Some Finite-Sample Results on the Hausman Test

Jinyong Hahn* UCLA

Zhipeng Liao† UCLA

Nan Liu‡ Xiamen University

Shuyang Sheng§ UCLA

November 17, 2023

Abstract

This paper shows that the endogeneity test using the control function approach in linear instrumental variable models is a variant of the Hausman test. Moreover, we find that the test statistics used in these tests can be numerically ordered, indicating their relative power properties in finite samples.

JEL Classification: C14, C31, C32

Keywords: Control Function; Endogeneity; Hausman Test; Specification Test

1 Introduction

This paper investigates the endogeneity test for potentially endogenous regressors in linear instrumental variable (IV) models that utilizes the control function (CF) approach. Specifically, we focus on the following model:

$$y_2 = y_1^\top \beta + z_1^\top \gamma + \varepsilon = x^\top \theta + \varepsilon, \tag{1}$$

$$y_1 = z_1^\top \pi_1 + z_2^\top \pi_2 + v = z^\top \pi + v, \tag{2}$$

where $x \equiv (y_1^\top, z_1^\top)^\top$, y_1 represents potentially endogenous regressors, ε and v are unobserved error terms in the structural and reduced-form equations respectively, and $z \equiv (z_1^\top, z_2^\top)^\top$ represents exogenous variables which satisfy:

$$E[z\varepsilon] = 0 \qquad \text{and} \qquad E[zv^\top] = 0. \tag{3}$$

*Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: hahn@econ.ucla.edu.

†Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: zhipeng.liao@econ.ucla.edu.

‡Wang Yanan Institute for Studies in Economics and School of Economics, Xiamen University, Xiamen, Fujian, 36100. Email: nanliu@xmu.edu.cn.

§Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: ssheng@econ.ucla.edu

Since z_2 is excluded from the structure equation (1) and satisfies the orthogonality conditions in (3), it serves as instruments for y_1.

One widely used method for testing the endogeneity of y_1 is the Hausman test (Hausman, 1978). This test compares the ordinary least squares (OLS) estimator $\hat{\beta}_{ols}$ of β against the two-stage least squares (2SLS) estimator $\hat{\beta}_{2sls}$. The test rejects the null hypothesis that y_1 is exogenous if the difference between $\hat{\beta}_{ols}$ and $\hat{\beta}_{2sls}$ exceeds a certain threshold determined by the significance level of the test.

As an alternative to the Hausman test in the IV approach, we can also apply the CF approach to obtain a consistent estimator of β and test for the endogeneity of y_1. Specifically, we run an OLS regression of the following model

$$y_2 = x^\top \theta + v^\top \rho + u, \tag{4}$$

where we replace v by the estimated residual $\hat{v}$ in the OLS regression of the reduced-form equation (2), and obtain estimators $\hat{\theta}_{cf}$ and $\hat{\rho}_{cf}$. Then we test the endogeneity of y_1 using the Wald test for the null hypothesis $H_0 : \rho = 0$.

This paper makes several contributions to the existing literature. First, we demonstrate that $\hat{\rho}_{cf}$ is a linear transformation of $\hat{\beta}_{ols} - \hat{\beta}_{2sls}$, thereby elucidating the connection between the Wald test for $H_0 : \rho = 0$ in the CF approach and the Hausman test in the IV approach. Second, we show that the Wald test differs from the Hausman test primarily in how the asymptotic variances of $\hat{\beta}_{2sls}$ and $\hat{\beta}_{ols}$ are estimated, thus representing a variant of the Hausman test. Third, our analysis reveals that the Wald test statistic is numerically larger than the Hausman test statistics, which rely on $\hat{\beta}_{ols}$ or $\hat{\beta}_{2sls}$ to obtain estimators of the asymptotic variances of $\hat{\beta}_{ols}$ and $\hat{\beta}_{2sls}$. Since the Wald test employs the same critical value as the Hausman test, it exhibits larger statistical power in finite samples. These findings are established without imposing restrictive assumptions on either the null or alternative hypotheses.

The remainder of the paper is structured as follows. Section 2 introduces the test statistics employed in both the Hausman test and the Wald test. Section 3 establishes a numerical order among the test statistics introduced in Section 2 and discusses its implications for relative power properties of the tests in finite samples. Section 4 concludes the paper. The proofs are presented in the Appendix.

Notation. We use $a \equiv b$ to indicate that a is defined as b. For any real vector a, d_a denotes the dimension of a. For any positive integer k, I_k denotes the $k \times k$ identity matrix. For any $k_1 \times k_2$ matrix A, $A^\top$ denotes the transpose of A, $P_A \equiv A(A^\top A)^{-1}A^\top$ and $M_A \equiv I_{k_1} - A(A^\top A)^{-1}A^\top$ as long as $A^\top A$ is non-singular. For any square matrix A, $A > 0$ means A is positive definite.

2 Testing for Endogeneity

We begin by formulating the Hausman test for assessing the endogeneity of y_1 in the model specified by (1) and (2). Suppose we have a random sample $\{y_{1,i}, y_{2,i}, z_i\}_{i=1}^n$, where $z_i \equiv (z_{1,i}^\top, z_{2,i}^\top)^\top$ for $i = 1, \ldots, n$. Recall that $x_i \equiv (y_{1,i}^\top, z_{1,i}^\top)^\top$, $i = 1, \ldots, n$, denotes the regressors in (1). Let $X \equiv (x_1, \ldots, x_n)^\top$, $Y_1 \equiv (y_{1,1}, \ldots, y_{1,n})^\top$, $Y_2 \equiv (y_{2,1}, \ldots, y_{2,n})^\top$, and $Z \equiv (z_1, \ldots, z_n)^\top$. The OLS and 2SLS estimators of θ in (1) are given by

$$\hat{\theta}_{ols} \equiv (X^\top X)^{-1} X^\top Y_2 \qquad \text{and} \qquad \hat{\theta}_{2sls} \equiv (X^\top P_Z X)^{-1} X^\top P_Z Y_2$$

respectively. Let $\hat{\beta}_{ols}$ and $\hat{\beta}_{2sls}$ denote the leading $d_{y_1} \times 1$ subvectors of $\hat{\theta}_{ols}$ and $\hat{\theta}_{2sls}$ respectively. The Hausman test statistic can be characterized by

$$t_{H,n}(\hat{\sigma}_1^2, \hat{\sigma}_2^2) \equiv (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top (\hat{\sigma}_1^2 (\hat{Y}_1^\top M_{Z_1} \hat{Y}_1)^{-1} - \hat{\sigma}_2^2 (Y_1^\top M_{Z_1} Y_1)^{-1})^{-1} (\hat{\beta}_{ols} - \hat{\beta}_{2sls}), \tag{5}$$

where $\hat{Y}_1 \equiv P_Z Y_1$, $\hat{\sigma}_1^2 (\hat{Y}_1^\top M_{Z_1} \hat{Y}_1)^{-1}$ and $\hat{\sigma}_2^2 (Y_1^\top M_{Z_1} Y_1)^{-1}$ are estimators of the asymptotic variances of $\hat{\beta}_{ols}$ and $\hat{\beta}_{2sls}$ respectively, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are (possibly different) estimates of the variance σ_ε^2 of the error term ε in the structural equation (1).

In practice, σ_ε^2 can be estimated by the sample variance of the fitted residual in the OLS estimation:

$$\hat{\sigma}_{ols}^2 \equiv n^{-1} (Y_2 - X\hat{\theta}_{ols})^\top (Y_2 - X\hat{\theta}_{ols}),$$

or through its counterpart in the 2SLS estimation:

$$\hat{\sigma}_{2sls}^2 \equiv n^{-1} (Y_2 - X\hat{\theta}_{2sls})^\top (Y_2 - X\hat{\theta}_{2sls}),$$

resulting in three popular versions of the Hausman test with test statistics $t_{H_1} \equiv t_{H,n}(\hat{\sigma}_{ols}^2, \hat{\sigma}_{ols}^2)$, $t_{H_2} \equiv t_{H,n}(\hat{\sigma}_{2sls}^2, \hat{\sigma}_{2sls}^2)$, and $t_{H_3} \equiv t_{H,n}(\hat{\sigma}_{2sls}^2, \hat{\sigma}_{ols}^2)$ respectively (Wooldridge, 2010, Section 6.3.1; Baum, Schaffer, and Stillman, 2003).

Alternatively, we can apply the CF approach and obtain the estimators of θ and ρ in (4) as

$$(\hat{\theta}_{cf}^\top, \hat{\rho}_{cf}^\top)^\top \equiv ((X, \hat{V})^\top (X, \hat{V}))^{-1} (X, \hat{V})^\top Y_2, \tag{6}$$

where $\hat{V} \equiv M_Z Y_1$, and then test the endogeneity of y_1 using the Wald test with the test statistic

$$t_{CF} \equiv \hat{\rho}_{cf}^\top \left(\widehat{Asv}(\hat{\rho}_{cf}) \right)^{-1} \hat{\rho}_{cf}, \tag{7}$$

where $\widehat{Asv}(\hat{\rho}_{cf})$ denotes an estimator of the asymptotic variance of $\hat{\rho}_{cf}$. Applying the partitioned regression formula (which is also known as the Frisch-Waugh-Lovell Theorem, see, e.g., Davidson and

MacKinnon (1993, Section 1.4)) to (6) yields

$$\hat{\rho}_{cf} = (\hat{V}^\top M_X \hat{V})^{-1}(\hat{V}^\top M_X Y_2) = \rho + (\hat{V}^\top M_X \hat{V})^{-1}[\hat{V}^\top M_X U - \hat{V}^\top M_X (\hat{V} - V)\rho], \tag{8}$$

where $U \equiv (u_1, \ldots, u_n)^\top$, $V \equiv (v_1, \ldots, v_n)^\top$, and the second equality follows by (4). Although the regressor v is estimated and hence the estimation error $\hat{V} - V$ should be taken into account when calculating $\widehat{Asv}(\hat{\rho}_{cf})$, the expansion in (8) shows that this estimation error can be ignored under the null that $\rho = 0$. Therefore, in the rest of the paper we use the estimator

$$\widehat{Asv}(\hat{\rho}_{cf}) = \hat{\sigma}_u^2 (\hat{V}^\top M_X \hat{V})^{-1}, \tag{9}$$

where $\hat{\sigma}_u^2$ is the sample variance of the fitted residual in the OLS regression of (4)

$$\hat{\sigma}_u^2 \equiv n^{-1}(Y_2 - X\hat{\theta}_{cf} - \hat{V}\hat{\rho}_{cf})^\top (Y_2 - X\hat{\theta}_{cf} - \hat{V}\hat{\rho}_{cf}). \tag{10}$$

As we shall see in the next section, choosing this estimator makes the Wald statistic t_{CF} comparable to $t_{H,n}(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$.

Throughout this paper, we assume that $X^\top X$, $X^\top P_Z X$, and $(X, \hat{V})^\top (X, \hat{V})$ are non-singular so that the estimators $\hat{\theta}_{ols}$, $\hat{\theta}_{2sls}$, $\hat{\theta}_{cf}$ and $\hat{\rho}_{cf}$ are well defined. Under these primitive conditions, we have $Y_1^\top M_{Z_1} Y_1 > 0$, $\hat{Y}_1^\top M_{Z_1} \hat{Y}_1 > 0$ and

$$Y_1^\top M_{Z_1} Y_1 - \hat{Y}_1^\top M_{Z_1} \hat{Y}_1 = Y_1^\top M_Z Y_1 > 0,$$

which further implies that

$$(\hat{Y}_1^\top M_{Z_1} \hat{Y}_1)^{-1} - (Y_1^\top M_{Z_1} Y_1)^{-1} > 0. \tag{11}$$

In view of (11) and the definitions of t_{H_j} $(j = 1, 2, 3)$ and t_{CF}, we also assume that $\hat{\sigma}_{ols}^2$, $\hat{\sigma}_{2sls}^2$ and $\hat{\sigma}_u^2$ are strictly positive so that these test statistics are well defined.

Under the null that y_1 is exogenous, along with other regularity conditions, one can establish that the asymptotic distributions of t_{H_j} $(j = 1, 2, 3)$ and t_{CF} are Chi-square with k_1 degrees of freedom (denoted as $\chi^2(k_1)$). Consequently, the Hausman tests and the Wald test reject the null hypothesis if the corresponding test statistic exceeds the $1 - \alpha$ quantile of $\chi^2(k_1)$, where α denotes the significance level. As we shall see in the next section, the test statistics t_{H_j} $(j = 1, 2, 3)$ and t_{CF} can be numerically ordered, indicating their relative rejection properties in finite samples.

3 Main Results

Our first objective is to establish a numerical relationship between the OLS and 2SLS estimators and the estimators in the CF approach. This result serves as a foundation for further investigating the numerical order among t_{H_j} $(j = 1, 2, 3)$ and t_{CF} in finite samples.

Lemma 1 *The estimators in the CF approach satisfy*

$$\begin{pmatrix} \hat{\theta}_{cf} \\ \hat{\rho}_{cf} \end{pmatrix} = \begin{pmatrix} \hat{\theta}_{2sls} \\ (Y_1^\top M_Z Y_1)^{-1}(Y_1^\top M_{Z_1} Y_1)(\hat{\beta}_{ols} - \hat{\beta}_{2sls}) \end{pmatrix}. \tag{12}$$

Moreover, the Wald test statistic t_{CF} satisfies

$$t_{CF} = (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top \left(\hat{\sigma}_u^2(\hat{Y}_1^\top M_{Z_1}\hat{Y}_1)^{-1} - \hat{\sigma}_u^2(Y_1^\top M_Z Y_1)^{-1}\right)^{-1}(\hat{\beta}_{ols} - \hat{\beta}_{2sls}), \tag{13}$$

where $\hat{\sigma}_u^2$ is defined in (10).

The lemma above carries two important implications. First, $\hat{\theta}_{cf}$ is numerically equivalent to $\hat{\theta}_{2sls}$, implying that $\hat{\theta}_{cf}$ shares the same standard error as $\hat{\theta}_{2sls}$. This finding has been well recognized in the literature since at least Hausman (1978) (see also Davidson and MacKinnon (1993, Section 7.9) and Wooldridge (2010, Problem 5.1)). Second, $\hat{\rho}_{cf}$ is a linear transformation of $\hat{\beta}_{ols} - \hat{\beta}_{2sls}$, establishing a connection between the Wald test based on t_{CF} and the Hausman tests based on $t_{H,n}(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$. To the best of our knowledge, this numerical relationship is a novel contribution to the literature. The expression of t_{CF} in (13) further suggests that t_{CF} is a special case of $t_{H,n}(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$ with $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \hat{\sigma}_u^2$.

Since t_{H_j} $(j = 1, 2, 3)$ and t_{CF} differ only in how the variance estimators $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ in (5) are calculated, their relative performances are determined by the differences of these variance estimators. Intuitively, $\hat{\sigma}_u^2$ should not be larger than $\hat{\sigma}_{ols}^2$ since, compared with the OLS estimation of (1), the CF approach includes an extra regressor v in (4), and the resulting R^2 should not be smaller. Moreover, we have $\hat{\sigma}_{ols}^2 \leq \hat{\sigma}_{2sls}^2$ by the definitions of OLS and 2SLS estimation. Therefore, we expect a weak order among these variance estimators: $\hat{\sigma}_u^2 \leq \hat{\sigma}_{ols}^2 \leq \hat{\sigma}_{2sls}^2$, which together with the definitions of t_{H_j} $(j = 1, 2, 3)$ and t_{CF} implies a weak numerical order among the test statistics: $t_{CF} \geq t_{H_1} \geq t_{H_2} \geq t_{H_3}$. Our next lemma establishes the exact relationships among the variance estimators $\hat{\sigma}_u^2$, $\hat{\sigma}_{ols}^2$ and $\hat{\sigma}_{2sls}^2$, enabling us to obtain a strong numerical order among them as well as among the test statistics t_{H_j} $(j = 1, 2, 3)$ and t_{CF}.

Lemma 2 *Let $H_n \equiv (n\hat{\sigma}_{2sls}^2)^{-1}(\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top(Y_1^\top M_{Z_1} Y_1)(\hat{\beta}_{ols} - \hat{\beta}_{2sls})$. Then*

$$\hat{\sigma}_u^2 = \hat{\sigma}_{ols}^2\left(1 - \frac{t_{H_1}}{n}\right) = \hat{\sigma}_{2sls}^2\left(1 - \frac{t_{H_2}}{n} - H_n\right). \tag{14}$$

Since t_{H_j} $(j = 1, 2, 3)$ and H_n are non-negative, from (14) we immediately obtain $\hat{\sigma}^2_{2sls} \geq \hat{\sigma}^2_{ols} \geq \hat{\sigma}^2_u$, which implies that $t_{CF} \geq t_{H_1} \geq t_{H_2} \geq t_{H_3}$. Moreover, when $\hat{\beta}_{ols} - \hat{\beta}_{2sls} \neq 0$, t_{H_j} and H_n are strictly positive. In this case, we can deduce from (14) that $\hat{\sigma}^2_{2sls} > \hat{\sigma}^2_{ols} > \hat{\sigma}^2_u$, and a strong numerical order among t_{H_j} $(j = 1, 2, 3)$ and t_{CF}, which is summarized in the lemma below.

Lemma 3 *Suppose that $\hat{\beta}_{ols} - \hat{\beta}_{2sls} \neq 0$. Then we have $t_{CF} > t_{H_1} > t_{H_2} > t_{H_3}$.*

Lemma 3 demonstrates that in finite samples, the endogeneity test based on t_{CF} has the largest rejection probability compared with t_{H_j} $(j = 1, 2, 3)$, although these four tests are asymptotically equivalent under the null hypothesis and local alternatives where $\hat{\beta}_{ols} - \hat{\beta}_{2sls} = o_p(1)$.

4 Conclusion

This paper explores the endogeneity test using the CF approach in linear IV models. We find that the OLS estimator of the coefficients of the generated CF is a linear transformation of the difference between the OLS and 2SLS estimators of the coefficients of endogenous regressors. This finding allows us to demonstrate that the commonly used endogeneity test using the CF approach is a variant of the Hausman test. In addition, we establish a numerical order among the test statistics employed in the Hausman tests and the endogeneity test using the CF approach. This numerical order highlights that the endogeneity test using the CF approach exhibits the highest rejection probability in finite samples.

References

Baum, C. F., M. E. Schaffer, and S. Stillman (2003): "Instrumental Variables and GMM: Estimation and Testing," *The Stata Journal*, 3(1), 1–31.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and inference in econometrics*, vol. 63. Oxford New York.

Hausman, J. A. (1978): "Specification tests in econometrics," *Econometrica: Journal of the econometric society*, pp. 1251–1271.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*. MIT press.

Appendix

A Proofs

Proof of Lemma 1. By partitioned regression/partialling out formula,

$$\begin{pmatrix} \hat{\theta}_{cf} \\ \hat{\rho}_{cf} \end{pmatrix} = \begin{pmatrix} \left(X^\top M_{\hat{V}} X\right)^{-1} X^\top M_{\hat{V}} Y_2 \\ (\hat{V}^\top M_X \hat{V})^{-1} \hat{V}^\top M_X Y_2 \end{pmatrix}. \tag{15}$$

Since $X^\top \hat{V} = X^\top M_Z Y_1 = (Y_1^\top M_Z Y_1, 0_{d_{y_1} \times d_{z_1}})^\top$ and $\hat{V}^\top \hat{V} = Y_1^\top M_Z Y_1$, we have

$$X^\top M_{\hat{V}} X = X^\top X - \begin{pmatrix} Y_1^\top M_Z Y_1 & 0_{d_{y_1} \times d_{z_1}} \\ 0_{d_{z_1} \times d_{y_1}} & 0_{d_{z_1} \times d_{z_1}} \end{pmatrix} = \begin{pmatrix} Y_1^\top P_Z Y_1 & Y_1^\top Z_1 \\ Z_1^\top Y_1 & Z_1^\top Z_1 \end{pmatrix} = X^\top P_Z X$$

and

$$X^\top M_{\hat{V}} Y_2 = X^\top Y_2 - \begin{pmatrix} Y_1^\top M_Z Y_2 \\ 0_{d_{z_1} \times 1} \end{pmatrix} = \begin{pmatrix} Y_1^\top P_Z Y_2 \\ Z_1^\top Y_2 \end{pmatrix} = X^\top P_Z Y_2,$$

which together with (15) implies that

$$\hat{\theta}_{cf} = (X^\top P_Z X)^{-1} X^\top P_Z Y_2 = \hat{\theta}_{2sls}. \tag{16}$$

By the definitions of $\hat{\theta}_{ols}$ and $\hat{\theta}_{2sls}$, it is easy to show that

$$\hat{\beta}_{ols} = (Y_1^\top M_{Z_1} Y_1)^{-1} Y_1^\top M_{Z_1} Y_2 \quad \text{and} \quad \hat{\beta}_{2sls} = (Y_1^\top (P_Z - P_{Z_1}) Y_1)^{-1} Y_1^\top (P_Z - P_{Z_1}) Y_2. \tag{17}$$

Moreover,

$$\begin{aligned} \hat{V}^\top M_X \hat{V} &= Y_1^\top M_Z Y_1 - (Y_1^\top M_Z Y_1, 0_{d_{y_1} \times d_{z_1}})(X^\top X)^{-1} (Y_1^\top M_Z Y_1, 0_{d_{y_1} \times d_{z_1}})' \\ &= Y_1^\top M_Z Y_1 - Y_1^\top M_Z Y_1 (Y_1^\top M_{Z_1} Y_1)^{-1} Y_1^\top M_Z Y_1 \\ &= Y_1^\top (P_Z - P_{Z_1}) Y_1 (Y_1^\top M_{Z_1} Y_1)^{-1} Y_1^\top M_Z Y_1, \end{aligned} \tag{18}$$

and

$$\begin{aligned} \hat{V}^\top M_X Y_2 &= Y_1^\top M_Z Y_2 - (Y_1^\top M_Z Y_1, 0_{d_{y_1} \times d_{z_1}})(X^\top X)^{-1} X^\top Y_2 \\ &= Y_1^\top M_Z Y_2 - Y_1^\top M_Z Y_1 (Y_1^\top M_{Z_1} Y_1)^{-1} Y_1^\top M_{Z_1} Y_2 \\ &= Y_1^\top (P_Z - P_{Z_1}) Y_1 (Y_1^\top M_{Z_1} Y_1)^{-1} Y_1^\top M_{Z_1} Y_2 - Y_1^\top (P_Z - P_{Z_1}) Y_2, \\ &= Y_1^\top (P_Z - P_{Z_1}) Y_1 (\hat{\beta}_{ols} - \hat{\beta}_{2sls}) \end{aligned} \tag{19}$$

where the second equality is by the partitioned regression formula on $(X^\top X)^{-1}X^\top Y_2$, and the last equality is by (17). Combining the expressions in (18) and (19) yields

$$\hat{\rho}_{cf} = (\hat{V}^\top M_X \hat{V})^{-1}\hat{V}^\top M_X Y_2 = (Y_1^\top M_Z Y_1)^{-1} Y_1^\top M_{Z_1} Y_1 (\hat{\beta}_{ols} - \hat{\beta}_{2sls}). \tag{20}$$

The claim in (12) follows from (16) and (20).

By the definition of $\widehat{Asv}(\hat{\rho}_{cf})$ in (9) and the expression in (18),

$$\widehat{Asv}(\hat{\rho}_{cf}) = \hat{\sigma}_u^2 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top M_{Z_1} Y_1 (Y_1^\top (P_Z - P_{Z_1}) Y_1)^{-1}. \tag{21}$$

Using (12) and (21), we have

$$\begin{aligned}
\hat{\rho}_{cf}^\top (\widehat{Asv}(\hat{\rho}_{cf}))^{-1}\hat{\rho}_{cf} &= \hat{\rho}_{cf}^\top \frac{Y_1^\top (P_Z - P_{Z_1}) Y_1 (Y_1^\top M_{Z_1} Y_1)^{-1} Y_1^\top M_Z Y_1}{\hat{\sigma}_u^2} \hat{\rho}_{cf} \\
&= (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top \frac{Y_1^\top M_{Z_1} Y_1 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top (P_Z - P_{Z_1}) Y_1}{\hat{\sigma}_u^2} (\hat{\beta}_{ols} - \hat{\beta}_{2sls}).
\end{aligned} \tag{22}$$

Since $M_Z = M_{Z_1} - P_Z + P_{Z_1}$ and $\hat{Y}_1 = P_Z Y_1$, we derive

$$\begin{aligned}
&(Y_1^\top (P_Z - P_{Z_1}) Y_1)^{-1} Y_1^\top M_Z Y_1 (Y_1^\top M_{Z_1} Y_1)^{-1} \\
&= (Y_1^\top (P_Z - P_{Z_1}) Y_1)^{-1} Y_1^\top (M_{Z_1} - P_Z + P_{Z_1}) Y_1 (Y_1^\top M_{Z_1} Y_1)^{-1} \\
&= (Y_1^\top (P_Z - P_{Z_1}) Y_1)^{-1} - (Y_1^\top M_{Z_1} Y_1)^{-1} \\
&= (\hat{Y}_1^\top M_{Z_1} \hat{Y}_1)^{-1} - (Y_1^\top M_{Z_1} Y_1)^{-1},
\end{aligned}$$

which together with (22) proves the second claim of the lemma. ■

Proof of Lemma 2. By the definition of $\hat{\theta}_{ols}$ and $\hat{\theta}_{2sls}$, we can write

$$\begin{aligned}
\hat{\theta}_{ols} - \hat{\theta}_{2sls} =& (X^\top X)^{-1} X^\top M_Z Y_2 + (X^\top X)^{-1} X^\top P_Z Y_2 - \hat{\theta}_{2sls} \\
=& (X^\top X)^{-1} X^\top M_Z Y_2 + (X^\top X)^{-1} (X^\top P_Z X - X^\top X)\hat{\theta}_{2sls} \\
=& (X^\top X)^{-1} (X^\top M_{Z_1} Y_2 - X^\top M_{Z_1} X \hat{\theta}_{2sls}) \\
& - (X^\top X)^{-1} (X^\top (P_Z - P_{Z_1}) Y_2 - X^\top (P_Z - P_{Z_1}) X \hat{\theta}_{2sls}) \\
=& (X^\top X)^{-1} (X^\top M_{Z_1} Y_2 - X^\top M_{Z_1} X \hat{\theta}_{2sls}).
\end{aligned} \tag{23}$$

Applying the formula for the inverse of block matrix to $X^\top X = (Y_1, Z_1)^\top (Y_1, Z_1)$ and elementary matrix operations yields

$$\begin{aligned}
&(X^\top X)^{-1} (X^\top M_{Z_1} Y_2 - X^\top M_{Z_1} X \hat{\theta}_{2sls}) \\
&= \begin{pmatrix} (Y_1^T M_{Z_1} Y_1)^{-1} \\ -(Z_1^\top Z_1)^{-1} Z_1^\top Y_1 (Y_1^\top M_{Z_1} Y_1)^{-1} \end{pmatrix} (Y_1^\top M_{Z_1} Y_1 - Y_1^\top M_{Z_1} Y_1 \hat{\beta}_{2sls}).
\end{aligned}$$

Note that

$$Y_1^\top M_{Z_1} Y_1 - Y_1^\top M_{Z_1} Y_1 \hat{\beta}_{2sls} = Y_1^\top M_{Z_1} Y_1 (\hat{\beta}_{ols} - \hat{\beta}_{2sls}).$$

Therefore, we can further derive

$$\hat{\theta}_{ols} - \hat{\theta}_{2sls} = \begin{pmatrix} I_{d_{y_1}} \\ -(Z_1^\top Z_1)^{-1} Z_1^\top Y_1 \end{pmatrix} (\hat{\beta}_{ols} - \hat{\beta}_{2sls}). \tag{24}$$

To show the first equality in (14), we write

$$\hat{u} = Y_2 - X\hat{\theta}_{cf} - \hat{V}\hat{\rho}_{cf} = Y_2 - X\hat{\theta}_{ols} + X(\hat{\theta}_{ols} - \hat{\theta}_{2sls}) - \hat{V}\hat{\rho}_{cf}. \tag{25}$$

Note that $X^T(Y_2 - X\hat{\theta}_{ols}) = 0$ by the definition of $\hat{\theta}_{ols}$. Some elementary matrix operations lead to

$$\begin{aligned} \hat{V}^T(Y_2 - X\hat{\theta}_{ols}) &= Y_1^\top M_Z Y_2 - Y_1^\top M_Z X \hat{\theta}_{ols} \\ &= Y_1^\top M_Z Y_2 - Y_1^\top M_Z Y_1 \hat{\beta}_{ols} \\ &= Y_1^\top M_{Z_1} Y_2 - Y_1^\top M_{Z_1} Y_1 \hat{\beta}_{ols} + Y_1^\top (P_Z - P_{Z_1}) Y_1 \hat{\beta}_{ols} - Y_1^\top (P_Z - P_{Z_1}) Y_2 \\ &= Y_1^\top (P_Z - P_{Z_1}) Y_1 (\hat{\beta}_{ols} - \hat{\beta}_{2sls}). \end{aligned}$$

Hence, by Lemma 1

$$\hat{\rho}_{cf}^T \hat{V}^T (Y_2 - X\hat{\theta}_{ols}) = (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top Y_1^\top M_{Z_1} Y_1 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top (P_Z - P_{Z_1}) Y_1 (\hat{\beta}_{ols} - \hat{\beta}_{2sls}). \tag{26}$$

Moreover, observe that by (24) and Lemma 1

$$\begin{aligned} X(\hat{\theta}_{ols} - \hat{\theta}_{2sls}) - \hat{V}\hat{\rho}_{cf} &= M_{Z_1} Y_1 (\hat{\beta}_{ols} - \hat{\beta}_{2sls}) - M_Z Y_1 (Y_1^\top M_Z Y_1)^{-1} (Y_1^\top M_{Z_1} Y_1)(\hat{\beta}_{ols} - \hat{\beta}_{2sls}) \\ &= (M_{Z_1} Y_1 - M_Z Y_1 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top M_{Z_1} Y_1)(\hat{\beta}_{ols} - \hat{\beta}_{2sls}) \end{aligned}$$

and by elementary matrix operations

$$\begin{aligned} &(M_{Z_1} Y_1 - M_Z Y_1 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top M_{Z_1} Y_1)^\top (M_{Z_1} Y_1 - M_Z Y_1 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top M_{Z_1} Y_1) \\ &= Y_1^\top M_{Z_1} Y_1 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top M_{Z_1} Y_1 - Y_1^\top M_{Z_1} Y_1 \\ &= Y_1^\top M_{Z_1} Y_1 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top (P_Z - P_{Z_1}) Y_1. \end{aligned}$$

Therefore, we have

$$\begin{aligned} &(X(\hat{\theta}_{ols} - \hat{\theta}_{2sls}) - \hat{V}\hat{\rho}_{cf})^T (X(\hat{\theta}_{ols} - \hat{\theta}_{2sls}) - \hat{V}\hat{\rho}_{cf}) \\ &= (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top Y_1^\top M_{Z_1} Y_1 (Y_1^\top M_Z Y_1)^{-1} Y_1^\top (P_Z - P_{Z_1}) Y_1 (\hat{\beta}_{ols} - \hat{\beta}_{2sls}). \end{aligned} \tag{27}$$

Collecting the results in (25), (26), and (27), we obtain

$$\begin{aligned}n^{-1}\hat{u}^\top\hat{u} =& n^{-1}(Y_2 - X\hat{\theta}_{ols})^T(Y_2 - X\hat{\theta}_{ols}) - 2n^{-1}\hat{\rho}_{cf}^T\hat{V}^T(Y_2 - X\hat{\theta}_{ols})\\ &+ n^{-1}(X(\hat{\theta}_{ols} - \hat{\theta}_{2sls}) - \hat{V}\hat{\rho}_{cf})^T(X(\hat{\theta}_{ols} - \hat{\theta}_{2sls}) - \hat{V}\hat{\rho}_{cf})\\ =&\hat{\sigma}^2_{ols} - n^{-1}(\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top Y_1^\top M_{Z_1}Y_1(Y_1^\top M_Z Y_1)^{-1}Y_1^\top(P_Z - P_{Z_1})Y_1(\hat{\beta}_{ols} - \hat{\beta}_{2sls})\\ =&\hat{\sigma}^2_{ols} - n^{-1}(\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top((Y_1^\top(P_Z - P_{Z_1})Y_1)^{-1} - (Y_1^\top M_{Z_1}Y_1)^{-1})^{-1}(\hat{\beta}_{ols} - \hat{\beta}_{2sls})\\ =&\hat{\sigma}^2_{ols}\left(1 - \frac{t_{H,n}\left(\hat{\sigma}^2_{ols}, \hat{\sigma}^2_{ols}\right)}{n}\right),\end{aligned}$$

which together with the definition of t_{H_1} proves the first equality in (14).

To show the second equality in (14), note that by Lemma 1

$$\hat{u} = Y_2 - X\hat{\theta}_{cf} - \hat{V}\hat{\rho}_{cf} = Y_2 - X\hat{\theta}_{2sls} - \hat{V}\hat{\rho}_{cf}. \tag{28}$$

By the definitions of $\hat{\beta}_{ols}$ and $\hat{\beta}_{2sls}$, we have

$$\begin{aligned}\hat{V}^T(Y_2 - X\hat{\theta}_{2sls}) =& Y_1^\top M_Z Y_2 - Y_1^\top M_Z X\hat{\theta}_{2sls}\\ =&Y_1^\top M_Z Y_2 - Y_1^\top M_Z Y_1\hat{\beta}_{2sls}\\ =&Y_1^\top M_{Z_1}Y_2 - Y_1^\top M_{Z_1}Y_1\hat{\beta}_{2sls} - Y_1^\top(P_Z - P_{Z_1})Y_2 + Y_1^\top(P_Z - P_{Z_1})Y_1\hat{\beta}_{2sls}\\ =&Y_1^\top M_{Z_1}Y_1(\hat{\beta}_{ols} - \hat{\beta}_{2sls}).\end{aligned} \tag{29}$$

Therefore,

$$\hat{\rho}_{cf}^T\hat{V}^T(Y_2 - X\hat{\theta}_{2sls}) = (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top Y_1^\top M_{Z_1}Y_1(Y_1^\top M_Z Y_1)^{-1}Y_1^\top M_{Z_1}Y_1(\hat{\beta}_{ols} - \hat{\beta}_{2sls}). \tag{30}$$

Moreover,

$$\hat{\rho}_{cf}^\top\hat{V}^T\hat{V}\hat{\rho}_{cf} = (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top Y_1^\top M_{Z_1}Y_1(Y_1^\top M_Z Y_1)^{-1}Y_1^\top M_{Z_1}Y_1(\hat{\beta}_{ols} - \hat{\beta}_{2sls}),$$

which combined with (28), (29) and (30) implies that

$$\begin{aligned}n^{-1}\hat{u}^\top\hat{u} =& n^{-1}(Y_2 - X\hat{\theta}_{2sls})^\top(Y_2 - X\hat{\theta}_{2sls}) + n^{-1}\hat{\rho}_{cf}^\top\hat{V}^T\hat{V}\hat{\rho}_{cf} - 2n^{-1}\hat{\rho}_{cf}^T\hat{V}^T(Y_2 - X\hat{\theta}_{2sls})\\ =&\hat{\sigma}^2_{2sls} - (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top Y_1^\top M_{Z_1}Y_1(Y_1^\top M_Z Y_1)^{-1}Y_1^\top M_{Z_1}Y_1(\hat{\beta}_{ols} - \hat{\beta}_{2sls})\\ =&\hat{\sigma}^2_{2sls} - n^{-1}(\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top Y_1^\top M_{Z_1}Y_1(Y_1^\top M_Z Y_1)^{-1}Y_1^\top(P_Z - P_{Z_1})Y_1(\hat{\beta}_{ols} - \hat{\beta}_{2sls})\\ &- n^{-1}(\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top Y_1^\top M_{Z_1}Y_1(Y_1^\top M_Z Y_1)^{-1}Y_1^\top M_Z Y_1(\hat{\beta}_{ols} - \hat{\beta}_{2sls})\\ =&\hat{\sigma}^2_{2sls} - \hat{\sigma}^2_{2sls}\frac{t_{H,n}\left(\hat{\sigma}^2_{2sls}, \hat{\sigma}^2_{2sls}\right)}{n} - (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top\frac{Y_1^\top M_{Z_1}Y_1}{n}(\hat{\beta}_{ols} - \hat{\beta}_{2sls}).\end{aligned}$$

This together with the definitions of t_{H_2} and H_n completes the proof. $\blacksquare$

Proof of Lemma 3. Note that by $\hat{\beta}_{ols} \neq \hat{\beta}_{2sls}$ and (11), we have

$$\hat{\sigma}^2_{ols} t_{H_1} = \hat{\sigma}^2_{2sls} t_{H_2} = \hat{\sigma}^2_u t_{CF} = (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top ((\hat{Y}_1^\top M_{Z_1} \hat{Y}_1)^{-1} - (Y_1^\top M_{Z_1} Y_1)^{-1})^{-1} (\hat{\beta}_{ols} - \hat{\beta}_{2sls}) > 0 \quad (31)$$

and

$$\hat{\sigma}^2_{2sls} H_n = n^{-1} (\hat{\beta}_{ols} - \hat{\beta}_{2sls})^\top (Y_1^\top M_{Z_1} Y_1)(\hat{\beta}_{ols} - \hat{\beta}_{2sls}) > 0. \quad (32)$$

Therefore, the equalities in (14) of Lemma 2 imply that

$$\hat{\sigma}^2_{2sls} > \hat{\sigma}^2_{ols} > \hat{\sigma}^2_u. \quad (33)$$

Combining (31) and (33) yields $t_{CF} > t_{H_1} > t_{H_2}$. By (33) and $(Y_1^\top M_{Z_1} Y_1)^{-1} > 0$, we can show that

$$\hat{\sigma}^2_{2sls} (\hat{Y}_1^\top M_{Z_1} \hat{Y}_1)^{-1} - \hat{\sigma}^2_{ols} (Y_1^\top M_{Z_1} Y_1)^{-1} > \hat{\sigma}^2_{2sls} [(\hat{Y}_1^\top M_{Z_1} \hat{Y}_1)^{-1} - (Y_1^\top M_{Z_1} Y_1)^{-1}],$$

which together with $\hat{\beta}_{ols} \neq \hat{\beta}_{2sls}$ and the definitions of t_{H_2} and t_{H_3} implies that $t_{H_2} > t_{H_3}$. $\blacksquare$