

Two-Sample Instrumental Variables Estimators

Atsushi Inoue

Gary Solon*

University of British Columbia

University of Michigan

North Carolina State University

December 2006

Abstract: Following an influential article by Angrist and Krueger (1992) on two-sample instrumental variables (TSIV) estimation, numerous empirical researchers have applied a computationally convenient two-sample two-stage least squares (TS2SLS) variant of Angrist and Krueger's estimator. In the two-sample context, unlike the single-sample situation, the IV and 2SLS estimators are numerically distinct. Our comparison of the properties of the two estimators demonstrates that the commonly used TS2SLS estimator is more asymptotically efficient than the TSIV estimator.

Classification Number: C300

*Atsushi Inoue is Associate Professor, Department of Economics, University of British Columbia, Vancouver, BC V6T 1Z1, Canada (email: ainoue@interchange.ubc.ca). Gary Solon is Professor, Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220 (email: gsolon@umich.edu). The authors are grateful for helpful comments from the editor, the referees, Joshua Angrist, and participants at the Hitotsubashi conference on economic statistics.

Two-Sample Instrumental Variables Estimators

I. Introduction

A familiar problem in econometric research is consistent estimation of the coefficient vector in the linear regression model

$$y = W\theta + \varepsilon \tag{1}$$

where y and ε are $n \times 1$ vectors and W is an $n \times k$ matrix of regressors, some of which are endogenous, i.e., contemporaneously correlated with the error term ε . As is well known, the ordinary least squares estimator of θ is inconsistent, but consistent estimation is still possible if there exists an $n \times q$ ($q \geq k$) matrix Z of valid instrumental variables. For example, in the case of exact identification with $q = k$, the conventional instrumental variables (IV) estimator is

$$\hat{\theta}_{IV} = (Z'W)^{-1}Z'y. \tag{2}$$

With exact identification, this estimator is identical to the two-stage least squares (2SLS) estimator

$$\hat{\theta}_{2SLS} = (\hat{W}'\hat{W})^{-1}\hat{W}'y \tag{3}$$

where $\hat{W} = Z(Z'Z)^{-1}Z'W$. If, in addition, ε is i.i.d. normal, this estimator is asymptotically efficient among “limited information” estimators.

An influential article by Angrist and Krueger (1992) has pointed out that, under certain conditions, consistent instrumental variables estimation is still possible even

when only y and Z (but not W) are observed in one sample and only W and Z (but not y) are observed in a second distinct sample. In that case, the same moment conditions that lead to the conventional IV estimator in equation (2) motivate the “two-sample instrumental variables” (TSIV) estimator

$$\hat{\theta}_{TSIV} = (Z_2'W_2/n_2)^{-1}(Z_1'y_1/n_1) \quad (4)$$

where Z_1 and y_1 contain the n_1 observations from the first sample and Z_2 and W_2 contain the n_2 observations from the second.

Of the many empirical researchers who have since used a two-sample approach (e.g., Bjorklund and Jantti, 1997; Currie and Yelowitz, 2000; Dee and Evans, 2003; Borjas, 2004), nearly all have used the “two-sample two-stage least squares” (TS2SLS) estimator

$$\hat{\theta}_{TS2SLS} = (\hat{W}_1'\hat{W}_1)^{-1}\hat{W}_1'y_1 \quad (5)$$

where $\hat{W}_1 = Z_1(Z_2'Z_2)^{-1}Z_2'W_2$. These researchers may not have been aware that the equivalence of IV and 2SLS estimation in a single sample does not carry over to the two-sample case. Instead, it is easy to show that, in the exactly identified case,

$$\hat{\theta}_{TS2SLS} = (Z_2'W_2/n_2)^{-1}C(Z_1'y_1/n_1) \quad (6)$$

where $C = (Z_2'Z_2/n_2)(Z_1'Z_1/n_1)^{-1}$. $\hat{\theta}_{TS2SLS}$ differs from $\hat{\theta}_{TSIV}$ by inserting the C matrix, which can be viewed as a sort of correction for differences between the two samples in their empirical covariance matrices for Z . Under Angrist and Krueger’s assumptions, the correction matrix C would converge in probability to the identity

matrix, and the TSIV and TS2SLS estimators therefore would have the same probability limit. In finite samples, however, the TSIV estimator originally proposed by Angrist and Krueger and the TS2SLS estimator typically used by practitioners are numerically distinct estimators.

The obvious question then becomes: Which estimator should be preferred? Our formal analysis, which considers overidentified as well as exactly identified models, demonstrates that the TS2SLS estimator is superior because its implicit correction for differences between the two samples in the distribution of Z yields a gain in asymptotic efficiency. After providing the formal demonstration, we develop intuition for the result by discussing simple examples.

II. Asymptotic Distributions of Two-Sample IV Estimators

We will compare two-sample IV estimators in a general single-equation framework:

$$y_{1i} = \beta'x_{1i} + \gamma'z_{1i}^{(1)} + \varepsilon_{1i} = \theta'w_{1i} + \varepsilon_{1i}, \quad (7)$$

$$x_{1i} = \Pi z_{1i} + \eta_{1i}, \quad (8)$$

$$x_{2i} = \Pi z_{2i} + \eta_{2i}, \quad (9)$$

where x_{1i} and x_{2i} are p -dimensional random vectors, $z_{1i} = [z_{1i}^{(1)'} \ z_{1i}^{(2)'}]'$ and z_{2i} are $q(=q^{(1)}+q^{(2)})$ -dimensional random vectors, w_{1i} is a $k(=p+q^{(1)})$ -dimensional random vector, and Π is a $p \times q$ matrix of parameters.

For efficiency comparison, it is useful to characterize these estimators as generalized method of moments (GMM) estimators. First the TSIV estimator is a GMM

estimator based on moment conditions

$$E \left[z_{1i}(y_{1i} - z_{1i}^{(1)'} \gamma) - z_{2i}x'_{2i}\beta \right] = 0. \quad (10)$$

Next the TS2SLS estimator is a GMM estimator based on

$$E[z_{1i}(y_{1i} - z'_{1i}\Pi'\beta - z_{1i}^{(1)'}\gamma)] = 0, \quad (11)$$

$$E[z_{2i} \otimes (x_{2i} - \Pi z_{2i})] = 0. \quad (12)$$

When Π is defined to be the coefficient on z_i in the population linear projection of x_i on z_i , (12) holds by definition of linear projections.

Finally we consider the two-sample limited-information maximum likelihood (TSLIML) estimator for efficiency comparison. Let $\sigma_{11} = E[(\varepsilon_{1i} + \beta'\eta_{1i})^2]$ and $\Sigma_{22} = E(\eta_{2i}\eta'_{2i})$. When $[\varepsilon_i \ \eta'_{1i}]'$ and η_{2i} are normally distributed the log of the likelihood function can be written as

$$\begin{aligned} \ln L = & -\frac{n}{2}\ln(2\pi) - \frac{n_1}{2}\ln(\sigma_{11}) - \frac{n_2}{2}\ln|\Sigma_{22}| \\ & - \frac{1}{2\sigma_{11}} \sum_{i=1}^{n_1} (y_{1i} - \beta'\Pi z_{1i} - \gamma'z_i^{(1)})^2 \\ & - \frac{1}{2} \sum_{i=1}^{n_2} (x_{2i} - \Pi z_{2i})' \Sigma_{22}^{-1} (x_{2i} - \Pi z_{2i}). \end{aligned} \quad (13)$$

The TSLIML estimator is asymptotically equivalent to a GMM estimator based on the population first-order conditions for the TSLIML estimator:

$$E[\Pi z_{1i}(y_{1i} - \beta'\Pi z_{1i} - \gamma'z_i^{(1)})] = 0, \quad (14)$$

$$E[z_i^{(1)}(y_{1i} - \beta'\Pi z_{1i} - \gamma'z_i^{(1)})] = 0, \quad (15)$$

$$E(z_{1i} \otimes \beta u_{1i}/\sigma_{11} + z_{2i} \otimes \Sigma_{22}^{-1}\eta_{2i}) = 0, \quad (16)$$

$$E(u_{1i}^2/\sigma_{11}^2 - 1/\sigma_{11}) = 0, \quad (17)$$

$$E[(\Sigma_{22}^{-1}\eta_{2i}) \otimes (\Sigma_{22}^{-1}\eta_{2i}) - |\Sigma_{22}|\text{tr}(\Sigma_{22}^{-1})]D_2 = 0, \quad (18)$$

where D_2 is a $p^2 \times p(p+1)/2$ matrix such that $\text{vec}(\Sigma_{22}) = D_2\text{vech}(\Sigma_{22})$.

To derive the asymptotic distributions of these estimators we assume the following conditions.

Assumptions.

- (a) $\{[y_{1i}, z'_{1i}]\}'_{i=1}^{n_1}$ and $\{[x_{2i}, z'_{2i}]\}'_{i=1}^{n_2}$ are i.i.d. random vectors with finite fourth moments and are independent.
- (b) $E(z_{1i}\varepsilon_{1i}) = 0$, $E(z_{1i}\eta_{1i}) = 0$ and $E(z_{2i}\eta_{2i}) = 0$.
- (c) ε_{1i} and η_{1i} are uncorrelated with third moments of z_{1i} .
- (d) Third moments of $[\varepsilon_{1i} \ \eta'_{1i}]$ and those of η_{2i} are all zero conditional on z_{1i} and z_{2i} , respectively.
- (e) $E(u_{1i}^2|z_{1i}) = \sigma_{11}$ and $E(\eta_{2i}\eta'_{2i}|z_{2i}) = \Sigma_{22}$ where $u_{1i} = \varepsilon_{1i} + \beta'\eta_{1i}$, $\sigma_{11} > 0$ and Σ_{22} is positive definite.
- (f) For the TSIV estimator

$$\text{rank} \begin{bmatrix} E(z_{2i}x'_{2i}) & 0 \\ 0 & E(z_{1i}z_{1i}^{(1)'}) \end{bmatrix} = \dim(\theta)$$

and for the TS2SLS and TSLIML estimators $\text{rank}[E(z_{1i}w'_{1i})] = \dim(\theta)$.

(g) $E(z_{1i}z'_{1i})$ and $E(z_{2i}z'_{2i})$ are nonsingular.

(h) $E(z_{1i}x'_{1i}) = E(z_{2i}x'_{2i})$ and second and fourth moments of z_{1i} and z_{2i} are identical.

(i) $\lim_{n_1, n_2 \rightarrow \infty} n_1/n_2 = \kappa$ for some $\kappa > 0$.

Remarks. Assumptions (c) and (d) are used to simplify the asymptotic covariance matrices of the TSIV and TSLIML estimators, respectively. Assumption (h) provides a basis for combining two samples.¹

The following proposition compares the asymptotic distributions of the three estimators. The proof is in the appendix.

Proposition. Suppose that Assumptions (a)–(i) hold. Then $\hat{\theta}_{TSIV}$, $\hat{\theta}_{TS2SLS}$ and $\hat{\theta}_{TSLIML}$ are $\sqrt{n_1}$ -consistent² and asymptotically normally distributed with asymptotic covariance matrices Σ_{TSIV} , Σ_{TS2SLS} and Σ_{TSLIML} , respectively, where

$$\Sigma_{TSIV} = \left\{ B' [(\sigma_{11} + \kappa\beta'\Sigma_{22}\beta)A + (1 + \kappa)C]^{-1} B \right\}^{-1}, \quad (19)$$

$$\Sigma_{TS2SLS} = \{ B' [(\sigma_{11} + \kappa\beta'\Sigma_{22}\beta)A]^{-1} B \}^{-1}, \quad (20)$$

$$\Sigma_{TSLIML} = \Sigma_{TS2SLS}, \quad (21)$$

¹One can show that the TS2SLS estimator does not require the second part of the assumption but requires a weaker condition $E(z_{1i}x'_{1i}) = cE(z_{2i}x'_{2i})$ and $E(z_{1i}z'_{1i}) = cE(z_{2i}z'_{2i})$ for some c . Because c does not have to be unity, the TS2SLS estimator is more robust than the TSIV estimator.

²Following Angrist and Krueger (1992), we scale the estimator by $\sqrt{n_1}$.

$A = E(z_{1i}z'_{1i}) = E(z_{2i}z'_{2i})$, $B = E(z_{1i}w'_{1i}) = E(z_{2i}w'_{2i})$, and $C = Cov(z_{ji}z'_{ji}\Pi\beta) = E(z_{ji}z'_{ji}\Pi\beta\beta'\Pi'z_{ji}z'_{ji}) - E(z_{ji}z'_{ji}\Pi\beta)E(\beta'\Pi'z_{ji}z'_{ji})$ for $j = 1, 2$.

Before discussing the proposition's important implications for the relative efficiency of alternative two-sample IV estimators, we wish to clarify the connections between our results on asymptotic covariance matrices and those in the previous literature. First, readers wishing to relate our Σ_{TSIV} to the equivalent expression in Angrist and Krueger (1992) should note that, in Angrist and Krueger's notation,

$$\phi_1 = \sigma_{11}A + C, \quad (22)$$

$$\omega_2 = \beta'\Sigma_{22}\beta A + C. \quad (23)$$

Second, in a subsequent paper on split-sample IV estimation as a method for avoiding finite-sample bias when the instruments are only weakly correlated with the endogenous regressors, Angrist and Krueger (1995) noted the distinction between TS2SLS and TSIV, but conjectured incorrectly that Σ_{TS2SLS} is the same as Σ_{TSIV} .³ Third, in the related literature on “generated regressors,” first-stage estimation is performed to create a proxy for an unobserved regressor in the second-stage equation, rather than

³The source of their error was the incorrectness of their claim near the beginning of p.228 that setting Φ to be $(Z'_2Z_2)^{-1}Z'_1Z_1(Z'_2Z_2)^{-1}$ would reproduce the TS2SLS estimator. It is unclear to us how much this error has affected inference in applied research. Bjorklund and Jantti (1997), for example, used a bootstrap method instead to obtain their standard error estimates. Dee and Evans (2003) noted that, in their exactly identified model, the TS2SLS estimator could be reinterpreted as an indirect least squares estimator, and they used that insight to motivate a straightforward delta method for estimating standard errors.

to treat the endogeneity of the regressor. Murphy and Topel (1985) explicitly discussed the instance in which the first-stage estimation is based on a different sample than the second-stage estimation. Our expression for Σ_{TS2SLS} follows from Murphy and Topel’s theorem 1, and their well-known discussion of how to estimate this covariance matrix is perfectly applicable here. Murphy and Topel’s analysis, however, says nothing about Σ_{TSIV} because $\hat{\theta}_{TSIV}$ cannot be written as a two-step regression estimator.

The main contributions of our proposition concern efficiency comparisons among alternative two-sample IV estimators. First, the asymptotic equivalence of TS2SLS and TSLIML implies, that, when the disturbance terms $[\varepsilon_i \ \eta_i']'$ are jointly normally distributed, the TS2SLS estimator is asymptotically efficient within the class of “limited information” estimators. Second, regardless of whether the disturbances are normal, since C is positive semidefinite, it follows that $\Sigma_{TSIV} - \Sigma_{TS2SLS}$ is positive semidefinite. Thus, the TS2SLS estimator is more asymptotically efficient than the TSIV estimator. The asymptotic efficiency gain comes from the implicit correction of the TS2SLS estimator for differences between the finite-sample distributions of z_{1i} and z_{2i} stemming from random sampling variation.⁴

⁴In Monte Carlo experiments, we have verified that these asymptotic results accurately characterize the finite-sample behavior of the TSIV, TS2SLS, and TSLIML estimators. The exception is that, when the instruments are very weakly correlated with the endogenous regressor, all three estimators appear to be biased towards zero. This corroborates an analytical result of Angrist and Krueger (1995) concerning TS2SLS. The Monte Carlo results suggest that the bias is most severe for

As an illustration, consider the case in which one endogenous variable is the only explanatory variable and we have one instrument, i.e., $p = q = q^{(2)} = 1$. In this case one can show that

$$\sqrt{n_1}(\hat{\beta}_{TSIV} - \beta) - \sqrt{n_1}(\hat{\beta}_{TS2SLS} - \beta) = \frac{n_1^{-1/2} \sum_{i=1}^{n_1} z_{1i}^2 - \sqrt{\kappa} n_1^{-1/2} \sum_{i=1}^{n_2} z_{2i}^2}{(1/n_2) \sum_{i=1}^{n_2} z_{2i} x_{2i}} \Pi \beta + o_p(1). \quad (24)$$

The first term on the RHS will be asymptotically independent of $\sqrt{n_1}(\hat{\beta}_{TS2SLS} - \beta)$ and have a positive variance even asymptotically and it follows from the proposition that its variance is given by

$$\frac{Var(z_{1i}^2 \Pi \beta) + \kappa Var(z_{2i}^2 \Pi \beta)}{[E(z_i x_i)]^2}. \quad (25)$$

The intuition for this illustration can be further developed by following a referee's suggestion to consider the special case in which the regressor is exogenous and therefore can serve as its own instrument. In this case, the TS2SLS estimator simplifies to OLS applied to the first sample while the TSIV estimator is the ratio of the sample covariance between the regressor and the dependent variable in the first sample to the sample variance of the regressor in the second sample. We know that OLS is the preferred estimator in this instance. The point becomes particularly clear if we further assume zero variance of the error term in the structural equation. In that case, the OLS/TS2SLS estimator is identically equal to θ , i.e., it estimates θ with TSIV and least so for TSLIML. The latter replicates a familiar finding in the one-sample literature (e.g., Angrist, Imbens, and Krueger, 1999). A summary of our Monte Carlo results is available at <http://www-personal.umich.edu/~gsolon/workingpapers.htm>.

zero variance. In contrast, the TSIV estimator is θ times the ratio of the regressor's sample variance in the first sample to its sample variance in the second sample. That ratio has probability limit 1, but random sampling variation causes it to have positive variance. What our proposition demonstrates is that this is a general phenomenon — random differences between the two samples in their empirical covariance matrices for the instruments cause the TSIV estimator to be less asymptotically efficient than the TS2SLS estimator.

To summarize, following Angrist and Krueger's (1992) influential work on two-sample instrumental variables (TSIV) estimation, many applied researchers have used a computationally convenient two-sample two-stage least squares (TS2SLS) variant of Angrist and Krueger's estimator. In the two-sample context, unlike the single-sample setting, the IV and 2SLS estimators are numerically distinct. We have shown that the commonly used TS2SLS approach is more asymptotically efficient because it implicitly corrects for differences in the empirical distributions of the instrumental variables between the two samples.⁵

⁵In Inoue and Solon (2005), we also show that the TS2SLS estimator remains consistent under a practically relevant type of stratified sampling that renders the TSIV estimator inconsistent. The longer paper also extends our analysis to the case of conditionally heteroskedastic disturbances.

Appendix: Proof of the proposition

The consistency and asymptotic normality of the GMM estimators follow from the standard arguments. Thus, we will focus on the derivation of asymptotic variances.

Let G_{TSIV} and V_{TSIV} denote the Jacobian and covariance matrix, respectively, of the moment condition (10). Under Assumptions 1(c), (e), and (i), we have

$$G_{TSIV} = -[E(z_{2i}x'_{2i}) E(z_{1i}z'_{1i})] = -B, \quad (26)$$

$$\begin{aligned} V_{TSIV} &= (\sigma_{11}E(z_{1i}z'_{1i}) + \kappa\beta'\Sigma_{22}\beta E(z_{2i}z'_{2i})) + Cov(z_{1i}z'_{1i}\Pi\beta) + \kappa Cov(z_{2i}z'_{2i}\Pi\beta) \\ &= (\sigma_{11} + \kappa\beta'\Sigma_{22}\beta) A + Cov(z_{1i}z'_{1i}\Pi\beta) + \kappa Cov(z_{2i}z'_{2i}\Pi\beta) \end{aligned} \quad (27)$$

from which (19) follows.

Let G_{TS2SLS} and V_{TS2SLS} denote the Jacobian and covariance matrix, respectively, of the moment conditions (11) and (12). Because the Jacobian and covariance matrices of the moment functions are given by

$$G_{TS2SLS} = - \begin{bmatrix} E(z_{1i}w'_{1i}) & E(z_{1i}z'_{1i}) \otimes \beta' \\ 0 & E(z_{2i}z'_{2i}) \otimes I_p \end{bmatrix} = - \begin{bmatrix} B & A \otimes \beta' \\ 0 & A \otimes I_p \end{bmatrix} \quad (28)$$

$$V_{TS2SLS} = \begin{bmatrix} \sigma_{11}E(z_{1i}z'_{1i}) & 0 \\ 0 & \kappa E(z_{2i}z'_{2i}) \otimes \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{11}A & 0 \\ 0 & \kappa A \otimes \Sigma_{22} \end{bmatrix} \quad (29)$$

respectively, the asymptotic covariance matrix of the TS2SLS estimator is the $k \times k$ upper-left submatrix of the inverse of

$$G'_{TS2SLS} V_{TS2SLS}^{-1} G_{TS2SLS} = \begin{bmatrix} \frac{1}{\sigma_{11}} B' A^{-1} B & B' \otimes \frac{\beta'}{\sigma_{11}} \\ B \otimes \frac{\beta}{\sigma_{11}} & A \otimes \left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa} \Sigma_{22}^{-1} \right) \end{bmatrix}. \quad (30)$$

Because the $k \times k$ upper-left submatrix of $(G'_{TS2SLS} V_{TS2SLS}^{-1} G_{TS2SLS})^{-1}$ is the inverse

of

$$\begin{aligned}
& \frac{1}{\sigma_{11}} B' A^{-1} B - B' \otimes \frac{\beta'}{\sigma_{11}} \left[A \otimes \left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa} \Sigma_{22}^{-1} \right) \right]^{-1} B \otimes \frac{\beta}{\sigma_{11}} \\
&= \left[\frac{1}{\sigma_{11}} - \frac{\beta'}{\sigma_{11}} \left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa} \Sigma_{22}^{-1} \right)^{-1} \frac{\beta}{\sigma_{11}} \right]^{-1} B' A^{-1} B
\end{aligned} \tag{31}$$

by Theorem 13 in Amemiya (1985, p. 460) and

$$\left[\frac{1}{\sigma_{11}} - \frac{\beta'}{\sigma_{11}} \left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa} \Sigma_{22}^{-1} \right)^{-1} \frac{\beta}{\sigma_{11}} \right]^{-1} = \sigma_{11} + \kappa \beta' \Sigma_{22} \beta \tag{32}$$

by Theorem 0.7.4 of Horn and Johnson (1985, p.19), (20) follows.

Under the assumptions, one can show that the asymptotic distribution of $\hat{\theta}_{TSLIML}$ and the one of the TSLIML estimator for $[\sigma_{11} \text{vech}(\Sigma_{22})']'$ are independent. Thus, we can focus on the moment conditions (14), (15) and (16). Under the stated assumptions, the negative of the Jacobian matrix G_{TSLIML} and the covariance matrix V_{TSLIML} of these moment conditions are the same and are given by

$$\begin{aligned}
& \begin{bmatrix} \Pi E(z_{1i} z'_{1i}) \Pi' / \sigma_{11} & \Pi E(z_{1i} z_{1i}^{(1)'}) / \sigma_{11} & \Pi E(z_{1i} z'_{1i}) \otimes \beta' / \sigma_{11} \\ E(z_{1i}^{(1)} z'_{1i}) \Pi' / \sigma_{11} & E(z_{1i}^{(1)} z_{1i}^{(1)'}) / \sigma_{11} & E(z_{1i}^{(1)} z'_{1i}) \otimes \beta' / \sigma_{11} \\ E(z_{1i} z'_{1i}) \Pi' \otimes \beta / \sigma_{11} & E(z_{1i} z_{1i}^{(1)'}) \otimes \beta / \sigma_{11} & E(z_{1i} z'_{1i}) \otimes \frac{\beta\beta'}{\sigma_{11}} + E(z_{2i} z'_{2i}) \otimes \frac{1}{\kappa} \Sigma_{22}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \Pi A \Pi' / \sigma_{11} & \Pi E(z_{1i} z_{1i}^{(1)'}) / \sigma_{11} & \Pi A \otimes \beta' / \sigma_{11} \\ E(z_{1i}^{(1)} z'_{1i}) \Pi' / \sigma_{11} & E(z_{1i}^{(1)} z_{1i}^{(1)'}) / \sigma_{11} & E(z_{1i}^{(1)} z'_{1i}) \otimes \beta' / \sigma_{11} \\ A \Pi' \otimes \beta / \sigma_{11} & E(z_{1i} z_{1i}^{(1)'}) \otimes \beta / \sigma_{11} & A \otimes \left(\frac{\beta\beta'}{\sigma_{11}} + \frac{1}{\kappa} \Sigma_{22}^{-1} \right) \end{bmatrix}.
\end{aligned} \tag{33}$$

Since $E(z_{1i} \eta_{1i}) = 0$, this matrix is the same as (30) from which (21) follows.

References

- Amemiya, Takeshi (1985), *Advanced Econometrics*, Harvard University Press: Cambridge, MA.
- Angrist, Joshua D., Guido W. Imbens, and Alan B. Krueger (1999), “Jackknife Instrumental Variables Estimation,” *Journal of Applied Econometrics*, 14, 57–67.
- Angrist, Joshua D. and Alan 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, Joshua D. and Alan B. Krueger (1995), “Split-Sample Instrumental Variables Estimates of the Return to Schooling,” *Journal of Business and Economic Statistics*, 13, 225–235.
- Björklund, Anders and Markus Jäntti (1997), “Intergenerational Income Mobility in Sweden Compared to the United States,” *American Economic Review*, 87, 1009–1018.
- Borjas, George J. (2004), “Food Insecurity and Public Assistance,” *Journal of Public Economics*, 88, 1421–1443.
- Currie, Janet and Aaron Yelowitz (2000), “Are Public Housing Projects Good

for Kids?” *Journal of Public Economics*, 75, 99–124.

Dee, Thomas S. and William N. Evans (2003), “Teen Drinking and Educational Attainment: Evidence from Two-Sample Instrumental Variables Estimates,” *Journal of Labor Economics*, 21, 178–209.

Horn, Roger A. and Charles R. Johnson (1985), *Matrix Analysis*, Cambridge University Press: Cambridge, United Kingdom.

Inoue, Atsushi and Gary Solon (2005), “Two-Sample Instrumental Variables Estimators,” Technical Working Paper No. 311, National Bureau of Economic Research.

Murphy, Kevin M. and Robert H. Topel (1985), “Estimation and Inference in Two-Step Econometric Models,” *Journal of Business and Economic Statistics*, 3, 370–379.