Never Stand Still   Business School   Economics

A Consistent Variance Estimator for 2SLS When Instruments Identify Different LATEs

Seojeong Lee

A Consistent Variance Estimator for 2SLS
When Instruments Identify Different LATEs

Seojeong Lee[*][†]

January 14, 2015

Abstract

Under treatment effect heterogeneity, an instrument identifies the instrument-specific local average treatment effect (LATE). If a regression model is estimated by the two-stage least squares (2SLS) using multiple instruments, then 2SLS is consistent for a weighted average of different LATEs. In practice, a rejection of the overidentifying restrictions test can indicate that there are more than one LATE. What is often overlooked in the literature is that the postulated moment condition evaluated at the 2SLS estimand does not hold unless those LATEs are the same. If so, the conventional heteroskedasticity-robust variance estimator would be inconsistent. However, 2SLS standard errors based on the conventional variance estimator have been reported even when the overidentifying restrictions test is rejected. I propose a consistent estimator for the asymptotic variance of 2SLS by using the result of Hall and Inoue (2003) on misspecified moment condition models. This can be used to correctly calculate the standard errors regardless of whether there are more than one LATE or not.

Keywords: local average treatment effect, treatment heterogeneity, two-stage least squares, variance estimator, model misspecification

JEL Classification: C13, C31, C36

1 Introduction

Since the series of seminal papers by Imbens and Angrist (1994), Angrist and Imbens (1995), and Angrist, Imbens, and Rubin (1996), the local average treatment effect (LATE) has played an important role in providing useful guidance to many policy questions. The key underlying assumption is treatment effect heterogeneity. Each individual has a different causal effect of treatment on outcome. Assume a binary treatment, D_i, and an outcome variable Y_i. Let Y_{di} denote the potential outcome of individual i given treatment status $D_i = 1$ or 0. Y_{1i} and Y_{0i} denote the response with and without the treatment, respectively. The individual treatment effect is $Y_{1i} - Y_{0i}$ which is assumed to be heterogeneous, but we never observe both values at the same time. Therefore, researchers focus on the average treatment effect (ATE), $E[Y_{1i} - Y_{0i}]$. However, unless the treatment status is randomly assigned, a naive estimate of ATE is likely to be biased because of selection into treatment.

Instrumental variables can be used to overcome this endogeneity problem. If an instrument Z_i which is independent of Y_{1i} and Y_{0i}, and correlated with the treatment D_i is available, then ATE of those whose treatment status can be changed by the instrument, thus the *local* ATE, can be identified. Assume Z_i is binary and define D_{1i} and D_{0i} be i's treatment status when $Z_i = 1$ and $Z_i = 0$, respectively. The LATE theorem of Imbens and Angrist (1994) shows that

$$\frac{Cov(Y_i, Z_i)}{Cov(D_i, Z_i)} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]. \qquad (1.1)$$

That is, the instrumental vaiables (IV) estimand (or the Wald estimand) is equal to the ATE for a subpopulation such that $D_{1i} > D_{0i}$, which is called compliers. Those who take the treatment regardless of the instrument status, $D_{1i} = D_{0i} = 1$, are always-takers, and those who do not take the treatment anyway, $D_{1i} = D_{0i} = 0$, are never-takers. We cannot identify ATE for always-takers and never-takers in general. By the monotonicity assumption of Imbens and Angrist (1994), we exclude defiers who behave in the opposite way with compliers, $D_{1i} = 0$ and $D_{0i} = 1$. Since the compliers are specific to the instrument Z_i, LATE is instrument-specific.[1]

The above setting can be generalized to multiple instruments. The two-stage

---

[1] Abadie (2002) shows that the marginal distributions of potential outcomes can be identified for compliers. These marginal distributions are also instrument-specific.

least squares (2SLS) estimator is commonly used to estimate the causal effect in such cases. Without loss of generality, consider mutually exclusive binary instruments, Z_i^j for $j = 1, ..., q$. Let D_{zi}^j be i's potential treatment status when $Z_i^j = z$ where $z = 0, 1$, and $j = 1, ..., q$. Each instrument identifies a version of LATE because compliers may differ for each Z_i^j. It is well known that the 2SLS estimator using multiple instruments is consistent for a weighted average of different LATEs. In other words, the 2SLS estimand is a weighted average of treatment effects for instrument-specific compliers (Angrist and Pischke, 2009; Kolesár, 2013):

$$\rho_a = \sum_{j=1}^{q} \xi_j \cdot E[Y_{1i} - Y_{0i} | D_{1i}^j > D_{0i}^j],$$

where ξ_j is a nonnegative number and $\sum_j \xi_j = 1$. For example, Angrist and Evans (1998) use twin births and same-sex sibships as instruments to estimate the effect of family size on mother's labor supply. The twins instrument identifies LATE of those who had more children than they otherwise would have had because of twinning, while the same-sex instrument identifies LATE of those whose fertility was affected by their children's sex mix. These two compliers need not be the same. Their result shows that the 2SLS estimate using both instruments is a weighted average of two IV estimates using one instrument at a time.

If the 2SLS estimand is a weighted average of more than one LATE, then the commonly conducted overidentifying restrictions test (the J test, hereinafter) would be rejected. What is less well known and often overlooked in the literature is that a rejection of the J test implies that the postulated moment condition is likely to be misspecified. If so, the conventional standard errors are no longer consistent regardless of how small the differences among LATEs are.[2] This fact has been neglected and the standard errors have been routinely calculated even with small p values of the J test, e.g. see Angrist and Krueger (1991), Angrist and Evans (1998) and Angrist, Lavy, and Schlosser (2010), among others.

In this paper, I propose a consistent estimator for the asymptotic variance of 2SLS robust to multiple LATEs. The standard errors based on the proposed vari-

---

[2]The J test can also be rejected due to invalid instruments. Kitagawa (2014) proposed a specification test for instrument validity in this framework. The 2SLS estimand would change, and the conventional standard errors would also be inconsistent if the instruments are invalid. I assume that the validity is justified either by a statistical test such as Kitagawa (2014) or by an economic reasoning.

Sample size = 486,926		Columns			
Independent variable		(2)	(4)	(6)	(8)
Years of education		0.0553	0.0948	0.0393	0.0779
	C	(0.0138)	(0.0221)	(0.0146)	(0.0238)
	MR	**(0.0166)**	**(0.0277)**	**(0.0175)**	**(0.0308)**
Race (1 = black)		—	—	-0.2266	-0.1787
	C			(0.0185)	(0.0298)
	MR			**(0.0221)**	**(0.0385)**
SMSA (1 = center city)		—	—	-0.1535	-0.1182
	C			(0.0136)	(0.0219)
	MR			**(0.0163)**	**(0.0283)**
Married (1 = married)		—	—	0.2442	0.2450
	C			(0.0025)	(0.0025)
	MR			**(0.0025)**	**(0.0026)**
9 Year-of-birth dummies		Yes	Yes	Yes	Yes
8 Region-of-residence dummies		No	No	Yes	Yes
Age		—	0.1326	—	0.1215
	C		(0.0487)		(0.0475)
	MR		**(0.0530)**		**(0.0527)**
Age-squared		—	-0.0016	—	-0.0015
	C		(0.0007)		(0.0007)
	MR		**(0.0008)**		**(0.0008)**
J test statistic (dof)		101.2 (29)	49.7 (27)	93.2 (29)	51.2 (27)
p-value		[0.0000]	[0.0049]	[0.0000]	[0.0033]
Number of instruments		40	40	51	51

Table 1: Comparison of the proposed multiple-LATEs-robust (MR) and the conventional (C) standard errors—Replication of Table VI in Angrist and Krueger (1991)

ance estimator can be substantially different from the one based on the conventional heteroskedasticity-robust standard errors even for a large sample size. Table 1 shows a replication result of Table VI in Angrist and Krueger (1991). The authors use quarter-of-birth as instruments to find a relationship between educational attainment and compulsory school attendance law. Since they used a full set of quarter-of-birth times year-of-birth interactions as instruments, the models are highly overidentified. The p-values of the J test are below any reasonable significance level. If we rule out the possibility of invalid instruments, then the rejection of the J test implies that there are more than one LATE and the conventional standard errors are no longer correct. The numbers in parentheses are standard errors, and those in bold are ones based on the proposed multiple-LATEs-robust variance estimator.

Using 2SLS with multiple instruments has been very popular. Stephens and Yang (2014) reexamine the relationship between compulsory education and returns to schooling. Siminski and Ville (2011) use Australian data to see long-run mortality effects of Vietnam-era army service. In health economics, 2SLS is used in Evans and Lien (2005), Evans and Garthwaite (2012), and Doyle Jr (2008). From a theoretical point of view, two recent papers cover similar topics with this paper. Kolesár (2013) shows that under treatment effect heterogeneity the 2SLS estimand is a convex combination of LATEs while the limited information maximum likelihood (LIML) estimand may not. Angrist and Fernandez-Val (2013) propose an estimand for new subpopulations by reweighting covariate-specific LATEs. However, neither of the two papers considers correct variance estimation.

In the next section, I show by example that the postulated moment condition is misspecified when there are more than one LATE. The asymptotic distribution of 2SLS estimators in such a case is derived, and a consistent variance estimator is proposed. Section 3 presents simulation results that show p-values of the J test are negatively correlated with the difference between the proposed multiple-LATEs-robust and the conventional standard errors. Section 4 concludes. The proofs of propositions are collected in Appendix.

2   Moment condition for 2SLS

Commonly used IV and 2SLS estimators can be derived from the corresponding moment conditions. Consider a linear model

$$Y_i = \alpha_0 + \rho_0 D_i + \varepsilon_i \equiv \mathbf{X}_i' \boldsymbol{\beta_0} + \varepsilon_i, \tag{2.1}$$

where $\mathbf{X}_i = (1, D_i)'$ and $\boldsymbol{\beta_0} = (\alpha_0, \rho_0)'$. Since D_i is endogeneous, $\boldsymbol{\beta_0}$ cannot be consistently estimated by OLS. If an instrument vector $\mathbf{Z}_i = (1, Z_i^1, Z_i^2, ..., Z_i^q)'$ such that $E[\mathbf{Z}_i \varepsilon_i] = 0$ exists, then $\boldsymbol{\beta_0}$ can be consistently estimated by the 2SLS estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}, \tag{2.2}$$

where $\mathbf{X} \equiv (\mathbf{X}_1', \cdots, \mathbf{X}_n')'$ is an $n \times 2$ matrix, $\mathbf{Z} \equiv (\mathbf{Z}_1', \cdots, \mathbf{Z}_n')'$ is an $n \times (q+1)$ matrix[3], and $\mathbf{Y} \equiv (Y_1, ..., Y_n)'$ is an $n \times 1$ vector. The 2SLS estimator is a special

---

[3]Note that a constant is included in the instrument vector.

case of a GMM estimator using $(n^{-1}\mathbf{Z}'\mathbf{Z})^{-1}$ as a weight matrix based on the moment condition

$$0 = E[\mathbf{Z}_i \varepsilon_i] = E[\mathbf{Z}_i(Y_i - \alpha_0 - \rho_0 D_i)]. \tag{2.3}$$

When (2.3) holds at a unique parameter vector $\boldsymbol{\beta_0} = (\alpha_0, \rho_0)$, the moment condition is correctly specified. The model is overidentified if the dimension of the moment condition is greater than that of the parameter vector and just-identified when they are equal. For example, if Z_i^1 is the only instrument available, the model is just-identified and the solution is given by

$$
\begin{aligned}
\alpha_0 &= E[Y_i] - \rho_0 \cdot E[D_i], \tag{2.4} \\
\rho_0 &= \frac{Cov(Y_i, Z_i^1)}{Cov(D_i, Z_i^1)} = E[Y_{1i} - Y_{0i} | D_{1i}^1 > D_{0i}^1].
\end{aligned}
$$

Thus, LATE with respect to Z_i^1 is identified, and can be consistently estimated by the IV estimator.

However, when multiple instruments are used so that the model is overidentified, the postulated moment condition becomes problematic because it restricts all the instrument-specific LATEs to be identical by construction. If those LATEs are different, then there may be no parameter that satisfies (2.3) simultaneously even if the instruments are valid. To see this, suppose that there are two instruments, Z_i^1 and Z_i^2. The moment condition is

$$0 = E[Y_i - \alpha_0 - \rho_0 D_i] = E[Z_i^1(Y_i - \alpha_0 - \rho_0 D_i)] = E[Z_i^2(Y_i - \alpha_0 - \rho_0 D_i)]. \tag{2.5}$$

Solving the first equation for α_0, substituting it into the second and third equations, and solving them for ρ_0, we have

$$\rho_0 = \frac{Cov(Y_i, Z_i^1)}{Cov(D_i, Z_i^1)} = \frac{Cov(Y_i, Z_i^2)}{Cov(D_i, Z_i^2)}. \tag{2.6}$$

But this implies that the two LATEs are the same, which is not true in general. Thus, (2.6) does not hold and the moment condition does not have a solution that satisfies the three equations simultaneously.

In general, if the moment condition does not hold for all possible values of parameter in the parameter space, the model is *misspecified*. In our case, the model is misspecified due to heterogeneous treatment effects across different complier groups

although all the instruments are valid. The J test is commonly used in practice to test whether the moment condition is correctly specified or not, which implies a homogeneous treatment effect in this framework. It is not surprising that researchers often face a significant J test statistic when multiple instruments are used because even a small difference in LATEs across complier groups will result in a rejection of the J test asymptotically.

Although it is well known that 2SLS with multiple instruments estimates a weighted average of different LATEs and a rejection of the J test can be merely due to heterogeneity of treatment effects, the consequence of misspecification on the asymptotic variance of 2SLS has been overlooked in the literature. This is surprising because the conventional heteroskedasticity-robust variance estimator would be inconsistent for the true asymptotic variance of 2SLS if the underlying moment condition is misspecified.

The above argument can be generalized to models with covariates, and situations where instruments or a treatment variable can take multiple values, or even be continuous. In such cases, the 2SLS estimand changes. Formulas and interpretations are given in Theorems 2 and 3 of Angrist and Imbens (1995), and Theorem 1 of Kolesár (2013). Despite the changes in interpretations of the 2SLS estimand, the fact that the postulated moment condition is misspecified does not change. The following proposition shows that the asymptotic distribution of 2SLS estimators when there are more than one LATE in a general setting.

Proposition 1. *Let $(Y_i, \mathbf{X}_i, \mathbf{Z}_i)_{i=1}^n$ be an iid sample, where $\mathbf{X}_i = (\mathbf{W}_i', D_i)'$, $\mathbf{Z}_i = (\mathbf{W}_i', Z_i^1, \cdots, Z_i^q)'$, and $\mathbf{W}_i$ be a vector of covariates including a constant. Suppose that the model is given by $Y_i = \mathbf{W}_i'\boldsymbol{\gamma} + D_i\rho + \varepsilon_i \equiv \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i$ where $\boldsymbol{\beta} \equiv (\boldsymbol{\gamma}', \rho)'$, and the 2SLS estimator (2.2) is used for estimation. Let $\boldsymbol{\beta_a} = (\boldsymbol{\gamma_a}', \rho_a)'$ be the 2SLS estimand where $\boldsymbol{\gamma_a}$ satisfies $E[Y_i] = E[\mathbf{W}_i]' \cdot \boldsymbol{\gamma_a} + E[D_i] \cdot \rho_a$ and ρ_a is a linear combination of different LATEs. Let $e_i \equiv Y_i - \mathbf{X}_i'\boldsymbol{\beta_a}$. The asymptotic distribution of 2SLS is*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_a}) \xrightarrow{d} N(\mathbf{0}, H^{-1}\Omega H^{-1}),$$

where $H = E[\mathbf{X}_i\mathbf{Z}_i'] \left(E[\mathbf{Z}_i\mathbf{Z}_i']\right)^{-1} E[\mathbf{Z}_i\mathbf{X}_i']$, $\Omega = E[\boldsymbol{\psi}_i\boldsymbol{\psi}_i']$, and

$$
\begin{aligned}
\boldsymbol{\psi_i} &= E[\mathbf{X}_i\mathbf{Z}_i'](E[\mathbf{Z}_i\mathbf{Z}_i'])^{-1} \left(\mathbf{Z}_i e_i - E[\mathbf{Z}_i e_i]\right) + (\mathbf{X}_i\mathbf{Z}_i' - E[\mathbf{X}_i\mathbf{Z}_i'])(E[\mathbf{Z}_i\mathbf{Z}_i'])^{-1}E[\mathbf{Z}_i e_i] \\
&\quad + E[\mathbf{X}_i\mathbf{Z}_i'](E[\mathbf{Z}_i\mathbf{Z}_i'])^{-1} \left(E[\mathbf{Z}_i\mathbf{Z}_i'] - \mathbf{Z}_i\mathbf{Z}_i'\right) (E[\mathbf{Z}_i\mathbf{Z}_i'])^{-1} E[\mathbf{Z}_i e_i].
\end{aligned}
$$

7

The next proposition proposes a consistent estimator for the asymptotic variance matrix of 2SLS robust to multiple-LATEs.

Proposition 2. *A multiple-LATEs-robust asymptotic variance estimator for 2SLS is given by*

$$\hat{\boldsymbol{\Sigma}}_{MR} = n \cdot \left(\mathbf{X}'\mathbf{Z} \left(\mathbf{Z}'\mathbf{Z} \right)^{-1} \mathbf{Z}'\mathbf{X} \right)^{-1} \left(\sum_i \hat{\boldsymbol{\psi}}_i \hat{\boldsymbol{\psi}}_i' \right) \left(\mathbf{X}'\mathbf{Z} \left(\mathbf{Z}'\mathbf{Z} \right)^{-1} \mathbf{Z}'\mathbf{X} \right)^{-1} \qquad (2.7)$$

where

$$
\begin{aligned}
\hat{\boldsymbol{\psi}}_i &= \frac{1}{n}\mathbf{X}'\mathbf{Z} \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} \right)^{-1} \left(\mathbf{Z}_i \hat{e}_i - \frac{1}{n}\mathbf{Z}'\hat{e} \right) \\
&+ \left(\mathbf{X}_i \mathbf{Z}_i' - \frac{1}{n}\mathbf{X}'\mathbf{Z} \right) \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} \right)^{-1} \frac{1}{n}\mathbf{Z}'\hat{e} \\
&+ \frac{1}{n}\mathbf{X}'\mathbf{Z} \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} \right)^{-1} \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} - \mathbf{Z}_i \mathbf{Z}_i' \right) \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} \right)^{-1} \frac{1}{n}\mathbf{Z}'\hat{e},
\end{aligned}
\qquad (2.8)
$$

$\hat{e}_i = Y_i - \mathbf{X}_i'\hat{\boldsymbol{\beta}}$, *and* $\hat{e} = (\hat{e}_1, \hat{e}_2, ..., \hat{e}_n)'$.

The formula of $\hat{\boldsymbol{\Sigma}}_{MR}$ is different from that of the conventional heteroskedasticity-robust variance estimator:

$$\hat{\boldsymbol{\Sigma}}_{C} = n \cdot \left(\mathbf{X}'\mathbf{Z} \left(\mathbf{Z}'\mathbf{Z} \right)^{-1} \mathbf{Z}'\mathbf{X} \right)^{-1} \left(\sum_i \mathbf{Z}_i \mathbf{Z}_i' \hat{e}_i^2 \right) \left(\mathbf{X}'\mathbf{Z} \left(\mathbf{Z}'\mathbf{Z} \right)^{-1} \mathbf{Z}'\mathbf{X} \right)^{-1}. \qquad (2.9)$$

Under treatment effect homogeneity, both $\hat{\boldsymbol{\Sigma}}_{MR}$ and $\hat{\boldsymbol{\Sigma}}_{C}$ converge in probability to the same limit, but they are generally different in finite sample. $\hat{\boldsymbol{\Sigma}}_{MR}$ is consistent for the true asymptotic variance matrix regardless of whether the postulated moment condition is misspecified or not, and thus can be used when there is one or more than one LATE. In contrast, $\hat{\boldsymbol{\Sigma}}_{C}$ is consistent only if the underlying LATEs are the same. This is also true for the standard errors based on $\hat{\boldsymbol{\Sigma}}_{MR}$ and $\hat{\boldsymbol{\Sigma}}_{C}$.

Remark 1 (Relation to existing studies) When there is a single endogenous variable without covariates, Proposition 1 coincides with the result in the proof of Theorem 3 of Imbens and Angrist (1994) when the first stage is known but needs to be estimated.[4] Intuituvely, this is because Imbens and Angrist do not use the assump-

---

[4]There are typos in the proof of Theorem 3 of Imbens and Angrist (1994). Their matrix Δ shoud

tion of $E[\mathbf{Z}_i\varepsilon_i] = 0$ in deriving their asymptotic distribution. Even in such cases, however, econometric softwares do not estimate their asymptotic variance, resulting in wrong standard errors. Angrist and Imbens (1995) and Kolesár (2013) show that the 2SLS estimand is a weighted average of LATEs and provide interpretations, but neither of them derives the asymptotic distribution of 2SLS. Thus, Proposition 1 complements their results. Hall and Inoue (2003) present the asymptotic distribution of GMM estimators under misspecification. Specifically, Proposition 1 is a special case of their Theorem 2 in the context of treatment effect heterogeneity. Thus, there is little surprise in terms of a proof technique. The marginal contribution of this paper is to show that 2SLS using multiple instruments under treatment effect heterogeneity is a special case of misspecified GMM so that the analysis of its asymptotic behavior can be significantly simplified.

Remark 2 (Invalid Instruments) The multiple-LATEs-robust variance estimator $\hat{\mathbf{\Sigma}}_{MR}$ is also robust to invalid instruments, i.e., instruments correlated with the error term. Consider a linear model $Y_i = \mathbf{X}_i'\boldsymbol{\beta_0} + \varepsilon_i$ where $\mathbf{X}_i$ is a $(k+p) \times 1$ vector of regressors. Among $k+p$ regressors, p are endogenous, i.e. $E[\mathbf{X}_i\varepsilon_i] \neq 0$. If a $k+q$ vector of instruments $\mathbf{Z}_i$ is available such that $E[\mathbf{Z}_i\varepsilon_i] = 0$ and $q \geq p$, then $\boldsymbol{\beta_0}$ can be consistently estimated by 2SLS or GMM. If any of the instruments is invalid, then $E[\mathbf{Z}_i\varepsilon_i] \neq 0$ and $\boldsymbol{\beta_0}$ may not be consistently estimated. Instead, a pseudo-true value that minimizes the corresponding GMM criterion is estimated.[5] Since the moment condition does not hold, the model is misspecified. There are two types of misspecification: (i) fixed or global misspecification such that $E[\mathbf{Z}_i\varepsilon_i] = \boldsymbol{\delta}$ where $\boldsymbol{\delta}$ is a constant vector containing at least one non-zero component, and (ii) local misspecification such that $E[\mathbf{Z}_i\varepsilon_i] = n^{-r}\boldsymbol{\delta}$ for some $r > 0$. A particular choice of $r = 1/2$ has beeen used to analyse the asymptotic behavior of 2SLS estimators with invalid instruments by Hahn and Hausman (2005), Bravo (2010), Berkowitz, Caner, and Fang (2008, 2012), Otsu (2011), Guggenberger (2012), and DiTraglia (2013). Under either fixed or local misspecification, $\hat{\mathbf{\Sigma}}_{MR}$ in Proposition 2 is consistent for the true asymptotic variance. However, the conventional variance estimator $\hat{\mathbf{\Sigma}}_C$ is inconsistent under fixed misspecification. Under local misspecification, $\hat{\mathbf{\Sigma}}_C$ is consistent but the rate of convergence is negatively affected.

read $\Delta = \begin{pmatrix} E[\psi(Z,D,\theta)\cdot\psi(Z,D,\theta)'] & E[\varepsilon\cdot\psi(Z,D,\theta)] & E[g(Z)\cdot\varepsilon\cdot\psi(Z,D,\theta)] \\ E[\varepsilon\cdot\psi(Z,D,\theta)]' & E[\varepsilon^2] & E[g(Z)\cdot\varepsilon^2] \\ E[g(Z)\cdot\varepsilon\cdot\psi(Z,D,\theta)]' & E[g(Z)\cdot\varepsilon^2] & E[g^2(Z)\cdot\varepsilon^2] \end{pmatrix}.$

[5]The 2SLS estimand $\boldsymbol{\beta_a}$ in Proposition 1 is an example of such pseudo-true values.

Remark 3 (Bootstrap) Bootstrapping can be used to get more accurate t tests and confidence intervals (CI's) based on $\hat{\boldsymbol{\beta}}$, in terms of having smaller errors in the rejection probabilities or coverage probabilities. This is called asymptotic refinements of the bootstrap. Since the model is overidentified and possibly misspecified, and 2SLS is a special case of GMM, the misspecification-robust bootstrap for GMM of Lee (2014) can be used. In contrast, the conventional bootstrap methods for overidentified GMM of Hall and Horowitz (1996), Brown and Newey (2002), and Andrews (2002) assume correctly specified moment conditions, in this case, treatment effect homogeneity. Therefore, they achieve neither asymptotic refinements nor consistency. Suppose one wants to test $H_0 : \beta_m = \beta_{a,m}$ or to construct a CI for $\beta_{a,m}$ where $\beta_{a,m}$ is the mth element of $\boldsymbol{\beta_a}$. The misspecification-robust bootstrap critical values for t tests and CI's are calculated from the simulated distribution of the bootstrap t statistic

$$T_n^* = \frac{\hat{\beta}_m^* - \hat{\beta}_m}{\sqrt{\hat{\Sigma}_{MR,m}^*/n}}$$

where $\hat{\beta}_m^*$ and $\hat{\beta}_m$ are the mth elements of $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{\beta}}$, respectively, $\hat{\Sigma}_{MR,m}^*$ is the mth diagonal element of $\hat{\boldsymbol{\Sigma}}_{MR}^*$, and $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{\Sigma}}_{MR}^*$ are the bootstrap versions of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Sigma}}_{MR}$ based on the same formula using the bootstrap sample rather than the original sample.

3 Simulation

A random coefficient model with two mutually exclusive instruments is considered for simulation. First, a single binary instrument Z_i^0 is randomly generated. Next, $\{u_i\}_{i=1}^n$ are generated from the uniform distribution $U[0,1]$ and individual compliance types with respect to Z_i^0 are assigned: An individual i is a never-taker if $0 \leq u_i < 0.2$, a complier if $0.2 \leq u_i < 0.8$, and an always-taker if $0.8 \leq u_i \leq 1$. Treatment status is determined accordingly. Then, Z_i^0 is decomposed into two mutually exclusive instruments randomly, Z_i^1 and Z_i^2, such that $Z_i^1 + Z_i^2 = Z_i^0$. Thus, in this setting, always-takers and never-takers are common to both instruments, but compliers for Z_i^0 is decomposed into three subgroups: Common compliers for both Z_i^1 and Z_i^2, compliers for Z_i^1 only, and compliers for Z_i^2 only.

n	100	1,000	10,000
mean($\hat{\rho}$)	1.8926	1.9213	1.9235
s.d.($\hat{\rho}$)	1.1770	0.3618	0.1140
mean($s.e._{\hat{\boldsymbol{\Sigma}}_{MR}}$)	1.1699	0.3607	0.1137
mean($s.e._{\hat{\boldsymbol{\Sigma}}_{C}}$)	1.1188	0.3485	0.1100
reject J test at 5%	32.6%	99.8%	100%

Table 2: The standard deviation and standard errors of $\hat{\rho}$ when the 2SLS estimand is a weighted average of different LATEs

The potential outomes with and without treatment are generated as

$$Y_{0i} \sim N(0, 3^2), \quad Y_{1i} = Y_{0i} + \rho_i. \tag{3.1}$$

Without loss of generality, ρ_i is assumed to be identical within each subgroup. Let $\rho_i = 4$ if i is an always-taker, $\rho_i = 1$ if i is a never-taker, $\rho_i = 2$ if i is a common complier, $\rho_i = 3$ if i is a complier for Z_i^1 only, and $\rho_i = 1$ if i is a complier for Z_i^2 only. LATEs for Z_i^1 and Z_i^2 are weighted averages of 2 and 3, and 2 and 1, respectively. Note that the values of ρ_i for always-takers and never-takers do not matter in calculating LATE. The 2SLS estimand ρ_a is a weighted average of the two LATEs. Thus, the postulated moment condition is misspecified.

Table 2 shows the mean and the standard deviation of the 2SLS estimator, the mean of multiple-LATEs-robust/conventional standard errors, and the rejection probability of the J test at 5% for different sample sizes. The number of Monte Carlo repetitions is 10,000. The result shows that the conventional standard error based on $\hat{\boldsymbol{\Sigma}}_C$ underestimates the standard deviation for any sample size n. In contrast, the proposed multiple-LATEs-robust standard error estimates the standard deviation more accurately. The mean of $\hat{\rho}$ is slightly less than two, because the number of compliers for Z_i^2 is larger than that for Z_i^1 in the simulation.

Figure 1 shows a negative relationship between the p-values of the J test and the percentage difference between the two standard errors $s.e._{\hat{\boldsymbol{\Sigma}}_{MR}}$ and $s.e._{\hat{\boldsymbol{\Sigma}}_{C}}$. When $n = 100$, it is quite possible that the J test does not reject the false null hypothesis at a usual significance level. However, as the p-value gets smaller, it becomes more likely that the two s.e.'s are different. Since only $\hat{\boldsymbol{\Sigma}}_{MR}$ is consistent for the true asymptotic variance when there are multiple LATEs, it is recommended to report

11

$s.e._{\cdot\hat{\boldsymbol{\Sigma}}_{MR}}$ especially when the p-value is small. Since the J test is consistent, the p-values become more concentrated around 0 as n increases. Around zero p-values, the difference of the two s.e.'s can be substantial.

4 Conclusion

Estimating a weighted average of LATEs with multiple instruments using 2SLS is a common practice for applied researchers. The resulting inferences and confidence intervals are often justified when estimated LATEs are similar and the overidentifying restrictions test does not reject the assumed model. However, when researchers face a rejection of the overidentifying restrictions test, there has been no guidance on how to proceed. Routinely reported standard errors generated by econometric softwares are likely to be incorrect because they do not take into account possible misspecification of the postulated moment condition model. This paper provided a solution to such dilemmas. The proposed variance estimator for 2SLS is consistent for the true asymptotic variance regardless of whether there are multiple LATEs or not. In addition, this estimator is robust to invalid instruments, and can be used for bootstrapping to achieve asymptotic refinements.

A Appendix: Proofs of Propositions

Proposition 1

Proof. Let $\boldsymbol{e} \equiv (e_1, ..., e_n)'$ be an $n \times 1$ vector where $e_i \equiv Y_i - \mathbf{X}_i'\boldsymbol{\beta_a}$. Evaluated at $\boldsymbol{\beta_a}$, the moment condition does not hold:

$$E[\mathbf{Z}_i(Y_i - \mathbf{X}_i'\boldsymbol{\beta_a})] \equiv E[\mathbf{Z}_i e_i] \neq 0, \tag{A.1}$$

if there are more than one LATE. This can be shown by the following argument. For simplicity, assume that we have two instruments, Z_i^1 and Z_i^2, such that each instrument satisfies regularity conditions for identifying the instrument-specific LATE. Let ρ^j be the LATE with respect to Z_i^j and $\boldsymbol{\beta}^j \equiv (\boldsymbol{\gamma}^{j\prime}, \rho^j)'$ be the parameter vector for $j = 1, 2$. By assumption, $\boldsymbol{\beta^1} \neq \boldsymbol{\beta^2}$. If we use each instrument at a time, $E[Z_i^1(Y_i - \mathbf{X}_i'\boldsymbol{\beta^1})] = E[Z_i^2(Y_i - \mathbf{X}_i'\boldsymbol{\beta^2})] = 0$. Now assume $E[\mathbf{Z}_i(Y_i - \mathbf{X}_i'\boldsymbol{\beta_a})] = 0$ holds. Then $E[Z_i^1(Y_i - \mathbf{X}_i'\boldsymbol{\beta_a})] = E[Z_i^2(Y_i - \mathbf{X}_i'\boldsymbol{\beta_a})] = 0$, but this implies $\boldsymbol{\beta_a} = \boldsymbol{\beta^1} = \boldsymbol{\beta^2}$.

This contradicts to the assumption. Thus, (A.1) holds.

From the first-order condition of GMM, we substitute $\mathbf{X}\boldsymbol{\beta_a} + \boldsymbol{e}$ for $\mathbf{Y}$, rearrange terms, and multiply $\sqrt{n}$ to have

$$
\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_a}) &= (\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\sqrt{n}\mathbf{Z}'\boldsymbol{e}, \qquad (A.2)\\
&= \left(\frac{1}{n}\mathbf{X}'\mathbf{Z}\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\frac{1}{n}\mathbf{Z}'\mathbf{X}\right)^{-1} \times\\
&\quad \left\{\frac{1}{n}\mathbf{X}'\mathbf{Z}\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}\sqrt{n}\left(\frac{1}{n}\mathbf{Z}'\boldsymbol{e} - E[\mathbf{Z}_i e_i]\right)\right.\\
&\quad +\sqrt{n}\left(\frac{1}{n}\mathbf{X}'\mathbf{Z} - E[\mathbf{X}_i\mathbf{Z}_i']\right)\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}E[\mathbf{Z}_i e_i]\\
&\quad \left. + E[\mathbf{X}_i\mathbf{Z}_i']\sqrt{n}\left(\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1} - (E[\mathbf{Z}_i\mathbf{Z}_i'])^{-1}\right)E[\mathbf{Z}_i e_i]\right\}.
\end{aligned}
$$

The second equality holds because the population first-order condition of GMM holds regardless of misspecification, i.e., $0 = E[\mathbf{X}_i\mathbf{Z}_i']E[\mathbf{Z}_i\mathbf{Z}_i]^{-1}E[\mathbf{Z}_i e_i]$. The expression (A.2) is different from the standard one because $E[\mathbf{Z}_i e_i] \neq 0$. As a result, the asymptotic variance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_a})$ includes additional terms, which are assumed to be zero in the standard asymptotic variance matrix of 2SLS. We use the fact that

$$
\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1} - E[\mathbf{Z}_i\mathbf{Z}_i']^{-1} = (E[\mathbf{Z}_i\mathbf{Z}_i'])^{-1}\left(E[\mathbf{Z}_i\mathbf{Z}_i'] - \frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)\left(\frac{1}{n}\mathbf{Z}'\mathbf{Z}\right)^{-1}, \qquad (A.3)
$$

and take the limit of the right-hand-side of (A.2). By the weak law of large numbers (WLLN), the continuous mapping theorem (CMT), and the central limit theorem (CLT),

$$
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta_a}) \xrightarrow{d} H^{-1} \cdot N(\mathbf{0}, \Omega). \qquad (A.4)
$$

Q.E.D.

Proposition 2

Proof. Since $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta_a}$, by WLLN and CMT, $n^{-1}\sum_i \hat{\boldsymbol{\psi}}_i \hat{\boldsymbol{\psi}}_i'$ is consistent for Ω. By using WLLN and CMT again, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{MR}}$ is consistent for $H^{-1}\Omega H^{-1}$. *Q.E.D.*

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. Journal of the American Statistical Association, 97(457), 284-292.

Andrews, D. W. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. Econometrica, 70(1), 119-162.

Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. Amerian Economic Review 88(3), 450-477.

Angrist, I. D. and Fernandez-Val, I. (2013, May). ExtrapoLATE-ing: External Validity and. In Advances in Economics and Econometrics: Tenth World Congress (Vol. 3, p. 401). Cambridge University Press.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American statistical Association, 90(430), 431-442.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American statistical Association, 91(434), 444-455.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings?. Quarterly Journal of Economics, 106(4), 979-1014.

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. Journal of Labor Economics, 28(4), 773-824.

Angrist, J. and Pischke, J. S. (2009). Mostly Harmless Econometrics: An Empiricists Companion. Princeton University Press.

Berkowitz, D., Caner, M., and Fang, Y. (2008). Are nearly exogenous instruments reliable? Economics Letters 101 (1), 20-23.

Berkowitz, D., Caner, M., and Fang, Y. (2012). The validity of instruments revisited. Journal of Econometrics 166 (2), 255-266.

Bravo, F. (2010). Efficient M-estimators with auxiliary information. Journal of Statistical Planning and Inference 140 (11), 3326-3342.

Brown, B. W. and Newey, W. K. (2002). Generalized method of moments, efficient bootstrapping, and improved inference. Journal of Business and Economic Statistics, 20(4), 507-517.

DiTraglia, F. J. (2013). Using invalid instruments on purpose: Focused moment selection and averaging for GMM. Working Paper. University of Pennsylvania.

Doyle Jr, J. J. (2008). Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care. Journal of Political Economy, 116(4), 746-770.

Evans, W. N., and Garthwaite, C. (2012). Estimating heterogeneity in the benefits of medical treatment intensity. Review of Economics and Statistics, 94(3), 635-649.

Evans, W. N. and Lien, D. S. (2005). The benefits of prenatal care: evidence from the PAT bus strike. Journal of Econometrics 125(1), 207-239.

Guggenberger, P. (2012). On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. Econometric Theory 28 (2), 387-421.

Hahn, J. and Hausman, J. (2005). Estimation with valid and invalid instruments. Annales d'Économie et de Statistique, 25-57.

Hall, A. R. and Inoue, A. (2003). The large sample behavior of the generalized method of moments estimator in misspecified models. Journal of Econometrics 114 (2), 361-394.

Hall, P. and Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica, 64(4), 891-916.

Imbens, G. W., and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. Econometrica 62(2), 467-475.

Kitagawa, T. (2014). A Test for Instrument Validity. CeMMAP Working Paper.

Kolesár, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. unpublished manuscript. Cowles Foundation.

Lee, S. (2014). Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators. Journal of Econometrics, 178(3), 398-413.

Otsu, T. (2011). Moderate deviations of generalized method of moments and empirical likelihood estimators. Journal of Multivariate Analysis 102 (8), 1203-1216.

Siminski, P. and Ville, S. (2011). Long-run mortality effects of Vietnam-era army service: evidence from Australia's conscription lotteries. American Economic Review, 345-349.

Stephens, M. Jr. and Yang, D-Y. (2014). Compulsory education and the benefits of schooling. American Economic Review, 104(6), 1777-92.
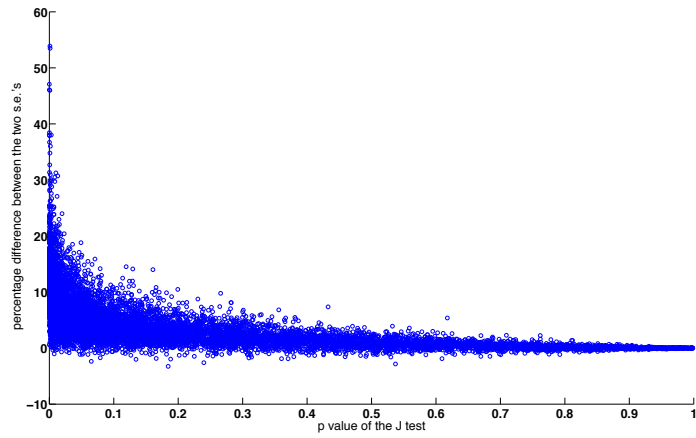
(a) $n = 100$
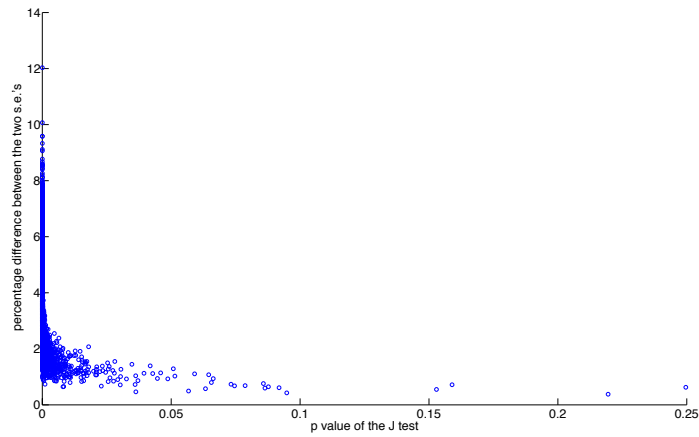
(b) $n = 1,000$

Figure 1: Relationship between p-values of the J test and percentage difference between two standard errors, $s.e._{\hat{\boldsymbol{\Sigma}}_{MR}}$ and $s.e._{\hat{\boldsymbol{\Sigma}}_C}$

16