LOGIT-BASED ALTERNATIVES TO TWO-STAGE LEAST SQUARES

DENIS CHETVERIKOV, JINYONG HAHN, ZHIPENG LIAO, AND SHUYANG SHENG

ABSTRACT. We propose logit-based IV and augmented logit-based IV estimators that serve as alternatives to the traditionally used 2SLS estimator in the model where both the endogenous treatment variable and the corresponding instrument are binary. Our novel estimators are as easy to compute as the 2SLS estimator but have an advantage over the 2SLS estimator in terms of causal interpretability. In particular, in certain cases where the probability limits of both our estimators and the 2SLS estimator take the form of weighted-average treatment effects, our estimators are guaranteed to yield non-negative weights whereas the 2SLS estimator is not.

1. INTRODUCTION

We study the problem of instrumental variable estimation in the setting with a binary treatment and a binary instrument in the presence of controls. Numerous parametric and nonparametric instrumental variable estimators have been proposed in the literature for this setting, and among all of them, perhaps the most important one is the 2SLS estimator. It is simple to compute and has straightforward motivation in the case of constant treatment effects. However, it has been recently demonstrated by Blandhol et al. (2022) that in the case of heterogeneous treatment effects, the 2SLS estimator has multiple issues unless saturated controls[1] are being used, which is rarely the case in practice. In this paper, we propose a new instrumental variable estimator that alleviates some of the problems of the 2SLS estimator and is as simple to compute as the 2SLS estimator itself.

Like the 2SLS estimator, our estimator consists of two steps. In the first step, we run a logit regression of the instrument on the set of controls. In the second step, we use a classic instrumental variable estimator for the linear regression of the outcome on the treatment using residuals from the logit regression calculated

Date: December 16, 2023.

[1]The vector of controls is said to be saturated if it consists of dummy variables such that for all its realizations, one and only one dummy takes value one.

on the first step as an instrument. We refer to this procedure as the *logit-based instrumental variable (IV) estimator*.

Under the standard monotonicity (no defiers) and conditional independence conditions, the probability limits of both 2SLS and logit-based IV estimators consist of the sum of complier, always-taker, never-taker, and non-causal terms. Without further conditions, complier, always-taker, and never-taker terms take the form of weighted-average treatment effects with the weights not necessarily integrating to one but the advantage of the logit-based IV estimator is that the corresponding weights in the complier term are always non-negative, which is not necessarily the case for the 2SLS estimator. This advantage is particularly important under additional conditions guaranteeing that the always-taker, never-taker, and non-causal terms vanish. Under these conditions, the probability limit of the logit-based IV estimator is represented by a convex combination of treatment effects for compliers and the probability limit of the 2SLS estimator is not. Thus, under these conditions, the logit-based IV estimator has a causal interpretation and the 2SLS estimator does not.

In addition, we develop an augmented logit-based IV estimator that has a causal interpretation under conditions that are more plausible than those underlying causal interpretability of the logit-based IV and 2SLS estimators. This estimator is similar to the logit-based IV estimator itself but contains an extra term in the logit regression used in the first step of the logit-based IV estimator. This term in turn originates from the binary regression model of the treatment variable on controls using a subsample of the data corresponding to an ex ante fixed value of the instrument.

Moreover, we construct a Hausman specification test that can be used to check whether the logit-based IV estimator has a causal interpretation. The test is based on the comparison of the logit-based IV and augmented logit-based IV estimators and is easy to perform.

Our paper contributes to the large literature discussing causal interpretability of various parametric IV estimators in the case of heterogeneous treatment effects. We therefore provide here only a few key references that are particularly relevant for our work. The literature has been started by Imbens and Angrist (1994), who gave the local average treatment effect interpretation of the instrumental variable estimator in the case of a binary treatment and a binary instrument without controls allowing for heterogeneous treatment effects under the monotonicity (no defiers) assumption. Angrist and Imbens (1995) showed that in a model with saturated controls, the 2SLS estimator that includes all interactions between the instrument and controls in the first step converges in probability to a weighted average of control-specific local average treatment effects. Abadie (2003), Kolesar (2013), and Sloczynski (2020) obtained a similar result for the same and other parametric instrumental variable estimators without saturated controls assuming that the conditional mean function of the instrument given controls is linear. Blandhol et al. (2022) demonstrated that parametric instrumental variable estimators generally lack a causal interpretation if this conditional mean function is not linear.

The rest of the paper is organized as follows. In the next section, we introduce the logit-based IV estimator and discuss its causal interpretation. In Section 3, we discuss the augmented logit-based IV estimator and compare its causal interpretability with that of the logit-based IV estimator itself. In Section 4, we derive asymptotic normality results for both estimators. In Section 5, we develop a Hausman test that can be used to check causal interpretability of the logit-based IV estimator. In the Appendix, we provide all the proofs.

2. Logit-Based IV Estimator

In this section, we propose a logit-based IV estimator and explain its advantages over the 2SLS estimator in the potential outcome model with a binary treatment and a binary instrument in the presence of controls. In particular, we derive a set of conditions under which our logit-based IV estimator has a causal interpretation and the 2SLS estimator does not.

Consider the potential outcome model with a binary treatment $T \in \{0, 1\}$ and a binary instrument $Z \in \{0, 1\}$:

$$Y = Y(1)T + Y(0)(1 - T) \quad \text{and} \quad T = T(1)Z + T(0)(1 - Z), \tag{1}$$

where $Y(0), Y(1) \in \mathbb{R}$ are potential outcome values and $T(0), T(1) \in \{0, 1\}$ are potential treatment values. In addition, let $X \in \mathcal{X} \subset \mathbb{R}^p$ be a vector of controls. We will assume that the instrument Z is independent of potential values $(Y(0), Y(1), T(0), T(1))$ conditional on X, which is a standard assumption in the program evaluation literature, e.g. see Chapter 4.5.2 in Angrist and Pischke (2009):

Assumption 2.1. $Z \perp (Y(0), Y(1), T(0), T(1)) \mid X$.

In this model, the group variable $G = (T(0), T(1))$ can take four values, $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$, that are typically thought to correspond to sub-populations of never-takers (NT), compliers (CP), defiers (DF), and always-takers (AT), respectively, and it is customary to use letter-based values instead of digit-based values. We will follow this tradition and will write, for example, $G = CP$ instead of $G = (0, 1)$.

For each sub-population $g \in \mathcal{G} = \{NT, CP, AT, AT\}$ and each $x \in \mathcal{X}$, define the x-conditional average treatment effect

$$\Delta_g(x) = \mathbb{E}[Y(1) - Y(0)|G = g, X = x].$$

For any estimator $\widehat{\beta}$, we will say that it has a causal interpretation if its probability limit takes the form of a weighted-average treatment effect

$$\sum_{g \in \mathcal{G}} \mathbb{E}[\Delta_g(X) w_g(X)], \tag{2}$$

where the weights $w_g(x)$ are non-negative for all $(g, x) \in \mathcal{G} \times \mathcal{X}$ and integrate to one: $\sum_{g \in \mathcal{G}} \mathbb{E}[w_g(X)] = 1$. In other words, the estimator has a causal interpretation if its probability limit can be represented as a convex combination of treatment effects. We will say that an estimator has a partially causal interpretation if takes the form of a weighted-average treatment effect (2) with non-negative weights that do not necessarily integrate to one.

In addition, we will impose the monotonicity condition as in Imbens and Angrist (1994):

Assumption 2.2. $\mathbb{P}(T(1) \geq T(0)) = 1$.

This assumption excludes the sub-population of defiers. In the model without controls X, Imbens and Angrist (1994) used this assumption to identify the LATE, the local average treatment effect for compliers:

$$\text{LATE} = \mathbb{E}[Y(1) - Y(0)|G = CP],$$

which is often a quantity of interest. Frolich (2007) extended this result and showed that the LATE is identified in the model with controls as well, as long as we impose Assumption 2.1 in addition to Assumption 2.2. Frolich (2007), as well as following papers, e.g. Belloni et al. (2017), developed nonparametric and machine learning estimators of the LATE in the model with controls.

In practice, however, empirical researchers often prefer simple parametric alternatives, such as the 2SLS estimator. As argued in Blandhol et al. (2022), this could be problematic. Indeed, not only may the 2SLS estimator not converge in probability to the LATE, it may not have a causal interpretation at all. In particular, it may not take the form of a weighted-average treatment effect and even if it does, the weights may take negative values and may not integrate to one.

To cope with some of these problems, we propose a simple alternative to the 2SLS estimator. Let $(Y_1, T_1, X_1, Z_1), \dots, (Y_n, T_n, X_n, Z_n)$ be a random sample from the distribution of (Y, T, X, Z). Also, let

$$\Lambda(t) = \frac{\exp(t)}{1 + \exp(t)}, \quad t \in \mathbb{R}$$

be the logit function. Our estimator, which we refer to as *the logit-based IV estimator*, takes the following form.

Algorithm 2.1 (Logit-Based IV Estimator). *Proceed in two steps:*

(1) *Run the logit estimator of Z on X,*

$$\widehat{\theta} = \arg\max_{\theta \in \mathbb{R}^p} \sum_{i=1}^{n} \left(Z_i X_i^\top \theta - \log(1 + \exp(X_i^\top \theta)) \right);$$

(2) *Compute the logit-based IV estimator as the following ratio:*

$$\widehat{\beta}_\Lambda = \frac{\sum_{i=1}^n Y_i (Z_i - \Lambda(X_i^\top \widehat{\theta}))}{\sum_{i=1}^n T_i (Z_i - \Lambda(X_i^\top \widehat{\theta}))}.$$

Note that this estimator differs from the 2SLS estimator by using the logit regression residuals $Z_i - \Lambda(X_i^\top \widehat{\theta})$ in the second step instead of the linear regression residuals used by the 2SLS estimator. For clarity of presentation, and since this will be helpful for our discussion below, we provide the formal algorithm for the 2SLS estimator as well.

Algorithm 2.2 (2SLS Estimator). *Proceed in two steps:*

(1) *Run the OLS estimator of Z on X,*

$$\widehat{\gamma} = \arg\min_{\gamma \in \mathbb{R}^p} \sum_{i=1}^{n} (Z_i - X_i^\top \gamma)^2;$$

(2) *Compute the 2SLS estimator as the following ratio:*

$$\widehat{\beta}_{2SLS} = \frac{\sum_{i=1}^n Y_i (Z_i - X_i^\top \widehat{\gamma})}{\sum_{i=1}^n T_i (Z_i - X_i^\top \widehat{\gamma})}. \tag{3}$$

Algorithm 2.2 may not be the usual way to define the 2SLS estimator but it is easy to verify that it does define the 2SLS estimator by applying the Frisch-Waugh-Lovell theorem.[2]

Without further assumptions, neither logit-based IV nor 2SLS estimators may have a causal interpretation. To see why this is so, we first need to introduce some additional notations. For all $x \in \mathcal{X}$, let

$$\omega_{CP}(x) = \mathbb{P}(G = CP|X = x), \quad \omega_{AT}(x) = \mathbb{P}(G = AT|X = x), \quad \omega_{NT}(x) = \mathbb{P}(G = NT|X = x)$$

denote the x-conditional fractions of compliers, always-takers, and never-takers in the population, respectively. Also, let

$$\theta_0 = \arg\max_{\theta \in \mathbb{R}^p} \mathbb{E}[ZX^\top\theta - \log(1 + \exp(X^\top\theta))] \tag{4}$$

and

$$\gamma_0 = \arg\min_{\gamma \in \mathbb{R}^p} \mathbb{E}[(Z - X^\top\gamma)^2] \tag{5}$$

be the probability limits of the estimators $\widehat{\theta}$ and $\widehat{\gamma}$ appearing in Algorithms 2.1 and 2.2. Moreover, for all $x \in \mathcal{X}$, let

$$h_\Lambda(x) = \Lambda(x^\top\theta_0) \quad \text{and} \quad h_{2SLS}(x) = x^\top\gamma_0.$$

The following theorem derives the probability limits of both $\widehat{\beta}_\Lambda$ and $\widehat{\beta}_{2SLS}$.

Theorem 2.1. *Suppose that Assumptions 2.1 and 2.2 are satisfied. Then under appropriate regularity conditions,[3] for any $s \in [0,1]$ and $a \in \{\Lambda, 2SLS\}$, we have that $\widehat{\beta}_a \to_p \beta_a$, where*

$$\beta_a = \frac{\mathbb{E}[\Delta_{CP}(X)\omega_{CP}(X)(s\mathbb{E}[Z|X] + (1-s)h_a(X) - h_a(X)\mathbb{E}[Z|X])]}{\mathbb{E}[T(Z - h_a(X))]} \tag{6}$$

$$+ \frac{s\mathbb{E}[\Delta_{AT}(X)\omega_{AT}(X)(\mathbb{E}[Z|X] - h_a(X))]}{\mathbb{E}[T(Z - h_a(X))]} \tag{7}$$

[2] Notably, our logit-based IV estimator is not related to the "forbidden regression" discussed in Chapter 4.6.1 of Angrist and Pischke (2009), which refers to the use of the fitted value $\hat{T}_i$ in the second stage OLS regression of Y_i on $\hat{T}_i$ and X_i obtained from a non-linear regression of T_i on Z_i and X_i. Instead, our approach replaces the linear predictor $X_i^\top\widehat{\gamma}$ of Z_i by a nonlinear predictor $\Lambda(X_i^\top\widehat{\theta})$, so we are "partialling out" X_i using a nonlinear model. Importantly, under the assumption of constant treatment effects, so that $Y(1) - Y(0) = \rho$, our logit-based IV estimator is consistent for ρ under the same conditions as those required for consistency of the 2SLS estimator, as opposed to the forbidden regression, which requires extra conditions. In particular, it is easy to check that the logit-based IV estimator is consistent for ρ under Assumption 2.1 as long as the function $x \mapsto \mathbb{E}[Y(0)|X = x]$ is linear.

[3] To avoid distractions, we provide the list of regularity conditions for this theorem and all other results in the Appendix.

$$+\frac{(s-1)\mathbb{E}[\Delta_{NT}(X)\omega_{NT}(X)(\mathbb{E}[Z|X]-h_a(X))]}{\mathbb{E}[T(Z-h_a(X))]} \tag{8}$$

$$+\frac{\mathbb{E}[\mathbb{E}[sY(0)+(1-s)Y(1)|X](\mathbb{E}[Z|X]-h_a(X))]}{\mathbb{E}[T(Z-h_a(X))]}. \tag{9}$$

The proof of this theorem, as well as of all other results in the main text, is rather simple and is provided in the Appendix. In fact, the expression for the probability limit of the 2SLS estimator $\widehat{\beta}_{2SLS}$ in this theorem is closely related to that in Proposition 1 of Blandhol et al. (2022).

Theorem 2.1 shows that without further assumptions, the probability limits of both logit-based IV and 2SLS estimators take the form of a sum of complier, always-taker, never-taker, and non-causal terms, appearing in expressions (6), (7), (8), and (9), respectively. The last term is referred to here as a non-causal term because it does not take the form of a functional of the treatment effect $Y(1)-Y(0)$. In particular, this term depends non-trivially on the level of potential outcomes $Y(0)$ and $Y(1)$. The representations for the probability limits given in Theorem 2.1 are not unique as different values of the parameter s give different representations. For example, it is always possible to get rid of the never-taker term by substituting $s=1$ and it is always possible to get rid of the always-taker term by substituting $s=0$. However, without further assumptions, it is not possible to get rid of both always-taker and never-taker terms at the same time.

Theorem 2.1 identifies at least three problems with both logit-based IV and 2SLS estimators. First, the presence of the non-causal term means that neither logit-based IV nor 2SLS estimators in general converge in probability to a weighted-average treatment effect. Second, the always-taker term in (7) shows that the x-conditional average treatment effect for always-takers $\Delta_{AT}(x)$ has the weight $\omega_{AT}(x)(\mathbb{E}[Z|X=x]-h_a(x))]/\mathbb{E}[T(Z-h_a(x))]$, which may be negative for both logit-based IV and 2SLS estimators, and the same applies to the never-taker term in (8). Third, even if the non-causal term in (9) is zero, so that the corresponding probability limits β_Λ and β_{2SLS} take the form of the weighted-average treatment effects, the weights may not integrate to one. These are all the problems discussed in Blandhol et al. (2022) in the case of the 2SLS estimator.

Theorem 2.1 also identifies the key advantage of the logit-based IV estimator in comparison with the 2SLS estimator: for the former estimator, all the weights in the complier term are non-negative and this is not necessarily the case for the latter estimator. To see this advantage more clearly, we now impose several additional assumptions.

Assumption 2.3. *For some vector $\eta_0 \in \mathbb{R}^p$, we have $\mathbb{E}[(Y(1)-Y(0))T(0)+Y(0)|X] = X^\top \eta_0$ with probability one.*

This assumption, although non-standard, seems rather attractive. Indeed, it specifies a linear regression model for the conditional mean of $(Y(1)-Y(0))T(0)+Y(0)$ given X, and linear regression models have a long tradition in econometrics. In empirical work, even if regression functions are not believed to be exactly linear, they are believed to be approximately linear. In this sense, Assumption 2.3 is in line with a traditional regression analysis in economics. Moreover, as long as the conditional mean function $x \mapsto \mathbb{E}[(Y(1)-Y(0))T(0)+Y(0)|X=x]$ is continuous, it can be well approximated by a linear combination of, say, polynomial transformations of x. In such a case, Assumption 2.3 can be made more plausible if we replace X by a set of polynomial, or other technical, transformations of X. In practice, this amounts to replacing all X_i's by, say, $q(X_i)$'s, where $q(\cdot) = (q_1(\cdot), \dots, q_k(\cdot))^\top$ is a vector of corresponding transformations. Moreover, under Assumption 2.1,

$$\begin{aligned}\mathbb{E}[(Y(1)-Y(0))T(0)+Y(0)|X] &= \mathbb{E}[(Y(1)-Y(0))T(0)+Y(0)|X, Z=0]\\ &= \mathbb{E}[Y|X, Z=0],\end{aligned}$$

which implies that Assumption 2.3 is testable.

Assumption 2.4. *For some vector $\psi_0 \in \mathbb{R}^p$, we have $\mathbb{E}[T(0)|X] = X^\top \psi_0$ with probability one.*

This assumption specifies a linear regression model for the conditional mean of $T(0)$ given X. Given that $T(0)$ is a binary random variable, the conditional mean $\mathbb{E}[T(0)|X]$ takes values in the $(0,1)$ interval, and so this assumption may be less plausible than Assumption 2.3. However, we will use this assumption mainly to make the comparison between the logit-based IV and 2SLS estimators particularly transparent. Without this assumption. our logit-based IV estimator will still have a partially causal interpretation. Also, under Assumption 2.1,

$$\mathbb{E}[T(0)|X] = \mathbb{E}[T(0)|X, Z=0] = \mathbb{E}[T|X, Z=0],$$

which implies that Assumption 2.4 is testable as well. In addition, like Assumption 2.3, it can be made more plausible if we replace X by a set of appropriate transformations of X. Finally, we will discuss in the next section how one can modify the logit-based IV estimator to accommodate more plausible versions of Assumption 2.4.

By combining Theorem 2.1 with Assumptions 2.3 and 2.4, we obtain the following corollary.

Corollary 2.1. *Suppose that Assumptions 2.1, 2.2, 2.3, and 2.4 are satisfied. Then under appropriate regularity conditions, the probability limits* β_Λ *and* β_{2SLS} *appearing in Theorem 2.1 take the following form:*

$$\beta_\Lambda = \mathbb{E}[\Delta_{CP}(X)w_\Lambda(X)] \quad \textit{and} \quad \beta_{2SLS} = \mathbb{E}[\Delta_{CP}(X)w_{2SLS}(X)],$$

where

$$w_\Lambda(x) = \frac{\omega_{CP}(x)\mathbb{E}[Z|X=x](1-\Lambda(x^\top\theta_0))}{\mathbb{E}[\omega_{CP}(X)\mathbb{E}[Z|X](1-\Lambda(X^\top\theta_0))]}$$

and

$$w_{2SLS}(x) = \frac{\omega_{CP}(x)\mathbb{E}[Z|X=x](1-x^\top\gamma_0)}{\mathbb{E}[\omega_{CP}(X)\mathbb{E}[Z|X](1-X^\top\gamma_0)]}$$

for all $x \in \mathcal{X}$.

This corollary provides a clean comparison between logit-based IV and 2SLS estimators. It shows that under our assumptions, both estimators converge in probability to weighted-average treatment effects for compliers, and the weights do integrate to one:

$$\mathbb{E}[w_\Lambda(X)] = \mathbb{E}[w_{2SLS}(X)] = 1.$$

However, in the case of the 2SLS estimator, some of the weights may take negative values, which happens whenever $x^\top\gamma_0$ exceeds one. At the same time, the weights of the logit-based IV estimator are always non-negative, as the logit function $\Lambda(\cdot)$ takes values in the $(0,1)$ interval. Thus, under our assumptions, the logit-based IV estimator has a causal interpretation and the 2SLS estimator does not. This explains the main advantage of the logit-based IV estimator relative to the traditionally used 2SLS estimator.

We emphasize here that using other binary choice models instead of logit in Algorithm 2.1 may not necessarily work. For example, if we were to replace the logit model by the probit model, under our assumptions, we would obtain an IV estimator whose probability limit does not necessarily take the form of the weighted-average treatment effect.

We now discuss two extensions of Corollary 2.1. First, without imposing Assumption 2.4, we obtain somewhat more convoluted expressions for the probability limits of the logit-based IV and 2SLS estimators, which are nonetheless useful to make comparisons between these two estimators.

Corollary 2.2. *Suppose that Assumptions 2.1, 2.2, and 2.3 are satisfied. Then under appropriate regularity conditions, the probability limits β_Λ and β_{2SLS} appearing in Theorem 2.1 take the following form:*

$$\beta_\Lambda = \mathbb{E}[\Delta_{CP}(X) w_\Lambda(X)] \quad \textit{and} \quad \beta_{2SLS} = \mathbb{E}[\Delta_{CP}(X) w_{2SLS}(X)],$$

where

$$w_\Lambda(x) = \frac{\omega_{CP}(x)\mathbb{E}[Z|X=x](1-\Lambda(x^\top\theta_0))}{\mathbb{E}[\omega_{CP}(X)\mathbb{E}[Z|X](1-\Lambda(X^\top\theta_0))] + \mathbb{E}[\omega_{AT}(X)(\mathbb{E}[Z|X] - \Lambda(X^\top\theta_0))]} \tag{10}$$

and

$$w_{2SLS}(x) = \frac{\omega_{CP}(x)\mathbb{E}[Z|X=x](1-x^\top\gamma_0)}{\mathbb{E}[\omega_{CP}(X)\mathbb{E}[Z|X](1-X^\top\gamma_0)] + \mathbb{E}[\omega_{AT}(X)(\mathbb{E}[Z|X] - X^\top\gamma_0)]} \tag{11}$$

for all $x \in \mathcal{X}$.

This corollary shows that without imposing Assumption 2.4, both logit-based IV and 2SLS estimators still converge in probability to weighted-average treatment effects for compliers but the weights now do not integrate to one. On the other hand, as long as the fraction of always-takers is not too large relative to the fraction of compliers, so that the denominators in (10) and (11) remain non-negative for all $x \in \mathcal{X}$, the main advantage of the logit-based IV estimator remains valid: its corresponding weights are still non-negative and the estimator has a partially causal interpretation, whereas the weights of the 2SLS estimator may be negative and the estimator does not have a partially causal interpretation.

Second, we can relax Assumptions 2.3 and 2.4 without losing causal interpretability of the logit-based IV estimator. Indeed, consider the following assumption.

Assumption 2.5. *For some constant $s \in [0,1]$ and some vectors η_0 and ψ_0 in $\mathbb{R}^p$, we have*

$$\mathbb{E}\Big[(Y(1)-Y(0))(sT(0)+(1-s)T(1)) + Y(0)|X\Big] = X^\top\eta_0 \tag{12}$$

and

$$\mathbb{E}[sT(0)+(1-s)T(1)|X] = X^\top\psi_0 \tag{13}$$

with probability one.

This assumption relaxes Assumptions 2.3 and 2.4 as it reduces to Assumptions 2.3 and 2.4 if we plugin $s = 1$. It requires that there exists a linear combination of the conditional mean functions $x \mapsto \mathbb{E}[(Y(1) - Y(0))T(1) + Y(0)|X = x]$ and $x \mapsto \mathbb{E}[(Y(1) - Y(0))T(0) + Y(0)|X = x]$ that is linear and that the linear combination of the conditional mean functions $x \mapsto \mathbb{E}[T(1)|X = x]$ and $x \mapsto \mathbb{E}[T(0)|X = x]$ with the same weights is linear as well. All the comments we made about Assumptions 2.3 and 2.4 apply to this assumption as well. In particular, one can show that equations (12) and (13) can be equivalently rewritten as

$$s\mathbb{E}[Y|X, Z = 0] + (1 - s)\mathbb{E}[Y|X, Z = 1] = X^\top \eta_0$$

and

$$s\mathbb{E}[T|X, Z = 0] + (1 - s)\mathbb{E}[T|X, Z = 1] = X^\top \psi_0,$$

respectively. Thus, Assumption 2.5 is testable.

Corollary 2.3. *Suppose that Assumptions 2.1, 2.2, and 2.5 are satisfied. Then under appropriate regularity conditions, the probability limit β_Λ appearing in Theorem 2.1 take the following form:*

$$\beta_\Lambda = \mathbb{E}[\Delta_{CP}(X) w_{\Lambda,s}(X)]$$

where

$$w_{\Lambda,s}(x) = \frac{\omega_{CP}(x)\Big(s\mathbb{E}[Z|X = x] + (1 - s)\Lambda(x^\top \theta_0) - \Lambda(x^\top \theta_0)\mathbb{E}[Z|X = x]\Big)}{\mathbb{E}\Big[\omega_{CP}(X)\Big(s\mathbb{E}[Z|X] + (1 - s)\Lambda(X^\top \theta_0) - \Lambda(X^\top \theta_0)\mathbb{E}[Z|X]\Big)\Big]}$$

for all $x \in \mathcal{X}$.

This corollary shows that under Assumptions 2.1 and 2.2, Assumption 2.5 is sufficient for causal interpretability of the logit-based IV estimator. The weights, however, take a more complicated form than those in Corollary 2.1. Note that a version of this corollary for the 2SLS estimator can be provided as well but, like in Corollary 2.1, the weights for the 2SLS estimator may take negative values.

To conclude this section, we demonstrate that the logit-based IV estimator has a causal interpretation even if Assumption 2.5 is not satisfied as long as it consistently estimates the conditional mean function $x \mapsto \mathbb{E}[Z|X = x]$. Indeed, consider the following assumption.

Assumption 2.6. *The conditional mean function $x \mapsto \mathbb{E}[Z|X = x]$ takes the logit form, i.e. $\mathbb{E}[Z|X] = \Lambda(X^\top \theta_0)$ with probability one.*

Note that in general, the assumption that the conditional mean function $x \mapsto \mathbb{E}[Z|X = x]$ takes the logit form means that there exists some $\theta \in \mathbb{R}^p$ such that

$\mathbb{E}[Z|X] = \Lambda(X^\top\theta)$ with probability one. However, given the definition of θ_0 in (4), this θ should be equal to θ_0, and so we simply assume that $\mathbb{E}[Z|X] = \Lambda(X^\top\theta_0)$.

Corollary 2.4. *Suppose that Assumptions 2.1, 2.2, and 2.6 are satisfied. Then under appropriate regularity conditions, the probability limit β_Λ appearing in Theorem 2.1 takes the following form:*

$$\beta_\Lambda = \mathbb{E}[\Delta_{CP}(X)w_0(X)],$$

where

$$w_0(x) = \frac{\omega_{CP}(x)\mathbb{E}[Z|X=x](1-\mathbb{E}[Z|X=x])}{\mathbb{E}[\omega_{CP}(X)\mathbb{E}[Z|X](1-\mathbb{E}[Z|X])]} \tag{14}$$

for all $x \in \mathcal{X}$.

Together with Corollary 2.3, this corollary shows that our logit-based IV estimator has a double robustness property, meaning that it has a causal interpretation if at least one of two conditions holds: either Assumption 2.5 or Assumption 2.6 is satisfied. Interestingly, however, the weights $w_{\Lambda,s}(\cdot)$ and $w_0(\cdot)$ appearing in Corollaries 2.3 and 2.4 may be different, which means that even though we have a causal interpretation in both cases, the parameter we are estimating depends on which condition is being satisfied.

Remark 2.1. When studying the 2SLS estimator, researchers often assume that the conditional mean function $x \mapsto \mathbb{E}[Z|X = x]$ takes the linear form, e.g. see Abadie (2003), Kolesar (2013), and Sloczynski (2020). In particular, with Assumptions 2.1 and 2.2 being maintained, as discussed in the Introduction, under this linear form condition, the 2SLS estimator has a causal interpretation. It is therefore useful to compare this linear form condition with our logit form condition in Assumption 2.6. With saturated controls, the linear form condition and the logit form condition are both satisfied, and so both logit-based IV and 2SLS estimators have a causal interpretation. Without saturated controls, however, it is unlikely that both conditions are satisfied simultaneously. In this case, given that Z is a binary random variable, so that the conditional mean function $x \mapsto \mathbb{E}[Z|X = x]$ takes values in the $(0, 1)$ interval, the logit form condition seems more reasonable that the linear form condition, and so our logit-based IV estimator is more likely to have a causal interpretation than the 2SLS estimator. □

3. Augmented Logit-Based IV Estimator

In this section, we replace Assumption 2.4 by a more plausible assumption and show how one can modify the logit-based IV estimator in order to obtain an estimator that still has a causal interpretation. Our modified estimator is similar to the original logit-based IV estimator but includes an extra covariate in the logit regression of Z on X.

Let $\Phi(\cdot)$ be a function mapping $\mathbb{R}$ to $[0,1]$ and consider the following alternative to Assumption 2.4.

Assumption 3.1. *For some vector $\psi_0 \in \mathbb{R}^p$, we have $\mathbb{E}[T(0)|X] = \Phi(X^\top \psi_0)$ with probability one.*

If we set $\Phi(t) = \min(\max(0,t),1)$ for all $t \in \mathbb{R}$, then Assumption 3.1 relaxes Assumption 2.4. Indeed, in this case, two assumptions are the same if the support of $X^\top \psi_0$ is contained in the $[0,1]$ interval but Assumption 3.1 does not actually require the support of $X^\top \psi_0$ to be contained in the $[0,1]$ interval. We are, however, primarily interested in the cases where the function $\Phi(\cdot)$ is nonlinear and smooth, e.g. $\Phi(\cdot)$ takes the logit or the probit form. In these cases, Assumption 3.1 seems more plausible than Assumption 2.4, as the function $x \mapsto \Phi(x^\top \psi_0)$ is smooth and automatically satisfies the constraint that the conditional mean of $T(0)$ given X takes values in the $(0,1)$ interval, without restricting the support of X. Like Assumptions 2.3 and 2.4, Assumption 3.1 is testable and can be made more plausible if we replace X by a set of appropriate transformations of X.

As it turns out, we can modify our logit-based IV estimator in a way so that it has a causal interpretation even if we replace Assumption 2.4 by Assumption 3.1. The modification, which yields *the augmented logit-based IV estimator*, is explained in the algorithm below.

Algorithm 3.1 (Augmented Logit-Based IV Estimator)**.** *Proceed in three steps:*

(1) *Run the following maximum likelihood estimator using the data with $Z = 0$ only,*

$$\widehat{\psi} = \arg\max_{\psi \in \mathbb{R}^p} \sum_{i=1}^{n} \mathbb{1}\{Z_i = 0\}\Big(T_i \log(\Phi(X_i^\top \psi)) + (1 - T_i)\log(1 - \Phi(X_i^\top \psi))\Big);$$

(2) *Run the logit estimator of Z on X and $\Phi(X^\top \widehat{\psi})$,*

$$(\widehat{\theta}, \widehat{\kappa}) = \arg\max_{\theta \in \mathbb{R}^p, \kappa \in \mathbb{R}} \sum_{i=1}^{n} \Big(Z_i(X_i^\top \theta + \widehat{C}_i \kappa) - \log(1 + \exp(X_i^\top \theta + \widehat{C}_i \kappa))\Big),$$

where we denoted $\widehat{C}_i = \Phi(X_i^\top \widehat{\psi})$ *for all* $i = 1, \ldots, n$*;*

(3) *Compute the augmented logit-based IV estimator as the following ratio:*

$$\widehat{\beta}_{A\Lambda} = \frac{\sum_{i=1}^n Y_i(Z_i - \Lambda(X_i^\top \widehat{\theta} + \widehat{C}_i \widehat{\kappa}))}{\sum_{i=1}^n T_i(Z_i - \Lambda(X_i^\top \widehat{\theta} + \widehat{C}_i \widehat{\kappa}))}.$$

This estimator requires that we know the link function Φ in the single-index structure $\Phi(X^\top \psi_0)$ used to model the conditional mean of $T(0)$ given X. However, from a practical point of view, different link functions, such as logit or probit, often lead to similar results, and the researcher can always check whether it is indeed the case by trying several link functions. Also, as pointed out above, the researcher can test whether a particular link function is consistent with the data. Finally, the researcher can estimate the link function Φ, as described in Chapter 2 of Horowitz (2009), for example.

To derive the probability limit of this estimator, define

$$\overline{\psi}_0 = \arg\max_{\psi \in \mathbb{R}^p} \mathbb{E}\Big[\mathbb{1}\{Z = 0\}\Big(T \log(\Phi(X^\top \psi)) + (1 - T)\log(1 - \Phi(X^\top \psi))\Big)\Big].$$

Note that $\bar{\psi}_0$ equals ψ_0 if Assumption 3.1 is satisfied, but may be different from ψ_0 otherwise. Also, define

$$(\overline{\theta}_0, \overline{\kappa}_0) = \arg\max_{\theta \in \mathbb{R}^p, \kappa \in \mathbb{R}} \mathbb{E}[Z(X^\top \theta + \Phi(X^\top \overline{\psi}_0)\kappa) - \log(\exp(X^\top \theta + \Phi(X^\top \overline{\psi}_0)\kappa))]. \quad (15)$$

Moreover, define

$$h_{A\Lambda}(x) = \Lambda(x^\top \overline{\theta}_0 + \Phi(x^\top \overline{\psi}_0)\overline{\kappa}_0)$$

for all $x \in \mathcal{X}$. The following theorem provides the probability limit of the estimator $\widehat{\beta}_{A\Lambda}$ imposing only Assumptions 2.1 and 2.2.

Theorem 3.1. *Suppose that Assumptions 2.1 and 2.2 are satisfied. Then under appropriate regularity conditions, we have that* $\widehat{\beta}_{A\Lambda} \to_p \beta_{A\Lambda}$*, where*

$$\begin{aligned}\beta_{A\Lambda} = & \frac{\mathbb{E}[\Delta_{CP}(X)\omega_{CP}(X)\mathbb{E}[Z|X](1 - h_{A\Lambda}(X))]}{\mathbb{E}[T(Z - h_{A\Lambda}(X))]} \\ & + \frac{\mathbb{E}[\Delta_{AT}(X)\omega_{AT}(X)(\mathbb{E}[Z|X] - h_{A\Lambda}(X))]}{\mathbb{E}[T(Z - h_{A\Lambda}(X))]} \\ & + \frac{\mathbb{E}[\mathbb{E}[Y(0)|X](\mathbb{E}[Z|X] - h_{A\Lambda}(X))]}{\mathbb{E}[T(Z - h_{A\Lambda}(X))]}.\end{aligned}$$

This theorem shows that the probability limit of the augmented logit-based IV estimator has the same structure as those of the logit-based IV and 2SLS estimators.

In particular, it shows that imposing only a minimal set of conditions used in the program evaluation literature is not sufficient to give the augmented logit-based IV estimator a causal interpretation, which is similar to conclusions in the previous section and in Blandhol et al. (2022) for the logit-based IV and 2SLS estimators, respectively. To obtain a causal interpretation, we need to impose additional conditions. The following corollary provides the probability limit of the estimator $\widehat{\beta}_{A\Lambda}$ under the condition that either Assumption 2.4 or Assumption 3.1 is satisfied, along with Assumptions 2.1, 2.2, and 2.3 used in the previous section.

Corollary 3.1. *Suppose that Assumptions 2.1, 2.2 and 2.3 are satisfied. In addition, suppose that either Assumption 2.4 or Assumption 3.1 is satisfied. Then under appropriate regularity conditions,* $\widehat{\beta}_{A\Lambda} \to_p \beta_{A\Lambda}$, *where*

$$\beta_{A\Lambda} = \mathbb{E}[\Delta_{CP}(X) w_{A\Lambda}(X)]$$

and

$$w_{A\Lambda}(x) = \frac{\omega_{CP}(x)\mathbb{E}[Z|X=x](1-\Lambda(x^\top\overline{\theta}_0 + \Phi(x^\top\overline{\psi}_0)\overline{\kappa}_0))}{\mathbb{E}[\omega_{CP}(X)\mathbb{E}[Z|X](1-\Lambda(X^\top\overline{\theta}_0 + \Phi(X^\top\overline{\psi}_0)\overline{\kappa}_0))]}$$

for all $x \in \mathcal{X}$.

This corollary demonstrates that the augmented logit-based IV estimator has a causal interpretation in a wider set of cases than the logit-based IV estimator itself. In particular, both have a causal interpretation if Assumption 2.4 is satisfied, along with other conditions, but the former has a causal interpretation even if Assumption 2.4 is not satisfied, as long as the correct link function Φ is being used. Interestingly, however, the weights $w_{A\Lambda}(\cdot)$ appearing in this theorem are generally different from the weights $w_{\Lambda}(\cdot)$ appearing in Corollary 2.1, which means that even when both estimators have a causal interpretation, they generally estimate different quantities.

In principle, we could extend Corollary 3.1 by relaxing Assumptions 2.3 and 3.1 in the same way Corollary 2.3 extends Corollary 2.1. However, this would make Step 1 of Algorithm 3.1 much more complicated. In particular, we would have to estimate the vector of parameters $\overline{\psi}_0 \in \mathbb{R}^p$ in the model asserting that

$$s\mathbb{E}[T|X, Z=1] + (1-s)\mathbb{E}[T|X, Z=0] = \Phi(X^\top\overline{\psi}_0)$$

for some $s \in \mathbb{R}$ and $\overline{\psi}_0 \in \mathbb{R}^p$. We therefore refrain from carrying out this extension. On the other hand, we could also consider an estimator that is similar to the one described in Algorithm 3.1 but using the data with $Z = 1$ on the first step. For such an estimator, it is clearly possible to derive a result like that In Theorem 3.1 as long as we replace $T(0)$ in Assumption 3.1 by $T(1)$.

To conclude this section, we note that as in the previous section, Assumptions 2.3, 2.4, and 3.1 are not needed for a causal interpretation of the augment logit-based IV estimator if the conditional mean function $x \mapsto \mathbb{E}[Z|X=x]$ takes the logit form, i.e. Assumption 2.6 is satisfied. Indeed, we have the following result.

Corollary 3.2. *Suppose that Assumptions 2.1, 2.2 and 2.6 are satisfied. Then under appropriate regularity conditions, $\widehat{\beta}_{A\Lambda} \to_p \beta_{A\Lambda}$, with the probability limit $\beta_{A\Lambda}$ taking the following form:*

$$\beta_{A\Lambda} = \mathbb{E}[\Delta_{CP}(X) w_0(X)],$$

where

$$w_0(x) = \frac{\omega_{CP}(x)\mathbb{E}[Z|X=x](1-\mathbb{E}[Z|X=x])}{\mathbb{E}[\omega_{CP}(X)\mathbb{E}[Z|X](1-\mathbb{E}[Z|X])]} \tag{16}$$

for all $x \in \mathcal{X}$.

Together with Corollary 3.1, this corollary shows that the augmented logit-based IV estimator has a triple robustness property, meaning that it has a causal interpretation if at least one of three conditions holds: either Assumptions 2.3 and 2.4 are satisfied, Assumptions 2.3 and 3.1 are satisfied, or Assumption 2.6 is satisfied. In the latter case, the logit-based IV and the augmented logit-based IV estimators have the same probability limits as the weights $w_0(\cdot)$ in (14) coincide with the weights $w_0(\cdot)$ in (16), i.e. $\widehat{\beta}_{\Lambda}$ and $\widehat{\beta}_{A\Lambda}$ estimate the same quantity.

4. Asymptotic Distribution Theory

In this section, we describe the asymptotic distribution of the logit-based IV estimator $\widehat{\beta}_{\Lambda}$ and of the augmented logit-based IV estimator $\widehat{\beta}_{A\Lambda}$. We do so without imposing Assumptions 2.3, 2.4, 2.5, and 3.1 and without assuming that the conditional mean function $x \mapsto \mathbb{E}[Z|X=x]$ takes the logit form (Assumption 2.6). We thus allow for general misspecification, with the probability limits β_{Λ} and $\beta_{A\Lambda}$ of the estimators being given by formulas in Theorems 2.1 and 3.1, respectively. For the sake of notational simplicity, we assume that $\Phi(\cdot) = \Lambda(\cdot)$ throughout the rest of this paper.

To describe the asymptotic distribution of $\widehat{\beta}_{\Lambda}$, let

$$\varphi_0 = (\mathbb{E}[\Lambda'(X^\top\theta_0)XX^\top])^{-1}\mathbb{E}[\Lambda'(X^\top\theta_0)X(Y - T\beta_{\Lambda})],$$

which is a vector of coefficients in the weighted projection of $Y - T\beta_\Lambda$ on X. Also, let

$$\ell_i^\Lambda = (Y_i - T_i\beta_\Lambda - X_i^\top\varphi_0)(Z_i - \Lambda(X_i^\top\theta_0))$$

for all $i = 1, \dots, n$ and let ℓ^Λ be defined analogously with (Y, T, X, Z) replacing (Y_i, T_i, X_i, Z_i). The following theorem derives the asymptotic distribution of $\widehat{\beta}_\Lambda$.

Theorem 4.1. *Suppose that Assumptions 2.1 and 2.2 are satisfied. Then under appropriate regularity conditions,*

$$\sqrt{n}(\widehat{\beta}_\Lambda - \beta_\Lambda) = \frac{n^{-1/2}\sum_{i=1}^n \ell_i^\Lambda}{\mathbb{E}[T(Z - \Lambda(X^\top\theta_0))]} + o_p(1) \to_d N(0, \sigma_\Lambda^2),$$

where $\sigma_\Lambda^2 = \mathbb{E}[(\ell^\Lambda)^2]/(\mathbb{E}[T(Z - \Lambda(X^\top\theta_0))])^2$.

To describe the asymptotic distribution of $\widehat{\beta}_{A\Lambda}$, let $C = \Lambda(X^\top\overline{\psi}_0)$ and $C_i = \Lambda(X_i^\top\overline{\psi}_0)$ for all $i = 1, \dots, n$. Also, let $W = (X^\top, C)^\top$ and $W_i = (X_i^\top, C_i)^\top$ for all $i = 1, \dots, n$. In addition, let e be the vector in $\mathbb{R}^{p+1}$ such that its last component is one and all other components are zero. Moreover, let

$$\xi_0 = \left(\mathbb{E}[\Lambda'(X^\top\overline{\theta}_0 + C\overline{\kappa}_0)WW^\top]\right)^{-1}\left(\mathbb{E}[\Lambda'(X^\top\overline{\theta}_0 + C\overline{\kappa}_0)W(Y - T\beta_{A\Lambda})]\right),$$

which is the vector of coefficients in the weighted projection of $Y - T\beta_{A\Lambda}$ on W,

$$A_1 = \mathbb{E}\left[\{(Z - \Lambda(X^\top\overline{\theta}_0 + C\overline{\kappa}_0))e - \overline{\kappa}_0\Lambda'(X^\top\overline{\theta}_0 + C\overline{\kappa}_0)W\}\Lambda'(X^\top\overline{\psi}_0)X^\top\right],$$

$$A_2 = \xi_o^\top A_1 + \overline{\kappa}_0\mathbb{E}\left[\Lambda'(X^\top\overline{\theta}_0 + C\overline{\kappa}_0)\Lambda'(X^\top\overline{\psi}_0)(Y - T\beta_{A\Lambda})\right].$$

Finally, let

$$\ell_{i,1}^{A\Lambda} = (Y_i - T_i\beta_{A\Lambda} - W_i^\top\xi_0)(Z_i - \Lambda(X_i^\top\overline{\theta}_0 + C_i\overline{\kappa}_0)),$$

$$\ell_{i,2}^{A\Lambda} = A_2\left(\mathbb{E}[\mathbb{1}\{Z = 0\}\Lambda'(X^\top\overline{\psi}_0)XX^\top]\right)^{-1}\mathbb{1}\{Z_i = 0\}(T_i - \Lambda(X_i^\top\overline{\psi}_0))X_i$$

for all $i = 1, \dots, n$ and let $\ell_1^{A\Lambda}$ and $\ell_2^{A\Lambda}$ be defined analogously with (Y, T, X, Z) replacing (Y_i, T_i, X_i, Z_i). The following theorem derives the asymptotic distribution of $\widehat{\beta}_{A\Lambda}$.

Theorem 4.2. *Suppose that Assumptions 2.1 and 2.2 are satisfied. Then under appropriate regularity conditions,*

$$\sqrt{n}(\widehat{\beta}_{A\Lambda} - \beta_{A\Lambda}) = \frac{n^{-1/2}\sum_{i=1}^n(\ell_{i,1}^{A\Lambda} - \ell_{i,2}^{A\Lambda})}{\mathbb{E}[T(Z - \Lambda(X^\top\overline{\theta}_0 + C\overline{\kappa}_0))]} + o_p(1) \to_d N(0, \sigma_{A\Lambda}^2),$$

where $\sigma_{A\Lambda}^2 = \mathbb{E}[(\ell_1^{A\Lambda} - \ell_2^{A\Lambda})^2]/(\mathbb{E}[T(Z - \Lambda(X^\top\overline{\theta}_0 + C\overline{\kappa}_0))])^2$.

In this theorem, the terms $\ell_{i,1}^{A\Lambda}$ are analogous to the terms ℓ_i^{Λ} in Theorem 4.1 and the terms $\ell_{i,2}^{A\Lambda}$ capture the extra noise appearing in Step 1 of Algorithm 3.1.

The asymptotic variances σ_{Λ}^2 and $\sigma_{A\Lambda}^2$ appearing in these theorems can clearly be estimated by a plugin method, and it is standard to provide conditions under which such estimators will be consistent. We omit more detailed discussion for the sake of paper brevity.

5. Hausman Specification Test

In Sections 2 and 3, we showed that one of the cases where the logit-based IV and augmented logit-based IV estimators $\widehat{\beta}_{\Lambda}$ and $\widehat{\beta}_{A\Lambda}$ have a causal interpretation is the case where the conditional mean function $x \mapsto \mathbb{E}[Z|X = x]$ takes the logit form, i.e. Assumption 2.6 is satisfied. In this section, we develop a Hausman test to check whether Assumption 2.6 is indeed satisfied, following the original work in Hausman (1978). Throughout this section, we will implicitly maintain Assumptions 2.1 and 2.2.

To describe the Hausman test, observe that under Assumption 2.6, it follows from Corollaries 2.4 and 3.2 that $\beta_{\Lambda} = \beta_{A\Lambda}$. Therefore, to test whether Assumption 2.6 is satisfied, it makes sense to check whether the estimators $\widehat{\beta}_{\Lambda}$ and $\widehat{\beta}_{A\Lambda}$ are sufficiently close to each other. In turn, the asymptotic distribution of the difference $\widehat{\beta}_{\Lambda} - \widehat{\beta}_{A\Lambda}$ can be obtained from the asymptotic expansions in Theorems 4.1 and 4.2. Indeed, under Assumption 2.6, we have $\overline{\theta}_0 = \theta_0$ and $\overline{\kappa}_0 = 0$, and so, by Theorems 4.1 and 4.2,

$$\sqrt{n}(\widehat{\beta}_{\Lambda} - \widehat{\beta}_{A\Lambda}) = \frac{n^{-1/2}\sum_{i=1}^{n}(\ell_i^{\Lambda} - \ell_{i,1}^{A\Lambda} + \ell_{i,2}^{A\Lambda})}{\mathbb{E}[T(Z - \Lambda(X^{\top}\theta_0))]} + o_p(1) \to N(0, \sigma_H^2),$$

where

$$\sigma_H^2 = \frac{\mathbb{E}[(\ell^{\Lambda} - \ell_1^{A\Lambda} + \ell_2^{A\Lambda})^2]}{(\mathbb{E}[T(Z - \Lambda(X^{\top}\theta_0))])^2}.$$

The Hausman test therefore rejects the null hypothesis that Assumption 2.6 is satisfied if the test statistic

$$\frac{|\sqrt{n}(\widehat{\beta}_{\Lambda} - \widehat{\beta}_{A\Lambda})|}{\widehat{\sigma}_H} \tag{17}$$

exceeds the critical value $z_{1-\alpha/2}$, where $\widehat{\sigma}_H$ is the plugin estimator of σ_H, α is the nominal level of the test and $z_{1-\alpha/2}$ is the number such that a standard normal random variable exceeds this number with probability $\alpha/2$.

Being parametric, the test we have just described may not have power against some alternatives. However, as long as Assumption 2.6 is not satisfied, we will generically have $(\bar{\theta}_0, \bar{\kappa}_0) \neq (\theta_0, 0)$, in which case the function $h_{A\Lambda}(\cdot)$ is different from the function $h_\Lambda(\cdot)$ and, as follows from Theorems 2.1 and 4.2, β_Λ is different from $\beta_{A\Lambda}$. Thus, the Hausman test will have power against most alternatives.

In addition, we can consider a split-sample version of the Hausman test. To describe it, randomly split the whole sample $\mathcal{I} = \{1, \dots, n\}$ into two subsamples $\mathcal{I}_1$ and $\mathcal{I}_2$ of equal size[4] and calculate the logit-based IV estimator using the subsample $\mathcal{I}_1$ and the augmented logit-based IV estimator using the subsample $\mathcal{I}_2$. Call these estimators $\widehat{\beta}_{\Lambda,1}$ and $\widehat{\beta}_{A\Lambda,2}$, respectively. Then, under Assumption 2.6, these two estimators have the same probability limits and, by Theorems 4.1 and 4.2,

$$\sqrt{n}(\widehat{\beta}_{\Lambda,1} - \widehat{\beta}_{A\Lambda,2}) = \frac{2n^{-1/2}(\sum_{i\in\mathcal{I}_1} \ell_i^\Lambda - \sum_{i\in\mathcal{I}_2}(\ell_{i,1}^{A\Lambda} + \ell_{i,2}^{A\Lambda}))}{\mathbb{E}[T(Z - \Lambda(X^\top\theta_0))]} \to_d N(0, \sigma_{H,2}^2),$$

where

$$\sigma_{H,2}^2 = \frac{2(\mathbb{E}[(\ell^\Lambda)^2] + \mathbb{E}[(\ell_1^{A\Lambda} + \ell_2^{A\Lambda})^2])}{(\mathbb{E}[T(Z - \Lambda(X^\top\theta_0))])^2}.$$

The split-sample Hausman test therefore rejects the null hypothesis that Assumption 2.6 is satisfied if the test statistic

$$\frac{|\sqrt{n}(\widehat{\beta}_{\Lambda,1} - \widehat{\beta}_{A\Lambda,2})|}{\widehat{\sigma}_{H,2}}$$

exceeds the critical value $z_{1-\alpha/2}$, where $\widehat{\sigma}_{H,2}$ is the plugin estimator of $\sigma_{H,2}$ and the rest is the same as before.

Being split-sample, this version of the Hausman test may be somewhat less powerful than the one described above. However, it is more robust in terms of size control if it incidentally happens that σ_H^2 is close to zero, in which case the distribution of the test statistic in (17) may not be well approximated by the standard normal distribution.

Remark 5.1. The main advantage of the Hausman tests we described in this section is their simplicity. We note, however, that there exist numerous nonparametric tests in the literature that are more complicated to implement but might have better power against some alternatives, e.g. see Bierens (1982), Hardle and Mammen (1993), Horowitz and Spokoiny (2001) for classical tests and Sorensen (2022) for

[4]If n is odd, split $\mathcal{I}$ so that the number of observations in $\mathcal{I}_1$ is equal to that in $\mathcal{I}_2$ plus one.

recent developments. On the other hand, it does not seem to be the case that the power of any of these tests uniformly dominates that of the Hausman tests. □

Appendix A. Regularity Conditions

In this section, we provide regularity conditions for all the theorems and corollaries in the paper, which were omitted in the main text.

Theorem 2.1 and Corollaries 2.1, 2.2, 2.3, and 2.4 use the following regularity conditions:

Assumption A.1. *We have (i)* $\mathbb{E}[Y^2] < \infty$*, (ii)* $\mathbb{E}[\|X\|^2] < \infty$ *and (iii)* $\mathbb{E}[\Lambda'(X^\top\theta_0)XX^\top]$ *is non-singular. In addition, (iv)* $\mathbb{E}[T(Z-\Lambda(X^\top\theta_0))] \neq 0$ *when we analyze the logit-based IV estimator; and* $\mathbb{E}[T(Z-X^\top\gamma_0)] \neq 0$ *when we analyze the 2SLS estimator.*

Theorems 3.1 and 4.1 and Corollaries 3.1 and 3.2, uses the following regularity conditions:

Assumption A.2. *We have (i)* $\mathbb{E}[Y^2] < \infty$*, (ii)* $\mathbb{E}[\|X\|^2] < \infty$*, (iii)* $\mathbb{P}(Z=0) > 0$*, (iv)* $\Phi(\cdot)$ *is Lipschitz-continuous, (v)* $\mathbb{E}[\Lambda'(X^\top\overline{\psi}_0)XX^\top|Z=0]$ *is non-singular, (vi)* $\mathbb{E}[\Lambda'(X^\top\overline{\theta}_0 + C\overline{\kappa}_0)WW^\top]$ *is non-singular for* $W = (X^\top, C)^\top$ *where* $C = \Phi(X^\top\overline{\psi}_0)$*, and (vii)* $\mathbb{E}[T(Z-\Lambda(X^\top\overline{\theta}_0 + C\overline{\kappa}_0))] \neq 0$.

Theorem 4.2 uses the following regularity conditions:

Assumption A.3. *We have (i)* $\mathbb{E}[Y^2] < \infty$*, (ii)* $\mathbb{E}[\|X\|^4] < \infty$*, (iii)* $\mathbb{P}(Z=0) > 0$*, (iv)* $\mathbb{E}[\Lambda'(X^\top\overline{\psi}_0)XX^\top|Z=0]$ *is non-singular, (v)* $\mathbb{E}[\Lambda'(X^\top\overline{\theta}_0 + C\overline{\kappa}_0)WW^\top]$ *is non-singular for* $W = (X^\top, C)^\top$ *where* $C = \Lambda(X^\top\overline{\psi}_0)$*, and (vi)* $\mathbb{E}[T(Z-\Lambda(X^\top\overline{\theta}_0 + C\overline{\kappa}_0))] \neq 0$.

Appendix B. Proofs for Section 2

Proof of Theorem 2.1. Observe that under Assumption A.1, we have $\widehat{\theta} \to_p \theta_0$ and $\widehat{\gamma} \to_p \gamma_0$, for example, by Theorem 2.7 in Newey and McFadden (1994). Thus, by standard arguments, again under Assumption A.1,

$$\frac{\sum_{i=1}^n Y_i(Z_i - \Lambda(X_i^\top\widehat{\theta}))}{\sum_{i=1}^n T_i(Z_i - \Lambda(X_i^\top\widehat{\theta}))} \to_p \frac{\mathbb{E}[Y(Z-\Lambda(X^\top\theta_0))]}{E[T(Z-\Lambda(X^\top\theta_0))]} \tag{18}$$

and

$$\frac{\sum_{i=1}^n Y_i(Z_i - X_i^\top\widehat{\gamma})}{\sum_{i=1}^n T_i(Z_i - X_i^\top\widehat{\gamma})} \to_p \frac{\mathbb{E}[Y(Z-X^\top\gamma_0)]}{\mathbb{E}[T(Z-X^\top\gamma_0)]}. \tag{19}$$

We thus only need to show that the probability limits appearing here coincide with those in the statement of Theorem 2.1, which is what we do below.

Fix $a \in \{\Lambda, 2SLS\}$ and $s \in [0, 1]$. Let $\Delta Y = Y(1) - Y(0)$ and $\Delta T = T(1) - T(0)$. Then Y can be decomposed as

$$Y = \Delta YT + Y(0) = \Delta Y(\Delta TZ + T(0)) + Y(0) = \Delta Y\Delta TZ + \Delta YT(0) + Y(0). \tag{20}$$

Here,

$$\begin{aligned}\mathbb{E}[\Delta Y\Delta TZ(Z - h_a(X))] &= \mathbb{E}[\Delta Y\Delta TZ] - \mathbb{E}[\Delta Y\Delta TZh_a(X)]\\ &= \mathbb{E}[\mathbb{E}[\Delta Y\Delta T|X]\mathbb{E}[Z|X]] - \mathbb{E}[\mathbb{E}[\Delta Y\Delta T|X]\mathbb{E}[Z|X]h_a(X)]\\ &= \mathbb{E}[\Delta_{CP}(X)\omega_{CP}(X)(\mathbb{E}[Z|X] - \mathbb{E}[Z|X]h_a(X))], \end{aligned}\tag{21}$$

where the first equality follows from the fact that $Z \in \{0, 1\}$, the second from the law of iterated expectations and Assumption 2.1, and the third from Lemma E.1. In addition,

$$\begin{aligned}\mathbb{E}[\Delta YT(0)(Z - h_a(X))] &= \mathbb{E}[\mathbb{E}[\Delta YT(0)|X]\mathbb{E}[Z - h_a(X)|X]]\\ &= \mathbb{E}[\Delta_{AT}(X)\omega_{AT}(X)(\mathbb{E}[Z|X] - h_a(X))], \end{aligned}\tag{22}$$

where the first equality follows from the law of iterated expectations and Assumption 2.1 and the second from Lemma E.1. Moreover,

$$\mathbb{E}[Y(0)(Z - h_a(X))] = \mathbb{E}[\mathbb{E}[Y(0)|X](\mathbb{E}[Z|X] - h_a(X))] \tag{23}$$

by the law of iterated expectations and Assumption 2.1. Combining (20), (21), (22), and (23) gives

$$\begin{aligned}\mathbb{E}[Y(Z - h_a(X))] = \mathbb{E}[\Delta_{CP}(X)\omega_{CP}(X)(\mathbb{E}[Z|X] - \mathbb{E}[Z|X]h_a(X))]\\ + \mathbb{E}[\Delta_{AT}(X)\omega_{AT}(X)(\mathbb{E}[Z|X] - h_a(X))]\\ + \mathbb{E}[\mathbb{E}[Y(0)|X](\mathbb{E}[Z|X] - h_a(X))]. \end{aligned}\tag{24}$$

Further, Y can also be decomposed as

$$\begin{aligned}Y &= Y(1) - \Delta Y(1 - T) = Y(1) - \Delta Y(1 - T(1) + \Delta T(1 - Z))\\ &= -\Delta Y\Delta T(1 - Z) - \Delta Y(1 - T(1)) + Y(1).\end{aligned}$$

Here,

$$\begin{aligned}\mathbb{E}[-\Delta Y\Delta T(1 - Z)(Z - h_a(X))] &= \mathbb{E}[\Delta Y\Delta Th_a(X)] - \mathbb{E}[\Delta Y\Delta TZh_a(X)]\\ &= \mathbb{E}[\mathbb{E}[\Delta Y\Delta T|X]h_a(X)] - \mathbb{E}[\mathbb{E}[\Delta Y\Delta T|X]\mathbb{E}[Z|X]h_a(X)]\\ &= \mathbb{E}[\Delta_{CP}(X)\omega_{CP}(X)(h_a(X) - \mathbb{E}[Z|X]h_a(X))], \end{aligned}\tag{25}$$

where the first equality follows from the fact that $Z \in \{0,1\}$, the second from the law of iterated expectations and Assumption 2.1, and the third from Lemma E.1. In addition,

$$\begin{aligned}\mathbb{E}[-\Delta Y(1-T(1))(Z-h_a(X))] &= \mathbb{E}[\mathbb{E}[-\Delta Y(1-T(1))|X]\mathbb{E}[Z-h_a(X)|X]] \\ &= \mathbb{E}[0\Delta_{NT}(X)\omega_{NT}(X)(\mathbb{E}[Z|X]-h_a(X))], \end{aligned} \tag{26}$$

where the first equality follows from the law of iterated expectations and Assumption 2.1 and the second from Lemma E.1. Moreover,

$$\mathbb{E}[Y(1)(Z-h_a(X))] = \mathbb{E}[\mathbb{E}[Y(1)|X](\mathbb{E}[Z|X]-h_a(X))] \tag{27}$$

by the law of iterated expectations and Assumption 2.1. Combining (24), (25), (26), and (27) gives

$$\begin{aligned}\mathbb{E}[Y(Z-h_a(X))] = \mathbb{E}[&\Delta_{CP}(X)\omega_{CP}(X)(h_a(X)-\mathbb{E}[Z|X]h_a(X))] \\ &- \mathbb{E}[\Delta_{NT}(X)\omega_{NT}(X)(\mathbb{E}[Z|X]-h_a(X))] \\ &+ \mathbb{E}[\mathbb{E}[Y(1)|X](\mathbb{E}[Z|X]-h_a(X))]. \end{aligned} \tag{28}$$

Writing $Y = sY + (1-s)Y$, using (24) and (28), and substituting the resulting expression into (18) and (19) gives the asserted claim. □

Proof of Corollary 2.1. By the law of iterated expectations and first-order conditions corresponding to the optimization problems (4) and (5), we have

$$\mathbb{E}[X(\mathbb{E}[Z|X]-\Lambda(X^\top\theta_0))] = \mathbb{E}[X(Z-\Lambda(X^\top\theta_0))] = 0 \tag{29}$$

and

$$\mathbb{E}[X(\mathbb{E}[Z|X]-X^\top\gamma_0)] = \mathbb{E}[X(Z-X^\top\gamma_0)] = 0, \tag{30}$$

respectively.

Now, fix $a \in \{\Lambda, 2SLS\}$ and apply Theorem 2.1 with $s = 1$ to obtain

$$\beta_a = \frac{\mathbb{E}[\Delta_{CP}(X)\omega_{CP}(X)(\mathbb{E}[Z|X]-h_a(X)\mathbb{E}[Z|X])]}{\mathbb{E}[T(Z-h_a(X))]} \tag{31}$$

$$+ \frac{\mathbb{E}[\Delta_{AT}(X)\omega_{AT}(X)(\mathbb{E}[Z|X]-h_a(X))]}{\mathbb{E}[T(Z-h_a(X))]} \tag{32}$$

$$+ \frac{\mathbb{E}[\mathbb{E}[Y(0)|X](\mathbb{E}[Z|X]-h_a(X))]}{\mathbb{E}[T(Z-h_a(X))]}. \tag{33}$$

Also, by Lemma E.1,

$$\Delta_{AT}(X)\omega_{AT}(X) = \mathbb{E}[(Y(1)-Y(0))T(0)|X].$$

Hence, the sum of terms in (32) and (33) is equal to

$$\mathbb{E}[\mathbb{E}[(Y(1)-Y(0))T(0)+Y(0)|X](\mathbb{E}[Z|X]-h_a(X))]=\mathbb{E}[\eta_0^\top X(\mathbb{E}[Z|X]-h_a(X))]=0$$

by Assumption 2.3. Therefore, it remains to derive an appropriate expression for the denominator in (31). To do so, note that

$$T=\Delta TZ+T(0), \tag{34}$$

where $\Delta T=T(1)-T(0)$. Also,

$$\begin{aligned}\mathbb{E}[\Delta TZ(Z-h_a(X))]&=\mathbb{E}[\Delta TZ]-\mathbb{E}[\Delta TZh_a(X)]\\&=\mathbb{E}[\mathbb{E}[\Delta T|X]\mathbb{E}[Z|X]]-\mathbb{E}[\mathbb{E}[\Delta T|X]\mathbb{E}[Z|X]h_a(X)]\\&=\mathbb{E}[\omega_{CP}(X)(\mathbb{E}[Z|X]-h_a(X)\mathbb{E}[Z|X])],\end{aligned} \tag{35}$$

where the first equality follows from the fact that $Z\in\{0,1\}$, the second from the law of iterated expectations and Assumption 2.1, and the third from Assumption 2.2. Also,

$$\begin{aligned}\mathbb{E}[T(0)(Z-h_a(X))]&=\mathbb{E}[\mathbb{E}[T(0)|X](\mathbb{E}[Z|X]-h_a(X))]\\&=\mathbb{E}[\psi_0^\top X(\mathbb{E}[Z|X]-h_a(X))]=0,\end{aligned} \tag{36}$$

where the first equality follows from the law of iterated expectations and Assumption 2.1, the second from Assumption 2.4, and the third from (29) and (30). Combining (34), (35), and (36) gives

$$\mathbb{E}[T(Z-h_a(X))]=\mathbb{E}[\omega_{CP}(X)(\mathbb{E}[Z|X]-h_a(X)\mathbb{E}[Z|X])],$$

and completes the proof of the corollary. □

Proof of Corollary 2.2. The proof is the same as that of Corollary 2.1 but instead of the simplification in (36), we use

$$\mathbb{E}[T(0)(Z-h_a(X))]=\mathbb{E}[\mathbb{E}[T(0)|X](\mathbb{E}[Z|X]-h_a(X))]=\mathbb{E}[\omega_{AT}(X)(\mathbb{E}[Z|X]-h_a(X))],$$

where the second equality follows from Assumption 2.2. □

Proof of Corollary 2.3. By Lemma E.1,

$$\Delta_{AT}(X)\omega_{AT}(X)=\mathbb{E}[(Y(1)-Y(0))T(0)|X] \tag{37}$$

and

$$\Delta_{NT}(X)\omega_{NT}(X)=\mathbb{E}[(Y(1)-Y(0))(1-T(1))|X]. \tag{38}$$

Also, for any $s \in [0,1]$,

$$(s-1)(Y(1)-Y(0)) + sY(0) + (1-s)Y(1) = Y(0). \tag{39}$$

Now, let s be the number in $[0,1]$ appearing in Assumption 2.5. Applying Theorem 2.1 with this s and using (37), (38), (39) gives

$$\begin{aligned}\beta_{\Lambda} = &\frac{\mathbb{E}[\Delta_{CP}(X)\omega_{CP}(X)(s\mathbb{E}[Z|X] + (1-s)\Lambda(X^{\top}\theta_0) - \Lambda(X^{\top}\theta_0)\mathbb{E}[Z|X])]}{\mathbb{E}[T(Z-\Lambda(X^{\top}\theta_0))]} \\ &+ \frac{\mathbb{E}[\mathbb{E}[(Y(1)-Y(0))(sT(0)+(1-s)T(1))|X](\mathbb{E}[Z|X]-\Lambda(X^{\top}\theta_0))]}{\mathbb{E}[T(Z-\Lambda(X^{\top}\theta_0))]}.\end{aligned} \tag{40}$$

However,

$$\begin{aligned}&\mathbb{E}[\mathbb{E}[(Y(1)-Y(0))(sT(0)+(1-s)T(1))|X](\mathbb{E}[Z|X]-\Lambda(X^{\top}\theta_0))] \\ &\qquad = \mathbb{E}[\eta_0^{\top}X(\mathbb{E}[Z|X]-\Lambda(X^{\top}\theta_0))] = 0,\end{aligned}$$

where the first equality follows from Assumption 2.5 and the second from (29) in the proof of Corollary 2.1. Therefore, it remains to derive an appropriate expression for the denominator in (40). To do so, denote $\Delta T = T(1) - T(0)$ and observe that

$$\mathbb{E}[\Delta T Z(Z-\Lambda(X^{\top}\theta_0))] = \mathbb{E}[\omega_{CP}(X)(\mathbb{E}[Z|X] - \Lambda(X^{\top}\theta_0)\mathbb{E}[Z|X])] \tag{41}$$

by (35) in the proof of Corollary 2.1. Also,

$$\begin{aligned}\mathbb{E}[-\Delta T(1-Z)(Z-\Lambda(X^{\top}\theta_0))] &= \mathbb{E}[\Delta T\Lambda(X^{\top}\theta_0)] - \mathbb{E}[\Delta T Z\Lambda(X^{\top}\theta_0)] \\ &= \mathbb{E}[\mathbb{E}[\Delta T|X]\Lambda(X^{\top}\theta_0)] - \mathbb{E}[\mathbb{E}[\Delta T|X]\mathbb{E}[Z|X]\Lambda(X^{\top}\theta_0)] \\ &= \mathbb{E}[\omega_{CP}(X)(\Lambda(X^{\top}\theta_0) - \Lambda(X^{\top}\theta_0)\mathbb{E}[Z|X])],\end{aligned} \tag{42}$$

where the first equation follows from the fact that $Z \in \{0,1\}$, the second from the law of iterated expectations and Assumption 2.1, and the third from Assumption 2.2. Further, note that $T = \Delta TZ + T(0)$ and $T = T(1) - \Delta T(1-Z)$, and so $T = s(\Delta TZ + T(0)) + (1-s)(T(1) - \Delta T(1-Z))$. Thus, given that $\mathbb{E}[sT(0) + (1-s)T(1)|X] = X^{\top}\psi_0$ by Assumption 2.5, we have

$$\begin{aligned}\mathbb{E}[T(Z-\Lambda(X^{\top}\theta_0))] &= \mathbb{E}\Big[\Big(s\Delta TZ + (s-1)\Delta T(1-Z)\Big)(Z-\Lambda(X^{\top}\theta_0))\Big] \\ &= \mathbb{E}\Big[\omega_{CP}(X)(s\mathbb{E}[Z|X] + (1-s)\Lambda(X^{\top}\theta_0) - \Lambda(X^{\top}\theta_0))\mathbb{E}[Z|X]\Big],\end{aligned} \tag{43}$$

where the first equality follows from the law of iterated expectations, Assumption 2.1, and (29) in the proof of Corollary 2.1, and the second from (41) and (42). The asserted claim follows. □

Proof of Corollary 2.4. Applying Theorem 2.1 with any $s \in [0,1]$ and using Assumption 2.6 gives

$$\beta_\Lambda = \frac{\mathbb{E}[\Delta_{CP}(X)\omega_{CP}(X)\mathbb{E}[Z|X](1-\mathbb{E}[Z|X])]}{\mathbb{E}[T(Z-\mathbb{E}[Z|X])]}.$$

Also, $T = \Delta TZ + T(0)$, where we denoted $\Delta T = T(1) - T(0)$. Moreover,

$$\mathbb{E}[\Delta TZ(Z-\mathbb{E}[Z|X])] = \mathbb{E}[\omega_{CP}(X)\mathbb{E}[Z|X](1-\mathbb{E}[Z|X])]$$

by (35) in the proof of Corollary 2.1 and Assumption 2.6. In addition,

$$\mathbb{E}[T(0)(Z-\mathbb{E}[Z|X])] = \mathbb{E}[\mathbb{E}[T(0)|X](\mathbb{E}[Z|X]-\mathbb{E}[Z|X])] = 0$$

by the law of iterated expectations and Assumption 2.1. Combining these equalities gives the asserted claim. □

Appendix C. Proofs for Section 3

Proof of Theorem 3.1. Observe that under Assumption A.2, we have $\widehat{\psi} \to_p \overline{\psi}_0$ by Theorem 2.7 in Newey and McFadden (1994). Thus, the function

$$(\theta,\kappa) \mapsto \widehat{Q}(\theta,\kappa) = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i(X_i^\top\theta + \widehat{C}_i\kappa) - \log(1+\exp(X_i^\top\theta + \widehat{C}_i\kappa))\right)$$

converges in probability point-wise for all $(\theta,\kappa) \in \mathbb{R}^p \times \mathbb{R}$ to the function

$$(\theta,\kappa) \mapsto Q_0(\theta,\kappa) = \mathbb{E}[Z(X^\top\theta + \Phi(X^\top\overline{\psi}_0)\kappa) - \log(\exp(X^\top\theta + \Phi(X^\top\overline{\psi}_0)\kappa))].$$

Hence, given that both functions are concave, under Assumption A.2, again by Theorem 2.7 in Newey and McFadden (1994), $(\widehat{\theta},\widehat{\kappa}) \to_p (\overline{\theta}_0,\overline{\kappa}_0)$. Therefore, by the standard arguments,

$$\widehat{\beta}_{A\Lambda} \to_p \frac{\mathbb{E}[Y(Z-\Lambda(X^\top\overline{\theta}_0 + \Phi(X^\top\overline{\psi}_0)\overline{\kappa}_0))]}{\mathbb{E}[T(Z-\Lambda(X^\top\overline{\theta}_0 + \Phi(X^\top\overline{\psi}_0)\overline{\kappa}_0))]}.$$

The rest of the proof coincides with that in Theorem 2.1. □

Proof of Corollary 3.1. By the first-order conditions corresponding to the optimization problem (15), we have

$$\mathbb{E}[X(\mathbb{E}[Z|X] - \Lambda(X^\top\overline{\theta}_0 + \Phi(X^\top\overline{\psi}_0)\overline{\kappa}_0))] = \mathbb{E}[X(Z-\Lambda(X^\top\theta_0 + \Phi(X^\top\overline{\psi}_0)\overline{\kappa}_0))] = 0 \quad (44)$$

and

$$\begin{aligned}&\mathbb{E}[\Phi(X^\top\overline{\psi}_0)(\mathbb{E}[Z|X] - \Lambda(X^\top\overline{\theta}_0 + \Phi(X^\top\overline{\psi}_0)\overline{\kappa}_0))]\\ &\qquad = \mathbb{E}[\Phi(X^\top\overline{\psi}_0)(Z-\Lambda(X^\top\theta_0 + \Phi(X^\top\overline{\psi}_0)\overline{\kappa}_0))] = 0. \quad (45)\end{aligned}$$

The proof therefore follows along the lines in the proof of Corollary 2.1 using the probability limit in Theorem 3.1 instead of the probability limit in Theorem 2.1 and relying on the first-order condition (44) if Assumption 2.4 holds and on (45) if Assumption 3.1 holds. □

Proof of Corollary 3.2. Under Assumption 2.6, it follows from the optimization problem in (15) that $\bar\kappa_0 = 0$ and $\bar\theta_0 = \theta_0$, so that $h_{A\Lambda}(X) = \Lambda(X^\top\theta_0) = \mathbb{E}[Z|X]$ with probability one. Hence, the asserted claim follows from Theorem 3.1. □

Appendix D. Proofs for Section 4

Proof of Theorem 4.1. By the standard asymptotic normality result for the logit estimator, under Assumption A.1,

$$\sqrt{n}(\widehat\theta - \theta_0) = \left(\mathbb{E}[\Lambda'(X^\top\theta_0)XX^\top]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n (Z_i - \Lambda(X_i^\top\theta_0))X_i + o_p(1).$$

Also,

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i - T_i\beta_\Lambda)(Z_i - \Lambda(X_i^\top\widehat\theta)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i - T_i\beta_\Lambda)(Z_i - \Lambda(X_i^\top\theta_0))\\ &\quad - \frac{1}{n}\sum_{i=1}^n (Y_i - T_i\beta_\Lambda)\Lambda'(X_i^\top\theta_0)X_i^\top\sqrt{n}(\widehat\theta - \theta_0) + o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i - T_i\beta_\Lambda - X_i^\top\varphi_0)(Z_i - \Lambda(X_i^\top\theta_0)) + o_p(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \ell_i^\Lambda + o_p(1).\end{aligned}$$

In addition,

$$\frac{1}{n}\sum_{i=1}^n T_i(Z_i - \Lambda(X_i^\top\widehat\theta)) \to_p \mathbb{E}[T(Z - \Lambda(X^\top\theta_0))].$$

Thus,

$$\begin{aligned}\sqrt{n}(\widehat\beta - \beta_\Lambda) &= \frac{n^{-1/2}\sum_{i=1}^n (Y_i - T_i\beta_\Lambda)(Z_i - \Lambda(X_i^\top\widehat\theta))}{n^{-1}\sum_{i=1}^n T_i(Z_i - \Lambda(X_i^\top\theta_0))}\\ &= \frac{n^{-1/2}\sum_{i=1}^n \ell_i^\Lambda}{\mathbb{E}[T(Z - \Lambda(X^\top\theta_0))]} + o_p(1) \to_d N(0, \sigma_\Lambda^2),\end{aligned}$$

yielding the asserted claim. □

Proof of Theorem 4.2. By the standard asymptotic normality result for the logit estimator, under Assumption A.3,

$$\sqrt{n}(\widehat{\psi}-\overline{\psi}_0) = \left(\mathbb{E}[\mathbb{1}\{Z=0\}\Lambda'(X^\top\overline{\psi}_0)XX^\top]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{1}\{Z_i=0\}(T_i-\Lambda(X_i^\top\overline{\psi}_0))X_i+o_p(1).$$

Similarly, denoting $A_0 = \mathbb{E}[\Lambda'(X^\top\overline{\theta}_0 + C\overline{\kappa}_0)WW^\top]$, we have

$$\sqrt{n}\left(\begin{pmatrix}\widehat{\theta}\\ \widehat{\kappa}\end{pmatrix} - \begin{pmatrix}\overline{\theta}_0\\ \overline{\kappa}_0\end{pmatrix}\right) = A_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(Z_i - \Lambda(X_i^\top\overline{\theta}_0 + \widehat{C}_i\overline{\kappa}_0)\right)\begin{pmatrix}X_i\\ \widehat{C}_i\end{pmatrix} + o_p(1).$$

Here,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(Z_i - \Lambda(X_i^\top\overline{\theta}_0 + \widehat{C}_i\overline{\kappa}_0)\right)\begin{pmatrix}X_i\\ \widehat{C}_i\end{pmatrix}\\
&\qquad = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(Z_i - \Lambda(X_i^\top\overline{\theta}_0 + C_i\overline{\kappa}_0)\right)\begin{pmatrix}X_i\\ C_i\end{pmatrix} + A_1\sqrt{n}(\widehat{\psi}-\overline{\psi}_0) + o_p(1).
\end{aligned}$$

Thus,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n(Y_i - T_i\beta_{A\Lambda})(Z_i - \Lambda(X_i^\top\widehat{\theta} + \widehat{C}_i\widehat{\kappa})) = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\ell_{i,1}^{A\Lambda} - \ell_{i,2}^{A\Lambda}) + o_p(1).$$

Also,

$$\frac{1}{n}\sum_{i=1}^n T_i(Z_i - \Lambda(X_i^\top\widehat{\theta} + \widehat{C}_i\widehat{\kappa})) \to_p \mathbb{E}[T(Z - \Lambda(X^\top\overline{\theta}_0 + C\overline{\kappa}_0))].$$

Hence,

$$\begin{aligned}
\sqrt{n}(\widehat{\beta}_{A\Lambda} - \beta_{A\Lambda}) &= \frac{n^{-1/2}\sum_{i=1}^n(Y_i - T_i\beta_{A\Lambda})(Z_i - \Lambda(X_i^\top\widehat{\theta} + \widehat{C}_i\widehat{\kappa}))}{n^{-1}\sum_{i=1}^n T_i(Z_i - \Lambda(X_i^\top\widehat{\theta} + \widehat{C}_i\widehat{\kappa}))}\\
&= \frac{n^{-1/2}\sum_{i=1}^n(\ell_{i,1}^{A\Lambda} - \ell_{i,2}^{A\Lambda})}{\mathbb{E}[T(Z - \Lambda(X^\top\overline{\theta}_0 + C\overline{\kappa}_0))]} + o_p(1) \to_d N(0, \sigma^2_{A\Lambda}),
\end{aligned}$$

yielding the asserted claim. □

Appendix E. Auxiliary Lemma

Lemma E.1. *Suppose that Assumption 2.2 is satisfied and that* $\mathbb{E}[|Y|] < \infty$. *Then*

$$\mathbb{E}[(Y(1) - Y(0))(T(1) - T(0))|X] = \Delta_{CP}(X)\omega_{CP}(X),$$
$$\mathbb{E}[(Y(1) - Y(0))T(0)|X] = \Delta_{AT}(X)\omega_{AT}(X),$$

$$\mathbb{E}[(Y(1) - Y(0))(1 - T(1))|X] = \Delta_{NT}(X)\omega_{NT}(X).$$

Proof. Let $\Delta Y = Y(1) - Y(0)$ and $\Delta T = T(1) - T(0)$. Then by the law of iterated expectations and Assumption 2.2,

$$\mathbb{E}[\Delta Y \Delta T|X] = \mathbb{E}[\Delta Y|X, \Delta T = 1]\mathbb{P}(\Delta T = 1|X) = \Delta_{CP}(X)\omega_{CP}(X),$$

$$\mathbb{E}[\Delta Y T(0)|X] = \mathbb{E}[\Delta Y|X, T(0) = 1]\mathbb{P}(T(0) = 1|X) = \Delta_{AT}(X)\omega_{AT}(X),$$

$$\mathbb{E}[\Delta Y(1 - T(1))|X] = \mathbb{E}[\Delta Y|X, T(1) = 0]\mathbb{P}(T(1) = 0|X) = \Delta_{NT}(X)\omega_{NT}(X).$$

The asserted claims follow. □

References

ABADIE, A. (2003): "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 113, 231–263.

ANGRIST, J., AND G. IMBENS (1995): "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of American Statistical Association*, 90, 431–442.

ANGRIST, J., AND J.-S. PISCHKE (2009): *Mostly harmless econometric: an empiricist's companion*: Princeton University Press.

BELLONI, A., V. CHERNOZHUKOV, I. FERNANDEZ-VAL, AND C. HANSEN (2017): "Program evaluation and causal inference with high-dimensional data," *Econometrica*, 85, 233–298.

BIERENS, H. (1982): "Consistent model specification tests," *Journal of Econometrics*, 20, 105–134.

BLANDHOL, C., J. BONNEY, M. MOGSTAD, AND A. TORGOVITSKY (2022): "When is TSLS actually LATE?" *working paper*.

FROLICH, M. (2007): "Nonparametric IV estimation of local average treatment effects with covariates," *Journal of Econometrics*, 139, 35–75.

HARDLE, W., AND E. MAMMEN (1993): "Comparing nonparametric versus parametric regression fits," *Annals of Statistics*, 21, 1926–1947.

HAUSMAN, J. (1978): "Specification tests in econometrics," *Econometrica*, 46, 1251–1271.

HOROWITZ, J. (2009): *Semiparametric and nonparametric methods in econometrics*: Spring Series in Statistics.

HOROWITZ, J., AND V. SPOKOINY (2001): "An adaptive rate-optimal test of a parametric mean-regression model against a nonparametric alternative," *Econometrica*, 69, 599–631.

IMBENS, G., AND J. ANGRIST (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

KOLESAR, M. (2013): "Estimation in an instrumental variables model with treatment effect heterogeneity," *working paper*.

NEWEY, W., AND D. MCFADDEN (1994): "Large sample estimation and hypothesis testing," *Handbook of Econometrics, Volume IV*, 2111–2245.

SLOCZYNSKI, T. (2020): "When should we (not) interpret linear IV estimands as LATE?" *working paper*.

SORENSEN, J. (2022): "Testing a class of semi- or nonparametric conditional moment restriction models using series methods," *Econometric Theory*, 0, 1–32.

(D. Chetverikov) DEPARTMENT OF ECONOMICS, UCLA, BUNCHE HALL, 8283, 315 PORTOLA PLAZA, LOS ANGELES, CA 90095, USA.

Email address: chetverikov@econ.ucla.edu

(J. Hahn) DEPARTMENT OF ECONOMICS, UCLA, BUNCHE HALL, 8283, 315 PORTOLA PLAZA, LOS ANGELES, CA 90095, USA.

Email address: hahn@econ.ucla.edu

(Z. Liao) DEPARTMENT OF ECONOMICS, UCLA, BUNCHE HALL, 8283, 315 PORTOLA PLAZA, LOS ANGELES, CA 90095, USA.

Email address: zhipeng.liao@econ.ucla.edu

(S. Sheng) DEPARTMENT OF ECONOMICS, UCLA, BUNCHE HALL, 8283, 315 PORTOLA PLAZA, LOS ANGELES, CA 90095, USA.

Email address: ssheng@econ.ucla.edu