The Stata Journal (2024)
24, Number 1, pp. 72–92 DOI: 10.1177/1536867X241233645

Ordinary least squares and instrumental-variables estimators for any outcome and heterogeneity

Myoung-jae Lee
Department of Economics
Korea University
Seoul, Republic of Korea
myoungjae@korea.ac.kr

Chirok Han
Department of Economics
Korea University
Seoul, Republic of Korea
chirokhan@korea.ac.kr

Abstract. Given an exogenous treatment d and covariates $\mathbf{x}$, an ordinary least-squares (OLS) estimator is often applied with a noncontinuous outcome y to find the effect of d, despite the fact that the OLS linear model is invalid. Also, when d is endogenous with an instrument z, an instrumental-variables estimator (IVE) is often applied, again despite the invalid linear model. Furthermore, the treatment effect is likely to be heterogeneous, say, $\mu_1(\mathbf{x})$, not a constant as assumed in most linear models. Given these problems, the question is then what kind of effect the OLS and IVE actually estimate. Under some restrictive conditions such as a "saturated model", the estimated effect is known to be a weighted average, say, $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$, but in general, OLS and the IVE applied to linear models with a noncontinuous outcome or heterogeneous effect fail to yield a weighted average of heterogeneous treatment effects. Recently, however, it has been found that $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$ can be estimated by OLS and the IVE without those restrictive conditions if the "propensity-score residual" $d - E(d|\mathbf{x})$ or the "instrument-score residual" $z - E(z|\mathbf{x})$ is used. In this article, we review this recent development and provide a command for OLS and the IVE with the propensity- and instrument-score residuals, which are applicable to any outcome and any heterogeneous effect.

Keywords: st0740, psr, OLS, IVE, propensity score, instrument score, overlap weight

1 Introduction

Given an exogenous binary treatment d, an outcome y, and covariates $\mathbf{x}$, consider the ordinary least-squares (OLS) estimator of y on $(\mathbf{x}, d)$ for a linear model $y = \boldsymbol{\beta}_{\mathbf{x}}'\mathbf{x} + \beta_d d + \text{error}$, where $\boldsymbol{\beta}_{\mathbf{x}}$ and β_d are parameters. In reality, the treatment effect may not be a constant (β_d), as is assumed in the linear model, but an unknown heterogeneous function, say, $\mu_1(\mathbf{x})$, that results in $\mu_1(\mathbf{x})d$ instead of $\beta_d d$. This is clear if one thinks of the COVID-19 vaccine effects, which can be highly heterogeneous; some individuals suffer from extremely negative effects (that is, death), while others benefit from fully positive effects. It is sometimes possible to account for $\mu_1(\mathbf{x})$ using interaction terms $d\mathbf{x}$, but approximating $\mu_1(\mathbf{x})$ with $d\mathbf{x}$ would generally be inadequate. Hence, an important question arises: What kind of effect does the OLS slope of d estimate when the true effect is an unknown function $\mu_1(\mathbf{x})$ of $\mathbf{x}$?

An answer is that the OLS d-slope is consistent for a $\text{Var}(d|\mathbf{x})$-weighted average of $\mu_1(\mathbf{x})$, where $\text{Var}(d|\mathbf{x})$ denotes the variance of d conditional on $\mathbf{x}$ (Angrist 1998; Angrist and Krueger 1999; Angrist and Pischke 2009; and Aronow and Samii 2016). Unfortunately, however, this answer requires that $E(d|\mathbf{x})$ be the same as the linear projection $L(d|\mathbf{x})$, that is, the "population OLS predictor" of d using $\mathbf{x}$.

Because $E(d|\mathbf{x}) = L(d|\mathbf{x})$ hardly holds in reality unless the model is saturated (that is, $\mathbf{x}$ is discrete, and a full set of dummy variables is used for all possible values of $\mathbf{x}$), another question emerges: What effect does the OLS d-slope estimate under $E(d|\mathbf{x}) \neq L(d|\mathbf{x})$, when the true effect is $\mu_1(\mathbf{x})$? In addition to this effect-heterogeneity problem, y is often a limited dependent variable (LDV) for which the linear model does not hold in general. This raises yet another question: What effect does the OLS d-slope estimate under $E(d|\mathbf{x}) \neq L(d|\mathbf{x})$, when the true effect is $\mu_1(\mathbf{x})$ or y is an LDV?

Going one step further, when d is endogenous with a binary instrument z, an instrumental-variables estimator (IVE) is often applied to an LDV y. Then an analogous question emerges: What effect does the IVE d-slope estimate under $E(d|\mathbf{x}) \neq L(d|\mathbf{x})$, when the true effect is $\mu_1(\mathbf{x})$ or y is an LDV? Instead of simply raising the "passive" or "negative" questions on the (biased) estimands of OLS and the IVE in difficult situations, a more "active" or "positive" question might be, Is there any way to do a valid OLS or IVE without $E(d|\mathbf{x}) = L(d|\mathbf{x})$ to estimate a meaningful treatment effect when the true effect is $\mu_1(\mathbf{x})$ or y is an LDV?

This article reviews the answers to the above questions as presented in Lee (2018, 2021) and Lee, Lee, and Choi (Forthcoming) and provides the command `psr`, which implements the positive approaches in Lee (2018) for OLS and Lee (2021) for the IVE. In essence, the approaches are that modified versions of OLS and the IVE are consistent for some specific weighted averages of $\mu_1(\mathbf{x})$ for any form of y. The generality of the approaches is that they allow for an arbitrary $\mu_1(\mathbf{x})$ and an arbitrary form of y, either of which would invalidate the usual linear model.

The approach by Lee (2018, 2021) is related to, but more generally applicable than, that of Angrist (2001), who advocates declaring the effect of interest first and then finding ways to identify and estimate it—a point also made by van der Laan and Rose (2011) later in the name "targeted learning". Angrist (2001) argues that simple linear estimators such as the IVE can be appropriately used even with noncontinuous outcome variables. However, such linear model estimators may exhibit bias unless $E(d|\mathbf{x}) = L(d|\mathbf{x})$, which hardly holds for general nonsaturated models. Lee, Lee, and Choi (Forthcoming) derive the bias in the context of linear probability models. Though Angrist (2001, 10) mentions the use of nonlinear least squares as a solution for LDVs, the implementation is practically inconvenient and often challenging because it may involve nuisance infinite weights. Readers are referred to the comments by Moffitt (2001), Imbens (2001), and Todd (2001) for other concerns.

In contrast, the approach by Lee (2018, 2021) presented below yields estimators that are always a weighted average of heterogeneous treatment effects, regardless of whether $E(d|\mathbf{x}) = L(d|\mathbf{x})$. Furthermore, their approach is applicable to any form of y and is straightforward to implement because it involves only linear regressions (for OLS and IVE) and widely used probit or logit.

In the remainder of this article, section 2 closely examines the above questions and answers in Lee (2018, 2021) and Lee, Lee, and Choi (Forthcoming). Section 3 presents the command `psr`, which performs the modified OLS and IVE for any y (continuous, binary, count, etc.) and estimates weighted averages of $\mu_1(\mathbf{x})$. Section 4 provides two empirical illustrations. Finally, section 5 concludes this article.

2 OLS and IVE valid for any outcome

2.1 Effect estimated by OLS

For the heterogeneous effect, $\mu_1(\mathbf{x}) \equiv E(y^1 - y^0|\mathbf{x})$, where y^1 and y^0 are the potential outcomes, Angrist and Pischke (2009, eq. 3.3.7) show that the slope estimator $\widehat{\beta}_d$ of d in the OLS regression of y on $(\mathbf{x}, d)$ is consistent for the weighted average

$$E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}, \;\; \omega(\mathbf{x}) \equiv \frac{\mathrm{Var}(d|\mathbf{x})}{E\{\mathrm{Var}(d|\mathbf{x})\}} = \frac{\pi_{\mathbf{x}}(1-\pi_{\mathbf{x}})}{E\{\pi_{\mathbf{x}}(1-\pi_{\mathbf{x}})\}}, \;\; \pi_{\mathbf{x}} \equiv E(d|\mathbf{x}) \qquad (1)$$

if $\pi_{\mathbf{x}} = \lambda_{\mathbf{x}} \equiv L(d|\mathbf{x}) \equiv L(d\mathbf{x}')\{E(\mathbf{x}\mathbf{x}')\}^{-1}\mathbf{x}$, where $\pi_{\mathbf{x}}$ is the propensity score (PS) and $\mathrm{Var}(d|\mathbf{x})$ is the variance of d conditional on $\mathbf{x}$.

The quadratic function $\pi_{\mathbf{x}}(1-\pi_{\mathbf{x}})$ reaches its maximum at $\pi_{\mathbf{x}} = 0.5$ and its minimum at $\pi_{\mathbf{x}} = 0, 1$. Considering the popular "PS matching" targeting for $E\{E(y^1 - y^0|\pi_{\mathbf{x}})\}$, because those subjects with $\pi_{\mathbf{x}} \simeq 0.5$ overlap well with subjects in the opposite group, whereas those with $\pi_{\mathbf{x}} \simeq 0, 1$ do not, Li, Morgan, and Zaslavsky (2018) called $\omega(\mathbf{x})$ the overlap weight (OW). That is, those subjects close to being randomized receive high weights in OW (Thomas, Li, and Pencina 2020). PS matching avoids the poor overlap problem by removing those with $\pi_{\mathbf{x}} \simeq 0, 1$, which amounts to targeting for $E\{\omega_{\mathrm{uni}}(\mathbf{x})E(y^1 - y^0|\pi_{\mathbf{x}})\}$, where $\omega_{\mathrm{uni}}(\mathbf{x})$ is a step-shaped "uniform weight" equal to 0 for $\pi_{\mathbf{x}} \simeq 0, 1$ and a positive constant otherwise. In this context, OW $\omega(\mathbf{x})$ can be viewed as a smoothed version of $\omega_{\mathrm{uni}}(\mathbf{x})$.

Using $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$ instead of $E\{\mu_1(\mathbf{x})\}$ as one marginal effect accords many benefits in causal analysis, such as 1) stabilizing "inverse probability weighting" estimators (Li and Greene 2013), 2) making "regression adjustment/imputation" estimators robust to misspecified outcome regression models (Vansteelandt and Daniel 2014), and 3) automatically ensuring the exact covariate balance when $\pi_{\mathbf{x}}$ is logistic (Li, Morgan, and Zaslavsky 2018). More advantages of OW can be seen in Choi and Lee (2023), who provide a review on OW.

OW is far more pervasive in the statistical, medical, and epidemiological literature than most researchers are aware. In recent years, in addition to the studies mentioned above, there appeared many studies advocating OW: Mao et al. (2018), Mao, Li, and Greene (2019), Li and Li (2019), Li, Thomas, and Li (2019), Mlcoch et al. (2019), Mao and Li (2020), and Cheng et al. (2022), among others. Bear in mind that OW $\omega(\mathbf{x})$ is not something artificial, because it is a smooth version of $\omega_{\text{uni}}(\mathbf{x})$ and it appears naturally in partialing out $\boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x}$ in $y = \boldsymbol{\beta}'_{\mathbf{x}}\mathbf{x} + \mu_1(\mathbf{x})d + \text{error}$.

Turning back to the Angrist and Pischke (2009) derivation, we see the convergence of the OLS estimator to the weighted average in (1) requires restrictively $\pi_{\mathbf{x}} = \lambda_{\mathbf{x}}$, where $\lambda_{\mathbf{x}}$ is the linear projection of d on $\mathbf{x}$ as defined above. To ensure $\pi_{\mathbf{x}} = \lambda_{\mathbf{x}}$, Angrist and Pischke (2009) assume a saturated model, but saturated models are rare.

In general, $\pi_{\mathbf{x}} \neq \lambda_{\mathbf{x}}$ (for example, probit for $\pi_{\mathbf{x}}$) and the OLS slope estimator may not converge to a weighted average of $\mu_1(\mathbf{x})$. For example, for binary y, Lee, Lee, and Choi (Forthcoming) show that

$$\widehat{\beta}_d \xrightarrow{p} a_{\text{ols}} + b_{\text{ols}}$$

where $\widehat{\beta}_d$ is the OLS d-slope estimator,

$$a_{\text{ols}} = E\left[\frac{\pi_{\mathbf{x}}(1-\pi_{\mathbf{x}}) + (\pi_{\mathbf{x}} - \lambda_{\mathbf{x}})^2}{E\{\pi_{\mathbf{x}}(1-\pi_{\mathbf{x}}) + (\pi_{\mathbf{x}} - \lambda_{\mathbf{x}})^2\}} \times \mu_1(\mathbf{x})\right]$$

is a weighted average of $\mu_1(\mathbf{x})$, and b_{ols} is a bias term. If $\pi_{\mathbf{x}} = \lambda_{\mathbf{x}}$, a_{ols} reduces to $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$ in (1) and $b_{\text{ols}} = 0$. Otherwise, $a_{\text{ols}} + b_{\text{ols}}$ is hard to interpret. The weighting function in a_{ols} makes little sense if $\pi_{\mathbf{x}} \neq \lambda_{\mathbf{x}}$, and the bias term b_{ols} involves $E(y^0|\mathbf{x})$ and $\mu_1(\mathbf{x})$ (see Lee, Lee, and Choi [Forthcoming, eq. 11]) for the exact formula of b_{ols}. In contrast, the methods proposed by Lee (2018, 2021) presented in the following sections always yield estimators that are consistent for a specific weighted average of treatment effects for any outcome and heterogeneity.

2.2 OLS with PS residual

For exogenous d, Lee (2018) shows that the OLS of y on $d - \pi_{\mathbf{x}}$ is consistent for $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$ for any form of y regardless of whether $\pi_{\mathbf{x}} = \lambda_{\mathbf{x}}$, as long as $y^1 - y^0$ makes sense. $y^1 - y^0$ makes sense for continuous, count, or binary y; for categorical y, turn each category to a dummy variable to use each dummy variable as an outcome. In fact, Lee's (2018) OLS was suggested much earlier by Robins, Mark, and Newey (1992) for a constant-effect semilinear model $y = \mu_0(\mathbf{x}) + \beta_d d + \text{error}$ with a continuous y and an unknown function $\mu_0(\mathbf{x})$.

Although $\pi_{\mathbf{x}}$ can be estimated nonparametrically, to make the OLS practical, Lee (2018) uses the probit $\pi_{\mathbf{x}}$ and proposed the OLS of $y - E(y|\pi_{\mathbf{x}})$ on $d - \pi_{\mathbf{x}}$. Using $y - E(y|\pi_{\mathbf{x}})$ instead of y makes the OLS robust to misspecifications in $\pi_{\mathbf{x}}$ to the extent that the OLS of $y - E(y|\pi_{\mathbf{x}})$ on $d - \pi_{\mathbf{x}}$ is close to the OLS of $y - E(y|\mathbf{x})$ on $d - \pi_{\mathbf{x}}$, which is "double debiased/orthogonalized" (Chernozhukov et al. 2017, 2018, 2022). Denote the OLS of $y - E(y|\pi_{\mathbf{x}})$ on the PS residual (PSR) $d - \pi_{\mathbf{x}}$ "OLS_{PSR}", and denote the OLS as "$\widehat{\beta}^q_{\text{PSR}}$"; q in $\widehat{\beta}^q_{\text{PSR}}$ is explained shortly.

To implement OLS_{PSR}, 1) obtain the probit of d on $\mathbf{x}$ to get the estimator $\widehat{\pi}_{\mathbf{x}} \equiv \Phi(\mathbf{x}'\widehat{\boldsymbol{\alpha}})$, where $\Phi(\cdot)$ is the standard normal distribution function and $\mathbf{x}'\boldsymbol{\alpha}$ is the probit regression function with a parameter $\boldsymbol{\alpha}$, 2) obtain the OLS $(\widehat{\gamma}_0, \ldots, \widehat{\gamma}_q)$ of y on $\{(\mathbf{x}'\widehat{\boldsymbol{\alpha}})^0, \ldots, (\mathbf{x}'\widehat{\boldsymbol{\alpha}})^q\}$ to get the predicted value $\sum_{j=0}^q \widehat{\gamma}_j(\mathbf{x}'\widehat{\boldsymbol{\alpha}})^j$ for $E(y|\pi_{\mathbf{x}})$—Lee (2018) suggested $q = 2$ or 3—and 3) do the OLS of $y - \sum_{j=0}^q \widehat{\gamma}_j(\mathbf{x}'\widehat{\boldsymbol{\alpha}})^j$ on $d - \widehat{\pi}_{\mathbf{x}}$:

$$\widehat{\beta}^q_{\text{PSR}} \equiv \frac{\sum_i (d_i - \widehat{\pi}_{\mathbf{x}_i}) \left\{ y_i - \sum_{j=0}^q \widehat{\gamma}_j (\mathbf{x}_i'\widehat{\boldsymbol{\alpha}})^j \right\}}{\sum_i (d_i - \widehat{\pi}_{\mathbf{x}_i})^2} \xrightarrow{p} \beta_{\text{PSR}} \equiv E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$$

Because $\Phi(\cdot)$ is one to one, $E(y|\pi_{\mathbf{x}}) = E(y|\mathbf{x}'\boldsymbol{\alpha})$ holds, which implies that step 2 above can be done equivalently by the OLS regression of y on the polynomials of the predicted probability $\widehat{\pi}_{\mathbf{x}}$ instead of the linear index $\mathbf{x}'\widehat{\boldsymbol{\alpha}}$. Although using $\mathbf{x}'\widehat{\boldsymbol{\alpha}}$ and using $\widehat{\pi}_{\mathbf{x}}$ are theoretically equivalent, they can differ in practice. On one hand, if there are outliers in $\mathbf{x}'\widehat{\boldsymbol{\alpha}}$, then it is better to use $\widehat{\pi}_{\mathbf{x}}$ because the outlier problem would be more subdued in $\widehat{\pi}_{\mathbf{x}}$. On the other hand, if $\widehat{\pi}_{\mathbf{x}}$ is too small, then $\widehat{\pi}^2_{\mathbf{x}}$ and $\widehat{\pi}^3_{\mathbf{x}}$ would be almost zero, in which case it is better to use $\mathbf{x}'\widehat{\boldsymbol{\alpha}}$.

For some Ω_{ols} and an independent and identically distributed sample of size n, we have $\sqrt{n}(\widehat{\beta}^q_{\text{PSR}} - \beta_{\text{PSR}}) \xrightarrow{d} N(0, \Omega_{\text{ols}})$,

$$\widehat{\Omega}_{\text{ols}} \equiv (\frac{1}{n}\sum_i \widehat{\varepsilon}_i^2)^{-2} \cdot \frac{1}{n}\sum_i (\widehat{v}_i\widehat{\varepsilon}_i + \widehat{\boldsymbol{\ell}}'\widehat{\boldsymbol{\eta}}_i)^2 \xrightarrow{p} \Omega_{\text{ols}}$$

$$\widehat{\varepsilon}_i \equiv d_i - \Phi(\mathbf{x}_i'\widehat{\boldsymbol{\alpha}}), \qquad \widehat{v}_i \equiv y_i - \sum_{j=0}^q \widehat{\gamma}_j(\mathbf{x}_i'\widehat{\boldsymbol{\alpha}})^j - \widehat{\beta}^q_{\text{psr}}\widehat{\varepsilon}_i,$$

$$\widehat{\boldsymbol{\ell}} \equiv -\frac{1}{n}\sum_i \mathbf{x}_i\widehat{v}_i\phi(\mathbf{x}_i'\widehat{\boldsymbol{\alpha}}), \quad \widehat{\boldsymbol{\eta}}_i \equiv (\frac{1}{n}\sum_i \widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i')^{-1}\widehat{\mathbf{s}}_i, \quad \widehat{\mathbf{s}}_i \equiv \frac{\widehat{\varepsilon}_i\phi(\mathbf{x}_i'\widehat{\boldsymbol{\alpha}})}{\Phi(\mathbf{x}_i'\widehat{\boldsymbol{\alpha}})\{1 - \Phi(\mathbf{x}_i'\widehat{\boldsymbol{\alpha}})\}}\mathbf{x}_i$$

The probit score function $\widehat{s}_i$ involves $h(t) \equiv \phi(t)/[\Phi(t)\{1 - \Phi(t)\}]$, which can be difficult to evaluate numerically when $|t|$ is large. In our experiments, $h(t)$ seems difficult to obtain precisely when $t < -20$ or $t > 5$. Hence, using the symmetry of $h(t)$, we compute $h(t)$ by $h(-|t|)$ so that $h(t)$ can be found reliably for $|t| < 20$. For extreme t values outside the range $|t| > 20$, using $h(t)/|t| \to 1$ as $|t| \to \infty$, we replace $h(t)$ with $|t|$. If logit were used for $\pi_{\mathbf{x}}$, then this kind of numerical problem would not appear, because the part in logit corresponding to $h(t)$ is equal to 1.

2.3 IVE with instrument-score residual

Generalizing Lee (2018), Lee (2021) allowed d to be endogenous with a binary instrument z. Defining the "instrument score" (IS) $\zeta_{\mathbf{x}} \equiv E(z|\mathbf{x})$, Lee (2021) proposed the IVE of $y - E(y|\zeta_{\mathbf{x}})$ on d with the IS residual (ISR) $z - \zeta_{\mathbf{x}}$ as the instrument.

Let (d^0, d^1) be the potential treatments of d for $z = 0, 1$, and define "compliers" (CPs) as those with $d^0 = 0$ and $d^1 = 1$. The IVE, dubbed "IVE$_{\text{ISR}}$", is consistent for

$$\beta_{\text{ISR}} \equiv E\{\omega_{\text{cp}}(\mathbf{x})\mu_{\text{cp}}(\mathbf{x})\}, \qquad \mu_{\text{cp}}(\mathbf{x}) \equiv E\left(y^1 - y^0|\mathbf{x}, \text{CP}\right)$$
$$\omega_{\text{cp}}(\mathbf{x}) \equiv \frac{\text{Cov}(z, d|\mathbf{x})}{E\{\text{Cov}(z, d|\mathbf{x})\}} = P(\text{CP}|\mathbf{x}) \times \zeta_{\mathbf{x}}(1 - \zeta_{\mathbf{x}}), \qquad \zeta_{\mathbf{x}} \equiv E(z|\mathbf{x})$$

The weight is high for subjects with a high $\text{Cov}(z, d|\mathbf{x})$, avoiding the "weak instrument problem" in the IVE. More importantly, the weight becomes an OW when z is taken as the underlying treatment, as in the "intent-to-treat effect" in clinical trials with an assignment z and the actual (not well complied) treatment d. Note that β_{ISR} becomes $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$ for exogenous d because then $d = z$ and everybody is a complier.

To implement IVE$_{\text{ISR}}$, 1) obtain the probit of z on $\mathbf{x}$ to get the estimator $\widehat{\zeta}_{\mathbf{x}} \equiv \Phi(\mathbf{x}'\widehat{\boldsymbol{\psi}})$ for $\zeta_{\mathbf{x}}$; 2) obtain the OLS $(\widehat{\gamma}_0, \ldots, \widehat{\gamma}_q)$ of y on $\{(\mathbf{x}'\widehat{\boldsymbol{\psi}})^0, \ldots, (\mathbf{x}'\widehat{\boldsymbol{\psi}})^q\}$ to get the predicted value $\sum_{j=0}^q \widehat{\gamma}_j(\mathbf{x}'\widehat{\boldsymbol{\psi}})^j$ for $E(y|\zeta_{\mathbf{x}})$ with $q = 2$ or 3; and 3) do the IVE of $y - \sum_{j=0}^q \widehat{\gamma}_j(\mathbf{x}'\widehat{\boldsymbol{\psi}})^j$ on d with the instrument $z - \widehat{\zeta}_{\mathbf{x}}$,

$$\widehat{\beta}_{\text{ISR}}^q \equiv \frac{\sum_i \left(z_i - \widehat{\zeta}_{\mathbf{x}_i}\right)\left\{y_i - \sum_{j=0}^q \widehat{\gamma}_j \left(\mathbf{x}_i'\widehat{\boldsymbol{\psi}}\right)^j\right\}}{\sum_i \left(z_i - \widehat{\zeta}_{\mathbf{x}_i}\right) d_i} \xrightarrow{p} \beta_{\text{ISR}}$$

As was the case for OLS$_{\text{PSR}}$, the linear index $\mathbf{x}'\widehat{\boldsymbol{\psi}}$ in step 2 can be replaced with the probability $\widehat{\zeta}_{\mathbf{x}}$.

For some Ω_{IVE}, $\sqrt{n}(\widehat{\beta}_{\text{ISR}}^q - \beta_{\text{ISR}}) \xrightarrow{d} N(0, \Omega_{\text{IVE}})$, where

$$\widehat{\Omega}_{ive} \equiv \left(\frac{1}{n}\sum_i \widehat{e}_i d_i\right)^{-2} \cdot \frac{1}{n}\sum_i (\breve{v}_i\widehat{e}_i + \breve{\boldsymbol{\ell}}'\breve{\boldsymbol{\eta}}_i)^2 \xrightarrow{p} \Omega_{ive},$$

$$\widehat{e}_i \equiv z_i - \Phi(\mathbf{x}_i'\widehat{\boldsymbol{\psi}}), \quad \breve{v}_i \equiv y_i - \sum_{j=0}^q \widehat{\gamma}_j(\mathbf{x}_i'\widehat{\boldsymbol{\psi}})^j - \widehat{\beta}_{isr}^q d_i,$$

$$\breve{\boldsymbol{\ell}} \equiv -\frac{1}{n}\sum_i \mathbf{x}_i\breve{v}_i\phi(\mathbf{x}_i'\widehat{\boldsymbol{\psi}}), \quad \breve{\boldsymbol{\eta}}_i = \left(\frac{1}{n}\sum_i \breve{\mathbf{s}}_i\breve{\mathbf{s}}_i'\right)^{-1}\breve{\mathbf{s}}_i, \quad \breve{\mathbf{s}}_i \equiv \frac{\widehat{e}_i\phi(\mathbf{x}_i'\widehat{\boldsymbol{\psi}})}{\Phi(\mathbf{x}_i'\widehat{\boldsymbol{\psi}})\{1 - \Phi(\mathbf{x}_i'\widehat{\boldsymbol{\psi}})\}}\mathbf{x}_i$$

2.4 Finding covariate slopes

After $\widehat{\beta}_{\text{ISR}}^q$ is obtained, one may desire to find the slopes of $\mathbf{x}$, as is typically done in linear models. The following shows what could be provided for the slopes of $\mathbf{x}$, although they are irrelevant for the treatment effect of interest.

Note the "linear-in-d representation" for y in Lee (2021),

$$y = \mu_0(\mathbf{x}) + \mu_{\text{cp}}(\mathbf{x})d + u, \quad E(u|\mathbf{x}, z) = 0$$
$$\mu_0(\mathbf{x}) \equiv E\{(y^1 - y^0)d^0 + y^0|\mathbf{x}\} - \mu_{\text{cp}}(\mathbf{x})E(d^0|\mathbf{x})$$

which holds under $d^0 \le d^1$ for any y as long as $y^1 - y^0$ makes sense, where $\mu_0(\mathbf{x})$ is an unknown function of $\mathbf{x}$ and u is an error term. For exogenous d, set $d = d^0 + (d^1 - d^0)z$ equal to z to obtain $(d^0 = 0,\ d^1 = 1)$ (that is, everybody is a complier for exogenous d) and

$$\mu_{\text{cp}}(\mathbf{x}) = \mu_1(\mathbf{x}) \equiv E(y^1 - y^0|\mathbf{x}) \quad \text{and} \quad \mu_0(\mathbf{x}) = E(y^0|\mathbf{x})$$

The linear-in-d representation is a "causal reduced form (RF)" because it is an RF, not an structural form (that is, the data-generating model), which contains a causal parameter $\mu_{\text{cp}}(\mathbf{x})$ of interest. The causal RF holds for any form of $\mu_{\text{cp}}(\mathbf{x})$ and y.

Rewrite the above linear-in-d representation as

$$y - \beta_{\text{ISR}}d = \mu_0(\mathbf{x}) + \{\mu_{\text{cp}}(\mathbf{x}) - \beta_{\text{ISR}}\}\, d + u$$

Then the estimand of the OLS of $y - \beta_{\text{ISR}}d$ on $\mathbf{x}$ is, with $E^{-1}(\cdot) \equiv \{E(\cdot)\}^{-1}$,

$$\begin{aligned} E^{-1}(\mathbf{x}\mathbf{x}')E\{\mathbf{x}(y - \beta_{\text{ISR}}d)\} &= E^{-1}(\mathbf{x}\mathbf{x}')E(\mathbf{x}[\mu_0(\mathbf{x}) + \{\mu_{\text{cp}}(\mathbf{x}) - \beta_{\text{ISR}}\}d]) \\ &= \boldsymbol{\beta}_{\mathbf{x}} \quad \text{if} \quad \mu_0(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}_{\mathbf{x}} \quad \text{and} \quad \mu_{\text{cp}}(\mathbf{x}) = \beta_{\text{ISR}} \end{aligned}$$

which is a sufficient condition for the slopes of $\mathbf{x}$ to be consistent for $\boldsymbol{\beta}_{\mathbf{x}}$. Note that under the sufficient condition, we have a linear model because

$$y = \mu_0(\mathbf{x}) + \mu_{\text{cp}}(\mathbf{x})d + u = \mathbf{x}'\boldsymbol{\beta}_{\mathbf{x}} + \beta_{\text{ISR}}d + u$$

Denote the OLS of $y - \widehat{\beta}_{\text{ISR}}d$ on $\mathbf{x}$ as $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$. Under the sufficient condition just mentioned, it holds that, with q in $\widehat{\beta}^q_{\text{ISR}}$ omitted,

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{\mathbf{x}} &\equiv \left(\frac{1}{n}\sum_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \times \frac{1}{n}\sum_i \mathbf{x}_i\left(y_i - \widehat{\beta}_{\text{ISR}}d_i\right) \\ &= \left(\frac{1}{n}\sum_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \times \frac{1}{n}\sum_i \mathbf{x}_i\left(\mathbf{x}_i'\boldsymbol{\beta}_{\mathbf{x}} + \beta_{\text{ISR}}d_i + u_i - \widehat{\beta}_{\text{ISR}}d_i\right) \\ &= \boldsymbol{\beta}_{\mathbf{x}} + \left(\frac{1}{n}\sum_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \times \left\{\frac{1}{n}\sum_i \mathbf{x}_iu_i - \frac{1}{n}\sum_i \mathbf{x}_id_i \times \left(\widehat{\beta}_{\text{ISR}} - \beta_{\text{ISR}}\right)\right\} \end{aligned}$$

Be aware that $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ here is not the same as the $\mathbf{x}$-slope in the OLS of y on $(\mathbf{x}, d)$, because the slope estimator of d in this OLS is not $\widehat{\beta}_{\text{ISR}}$.

Let the influence function of $\widehat{\beta}_{\text{ISR}}$ be θ_i: $\sqrt{n}(\widehat{\beta}_{\text{ISR}} - \beta_{\text{ISR}}) = n^{-1/2}\sum_i \theta_i + o_p(1)$; the form of θ_i is shown shortly. Then, using the preceding display, we obtain

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta}_{\mathbf{x}}\right) = \left(\frac{1}{n}\sum_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_i \left(\mathbf{x}_iu_i - \overline{\mathbf{x}d} \times \theta_i\right) + o_p(1)$$

where $\overline{\mathbf{x}d} \equiv n^{-1} \sum_i \mathbf{x}_i d_i$. Hence, $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{\mathbf{x}} - \boldsymbol{\beta}_{\mathbf{x}})$ is asymptotically normal with the variance estimated with

$$\left(\frac{1}{n}\sum_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1} \times \frac{1}{n}\sum_i \left(\mathbf{x}_i\widehat{u}_i - \overline{\mathbf{x}d}\widehat{\theta}_i\right)\left(\mathbf{x}_i\widehat{u}_i - \overline{\mathbf{x}d}\widehat{\theta}_i\right)' \times \left(\frac{1}{n}\sum_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}$$

$$\widehat{u}_i \equiv y_i - \widehat{\beta}_{\mathrm{ISR}} d_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}_{\mathbf{x}}, \quad \widehat{\theta}_i \equiv \frac{\breve{v}_i\widehat{e}_i + \breve{\boldsymbol{\ell}}'\breve{\eta}_i}{n^{-1}\sum_i \widehat{e}_i d_i}$$

Bear in mind that $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ is not of primary interest. Rather, $\widehat{\boldsymbol{\beta}}_{\mathbf{x}}$ and its asymptotic distribution derived under the restrictive conditions $[\mu_0(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}_{\mathbf{x}}$ and $\mu_{\mathrm{cp}}(\mathbf{x}) = \beta_{\mathrm{ISR}}]$ are just to give some idea on the contribution of $\mathbf{x}$ to $\mu_0(\mathbf{x})$ because having only one estimate $\widehat{\beta}_{\mathrm{ISR}}$ at the end of data analysis may leave the researcher feeling somewhat "unfulfilled".

3 The psr command

The command `psr` implements OLS with a PSR and an IVE with an ISR.

3.1 Syntax

The syntax for OLS with a PSR is

`psr` *depvar treatment_var covariates* [*if*] [*in*] [`, logit order(#) useprob auxiliary verbose vverbose`]

when the binary treatment variable is exogenous. The syntax for an IVE with ISRs is

`psr` *depvar* `(`*treatment_var* `=` *instrument*`)` *covariates* [*if*] [*in*] [`, logit order(#) useprob auxiliary verbose vverbose`]

when the endogenous binary treatment variable is instrumented by an exogenous binary instrument. The covariates should be exogenous always.

3.2 Options

`logit` specifies to use `logit` instead of `probit` for propensity or IS regression.

`order(#)` specifies the maximum polynomial order q for predicting the dependent variable by the fitted linear index or the fitted probability. The default is `order(2)`.

`useprob` uses the fitted probability for the prediction of the dependent variable instead of the fitted linear index. By default, the linear index is used.

`auxiliary` reports results from auxiliary regression for covariate slopes with the standard errors.

`verbose` displays all intermediate step results.

`vverbose` is the same as the `auxiliary` and `verbose` options used together.

3.3 Stored results

`psr` stores the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(q)`	polynomial order for outcome prediction
Macros	
`e(cmd)`	`psr`
`e(depvar)`	name of outcome variable
`e(model)`	`ols` or `iv`
`e(pscmd)`	binary model classifier (`probit` or `logit`)
`e(predictor)`	`xb` or `pr`
Matrices	
`e(b)`	coefficient vector (average treatment effect)
`e(V)`	variance of the average treatment effect estimator
`e(b_bin)`	coefficient vector from probit or logit regression
`e(V_bin)`	variance–covariance matrix of the estimators from probit or logit regression
`e(b_aux)`	coefficient vector from auxiliary regression
`e(V_aux)`	variance–covariance matrix from auxiliary regression
Functions	
`e(sample)`	marks estimation sample

4 Applications

We present two applications in this section: flu vaccine effect on hospitalization or not (Lee 2021; Ding and Lu 2017) and education effect on wage (Lee 2021).

4.1 Flu vaccine effect on hospitalization

The first application is flu vaccine effect on hospitalization or not. `flu.dta` is prepared using the Ding and Lu (2017) data. The outcome variable is `outcome`, the treatment variable is `receive`, and the exogenous covariates are `age`, `female`, `white`, `copd`, `heartd`, and `renal`. Lee (2021) allows the treatment to be endogenous and instruments it with `assign`. We consider exogenous and endogenous treatments for the sake of illustration.

With the treatment assumed to be exogenous, the `psr` command invoked with the default options gives the following results, where the OW average treatment effect is small in magnitude and is statistically insignificant:

```
. use flu

. psr outcome receive age female white copd heartd renal
(intermediate outputs suppressed; use verbose option to override)
```

outcome	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
receive	-.0033608	.0120239	-0.28	0.780	-.0269271	.0202056

```
Polynomial order of xb for output prediction =  2            obs =      2861
Treatment: receive (exogenous, fitted by probit)
Exogenous: age female white copd heartd renal
```

In this OLS result, the treatment effect (-0.0033608) is obtained as follows: The treatment variable (`receive`) is regressed on the covariates (`age`, `female`, `white`, `copd`, `heartd`, and `renal`) by probit, and the outcome prediction error is obtained by regressing the outcome variable (`outcome`) on the linear index (`xb`) and its square. The OW average treatment effect is then obtained by the OLS regression of the outcome prediction error on the PSR.

A third-order polynomial of the linear index can be added to the outcome prediction regression, which yields the following result:

```
. psr outcome receive age female white copd heartd renal, order(3)
(intermediate outputs suppressed; use verbose option to override)
```

outcome	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
receive	-.00328	.0120184	-0.27	0.785	-.0268356	.0202755

```
Polynomial order of xb for output prediction =  3            obs =      2861
Treatment: receive (exogenous, fitted by probit)
Exogenous: age female white copd heartd renal
```

The results change only marginally.

Logit can replace probit by using the `logit` option, as the following example illustrates:

```
. psr outcome receive age female white copd heartd renal, logit
(intermediate outputs suppressed; use verbose option to override)
```

outcome	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
receive	-.0033386	.0120322	-0.28	0.781	-.0269213	.0202441

```
Polynomial order of xb for output prediction =  2            obs =      2861
Treatment: receive (exogenous, fitted by logit)
Exogenous: age female white copd heartd renal
```

The results are almost identical to those by the probit regression.

The auxiliary covariate slopes (derived under restrictive conditions in section 2.4 to make the y model linear) are reported as follows when the `auxiliary` option is used:

```
. psr outcome receive age female white copd heartd renal, auxiliary
(intermediate outputs suppressed; use verbose option to override)

------------------------------------------------------------------------------
     outcome | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     receive |  -.0033608   .0120239    -0.28   0.780    -.0269271    .0202056
------------------------------------------------------------------------------
Polynomial order of xb for output prediction =  2             obs =       2861
Treatment: receive (exogenous, fitted by probit)
Exogenous: age female white copd heartd renal
Auxiliary results from OLS of Y - teffect*D on covariates (with robust
standard errors taking into account the first-stage estimation error for
teffect in Y - teffect*D):
------------------------------------------------------------------------------
  Y-teffect*D | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |  -.0006199    .000426    -1.46   0.146    -.0014549    .0002151
      female |   -.016198   .0120737    -1.34   0.180     -.039862     .007466
       white |  -.0191919   .0115777    -1.66   0.097    -.0418837       .0035
        copd |    .023678   .0136739     1.73   0.083    -.0031224    .0504785
      heartd |   .0509577   .0100869     5.05   0.000     .0311877    .0707278
       renal |   .2234275   .0746846     2.99   0.003     .0770484    .3698066
       _cons |   .1111083   .0317467     3.50   0.000     .0488859    .1733306
------------------------------------------------------------------------------
Caution: Linearity and effect homogeneity are assumed.
Treatment effect is valid without those restrictive assumptions.
```

The results from the auxiliary regression of $y - \widehat{\beta}d$ on the covariates are displayed after the estimated OW treatment effect. The reported standard errors are corrected for the first-stage estimation error, and cautionary warnings are noted.

All the intermediate-step results are displayed if the `verbose` option is used:

```
. psr outcome receive age female white copd heartd renal, verbose
Step 1: probit regression of receive
Iteration 0:  Log likelihood = -1609.6643
Iteration 1:  Log likelihood = -1588.4349
Iteration 2:  Log likelihood = -1588.3905
Iteration 3:  Log likelihood = -1588.3905
Probit regression                                       Number of obs   =       2861
                                                        LR chi2(6)      =      42.55
                                                        Prob > chi2     =     0.0000
Log likelihood = -1588.3905                             Pseudo R2       =     0.0132
------------------------------------------------------------------------------
     receive | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |   .0104581   .0021883     4.78   0.000     .0061692     .014747
      female |  -.0553546   .0560536    -0.99   0.323    -.1652177    .0545085
       white |  -.0049985   .0555472    -0.09   0.928     -.113869    .1038721
        copd |   .2618746   .0607797     4.31   0.000     .1427486    .3810007
      heartd |   .0457934   .0523338     0.88   0.382    -.0567789    .1483657
       renal |   .0216417   .2222316     0.10   0.922    -.4139243    .4572077
       _cons |  -1.426146   .1621891    -8.79   0.000    -1.744031   -1.108261
------------------------------------------------------------------------------
```

```
Step 2: regress outcome on xb's polynomials of order 2
```

outcome	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
xb^1	.2972206	.1482517	2.00	0.045	.0065294	.5879118
xb^2	.1648161	.1021603	1.61	0.107	-.0354993	.3651316
_cons	.2066716	.0530624	3.89	0.000	.1026271	.3107162

```
Note. Standard errors not adjusted for generated regressors.
Step 3: Regression of outcome prediction error on PSR
```

outcome	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
receive	-.0033608	.0120239	-0.28	0.780	-.0269271	.0202056

```
Polynomial order of xb for output prediction =  2            obs =       2861
Treatment: receive (exogenous, fitted by probit)
Exogenous: age female white copd heartd renal
```

The displayed results are self-evident. The `vverbose` option is identical to `verbose` and `auxiliary` used together.

We have so far assumed that the treatment (`receive`) is exogenous. When it is possibly endogenous and is instrumented by `assign`, the results (using the `vverbose` option to display all intermediate steps and the auxiliary covariate slopes) are as follows:

```
. use flu, clear
. psr outcome (receive = assign) age female white copd heartd renal, vverbose
Step 1: probit regression of assign
Iteration 0:  Log likelihood =   -1981.89
Iteration 1:  Log likelihood = -1978.6418
Iteration 2:  Log likelihood = -1978.6417
Probit regression                                 Number of obs   =       2861
                                                  LR chi2(6)      =       6.50
                                                  Prob > chi2     =     0.3699
Log likelihood = -1978.6417                       Pseudo R2       =     0.0016
```

assign	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
age	.0016601	.0019013	0.87	0.383	-.0020664	.0053866
female	.0867887	.0515592	1.68	0.092	-.0142656	.1878429
white	.027004	.0508958	0.53	0.596	-.07275	.126758
copd	-.004955	.0563516	-0.09	0.930	-.1154022	.1054921
heartd	.0402458	.0478214	0.84	0.400	-.0534824	.1339739
renal	.2083722	.2082433	1.00	0.317	-.1997772	.6165216
_cons	-.1720573	.1405214	-1.22	0.221	-.4474743	.1033596

```
Step 2: regress outcome on xb's polynomials of order 2
```

outcome	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
xb^1	-.169189	.1115664	-1.52	0.130	-.3879478	.0495697
xb^2	2.739879	.7020163	3.90	0.000	1.36337	4.116389
_cons	.0779886	.0061648	12.65	0.000	.0659007	.0900765

```
Note. Standard errors not adjusted for generated regressors.
Step 3: Regression of outcome prediction error on D using ISR as instrument
```

outcome	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
receive	-.1265733	.0895777	-1.41	0.158	-.3021424	.0489957

```
Polynomial order of xb for output prediction =  2            obs =       2861
Treatment: receive (instrumented by assign, fitted by probit)
Exogenous: age female white copd heartd renal
Auxiliary results from OLS of Y - teffect*D on covariates (with robust
standard errors taking into account the first-stage estimation error for
teffect in Y - teffect*D):
```

Y-teffect*D	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
age	-.0002391	.0004316	-0.55	0.580	-.0010851	.0006069
female	-.0184638	.0122571	-1.51	0.132	-.0424872	.0055596
white	-.0193098	.0117595	-1.64	0.101	-.0423579	.0037384
copd	.0339783	.0138496	2.45	0.014	.0068336	.061123
heartd	.0526859	.0102918	5.12	0.000	.0325144	.0728574
renal	.2243149	.0751364	2.99	0.003	.0770503	.3715795
_cons	.1147726	.0321669	3.57	0.000	.0517266	.1778185

```
Caution: Linearity and effect homogeneity are assumed.
Treatment effect is valid without those restrictive assumptions.
```

First, the instrument (`assign`) is regressed on the covariates. Second, the outcome variable (`outcome`) is fit using the polynomials of the linear index up to the second order. Finally, the outcome prediction error is regressed on the treatment variable (`receive`) using the ISR as the instrument. The estimated OW average treatment effect is -0.1265733, which is much larger than the one obtained under the treatment exogeneity assumption, though it is still statistically insignificant at the 10% level (p-value $= 0.158$). Lee (2021, 626) shows that the usual linear-model IVE yields almost the same effect estimate, -0.13, with or without $\mathbf{x}$ controlled for. The estimated covariate slopes are presented as the final output.

4.2 Education effect on wage

Our second empirical analysis is for education effect on wage (Lee 2021) based on Card (1995). `nlsdat.dta` is prepared using files at https://davidcard.berkeley.edu/data_sets.html. The variable names to appear below are taken from the codebook provided in this URL.

The outcome variable is ln(wage) in 1976 (`lwage76`), the treatment variable (d) is the binary indicator for the schooling years in 1976 (`ed76`) to exceed 12, the instrument (z) is for whether one grew up near a four-year college (`nearc4`), and the covariates ($\mathbf{x}$) consist of 1, age (`age76`), dummy for black (`black`), dummies for nine residence regions in 1966 (`reg662`–`reg669`, with `reg661` as the base category), dummy for living in a standard metropolitan statistical area in 1966 (`smsa66r`), dummy for living in a standard metropolitan statistical area in 1976 (`smsa76r`), and dummy for living in the South in 1976 (`reg76r`).

Lee (2021) presented two results with and without `smsa76r` and `reg76r` because of the possibility of them being affected by d. With `smsa76r` and `reg76r` included, the IVE regression of the outcome prediction error on the treatment using the instrument propensity residual as the instrument gives the following results (with the `vverbose` option used to show all the intermediate-step results and the auxiliary covariate slopes):

```
. use nlsdat, clear
. generate d = ed76 > 12
. global X0 age76 black reg662-reg669 smsa66r
. global X1 ${X0} smsa76r reg76r
. psr lwage76 (d = nearc4) ${X1}, vverbose
Step 1: probit regression of nearc4
Iteration 0:  Log likelihood = -1882.1743
Iteration 1:  Log likelihood = -1495.0081
Iteration 2:  Log likelihood = -1488.4087
Iteration 3:  Log likelihood = -1488.3888
Iteration 4:  Log likelihood = -1488.3888
Probit regression                                 Number of obs   =      3010
                                                  LR chi2(13)     =    787.57
                                                  Prob > chi2     =    0.0000
Log likelihood = -1488.3888                       Pseudo R2       =    0.2092
```

nearc4	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
age76	.0118223	.008551	1.38	0.167	-.0049374	.028582
black	.1820022	.068917	2.64	0.008	.0469274	.317077
reg662	.0499826	.1602549	0.31	0.755	-.2641112	.3640764
reg663	-.4619476	.1518162	-3.04	0.002	-.7595019	-.1643933
reg664	-.3846031	.1678376	-2.29	0.022	-.7135589	-.0556474
reg665	-.5841955	.1725978	-3.38	0.001	-.922481	-.2459101
reg666	-.9552103	.1831986	-5.21	0.000	-1.314273	-.5961477
reg667	-.82846	.1835112	-4.51	0.000	-1.188135	-.4687847
reg668	-.6587352	.2023066	-3.26	0.001	-1.055249	-.2622215
reg669	-.2381267	.1696908	-1.40	0.161	-.5707146	.0944612
smsa66r	.9892418	.0697661	14.18	0.000	.8525026	1.125981
smsa76r	.2880253	.0716101	4.02	0.000	.147672	.4283786
reg76r	-.0816956	.0972298	-0.84	0.401	-.2722626	.1088713
_cons	-.1616815	.2850832	-0.57	0.571	-.7204343	.3970712

Step 2: regress lwage76 on xb's polynomials of order 2

lwage76	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
xb^1	.2122322	.0183157	11.59	0.000	.1763196	.2481448
xb^2	-.050734	.017508	-2.90	0.004	-.0850628	-.0164052
_cons	6.182557	.0118082	523.58	0.000	6.159404	6.20571

Note. Standard errors not adjusted for generated regressors.

Step 3: Regression of outcome prediction error on D using ISR as instrument

lwage76	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
d	.4102675	.2511422	1.63	0.102	-.0819622	.9024973

Polynomial order of xb for output prediction = 2 obs = 3010
Treatment: d (instrumented by nearc4, fitted by probit)
Exogenous: age76 black reg662 reg663 reg664 reg665 reg666 reg667 reg668 reg669
> smsa66r smsa76r reg76r

Auxiliary results from OLS of Y - teffect*D on covariates (with robust
standard errors taking into account the first-stage estimation error for
teffect in Y - teffect*D):

Y-teffect*D	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
age76	.0413973	.0024043	17.22	0.000	.036685	.0461096
black	-.1540869	.0187519	-8.22	0.000	-.19084	-.1173339
reg662	.0747168	.0386603	1.93	0.053	-.0010561	.1504897
reg663	.1183685	.0377422	3.14	0.002	.044395	.1923419
reg664	.0217221	.045415	0.48	0.632	-.0672897	.1107339
reg665	.1180115	.0470947	2.51	0.012	.0257076	.2103155
reg666	.1179829	.0494027	2.39	0.017	.0211553	.2148104
reg667	.118484	.0493607	2.40	0.016	.0217389	.2152292
reg668	-.1347822	.0537282	-2.51	0.012	-.2400876	-.0294768
reg669	.0627177	.0427566	1.47	0.142	-.0210836	.1465191
smsa66r	.0319594	.0198708	1.61	0.108	-.0069867	.0709055
smsa76r	.1020951	.0207117	4.93	0.000	.0615009	.1426893
reg76r	-.1818059	.0304977	-5.96	0.000	-.2415803	-.1220314
_cons	4.818898	.0781723	61.64	0.000	4.665683	4.972113

Caution: Linearity and effect homogeneity are assumed.
Treatment effect is valid without those restrictive assumptions.

The OW average treatment effect is estimated to be 0.4102675 (with the robust standard error 0.2511422), as is shown in the `Step 3` part of the output. Lee (2021, 628) shows that the usual linear-model IVE with $\mathbf{x}$ controlled yields the effect estimate 0.43 (with the robust standard error 0.24), which differs little from the OW average treatment effect result.

With smsa76r and reg76r excluded from the covariate list, the results change to the following:

```
. psr lwage76 (d = nearc4) ${X0}, vverbose
Step 1: probit regression of nearc4
Iteration 0:  Log likelihood = -1882.1743
Iteration 1:  Log likelihood = -1503.6449
Iteration 2:  Log likelihood = -1497.3817
Iteration 3:  Log likelihood = -1497.3625
Iteration 4:  Log likelihood = -1497.3625
Probit regression                                 Number of obs   =       3010
                                                  LR chi2(11)     =     769.62
                                                  Prob > chi2     =     0.0000
Log likelihood = -1497.3625                       Pseudo R2       =     0.2045

------------------------------------------------------------------------------
      nearc4 | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       age76 |   .0099669   .0085142     1.17   0.242    -.0067206    .0266544
       black |   .2005843   .0685613     2.93   0.003     .0662067    .3349618
      reg662 |   .0869013   .1589232     0.55   0.585    -.2245824     .398385
      reg663 |  -.4377081   .1504965    -2.91   0.004    -.7326758   -.1427404
      reg664 |  -.3763224   .1669333    -2.25   0.024    -.7035057   -.0491391
      reg665 |  -.6445885   .1499743    -4.30   0.000    -.9385328   -.3506442
      reg666 |  -.9993765   .1621363    -6.16   0.000    -1.317158   -.6815952
      reg667 |   -.872297   .1589835    -5.49   0.000    -1.183899    -.560695
      reg668 |   -.651649   .2012587    -3.24   0.001    -1.046109   -.2571892
      reg669 |  -.2106687   .1685628    -1.25   0.211    -.5410457    .1197083
     smsa66r |   1.157478   .0555694    20.83   0.000     1.048564    1.266392
       _cons |  -.0437532   .2825242    -0.15   0.877    -.5974904    .5099841
------------------------------------------------------------------------------
Step 2: regress lwage76 on xb's polynomials of order 2
------------------------------------------------------------------------------
     lwage76 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        xb^1 |   .1862023   .0196759     9.46   0.000     .1476227    .2247818
        xb^2 |  -.0319933   .0186826    -1.71   0.087    -.0686253    .0046387
       _cons |    6.18133   .0118389   522.12   0.000     6.158117    6.204543
------------------------------------------------------------------------------
Note. Standard errors not adjusted for generated regressors.
```

```
Step 3: Regression of outcome prediction error on D using ISR as instrument
```

lwage76	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
d	.5276801	.2259549	2.34	0.020	.0848165	.9705436

```
Polynomial order of xb for output prediction =  2            obs =       3010
Treatment: d (instrumented by nearc4, fitted by probit)
Exogenous: age76 black reg662 reg663 reg664 reg665 reg666 reg667 reg668 reg669
> smsa66r
Auxiliary results from OLS of Y - teffect*D on covariates (with robust
standard errors taking into account the first-stage estimation error for
teffect in Y - teffect*D):
```

Y-teffect*D	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
age76	.0413565	.00255	16.22	0.000	.0363585	.0463544
black	-.1189603	.0197279	-6.03	0.000	-.1576262	-.0802943
reg662	.0727586	.0415344	1.75	0.080	-.0086473	.1541645
reg663	.1098763	.0408153	2.69	0.007	.0298798	.1898729
reg664	.0068656	.0486819	0.14	0.888	-.0885491	.1022803
reg665	-.0407721	.040892	-1.00	0.319	-.1209189	.0393748
reg666	-.0386979	.0453898	-0.85	0.394	-.1276603	.0502644
reg667	-.0469446	.043589	-1.08	0.281	-.1323775	.0384882
reg668	-.1634314	.0570261	-2.87	0.004	-.2752005	-.0516622
reg669	.0505569	.0457973	1.10	0.270	-.039204	.1403179
smsa66r	.0824573	.0168235	4.90	0.000	.049484	.1154307
_cons	4.790264	.0826818	57.94	0.000	4.628211	4.952317

```
Caution: Linearity and effect homogeneity are assumed.
Treatment effect is valid without those restrictive assumptions.
```

The estimated OW average treatment effect is 0.5276801 (with the robust standard error 0.2259549), which is statistically significant at the 5% level (p-value = 0.020). Lee (2021, 628) shows that the usual linear-model IVE with $\mathbf{x}$ controlled yields the effect estimate 0.55 (with the robust standard error 0.22), which differs little from the OW average treatment effect result.

5 Conclusions

In finding the effect of a binary treatment d, practitioners often apply OLS to LDVs y, despite the fact that the linear model is untenable. Also, COVID-19 recently demonstrated how heterogeneous treatment effects can be, yet linear models typically assume a constant effect. This brought up an important question: What does the OLS estimate when the effect is $\mathbf{x}$-heterogeneous, being a function $\mu_1(\mathbf{x})$ of $\mathbf{x}$, and the linear model is invalid?

The answer in the literature is that the effect estimated by the OLS is a weighted average $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$, where the weight $\omega(\mathbf{x})$ is high for those with their PS $\pi_{\mathbf{x}} \equiv E(d|\mathbf{x}) \simeq 0.5$ and low for those with $\pi_{\mathbf{x}} \simeq 0, 1$; $\omega(\mathbf{x})$ is called the OW. However, this answer requires that $\pi_{\mathbf{x}}$ be equal to the linear projection of d on $\mathbf{x}$, which unfortunately rarely holds in reality. Without this condition, the OLS can be consistent for a nonzero number even when $\mu_1(\mathbf{x}) = 0$. This brought up another question: Can $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$ be estimated without the restrictive condition?

To this question, Lee (2018) showed that the OLS of y on $d - \pi_{\mathbf{x}}$ ("OLS_{PSR}") is consistent for $E\{\omega(\mathbf{x})\mu_1(\mathbf{x})\}$ without the restrictive condition. Going further, Lee (2021) showed that when d is endogenous with a binary instrument z, the IVE of y on d with instrument $z - \zeta_{\mathbf{x}}$ ("IVE_{ISR}"), where $\zeta_{\mathbf{x}} \equiv E(z|\mathbf{x})$ is the instrument score (IS), is consistent for a modified OW average of $\mu_1(\mathbf{x})$. This article reviewed OLS_{PSR} and IVE_{ISR} and then provided the command `psr` to implement OLS_{PSR} and IVE_{ISR}, which are applicable to any $\mu_1(\mathbf{x})$ and any y (continuous, count, binary, etc.).

Because both OLS_{PSR} and IVE_{ISR} are implemented by specifying $\pi_{\mathbf{x}}$ and $\zeta_{\mathbf{x}}$ as probit (or logit), one concern in OLS_{PSR} and IVE_{ISR} is misspecifications in the probit or logit for $\pi_{\mathbf{x}}$ or $\zeta_{\mathbf{x}}$. To make OLS_{PSR} and IVE_{ISR} robust to such misspecifications, the actually implemented version of OLS_{PSR} and IVE_{ISR} uses, respectively, $y - E(y|\pi_{\mathbf{x}})$ and $y - E(y|\zeta_{\mathbf{x}})$ instead of y. This idea is closely related to the recent "double debiasing" idea for robustness to misspecifications in nuisance functions such as $\pi_{\mathbf{x}}$ or $\zeta_{\mathbf{x}}$, where $y - E(y|\mathbf{x})$ is used instead of y and then a machine-learning method estimates $E(y|\mathbf{x})$.

In our two empirical illustrations where one outcome is binary and the other is continuous, OLS_{PSR} and IVE_{ISR} yielded estimates very close to those of the usual linear-model OLS and IVE. This demonstrates that OLS_{PSR} and IVE_{ISR} reduce to the usual linear-model OLS and IVE when $\pi_{\mathbf{x}}$ is the same as the linear projection; otherwise, OLS_{PSR} and IVE_{ISR} would still provide the OW average treatment effect. By providing the command, we hope that OLS_{PSR} and IVE_{ISR} are more widely applied because they require neither any linear approximation argument nor constant effect assumption.

As far as we can see, there are two remaining issues in OLS_{PSR} and IVE_{ISR}; one is easy to address, but the other is not. The first issue is the robustness to misspecifications in $\pi_{\mathbf{x}}$ and $\zeta_{\mathbf{x}}$ because it is not "full" yet. If the $\mathbf{x}$-centered variables $\{y - E(y|\mathbf{x}), d - \pi_{\mathbf{x}}, z - \zeta_{\mathbf{x}}\}$ are used in the OLS_{PSR} and IVE_{ISR}, then OLS_{PSR} and IVE_{ISR} will be fully double debiasing. This will be done in the near future, and a command will become available before long because the research is underway. The other issue is the OW: If one objects to OW, are there any other sensible weighting schemes to adopt? The answer would be yes, but it is unlikely that any sensible weighting scheme other than OW is compatible with OLS- and IVE-based approaches.

6 Acknowledgments

The authors thank the editor and an anonymous referee for helpful comments. The research of Myoung-jae Lee has been supported by the National Research Foundation of Korea grant funded by the Korea government (MSIT) (2020R1A2C1A01007786).

7 Programs and supplemental material

To install the software files as they existed at the time of publication of this article, type

```
. net sj 24-1
. net install st0740      (to install program files, if available)
. net get st0740          (to install ancillary files, if available)
```

8 References

Angrist, J. D. 1998. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica* 66: 249–288. https://doi.org/10.2307/2998558.

———. 2001. Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business and Economic Statistics* 19: 2–28. https://doi.org/10.1198/07350010152472571.

Angrist, J. D., and A. B. Krueger. 1999. Empirical strategies in labor economics. In Vol. 3A of *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card, 1277–1366. Amsterdam: Elsevier. https://doi.org/10.1016/S1573-4463(99)03004-7.

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press. https://doi.org/10.2307/j.ctvcm4j72.

Aronow, P. M., and C. Samii. 2016. Does regression produce representative estimates of causal effects? *American Journal of Political Science* 60: 250–267. https://doi.org/10.1111/ajps.12185.

Card, D. E. 1995. Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. L. N. Christofides, E. K. Grant, and R. Swidinsky, 201–222. Toronto, Canada: University of Toronto Press.

Cheng, C., F. Li, L. E. Thomas, and F. Li. 2022. Addressing extreme propensity scores in estimating counterfactual survival functions via the overlap weights. *American Journal of Epidemiology* 191: 1140–1151. https://doi.org/10.1093/aje/kwac043.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey. 2017. Double/debiased/Neyman machine learning of treatment effects. *American Economic Review: Papers and Proceedings* 107: 261–265. https://doi.org/10.1257/aer.p20171038.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. 2018. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal* 21: C1–C68. https://doi.org/10.1111/ectj.12097.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins. 2022. Locally robust semiparametric estimation. *Econometrica* 90: 1501–1535. https://doi.org/10.3982/ECTA16294.

Choi, J.-Y., and M.-J. Lee. 2023. Overlap weight and propensity score residual for heterogeneous effects: A review with extensions. *Journal of Statistical Planning and Inference* 222: 22–37. https://doi.org/10.1016/j.jspi.2022.04.003.

Ding, P., and J. Lu. 2017. Principal stratification analysis using principal scores. *Journal of the Royal Statistical Society,* B ser., 79: 757–777. https://doi.org/10.1111/rssb.12191.

Imbens, G. W. 2001. Comment: Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business and Economic Statistics* 19: 17–20. https://doi.org/10.1198/07350010152472571.

Lee, M.-J. 2018. Simple least squares estimator for treatment effects using propensity score residuals. *Biometrika* 105: 149–164. https://doi.org/10.1093/biomet/asx062.

———. 2021. Instrument residual estimator for any response variable with endogenous binary treatment. *Journal of the Royal Statistical Society,* B ser., 83: 612–635. https://doi.org/10.1111/rssb.12442.

Lee, M.-J., G. Lee, and J.-Y. Choi. Forthcoming. Linear probability model revisited: Why it works and how it should be specified. *Sociological Methods and Research*. https://doi.org/10.1177/00491241231176850.

Li, F., and F. Li. 2019. Propensity score weighting for causal inference with multiple treatments. *Annals of Applied Statistics* 13: 2389–2415. https://doi.org/10.1214/19-AOAS1282.

Li, F., K. L. Morgan, and A. M. Zaslavsky. 2018. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association* 113: 390–400. https://doi.org/10.1080/01621459.2016.1260466.

Li, F., L. E. Thomas, and F. Li. 2019. Addressing extreme propensity scores via the overlap weights. *American Journal of Epidemiology* 188: 250–257. https://doi.org/10.1093/aje/kwy201.

Li, L., and T. Greene. 2013. A weighting analogue to pair matching in propensity score analysis. *International Journal of Biostatistics* 9: 215–234. https://doi.org/10.1515/ijb-2012-0030.

Mao, H., and L. Li. 2020. Flexible regression approach to propensity score analysis and its relationship with matching and weighting. *Statistics in Medicine* 39: 2017–2034. https://doi.org/10.1002/sim.8526.

Mao, H., L. Li, and T. Greene. 2019. Propensity score weighting analysis and treatment effect discovery. *Statistical Methods in Medical Research* 28: 2439–2454. https://doi.org/10.1177/0962280218781171.

Mao, H., L. Li, W. Yang, and Y. Shen. 2018. On the propensity score weighting analysis with survival outcome: Estimands, estimation, and inference. *Statistics in Medicine* 37: 3745–3763. https://doi.org/10.1002/sim.7839.

Mlcoch, T., T. Hrnciarova, J. Tuzil, J. Zadak, M. Marian, and T. Dolezal. 2019. Propensity score weighting using overlap weights: A new method applied to regorafenib clinical data and a cost-effectiveness analysis. *Value in Health* 22: 1370–1377. https://doi.org/10.1016/j.jval.2019.06.010.

Moffitt, R. A. 2001. Comment: Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business and Economic Statistics* 19: 20–23. https://doi.org/10.1198/07350010152472571.

Robins, J. M., S. D. Mark, and W. K. Newey. 1992. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* 48: 479–495. https://doi.org/10.2307/2532304.

Thomas, L. E., F. Li, and M. J. Pencina. 2020. Overlap weighting: A propensity score method that mimics attributes of a randomized clinical trial. *Journal of the American Medical Association* 323: 2417–2418. https://doi.org/10.1001/jama.2020.7819.

Todd, P. 2001. Comment: Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business and Economic Statistics* 19: 25–27. https://doi.org/10.1198/07350010152472571.

van der Laan, M. J., and S. Rose. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York: Springer. https://doi.org/10.1007/978-1-4419-9782-1.

Vansteelandt, S., and R. M. Daniel. 2014. On regression adjustment for the propensity score. *Statistics in Medicine* 33: 4053–4072. https://doi.org/10.1002/sim.6207.

About the authors

Myoung-jae Lee is a professor in the Department of Economics at Korea University, Seoul, Republic of Korea.

Chirok Han is a professor in the Department of Economics at Korea University, Seoul, Republic of Korea.