On Variance Estimation for 2SLS
When Instruments Identify Different LATEs

Seojeong Lee†

June 30, 2014

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 this case, the conventional heteroskedasticity-robust variance estimator for 2SLS is incorrect because the postulated moment condition evaluated at the 2SLS estimand does not hold unless those LATEs are the same. I provide a correct formula for the asymptotic variance of 2SLS by using the result of Hall and Inoue (2003).

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. That is, each individual has a different causal effect of treatment on outcome. If an instrument

---

†School of Economics, Australian School of Business, UNSW, Sydney NSW 2052 Australia, email: jay.lee@unsw.edu.au, homepage: https://sites.google.com/site/misspecified/

I thank Tue Gorgens for stimulating my interest on this topic. I also thank Bruce Hansen, Han Hong, and Toru Kitagawa as well as session participants at SETA 2014 conference for their helpful discussions and suggestions.
is available, then the average treatment effect (ATE) of those whose treatment status can be changed by the instrument, thus the local ATE, can be identified. Hence, LATE is instrument-specific. When multiple instruments are available, the two-stage least squares (2SLS) estimator is commonly used to estimate the causal effect. It is well known that the 2SLS estimand is a weighted average of different LATEs and a rejection of the overidentification test (the J test, hereinafter) is merely due to treatment effect heterogeneity, rather than due to invalid instruments.

What is less well known and often overlooked in the literature is that the rejection of the J test implies that the postulated moment condition model is likely to be misspecified. If so, the conventional standard errors are no longer consistent regardless of how small the differences among individual LATEs are. This fact has been neglected and the conventional standard errors have been routinely calculated even with small \( p \) values of the J test, e.g. see Angrist and Evans (1998) and Angrist, Lavy, and Schlosser (2010), among others.

In this paper, I propose a consistent estimator for the correct asymptotic variance of 2SLS under treatment effect heterogeneity. Recently, 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) constructs an interesting estimand for new subpopulations by reweighting covariate-specific LATEs. However, neither of the two papers considers correct variance estimation.

## 2 2SLS with multiple instruments

### 2.1 Random coefficient model

The notations follow Imbens and Angrist (1994). Assume a binary treatment, \( D_i \) and an outcome variable \( Y_i \). Let \( Y_{1i} \) 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 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. To overcome this endogeneity problem, researchers often rely on instrumental variables. An instrument, \( Z_i \), is independent of \( Y_{1i} \) and \( Y_{0i} \), and
correlated with the treatment, $D_i$. 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. Since $D_i \equiv D_{0i} + (D_{1i} - D_{0i})Z_i$ and $Y_i \equiv Y_{0i} + (Y_{1i} - Y_{0i})D_i$, we can write a random coefficient model

$$D_i = \pi_0 + \pi_1 Z_i + \xi_i,$$

$$Y_i = \alpha_0 + \rho_i D_i + \eta_i,$$

where $\pi_0 \equiv E[D_{0i}]$, $\pi_1 \equiv D_{1i} - D_{0i}$, $\alpha_0 \equiv E[Y_{0i}]$ and $\rho_i \equiv Y_{1i} - Y_{0i}$. Note that $\pi_1i$ and $\rho_i$ are heterogeneous across individuals. Imbens and Angrist (1994) establish that under regularity conditions,

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

That is, the IV estimand (or the Wald estimand) is equal to ATE for a subpopulation such that $D_{1i} > D_{0i}$. This is called LATE. The subpopulation whose treatment status can be changed by the instrument is 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 $D_{0i} = 1$.

The above setting can be generalized to multiple instruments. Without loss of generality, consider mutually exclusive binary instruments, $Z_ji$ for $j = 1, \ldots, q$. Let $D_{zi}^j$ be $i$'s potential treatment status when $Z_ji = z$ where $z = 0, 1$, and $j = 1, \ldots, q$. Each instrument identifies a version of LATE because compliers may differ for each $Z_ji$. It is well known that the 2SLS estimator using multiple instruments is consistent for a weighted average of different LATEs. That is, the 2SLS estimand is a weighted average of treatment effects for instrument-specific compliers (Angrist and Pischke, 2009; Kolesár, 2013):

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

where $\psi_j$ is a nonnegative number and $\sum_j \psi_j = 1$. 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 ATE of those
who had more children than they otherwise would have had because of twinning, while the same-sex instrument identifies ATE 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.

Whether this weighted average of different LATEs is an interesting parameter or not is context-specific. Complier groups may or may not overlap. If a complier group is of researchers’ interest and other complier groups similarly overlap, then the resulting 2SLS estimand would be the parameter of interest as it is similar to the individual LATEs. In practice, this is often justified by a failure to reject the null of the J test. But even a small difference in those LATEs would lead to a rejection of the null hypothesis of homogenous LATEs asymptotically. If complier groups do not overlap much and exhibit distinct characteristics, then combining them may make the combined subpopulation more representative of the whole population. As a result, the 2SLS estimate would be more informative than a single IV estimate. This case is often followed by a significant J test statistic.

2.2 Moment condition model

While the random coefficient model provides a theoretical ground for LATE, the IV and 2SLS estimators can be derived from a slightly different but more general framework, the moment condition model. Consider a linear model

$$Y_i = \alpha_0 + \rho_0 D_i + \eta_i \equiv X'_i \beta_0 + \eta_i,$$

(2.3)

where $X_i = (1, D_i)'$ and $\beta_0 = (\alpha_0, \rho_0)'$. If $D_i$ is endogeneous, i.e., $D_i$ is correlated with $\eta_i$, $\beta_0$ cannot be consistently estimated by OLS. If an instrument vector $Z_i = (1, Z^1_i, Z^2_i, \ldots, Z^q_i)'$ such that $E[Z_i \eta_i] = 0$ exists, then $\beta_0$ can be consistently estimated by the 2SLS estimator

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

(2.4)

where $X \equiv (X'_1, \ldots, X'_n)'$ is an $n \times 2$ matrix, $Z \equiv (Z'_1, \ldots, Z'_n)'$ is an $n \times (q + 1)$ matrix$^1$, and $Y \equiv (Y_1, \ldots, Y_n)'$ is an $n \times 1$ vector. The 2SLS estimator is a special

$^1$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
\[
0 = E[\mathbf{Z}_i\eta_i] = E[\mathbf{Z}_i(Y_i - \alpha_0 - \rho_0 D_i)].
\] (2.5)

When (2.5) 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 moment condition is greater than that of the parameter vector and just-identified when they are equal. For example, if \(Z^1_i\) is the only instrument available, the model is just-identified and the solution is given by
\[
\alpha_0 = E[Y_i] - \rho_0 \cdot E[D_i],
\]
\[
\rho_0 = \frac{Cov(Y_i, Z^1_i)}{Cov(D_i, Z^1_i)} = E[\rho_i | D^1_i > D^1_{0i}].
\] (2.6)

Thus, LATE for \(Z^1_i\) is identified.

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 (2.5) simultaneously even if the instrument vector is uncorrelated with the error term \(\eta_i\). To see this, suppose that there are two instruments, \(Z^1_i\) and \(Z^2_i\). The moment condition is
\[
0 = E[Y_i - \alpha_0 - \rho_0 D_i],
\]
\[
0 = E[Z^1_i(Y_i - \alpha_0 - \rho_0 D_i)],
\]
\[
0 = E[Z^2_i(Y_i - \alpha_0 - \rho_0 D_i)].
\] (2.7)

Solving the first equation for \(\alpha_0\) and substituting it into the second and third equations, we have
\[
\rho_0 = \frac{Cov(Y_i, Z^1_i)}{Cov(D_i, Z^1_i)} = \frac{Cov(Y_i, Z^2_i)}{Cov(D_i, Z^2_i)}. \] (2.8)

But this implies that the two LATEs are the same, which is not true in general. Thus, (2.8) 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 effect 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 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 variance estimation for 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. Since 2SLS is a special case of GMM, the argument can be easily shown by using that of Hall and Inoue (2003).

**Proposition 1.** Let \((Y_i, X_i, Z_i)_{i=1}^{n}\) be an iid sample, where \(X_i = (1, D_i)'\) and \(Z_i = (1, Z_i^1, Z_i^2, \ldots, Z_i^q)'\). Suppose Theorem 2 (the LATE theorem) of Imbens and Angrist (1994) holds for each \(Z_i^j\) and the model is given by (2.3). Then \(\hat{\beta}\) is consistent for \(\beta_a = (\alpha_a, \rho_a)'\) where \(\alpha_a = E[Y_i] - \rho_a E[D_i]\) and \(\rho_a\) is a linear combination of individual LATE parameters. In addition, the asymptotic distribution is

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

where \(H = E[X_i Z_i'] (E[Z_i Z_i'])^{-1} E[Z_i X_i']\) and \(\Omega\) is a \((k + 2q + 2) \times (k + 2q + 2)\) variance-covariance matrix of the following process

\[
\sqrt{n}^{-1} \sum_{i=1}^{n} \begin{pmatrix}
Z_i (Y_i - X_i' \beta_a) - E[Z_i (Y_i - X_i' \beta_a)] \\
(X_i Z_i' - E[X_i Z_i']) (E[Z_i Z_i'])^{-1} E[Z_i (Y_i - X_i' \beta_a)] \\
E[Z_i Z_i']^{-1} (E[Z_i Z_i'] - Z_i Z_i') (\frac{1}{n} \sum_{i=1}^{n} Z_i Z_i')^{-1}
\end{pmatrix}.
\]

**Proof.** By the Weak Law of Large Numbers (WLLN), the continuous mapping theorem (CMT) and the LATE theorem, consistency of \(\hat{\beta}\) for \(\beta_a\) immediately follows.\(^2\)

Let \(\varepsilon \equiv (\varepsilon_1, \ldots, \varepsilon_n)'\) be an \(n \times 1\) vector where \(\varepsilon_i \equiv Y_i - X_i' \beta_a\). Note that \(E[Z_i \varepsilon_i] \neq 0\).

\(^2\)This is the 2SLS estimand and also called the pseudo-true value in a general context.
From the first-order condition, substitute $X\beta_a + \varepsilon$ for $Y$ and rearrange terms to have

$$
\hat{\beta} - \beta_a = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon, \quad (2.9)
$$

The second equality holds because the population first-order condition of GMM holds regardless of misspecification, $0 = E[X_iZ_i]E[Z_iZ_i]^{-1}E[Z_i\varepsilon_i]$. The expression (2.9) is different from the usual one because $E[Z_i\varepsilon_i] \neq 0$. As a result, the asymptotic variance matrix of $\sqrt{n}(\hat{\beta} - \beta)$ includes additional terms, which are assumed to be zero in the standard asymptotic variance matrix of 2SLS. By taking the limit of the right-hand-side of (2.9), the Proposition follows. Q.E.D.

**Remark 1.** Since the proof of Proposition 1 deriving the asymptotic distribution is essentially the same as that of Hall and Inoue (2003), there is little technical contribution. 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. Thus, the analysis of its asymptotic behavior can be significantly simplified.

The next Proposition proposes a consistent estimator for the asymptotic variance matrix of 2SLS when $0 = E[Z_i\varepsilon_i]$ may not hold. The proof is straightforward by using WLLN and CMT, and thus omitted.

**Proposition 2.** A treatment-heterogeneity-robust asymptotic variance estimator for 2SLS is given by

$$
\hat{\Sigma}_{MR} = n \cdot (X'Z(Z'Z)^{-1}Z'X)^{-1} \left( \sum_i \psi_i \psi_i' \right) \left( X'Z(Z'Z)^{-1}Z'X \right)^{-1} \quad (2.10)
$$
where \( \hat{\varepsilon}_i = Y_i - X_i'\hat{\beta} \), \( \hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n)' \), and

\[
\psi_i = \frac{1}{n} X'Z \left( \frac{1}{n} Z'Z \right)^{-1} \left( Z_i \hat{\varepsilon}_i - \frac{1}{n} Z'\hat{\varepsilon} \right)
+ \left( X_i Z_i' - \frac{1}{n} X'Z \right) \left( \frac{1}{n} Z'Z \right)^{-1} \frac{1}{n} Z'\hat{\varepsilon}
- \frac{1}{n} X'Z \left( \frac{1}{n} Z'Z \right)^{-1} \left( Z_i Z_i' - \frac{1}{n} Z'Z \right) \left( \frac{1}{n} Z'Z \right)^{-1} \frac{1}{n} Z'\hat{\varepsilon}.
\]

The formula of \( \hat{\Sigma}_{MR} \) is different from that of the conventional heteroskedasticity-robust variance estimator for 2SLS:

\[
\hat{\Sigma}_C = n \cdot \left( X'Z \left( Z'Z \right)^{-1} Z'X \right)^{-1} \left( \sum_i Z_i Z_i' \hat{\varepsilon}_i^2 \right) \left( X'Z \left( Z'Z \right)^{-1} Z'X \right)^{-1}.
\] (2.12)

Although both \( \hat{\Sigma}_{MR} \) and \( \hat{\Sigma}_C \) converge in probability to the same limit under treatment effect homogeneity, but they are generally different in finite sample. \( \hat{\Sigma}_{MR} \) is consistent for the true asymptotic variance matrix regardless of whether the postulated moment condition is misspecified or not, while \( \hat{\Sigma}_C \) is consistent only if the underlying LATEs are the same. This is also true for the standard errors based on \( \hat{\Sigma}_{MR} \) and \( \hat{\Sigma}_C \).

**Remark 2 (Generalization of LATE)** The conclusions of Propositions 1 and 2 do not change with covariates. Let \( W_i \) be the vector of covariates including a constant. Then \( X_i = (W_i', D_i)' \) and \( Z_i = (W_i', Z_{i1}', \ldots, Z_{iq}')' \) in Propositions 1 and 2. In addition, it is possible that \( Z_i^j \) or \( D_i \) can take multiple values, or even continuous. With covariates or with variable treatment/instrument intensity, only the interpretation of LATE parameters changes.

**Remark 3 (Invalid Instruments)** The treatment-heterogeneity-robust variance estimator \( \hat{\Sigma}_{MR} \) is also robust to invalid instruments, i.e. instruments correlated with the error term. Consider a generic linear model

\[
Y_i = X_i'\beta_0 + \eta_i,
\] (2.13)

where \( X_i \) is a \((k + p) \times 1\) vector of regressors. Among \( k + p \) regressors, \( p \) are endogenous, i.e. \( E[X_i\eta_i] \neq 0 \). If a \((k + q)\) vector of instruments \( Z_i \) is available such that
$E[Z_i \eta_i] = 0$ and $q \geq p$, then $\beta_0$ can be consistently estimated by 2SLS or GMM. If any of the instruments is invalid, then $E[Z_i \eta_i] \neq 0$ and $\beta_0$ may not be consistently estimated. Since the moment condition does not hold, the model is misspecified. There are two types of misspecification: (i) fixed or global misspecification such that $E[Z_i \eta_i] = \delta$ where $\delta$ is a constant vector containing at least one non-zero component, and (ii) local misspecification such that $E[Z_i \eta_i] = n^{-r} \delta$ for some $r > 0$. With a particular choice of $r = 1/2$, this framework has been used to analyse the asymptotic behavior of the 2SLS estimator with invalid instruments by Hahn and Hausman (2005), Bravo (2010), Berkowitz, Caner, and Fang (2008, 2012), Otsu (2011), Guggenberger (2012), DiTraglia (2013), among others. Either under fixed misspecification or under local misspecification, $\hat{\Sigma}_{MR}$ in Proposition 2 is consistent for the true asymptotic variance matrix. However, the conventional variance estimator $\hat{\Sigma}_C$ is inconsistent under fixed misspecification. Under local misspecification, $\hat{\Sigma}_C$ is consistent for the true asymptotic variance matrix but its finite sample performance can be poor.

**Remark 4 (Bootstrap)** Bootstrapping can be used to get more accurate $t$ tests and confidence intervals (CI’s) based on $\hat{\beta}$, 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, the misspecification-robust bootstrap of Lee (2014) should be used. In contrast, the conventional bootstrap methods for overidentified moment condition models of Hall and Horowitz (1996), Brown and Newey (2002), and Andrews (2002) assume correctly specified model, 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 $m$th element of $\beta_a$. 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 $m$th elements of $\hat{\beta}^*$ and $\hat{\beta}$, respectively, $\hat{\Sigma}^*_{MR,m}$ is the $m$th diagonal element of $\hat{\Sigma}^*_{MR}$, and $\hat{\beta}^*$ and $\hat{\Sigma}^*_{MR}$ are the bootstrap versions of $\hat{\beta}$ and $\hat{\Sigma}_{MR}$ based on the same formula using the bootstrap sample rather than the original sample.
3 Simulation

The random coefficient model with two mutually exclusive instruments is considered for simulation. First, a single binary instrument $Z^0_i$ 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 $Z^0_i$: 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, $Z^0_i$ is decomposed into two mutually exclusive instruments randomly, $Z^1_i$ and $Z^2_i$, such that $Z^1_i + Z^2_i = Z^0_i$. Thus, in this setting, always-takers and never-takers are common to both instruments, but compliers for $Z^0_i$ is decomposed into three subgroups: Common compliers for both $Z^1_i$ and $Z^2_i$, compliers for $Z^1_i$ only, and compliers for $Z^2_i$ only.

The potential outcomes with and without treatment are generated as

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

Without loss of generality, $\rho_i$ 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 $Z^1_i$ only, and $\rho_i = 1$ if $i$ is a complier for $Z^2_i$ only. LATEs for $Z^1_i$ and $Z^2_i$ are weighted averages of 2 and 3, and 1 and 2, respectively. Note that $\rho_i$ of always-takers and never-takers do not matter in calculating LATE. The 2SLS estimand $\rho_a$ is a weighted average of the two LATEs, which are different. Thus, the postulated moment condition is misspecified.

Table 1 shows the mean and the standard deviation of the 2SLS estimator, the mean of conventional/misspecification-robust standard errors, and the rejection prob-

<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>1,000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean($\hat{\rho}$)</td>
<td>1.8926</td>
<td>1.9213</td>
<td>1.9235</td>
</tr>
<tr>
<td>s.d.($\hat{\rho}$)</td>
<td>1.1770</td>
<td>0.3618</td>
<td>0.1140</td>
</tr>
<tr>
<td>mean(s.e.$\hat{\Sigma}_{MR}$)</td>
<td>1.1699</td>
<td>0.3607</td>
<td>0.1137</td>
</tr>
<tr>
<td>mean(s.e.$\hat{\Sigma}_{C}$)</td>
<td>1.1188</td>
<td>0.3485</td>
<td>0.1100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>100</th>
<th>1,000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>reject J test at 5%</td>
<td>32.6%</td>
<td>99.8%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 1: The standard deviation and standard errors of $\hat{\rho}$ when the 2SLS estimand is a weighted average of different LATEs
ability of the $J$ test at 5% for different sample sizes. The number of Monte Carlo repetition is 10,000. The result clearly shows that the conventional standard error based on $\hat{\Sigma}_C$ is inconsistent for the standard deviation of the estimator, and underestimate it for any sample size $n$. In contrast, the misspecification-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 $Z^2_i$ is larger than that for $Z^1_i$.

Figure 1 shows a negative relationship between the p-values of the $J$ test and the percentage difference of the two standard errors $s.e.\hat{\Sigma}_{MR}$ and $s.e.\hat{\Sigma}_C$ for different sample sizes. 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 $\hat{\Sigma}_{MR}$ is consistent for the true asymptotic variance under treatment effect heterogeneity, it is recommended to report $s.e.\hat{\Sigma}_{MR}$ 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 by a statistical program are likely to be inconsistent because they do not take into account the consequence of possible misspecification of the postulated moment condition model. This paper provided a solution to such dilemmas. The proposed formula of the variance estimator for 2SLS is consistent for the true asymptotic variance regardless of whether the underlying LATEs are heterogenous or not. In addition, this variance estimator is robust to invalid instruments, and can be used for bootstrapping to achieve asymptotic refinements.
References


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