`ivreg`

Michal Kolesár[*]

Version 40501b6 (This version is not tagged), April 9, 2013[†]

Abstract

This note provides additional documentation and background for the `ivreg.m` Matlab function. `ivreg` computes a number of estimators of the causal effect in an instrumental variables regression, and provides several estimators of standard errors based on different asymptotic approximations. It also computes additional statistics and diagnostic tests.

ToDo: derive and add standard error estimators for `ujive` and `rtsls`

1. Instrumental variables model and notation

The standard instrumental variables model is given by

$$y_i = T_i\beta + W_i'\psi + \epsilon_i, \tag{1}$$

where $y_i \in \mathbb{R}$ is the outcome variable, $T_i \in \mathbb{R}$ is a single endogenous regressor, $W_i \in \mathbb{R}^L$ is a vector of exogenous regressors (covariates), and ϵ_i is a structural error. The parameter of interest is β. The identifying assumption is that the structural error ϵ_i is uncorrelated with the covariates W_i and a given vector of instruments $Z_i \in \mathbb{R}^K$:

$$\mathbb{E}\left[\epsilon_i \begin{pmatrix} Z_i \\ W_i \end{pmatrix}\right] = 0. \tag{2}$$

We observe an i.i.d. sample $\{Y_i, T_i, W_i, Z_i\}_{i=1}^n$. The arguments of `ivreg` are the matrices $\mathbf{Y}$, $\mathbf{T}$, $\mathbf{Z}$, and $\mathbf{W}$, with rows y_i, T_i, Z_i' and W_i'.

For any full-rank $n \times m$ matrix $\mathbf{A}$, let $\mathbf{H_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ denote the associated $n \times n$ projection matrix (also known as the hat matrix), and let $\mathbf{D_A}$ be an $n \times n$ diagonal matrix with $(\mathbf{H_A})_{ii}$ on the diagonal. Let $\mathbf{I}_m$ denote the $m \times m$ identity matrix, and let $\mathbf{M_A} = \mathbf{I}_n - \mathbf{H_A}$ denote the annihilator matrix. Let $\mathbf{A}_\perp = \mathbf{M_W}\mathbf{A}$ denote the residual from the sample projection of $\mathbf{A}$ onto $\mathbf{W}$.

2. Estimation

The command

```
betahat = ivreg(Y, T, Z, W);
```

[*]kolesarmi@googlemail.com

[†]Version hash 40501b6, compiled May 19, 2013

returns a vector `betahat` of different estimators of the causal effect β in Equation (1). In particular:

```
betahat = [ols, tsls, liml, mbtsls, jive, ujive, rtsls]
```

where `ols` is the least-squares estimator, `tsls` is the two-stage least squares estimator, `liml` is the limited information maximum likelihood estimator (Anderson and Rubin, 1949), `mbtsls` is the modified bias-corrected two-stage least squares estimator (Anatolyev, 2011; Kolesár, Chetty, Friedman, Glaeser and Imbens, 2011), `jive` is the jackknife iv estimator (Phillips and Hale, 1977; Angrist, Imbens and Krueger, 1999), also called JIVE1, `ujive` is the Kolesár (2012a) version of the jackknife iv estimator, and `rtsls` is the reverse two-stage least squares estimator.

Expressions for estimators of β All estimators can be written as

$$\hat{\beta} = \frac{\hat{\mathbf{P}}'\mathbf{Y}}{\hat{\mathbf{P}}'\mathbf{T}}, \tag{3}$$

where the form of $\hat{\mathbf{P}}$ depends on the estimator. In particular,

$$\begin{aligned}
\hat{\mathbf{P}}_{\texttt{ols}} &= \mathbf{T}_{\perp} = (\mathbf{I} - 0 \cdot \mathbf{M}_{\mathbf{Z}_{\perp}})\mathbf{T}_{\perp}, \\
\hat{\mathbf{P}}_{\texttt{tsls}} &= \mathbf{H}_{\mathbf{Z}_{\perp}}\mathbf{T} = (\mathbf{I} - 1 \cdot \mathbf{M}_{\mathbf{Z}_{\perp}})\mathbf{T}_{\perp}, \\
\hat{\mathbf{P}}_{\texttt{liml}} &= \mathbf{M}_{\mathbf{W}}(\mathbf{I} - k_{\texttt{liml}}\mathbf{M}_{\mathbf{W},\mathbf{Z}})\mathbf{T} = (\mathbf{I} - k_{\texttt{liml}}\mathbf{M}_{\mathbf{Z}_{\perp}})\mathbf{T}_{\perp}, \\
\hat{\mathbf{P}}_{\texttt{mbtsls}} &= \mathbf{M}_{\mathbf{W}}(\mathbf{I} - k_{\texttt{mbtsls}}\mathbf{M}_{\mathbf{W},\mathbf{Z}})\mathbf{T} = (\mathbf{I} - k_{\texttt{mbtsls}}\mathbf{M}_{\mathbf{Z}_{\perp}})\mathbf{T}_{\perp}, \\
\hat{\mathbf{P}}_{\texttt{jive}} &= \mathbf{M}_{\mathbf{W}}(\mathbf{I}_n - \mathbf{D}_{\mathbf{Z},\mathbf{W}})^{-1}(\mathbf{H}_{\mathbf{Z},\mathbf{W}} - \mathbf{D}_{\mathbf{Z},\mathbf{W}})\mathbf{T} = \mathbf{M}_{\mathbf{W}}(\mathbf{I}_n - (\mathbf{I}_n - \mathbf{D}_{(\mathbf{Z},\mathbf{W})})^{-1}\mathbf{M}_{(\mathbf{Z},\mathbf{W})})\mathbf{T}, \\
\hat{\mathbf{P}}_{\texttt{ujive}} &= \left[(\mathbf{I}_n - \mathbf{D}_{(\mathbf{Z},\mathbf{W})})^{-1}(\mathbf{H}_{(\mathbf{Z},\mathbf{W})} - \mathbf{D}_{(\mathbf{Z},\mathbf{W})}) - (\mathbf{I}_n - \mathbf{D}_{\mathbf{W}})^{-1}(\mathbf{H}_{\mathbf{W}} - \mathbf{D}_{\mathbf{W}})\right]\mathbf{T}, \\
\hat{\mathbf{P}}_{\texttt{rtsls}} &= \mathbf{H}_{\mathbf{Z}_{\perp}}\mathbf{Y},
\end{aligned}$$

where

$$k_{\texttt{liml}} = \min \operatorname{eig}\left[\left(\begin{pmatrix}\mathbf{Y} & \mathbf{T}\end{pmatrix}'\mathbf{M}_{\mathbf{Z},\mathbf{W}}\begin{pmatrix}\mathbf{Y} & \mathbf{T}\end{pmatrix}\right)^{-1}\begin{pmatrix}\mathbf{Y} & \mathbf{T}\end{pmatrix}'\mathbf{M}_{\mathbf{W}}\begin{pmatrix}\mathbf{Y} & \mathbf{T}\end{pmatrix}\right], \tag{4}$$

and $k_{\texttt{mbtsls}} = (1 - L/n)/(1 - (K-1)/n - L/n)$. The version of `mbtsls` proposed in Kolesár *et al.* (2011) uses $k_{\texttt{mbtsls}} = (1 - L/n)/(1 - K/n - L/n)$—the modification used here ensures that when $K = 1$ (there is a single instrument), the `tsls` and `mbtsls` estimators will coincide.

3. Standard Errors

The command

```
[betahat, se] = ivreg(Y, T, Z, W);
```

returns an estimate of standard errors for the different estimators of β. `se` is a 4×7 matrix with rows computed as follows

- The first row gives estimates of the asymptotic standard errors under homoscedasticity ($\mathbb{E}[\epsilon_i^2 \mid z_i, w_i] = \mathbb{E}[\epsilon_i^2]$) and standard asymptotics, which hold the distribution of the observed data (Y_i, T_i, W_i, Z_i) fixed as $n \to \infty$;

- The second row gives heteroscedasticity-robust standard errors;
- The third row gives standard errors which are valid under many-instrument asymptotics. Like in Bekker (1994), these errors are valid when the number of instruments, K, is allowed to grow in proportion with the sample size, $K/n \to \kappa$. In addition, the number of exogenous covariates, W_i, is also allowed to grow with the sample size, $L/n \to \lambda$, as in Anatolyev (2011) and Chetty, Friedman, Hilger, Saez, Schanzenbach and Yagan (2011). As in Bekker (1994), the structural error ϵ_i is assumed to be Normally distributed. Since `tsls` and `jive` are not consistent under these asymptotics, elements of the `se` matrix that correspond to them return `NaN`;
- The fourth row gives standard errors that are valid under the many invalid instrument asymptotic sequence of Kolesár *et al.* (2011). Only `mbtsls` and `ujive` are consistent for β under this sequence, so element of the `se` matrix that correspond to other estimators return `NaN`.

Standard errors under homoscedasticity and standard asymptotics Let

$$\hat{\sigma}_{\epsilon}^2(\hat{\beta}) = (\mathbf{Y}_{\perp} - \mathbf{T}_{\perp}\hat{\beta})'(\mathbf{Y}_{\perp} - \mathbf{T}_{\perp}\hat{\beta})/n. \tag{5}$$

The standard error for `ols` is given by

$$\widehat{se}(\hat{\beta}_{\texttt{ols}}) = \sqrt{\frac{\hat{\sigma}_{\epsilon}^2(\hat{\beta})}{\mathbf{T}_{\perp}'\mathbf{T}_{\perp}}}.$$

The estimator of the standard error of $\hat{\beta}$ for `tsls`, `mbtsls` and `liml` is given by

$$\widehat{se}(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}_{\epsilon}^2(\hat{\beta})}{\mathbf{T}'\mathbf{H}_{\mathbf{Z}_{\perp}}\mathbf{T}}}.$$

The standard errors will be different if they produce different estimates of β. These expressions correspond to the standard errors for `tsls` in Stata, as well as to the standard errors for `mbtsls` and `liml` when the option "`coviv`" is given (Baum, Schaffer and Stillman, 2007). The standard error for `jive` and `ujive` is given by:

$$\widehat{se}(\hat{\beta}) = \frac{\sqrt{\hat{\sigma}_{\epsilon}^2(\hat{\beta}) \cdot \hat{\mathbf{P}}'\hat{\mathbf{P}}}}{\hat{\mathbf{P}}'\mathbf{T}}.$$

This estimator matches the standard error estimator for `jive` in Stata.[1]

Standard errors under heteroscedasticity Let

$$\hat{\epsilon}_{\hat{\beta}} = \mathbf{Y} - \mathbf{T}\hat{\beta} - \mathbf{W}\hat{\psi}(\hat{\beta}) = \mathbf{Y}_{\perp} - \mathbf{T}_{\perp}\hat{\beta}$$

[1]The Angrist *et al.* (1999) JIVE1 esimator is referred to as UJIVE1 in Stata, and is computed by the `jive` command with the `ujive1` option, which is also the default option.

be an estimate of the structural error in Equation (1), where $\hat{\psi}(\hat{\beta}) = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}(\mathbf{Y} - \mathbf{T}\hat{\beta})$. The heteroscedasticity-robust standard error is given by

$$\widehat{se}(\hat{\beta}) = \frac{\sqrt{\sum_{i=1}^{n} \hat{\epsilon}_{\beta,i}^2 \hat{\mathbf{P}}_i^2}}{\hat{\mathbf{P}}'\mathbf{T}},$$

where for `ols`, $\hat{\mathbf{P}} = \mathbf{T}_{\perp}$, for `tsls`, `liml`, and `mjive`, $\hat{\mathbf{P}} = \mathbf{H}_{\mathbf{Z}_{\perp}}\mathbf{T}$, for `jive` $\hat{\mathbf{P}} = \hat{\mathbf{P}}_{\texttt{jive}}$, and for `ujive`, $\hat{\mathbf{P}} = \hat{\mathbf{P}}_{\texttt{ujive}}$.

Standard errors under many instrument asymptotics The estimator of the standard error for `liml` corresponds to $\sqrt{-\hat{\mathcal{H}}_{\text{RE}}^{11}}$, where $\hat{\mathcal{H}}_{\text{RE}}^{11}$ is the (1,1) element of the inverse Hessian of the random effects likelihood (Kolesár, 2012b),

$$\hat{\mathcal{H}}_{\text{RE}}^{11} = \frac{\hat{b}'\hat{\Omega}_{\text{RE}}\hat{b}(\hat{\lambda}_{\text{RE}} + K/n)}{n\hat{\lambda}_{\text{RE}}} \left(\hat{Q}_{\mathcal{S}}\hat{\Omega}_{\text{RE},22} - S_{22} + \frac{\hat{c}}{1-\hat{c}} \frac{\hat{Q}_{\mathcal{S}}}{\hat{a}'\hat{\Omega}_{\text{RE}}^{-1}\hat{a}} \right)^{-1},$$

where

$$S_{\perp} = \begin{pmatrix} \mathbf{Y} & \mathbf{T} \end{pmatrix}' \mathbf{M}_{\mathbf{Z},\mathbf{W}} \begin{pmatrix} \mathbf{Y} & \mathbf{T} \end{pmatrix} / (n-K-L), \qquad S = \begin{pmatrix} \mathbf{Y} & \mathbf{T} \end{pmatrix}' \mathbf{H}_{\mathbf{Z}_{\perp}} \begin{pmatrix} \mathbf{Y} & \mathbf{T} \end{pmatrix} / n, \tag{6}$$

$$\hat{\lambda}_{\text{RE}} = \max \operatorname{eig}(S_{\perp}^{-1}S) - K/n, \tag{7}$$

$$\hat{\Omega}_{\text{RE}} = \frac{n-K-L}{n-L} S_{\perp} + \frac{n}{n-L} \left(S - \frac{\hat{\lambda}_{\text{RE}}}{\hat{a}'S_{\perp}^{-1}\hat{a}} \hat{a}\hat{a}' \right), \qquad \hat{a} = \begin{pmatrix} \hat{\beta}_{\texttt{liml}} \\ 1 \end{pmatrix}, \tag{8}$$

$$\hat{Q}_{\mathcal{S}} = \frac{\hat{b}'S\hat{b}}{\hat{b}'\hat{\Omega}_{\text{RE}}\hat{b}}, \qquad \hat{b} = \begin{pmatrix} 1 \\ -\hat{\beta}_{\texttt{liml}} \end{pmatrix}, \tag{9}$$

$$\hat{c} = \frac{\hat{\lambda}_{\text{RE}}\hat{Q}_{\mathcal{S}}}{(K/n + \hat{\lambda}_{\text{RE}})(1 - L/n)}. \tag{10}$$

The estimator for the standard error of `mbtsls` is based on the Hessian of the uncorrelated random effects likelihood with the estimator of the variance of the direct effects set to zero (Kolesár, 2012b)

$$\hat{\mathcal{H}}_{\text{URE}} = \frac{1}{\hat{\Lambda}_{22}^2} \left(\hat{\Lambda}_{22}\hat{\Sigma}_{11} + \frac{(1-L/n)K/n}{(1-K/n-L/n)} (\hat{\Sigma}_{11}\hat{\Sigma}_{22} + \hat{\Sigma}_{12}^2) \right),$$

where

$$\hat{\Lambda}_{22} = \begin{cases} S_{22} - \frac{K}{n}\hat{\Omega}_{22,\text{URE}} & \text{if } (\min \operatorname{eig}(S_{\perp}^{-1}S)) \geq K/n, \\ \frac{\hat{\lambda}_{\text{RE}}}{\hat{a}_{\text{RE}}'\hat{\Omega}_{\text{RE}}^{-1}\hat{a}_{\text{RE}}} & \text{otherwise.} \end{cases},$$

$$\hat{\Omega}_{\text{URE}} = \begin{cases} S_{\perp} & \text{if } (\min \operatorname{eig}(S_{\perp}^{-1}S)) \geq K/n, \\ \hat{\Omega}_{\text{RE}} & \text{otherwise.} \end{cases},$$

$$\hat{\Sigma}_{\text{RE}} = \hat{\Gamma}_{\text{RE}}^{-1}\hat{\Omega}_{\text{RE}}\hat{\Gamma}_{\text{RE}}^{-1}, \qquad \hat{\Gamma}_{\text{RE}} = \begin{pmatrix} 1 & \hat{\beta}_{\texttt{mbtsls}} \\ 0 & 1 \end{pmatrix}.$$

Many invalid instruments The estimator for the standard error of `mbtsls` is based on the Hessian of the uncorrelated random effects likelihood,

$$\hat{\mathcal{H}}_{\text{URE}} = \frac{1}{\hat{\Lambda}_{22}^2}\left(\hat{\Lambda}_{22}\hat{\Sigma}_{11} + \frac{(1-L/n)K/n}{(1-K/n-L/n)}(\hat{\Sigma}_{11}\hat{\Sigma}_{22} + \hat{\Sigma}_{12}^2) + \hat{\Lambda}_{11}\hat{\Sigma}_{22} + \hat{\Lambda}_{11}\Lambda_{22}\cdot n/K\right),$$

where $\hat{\Lambda}_{22}$ and $\hat{\Sigma}_{22}$ are computed as before, and $\hat{\Lambda}_{11} = \max\left\{\hat{b}'_{\texttt{mbtsls}}(S - \frac{K}{n}S_{\perp})\hat{b}_{\texttt{mbtsls}}, 0\right\}$,

4. Other outputs

The command

```
[betahat, se, stats] = ivreg(Y, T, Z, W);
```

returns a structure `stats`, which contains

```
stats =

         F: 5.5275
     Omega: [2x2 double]
        Xi: [2x2 double]
    Sargan: [42.3602 5.7180e-05]
        CD: [48.9588 8.7507e-06]
```

`stats.F` refers to the first-stage F-statistic,

$$\texttt{stats\{1,1\}} = \frac{\mathbf{T}'\mathbf{H}_{\mathbf{Z}_\perp}\mathbf{T}}{K\cdot\mathbf{T}'\mathbf{M}_{\mathbf{Z},\mathbf{W}}\mathbf{T}/(n-K-L)}.$$

`stats.Omega` is an estimator of the covariance matrix of reduced-form errors

$$\texttt{stats\{2,1\}} = \begin{pmatrix}\mathbf{Y} & \mathbf{T}\end{pmatrix}'\mathbf{M}_{\mathbf{Z},\mathbf{T}}\begin{pmatrix}\mathbf{Y} & \mathbf{T}\end{pmatrix}/(n-K-L).$$

`Stats.Xi` is an estimator of the covariance matrix of the reduced form coefficients, $\Xi = \Pi'\mathbf{Z}_\perp'\mathbf{Z}_\perp\Pi/n$ (see Kolesár (2012a)), where Π is the coefficient on Z_i in the linear predictor $\mathbb{E}^*[(Y_i, T_i) \mid Z_i, W_i]$:

$$\texttt{stats\{3,1\}} = \begin{pmatrix}\mathbf{Y} & \mathbf{T}\end{pmatrix}'\mathbf{H}_{\mathbf{Z}_\perp}\begin{pmatrix}\mathbf{Y} & \mathbf{T}\end{pmatrix}/n - K/n\cdot\texttt{stats\{2,1\}}.$$

The estimator is consistent under the many-instrument asymptotic sequence and homoscedasticity.

`stats.Sargan` gives the test statistic and p-value for the Sargan (1958) test of overidentifying restrictions. `stats{5,1}` a 2-by-1 vector, with the first element equal to the test statistic, given by $n\min\operatorname{eig}(S_\perp^{-1}S)/(1 - K/n - L/n + \min\operatorname{eig}(S_\perp^{-1}S))$, where S and $S_\perp$ are defined in Equation (6), and the second element equal to the p-value.

`Stats.CD` gives the test statistic and p-value for the Cragg and Donald (1993) test of overidentifying restrictions. The p-value contains a size-correction derived in Kolesár (2012b) that ensures correct coverage under many-instrument asymptotics, and it is given by

$$1-\Phi(\sqrt{(n-K-L)/(n-L)}\Phi^{-1}(F_{K-1}(J_S))),$$

where $J_S = \min \operatorname{eig}(S_{\perp}^{-1}S)$ is the Sargan test statistic , Φ is the cdf of a standard Normal distribution, and F_{K-1} is the cdf of a χ^2 distribution with $K-1$ degrees of freedom. If If there is only one instrument, `stats{5,1}` and `stats{6,1}` return **`NaN`**.

Appendices

The appendix derives the expressions for different point estimators as well as estimators of the asymptotic variance. Throughout the appendix, let $\mathbf{X} = (\mathbf{T}, \mathbf{W})$ denote the full matrix of covariates in the structural equation.

A. Estimators

If an estimator of $(\beta, \psi')'$ can be written as $(\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{Y}$, where $\hat{\mathbf{X}} = (\hat{\mathbf{T}}, \mathbf{W})$ for some $\hat{\mathbf{T}}$, we have that

$$\begin{pmatrix} \hat{\beta} \\ \hat{\psi} \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{T}}'\mathbf{T} & \hat{\mathbf{T}}'\mathbf{W} \\ \mathbf{W}'\mathbf{T} & \mathbf{W}'\mathbf{W} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\mathbf{T}}'\mathbf{Y} \\ \mathbf{W}'\mathbf{Y} \end{pmatrix} = \begin{pmatrix} (\hat{\mathbf{T}}'\mathbf{M_W}\mathbf{T})^{-1}\hat{\mathbf{T}}'\mathbf{M_W}\mathbf{Y} \\ (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'(\mathbf{Y}-\mathbf{T}\hat{\beta}) \end{pmatrix}, \tag{11}$$

so that letting $\hat{\mathbf{P}} = \mathbf{M_W}\hat{\mathbf{T}}$, an estimator of β is given by Equation (3). Now:

$$\begin{aligned} \hat{\mathbf{X}}_{\mathtt{ols}} &= \mathbf{X}, \\ \hat{\mathbf{X}}_{\mathtt{tsls}} &= \mathbf{H_{Z,W}}\mathbf{X} = \begin{pmatrix} \mathbf{H_{Z,W}}\mathbf{T} & \mathbf{W} \end{pmatrix}, \end{aligned}$$

which, using the identity

$$\mathbf{M_W}\mathbf{H_{Z,W}} = \mathbf{M_W}(\mathbf{H}_{\mathbf{Z}_\perp} + \mathbf{H_W}) = \mathbf{H}_{\mathbf{Z}_\perp}, \tag{12}$$

leads to the expressions for the `ols` and `tsls` estimators of β. The espressions for `liml`, `mbtsls`, and `jive` follow similarly. The expression for `ujive` and `rtsls` are derived in Kolesár (2012a).

B. Standard Errors

Homoscedasticity The homoscedastic case strengthens the moment condition (2) to

$$\mathbb{E}[\epsilon_i \mid Z_i, W_i] = 0 \qquad \mathbb{E}[\epsilon_i^2 \mid Z_i, W_i] = \mathbb{E}[\epsilon_i^2] = \sigma_\epsilon^2$$

An estimator of variance of the structural error is then given by $\hat{\sigma}_\epsilon^2 = \hat{\epsilon}'\hat{\epsilon}/n$, where $\hat{\epsilon} = \mathbf{Y} - \mathbf{T}\hat{\beta} - \mathbf{W}'\hat{\psi}$. If an estimator admits the representation (11), then $\hat{\epsilon}$ can be written as $\hat{\epsilon} = \mathbf{M_W}(\mathbf{Y} - \mathbf{T}\hat{\beta})$, which leads to the expression for $\hat{\sigma}_\epsilon^2$ given by Equation (5).

The asymptotic variance of the `ols` estimator of the best linear predictor, $(\mathbb{E}[X_iX_i']^{-1}\mathbb{E}[X_i'y_i])$, is given by $\sigma_\epsilon^2 \mathbb{E}[X_i X_i']^{-1}$. A the standard White estimator of this variance is given by

$$\hat{\sigma}_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

Since the (1,1) element of $(\mathbf{X}'\mathbf{X})^{-1}$ is given by $(\mathbf{T}\mathbf{M}_\mathbf{W}\mathbf{T})^{-1}$, the expression for the `ols` standard error follows. The asymptotic variance for `tsls`, `liml`, and `mbtsls` estimators of (β, ψ) is given by[2]

check the Davidson and MacKinnon (1993) reference for asymptotic variances

$$\sigma_\epsilon^2 \left\{ \mathbb{E}[X_i \tilde{Z}_i'] \mathbb{E}[\tilde{Z}_i \tilde{Z}_i']^{-1} \mathbb{E}[\tilde{Z}_i X_i'] \right\}^{-1},$$

where $\tilde{Z}_i = (Z_i', W_i')'$. This leads to the plug-in estimator

$$\hat{\sigma}_\epsilon^2(\hat{\beta}) (\mathbf{X}'\mathbf{H}_{\mathbf{Z},\mathbf{W}}\mathbf{X})^{-1}.$$

Now, since

$$\begin{aligned}
(\mathbf{X}'\mathbf{H}_{\mathbf{Z},\mathbf{W}}\mathbf{X})^{-1} &= \begin{pmatrix} \mathbf{T}'\mathbf{H}_{\mathbf{Z},\mathbf{W}}\mathbf{T} & \mathbf{T}'\mathbf{W} \\ \mathbf{W}'\mathbf{T} & \mathbf{W}'\mathbf{W} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} (\mathbf{T}'\mathbf{H}_{\mathbf{Z}_\perp}\mathbf{T})^{-1} & -(\mathbf{T}'\mathbf{H}_{\mathbf{Z}_\perp}\mathbf{T})^{-1}\mathbf{T}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1} \\ -(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{T}(\mathbf{T}'\mathbf{H}_{\mathbf{Z}_\perp}\mathbf{T})^{-1} & (\mathbf{W}'\mathbf{W})^{-1} + (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{T}(\mathbf{T}'\mathbf{H}_{\mathbf{Z}_\perp}\mathbf{T})^{-1}\mathbf{T}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1} \end{pmatrix},
\end{aligned}$$

the expression for the `tsls`, `liml`, and `mbtsls` standard errors follows.

`ivreg` uses the Stata estimator of the variance of `jive` (Poi, 2006), given by

$$\hat{\sigma}_\epsilon^2(\hat{\beta}) (\hat{\mathbf{X}}'\mathbf{X})^{-1} \hat{\mathbf{X}}'\hat{\mathbf{X}} (\mathbf{X}'\hat{\mathbf{X}})^{-1},$$

where $\hat{\mathbf{X}} = (\hat{\mathbf{T}}, \mathbf{W})$, $\hat{\mathbf{T}} = (\mathbf{I}_n - \mathbf{D}_{\mathbf{Z},\mathbf{W}})^{-1}(\mathbf{H}_{\mathbf{Z},\mathbf{W}} - \mathbf{D}_{\mathbf{Z},\mathbf{W}})\mathbf{T}$. Therefore,

$$(\hat{\mathbf{X}}'\mathbf{X})^{-1} = \begin{pmatrix} (\hat{\mathbf{P}}'\mathbf{T})^{-1} & -(\hat{\mathbf{P}}'\mathbf{T})^{-1}\hat{\mathbf{T}}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1} \\ -(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{T}(\hat{\mathbf{P}}'\mathbf{T})^{-1} & (\mathbf{W}'\mathbf{W})^{-1} + (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{T}(\hat{\mathbf{P}}'\mathbf{T})^{-1}\hat{\mathbf{T}}'\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1} \end{pmatrix},$$

where $\hat{\mathbf{P}} = \mathbf{M}_\mathbf{W}\hat{\mathbf{T}} = \hat{\mathbf{P}}_{\texttt{jive}}$ so that the (1,1) element of the estimator of the variance evaluates as

$$\hat{\sigma}_\epsilon^2(\hat{\beta}) (\hat{\mathbf{P}}'\mathbf{T})^{-1} (\hat{\mathbf{T}}'\mathbf{M}_\mathbf{W}\hat{\mathbf{T}}) (\mathbf{T}'\hat{\mathbf{P}})^{-1}.$$

The expression for the `jive` standard error estimator follows.

Heteroscedasticity Without the additional homoscedasticity restrictions on the structural error, the asymptotic variance of the `ols` estimator is given by $\mathbb{E}[X_i X_i']^{-1} \mathbb{E}[\epsilon_i^2 X_i X_i'] \mathbb{E}[X_i X_i']^{-1}$, and the standard White estimator of the `ols` variance is given by

$$(\mathbf{X}'\mathbf{X})^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2 X_i X_i' (\mathbf{X}'\mathbf{X})^{-1}$$

[2]see Wooldridge (2002, Equation 5.24) and Davidson and MacKinnon (1993).

A robust estimator of the variance of `tsls`, `liml`, and `mbtsls` is given by (Wooldridge, 2002, Equation 5.34)

$$(\hat{\mathbf{X}}'\hat{\mathbf{X}}')^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i^2(\hat{\mathbf{X}})_i(\hat{\mathbf{X}})_i'(\hat{\mathbf{X}}'\hat{\mathbf{X}}')^{-1},$$

where $\hat{\mathbf{X}} = (\mathbf{H}_{\mathbf{Z},\mathbf{W}}\mathbf{T}, \mathbf{W})$. For `jive`, I use the Stata implementation (Poi, 2006)

$$(\hat{\mathbf{X}}'\mathbf{X})^{-1}\sum_{i}\hat{\epsilon}_i^2\hat{\mathbf{X}}_i\hat{\mathbf{X}}_i'(\mathbf{X}'\hat{\mathbf{X}})^{-1},$$

wherewhere $\hat{\mathbf{X}} = (\hat{\mathbf{T}}, \mathbf{W})$, $\hat{\mathbf{T}} = (\mathbf{I}_n - \mathbf{D}_{\mathbf{Z},\mathbf{W}})^{-1}(\mathbf{H}_{\mathbf{Z},\mathbf{W}} - \mathbf{D}_{\mathbf{Z},\mathbf{W}})\mathbf{T}$.All of these estimators have the form

$$(\hat{\mathbf{X}}'\mathbf{X})^{-1}\sum_{i}\hat{\epsilon}_i^2\hat{\mathbf{X}}_i\hat{\mathbf{X}}_i'(\mathbf{X}'\hat{\mathbf{X}})^{-1}, \tag{13}$$

where $\hat{\mathbf{X}} = (\hat{\mathbf{T}}, \mathbf{W})$ for some $\hat{\mathbf{T}}$. The (1,1) element of (13) can be written as

$$\begin{aligned}(\mathbf{T}'\mathbf{M}_{\mathbf{W}}\hat{\mathbf{T}})^{-1}&\left[\sum_{i=1}^{n}\hat{\epsilon}_i^2\hat{\mathbf{T}}_i^2 - \sum_{i=1}^{n}\hat{\epsilon}_i^2\hat{\mathbf{T}}_iW_i'\hat{\phi} - \hat{\phi}'\sum_{i=1}^{n}\hat{\epsilon}_i^2W_i\hat{\mathbf{T}}_i + \hat{\phi}'\sum_{i=1}^{n}\hat{\epsilon}_i^2W_iW_i'\hat{\phi}\right](\hat{\mathbf{T}}'\mathbf{M}_{\mathbf{W}}\mathbf{T})^{-1}\\ &= (\mathbf{T}_\perp'\mathbf{T}_\perp)^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i^2\left(\hat{\mathbf{T}}_i - W_i'\hat{\phi}\right)^2(\mathbf{T}_\perp'\mathbf{T}_\perp)^{-1},\end{aligned}$$

where $\hat{\phi} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\hat{\mathbf{T}}$. Since $(\mathbf{M}_{\hat{\mathbf{T}}})_i = \hat{\mathbf{T}}_i - W_i'\hat{\phi}$, the expressions for the standard errors given in the test follow.

References

ANATOLYEV, S. (2011). Instrumental variables estimation and inference in the presence of many exogenous regressors. 2, 3

ANDERSON, T. W. and RUBIN, H. (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *The Annals of Mathematical Statistics*, **20** (1), 46–63. 2

ANGRIST, J. D., IMBENS, G. W. and KRUEGER, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, **14** (1), 57–67. 2, 3

BAUM, C. F., SCHAFFER, M. E. and STILLMAN, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *The Stata Journal*, **7** (4), 465–506. 3

BEKKER, P. A. (1994). Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica*, **62** (3), 657–681. 3

CHETTY, R., FRIEDMAN, J. N., HILGER, N., SAEZ, E., SCHANZENBACH, D. W. and YAGAN, D. (2011). How does your kindergarten classroom affect your earnings? *Quarterly Journal of Economics*, **126** (4), 1593–1660. 3

CRAGG, J. G. and DONALD, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, **9** (2), 222–240. 5

DAVIDSON, R. and MACKINNON, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford: Oxford University Press. 7

KOLESÁR, M. (2012a). Estimation in instrumental variables models with heterogeneous treatment effects. 2, 5, 6

KOLESÁR, M. (2012b). Random-Effects Approach to Inference With Many Instruments. 4, 5

Kolesár, M., Chetty, R., Friedman, J. N., Glaeser, E. and Imbens, G. W. (2011). Identification and Inference with Many Invalid Instruments. 2, 3

Phillips, G. D. A. and Hale, C. (1977). The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems. *International Economic Review*, **18** (1), 219–228. 2

Poi, B. P. (2006). Jackknife instrumental variables estimation in Stata. *The Stata Journal*, **6** (3), 364–376. 7, 8

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, **26** (3), 393–415. 5

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press. 7, 8