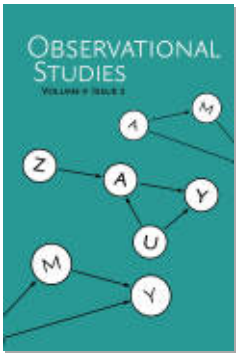
ivmte: An R Package for Extrapolating Instrumental Variable Estimates Away From Compliers*



Joshua Shea, Alexander Torgovitsky

Observational Studies, Volume 9, Issue 2, 2023, pp. 1-42 (Article)

Published by University of Pennsylvania Press

Observational Studies 9.2 (2023) 1-42 Submitted Jun 16 2022; Published Mar 2023

`ivmte`: An R Package for Extrapolating Instrumental Variable Estimates Away From Compliers*

Joshua Shea **jkcshea@illinois.edu**
Department of Economics
University of Illinois, Urbana-Champaign
Urbana, IL 61801 USA

Alexander Torgovitsky **torgovitsky@uchicago.edu**
Kenneth C. Griffin Department of Economics
University of Chicago
Chicago, IL 60637 USA

Abstract

Instrumental variable (IV) strategies are widely used to estimate causal effects in economics, political science, epidemiology, sociology, psychology, and other fields. When there is unobserved heterogeneity in causal effects, standard linear IV estimators only represent effects for complier subpopulations (Imbens and Angrist, 1994). Marginal treatment effect (MTE) methods (Heckman and Vytlacil, 1999, 2005) allow researchers to use additional assumptions to extrapolate beyond complier subpopulations. We discuss a flexible framework for MTE methods based on linear regression and the generalized method of moments. We show how to implement the framework using the `ivmte` package for `R`.

Keywords: instrumental variables, marginal treatment effects, local average treatment effect, partial identification

1. Introduction

A central task in many empirical fields is to determine the effect (the *causal* effect) of one variable on another. The task is often complicated by the fact that the effecting variable (the treatment) is not only not randomly assigned, it is *chosen* by an agent with information unavailable to the researcher. For example, in the application discussed later, the treatment is the number of children a family decides to have, and the empirical question is the effect that bearing more children has on the mother's labor force participation. Since having a child and working are joint decisions a family makes using their own private information, strategies such as propensity score matching are unlikely to eliminate systematic unobserved differences between families with more children and those with fewer children. Different empirical strategies are needed to credibly identify a causal effect.

Instrumental variables (IVs) are one extremely popular strategy (e.g. Heckman and Robb, 1985; Bollen, 2012; Baiocchi et al., 2014; Imbens, 2014). An IV (or instrument) is an observed variable that is correlated with the treatment variable, but uncorrelated with confounding

unobservable differences. A well-known example of an instrument for fertility is the same-sex instrument introduced by Angrist and Evans (1998). This instrument is a binary variable that is equal to 1 if a family's first two children had the same sex (female-female or male-male) and is 0 otherwise.[1] The key assumption of an IV model is that the sex of the second child is as good as randomly assigned—and therefore independent of any confounding unobservable differences across families—while still impacting a family's decision to have a third child due to a preference for having both a male and female child. The intuition is that by comparing the labor supply decisions of families whose first two children were the same sex to families whose children were mixed sex, one picks up only the differences that are caused by the decision to have additional children.

While IV strategies have been widely studied and applied for many decades (see Stock and Trebbi, 2003, for a history), it wasn't until the 1980s that researchers started to focus on IV models with unobserved heterogeneity in treatment effects.[2] In an influential paper, Imbens and Angrist (1994) provided nonparametric conditions under which a simple linear IV estimator can be interpreted as estimating the average causal effect (the "local average treatment effect," or LATE) among a subpopulation described as the compliers. The compliers are the individuals whose treatment choice would have been different had their instrument been different. In the fertility application, they are the families who would have had a third child if and only if their first two children had the same sex.

An important implication is that the interpretation of a linear IV estimator depends on the instrument used. If there is unobserved treatment effect heterogeneity, then linear IV estimators cannot in general be interpreted as providing estimates of conventional parameters such as the average treatment effect (ATE) or the average treatment effect on the treated (ATT). One response to this finding is to continue to use linear IV estimators and change the research question to serve the definition of the complier group, a practice espoused by Angrist and Krueger (2001) and Angrist and Pischke (2009, 2010). A number of authors in multiple disciplines have criticized this practice (e.g. Robins and Greenland, 1996; Heckman, 1997; Deaton, 2010; Pearl, 2011; Swanson and Hernán, 2014, among many others).[3] Another response is to change the estimator and extrapolate from the compliers to the subpopulation that better answers the researchers' empirical question (see Mogstad and Torgovitsky, 2018, for a discussion of different approaches).

1. Using the same-sex instrument requires restricting the analysis to families with two or more children.
2. An early example is Heckman (1976). See also Heckman and Robb (1985), Björklund and Moffitt (1987), and Manski (1990). Taking a more expansive view of unobserved heterogeneity in "treatment effects" to also encompass random coefficient models, one can trace interest back to the Cowles Foundation (Hurwicz, 1950; Rubin, 1950) as well as foundational economic analyses like Becker and Chiswick (1966).
3. More recently, Blandhol et al. (2022) have argued that the LATE interpretation typically does not even apply to the types of linear IV specifications used in practice.

In a series of papers, Heckman and Vytlacil (1999, 2005, 2007a,b) developed the concept of the marginal treatment effect (MTE) and showed how it can be used to nonparametrically model this type of extrapolation under the same "monotonicity" condition used by Imbens and Angrist (1994). Carneiro et al. (2011) and Brinch et al. (2012, 2017) showed how to apply their idea to identify and estimate semiparametric MTE models. These methods have now been applied across a wide range of topics in empirical economics including the returns to schooling (Moffitt, 2008; Carneiro et al., 2011, 2016; Nybom, 2017; Heinesen and Stenholt Lange, 2022), and its impacts on wage inequality (Carneiro and Lee, 2009), discrimination (Arnold et al., 2018, 2020), the returns to citizenship (Gathmann et al., 2021), and to agricultural technology (Mellon Bedi et al., 2021; Sarr et al., 2021), the effects of foster care (Doyle Jr., 2007), the impacts of welfare (Moffitt, 2019) and disability insurance (Maestas et al., 2013; French and Song, 2014; Autor et al., 2019) programs on labor supply, the performance of charter schools (Walters, 2018), health care (Kowalski, 2018; Depalo, 2020), marketing (Daljord et al., 2021), nonresponse bias in social surveys (Dutz et al., 2021), the effects of early childhood programs (Kline and Walters, 2016; Cornelissen et al., 2018; Felfe and Lalive, 2018), the efficacy of preventative health products (Mogstad et al., 2017), the quantity–quality theory of fertility (Brinch et al., 2017), the demand for electricity (Ito et al., 2022), and the effects of fines (Goncalves and Mello, 2022; Possebom, 2022), misdemeanor prosecution (Agan et al., 2021), and incarceration (Bhuller et al., 2020; Rose and Shem-Tov, 2021), among many others.

Mogstad et al. (2018) developed a general moment-based framework for implementing MTE approaches that allows for partial identification (bounds) in cases when the researcher's assumptions are not strong enough (or the data is not rich enough) to pin down a unique conclusion.[4] In this paper, we discuss implementation of this framework, as well as a related regression-based framework, that minimizes a least squares criterion.[5] Instead of focusing on fitting specific moments, the regression framework minimizes a least squares criterion. As we discuss, this has both benefits and drawbacks that depend on the researcher's empirical setting and goals. We then describe the `R` package `ivmte`, which can be used to implement both the moment and regression frameworks. The package provides a flexible environment for using IV strategies to conduct rigorous policy evaluation in the presence of unobserved heterogeneity.

4. Rose and Shem-Tov (2021) extended and applied the framework in their study of the impact of incarceration on recidivism.
5. Brinch et al. (2012, 2017) introduced the regression-based framework for point-identified cases.

2. Model and identification

2.1 Potential outcomes and choices

The model is about the impact of a binary treatment $D_i \in \{0,1\}$ on individual i's observed outcome variable, Y_i. Let $Y_i(0)$ and $Y_i(1)$ denote the unobserved potential outcomes for Y_i if individual i had received $D_i = 0$ or 1, respectively, so that $Y_i = D_i Y_i(1) + (1 - D_i)Y_i(0)$. The researcher is interested in features of the distribution of the causal effect, $Y_i(1) - Y_i(0)$. The researcher has access to some observable covariates, X_i, but they are concerned that D_i is still dependent with $Y_i(0)$ or $Y_i(1)$ even after conditioning on X_i, so that the unconfoundedness (selection on observables) assumption (e.g. Barnow et al. (1980), Rosenbaum and Rubin (1983), Heckman et al. (1996)) does not hold.

However, the researcher also has access to an instrumental variable, Z_i. The instrument influences individual i's treatment choice with $D_i(z)$ denoting their unobserved potential treatment choice if Z_i were set to z. Their observed treatment choice is related to these potential choices via $D_i = \sum_{z \in \mathcal{Z}} \mathbb{1}[Z_i = z] D_i(z)$, where $\mathcal{Z}$ is the support of Z_i. In contrast to D_i, the instrument is assumed to be as good as randomly assigned, conditional on X_i, in the sense that Z_i is independent of $(Y_i(0), Y_i(1), \{D_i(z)\}_{z \in \mathcal{Z}})$, conditional on X_i. Given that the potential outcomes $Y_i(d)$ are indexed by d only, this assumption also implies that the instrument also has no direct causal effect on Y_i, an assumption typically referred to as the exclusion restriction.

2.2 The selection model

Imbens and Angrist (1994) introduced an additional assumption that they described as monotonicity. The monotonicity assumption says that for any pair of instrument values z and z', either $D_i(z) \geq D_i(z')$ for all individuals i, or else $D_i(z') \geq D_i(z)$ for all individuals i. That is, a shift from z to z' either pushes every individual towards treatment, or else pushes every individual away from treatment. Whichever direction holds, the monotonicity condition maintains that there are no individuals who deviate from this ordering, a requirement sometimes described as "no defiers."[6]

Vytlacil (2002) showed that the monotonicity condition is equivalent to the latent variable selection model

$$D_i = \mathbb{1}[U_i \leq p(X_i, Z_i)], \tag{1}$$

6. Despite the name "monotonicity," the condition would be more accurately described as "uniformity," since it restricts heterogeneity in how the instrument impacts treatment choice (Heckman et al., 2006; Mogstad et al., 2021).

where U_i is a continuously distributed unobserved random variable, and $p(x,z) \equiv \mathbb{P}[D_i = 1|X_i = x, Z_i = z]$ is the propensity score. The latent variable U_i is independent of Z_i, conditional on X_i, and is customarily normalized to be uniformly distributed on $[0,1]$.[7] It can be interpreted as the individual i's rank (quantile) of latent willingness to choose $D_i = 1$, with smaller values of U_i corresponding to more-willing individuals. When D_i is a variable chosen by an agent, such as in the fertility example, we expect that U_i will be dependent with $Y_i(0)$ and $Y_i(1)$ if these potential outcomes themselves either directly influence the agent's choice or are correlated with other factors that do.

2.3 Marginal treatment response and effect functions

The advantage of the latent variable model (1) is that it facilitates modeling unobserved heterogeneity in the effect of D_i on Y_i. The key object for this purpose is the marginal treatment response (MTR) function

$$m(d|u,x) \equiv \mathbb{E}[Y_i(d)|U_i = u, X_i = x] \tag{2}$$

The MTR function describes how expected treated and untreated outcomes vary conditional on both observed covariates, X_i, and the unobserved latent propensity to take treatment, U_i. The marginal treatment effect (MTE) of Heckman and Vytlacil (1999, 2005, 2007a,b) is the difference of the MTR function between treatment states: $m(1|u,x) - m(0|u,x)$. For example, if the MTE is declining in u, then individuals who are less likely to choose treatment (larger u) would experience smaller treatment effects than those who are more likely to choose treatment. Thus, the MTE captures the idea of selection on *unobservables*, where the unobservable in question is an individual's latent propensity to take treatment, U_i.

2.4 Target parameters

Many treatment effect parameters can be written as weighted averages of the MTR function. For example, the average treatment effect (ATE) can be written as

$$\begin{aligned}\underbrace{\mathbb{E}[Y_i(1) - Y_i(0)]}_{\equiv \text{ATE}} &= \mathbb{E}\left[\mathbb{E}[Y_i(1)|U_i, X_i] - \mathbb{E}[Y_i(0)|U_i, X_i]\right] \\ &= \mathbb{E}[m(1|U_i, X_i) - m(0|U_i, X_i)] = \mathbb{E}\left[\int_0^1 m(1|u, X_i) - m(0|u, X_i)\, du\right],\end{aligned} \tag{3}$$

7. See Heckman and Vytlacil (2005), Matzkin (2007), or Mogstad and Torgovitsky (2018) for a detailed discussion of the normalization.

where the final equality used the normalization on the distribution of U_i to be uniform and independent of X_i. Similarly, the average treatment effect on the treated (ATT) can be written as

$$\underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]}_{\equiv \text{ATT}} = \mathbb{E}\left[\int_0^1 (m(1|u, X_i) - m(0|u, X_i)) \times \frac{\mathbb{1}\left[u \leq p(X_i, Z_i)\right]}{\mathbb{P}[D_i = 1]}\, du\right], \quad (4)$$

see e.g. Heckman and Vytlacil (2005). As in Mogstad et al. (2018), we view both (3) and (4) as examples of *target parameters* τ with the general form

$$\tau(m) \equiv \sum_{d \in \{0,1\}} \mathbb{E}\left[\int_0^1 m(d|u, X_i)\omega_\tau(d|u, X_i, Z_i)\, du\right], \quad (5)$$

where ω_τ is a weighting function that is either known to the researcher (as in (3)) or point identified from the distribution of (D_i, X_i, Z_i) (as in (4)). Heckman and Vytlacil (2005) and Mogstad et al. (2018) provide extensive discussions and many examples of target parameters, along with their weighting functions, ω_τ.

2.5 Implied observable quantities

The model implies a relationship between the MTR function and moments of the observed outcome, Y_i. In particular, Mogstad et al. (2018, Proposition 1) show that for any (measurable) function s of (D_i, X_i, Z_i)

$$\begin{aligned}\mathbb{E}[Y_i s(D_i, X_i, Z_i)] = \quad & \mathbb{E}\left[s(0, X_i, Z_i)\int_0^1 m(0|u, X_i)\mathbb{1}[u > p(X_i, Z_i)]\, du\right] \\ & + \mathbb{E}\left[s(1, X_i, Z_i)\int_0^1 m(1|u, X_i)\mathbb{1}[u \leq p(X_i, Z_i)]\, du\right] \equiv \gamma_s(m). \quad (6)\end{aligned}$$

A similar expression can be derived for the conditional moments of Y_i:

$$\begin{aligned}&\mathbb{E}[Y_i|D_i = 1, X_i = x, Z_i = z] \\ &\quad = \mathbb{E}[Y_i(1)|U_i \leq p(x, z), X_i = x] = \int_0^1 m(1|u, x)\frac{\mathbb{1}[u \leq p(x, z)]}{p(x, z)}\, du,\end{aligned}$$

and, symmetrically,

$$\mathbb{E}[Y_i|D_i = 0, X_i = x, Z_i = z] = \int_0^1 m(0|u, x)\frac{\mathbb{1}[u > p(x, z)]}{(1 - p(x, z))},\ du.$$

We combine the right-hand side of these two relationships using the notation

$$\gamma_{\text{cm}}(m|d,x,z) = \int_0^1 m(0|u,x)(1-d)\frac{\mathbb{1}[u > p(x,z)]}{(1-p(x,z))} + m(1|u,x)d\frac{\mathbb{1}[u \leq p(x,z)]}{p(x,z)}\,du. \quad (7)$$

2.6 Identification

We use expressions (6) and (7) to define two identified sets for the target parameter, τ. To do this, we first define identified sets for the MTR function. We assume that m lives in some set $\mathcal{M}$ contained in a vector space, where $\mathcal{M}$ encodes any additional assumptions we might want to place on m, such as parameterizations or shape restrictions.

One identified set matches a collection of unconditional moments (6) formed by a collection of functions $s \in \mathcal{S}$:

$$\mathcal{M}^\star_{\mathcal{S}} = \{m \in \mathcal{M} : \gamma_s(m) = \mathbb{E}[Y_i s(D_i, X_i, Z_i)] \text{ for all } s \in \mathcal{S}\}\,. \quad (8)$$

The moment approach is based on $\mathcal{M}^\star_{\mathcal{S}}$. Another identified set matches the conditional mean of the observed outcome:

$$\mathcal{M}^\star_{\text{cm}} = \{m \in \mathcal{M} : \gamma_{\text{cm}}(m|d,x,z) = \mathbb{E}[Y_i|D_i = d, X_i = x, Z_i = z] \text{ for almost every } d,x,z\}\,. \quad (9)$$

The regression approach is based on $\mathcal{M}^\star_{\text{cm}}$. If m satisfies the set of conditional moment equalities in $\mathcal{M}^\star_{\text{cm}}$, then it also satisfies the unconditional moment equalities in $\mathcal{M}^\star_{\mathcal{S}}$ for any choice of $\mathcal{S}$, and thus $\mathcal{M}^\star_{\text{cm}} \subseteq \mathcal{M}^\star_{\mathcal{S}}$. However, there are some practical, statistical, and conceptual considerations that may nevertheless favor the moment approach (see Section 3.4).

Our object of interest is not an identified set for the MTR function, but rather an identified set for the target parameter, τ. An identified set for the target parameter can be formed by taking the image of either $\mathcal{M}^\star_{\mathcal{S}}$ or $\mathcal{M}^\star_{\text{cm}}$ under τ:

$$\mathcal{T}^\star_{\mathcal{S}} \equiv \{\tau(m) : m \in \mathcal{M}^\star_{\mathcal{S}}\} \quad \text{and} \quad \mathcal{T}^\star_{\text{cm}} \equiv \{\tau(m) : m \in \mathcal{M}^\star_{\text{cm}}\}\,. \quad (10)$$

The set $\mathcal{T}^\star_{\mathcal{S}}$ gives the values of the target parameter that are consistent with the assumptions of the model and the unconditional moments (6) for $s \in \mathcal{S}$. The set $\mathcal{T}^\star_{\text{cm}}$ is the subset of $\mathcal{T}^\star_{\mathcal{S}}$ that is consistent with the entire conditional mean of the observed outcome.

2.7 Point identification vs. partial identification

The formulation in the previous section allows the identified sets $\mathcal{M}^\star_{\mathcal{S}}, \mathcal{M}^\star_{\text{cm}}$, $\mathcal{T}^\star_{\mathcal{S}}$, and $\mathcal{T}^\star_{\text{cm}}$ to be either singletons or proper non-singleton sets. In the first case, we say that m or τ is point identified, while in the second case we say that they are partially identified. Point identification of m implies point identification of τ. When τ is not point identified, its identified sets $\mathcal{T}^\star_{\mathcal{S}}$ and $\mathcal{T}^\star_{\text{cm}}$ will still be closed intervals under weak conditions (see Mogstad et al., 2018, for a precise statement). One can thus describe the partial identification case as providing bounds on the target parameter. It is also possible for the identified sets to be empty, in which case the model is said to be misspecified.

The size and cardinality of the identified sets depend on a few factors. Having a smaller parameter space $\mathcal{M}$—that is, maintaining more restrictive assumptions—mechanically shrinks the identified sets. Making $\mathcal{S}$ a larger set of functions also mechanically shrinks the moment-based identified set. The number of distinct functions one can potentially include in $\mathcal{S}$ is determined by the supports of Z_i and X_i. Richer supports of Z_i allow for smaller identified sets and thus tighter conclusions; richer supports of X_i can also be helpful if $\mathcal{M}$ is such that $m(d|u,x)$ depends on x in a restricted way. These richer supports get automatically incorporated into the conditional mean identified set $\mathcal{T}^\star_{\text{cm}}$, so that it necessarily shrinks with additional support points.

2.8 Criterion functions

For implementation, it is useful to have a scalar function that determines if a candidate MTR function m is in either $\mathcal{M}^\star_{\mathcal{S}}$ or $\mathcal{M}^\star_{\text{cm}}$.

For the moment approach, we let $c_s \equiv \mathbb{E}[Y_i s(D_i, X_i, Z_i)]$, and stack both c_s and $\gamma_s(m)$ across $s \in \mathcal{S}$ into vectors $c_{\mathcal{S}}$ and $\gamma_{\mathcal{S}}(m)$. Define

$$Q_{\mathcal{S}}(m) = \|\gamma_{\mathcal{S}}(m) - c_{\mathcal{S}}\|,$$

for some choice of norm $\|\cdot\|$. Then $m \in \mathcal{M}^\star_{\mathcal{S}}$ if and only if $Q_{\mathcal{S}}(m) = 0$, so that

$$\mathcal{T}^\star_{\mathcal{S}} = \{\tau(m) : Q_{\mathcal{S}}(m) = 0\}.$$

The minimum value of $Q_{\mathcal{S}}(m)$ over $m \in \mathcal{M}$ can be used as the basis for a specification test; if the model is correctly specified, then it should be 0.

For the regression approach, we define the least squares criterion:

$$Q_{\text{cm}}(m) \equiv \mathbb{E}\left[\left(Y_i - \gamma_{\text{cm}}(m|D_i, X_i, Z_i)\right)^2\right].$$

If $\mathcal{M}^\star_{\text{cm}}$ is not empty—that is, if the model is correctly specified—then $m \in \mathcal{M}^\star_{\text{cm}}$ if and only if $m \in \arg\min_{m' \in \mathcal{M}} Q_{\text{cm}}(m')$, so that

$$\mathcal{T}^\star_{\text{cm}} = \left\{ \tau(m) : Q_{\text{cm}}(m) = \min_{m' \in \mathcal{M}} Q_{\text{cm}}(m') \right\}.$$

Unlike $\mathcal{T}^\star_{\mathcal{S}}$, which can be empty, $\mathcal{T}^\star_{\text{cm}}$ is necessarily non-empty, reflecting the fact that the minimum of the least squares criterion cannot be used as the basis for a specification test.

3. Estimation and computation

3.1 Linear basis representation

Implementation requires evaluating the functions τ and γ_s or γ_{cm} at candidate choices of the MTR function. Some dimension reduction is needed for computation. Let

$$m_\theta(d|u,x) = \sum_{k=1}^{K} \theta_k b_k(d|u,x) \quad \text{for some } \theta \in \mathbb{R}^K, \tag{11}$$

where b_k are known basis functions and θ_k are unknown parameters. We assume that the parameter space is $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$ for some subset Θ of $\mathbb{R}^K$. That is, the MTR function is assumed to be a member of the class of functions formed by taking linear combinations of the basis functions. This reduces the dimension of the function m to a K-dimensional real vector θ.

Linear-in-parameters specifications like (11) are commonplace in statistical models. For example, if x is scalar, one could specify

$$\begin{aligned} m_\theta(d|u,x) = \; & \overbrace{\theta_1(1-d) + \theta_2(1-d)u + \theta_3(1-d)x + \theta_4(1-d)ux}^{m_\theta(0|u,x)} \\ & + \underbrace{\theta_5 d + \theta_6 du + \theta_7 du^2 + \theta_8 dx + \theta_9 dx^2}_{m_\theta(1|u,x)} \end{aligned} \tag{12}$$

which corresponds to $K = 9$ parameters with basis functions e.g. $b_3(d|u,x) = (1-d)x$ and $b_7(d|u,x) = du^2$. The assumption used in the `ivmte` package is that $\mathcal{M}$ contains only MTR functions such that both $m_\theta(0|u,x)$ and $m_\theta(1|u,x)$ are either polynomials or polynomial B–splines in u. This is certainly a special case of (11), but one that is popular both as a parametric restriction (e.g. a polynomial, like (12)) and as an approximating basis for nonparametric sieve estimation (e.g. Chen, 2007).

3.2 Sample analogs

The benefit of using the linear-in-parameters specification (11) is that it preserves the linearity of τ, γ_s, and γ_{cm} as functions of m. In particular,

$$\tau(m_\theta) = \theta' T, \tag{13}$$

where T is a K-dimensional vector with kth element

$$\tau(b_k) = \sum_{d\in\{0,1\}} \mathbb{E}\left[\int_0^1 b_k(d|u,X_i)\omega_\tau(d|u,X_i,Z_i)\,du\right]. \tag{14}$$

Given a sample of data $\{(Y_i,D_i,X_i,Z_i)\}_{i=1}^n$, each component of T can be estimated by its sample analog

$$\hat{\tau}(b_k) \equiv \frac{1}{n}\sum_{i=1}^n \sum_{d\in\{0,1\}} \int_0^1 b_k(d|u,X_i)\hat{\omega}_\tau(d|u,X_i,Z_i)\,du, \tag{15}$$

where $\hat{\omega}_\tau$ is an estimate of ω_τ. Requiring $m_\theta(d|u,x)$ to be a polynomial or B–spline in u means that the integral in $\hat{\tau}(b_k)$ can be computed analytically as long as $\hat{\omega}_\tau(u,x,z)$ has a tractable form, which it does for all conventional target parameters.[8] By the same reasoning,

$$\gamma_s(m_\theta) = \theta'\Gamma_s \quad \text{and} \quad \gamma_{\text{cm}}(m_\theta|d,x,z) = \theta'\Gamma_{\text{cm}}(d,x,z),$$

where Γ_s and $\Gamma_{\text{cm}}(d,x,z)$ are K-dimensional vectors with kth elements given by $\gamma_s(b_k)$ and $\gamma_{\text{cm}}(b_k|d,x,z)$. A sample analog estimator of $\gamma_s(b_k)$ is

$$\begin{aligned}\hat{\gamma}_s(b_k) \equiv \quad & \frac{1}{n}\sum_{i=1}^n \hat{s}(0,X_i,Z_i)\int_0^1 b_k(0|u,X_i)\mathbb{1}[u>\hat{p}(X_i,Z_i)]\,du \\ & + \frac{1}{n}\sum_{i=1}^n \hat{s}(1,X_i,Z_i)\int_0^1 b_k(1|u,X_i)\mathbb{1}[u\le\hat{p}(X_i,Z_i)]\,du,\end{aligned}$$

where $\hat{s}$ is an estimator of s, and $\hat{p}$ is an estimator of p. A sample analog estimator of $\gamma_{\text{cm}}(b_k|d,x,z)$ is

$$\hat{\gamma}_{\text{cm}}(b_k|d,x,z) = \int_0^1 b_k(0|u,x)(1-d)\frac{\mathbb{1}[u>\hat{p}(x,z)]}{(1-\hat{p}(x,z))} + b_k(1|u,x)d\frac{\mathbb{1}[u\le\hat{p}(x,z)]}{\hat{p}(x,z)}\,du.$$

8. `ivmte` allows for any target parameter for which $\hat{\omega}_\tau(u,x,z)$ can be written as a constant spline in u—see Section 4.3.

We use these sample analogs to define sample criterion functions. The moment-based sample criterion is

$$\hat{Q}_{\mathcal{S}}(m_\theta) \equiv \left\| \hat{\Gamma}_{\mathcal{S}}\theta - \hat{c}_{\mathcal{S}} \right\|,$$

where $\hat{\Gamma}_{\mathcal{S}}$ is an $|\mathcal{S}| \times K$ matrix with rows $\hat{\Gamma}_s' \equiv [\hat{\gamma}_s(b_1), \ldots, \hat{\gamma}_s(b_K)]$, and $\hat{c}_{\mathcal{S}}$ is a vector with elements

$$\hat{c}_s \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i \hat{s}(D_i, X_i, Z_i).$$

The regression-based sample criterion is

$$\hat{Q}_{\text{cm}}(m_\theta) \equiv \frac{1}{n}\sum_{i=1}^{n} \left(Y_i - \theta' \hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i) \right)^2,$$

where $\hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)$ is a K-dimensional vector with kth element $\hat{\gamma}_{\text{cm}}(b_k | D_i, X_i, Z_i)$.

3.3 Estimation

Estimation differs for point and partially identified cases.

3.3.1 Point identification

In the point identified case, we assume that the parameter space is $\Theta = \mathbb{R}^K$. This simplification allows for closed-form estimators.

For the moment criterion, we use the generalized method of moments (Hansen, 1982, "GMM") estimator of θ:

$$\hat{\theta} = \text{argmin}_{\theta \in \mathbb{R}^K} \left(\hat{\Gamma}_{\mathcal{S}}\theta - \hat{c}_{\mathcal{S}} \right)' \hat{\Omega} \left(\hat{\Gamma}_{\mathcal{S}}\theta - \hat{c}_{\mathcal{S}} \right), \tag{16}$$

where $\hat{\Omega}$ is a positive semi-definite weighting matrix. The minimizer $\hat{\theta}$ of (16) is the minimizer of $\hat{Q}_{\mathcal{S}}$ when $\|\cdot\|$ is taken to be the Euclidean norm weighted by $\hat{\Omega}$. In a point identified case, one would expect that (16) has a unique solution, in which case it can be solved for analytically.

The regression sample criterion $\hat{Q}_{\text{cm}}$ is simply the ordinary least squares criterion for a linear regression of Y_i onto the vector of generated regressors $\hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)$. It has a unique minimizer if and only if the matrix

$$\sum_{i=1}^{n} \hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i) \hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)' \tag{17}$$

is invertible. If the researcher believes that point identification holds, then a simple estimator $\hat{\theta}$ of θ is the ordinary least squares estimator from a regression of Y_i on $\hat{\Gamma}_{\text{cm}}(D_i, X_i, Z_i)$.

For both the moment and regression approaches, we then set

$$\hat{\tau}^\star \equiv \hat{\theta}'\hat{T} \tag{18}$$

where $\hat{T}$ is the K-dimensional vector with kth element $\hat{\tau}(b_k)$. Then $\hat{\tau}^\star$ is our point estimator of the (assumed singleton) identified set for the target parameter, $\mathcal{T}_{\mathcal{S}}^\star$ or $\mathcal{T}_{\text{cm}}^\star$, depending on the criterion function used.

3.3.2 Partial identification

For partially identified cases we use a two-step estimator developed by Mogstad et al. (2018) for both the moment and regression approaches. In the first step, we minimize the criterion function to find

$$\hat{Q}^\star \equiv \min_{\theta \in \Theta} \hat{Q}(m_\theta), \tag{19}$$

where $\hat{Q}$ could be either $\hat{Q}_{\mathcal{S}}$ or $\hat{Q}_{\text{cm}}$, and now the parameter space Θ is allowed to be a proper subset of $\mathbb{R}^K$. In the second step, we then minimize and maximize the target parameter over the set of $\theta \in \Theta$ that produce sample criteria close to the best possible value, $\hat{Q}^\star$. That is, we solve for

$$\hat{\tau}_{\text{lb}}^\star / \hat{\tau}_{\text{ub}}^\star \equiv \min_{\theta \in \Theta} / \max_{\theta \in \Theta} \theta'\hat{T} \qquad \text{subject to} \quad \hat{Q}(m_\theta) \leq (1+\sigma)\hat{Q}^\star, \tag{20}$$

where $\sigma \geq 0$ is a tuning parameter used in the asymptotic theory (see Mogstad et al., 2018, for more detail). The feasible set in (20) is always non-empty due to the definition of $\hat{Q}^\star$, so that both $\hat{\tau}_{\text{lb}}^\star$ and $\hat{\tau}_{\text{ub}}^\star$ are always well-defined. Mogstad et al. (2018) provide conditions under which $[\hat{\tau}_{\text{lb}}^\star, \hat{\tau}_{\text{ub}}^\star]$ is a consistent set estimator of $\mathcal{T}_{\mathcal{S}}^\star$ if $\hat{Q} = \hat{Q}_{\mathcal{S}}$, and of $\mathcal{T}_{\text{cm}}^\star$ if $\hat{Q} = \hat{Q}_{\text{cm}}$.

To facilitate computation, we assume that the constraint set Θ can be written in terms of linear inequality constraints:

$$\Theta = \left\{ \theta \in \mathbb{R}^K : r_{\text{lb}} \leq R\theta \leq r_{\text{ub}} \right\}, \tag{21}$$

for vectors $r_{\text{lb}}, r_{\text{ub}}$, and a conformable matrix R. In practice, these constraints typically represent bounds on levels and/or derivatives of $m(0|\cdot, x)$ and $m(1|\cdot, x)$ and/or $m(1|\cdot, x) - m(0|\cdot, x)$ on a large grid of evaluation points. We discuss shape constraints in more detail in Sections 4.5 and 5.

For the moment approach, the structure of the first and second step programs depends on the choice of norm $\|\cdot\|$. In the `ivmte` module, we take $\|\cdot\|$ to be the ℓ_1 norm so that

$$\hat{Q}_{\mathcal{S}}(m_\theta) = \sum_{s \in \mathcal{S}} \left| \hat{\Gamma}_s \theta - \hat{c}_s \right|. \tag{22}$$

This choice is attractive given (21) because one can then reformulate the first and second step problems (19) and (20) as linear programs by replacing absolute values with appropriate slack variables. Linear programs scale quite well with the number of parameters and constraints.

Given (21), the program defining $\hat{Q}^\star$ in the regression approach is a convex quadratic program. The second step programs in (20) are convex quadratically-constrained quadratic programs (QCQPs). Mature algorithms exist for solving both types of programs to global optimality. As one might expect, QCQPs do not tend to scale as well as LPs, and can be more sensitive to numerical issues.

3.4 Trade-offs between the moment and regression approaches

As already noted, the identified set for the regression approach is always weakly smaller than in the moment approach: $\mathcal{T}^\star_{\text{cm}} \subseteq \mathcal{T}^\star_{\mathcal{S}}$. Not only that, but the researcher does not need to specify the set $\mathcal{S}$, as they would in the moment approach. The number of options for elements of $\mathcal{S}$ can be large, especially with covariates, and removing this subjective element may be attractive. In point identified cases, the regression approach has the additional benefit of being implementable through ordinary least squares, which is computationally trivial and can be expected to have good statistical properties. These are certainly strong points in favor of the regression approach.

There are, however, also some benefits to the moment approach. Being able to choose the set of moments $\mathcal{S}$ that are fit can be attractive since it draws a clear line between the portions of the observed data that are used in inference and the portions that are not. For example, Mogstad and Torgovitsky (2018) suggest reporting common linear IV model estimates such as various two-stage least squares specifications—which do not in general estimate an interesting target parameter—together with bounds on the target parameter that incorporate the same information by using the same linear IV estimands as functions in $\mathcal{S}$. The moment-based criterion can also be easier to interpret; it simply measures the distance to satisfying the moments, so if the number of moments is small and the MTR function is flexible, it can be exactly zero, indicating that all the moments can be reproduced. A related consequence is that the moment-based approach can be used for specification testing. The other primary benefit of the moment approach is computation in partially

identified cases, where it produces an LP implementation that can usually be expected to be easier to compute than the QCQPs required in the regression approach.

4. The `ivmte` package

4.1 Installation and requirements

The `ivmte` package is available in CRAN, and can be installed and loaded as usual:

```
install.packages("ivmte")
library("ivmte")
```

The most up-to-date version can be installed directly from the GitHub repository:

```
devtools::install_github("jkcshea/ivmte")
```

The `splines2` package, which is available on CRAN, is required for implementing specifications in which the MTR function contains polynomial splines (see Section 4.4). No additional packages are required for implementing the point estimators discussed in Section 3.3.1.

For the partially identified cases, `ivmte` requires a solver package. If using the moment approach, the options are `gurobi`, `Rmosek`, `cplexAPI` (Roettger et al., 2019), or `lpSolveAPI` (Konis, 2019). The first package requires a Gurobi (Gurobi Optimization, Inc., 2015) license, the second requires a MOSEK (MOSEK ApS, 2021) license, while the third requires a CPLEX (IBM, 2010) license. These are available at no cost for academic researchers. Alternatively, `lpSolveAPI` is freely available through CRAN and does not require a license. For the regression approach with partial identification, `ivmte` requires either `gurobi` or `Rmosek`, since the other solvers cannot solve QCQPs.[9]

All of the examples shown ahead in Section 5 were computed using Gurobi.

4.2 Basic syntax

The main command in `ivmte` is called `ivmte`. It requires the following arguments

```
ivmte(data, target, m0, m1, outcome, propensity)
```

where `data` is the usual `dataframe` and

- `target` specifies the target parameter, τ.

9. CPLEX can solve QCQPs, but its `R` API does not appear to allow for it.

- `m0` and `m1` are formulas indicating the specification for the MTR function m broken up into treatment arms $m(0|u,x)$ and $m(1|u,x)$.
- `outcome` indicates the outcome variable, Y_i, and implements the regression criterion. To use moment criterion, one instead passes `ivlike`, which contains a list of formula that determine the set of functions $\mathcal{S}$ that define the moment conditions.[10]
- `propensity` is a formula that specifies how the propensity score is estimated.

In the remainder of this section we discuss how to specify these arguments to implement the methodology previously described. Along the way, we cover some additional options that provide extra functionality. More detail and discussions of some lesser-used options are provided in a vignette at the GitHub repository (https://github.com/jkcshea/ivmte).

4.3 Specifying the target parameter

The `target` option can be set to one of `"ate"`, `"att"`, `"atu"`, `"late"`, or `"genlate"`, which correspond respectively to the average treatment effect (ATE), the average treatment on the treated (ATT), the average treatment on the untreated (ATU), the local average treatment effect (LATE; Imbens and Angrist, 1994), and the generalized LATE (Heckman and Vytlacil, 2005; Mogstad et al., 2018). The choice of this argument specifies the form of the target parameter τ via its weighting function ω_τ in (5). Nothing else has to be specified for `"ate"`, `"att"`, and `"atu"`. It is also possible to specify a custom parameter by specifying the weights ω_τ.

4.3.1 LATE

The local average treatment effect (LATE) from shifting the instrument Z_i from z_0 to z_1 is defined as

$$\text{LATE}(z_0 \to z_1) \equiv \mathbb{E}[Y_i(1) - Y_i(0) | D_i(z_0) = 0, D_i(z_1) = 1].$$

In terms of the equivalent selection model (1),

$$\text{LATE}(z_0 \to z_1) = \mathbb{E}\left[\int_{p(X_i,z_0)}^{p(X_i,z_1)} (m(1|u,X_i) - m(0|u,X_i)) \left(\frac{1}{\mathbb{E}\left[p(X_i,z_1) - p(X_i,z_0)\right]}\right) du\right],$$

10. The terminology comes from Mogstad et al. (2018), who described the class of cross-moments $c_s \equiv \mathbb{E}[Y_i s(D_i, X_i, Z_i)]$ as "IV–like" estimands because they nest standard linear IV estimands via particular choices of s.

which takes the form (5) with weighting function

$$\omega_\tau(d|x,z) = (-1)^{d+1}\frac{\mathbb{1}[p(x,z_0) < u \leq p(x,z_1)]}{\mathbb{E}\left[p(X_i,z_1) - p(X_i,z_0)\right]}.$$

This is the form of ω_τ used if `target = "late"`. The user must pass `late.from` and `late.to`, which should be named lists indicating the identity and value of z_0 and z_1, respectively.

As defined, the LATE parameter averages over all covariates. The `ivmte` package also allows for "effect modification," where the LATE is computed conditional on $V_i = v$, for some function V_i of the covariate vector X_i (e.g. Ogburn et al., 2015; Kennedy et al., 2019):

$$\begin{aligned}&\text{LATE}(z_0 \to z_1|v)\\ &\quad\equiv \mathbb{E}\left[\int_{p(X_i,z_0)}^{p(X_i,z_1)} (m(1|u,X_i) - m(0|u,X_i))\left(\frac{1}{\mathbb{E}\left[p(X_i,z_1) - p(X_i,z_0)|V_i = v\right]}\right) du|V_i = v\right].\end{aligned}$$

To do this, set `target = "late"` and `late.from`, `late.to` as above, but also pass a named list `late.X` to indicate the variable V_i and value v. Note that no smoothing is done for the conditional expectation, so V_i should be a discrete variable.

4.3.2 Generalized LATE

The selection model (1) allows conceptualizing a generalized LATE where instead of choosing instrument values z_0 and z_1, we choose values u_0 and u_1 for the latent propensity variable U_i (Heckman and Vytlacil, 2005). This can be useful for diagnosing the robustness of a standard LATE to broadening the complier subpopulation (Mogstad and Torgovitsky, 2018). The formal definition is

$$\text{GenLATE}(u_0,u_1) \equiv \mathbb{E}\left[\int_{u_0}^{u_1} (m(1|u,X_i) - m(0|u,X_i))\frac{1}{(u_1-u_0)}\,du\right], \tag{23}$$

for values $u_0, u_1 \in [0,1]$ with $u_0 < u_1$. To set the target parameter to (23) in `ivmte`, pass `target = "genlate"`, `genlate.lb` and `genlate.ub`, where the latter two parameters correspond to u_0 and u_1. Effect modification can also be incorporated by passing `late.X`, in the same way as for the usual LATE discussed in the previous section.

4.3.3 Custom target parameters

The user can define their own target parameters by directly specifying the weight function ω_τ in (5). To facilitate computation, these weight functions are required to be constant

splines in u, i.e.

$$\omega_\tau(d|u, X_i, Z_i) = \sum_{j=1}^{J_d} \mathbb{1}\left[\kappa_{j-1}(d|X_i, Z_i) < u \leq \kappa_j(d|X_i, Z_i)\right] \bar{\omega}_{\tau,j}(d|X_i, Z_i),$$

where $\kappa_0(d|X_i, Z_i) \equiv 0$, and $k_{J_d}(d|X_i, Z_i) \equiv 1$. The user sets these weights by passing the $J_0 - 1$ knot functions $(\kappa_1(0|\cdot,\cdot), \ldots, \kappa_{J_0-1}(0|\cdot,\cdot))$ as a list via `target.knots0` and the J_0 weight functions $(\bar{\omega}_{\tau,1}(0|\cdot,\cdot), \ldots, \bar{\omega}_{\tau,J_0}(0|\cdot,\cdot))$ as a list via `target.weight0`. The analogous options `target.knots1` and `target.weight1` for the treated ($d = 1$) weights also need to be specified. For any component of these lists, a constant (scalar numeric) can be passed instead of a function to indicate a function that does not vary with (x, z). Note that the option `target` is ignored when any of the custom `target.*` options are passed.

4.4 Specifying the MTR function

The required `m0` and `m1` arguments accept specifications for two treatment arms of the MTR function using the standard `R` formula syntax familiar from functions like `lm` or `glm`. However, these formulas involve an unobservable variable whose default name is `u`.[11] Typical specifications will involve combinations of `u` and other covariates. For example,

```
m0 <- ~ var1 + u + I(var1 * u) + I(u^2)
```

specifies $m(0|u, x)$ to be quadratic in the unobservable `u` (u) and linear in `var1` (a subcomponent of x), with a first order interaction between `u` and `var1`. Note that the left-hand side of these formulas is empty. Also note the use of `I()` to inhibit the interpretation of `*` and `^` as formula operators.

Currently, `ivmte` requires specifications of `m0` and `m1` to either be polynomials or B-splines in `u`. B-splines are incorporated using the function `uSplines`, which is an interpreter that utilizes the `splines2` package (Wang and Yan, 2018). An example of the syntax is

```
m1 <- ~ var1 + uSplines(degree = 0, knots = c(0.2, 0.5, 0.8))
```

which would specify $m(1|u, x)$ to be linear in `var1` and piecewise constant in `u` with jumps at the specified knot points.[12] Splines can be interacted with other variables and intermingled with other polynomials, for example

11. The name can be changed with the `uname` option.
12. As Wang and Yan (2018) describe in their vignette, the only difference between the `bSpline` function in `splines2` and the `bs` function in the core package `splines` is that `bSpline` allows for degree 0 splines, i.e. piecewise constant functions. This turns out to be particularly useful for our purposes because piecewise constant functions have a special place in the theory developed by Mogstad et al. (2018); see their Proposition 4.

```
m0 <- ~ u + I(u^2) + var1:uSplines(degree = 2, knots = c(0.3, 0.4, 0.5, 0.7))
```

would specify a quadratic function of `u` and a linear function of `var1` whose slope varies with u according to a quadratic B-spline with knot points at `.3`, `.4`, `.5`, and `.7`.

4.5 Imposing shape constraints

For partial identification cases, `ivmte` also allows the user to require the MTR and/or MTE functions to be bounded and/or monotone in u. The bounds are imposed through the arguments `m0.lb`, `m0.ub`, `m1.lb`, `m1.ub`, `mte.lb`, and `mte.ub`. Note that the default action of `ivmte` is to set the upper and lower bounds on `m0` and `m1` to the largest and smallest values of the response variable observed in the data, which also implies values for `mte.lb` and `mte.ub`. Monotonicity in u, in either an increasing or decreasing sense, is set through the boolean arguments `m0.dec`, `m0.inc`, `m1.dec`, `m1.inc`, `mte.dec`, and `mte.inc`. These arguments are set to `FALSE` by default.

These shape constraints (boundedness and monotonicity) are enforced through an "auditing" procedure. The procedure is designed to circumvent the difficulty of determining whether a polynomial function is bounded or monotone on its domain. It starts by imposing the desired shape constraints on the MTR function at all points on a well-spaced, relatively coarse *constraint grid*. After producing the bound estimates $\hat{\tau}^{\star}_{\text{lb}}$ and $\hat{\tau}^{\star}_{\text{ub}}$, the solution MTR functions at these bounds are checked ("audited") for shape restrictions on a much finer *audit grid*. If the solutions satisfy the shape restrictions across the entire audit grid, then the process ends. Otherwise, the estimator is recomputed with an expanded constraint grid that contains some of the points in the audit grid where the restrictions were violated. The procedure repeats until the solutions pass the audit, or until a maximum number of iterations are reached.

The user can adjust the size of the initial constraint grid through the arguments `initgrid.nu` and `initgrid.nx`. These arguments control the initial number of points at which to impose the constraints for `u`, via `initgrid.nu`, and all other variables included in the specification of `m0` and `m1`, via `initgrid.nx`. For the latter, the points are drawn randomly from the empirical distribution in `data`. The default values of `initgrid.nu` and `initgrid.nx` are both 20, so that the total initial constraint grid size is 400.

The user can also adjust the size of audit grid through the arguments `audit.nu` and `audit.nx`. The default for `audit.nu` is 25, while the default for `audit.nx` is set at 2500. By default then, the solution MTR function must satisfy the shape constraints on an audit grid with 62,500 points.[13] When a solution MTR function fails an audit, the number of

13. Assuming of course that there are at least 2500 unique values of X_i in the data. Otherwise, the entire empirical support of X_i is used.

violating points that are added to the constraint grid from the audit grid (for each shape constraint) is given by `audit.add`, which has a default of 100.

The audit is terminated after the solution MTR functions satisfy the constraints on the entire audit grid, or after `audit.max` rounds of the audit procedure, which has a default of 25 rounds. If `audit.max` is hit, the user should investigate the `audit.grid$violations` field of the list that `ivmte` returns. This reports the points of the audit grid at which the shape restrictions are violated. Small regions of violation can likely be ignored without seriously affecting the estimated bounds $\hat{\tau}^{\star}_{\text{lb}}$ and $\hat{\tau}^{\star}_{\text{ub}}$. If the violations occur on a large region, the user can let the audit procedure run for more rounds by increasing `audit.max`.

4.6 Specifying the criterion function

To use the regression approach, simply leave the `ivlike` input empty and indicate the outcome variable Y_i as `outcome = y`.

To use the moment approach, one needs to specify the collection of functions $\mathcal{S}$ via `ivlike` by passing a vector of formulas, each of which has the same outcome variable (Y_i) on the left-hand side. For example,

```
ivlike <- c(
  y ~ d,
  y ~ d | z,
  y ~ d + x | z + I(z^2) + x
)
```

has three formulas, with the second two specified using the `|` syntax familiar from the `ivreg` command in the `AER` package (Kleiber and Zeileis, 2018). The first formula is an OLS regression of `y` on `d` and a constant. The second formula uses `z` as an instrument for `d`, as in just-identified IV regressions. The third formula uses `z` and `z^2` as instruments for `d`, with `x` serving as a covariate that instruments for itself.

The default behavior of `ivmte` is to include all of the estimated coefficients from each specification as functions $s \in \mathcal{S}$. In the example above, this would mean the coefficients on the constant and `d` in the first and second specifications, and the coefficients on the constant, `d` and `x` in the third, for a total of 7 moments to match. The user can change this behavior with the `components` argument. This argument expects a list of the same length as `ivlike`, with the jth component of the list being a vector that indicates which coefficients should be included from the jth IV–like specification in `ivlike`. In the example above, we could have used

```
components <- l(intercept, d, c(d, x))
```

to indicate that we want only the coefficient on the constant (intercept) from the first specification, only the coefficient on `d` in the second, and both the coefficients on `d` and `x` in the third, for a total of 4 moments. Note that `intercept` is used to refer to the implied constant term in the formula specifications, and so should be viewed as a reserved word when it comes to naming data columns.[14]

Conditioning subsets for the IV–like specification can be set through the optional `subset` argument. This option expects a list of the same length as `ivlike`, with each component of the list representing a logical statement. For example,

```
subset <- l(z == 1, , x %in% c(2, 3))
```

would estimate the first IV–like specification only on the subset with `z == 1`, the second for all observations, and the third only for the subset for which either `x == 2` or `x == 3`. This provides an easy way to specify conditional moments as components of $\mathcal{S}$, e.g. via

$$\mathbb{E}[Y_i|Z_i=1] = \mathbb{E}\left[\underbrace{Y_i \frac{\mathbb{1}[Z_i=1]}{\mathbb{P}[Z_i=1]}}_{\text{example of } s(D_i,X_i,Z_i)}\right]. \tag{24}$$

4.7 Propensity score estimation

Estimating $\hat{\gamma}_s$ and $\hat{\gamma}_{\text{cm}}$, as well as $\hat{\tau}$ for many choices of target parameter requires first estimating the propensity score, $p(x,z) \equiv \mathbb{P}[D_i=1|X_i=x, Z_i=z]$. This is communicated through the `propensity` argument. Typically, the user will pass a formula for `propensity` in which the treatment variable appears on the left-hand side, for example

```
propensity <- d ~ x + z
```

By default, this estimates a logit model using `glm` with the specified right-hand side variables, but the user can change this to probit or linear by passing `link = "probit"` or `link = "linear"`.

Alternatively, the user can estimate the propensity score before running `ivmte`, save estimates of $p(X_i, Z_i)$ in their dataframe as a new column, say `p`, and then pass `propensity = p`. When this is done, the user must also indicate the name of the treatment variable through the argument `treat`. When a formula is passed for `propensity`, the treatment variable is inferred to be the response variable of the formula, and the `treat` argument is ignored unless it does not match the inferred variable, in which case an error is thrown.

14. The `l` function is a generalization of the `list` function, and allows the user to list variables and expressions without having to enclose them by quotation marks.

4.8 Solving

By default, `ivmte` attempts to determine whether there is a unique solution to either the moment-based or regression-based criteria (depending on the user's specification of `ivlike`) by checking the rank of their (unconstrained) first-order equations. If it determines that there is a unique solution, and `point` is either not passed, or passed as `point = TRUE`, then it proceeds to solve for the unique solution and form a point estimate of the target parameter as described in Section 3.3.1. For the moment criterion, the default behavior is to use the optimal two-step weighting for $\hat{\Omega}$, but this can be changed to the identity weighting by passing `point.eyeweight = TRUE`.

If `ivmte` determines there is not a unique solution, or if `point = FALSE` is passed, then it proceeds with the two-step bounds estimator described in Section 3.3.2 The solver package for these problems is set using the option `solver`, which currently accepts the following values: `gurobi`, `Rmosek`, `cplexAPI`, and `lpSolveAPI`.[15] If no value is passed for `solver`, then `ivmte` searches for a solver in the order given above and uses the first one that is found. The value of the tuning parameter, σ, in (20) is set to 10^{-4} by default, and can be changed with the `criterion.tol` argument.

4.9 Confidence intervals

The `ivmte` command can construct confidence intervals by resampling (bootstrapping or subsampling). The number of replications is determined by the argument `bootstraps`, which is set to 0 by default so that confidence intervals are not computed. The size of the resampled dataset is determined by `bootstraps.m`, which is set to the sample size of `data` by default. The default behavior is to draw the resampled data with replacement from `data`, but this can be toggled with the boolean argument `bootstraps.replace`. Confidence intervals are reported for all levels in `levels`, which has the default of `c(.99, .95, .90)`.

For the point-identified case, the reported intervals are formed from the resampled distribution of (18).[16] Conducting statistical inference on bounds in the partially identified case is more delicate due to their potentially non-standard asymptotic distributions.[17] There does not currently exist a solution for the MTE framework that is both theoretically satisfactory and computationally tractable. Instead, `ivmte` implements the forward and

15. The `gurobi` package is included with Gurobi, while `cplexAPI` and `lpSolveAPI` are available from CRAN. The version requirements of `ivmte` as of this writing are: `gurobi` 8.1-0 or later, `Rmosek` 9.2.38 or later, `cplexAPI` 1.3.3 or later, and `lpSolveAPI` 5.5.2 or later.
16. The moment-based criterion uses re-centered moment conditions (Hall and Horowitz, 1996; Brown and Newey, 2002).
17. See Canay and Shaikh (2017) for a recent survey on inference in partially identified models.

reverse bootstrap procedures discussed by Andrews and Han (2009).[18] While these are known to *not* be valid in general, they may still provide a reasonable indication of statistical uncertainty for the user. In addition to confidence intervals for each level in `levels`, `imvte` also returns a p-value, computed as the smallest level a such that a $1-a$ confidence interval would not contain 0.

4.10 Specification tests

If using the moment-based criterion function, `ivmte` will also conduct a bootstrap test of the null hypothesis that the model is correctly specified (i.e. of the null hypothesis that the minimum value of the population criterion is zero) whenever `bootstraps` is a positive number. In the point-identified case, the test used is the well-known Hansen (1982) overidentification test for GMM using the adjustment for bootstrapping discussed by Hall and Horowitz (1996). In the partially-identified case, the test used is the "re-sampling" test of Bugni et al. (2015). In either case, `ivmte` returns a p-value for the null hypothesis of correct specification. The user can turn off the specification test by passing `specification.test = FALSE`.

4.11 Output

The return of `ivmte` is a named list with a large number of fields.[19] The most important fields are `pointestimate` and `bounds`, which return (18) or (20), depending on whether `point` is `TRUE` or `FALSE`. If confidence intervals are being computed, these are returned in the fields `pointestimate.ci` or `bounds.ci`, with the p-value returned in the field `pvalue`. Other fields that may be useful for diagnostics or debugging are `s.set`, which contains the results of running the IV–like specifications, `propensity`, which contains the results of the propensity score estimation, `audit.criterion`, which gives the value $\hat{Q}^\star$ in (19), and `audit.grid$violations`, which reports points at which the audit procedure failed to secure compliance with the desired shape restrictions.

5. Empirical illustration

5.1 Data and motivation

We illustrate the motivation and usage of `ivmte` by revisiting Angrist and Evans's (1998) analysis of the relationship between fertility on labor supply. The data comes from the 1980 Census Public Use Micro Samples (PUMS); a detailed description can be found in Angrist

18. The default is to compute and report the results from both backward and forward procedures. This behavior can be changed by passing `ci.type = "backward"` or `ci.type = "forward"`.
19. In case memory usage is an important issue to the user, we have included an option `smallreturnlist` that can be set to `TRUE` to limit the number of objects that are returned.

and Evans (1998).[20] Our illustration uses three main variables: `worked` is an indicator for whether a woman worked for pay in the year prior to the survey, `morekids` is an indicator for whether a woman has exactly two children (`morekids = 0`) or three or more children (`morekids = 1`), and `samesex` is an indicator that is 1 if the first two children had the same sex. Later, we will also use the woman's year of birth (`yob`) and indicators for her race (`hisp`, `black`, `other`) to demonstrate specifications with covariates. Our interest is in the effect of having more than two children (`morekids`) on labor supply (`worked`).

A simple linear regression of `worked` on `morekids` shows that 58% of women with two children work, compared to only 44% of those with three or more children:

```
lm(data = AE, worked ~ morekids)
```

```
##
## Call:
## lm(formula = worked ~ morekids, data = AE)
##
## Coefficients:
## (Intercept)     morekids
##      0.5822      -0.1423
```

The coefficient on `morekids` of $-.14$ probably overstates the causal impact of fertility on labor supply, since women who choose to have more children likely do so in part because their labor market prospects are weaker. An IV regression using `samesex` as an instrument for `morekids` returns a coefficient on `morekids` that is substantially smaller in magnitude:

```
library("AER")
ivreg(data = AE, worked ~ morekids | samesex)$coeff["morekids"]
```

```
##    morekids
## -0.08484221
```

If there is heterogeneity in the effect of fertility on working, then this IV estimate only reflects the same-sex compliers, that is, those women who would have a third child if and only if their first two had the same sex. The "first stage" regression of `morekids` on `samesex` shows that this group is rather small, comprising less than 6% of the population.

20. The original data can be downloaded from https://economics.mit.edu/files/1199 or from http://sites.bu.edu/ivanf/files/2014/03/m_d_806.dta_.zip. The data we use is restricted to women who were at least 20 years old at their first birth. The cleaned subsample data with only the variables relevant to the current analysis is included as data with `ivmte`.

```
lm(data = AE, morekids ~ samesex)$coeff["samesex"]
```

```
##    samesex
## 0.05886826
```

If our research question requires knowing a quantity involving the entire population, such as the ATE or the ATT, then this linear IV estimate is not particularly helpful.

5.2 Extrapolation to the ATE under different assumptions

The `ivmte` package can be used to extrapolate from the small complier group to larger groups by providing a coherent framework under which additional assumptions can be imposed. Suppose for example that we assume that the MTR function is quadratic in u for both treatment states, so that the pair is characterized by six parameters. Since both `morekids` and `samesex` are binary, we only have four moments at our disposal to identify these six parameters, so the model is not point identified. However, we can use `ivmte` to estimate bounds on the ATE:[21]

```
ivmte(
  data = AE,
  ivlike = c(worked ~ morekids + samesex + morekids * samesex),
  target = "ate",
  m0 = ~ u + I(u^2),
  m1 = ~ u + I(u^2),
  propensity = morekids ~ samesex
)
```

```
##
## Bounds on the target parameter: [-0.2862919, 0.1050867]
## Audit terminated successfully after 1 round
```

As a comparison, Manski's (1990) nonparametric IV bounds on the ATE are $[-.548, .393]$. The bounds produced by `ivmte` are much tighter because they impose a parametric assumption on the model primitives which smooths out the extreme cases at which Manski's bounds are obtained. The parametric assumption says that if we line up families by their latent propensity to have a third child, then families who are close to having the same propensity (similar u) are, on average, not too dissimilar in their potential work outcomes. A weaker

21. This call illustrates the basic syntax introduced in Section 4.2. The specification of `ivlike` indicates four moments in $\mathcal{S}$ (see Section 4.6). The specifications of `m0` and `m1` follow the syntax discussed in Section 4.4 with the `u` syntax for the unobserved variable. The `propensity` option indicates a simple logit regression of `morekids` on `samesex` and a constant (Section 4.7), which in this case is numerically equivalent to a binned estimator of the conditional probability of treatment.

parameterization for the MTR, such as a polynomial of higher order than 2, would allow families with similar fertility propensities to be more different, since such a function could "wiggle" more quickly between the natural bounds of 0 and 1 for Y_i.

While narrower than Manski's nonparametric bounds, the bounds under a quadratic parameterization are still quite wide; they are consistent with both large negative and modest positive causal effects. However, because the outcome is binary and all four potential s functions have been incorporated into the saturated specification `ivlike = c(worked ~ morekids + samesex + morekids*samesex)`, we know from Proposition 3 of Mogstad et al. (2018) that the bounds are sharp (best possible) in the sense of fully exhausting the information contained in the model and the data. Thus, if the researcher is unsatisfied with the width of the bounds, they have two paths to satisfaction: (i) make stronger (or different) assumptions, or (ii) ask a less ambitious question by changing the target parameter.

A natural way to strengthen the assumptions is to eliminate the quadratic terms in the MTR specifications, so that there are only four parameters:

```
ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~u,
  m1 = ~u,
  propensity = morekids ~ samesex
)
```

```
## Warning: MTR is point identified via GMM.

##
## Point estimate of the target parameter: -0.07791036
```

The bounds have collapsed to a point. This makes sense since we have not changed `ivlike`, so we still have four moments, but relative to the quadratic case we have reduced the number of parameters from six to four (Brinch et al., 2012, 2017). If we had done this moment-counting exercise ahead of time, we could have added `point = TRUE` to the call (Section 4.8):

```
ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~u,
  m1 = ~u,
```

```
  propensity = morekids ~ samesex,
  point = TRUE
)
```

```
##
## Point estimate of the target parameter: -0.07791036
```

Linearity is a restrictive parameterization, and one might be uncomfortable with the fact that it allows for complete extrapolation from the 6% of the population represented in the LATE to the entire population represented in the ATE. As an alternative, consider combining the quadratic case with shape restrictions (Section 4.5). For example, we could assume that the MTRs must generate an MTE curve that is negative and increasing:

```
ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~ u + I(u^2),
  m1 = ~ u + I(u^2),
  mte.inc = TRUE,
  mte.ub = 0,
  propensity = morekids ~ samesex
)
```

```
##
## Bounds on the target parameter: [-0.08484221, -0.06323574]
## Audit terminated successfully after 1 round
```

The assumption behind this shape restriction is that the effect of having another child on working is negative ($m(1|u,x) - m(0|u,x) \leq 0$, imposed via `mte.ub = 0`), and is more negative for women who are more likely to have more children ($m(1|u,x) - m(0|u,x)$ increasing as a function of u, imposed via `mte.inc = TRUE`). Adding the assumption narrows the bounds considerably, from $[-.286, .105]$ to $[-.085, -.063]$. In this case, the resulting bounds happen to be similar to the original linear IV estimate for compliers, but they are the product of a formally-justified theoretical framework for extrapolation, rather than verbal extrapolation and wishful thinking.

5.3 Easier extrapolation problems

Extrapolating from the complier group represented in the LATE (6%) to the entire population represented in the ATE is a heroic challenge. Changing the target parameter to something less ambitious makes the extrapolation problem easier. For example, one could consider

extrapolated LATEs, i.e. generalized LATEs (23) with $u_{\text{lb}} = \max\{p(0) - \alpha, 0\}$ and $u_{\text{ub}} = \min\{p(1) + \alpha, 1\}$ for different non-negative values of α (Mogstad et al., 2018, Section 4.2). For $\alpha = 0$, the extrapolated LATE is equivalent to the usual LATE, while as $\alpha \to \max\{p(0), 1 - p(1)\}$, it returns to the ATE.[22]

```
# Set up ivmte arguments as a list so they can be easily changed
args <- list(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "genlate",
  m0 = ~ u + I(u^2),
  m1 = ~ u + I(u^2),
  propensity = morekids ~ samesex,
  audit.nu = 200
)

# Get propensity score and construct alpha list
p <- predict(lm(data = AE, morekids ~ samesex),
  newdata = data.frame(samesex = c(0, 1)),
  type = "response"
)
alphalist <- seq(from = 0, to = max(p[1], (1 - p[2])), by = .01)

# Function for computing genlate bounds at different values
loopivmte <- function(args, alphalist) {
  df.lb <- data.frame(alpha = alphalist, value = NA, type = "lb")
  df.ub <- data.frame(alpha = alphalist, value = NA, type = "ub")
  for (i in 1:length(alphalist)) {
    args[["genlate.lb"]] <- max(p[1] - alphalist[i], 0)
    args[["genlate.ub"]] <- min(p[2] + alphalist[i], 1)
    r <- do.call(ivmte, args)
    df.lb$value[i] <- r$bound[1]
    df.ub$value[i] <- r$bound[2]
  }
  return(rbind(df.lb, df.ub))
}

# Run the quadratic case
plotquadratic <- loopivmte(args, alphalist)
plotquadratic$name <- "Quadratic"

# Run the quartic case
```

22. These calls use the `genlate`, `genlate.lb`, and `genlate.ub` parameters, which are discussed in Section 4.3.2, as well as the spline functionality discussed in Section 4.4.

```
args[["m0"]] <- ~ u + I(u^2) + I(u^3) + I(u^4)
args[["m1"]] <- args[["m0"]]
plotquartic <- loopivmte(args, alphalist)
plotquartic$name <- "Quartic"

# Run the spline case
args[["m0"]] <- ~ uSplines(degree = 3, knots = seq(from = .1, to = .9, by = .1))
args[["m1"]] <- args[["m0"]]
plotspline <- loopivmte(args, alphalist)
plotspline$name <- "Cubic spline"
```

```
library("ggplot2")
plotdf <- rbind(plotquadratic, plotquartic, plotspline)
ggplot(plotdf, aes(x = alpha, y = value, color = name)) +
  geom_line(data = subset(plotdf, type == "lb")) +
  geom_line(data = subset(plotdf, type == "ub")) +
  labs(
    x = expression(paste("Extrapolation distance (", alpha, ")")),
    y = "Bounds",
    color = "MTR function"
  ) +
  theme(legend.position = "bottom")
```

Figure 1: *Bounds as a function of assumptions and extrapolation difficulty*

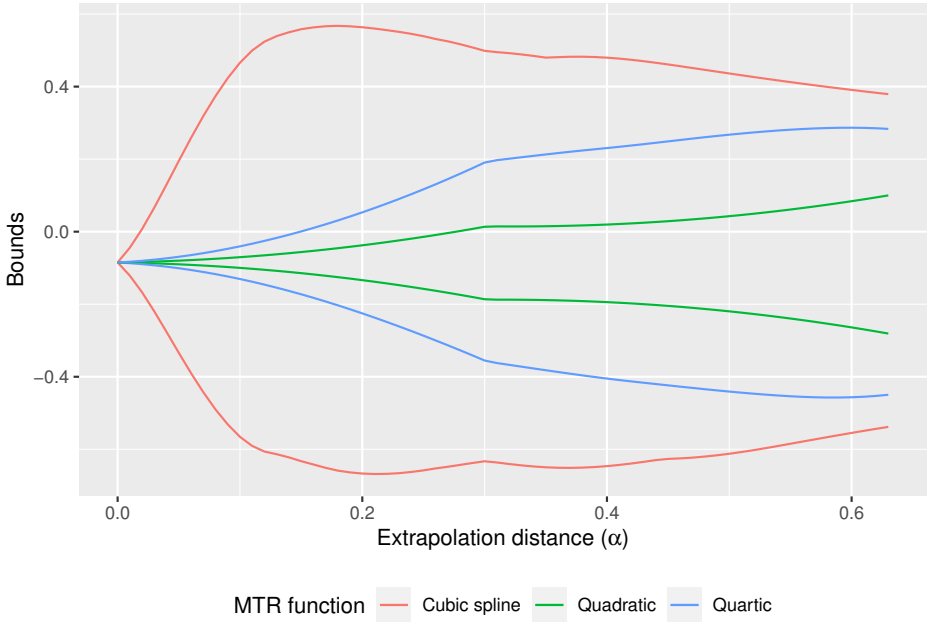


Figure 1 reports bounds on extrapolated LATEs as a function of α for three different specifications of the MTR function. The tightest specification is the unconstrained quadratic used above. The quartic specification takes the quadratic and adds third and fourth order terms to the MTR function for both treatment states. The spline specification is a flexible cubic spline with nine knots.

As expected, the bounds are always ordered in width with quadratic being narrowest and the cubic spline being widest. For all specifications, the bounds start as a point at $\alpha = 0$ (the LATE) and tend towards the ATE bounds as $\alpha \to 1$. This shows how the MTE framework allows the researcher to achieve bounds of any width they desire, while still being constrained by the reality that stronger conclusions require stronger assumptions. Given this freedom, it is unlikely that the researcher's preferred trade-off between assumptions and conclusions is the corner solution of reporting only nonparametrically point-identified parameters such as the LATE, which reflect both the weakest assumptions and the weakest conclusions.

5.4 Covariates

Covariates (X_i) serve two roles in all IV strategies. First, they can increase the credibility of the assumption that the instrument is as good as randomly assigned by making that assumption conditional on other observables. Second, covariates can reduce sampling uncertainty to the extent that they soak up residual variation in the outcome and/or treatment variables. In the MTE framework, covariates can also be used in a third role to provide identifying content through separability (e.g. Carneiro et al., 2011; Brinch et al., 2012, 2017).

To demonstrate this, we return to the quadratic specification with the ATE as the target parameter, but now we fully interact the MTE specification in `yob`, viewed here as X_i:

```
set.seed(1234) # the covariate part of the audit grid is stochastic
ivmte(
  data = AE,
  ivlike = worked ~ (morekids + samesex + morekids * samesex) * yob,
  target = "ate",
  m0 = ~ u + yob + u * yob + I(u^2) + I(u^2) * yob,
  m1 = ~ u + yob + u * yob + I(u^2) + I(u^2) * yob,
  propensity = morekids ~ yob + samesex + samesex * yob
)
```

```
##
## Bounds on the target parameter: [-0.2790478, 0.09365855]
## Audit terminated successfully after 1 round
```

The bounds are quite similar to the previous bounds that we obtained without covariates. This is expected because for each new interacted moment that is being matched we are adding an interaction term in the MTR that must be fit. Eliminating one or more of these interaction terms imposes *separability*, that is, the assumption that unobserved heterogeneity in potential outcomes operates similarly for different values of the covariate. Here we eliminate the quadratic interaction and see that the bounds narrow considerably.

```
set.seed(1234)
ivmte(
  data = AE,
  ivlike = worked ~ (morekids + samesex + morekids * samesex) * yob,
  target = "ate",
  m0 = ~ u + yob + u * yob + I(u^2),
  m1 = ~ u + yob + u * yob + I(u^2),
  propensity = morekids ~ yob + samesex + samesex * yob
)
```

```
##
## Bounds on the target parameter: [-0.1206799, 0.03139476]
## Audit terminated successfully after 1 round
```

With multiple types of assumptions to impose there is naturally a trade-off. For example, we might want to use the information we obtain with separability to buy a more flexible functional form.

```
set.seed(1234)
ivmte(
  data = AE,
  ivlike = worked ~ (morekids + samesex + morekids * samesex) * yob,
  target = "ate",
  m0 = ~ 0 + uSplines(degree = 3, knots = seq(from = .25, to = .75, by = .25)) + yob,
  m1 = ~ 0 + uSplines(degree = 3, knots = seq(from = .25, to = .75, by = .25)) + yob,
  propensity = morekids ~ yob + samesex + samesex * yob
)
```

```
##
## Bounds on the target parameter: [-0.2945988, 0.1704043]
## Audit terminated successfully after 2 rounds
```

The regression approach starts to become particularly attractive in rich specifications with multiple different covariates. This is because the number of possible moments that could be matched blows up; using all of them is potentially unwise due to small-sample bias, but it is also not necessarily clear how to choose which ones to use. The regression approach removes this choice through the usual least squares weighting. For example:

```
set.seed(1234)
ivmte(
  data = AE,
  outcome = worked,
  target = "ate",
  m0 = ~ 0 + uSplines(degree = 3, knots = seq(from = .25, to = .75, by = .25)) + yob +
    black + hisp + other,
  m1 = ~ 0 + uSplines(degree = 3, knots = seq(from = .25, to = .75, by = .25)) + yob +
    black + hisp + other,
  propensity = morekids ~ samesex + yob + black + hisp + other,
  solver = "gurobi"
)
```

```
##
## Bounds on the target parameter: [-0.2958156, 0.1643243]
## Audit terminated successfully after 2 rounds
```

5.5 Run time

In this section, we provide a sense of the run time involved in `ivmte`. The following benchmarks were performed with a Intel Xeon W-2125 processor. The `AE` dataset has 209,133 observations.

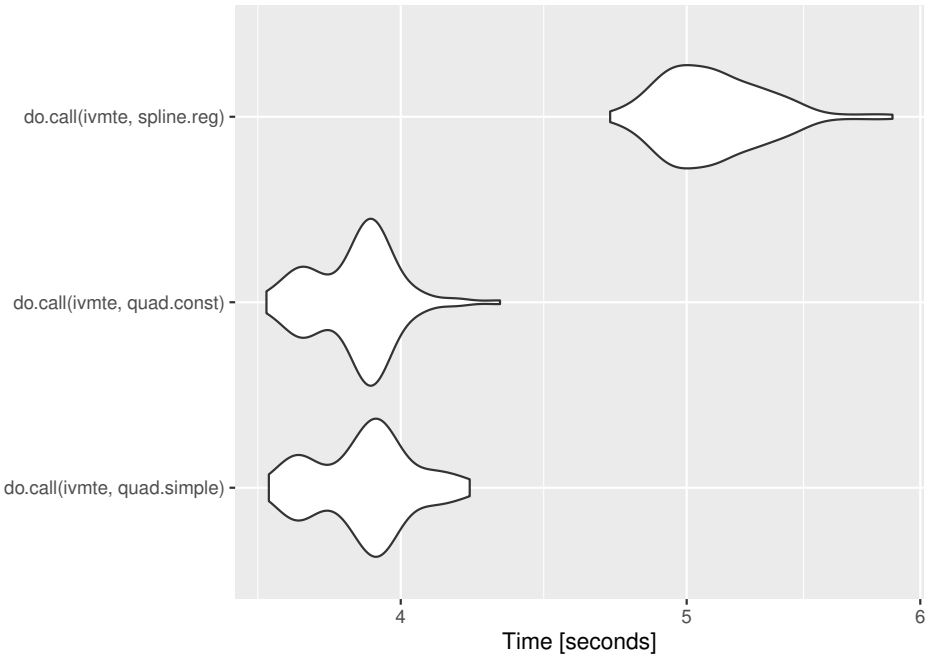
```
library("microbenchmark")
```

```
quad.simple <-
  list(
    data = AE,
    ivlike = c(worked ~ morekids + samesex + morekids * samesex),
    target = "ate",
    propensity = morekids ~ samesex
  )
quad.simple[["m0"]] <- ~ u + I(u^2)
quad.simple[["m1"]] <- quad.simple[["m0"]]
quad.const <- quad.simple
quad.const[["mte.inc"]] <- TRUE
quad.const[["mte.ub"]] <- 0
spline.reg <-
  list(
    data = AE,
    outcome = "worked",
    target = "ate",
    m0 = ~ 0 + uSplines(degree = 3, knots = seq(from = .25, to = .75, by = .25)) +
      yob + black + hisp + other,
```

```
    propensity = morekids ~ samesex + yob + black + hisp + other,
    solver = "gurobi",
    audit.nu = 50
  )
spline.reg[["m1"]] <- spline.reg[["m0"]]
m <- microbenchmark(
  do.call(ivmte, quad.simple), # simple, unconstrained
  do.call(ivmte, quad.const), # add some constraints
  do.call(ivmte, spline.reg), # more complex specification
  times = 100
)
```

```
autoplot(m)
```

Figure 2: *Run time distributions for three specifications*



5.6 Confidence intervals

To conclude, we demonstrate how `ivmte` constructs confidence intervals (Section 4.9). If we estimate the model assuming point identification, as with a linear specification, `ivmte` returns:

```
set.seed(1234) # the bootstrap is stochastic
r <- ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~u,
  m1 = ~u,
  point = TRUE,
  bootstraps = 100,
  propensity = morekids ~ samesex
)
summary(r)
```

```
##
## Point estimate of the target parameter: -0.07791036
## MTR coefficients: 4
## Independent/total moments: 4/4
##
## Bootstrapped confidence intervals (nonparametric):
##     90%: [-0.150156, -0.006348619]
##     95%: [-0.161036, 0.003467422]
##     99%: [-0.1650715, 0.02561266]
## p-value: 0.08
## Number of bootstraps: 100
```

While for the general case of bound estimation, `ivmte` returns:

```
set.seed(1234)
r <- ivmte(
  data = AE,
  ivlike = worked ~ morekids + samesex + morekids * samesex,
  target = "ate",
  m0 = ~ u + I(u^2),
  m1 = ~ u + I(u^2),
  mte.inc = TRUE,
  mte.ub = 0,
  bootstraps = 100,
  propensity = morekids ~ samesex
)
summary(r)
```

```
##
## Bounds on the target parameter: [-0.08484221, -0.06323574]
## Audit terminated successfully after 1 round
## MTR coefficients: 6
```

```
## Independent/total moments: 4/4
## Minimum criterion: 0
## Solver: Gurobi ('gurobi')
##
## Bootstrapped confidence intervals (backward):
##     90%: [-0.1409956, -0.01898701]
##     95%: [-0.14766, -0.009421106]
##     99%: [-0.1587216, 2.775558e-17]
## p-value: 0.02
## Number of bootstraps: 100
```

Acknowledgments

This research was supported in part by National Science Foundation grants SES-1530538 and SES-1846832. We thank Christine Blandhol and John Bonney for testing the `ivmte` package. We also thank two anonymous referees for their helpful comments.

References

Amanda Agan, Jennifer Doleac, and Anna Harvey. Misdemeanor Prosecution. Technical Report w28600, National Bureau of Economic Research, Cambridge, MA, March 2021. 3

Donald W. K. Andrews and Sukjin Han. Invalidity of the bootstrap and the m out of n bootstrap for confidence interval endpoints defined by moment inequalities. *Econometrics Journal*, 12:S172–S199, 2009. ISSN 1368-423X. 22

Joshua D. Angrist and William N. Evans. Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *The American Economic Review*, 88 (3):450–477, June 1998. ISSN 00028282. 2, 22

Joshua D. Angrist and Alan B. Krueger. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives*, 15(4):69–85, October 2001. ISSN 08953309. 2

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009. 2

Joshua D Angrist and Jrn-Steffen Pischke. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2):3–30, May 2010. doi: 10.1257/jep.24.2.3. 2

David Arnold, Will Dobbie, and Crystal S Yang. Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, 133(4):1885–1932, May 2018. doi: 10.1093/qje/qjy012. 3

David Arnold, Will S Dobbie, and Peter Hull. Measuring Racial Discrimination in Bail Decisions. Working Paper 26999, National Bureau of Economic Research, April 2020. 3

David Autor, Andreas Kostøl, Magne Mogstad, and Bradley Setzler. Disability benefits, consumption insurance, and household labor supply. *American Economic Review*, 109(7): 2613–54, July 2019. doi: 10.1257/aer.20151231. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151231. 3

Michael Baiocchi, Jing Cheng, and Dylan S. Small. Instrumental variable methods for causal inference. *Statistics in Medicine*, 33(13):2297–2340, 2014. ISSN 1097-0258. doi: 10.1002/sim.6128. 1

Burt S Barnow, Glen George Cain, Arthur Stanley Goldberger, et al. *Issues in the Analysis of Selectivity Bias*. University of Wisconsin, Inst. for Research on Poverty, 1980. 4

Gary S. Becker and Barry R. Chiswick. Education and the Distribution of Earnings. *The American Economic Review*, 56(1/2):358–369, 1966. ISSN 00028282. 2

Manudeep Bhuller, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad. Incarceration, Recidivism, and Employment. *Journal of Political Economy*, 128(4):1269–1324, April 2020. ISSN 0022-3808. doi: 10.1086/705330. 3

Anders Björklund and Robert Moffitt. The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. *The Review of Economics and Statistics*, 69(1):42–49, February 1987. ISSN 00346535. 2

Christine Blandhol, John Bonney, Magne Mogstad, and Alexander Torgovitsky. When is TSLS Actually LATE? Technical Report w29709, National Bureau of Economic Research, Cambridge, MA, January 2022. 2

Kenneth A. Bollen. Instrumental Variables in Sociology and the Social Sciences. *Annual Review of Sociology*, 38(1):37–72, August 2012. ISSN 0360-0572, 1545-2115. doi: 10.1146/annurev-soc-081309-150141. 1

Christian N. Brinch, Magne Mogstad, and Matthew Wiswall. Beyond LATE with a Discrete Instrument. *Working paper*, 2012. 3, 25, 29

Christian N. Brinch, Magne Mogstad, and Matthew Wiswall. Beyond LATE with a Discrete Instrument. *Journal of Political Economy*, 125(4):985–1039, August 2017. doi: 10.1086/692712. 3, 25, 29

Bryan W. Brown and Whitney K. Newey. Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference. *Journal of Business & Economic Statistics*, 20 (4):507–517, October 2002. ISSN 0735-0015. doi: 10.1198/073500102288618649. 21

Federico A. Bugni, Ivan A. Canay, and Xiaoxia Shi. Specification tests for partially identified models defined by moment inequalities. *Journal of Econometrics*, 185(1):259–282, March 2015. ISSN 0304-4076. 22

Ivan A. Canay and Azeem M. Shaikh. Practical and Theoretical Advances in Inference for Partially Identified Models. In Bo Honore, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, editors, *Advances in Economics and Econometrics*, pages 271–306. Cambridge University Press, 2017. doi: 10.1017/9781108227223.009. 21

Pedro Carneiro and Sokbae Lee. Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics*, 149(2):191–208, April 2009. ISSN 0304-4076. 3

Pedro Carneiro, James J. Heckman, and Edward J. Vytlacil. Estimating Marginal Returns to Education. *American Economic Review*, 101(6):2754–81, 2011. doi: 10.1257/aer.101.6.2754. 3, 29

Pedro Carneiro, Michael Lokshin, and Nithin Umapathi. Average and Marginal Returns to Upper Secondary Schooling in Indonesia. *Journal of Applied Econometrics*, 32(1):16–36, May 2016. doi: 10.1002/jae.2523. 3

Xiaohong Chen. Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume Volume 6, Part 2, pages 5549–5632. Elsevier, 2007. 9

Thomas Cornelissen, Christian Dustmann, Anna Raute, and Uta Schönberg. Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy*, 126(6):2356–2409, December 2018. ISSN 0022-3808, 1537-534X. doi: 10.1086/699979. 3

Øystein Daljord, Carl F. Mela, Jason M.T. Roos, Jim Sprigg, and Song Yao. The Design and Targeting of Compliance Promotions. *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.4006675. 3

Angus Deaton. Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48(2):424–455, June 2010. ISSN 0022-0515. doi: 10.1257/jel.48.2.424. 2

Domenico Depalo. Explaining the causal effect of adherence to medication on cholesterol through the marginal patient. *Health Economics*, n/a(n/a), 2020. ISSN 1099-1050. doi: 10.1002/hec.4030. 3

Joseph J. Doyle Jr. Child Protection and Child Outcomes: Measuring the Effects of Foster Care. *The American Economic Review*, 97(5):1583–1610, 2007. 3

Deniz Dutz, Ingrid Huitfeldt, Santiago Lacouture, Magne Mogstad, Alexander Torgovitsky, and Winnie van Dijk. Selection in Surveys. Technical Report w29549, National Bureau of Economic Research, Cambridge, MA, December 2021. 3

Christina Felfe and Rafael Lalive. Does early child care affect children's development? *Journal of Public Economics*, 159:33–53, March 2018. ISSN 0047-2727. doi: 10.1016/j.jpubeco.2018.01.014. 3

Eric French and Jae Song. The Effect of Disability Insurance Receipt on Labor Supply. *American Economic Journal: Economic Policy*, 6(2):291–337, May 2014. ISSN 1945-774X. doi: 10.1257/pol.6.2.291. 3

Christina Gathmann, Christina Vonnahme, Anna Busse, and Jongoh Kim. *Marginal Returns to Citizenship and Educational Performance*. RWI, DE, October 2021. ISBN 978-3-96973-066-9. 3

Felipe Goncalves and Steven Mello. Should the Punishment Fit the Crime? Deterrence and Retribution in Law Enforcement. *Working Paper*, 2022. 3

Gurobi Optimization, Inc. Gurobi Optimizer Reference Manual. 2015. 14

Peter Hall and Joel L. Horowitz. Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators. *Econometrica*, 64(4):891–916, July 1996. ISSN 00129682. 21, 22

Lars Peter Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, July 1982. ISSN 00129682. 11, 22

James Heckman. Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *The Journal of Human Resources*, 32(3):441–462, 1997. ISSN 0022166X. 2

James J. Heckman. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 1976. 2

James J. Heckman and Richard Robb. Alternative methods for evaluating the impact of interventions. In James J. Heckman and Burton Singer, editors, *Longitudinal Analysis of Labor Market Data*. Cambridge University Press, 1985. 1, 2

James J. Heckman and Edward Vytlacil. Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3):669–738, 2005. ISSN 1468-0262. 1, 3, 5, 6, 15, 16

James J. Heckman and Edward J. Vytlacil. 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, April 1999. ISSN 00278424. 1, 3, 5

James J. Heckman and Edward J. Vytlacil. Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume Volume 6, Part 2, pages 4779–4874. Elsevier, 2007a. 3, 5

James J. Heckman and Edward J. Vytlacil. 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 James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume Volume 6, Part 2, pages 4875–5143. Elsevier, 2007b. 3, 5

James J. Heckman, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation. *Proceedings of the National Academy of Sciences*, 93(23):13416–13420, 1996. 4

James J Heckman, Sergio Urzua, and Edward Vytlacil. Understanding Instrumental Variables in Models with Essential Heterogeneity. *Review of Economics and Statistics*, 88(3):389–432, 2006. doi: 10.1162/rest.88.3.389. 4

Eskil Heinesen and Elise Stenholt Lange. Vocational versus General Upper Secondary Education and Earnings. *Journal of Human Resources*, pages 0221–11497R2, May 2022. ISSN 0022-166X, 1548-8004. doi: 10.3368/jhr.0221-11497R2. 3

Leonid Hurwicz. Systems with Nonadditive Disturbances. In T.C. Koopmans, editor, *Cowles 10*, number 10 in Cowles Commission Monographs, pages 410–418. 1950. 2

IBM. *IBM ILOG AMPL Version 12.2*. International Business Machines Corporation, 2010. 14

Guido W. Imbens. Instrumental Variables: An Econometrician's Perspective. *Statistical Science*, 29(3):323–358, August 2014. ISSN 0883-4237, 2168-8745. doi: 10.1214/14-STS480. 1

Guido W. Imbens and Joshua D. Angrist. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475, March 1994. ISSN 00129682. 1, 2, 3, 4, 15

Koichiro Ito, Takanori Ida, and Makoto Tanaka. Selection on Welfare Gains: Experimental Evidence from Electricity Plan Choice. *Working Paper*, 2022. 3

Edward H. Kennedy, Scott Lorch, and Dylan S. Small. Robust causal inference with continuous instruments using the local instrumental variable curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(1):121–143, 2019. ISSN 1467-9868. doi: 10.1111/rssb.12300. 16

Christian Kleiber and Achim Zeileis. *AER: Applied Econometrics with R*, 2018. URL https://CRAN.R-project.org/package=AER. R package version 1.2-6. 19

Patrick Kline and Christopher R. Walters. Evaluating Public Programs with Close Substitutes: The Case of Head Start*. *The Quarterly Journal of Economics*, 131(4):1795–1848, November 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw027. 3

Kjell Konis. *lpSolveAPI: R Interface to 'lp_solve'*, 2019. URL https://CRAN.R-project.org/package=lpSolveAPI. R package version 5.5.2.0-17.4. 14

Amanda E Kowalski. Behavior within a Clinical Trial and Implications for Mammography Guidelines. Working Paper 25049, National Bureau of Economic Research, September 2018. 3

Nicole Maestas, Kathleen J. Mullen, and Alexander Strand. Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. *The American Economic Review*, 103(5):1797–1829, 2013. ISSN 00028282. 3

Charles F. Manski. Nonparametric Bounds on Treatment Effects. *The American Economic Review*, 80(2):319–323, May 1990. ISSN 00028282. 2, 24

Rosa L. Matzkin. Chapter 73 Nonparametric identification. In James J. Heckman and Edward E. Leamer, editors, *Handbook of Econometrics*, volume Volume 6, Part 2, pages 5307–5368. Elsevier, 2007. 5

Shaibu Mellon Bedi, Carlo Azzarri, Bekele Hundie Kotu, Lukas Kornher, and Joachim von Braun. Scaling-up agricultural technologies: Who should be targeted? *European Review of Agricultural Economics*, page jbab054, December 2021. ISSN 0165-1587, 1464-3618. doi: 10.1093/erae/jbab054. 3

robert Moffitt. Estimating Marginal Treatment Effects in Heterogeneous Populations. *Annales d'Economie et de Statistique*, (91/92):239–261, 2008. ISSN 0769489X, 22726497. 3

Robert A Moffitt. The Marginal Labor Supply Disincentives of Welfare Reforms. Working Paper 26028, National Bureau of Economic Research, June 2019. 3

Magne Mogstad and Alexander Torgovitsky. Identification and Extrapolation of Causal Effects with Instrumental Variables. *Annual Review of Economics*, 10(1), May 2018. doi: 10.1146/annurev-economics-101617-041813. 2, 5, 13, 16

Magne Mogstad, Andres Santos, and Alexander Torgovitsky. Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters. *NBER Working Paper*, 2017. 3

Magne Mogstad, Andres Santos, and Alexander Torgovitsky. Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters. *Econometrica*, 86(5):1589–1619, 2018. doi: 10.3982/ecta15463. 3, 6, 8, 12, 15, 17, 25, 27

Magne Mogstad, Alexander Torgovitsky, and Christopher R. Walters. The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables. *American Economic Review*, 111(11):3663–3698, November 2021. ISSN 0002-8282. doi: 10.1257/aer.20190221. 4

MOSEK ApS. MOSEK Licensing Guide. 2021. 14

Martin Nybom. The Distribution of Lifetime Earnings Returns to College. *Journal of Labor Economics*, pages 000–000, June 2017. doi: 10.1086/692475. 3

Elizabeth L. Ogburn, Andrea Rotnitzky, and James M. Robins. Doubly robust estimation of the local average treatment effect curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):373–396, 2015. ISSN 1467-9868. doi: 10.1111/rssb.12078. 16

Judea Pearl. Principal Stratification – a Goal or a Tool? *The International Journal of Biostatistics*, 7(1):1–13, March 2011. ISSN 1557-4679. doi: 10.2202/1557-4679.1322. 2

Vitor Possebom. Crime and Mismeasured Punishment: Marginal Treatment Effect with Misclassification, July 2022. 3

James M. Robins and Sander Greenland. Identification of Causal Effects Using Instrumental Variables: Comment. *Journal of the American Statistical Association*, 91(434):456–458, 1996. ISSN 0162-1459. doi: 10.2307/2291630. 2

Mayo Roettger, Gabriel Gelius-Dietrich, and Jonathan C. Fritzemeier. *cplexAPI: R Interface to C API of IBM ILOG CPLEX*, 2019. URL https://CRAN.R-project.org/package=cplexAPI. R package version 1.3.6. 14

Evan K. Rose and Yotam Shem-Tov. How Does Incarceration Affect Reoffending? Estimating the Dose-Response Function. *Journal of Political Economy*, pages 000–000, July 2021. ISSN 0022-3808. doi: 10.1086/716561. 3

Paul R. Rosenbaum and Donald B. Rubin. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1):41–55, April 1983. ISSN 00063444. 4

Herman Rubin. Note on Random Coefficients. In T.C. Koopmans, editor, *Cowles 10*, number 10 in Cowles Commission Monographs, pages 419–421. 1950. 2

Mare Sarr, Mintewab Bezabih Ayele, Mumbi E. Kimani, and Remidius Ruhinduka. Who benefits from climate-friendly agriculture? The marginal returns to a rainfed system of rice intensification in Tanzania. *World Development*, 138:105160, February 2021. ISSN 0305750X. doi: 10.1016/j.worlddev.2020.105160. 3

James H Stock and Francesco Trebbi. Retrospectives: Who Invented Instrumental Variable Regression? *Journal of Economic Perspectives*, 17(3):177–194, August 2003. doi: 10.1257/089533003769204416. 2

Sonja A. Swanson and Miguel A. Hernán. Think Globally, Act Globally: An Epidemiologist's Perspective on Instrumental Variable Estimation. *Statistical Science*, 29(3):371–374, August 2014. ISSN 0883-4237, 2168-8745. doi: 10.1214/14-STS491. 2

Edward Vytlacil. Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica*, 70(1):331–341, January 2002. ISSN 00129682. 4

Christopher R. Walters. The Demand for Effective Charter Schools. *Journal of Political Economy*, 126(6):2179–2223, December 2018. ISSN 0022-3808. doi: 10.1086/699980. 3

Wenjie Wang and Jun Yan. *splines2: Regression Spline Functions and Classes*, 2018. URL https://CRAN.R-project.org/package=splines2. R package version 0.2.8. 17