# Robust Inference for the Two-Sample 2SLS Estimator

David Pacini
Frank Windmeijer

Discussion Paper 16 / 676

20 June 2016

Department of Economics
University of Bristol
8 Woodland Road
Bristol BS8 1TN
United Kingdom

# Robust Inference for the Two-Sample 2SLS Estimator*

David Pacini[a,c], Frank Windmeijer[a,b,c†]

[a]Department of Economics
University of Bristol, UK
[b]IEU, Bristol, UK
[c]Cemmap, London, UK

June 2016

**Abstract**

The Two-Sample Two-Stage Least Squares (TS2SLS) data combination estimator is a popular estimator for the parameters in linear models when not all variables are observed jointly in one single data set. Although the limiting normal distribution has been established, the asymptotic variance formula has only been stated explicitly in the literature for the case of conditional homoskedasticity. By using the fact that the TS2SLS estimator is a function of reduced form and first-stage OLS estimators, we derive the variance of the limiting normal distribution under conditional heteroskedasticity. A robust variance estimator is obtained, which generalises to cases with more general patterns of variable (non-)availability. Stata code and some Monte Carlo results are provided in an Appendix. Stata code for a nonlinear GMM estimator that is identical to the TS2SLS estimator in just identified models and asymptotically equivalent to the TS2SLS estimator in overidentified models is also provided there.

**JEL Classification:** C12, C13, C26
**Key Words:** Linear Model, Data Combination, Instrumental Variables, Robust Inference, Nonlinear GMM

*We would like to thank Helmut Farbmacher, Costas Meghir, Tom Palmer, Mark Schaffer, Jon Temple and Kate Tilling for helpful comments. Windmeijer acknowledges funding by the Medical Research Council, grant no. MC_UU_12013/9

†Corresponding author: Department of Economics, University of Bristol, UK. E-mail address: f.windmeijer@bristol.ac.uk.

# 1 Introduction

The Two-Sample Two-Stage Least Squares (TS2SLS) estimator was introduced by Klevmarken (1982) and applies in cases where one wants to estimate the effects of possibly endogenous explanatory variables $x$ on outcome $y$, but where $y$ and $x$ are not observed in the same data set. Instead, one has observations on outcomes $y$ and instruments $z$ in one sample (sample 1) and on $x$ and $z$ in another (sample 2). Related Two-Sample IV (TSIV) estimators were proposed by Arellano and Meghir (1992) and Angrist and Krueger (1992). Furthermore, Angrist and Krueger (1995) proposed the TS2SLS estimator as a Split-Sample IV (SSIV) estimator. Inoue and Solon (2010) show that the TS2SLS estimator is more efficient than the TSIV estimator of Angrist and Krueger (1992). For further details, see Angrist and Pischke (2009) and the review of Ridder and Moffitt (2007).

This type of data combination estimation method is popular in economics. It is for example used in research on intergenerational mobility, as earnings of different generations are often not observed in the same data set, see the extensive list of references in Jerrim, Choi and Rodriguez (2014). A further recent application is Van den Berg, Pinger and Schoch (2015), who investigate the effect of early-life hunger on late-life health and use the two-sample IV approach to deal with imperfect recollection of conditions early in life. Pierce and Burgess (2013) propose the use of the TS2SLS estimator in epidemiology, in particular when estimating the causal relationship between an exposure and an outcome using genetic factors as instrumental variables, so-called Mendelian randomisation, and where obtaining complete exposure data may be difficult due to high measurement costs.

Under certain assumptions, as stated below, the TS2SLS estimator is consistent and has a limiting normal distribution, see e.g. Klevmarken (1982) and Inoue and Solon (2010). Here we derive the limiting distribution of the TS2SLS estimator under general, unspecified, forms of conditional heteroskedasticity. As the TS2SLS estimator is a simple function of the reduced form parameters for $y$ in sample 1, and the first-stage parameters for $x$ in sample 2, its asymptotic variance is a function of the variances and covariances of these OLS estimators.

The variance of the limiting normal distribution of the TS2SLS estimator is given in (10) below and the formula for a robust estimator of the asymptotic variance is presented in (12). Neither of these have been derived and/or proposed in the literature before. The result in Inoue and Solon (2010) for the conditionally homoskedastic case is similar to our result for that case. They derive the limiting variance of the TS2SLS estimator from the optimal nonlinear GMM estimator. For overidentified models, these two estimators are not the same, but they have the same limiting distribution. Inoue and Solon (2010) did not derive the limiting robust variance for this GMM estimator, but did derive the limiting variance of the efficient two-step GMM estimator under general forms of conditional heteroskedasticity in Inoue and Solon (2005), which is also the approach presented in Arellano and Meghir (1992). Our derivation is different as we focus solely on the TS2SLS estimator as defined below in (5). For the conditional homoskedastic case, our variance estimator differs from the one proposed by Inoue and Solon (2010), as it uses the information from the two samples differently.

Applied researchers have constructed robust standard errors for the just-identified single endogenous regressor case by means of the delta method, see e.g. Dee and Evans (2003). Our result can be seen as a generalisation of this method to situations with multiple regressors and overidentification. Although we consider here a simple cross-sectional setup, other sampling designs can be accommodated and the result is straightforwardly extended to compute, for example, cluster-robust standard errors.

Our result also generalises to situations outside the standard TS2SLS setup. For example, it can accommodate a model with three explanatory variables where one endogenous variable is observed with the outcome variable in sample 1, but not in sample 2, one explanatory variable is only observed in sample 2 and one endogenous variable is observed in both samples 1 and 2. This is discussed in Section 5 below and we present Stata code for this example and for the standard TS2SLS setup in the Appendix.

In the next section we present the model, assumptions and the TS2SLS estimator. In Section 3, we present our main results. Section 4 compares our results to those derived for nonlinear GMM. The Appendix also presents Stata code for the GMM estimator.

# 2 Model, Assumptions and TS2SLS Estimator

The structural linear model of interest is given by

$$y_i = x_i'\beta + \varepsilon_i, \tag{1}$$

but we cannot estimate this model as $y_i$ and $x_i$ are not jointly observed. Instead, we have two independent samples. In sample 1 we have observations on $y$ and $k_z$ exogenous instruments $z$. Sample 2 contains observations on the $k_x$ explanatory variables $x$ and $z$. Denoting by subscripts 1 and 2 whether the variables are observed in sample 1 or sample 2, in the first sample we observe $\{y_{1i}, z'_{1i}\}$ for $i = 1, ..., n_1$, and in the second sample we observe $\left\{x'_{2j}, z'_{2j}\right\}$ for $j = 1, ..., n_2$. Throughout we assume that $k_z \geq k_x$. Other explanatory variables that enter model (1), but that are observed in both samples and are exogenous, including the constant, have been partialled out.

The TS2SLS estimator is derived as follows. From the information in sample 1, we can estimate the reduced form model for $y_{1i}$, given by

$$y_{1i} = z'_{1i}\pi_{y1} + u_{1i}. \tag{2}$$

From sample 2, we can estimate the linear projections

$$x_{2j} = \Pi'_{x2} z_{2j} + v_{2j}, \tag{3}$$

with $\Pi_{x2} = E\left(z_{2j}z'_{2j}\right)^{-1} E\left(z_{2j}x'_{2j}\right)$, a $k_z \times k_x$ matrix of rank $k_x$ by assumption. As (3) is a linear projection, it follows that $E\left(z_{2j}v'_{2j}\right) = 0$. Although the $x_{1i}$ are not observed, the data generating process for $y_{1i}$ is given by the structural model (1) and hence it and its reduced form are given by

$$\begin{aligned} y_{1i} &= x'_{1i}\beta + \varepsilon_{1i} = \left(z'_{1i}\Pi_{x1} + v'_{1i}\right)\beta + \varepsilon_{1i} \\ &= z'_{1i}\Pi_{x1}\beta + \varepsilon_{1i} + v'_{1i}\beta, \end{aligned} \tag{4}$$

with the linear projection parameters $\Pi_{x1} = E\left(z_{1i}z'_{1i}\right)^{-1} E\left(z_{1i}x'_{1i}\right)$. Again, $E\left(z_{1i}v'_{i1}\right) = 0$. From (2) and (4) it follows that $\pi_{y1} = \Pi_{x1}\beta$ and $u_{1i} = \varepsilon_{1i} + v'_{1i}\beta$. Clearly, knowledge of $\pi_{y1}$ and $\Pi_{x1}$ identifies the structural parameters $\beta$, and the standard 2SLS estimator in a sample with $y_{1i}$, $x_{1i}$ and $z_{1i}$ all observed combines the information contained in the OLS estimators for $\pi_{y1}$ and $\Pi_{x1}$, denoted by $\widehat{\pi}_{y1}$ and $\widehat{\Pi}_{x1}$ as follows

$$\widehat{\beta}_{2sls} = \left(\widehat{\Pi}'_{x1}Z'_1Z_1\widehat{\Pi}_{x1}\right)^{-1}\widehat{\Pi}'_{x1}Z'_1Z_1\widehat{\pi}_{y1},$$

with $Z_1$ the $n_1 \times k_z$ matrix $[z'_{1i}]$.

As $x_{1i}$ is not observed, we cannot estimate $\Pi_{x1}$, but we can estimate $\Pi_{x2}$ using the second sample. Denoting the OLS estimator for $\Pi_{x2}$ by $\widehat{\Pi}_{x2}$, the Two-Sample 2SLS estimator is given by

$$\begin{aligned}
\widehat{\beta}_{ts2sls} &= \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\widehat{X}_1'y_1 = \left(\widehat{\Pi}_{x2}'Z_1'Z_1\widehat{\Pi}_{x2}\right)^{-1}\widehat{\Pi}_{x2}'Z_1'y_1 \\
&= \left(\widehat{\Pi}_{x2}'Z_1'Z_1\widehat{\Pi}_{x2}\right)^{-1}\widehat{\Pi}_{x2}'Z_1'Z_1\widehat{\pi}_{y1}. \qquad (5)
\end{aligned}$$

We make the following assumptions:

A1: $\{y_{1i}, z_{1i}'\}_{i=1}^{n_1}$ and $\left\{x_{2j}', z_{2j}'\right\}_{j=1}^{n_2}$ are i.i.d. random samples from the same population with finite fourth moments and are independent.
A2: $E\left(z_{1i}z_{1i}'\right) = Q_{zz1}$; $E\left(z_{2j}z_{2j}'\right) = Q_{zz2}$. $Q_{zz1}$ and $Q_{zz2}$ are nonsingular.
A3: $E\left(z_{1i}x_{1i}'\right)$ and $E\left(z_{2i}x_{2i}'\right)$ both have rank $k_x$.
A4: $E\left(z_{1i}\varepsilon_{1i}\right) = 0$.
A5: $E\left(u_{1i}^2 z_{1i}z_{i1}'\right) = \Omega_{y1}$, a finite and positive definite matrix.
A6: $E\left[\left(I_{k_x}\otimes z_{2j}\right)v_{2j}v_{2j}'\left(I_{k_x}\otimes z_{2j}'\right)\right] = E\left(v_{2j}v_{2j}'\otimes z_{2j}z_{2j}'\right) = \Omega_{x2}$, a finite and positive definite matrix. $I_{k_x}$ is the identity matrix of order $k_x$.
A7: $\lim_{n_1\to\infty, n_2\to\infty}\frac{n_1}{n_2} = \alpha$ for some $\alpha > 0$.

Assumptions A1-A3 and A7 are standard data combination assumptions, see e.g. Inoue and Solon (2010). Assumptions A2 and A3, combined with A1, result in $E\left(z_{1i}z_{1i}'\right) = E\left(z_{2j}z_{2j}'\right)$ and $E\left(z_{1i}x_{1i}'\right) = E\left(z_{2i}x_{2i}'\right)$, and hence $\Pi_{x1} = \Pi_{x2}$. A1-A3 are clearly sufficient, but not necessary conditions for $\Pi_{x1}$ to be equal to $\Pi_{x2}$. The condition $\Pi_{x1} = \Pi_{x2}$ itself is sufficient for consistency of $\widehat{\beta}_{ts2sls}$, and necessary for the limiting normal distribution of $\sqrt{n_1}\left(\widehat{\beta}_{ts2sls} - \beta\right)$ to have a mean of zero. In the derivations below we do not (need to) impose $Q_{zz1} = Q_{zz2}$. The resulting estimator of the variance of $\widehat{\beta}_{ts2sls}$ is a simple function of the variances of $\widehat{\pi}_{y1}$ and $\operatorname{vec}\left(\widehat{\Pi}_{x2}\right)$, and this function is unambiguous about which information from which sample is being utilised.

Assumptions A5 and A6 explicitly allow for general forms of heteroskedasticity. The robust variance estimator for $\widehat{\beta}_{ts2sls}$ is obtained incorporating robust variance estimators for $\widehat{\pi}_{y1}$ and $\operatorname{vec}\left(\widehat{\Pi}_{x2}\right)$. This was done by Dee and Evans (2003) using the delta method for the just identified single regressor case, i.e. $k_x = k_z = 1$. The result derived below can be seen as a generalisation of this to multiple regressors and overidentified settings.

# 3 Limiting Distribution and Variance Estimator

The OLS estimators for $\pi_{y1}$ and $\Pi_{x2}$ are given by

$$\begin{aligned}
\widehat{\pi}_{y1} &= \left(Z_1'Z_1\right)^{-1} Z_1'y_1 \\
\widehat{\Pi}_{x2} &= \left(Z_2'Z_2\right)^{-1} Z_2'X_2,
\end{aligned}$$

with $Z_1$ the $n_1 \times k_z$ matrix $[z_{1i}']$; $Z_2$ the $n_2 \times k_z$ matrix $\left[z_{2j}'\right]$; $y_1$ the $n_1$ vector $(y_{1i})$ and $X_2$ the $n_2 \times k_x$ matrix $\left[x_{2j}'\right]$. Under Assumptions A1-A4 and A7 we obtain

$$\begin{aligned}
\operatorname{plim}\left(\widehat{\pi}_{y1}\right) &= E\left(z_{1i}z_{1i}'\right)^{-1} E\left(z_{1i}x_{1i}'\right)\beta \\
&= \pi_{y1} = \Pi_{x1}\beta = \Pi_{x2}\beta; \\
\operatorname{plim}\left(\widehat{\Pi}_{x2}\right) &= E\left(z_{2j}z_{2j}'\right)^{-1} E\left(z_{2j}x_{2j}'\right) = \Pi_{x2},
\end{aligned}$$

and hence the TS2SLS estimator is consistent as

$$\begin{aligned}
\operatorname{plim}\left(\widehat{\beta}_{ts2sls}\right) &= \operatorname{plim}\left(\frac{1}{n_1}\widehat{\Pi}_{x2}'Z_1'Z_1\widehat{\Pi}_{x2}\right)^{-1}\frac{1}{n_1}\widehat{\Pi}_{x2}'Z_1'Z_1\widehat{\pi}_{y1} \\
&= \left(\Pi_{x2}'Q_{zz1}\Pi_{x2}\right)^{-1}\Pi_{x2}'Q_{zz1}\pi_{y1} = \beta. \qquad (6)
\end{aligned}$$

Note that the probability limits obtained here and the limiting distributions derived below are for both $n_1 \to \infty$ and $n_2 \to \infty$.

For the derivation of the asymptotic distribution of $\widehat{\beta}_{ts2sls}$, denote $\pi_{x2} = \operatorname{vec}\left(\Pi_{x2}\right)$; $\widehat{\pi}_{x2} = \operatorname{vec}\left(\widehat{\Pi}_{x2}\right)$; $\theta = \begin{pmatrix} \pi_{y1}' & \pi_{x2}' \end{pmatrix}'$ and $\widehat{\theta} = \begin{pmatrix} \widehat{\pi}_{y1}' & \widehat{\pi}_{x2}' \end{pmatrix}'$. Under Assumptions A1-A7

$$\sqrt{n_1}\left(\widehat{\pi}_{y1} - \pi_{y1}\right) \xrightarrow{d} N\left(0, V_{\pi_{y1}}\right); \qquad (7)$$

$$\sqrt{n_2}\left(\widehat{\pi}_{x2} - \pi_{x2}\right) \xrightarrow{d} N\left(0, V_{\pi_{x2}}\right), \qquad (8)$$

where

$$\begin{aligned}
V_{\pi_{y1}} &= Q_{zz1}^{-1}\Omega_{y1}Q_{zz1}^{-1}; \\
V_{\pi_{x2}} &= \left(I_{k_x} \otimes Q_{zz2}^{-1}\right)\Omega_{x2}\left(I_{k_x} \otimes Q_{zz2}^{-1}\right).
\end{aligned}$$

Hence

$$\sqrt{n_1}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} N\left(0, V_\theta\right), \qquad (9)$$

with

$$V_\theta = \begin{bmatrix} V_{\pi_{y1}} & 0 \\ 0 & \alpha V_{\pi_{x2}} \end{bmatrix}.$$

From the limiting distribution of $\widehat{\theta}$, the limiting distribution of $\widehat{\beta}_{ts2sls}$ is readily obtained and we give a simple proof in the Appendix. Our main result is:

Under Assumptions A1-A7, the limiting distribution of $\widehat{\beta}_{ts2sls}$ is given by

$$\sqrt{n_1}\left(\widehat{\beta}_{ts2sls}-\beta\right)\xrightarrow{d}N\left(0,V_\beta\right);$$

$$\begin{aligned} V_\beta &= C\left(V_{\pi_{y1}}+\alpha\left(\beta'\otimes I_{k_z}\right)V_{\pi_{x2}}\left(\beta\otimes I_{k_z}\right)\right)C' \\ &= CV_{\pi_{y1}}C'+\alpha\left(\beta'\otimes C\right)V_{\pi_{x2}}\left(\beta\otimes C'\right), \end{aligned} \tag{10}$$

where

$$C=\left(\Pi_{x2}'Q_{zz1}\Pi_{x2}\right)^{-1}\Pi_{x2}'Q_{zz1}. \tag{11}$$

We can obtain an estimator for the asymptotic variance of $\widehat{\beta}_{ts2sls}$ as follows. Let $V\widehat{a}r\left(\widehat{\pi}_{y1}\right)$ and $V\widehat{a}r\left(\widehat{\pi}_{x2}\right)$ be estimators of the asymptotic variances of $\widehat{\pi}_{y1}$ and $\widehat{\pi}_{x2}$, in the sense that $\operatorname{plim}\left(n_1V\widehat{a}r\left(\widehat{\pi}_{y1}\right)\right)=V_{\pi_{y1}}$ and $\operatorname{plim}\left(n_2V\widehat{a}r\left(\widehat{\pi}_{x2}\right)\right)=V_{\pi_{x2}}$. Let $\widehat{C}$ be the matrix of least squares coefficients from the regressions of the columns of $Z_1$ on $\widehat{X}_1$. As $\operatorname{plim}\left(\widehat{C}\right)=\operatorname{plim}\left(\left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\widehat{X}_1'Z_1\right)=C$, an estimator of the asymptotic variance of $\widehat{\beta}_{ts2sls}$ is given by

$$V\widehat{a}r\left(\widehat{\beta}_{ts2sls}\right)=\widehat{C}V\widehat{a}r\left(\widehat{\pi}_{y1}\right)\widehat{C}'+\left(\widehat{\beta}_{ts2sls}'\otimes\widehat{C}\right)V\widehat{a}r\left(\widehat{\pi}_{x2}\right)\left(\widehat{\beta}_{ts2sls}\otimes\widehat{C}'\right), \tag{12}$$

as

$$\begin{aligned} n_1V\widehat{a}r\left(\widehat{\beta}_{ts2sls}\right) &= \widehat{C}\left(n_1V\widehat{a}r\left(\widehat{\pi}_{y1}\right)\right)\widehat{C}'+\frac{n_1}{n_2}\left(\widehat{\beta}_{ts2sls}'\otimes\widehat{C}\right)\left(n_2V\widehat{a}r\left(\widehat{\pi}_{x2}\right)\right)\left(\widehat{\beta}_{ts2sls}\otimes\widehat{C}'\right) \\ &\xrightarrow{p} V_\beta. \end{aligned}$$

When the model is just identified, $k_z=k_x$, then $\widehat{C}=\widehat{\Pi}_{x2}^{-1}$. When furthermore $k_x=k_z=1$, (12) reduces to the simple expression

$$V\widehat{a}r\left(\widehat{\beta}_{ts2sls}\right)=\left(V\widehat{a}r\left(\widehat{\pi}_{y1}\right)+\widehat{\beta}_{ts2sls}^2V\widehat{a}r\left(\widehat{\pi}_{x2}\right)\right)/\widehat{\pi}_{x2}^2,$$

with $\widehat{\beta}_{ts2sls}=\frac{\widehat{\pi}_{y1}}{\widehat{\pi}_{x2}}$, which is identical to the expression obtained using the delta method as in Dee and Evans (2003).

Specifying $V\widehat{a}r\left(\widehat{\pi}_{y1}\right)$ and $V\widehat{a}r\left(\widehat{\pi}_{x2}\right)$ in (12) as being robust to general forms of heteroskedasticity results in a robust variance estimator for $\widehat{\beta}_{ts2sls}$. A small Monte Carlo exercise reported in the Appendix confirms that our asymptotic results reflect the behaviour of the TS2SLS estimator. Although we have here an i.i.d. cross-sectional setup, the results generalise to e.g. cluster-robust variances straightforwardly.

# 4 GMM

Assuming conditional homoskedasticity for both $u_{1i}$ and $v_{2j}$ such that

$$E\left(u_{1i}^2|z_{1i}\right)=\sigma_u^2 \quad \text{and} \quad E\left(v_{2j}v_{2j}'|z_{2j}\right)=\Sigma_v,$$

we have that

$$V_{\pi_{y1}}=\sigma_u^2 Q_{zz1} \quad \text{and} \quad V_{\pi_{x2}}=\Sigma_v\otimes Q_{zz2}^{-1},$$

and hence

$$V_\beta=\sigma_u^2\left(\Pi_{x2}'Q_{zz1}\Pi_{x2}\right)^{-1}+\alpha\beta'\Sigma_v\beta C Q_{zz2}^{-1}C'.$$

The variance estimator (12) is then

$$V\widehat{a}r\left(\widehat{\beta}_{ts2sls}\right)=\widehat{\sigma}_u^2\left(\widehat{X}_1'\widehat{X}_1\right)^{-1}+\widehat{\beta}_{ts2sls}'\widehat{\Sigma}_v\widehat{\beta}_{ts2sls}\widehat{C}\left(Z_2'Z_2\right)^{-1}\widehat{C}', \tag{13}$$

with $\widehat{\sigma}_u^2=\left(y_1-Z_1\widehat{\pi}_{y1}\right)'\left(y_1-Z_1\widehat{\pi}_{y1}\right)/n_1$ and $\widehat{\Sigma}_v=\left(X_2-Z_2\widehat{\Pi}_{x2}\right)'\left(X_2-Z_2\widehat{\Pi}_{x2}\right)/n_2$.

Inoue and Solon (2010) derive $V_\beta$ from the limiting distribution of the optimal GMM estimator using moment conditions

$$E\left[z_{1i}\left(y_{1i}-z_{1i}'\Pi_{x2}\beta\right)\right] = 0; \tag{14}$$

$$E\left[z_{2j}\otimes\left(x_{2j}-\Pi_{x2}'z_{2j}\right)\right] = 0, \tag{15}$$

and weight matrix

$$\begin{bmatrix} V\widehat{a}r\left(\widehat{\pi}_{y1}\right) & 0 \\ 0 & V\widehat{a}r\left(\widehat{\pi}_{x2}\right)\end{bmatrix}=\begin{bmatrix}\widehat{\sigma}_u^2\left(Z_1'Z_1\right)^{-1} & 0 \\ 0 & \widehat{\Sigma}_v\otimes\left(Z_2'Z_2\right)^{-1}\end{bmatrix}.$$

Let $\psi=\begin{pmatrix}\beta' & \pi_{x2}'\end{pmatrix}'$, then this GMM estimator is the same as the minimum distance estimator

$$\widetilde{\psi}=\arg\min_{\beta,\pi_{x2}}\begin{pmatrix}\widehat{\pi}_{y1}-\Pi_{x2}\beta \\ \widehat{\pi}_{x2}-\pi_{x2}\end{pmatrix}'\begin{bmatrix}\left(V\widehat{a}r\left(\widehat{\pi}_{y1}\right)\right)^{-1} & 0 \\ 0 & \left(V\widehat{a}r\left(\widehat{\pi}_{x2}\right)\right)^{-1}\end{bmatrix}\begin{pmatrix}\widehat{\pi}_{y1}-\Pi_{x2}\beta \\ \widehat{\pi}_{x2}-\pi_{x2}\end{pmatrix}.$$

Unless the model is just identified, $\widetilde{\beta}\neq\widehat{\beta}_{ts2sls}$, but their limiting distributions are the same. This is a situation similar to that of the LIML and 2SLS estimators in the standard

IV model. When the model is overidentified, the TS2SLS estimator itself cannot be obtained as a GMM estimator. The limiting variance of $\sqrt{n_1}\left(\widetilde{\beta}-\beta\right)$ is obtained from the limiting variance of $\sqrt{n_1}\left(\widetilde{\psi}-\psi\right)$. Inoue and Solon (2010) imposed $Q_{zz1}=Q_{zz2}$ and obtained the variance as

$$V_{\beta,IS}=\left(\sigma_u^2+\alpha\beta'\Sigma_v\beta\right)\left(\Pi_{x2}'Q_{zz1}\Pi_{x2}\right)^{-1}$$

and their variance estimator is given by

$$V\widehat{a}r_{IS}\left(\widehat{\beta}_{ts2sls}\right)=\left(\widetilde{\sigma}_u^2+\frac{n_1}{n_2}\widehat{\beta}_{ts2sls}'\widehat{\Sigma}_v\widehat{\beta}_{ts2sls}\right)\left(\widehat{X}_1'\widehat{X}_1\right)^{-1},$$

where $\widetilde{\sigma}_u^2=\left(y_1-\widehat{X}_1\widehat{\beta}_{ts2sls}\right)'\left(y_1-\widehat{X}_1\widehat{\beta}_{ts2sls}\right)/n_1$. Apart from this difference in the estimation of $\sigma_u^2$, the main difference is the imposition that $Q_{zz1}=Q_{zz2}$. Although this is justified asymptotically given the assumptions A1-A3, the finite sample variance of $\widehat{\pi}_{x2}$ in (12) is clearly more naturally estimated by $\widehat{\Sigma}_v\otimes\left(Z_2'Z_2\right)^{-1}$ than by $\widehat{\Sigma}_v\otimes\left(\frac{n_2}{n_1}Z_1'Z_1\right)^{-1}$. Also, for the example in footnotes 3 and 2 in Inoue and Solon (2010) and (2005) respectively, when $E\left(z_{1i}x_{1i}'\right)=cE\left(z_{2j}x_{2j}\right)'$ and $E\left(z_{1i}z_{1i}'\right)=cE\left(z_{2j}z_{2j}\right)'$, with $c\neq 1$, then the TS2SLS estimator is consistent and asymptotically normally distributed but $n_1V\widehat{a}r_{IS}\left(\widehat{\beta}_{ts2sls}\right)$ is no longer a consistent estimator of the variance of the limiting distribution, whereas $n_1V\widehat{a}r\left(\widehat{\beta}_{ts2sls}\right)$ is.

Inoue and Solon (2010) did not derive the robust variance of $\widetilde{\beta}$. Although this can be obtained from the robust variance of $\widetilde{\psi}$, the matrix expressions involved are quite cumbersome. Arellano and Meghir (1992) similarly considered the robust variance of the GMM estimator $\widetilde{\psi}$ but also did not derive a variance estimator for $\widetilde{\beta}$ separately. One can of course simply obtain robust standard errors for $\widetilde{\psi}$ and hence $\widetilde{\beta}$ using GMM routines that can estimate the parameters using the nonlinear and linear moment conditions (14) and (15). These estimates are then obtained using iterative methods, and for just-identified models this produces the TS2SLS estimator with robust standard errors. For overidentified models, the efficient two-step GMM estimator for $\psi$ can then also be obtained together with a Hansen test for the validity of the moment conditions. We present Stata code for this GMM estimation procedure in the Appendix.

# 5 Generalising the Result

Although we derived the results in Section 3 for the standard TS2SLS estimator, the limiting distribution results (17) and (18) in the Appendix apply more generally. Indeed, the only aspect in $V_\theta$ that is particular to this specific two-sample setup is the zero covariance between $\widehat{\pi}_{y1}$ and $\widehat{\pi}_{x2}$, due to the samples being independent.

Consider as a generalisation a model with three explanatory variables $x_1$, $x_2$ and $x_3$. Using the same notational convention as before, in sample 1 we observe $\{y_{1i}, x_{11i}, x_{31i}, z'_{1i}\}_{i=1}^{n_1}$. In sample 2 we observe $\left\{x_{22j}, x_{32j}, z'_{2j}\right\}_{j=1}^{n_2}$. In this case, $x_1$ is only observed in sample 1, $x_2$ is only observed in sample 2, whereas $x_3$ is observed in both samples. Let $Z = \begin{pmatrix} Z'_1 & Z'_2 \end{pmatrix}'$ and $x_3 = \begin{pmatrix} x'_{31} & x'_{32} \end{pmatrix}'$, then the reduced form and first-stage OLS estimators are given by

$$\begin{aligned} \widehat{\pi}_{y1} &= \left(Z'_1 Z_1\right)^{-1} Z'_1 y_1; \;\; \widehat{\pi}_{x11} = \left(Z'_1 Z_1\right)^{-1} Z'_1 x_{11} \\ \widehat{\pi}_{x22} &= \left(Z'_2 Z_2\right)^{-1} Z'_2 x_{22}; \;\; \widehat{\pi}_{x3} = \left(Z' Z\right)^{-1} Z' x_3. \end{aligned}$$

Let $\widehat{\Pi}_x = \left[ \begin{array}{ccc} \widehat{\pi}_{x11} & \widehat{\pi}_{x22} & \widehat{\pi}_3 \end{array} \right]$, then the two-sample IV estimator is given by

$$\widehat{\beta}_{2s} = \left(\widehat{\Pi}'_x Z'_1 Z_1 \widehat{\Pi}_x\right)^{-1} \widehat{\Pi}'_x Z'_1 Z_1 \widehat{\pi}_{y1}.$$

We differentiate this estimator from the standard two-sample setup above and reserve the name $\widehat{\beta}_{ts2sls}$ for that particular setup. Under Assumptions A1-A7, the limiting distribution is as in (17), but as $\widehat{\theta} = \left( \begin{array}{cc} \widehat{\pi}'_{y1} & \operatorname{vec}\left(\widehat{\Pi}_x\right)' \end{array} \right)$, the variance $V_\theta$ differs from the standard setup as there is a different covariance structure. There are non-zero covariances between $\widehat{\pi}_{y1}$ and $\widehat{\pi}_{x11}$; $\widehat{\pi}_{x11}$ and $\widehat{\pi}_{x3}$; and $\widehat{\pi}_{x11}$ and $\widehat{\pi}_{x3}$, whereas the covariances between $\widehat{\pi}_{y1}$ and $\widehat{\pi}_{x22}$; and $\widehat{\pi}_{x11}$ and $\widehat{\pi}_{x22}$ are zero. From (18), an estimator for the asymptotic variance is given by

$$V\widehat{a}r\left(\widehat{\beta}_{2s}\right) = \left(\widehat{\delta}' \otimes \widehat{C}\right) V\widehat{a}r\left(\widehat{\theta}\right)\left(\widehat{\delta} \otimes \widehat{C}'\right), \tag{16}$$

where $\widehat{\delta} = \begin{pmatrix} 1 & -\widehat{\beta}'_{2s} \end{pmatrix}'$ and $\widehat{C} = \left(\widehat{X}'_1 \widehat{X}_1\right)^{-1} \widehat{X}'_1 Z_1 = \left(\widehat{\Pi}'_x Z'_1 Z_1 \widehat{\Pi}_x\right)^{-1} \widehat{\Pi}'_x Z'_1 Z_1$.

For the standard TS2SLS setup and the more general structures, one can obtain the robust variance estimates using standard routines. We give Stata code for two examples in the Appendix. The structure of the algorithm for the general case is:

1. Estimate the reduced form and first-stage parameters by OLS, obtain the predicted values $\widehat{X}_1$ and a robust variance estimate for $\widehat{\theta} = \begin{pmatrix} \widehat{\pi}'_{y1} & \widehat{\pi}'_{x2} \end{pmatrix}'$, the matrix $V\widehat{a}r\left(\widehat{\theta}\right)$. In Stata, the latter can be obtained using the 'gmm' or the 'suest' routine.
2. Regress $y_1$ on $\widehat{X}_1$ to obtain the TS2SLS estimator.
3. Regress the columns of $Z_1$ on $\widehat{X}_1$ and collect the parameter estimates in the matrix $\widehat{C}$.
4. Calculate $V\widehat{a}r\left(\widehat{\beta}_{2s}\right)$ by the matrix expression in (16).
5. Some adjustments have to be made when parameters on exogenous variables and the constant are included in the estimation. These are detailed in the code in the Appendix.

# 6 Conclusions

In this note, we have derived the variance of the limiting normal distribution of the Two-Sample Two-Stage Least Squares (TS2SLS) estimator under general, unspecified, forms of heteroskedasticity, and have proposed a new robust variance estimator. This estimator is a simple function of the robust variance estimates of the reduced-form and first-stage OLS estimates in the two samples, and only requires linear projections for its calculation. We provide Stata code for the calculation of the TS2SLS estimator and its robust variance estimator in the Appendix. It is also straigthforward to obtain a cluster-robust variance estimator.

Under conditional homoskedasticity, our variance estimator differs from the one derived by Inoue and Solon (2010) in the way it uses the information from the two samples, as, unlike Inoue and Solon (2010), we don't impose $\text{plim}\left(\frac{1}{n_1}Z_1'Z_1\right) = Q_{zz1} = Q_{zz2} = \text{plim}\left(\frac{1}{n_2}Z_2'Z_2\right)$ in the derivation of the variance of the limiting distribution. Inoue and Solon (2010) obtained the variance of the limiting distribution of the TS2SLS estimator from that of an optimal nonlinear GMM estimator. This estimator is identical to the TS2SLS estimator when the model is just identified, and is asymptotically equivalent when the model is overidentified. We provide Stata code for this nonlinear GMM estimator in the Appendix, naturally leading to an efficient two-step GMM estimator and the Hansen test for overidentification. We therefore provide the tools needed for estimation and robust inference for this type of two-sample data combination analysis.

# References


Angrist, J.D., and A.B. Krueger (1992), The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples, *Journal of the American Statistical Association* 87, 328-336.

Angrist, J.D., and A.B. Krueger (1995), Split-sample instrumental variables estimates of the return to schooling, *Journal of Business & Economic Statistics* 13, 225-235.

Angrist, J.D. and J.-S. Pischke (2009), *Mostly Harmless Econometrics. An Empiricist's Companion.* Princeton University Press, Princeton.

Arellano, M., and C. Meghir (1992), Female labour supply and on-the-job search: An empirical model estimated using complementary data sets, *Review of Economic Studies* 59, 537-559.

van den Berg, G.J., P.R. Pinger and J. Schoch (2015), Instrumental variable estimation of the causal effect of hunger early in life on health later in life, *The Economic Journal.* In Press, doi:10.1111/ecoj.12250.

Dee, T.S., and W.N. Evans (2003), Teen drinking and educational attainment: Evidence from two-sample instrumental variables estimates, *Journal of Labour Economics* 21, 178-209.

Inoue, A., and G. Solon (2005), Two-sample instrumental variables estimators, *NBER Technical Working Paper* 311.

Inoue, A., and G. Solon (2010), Two-sample instrumental variables estimators, *Review of Economics and Statistics* 92, 557-561.

Jerrim, J., A. Choi and R.S. Rodriguez (2014), Two-Sample Two-Stage Least Squares (TSTSLS) estimates of earnings mobility: how consistent are they?, Working Paper No. 14-17, Institute of Education, University of London.

Klevmarken, N.A. (1982), Missing variables and two-stage least squares estimation from more than one data set, Working Paper Series No. 62, Research Institute of Industrial Economics, Stockholm, Sweden.

Newey, W.K. (1984), A method of moments interpretation of sequential estimators, *Economics Letters* 14, 201-206.

Pierce, B.L. and S. Burgess (2013), Efficient design for Mendelian randomization studies: Subsample and 2-sample instrumental variables estimators, *American Journal of*

*Epidemiology* 178, 1177-1184.

Ridder, G., and R. Moffitt (2007), The econometrics of data combination, in: J.J. Heckman and H.E. Leamer (eds.), *Handbook of Econometrics* Volume 6, Part B, Chapter 75, 5469-5547.

# A Appendix

## A.1 Derivation of Limiting Distribution

The limiting distribution of $\widehat{\beta}_{ts2sls}$ can be obtained as follows, alternative methods like the delta method or Newey (1984) give identical results. Rewrite the reduced form model for $y_{1i}$ as

$$\begin{aligned} y_{1i} &= z_{1i}'\pi_{y1} + u_{1i} = z_{1i}'\Pi_{x1}\beta + u_{1i} \\ &= z_{1i}'\widehat{\Pi}_{x2}\beta + u_{1i} - z_{1i}'\left(\widehat{\Pi}_{x2} - \Pi_{x1}\right)\beta \\ &= \widehat{x}_{1i}'\beta + u_{1i} - z_{1i}'\left(\widehat{\Pi}_{x2} - \Pi_{x2}\right)\beta, \end{aligned}$$

where the last equality holds as $\Pi_{x2} = \Pi_{x1}$ by the assumptions. Then

$$\begin{aligned} \widehat{\beta}_{ts2sls} &= \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\widehat{X}_1 y_1 \\ &= \beta + \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\widehat{X}_1'\left(u_1 - Z_1\left(\widehat{\Pi}_{x2} - \Pi_{x2}\right)\beta\right), \end{aligned}$$

where $u_1$ is the $n_1$ vector $(u_{1i})$. For the limiting distribution,

$$\sqrt{n_1}\left(\widehat{\beta}_{ts2sls} - \beta\right) = \left(\frac{1}{n_1}\widehat{X}_1'\widehat{X}_1\right)^{-1}\frac{1}{\sqrt{n_1}}\widehat{X}_1'\left(u_1 - Z_1\left(\widehat{\Pi}_{x2} - \Pi_{x2}\right)\beta\right).$$

As

$$\begin{aligned} &\frac{1}{\sqrt{n_1}}\widehat{X}_1'\left(u_1 - Z_1\left(\widehat{\Pi}_{x2} - \Pi_{x2}\right)\beta\right) \\ =\ &\frac{1}{\sqrt{n_1}}\widehat{\Pi}_{x2}'Z_1'\left(u_1 - Z_1\left(\widehat{\Pi}_{x2} - \Pi_{x2}\right)\beta\right) \\ =\ &\frac{1}{\sqrt{n_1}}\widehat{\Pi}_{x2}'Z_1'Z_1\left(\left(Z_1'Z_1\right)^{-1}Z_1'u_1 - \left(\widehat{\Pi}_{x2} - \Pi_{x2}\right)\beta\right) \\ =\ &\widehat{\Pi}_{x2}'\left(\frac{1}{n_1}Z_1'Z_1\right)\left(\sqrt{n_1}\left(\widehat{\pi}_{y1} - \pi_{y1}\right) - \sqrt{n_1}\left(\widehat{\Pi}_{x2} - \Pi_{x2}\right)\beta\right), \end{aligned}$$

and

$$
\begin{aligned}
\left(\widehat{\Pi}_{x2}-\Pi_{x2}\right)\beta &= \operatorname{vec}\left(\left(\widehat{\Pi}_{x2}-\Pi_{x2}\right)\beta\right) \\
&= \left(\beta'\otimes I_{k_z}\right)\operatorname{vec}\left(\widehat{\Pi}_{x2}-\Pi_{x2}\right) \\
&= \left(\beta'\otimes I_{k_z}\right)\left(\widehat{\pi}_{x2}-\pi_{x2}\right),
\end{aligned}
$$

it follows that

$$
\sqrt{n_1}\left(\widehat{\beta}_{ts2sls}-\beta\right)=\left(\frac{1}{n_1}\widehat{X}_1'\widehat{X}_1\right)^{-1}\widehat{\Pi}_{x2}'\left(\frac{1}{n_1}Z_1'Z_1\right)\left(\delta'\otimes I_{k_z}\right)\sqrt{n_1}\left(\widehat{\theta}-\theta\right),
$$

where $\delta=\begin{pmatrix}1 & -\beta'\end{pmatrix}'$.

As

$$
\operatorname{plim}\left(\frac{1}{n_1}\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{\Pi}_{x2}'\left(\frac{1}{n_1}Z_1'Z_1\right)\right)=\left(\Pi_{x2}'Q_{zz1}\Pi_{x2}\right)^{-1}\Pi_{x2}'Q_{zz1}=C,
$$

it follows that

$$
\sqrt{n_1}\left(\widehat{\beta}_{ts2sls}-\beta\right)\xrightarrow{d}N\left(0,V_\beta\right), \tag{17}
$$

where

$$
\begin{aligned}
V_\beta &= C\left(\delta'\otimes I_{k_z}\right)V_\theta\left(\delta\otimes I_{k_z}\right)C' \\
&= \left(\delta'\otimes C\right)V_\theta\left(\delta\otimes C'\right) \qquad (18)\\
&= C\left(V_{\pi_{y1}}+\alpha\left(\beta'\otimes I_{k_z}\right)V_{\pi_{x2}}\left(\beta\otimes I_{k_z}\right)\right)C' \\
&= CV_{\pi_{y1}}C'+\alpha\left(\beta'\otimes C\right)V_{\pi_{x2}}\left(\beta\otimes C'\right).
\end{aligned}
$$

## A.2 Some Monte Carlo Results

We generate data according to the standard setup for the TS2SLS estimator above. The parameters in model (19) are set to $\beta_1=0.3$, $\beta_2=-0.1$, $\beta_w=0.1$ and $\beta_0=0.2$. Further, $w_i\sim N(0,1)$, and

$$
\begin{aligned}
x_1 &= Z\pi_1+w\pi_{w1}+\pi_{01}+v_1 \\
x_2 &= Z\pi_2+w\pi_{w2}+\pi_{02}+v_2
\end{aligned}
$$

where $Z=\begin{bmatrix} z_1 & z_2 & z_3\end{bmatrix}$ and $z_i\sim N(0,I_3)$. The parameters for $x_1$ are given by $(0.4,0.6,-0.2,0.4,0.2)$, those for $x_2$ by $(0.2,-0.2,0.6,0.4,-0.6)$. We draw

$$
u_i=\begin{pmatrix}u_{1i}\\u_{2i}\\u_{3i}\end{pmatrix}\sim N\left(\begin{pmatrix}0\\0\\0\end{pmatrix},\begin{pmatrix}1 & \rho_1 & \rho_2\\ \rho_1 & 1 & \rho_1\rho_2\\ \rho_2 & \rho_1\rho_2 & 1\end{pmatrix}\right)
$$

with $\rho_1 = 0.3$ and $\rho_2 = -0.2.$, and set $\varepsilon_i = u_{1i}\sqrt{\exp(\gamma_\varepsilon z_1)}$, $v_{1i} = u_{2i}\sqrt{\exp(z_i'\gamma)}$ and $v_{2i} = u_{3i}\sqrt{\exp(z_i'\gamma)}$.

The first design is homoskedastic and hence $\gamma_\varepsilon = 0$ and $\gamma = 0$. The second design is heteroskedastic, with $\gamma_\varepsilon = 1.5$, and $\gamma = (0.5, 0.8, -0.3)'$. We standardise the $n$-vectors $\varepsilon$, $v_1$ and $v_2$ such the square of their $L_2$-norms are equal to $n$, where $n = n_1 + n_2$. Estimation results for these two designs are presented in Table 1. Sample sizes are $n_1 = 500$ and $n_2 = 1000$ for both designs. Results are presented from 10,000 MC replications. The table reports the means and standard deviations of the TS2SLS estimates for $\beta_1$ and $\beta_2$, the means of the non-robust and robust standard errors, plus the rejection frequencies of the Wald tests, testing $H_0 : \beta_1 = 0.3$, and $H_0 : \beta_2 = -0.1$ respectively, at the 5% nominal size. The results clearly show that the means of the robust standard errors are very close to the standard deviation of the TS2SLS estimates for both designs, whereas the non-robust standard errors underestimate the variability of the estimates in the heteroskedastic design. This is reflected in the behaviour of the Wald tests. Those based on the robust variance estimates have correct size, whereas those based on the non-robust variance estimates overreject the null in the heteroskedastic design.

Table 1. Monte Carlo results for the TS2SLS estimator

| Design | mean | std dev | mean se | mean rob se | Wald | rob Wald |
|---|---|---|---|---|---|---|
| Homoskedastic | | | | | | |
| $\beta_1$ | 0.300 | 0.075 | 0.074 | 0.074 | 0.049 | 0.051 |
| $\beta_2$ | -0.099 | 0.086 | 0.083 | 0.083 | 0.054 | 0.055 |
| Heteroskedastic | | | | | | |
| $\beta_1$ | 0.301 | 0.102 | 0.072 | 0.099 | 0.155 | 0.052 |
| $\beta_2$ | -0.099 | 0.099 | 0.082 | 0.096 | 0.102 | 0.054 |

Notes: results from 10,000 MC replications. Rej. freq. of Wald tests at 5% nominal size. Sample sizes $n_1$= 500, $n_2$= 1000

## A.3 Stata Code

### A.3.1 TS2SLS

In the first example, we want to estimate the following model

$$y = x_1\beta_1 + x_2\beta_2 + w\beta_w + \beta_0 + \varepsilon. \tag{19}$$

We are in the standard setup for the TS2SLS estimator. In sample 1, we have observations on the variables $y$, $w$, $z_1$, $z_2$ and $z_3$. In sample 2, we have observations on the variables

$x_1$, $x_2$, $w$, $z_1$, $z_2$ and $z_3$. Simple Stata syntax to compute the TS2SLS estimator and the non-robust and robust standard errors is given below.

```
use sample2.dta, clear
gen const = 1

qui gmm (x1 - {xb1: z1 z2 z3 w const}) ///
       (x2 - {xb2:z1 z2 z3 w const}), ///
       instruments(1 2: z1 z2 z3 w) ///
       winit(unadjusted,independent) onestep  ///
       deriv(1/xb1 = -1) ///
       deriv(2/xb2 = -1)
mat Vx2het = e(V)      /*Robust variance estimate of pix2*/

qui sureg (x1 x2 = z1 z2 z3 w )
mat Vx2hom = e(V)      /*Non-robust variance estimate of pix2*/

use sample1.dta, clear

/*Generating predicted X*/
qui predict x1h, equation(x1)
qui predict x2h, equation(x2)

scalar kx = 2       /*Number of predicted variables, here x1 and x2*/
scalar ke = 2       /*Number of exogenous variables, here w and constant*/

qui reg y z1 z2 z3 w
mat Vy1hom = e(V)*e(df_r)/_N   /*Non-robust variance estimate of piy1,*/
                              /*without degrees of freedom correction*/
qui reg y z1 z2 z3 w, rob
mat Vy1het = e(V)*e(df_r)/_N   /*Robust variance estimate of piy1,*/
                              /*without degrees of freedom correction*/

/*TS2SLS estimator*/
qui reg y x1h x2h w
mat b2s = e(b)
mat b2sx = b2s[1,1..kx]'         /*Selecting beta for predicted X only*/

/*Constructing C hat*/
qui reg z1 x1h x2h w
mat ch = e(b)'
qui reg z2 x1h x2h w
mat ch = ch,e(b)'
qui reg z3 x1h x2h w
mat ch = ch,e(b)'
mat ch = ch,(J(kx,ke,0)\I(ke))        /*Adjusting ch for the exogenous variables*/


/*Calculating non-robust standard errors*/
mat var1hom =  ch*Vy1hom*ch' + (b2sx' # ch)*Vx2hom*(b2sx # ch')
mat seb2shom = vecdiag(cholesky(diag(vecdiag(var1hom))))'
```

```
/*Calculating robust standard errors*/
mat var1het =  ch*Vy1het*ch' + (b2sx' # ch)*Vx2het*(b2sx # ch')
mat seb2shet = vecdiag(cholesky(diag(vecdiag(var1het))))'

/*Displaying the results*/
mat res = b2s',seb2shom,seb2shet
mat colnames res = b_ts2sls se "rob se"
mat rownames res = x1 x2 w _cons
matlist res
```

### A.3.2 A Generalisation

In the second example, we are interested in estimating the model

$$y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + w\beta_w + \beta_0 + \varepsilon.$$

We now observe in sample 1 the variables $y$, $x_1$, $x_3$, $w$, $z_1$, $z_2$, $z_3$ and $z_4$. In sample 2 we observe the variables $x_2$, $x_3$, $w$, $z_1$, $z_2$, $z_3$ and $z_4$. The Stata code for computing the Two-Sample estimator and the robust standard errors is as follows.

```
/*Merging Sample1 and Sample2 data on identifier id*/
use sample1.dta, clear
qui merge 1:1 id using sample2.dta

/*Generating Var(theta) and predicted X*/
qui reg y z1 z2 z3 z4 w
est store eqn_y
qui reg x1 z1 z2 z3 z4 w
est store eqn_x1
qui predict x1h
qui reg x2 z1 z2 z3 z4 w
est store eqn_x2
qui predict x2h
qui reg x3 z1 z2 z3 z4 w
est store eqn_x3
qui predict x3h

qui suest eqn_y eqn_x1 eqn_x2 eqn_x3

mat var = e(V)*(_N-1)/_N     /*Robust variance estimate of theta*/
                            /*without degrees of freedom correction*/

/*Selecting rows and columns from var associated with theta*/
mata
kz = 6       /*Total number of instruments, here z1, z2, z3, z4, w and const*/
kyx = 4      /*Number of variables in X, here x1, x2 and x3, plus 1 for y*/
sel = range(1,kz,1)
```

```
j = 2
while (j<=kyx)
{
ss = range((j-1)*kz+j,j*kz+(j-1),1)
sel = sel\ss
j = j+1
}

var = st_matrix("var")
var = var[sel,sel]
st_matrix("Vthetahet",var)
end

/*Selecting Sample1 data and variables*/
/*to compute TS estimator and variance*/
drop if y==.
keep y x1h x2h x3h w z1 z2 z3 z4

scalar kx = 3       /*Number of predicted variables, here x1, x2 and x3*/
scalar ke = 2       /*Number of exogenous variables, here w and constant*/

/*TS Estimator*/
qui reg y x1h x2h x3h w
mat b2s = e(b)
mat b2sx = b2s[1,1..kx]'           /*Selecting beta for predicted X only*/

/*Constructing C hat*/
qui reg z1 x1h x2h x3h w
mat ch = e(b)'
qui reg z2 x1h x2h x3h w
mat ch = ch,e(b)'
qui reg z3 x1h x2h x3h w
mat ch = ch,e(b)'
qui reg z4 x1h x2h x3h w
mat ch = ch,e(b)'
mat ch = ch,(J(kx,ke,0)\I(ke))        /*Adjusting ch for the exogenous variables*/

/*Calculating robust standard errors*/
mat delta = 1\-b2sx
mat var1het = (delta' # ch)*Vthetahet*(delta # ch')
mat seb2shet = vecdiag(cholesky(diag(vecdiag(var1het))))'

/*Displaying the results*/
local names = "x1 x2 x3 w _cons"
mat colnames b2s = `names'
mat colnames var1het = `names'
mat colnames var1het = _:
mat rownames var1het = `names'
mat rownames var1het = _:

cap prog drop output2s
prog output2s, eclass
```

```
eret post b2s var1het
eret local depvar y
eret local vcetype Robust
eret dis
end

output2s
```

### A.3.3 GMM

Stata code for the nonlinear GMM estimator for the example as in Section A.3.1.

```
use sample2.dta, clear
qui sureg (x1 x2 = z1 z2 z3 w)
mat Vpx2 = e(V)

use sample1.dta, clear
qui reg y z1 z2 z3 w
matrix Vpy1 = e(V)

/*vyvxmat is weightmatrix for optimal GMM estimator under conditional homoskedasticity*/
mat vyvxmat = Vpy1,J(rowsof(Vpy1),colsof(Vpx2),0)
mat vyvxmat = vyvxmat\(J(rowsof(Vpx2),colsof(Vpy1),0),Vpx2)

/*s1 is sample 1 identifier*/
gen s1 = 1

/*Merging Sample1 and Sample2 data on identifier id*/
/*No overlap in id here*/
qui merge 1:1 id using sample2.dta

/*s2 is sample 2 identifier*/
replace s1 = 0 if s1==.
gen s2 = 1-s1

/*setting values of y and x to zero in samples where missing*/
replace y = 0 if s2==1
replace x1 = 0 if s1==1
replace x2 = 0 if s1==1

/*instruments with different names for different samples, with zero values elsewhere*/
gen z11 = z1*s1
gen z12 = z1*s2
gen z21 = z2*s1
gen z22 = z2*s2
gen z31 = z3*s1
gen z32 = z3*s2
gen w1 = w*s1
gen w2 = w*s2

#delimit ;
```

```
gmm (y - {b1}*({p11}*z11+{p12}*z21+{p13}*z31+{p1w}*w1+{p10}*s1)
     - {b2}*({p21}*z11+{p22}*z21+{p23}*z31+{p2w}*w1+{p20}*s1)
     - {bw}*w1-{b0}*s1)
  (x1 - {p11}*z12-{p12}*z22-{p13}*z32-{p1w}*w2-{p10}*s2)
  (x2 - {p21}*z12-{p22}*z22-{p23}*z32-{p2w}*w2-{p20}*s2),
   instruments(1:z11 z21 z31 w1 s1, nocons)
   instruments(2 3:z12 z22 z32 w2 s2, nocons)
   winit(vyvxmat) onestep
   deriv(1/b1 = -({p10}*s1+{p1w}*w1+{p11}*z11+{p12}*z21+{p13}*z31))
   deriv(1/b2 = -({p20}*s1+{p2w}*w1+{p21}*z11+{p22}*z21+{p23}*z31))
   deriv(1/bw = -w1)
   deriv(1/b0 = -s1)
   deriv(1/p11 = -{b1}*z11)
   deriv(1/p12 = -{b1}*z21)
   deriv(1/p13 = -{b1}*z31)
   deriv(1/p1w = -{b1}*w1)
   deriv(1/p10 = -{b1}*s1)
   deriv(1/p21 = -{b2}*z11)
   deriv(1/p22 = -{b2}*z21)
   deriv(1/p23 = -{b2}*z31)
   deriv(1/p2w = -{b2}*w1)
   deriv(1/p20 = -{b2}*s1)
   deriv(2/p11 = -z12)
   deriv(2/p12 = -z22)
   deriv(2/p13 = -z32)
   deriv(2/p1w = -w2)
   deriv(2/p10 = -s2)
   deriv(3/p21 = -z12)
   deriv(3/p22 = -z22)
   deriv(3/p23 = -z32)
   deriv(3/p2w = -w2)
   deriv(3/p20 = -s2);

/*use twostep option and ‘estat overid’ command for efficient GMM and Hansen test*/
```