_k} \phi_{n,k}(W_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_P(W_i) + o_P(1).$$

Combining this with (B.9) and $\max_k |b_{n,k}| = o_P(n^{-1/2})$, we have

$$\sqrt{n}(\hat{\bar{\theta}}_n^{(\mathrm{CF})} - \bar{\theta}(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_P(W_i) + o_P(1) \ ,$$

and the result now follows immediately from the central limit theorem, since ϕ_P is bounded and has mean zero. ■

B.1.7. *Proof of Lemma 5.2.*

Proof. Applying Lemma B.11 to the given path $(P_t)_t \subseteq \mathfrak{P}_{\mathbf{S}}$ establishes that the required expansions hold in $\mathfrak{P} \supseteq \mathfrak{P}_{\mathbf{S}}$ and that the claimed derivatives are continuous and linear on $L_0^2(P)$, and therefore on $\dot{\mathfrak{P}}_{\mathbf{S}}$, which immediately implies the result, by definition of path derivative. ■

B.1.8. *Proof of Lemma 5.3.*

Proof. Denote the tangent set of $\mathfrak{P}_{\mathbf{S}}$ by $T_{\mathbf{S}}^P$ and note that $\dot{\mathfrak{P}}_{\mathbf{S}}^P := \overline{\text{span}}^{L^2(P)}\left(T_{\mathbf{S}}^P\right)$ by definition.

First, for any $a : \mathcal{Z} \to \mathbb{R}$ with $\mathbb{E}_P[a(Z)] = 0$,

$$t \mapsto r_t^a(y,d,z) := (1 + ta(z))\, r_0(y,d,z) = (1 + ta(z))\, P_Z(z)\, p_{d|z}^P(y)$$

defines a path $(P_t^a)_t$ in $\mathfrak{P}_{\mathbf{S}}$ via $\frac{dP_t^a}{d\nu} := r_t^a$ since (i) $1 + ta(z) > 0$ for all $z \in \mathcal{Z}$ when t is sufficiently small and r_t^a integrates to 1 over $\mathcal{W}$ since $r_0(y,d,z) = P_Z(z)\, p_{d|z}^P(y)$ is a density on $\mathcal{W}$ and $a(Z)$ has mean zero, and (ii):

$$P_{t;Z}^a(z) = \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} r_t^a(y,d,z)\, d\mu(y) = (1 + ta(z))\, P_Z(z) > 0\,,$$

$$p_{d|z}^{P_t^a}(y) = \frac{r_t^a(y,d,z)}{P_{t;Z}^a(z)} = p_{d|z}^P(y)\,,$$

which implies the conditional (Y,D) distribution is unchanged along this path and, for sufficiently small t, $P_{t;Z}^a(z) > P_Z(z)/2$. Therefore $\left(P_{t;Z}^a\right)_t \subseteq \mathfrak{P}_{\mathbf{S}}$ for sufficiently small t. We now have the following expansion around $t = 0$,

$$\begin{aligned}\left\| \sqrt{r_t^a} - \sqrt{r_0} - \frac{t}{2} a \sqrt{r_0} \right\|_{L^2(\nu)}^2 &= \sum_{z \in \mathcal{Z}} P_Z(z) \left(\sqrt{1 + ta(z)} - 1 - \frac{t}{2} a(z) \right)^2 \\ &\leq C^2 t^4 \left(\sum_{z \in \mathcal{Z}} P_Z(z) a(z)^4 \right) = O(t^4) = o(t^2)\,.\end{aligned}$$

and so $(P_t^a)_t$ is QMD with score a, therefore $a \in T_{\mathbf{S}}^P \subseteq \dot{\mathfrak{P}}_{\mathbf{S}}$.

Next, we have that the linear map $s \mapsto \dot{\boldsymbol{p}}(s)$ satisfies, for any $u \in L_0^2(P)$,

$$\|\dot{\boldsymbol{p}}(u)\|_P = \|u - \mathbb{E}_P[u(W) \mid Z]\|_{L^2(P)} \leq \|u\|_{L^2(P)}\,,$$

where the first equality follows from the isometry established in the main text. This implies

$$\dot{\mathfrak{P}}_{\mathbf{S}}^P \subseteq \left\{ s \in L_0^2(P) \,\middle|\, \dot{\boldsymbol{p}}^P(s) \in \dot{\mathscr{C}}_{\mathbf{S}}^P \right\}\,,$$

since if $s \in \dot{\mathfrak{P}}_{\mathbf{S}}^P$, then there exists a sequence QMD paths in $\mathfrak{P}_{\mathbf{S}}$ with scores s_n such that

$$\|\dot{\boldsymbol{p}}(s_n) - \dot{\boldsymbol{p}}(s)\|_P = \|\dot{\boldsymbol{p}}(s_n - s)\|_P \leq \|s_n - s\|_{L^2(P)} \to 0\,,$$

which implies $\dot{\boldsymbol{p}}(s) \in \dot{\mathscr{C}}_{\mathbf{S}}^P$ since $\dot{\mathscr{C}}_{\mathbf{S}}^P$ is closed in $\mathbb{H}_P$ and each $s_n \in \dot{\mathscr{C}}_{\mathbf{S}}^P$, by definition.

Finally, let $s \in \dot{\mathscr{C}}_{\mathbf{S}}^{P}$. Then, there exists $s_n \in \operatorname{span}\left(T_{\mathbf{S}}^{P}\right)$ such that

$$\|h_{s_n} - h_s\|_{L^2(P)} = \|\dot{\boldsymbol{p}}(s_n) - \dot{\boldsymbol{p}}(s)\|_P \to 0\,,$$

and since $s_n \in \dot{\mathfrak{P}}_{\mathbf{S}}^{P}$, and $z \mapsto \mathbb{E}_P[s_n(W) \mid Z = z], z \mapsto \mathbb{E}_P[s(W) \mid Z = z] \in T_{\mathbf{S}}^{P} \subseteq \dot{\mathfrak{P}}_{\mathbf{S}}^{P}$ by the earlier argument, we have $h_{s_n} \in \dot{\mathfrak{P}}_{\mathbf{S}}^{P}$, and the result now follows immediately by noting that $\dot{\mathfrak{P}}_{\mathbf{S}}^{P}$ is the closed linear span of $T_{\mathbf{S}}^{P}$. ■

B.1.9. *Proof of Theorem 5.6.*

Proof. By the definitions in Lemmas B.12 and B.13, the asserted right-hand side is $\mathscr{U}_P$. For every two-sided QMD path through P in $\mathfrak{P}_{\mathbf{S}}$ with score s, Lemma B.17 gives $\dot{\boldsymbol{p}}^P(s) \in \mathscr{U}_P$. By Lemma B.13, $\mathscr{U}_P$ is a closed linear subspace of $\mathbb{H}_P$, and so the definition of $\dot{\mathscr{C}}_{\mathbf{S}}^{P}$ gives

$$\dot{\mathscr{C}}_{\mathbf{S}}^{P} \subseteq \mathscr{U}_P.$$

Conversely, let $\boldsymbol{b} \in \mathscr{U}_P$. Choose $\dot{\boldsymbol{q}}_d \in L^1(\mu; \mathbb{R}^{\boldsymbol{\tau}})$ and $\dot{\boldsymbol{\pi}} \in \ker(\mathbf{A}_{\mathbf{S},J_{\mathbf{S}}^P})$ such that

$$\begin{aligned}
\mathbf{R}_d \dot{\boldsymbol{q}}_d &= \boldsymbol{b}_d, \\
\int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d \, d\mu &= \dot{\boldsymbol{\pi}} \qquad \forall d, \\
\dot{q}_{d|\boldsymbol{\tau}}(y) &= 0 \qquad \forall d, y, \boldsymbol{\tau} \notin S_d(y).
\end{aligned}$$

Take $(\boldsymbol{\pi}^\dagger, \boldsymbol{q}^\dagger)$ from Lemma B.12, and note that its positivity sets are exactly $S_d(y)$, and the rows of $\mathbf{A}_{\mathbf{S}}\boldsymbol{\pi} \leq \mathbf{a}_{\mathbf{S}}$ binding at $\boldsymbol{\pi}^\dagger$ are exactly $J_{\mathbf{S}}^P$.

Apply Lemma B.14 with $(\boldsymbol{\pi}^0, \boldsymbol{q}^0) = (\boldsymbol{\pi}^\dagger, \boldsymbol{q}^\dagger)$ and $J = J_{\mathbf{S}}^P$, to get $\dot{\boldsymbol{q}}_d^{(n)}$, $\dot{\boldsymbol{\pi}}_n \in \ker(\mathbf{A}_{\mathbf{S},J_{\mathbf{S}}^P})$, and $C_n < \infty$ such that

$$\begin{aligned}
\int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d^{(n)} \, d\mu &= \dot{\boldsymbol{\pi}}_n \qquad \forall d, n, \\
|\dot{q}_{d|\boldsymbol{\tau}}^{(n)}(y)| &\leq C_n q_{d|\boldsymbol{\tau}}^{\dagger}(y) \qquad \forall d, \boldsymbol{\tau}, y, n,
\end{aligned}$$

and with $\boldsymbol{b}_n := (\mathbf{R}_d \dot{\boldsymbol{q}}_d^{(n)})_{d \in \mathcal{D}}$, we have $\|\boldsymbol{b}_n - \boldsymbol{b}\|_P \to 0$ For each n, Lemma B.15 therefore supplies $\epsilon_n > 0$ and a two-sided QMD path $(P_{n,t})_{|t|<\epsilon_n} \subseteq \mathfrak{P}_{\mathbf{S}}$ through P, with score s_n, satisfying

$$\begin{aligned}
\boldsymbol{p}_d^{P_{n,t}} &= \boldsymbol{p}_d^P + t\mathbf{R}_d \dot{\boldsymbol{q}}_d^{(n)} \qquad \forall d,\ |t| < \epsilon_n, \\
\mathbb{E}_P[s_n \mid Z] &= 0, \\
\dot{\boldsymbol{p}}^P(s_n) &= \boldsymbol{b}_n.
\end{aligned}$$

Thus $\boldsymbol{b}_n \in \dot{\mathscr{C}}_{\mathbf{S}}^{P}$ for every n. Since this space is closed in $\|\cdot\|_P$, $\boldsymbol{b} \in \dot{\mathscr{C}}_{\mathbf{S}}^{P}$, proving the reverse inclusion. ■

B.1.10. *Proof of Lemma 5.4.*

Proof. For every stipulated QMD path, Corollary B.2 gives the asserted right derivative. The maximum over conditioning probabilities equals

$$\begin{cases} \dfrac{\delta_P(\dot{\boldsymbol{p}}^P(s))}{\underline{p}^\star_A(P)} & \text{if } \delta_P(\dot{\boldsymbol{p}}^P(s)) \geq 0, \\ \dfrac{\delta_P(\dot{\boldsymbol{p}}^P(s))}{\overline{p}^\star_A(P)} & \text{if } \delta_P(\dot{\boldsymbol{p}}^P(s)) < 0. \end{cases}$$

By Lemma B.21, the quadratic problem (5.14) has a unique solution, and pathwise differentiability is equivalent to (5.13). ∎

B.1.11. *Proof of Theorem 5.7.*

Proof. Let $\boldsymbol{w} := \dot{\boldsymbol{p}}^{P,\star}$ and $a_- := \underline{p}^\star_A(P) > 0$, and define

$$\phi^*_P(y,d,z) := \begin{cases} \dfrac{\boldsymbol{w}_{d;(z)}(y)}{a_- p^P_{d|z}(y)} & \text{if } p^P_{d|z}(y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Every $\boldsymbol{b} \in \dot{\mathscr{C}}^P_{\mathbf{S}}$ satisfies $\sum_d \int_{\mathcal{Y}} \boldsymbol{b}_{d;(z)}\, d\mu = 0$ for every z: this holds for each generating conditional-density derivative by Lemma 5.2, and extends to the closed span because

$$\left| \sum_d \int_{\mathcal{Y}} \boldsymbol{b}_{d;(z)}\, d\mu \right| \leq \frac{1}{\sqrt{P_Z(z)}} \|\boldsymbol{b}\|_P.$$

Applying this to $\boldsymbol{w}$ gives

$$\mathbb{E}_P[\phi^*_P(W) \mid Z = z] = \frac{1}{a_-} \sum_d \int_{\mathcal{Y}} \boldsymbol{w}_{d;(z)}\, d\mu = 0,$$

$$\|\phi^*_P\|^2_{L^2(P)} = \frac{\|\boldsymbol{w}\|^2_P}{a_-^2}.$$

Moreover, $\dot{\boldsymbol{p}}^P(\phi^*_P) = \boldsymbol{w}/a_-$, and so Lemma 5.3 gives $\phi^*_P \in \dot{\mathfrak{P}}^P_{\mathbf{S}}$.

For every path score s, Lemma B.21 gives

$$\begin{aligned} \dot{\theta}_P(s) &= \frac{1}{a_-} \left\langle \boldsymbol{w}, \dot{\boldsymbol{p}}^P(s) \right\rangle_P \\ &= \frac{1}{a_-} \sum_{d,z} P_Z(z) \int_{\mathcal{Y}} \boldsymbol{w}_{d;(z)}(y) \left(s(y,d,z) - \mathbb{E}_P[s(W) \mid Z = z] \right) d\mu(y) \\ &= \mathbb{E}_P[\phi^*_P(W) s(W)]. \end{aligned}$$

The last equality uses the conditional mean-zero property proved above. Thus ϕ^*_P is the tangent-space representative of the derivative, and hence the EIF. Its squared norm gives the stated efficiency bound. The same argument applies with known assignment probabilities, since ϕ^*_P is conditionally centred. ∎

B.1.12. *Proof of Corollary 5.1.*

Proof. Let $a_* := p_A^\dagger(P) = \underline{p}_A^\star(P) = \overline{p}_A^\star(P)$. By Lemma B.21, pathwise differentiability is equivalent to

$$\delta_P(\boldsymbol{b}) = \left\langle \dot{\boldsymbol{p}}^{P,\star}, \boldsymbol{b} \right\rangle_P \qquad \forall \boldsymbol{b} \in \dot{\mathscr{C}}_{\mathbf{S}}^P.$$

This implies linearity. Conversely, if δ_P is linear, its continuity and the Riesz representation theorem give $\delta_P(\boldsymbol{b}) = \langle \boldsymbol{a}, \boldsymbol{b} \rangle_P$ for some $\boldsymbol{a} \in \dot{\mathscr{C}}_{\mathbf{S}}^P$. Completing the square gives the unique quadratic maximizer $\dot{\boldsymbol{p}}^{P,\star} = \boldsymbol{a}$, proving the equivalence of the first two statements.

If δ_P is linear, then, for every $m \in \mathfrak{M}_P^\star$ and $\boldsymbol{b} \in \dot{\mathscr{C}}_{\mathbf{S}}^P$,

$$\delta_P(\boldsymbol{b}) \le \sum_{d,z} \int_{\mathcal{Y}} m \boldsymbol{b}_{d;(z)} \, d\mu,$$

$$-\delta_P(\boldsymbol{b}) = \delta_P(-\boldsymbol{b}) \le -\sum_{d,z} \int_{\mathcal{Y}} m \boldsymbol{b}_{d;(z)} \, d\mu.$$

Thus

$$\delta_P(\boldsymbol{b}) = \sum_{d,z} \int_{\mathcal{Y}} m \boldsymbol{b}_{d;(z)} \, d\mu \qquad \forall m \in \mathfrak{M}_P^\star,\ \boldsymbol{b} \in \dot{\mathscr{C}}_{\mathbf{S}}^P.$$

This gives the third statement. Conversely, if all these linear functionals agree, their infimum δ_P equals any one of them and is linear.

Under the equivalent conditions, substituting the last display into (5.14) gives the asserted quadratic objective for every $m \in \mathfrak{M}_P^\star$. The EIF and variance formulas follow from Theorem 5.7 with $a_- = a_*$. ∎

B.1.13. *Proof of Lemma 5.5.*

Proof. Take $Q^\dagger$ from Assumption 5.8, and let $\boldsymbol{\pi}^\dagger := \boldsymbol{\pi}^{Q^\dagger}$ and $\boldsymbol{q}_d^\dagger := \boldsymbol{q}_d^{Q^\dagger}$. Since $\boldsymbol{\pi}^\dagger \in \operatorname{relint} \mathcal{R}_{\boldsymbol{\pi}}$,

$$\left\{ j \in [m_{\mathbf{S}}] \,\middle|\, \mathbf{A}_{\mathbf{S};(j,\cdot)} \boldsymbol{\pi}^\dagger = \mathbf{a}_{\mathbf{S};(j)} \right\} = E_{\mathbf{S}}.$$

Indeed, a valid linear inequality binds at a relative-interior point of a convex set if and only if it binds throughout that set. Since $\boldsymbol{\pi}^\dagger \in \boldsymbol{\Pi}(P) \subseteq \mathcal{R}_{\boldsymbol{\pi}}$, it follows that $J_{\mathbf{S}}^P = E_{\mathbf{S}}$.

By Theorem 5.6, every $\boldsymbol{b} \in \dot{\mathscr{C}}_{\mathbf{S}}^P$ therefore has a representation $\boldsymbol{b}_d = \mathbf{R}_d \dot{\boldsymbol{q}}_d$ with common integral $\dot{\boldsymbol{\pi}} \in \ker(\mathbf{A}_{\mathbf{S},E})$. In particular,

$$\boldsymbol{b}_d(y) \in \operatorname{col}(\mathbf{R}_d),$$

$$\int_{\mathcal{Y}} \boldsymbol{b}_d \, d\mu = \mathbf{R}_d \int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d \, d\mu = \mathbf{R}_d \dot{\boldsymbol{\pi}},$$

which proves one inclusion.

Conversely, let $\boldsymbol{b} \in \mathbb{H}_P$ satisfy

$$\boldsymbol{b}_d(y) \in \operatorname{col}(\mathbf{R}_d) \qquad \forall d, y,$$

$$\int_{\mathcal{Y}} \boldsymbol{b}_d \, d\mu = \mathbf{R}_d \dot{\boldsymbol{\pi}} \qquad \forall d$$

for some $\dot{\boldsymbol{\pi}} \in \ker(\mathbf{A}_{\mathbf{S},E})$. Define

$$S_d^\dagger(y) := \left\{ \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \,\middle|\, q^\dagger_{d|\boldsymbol{\tau}}(y) > 0 \right\}.$$

Since $\boldsymbol{b}_{d;(z)}(y) = 0$ whenever $p^P_{d|z}(y) = 0$, Assumption 5.8(iii) gives

$$\boldsymbol{b}_d(y) \in \operatorname{col}\left(\mathbf{R}_d(y; Q^\dagger)\right) = \operatorname{col}\left(\mathbf{R}_{d;(\cdot, S_d^\dagger(y))}\right).$$

Use the fixed linear maps $\boldsymbol{L}_{d,S}$ constructed in the proof of Lemma B.13, and set

$$\dot{\boldsymbol{q}}_d^0(y) := \boldsymbol{L}_{d,S_d^\dagger(y)} \boldsymbol{b}_d(y).$$

These functions are measurable and satisfy

$$\begin{aligned} &\mathbf{R}_d \dot{\boldsymbol{q}}_d^0 = \boldsymbol{b}_d, \\ &\dot{q}^0_{d|\boldsymbol{\tau}}(y) = 0 \qquad \forall d, \boldsymbol{\tau}, y \text{ with } q^\dagger_{d|\boldsymbol{\tau}}(y) = 0. \end{aligned}$$

There are finitely many maps $\boldsymbol{L}_{d,S}$. Writing $C_L := \max_{d,S} \|\boldsymbol{L}_{d,S}\|_{1\leftarrow 1}$, the bound in Lemma B.13 gives

$$\begin{aligned} \sum_d \int_{\mathcal{Y}} \|\dot{\boldsymbol{q}}_d^0\|_1 \, d\mu &\le C_L \|\boldsymbol{b}\|_{L^1} \\ &\le C_L \left(\sum_z \frac{1}{P_Z(z)} \right)^{1/2} \|\boldsymbol{b}\|_P < \infty. \end{aligned}$$

It remains to correct the integrals without changing the response images. Define

$$\boldsymbol{e}_d := \dot{\boldsymbol{\pi}} - \int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d^0 \, d\mu.$$

Then

$$\mathbf{R}_d \boldsymbol{e}_d = \mathbf{R}_d \dot{\boldsymbol{\pi}} - \int_{\mathcal{Y}} \boldsymbol{b}_d \, d\mu = \mathbf{0}.$$

For each $d \in \mathcal{D}$, define

$$\rho_d(y) := \frac{\min_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d} q^\dagger_{d|\boldsymbol{\tau}}(y)}{\displaystyle\int_{\mathcal{Y}} \min_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d} q^\dagger_{d|\boldsymbol{\tau}} \, d\mu}.$$

The denominator is positive by Assumption 5.8(ii), and $\int_{\mathcal{Y}} \rho_d \, d\mu = 1$.

Every $\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}$ belongs to $\boldsymbol{\mathcal{T}}_d$ for some d, and for such a d,

$$\begin{aligned} \boldsymbol{\pi}^\dagger_{(\boldsymbol{\tau})} &= \int_{\mathcal{Y}} q^\dagger_{d|\boldsymbol{\tau}} \, d\mu \\ &\ge \int_{\mathcal{Y}} \min_{\boldsymbol{\tau}' \in \boldsymbol{\mathcal{T}}_d} q^\dagger_{d|\boldsymbol{\tau}'} \, d\mu > 0. \end{aligned}$$

We can therefore define

$$\boldsymbol{k}_{d;(\boldsymbol{\tau})}(y) := \begin{cases} \boldsymbol{e}_{d;(\boldsymbol{\tau})}\rho_d(y) & \text{if } \boldsymbol{\tau} \in \mathcal{T}_d, \\ \boldsymbol{e}_{d;(\boldsymbol{\tau})} \dfrac{q^\dagger_{d|\boldsymbol{\tau}}(y)}{\boldsymbol{\pi}^\dagger_{(\boldsymbol{\tau})}} & \text{otherwise.} \end{cases}$$

Each density used in this definition integrates to one, and the inactive columns of $\mathbf{R}_d$ are zero, and so

$$\int_{\mathcal{Y}} \boldsymbol{k}_d \, d\mu = \boldsymbol{e}_d, \qquad \int_{\mathcal{Y}} \|\boldsymbol{k}_d\|_1 \, d\mu = \|\boldsymbol{e}_d\|_1 < \infty,$$
$$\mathbf{R}_d \boldsymbol{k}_d(y) = \rho_d(y) \mathbf{R}_d \boldsymbol{e}_d = \mathbf{0}.$$

Also, $\boldsymbol{k}_{d;(\boldsymbol{\tau})}(y) = 0$ whenever $q^\dagger_{d|\boldsymbol{\tau}}(y) = 0$: on active coordinates this follows from the minimum defining ρ_d, and on inactive coordinates it follows from the displayed formula for $\boldsymbol{k}_d$.

Set $\dot{\boldsymbol{q}}_d := \dot{\boldsymbol{q}}^0_d + \boldsymbol{k}_d$. Then

$$\mathbf{R}_d \dot{\boldsymbol{q}}_d = \boldsymbol{b}_d,$$
$$\int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d \, d\mu = \dot{\boldsymbol{\pi}} \in \ker(\mathbf{A}_{\mathbf{S},E}) = \ker(\mathbf{A}_{\mathbf{S},J^P_{\mathbf{S}}}),$$
$$\dot{q}_{d|\boldsymbol{\tau}}(y) = 0 \qquad \forall d, \boldsymbol{\tau}, y \text{ with } q^\dagger_{d|\boldsymbol{\tau}}(y) = 0.$$

By Lemma B.12, $S^\dagger_d(y) \subseteq S_d(y)$, and so $\dot{q}_{d|\boldsymbol{\tau}}(y) = 0$ whenever $\boldsymbol{\tau} \notin S_d(y)$. Therefore, Theorem 5.6 gives $\boldsymbol{b} \in \dot{\mathscr{C}}^P_{\mathbf{S}}$ and completes the proof. ∎

B.1.14. *Proof of Theorem 5.8.*

Proof. By Theorem 5.5,

$$\sqrt{n}(\hat{\bar{\theta}}^{(\mathrm{CF})}_n - \bar{\theta}(P)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_P(W_i) + o_P(1), \tag{B.2}$$

where, for any $m \in \mathfrak{M}^\star_P$,

$$\phi_P(W) = \frac{m(W) - \mathbb{E}_P[m(W) \mid Z]}{P_Z(Z) p^\dagger_A(P)}.$$

In particular, ϕ_P is bounded and $\mathbb{E}_P[\phi_P(W) \mid Z] = 0$.

For every $\boldsymbol{b} \in \dot{\mathscr{C}}^P_{\mathbf{S}}$, $\sum_d \int_{\mathcal{Y}} \boldsymbol{b}_{d;(z)} \, d\mu = 0$ for each z, as established in the proof of Theorem 5.7. By Assumption 5.6, every $m_1, m_2 \in \mathfrak{M}^\star_P$ therefore satisfies

$$\sum_{d,z} \int_{\mathcal{Y}} (m_1 - m_2) \boldsymbol{b}_{d;(z)} \, d\mu = \sum_{d,z} \int_{\mathcal{Y}} (\mathbb{E}_P[m_1(W) \mid Z = z] - \mathbb{E}_P[m_2(W) \mid Z = z]) \, \boldsymbol{b}_{d;(z)} \, d\mu$$
$$= 0.$$

By Corollary 5.1, $\bar{\theta}$ is pathwise differentiable.

For any QMD path score s, put $h_s := s - \mathbb{E}_P[s(W) \mid Z]$. Using Lemma 5.2 and Corollary 5.1, and writing $a_* := p_A^\dagger(P)$, gives

$$\begin{aligned}\dot{\bar{\theta}}_P(s) &= \frac{1}{a_*}\sum_{d,z}\int_{\mathcal{Y}} m(y,d,z)p^P_{d|z}(y)h_s(y,d,z)\,d\mu(y)\\ &= \frac{1}{a_*}\sum_{d,z}\int_{\mathcal{Y}} \left(m(y,d,z) - \mathbb{E}_P[m(W)\mid Z=z]\right)p^P_{d|z}(y)s(y,d,z)\,d\mu(y)\\ &= \mathbb{E}_P[\phi_P(W)s(W)].\end{aligned}$$

The second equality uses $\mathbb{E}_P[h_s(W)\mid Z]=0$ and $\mathbb{E}_P[m(W)-\mathbb{E}_P[m(W)\mid Z]\mid Z]=0$.

To show regularity, fix a two-sided QMD path $(P_t)_{|t|<\epsilon}\subseteq \mathfrak{P}_{\mathbf{S}}$ through P with score s, and fix $h\in\mathbb{R}$. Writing $r_t := dP_t/d\nu$, local asymptotic normality gives, under P,

$$\sum_{i=1}^n \log\frac{r_{h/\sqrt{n}}(W_i)}{r_0(W_i)} = \frac{h}{\sqrt{n}}\sum_{i=1}^n s(W_i) - \frac{h^2}{2}\|s\|^2_{L^2(P)} + o_P(1);$$

see Van Der Vaart (1998, Chapter 7). Independent randomization used for the folds or fitting has the same law under the local alternatives and leaves this likelihood ratio unchanged.

The joint central limit theorem gives

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\begin{bmatrix}\phi_P(W_i)\\ s(W_i)\end{bmatrix} \rightsquigarrow \mathcal{N}\left(\mathbf{0},\begin{bmatrix}\mathbb{E}_P[\phi_P^2] & \mathbb{E}_P[\phi_P s]\\ \mathbb{E}_P[\phi_P s] & \mathbb{E}_P[s^2]\end{bmatrix}\right).$$

Contiguity transfers the remainder in (B.2) to $P_{h/\sqrt{n}}$, and Le Cam's third lemma gives, under these local alternatives,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_P(W_i) \rightsquigarrow \mathcal{N}\left(h\mathbb{E}_P[\phi_P s],\, \mathbb{E}_P[\phi_P^2]\right);$$

see Van Der Vaart and Wellner (1996, Section 3.10). Pathwise differentiability also gives

$$\sqrt{n}\left(\bar{\theta}(P_{h/\sqrt{n}}) - \bar{\theta}(P)\right) \to h\dot{\bar{\theta}}_P(s) = h\mathbb{E}_P[\phi_P s].$$

Subtracting this deterministic shift gives

$$\sqrt{n}\left(\hat{\bar{\theta}}_n^{(\mathrm{CF})} - \bar{\theta}(P_{h/\sqrt{n}})\right) \rightsquigarrow \mathcal{N}(0, \mathbb{E}_P[\phi_P^2]),$$

as needed.

The EIF is the orthogonal projection of ϕ_P onto $\dot{\mathfrak{P}}^P_{\mathbf{S}}$. Therefore

$$\|\phi_P\|^2_{L^2(P)} = \|\phi_P^*\|^2_{L^2(P)} + \|\phi_P - \phi_P^*\|^2_{L^2(P)},$$

and efficiency is equivalent to $\phi_P \in \dot{\mathfrak{P}}^P_{\mathbf{S}}$. Since ϕ_P is conditionally centred, Lemma 5.3 makes this equivalent to

$$\left(\left(p^P_{d|z}\phi_P(\cdot,d,z)\right)_{z\in\mathcal{Z}}\right)_{d\in\mathcal{D}} \in \dot{\mathscr{C}}^P_{\mathbf{S}}.$$

Under Assumption 5.8, Lemma 5.5 identifies this membership with the two requirements

$$\left(p^P_{d|z}(y)\phi_P(y,d,z)\right)_{z\in\mathcal{Z}} \in \operatorname{col}(\mathbf{R}_d) \qquad \forall d, y,$$
$$\mathbf{R}_d\dot{\boldsymbol{\pi}} = \left(\mathbb{E}_P[\mathbb{1}\{D=d\}\phi_P(W) \mid Z=z]\right)_{z\in\mathcal{Z}} \qquad \forall d$$

for some $\dot{\boldsymbol{\pi}} \in \ker(\mathbf{A}_{\mathbf{S},E})$. By Corollary B.3, they hold under (5.17), with the stated simplification under support restrictions alone.

Finally, conditioning ϕ^2_P on Z gives

$$\begin{aligned}\mathbb{E}_P[\phi_P(W)^2] &= \frac{1}{p^\dagger_A(P)^2}\sum_{z\in\mathcal{Z}} \frac{\mathbb{E}_P\left[(m(W)-\mathbb{E}_P[m(W)\mid Z=z])^2 \,\middle|\, Z=z\right]}{P_Z(z)} \\ &= \frac{1}{p^\dagger_A(P)^2}\sum_{z\in\mathcal{Z}} \frac{\operatorname{Var}_P(m(W)\mid Z=z)}{P_Z(z)}.\end{aligned}$$

When the estimator is efficient, $\phi_P = \phi^*_P$ in $L^2(P)$, and this is the efficiency bound. ■

B.2. **Auxiliary Estimation Lemmas**

Lemma B.1. *Suppose* $\mathbf{S}$ *satisfies Condition CP. For each* $d \in \mathcal{D}$, *let* $\boldsymbol{C}_d : \mathcal{Y} \to \mathbb{R}^{\mathcal{T}\times\mathcal{T}}$ *be measurable and diagonal, with diagonal entries in* $\{0,1\}$ *and zero entries on* $\mathcal{T}\setminus\mathcal{T}_d$. *Fix* $\boldsymbol{E} \in \mathbb{R}^{r\times\mathcal{T}}$ *and* $\boldsymbol{e}\in\mathbb{R}^r$, *allowing* $r=0$. *For* P' *satisfying Condition D, define*

$$\mathcal{F}^{P'}_{\mathbf{S};\boldsymbol{C},\boldsymbol{E},\boldsymbol{e}} := \left\{ (\boldsymbol{\pi}',\boldsymbol{q}') \in \mathcal{F}^{P'}_{\mathbf{S}} \;\middle|\; \begin{array}{l} \boldsymbol{E}\boldsymbol{\pi}' = \boldsymbol{e}, \\ \boldsymbol{C}_d(y)\boldsymbol{q}'^a_d(y) = \mathbf{0} \quad \forall d\in\mathcal{D},\ y\in\mathcal{Y} \end{array} \right\}.$$

There is $C_0 < \infty$, *depending only on* $(\mathbf{R}_d)_{d\in\mathcal{D}}$, $\mathbf{A}_{\mathbf{S}}$, *and* $\boldsymbol{E}$, *such that whenever this set is nonempty, every* P *satisfying Condition D and* $(\boldsymbol{\pi},\boldsymbol{q}) \in \mathcal{F}^P_{\mathbf{S}}$ *admit* $(\boldsymbol{\pi}',\boldsymbol{q}') \in \mathcal{F}^{P'}_{\mathbf{S};\boldsymbol{C},\boldsymbol{E},\boldsymbol{e}}$ *with*

$$\|(\boldsymbol{\pi}',\boldsymbol{q}') - (\boldsymbol{\pi},\boldsymbol{q})\|_{\mathcal{F},1} \le C_0\left[\Delta(P',P) + \sum_{d\in\mathcal{D}}\int_{\mathcal{Y}} \|\boldsymbol{C}_d\boldsymbol{q}^a_d\|_1\, d\mu + \|\boldsymbol{E}\boldsymbol{\pi}-\boldsymbol{e}\|_1\right]. \tag{B.3}$$

In particular, C_0 *does not depend on the laws,* $\boldsymbol{e}$, *or the measurable locations at which the entries of* $\boldsymbol{C}_d$ *equal one.*

Proof. Fix $(\boldsymbol{\pi},\boldsymbol{q})$ and a target witness $(\bar{\boldsymbol{\pi}},\bar{\boldsymbol{q}}) \in \mathcal{F}^{P'}_{\mathbf{S};\boldsymbol{C},\boldsymbol{E},\boldsymbol{e}}$. We encode the new active density as $\boldsymbol{t}_d + \boldsymbol{u}_d$, with $\boldsymbol{t}_d + \boldsymbol{v}_d = \boldsymbol{q}^a_d$. This makes its change equal to $\boldsymbol{u}_d - \boldsymbol{v}_d$.

Let $\boldsymbol{I}$ be the identity on $\mathbb{R}^{\mathcal{T}}$, and define

$$\mathscr{D}_d := \left\{\operatorname{diag}(\boldsymbol{c}) \,\middle|\, \boldsymbol{c}\in\{0,1\}^{\mathcal{T}},\ \boldsymbol{c}_{(\mathcal{T}\setminus\mathcal{T}_d)} = \mathbf{0}\right\}.$$

For $\boldsymbol{C}\in\mathscr{D}_d$, define

$$\boldsymbol{H}_{d,\boldsymbol{C}} := \begin{bmatrix} \mathbf{R}_d & \mathbf{R}_d & \mathbf{0} \\ \boldsymbol{I} & \mathbf{0} & \boldsymbol{I} \\ \mathbf{0} & \boldsymbol{I}_{(\mathcal{T}\setminus\mathcal{T}_d,\cdot)} & \mathbf{0} \\ \boldsymbol{C} & \boldsymbol{C} & \mathbf{0} \end{bmatrix},$$

$$\boldsymbol{b}_d(y) := \begin{bmatrix} \boldsymbol{p}_d^{P'}(y) \\ \boldsymbol{q}_d^a(y) \\ \mathbf{0}_{\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d} \\ \mathbf{0}_{\boldsymbol{\mathcal{T}}} \end{bmatrix}, \qquad A_{d,\boldsymbol{C}} := \{y \in \mathcal{Y} \,|\, \boldsymbol{C}_d(y) = \boldsymbol{C}\}.$$

The binary matrix $\boldsymbol{H}_{d,\boldsymbol{C}}$ has no zero columns: the identity blocks cover its first and third column blocks, and $\mathbf{R}_d$ and $\boldsymbol{I}_{(\boldsymbol{\mathcal{T}}\setminus\boldsymbol{\mathcal{T}}_d,\cdot)}$ cover its second. For every $y \in A_{d,\boldsymbol{C}}$,

$$\boldsymbol{H}_{d,\boldsymbol{C}} \begin{bmatrix} \mathbf{0} \\ \overline{\boldsymbol{q}_d}^a(y) \\ \boldsymbol{q}_d^a(y) \end{bmatrix} = \boldsymbol{b}_d(y).$$

Therefore $\mathbb{1}_{A_{d,\boldsymbol{C}}}\boldsymbol{b}_d$ is integrable and takes values in $\operatorname{cone}\boldsymbol{H}_{d,\boldsymbol{C}}$. Apply Lemma A.3 to this function and matrix, and intersect the resulting partition with $A_{d,\boldsymbol{C}}$, discarding empty cells. Denote this partition of $A_{d,\boldsymbol{C}}$ by $\mathcal{B}_{d,\boldsymbol{C}}$, and let

$$\mathcal{B}_d := \bigcup_{\boldsymbol{C}\in\mathscr{D}_d} \mathcal{B}_{d,\boldsymbol{C}}.$$

In particular, $|\mathcal{B}_{d,\boldsymbol{C}}| \le N_{\boldsymbol{H}_{d,\boldsymbol{C}}}$.

Consider the finite system

$$\begin{aligned} \boldsymbol{\pi}' \ge \mathbf{0}, \qquad \mathbf{A}_{\mathbf{S}}\boldsymbol{\pi}' \le \mathbf{a}_{\mathbf{S}}, \qquad \boldsymbol{E}\boldsymbol{\pi}' = \boldsymbol{e}, \\ \boldsymbol{t}_{d,B}, \boldsymbol{u}_{d,B}, \boldsymbol{v}_{d,B} \ge \mathbf{0}, \\ \boldsymbol{H}_{d,\boldsymbol{C}} \begin{bmatrix} \boldsymbol{t}_{d,B} \\ \boldsymbol{u}_{d,B} \\ \boldsymbol{v}_{d,B} \end{bmatrix} = \int_B \boldsymbol{b}_d \, d\mu \quad \forall d \in \mathcal{D},\ \boldsymbol{C} \in \mathscr{D}_d,\ B \in \mathcal{B}_{d,\boldsymbol{C}}, \\ \sum_{B\in\mathcal{B}_d} (\boldsymbol{t}_{d,B} + \boldsymbol{u}_{d,B})_{(\boldsymbol{\mathcal{T}}_d)} = \boldsymbol{\pi}'_{(\boldsymbol{\mathcal{T}}_d)} \qquad \forall d \in \mathcal{D}. \end{aligned} \tag{B.4}$$

This system is feasible: take $\boldsymbol{\pi}' = \bar{\boldsymbol{\pi}}$ and

$$(\boldsymbol{t}_{d,B}, \boldsymbol{u}_{d,B}, \boldsymbol{v}_{d,B}) = \left(\mathbf{0}, \int_B \overline{\boldsymbol{q}_d}^a \, d\mu, \int_B \boldsymbol{q}_d^a \, d\mu\right).$$

Let $\boldsymbol{w}$ collect the variables in (B.4), and define the source vector

$$\boldsymbol{w}^0 := \left(\boldsymbol{\pi}, \left(\int_B \boldsymbol{q}_d^a \, d\mu, \mathbf{0}, \mathbf{0}\right)_{d\in\mathcal{D},\ B\in\mathcal{B}_d}\right).$$

At $\boldsymbol{w}^0$, all inequalities and active-mass equations in (B.4) hold. The remaining equality residuals are

$$\begin{bmatrix} \int_B (\boldsymbol{p}_d^P - \boldsymbol{p}_d^{P'})\, d\mu \\ \mathbf{0} \\ \mathbf{0} \\ \int_B \boldsymbol{C}\boldsymbol{q}_d^a\, d\mu \end{bmatrix}, \qquad \boldsymbol{E}\boldsymbol{\pi} - \boldsymbol{e}.$$

Their total ℓ^1 norm is at most

$$\eta := \Delta(P', P) + \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} \|\boldsymbol{C}_d \boldsymbol{q}_d^a\|_1\, d\mu + \|\boldsymbol{E}\boldsymbol{\pi} - \boldsymbol{e}\|_1.$$

Hoffman's error bound applies to the nonempty finite system (B.4), and supplies a feasible $\boldsymbol{w}$ with

$$\|\boldsymbol{w} - \boldsymbol{w}^0\|_1 \le C_H \eta.$$

The coefficient matrices of that system vary only with the finitely many possible cell counts $(|\mathcal{B}_{d,\boldsymbol{C}}|)_{d,\boldsymbol{C}}$, each bounded by $N_{\boldsymbol{H}_{d,\boldsymbol{C}}}$. Taking the maximum of their Hoffman constants gives one C_H depending only on $(\mathbf{R}_d)_{d\in\mathcal{D}}$, $\mathbf{A}_{\mathbf{S}}$, and $\boldsymbol{E}$. Since the source $\boldsymbol{u}$ and $\boldsymbol{v}$ coordinates are zero,

$$\|\boldsymbol{\pi}' - \boldsymbol{\pi}\|_1 + \sum_{d\in\mathcal{D}} \sum_{B\in\mathcal{B}_d} \mathbf{1}^{\mathsf{T}}(\boldsymbol{u}_{d,B} + \boldsymbol{v}_{d,B}) \le C_H \eta. \tag{B.5}$$

By Lemma A.3, the nonnegative cell vectors in (B.4) can be realized by nonnegative integrable functions $\boldsymbol{t}_d, \boldsymbol{u}_d, \boldsymbol{v}_d$ satisfying

$$\boldsymbol{H}_{d,\boldsymbol{C}} \begin{bmatrix} \boldsymbol{t}_d(y) \\ \boldsymbol{u}_d(y) \\ \boldsymbol{v}_d(y) \end{bmatrix} = \boldsymbol{b}_d(y) \qquad \forall y \in B,$$

$$\int_B \begin{bmatrix} \boldsymbol{t}_d \\ \boldsymbol{u}_d \\ \boldsymbol{v}_d \end{bmatrix} d\mu = \begin{bmatrix} \boldsymbol{t}_{d,B} \\ \boldsymbol{u}_{d,B} \\ \boldsymbol{v}_{d,B} \end{bmatrix}$$

for every $d \in \mathcal{D}$, $\boldsymbol{C} \in \mathscr{D}_d$, and $B \in \mathcal{B}_{d,\boldsymbol{C}}$. The block equations give

$$\begin{aligned} \boldsymbol{t}_d + \boldsymbol{v}_d &= \boldsymbol{q}_d^a, \\ \mathbf{R}_d(\boldsymbol{t}_d + \boldsymbol{u}_d) &= \boldsymbol{p}_d^{P'}, \\ (\boldsymbol{u}_d)_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} &= \mathbf{0}, \\ \boldsymbol{C}_d(\boldsymbol{t}_d + \boldsymbol{u}_d) &= \mathbf{0}. \end{aligned}$$

Nonnegativity and $\boldsymbol{t}_d + \boldsymbol{v}_d = \boldsymbol{q}_d^a$ also force the inactive coordinates of $\boldsymbol{t}_d$ and $\boldsymbol{v}_d$ to be zero. Set $\boldsymbol{q}_d'^a := \boldsymbol{t}_d + \boldsymbol{u}_d$. The last equations in (B.4) give its required active masses, and

$$\begin{aligned}\int_{\mathcal{Y}} \|\boldsymbol{q}_d'^a - \boldsymbol{q}_d^a\|_1 \, d\mu &= \int_{\mathcal{Y}} \|\boldsymbol{u}_d - \boldsymbol{v}_d\|_1 \, d\mu \\ &\le \sum_{B \in \mathcal{B}_d} \mathbf{1}^\intercal (\boldsymbol{u}_{d,B} + \boldsymbol{v}_{d,B}).\end{aligned}$$

Let ρ_0 be the probability density from Lemma A.1. For $\boldsymbol{\tau} \notin \boldsymbol{\mathcal{T}}_d$, define

$$q'_{d|\boldsymbol{\tau}}(y) := \begin{cases} \dfrac{\boldsymbol{\pi}'_{(\boldsymbol{\tau})}}{\boldsymbol{\pi}_{(\boldsymbol{\tau})}} q_{d|\boldsymbol{\tau}}(y) & \text{if } \boldsymbol{\pi}_{(\boldsymbol{\tau})} > 0, \\ \boldsymbol{\pi}'_{(\boldsymbol{\tau})} \rho_0(y) & \text{otherwise.} \end{cases}$$

In either case,

$$\int_{\mathcal{Y}} q'_{d|\boldsymbol{\tau}} \, d\mu = \boldsymbol{\pi}'_{(\boldsymbol{\tau})},$$

$$\int_{\mathcal{Y}} \left| q'_{d|\boldsymbol{\tau}} - q_{d|\boldsymbol{\tau}} \right| \, d\mu = \left| \boldsymbol{\pi}'_{(\boldsymbol{\tau})} - \boldsymbol{\pi}_{(\boldsymbol{\tau})} \right|.$$

These coordinates do not enter $\mathbf{R}_d \boldsymbol{q}'_d$. The resulting tuple belongs to $\mathcal{F}^{P'}_{\mathbf{S};\boldsymbol{C},\boldsymbol{E},\boldsymbol{e}}$, and

$$\begin{aligned}\|(\boldsymbol{\pi}', \boldsymbol{q}') - (\boldsymbol{\pi}, \boldsymbol{q})\|_{\mathcal{F},1} &\le (1 + |\mathcal{D}|) \|\boldsymbol{\pi}' - \boldsymbol{\pi}\|_1 + \sum_{d \in \mathcal{D}} \sum_{B \in \mathcal{B}_d} \mathbf{1}^\intercal (\boldsymbol{u}_{d,B} + \boldsymbol{v}_{d,B}) \\ &\le (1 + |\mathcal{D}|) C_H \eta,\end{aligned}$$

by (B.5). Take $C_0 = (1 + |\mathcal{D}|) C_H$. ■

Lemma B.2. *Suppose* $\mathbf{S}$ *satisfies Condition CP. There is* $C < \infty$*, depending only on* $(\mathbf{R}_d)_{d \in \mathcal{D}}$ *and* $\mathbf{A}_{\mathbf{S}}$*, such that, for any* P, P' *satisfying Condition D with* $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$ *and* $\mathcal{Q}_{\mathbf{S}}(P') \neq \emptyset$*, every* $(\boldsymbol{\pi}, \boldsymbol{q}) \in \mathcal{F}^P_{\mathbf{S}}$ *admits* $(\boldsymbol{\pi}', \boldsymbol{q}') \in \mathcal{F}^{P'}_{\mathbf{S}}$ *satisfying*

$$\|(\boldsymbol{\pi}', \boldsymbol{q}') - (\boldsymbol{\pi}, \boldsymbol{q})\|_{\mathcal{F},1} \le C \Delta(P', P). \tag{B.6}$$

For $b_0 \in \mathbb{R}$*,* $\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$*, and bounded measurable* $\boldsymbol{g}_d : \mathcal{Y} \to \mathbb{R}^{\boldsymbol{\mathcal{T}}}$*, define*

$$T(\boldsymbol{\pi}, \boldsymbol{q}) := b_0 + \boldsymbol{b}^\intercal \boldsymbol{\pi} + \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} \boldsymbol{g}_d^\intercal \boldsymbol{q}_d \, d\mu,$$

$$L_T := \max \left\{ \|\boldsymbol{b}\|_\infty, \max_{d \in \mathcal{D}} \sup_{y \in \mathcal{Y}} \|\boldsymbol{g}_d(y)\|_\infty \right\}.$$

The same $(\boldsymbol{\pi}', \boldsymbol{q}')$ *satisfies*

$$|T(\boldsymbol{\pi}', \boldsymbol{q}') - T(\boldsymbol{\pi}, \boldsymbol{q})| \le C L_T \Delta(P', P).$$

Further,

$$\left|\inf_{(\boldsymbol{\pi}',\boldsymbol{q}')\in\mathcal{F}_{\mathbf{S}}^{P'}} T(\boldsymbol{\pi}',\boldsymbol{q}') - \inf_{(\boldsymbol{\pi},\boldsymbol{q})\in\mathcal{F}_{\mathbf{S}}^{P}} T(\boldsymbol{\pi},\boldsymbol{q})\right| \le CL_T\Delta(P',P),$$

$$\left|\sup_{(\boldsymbol{\pi}',\boldsymbol{q}')\in\mathcal{F}_{\mathbf{S}}^{P'}} T(\boldsymbol{\pi}',\boldsymbol{q}') - \sup_{(\boldsymbol{\pi},\boldsymbol{q})\in\mathcal{F}_{\mathbf{S}}^{P}} T(\boldsymbol{\pi},\boldsymbol{q})\right| \le CL_T\Delta(P',P).$$

Proof. Apply Lemma B.1 with $\boldsymbol{C}_d \equiv \mathbf{0}$, $\boldsymbol{E} \in \mathbb{R}^{0\times\mathcal{T}}$, and $\boldsymbol{e} \in \mathbb{R}^0$. Its target set is $\mathcal{F}_{\mathbf{S}}^{P'}$, which is nonempty by the construction in the proof of Theorem 3.1. Both additional residuals in (B.3) vanish, giving (B.6).

For the resulting pair,

$$\begin{aligned}|T(\boldsymbol{\pi}',\boldsymbol{q}') - T(\boldsymbol{\pi},\boldsymbol{q})| &\le \|\boldsymbol{b}\|_\infty\|\boldsymbol{\pi}'-\boldsymbol{\pi}\|_1 \\ &\quad + \sum_{d\in\mathcal{D}} \sup_{y\in\mathcal{Y}} \|\boldsymbol{g}_d(y)\|_\infty \int_{\mathcal{Y}} \|\boldsymbol{q}'_d - \boldsymbol{q}_d\|_1 \, d\mu \\ &\le L_T\|(\boldsymbol{\pi}',\boldsymbol{q}') - (\boldsymbol{\pi},\boldsymbol{q})\|_{\mathcal{F},1} \le CL_T\Delta(P',P).\end{aligned}$$

In particular,

$$\begin{aligned}\inf_{(\boldsymbol{\pi}',\boldsymbol{q}')\in\mathcal{F}_{\mathbf{S}}^{P'}} T(\boldsymbol{\pi}',\boldsymbol{q}') &\le \inf_{(\boldsymbol{\pi},\boldsymbol{q})\in\mathcal{F}_{\mathbf{S}}^{P}} T(\boldsymbol{\pi},\boldsymbol{q}) + CL_T\Delta(P',P), \\ \sup_{(\boldsymbol{\pi},\boldsymbol{q})\in\mathcal{F}_{\mathbf{S}}^{P}} T(\boldsymbol{\pi},\boldsymbol{q}) &\le \sup_{(\boldsymbol{\pi}',\boldsymbol{q}')\in\mathcal{F}_{\mathbf{S}}^{P'}} T(\boldsymbol{\pi}',\boldsymbol{q}') + CL_T\Delta(P',P).\end{aligned}$$

Interchanging P and P' gives the two asserted absolute-value bounds. ■

Corollary B.1. *Suppose* $\mathbf{S}$ *satisfies Condition CP. Fix* $\boldsymbol{b}_N, \boldsymbol{b}_D \in \mathbb{R}^{\mathcal{T}}$ *and bounded measurable functions* $\boldsymbol{g}_{N,d}, \boldsymbol{g}_{D,d} : \mathcal{Y} \to \mathbb{R}^{\mathcal{T}}$, $d \in \mathcal{D}$. *Define*

$$\mathsf{N}(\boldsymbol{\pi},\boldsymbol{q}) := \boldsymbol{b}_N^\intercal\boldsymbol{\pi} + \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} \boldsymbol{g}_{N,d}^\intercal \boldsymbol{q}_d \, d\mu,$$

$$\mathsf{D}(\boldsymbol{\pi},\boldsymbol{q}) := \boldsymbol{b}_D^\intercal\boldsymbol{\pi} + \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} \boldsymbol{g}_{D,d}^\intercal \boldsymbol{q}_d \, d\mu,$$

$$D_*(R) := \inf_{(\boldsymbol{\pi},\boldsymbol{q})\in\mathcal{F}_{\mathbf{S}}^{R}} \mathsf{D}(\boldsymbol{\pi},\boldsymbol{q}).$$

For laws with $D_*(R) > 0$, *define*

$$\theta_-(R) := \inf_{(\boldsymbol{\pi},\boldsymbol{q})\in\mathcal{F}_{\mathbf{S}}^{R}} \frac{\mathsf{N}(\boldsymbol{\pi},\boldsymbol{q})}{\mathsf{D}(\boldsymbol{\pi},\boldsymbol{q})}, \qquad \theta_+(R) := \sup_{(\boldsymbol{\pi},\boldsymbol{q})\in\mathcal{F}_{\mathbf{S}}^{R}} \frac{\mathsf{N}(\boldsymbol{\pi},\boldsymbol{q})}{\mathsf{D}(\boldsymbol{\pi},\boldsymbol{q})}.$$

Fix P *satisfying Condition D with* $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$ *and* $D_*(P) > 0$, *and define*

$$\mathcal{U}_\delta(P) := \left\{ R \;\middle|\; \begin{array}{l} R \text{ satisfies Condition D}, \\ \mathcal{Q}_{\mathbf{S}}(R) \neq \emptyset, \quad \Delta(R,P) < \delta \end{array} \right\}.$$

There exist $\delta, a > 0$ and finite C_D, C_θ such that, for every $R, S \in \mathcal{U}_\delta(P)$,

$$D_*(R) \geq a,$$
$$|D_*(R) - D_*(S)| \leq C_D \Delta(R, S),$$
$$\max_{s \in \{-,+\}} |\theta_s(R) - \theta_s(S)| \leq C_\theta \Delta(R, S).$$

In the conditional-mean case

$$\mathsf{N}(\boldsymbol{\pi}, \boldsymbol{q}) = \int_{\mathcal{Y}} \kappa \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{q}_{d^\star} \, d\mu, \qquad \mathsf{D}(\boldsymbol{\pi}, \boldsymbol{q}) = \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi},$$

with $d^\star \in \mathcal{D}$, $\emptyset \neq A \subseteq \boldsymbol{\mathcal{T}}_{d^\star}$, and bounded measurable $\kappa : \mathcal{Y} \to \mathbb{R}$, both endpoints are attained at every $R \in \mathcal{U}_\delta(P)$.

Proof. With C from Lemma B.2, set

$$L_r := C \max \left\{ \|\boldsymbol{b}_r\|_\infty, \max_{d \in \mathcal{D}} \sup_{y \in \mathcal{Y}} \|\boldsymbol{g}_{r,d}(y)\|_\infty \right\} \qquad \forall r \in \{N, D\}.$$

By Lemma B.2,

$$|D_*(R) - D_*(S)| \leq L_D \Delta(R, S).$$

Take

$$a := \frac{D_*(P)}{2}, \qquad \delta := \frac{a}{1 + L_D}.$$

For $R \in \mathcal{U}_\delta(P)$,

$$\begin{aligned} D_*(R) &\geq D_*(P) - L_D \Delta(R, P) \\ &\geq D_*(P) - L_D \delta \geq a. \end{aligned}$$

Every feasible tuple satisfies $\|\boldsymbol{\pi}\|_1 = \int_{\mathcal{Y}} \|\boldsymbol{q}_d\|_1 \, d\mu = 1$ for every d, and so

$$|\mathsf{N}(\boldsymbol{\pi}, \boldsymbol{q})| \leq \|\boldsymbol{b}_N\|_\infty + \sum_{d \in \mathcal{D}} \sup_{y \in \mathcal{Y}} \|\boldsymbol{g}_{N,d}(y)\|_\infty =: B.$$

For $R, S \in \mathcal{U}_\delta(P)$ and $(\boldsymbol{\pi}, \boldsymbol{q}) \in \mathcal{F}_{\mathbf{S}}^R$, choose $(\boldsymbol{\pi}', \boldsymbol{q}') \in \mathcal{F}_{\mathbf{S}}^S$ using Lemma B.2, and let

$$N := \mathsf{N}(\boldsymbol{\pi}, \boldsymbol{q}), \qquad D := \mathsf{D}(\boldsymbol{\pi}, \boldsymbol{q}), \qquad N' := \mathsf{N}(\boldsymbol{\pi}', \boldsymbol{q}'), \qquad D' := \mathsf{D}(\boldsymbol{\pi}', \boldsymbol{q}').$$

Then

$$\begin{aligned} |N' - N| &\leq L_N \Delta(R, S), \\ |D' - D| &\leq L_D \Delta(R, S), \qquad D, D' \geq a, \end{aligned}$$

and

$$\frac{N'}{D'} - \frac{N}{D} = \frac{N' - N}{D'} + \frac{N(D - D')}{DD'},$$

$$\left|\frac{N'}{D'}-\frac{N}{D}\right| \le \left(\frac{L_N}{a}+\frac{BL_D}{a^2}\right)\Delta(R,S).$$

Taking infima or suprema over $(\boldsymbol{\pi},\boldsymbol{q})$ gives

$$\theta_-(S) \le \theta_-(R) + C_\theta \Delta(R,S),$$
$$\theta_+(R) \le \theta_+(S) + C_\theta \Delta(R,S),$$

where $C_\theta := L_N/a + BL_D/a^2$. Interchanging R and S gives the asserted absolute-value bounds, with $C_D = L_D$.

In the conditional-mean case, Lemma 3.1 and the construction in the proof of Theorem 3.1 give

$$\left\{(\mathsf{N}(\boldsymbol{\pi},\boldsymbol{q}),\mathsf{D}(\boldsymbol{\pi},\boldsymbol{q})) \,\middle|\, (\boldsymbol{\pi},\boldsymbol{q}) \in \mathcal{F}_{\mathbf{S}}^{R}\right\} = K_R,$$

where K_R is the set in Lemma A.10. That lemma makes K_R compact, and $(N,D) \mapsto N/D$ is continuous on K_R because $D \ge a$. Both extrema are therefore attained. ∎

Lemma B.3. *Suppose θ is the conditional-mean parameter (4.1), with $\emptyset \neq A \subseteq \mathcal{T}_{d^\star}$, and Condition T3 holds. Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and Assumption 5.1 holds.*

There exist bounded measurable $\boldsymbol{\ell}_d : \mathcal{Y} \to \mathbb{R}^{\mathcal{Z}}_{\ge 0}$, vectors $\boldsymbol{v}_d \in \mathbb{R}^{\mathcal{T}}$, and $\boldsymbol{\lambda} \ge \mathbf{0}$ such that

$$\sum_{d\in\mathcal{D}} \boldsymbol{v}_d = \mathbf{A}_{\mathbf{S}}^{\intercal}\boldsymbol{\lambda},$$
$$\boldsymbol{r}_d(y) := \mathbf{R}_d^{\intercal}\boldsymbol{\ell}_d(y) - \mathbb{1}\{d = d^\star\}(\kappa(y) - \bar\theta(P))\boldsymbol{c}_A + \boldsymbol{v}_d \ge \mathbf{0},$$

and every $(\boldsymbol{\pi},\boldsymbol{q}) \in \mathcal{F}_{\mathbf{S}}^{P}$ satisfies

$$\int_{\mathcal{Y}} (\bar\theta(P) - \kappa)\boldsymbol{c}_A^{\intercal}\boldsymbol{q}_{d^\star}\, d\mu = \boldsymbol{\lambda}^{\intercal}(\mathbf{a}_{\mathbf{S}} - \mathbf{A}_{\mathbf{S}}\boldsymbol{\pi}) + \sum_d \int_{\mathcal{Y}} \boldsymbol{r}_d^{\intercal}\boldsymbol{q}_d\, d\mu \ge 0. \tag{B.7}$$

The functions can be chosen using the matrices $\boldsymbol{B}_{d,J}$ and finite sets $\mathscr{J}_d$ in Lemma A.7, so that

$$\boldsymbol{r}_{d;(\boldsymbol{\tau})}(y) \in \left\{\alpha_{d,\boldsymbol{\tau},J}\kappa(y) + \beta_{d,\boldsymbol{\tau},J} \,\middle|\, J \in \mathscr{J}_d\right\},$$

where, with $\boldsymbol{I}$ the identity on $\mathbb{R}^{\mathcal{T}}$,

$$\alpha_{d,\boldsymbol{\tau},J} := \mathbb{1}\{d = d^\star\}\left((\mathbf{R}_d^{\intercal}\boldsymbol{B}_{d,J} - \boldsymbol{I})\boldsymbol{c}_A\right)_{(\boldsymbol{\tau})},$$
$$\beta_{d,\boldsymbol{\tau},J} := -\left((\mathbf{R}_d^{\intercal}\boldsymbol{B}_{d,J} - \boldsymbol{I})\left[\mathbb{1}\{d = d^\star\}\bar\theta\boldsymbol{c}_A + \boldsymbol{v}_d\right]\right)_{(\boldsymbol{\tau})}.$$

In particular, $\alpha_{d,\boldsymbol{\tau},J} = 0$ whenever $d \neq d^\star$.

Proof. By Lemma A.12, choose an optimal $(\boldsymbol{u}, \boldsymbol{\lambda})$ at $\gamma = \bar{\theta}(P)$ with $\boldsymbol{u}_{d^\star} = \mathbf{0}$. Set

$$\boldsymbol{v}_{d^\star} := \mathbf{A}_{\mathbf{S}}^{\intercal}\boldsymbol{\lambda} + \sum_{d \neq d^\star} \boldsymbol{u}_d, \qquad \boldsymbol{v}_d := -\boldsymbol{u}_d \qquad \forall d \neq d^\star.$$

Choose the bounded basic dual solutions from Lemma A.7 at coefficients

$$\mathbb{1}\{d = d^\star\}(\kappa(y) - \bar{\theta}(P))\boldsymbol{c}_A - \boldsymbol{v}_d.$$

Their feasibility gives $\boldsymbol{r}_d \geq \mathbf{0}$, and their objective equalities give

$$\sum_d \int_{\mathcal{Y}} \boldsymbol{\ell}_d^{\intercal} \boldsymbol{p}_d^P \, d\mu + \boldsymbol{\lambda}^{\intercal}\mathbf{a}_{\mathbf{S}} = 0.$$

For every feasible pair, therefore,

$$\begin{aligned}\sum_d \int_{\mathcal{Y}} \boldsymbol{r}_d^{\intercal}\boldsymbol{q}_d \, d\mu &= \sum_d \int_{\mathcal{Y}} \boldsymbol{\ell}_d^{\intercal}\boldsymbol{p}_d^P \, d\mu - \int_{\mathcal{Y}} (\kappa - \bar{\theta}(P))\boldsymbol{c}_A^{\intercal}\boldsymbol{q}_{d^\star} \, d\mu + \sum_d \boldsymbol{v}_d^{\intercal}\boldsymbol{\pi} \\ &= -\boldsymbol{\lambda}^{\intercal}\mathbf{a}_{\mathbf{S}} - \int_{\mathcal{Y}} (\kappa - \bar{\theta}(P))\boldsymbol{c}_A^{\intercal}\boldsymbol{q}_{d^\star} \, d\mu + \boldsymbol{\lambda}^{\intercal}\mathbf{A}_{\mathbf{S}}\boldsymbol{\pi}.\end{aligned}$$

Rearranging proves (B.7); every term on its right is nonnegative.

Finally, the result now follows by noting that, wherever the selected basis is J, we have

$$\boldsymbol{r}_d(y) = (\mathbf{R}_d^{\intercal}\boldsymbol{B}_{d,J} - \boldsymbol{I})\left[\mathbb{1}\{d = d^\star\}(\kappa(y) - \bar{\theta}(P))\boldsymbol{c}_A - \boldsymbol{v}_d\right]. \qquad \blacksquare$$

Lemma B.4. *Maintain the assumptions and fixed choices of Lemma B.3, and let*

$$r_d(y) := \sum_{z \in \mathcal{Z}} p_{d|z}^P(y).$$

There exists $C_P < \infty$ such that, for every $t > 0$ and $(\boldsymbol{\pi}, \boldsymbol{q}) \in \mathcal{F}_{\mathbf{S}}^P$, there is $(\boldsymbol{\pi}_*, \boldsymbol{q}_*) \in \mathcal{A}_\theta(P)$ satisfying

$$\begin{aligned}\|(\boldsymbol{\pi}, \boldsymbol{q}) - (\boldsymbol{\pi}_*, \boldsymbol{q}_*)\|_{\mathcal{F},1} \leq C_P \Bigg[&\left(1 + \frac{1}{t}\right)\int_{\mathcal{Y}} (\bar{\theta}(P) - \kappa)\boldsymbol{c}_A^{\intercal}\boldsymbol{q}_{d^\star} \, d\mu \\ &+ \sum_{d \in \mathcal{D}} \sum_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d} \int_{\mathcal{Y}} r_d(y)\mathbb{1}\{0 < \boldsymbol{r}_{d;(\boldsymbol{\tau})}(y) \leq t\} \, d\mu(y)\Bigg].\end{aligned}$$

Proof. Fix a feasible pair and set

$$\epsilon := \int_{\mathcal{Y}} (\bar{\theta}(P) - \kappa)\boldsymbol{c}_A^{\intercal}\boldsymbol{q}_{d^\star} \, d\mu \geq 0.$$

For $\boldsymbol{\tau} \notin \boldsymbol{\mathcal{T}}_d$, the corresponding response column is $\mathbf{R}_{d;(\cdot,\boldsymbol{\tau})} = \mathbf{0}$ and the target coefficient is zero, and so $\boldsymbol{r}_{d;(\boldsymbol{\tau})}(y) = \boldsymbol{v}_{d;(\boldsymbol{\tau})} \geq 0$. Define

$$I_0 := \left\{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \,\middle|\, \exists d \in \mathcal{D} : \boldsymbol{\tau} \notin \boldsymbol{\mathcal{T}}_d,\ \boldsymbol{v}_{d;(\boldsymbol{\tau})} > 0\right\}.$$

For each $\boldsymbol{\tau} \in I_0$, choose one such $d(\boldsymbol{\tau})$. By (B.7),

$$\sum_{\boldsymbol{\tau} \in I_0} \boldsymbol{\pi}_{(\boldsymbol{\tau})} \leq \epsilon \sum_{\boldsymbol{\tau} \in I_0} \frac{1}{\boldsymbol{v}_{d(\boldsymbol{\tau});(\boldsymbol{\tau})}}.$$

Let $\boldsymbol{C}_d(y)$ be diagonal, with

$$\boldsymbol{C}_{d;(\boldsymbol{\tau},\boldsymbol{\tau})}(y) := \mathbb{1}\{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d,\ \boldsymbol{r}_{d;(\boldsymbol{\tau})}(y) > 0\}.$$

For active coordinates, $q_{d|\boldsymbol{\tau}} \leq r_d$, and so

$$\sum_d \int_{\mathcal{Y}} \|\boldsymbol{C}_d \boldsymbol{q}_d^a\|_1 \, d\mu \leq \frac{\epsilon}{t} + \sum_d \sum_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d} \int_{\mathcal{Y}} r_d \mathbb{1}\{0 < \boldsymbol{r}_{d;(\boldsymbol{\tau})} \leq t\} \, d\mu.$$

Here the contribution from $\{\boldsymbol{r}_{d;(\boldsymbol{\tau})} > t\}$ is bounded using $q_{d|\boldsymbol{\tau}} \leq \boldsymbol{r}_{d;(\boldsymbol{\tau})} q_{d|\boldsymbol{\tau}}/t$ and (B.7).

With $\boldsymbol{I}$ the identity on $\mathbb{R}^{\boldsymbol{\mathcal{T}}}$, set

$$\boldsymbol{E} := \begin{bmatrix} \boldsymbol{I}_{(I_0,\cdot)} \\ \boldsymbol{\lambda}^{\intercal} \mathbf{A}_{\mathbf{S}} \end{bmatrix}, \qquad \boldsymbol{e} := \begin{bmatrix} \mathbf{0}_{I_0} \\ \boldsymbol{\lambda}^{\intercal} \mathbf{a}_{\mathbf{S}} \end{bmatrix}.$$

The preceding mass bound and (B.7) imply

$$\|\boldsymbol{E}\boldsymbol{\pi} - \boldsymbol{e}\|_1 \leq \left(1 + \sum_{\boldsymbol{\tau} \in I_0} \frac{1}{\boldsymbol{v}_{d(\boldsymbol{\tau});(\boldsymbol{\tau})}}\right) \epsilon.$$

Every term in (B.7) is nonnegative. Its left-hand side is therefore zero exactly when

$$\boldsymbol{C}_d \boldsymbol{q}_d^a = \mathbf{0} \quad \forall d, \qquad\qquad \boldsymbol{E}\boldsymbol{\pi} = \boldsymbol{e}.$$

Thus $\mathcal{F}^P_{\mathbf{S};\boldsymbol{C},\boldsymbol{E},\boldsymbol{e}} = \mathcal{A}_\theta(P)$, which is nonempty by Corollary B.1. Apply Lemma B.1 with $P' = P$ and the displayed $\boldsymbol{C}, \boldsymbol{E}, \boldsymbol{e}$. Its term $\Delta(P', P)$ is zero, and substitution of the two bounds above proves the claim. ■

Lemma B.5. *Maintain the assumptions and basic dual choices of Lemma B.3, and suppose Assumption 5.3 holds. Let $r_d := \sum_z p^P_{d|z}$. There exist $t_P > 0$ and $C_P < \infty$ such that*

$$\sum_{d \in \mathcal{D}} \sum_{\boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d} \int_{\mathcal{Y}} r_d(y) \mathbb{1}\{0 < \boldsymbol{r}_{d;(\boldsymbol{\tau})}(y) \leq t\} \, d\mu(y) \leq C_P t \qquad \forall t \in (0, t_P). \tag{B.8}$$

Proof. Use the finite coefficients $\alpha_{d,\boldsymbol{\tau},J}, \beta_{d,\boldsymbol{\tau},J}$ from Lemma B.3. For every $d, \boldsymbol{\tau}$,

$$\left\{y \,\middle|\, 0 < \boldsymbol{r}_{d;(\boldsymbol{\tau})}(y) \leq t\right\} \subseteq \bigcup_{J \in \mathscr{J}_d} \left\{y \,\middle|\, 0 < \alpha_{d,\boldsymbol{\tau},J} \kappa(y) + \beta_{d,\boldsymbol{\tau},J} \leq t\right\}.$$

Let F_κ be the finite atomic support and f_κ the bounded density of the absolutely continuous part of the conditional law of $\kappa(Y)$ given $D = d^\star$, as supplied by Assumption 5.3.

There are finitely many positive values among

$$\{\beta_{d,\boldsymbol{\tau},J} \,|\, d \in \mathcal{D},\ \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}}_d,\ J \in \mathscr{J}_d,\ \alpha_{d,\boldsymbol{\tau},J} = 0\},$$

$$\{\alpha_{d^\star,\boldsymbol{\tau},J}x+\beta_{d^\star,\boldsymbol{\tau},J}\,|\,\boldsymbol{\tau}\in\boldsymbol{\mathcal{T}}_{d^\star},\ J\in\mathscr{J}_{d^\star},\ x\in F_\kappa\}\,.$$

Choose $t_P>0$ smaller than all those positive values. Then, for $0<t<t_P$, no positive constant term and no positive value at an atom can belong to $(0,t]$. For $\alpha\neq 0$, the interval $\{x\in\mathbb{R}\,|\,0<\alpha x+\beta\le t\}$ has length $t/|\alpha|$, and so

$$\int_{\mathbb{R}}\mathbb{1}\{0<\alpha x+\beta\le t\}f_\kappa(x)\,dx\le\frac{\|f_\kappa\|_\infty}{|\alpha|}t.$$

Set $p_{\min}:=\min_z P_Z(z)>0$. For every Borel $B\subseteq\mathbb{R}$,

$$\begin{aligned}\int_{\mathcal{Y}} r_{d^\star}(y)\mathbb{1}\{\kappa(y)\in B\}\,d\mu(y)&\le\frac{1}{p_{\min}}\sum_z P_Z(z)\int_{\mathcal{Y}}p^P_{d^\star|z}(y)\mathbb{1}\{\kappa(y)\in B\}\,d\mu(y)\\&=\frac{P(D=d^\star)}{p_{\min}}P(\kappa(Y)\in B\,|\,D=d^\star).\end{aligned}$$

All slopes with $d\neq d^\star$ are zero. Combining the finite union above with these bounds proves (B.8), for example with

$$C_P:=\frac{\|f_\kappa\|_\infty}{p_{\min}}\sum_{\boldsymbol{\tau}\in\boldsymbol{\mathcal{T}}_{d^\star}}\sum_{\substack{J\in\mathscr{J}_{d^\star}\\ \alpha_{d^\star,\boldsymbol{\tau},J}\neq 0}}\frac{1}{|\alpha_{d^\star,\boldsymbol{\tau},J}|}.$$

■

Lemma B.6. *Maintain the assumptions of Lemma B.5. There exist $\epsilon_P,\delta_P>0$ and $C_P<\infty$ such that the following statements hold.*

For every $(\boldsymbol{\pi},\boldsymbol{q})\in\mathcal{F}^P_{\mathbf{S}}$ with

$$0\le\int_{\mathcal{Y}}(\bar\theta(P)-\kappa)\boldsymbol{c}_A^{\intercal}\boldsymbol{q}_{d^\star}\,d\mu<\epsilon_P,$$

there is $(\boldsymbol{\pi}_*,\boldsymbol{q}_*)\in\mathcal{A}_\theta(P)$ with

$$\|(\boldsymbol{\pi},\boldsymbol{q})-(\boldsymbol{\pi}_*,\boldsymbol{q}_*)\|_{\mathcal{F},1}\le C_P\left[\int_{\mathcal{Y}}(\bar\theta(P)-\kappa)\boldsymbol{c}_A^{\intercal}\boldsymbol{q}_{d^\star}\,d\mu\right]^{1/2}.$$

Further, for every $R\in\mathcal{U}_{\delta_P}(P)$ and $(\boldsymbol{\pi}',\boldsymbol{q}')\in\mathcal{A}_\theta(R)$,

$$\operatorname{dist}\left(\boldsymbol{c}_A^{\intercal}\boldsymbol{\pi}',[\underline{p}^\star_A(P),\overline{p}^\star_A(P)]\right)\le C_P\sqrt{\Delta(R,P)}.$$

Proof. For a feasible pair, let

$$\epsilon:=\int_{\mathcal{Y}}(\bar\theta(P)-\kappa)\boldsymbol{c}_A^{\intercal}\boldsymbol{q}_{d^\star}\,d\mu.$$

Let C_4 be the constant from Lemma B.4, and let C_5,t_P be those from Lemma B.5. If $0<\epsilon<\min\{1,t_P^2\}$, apply these results with $t=\sqrt{\epsilon}$ to obtain an attaining pair satisfying

$$\|(\boldsymbol{\pi},\boldsymbol{q})-(\boldsymbol{\pi}_*,\boldsymbol{q}_*)\|_{\mathcal{F},1}\le C_4\left[\epsilon+\frac{\epsilon}{\sqrt{\epsilon}}+C_5\sqrt{\epsilon}\right]$$

$$\leq C_4(2 + C_5)\sqrt{\epsilon}.$$

If $\epsilon = 0$, the original pair is in $\mathcal{A}_\theta(P)$.

Now fix $(\boldsymbol{\pi}', \boldsymbol{q}') \in \mathcal{A}_\theta(R)$. By Lemma B.2, choose $(\boldsymbol{\pi}, \boldsymbol{q}) \in \mathcal{F}_{\mathbf{S}}^P$ with

$$\|(\boldsymbol{\pi}, \boldsymbol{q}) - (\boldsymbol{\pi}', \boldsymbol{q}')\|_{\mathcal{F},1} \leq C\Delta(R, P).$$

Since the primed pair attains $\bar{\theta}(R)$,

$$\epsilon = [\bar{\theta}(P) - \bar{\theta}(R)]\boldsymbol{c}_A^\intercal \boldsymbol{\pi}' + \bar{\theta}(P)\boldsymbol{c}_A^\intercal(\boldsymbol{\pi} - \boldsymbol{\pi}')$$
$$- \int_{\mathcal{Y}} \kappa \boldsymbol{c}_A^\intercal(\boldsymbol{q}_{d^\star} - \boldsymbol{q}'_{d^\star})\, d\mu.$$

By Corollary B.1, and since $\boldsymbol{c}_A^\intercal \boldsymbol{\pi}' \leq 1$, there is $C' < \infty$ such that

$$0 \leq \epsilon \leq C'\Delta(R, P)$$

throughout a sufficiently small neighbourhood of P. Choose $\delta_P \leq 1$ small enough that this upper bound is less than ϵ_P.

The first claim now supplies $(\boldsymbol{\pi}_*, \boldsymbol{q}_*) \in \mathcal{A}_\theta(P)$ with

$$|\boldsymbol{c}_A^\intercal(\boldsymbol{\pi}' - \boldsymbol{\pi}_*)| \leq \|\boldsymbol{c}_A\|_\infty \left(C\Delta(R, P) + C_P\sqrt{C'\Delta(R, P)}\right).$$

By Lemma A.10, the conditioning probabilities of pairs in $\mathcal{A}_\theta(P)$ form exactly $[\underline{p}_A^\star(P), \overline{p}_A^\star(P)]$. Taking the distance to this interval and enlarging C_P proves the second claim. ■

Let $I_1, \ldots, I_K$ be independent folds, as defined in Section 5, with fixed K and $n^{(k)}/n \xrightarrow{P} c_k \in (0, 1)$. Let $\mathcal{F}_{n,k}$ be the sigma-algebra generated by the observations outside I_k

Lemma B.7. *Consider the conditional-mean parameter (4.1), with bounded κ and $\emptyset \neq A \subseteq \mathcal{T}_{d^\star}$. Suppose $\mathbf{S}$ satisfies Condition CP, and both P and R satisfy Condition D and Assumption 5.1. For each $S \in \{P, R\}$, let*

$$(\boldsymbol{\pi}_S, \boldsymbol{q}_S) \in \mathcal{A}_\theta(S), \qquad m_S \in \mathfrak{M}_S^\star, \qquad a_S := \boldsymbol{c}_A^\intercal \boldsymbol{\pi}_S.$$

Then

$$\frac{(\bar{\theta}(R) - \bar{\theta}(P))(a_R - a_P)}{a_R} \leq \bar{\theta}(R) - \bar{\theta}(P) + \frac{1}{a_R}\sum_{z \in \mathcal{Z}} \mathbb{E}_P[m_R(W) \mid Z = z]$$
$$\leq \frac{1}{a_R}\sum_{d \in \mathcal{D}, z \in \mathcal{Z}} \int_{\mathcal{Y}} [p_{d|z}^P(y) - p_{d|z}^R(y)][m_R(y, d, z) - m_P(y, d, z)]\, d\mu(y).$$

Proof. Let $\boldsymbol{q}_T = (\boldsymbol{q}_{d,T})_{d \in \mathcal{D}}$. For $S, T \in \{P, R\}$, take $(\boldsymbol{u}_S, \boldsymbol{\lambda}_S, \boldsymbol{\ell}_{0,S}, (\boldsymbol{\ell}_{d,S})_{d \in \mathcal{D}})$ satisfying the defining conditions for m_S. Those conditions and $(\boldsymbol{\pi}_T, \boldsymbol{q}_T) \in \mathcal{A}_\theta(T)$ give

$$\sum_z \mathbb{E}_T[m_S(W) \mid Z = z] = \int_{\mathcal{Y}} \boldsymbol{\ell}_{0,S}^\intercal \boldsymbol{p}_{d^\star}^T\, d\mu + \sum_d \int_{\mathcal{Y}} \boldsymbol{\ell}_{d,S}^\intercal \boldsymbol{p}_d^T\, d\mu + \boldsymbol{\lambda}_S^\intercal \mathbf{a}_\mathbf{S}$$

$$\geq \int_{\mathcal{Y}} (\kappa - \bar{\theta}(S)) \mathbf{c}_A^\intercal \mathbf{q}_{d^\star,T}\, d\mu + \boldsymbol{\lambda}_S^\intercal (\mathbf{a_S} - \mathbf{A_S}\boldsymbol{\pi}_T)$$
$$\geq (\bar{\theta}(T) - \bar{\theta}(S)) a_T.$$

Also, $\sum_z \mathbb{E}_S[m_S(W) \mid Z = z] = 0$, so

$$\sum_{d,z} \int_{\mathcal{Y}} [p^P_{d|z}(y) - p^R_{d|z}(y)][m_R(y,d,z) - m_P(y,d,z)]\, d\mu(y)$$
$$= \sum_z \mathbb{E}_P[m_R(W) \mid Z = z] + \sum_z \mathbb{E}_R[m_P(W) \mid Z = z]$$
$$\geq \sum_z \mathbb{E}_P[m_R(W) \mid Z = z] + (\bar{\theta}(R) - \bar{\theta}(P)) a_R.$$

Dividing the last inequality by $a_R > 0$ gives the upper bound. The first display with $(S,T) = (R,P)$ gives the lower bound. ■

Lemma B.8. *For every k, let $\eta_{n,k}$ and $a_{n,k}$ be $\mathcal{F}_{n,k}$-measurable, and let $m_{n,k} : \Omega \times \mathcal{W} \to \mathbb{R}$ be jointly $\mathcal{F}_{n,k} \otimes \mathscr{B}_{\mathcal{W}}$-measurable. Suppose that, for fixed $a > 0$ and $M < \infty$,*

$$\min_{k \in [K]} a_{n,k} \geq a, \qquad \max_{k \in [K]} \|m_{n,k}\|_\infty \leq M$$

with probability tending to one. For a fixed $\vartheta \in \mathbb{R}$, define

$$\widehat{\vartheta}_n := \sum_{k=1}^K \frac{n^{(k)}}{n} \left[\eta_{n,k} + \frac{1}{a_{n,k}} \sum_{z \in \mathcal{Z}} \frac{1}{n_z^{(k)}} \sum_{\substack{i \in I_k \\ Z_i = z}} m_{n,k}(W_i) \right],$$
$$b_{n,k} := \eta_{n,k} - \vartheta + \frac{1}{a_{n,k}} \sum_{z \in \mathcal{Z}} \mathbb{E}_P[m_{n,k}(W) \mid Z = z],$$
$$\phi_{n,k}(W) := \frac{m_{n,k}(W) - \mathbb{E}_P[m_{n,k}(W) \mid Z]}{P_Z(Z) a_{n,k}}.$$

Then

$$\widehat{\vartheta}_n - \vartheta = \sum_{k=1}^K \frac{n^{(k)}}{n} b_{n,k} + \frac{1}{n} \sum_{k=1}^K \sum_{i \in I_k} \phi_{n,k}(W_i) + O_P(n^{-1}), \tag{B.9}$$

$$\frac{1}{n} \sum_{k=1}^K \sum_{i \in I_k} \phi_{n,k}(W_i) = O_P(n^{-1/2}). \tag{B.10}$$

Further, suppose $\xi_{n,k} : \Omega \times \mathcal{W} \to \mathbb{R}$ is jointly $\mathcal{F}_{n,k} \otimes \mathscr{B}_{\mathcal{W}}$-measurable, $\mathbb{E}_P[\xi_{n,k}(W)] = 0$, and $\max_k \|\xi_{n,k}\|_{L^2(P)} = o_P(1)$. Then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^K \sum_{i \in I_k} \xi_{n,k}(W_i) = o_P(1). \tag{B.11}$$

Proof. The stipulated bounds and $n_z^{(k)}/n^{(k)} \geq P_Z(z)/2$ for every k, z hold w.p.a.1, and so we assume them throughout.

Set

$$\mu_{n,k,z} := \mathbb{E}_P[m_{n,k}(W) \mid Z = z],$$
$$A_{n,k,z} := \frac{1}{n^{(k)}} \sum_{i \in I_k} \mathbb{1}\{Z_i = z\} \left[m_{n,k}(W_i) - \mu_{n,k,z}\right].$$

Conditional on $\mathcal{F}_{n,k}$, these summands are independent and have mean zero. On training outcomes satisfying $\|m_{n,k}\|_\infty \leq M$,

$$\mathbb{E}[A_{n,k,z}^2 \mid \mathcal{F}_{n,k}] \leq \frac{4M^2}{n^{(k)}}.$$

Therefore $A_{n,k,z} = O_P(n^{-1/2})$. Also, $n_z^{(k)}/n^{(k)} - P_Z(z) = O_P(n^{-1/2})$, and so

$$\begin{aligned}
&\frac{1}{n_z^{(k)}} \sum_{\substack{i \in I_k \\ Z_i = z}} m_{n,k}(W_i) - \mu_{n,k,z} \\
&= \frac{A_{n,k,z}}{n_z^{(k)}/n^{(k)}} \\
&= \frac{A_{n,k,z}}{P_Z(z)} + A_{n,k,z} \frac{P_Z(z) - n_z^{(k)}/n^{(k)}}{P_Z(z) n_z^{(k)}/n^{(k)}} \\
&= \frac{A_{n,k,z}}{P_Z(z)} + O_P(n^{-1}).
\end{aligned}$$

Summing over z, dividing by $a_{n,k} \geq a$, and weighting by $n^{(k)}/n$ gives (B.9). The conditional variance bound also gives (B.10), since K is fixed.

For the last claim, set

$$V_{n,k} := \frac{1}{\sqrt{n}} \sum_{i \in I_k} \xi_{n,k}(W_i).$$

Conditional independence and conditional mean zero give

$$\mathbb{E}[V_{n,k}^2 \mid \mathcal{F}_{n,k}] = \frac{n^{(k)}}{n} \|\xi_{n,k}\|_{L^2(P)}^2.$$

For every $\epsilon > 0$, conditional Chebyshev therefore gives

$$\Pr(|V_{n,k}| > \epsilon) \leq \mathbb{E}\left[\min\left\{1, \frac{n^{(k)}}{n\epsilon^2} \|\xi_{n,k}\|_{L^2(P)}^2\right\}\right] \to 0.$$

Summing over the fixed number of folds proves (B.11). ■

Lemma B.9. *Consider the conditional-mean parameter in* (4.1)*, with bounded* κ*. Suppose* P *satisfies Condition D,* **S** *satisfies Condition CP, and Assumptions 5.1 and 5.5 hold.*

For every $\epsilon > 0$, there exists $\delta > 0$ such that, for every $P' \in \mathcal{U}_\delta(P)$, with $\mathcal{U}_\delta(P)$ as defined in Corollary B.1, all feasible conditioning probabilities are positive, and every $(\boldsymbol{\pi}', \boldsymbol{q}') \in \mathcal{F}_{\mathbf{S}}^{P'}$ satisfying

$$\frac{\int_{\mathcal{Y}} \kappa(y) \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{q}'_{d^\star}(y)\, d\mu(y)}{\boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}'} = \bar{\theta}(P')$$

also satisfies

$$\left| \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}' - p_A^\dagger(P) \right| < \epsilon.$$

In particular, if measurable random laws P'_n and measurable pairs $(\boldsymbol{\pi}'_n, \boldsymbol{q}'_n)$ satisfy these compatibility and attainment requirements with probability tending to one, and $\Delta(P'_n, P) \xrightarrow{P} 0$, then

$$\boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}'_n \xrightarrow{P} p_A^\dagger(P).$$

Proof. The positivity assertion follows from Corollary B.1. Suppose the uniform assertion fails. Then there exist $\epsilon > 0$, laws $P'_n \in \mathcal{U}_{1/n}(P)$, and attaining pairs $(\boldsymbol{\pi}'_n, \boldsymbol{q}'_n) \in \mathcal{F}_{\mathbf{S}}^{P'_n}$ such that

$$\left| \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}'_n - p_A^\dagger(P) \right| \geq \epsilon \qquad \forall n.$$

By Lemma B.2, choose $(\boldsymbol{\pi}_n, \boldsymbol{q}_n) \in \mathcal{F}_{\mathbf{S}}^{P}$ with

$$\left\| (\boldsymbol{\pi}_n, \boldsymbol{q}_n) - (\boldsymbol{\pi}'_n, \boldsymbol{q}'_n) \right\|_{\mathcal{F},1} \leq C \Delta(P'_n, P) \to 0.$$

Define

$$a_n := \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}_n, \qquad a'_n := \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{\pi}'_n,$$
$$\mathsf{N}_n := \int_{\mathcal{Y}} \kappa(y) \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{q}_{d^\star,n}(y)\, d\mu(y), \qquad \mathsf{N}'_n := \int_{\mathcal{Y}} \kappa(y) \boldsymbol{c}_A^{\mathsf{T}} \boldsymbol{q}'_{d^\star,n}(y)\, d\mu(y),$$

where $\boldsymbol{q}_{d,n}$ and $\boldsymbol{q}'_{d,n}$ are the treatment-d components of $\boldsymbol{q}_n$ and $\boldsymbol{q}'_n$, respectively. Boundedness of κ gives

$$|a_n - a'_n| + |\mathsf{N}_n - \mathsf{N}'_n| \leq (1 + \|\kappa\|_\infty) \|\boldsymbol{c}_A\|_\infty \left\| (\boldsymbol{\pi}_n, \boldsymbol{q}_n) - (\boldsymbol{\pi}'_n, \boldsymbol{q}'_n) \right\|_{\mathcal{F},1} \to 0.$$

Since $\mathsf{N}'_n = \bar{\theta}(P'_n) a'_n$, $0 \leq a'_n \leq 1$, and $\bar{\theta}(P'_n) \to \bar{\theta}(P)$ by Corollary B.1,

$$\left| \mathsf{N}_n - \bar{\theta}(P) a_n \right| \leq |\mathsf{N}_n - \mathsf{N}'_n| + |\bar{\theta}(P'_n) - \bar{\theta}(P)| a'_n + |\bar{\theta}(P)|\, |a'_n - a_n| \to 0.$$

The construction in Theorem 3.1 and Lemma A.10 give $(\mathsf{N}_n, a_n) \in K_P$, where K_P is compact. Choose increasing indices n_k such that

$$(\mathsf{N}_{n_k}, a_{n_k}) \to (\mathsf{N}, a) \in K_P.$$

Then $\mathsf{N} = \bar{\theta}(P) a$, and so Assumption 5.5 gives $a = p_A^\dagger(P)$. Since $a'_{n_k} - a_{n_k} \to 0$, this contradicts $|a'_{n_k} - p_A^\dagger(P)| \geq \epsilon$.

Finally, the uniform assertion gives, for each $\epsilon > 0$ and its corresponding $\delta > 0$,

$$\mathbb{P}\left(\left|\boldsymbol{c}_A^{\intercal}\boldsymbol{\pi}'_n - p_A^{\dagger}(P)\right| \geq \epsilon\right) \leq \mathbb{P}\left(\Delta(P'_n, P) \geq \delta\right) + o(1) \to 0.$$

■

Lemma B.10. *Suppose θ is the conditional-mean parameter (4.1), with $\emptyset \neq A \subseteq \mathcal{T}_{d^\star}$, and Condition T3 holds. Suppose also that P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and Assumptions 5.1 and 5.6 hold. Fix $m_* \in \mathfrak{M}_P^\star$.*

For every $\epsilon > 0$, there exists $\delta > 0$ such that every $R \in \mathcal{U}_\delta(P)$ and $m \in \mathfrak{M}_R^\star$ satisfy

$$\|m - \mathbb{E}_P[m(W) \mid Z] - m_* + \mathbb{E}_P[m_*(W) \mid Z]\|_{L^2(P)} < \epsilon.$$

In particular, for jointly measurable random selections $m_n \in \mathfrak{M}_{P'_n}^\star$ with $\Delta(P'_n, P) \xrightarrow{P} 0$,

$$\|m_n - \mathbb{E}_P[m_n(W) \mid Z] - m_* + \mathbb{E}_P[m_*(W) \mid Z]\|_{L^2(P)} \xrightarrow{P} 0.$$

Proof. Suppose the uniform assertion fails. Then there exist $\epsilon > 0$, laws $R_n \in \mathcal{U}_{1/n}(P)$, and $m_n \in \mathfrak{M}_{R_n}^\star$ for which the displayed norm is at least ϵ.

Let $\gamma_n := \bar{\theta}(R_n)$ and $\gamma := \bar{\theta}(P)$. Choose witnesses $(\boldsymbol{u}_n, \boldsymbol{\lambda}_n, \boldsymbol{\ell}_{n,0}, (\boldsymbol{\ell}_{n,d})_{d\in\mathcal{D}})$ for m_n. By compactness of the bounded outer domain, there exist $1 \leq n_1 < n_2 < \cdots$ and $(\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{D}_\theta$ such that

$$\|\boldsymbol{u}_{n_k} - \boldsymbol{u}\|_\infty + \|\boldsymbol{\lambda}_{n_k} - \boldsymbol{\lambda}\|_\infty \to 0.$$

Also, Corollary B.1 gives $\gamma_n \to \gamma$.

Enumerate $\mathcal{D} = \{d_1, \ldots, d_J\}$ and also denote $d_0 := d^\star$. For $i \in [J]$, use $\boldsymbol{\ell}_{n,i} := \boldsymbol{\ell}_{n,d_i}$, and define

$$\boldsymbol{c}_{n,0}(y) := (\kappa(y) - \gamma_n)\boldsymbol{c}_A - \mathbf{A}_{\mathbf{S}}^{\intercal}\boldsymbol{\lambda}_n - \sum_d \boldsymbol{u}_{n,d}, \qquad \boldsymbol{c}_{n,i}(y) := \boldsymbol{u}_{n,d_i},$$

$$\boldsymbol{c}_0(y) := (\kappa(y) - \gamma)\boldsymbol{c}_A - \mathbf{A}_{\mathbf{S}}^{\intercal}\boldsymbol{\lambda} - \sum_d \boldsymbol{u}_d, \qquad \boldsymbol{c}_i(y) := \boldsymbol{u}_{d_i}.$$

In particular,

$$\sup_y \|\boldsymbol{c}_{n_k,0}(y) - \boldsymbol{c}_0(y)\|_\infty \leq |\gamma_{n_k} - \gamma|\|\boldsymbol{c}_A\|_\infty + \|\mathbf{A}_{\mathbf{S}}^{\intercal}(\boldsymbol{\lambda}_{n_k} - \boldsymbol{\lambda})\|_\infty + \sum_d \|\boldsymbol{u}_{n_k,d} - \boldsymbol{u}_d\|_\infty \to 0,$$

$$\max_{i\in[J]} \sup_y \|\boldsymbol{c}_{n_k,i}(y) - \boldsymbol{c}_i(y)\|_\infty \to 0.$$

By Corollary A.4, $\|m_n\|_\infty \leq C_m$, and so

$$\left|\sum_z \mathbb{E}_P[m_n(W) \mid Z = z]\right| = \left|\sum_{d,z} \int_{\mathcal{Y}} m_n (p_{d|z}^P - p_{d|z}^{R_n})\, d\mu\right|$$

$$\leq C_m \Delta(R_n, P) \to 0.$$

Inner dual feasibility gives

$$\overline{\Psi}(\gamma_n; P) \leq \mathscr{J}_P(\gamma_n, \boldsymbol{u}_n, \boldsymbol{\lambda}_n) \leq \sum_z \mathbb{E}_P[m_n(W) \mid Z = z].$$

Since conditioning probabilities belong to $[0, 1]$,

$$|\overline{\Psi}(\gamma_n; P)| \leq |\gamma_n - \gamma| \to 0.$$

Therefore $\mathscr{J}_P(\gamma_n, \boldsymbol{u}_n, \boldsymbol{\lambda}_n) \to 0$.

The integrated-support inequality in Lemma A.2 gives, here,

$$\begin{aligned} &|\mathscr{J}_P(\gamma_{n_k}, \boldsymbol{u}_{n_k}, \boldsymbol{\lambda}_{n_k}) - \mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda})| \\ &\leq \sum_{i=0}^{J} \int_{\mathcal{Y}} \|\boldsymbol{p}_{d_i}^P(y)\|_1 \|\boldsymbol{c}_{n_k,i}(y) - \boldsymbol{c}_i(y)\|_\infty \, d\mu(y) \\ &\quad + \|\mathbf{a}_\mathbf{S}\|_1 \|\boldsymbol{\lambda}_{n_k} - \boldsymbol{\lambda}\|_\infty \to 0. \end{aligned}$$

Thus $\mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}) = 0$.

For $i \in \{0, \ldots, J\}$, set

$$g_{n,i}(y) := \boldsymbol{p}_{d_i}^P(y)^\intercal \boldsymbol{\ell}_{n,i}(y) - h_{d_i}(\boldsymbol{c}_{n,i}(y); \boldsymbol{p}_{d_i}^P(y)) \geq 0.$$

Expanding the two objectives gives

$$\sum_{i=0}^{J} \int_{\mathcal{Y}} g_{n,i} \, d\mu = \sum_z \mathbb{E}_P[m_n(W) \mid Z = z] - \mathscr{J}_P(\gamma_n, \boldsymbol{u}_n, \boldsymbol{\lambda}_n) \to 0.$$

Choose $1 \leq k_1 < k_2 < \cdots$ such that

$$\sum_{i=0}^{J} \int_{\mathcal{Y}} g_{n_{k_j},i} \, d\mu \leq 2^{-j}.$$

By monotone convergence,

$$\int_{\mathcal{Y}} \sum_{j=1}^{\infty} \sum_{i=0}^{J} g_{n_{k_j},i}(y) \, d\mu(y) \leq \sum_{j=1}^{\infty} 2^{-j} = 1,$$

and so $g_{n_{k_j},i}(y) \to 0$ for every i, outside one μ-null set.

Define

$$\mathcal{L}_i(y) := \left\{ \boldsymbol{\ell} \in \mathbb{R}_{\geq 0}^{\mathcal{Z}} \;\middle|\; \begin{array}{l} \mathbf{R}_{d_i}^\intercal \boldsymbol{\ell} \geq \boldsymbol{c}_i(y), \\ \|\boldsymbol{\ell}\|_\infty \leq \|\boldsymbol{c}_i(y)\|_\infty, \\ \boldsymbol{p}_{d_i}^P(y)^\intercal \boldsymbol{\ell} = h_{d_i}(\boldsymbol{c}_i(y); \boldsymbol{p}_{d_i}^P(y)) \end{array} \right\}.$$

These sets are nonempty by Lemma 3.2, and are compact and convex. Fix y outside the null set. Every convergent subsequence $\boldsymbol{\ell}_{n_{k_{j_r}},i}(y) \to \boldsymbol{\ell}$ has $\boldsymbol{\ell} \in \mathcal{L}_i(y)$: the feasibility and norm inequalities pass to the limit, and $g_{n_{k_{j_r}},i}(y) \to 0$ gives the objective equality. The vectors are

uniformly bounded, and so

$$\operatorname{dist}_2\left(\boldsymbol{\ell}_{n_{k_j},i}(y), \mathcal{L}_i(y)\right) \to 0.$$

The constraints defining $\mathcal{L}_i(y)$ are measurable in y and continuous in $\boldsymbol{\ell}$. By Theorems 14.36-14.37 in Rockafellar and Wets (1998), the unique Euclidean projections

$$\boldsymbol{\ell}^{\circ}_{j,i}(y) := \operatorname*{argmin}_{\boldsymbol{\ell}\in\mathcal{L}_i(y)} \|\boldsymbol{\ell} - \boldsymbol{\ell}_{n_{k_j},i}(y)\|_2^2$$

are measurable. Define

$$m^{\circ}_j(y,d,z) := \mathbb{1}\{d = d^{\star}\}\boldsymbol{\ell}^{\circ}_{j,0;(z)}(y) + \sum_{i=1}^{J} \mathbb{1}\{d = d_i\}\boldsymbol{\ell}^{\circ}_{j,i;(z)}(y) + \frac{\boldsymbol{\lambda}^{\intercal}\mathbf{a}_{\mathbf{S}}}{|\mathcal{Z}|}.$$

Its inner objective equalities and $\mathscr{J}_P(\gamma, \boldsymbol{u}, \boldsymbol{\lambda}) = 0$ give $m^{\circ}_j \in \mathfrak{M}^{\star}_P$. Pointwise convergence of the projection distances, convergence of $\boldsymbol{\lambda}_{n_{k_j}}$, and the common bounded envelope give

$$\|m_{n_{k_j}} - m^{\circ}_j\|_{L^2(P)} \to 0$$

by dominated convergence.

By Assumption 5.6, $m^{\circ}_j - \mathbb{E}_P[m^{\circ}_j(W) \mid Z] = m_* - \mathbb{E}_P[m_*(W) \mid Z]$ in $L^2(P)$. Conditional centring is an $L^2(P)$ contraction, and so

$$\begin{aligned}
&\left\| m_{n_{k_j}} - \mathbb{E}_P[m_{n_{k_j}}(W) \mid Z] - m_* + \mathbb{E}_P[m_*(W) \mid Z] \right\|_{L^2(P)} \\
&\leq \|m_{n_{k_j}} - m^{\circ}_j\|_{L^2(P)} \to 0,
\end{aligned}$$

contradicting the choice of m_n. This proves the uniform assertion, which immediately gives the random-sequence assertion because $\Pr(\Delta(P'_n, P) \geq \delta) \to 0$. ■

Lemma B.11. *Let ν be the product measure on $\mathcal{Y} \times \mathcal{D} \times \mathcal{Z}$ produced by μ on $\mathcal{Y}$ and the counting measure on $\mathcal{D} \times \mathcal{Z}$. Let $(P_t)_t$ be a two-sided QMD path through P in $\mathfrak{P}$ with score $s \in L_0^2(P)$. Let $\bar{s}_z := \mathbb{E}_P[s(W) \mid Z = z]$, $h_s(y,d,z) = s(y,d,z) - \bar{s}_z$, and $p^{P_t}_{d|z}(y) = r_t(y,d,z)/P_{t;Z}(z)$ for sufficiently small $|t|$. Then, as $t \to 0$,*

$$\begin{aligned}
\|r_t - r_0 - t r_0 s\|_{L^1(\nu)} &= o(|t|)\ , \\
\sum_{z\in\mathcal{Z}} |P_{t;Z}(z) - P_Z(z) - t P_Z(z)\,\bar{s}_z| &= o(|t|)\ , \\
\sum_{d\in\mathcal{D}} \|p^{P_t}_{d|z} - p^{P}_{d|z} - t p^{P}_{d|z} h_s(\cdot,d,z)\|_{L^1(\mu)} &= o(|t|)\ ,
\end{aligned}$$

and

$$\sum_{d,z} P_Z(z) \left\| \frac{\sqrt{p^{P_t}_{d|z}} - \sqrt{p^{P}_{d|z}}}{t} - \frac{1}{2} h_s(\cdot,d,z)\sqrt{p^{P}_{d|z}} \right\|^2_{L^2(\mu)} \to 0\ .$$

Finally, the following maps, defined for $u \in L^2(P)$, are linear and continuous over $L_0^2(P)$:

$$\begin{aligned}
\dot{r}_0(u) &:= r_0 u \,, \\
\dot{\boldsymbol{p}}^P(u) &:= \left(p^P_{d|z} h_u(\cdot, d, z)\right)_{d,z} \,, \\
\dot{P}_Z(u) &:= \left(P_Z(z)\, \mathbb{E}_P[u(W) \,|\, Z = z]\right)_z \,, \\
\dot{u}_P(u) &:= \left(\frac{1}{2} h_u(\cdot, d, z) \sqrt{p^P_{d|z}}\right)_{d,z} \,,
\end{aligned}$$

where $\dot{u}_P$ has the finite product space $\prod_{d\in\mathcal{D}, z\in\mathcal{Z}} L^2(\mu)$ endowed with the norm $\|v\| = \|(v_{d,z})_{d,z}\| := (\sum_{d\in\mathcal{D}} \sum_{z\in\mathcal{Z}} P_Z(z) \|v_{d,z}\|^2_{L^2(\mu)})^{1/2}$ as its codomain.

Proof. Let $r := r_0$. Define

$$e_t := \sqrt{r_t} - \sqrt{r} - \frac{t}{2} s \sqrt{r},$$

so that $\frac{\|e_t\|_{L^2(\nu)}}{|t|} \to 0$ and

$$\|\sqrt{r_t} - \sqrt{r}\|_{L^2(\nu)} \leq \frac{|t|}{2} \|s\|_{L^2(P)} + \|e_t\|_{L^2(\nu)} = O(|t|) \,,$$

with $\|\sqrt{r_t}\|_{L^2(\nu)} = \|\sqrt{r}\|_{L^2(\nu)} = 1$.

Let $A_t = e_t(\sqrt{r_t} + \sqrt{r})$, $B_t = (t/2) s \sqrt{r}(\sqrt{r_t} - \sqrt{r})$ and $R_t = r_t - r - tsr$, and note that

$$\begin{aligned}
R_t &= (\sqrt{r_t} - \sqrt{r})(\sqrt{r_t} + \sqrt{r}) - tsr \\
&= e_t(\sqrt{r_t} + \sqrt{r}) + \frac{t}{2} s \sqrt{r}(\sqrt{r_t} + \sqrt{r} - 2\sqrt{r}) = A_t + B_t.
\end{aligned}$$

By Cauchy-Schwarz,

$$\begin{aligned}
\|A_t\|_{L^1(\nu)} &\leq \|e_t\|_{L^2(\nu)} \|\sqrt{r_t} + \sqrt{r}\|_{L^2(\nu)} \leq 2\|e_t\|_{L^2(\nu)} = o(|t|), \\
\|B_t\|_{L^1(\nu)} &\leq \frac{|t|}{2} \|s\|_{L^2(P)} \|\sqrt{r_t} - \sqrt{r}\|_{L^2(\nu)} \\
&\leq \frac{t^2}{4} \|s\|^2_{L^2(P)} + \frac{|t|}{2} \|s\|_{L^2(P)} \|e_t\|_{L^2(\nu)} = O(t^2) \,,
\end{aligned}$$

and so,

$$\|R_t\|_{L^1(\nu)} \leq \|A_t\|_{L^1(\nu)} + \|B_t\|_{L^1(\nu)} = o(|t|) \,,$$

as required.

Now, for each z, integrating $r_t - r = tsr + R_t$ over $y \in \mathcal{Y}$ and summing over $d \in \mathcal{D}$ gives

$$\begin{aligned}
P_{t;Z}(z) - P_Z(z) &= t \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} s(y, d, z) r(y, d, z) \, d\mu(y) + \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} R_t(y, d, z) \, d\mu(y) \\
&= t P_Z(z) \, \bar{s}_z + \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} R_t(y, d, z) \, d\mu(y)
\end{aligned}$$

$$= O\left(|t|\right) \ ,$$

and so, $P_{t;Z}\left(z\right) > P_Z\left(z\right)/2$ for sufficiently small t, since

$$\left|\sum_{d\in\mathcal{D}}\int_{\mathcal{Y}} R_t(y,d,z)\,d\mu(y)\right| \le \sum_{d\in\mathcal{D}}\int_{\mathcal{Y}} |R_t(y,d,z)|\,d\mu(y) \le \|R_t\|_{L^1(\nu)} = o(|t|) \ ,$$

which immediately gives the second claim since

$$\begin{aligned}\sum_{z\in\mathcal{Z}} |P_{t;Z}(z) - P_Z(z) - tP_Z(z)\bar{s}_z| &= \sum_{z\in\mathcal{Z}}\left|\sum_{d\in\mathcal{D}}\int_{\mathcal{Y}} R_t(y,d,z)\,d\mu(y)\right| \\ &\le \|R_t\|_{L^1(\nu)}\end{aligned}$$

Next, fix (d,z), and define the shorthands $p := p^P_{d|z}$, $p_t := p^{P_t}_{d|z}$, $s_{d,z}(y) := s(y,d,z)$, $\dot{p} := p(s_{d,z} - \bar{s}_z)$ and $p_{\min} := \min_{z\in\mathcal{Z}} P_Z(z) > 0$, so that

$$\begin{aligned}P_{t;Z}(z)(p_t - p - t\dot{p}) &= r_t(\cdot,d,z) - P_{t;Z}(z)p - tP_{t;Z}(z)\dot{p} \\ &= tP_Z(z)s_{d,z}p + R_t(\cdot,d,z) - \left(tP_Z(z)\bar{s}_z + \sum_{d\in\mathcal{D}}\int_{\mathcal{Y}} R_t(y,d,z)\,d\mu(y)\right)p - tP_{t;Z}(z)\dot{p} \\ &= R_t(\cdot,d,z) - p\sum_{d\in\mathcal{D}}\int_{\mathcal{Y}} R_t(y,d,z)\,d\mu(y) - t\dot{p}(P_{t;Z}(z) - P_Z(z)),\end{aligned}$$

Note that $\|p\|_{L^1(\mu)} = P(D = d \mid Z = z) \le 1$ and

$$\begin{aligned}\|h_s\|^2_{L^2(P)} &= \mathbb{E}_P s^2 - \mathbb{E}_P\{\mathbb{E}_P[s \mid Z]^2\} \le \|s\|^2_{L^2(P)} \ , \\ \sum_{\substack{d'\in\mathcal{D}\\ z'\in\mathcal{Z}}} \|p^P_{d'|z'} h_s(\cdot,d',z')\|_{L^1(\mu)} &= \sum_{\substack{d'\in\mathcal{D}\\ z'\in\mathcal{Z}}} \frac{1}{P_Z(z')}\int_{\mathcal{Y}} |h_s(y,d',z')| r(y,d',z')\,d\mu(y) \\ &\le \frac{1}{p_{\min}}\|h_s\|_{L^1(P)} \le \frac{1}{p_{\min}}\|h_s\|_{L^2(P)} \\ &\le \frac{1}{p_{\min}}\|s\|_{L^2(P)} < \infty \ ,\end{aligned} \tag{B.12}$$

so, $\|\dot{p}\|_{L^1(\mu)} < \infty$ and we have

$$\begin{aligned}&\|p_t - p - t\dot{p}\|_{L^1(\mu)} \\ &\le \frac{2}{P_Z(z)}\left\{\|R_t(\cdot,d,z)\|_{L^1(\mu)} + \left|\sum_{d\in\mathcal{D}}\int_{\mathcal{Y}} R_t(y,d,z)\,d\mu(y)\right| \|p\|_{L^1(\mu)} + |t|\ |P_{t;Z}(z) - P_Z(z)|\,\|\dot{p}\|_{L^1(\mu)}\right\} \\ &\le \frac{2}{P_Z(z)}\{o(|t|) + o(|t|) + O(t^2)\} = o(|t|),\end{aligned}$$

and, for each $z \in \mathcal{Z}$, summing over $d \in \mathcal{D}$ gives the third claim.

For $z \in \mathcal{Z}$ let $a_{t,z} = P_{t;Z}(z)^{-1/2}$ and $a_z = P_Z(z)^{-1/2}$. For $f(x) = x^{-1/2}$, $f'(P_Z(z)) = -\frac{1}{2}P_Z(z)^{-3/2}$ and $\sup_{x\in[P_Z(z)/2,3P_Z(z)/2]} |f''(x)| < \infty$. We now have the expansion,

$$\begin{aligned}\alpha_{t,z} &:= a_{t,z} - a_z + \frac{t}{2} a_z \bar{s}_z \\ &= -\tfrac{1}{2} P_Z(z)^{-3/2} \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} R_t(y,d,z)\, d\mu(y) + \zeta_{t,z},\end{aligned}$$

with Taylor remainder $|\zeta_{t,z}| \le \frac{1}{2}\sup_{x\in[P_Z(z)/2,3P_Z(z)/2]} |f''(x)|\,|P_{t;Z}(z) - P_Z(z)|^2$ and so, $\alpha_{t,z} = o(|t|)$, $a_{t,z} - a_z = O(|t|)$, and $a_{t,z} \le (P_Z(z)/2)^{-1/2}$.

Next, for each (d,z), substituting $\sqrt{p_t} = a_{t,z}\sqrt{r_t(\cdot,d,z)}$ into the QMD expansion gives

$$\begin{aligned}&\sqrt{p^{P_t}_{d|z}} - \sqrt{p^{P}_{d|z}} - \frac{t}{2} h_s(\cdot,d,z)\sqrt{p^{P}_{d|z}} \\ &= a_{t,z} e_t(\cdot,d,z) + \frac{t}{2}(a_{t,z} - a_z) s(\cdot,d,z)\sqrt{r(\cdot,d,z)} + \alpha_{t,z}\sqrt{r(\cdot,d,z)},\end{aligned}$$

where these three terms have the following $L^2(\mu)$ bounds:

$$\begin{aligned}\|a_{t,z} e_t(\cdot,d,z)\|_2 &\le (P_Z(z)/2)^{-1/2}\|e_t\|_{L^2(\nu)} = o(|t|), \\ \left\|\tfrac{t}{2}(a_{t,z} - a_z) s(\cdot,d,z)\sqrt{r(\cdot,d,z)}\right\|_2 &\le \tfrac{|t|}{2}|a_{t,z} - a_z|\|s\|_{L^2(P)} = O(t^2), \\ \|\alpha_{t,z}\sqrt{r(\cdot,d,z)}\|_2 &= |\alpha_{t,z}|\sqrt{P(D=d, Z=z)} = o(|t|).\end{aligned}$$

Applying triangle inequality, dividing by $|t|$, squaring, and taking the finite sum over (d,z) gives the fourth claim.

Finally, linearity of the four maps follows immediately from linearity of expectation and so, it suffices to show that they are bounded (hence continuous), which we get as follows,

$$\begin{aligned}\|\dot{r}_0(u)\|_{L^1(\nu)} &= \int_{\mathcal{W}} |u(w)| r_0(w)\, d\nu(w) = \mathbb{E}_P[|u|] \le \|u\|_{L^2(P)}, \\ \|\dot{P}_Z(u)\|_1 = \sum_{z\in\mathcal{Z}} P_Z(z)\,|\mathbb{E}_P[u \mid Z=z]| &\le \sum_{z\in\mathcal{Z}} P_Z(z)\,\mathbb{E}_P[|u| \mid Z=z] \le \|u\|_{L^2(P)}, \\ \sum_{d\in\mathcal{D},z\in\mathcal{Z}} \|p^P_{d|z} h_u(\cdot,d,z)\|_{L^1(\mu)} &\le \frac{1}{p_{\min}}\|u\|_{L^2(P)}, \\ \|\dot{u}_P(u)\| = \tfrac{1}{2}\|h_u\|_{L^2(P)} &\le \tfrac{1}{2}\|u\|_{L^2(P)},\end{aligned}$$

where the last two lines follow from applying (B.12) with $s = u$. ∎

Lemma B.12. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. Define*

$$S_d(y) := \left\{ \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \,\middle|\, \sup_{\boldsymbol{x}\in\mathscr{F}^P_d(y)} \boldsymbol{x}_{(\boldsymbol{\tau})} > 0 \right\}.$$

Then $\{y : \boldsymbol{\tau} \in S_d(y)\}$ *is measurable for every* $d, \boldsymbol{\tau}$*, and every* $(\boldsymbol{\pi}, \boldsymbol{q}) \in \mathcal{F}_{\mathbf{S}}^P$ *satisfies*

$$\boldsymbol{q}_d(y) \in \mathscr{F}_d^P(y), \qquad q_{d|\boldsymbol{\tau}}(y) = 0 \qquad \forall \boldsymbol{\tau} \notin S_d(y).$$

Moreover, there exists $(\boldsymbol{\pi}^\dagger, \boldsymbol{q}^\dagger) \in \mathcal{F}_{\mathbf{S}}^P$ *such that*

$$\boldsymbol{\pi}^\dagger \in \operatorname{relint} \boldsymbol{\Pi}(P),$$
$$q^\dagger_{d|\boldsymbol{\tau}}(y) > 0 \quad \Longleftrightarrow \quad \boldsymbol{\tau} \in S_d(y),$$
$$\left\{ j \in [m_{\mathbf{S}}] \,\middle|\, \mathbf{A}_{\mathbf{S};(j,\cdot)} \boldsymbol{\pi}^\dagger = \mathbf{a}_{\mathbf{S};(j)} \right\} = J_{\mathbf{S}}^P.$$

Proof. Choose $\boldsymbol{\pi}^\dagger \in \operatorname{relint} \boldsymbol{\Pi}(P)$, which is possible by Theorem 3.1 and lemma 3.3, and define

$$Z^0 := \left\{ \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \,\middle|\, \max_{\boldsymbol{\pi} \in \boldsymbol{\Pi}(P)} \boldsymbol{\pi}_{(\boldsymbol{\tau})} = 0 \right\},$$
$$C_d^\dagger := \left\{ \boldsymbol{u} \in \boldsymbol{\Pi}_d \,\middle|\, \begin{array}{l} \mathbf{A}_{d,J_d^P} \boldsymbol{u} = \displaystyle\int_{\mathcal{Y}} \left(h_d\left(\boldsymbol{a}_{d,j}; \boldsymbol{p}_d^P(y)\right) \right)_{j \in J_d^P} d\mu(y), \\ \boldsymbol{u}_{(Z^0)} = \mathbf{0} \end{array} \right\}.$$

A valid linear inequality binds at a relative-interior point of a convex set if and only if it binds throughout that set. Therefore, by the finite inequality representation in Lemma 3.3, $C_d^\dagger$ is the minimal face of $\boldsymbol{\Pi}_d$ containing $\boldsymbol{\pi}^\dagger$, and $\boldsymbol{\pi}^\dagger \in \operatorname{relint} C_d^\dagger$. The same observation applied to $\mathbf{A}_{\mathbf{S}} \boldsymbol{\pi} \leq \mathbf{a}_{\mathbf{S}}$ gives the asserted equality of binding row sets.

Every response inequality $\mathbf{R}_d \boldsymbol{u} \leq \int_{\mathcal{Y}} \boldsymbol{p}_d^P \, d\mu$ is valid on $\boldsymbol{\Pi}_d$ and binds at $\boldsymbol{\pi}^\dagger$ by Lemma 3.1. It therefore binds throughout $C_d^\dagger$. If an integrable selection $\boldsymbol{x}(y) \in \operatorname{Fibre}_d(\boldsymbol{p}_d^P(y))$ has integral in $C_d^\dagger$, then

$$\int_{\mathcal{Y}} \left(\boldsymbol{p}_d^P - \mathbf{R}_d \boldsymbol{x} \right) d\mu = \mathbf{0},$$
$$\int_{\mathcal{Y}} \left[h_d\left(\boldsymbol{a}_{d,j}; \boldsymbol{p}_d^P(y)\right) - \boldsymbol{a}_{d,j}^\intercal \boldsymbol{x}(y) \right] d\mu(y) = 0 \qquad \forall j \in J_d^P,$$
$$\int_{\mathcal{Y}} \boldsymbol{x}_{(Z^0)} \, d\mu = \mathbf{0}.$$

All these integrands are nonnegative, and so their coordinates vanish almost everywhere. These are exactly the equations defining $\mathscr{F}_d^P(y)$. The reverse implication follows by integrating those equations, and so, by the definition of Minkowski integral,

$$C_d^\dagger = \int_{\mathcal{Y}} \mathscr{F}_d^P(y) \, d\mu(y).$$

Since $\boldsymbol{\Pi}(P) \subseteq C_d^\dagger$, this also proves the assertion about every feasible pair.

It remains to construct a selection that is positive on every permitted coordinate. Let $\boldsymbol{I}$ be the identity matrix on $\mathbb{R}^{\boldsymbol{\mathcal{T}}}$. The active-coordinate projection of $\mathscr{F}_d^P(y)$ is the nonempty

compact polytope

$$X_d(y) := \left\{ \boldsymbol{x} \in \mathbb{R}^{\mathcal{T}_d}_{\geq 0} \,\middle|\, \boldsymbol{H}_d \boldsymbol{x} = \boldsymbol{b}_d(y) \right\},$$

$$\boldsymbol{H}_d := \begin{bmatrix} \mathbf{R}_{d;(\cdot,\mathcal{T}_d)} \\ \mathbf{A}_{d;(J_d^P,\mathcal{T}_d)} \\ \boldsymbol{I}_{(Z^0,\mathcal{T}_d)} \end{bmatrix}, \qquad \boldsymbol{b}_d(y) := \begin{bmatrix} \boldsymbol{p}_d^P(y) \\ \left(h_d(\boldsymbol{a}_{d,j}; \boldsymbol{p}_d^P(y))\right)_{j \in J_d^P} \\ \mathbf{0}_{Z^0} \end{bmatrix}.$$

Its right-hand side is measurable by Lemma A.2. Choose independent rows I_d spanning the row space of $\boldsymbol{H}_d$, and enumerate the column sets $J \subseteq \mathcal{T}_d$ with $|J| = |I_d|$ and $\boldsymbol{H}_{d;(I_d,J)}$ nonsingular. For each such J, define the basic formula

$$\boldsymbol{x}_{d,J;(J)}(y) := \boldsymbol{H}^{-1}_{d;(I_d,J)} \boldsymbol{b}_{d;(I_d)}(y), \qquad \boldsymbol{x}_{d,J;(\mathcal{T}_d \setminus J)}(y) := \mathbf{0}.$$

Let $\mathcal{J}_d(y)$ consist of those J for which $\boldsymbol{x}_{d,J}(y) \geq \mathbf{0}$ and $\boldsymbol{H}_d \boldsymbol{x}_{d,J}(y) = \boldsymbol{b}_d(y)$, and set

$$\boldsymbol{x}^+_{d,\mathrm{act}}(y) := \frac{1}{|\mathcal{J}_d(y)|} \sum_{J \in \mathcal{J}_d(y)} \boldsymbol{x}_{d,J}(y).$$

There is at least one feasible basic solution. Each basic formula is a fixed linear map of $\boldsymbol{b}_d(y)$, and its feasibility is determined by finitely many measurable equalities and inequalities, and so this average is measurable.

Every vertex of $X_d(y)$ occurs among these formulas. Since $X_d(y)$ is their convex hull,

$$\boldsymbol{x}^+_{d,\mathrm{act};(\boldsymbol{\tau})}(y) > 0 \quad \Longleftrightarrow \quad \boldsymbol{\tau} \in S_d(y) \qquad \forall \boldsymbol{\tau} \in \mathcal{T}_d,$$

$$\|\boldsymbol{x}^+_{d,\mathrm{act}}(y)\|_1 \leq \mathbf{1}^{\intercal} \boldsymbol{p}_d^P(y).$$

The response and support rows are zero on $\mathcal{T} \setminus \mathcal{T}_d$, by the definition of $\mathbf{R}_d$ and the construction in Lemma 3.3. Therefore, using ρ_0 from Lemma A.1, complete this vector by

$$\boldsymbol{x}^+_{d;(\boldsymbol{\tau})}(y) := \rho_0(y) \mathbb{1}\{\boldsymbol{\tau} \notin Z^0\} \qquad \forall \boldsymbol{\tau} \notin \mathcal{T}_d.$$

Then $\boldsymbol{x}_d^+$ is integrable, belongs to $\mathscr{F}_d^P(y)$, and is positive exactly on $S_d(y)$. This also proves the asserted measurability of S_d.

Let $\boldsymbol{u}_d^+ := \int_{\mathcal{Y}} \boldsymbol{x}_d^+ \, d\mu$. Since $\boldsymbol{u}_d^+ \in C_d^\dagger$ and $\boldsymbol{\pi}^\dagger \in \operatorname{relint} C_d^\dagger$, choose $\epsilon_d \in (0,1)$ such that

$$\boldsymbol{u}_d^- := \frac{\boldsymbol{\pi}^\dagger - \epsilon_d \boldsymbol{u}_d^+}{1 - \epsilon_d} \in C_d^\dagger.$$

The Minkowski-integral equality above supplies $\boldsymbol{x}_d^-(y) \in \mathscr{F}_d^P(y)$ with $\int_{\mathcal{Y}} \boldsymbol{x}_d^- \, d\mu = \boldsymbol{u}_d^-$. Setting

$$\boldsymbol{q}_d^\dagger := \epsilon_d \boldsymbol{x}_d^+ + (1 - \epsilon_d) \boldsymbol{x}_d^- \,,$$

the result follows immediately by noting that

$$\int_{\mathcal{Y}} \boldsymbol{q}_d^\dagger \, d\mu = \boldsymbol{\pi}^\dagger, \qquad \mathbf{R}_d \boldsymbol{q}_d^\dagger = \boldsymbol{p}_d^P, \qquad q^\dagger_{d|\boldsymbol{\tau}}(y) > 0 \quad \Longleftrightarrow \quad \boldsymbol{\tau} \in S_d(y) \qquad \blacksquare$$

Lemma B.13. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. With S_d as defined in Lemma B.12, define*

$$\mathscr{E}_P := \left\{ \boldsymbol{u} \in \prod_{d\in\mathcal{D}} L^1(\mu;\mathbb{R}^{\boldsymbol{\mathcal{T}}}) \,\middle|\, \boldsymbol{u}_{d;(\boldsymbol{\tau})}(y) = 0 \;\forall d, y, \boldsymbol{\tau} \notin S_d(y) \right\},$$

$$\mathscr{U}_{P,1} := \left\{ (\mathbf{R}_d \boldsymbol{u}_d)_{d\in\mathcal{D}} \,\middle|\, \begin{array}{l} \boldsymbol{u} \in \mathscr{E}_P, \\ \exists \dot{\boldsymbol{\pi}} \in \ker \mathbf{A}_{\mathbf{S},J_{\mathbf{S}}^P} : \displaystyle\int_{\mathcal{Y}} \boldsymbol{u}_d \, d\mu = \dot{\boldsymbol{\pi}} \;\forall d \end{array} \right\},$$

$$\mathscr{U}_P := \mathscr{U}_{P,1} \cap \mathbb{H}_P.$$

Then $\mathscr{U}_{P,1}$ is a closed linear subspace of $\prod_{d\in\mathcal{D}} L^1(\mu;\mathbb{R}^{\mathcal{Z}})$, equipped with

$$\|\boldsymbol{b}\|_{L^1} := \sum_{d\in\mathcal{D}} \int_{\mathcal{Y}} \|\boldsymbol{b}_d(y)\|_1 \, d\mu(y),$$

and $\mathscr{U}_P$ is a closed linear subspace of $(\mathbb{H}_P, \|\cdot\|_P)$.

Proof. For $S \subseteq \boldsymbol{\mathcal{T}}$, let $B_{d,S} := \{y : S_d(y) = S\}$, which is measurable by Lemma B.12. Choose a fixed linear map $\boldsymbol{L}_{d,S} : \mathbb{R}^{\mathcal{Z}} \to \mathbb{R}^{\boldsymbol{\mathcal{T}}}$ such that

$$\begin{aligned} (\boldsymbol{L}_{d,S}\boldsymbol{b})_{(\boldsymbol{\mathcal{T}}\setminus S)} &= \mathbf{0} \qquad \forall \boldsymbol{b} \in \mathbb{R}^{\mathcal{Z}}, \\ \mathbf{R}_d \boldsymbol{L}_{d,S} \boldsymbol{b} &= \boldsymbol{b} \qquad \forall \boldsymbol{b} \in \operatorname{col} \mathbf{R}_{d;(\cdot,S)}. \end{aligned}$$

For example, choose row indices I and column indices $J \subseteq S$ such that $|I| = |J| = \operatorname{rank} \mathbf{R}_{d;(\cdot,S)}$ and $\mathbf{R}_{d;(I,J)}$ is nonsingular, and set

$$(\boldsymbol{L}_{d,S}\boldsymbol{b})_{(J)} := \mathbf{R}_{d;(I,J)}^{-1} \boldsymbol{b}_{(I)}, \qquad (\boldsymbol{L}_{d,S}\boldsymbol{b})_{(\boldsymbol{\mathcal{T}}\setminus J)} := \mathbf{0}.$$

For rank zero, take $\boldsymbol{L}_{d,S} = \mathbf{0}$.

For any $\boldsymbol{b} \in \prod_d L^1(\mu;\mathbb{R}^{\mathcal{Z}})$, define

$$\boldsymbol{u}_d^0(\boldsymbol{b})(y) := \sum_{S\subseteq\boldsymbol{\mathcal{T}}} \mathbb{1}_{B_{d,S}}(y) \boldsymbol{L}_{d,S} \boldsymbol{b}_d(y). \tag{B.13}$$

There are finitely many matrices $\boldsymbol{L}_{d,S}$, and so, with $C_L := \max_{d,S} \|\boldsymbol{L}_{d,S}\|_{1\leftarrow 1}$,

$$\sum_d \int_{\mathcal{Y}} \|\boldsymbol{u}_d^0(\boldsymbol{b}) - \boldsymbol{u}_d^0(\boldsymbol{c})\|_1 \, d\mu \leq C_L \|\boldsymbol{b} - \boldsymbol{c}\|_{L^1}.$$

Define the finite-dimensional subspaces

$$\mathscr{K}_d := \sum_{\substack{S\subseteq\boldsymbol{\mathcal{T}} \\ \mu(B_{d,S})>0}} \left\{ \boldsymbol{k} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}} \,\middle|\, \boldsymbol{k}_{(\boldsymbol{\mathcal{T}}\setminus S)} = \mathbf{0},\ \mathbf{R}_d \boldsymbol{k} = \mathbf{0} \right\},$$

$$V := \left\{ (\dot{\boldsymbol{\pi}} - \boldsymbol{k}_d)_{d\in\mathcal{D}} \,\middle|\, \dot{\boldsymbol{\pi}} \in \ker \mathbf{A}_{\mathbf{S},J_{\mathbf{S}}^P},\ \boldsymbol{k}_d \in \mathscr{K}_d \;\forall d \right\}.$$

The space $\mathscr{K}_d$ is exactly the set of integrals of integrable functions $\boldsymbol{k}_d(y)$ satisfying

$$\boldsymbol{k}_{d;(\boldsymbol{\tau}\setminus S_d(y))}(y) = \mathbf{0}, \qquad\qquad \mathbf{R}_d \boldsymbol{k}_d(y) = \mathbf{0}.$$

One inclusion follows by integrating over $B_{d,S}$. For the other, every vector $\boldsymbol{k}$ in the summand associated with $B_{d,S}$ is realized by

$$y \mapsto \boldsymbol{k}\, \frac{\rho_0(y)\mathbb{1}_{B_{d,S}}(y)}{\int_{B_{d,S}} \rho_0 \, d\mu},$$

using ρ_0 from Lemma A.1.

Set

$$T\boldsymbol{b} := \left(\int_{\mathcal{Y}} \boldsymbol{u}_d^0(\boldsymbol{b})\, d\mu \right)_{d\in\mathcal{D}}.$$

The preceding characterization of $\mathscr{K}_d$ gives

$$\boldsymbol{b} \in \mathscr{U}_{P,1} \quad \Longleftrightarrow \quad \begin{cases} \boldsymbol{b}_d = \mathbf{R}_d \boldsymbol{u}_d^0(\boldsymbol{b}) & \forall d, \\ T\boldsymbol{b} \in V. \end{cases}$$

Indeed, any witnessing $\boldsymbol{u}$ differs from $\boldsymbol{u}^0(\boldsymbol{b})$ by functions whose integrals belong to $\mathscr{K}_d$. Conversely, if $T\boldsymbol{b} = (\dot{\boldsymbol{\pi}} - \boldsymbol{k}_d)_d \in V$, add the kernel functions just constructed to $\boldsymbol{u}_d^0(\boldsymbol{b})$ to obtain common integral $\dot{\boldsymbol{\pi}}$.

The maps $\boldsymbol{b} \mapsto T\boldsymbol{b}$ and $\boldsymbol{b} \mapsto (\boldsymbol{b}_d - \mathbf{R}_d \boldsymbol{u}_d^0(\boldsymbol{b}))_d$ are continuous in $\|\cdot\|_{L^1}$ by the displayed bound, and V is a finite-dimensional linear subspace. Their inverse images in the last display are therefore closed, proving that $\mathscr{U}_{P,1}$ is closed.

Finally, Cauchy–Schwarz gives

$$\|\boldsymbol{b}\|_{L^1} \le \left(\sum_{z\in\mathcal{Z}} \frac{1}{P_Z(z)} \right)^{1/2} \|\boldsymbol{b}\|_P \qquad \forall \boldsymbol{b} \in \mathbb{H}_P,$$

and so $\mathscr{U}_P = \mathscr{U}_{P,1} \cap \mathbb{H}_P$ is closed in $\|\cdot\|_P$. ∎

Fix $d_0 \in \mathcal{D}$. For $J \subseteq [m_{\mathbf{S}}]$ and $\boldsymbol{u} \in \prod_{d\in\mathcal{D}} L^1(\mu;\mathbb{R}^{\boldsymbol{\tau}})$, define

$$\mathcal{A}_J(\boldsymbol{u}) := \begin{bmatrix} \left(\int_{\mathcal{Y}} (\boldsymbol{u}_d - \boldsymbol{u}_{d_0})\, d\mu \right)_{d\in\mathcal{D}\setminus\{d_0\}} \\ \mathbf{A}_{\mathbf{S},J} \int_{\mathcal{Y}} \boldsymbol{u}_{d_0}\, d\mu \end{bmatrix}. \tag{B.14}$$

Thus $\mathcal{A}_J(\boldsymbol{u}) = \mathbf{0}$ if and only if the integrals of $\boldsymbol{u}_d$ equal a common vector in $\ker \mathbf{A}_{\mathbf{S},J}$.

Lemma B.14. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and $(\boldsymbol{\pi}^0, \boldsymbol{q}^0) \in \mathcal{F}_{\mathbf{S}}^P$. Let $J \subseteq [m_{\mathbf{S}}]$ and $\dot{\boldsymbol{q}}_d \in L^1(\mu;\mathbb{R}^{\boldsymbol{\tau}})$ satisfy*

$$\dot{q}_{d|\boldsymbol{\tau}}(y) = 0 \qquad \forall d, \boldsymbol{\tau}, y \text{ with } q^0_{d|\boldsymbol{\tau}}(y) = 0,$$

$$\int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d \, d\mu = \dot{\boldsymbol{\pi}} \in \ker \mathbf{A}_{\mathbf{S},J} \qquad \forall d.$$

Suppose $\boldsymbol{b} := (\mathbf{R}_d \dot{\boldsymbol{q}}_d)_{d\in\mathcal{D}} \in \mathbb{H}_P$. *Then there exist* $\dot{\boldsymbol{q}}_d^{(n)} \in L^1(\mu;\mathbb{R}^{\boldsymbol{\mathcal{T}}})$, $\dot{\boldsymbol{\pi}}_n \in \ker \mathbf{A}_{\mathbf{S},J}$, *and* $C_n < \infty$ *such that*

$$\int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d^{(n)}\, d\mu = \dot{\boldsymbol{\pi}}_n \qquad \forall d, n,$$

$$|\dot{q}_{d|\boldsymbol{\tau}}^{(n)}(y)| \le C_n q_{d|\boldsymbol{\tau}}^0(y) \qquad \forall d, \boldsymbol{\tau}, y, n,$$

$$\sum_d \int_{\mathcal{Y}} \|\dot{\boldsymbol{q}}_d^{(n)} - \dot{\boldsymbol{q}}_d\|_1\, d\mu \to 0,$$

$$\left\| (\mathbf{R}_d \dot{\boldsymbol{q}}_d^{(n)})_d - \boldsymbol{b} \right\|_P \to 0.$$

Proof. Define

$$t_{d,n}(y) := \min\left(\{1\} \cup \left\{ \frac{n q_{d|\boldsymbol{\tau}}^0(y)}{|\dot{q}_{d|\boldsymbol{\tau}}(y)|} \,\middle|\, \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}},\ \dot{q}_{d|\boldsymbol{\tau}}(y) \neq 0 \right\} \right).$$

Then $0 < t_{d,n} \le 1$, $t_{d,n} \uparrow 1$, and $|t_{d,n}\dot{q}_{d|\boldsymbol{\tau}}| \le n q_{d|\boldsymbol{\tau}}^0$. Dominated convergence gives

$$\sum_d \int_{\mathcal{Y}} (1 - t_{d,n}) \|\dot{\boldsymbol{q}}_d\|_1\, d\mu \to 0,$$

$$\|((t_{d,n} - 1)\boldsymbol{b}_d)_d\|_P^2 = \sum_{d,z} P_Z(z) \int_{\{p_{d|z}^P > 0\}} (1 - t_{d,n})^2 \frac{b_{d|z}^2}{p_{d|z}^P}\, d\mu \to 0.$$

The respective dominators are $\sum_d \|\dot{\boldsymbol{q}}_d\|_1$ and the integrable coordinates defining $\|\boldsymbol{b}\|_P^2$.

Let

$$\mathscr{B}^0 := \left\{ \boldsymbol{u} \in \prod_d L^1(\mu;\mathbb{R}^{\boldsymbol{\mathcal{T}}}) \,\middle|\, \exists C < \infty : |\boldsymbol{u}_{d;(\boldsymbol{\tau})}(y)| \le C q_{d|\boldsymbol{\tau}}^0(y)\ \forall d, \boldsymbol{\tau}, y \right\}.$$

Choose $\boldsymbol{u}^{(1)}, \dots, \boldsymbol{u}^{(r)} \in \mathscr{B}^0$ such that $\mathcal{A}_J(\boldsymbol{u}^{(1)}), \dots, \mathcal{A}_J(\boldsymbol{u}^{(r)})$ form a basis of $\mathcal{A}_J(\mathscr{B}^0)$. Let

$$\boldsymbol{G} := \left[\mathcal{A}_J(\boldsymbol{u}^{(1)}) \ \cdots\ \mathcal{A}_J(\boldsymbol{u}^{(r)}) \right],$$

$$\boldsymbol{\alpha}_n := (\boldsymbol{G}^\intercal \boldsymbol{G})^{-1} \boldsymbol{G}^\intercal \mathcal{A}_J((t_{d,n}\dot{\boldsymbol{q}}_d)_d).$$

The columns of $\boldsymbol{G}$ are independent. Moreover, $\mathcal{A}_J((\dot{\boldsymbol{q}}_d)_d) = \mathbf{0}$, and the first displayed convergence gives $\boldsymbol{\alpha}_n \to \mathbf{0}$. If $r = 0$, omit the coefficient vector and use empty sums below.

Set

$$\dot{\boldsymbol{q}}_d^{(n)} := t_{d,n}\dot{\boldsymbol{q}}_d - \sum_{j=1}^{r} \boldsymbol{\alpha}_{n;(j)} \boldsymbol{u}_d^{(j)}.$$

Then $\mathcal{A}_J((\dot{\boldsymbol{q}}_d^{(n)})_d) = \mathbf{0}$, which gives the required common integral $\dot{\boldsymbol{\pi}}_n \in \ker \mathbf{A}_{\mathbf{S},J}$. Choose $C_j < \infty$ with $|\boldsymbol{u}_{d;(\boldsymbol{\tau})}^{(j)}| \le C_j q_{d|\boldsymbol{\tau}}^0$. The relative bound holds with

$$C_n := n + \sum_{j=1}^{r} |\boldsymbol{\alpha}_{n;(j)}| C_j.$$

Also, $|(\mathbf{R}_d \boldsymbol{u}_d^{(j)})_{(z)}| \le C_j p_{d|z}^P$, and so

$$\|(\mathbf{R}_d \boldsymbol{u}_d^{(j)})_d\|_P^2 \le C_j^2 \sum_{d,z} P_Z(z) \int_{\mathcal{Y}} p_{d|z}^P \, d\mu = C_j^2.$$

Therefore,

$$\begin{aligned} &\sum_d \int_{\mathcal{Y}} \|\dot{\boldsymbol{q}}_d^{(n)} - \dot{\boldsymbol{q}}_d\|_1 \, d\mu + \left\| (\mathbf{R}_d \dot{\boldsymbol{q}}_d^{(n)})_d - \boldsymbol{b} \right\|_P \\ &\le \sum_d \int_{\mathcal{Y}} (1 - t_{d,n}) \|\dot{\boldsymbol{q}}_d\|_1 \, d\mu + \|((t_{d,n} - 1)\boldsymbol{b}_d)_d\|_P \\ &\quad + \sum_{j=1}^{r} |\boldsymbol{\alpha}_{n;(j)}| \left(\sum_d \int_{\mathcal{Y}} \|\boldsymbol{u}_d^{(j)}\|_1 \, d\mu + C_j \right) \to 0, \end{aligned}$$

which proves both convergence assertions. ■

Lemma B.15. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and $(\boldsymbol{\pi}^0, \boldsymbol{q}^0) \in \mathcal{F}_{\mathbf{S}}^P$. Define*

$$J := \left\{ j \in [m_{\mathbf{S}}] \,\middle|\, \mathbf{A}_{\mathbf{S};(j,\cdot)} \boldsymbol{\pi}^0 = \mathbf{a}_{\mathbf{S};(j)} \right\}.$$

Suppose, for some $C < \infty$,

$$|\dot{q}_{d|\boldsymbol{\tau}}(y)| \le C q_{d|\boldsymbol{\tau}}^0(y) \qquad \forall d, \boldsymbol{\tau}, y,$$

$$\int_{\mathcal{Y}} \dot{\boldsymbol{q}}_d \, d\mu = \dot{\boldsymbol{\pi}} \in \ker \mathbf{A}_{\mathbf{S},J} \qquad \forall d.$$

Then there exist $\epsilon > 0$ and laws $(P_t)_{|t|<\epsilon}$ satisfying Condition D, with

$$P_0 = P, \qquad \mathcal{Q}_{\mathbf{S}}(P_t) \ne \emptyset, \qquad P_{t;Z} = P_Z,$$

$$\boldsymbol{p}_d^{P_t} = \boldsymbol{p}_d^P + t \mathbf{R}_d \dot{\boldsymbol{q}}_d \qquad \forall d, \ |t| < \epsilon.$$

This path is QMD at zero with score

$$h(y, d, z) := \begin{cases} \dfrac{(\mathbf{R}_d \dot{\boldsymbol{q}}_d(y))_{(z)}}{p_{d|z}^P(y)} & \text{if } p_{d|z}^P(y) > 0, \\ 0 & \text{otherwise}, \end{cases}$$

and $\mathbb{E}_P[h \mid Z] = 0$. If $P \in \mathfrak{P}_{\mathbf{S}}$, ϵ can be chosen so that $P_t \in \mathfrak{P}_{\mathbf{S}}$ for every $|t| < \epsilon$.

Proof. Choose $\epsilon_0 > 0$ such that

$$\epsilon_0 C \le \frac{1}{2},$$
$$\epsilon_0 |\mathbf{A}_{\mathbf{S};(j,\cdot)} \dot{\boldsymbol{\pi}}| \le \frac{\mathbf{a}_{\mathbf{S};(j)} - \mathbf{A}_{\mathbf{S};(j,\cdot)} \boldsymbol{\pi}^0}{2} \qquad \forall j \notin J.$$

Such a choice exists because there are finitely many rows and each numerator in the second line is positive. For $|t| < \epsilon_0$, set

$$\boldsymbol{q}_d^t := \boldsymbol{q}_d^0 + t\dot{\boldsymbol{q}}_d, \qquad \boldsymbol{\pi}^t := \boldsymbol{\pi}^0 + t\dot{\boldsymbol{\pi}}.$$

Then

$$q^t_{d|\boldsymbol{\tau}} \ge (1 - |t|C) q^0_{d|\boldsymbol{\tau}} \ge \frac{1}{2} q^0_{d|\boldsymbol{\tau}},$$
$$\int_{\mathcal{Y}} \boldsymbol{q}_d^t \, d\mu = \boldsymbol{\pi}^t,$$
$$\mathbf{A}_{\mathbf{S},J} \boldsymbol{\pi}^t = \mathbf{a}_{\mathbf{S},J},$$
$$\mathbf{a}_{\mathbf{S};(j)} - \mathbf{A}_{\mathbf{S};(j,\cdot)} \boldsymbol{\pi}^t \ge \frac{\mathbf{a}_{\mathbf{S};(j)} - \mathbf{A}_{\mathbf{S};(j,\cdot)} \boldsymbol{\pi}^0}{2} \qquad \forall j \notin J.$$

In particular, $\boldsymbol{\pi}^t \in \mathcal{R}_{\boldsymbol{\pi}}$.

Define

$$r_t(y, d, z) := P_Z(z) (\mathbf{R}_d \boldsymbol{q}_d^t(y))_{(z)}.$$

This function is nonnegative, and

$$\sum_d \int_{\mathcal{Y}} r_t(y, d, z) \, d\mu(y) = P_Z(z) \sum_d (\mathbf{R}_d \boldsymbol{\pi}^t)_{(z)} = P_Z(z) \mathbf{1}^\intercal \boldsymbol{\pi}^t = P_Z(z).$$

It therefore defines a probability law P_t with $P_{t;Z} = P_Z$. The common-parent construction in Theorem 3.1, applied to $(\boldsymbol{\pi}^t, (\boldsymbol{q}_d^t)_d)$, gives $\mathcal{Q}_{\mathbf{S}}(P_t) \neq \emptyset$ and the asserted conditional subdensities.

Nonnegativity of $\mathbf{R}_d$ gives

$$|(\mathbf{R}_d \dot{\boldsymbol{q}}_d)_{(z)}| \le C p^P_{d|z},$$

and so $|h| \le C$ and $r_t = r_0(1 + th)$. Since $\mathbf{1}^\intercal \boldsymbol{\pi}^t = 1$ for all these t, $\mathbf{1}^\intercal \dot{\boldsymbol{\pi}} = 0$, and

$$\mathbb{E}_P[h \mid Z = z] = \sum_d \int_{\mathcal{Y}} (\mathbf{R}_d \dot{\boldsymbol{q}}_d)_{(z)} \, d\mu = \mathbf{1}^\intercal \dot{\boldsymbol{\pi}} = 0.$$

For an absolute constant $C_T < \infty$,

$$\left| \sqrt{1 + u} - 1 - \frac{u}{2} \right| \le C_T u^2 \qquad \forall |u| \le \frac{1}{2}.$$

Therefore,

$$\left\| \sqrt{r_t} - \sqrt{r_0} - \frac{t}{2} h \sqrt{r_0} \right\|_{L^2(\nu)}^2 \leq C_T^2 t^4 \mathbb{E}_P[h^4] \leq C_T^2 C^4 t^4 = o(t^2),$$

which proves QMD.

Finally,

$$\Delta(P_t, P) = |t| \sum_d \int_{\mathcal{Y}} \|\mathbf{R}_d \dot{\boldsymbol{q}}_d\|_1 \, d\mu \to 0.$$

If $P \in \mathfrak{P}_{\mathbf{S}}$, Corollary B.1 therefore keeps the minimum conditioning probability positive after reducing ϵ_0. The assignment probabilities remain unchanged, and so this also applies to the model with known assignment probabilities. ■

Lemma B.16. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. There exist bounded measurable $\boldsymbol{f}_d^\dagger : \mathcal{Y} \to \mathbb{R}^{\mathcal{Z}}$, vectors $\boldsymbol{v}_d^\dagger \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$, and $\boldsymbol{\lambda}^\dagger \in \mathbb{R}^{m_{\mathbf{S}}}_{\geq 0}$ such that, defining*

$$\boldsymbol{a}_d^\dagger(y) := \mathbf{R}_d^\intercal \boldsymbol{f}_d^\dagger(y) + \boldsymbol{v}_d^\dagger,$$

we have

$$\begin{aligned}
\boldsymbol{a}_{d;(\boldsymbol{\tau})}^\dagger(y) &= 0 \qquad \forall d, y, \boldsymbol{\tau} \in S_d(y), \\
\boldsymbol{a}_{d;(\boldsymbol{\tau})}^\dagger(y) &\geq 1 \qquad \forall d, y, \boldsymbol{\tau} \notin S_d(y), \\
\boldsymbol{\lambda}_{(j)}^\dagger &\geq 1 \qquad \forall j \in J_{\mathbf{S}}^P, \\
\boldsymbol{\lambda}_{(j)}^\dagger &= 0 \qquad \forall j \notin J_{\mathbf{S}}^P, \\
\sum_d \boldsymbol{v}_d^\dagger &= \mathbf{A}_{\mathbf{S}}^\intercal \boldsymbol{\lambda}^\dagger, \\
\sum_d \int_{\mathcal{Y}} (\boldsymbol{f}_d^\dagger)^\intercal \boldsymbol{p}_d^P \, d\mu + (\boldsymbol{\lambda}^\dagger)^\intercal \mathbf{a}_{\mathbf{S}} &= 0.
\end{aligned}$$

Proof. For each d, apply Lemma A.3 to $\mathbf{R}_{d;(\cdot,\boldsymbol{\mathcal{T}}_d)}$ and $\boldsymbol{p}_d^P$, and denote its partition by $\mathcal{B}_d$. Let

$$\mathcal{B} := \left\{ \bigcap_{d \in \mathcal{D}} B_d \,\middle|\, B_d \in \mathcal{B}_d \, \forall d, \; \mu\left(\bigcap_{d \in \mathcal{D}} B_d \right) > 0 \right\},$$

$$\boldsymbol{b}_{d,B} := \int_B \boldsymbol{p}_d^P \, d\mu.$$

The omitted cells form a null set. Because Lemma A.3 applies to every measurable subset of each of its cells, its realization conclusion holds on every $B \in \mathcal{B}$.

Consider the finite system

$$\boldsymbol{x}_{d,B} \geq \mathbf{0}, \qquad \mathbf{R}_d \boldsymbol{x}_{d,B} = \boldsymbol{b}_{d,B} \qquad \forall d, B,$$

$$\sum_{B\in\mathcal{B}} \boldsymbol{x}_{d,B} = \boldsymbol{\pi} \qquad \forall d, \qquad\qquad \mathbf{A_S}\boldsymbol{\pi} + \boldsymbol{s} = \mathbf{a_S}, \qquad \boldsymbol{s} \geq \mathbf{0}. \tag{B.15}$$

It is nonempty by compatibility. Its feasible set is compact: $\boldsymbol{\pi} \in \mathcal{R}_{\boldsymbol{\pi}}$, $0 \leq \boldsymbol{x}_{d,B;(\boldsymbol{\tau})} \leq \boldsymbol{\pi}_{(\boldsymbol{\tau})} \leq 1$, and $\boldsymbol{s} = \mathbf{a_S} - \mathbf{A_S}\boldsymbol{\pi}$.

Every feasible array in (B.15) can be realized by nonnegative functions $\boldsymbol{q}_{d,B}$, zero outside B, satisfying

$$\int_B \boldsymbol{q}_{d,B}\, d\mu = \boldsymbol{x}_{d,B},$$

$$\mathbf{R}_d \boldsymbol{q}_{d,B}(y) = \boldsymbol{p}_d^P(y) \qquad \forall y \in B,$$

$$\boldsymbol{q}_{d,B;(\boldsymbol{\tau})}(y) > 0 \quad \Longleftrightarrow \quad \boldsymbol{x}_{d,B;(\boldsymbol{\tau})} > 0 \qquad \forall y \in B, \boldsymbol{\tau}.$$

For active coordinates this is Lemma A.3. For inactive coordinates, use

$$\boldsymbol{q}_{d,B;(\boldsymbol{\tau})}(y) := \boldsymbol{x}_{d,B;(\boldsymbol{\tau})} \frac{\rho_0(y)\mathbb{1}_B(y)}{\int_B \rho_0\, d\mu},$$

where ρ_0 is supplied by Lemma A.1. Then $\boldsymbol{q}_d := \sum_B \boldsymbol{q}_{d,B}$ has common integral $\boldsymbol{\pi}$, and $(\boldsymbol{\pi}, (\boldsymbol{q}_d)_d) \in \mathcal{F}_{\mathbf{S}}^P$.

Take $(\boldsymbol{\pi}^\dagger, \boldsymbol{q}^\dagger)$ from Lemma B.12, and set

$$\boldsymbol{x}_{d,B}^\dagger := \int_B \boldsymbol{q}_d^\dagger\, d\mu, \qquad\qquad \boldsymbol{s}^\dagger := \mathbf{a_S} - \mathbf{A_S}\boldsymbol{\pi}^\dagger,$$

$$S_{d,B} := \left\{ \boldsymbol{\tau} \in \boldsymbol{\mathcal{T}} \,\middle|\, \boldsymbol{x}_{d,B;(\boldsymbol{\tau})}^\dagger > 0 \right\}.$$

If $\boldsymbol{x}_{d,B;(\boldsymbol{\tau})}^\dagger = 0$, then $q_{d|\boldsymbol{\tau}}^\dagger = 0$ almost everywhere on B, and so $\boldsymbol{\tau} \notin S_d(y)$ there. If $\boldsymbol{x}_{d,B;(\boldsymbol{\tau})}^\dagger > 0$, the preceding realization of the array $\boldsymbol{x}^\dagger$ gives a feasible pair positive in this coordinate throughout B. The universal support assertion in Lemma B.12 then gives $\boldsymbol{\tau} \in S_d(y)$ almost everywhere on B. Therefore,

$$S_d(y) = S_{d,B} \qquad \forall B \in \mathcal{B},\ y \in B.$$

Define

$$\boldsymbol{c}_{d,B} := \left(\mathbb{1}\{\boldsymbol{\tau} \notin S_{d,B}\}\right)_{\boldsymbol{\tau}\in\boldsymbol{\mathcal{T}}}, \qquad\qquad \boldsymbol{e} := \left(\mathbb{1}\{j \in J_{\mathbf{S}}^P\}\right)_{j\in[m_{\mathbf{S}}]}.$$

Maximize

$$\sum_{d,B} \boldsymbol{c}_{d,B}^\intercal \boldsymbol{x}_{d,B} + \boldsymbol{e}^\intercal \boldsymbol{s}$$

subject to (B.15). Every feasible array has $\boldsymbol{x}_{d,B;(\boldsymbol{\tau})} = 0$ for $\boldsymbol{\tau} \notin S_{d,B}$, by its realization and Lemma B.12, and $\boldsymbol{s}_{(j)} = 0$ for $j \in J_{\mathbf{S}}^P$, by the definition of $J_{\mathbf{S}}^P$.

The dual minimizes

$$\sum_{d,B} \boldsymbol{f}_{d,B}^\intercal \boldsymbol{b}_{d,B} + \boldsymbol{\lambda}^\intercal \mathbf{a_S}$$

over $\boldsymbol{f}_{d,B} \in \mathbb{R}^{\mathcal{Z}}$, $\boldsymbol{v}_d \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$, and $\boldsymbol{\lambda} \in \mathbb{R}^{m_{\mathbf{S}}}$ satisfying

$$\mathbf{R}_d^{\intercal} \boldsymbol{f}_{d,B} + \boldsymbol{v}_d \geq \boldsymbol{c}_{d,B} \qquad \forall d, B,$$
$$\boldsymbol{\lambda} \geq \boldsymbol{e}, \qquad\qquad \sum_d \boldsymbol{v}_d = \mathbf{A}_{\mathbf{S}}^{\intercal} \boldsymbol{\lambda}.$$

Here the first inequalities correspond to $\boldsymbol{x}_{d,B} \geq \mathbf{0}$, the second to $\boldsymbol{s} \geq \mathbf{0}$, and the equality to the unrestricted vector $\boldsymbol{\pi}$.

Strong duality supplies an optimal $(\boldsymbol{f}^{\dagger}_{d,B}, \boldsymbol{v}^{\dagger}_d, \boldsymbol{\lambda}^{\dagger})$ with value zero. Substitution of the feasible $(\boldsymbol{\pi}^{\dagger}, \boldsymbol{x}^{\dagger}, \boldsymbol{s}^{\dagger})$ gives

$$0 = \sum_{d,B} \left(\mathbf{R}_d^{\intercal} \boldsymbol{f}^{\dagger}_{d,B} + \boldsymbol{v}^{\dagger}_d \right)^{\intercal} \boldsymbol{x}^{\dagger}_{d,B} + (\boldsymbol{\lambda}^{\dagger})^{\intercal} \boldsymbol{s}^{\dagger}.$$

Since $\boldsymbol{x}^{\dagger}_{d,B;(\boldsymbol{\tau})} > 0$ for $\boldsymbol{\tau} \in S_{d,B}$, the corresponding dual coefficient must be zero. Also, $\boldsymbol{s}^{\dagger}_{(j)} > 0$ for $j \notin J^P_{\mathbf{S}}$ by Lemma B.12, and so $\boldsymbol{\lambda}^{\dagger}_{(j)} = 0$ for those rows.

Finally, define

$$\boldsymbol{f}^{\dagger}_d(y) := \sum_{B \in \mathcal{B}} \mathbb{1}_B(y) \boldsymbol{f}^{\dagger}_{d,B} \,,$$

and note that it is measurable and bounded since $\mathcal{B}$ is finite, as required. ■

Lemma B.17. *Suppose P satisfies Condition D, $\mathbf{S}$ satisfies Condition CP, and $\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset$. Let $(P_t)_{|t|<\epsilon}$ be a two-sided QMD path through P with score $s \in L^2_0(P)$, such that P_t satisfies Condition D and $\mathcal{Q}_{\mathbf{S}}(P_t) \neq \emptyset$ for every $|t| < \epsilon$. Then*

$$\dot{\boldsymbol{p}}^P(s) \in \mathscr{U}_P,$$

where $\mathscr{U}_P$ is defined in Lemma B.13. In particular, this conclusion holds for every two-sided QMD path through P in $\mathfrak{P}_{\mathbf{S}}$.

Proof. Let $\boldsymbol{b} := \dot{\boldsymbol{p}}^P(s)$. By Lemma B.11, $\boldsymbol{b} \in \mathbb{H}_P$ and

$$\left\| \left(\frac{\boldsymbol{p}_d^{P_t} - \boldsymbol{p}_d^P}{t} \right)_d - \boldsymbol{b} \right\|_{L^1} \to 0.$$

Take $(\boldsymbol{\pi}^{\dagger}, \boldsymbol{q}^{\dagger})$ from Lemma B.12, and $\boldsymbol{f}^{\dagger}_d, \boldsymbol{v}^{\dagger}_d, \boldsymbol{\lambda}^{\dagger}, \boldsymbol{a}^{\dagger}_d$ from Lemma B.16. Define

$$F(t) := \sum_d \int_{\mathcal{Y}} (\boldsymbol{f}^{\dagger}_d)^{\intercal} \boldsymbol{p}_d^{P_t} \, d\mu + (\boldsymbol{\lambda}^{\dagger})^{\intercal} \mathbf{a}_{\mathbf{S}}.$$

For each t, choose any $(\boldsymbol{\pi}^t, \boldsymbol{q}^t) \in \mathcal{F}^{P_t}_{\mathbf{S}}$. The response equations and common integrals give

$$F(t) = \sum_d \int_{\mathcal{Y}} (\boldsymbol{a}^{\dagger}_d)^{\intercal} \boldsymbol{q}^t_d \, d\mu + (\boldsymbol{\lambda}^{\dagger})^{\intercal} (\mathbf{a}_{\mathbf{S}} - \mathbf{A}_{\mathbf{S}} \boldsymbol{\pi}^t)$$

$$\geq \sum_d \int_{\mathcal{Y}} \sum_{\boldsymbol{\tau} \notin S_d(y)} q^t_{d|\boldsymbol{\tau}}(y)\, d\mu(y) + \sum_{j \in J^P_{\mathbf{S}}} \left(\mathbf{a}_{\mathbf{S};(j)} - \mathbf{A}_{\mathbf{S};(j,\cdot)}\boldsymbol{\pi}^t\right) \geq 0.$$

Also, $F(0) = 0$ by Lemma B.16. Boundedness of $\boldsymbol{f}^\dagger_d$ and the L^1 derivative above yield

$$F(t) = t \sum_d \int_{\mathcal{Y}} (\boldsymbol{f}^\dagger_d)^\intercal \boldsymbol{b}_d\, d\mu + o(|t|).$$

Since $F(t) \geq F(0) = 0$ for both signs of t, the coefficient of t is zero, and so $F(t) = o(|t|)$.

For $t \neq 0$, define

$$\boldsymbol{e}_{t,d;(\boldsymbol{\tau})}(y) := q^t_{d|\boldsymbol{\tau}}(y) \mathbb{1}\{\boldsymbol{\tau} \notin S_d(y)\},$$

$$\boldsymbol{u}_{t,d} := \frac{\boldsymbol{q}^t_d - \boldsymbol{e}_{t,d} - \boldsymbol{q}^\dagger_d}{t}.$$

Then $\boldsymbol{u}_t \in \mathscr{E}_P$, and the preceding inequality gives

$$\sum_d \int_{\mathcal{Y}} \|\boldsymbol{e}_{t,d}\|_1\, d\mu \leq F(t),$$

$$\left\| \mathbf{A}_{\mathbf{S},J^P_{\mathbf{S}}}(\boldsymbol{\pi}^t - \boldsymbol{\pi}^\dagger) \right\|_1 \leq F(t).$$

Let $C_R := \max_d \|\mathbf{R}_d\|_{1\leftarrow 1}$. Then

$$\|(\mathbf{R}_d \boldsymbol{u}_{t,d})_d - \boldsymbol{b}\|_{L^1} \leq \left\| \left(\frac{\boldsymbol{p}^{P_t}_d - \boldsymbol{p}^P_d}{t} \right)_d - \boldsymbol{b} \right\|_{L^1} + \frac{C_R F(t)}{|t|} \to 0.$$

Set $J := J^P_{\mathbf{S}}$ and use the map $\mathcal{A}_J$ from (B.14). The common integrals of $\boldsymbol{q}^t_d$ and $\boldsymbol{q}^\dagger_d$ give

$$\int_{\mathcal{Y}} (\boldsymbol{u}_{t,d} - \boldsymbol{u}_{t,d_0})\, d\mu = -\frac{1}{t} \int_{\mathcal{Y}} (\boldsymbol{e}_{t,d} - \boldsymbol{e}_{t,d_0})\, d\mu,$$

$$\mathbf{A}_{\mathbf{S},J} \int_{\mathcal{Y}} \boldsymbol{u}_{t,d_0}\, d\mu = \frac{1}{t} \left[\mathbf{A}_{\mathbf{S},J}(\boldsymbol{\pi}^t - \boldsymbol{\pi}^\dagger) - \mathbf{A}_{\mathbf{S},J} \int_{\mathcal{Y}} \boldsymbol{e}_{t,d_0}\, d\mu \right].$$

Therefore, with $C_A := |\mathcal{D}| + 1 + \|\mathbf{A}_{\mathbf{S},J}\|_{1\leftarrow 1}$,

$$\|\mathcal{A}_J(\boldsymbol{u}_t)\|_1 \leq \frac{C_A F(t)}{|t|} \to 0.$$

The image $\mathcal{A}_J(\mathscr{E}_P)$ is finite-dimensional. Choose a basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$ of this image and preimages $\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(r)} \in \mathscr{E}_P$ with $\mathcal{A}_J(\boldsymbol{w}^{(j)}) = \boldsymbol{a}_j$. Define the linear map

$$B\left(\sum_{j=1}^r c_j \boldsymbol{a}_j \right) := \sum_{j=1}^r c_j \boldsymbol{w}^{(j)}.$$

Then $\mathcal{A}_J B$ is the identity on $\mathcal{A}_J(\mathscr{E}_P)$, and, for some $C_B < \infty$,

$$\sum_d \int_{\mathcal{Y}} \|(B\boldsymbol{a})_d\|_1\, d\mu \leq C_B \|\boldsymbol{a}\|_1 \qquad \forall \boldsymbol{a} \in \mathcal{A}_J(\mathscr{E}_P).$$

If the image is $\{\mathbf{0}\}$, take $B = 0$.

Set

$$\widetilde{\boldsymbol{u}}_t := \boldsymbol{u}_t - B\mathcal{A}_J(\boldsymbol{u}_t).$$

Then $\widetilde{\boldsymbol{u}}_t \in \mathscr{E}_P$ and $\mathcal{A}_J(\widetilde{\boldsymbol{u}}_t) = \mathbf{0}$, and so $(\mathbf{R}_d \widetilde{\boldsymbol{u}}_{t,d})_d \in \mathscr{U}_{P,1}$. Moreover,

$$\|(\mathbf{R}_d \widetilde{\boldsymbol{u}}_{t,d})_d - \boldsymbol{b}\|_{L^1} \le \|(\mathbf{R}_d \boldsymbol{u}_{t,d})_d - \boldsymbol{b}\|_{L^1} + C_R C_B \|\mathcal{A}_J(\boldsymbol{u}_t)\|_1 \to 0.$$

Closedness in Lemma B.13 gives $\boldsymbol{b} \in \mathscr{U}_{P,1}$. Since $\boldsymbol{b} \in \mathbb{H}_P$, it follows that $\boldsymbol{b} \in \mathscr{U}_P$, as required. ∎

Lemma B.18. *Suppose θ is the conditional-mean parameter (4.1), with $\emptyset \neq A \subseteq \mathcal{T}_{d^\star}$, and Condition T3 holds. Suppose $\mathbf{S}$ satisfies Condition CP. Fix $\gamma \in [\underline{\kappa} - 1, \overline{\kappa} + 1]$, and use $\mathcal{D}_\theta, \boldsymbol{c}_0$ from Lemma A.11 and B_θ from Lemma A.12.*

Let $\mathscr{V}_\theta(\gamma)$ be the class of measurable functions

$$m(y, d, z) = \mathbb{1}\{d = d^\star\}\boldsymbol{\ell}_{0;(z)}(y) + \boldsymbol{\ell}_{d;(z)}(y) + \frac{\boldsymbol{\lambda}^\intercal \mathbf{a}_{\mathbf{S}}}{|\mathcal{Z}|}$$

admitting witnesses $(\boldsymbol{u}, \boldsymbol{\lambda}) \in \mathcal{D}_\theta$ and measurable $\boldsymbol{\ell}_0, \boldsymbol{\ell}_d : \mathcal{Y} \to \mathbb{R}^{\mathcal{Z}}_{\ge 0}$ such that

$$\begin{aligned}
&\|(\boldsymbol{u}, \boldsymbol{\lambda})\|_\infty \le B_\theta, \\
&\mathbf{R}_{d^\star}^\intercal \boldsymbol{\ell}_0(y) \ge \boldsymbol{c}_0(y; \gamma, \boldsymbol{u}, \boldsymbol{\lambda}), && \|\boldsymbol{\ell}_0(y)\|_\infty \le \|\boldsymbol{c}_0(y; \gamma, \boldsymbol{u}, \boldsymbol{\lambda})\|_\infty, \\
&\mathbf{R}_d^\intercal \boldsymbol{\ell}_d(y) \ge \boldsymbol{u}_d, && \|\boldsymbol{\ell}_d(y)\|_\infty \le \|\boldsymbol{u}_d\|_\infty \qquad \forall d \in \mathcal{D}.
\end{aligned}$$

Every member satisfies $\|m\|_\infty \le C_m$, with C_m from Corollary A.4. For every finite list $f_1, \dots, f_r \in L^1(\nu)$, the set

$$\left\{ \left(\int_{\mathcal{W}} f_j(w) m(w)\, d\nu(w) \right)_{j \in [r]} \,\middle|\, m \in \mathscr{V}_\theta(\gamma) \right\}$$

is nonempty and compact.

For every R satisfying Condition D with $\mathcal{Q}_{\mathbf{S}}(R) \neq \emptyset$,

$$\overline{\Psi}(\gamma; R) = \min_{m \in \mathscr{V}_\theta(\gamma)} \sum_z \mathbb{E}_R[m(W) \mid Z = z]. \tag{B.16}$$

If, in addition, R satisfies Assumption 5.1, then

$$\mathfrak{M}_R^\star = \left\{ m \in \mathscr{V}_\theta(\bar{\theta}(R)) \,\middle|\, \sum_z \mathbb{E}_R[m(W) \mid Z = z] = 0 \right\}.$$

Proof. The bounds defining the class give $\|m\|_\infty \le C_m$ by the calculation in Corollary A.4. The class is nonempty: take $(\boldsymbol{u}, \boldsymbol{\lambda}) = \mathbf{0}$ and choose the bounded inner duals supplied

by Lemma 3.2. For any feasible pair at R, direct integration of the dual inequalities gives

$$\sum_z \mathbb{E}_R[m(W) \mid Z=z] \geq \int_{\mathcal{Y}} (\kappa-\gamma)\boldsymbol{c}_A^{\mathsf{T}}\boldsymbol{q}_{d^\star}\, d\mu + \boldsymbol{\lambda}^{\mathsf{T}}(\mathbf{a}_{\mathbf{S}} - \mathbf{A}_{\mathbf{S}}\boldsymbol{\pi})$$
$$\geq \int_{\mathcal{Y}} (\kappa-\gamma)\boldsymbol{c}_A^{\mathsf{T}}\boldsymbol{q}_{d^\star}\, d\mu.$$

Taking the supremum over $\mathcal{F}_{\mathbf{S}}^R$ gives the lower bound in (B.16). Equality is attained by Lemma A.12 and Corollary A.4. At $\gamma = \bar{\theta}(R)$ the minimum is zero, and the remaining assertion follows from the definition of $\mathfrak{M}_R^\star$.

To prove compactness, take any sequence $m_n \in \mathscr{V}_\theta(\gamma)$ and choose witnesses. Enumerate $\mathcal{D} = \{d_1, \dots, d_J\}$, and also define $d_0 = d^\star$. Compactness of the bounded outer domain gives $n_1 < n_2 < \cdots$ and a feasible limit $(\boldsymbol{u}, \boldsymbol{\lambda})$ such that

$$\|(\boldsymbol{u}_{n_k}, \boldsymbol{\lambda}_{n_k}) - (\boldsymbol{u}, \boldsymbol{\lambda})\|_\infty \to 0.$$

Writing $\boldsymbol{c}_{n_k,i}$ and $\boldsymbol{c}_i$ for the corresponding coefficients, their formulas give

$$\max_{0\leq i\leq J} \sup_y \|\boldsymbol{c}_{n_k,i}(y) - \boldsymbol{c}_i(y)\|_\infty$$
$$\leq \left(1 + |\mathcal{D}| + \|\mathbf{A}_{\mathbf{S}}^{\mathsf{T}}\|_{\infty\to\infty}\right) \|(\boldsymbol{u}_{n_k}, \boldsymbol{\lambda}_{n_k}) - (\boldsymbol{u}, \boldsymbol{\lambda})\|_\infty \to 0.$$

Take the positive probability density ρ_0 from Lemma A.1. The vectors $\rho_0 \boldsymbol{\ell}_{n_k,i}$ have a common integrable envelope $C\rho_0$. Apply Lemma A.6 jointly to their finitely many coordinates. It supplies $k_1 < k_2 < \cdots$ and bounded measurable $\boldsymbol{\ell}_i$ such that

$$\int_{\mathcal{Y}} f(y) \left(\boldsymbol{\ell}_{n_{k_j},i}(y) - \boldsymbol{\ell}_i(y)\right) d\mu(y) \to \mathbf{0} \qquad \forall f \in L^1(\mu).$$

Indeed, the test function in Lemma A.6 is f/ρ_0, and its required weighted integral is $C\int_{\mathcal{Y}} |f|\, d\mu < \infty$.

For every nonnegative $f \in L^1(\mu)$, we now have

$$\int_{\mathcal{Y}} f\left(\mathbf{R}_{d_i}^{\mathsf{T}}\boldsymbol{\ell}_i - \boldsymbol{c}_i\right) d\mu = \lim_j \int_{\mathcal{Y}} f\left(\mathbf{R}_{d_i}^{\mathsf{T}}\boldsymbol{\ell}_{n_{k_j},i} - \boldsymbol{c}_{n_{k_j},i}\right) d\mu \geq \mathbf{0},$$
$$\int_{\mathcal{Y}} f\boldsymbol{\ell}_i\, d\mu \geq \mathbf{0},$$
$$\int_{\mathcal{Y}} f\left(\|\boldsymbol{c}_i(y)\|_\infty \mathbf{1} - \boldsymbol{\ell}_i(y)\right) d\mu(y) \geq \mathbf{0}.$$

Since this holds for every such test function f (and μ is σ-finite), these inequalities must hold pointwise for a.e. $y \in \mathcal{Y}$, and so $m \in \mathscr{V}_\theta(\gamma)$, as required. ■

Lemma B.19. *Maintain the parameter and model assumptions of Lemma B.18. For all compatible R, S satisfying Condition D and all $\gamma \in [\underline{\kappa} - 1, \overline{\kappa} + 1]$,*

$$|\overline{\Psi}(\gamma; R) - \overline{\Psi}(\gamma; S)| \leq C_m \Delta(R, S). \tag{B.17}$$

Fix $P \in \mathfrak{P}_{\mathbf{S}}$ and a two-sided QMD path $(P_t)_{|t|<\epsilon} \subseteq \mathfrak{P}_{\mathbf{S}}$ through P with score s. Then

$$\overline{\Psi}(\bar{\theta}(P); P_t) = t\delta_P\left(\dot{\boldsymbol{p}}^P(s)\right) + o(t) \qquad (t \downarrow 0). \tag{B.18}$$

Proof. For fixed γ, choose a minimizer m_R at R in (B.16). It is also feasible in the minimization at S, and so

$$\begin{aligned}\overline{\Psi}(\gamma; S) - \overline{\Psi}(\gamma; R) &\le \sum_{d,z} \int_{\mathcal{Y}} m_R(y,d,z)(p^S_{d|z} - p^R_{d|z})\, d\mu \\ &\le C_m \Delta(R,S).\end{aligned}$$

Interchanging R, S proves (B.17).

Let $\bar{\theta} := \bar{\theta}(P)$ and $\dot{\boldsymbol{p}} := \dot{\boldsymbol{p}}^P(s)$. By Lemma B.11,

$$\sum_{d,z} \|p^{P_t}_{d|z} - p^P_{d|z} - t\dot{p}_{d|z}\|_{L^1(\mu)} = o(|t|).$$

Hence

$$\begin{aligned}&\sup_{m \in \mathscr{V}_\theta(\bar{\theta})} \left| \sum_z \mathbb{E}_{P_t}[m(W) \mid Z = z] - \sum_z \mathbb{E}_P[m(W) \mid Z = z] - t \sum_{d,z} \int_{\mathcal{Y}} m\dot{p}_{d|z}\, d\mu \right| \\ &\le C_m \sum_{d,z} \|p^{P_t}_{d|z} - p^P_{d|z} - t\dot{p}_{d|z}\|_{L^1(\mu)} = o(|t|).\end{aligned}$$

By Lemma B.18, the set

$$K := \left\{ \left(\sum_z \mathbb{E}_P[m(W) \mid Z = z], \sum_{d,z} \int_{\mathcal{Y}} m\dot{p}_{d|z}\, d\mu \right) \,\middle|\, m \in \mathscr{V}_\theta(\bar{\theta}) \right\}$$

is nonempty and compact. Its minimum first coordinate is zero. Among points with first coordinate zero, the minimum second coordinate is $\delta_P(\dot{\boldsymbol{p}})$, because their moments are exactly $\mathfrak{M}^\star_P$.

Danskin's theorem, applied to $v(t) := \min_{(a,b)\in K}(a + tb)$, therefore gives

$$v(t) = t\delta_P(\dot{\boldsymbol{p}}) + o(t) \qquad (t \downarrow 0).$$

Taking minima in the uniform remainder bound above gives $\overline{\Psi}(\bar{\theta}; P_t) = v(t) + o(|t|)$, proving (B.18). ∎

Lemma B.20. *Maintain the parameter and model assumptions of Lemma B.18. Fix $P \in \mathfrak{P}_{\mathbf{S}}$ and a two-sided QMD path $(P_t)_{|t|<\epsilon} \subseteq \mathfrak{P}_{\mathbf{S}}$ through P with score s. Let $\dot{\boldsymbol{p}} := \dot{\boldsymbol{p}}^P(s)$. For every $b \in \mathbb{R}$,*

$$\overline{\Psi}(\bar{\theta}(P) + tb; P_t) = t\left[\delta_P(\dot{\boldsymbol{p}}) + \max_{a \in [\underline{p}^\star_A(P), \overline{p}^\star_A(P)]} (-ba)\right] + o(t) \qquad (t \downarrow 0).$$

Proof. Let $\bar{\theta} := \bar{\theta}(P)$ and $I_P := [\underline{p}^\star_A(P), \overline{p}^\star_A(P)]$. By Lemma A.10, choose a feasible pair at P_t attaining $\overline{\Psi}(\bar{\theta} + tb; P_t)$, and denote its numerator and conditioning probability by N_t and a_t. Evaluating the same pair at threshold $\bar{\theta}$ gives

$$\begin{aligned}\overline{\Psi}(\bar{\theta} + tb; P_t) &= \mathsf{N}_t - \bar{\theta} a_t - tba_t \\ &\le \overline{\Psi}(\bar{\theta}; P_t) - tba_t. \end{aligned} \tag{B.19}$$

By Lemma B.11, $\Delta(P_t, P) = O(t)$. Use Lemma B.2 to match the chosen pair to a pair in $\mathcal{F}^P_{\mathbf{S}}$ within $O(t)$ in $\|\cdot\|_{\mathcal{F},1}$, and denote the latter's numerator and mass by N^0_t, a^0_t. Boundedness of κ gives

$$|\mathsf{N}^0_t - \mathsf{N}_t| + |a^0_t - a_t| = O(t).$$

By Lemma A.10,

$$\overline{\Psi}(\gamma; P_t) = \max_{(\mathsf{N},a)\in K_{P_t}} \{\mathsf{N} - \gamma a\}, \qquad \gamma \in \mathbb{R}\,,$$

and since $0 \le a \le 1$ for every $(\mathsf{N}, a) \in K_{P_t}$,

$$\left|\overline{\Psi}(\bar{\theta} + tb; P_t) - \overline{\Psi}(\bar{\theta}; P_t)\right| \le \max_{(\mathsf{N},a)\in K_{P_t}} |tba| \le t|b|.$$

Combining this inequality with $\overline{\Psi}(\bar{\theta}; P) = 0$ and Lemma B.19 gives

$$|\overline{\Psi}(\bar{\theta} + tb; P_t)| \le t|b| + C_m \Delta(P_t, P) = O(t)\,,$$

and so, $\mathsf{N}^0_t - \bar{\theta} a^0_t = O(t)$.

By Lemma A.10, compactness of K_P, together with $\mathsf{N}^0_t - \bar{\theta} a^0_t \to 0$ and $a_t - a^0_t \to 0$, implies that every accumulation point of a_t belongs to I_P, which gives

$$\limsup_{t\downarrow 0} (-ba_t) \le \max_{a\in I_P} (-ba).$$

Now, dividing (B.19) by t and applying Lemma B.19 gives

$$\limsup_{t\downarrow 0} \frac{\overline{\Psi}(\bar{\theta} + tb; P_t)}{t} \le \delta_P(\dot{\boldsymbol{p}}) + \max_{a\in I_P} (-ba).$$

For the reverse bound, fix $a_0 \in I_P$ and an endpoint-attaining pair $(\boldsymbol{\pi}^0, \boldsymbol{q}^0)$ at P with conditioning probability a_0. Fix $\alpha \in (0,1)$. For each sufficiently small $u > 0$, choose $(\widehat{\boldsymbol{\pi}}_u, \widehat{\boldsymbol{q}}_u) \in \mathcal{F}^{P_u}_{\mathbf{S}}$ attaining $\overline{\Psi}(\bar{\theta}; P_u)$, and let $\widehat{a}_u := \boldsymbol{c}_A^{\mathsf{T}} \widehat{\boldsymbol{\pi}}_u$.

The pair

$$(\boldsymbol{\pi}_{t,\alpha}, \boldsymbol{q}_{t,\alpha}) := (1-\alpha)(\boldsymbol{\pi}^0, \boldsymbol{q}^0) + \alpha(\widehat{\boldsymbol{\pi}}_{t/\alpha}, \widehat{\boldsymbol{q}}_{t/\alpha})$$

is feasible for the law $R_{t,\alpha}$ with assignment probabilities P_Z and conditional sub-densities

$$\boldsymbol{p}_d^{R_{t,\alpha}} := (1-\alpha)\boldsymbol{p}_d^P + \alpha \boldsymbol{p}_d^{P_{t/\alpha}}.$$

Indeed, its marginal integrals agree, its type vector is in the convex set $\mathcal{R}_{\boldsymbol{\pi}}$, and its response equations are those displayed above; Lemma 3.1 gives compatibility.

Let $\boldsymbol{e}_t := \boldsymbol{p}^{P_t} - \boldsymbol{p}^P - t\dot{\boldsymbol{p}}$, with $\|\boldsymbol{e}_t\|_1 := \sum_{d,z} \int_{\mathcal{Y}} |e_{t,d|z}|\, d\mu$. By Lemma B.11, $\|\boldsymbol{e}_t\|_1 = o(|t|)$. For fixed α,

$$\begin{aligned}\Delta(R_{t,\alpha}, P_t) &= \|\alpha \boldsymbol{e}_{t/\alpha} - \boldsymbol{e}_t\|_1 \\ &\le \alpha\|\boldsymbol{e}_{t/\alpha}\|_1 + \|\boldsymbol{e}_t\|_1 = o(t).\end{aligned}$$

Evaluate the residual at the mixed feasible pair and use (B.17) to get

$$\begin{aligned}\overline{\Psi}(\bar{\theta} + tb; P_t) \ge \alpha \overline{\Psi}(\bar{\theta}; P_{t/\alpha}) - tb\left[(1-\alpha)a_0 + \alpha \widehat{a}_{t/\alpha}\right] \\ - C_m \Delta(R_{t,\alpha}, P_t).\end{aligned}$$

Since $a_0, \widehat{a}_{t/\alpha} \in [0,1]$, division by t and Lemma B.19 give

$$\liminf_{t \downarrow 0} \frac{\overline{\Psi}(\bar{\theta} + tb; P_t)}{t} \ge \delta_P(\dot{\boldsymbol{p}}) - ba_0 - |b|\alpha.$$

Let $\alpha \downarrow 0$, then maximize over $a_0 \in I_P$. This matches the upper bound and proves the result. ■

Corollary B.2. *Under the assumptions and notation of Lemma B.20,*

$$\lim_{t \downarrow 0} \frac{\bar{\theta}(P_t) - \bar{\theta}(P)}{t} = \max_{a \in [\underline{p}^\star_A(P), \overline{p}^\star_A(P)]} \frac{\delta_P(\dot{\boldsymbol{p}})}{a}.$$

Proof. Let $a_- := \underline{p}^\star_A(P) > 0$, $a_+ := \overline{p}^\star_A(P)$, and

$$F(b) := \delta_P(\dot{\boldsymbol{p}}) + \max_{a \in [a_-, a_+]} (-ba),$$

$$b_* := \max_{a \in [a_-, a_+]} \frac{\delta_P(\dot{\boldsymbol{p}})}{a}.$$

Direct substitution gives $F(b_*) = 0$. For $b_2 > b_1$,

$$F(b_2) \le F(b_1) - a_-(b_2 - b_1).$$

Thus, for every $\eta > 0$, $F(b_* - \eta) > 0$ and $F(b_* + \eta) < 0$. By Lemma B.20, for all sufficiently small $t > 0$,

$$\overline{\Psi}(\bar{\theta}(P) + t(b_* - \eta); P_t) > 0,$$

$$\overline{\Psi}(\bar{\theta}(P) + t(b_* + \eta); P_t) < 0.$$

The residual-sign characterization in Lemma A.10 therefore gives

$$b_* - \eta < \frac{\bar{\theta}(P_t) - \bar{\theta}(P)}{t} < b_* + \eta.$$

■

Lemma B.21. *Suppose θ is the conditional-mean parameter (4.1), with $\emptyset \neq A \subseteq \mathcal{T}_{d^\star}$, and Condition T3 holds. Suppose $\mathbf{S}$ satisfies Condition CP and $P \in \mathfrak{P}_{\mathbf{S}}$.*

The functional δ_P is finite, Lipschitz, concave, and positively homogeneous on $\dot{\mathscr{C}}^P_{\mathbf{S}}$. The problem

$$\max_{\boldsymbol{b} \in \dot{\mathscr{C}}^P_{\mathbf{S}}} \left\{ 2\delta_P(\boldsymbol{b}) - \|\boldsymbol{b}\|_P^2 \right\}$$

has a unique solution $\boldsymbol{w} := \dot{\boldsymbol{p}}^{P,\star}$.

Let $a_- := \underline{p}^\star_A(P)$ and $a_+ := \overline{p}^\star_A(P)$. Then $\bar{\theta}$ is pathwise differentiable at P if and only if

$$\delta_P(\boldsymbol{b}) = \begin{cases} \langle \boldsymbol{w}, \boldsymbol{b} \rangle_P & \text{if } \langle \boldsymbol{w}, \boldsymbol{b} \rangle_P \geq 0, \\ \dfrac{a_+}{a_-} \langle \boldsymbol{w}, \boldsymbol{b} \rangle_P & \text{if } \langle \boldsymbol{w}, \boldsymbol{b} \rangle_P < 0 \end{cases} \qquad \forall \boldsymbol{b} \in \dot{\mathscr{C}}^P_{\mathbf{S}}.$$

When this condition holds, the derivative at score s is

$$\dot{\bar{\theta}}_P(s) = \frac{1}{a_-} \left\langle \boldsymbol{w}, \dot{\boldsymbol{p}}^P(s) \right\rangle_P.$$

Proof. Let $H := \dot{\mathscr{C}}^P_{\mathbf{S}}$. For $m \in \mathfrak{M}^\star_P$ and $\boldsymbol{b} \in \mathbb{H}_P$, weighted Cauchy-Schwarz gives

$$\left| \sum_{d,z} \int_{\mathcal{Y}} m(y,d,z) \boldsymbol{b}_{d;(z)}(y) \, d\mu(y) \right| \leq \left(\sum_{d,z} \frac{1}{P_Z(z)} \int_{\mathcal{Y}} m(y,d,z)^2 p^P_{d|z}(y) \, d\mu(y) \right)^{1/2} \|\boldsymbol{b}\|_P$$
$$\leq C_m \left(\sum_z \frac{1}{P_Z(z)} \right)^{1/2} \|\boldsymbol{b}\|_P.$$

The bound is uniform over m. Taking infima in the definition of δ_P therefore gives

$$|\delta_P(\boldsymbol{b}) - \delta_P(\boldsymbol{b}')| \leq C_m \left(\sum_z \frac{1}{P_Z(z)} \right)^{1/2} \|\boldsymbol{b} - \boldsymbol{b}'\|_P.$$

Finiteness follows by taking $\boldsymbol{b}' = \mathbf{0}$. Concavity and positive homogeneity follow by taking the infimum of the same continuous linear functionals: for $\alpha \in [0,1]$ and $t \geq 0$,

$$\delta_P(\alpha \boldsymbol{b} + (1-\alpha)\boldsymbol{b}') \geq \alpha \delta_P(\boldsymbol{b}) + (1-\alpha)\delta_P(\boldsymbol{b}'),$$
$$\delta_P(t\boldsymbol{b}) = t\delta_P(\boldsymbol{b}).$$

For every $m \in \mathfrak{M}^\star_P$, let

$$\boldsymbol{a}_m := \Pi_H \left(\left(\frac{p^P_{d|z}(\cdot) m(\cdot,d,z)}{P_Z(z)} \right)_{z \in \mathcal{Z}} \right)_{d \in \mathcal{D}},$$
$$C := \overline{\operatorname{conv}} \left\{ \boldsymbol{a}_m \,|\, m \in \mathfrak{M}^\star_P \right\},$$

where Π_H is orthogonal projection in $\mathbb{H}_P$ and the closure is in H. The preceding bound shows that C is nonempty, closed, convex, and bounded. For every $\boldsymbol{b} \in H$,

$$\langle \boldsymbol{a}_m, \boldsymbol{b} \rangle_P = \sum_{d,z} \int_{\mathcal{Y}} m(y,d,z) \boldsymbol{b}_{d;(z)}(y)\, d\mu(y),$$

$$\delta_P(\boldsymbol{b}) = \inf_{\boldsymbol{a} \in C} \langle \boldsymbol{a}, \boldsymbol{b} \rangle_P.$$

Let $\boldsymbol{w} := \Pi_C \mathbf{0}$. The Hilbert projection theorem gives $\langle \boldsymbol{a} - \boldsymbol{w}, \boldsymbol{w} \rangle_P \geq 0$ for every $\boldsymbol{a} \in C$, and so

$$\delta_P(\boldsymbol{w}) = \|\boldsymbol{w}\|_P^2, \qquad \delta_P(\boldsymbol{b}) \leq \langle \boldsymbol{w}, \boldsymbol{b} \rangle_P.$$

It follows that

$$\begin{aligned} 2\delta_P(\boldsymbol{b}) - \|\boldsymbol{b}\|_P^2 &\leq 2\langle \boldsymbol{w}, \boldsymbol{b} \rangle_P - \|\boldsymbol{b}\|_P^2 \\ &= \|\boldsymbol{w}\|_P^2 - \|\boldsymbol{b} - \boldsymbol{w}\|_P^2. \end{aligned}$$

Equality holds at $\boldsymbol{b} = \boldsymbol{w}$, proving existence and uniqueness of the asserted maximizer.

Suppose now that $\bar{\theta}$ is pathwise differentiable, with derivative represented by $g \in L_0^2(P)$. The map

$$\boldsymbol{b} \mapsto \mathbb{E}_P \left[g(W) \frac{\boldsymbol{b}_{D;(Z)}(Y)}{p^P_{D|Z}(Y)} \right]$$

is continuous and linear on H, with the quotient defined as zero where its denominator is zero, because

$$\left\| \frac{\boldsymbol{b}_{D;(Z)}(Y)}{p^P_{D|Z}(Y)} \right\|_{L^2(P)} = \|\boldsymbol{b}\|_P.$$

By the Riesz representation theorem, there is $\boldsymbol{a} \in H$ representing this map.

Fix $\boldsymbol{b} \in H$. By Theorem 5.6 and lemmas B.12 and B.14, there are $\boldsymbol{b}_n \in H$ with $\|\boldsymbol{b}_n - \boldsymbol{b}\|_P \to 0$ that satisfy the relative bounds in Lemma B.15. That result supplies two-sided QMD paths with fixed assignment probabilities and scores

$$s_n(y,d,z) := \begin{cases} \dfrac{\boldsymbol{b}_{n,d;(z)}(y)}{p^P_{d|z}(y)} & \text{if } p^P_{d|z}(y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Their endpoint derivatives equal $\mathbb{E}_P[g s_n] = \langle \boldsymbol{a}, \boldsymbol{b}_n \rangle_P$. By Corollary B.2,

$$\delta_P(\boldsymbol{b}_n) = \begin{cases} a_- \langle \boldsymbol{a}, \boldsymbol{b}_n \rangle_P & \text{if } \langle \boldsymbol{a}, \boldsymbol{b}_n \rangle_P \geq 0, \\ a_+ \langle \boldsymbol{a}, \boldsymbol{b}_n \rangle_P & \text{if } \langle \boldsymbol{a}, \boldsymbol{b}_n \rangle_P < 0. \end{cases}$$

Continuity extends this identity to every $\boldsymbol{b} \in H$. Since $a_+ \geq a_-$, it implies

$$2\delta_P(\boldsymbol{b}) - \|\boldsymbol{b}\|_P^2 \leq a_-^2 \|\boldsymbol{a}\|_P^2 - \|\boldsymbol{b} - a_- \boldsymbol{a}\|_P^2,$$

with equality at $\boldsymbol{b} = a_- \boldsymbol{a}$. Uniqueness of the maximizer gives $\boldsymbol{w} = a_- \boldsymbol{a}$, which yields the stated condition.

Conversely, suppose that condition holds. By Corollary B.2, the right derivative along every QMD path with score s is

$$\frac{1}{a_-} \left\langle \boldsymbol{w}, \dot{\boldsymbol{p}}^P(s) \right\rangle_P .$$

For the reversed path the score is $-s$, and so its right derivative is the negative of this expression. The two-sided derivative therefore exists. By Lemma B.11,

$$\left| \frac{1}{a_-} \left\langle \boldsymbol{w}, \dot{\boldsymbol{p}}^P(s) \right\rangle_P \right| \leq \frac{\|\boldsymbol{w}\|_P}{a_-} \|s\|_{L^2(P)}.$$

The derivative is continuous and linear, proving pathwise differentiability. ■

Corollary B.3. *Suppose* $\mathbf{S}$ *satisfies Condition CP,* $\mathcal{R}_{\boldsymbol{\pi}} \neq \emptyset$*, and* P *satisfies Condition D. Use the globally binding row set* $E_{\mathbf{S}}$ *defined in Section 5. If (5.17) holds, then every* $g \in L^2(P)$ *with* $\mathbb{E}_P[g(W) \mid Z] = 0$ *satisfies*

$$\left(p^P_{d|z}(y) g(y, d, z) \right)_{z \in \mathcal{Z}} \in \operatorname{col}(\mathbf{R}_d) \qquad \forall d, y,$$

and there exists $\dot{\boldsymbol{\pi}} \in \ker(\mathbf{A}_{\mathbf{S},E})$ *such that*

$$\mathbf{R}_d \dot{\boldsymbol{\pi}} = \left(\mathbb{E}_P[\mathbb{1}\{D = d\} g(W) \mid Z = z] \right)_{z \in \mathcal{Z}} \qquad \forall d.$$

Under restrictions only on the support of $\boldsymbol{T}$*, (5.17) is equivalent to*

$$\operatorname{rank} \begin{bmatrix} \mathbf{R}_{d_1} \\ \vdots \\ \mathbf{R}_{d_{|\mathcal{D}|}} \end{bmatrix} = 1 + |\mathcal{Z}|(|\mathcal{D}| - 1).$$

Proof. Set

$$\boldsymbol{A} := \mathbf{A}_{\mathbf{S},E},$$

$$\boldsymbol{R} := \begin{bmatrix} \mathbf{R}_{d_1} \\ \vdots \\ \mathbf{R}_{d_{|\mathcal{D}|}} \end{bmatrix},$$

$$V_0 := \left\{ (\boldsymbol{v}_d)_{d \in \mathcal{D}} \in (\mathbb{R}^{\mathcal{Z}})^{\mathcal{D}} \,\middle|\, \sum_d \boldsymbol{v}_d = \mathbf{0} \right\}.$$

Since $\ker \boldsymbol{A} = \operatorname{span}(\mathcal{R}_{\boldsymbol{\pi}} - \mathcal{R}_{\boldsymbol{\pi}})$ and $\mathbf{1}^\intercal \boldsymbol{\pi} = 1$ on $\mathcal{R}_{\boldsymbol{\pi}}$, every $\dot{\boldsymbol{\pi}} \in \ker \boldsymbol{A}$ satisfies $\mathbf{1}^\intercal \dot{\boldsymbol{\pi}} = 0$. Also, $\sum_d \mathbf{R}_d = \mathbf{1}_{\mathcal{Z}} \mathbf{1}_{\mathcal{T}}^\intercal$. Therefore $\boldsymbol{R}(\ker \boldsymbol{A}) \subseteq V_0$. By rank-nullity,

$$\dim \boldsymbol{R}(\ker \boldsymbol{A}) = \dim \ker \boldsymbol{A} - \dim(\ker \boldsymbol{A} \cap \ker \boldsymbol{R})$$

$$= \operatorname{rank} \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{R} \end{bmatrix} - \operatorname{rank} \boldsymbol{A}$$

$$= |\mathcal{Z}|(|\mathcal{D}| - 1) = \dim V_0.$$

Thus $\boldsymbol{R}(\ker \boldsymbol{A}) = V_0$.

For any d and $\boldsymbol{v} \in \mathbb{R}^{\mathcal{Z}}$, choose $e \neq d$. The block vector with blocks $\boldsymbol{v}$ at d, $-\boldsymbol{v}$ at e, and zero elsewhere belongs to V_0. Hence $\operatorname{col}(\mathbf{R}_d) = \mathbb{R}^{\mathcal{Z}}$, which proves the pointwise assertion.

Set

$$\boldsymbol{b}_d := \left(\mathbb{E}_P[\mathbb{1}\{D = d\} g(W) \mid Z = z]\right)_{z \in \mathcal{Z}}.$$

Conditional centring gives $\sum_d \boldsymbol{b}_d = \mathbf{0}$, and so $(\boldsymbol{b}_d)_d \in V_0 = \boldsymbol{R}(\ker \boldsymbol{A})$. This supplies the required $\dot{\boldsymbol{\pi}}$.

Under support restrictions alone, $\mathcal{R}_{\boldsymbol{\pi}}$ is the simplex on the retained types, and so $\operatorname{row}(\boldsymbol{A}) = \operatorname{span}\{\mathbf{1}^\intercal\}$. For every z, summing the zth rows of $\mathbf{R}_d$ over d gives $\mathbf{1}^\intercal$. Thus this row already belongs to $\operatorname{row}(\boldsymbol{R})$, and the final claim follows from

$$\operatorname{rank} \begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{R} \end{bmatrix} - \operatorname{rank} \boldsymbol{A} = \operatorname{rank} \boldsymbol{R} - 1.$$

■

Appendix C: Inference Results and Proofs

C.1. Uniformly Valid Inference on Specification and Parameter-Inclusion Tests

Let $\theta = \frac{\mathsf{N}^Q}{\mathsf{D}^Q}$ be represented as in (†). Let Θ_0, $\Theta_{\mathbf{S}}(P)$ and $\tilde{\Theta}_{\mathbf{S}}(P)$ be defined as in Appendix A.3. In this section, we develop inference procedures that implement the inclusion test proposed in Theorem A.1 and the specification test proposed in Theorem 3.3, and control size uniformly over null distributions in a very broad class of data distributions $\boldsymbol{\mathcal{P}}$, while remaining consistent against any fixed alternative in this class. We will only impose the following minimal requirement on $\boldsymbol{\mathcal{P}}$, which only requires that its assignment probabilities are uniformly bounded away from zero.

Assumption C.1. *The class $\boldsymbol{\mathcal{P}}$ consists of data distributions satisfying Condition D, such that*

$$\inf_{P \in \boldsymbol{\mathcal{P}}} \min_{z \in \mathcal{Z}} P(Z = z) > 0 .$$

Note that when assignment probabilities are known exactly, as in a randomized controlled trial, $\boldsymbol{\mathcal{P}}$ includes *all* data distributions P with those assignment probabilities.

In the remainder of this section, assume that we only have access to a finite sample

$$\{(Y_i, D_i, Z_i)\}_{i=1}^n = \{W_i\}_{i=1}^n \overset{i.i.d.}{\sim} P$$

of independently and identically distributed draws from an unknown distribution $P \in \boldsymbol{\mathcal{P}}$.

C.1.1. *The Test.* We will start from the generalized inclusion test provided in Theorem A.1, which suggests using an infimal convolution statistic based on the criterion $F_P(\boldsymbol{v}, t; \gamma)$, defined in (A.4). As written there, however, $F_P(\boldsymbol{v}, t; \gamma)$ is expressed in terms of the support function of $\mathbf{\Pi}$, which depends on P. We will instead work with the following criterion that replaces $\mathbf{\Pi}$ with $\mathcal{R}_{\boldsymbol{\pi}}$, and is obtained by substituting the expression given in Theorem 3.3 in place of $h_{\mathbf{\Pi}}$, simplifying, and reparametrizing.

Recall that $\mathcal{C} := \prod_{d \in \mathcal{D}} \mathcal{C}_d$, with $\mathcal{C}_d := \left\{ \boldsymbol{u} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}} \,\middle|\, \boldsymbol{u}_{(\boldsymbol{\mathcal{T}} \setminus \boldsymbol{\mathcal{T}}_d)} = \mathbf{0} \right\}$, and the shorthand $\boldsymbol{v} := (\boldsymbol{v}_d)_{d \in \mathcal{D}}$ is used to split $\boldsymbol{v}$ into its $\mathcal{C}_d$ components. For any $\boldsymbol{b} \in \mathbb{R}^{\boldsymbol{\mathcal{T}}}$, denote by $\boldsymbol{b}_{\restriction d}$ the vector generated from $\boldsymbol{b}$ by setting its d-Never-Taker coordinates to 0. Define the criterion

$$L_P(\gamma; \boldsymbol{v}, t) := \sum_{d \in \mathcal{D}} \int_{\mathcal{Y}} h_d \left(t \left(\kappa_d(y) \, \boldsymbol{c}_d - \gamma \varkappa_d(y) \, \boldsymbol{a}_d \right)_{\restriction d} - \boldsymbol{v}_d ; \boldsymbol{p}_d^P(y) \right) d\mu(y) + h_{\mathcal{R}_{\boldsymbol{\pi}}} \left(\sum_{d \in \mathcal{D}} \boldsymbol{v}_d + t(\boldsymbol{c}_\pi - \gamma \boldsymbol{a}_\pi) + \boldsymbol{\zeta}_\gamma(t) \right), \quad \text{(C.1)}$$

where $\boldsymbol{\zeta}_\gamma$ is the never-taker adjustment term defined in Appendix A.3. Now, under the conditions of Theorem A.1, we still have

$$\tilde{\Theta}_{\mathbf{S}}(P) = \left\{ \gamma \in \Theta_0 \,\middle|\, \inf_{\substack{\boldsymbol{v} \in \mathcal{C}, t \in \mathbb{R} \\ \|(\boldsymbol{v}, t)\|_2 \leq 1}} L_P(\gamma; \boldsymbol{v}, t) \geq 0 \right\}.$$

Further, we also have, by Theorem 3.3,

$$\mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset \iff \inf_{\substack{\boldsymbol{v} \in \mathcal{C} \\ \|\boldsymbol{v}\|_2 \leq 1}} L_P(\boldsymbol{v}, 0) \geq 0,$$

where I note that L_P does not depend on γ when $t = 0$ and so omit it above. Therefore, the specification test can be viewed as a specialization of the same criterion.

C.1.2. *The Test Statistic and Critical Value.* In moment form, we can write

$$L_P(\gamma; \boldsymbol{v}, t) := \sum_{z \in \mathcal{Z}} \mathbb{E}_P \left[\boldsymbol{m}_{L(\theta)}^P(\gamma; \boldsymbol{v}, t; W) \,\middle|\, Z = z \right], \quad \text{(C.2)}$$

where

$$\boldsymbol{m}_{L(\theta)}^P(\gamma; \boldsymbol{v}, t; Y, D, Z) := \sum_{d \in \mathcal{D}} \boldsymbol{m}_d^P \left(t \left(\kappa_d(Y) \, \boldsymbol{c}_d - \gamma \varkappa_d(Y) \, \boldsymbol{a}_d \right)_{\restriction d} - \boldsymbol{v}_d ; Y, D, Z \right) + \frac{1}{|\mathcal{Z}|} h_{\mathcal{R}_{\boldsymbol{\pi}}} \left(\sum_{d \in \mathcal{D}} \boldsymbol{v}_d + t(\boldsymbol{c}_\pi - \gamma \boldsymbol{a}_\pi) + \boldsymbol{\zeta}_\gamma(t) \right).$$

Partition the sample into training and evaluation folds $I_T, I_E \subseteq [n]$ and define $n^{(k)} := |I_k|$ and $n_z^{(k)} := |\{ i \in I_k \,|\, Z_i = z \}|$, for $k \in \{T, E\}$. As before, denote the empirical distribution by $\hat{P}_n$ and denote by $\check{P}_n$ a (potentially regularized) estimator of $\hat{P}_n$ such that $\check{P}_n$ induces a μ-dominated distribution of Y; in the remainder, we will generally suppress their dependence

on n. Let $\check{P}^{(k)}$ denote the version of $\check{P}$ estimated on fold k, and extend the remaining notation accordingly. In the remainder, I assume that the folds are selected independently of the data, that $|I_k| > 0$ for each $k \in \{T, E\}$, and $\frac{n^{(T)}}{n} \to c_T$ for some $c_T \in (0,1)$, with $\frac{n^{(T)}}{n} + \frac{n^{(E)}}{n} = 1$.

Now, let $\check{\boldsymbol{m}}_{L(\theta)}$ be the version of $\boldsymbol{m}^P_{L(\theta)}$ obtained by replacing $\boldsymbol{m}^P_d$ with $\check{\boldsymbol{m}}_d$, where, for each $d \in \mathcal{D}$, $\check{\boldsymbol{m}}_d$ is the training fold estimated moment function,

$$\check{\boldsymbol{m}}_d = \check{\boldsymbol{m}}_d^{(T)} := \boldsymbol{m}_d^{\check{P}^{(T)}} ,$$

which we will assume is obtained via Algorithm 3.1.[38] Next, letting $\boldsymbol{w} := (\boldsymbol{v}, t)$, define the evaluation fold quantities

$$\bar{\boldsymbol{m}}_z(\gamma; \boldsymbol{w}) := \frac{1}{n_z^{(E)}} \sum_{i \in I_E : Z_i = z} \check{\boldsymbol{m}}_{L(\theta)}(\gamma; \boldsymbol{w}; W_i) , \qquad \hat{L}_n(\gamma; \boldsymbol{w}) := \sum_{z \in \mathcal{Z}} \bar{\boldsymbol{m}}_z(\gamma, \boldsymbol{w}) ,$$

$$\hat{\psi}_i(\gamma; \boldsymbol{w}) := \frac{n^{(E)}}{n_{Z_i}^{(E)}} \left(\check{\boldsymbol{m}}_{L(\theta)}(\gamma; \boldsymbol{w}; W_i) - \bar{\boldsymbol{m}}_{Z_i}(\gamma; \boldsymbol{w}) \right) , \qquad \hat{\sigma}^2(\gamma; \boldsymbol{w}) := \frac{1}{n^{(E)}} \sum_{i \in I_E} \hat{\psi}_i(\gamma; \boldsymbol{w})^2$$

and, for i.i.d. multipliers $\{\xi_i\}_{i \in I_E} \sim N(0,1)$ drawn independently of the data, let

$$G_n^*(\gamma; \boldsymbol{w}) := \frac{1}{\sqrt{n^{(E)}}} \sum_{i \in I_E} \xi_i \hat{\psi}_i(\gamma; \boldsymbol{w}) .$$

Define $\mathscr{V} := \{\boldsymbol{w} \in \mathcal{C} \times \mathbb{R} \,|\, \|\boldsymbol{w}\|_2 \leq 1\}$ and, for any $\tau_n := cn^{-\beta}$ with $c > 0$ and $0 < \beta < \frac{1}{2}$, define the two "screening" sets,

$$\hat{\mathscr{V}}_{n,\gamma}^{\text{cal}} := \{\boldsymbol{w} \in \mathscr{V} \,|\, \hat{\sigma}(\gamma; \boldsymbol{w}) > \tau_n (1 + |\gamma|) \|\boldsymbol{w}\|_2\} ,$$
$$\hat{\mathscr{V}}_{n,\gamma}^{\text{test}} := \{\boldsymbol{w} \in \mathscr{V} \,|\, \hat{\sigma}(\gamma; \boldsymbol{w}) > 2\tau_n (1 + |\gamma|) \|\boldsymbol{w}\|_2\} ,$$

where the factor $(1 + |\gamma|)$ is included to accommodate unbounded Θ_0, but can be omitted when Θ_0 is compact.

For $\alpha \in (0,1)$, define the critical value $k_{n,\gamma}(1-\alpha)$ as the (conditional on both folds) $1-\alpha$ quantile of[39]

$$\sup_{\boldsymbol{w} \in \hat{\mathscr{V}}_{n,\gamma}^{\text{cal}}} \frac{-G_n^*(\gamma; \boldsymbol{w})}{\hat{\sigma}(\gamma; \boldsymbol{w})} .$$

[38] This fact is important for the empirical process complexity control result used in the proof of Theorem C.1, but other constructions could be accommodated, provided a suitable empirical process complexity control argument is applicable.

[39] Set $k_{n,\gamma}(1-\alpha) = \infty$ whenever $\hat{\mathscr{V}}_{n,\gamma}^{\text{cal}} = \emptyset$.

Consider the inclusion test $\phi_n(\gamma)$ defined by

$$\phi_n(\gamma) := \begin{cases} 1 \text{ if } \sup_{\boldsymbol{w} \in \hat{\mathscr{V}}^{\text{test}}_{n,\gamma}} \dfrac{-\sqrt{n^{(E)}}\hat{L}_n(\gamma;\boldsymbol{w})}{\hat{\sigma}(\gamma;\boldsymbol{w})} > k_{n,\gamma}(1-\alpha) \ , \\ 1 \text{ if } \inf_{\boldsymbol{w} \in \mathscr{V}} \left\{ \hat{L}_n(\gamma;\boldsymbol{w}) + \dfrac{\log\left(n^{(E)}\right)}{\sqrt{n^{(E)}}}(1+|\gamma|)\|\boldsymbol{w}\|_2 \right\} < 0 \ , \\ 0 \text{ otherwise.} \end{cases}$$

Define the ϕ_n^{spec} as the specification test defined analogously using $\hat{L}_n(\boldsymbol{v},0)$ in place of $\hat{L}_n(\gamma;\boldsymbol{v},t)$ and imposing $\gamma = t = 0$ throughout the construction just specified, including in the rejection criteria, critical value, and the definitions of $\mathscr{V}$, $\hat{\mathscr{V}}^{\text{cal}}_n$ and $\hat{\mathscr{V}}^{\text{test}}_n$.[40]

We now have the following result, which establishes the uniform validity and fixed-alternative consistency of these two procedures.

Theorem C.1. *Suppose* $\mathbf{S}$ *satisfies Condition C and* $\boldsymbol{\mathcal{P}}$ *satisfies Assumption C.1. Then, the following statements hold, for any given* $\alpha \in (0,1)$, $c > 0$, *and* $\beta \in (0, 1/2)$.

(i) *(Specification Test.)*

a) $\limsup_{n\to\infty} \sup_{\substack{P \in \boldsymbol{\mathcal{P}}: \\ \mathcal{Q}_{\mathbf{S}}(P) \neq \emptyset}} P(\phi_n^{\text{spec}} = 1) \leq \alpha$ *and,*

b) *whenever* $P \in \boldsymbol{\mathcal{P}}$ *satisfies* $\Delta\left(\check{P}^{(T)}, P\right) \xrightarrow{P} 0$ *and* $\mathcal{Q}_{\mathbf{S}}(P) = \emptyset$,

$$P(\phi_n^{\text{spec}} = 1) \to 1 \ .$$

(ii) *(Parameter-Inclusion Test.) Suppose, in addition, that* θ *satisfies Condition T3 and either Condition T1, or Condition T2 together with Assumption A.1 for each* $P \in \boldsymbol{\mathcal{P}}$.[41] *Then,*

a) $\limsup_{n\to\infty} \sup_{\substack{P \in \boldsymbol{\mathcal{P}} \\ \gamma \in \tilde{\Theta}_{\mathbf{S}}(P)}} P(\phi_n(\gamma) = 1) \leq \alpha$ *and,*

b) *whenever* $P \in \boldsymbol{\mathcal{P}}$ *satisfies* $\Delta\left(\check{P}^{(T)}, P\right) \xrightarrow{P} 0$ *and* $\gamma \in \Theta_0 \setminus \tilde{\Theta}_{\mathbf{S}}(P)$,

$$P(\phi_n(\gamma) = 1) \to 1 \ .$$

Appendix D: Equivalence of Structural and Potential Outcomes IV Models

The potential outcomes framework with a randomly assigned instrument, is summarized by the following condition on P.

Condition PO (Potential Outcomes IV Model)**.** *There exists random variables* $((Y_d : d \in \mathcal{D}), (D_z : z \in \mathcal{Z}))$ *that satisfy the following properties:*

[40] If $n_z^{(E)} = 0$ for any $z \in \mathcal{Z}$, then define both tests as failing to reject.

[41] Note that this non-zero denominator condition, only needed for effects ratios, allows for the separation of the denominator from 0 to depend on P, it need not be uniform over $\boldsymbol{\mathcal{P}}$.

(i) *Almost surely,*

$$D = \sum_{z \in \mathcal{Z}} \mathbb{1}\{Z = z\} D_z \text{ and } Y = \sum_{d \in \mathcal{D}} \mathbb{1}\{D = d\} Y_d . \tag{D.1}$$

(ii) *For each $d \in \mathcal{D}$, Y_d has a density with respect to μ.*

(iii) $((Y_d : d \in \mathcal{D}), (D_z : z \in \mathcal{Z})) \perp Z$

In fact, this model is equivalent to the model of Condition M, in the sense formalized in the following result.

Proposition D.1. *Suppose P satisfies Condition D. Then, the following statements hold:*

(i) *Condition PO holds if and only if Condition M holds.*

(ii) *Given any random variables $\boldsymbol{V}$ and $\boldsymbol{\varepsilon}_Y$, and maps g_D and g_Y such that $(\boldsymbol{V}, \boldsymbol{\varepsilon}_Y, g_D, g_Y)$ witnesses Condition M, $((g_Y(d, \boldsymbol{V}, \boldsymbol{\varepsilon_Y}) : d \in \mathcal{D}), (g_D(z, \boldsymbol{V}) : z \in \mathcal{Z}))$ witnesses Condition PO.*

(iii) *Given any $((Y_d : d \in \mathcal{D}), (D_z : z \in \mathcal{Z}))$ that witnesses Condition PO, there exists random variables $\boldsymbol{V}$ and $\boldsymbol{\varepsilon}_Y$, and maps g_D and g_Y such that*

$$\text{Law}((g_Y(d, \boldsymbol{V}, \boldsymbol{\varepsilon_Y}) : d \in \mathcal{D}), (g_D(z, \boldsymbol{V}) : z \in \mathcal{Z})) = \text{Law}((Y_d : d \in \mathcal{D}), (D_z : z \in \mathcal{Z})) ,$$

and $(\boldsymbol{V}, \boldsymbol{\varepsilon}_Y, g_D, g_Y)$ witnesses Condition M.

Proof. Part (ii) follows from (2.4), (2.5), and the independence and μ-domination requirements imposed by Condition M. For (iii), set $\boldsymbol{V} := \boldsymbol{D} = (D_z)_{z \in \mathcal{Z}}$. As in the proof of Theorem 2.1, take mutually independent random vectors

$$\boldsymbol{e}(\boldsymbol{v}) \sim \text{Law}((Y_d : d \in \mathcal{D}) \mid \boldsymbol{V} = \boldsymbol{v}), \qquad \boldsymbol{v} \in \text{supp}(\boldsymbol{V}),$$

jointly independent of $(\boldsymbol{V}, Z)$. Define $\boldsymbol{\varepsilon}_Y := (\boldsymbol{e}(\boldsymbol{v}))_{\boldsymbol{v} \in \text{supp}(\boldsymbol{V})}$ and

$$g_D(z, \boldsymbol{v}) := \boldsymbol{v}_{(z)},$$
$$g_Y(d, \boldsymbol{v}, \boldsymbol{e}') := \left(\boldsymbol{e}'_{(\boldsymbol{v})}\right)_{(d)} .$$

By Condition PO and construction,

$$(\boldsymbol{e}(\boldsymbol{V}), \boldsymbol{V}, Z) \overset{\text{d}}{=} ((Y_d : d \in \mathcal{D}), \boldsymbol{D}, Z).$$

The maps are measurable and $Z, \boldsymbol{V}, \boldsymbol{\varepsilon}_Y$ are mutually independent. Now, the structural equations generate $((Y_d : d \in \mathcal{D}), \boldsymbol{D}, Z)$, and hence generates P and satisfies the required μ-domination, proving (iii). Part (i) follows immediately from (ii)–(iii). ■

Appendix E: Computational Appendix

<coming soon>