Finite Sample Properties and Empirical Applicability of Two-Sample Two-Stage Least Squares*

Christopher Khawand[†], Wei Lin[‡]

July 9, 2015

Abstract

The two-sample two-stage least squares (TS2SLS) estimator allows instrumental variables estimation of a linear causal effect when two different samples contain the data to estimate the first-stage parameters and second-stage (or "reduced form") parameters. We express the TS2SLS estimator as a weighted average of 2SLS and split-sample 2SLS (SS2SLS) estimators which use the underlying data composing the sample used by the TS2SLS estimator while generally allowing the first- and second-stage samples to "overlap" (i.e., contain data from the same population units). The weight on the 2SLS component is increasing in the level of overlap between the first-stage and second-stage samples. A first-order approximation of the bias of the TS2SLS estimator reflects the mixture of biases from the 2SLS estimator (towards the probability limit of OLS) and the SS2SLS estimator (an attenuation bias). The asymptotic variance of TS2SLS also demonstrates the mixture property, with variance decreasing in the proportion of overlapping observations. We present Monte Carlo evidence of each notable property, and demonstrate an empirical example of the behavior of TS2SLS estimates relative to 2SLS based on subsamples of data from Angrist and Evans (1998).

*We thank Gary Solon, Jeff Wooldridge, and Soren Anderson for helpful comments. This research was supported in part by Institute for Education Sciences Grant R305B090011 to Michigan State University; the views expressed are those of the author and do not represent the views of the U.S. Department of Education.

[†]Department of Economics, Michigan State University. Email: khawandc@msu.edu

[‡]Department of Economics, Michigan State University. Email: linwei@msu.edu

1 Introduction

Instrumental variables (IV) methods enable the consistent estimation of endogenous variables' causal effects but suffer from poor finite-sample properties and data availability constraints. Bound, Jaeger, and Baker (1995) establish that estimation with weak instruments can lead to large inconsistencies and finite sample bias. IV estimates also tend to have relatively large standard errors, often inhibiting the interpretability of differences between IV and non-IV point estimates. Lastly, the idiosyncratic nature of valid instrumental variables reduces their availability in data sets alongside outcome and other variables of interest. Beginning with Klevmarken (1982), some researchers have sought to address the problem of data availability by using two-sample IV methods (TSIVM), which combine parameter estimates from multiple data sets into a final IV estimate. Under a set of ideal conditions, a TSIVM produces an estimate with identical bias to the otherwise inaccessible traditional IV estimate. However, the finite-sample properties of TSIVM estimators are generally unknown, and prior literature lacks clear guidelines for how researchers should interpret them. The potential for researchers to introduce additional data and estimate models by TSIVM to produce estimates superior to available single-sample estimates has also not been explored.

We establish some insights into the finite-sample properties of the two-sample two-stage least squares (TS2SLS) estimator. Likely owing to its ease of implementation and interpretation, the TS2SLS estimator is the most commonly-used TSIVM estimator in empirical applications (e.g., Arellano and Meghir 1992; den Berg et al. 2015; Devereux and Hart 2014; Nicoletti and Ermisch 2014; Rothstein and Wozny 2014). We broaden the set of potential applications of TS2SLS by demonstrating that even where a one-sample 2SLS estimate is available, a TS2SLS estimator may sometimes be preferred or worth reporting alongside the one-sample estimate because of greater precision and smaller bias. We propose approximations for the bias and variance of the TS2SLS estimator that are dependent on both the typical set of parameters in the "weak instruments" literature for one-sample IV estimators and on three parameters unique to two-sample estimators: the distinct sample sizes of the first- and second-stage samples and the proportion of observations which "overlap" between them (i.e., the fraction of real population units from the first-stage sample which are also in the second-stage sample). To test the approximations, we conduct a series of Monte Carlo simulations and compute the average bias and standard errors across simulations.

We develop a data framework in which the TS2SLS estimator is computed from a complete theoretical sample of population units (the "super-sample") of which the first- and second-stage samples used to compute the TS2SLS estimate of interest are subsets containing potentially overlapping units. This approach formally reconciles two-sample and split-sample IV estimators by making them identical; for example, it makes equivalent the no-overlap TS2SLS estimator studied in Inoue and Solon (2008) and split-sample 2SLS (SS2SLS) estimator analogous to Angrist and Krueger (1995)'s split-sample IV (SSIV). I find that the TS2SLS estimator can be written as a convex linear combination of the 2SLS estimator computed using one subsample of the super-sample, and the SS2SLS estimator computed using the remaining units. The weight on the 2SLS component is a function of sampling variation in the first-stage estimates for each subsample, and it converges asymptotically to the "overlap" parameter, representing the proportion of units in the second sample which are also in the first sample. This linear partitioning of the TS2SLS estimator frames it as a function of two estimators with known properties, simplifying the development of approximations of bias and variance.

We find that the TS2SLS estimator has bias behavior parallel to the 2SLS estimator, with all bias dependent on the degree of sampling error in the first stage parameter relative to the strength of the endogeneity. This result only holds for the finite-sample bias from first-stage sampling error, not from invalid instruments for from violations of the key TS2SLS assumption (that the first-stage coefficients on the instruments are identical for both primary and secondary samples). The TS2SLS estimator has gains in precision from increases in each of its corresponding sample sizes, with gains diminishing more rapidly from increasing first-stage sample size.

We demonstrate a hypothetical empirical application of TS2SLS using data from Angrist and Evans (1998), using their actual sample as the "super-sample" and examining the estimates they could have recovered had they been forced to use TS2SLS with subsets of their sample instead of 2SLS with their entire sample. Using TS2SLS with half the data for the first stage and half for the second stage (effectively a SS2SLS estimate), the estimate is closer to zero than the super-sample 2SLS estimate and less precise. We find that a TS2SLS estimate using only half of their observations for the first stage estimation but their entire sample for the second stage almost exactly recovers the super-sample 2SLS estimate with equivalent precision, providing evidence of the strength of the instrument used. This exercise suggests one situation in practice in which TS2SLS which is most likely to yield a high return: when the researcher can estimate the first stage precisely with one set of data, but also has access to more observations containing the outcome and the instrument but not the endogenous variable.

The literature formally concerning two-sample IV estimators has explored the computation and asymptotic properties of various two-sample IV estimators. Angrist and Krueger (1995) provide the finite-sample properties of the split-sample IV estimator, only alluding to its relationship to the two-sample IV estimator, which they use in another study (Angrist and Krueger 1993). Inoue and Solon (2008) computationally distinguish the two-sample IV estimator, which is calculated explicitly using the ratio of covariance matrices each estimated from different data sets, from the TS2SLS estimator, which is calculated using ordinary least squares of the outcome against cross-sample first stage fitted values. Their main finding is that the TS2SLS approach is asymptotically more efficient than the TSIV approach because TS2SLS takes into account differences in the sampling distribution of the instrument between the primary and secondary samples, while TSIV does not. Both Angrist and Krueger (1995) and Inoue and Solon (2008) only consider two-sample estimators which use fully independent samples. This paper fits into this literature by developing additional insight into the finite-sample behavior of the TS2SLS estimator while also generalizing it to potentially non-independent first- and second-stage samples.

2 Properties of the TS2SLS Estimator

2.1 Model

We follow a general single-equation framework in which all predetermined variables (including constants) are partialled out of the specification:

$$
\begin{aligned}
y_{1i} &= x_{1i}\beta + \varepsilon_i && (1)\\
&= \mathbf{z}_{1i}\gamma\beta + \beta v_{1i} + \varepsilon_i && (2)\\
&= \mathbf{z}_{1i}\gamma\beta + u_{1i} && (3)\\
x_{1i} &= \mathbf{z}_{1i}\gamma_1 + v_{1i} && (4)\\
x_{2j} &= \mathbf{z}_{2j}\gamma_2 + v_{2j}. && (5)
\end{aligned}
$$

The data sets available for use in estimation consist of two subsamples, $S1$ and $S2$ of a broader data set of N units. $S1$ and $S2$ generally may partially or entirely overlap in terms of the underlying units for which they have data. The assumptions for this model are as follows:

Assumptions:

1. S1 $\{(y_{1i}, \mathbf{z}_{1i})\}_{i=1}^{N_1}$ and S2 $\left\{(x_{2j}, \mathbf{z}_{2j})\right\}_{j=1}^{N_2}$ are i.i.d. random vectors from the same underlying population, where $\mathbf{z}_{1i}^{'}$ and $\mathbf{z}_{2j}^{'}$ are $K \times 1$ vectors and y_{1i} and x_{2j} are scalars. This implies:
    (a) $E(\mathbf{z}_{1i}'\mathbf{z}_{1i}) = E(\mathbf{z}_{2j}'\mathbf{z}_{2j}) = \Omega_{\mathbf{z}}$.
    (b) The "first stage" linear projection coefficients are the same for equation (4) and equation (5) in the population: $\gamma_1 = \gamma_2 = \gamma$.
    (c) $E(\mathbf{z}_{1i}'x_{1i}) = E(\mathbf{z}_{2j}'x_{2j}) = \Omega_{x\mathbf{z}}$.
    (d) Rank $E\left(\mathbf{z}_{1i}^{'}\mathbf{z}_{1i}\right) =$ Rank $E\left(\mathbf{z}_{2j}^{'}\mathbf{z}_{2j}\right) = K$
    (e) Rank $E\left(\mathbf{z}_{1i}^{'}x_{1i}\right) =$ Rank $E\left(\mathbf{z}_{2j}^{'}x_{2j}\right) = 1$
2. The number of units in N_2 also in N_1 is ρN_2, with $0 \leq \rho \leq 1$.

(a) $N = N_1 + (1-\rho)N_2$

3. The data are ordered so that sequence of the first ρN_2 number of observations from S2 corresponds to the sequence of the first ρN_2 number of observations in S1 $\left\{\left(x_{2j}, \mathbf{z}_{2j}\right)\right\}_{j=1}^{\rho N_2} = \{(x_{1i}, \mathbf{z}_{1i})\}_{i=1}^{\rho N_2}$.

4. Equation (1) is the structural equation with structural error ε_i, and equation (3) is the reduced form equation with reduced-form error u_{1i}. The error terms are homoskedastic and normally distributed,

$$\begin{pmatrix} \varepsilon_i \\ v_{1i} \\ v_{2j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon v} & 0 \\ \sigma_{\varepsilon v} & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_v^2 \end{pmatrix}\right].$$

Note v_{2j} is independent of ε_i, v_{1i} due to the fact that i and j are different observations. Following are the implications:

(a) $u_{1i} \sim N\left(0, \beta^2\sigma_v^2 + \sigma_\varepsilon^2 + 2\beta\sigma_{\varepsilon v}\right)$ and let $\sigma_u^2 \equiv \beta^2\sigma_v^2 + \sigma_\varepsilon^2 + 2\beta\sigma_{\varepsilon v}$.

(b) $\varepsilon_i = \theta v_{1i} + r_{1i}$, where $\theta = \frac{\sigma_{\varepsilon v}}{\sigma_v^2}$, and r_{1i} is independent of v_{1i}.

(c) $E\left(x_{2j}\varepsilon_i\right) = 0$.

5. $\mathbf{z}_{1i}$ are valid and relevant excluded instruments for x; that is,

$$E(\varepsilon_i|\mathbf{z}_{1i}) = E(\varepsilon_i|\mathbf{z}_{2i}) = 0, E(v_{1i}|\mathbf{z}_{1i}) = 0, E(v_{2j}|\mathbf{z}_{2j}) = 0, \gamma \neq 0$$

6. The ratio of S1 and S2 converges to a fixed positive number in large samples, that is,

$$plim_{N_1,N_2\to\infty}\frac{N_1}{N_2} = \alpha.$$

2.2 Definitions of Estimators

In practice, the TS2SLS estimator involves generating an estimate of the first stage parameter γ, $\hat{\gamma}_2$, using N_2 observations with nonmissing values of x and z, generating N_1 cross-sample fitted values $\hat{x}_{1i} = z_{1i}\hat{\gamma}_2$, and then regressing y_1 on $\hat{x}_1$ via OLS to estimate β. To facilitate a clearer understanding of the estimator, we express TS2SLS as a weighted combination of 2SLS and SS2SLS estimators on different subsets of the data, whose sizes depend on the degree of overlap between samples $S1$ and $S2$. Intuitively, the 2SLS component of the estimator is estimated using all units which are shared between S1 and S2; the SS2SLS component is estimated using all units which lie exclusively within S1 or S2. Accordingly, the TS2SLS estimator is equivalent to 2SLS for $\rho = 1$

and SS2SLS for $\rho = 0$ when $N_1 = N_2$. Note that these individual estimates may rely on data that is not observed in the practical setting in which we are considering estimators; the expression of a weighted average of estimators primarily serves the purpose of providing a more interpetable and algebraically convenient starting point for deriving the estimator's properties.

Let $Y_1 \equiv \begin{pmatrix} y_{11} \\ \vdots \\ y_{1N_1} \end{pmatrix} \equiv \begin{pmatrix} Y_{11} \\ Y_{12} \end{pmatrix}, \widehat{X}_1 \equiv \begin{pmatrix} \widehat{x}_{11} \\ \vdots \\ \widehat{x}_{1N_1} \end{pmatrix} \equiv \begin{pmatrix} \widehat{z}_{11}\hat{\gamma}_2 \\ \vdots \\ \widehat{z}_{1N_1}\hat{\gamma}_2 \end{pmatrix} \equiv \begin{pmatrix} \widehat{X}_{11} \\ \widehat{X}_{12} \end{pmatrix}, \mathbf{Z}_1 \equiv \begin{pmatrix} \mathbf{z}_{11} \\ \vdots \\ \mathbf{z}_{1N_1} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{Z}_{11} \\ \mathbf{Z}_{12} \end{pmatrix}$, $\Xi_1 \equiv \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{N_1} \end{pmatrix} \equiv \begin{pmatrix} \Xi_{11} \\ \Xi_{12} \end{pmatrix}$ and $V_1 \equiv \begin{pmatrix} v_{11} \\ \vdots \\ v_{1N_1} \end{pmatrix} \equiv \begin{pmatrix} V_{11} \\ V_{12} \end{pmatrix}$ be the vectors in S1, where vectors Y_{11}, $\widehat{X}_{11}$, $\mathbf{Z}_{11}$, Ξ_{11}, V_{11} are the first ρN_2 rows, and vectors Y_{12}, $\widehat{X}_{12}$, $\mathbf{Z}_{12}$, Ξ_{12}, V_{12} are the remaining $N_1 - \rho N_2$ rows. In particular, let $X_1 \equiv \begin{pmatrix} x_{11} \\ \vdots \\ x_{1N_1} \end{pmatrix} \equiv \begin{pmatrix} X_{11} \\ X_{12} \end{pmatrix}$, where we observe X_{11} but not X_{12}. Similarly, let $X_2 \equiv \begin{pmatrix} x_{21} \\ \vdots \\ x_{2N_2} \end{pmatrix} \equiv \begin{pmatrix} X_{11} \\ X_{22} \end{pmatrix}$, $\mathbf{Z}_2 \equiv \begin{pmatrix} \mathbf{z}_{21} \\ \vdots \\ \mathbf{z}_{2N_2} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{Z}_{11} \\ \mathbf{Z}_{22} \end{pmatrix}$ and $V_2 \equiv \begin{pmatrix} v_{21} \\ \vdots \\ v_{2N_2} \end{pmatrix} \equiv \begin{pmatrix} V_{11} \\ V_{22} \end{pmatrix}$ be the vectors in S2, where vectors X_{11}, $\mathbf{Z}_{11}$ are the same as the first ρN_2 rows in S1, and vectors X_{22} and $\mathbf{Z}_{22}$ are the remaining $(1-\rho)N_2$ rows in S2.

The TS2SLS estimator is defined by

$$\widehat{\beta}_O = \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\widehat{X}_1'Y_1.$$

Proposition 1. *By the proof in the appendix,* $\widehat{\beta}_O$ *can be rewritten as*

$$\begin{aligned}\widehat{\beta}_O &= \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{11}'\widehat{X}_{11}\right)\left(\widehat{X}_{11}'\widehat{X}_{11}\right)^{-1}\left(\widehat{X}_{11}'Y_{11}\right) \\ &\quad + \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{12}'\widehat{X}_{12}\right)\left(\widehat{X}_{12}'\widehat{X}_{12}\right)^{-1}\left(\widehat{X}_{12}'Y_{12}\right) \\ &= \widehat{W}\widehat{\beta}^{(1)}_{2SLS} + \left(1-\widehat{W}\right)\widehat{\beta}^{(2)}_{SS2SLS} \qquad (6)\end{aligned}$$

where

$$\hat{W} \equiv \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{11}'\widehat{X}_{11}\right),$$

$$\widehat{\beta}^{(1)}_{2SLS} = \left(\widehat{X}'_{11}\widehat{X}_{11}\right)^{-1}\left(\widehat{X}'_{11}Y_{11}\right),$$

and

$$\widehat{\beta}^{(2)}_{SS2SLS} = \left(\widehat{X}'_{12}\widehat{X}_{12}\right)^{-1}\left(\widehat{X}'_{12}Y_{12}\right).$$

Proposition 2. *The probability limit of $\hat{W}$ as both samples approach infinity is the ratio of the overlap parameter to the asymptotic ratio of sample sizes.*

$$plim\widehat{W} \equiv W = \frac{\rho}{\alpha}.$$

Remark 3. The expected value of $\hat{W}$ can be approximated as follows (see appendix):

$$E\left(\widehat{W}\right) \approx \frac{\rho N_2 \gamma' \Omega_z \gamma + K \cdot \sigma_v^2}{N_1 \gamma' \Omega_z \gamma + K \cdot \sigma_v^2 + (N_1 - \rho N_2) / (1-\rho) N_2 \sigma_v^2}.$$

Proposition 1 formalizes the partition of $\widehat{\beta}_O$ into a weighted average of $\widehat{\beta}^{(1)}_{2SLS}$ and $\widehat{\beta}^{(2)}_{SS2SLS}$. The weights represent the sum of squares used by each estimator relative to the total sum of squares in the entire vector of data. This variation has two components: variation explicitly from instruments Z, and variation from first stage error vectors V which manifests through the estimated first stage coefficients. Asymptotically, the weights are a function of the degree of overlap between samples and the ratio of sample sizes.

2.3 First-order bias approximation

There is an expansive literature on the finite-sample bias of IV estimators, with papers such as Nagar (1959), Bekker (1994), Staiger and Stock (1997), and Bun and Windmeijer (2010) offering approximations of various forms. For simplicity, we consider first-order approximations of the bias of 2SLS and SS2SLS and develop a bias approximation for TS2SLS. Hahn and Hausman (2002) offer a simple approximation of the bias of 2SLS, using the product of the inverse expected value of the variance of fitted values to the expected value of the covariance between the fitted values and the outcome. In a scalar case, this is in effect using the ratio of expectations in place of the expected value of the ratio. This approximation corresponds to a first-order Taylor Series expansion of the 2SLS estimator about each of these "numerator" and "denominator" terms. For 2SLS, this approximation sufficiently characterizes the directional responses of bias to sample size, number of instruments, error covariance, and instrument strength. The same holds for SS2SLS, which has a comparable shape of response but has bias characterized enitrely by first-stage sampling error.

Proposition 4. *The finite-sample bias for the 2SLS estimator on the overlapping portion of the sample is approximated by a first-order Taylor expansion:*

$$E(\widehat{\beta}^{(1)}_{2SLS} - \beta) \approx \frac{K \cdot \sigma_{\varepsilon v}}{\rho N_2 \gamma_1' \Omega_z \gamma_1 + K \cdot \sigma_v^2}. \tag{7}$$

The approximate finite-sample bias for the SS2SLS estimator on the non-overlap portion of the sample is given by

$$\begin{aligned} E(\widehat{\beta}^{(2)}_{SS2SLS} - \beta) &\approx -\frac{\sigma_v^2 \beta / (1-\rho) N_2}{\gamma' \Omega_{\mathbf{z}} \gamma + \sigma_v^2 / (1-\rho) N_2} \\ &= -\frac{\beta}{(1-\rho) N_2 \gamma' \Omega_{\mathbf{z}} \gamma / \sigma_v^2 + 1} \end{aligned} \tag{8}$$

Equation (7) and equation (8) show that the bias of both the 2SLS estimator and SS2SLS estimator approach zero as N_2 gets large: they are asymptotically unbiased in first stage sample size. This conforms with the intuition that all finite-sample bias in 2SLS (under valid instruments) originates from first-stage sampling error. Both the direction and magnitude of bias in the 2SLS estimator depends on $\sigma_{\varepsilon v}$, the covariance of error terms representing the endogenous portion of x. SS2SLS has a attenuation bias that is inversely proportional to $(1-\rho)N_2\gamma'\Omega_{\mathbf{z}}\gamma/\sigma_v^2$, the first-stage "concentration parameter," intuitively similar to a bias from measurement error. Note that this attenuation, found similarly in Angrist and Krueger (1995), is dependent on the assumption of a linear conditional expectation function in the first stage. An application of Jensen's inequality to the SS2SLS estimator suggests that the attenuation bias may not generally hold (see appendix).

Proposition 5. *For $0 < \rho < 1$ and $N_2 \neq N_1$, the finite sample bias of $\widehat{\beta}_O$ is approximated by*

$$\begin{aligned} E\left(\widehat{\beta}_O - \beta\right) &= E\left[\widehat{W}\left(\widehat{\beta}^{(1)}_{2SLS} - \beta\right) + \left(1 - \widehat{W}\right)\left(\widehat{\beta}^{(2)}_{SS2SLS} - \beta\right)\right] \\ &\approx \frac{K \cdot \sigma_{\varepsilon v} - ((N_1 - \rho N_2) / (1-\rho) N_2) \sigma_v^2 \beta}{N_1 \gamma' \Omega_z \gamma + K \cdot \sigma_v^2 + ((N_1 - \rho N_2) / (1-\rho) N_2) \sigma_v^2}. \end{aligned} \tag{9}$$

$\widehat{\beta}_O$ is approximately unbiased when the overlap proportion ρ is set according to the following formula (noting $\theta = \frac{\sigma_{\varepsilon v}}{\sigma_v^2}$ and K is the number of instruments):

For $\frac{N_2}{N_1} < 1$ and $\beta < K\theta \frac{N_2}{N_1}$,

$$\rho = \frac{K\theta - \beta \frac{N_1}{N_2}}{K\theta - \beta} \in (0,1),$$

For $\frac{N_2}{N_1} > 1$ and $\beta > K\theta\frac{N_2}{N_1}$, set the overlap proportion to

$$\rho = \frac{\beta\frac{N_1}{N_2} - K\theta}{\beta - K\theta} \in (0,1)$$

Proposition 5 follows from a first-order Taylor series approximation of $E(\widehat{\beta}_O)$, derived in the appendix. The second part is shown by setting the numerator in equation (9) to zero and solving for ρ. The net bias of TS2SLS is dependent on the overlap parameter ρ and sampling variation in the first stage parameters estimated from the units used in TS2SLS estimation expressed through the ratio $\hat{W}$.

Because of the nature of the approximation used, the cases with $\rho = 1$, $\rho = 0$, or $N_1 = N_2$ are poorly characterized or undefined. We provide conjectures for a subset of these cases. With $N_1 = N_2$, the no-overlap ($\rho = 0$) case corresponds exactly to the "split-sample" estimator and is biased toward zero, a finding consistent with Angrist and Krueger (1995). The total overlap ($\rho = 1$) case with $N_1 = N_2$ results in an estimate computationally identical to 2SLS; the two-step nature of 2SLS means that the ability to explicitly link the units used in the steps has no bearing on the estimate if the units used in the steps are indeed the same. When the combination of β and $\sigma_{\varepsilon v}$ makes the direction of bias for the two estimators have different signs, the overlap parameterρ can be tuned to yield an approximately unbiased estimator $\widehat{\beta}_O$.

2.4 Asymptotic Variance of TS2SLS

Proposition 6. $\widehat{\beta}_O$ *is consistent and asymptotically normally distributed with asymptotic variance*

$$\sqrt{N_1 + (1-\rho)N_2}\left(\widehat{\beta}_O - \beta\right) \overset{a}{\sim} N\left\{0, \frac{(1+\alpha-\rho)\rho}{\alpha^2}\sigma_v^2\left[\Omega_{x\mathbf{z}}\Omega_{\mathbf{z}}^{-1}\Omega_{x\mathbf{z}}\right]^{-1} + \frac{(1+\alpha-\rho)(\alpha-\rho)}{\alpha^2}\left[\Omega_{x\mathbf{z}}'\left[\left(\sigma_u^2 + \frac{\alpha-\rho}{1-\rho}\beta'\sigma_v^2\beta\right)\Omega_{\mathbf{z}}\right]^{-1}\Omega_{x\mathbf{z}}\right]^{-1}\right\} \quad (10)$$

This result follows from the fact that the limiting distribution of $\widehat{\beta}_O$ is a linear combination of the two estimators $\widehat{\beta}^{(1)}_{2SLS}$ and $\widehat{\beta}^{(2)}_{SS2SLS}$ with weight $\frac{\rho}{\alpha}$ (see appendix for proof).

TS2SLS has an asymptotic variance that accounts for variation in the final estimate $\widehat{\beta}_O$ due to first-stage sampling error, but only due to the error originating in the SS2SLS component. The conventional 2SLS asymptotic variance treats the first stage parameter as known (Wooldridge 2010), and so any variability from the first stage of the 2SLS component of the TS2SLS estimator is ignored. The stabilizing factor implies that additional first stage observations have a discounted return to precision relative to N_1.

Inoue and Solon (2008) derive the asymptotic variance of TS2SLS, but do so only for the case of independent primary and auxiliary samples (i.e., $\rho = 0$). Allowing the samples to generally overlap, we find that the asymptotic variance is decreasing in ρ. The basic intuition underlying this property is that sampling variation in the first stage parameter coming from a secondary, independent sample is additional noise originating with error term v_2 unrelated to the outcome y_1. First stage estimate sampling variation coming from the same units used in second stage estimation (as they would in the typical 2SLS computation implied by $\rho = 1$) have a component with explanatory power for outcome y through the error component of x_1, v_1.

3 Simulation Evidence

We conduct a series of Monte Carlo simulations of the model characterized in Section 2.1, setting the following parameters:

$$\beta = 1$$

$$\begin{pmatrix} \varepsilon \\ v \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} .5 & .4 \\ .4 & .5 \end{pmatrix})$$

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, K = 2$$

$$z_1 \sim N(0, 2.5), z_2 \sim N(0, 2.5)$$

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} 0.0316 \\ 0.0316 \end{pmatrix}$$

Figure 1 plots the average parameter estimate across simulation repetitions for different first stage sample sizes N_2, approximating the expected values of 2SLS and TS2SLS estimators. The 2SLS estimator uses $N = N_2$ observations, while the TS2SLS estimator uses two non-overlapping data sets ($\rho = 0$) of $N_1 = 200$ and N_2observations. In this setup, the OLS estimate is biased upward by about 80%, and the 2SLS estimate is biased toward the OLS estimate, while the TS2SLS estimate is biased towards zero. For both estimators, as N_2 increases the bias decreases. This simulation illustrates one of the potential bias advantages of using a second sample even when a single-sample estimate is available. A TS2SLS estimate using $N_1 = 200$ and $N_2 \geq 400$ is approximately unbiased while the 2SLS estimate with $N = 200$ is biased by around 6%, suggesting the TS2SLS estimate is preferable (notwithstanding precision).

Figure 2 plots the average standard deviation across simulation repetitions for first stage sample sizes N_2, revealing that the variance of TS2SLS is decreasing in the number of first stage observations but eventually flattening as the first stage becomes precisely estimated while the 2SLS estimator continues to grow in precision (because of additional data for the second stage).

The bias consequences of overlap dovetail with the description of TS2SLS as a linear combination of 2SLS and SS2SLS estimators. Figure 4 plots the mean simulated point estimates for the TS2SLS estimator as a function of the percentage of overlap between samples for $N_1 = N_2 = 200$. Increasing overlap between samples weighs the estimate towards 2SLS, which is biased towards the probability limit of OLS. In this calibration, the sample-varying portion of $\hat{W}$ is small enough such that the weight is nearly the probability limit of $\hat{W}$, $\frac{\rho}{\alpha} = \rho$. The point of zero bias occurs approximately at 57 percent overlap, which the approximation from proposition 5 is unable to predict. On the other hand, the effect on estimate variance of overlap is predicted by the asymptotic variance expression in proposition 6. Figure 5 shows a plot of the simulated standard error against the percentage overlap parameter, demonstrating the anticipated downward-sloping relationship. In this setting, 2SLS is always more efficient than SS2SLS using an equivalent number of observations per stage. Overlap results in a mixture of the two estimators, with more observations used in (and thus greater weight placed on) the 2SLS component when the degree of overlap increases.

4 Application

4.1 TS2SLS in Practice: Synthetic Example from Angrist and Evans (1998)

Angrist and Evans (1998) estimate the effect of fertility on parents' labor supply decisions. They use gender composition indicators of a family's first two children as instruments for whether the family has more than two children. In one set of estimates, they use Census microdata to estimate IV regressions of various labor market outcomes (e.g., whether the mother worked, how much she worked, and her labor income) on an indicator for more than two children, using indicators for the first two children being boys and first two children being girls as instruments.[1] We create hypothetical situations in which some data are missing and examine whether the study's findings could have been recreated using TS2SLS estimates under these conditions. The hypothetical TS2SLS estimates mimic what the authors might have needed to run if their data were only available in two samples, as they might have been if the Census had decided to release their first-stage variables in one anonymized microdata set and their second-stage variables in another.

We attempt this exercise for Angrist and Evans' estimates for married women ages 21-35 with 2 or more children. We develop hypothetical scenarios in which the authors receive their data in

[1]There is some controversy as to the validity of the family gender-composition instruments they use (e.g., Rosenzweig and Wolpin 2000). This exercise takes no position in this debate and only focuses on the comparison of the bias and variance mechanics of the estimators used as they relate to one-sample vs. two-sample 2SLS.

two data subsets of their true data on $\{y, x, z\}$. We simulate multiple draws of data sets for each estimator in order to isolate the expected value of each estimator across possible data draws the hypothetical data-constrained Angrist and Evans might have faced. For example, we randomly draw two mutually exclusive and exhaustive 50-percent subsets of the 254,654 observations and compute the SS2SLS estimate with each sample, repeating the procedure a total of 1000 times. In reality, practitioners would face a single draw of data, and sampling variation could result in a pattern of results that does not conform with theoretical predictions.

Table 6 shows a series of estimates for various hypothetical situations in which the authors have restricted access to some part of their data. First, we exactly replicate the estimates from Table 7, columns 4 and 6 (p. 465) in the original paper. As a baseline, we show what the 2SLS estimates would have been had the authors only had access to a 50 percent subsample of their data. Column 5 shows what the split-sample 2SLS estimate using 50% of their sample for each stage would have been, had they only had access to x and z in one set and y and z in another. Column 6 shows the estimate if they had access to 100% of their original sample for the first stage, but only 50% for the second stage. Finally, column 6 shows the TS2SLS estimate if they had 50% of their original sample for the first stage, but 100% of their sample for the second stage (y and z).

4.2 Other Considerations for Applications

There are several meaningful practical applications of two-sample estimators. The typical application empirical literature using two-sample estimators is when the researcher only has access to one data set with the outcome and instrument and one data set with the endogenous variable and instrument; in this case, the researcher's only option to generate an IV estimate is to use a TSIVM. The second application, strongly suggested by the finite-sample findings here, is when researchers have access to a single-sample IV estimator but have additional observations usable to estimate the first stage (endogenous variable and insturment) or second stage (outcome and instrument) relationships.

The simulation evidence and synthetic working example from Angrist and Evans (1998) suggest that reporting a TS2SLS estimate may be the most rewarding when practitioners have access to additional second-stage observations. The TS2SLS estimate using all available observations could then be substantially more precise without a need for having equal numbers of observations for the first and second-stage samples. In that case, the strength of the instrument is likely to be established, and practitioners need only justify that their additional observations come from a population with the same first-stage projection coefficients. Even when that justification is lacking, the two-sample estimates can simply be presented alongside the single-sample estimates as additional evidence for a causal hypothesis, allowing the reader to decide the weight of evidence to assign to the TS2SLS estimate. A secondary case of interest is when practitioners have a single sample 2SLS estimate with weak instruments (i.e., an imprecisely-estimated first stage) but have access to additional observations with which to estimate the first stage. In this case, the TS2SLS estimate can potentially "solve" a weak instruments problem, but then the justification of first-stage coefficient equivalence between the two samples is of much more important, since the primary causal inference in the paper comes from the TS2SLS estimate.

In practice, these situations occur because of real-world limitations on the way data can be collected. Data on the endogenous variable may be costly to collect and thus limited (e.g., environmental monitoring of pollutants, accurate personal income measures), but sufficient to estimate a strong, representative first stage relationship with an instrument. With weak instruments bias not a concern, a study could then be scaled to have adequate power only through the expansion of data collection on the second-stage varaibles.

Practical issues with data linkage also make TSIVM potentially useful, especially where the providers of data can perfectly manipulate what data is available to the public. The complete interchangeability of two-sample estimators with one-sample estimators when the data sampling process is exactly known is a useful property for confidentiality applications. Consider a situation in which the combination of information on y, x, and z for subjects in a randomized study could allow anonymous subjects to be identified, but providing information on just x and z or just y and z reduce that risk significantly. Data releases could then offer two data sets with those elements separate and unlinkable, but researchers could still generate IV estimates using TSIVM at little cost to their inferences. In this setting, two of the greatest sources of uncertainty in the validity of TS2SLS estimates are obviated: the populations constituting each of the samples are known to be equivalent and the level of overlap between the primary and secondary samples is generated by the study designers and passed on to the practitioner.

Lastly, as a generalization of split-sample IV methods, the results presented in this paper could also be used as part of an "eyeball test" for weak instruments. Practitioners can subdivide their samples arbitrarily and run a series of TS2SLS estimates using the subsamples, examining the sensitivity of their coefficient estimates to smaller first stage sample sizes or varying levels of overlap. This methodology may be superseded by other weak-instruments tests or bias-robust IV estimation methods such as Jackknife IV (Angrist et al. 1999), but it is also easy to implement and intuitively present to an audience using plots of estimates across researcher-manipulated overlap or sample size parameters.

5 Conclusion

This paper introduces new considerations for applied researchers using TS2SLS. First, we offer a new way of expressing the TS2SLS estimator as a weighted average of 2SLS and SS2SLS estimators using the underlying sample units. The weights structurally depend on the degree to which the researcher's auxiliary sample overlap with those in the primary sample, overlapping units used in the 2SLS component and only non-overlapping units used in the SS2SLS component. A first-order approximation characterizes the behavior of the TS2SLS estimator in finite samples in conformity with the typical intuition from the weak instruments literature for 2SLS: the magnitude of bias is always related to the degree of sampling error in the first stage parameter estimate. However, the bias is pulled in two competing directions: toward zero to the extent that the two samples are non-overlapping, and toward the probability limit of OLS to the extent that the two samples are overlapping. We also show that the variance of the TS2SLS estimator is decreasing in both first- and second-stage sample sizes as well as the degree of overlap. We replicate our theoretical findings using both simulation methods and an empirical example from Angrist and Evans (1998), noting that TS2SLS can perform as well as 2SLS with equivalent sample size when instruments are strong.

These results show the potential for TS2SLS to be useful in empirical studies beyond its traditional use as a solution to missing data, while also suggesting that TS2SLS estimates should be interpreted with caution when samples have an unknown level of overlap. Researchers may have access to many different data sets with multiple options for how to combine them into IV estimates, also potentially having access to both single-sample and two-sample estimates. Any uses of TS2SLS, whether to present alone or alongside single-sample estimates, must carefully consider the populations represented by each sample.

References

[1] Anderson, M. L., & Matsa, D. A. (2011). Are restaurants really supersizing America? American Economic Journal: Applied Economics, 152–188.

[2] Angrist, J. D., & Evans, W. N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. The American Economic Review, 88(3), 450–477.

[3] Angrist, J. D., Imbens, G. W., & Krueger, A. B. (1999). Jackknife instrumental variables estimation. Journal of Applied Econometrics, 14(1), 57–67. http://doi.org/10.1002/(SICI)1099-1255(199901/02)14:1<57::AID-JAE501>3.0.CO;2-G

[4] Angrist, J. D., & Krueger, A. B. (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(418), 328–336. http://doi.org/10.2307/2290263

[5] Angrist, J. D., & Krueger, A. B. (1995). Split-Sample Instrumental Variables Estimates of the Return to Schooling. Journal of Business & Economic Statistics, 13(2), 225–235. http://doi.org/10.2307/1392377

[6] Arellano, M., & Meghir, C. (1992). Female Labour Supply and On-the-Job Search: An Empirical Model Estimated Using Complementary Data Sets. The Review of Economic Studies, 59(3), 537–559. http://doi.org/10.2307/2297863

[7] Bekker, P. A. (1994). Alternative Approximations to the Distributions of Instrumental Variable Estimators. Econometrica, 62(3), 657–681. http://doi.org/10.2307/2951662

[8] Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. Journal of the American Statistical Association, 90(430), 443–450. http://doi.org/10.2307/2291055

[9] Bun, M. J. G., & Windmeijer, F. (2011). A comparison of bias approximations for the two-stage least squares (2SLS) estimator. Economics Letters, 113(1), 76–79. http://doi.org/10.1016/j.econlet.2011.05.047

[10] Den Berg, G. J., Pinger, P. R., & Schoch, J. (2015). Instrumental variable estimation of the causal effect of hunger early in life on health later in life. The Economic Journal. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/ecoj.12250/abstract

[11] Devereux, P. J., & Hart, R. A. (2010). Forced to be Rich? Returns to Compulsory Schooling in Britain. The Economic Journal, 120(549), 1345–1364.

[12] Gong, H., Leigh, A., & Meng, X. (2012). Intergenerational income mobility in urban China. Review of Income and Wealth, 58(3), 481–503.

[13] Hahn, J., & Hausman, J. (2002). Notes on bias in estimators for simultaneous equation models. Economics Letters, 75(2), 237–241. http://doi.org/10.1016/S0165-1765(01)00602-4

[14] Hahn, J., Hausman, J., & Kuersteiner, G. (2004). Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. The Econometrics Journal, 7(1), 272–306.

[15] Inoue, A., & Solon, G. (2010). Two-Sample Instrumental Variables Estimators. Review of Economics and Statistics, 92(3), 557–561. http://doi.org/10.1162/REST_a_00011

[16] Klevmarken, N. A. (1982). On the Stability of Age-Earnings Profiles. The Scandinavian Journal of Economics, 84(4), 531–554. http://doi.org/10.2307/3439516

[17] Nagar, A. L. (1959). The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations. Econometrica, 27(4), 575–595. http://doi.org/10.2307/1909352

[18] Nicoletti, C., & Ermisch, J. F. (2008). Intergenerational earnings mobility: changes across cohorts in Britain. The BE Journal of Economic Analysis & Policy, 7(2). Retrieved from http://www.degruyter.com/view/j/bejeap.2007.7.2/bejeap.2007.7.2.1755/bejeap.2007.7.2.1755.xml

[19] Olivetti, C., & Paserman, M. D. (2014). In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940. Retrieved from http://people.bu.edu/olivetti/papers/Olivetti-Paserman_NameOfTheSon_July2014.pdf

[20] Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural“ natural experiments” in economics. Journal of Economic Literature, 827–874.

[21] Rothstein, J., & Wozny, N. (2013). Permanent income and the black-white test score gap. Journal of Human Resources, 48(3), 510–544.

[22] Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.

6 Figures and Tables

Figure 1: Mean Simulated TS2SLS Point Estimate by First Stage Sample Size N_2

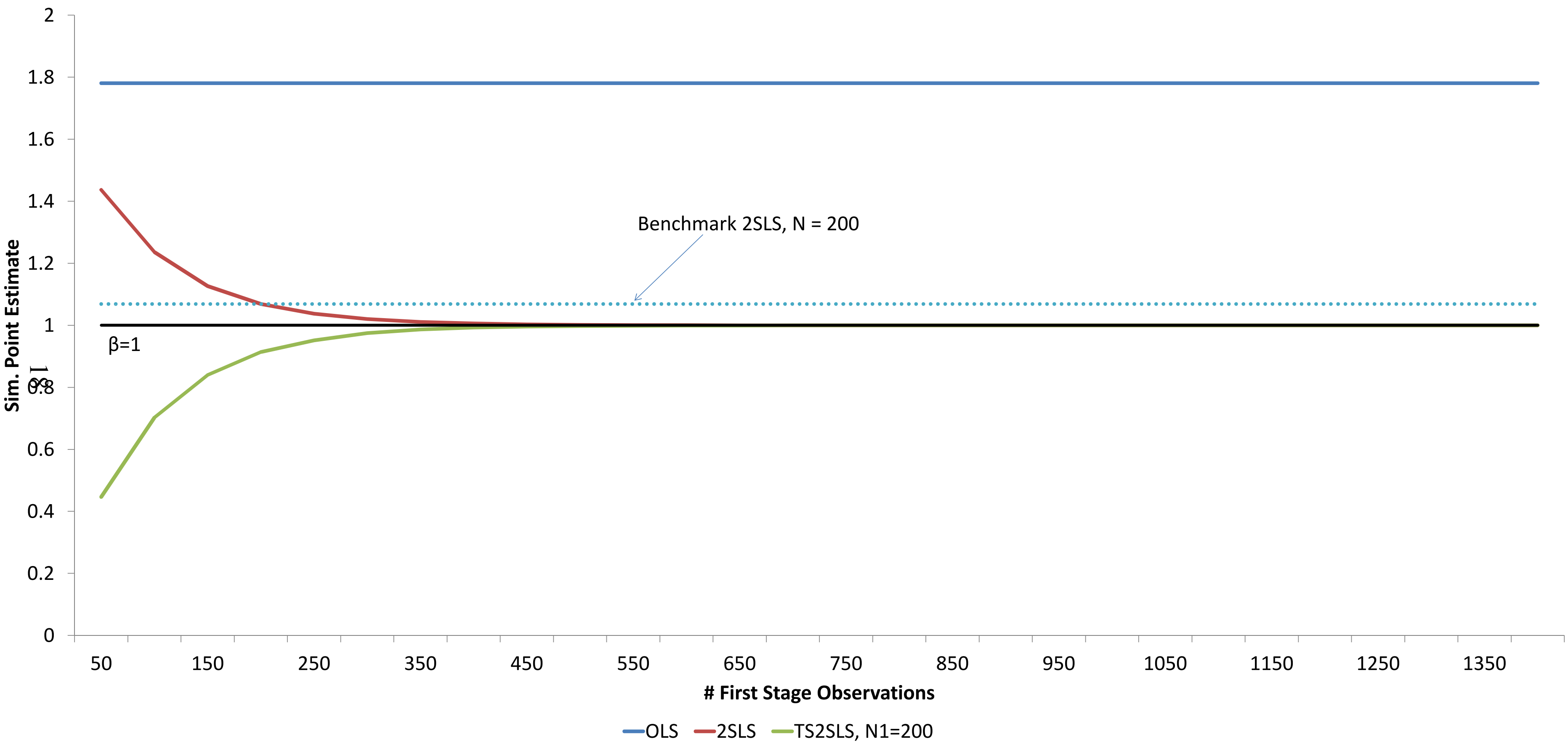

Figure 2: Simulated TS2SLS Standard Error by First Stage Sample Size N_2

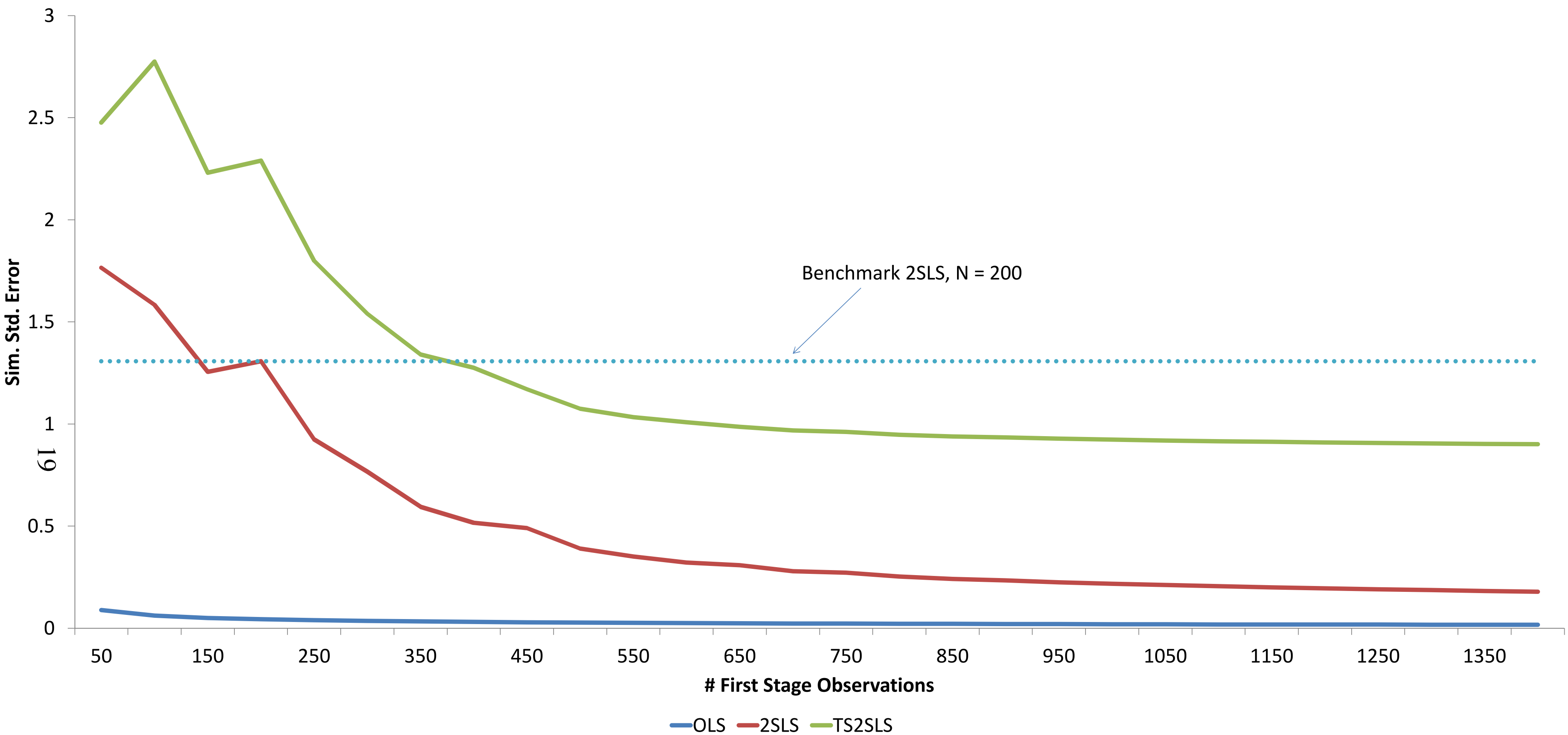

Figure 3: Simulated TS2SLS Standard Error by Second Stage Sample Size N_1

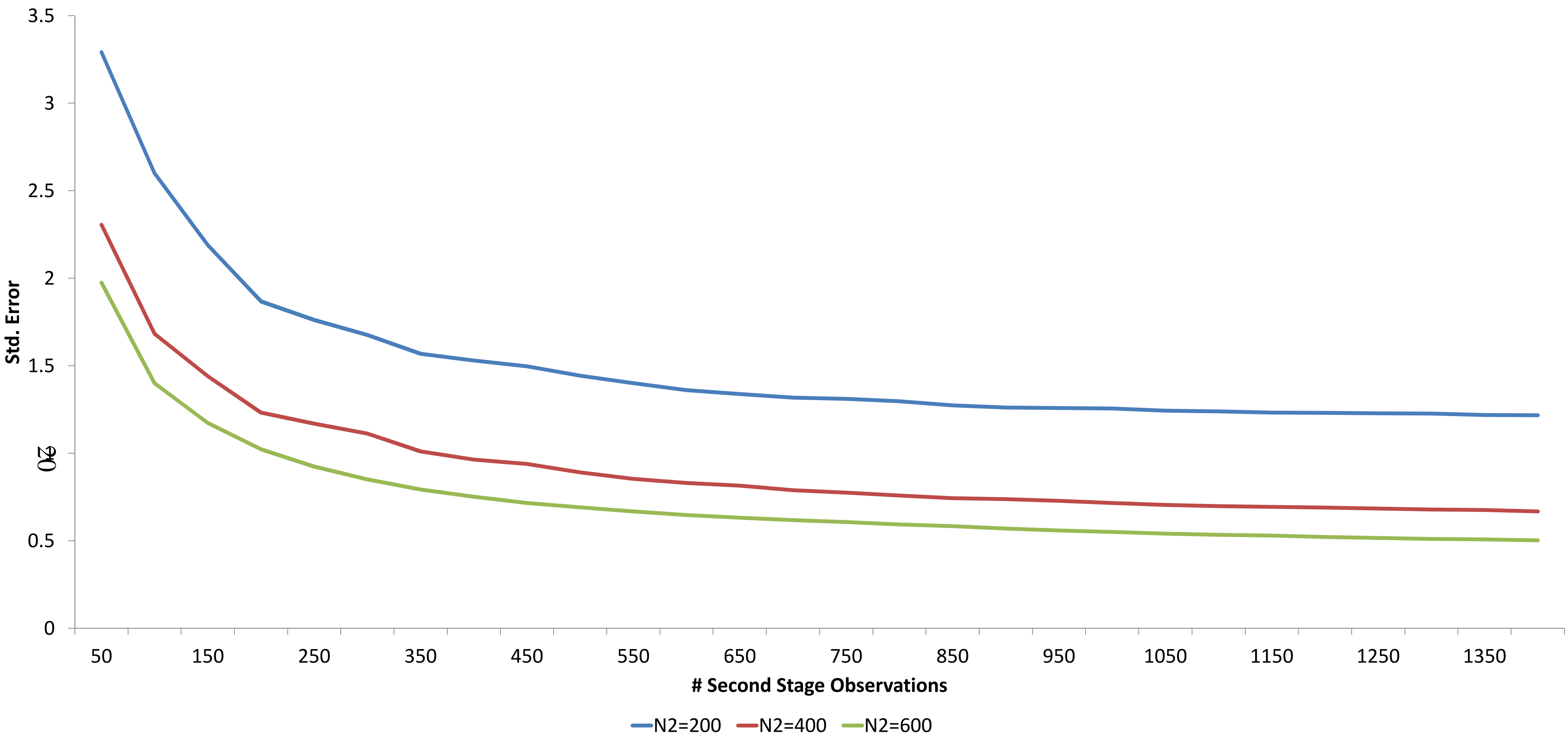

Figure 4: Mean Simulated TS2SLS Point Estimate by Proportion of Overlap Between Samples

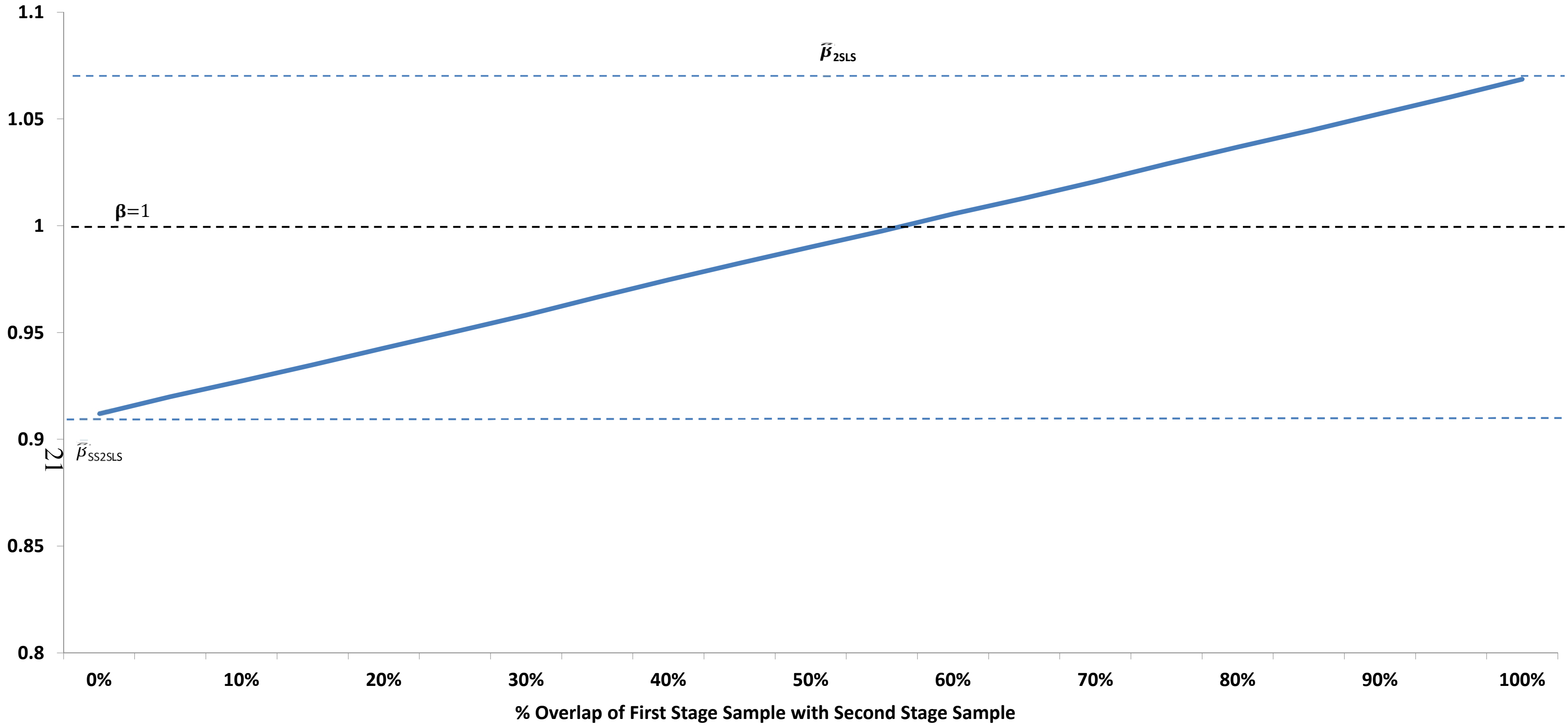

Figure 5: TS2SLS Simulated Standard Error by Proportion of Overlap Between Samples

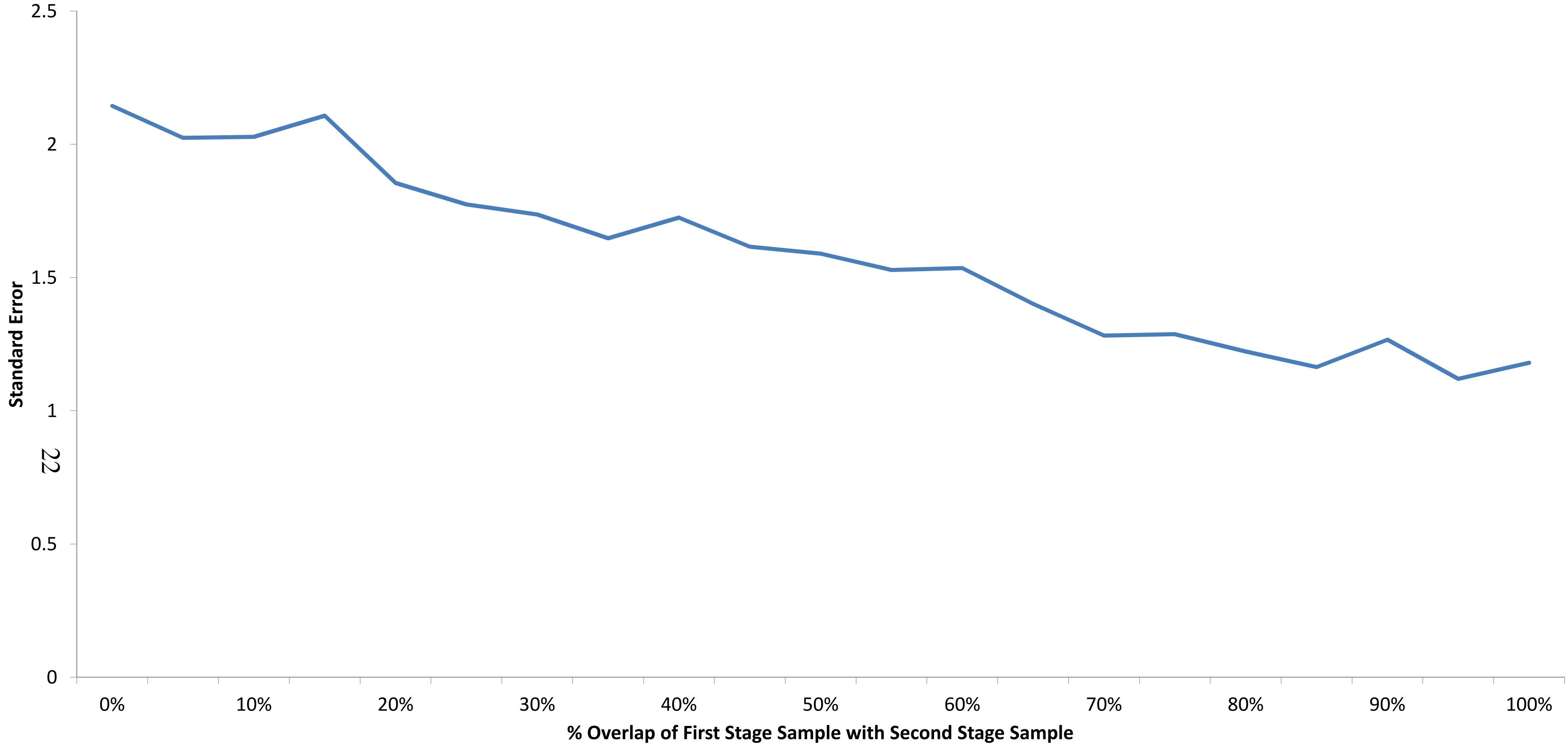

Figure 6: Hypothetical Two-Sample Estimates for Angrist and Evans (1998), Effects of 2 or More Children on Labor Supply for Married Women, 21-35

	Original Estimates		**Split-Sample Estimates (Bootstrapped)**			
	OLS	**2SLS**	**50% Sample 2SLS**	**50-50 SS/TS2SLS**	**50% Sample + 100% First Stage**	**50% Sample + 100% Reduced-Form**
Worked?	-0.167	-0.113	-0.086	-0.092	-0.091	-0.114
	(0.002)	(0.028)	(0.039)	(0.041)	(0.041)	(0.028)
# Weeks Worked	-8.043	-5.164	-4.461	-5.037	-4.981	-5.242
	(0.086)	(1.156)	(1.694)	(1.716)	(1.694)	(1.158)
# Hours/Week	-6.021	-4.613	-3.678	-2.972	-2.938	-4.682
	(0.075)	(1.023)	(1.412)	(1.521)	(1.5)	(1.038)
Labor Income	-3165.4	-1321.2	-836.5	-567.6	-569.7	-1344.3
	(39.41)	(550)	(793.53)	(819.14)	(813.72)	(550.03)
N1	254,654	254,654	127,327	127,327	127,327	254,654
N2	--	--	--	127,327	254,654	127,327

7 Appendix

Proof for proposition 1

Under the regularity condition that the data matrix $\widehat{X}_1'\widehat{X}_1$ is nonsingular w.p.a. 1,

$$\begin{aligned}
\widehat{\beta}_O &= \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\widehat{X}_1'Y_1 \\
&= \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left[\left(\widehat{X}_{11}',\widehat{X}_{12}'\right)\begin{pmatrix}Y_{11}\\Y_{12}\end{pmatrix}\right] \\
&= \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{11}'Y_{11}+\widehat{X}_{12}'Y_{12}\right) \\
&= \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{11}'\widehat{X}_{11}\right)\left(\widehat{X}_{11}'\widehat{X}_{11}\right)^{-1}\left(\widehat{X}_{11}'Y_{11}\right) \\
&\quad+\left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{12}'\widehat{X}_{12}\right)\left(\widehat{X}_{12}'\widehat{X}_{12}\right)^{-1}\left(\widehat{X}_{12}'Y_{12}\right)
\end{aligned}$$

Denote $W \equiv \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{11}'\widehat{X}_{11}\right),\ 1-W \equiv \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{12}'\widehat{X}_{12}\right)$, and that

$$\begin{aligned}
\widehat{\beta}_{2SLS}^{(1)} &= \left(\widehat{X}_{11}'\widehat{X}_{11}\right)^{-1}\left(\widehat{X}_{11}'Y_{11}\right), \\
\widehat{\beta}_{SS2SLS}^{(2)} &= \left(\widehat{X}_{12}'\widehat{X}_{12}\right)^{-1}\left(\widehat{X}_{12}'Y_{12}\right),
\end{aligned}$$

from which $\widehat{\beta}_O = \widehat{W}\widehat{\beta}_{2SLS}^{(1)}+\left(1-\widehat{W}\right)\widehat{\beta}_{SS2SLS}^{(2)}$ follows.

Proof for proposition 2

Decompose $\widehat{W}$ into the following block vectors,

$$\begin{aligned}
\widehat{W} &= \left(\widehat{X}_1'\widehat{X}_1\right)^{-1}\left(\widehat{X}_{11}'\widehat{X}_{11}\right) \\
&= \left(\widehat{\gamma}_1'\mathbf{Z}_{11}'\mathbf{Z}_{11}\widehat{\gamma}_1+\widehat{\gamma}_2'\mathbf{Z}_{12}'\mathbf{Z}_{12}\widehat{\gamma}_2\right)^{-1}\left(\widehat{\gamma}_1'\mathbf{Z}_{11}'\mathbf{Z}_{11}\widehat{\gamma}_1\right) \\
&= \left[\rho N_2\widehat{\gamma}_1'\frac{\mathbf{Z}_{11}'\mathbf{Z}_{11}}{\rho N_2}\widehat{\gamma}_1+(N_1-\rho N_2)\widehat{\gamma}_2'\frac{\mathbf{Z}_{12}'\mathbf{Z}_{12}}{N_1-\rho N_2}\widehat{\gamma}_2\right]^{-1}\left(\rho N_2\widehat{\gamma}_1'\frac{\mathbf{Z}_{11}'\mathbf{Z}_{11}}{\rho N_2}\widehat{\gamma}_1\right)
\end{aligned}$$

By assumption 1.(a) $\text{plim}\frac{\mathbf{Z}_{11}'\mathbf{Z}_{11}}{\rho N_2} = E(\mathbf{z}_{1i}'\mathbf{z}_{1i}) = \text{plim}\ \frac{\mathbf{Z}_{12}'\mathbf{Z}_{12}}{N_1-\rho N_2} = E(\mathbf{z}_{2i}'\mathbf{z}_{2i}) = \Omega_{\mathbf{z}}$, and assumption 1.(b), plim $\widehat{\gamma}_1 = $ plim $\widehat{\gamma}_2 = \gamma$. Therefore, by Slutsky's theorem,

$$\begin{aligned}
\text{plim}\widehat{W} &= \left[\rho N_2\gamma'\Omega_{\mathbf{z}}\gamma_1+(N_1-\rho N_2)\gamma'\Omega_{\mathbf{z}}\gamma_1\right]^{-1}\left(\rho N_2\gamma'\Omega_{\mathbf{z}}\gamma_1\right) \\
&= \left[\rho N_2+(N_1-\rho N_2)\right]^{-1}\rho N_2\left(\gamma'\Omega_{\mathbf{z}}\gamma_1\right)^{-1}\left(\gamma'\Omega_{\mathbf{z}}\gamma_1\right) \\
&= \frac{\rho}{\alpha}.
\end{aligned}$$

Derivation of Remark 3

Use the first order Taylor expansion of $\widehat{W}$ at the point of $\left[E\left(\widehat{X}_{11}'\widehat{X}_{11}\right),E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)\right]$

$$\begin{aligned}
E\left(\widehat{W}\right) &= E\left[\left(\widehat{X}_{11}'\widehat{X}_{11}+\widehat{X}_{12}'\widehat{X}_{12}\right)^{-1}\left(\widehat{X}_{11}'\widehat{X}_{11}\right)\right] \\
&\equiv E\left[g\left(\widehat{X}_{11}'\widehat{X}_{11},\widehat{X}_{12}'\widehat{X}_{12}\right)\right] \\
&\approx E\left\{\left[E\left(\widehat{X}_{11}'\widehat{X}_{11}\right)+E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)\right]^{-1}E\left(\widehat{X}_{11}'\widehat{X}_{11}\right)+\frac{E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)\left[\widehat{X}_{11}'\widehat{X}_{11}-E\left(\widehat{X}_{11}'\widehat{X}_{11}\right)\right]}{E\left(\widehat{X}_{11}'\widehat{X}_{11}\right)+E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)}\right. \\
&\qquad\left.-\frac{E\left(\widehat{X}_{11}'\widehat{X}_{11}\right)\left[\widehat{X}_{12}'\widehat{X}_{12}-E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)\right]}{\left[E\left(\widehat{X}_{11}'\widehat{X}_{11}\right)+E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)\right]^{2}}\right\} \\
&= \left[E\left(\widehat{X}_{11}'\widehat{X}_{11}\right)+E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)\right]^{-1}E\left(\widehat{X}_{11}'\widehat{X}_{11}\right)
\end{aligned}$$

The numerator is equal to

$$\begin{aligned}
E\left(\widehat{X}_{11}'\widehat{X}_{11}\right) &= E\left(X_{11}'P_{\mathbf{Z}_{11}}X_{11}\right) \\
&= E\left[\left(\gamma'\mathbf{Z}_{11}'+V_{11}'\right)P_{\mathbf{Z}_{11}}\left(\mathbf{Z}_{11}\gamma+V_{11}\right)\right] \\
&= E\left(\gamma'\mathbf{Z}_{11}'\mathbf{Z}_{11}\gamma\right)+E\left(V_{11}'P_{\mathbf{Z}_{11}}V_{11}'\right) \\
&= \rho N_2\gamma'\Omega_z\gamma+K\cdot\sigma_v^2.
\end{aligned}$$

where $P_{\mathbf{Z}_{11}}\equiv\mathbf{Z}_{11}\left(\mathbf{Z}_{11}'\mathbf{Z}_{11}\right)^{-1}\mathbf{Z}_{11}'$ is the projection matrix. By assumption 1, $E(\mathbf{z}_{1i}'\mathbf{z}_{1i})=\Omega_{\mathbf{z}}$. Since $rank\left(P_{\mathbf{Z}_{11}}\right)=K$, w.p.a 1, we have $\frac{V_{11}'}{\sigma_v}P_{\mathbf{Z}_{11}}\frac{V_{11}}{\sigma_v}\sim\chi_K^2$, and $E\left(\chi_K^2\right)=K$ for the last equality.

Another component of the denominator is

$$\begin{aligned}
E\left(\widehat{X}_{12}'\widehat{X}_{12}\right) &= E\left\{\left[\mathbf{Z}_{12}\gamma+\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right]'\left[\mathbf{Z}_{12}\gamma+\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right]\right\} \\
&= E\left(\gamma'\mathbf{Z}_{12}'\mathbf{Z}_{12}\gamma\right)+E\left(\gamma'\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right)+E\left(V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\gamma\right) \\
&\quad+E\left(V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right) \\
&= \left(N_1-\rho N_2\right)\gamma'\Omega_{\mathbf{z}}\gamma+E\left(V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right)
\end{aligned}$$

The third equality follows because $\mathbf{Z}_{12}$ and $\mathbf{Z}_{22}$ are independent and by Assumption 5, $E\left(V_{22}|\mathbf{Z}_{22}\right)=0$.

Because $E\left(V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right)$ is a scalar,

$$\begin{aligned}
& E\left(V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right) \\
= & E\left[tr\left(V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right)\right] \\
= & E\left[tr\left(\mathbf{Z}_{22}'V_{22}V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\right)\right] \\
= & E\left\{tr\left[\mathbf{Z}_{22}'E\left(V_{22}V_{22}'|\mathbf{Z}_{22}\right)\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\right]\right\} \\
= & E\left\{tr\left[\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\right]\right\}\sigma_v^2 \\
= & tr\left\{E\left[\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\right]\right\}\sigma_v^2 \\
\approx & \frac{(N_1-\rho N_2)}{(1-\rho)N_2}\sigma_v^2
\end{aligned}$$

The third equality follows by the law of iterated expectations. The independence between $\mathbf{Z}_{12}$ and $\mathbf{Z}_{22}$ and Assumption 1.(a) and 5 enables us to pull $E\left(V_{22}V_{22}'|\mathbf{Z}_{22}\right)=E\left(V_{22}V_{22}'\right)=\sigma_v^2 I_n$ out of the expectation. The last equality follows because $\mathbf{Z}_{12}'\mathbf{Z}_{12}$ is independent of $\mathbf{Z}_{22}'\mathbf{Z}_{22}$ and $E\left(\mathbf{Z}_{12}'\mathbf{Z}_{12}\right)=(N_1-\rho N_2)\Omega_{\mathbf{z}}$, $E\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)=(1-\rho)N_2\Omega_{\mathbf{z}}$, and the first order Taylor expansion gives $tr\left\{E\left[\mathbf{Z}_{12}'\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\right]\right\}\approx\frac{(N_1-\rho N_2)}{(1-\rho)N_2}$.

Therefore,

$$E\left(\widehat{W}\right)\approx\frac{\rho N_2\gamma'\Omega_z\gamma+K\cdot\sigma_v^2}{N_1\gamma'\Omega_z\gamma+K\cdot\sigma_v^2+(N_1-\rho N_2)/(1-\rho)N_2\sigma_v^2}$$

Proof for proposition 4

$$\widehat{\beta}_{2SLS}^{(1)}-\beta=\left(\widehat{X}_{11}'\widehat{X}_{11}\right)^{-1}\left(\widehat{X}_{11}'\Xi_{11}\right)$$

We have derived the expectation of the denominator in remark 3. Similarly, the expectation of the numerator is

$$\begin{aligned}
E\left(\widehat{X}_{11}'\Xi_{11}\right) &= E\left(X_{11}'P_{\mathbf{Z}_{11}}E_{11}\right) \\
&= E\left[\left(\gamma'\mathbf{Z}_{11}'+V_{11}'\right)P_{\mathbf{Z}_{11}}\left(\theta V_{11}+R_{11}\right)\right] \\
&= \theta E\left(V_{11}'P_{\mathbf{Z}_{11}}V_{11}\right)+E\left(V_{11}'P_{\mathbf{Z}_{11}}R_{11}\right) \\
&= \theta\sigma_v^2\cdot E\left(\chi_1^2\right) \\
&= K\cdot\sigma_{\varepsilon v},
\end{aligned}$$

The second equality follows from assumption 4, that is $E\left(\gamma'\mathbf{Z}_{11}'P_{\mathbf{Z}_{11}}\Xi_{11}\right)=E\left(\gamma'\mathbf{Z}_{11}'\Xi_{11}\right)=$

0. Also, due to assumption 3.(c), write $\Xi_{11} = \theta V_{11} + R_{11}$, and R_{11} is independent of V_{11}. So $E\left(V_{11}' P_{\mathbf{Z}_{11}} R_{11}\right) = 0$ for the fourth equality to hold.

The first-order Taylor expansion of $g\left(X_{11}' P_{\mathbf{Z}_{11}} X_{11}, X_{11}' P_{\mathbf{Z}_{11}} \Xi_{11}\right)$ at point $\left[E\left(X_{11}' P_{\mathbf{Z}_{11}} X_{11}\right), E\left(X_{11}' P_{\mathbf{Z}_{11}} E_{11}\right)\right]$ is

$$
\begin{aligned}
E\left(\widehat{\beta}_{2SLS}^{(1)} - \beta\right) &\approx E\left(\widehat{X}_{11}' \widehat{X}_{11}\right)^{-1} E\left(\widehat{X}_{11}' \Xi_{11}\right) \\
&= \frac{K \cdot \sigma_{\varepsilon v}}{\rho N_2 \gamma_1' \Omega_z \gamma_1 + K \cdot \sigma_v^2}.
\end{aligned}
$$

Following Angrist and Kruger (1995),

$$
\begin{aligned}
E\left(\widehat{\beta}_{SS2SLS}^{(2)} - \beta\right) &= E\left[\left(\widehat{X}_{12}' \widehat{X}_{12}\right)^{-1} \left(\widehat{X}_{12}' Y_{12}\right)\right] - \beta \\
&= E\left\{\left[\widehat{X}_{12}' \widehat{X}_{12}\right]^{-1} \left[\widehat{X}_{12}' \left(X_{12}\beta + V_{12}\right)\right]'\right\} - \beta \\
&= E\left[\left(\widehat{X}_{12}' \widehat{X}_{12}\right)^{-1} \left(\widehat{X}_{12}' X_{12}\right)\right] \beta - \beta
\end{aligned}
$$

The second equality follows from Assumption 3 ($E(v_{2j}|\mathbf{z}_{2j}) = 0$) and the independence between V_{12} and $(X_{22}\,, \mathbf{Z}_{12})$,

$$
\begin{aligned}
& E\left[\left(\widehat{X}_{12}' \widehat{X}_{12}\right)^{-1} \left(\widehat{X}_{12}' \Xi_{12}\right)\right] \\
= \; & E\left\{\left\{\left[\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}' X_{22}\right]' \left[\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}' X_{22}\right]\right\}^{-1} X_{22}' \mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}' V_{12}\right\} \\
= \; & 0
\end{aligned}
$$

The following shows that $E(x_{1i}|\widehat{x}_{1i}) = \widehat{x}_{1i}\left\{E\left[\widehat{x}_{1i}'\widehat{x}_{1i}\right]^{-1} E\left[\widehat{x}_{1i}' x_{1i}\right]\right\}$ for $i = 1, ..., \rho N_2$:

$$
\begin{aligned}
E(x_{1i}|\widehat{x}_{1i}) &= E\left(z_{1i}\gamma + v_{1i}|z_{1i}\widehat{\gamma}_2\right) \\
&= E\left(z_{1i}\gamma|z_{1i}\widehat{\gamma}_2\right) \\
&= z_{1i}\widehat{\gamma}_2 \frac{E\left(\widehat{\gamma}_2' z_{1i}' z_{1i}\gamma\right)}{E\left(\widehat{\gamma}_2' z_{1i}' z_{1i}\widehat{\gamma}_2\right)} \\
&= z_{1i}\widehat{\gamma}_2 \frac{E\left(\widehat{\gamma}_2' z_{1i}' z_{1i}\gamma\right) + E\left(\widehat{\gamma}_2' z_{1i}' v_{1i}\right)}{E\left(\widehat{\gamma}_2' z_{1i}' z_{1i}\widehat{\gamma}_2\right)} \\
&= \widehat{x}_{1i} \frac{E\left(\widehat{x}_{1i}' x_{1i}\right)}{E\left(\widehat{x}_{1i}'\widehat{x}_{1i}\right)},
\end{aligned}
$$

Because $\widehat{\gamma}_2 = \left(z_{2j}' z_{2j}\right)^{-1}\left(z_{2j}' x_{2j}\right)$ and the fact that v_{1i} is independent of z_{1i}, z_{2j}, x_{2j}, the second equality follows from $E\left(v_{1i}|z_{1i}\widehat{\gamma}_2\right)=0$. The third equality is clear since $\widehat{\gamma}_2$ can be seen as a constant so that $z_{1i}\gamma$ is linear in $z_{1i}\widehat{\gamma}_2$.

By stacking the number of the observations, it is follows that $E(X_{12}|\widehat{X}_{12})$ is linear as well and that $E(X_{12}|\widehat{X}_{12})=\widehat{X}_{12}\left\{E\left[\widehat{X}_{12}'\widehat{X}_{12}\right]^{-1}E\left[\widehat{X}_{12}'X_{12}\right]\right\}$. Once more, by the law of iterated expectations,

$$E\left[\left(\widehat{X}_{12}'\widehat{X}_{12}\right)^{-1}\left(\widehat{X}_{12}'X_{12}\right)\right]=E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)^{-1}E\left(\widehat{X}_{12}'X_{12}\right)$$

The numerator simplifies to

$$\begin{aligned}
E\left(\widehat{X}_{12}'X_{12}\right) &= E\left\{\left[\mathbf{Z}_{12}\gamma+\mathbf{Z}_{12}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{22}'V_{22}\right]'\left(\mathbf{Z}_{12}\gamma+V_{12}\right)\right\} \\
&= E\left(\gamma'\mathbf{Z}_{12}'\mathbf{Z}_{12}\gamma\right)+E(V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'\mathbf{Z}_{12}\gamma)+E\left(\gamma'\mathbf{Z}_{12}'V_{12}\right) \\
&\quad +E\left(V_{22}'\mathbf{Z}_{22}\left(\mathbf{Z}_{22}'\mathbf{Z}_{22}\right)^{-1}\mathbf{Z}_{12}'V_{12}\right) \\
&= E\left(\gamma'\mathbf{Z}_{12}'\mathbf{Z}_{12}\gamma\right) \\
&= \left(N_1-\rho N_2\right)\gamma'\Omega_{\mathbf{z}}\gamma
\end{aligned}$$

The third equality follows from assumption 5 and V_{22} and V_{12} coming from independent samples. Also, recall $E(\mathbf{z}_{2i}'\mathbf{z}_{2i})=\Omega_{\mathbf{z}}$.

We already derived in remark 3

$$E\left(\widehat{X}_{12}'\widehat{X}_{12}\right)=\left(N_1-\rho N_2\right)\gamma'\Omega_{\mathbf{z}}\gamma+\left(\left(N_1-\rho N_2\right)/\left(1-\rho\right)N_2\right)\sigma_v^2,$$

so the approximate bias of SS2SLS follows as,

$$E\left(\widehat{\beta}_{SS2SLS}^{(2)}-\beta\right)=-\frac{\sigma_v^2\beta/\left(\left(1-\rho\right)N_2\right)}{\gamma'\Omega_{\mathbf{z}}\gamma+\sigma_v^2/\left(\left(1-\rho\right)N_2\right)}.$$

Proof for Proposition 5

$$\begin{aligned}
& E\left[\widehat{W}\left(\widehat{\beta}^{(1)}_{2SLS}-\beta\right)+\left(1-\widehat{W}\right)\left(\widehat{\beta}^{(2)}_{SS2SLS}-\beta\right)\right] \\
= \; & E\left[\widehat{W}\left(\widehat{\beta}^{(1)}_{2SLS}-\beta\right)\right]+E\left[\left(1-\widehat{W}\right)\widehat{\beta}^{(2)}_{SS2SLS}\right]-E\left(1-\widehat{W}\right)\beta \\
= \; & E\left(\frac{\widehat{X}'_{11}\Xi_{11}}{\widehat{X}'_{11}\widehat{X}_{11}+\widehat{X}'_{12}\widehat{X}_{12}}\right)+E\left(\frac{\widehat{X}'_{12}X_{12}}{\widehat{X}'_{11}\widehat{X}_{11}+\widehat{X}'_{12}\widehat{X}_{12}}\right)\beta-E\left(\frac{\widehat{X}'_{12}\widehat{X}_{12}}{\widehat{X}'_{11}\widehat{X}_{11}+\widehat{X}'_{12}\widehat{X}_{12}}\right)\beta \\
\approx \; & \frac{E\left(\widehat{X}'_{11}\Xi_{11}\right)+E\left(\widehat{X}'_{12}X_{12}\right)\beta-E\left(\widehat{X}'_{12}\widehat{X}_{12}\right)\beta}{E\left(\widehat{X}'_{11}\widehat{X}_{11}+\widehat{X}'_{12}\widehat{X}_{12}\right)} \\
= \; & \frac{K\cdot\sigma_{\varepsilon v}-\left(\left(N_1-\rho N_2\right)/\left(1-\rho\right)N_2\right)\sigma_v^2\beta}{N_1\gamma'\Omega_z\gamma+K\cdot\sigma_v^2+\left(\left(N_1-\rho N_2\right)/\left(1-\rho\right)N_2\right)\sigma_v^2}.
\end{aligned}$$

The third approximation uses a first-order Taylor expansion. All the moments in this approximation can be found in the proofs of proposition 1 and 2.

Proof for Proposition 6

The asymptotic variance of 2SLS estimator is

$$\sqrt{\rho N_2}\left(\widehat{\beta}^{(1)}_{2SLS}-\beta\right)\overset{a}{\sim}N\left[0,\sigma_v^2E\left(x^{*\prime}_{1i}x^*_{1i}\right)^{-1}\right]=N\left\{0,\sigma_v^2\left[\Omega_{x\mathbf{z}}\Omega_{\mathbf{z}}^{-1}\Omega_{x\mathbf{z}}\right]^{-1}\right\}$$

where $x^*_{1i}=\mathbf{z}_{1i}\gamma=\mathbf{z}_{1i}E\left(\mathbf{z}'_{1i}\mathbf{z}_{1i}\right)^{-1}E\left(\mathbf{z}'_{1i}x_{1i}\right)$ (Wooldridge 2010).

The asymptotic variance of the SS2SLS estimator is adapted from Inoue and Solon (2010) in this context as a special case of TS2SLS:

$$\sqrt{N_1-\rho N_2}\left(\widehat{\beta}^{(2)}_{SS2SLS}-\beta\right)\overset{a}{\sim}N\left\{0,\left[\Omega'_{x\mathbf{z}}\left[\left(\sigma_u^2+\frac{\alpha-\rho}{1-\rho}\beta'\sigma_v^2\beta\right)\Omega_{\mathbf{z}}\right]^{-1}\Omega_{x\mathbf{z}}\right]^{-1}\right\}.$$

By the asymptotic equivalence theorem,

$$\begin{aligned}
& \sqrt{N_1+(1-\rho)N_2}\left(\widehat{\beta}_O-\beta\right) \\
= \; & \sqrt{\frac{N_1+(1-\rho)N_2}{\rho N_2}}\sqrt{\rho N_2}\widehat{W}\left(\widehat{\beta}^{(1)}_{2SLS}-\beta\right)+\sqrt{\frac{N_1+(1-\rho)N_2}{\alpha-\rho N_2}}\sqrt{N_1-\rho N_2}\left(1-\widehat{W}\right)\left(\widehat{\beta}^{(2)}_{SS2SLS}-\beta\right) \\
& \overset{p}{\to}\frac{\sqrt{(1+\alpha-\rho)\rho}}{\alpha}\sqrt{\rho N_2}\left(\widehat{\beta}^{(1)}_{2SLS}-\beta\right)+\frac{\sqrt{(1+\alpha-\rho)(\alpha-\rho)}}{\alpha}\sqrt{N_1-\rho N_2}\left(\widehat{\beta}^{(2)}_{SS2SLS}-\beta\right) \\
& \overset{a}{\sim}N\left[0,\frac{(1+\alpha-\rho)\rho}{\alpha^2}\sigma_v^2\left[\Omega_{x\mathbf{z}}\Omega_{\mathbf{z}}^{-1}\Omega_{x\mathbf{z}}\right]^{-1}+\frac{(1+\alpha-\rho)(\alpha-\rho)}{\alpha^2}\left[\Omega'_{x\mathbf{z}}\left[\left(\sigma_u^2+\frac{\alpha-\rho}{1-\rho}\beta'\sigma_v^2\beta\right)\Omega_{\mathbf{z}}\right]^{-1}\Omega_{x\mathbf{z}}\right]^{-1}\right]
\end{aligned}$$