7-1-2015

Inference in Near Singular Regression

Peter C.B. Phillips

INFERENCE IN NEAR SINGULAR REGRESSION

By

Peter C. B. Phillips

July 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 2009

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles. yale.edu/

Inference in Near Singular Regression*

Peter C. B. Phillips
Yale University, University of Auckland,
Singapore Management University & University of Southampton

June 27, 2015

Abstract

This paper considers stationary regression models with near-collinear regressors. Limit theory is developed for regression estimates and test statistics in cases where the signal matrix is nearly singular in finite samples and is asymptotically degenerate. Examples include models that involve evaporating trends in the regressors that arise in conditions such as growth convergence. Structural equation models are also considered and limit theory is derived for the corresponding instrumental variable estimator, Wald test statistic, and overidentification test when the regressors are endogenous.

Keywords: Endogeneity, Instrumental variable, Overidentification test, Regression, Singular Signal Matrix, Structural equation.

JEL classification: C23

1 Introduction

Near-collinear regressors arise frequently in empirical work in both time series and cross section data. The case of co-moving regressors is particularly well known and and has been extensively studied (Park and Philllips, 1988, 1989; Phillips, 1988, 1989; Sims, Stock and Watson, 1990; Toda and Phillips, 1993; Phillips, 1995) in the context of time series regression with some unit roots and possibly cointegrated regressors. Related problems of partial identification and weak instrumentation in structural model estimation have also proved to be relevant in applications and have been studied in a large literature following initial research on the asymptotic theory of these models by Phillips (1989) and Staiger and Stock (1997). Earlier important work by Sargan (1958, 1983) also considered some aspects of the impact of nearly unidentified models on estimation and inference. More recent work on common explosive roots has

*This paper was written during a cross-Canada rail journey during June 2015. It originated in a Yale Take Home Examination given in the Fall, 2014. The author acknowledges support of the NSF under Grant SES 12-58258.

shown that near collinearity can produce inconsistencies even in the presence of extremely strong regressor signals (Phillips and Magdalinos, 2013).

While this research primarily involves parametric models and linear systems of equations, nonlinear regressions are also affected by near collinearity in the regressors, weak identification (Stock and Wright, 2000), and singularities in the limit theory that can produce inconsistencies and differing rates of convergence (Park and Phillips, 2000). It has recently been discovered that nonparametric kernel regression, an area of econometrics to which Aman Ullah has made many lasting contributions including a foundational text (Pagan and Ullah, 1999), is also affected by singularities and differing convergence rates when the regressors are nonstationary (Phillips et al, 2014; Li et al, 2015).

The present work considers analogous problems associated with near-collinear regressors that arise in stationary regression. To illustrate, we study the case of a near-singular signal matrix where there is degeneracy in the limit. Such cases occur in practical econometric work when there are evaporating trends or decay effects in the data that produce asymptotic co-movement, as in growth convergence modeling (Phillips and Sul, 2007 and 2009), or when power law time trends need to be estimated (Phillips, 2007; Robinson, 2012).

We develop stationary asymptotics for estimates and tests in regressions where signal matrix singularities that arise in the limit produce inconsistencies in estimation and failures in central limit theory. We also provide limit theory for instrumental variable (IV) regression and the associated Wald test statistic and overidentification test when the regressor is endogenous. The limit theory is developed for stationary regressions with martingale difference errors.

The remainder of the paper is organized as follows. Section 2 examines a prototypical stationary linear regression model with asymptotically collinear regressors and develops limit theory for the coefficient estimates and block Wald test. Although the coefficient estimates are generally inconsistent, some linear functions as well as the equation error variance are shown to be consistently estimable. Section 3 develops similar limit theory for instrumental variable estimates and test statistics in the structural model case with endogenous regressors. Section 4 concludes and discusses extensions. Proofs are given in the Appendix.

2 Singular Regression Models and Limit Theory

2.1 A Prototypical Model

We study the linear model

$$y_t = x_t' \beta + u_{0t}, \quad t = 1, ..., n \tag{1}$$

where β is an unknown $k \times 1$ vector of parameters and the errors u_{0t} are martingale differences with respect to the filtration $\mathcal{F}_t = \sigma \left\{ u_{0t}, u_{0t-1}, ...; x_{t+1}, x_t, ... \right\}$ and with conditional variance $\mathbb{E} \left\{ u_{0t}^2 | \mathcal{F}_{t-1} \right\} = \sigma_{00}$ *a.s.*. The regressor x_t in (1) is assumed to have components with differing asymptotic characteristics that lead to a limiting singular system. In particular, upon transformation by some (unknown) nonsingular matrix $G' = [G_a, G_b]'$, we have the partitioned system

$$y_t = x_t' G G^{-1} \beta + u_{0t} = w_{at}' \alpha_a + w_{bt}' \alpha_b + u_{0t}, \tag{2}$$

with

$$w_t = \begin{bmatrix} w_{at} \\ w_{bt} \end{bmatrix} := G' x_t = \begin{bmatrix} G_a' x_t \\ G_b' x_t \end{bmatrix}, \ \alpha := \begin{bmatrix} \alpha_a \\ \alpha_b \end{bmatrix} := \begin{bmatrix} G^{a\prime} \beta \\ G^{b\prime} \beta \end{bmatrix} =: G^{-1}\beta,$$

involving a k_a- vector of stationary, ergodic variates w_{at} and a k_b- vector w_{bt} which satifies $\sum_{t=1}^n w_{bt} w_{bt}' \rightarrow_{a.s.} \Sigma_{bb}$. Sample moments of the components w_{at} and vector w_{bt} therefore have different orders of magnitude. Let $X' = [x_1, ..., x_n]'$, $W' = [w_1, ..., w_n]'$, $u_0 = [u_{01}, ..., u_{0n}]$, and $y' = [y_1, ..., y_n]$. In observation matrix form, (2) then takes the form

$$y = X\beta + u_0 = W\alpha + u_0. \tag{3}$$

Upon standardization with the matrix $D_n = diag\left[\sqrt{n} I_{k_a}, I_{k_b}\right]$ the sample moment matrix $X'X = \sum_{t=1}^n x_t x_t'$ satisfies, as shown in (9) below,

$$D_n^{-1} G' X' X G D_n^{-1} \rightarrow_{a.s.} \Sigma = \begin{bmatrix} \Sigma_{aa} & 0 \\ 0 & \Sigma_{bb} \end{bmatrix} > 0, \tag{4}$$

leading to

$$\frac{\sum_{t=1}^n w_{at} w_{at}'}{\sum_{t=1}^n u_{0t}^2} \rightarrow_{a.s.} \frac{\Sigma_{aa}}{\sigma_{00}}, \text{ and } \frac{\sum_{t=1}^n w_{bt} w_{bt}'}{\sum_{t=1}^n u_{0t}^2} = O_{a.s}\left(n^{-1}\right).$$

So signal to noise ratios differ by an order of magnitude in the directions w_{at} and w_{bt}.

To fix ideas, we henceforth assume that the regressors x_t in (2) have the partitioned form

$$x_t = \begin{bmatrix} x_{at} \\ x_{bt} \end{bmatrix} = \begin{bmatrix} x_t^0 \\ \Pi x_t^0 + a_t v_t \end{bmatrix}, \tag{5}$$

where $\left(x_t^{0\prime}, v_t'\right)'$ is a $k_a + k_b$ vector of stationary ergodic time series, Π is an unknown constant matrix of dimension $k_b \times k_a$, and a_t is a deterministic sequence with $a_t \to 0$ as $t \to \infty$. The regressors x_{at} and x_{bt} may then be interpreted as asymptotically co-moving stationary regressors. For instance, when $a_t = 1/t$, we have $x_{bt} = \Pi x_{at} + O_{a.s}\left(\frac{1}{t}\right) \sim \Pi x_{at}$ as $t \to \infty$.

With this structure the system (1) has the partitioned form

$$y_t = x_{at}' \beta_a + x_{bt}' \beta_b + u_{0t}, \tag{6}$$

where $\beta' = (\beta_a, \beta_b)$ is a conformable partition of β. The block triangular transform matrix

$$G = \begin{bmatrix} I & -\Pi' \\ 0 & I \end{bmatrix}, G^{-1} = \begin{bmatrix} I & \Pi' \\ 0 & I \end{bmatrix} =: \begin{bmatrix} G^{a\prime} \\ G^{b\prime} \end{bmatrix} \tag{7}$$

leads to the transformed parametric structure $\alpha = G^{-1}\beta$ written in partitioned form as

$$\begin{bmatrix} \alpha_a \\ \alpha_b \end{bmatrix} = \begin{bmatrix} G^{a'}\beta \\ G^{b\prime}\beta \end{bmatrix} = \begin{bmatrix} \beta_a + \Pi'\beta_b \\ \beta_b \end{bmatrix}$$

and corresponding regressor structure

$$w_t = G'x_t = \begin{bmatrix} w_{at} \\ w_{bt} \end{bmatrix} = \begin{bmatrix} x_t^0 \\ a_t v_t \end{bmatrix}. \tag{8}$$

Here, $w_{bt} = a_t v_t$ involves a stationary component v_t and an evaporating deterministic trend factor, $a_t = o(1)$ as $t \to \infty$, of the type that arises in the study of growth convergence (Phillips and Sul, 2007, 2009). The regression components (x_{at}, x_{bt}) in the untransformed model (6) are therefore asymptotically collinear because $w_{bt} = a_t v_t = o_{a.s.}(1)$ as $t \to \infty$.

Let $s_t = \left(x_t^{0\prime}, v_t'\right)'$ and $q_t = s_t u_{0t} = \left(q_{xt}', q_{vt}'\right)$, partitioned conformably with s_t. We make the following conditions on these components to facilitate the development of the limit theory.

Assumption A **(i)** *u_{0t} is a martingale difference sequence (mds) with respect to the filtration $\mathcal{F}_t = \sigma\{u_{0t}, u_{0t-1}, ...; s_{t+1}, s_t, ...\}$ and with conditional variance $\mathbb{E}\left\{u_{0t}^2|\mathcal{F}_{t-1}\right\} = \sigma_{00}$ a.s.*

(ii) *$r_t = (s_t, u_{0t})'$ is strictly stationary and ergodic with $E\left(\|r_t\|^{2+\delta}\right) < \infty$ for some $\delta > 0$, and variance matrix $\Sigma_{rr} = diag\left[\Sigma_{ss}, \sigma_{00}\right] > 0$.*

Assumption B *a_t is a deterministic sequence for which either*

(i) $\sum_{t=1}^{\infty} |a_t|^{1+\eta} < \infty$ *for some (possibly small) $\eta \in (0,1)$, or*

(ii) $\sum_{t=1}^{\infty} |a_t| < \infty$.

As shown in Lemma A in the Appendix, Assumptions A(i) and (ii) ensure that a functional law applies to partial sums of the mds $q_t = \left(q_{xt}', q_{vt}'\right)$, so that $n^{-1/2}\sum_{t=1}^{\lfloor n\cdot \rfloor} q_t \Rightarrow B_q(\cdot)$, with limiting Brownian motion vector B_q and covariance matrix $\Omega_{qq} = \sigma_{00}\Sigma_{ss}$ where $B_q = \left(B_{q_x}', B_{q_v}'\right)'$ and

$$\Sigma_{ss} = \mathbb{E}\left(s_t s_t'\right) = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xv} \\ \Sigma_{vx} & \Sigma_{vv} \end{bmatrix} > 0$$

are conformably partitioned with q_t. Assumption B requires absolute summability of the deterministic sequence $\{a_t\}$ in B(ii) or the alternate $(1+\eta)$ absolute summability in B(i). These conditions imply that a_t is an evaporating sequence, so that $a_t \to 0$, and they are sufficient to ensure $a.s.$ summability of certain sums that appear in the limit theory such $\sum_{t=1}^{\infty} a_t^2 v_t v_t'$ and $\sum_{t=1}^{\infty} a_t x_t^0 v_t'$ in the following analysis. For example, $a_t = t^{-1}$ satisfies B(i) for any $\eta > 0$, and $a_t = t^{-1}(\log t)^{-1-\epsilon}$ satisfies B(ii) for any $\epsilon > 0$.

Under Assumptions A and B we have the following explicit form for the limit behavior of the standardized signal matrix in (4)

$$D_n^{-1}G'X'XGD_n^{-1} \rightarrow_{a.s.} \Sigma = \begin{bmatrix} \Sigma_{aa} = \mathbb{E}\left(x_t^0 x_t^{0\prime}\right) & 0 \\ 0 & \Sigma_{bb} = \sum_{t=1}^{\infty} a_t^2 v_t v_t' \end{bmatrix}. \quad (9)$$

Observe that the sum $\sum_{t=1}^{\infty} a_t^2 v_t v_t' < \infty$ *a.s.* since

$$\mathbb{E}\left(\sum_{t=1}^{\infty} a_t^2 v_t v_t'\right) = \sum_{t=1}^{\infty} a_t^2 \mathbb{E}\left(v_t v_t'\right) = \Sigma_{vv} \sum_{t=1}^{\infty} a_t^2 < \infty,$$

under both B(i) and (ii). The off-diagonal block in (9) is a zero matrix because: under B(ii), $\mathbb{E}\left(\sum_{t=1}^{\infty} \left\|x_t^0 v_t'\right\| |a_t|\right) = \mathbb{E}\left(\left\|x_t^0 v_t'\right\|\right) \sum_{t=1}^{\infty} |a_t| < \infty$, in which case $\sum_{t=1}^{\infty} x_t^0 v_t' a_t$ converges almost surely and $n^{-1/2} \sum_{t=1}^{n} x_t^0 v_t a_t = O_{a.s.}\left(n^{-1/2}\right)$; alternatively, under B(i), we have by Hölder's inequality

$$\begin{aligned} \frac{1}{\sqrt{n}} \mathbb{E}\left(\sum_{t=1}^{n} \left\|x_t^0 v_t'\right\| |a_t|\right) &= \mathbb{E}\left(\left\|x_t^0 v_t'\right\|\right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |a_t| \\ &\leq \mathbb{E}\left(\left\|x_t^0 v_t'\right\|\right) \frac{n^{\frac{\eta}{1+\eta}}}{\sqrt{n}} \left(\sum_{t=1}^{n} |a_t|^{1+\eta}\right)^{\frac{1}{1+\eta}} \\ &= O\left(\frac{1}{n^{\frac{1}{2} - \frac{\eta}{1+\eta}}}\right) = o(1) \text{ for all } \eta \in (0,1), \end{aligned}$$

and then $n^{-1/2} \sum_{t=1}^{n} x_t^0 v_t a_t \rightarrow_{L_1} 0$.

The standardized signal matrix therefore has a random limit and no invariance principle applies because $\sum_{t=1}^{\infty} a_t^2 v_t v_t'$ depends on the distribution of v_t. Further,

$$D_n^{-1} W' u_0 = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} w_{at} u_{0t} \\ \sum_{t=1}^{n} w_{bt} u_{0t} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t^0 u_{0t} \\ \sum_{t=1}^{n} a_t v_t u_{0t} \end{bmatrix} \Rightarrow \begin{bmatrix} B_{q_x}(1) \\ Q_v \end{bmatrix}, \quad (10)$$

where $Q_v := \sum_{t=1}^{\infty} a_t v_t u_{0t}$ converges almost surely since $\sum_{t=1}^{m} a_t v_t u_{0t}$ is an L_2 martingale with $\sum_{t=1}^{\infty} a_t^2 \mathbb{E}\left\|v_t v_t' u_{0t}^2\right\| = \sigma_{00} \mathbb{E}\left\|v_t v_t'\right\| \sum_{t=1}^{\infty} a_t^2 < \infty$. So $D_n^{-1} W' u_0$ converges weakly but does not satisfy an invariance principle, the distribution of the limit component Q_v depending on the distribution of the component variates (v_t, u_{0t}).

2.2 Near-Singular Least Squares Regression

The parameter vector β in (1) is estimated by ordinary least squares regression and the null hypothesis $\mathbb{H}_0 : \beta = 0$ is tested using the Wald statistic $W_n = \hat{\beta}' X' X \hat{\beta} / \hat{\sigma}^2$, where $\hat{\sigma}^2 = n^{-1} y' \left(I_n - X\left(X'X\right)^{-1} X\right) y$ is the usual sample variance of the regression residuals. The limit behavior of the regression components $\left\{\hat{\beta}, \hat{\sigma}^2, W_n\right\}$ is as follows.

Theorem 1 Under Assumptions A and B

(i) $\begin{bmatrix} \hat{\beta}_a - \beta_a \\ \hat{\beta}_b - \beta_b \end{bmatrix} \Rightarrow \begin{bmatrix} -\Pi' \\ I_{k_b} \end{bmatrix} \xi_b$, where $\xi_b = \left(\sum_{t=1}^{\infty} a_t^2 v_t v_t'\right)^{-1} \sum_{t=1}^{\infty} a_t v_t u_{0t}$,

(ii) $\hat{\sigma}^2 \to_p \sigma_{00}$,

(iii) $W_n \Rightarrow \chi^2_{k_a} + \zeta_b' \zeta_b$, where $\zeta_b := \Sigma_{bb}^{1/2} \xi_b / \sigma_{00}^{1/2} = \Sigma_{bb}^{-1/2} \sum_{t=1}^{\infty} a_t v_t u_{0t} / \sigma_{00}^{1/2}$.

It follows from (i) that both estimates $\hat{\beta}_a$ and $\hat{\beta}_b$ are inconsistent and converge to random quantities dependent on ξ_b. No invariance principle applies because the distribution of ξ_b depends on the distribution of the data through the inputs $\{v_t, u_{0t}\}_{t=1}^{\infty}$. The limit theory also has degenerate dimension k_b because $\hat{\beta}_a - \beta_a$ is asymptotically proportional to $\hat{\beta}_b - \beta_b$. Thus, the asymptotic singularity in the signal matrix leads to inconsistency in the regression coefficients and degeneracy in their limit distribution. As noted above, the weak signal is in the direction w_{bt} for which the sample excitation matrix $\sum_{t=1}^{n} a_t^2 v_t v_t' \to \sum_{t=1}^{\infty} a_t^2 v_t v_t'$ does not diverge as the sample size $n \to \infty$, leading to the inconsistency and a singular limit distribution that depends on the limit regression coefficient $\xi_b = \left(\sum_{t=1}^{\infty} a_t^2 v_t v_t'\right)^{-1} \sum_{t=1}^{\infty} a_t v_t u_{0t}$ in this direction.

Nonetheless, there are identifiable and estimable functions of the coefficients. In particular, as shown in the proof of (i), the linear combination $\beta_a + \Pi' \beta_b$ is consistently estimated by $\hat{\beta}_a + \Pi' \hat{\beta}_b$ at a $\sqrt{n}$ rate, giving a consistently estimable function of the original coordinates with the normal limit distribution

$$\sqrt{n}\left(\hat{\beta}_a + \Pi' \hat{\beta}_b - \beta_a - \Pi' \beta_b\right) \Rightarrow N\left(0, \sigma_{00} \Sigma_{aa}^{-1}\right). \tag{11}$$

The matrix Π is generally unknown but it can be consistently estimated at an $O(n)$ rate. In particular, if the partition structure of $x_t = (x_{at}', x_{bt}')'$ is known, least squares regression of x_{bt} on x_{at} gives $\hat{\Pi} = \left(\sum_{t=1}^{n} x_{bt} x_{at}'\right)\left(\sum_{t=1}^{n} x_{at} x_{at}'\right)^{-1}$ and simple manipulations reveal that $n\left(\hat{\Pi} - \Pi\right) \to_{a.s.} \left(\sum_{t=1}^{\infty} a_t v_t x_t^{0\prime}\right) \Sigma_{aa}^{-1}$. Then $\hat{\beta}_a + \hat{\Pi}' \hat{\beta}_b$ is consistent for $\beta_a + \Pi' \beta_b$ with the same $\sqrt{n}$ rate of convergence and asymptotic distribution as (11).

Curiously, as shown in (ii), the least squares error variance estimate $\hat{\sigma}^2$ is consistent even though the regression coefficients are inconsistent. The reason is that asymptotic collinearity in the regressor vector x_t does not prevent consistency of the residual variance. In particular, the fitted residual is

$$\begin{aligned} \hat{u}_{0t} &= y_t - x_t' \hat{\beta} = u_{0t} - x_t' \left(\hat{\beta} - \beta\right) = u_{0t} - w_t' (\hat{\alpha} - \alpha) \\ &= u_{0t} - x_t^{0\prime} (\hat{\alpha}_a - \alpha_a) - a_t v_t (\hat{\alpha}_b - \alpha_b) \\ &= u_{0t} - w_t' D_n^{-1} D_n (\hat{\alpha} - \alpha) \end{aligned}$$

and, since $D_n(\hat{\alpha}-\alpha)$ and $D_n^{-1}W'WD_n^{-1}$ are both $O_p(1)$ from (15) and (17) in the proof of Theorem 1, we find that

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}\sum_{t=1}^{n}\hat{u}_{0t}^2 = \frac{1}{n}\sum_{t=1}^{n}u_{0t}^2 - \frac{1}{n}(\hat{\alpha}-\alpha)' D_n\left(D_n^{-1}W'WD_n^{-1}\right)D_n(\hat{\alpha}-\alpha) \\ &= \frac{1}{n}\sum_{t=1}^{n}u_{0t}^2 + o_p(1) \to_p \sigma_{00}.\end{aligned}$$

From (iii), W_n is a limiting mixture of a chi square variate and the squared length of the vector variate ζ_b. No invariance principle holds because ζ_b depends on the data distribution through $\{v_t, u_{0t}\}_{t=1}^{\infty}$. However, when (v_t, u_{0t}) is Gaussian, then $u_{0t} \sim iid\, N(0,\sigma_{00})$ is independent of $\{v_t\}$ because $E(v_t u_{0t}) = 0$ in view of Assumption A(ii). Then $\zeta_b =_d N(0, I_{k_b})$ and $\zeta_b'\zeta_b \sim_d \chi^2_{k_b}$, so that $W_n \Rightarrow \chi^2_k$. Thus, the usual limit theory for the test statistic W_n applies when the input variates are Gaussian.

3 Singular Structural Model and IV Estimation

3.1 Model Formulation and Limit Theory

We now consider the structural equation case where the regressor x_t in (1) is endogenous. The asymptotic characteristics of x_t are assumed to be the same as those given earlier, so that (4) and (5) continue to hold but now $\mathbb{E}(x_t u_{0t}) = \Sigma_{x0} \neq 0$. Let z_t be a $K \times 1$ vector of instruments with $K \geq k+1$. The IV estimator is $\beta_{IV} = (X'P_Z X)^{-1}(X'P_Z y)$ and the estimation error has the form

$$\begin{aligned}\beta_{IV} - \beta &= (X'P_Z X)^{-1} X'P_Z u_0 = \left(G'^{-1}W'P_Z W G^{-1}\right)^{-1} G'^{-1}W'P_Z u_0 \\ &= G(W'P_Z W)^{-1} W'P_Z u_0,\end{aligned}$$

with G and W defined as in (7 & 8) and corresponding coefficient estimates $\alpha_{IV} = G^{-1}\beta_{IV}$ with estimation error

$$\alpha_{IV} - \alpha = G^{-1}(\beta_{IV} - \beta) = (W'P_Z W)^{-1}(W'P_Z u_0).$$

We replace Assumption A with the following.

Assumption A′ (i) *u_{0t} is a martingale difference sequence (mds) with respect to the filtration $\mathcal{F}_t = \sigma\{u_{0t}, u_{0t-1}, ...; z_{t+1}, z_t, ...\}$ and with conditional variance $\mathbb{E}\{u_{0t}^2|\mathcal{F}_{t-1}\} = \sigma_{00}$ a.s.*

(ii) *$r_t = \left(x_t^0, v_t, z_t, u_{0t}\right)'$ is strictly stationary and ergodic with $E\left(\|r_t\|^{2+\delta}\right) < \infty$ for some $\delta > 0$, and variance matrix*

$$\Sigma_{rr} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xv} & \Sigma_{xz} & \Sigma_{x0} \\ \Sigma_{vx} & \Sigma_{vv} & \Sigma_{vz} & \Sigma_{v0} \\ \Sigma_{zx} & \Sigma_{zv} & \Sigma_{zz} & 0 \\ \Sigma_{0x} & \Sigma_{0v} & 0 & \sigma_{00} \end{bmatrix} > 0$$

with Σ_{xz} having full rank $k_a < K$.

Assumption A$'$(i) ensures that the orthogonality condition $\mathbb{E}\{z_t u_{0t}\} = 0$ holds, giving instrument validity to z_t, and A$'$(ii) imposes the partial relevance rank condition that $\text{rank}(\Sigma_{zx}) = k_a < K$. The full relevance condition $\text{rank}[\Sigma_{zx}, \Sigma_{zv}] = k$ with respect to x_t, or equivalently the pair $\left(x_t^0, v_t\right)$, is not required in what follows as the regressor singularity dominates the asymptotics.

The parameter vector β in (1) is estimated by instrumental variables regression using the instruments z_t. The null hypothesis $\mathbb{H}_0 : \beta = 0$ is block tested using the corresponding Wald statistic $\tilde{W}_n = \beta_{IV}' X' P_Z X \beta_{IV} / \tilde{\sigma}^2$, where $\tilde{\sigma}^2 = n^{-1}\tilde{u}'\tilde{u}$ is the usual sample variance of the regression residuals $\tilde{u} = y - X\beta_{IV}$. We also consider the Sargan overidentification test statistic for testing the validity of the instruments. Using the IV residuals

$$\begin{aligned} \tilde{u} &= y - X\beta_{IV} = u_0 - X\left(X'P_Z X\right)^{-1} X' P_Z u_0 \\ &= u_0 - W\left(W'P_Z W\right)^{-1} W' P_Z u_0, \end{aligned}$$

we write the projection

$$P_Z \tilde{u} = \left\{P_z - P_z X\left(X'P_Z X\right)^{-1} X' P_Z\right\} u_0 = \left\{P_z - P_z W\left(W'P_Z W\right)^{-1} W' P_Z\right\} u_0.$$

Then the Sargan test for overidentification has the form

$$\begin{aligned} S_n &= \tilde{u}' P_Z \tilde{u} / \tilde{\sigma}^2 = u_0'\left\{P_z - P_z X\left(X'P_Z X\right)^{-1} X' P_Z\right\} u_0 / \tilde{\sigma}^2 \\ &= u_0'\left\{P_z - P_z W\left(W'P_Z W\right)^{-1} W' P_Z\right\} u_0 / \tilde{\sigma}^2. \end{aligned}$$

The limit behavior of the IV regression components $\left\{\beta_{IV}, \tilde{\sigma}^2, \tilde{W}_n, S_n\right\}$ is given in the following result where $MN(0, V)$ signifies a mixed normal distribution with zero mean and mixing variance matrix V.

Theorem 2 *Under Assumptions A$'$ and B*

(i) $F_n G^{-1}\left(\beta_{IV} - \beta\right) \Rightarrow MN\left(0, \sigma_{00}\begin{bmatrix} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xz}\Sigma_{zz}^{-1}A_z \\ A_z'\Sigma_{zz}^{-1}\Sigma_{zx} & A_z'\Sigma_{zz}^{-1}A_z \end{bmatrix}^{-1}\right)$,
where $A_z = \sum_{t=1}^{\infty} a_t z_t v_t'$, and in partitioned component form

$$\frac{1}{\sqrt{n}}\begin{bmatrix} \beta_{a,IV} - \beta_a \\ \beta_{b,IV} - \beta_b \end{bmatrix} \Rightarrow \begin{bmatrix} -\Pi' \\ I_{k_b} \end{bmatrix} \times MN\left(0, \sigma_{00} H^{-1}\right), \tag{12}$$

where $H = A_z'\Sigma_{zz}^{-1}A_z - A_z'\Sigma_{zz}^{-1}\Sigma_{zx}\left(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right)^{-1}\Sigma_{xz}\Sigma_{zz}^{-1}A_z$.

(ii) $\tilde{\sigma}^2 \to_p \sigma_{00}\{1 + \omega_{zz}\}$, *where $\omega_{zz} = \psi_b' H^{-1/2}\left(\sum_{t=1}^{\infty} a_t^2 v_t v_t'\right) H^{-1/2}\psi_b$ and $\psi_b \sim_d N(0, I_{k_b})$.*

(iii) $\tilde{W}_n \Rightarrow \chi_k^2 / \{1 + \omega_{zz}\}$.

(iv) $S_n \Rightarrow \chi^2_{K-k} / \{1 + \omega_{zz}\}$

The standardized and centred IV estimate $F_n G^{-1} (\beta_{IV} - \beta) = F_n (\alpha_{IV} - \alpha)$ has a mixed normal limit, where the mixing variance matrix depends on the matrix $A_z = \sum_{t=1}^{\infty} a_t z_t v_t'$, which in turn depends on the distribution of (z_t, v_t) and the deterministic sequence (a_t). This random matrix A_z is a measure of the importance of the near-collinearity in the system between the component regressors $x_{at} = x_t^0$ and $x_{bt} = \Pi x_t^0 + a_t v_t$ when the system is estimated using instrumental variables z_t. Importantly, the series $\sum_{t=1}^{\infty} a_t z_t v_t' < \infty$ *a.s.*, so that A_z is a well defined random matrix.

As is apparent from (12), the individual IV component vectors $\beta_{a,IV}$ and $\beta_{b,IV}$ both have divergent behavior at the $\sqrt{n}$ rate. Hence, the effects of the weak signal arising from the near collinearity in the regressors that is evident in least squares regression under exogeneity, is exacerbated by endogeneity, even when the instruments are valid, satisfying both orthogonality and strong relevance conditions. Thus, near-collinearity in the presence of endogeneity, even with strong instruments in regression, leads to divergent behavior in the estimates.

On the other hand, as in the case of exogenous x_t and as shown in the proof of (i), there are some estimable components. In particular, the linear combination $\beta_a + \Pi' \beta_b$ is again consistently estimated, here by $\beta_{a,IV} + \Pi' \beta_{b,IV}$ and at a $\sqrt{n}$ rate, giving a consistently estimable function of the original coordinates with the mixed normal limit distribution

$$\sqrt{n} \left(\beta_{a,IV} + \Pi' \beta_{b,IV} - \beta_a - \Pi' \beta_b \right) \Rightarrow MN \left(0, \sigma_{00} H_{\beta\beta}^{-1} \right), \tag{13}$$

where $H_{\beta\beta} = \left[\Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx} - \Sigma_{xz} \Sigma_{zz}^{-1} A_z \left(A_z' \Sigma_{zz}^{-1} A_z \right)^{-1} A_z' \Sigma_{zz}^{-1} \Sigma_{zx} \right]$ and with the mixing matrix A_z again influencing the asymptotics. The matrix Π is generally unknown but, as earlier in the regression model case, it can be consistently estimated at an $O(n)$ rate by least squares regression of x_{bt} on x_{at}. In the same way, the estimate $\hat{\Pi} = \left(\sum_{t=1}^{n} x_{bt} x_{at}' \right) \left(\sum_{t=1}^{n} x_{at} x_{at}' \right)^{-1} \to_{a.s} \Pi$ with limit distribution $n \left(\hat{\Pi} - \Pi \right) \to_{a.s.} \left(\sum_{t=1}^{\infty} a_t v_t x_t^{0\prime} \right) \Sigma_{aa}^{-1}$. So, $\beta_{a,IV} + \hat{\Pi}' \beta_{b,IV}$ is again consistent for $\beta_a + \Pi' \beta_b$ with the same $\sqrt{n}$ rate of convergence and asymptotic distribution as (13).

Part (ii) shows that the usual error variance estimate is inconsistent and asymptotically overestimates σ_{00} by the asymptotic bias expression $\sigma_{00} \omega_{zz} = \psi_b' H^{-1/2} \left(\sum_{t=1}^{\infty} a_t^2 v_t v_t' \right) H^{-1/2} \psi_b$. As shown in the proof, this asymptotic bias arises in the residual variance estimate from the limit of the following component involving a quadratic form in the estimation error $(\alpha_{IV} - \alpha)$

$$\begin{aligned} & (\alpha_{IV} - \alpha)' F_n \left(\frac{1}{n} F_n^{-1} W' W F_n^{-1} \right) F_n (\alpha_{IV} - \alpha) \\ = \; & \left\{ \frac{1}{\sqrt{n}} (\alpha_{b,IV} - \alpha_b)' \right\} \left(\sum_{t=1}^{n} a_t^2 v_t v_t' \right) \left\{ \frac{1}{\sqrt{n}} (\alpha_{b,IV} - \alpha_b) \right\} + o_p(1). \end{aligned}$$

Thus, in contrast to the linear regression case, the estimation error is not negligible when estimating the error variance and produces error variance estimation bias in the limit.

It follows from Part (iii) that the limit distribution of the Wald test of the block hypothesis $H_0 : \beta = 0$ is a mixed chi-square distribution with degrees of freedom k and scale mixing coefficient $\{1 + \omega_{zz}\}^{-1} < 1$ $a.s.$. In particular, $\tilde{W}_n \Rightarrow \chi^2_k / \{1 + \omega_{zz}\} \leq \chi^2_k$. Tail significance in the limit occurs when $\chi^2_k / \{1 + \omega_{zz}\} > cv_\alpha$ for the test critical value cv_α and this inequality implies that $\chi^2_k > cv_\alpha$ so that

$$P\left[\tilde{W}_n > cv_\alpha\right] \to \mathbb{P}\left[\chi^2_k / \{1 + \omega_{zz}\} > cv_\alpha\right] < \mathbb{P}\left[\chi^2_k > cv_\alpha\right].$$

Test based on $\tilde{W}_n$ with the usual χ^2_k critical value are therefore conservative asymptotically. The reason is that the IV error variance estimate $\tilde{\sigma}^2 \Rightarrow \sigma_{00}\{1 + \omega_{zz}\} > \sigma_{00}$ so that $\tilde{\sigma}^2$ overestimates σ_{00} and hence the Wald statistic $\tilde{W}_n$ is biased downwards, thereby favoring the null and leading to a conservative test.

This is a curious finding that implies size-controlled tests of $\beta = 0$ exist even when the regression coefficient β cannot be consistently estimated. Lack of asymptotic identifiability means that the equation error variance estimate is larger than the error variance in the limit, which then biases the test in favor of the null hypothesis, thereby reducing power. The impact on test power may be further investigated by doing an asymptotic power analysis for local and distant alternatives in various directions, a topic that is not pursued here.

The mixed normal limit distribution given in Part (i) of Theorem 2 presumes the invertibility of the (conditional) covariance matrix

$$\begin{bmatrix} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xz}\Sigma_{zz}^{-1}A_z \\ A_z'\Sigma_{zz}^{-1}\Sigma_{zx} & A_z'\Sigma_{zz}^{-1}A_z \end{bmatrix} = \begin{bmatrix} \Sigma_{xz} \\ A_z' \end{bmatrix} \Sigma_{zz}^{-1} \begin{bmatrix} \Sigma_{zx} & A_z \end{bmatrix}. \tag{14}$$

This matrix is nonsingular if the matrix $[\Sigma_{zx}, A_z]$ has full column rank. By assumption A$'$(ii) Σ_{zx} has full column rank k_a. The second component in the partition, $A_z = \sum_{t=1}^{\infty} a_t z_t v_t'$, is a random matrix. We take a leading case for analysis. In particular, if $(z_t, v_t) \sim_d iid\ N(0, \text{diag}(\Sigma_{zz}, \Sigma_{vv}))$, then

$$A_z = \sum_{t=1}^{\infty} a_t z_t v_t' \sim_d MN\left(0, \Sigma_{zz} \otimes \sum_{t=1}^{\infty} a_t^2 v_t v_t'\right),$$

which is a nondegenerate mixed normal distribution since $\sum_{t=1}^{\infty} a_t^2 v_t v_t' > 0$ $a.s.$, and Σ_{zz} is positive definite, by assumption. Deficient rank of (14) means that $[\Sigma_{zx}, A_z]\, g = \Sigma_{zx} g_a + A_z g_b = 0$ $a.s.$ for some $g' = (g_a', g_b) \neq 0$. That is, $A_z g_b = -\Sigma_{zx} g_a$, a constant vector $a.s.$. Note that $g_b \neq 0$, otherwise $\Sigma_{zx} g_a = 0$ which further implies $g_a = 0$ because Σ_{zx} has full rank by assumption. Since A_z has a full rank mixed normal distribution, it follows that for $g_b \neq 0$ we have $\mathbb{P}(A_z g_b = -\Sigma_{zx} g_a) = 0$. So the conditional covariance matrix (14) almost surely has full rank.

The final part of Theorem 2 considers the behavior of the Sargan overidentification test statistic for testing the validity of the instruments, showing that the Sargan statistic S_n is distributed in the limit as $\chi^2_{K-k}/\{1+\omega_{zz}\}$, which is proportional to a chi-squared variate with degrees of freedom $K-k$ corresponding to the degree of overidentification. This limit theory involves the error variance estimation bias through the presence of the scale factor $\{1+\omega_{zz}\}^{-1}$, which leads to a mixed chi-square limit. Thus, even though the estimates of the structural coefficients are inconsistent, the overidentification test is proportional to chi-square with the usual degrees of freedom. In consequence, like the Wald test, the overidentification test statistic is biased in favor of the null, leading to a conservative test of instrument validity.

4 Conclusion and Extension

In order to explore the implications for inference of asymptotic singularity in stationary regressors, it has been convenient to use the partitioned structure $x_t = (x'_{at}, x'_{bt})'$ given in (5). This structure leads to a triangular model in which the components of x_t are related according to the linear system $x_{bt} = \Pi x_{at} + a_t v_t$. In practical work, theory may sometimes suggest such a relationship in which variables are asymptotically stationary and co-related. In general, however, near-collinearity in stationary regressors may be suspected without knowledge of a particular functional relation. In such cases, it will be of practical interest to develop methods that enable inference about possible asymptotic singularities when the form of the dependence between the components of x_t is completely unknown. This topic of investigation is now being explored.

5 Appendix

The following preliminary result is useful.

Lemma A

(a) Under Assumptions A(i), A(ii) and with $s_t = (x_t, v_t)$, p*artial sums of* $q_t = s_t u_{0t} = (q'_{xt}, q'_{vt})$, *partitioned conformably with* $s_t = \left(x_t^{0\prime}, v_t'\right)'$, *satisfy the functional law* $n^{-1/2}\sum_{t=1}^{\lfloor n\cdot\rfloor} q_t \Rightarrow B_q(\cdot)$ *with limiting Brownian motion vector* $B_q = \left(B'_{q_x}, B'_{q_v}\right)'$, *conformably partitioned with* q_t, *and covariance matrix*

$$\Omega_{qq} = \sigma_{00}\Sigma_{ss} = \sigma_{00}\begin{bmatrix} \Sigma_{xx} & \Sigma_{xv} \\ \Sigma_{vx} & \Sigma_{vv} \end{bmatrix} > 0.$$

(b) *Under Assumptions A′(i) and A′(ii),* p*artial sums of* $z_t u_{0t}$ *satisfy the functional law* $n^{-1/2}\sum_{t=1}^{\lfloor n\cdot\rfloor} z_t u_{0t} \Rightarrow B_{zu}(\cdot)$ *with limiting Brownian motion* B_{zu} *with covariance matrix* $\sigma_{00}\Sigma_{zz}$.

Proof

Part (a) The CLT follows from Assumptions A(i) and A(ii) since $n^{-1/2}\sum_{t=1}^{n} q_t$ satisfies the stability and Lindeberg conditions. In particular, the martingale conditional variance matrix $n^{-1}\sum_{t=1}^{n} s_t s_t' \mathbb{E}\left\{u_{0t}^2|\mathcal{F}_{t-1}\right\} \rightarrow_{a.s} \sigma_{00}\Sigma_{ss}$ as $n \rightarrow \infty$, ensuring stability. The Lindeberg condition

$$n^{-1}\sum_{t=1}^{n}\mathbb{E}\left[\left\|s_t s_t'\right\| u_{0t}^2 1\left\{\left\|s_t s_t'\right\|^{1/2}|u_{0t}| > \sqrt{n}\epsilon\right\}|\mathcal{F}_{t-1}\right] \rightarrow_p 0, \text{ for all } \epsilon > 0$$

holds by standard manipulations since

$$1\left\{\left\|s_t s_t'\right\|^{1/2}|u_{0t}| > \sqrt{n}\epsilon\right\} \leq 1\left\{\left\|s_t s_t'\right\|^{1/2} > n^{1/4}\epsilon^{1/2}\right\} + 1\left\{|u_{0t}| > n^{1/4}\epsilon^{1/2}\right\},$$

and

$$\begin{aligned} & n^{-1}\sum_{t=1}^{n}\mathbb{E}\left[\left\|s_t s_t'\right\| u_{0t}^2 1\left\{\left\|s_t s_t'\right\|^{1/2}|u_{0t}| > \sqrt{n}\epsilon\right\}|\mathcal{F}_{t-1}\right] \\ \leq\ & n^{-1}\sum_{t=1}^{n}\left\|s_t s_t'\right\| 1\left\{\left\|s_t s_t'\right\|^{1/2} > n^{1/4}\epsilon^{1/2}\right\}\mathbb{E}\left\{u_{0t}^2|\mathcal{F}_{t-1}\right\} \\ & + n^{-1}\sum_{t=1}^{n}\left\|s_t s_t'\right\|\mathbb{E}\left[u_{0t}^2 1\left\{|u_{0t}| > n^{1/4}\epsilon^{1/2}\right\}|\mathcal{F}_{t-1}\right] \\ =\ & \sigma_{00} n^{-1}\sum_{t=1}^{n}\left\|s_t s_t'\right\| 1\left\{\left\|s_t s_t'\right\|^{1/2} > n^{1/4}\epsilon^{1/2}\right\} \\ & + \left(n^{-1}\sum_{t=1}^{n}\left\|s_t s_t'\right\|\right)\mathbb{E}\left[u_{01}^2 1\left\{|u_{01}| > n^{1/4}\epsilon^{1/2}\right\}|\mathcal{F}_0\right] \\ \rightarrow_{L_1}\ & 0. \end{aligned}$$

The functional law $n^{-1/2}\sum_{t=1}^{\lfloor n\cdot\rfloor} q_t \Rightarrow B_q(\cdot)$ then follows directly by Hall and Heyde (1980, theorem 4.1).

Part (b) The CLT follows in the same way from Assumptions A′(i) and A′(ii): $n^{-1/2}\sum_{t=1}^{n} z_t u_{0t}$ has martingale conditional variance matrix $n^{-1}\sum_{t=1}^{n} z_t z_t' \mathbb{E}\left\{u_{0t}^2|\mathcal{F}_{t-1}\right\} \rightarrow_{a.s} \sigma_{00}\Sigma_{zz}$ as $n \rightarrow \infty$, and the Lindeberg condition

$$n^{-1}\sum_{t=1}^{n}\mathbb{E}\left[\left\|z_t z_t'\right\| u_{0t}^2 1\left\{\left\|z_t z_t'\right\|^{1/2}|u_{0t}| > \sqrt{n}\epsilon\right\}|\mathcal{F}_{t-1}\right] \rightarrow_p 0, \quad \text{for all } \epsilon > 0$$

holds by the same argument given in part (a). The functional law again follows.

Proof of Theorem 1

Part (i) We start by considering the transformed system (3) and corresponding least squares estimate $\hat{\alpha} = (W'W)^{-1} W'y$. In view of (9) and (10) we have

$$D_n^{-1} W'W D_n^{-1} \rightarrow_{a.s.} \begin{bmatrix} \Sigma_{aa} = \mathbb{E}\left(x_t^0 x_t^{0\prime}\right) & 0 \\ 0 & \Sigma_{bb} = \sum_{t=1}^{\infty} a_t^2 v_t v_t' \end{bmatrix}, \quad (15)$$

and

$$D_n^{-1}W'u_0 = \left[\begin{array}{c} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} w_{at}u_{0t} \\ \sum_{t=1}^{n} w_{bt}u_{0t} \end{array}\right] \Rightarrow \left[\begin{array}{c} B_{q_x}(1) \\ Q_v := \sum_{t=1}^{\infty} a_t v_t u_{0t} \end{array}\right], \qquad (16)$$

so that

$$\begin{aligned} D_n\left(\hat{\alpha}-\alpha\right) &= \left[D_n^{-1}W'WD_n^{-1}\right]^{-1}\left[D_n^{-1}W'u_0\right] \\ &\Rightarrow \left[\begin{array}{c} \Sigma_{aa}^{-1}B_{q_x}(1) \\ \Sigma_{bb}^{-1}\sum_{t=1}^{\infty} a_t v_t u_{0t} \end{array}\right] =: \left[\begin{array}{c} \xi_a \\ \xi_b \end{array}\right], \end{aligned} \qquad (17)$$

where $\xi_a \equiv N\left(0, \sigma_{00}\Sigma_{aa}^{-1}\right)$ and $\xi_b = \left(\sum_{t=1}^{\infty} a_t^2 v_t v_t'\right)^{-1}\sum_{t=1}^{\infty} a_t v_t u_{0t}$. Next note that $\hat{\beta}-\beta = G\left(\hat{\alpha}-\alpha\right) = GD_n^{-1}D_n\left(\hat{\alpha}-\alpha\right)$, so that

$$\left(GD_n^{-1}\right)^{-1}\left(\hat{\beta}-\beta\right) = D_n\left(\hat{\alpha}-\alpha\right) \Rightarrow \left(\xi_a', \xi_b'\right)'.$$

Now

$$\left[\begin{array}{c} \hat{\beta}_a-\beta_a \\ \hat{\beta}_b-\beta_b \end{array}\right] = GD_n^{-1}D_n\left[\begin{array}{c} \hat{\alpha}_a-\alpha_a \\ \hat{\alpha}_b-\alpha_b \end{array}\right] = \left[\begin{array}{cc} \frac{1}{\sqrt{n}}I_{k_a} & -\Pi' \\ 0 & I_{k_b} \end{array}\right] D_n\left[\begin{array}{c} \hat{\alpha}_a-\alpha_a \\ \hat{\alpha}_b-\alpha_b \end{array}\right],$$

so that

$$D_nG^{-1}\left[\begin{array}{c} \hat{\beta}_a-\beta_a \\ \hat{\beta}_b-\beta_b \end{array}\right] = D_n\left[\begin{array}{c} \hat{\alpha}_a-\alpha_a \\ \hat{\alpha}_b-\alpha_b \end{array}\right] \Rightarrow \left[\begin{array}{c} \xi_a \\ \xi_b \end{array}\right].$$

That is

$$\left[\begin{array}{cc} \sqrt{n}I_{k_a} & \sqrt{n}\Pi' \\ 0 & I_{k_b} \end{array}\right]\left[\begin{array}{c} \hat{\beta}_a-\beta_a \\ \hat{\beta}_b-\beta_b \end{array}\right] \Rightarrow \left[\begin{array}{c} \xi_a \\ \xi_b \end{array}\right],$$

It follows that

$$\begin{aligned} \hat{\beta}_b-\beta_b &\Rightarrow \xi_b, \\ \sqrt{n}\left(\hat{\beta}_a-\beta_a\right)+\sqrt{n}\Pi'\left(\hat{\beta}_b-\beta_b\right) &\Rightarrow N\left(0, \sigma_{00}\Sigma_{aa}^{-1}\right), \end{aligned}$$

which leads to

$$\left[\begin{array}{c} \hat{\beta}_a-\beta_a \\ \hat{\beta}_b-\beta_b \end{array}\right] = \left[\begin{array}{c} -\Pi' \\ I_{k_b} \end{array}\right]\left(\hat{\beta}_b-\beta_b\right)+o_p(1) \Rightarrow \left[\begin{array}{c} -\Pi' \\ I_{k_b} \end{array}\right]\xi_b.$$

Hence, both $\hat{\beta}_a$ and $\hat{\beta}_b$ are inconsistent with limits that are random, dependent on $\xi_b = \left(\sum_{t=1}^{\infty} a_t^2 v_t v_t'\right)^{-1}\sum_{t=1}^{\infty} a_t v_t u_{0t}$, and of degenerate dimension k_b because $\hat{\beta}_a-\beta_a$ is asymptotically proportional to $\hat{\beta}_b-\beta_b$. No invariance principle applies because the distribution of ξ_b depends on the distribution of the data.

Part (ii) Note that

$$\begin{aligned}
\hat{\sigma}^2 &= n^{-1}u_0'\left(I_n - X\left(X'X\right)^{-1}X'\right)u_0 = \frac{u_0'u_0}{n} - \frac{1}{n}u_0'XG\left(G'X'XG\right)^{-1}G'X'u_0 \\
&= \frac{u_0'u_0}{n} - \frac{1}{n}u_0'W\left(W'W\right)^{-1}W'u_0 = \frac{u_0'u_0}{n} - \frac{1}{n}u_0'WD_n^{-1}\left(D_n^{-1}W'WD_n^{-1}\right)^{-1}D_n^{-1}W'u_0 \\
&\to_p \mathbb{E}\left(u_{0t}^2\right) = \sigma_{00},
\end{aligned}$$

since $n^{-1}u_0'u_0 \to_{a.s} \mathbb{E}\left(u_{0t}^2\right)$ by the ergodic theorem, $D_n^{-1}W'WD_n^{-1} \to_{a.s.} \operatorname{diag}\left\{\Sigma_{aa}, \Sigma_{bb}\right\} > 0$ by (15), and $D_n^{-1}W'u_0 = O_p(1)$ by (16). Hence, $\hat{\sigma}^2$ is consistent for σ_{00}.

Part (iii) Under the null $\mathbb{H}_0 : \beta = 0$, we have $\alpha = G^{-1}\beta = 0$ and

$$\begin{aligned}
W_n &= \hat{\beta}'X'X\hat{\beta}/\hat{\sigma}^2 = \hat{\beta}'G'^{-1}G'X'XGG^{-1}\hat{\beta}/\hat{\sigma}^2 = \hat{\alpha}'W'W\hat{\alpha}/\hat{\sigma}^2 \\
&= \hat{\alpha}'D_nD_n^{-1}W'WD_n^{-1}D_n\hat{\alpha}/\hat{\sigma}^2 \\
&\Rightarrow \left(\xi_a', \xi_b'\right)\begin{bmatrix} \Sigma_{aa} & 0 \\ 0 & \Sigma_{bb} \end{bmatrix}\begin{bmatrix} \xi_a \\ \xi_b \end{bmatrix}/\sigma_{00} = \left\{\xi_a'\Sigma_{aa}\xi_a + \xi_b'\Sigma_{bb}\xi_b\right\}/\sigma_{00} \\
&= \zeta_a'\zeta_a + \zeta_b'\zeta_b,
\end{aligned}$$

where

$$D_n\left(\hat{\alpha} - \alpha\right) \Rightarrow \begin{bmatrix} \xi_a \\ \xi_b \end{bmatrix} = \begin{bmatrix} N\left(0, \sigma_{00}\Sigma_{aa}^{-1}\right) \\ \Sigma_{bb}^{-1}\sum_{t=1}^{\infty} a_tv_tu_{0t} \end{bmatrix},$$

using (17) and setting $\zeta_a := \Sigma_{aa}^{1/2}\xi_a/\sigma_{00}^{1/2} =_d N\left(0, I_m\right)$ and $\zeta_b := \Sigma_{bb}^{1/2}\xi_b/\sigma_{00}^{1/2} = \Sigma_{bb}^{-1/2}\sum_{t=1}^{\infty} a_tv_tu_{0t}/\sigma_{00}^{1/2}$. We deduce that $W_n \Rightarrow \chi_{k_a}^2 + \zeta_b'\zeta_b$, a mixture of a chi square distribution and the squared length of the vector variate ζ_b. No invariance principle holds because ζ_b depends on the data distribution through $\left\{v_t, u_{0t}\right\}_{t=1}^{\infty}$. However, note that when $\left(v_t, u_{0t}\right)$ is Gaussian, then $u_{0t} \sim iid\ N\left(0, \sigma_{00}\right)$ is independent of $\left\{v_t\right\}$ because $E\left(v_tu_{0t}\right) = 0$ in view of Assumption A(ii). Then $\zeta_b =_d N\left(0, I_{k_b}\right)$ and $\zeta_b'\zeta_b \sim_d \chi_{k_b}^2$ so that $W_n \Rightarrow \chi_k^2$.

Proof of Theorem 2

Part (i) We start the analysis by considering the behavior of the sample moment matrix of w_t and the instruments z_t, viz.,

$$W'Z = \begin{bmatrix} \sum_{t=1}^{n} x_t^0z_t' \\ \sum_{t=1}^{n} a_tv_tz_t' \end{bmatrix}.$$

Under Assumption A$'$(i), A$'$(ii), and B(ii) $\sum_{t=1}^{n} a_tv_tz_t' \to_{a.s} \sum_{t=1}^{\infty} a_tv_tz_t'$, which is convergent $a.s.$ because $\sum_{t=1}^{\infty}\left|a_t\right|\mathbb{E}\left\|v_tz_t'\right\| < \infty$. It follows that

$n^{-1/2}\sum_{t=1}^{n} a_t v_t z_t' \to_{a.s} 0$ and then

$$\begin{aligned}
D_n^{-1}W'P_Z W D_n^{-1} &= \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t^0 z_t' \\ \sum_{t=1}^{n} a_t v_t z_t' \end{bmatrix} \left(\sum_{t=1}^{n} z_t z_t'\right)^{-1} \left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t x_t^{0\prime}, \sum_{t=1}^{n} a_t v_t z_t'\right] \\
&= \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} x_t^0 z_t' \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} a_t v_t z_t' \end{bmatrix} \left(\frac{\sum_{t=1}^{n} z_t z_t'}{n}\right)^{-1} \left[\frac{1}{n}\sum_{t=1}^{n} z_t x_t^{0\prime}, \frac{1}{\sqrt{n}}\sum_{t=1}^{n} a_t v_t z_t'\right] \\
&\to_{a.s} \begin{bmatrix} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & 0 \\ 0 & 0 \end{bmatrix},
\end{aligned}$$

which is singular. Applying the martingale CLT (see Lemma A) we have $n^{-1/2}\sum_{t=1}^{n} z_t u_{0t} \Rightarrow N\left(0, \sigma_{00}\Sigma_{zz}\right)$, and by ergodicity $n^{-1}\left[\sum_{t=1}^{n} z_t x_t^{0\prime}, \sum_{t=1}^{n} z_t v_t', \sum_{t=1}^{n} z_t z_t'\right] \to_{a.s.} \begin{bmatrix} \Sigma_{zx} & \Sigma_{zv} & \Sigma_{zz} \end{bmatrix}$, which leads to

$$\begin{aligned}
D_n^{-1}W'P_Z u_0 &= \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} x_t^0 z_t' \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} a_t v_t z_t' \end{bmatrix} \left(\frac{\sum_{t=1}^{n} z_t z_t'}{n}\right)^{-1} \left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t u_{0t}\right] \\
&\Rightarrow \begin{bmatrix} \Sigma_{xz} \\ 0 \end{bmatrix} \Sigma_{zz}^{-1} \times N\left(0, \sigma_{00}\Sigma_{zz}\right) = \begin{bmatrix} N\left(0, \sigma_{00}\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right) \\ 0 \end{bmatrix}.
\end{aligned}$$

Now define $F_n = diag\left(\sqrt{n} I_{k_a}, \frac{1}{\sqrt{n}} I_{k_b}\right)$ and note that

$$\begin{aligned}
F_n^{-1}W'P_Z W F_n^{-1} &= \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} x_t^0 z_t' \\ \sum_{t=1}^{n} a_t v_t z_t' \end{bmatrix} \left(\frac{\sum_{t=1}^{n} z_t z_t'}{n}\right)^{-1} \left[\frac{1}{n}\sum_{t=1}^{n} z_t x_t^{0\prime}, \sum_{t=1}^{n} a_t z_t v_t'\right] \\
&\to_{a.s} \begin{bmatrix} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xz}\Sigma_{zz}^{-1}\left(\sum_{t=1}^{\infty} a_t z_t v_t'\right) \\ \left(\sum_{t=1}^{\infty} a_t v_t z_t'\right)\Sigma_{zz}^{-1}\Sigma_{zx} & \left(\sum_{t=1}^{\infty} a_t v_t z_t'\right)\Sigma_{zz}^{-1}\left(\sum_{t=1}^{\infty} a_t z_t v_t'\right) \end{bmatrix} \\
&=: \begin{bmatrix} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xz}\Sigma_{zz}^{-1}A_z \\ A_z'\Sigma_{zz}^{-1}\Sigma_{zx} & A_z'\Sigma_{zz}^{-1}A_z \end{bmatrix} = M, \qquad (18)
\end{aligned}$$

where $A_z = \sum_{t=1}^{\infty} a_t z_t v_t'$, which is convergent almost surely because $\sum_{t=1}^{\infty} |a_t|\,\mathbb{E}\left\|z_t v_t'\right\| < \infty$ in view of B(ii) and A$'$(ii). Also

$$\begin{aligned}
F_n^{-1}W'P_Z u_0 &= \begin{bmatrix} \frac{1}{n}\sum_{t=1}^{n} x_t^0 z_t' \\ \sum_{t=1}^{n} a_t v_t z_t' \end{bmatrix} \left(\frac{\sum_{t=1}^{n} z_t z_t'}{n}\right)^{-1} \left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t u_{0t}\right] \\
&\Rightarrow \begin{bmatrix} \Sigma_{xz} \\ A_z' \end{bmatrix} \Sigma_{zz}^{-1} \times N\left(0, \sigma_{00}\Sigma_{zz}\right) = MN\left(0, \sigma_{00}M\right).
\end{aligned}$$

Note that the matrix variate $A_z = \sum_{t=1}^{\infty} a_t z_t v_t'$ is independent of the limit of $\left(\frac{1}{n}\sum_{t=1}^{n} z_t z_t'\right)^{-1/2} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t u_{0t} \Rightarrow N\left(0, I_K\right)$, since this Gaussian limit does not depend on $\{z_t, v_t\}_{t=1}^{\infty}$. Hence, we have the mixed normal (MN) limit theory

$$\begin{aligned}
F_n\left(\alpha_{IV} - \alpha\right) &= \left(F_n^{-1}W'P_Z W F_n^{-1}\right)^{-1}\left(F_n^{-1}W'P_Z u_0\right) \\
&\Rightarrow MN\left(0, \sigma_{00}M^{-1}\right). \qquad (19)
\end{aligned}$$

In partitioned form, we have

$$\begin{aligned} F_n\left(\alpha_{IV}-\alpha\right) &= \left\{\sqrt{n}\left(\alpha_{a,IV}-\alpha_a\right), \frac{1}{\sqrt{n}}\left(\alpha_{b,IV}-\alpha_b\right)\right\} \\ &\Rightarrow MN\left(0, \sigma_{00}\left[\begin{array}{cc} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xz}\Sigma_{zz}^{-1}A_z \\ A_z'\Sigma_{zz}^{-1}\Sigma_{zx} & A_z'\Sigma_{zz}^{-1}A_z \end{array}\right]^{-1}\right) \end{aligned} \quad (20)$$

and so $\alpha_{a,IV} \to_p \alpha_a$ but $\alpha_{b,IV}$ diverges at a $\sqrt{n}$ rate. Transforming to the original coordinates, we have

$$\beta_{IV}-\beta = G\left(\alpha_{IV}-\alpha\right) = \left[\begin{array}{cc} I_{k_a} & -\Pi' \\ 0 & I_{k_b} \end{array}\right]\left(\alpha_{IV}-\alpha\right), \quad (21)$$

and then

$$\beta_{IV}-\beta = \left[\begin{array}{c} \beta_{a,IV}-\beta_a \\ \beta_{b,IV}-\beta_b \end{array}\right] = G\left[\begin{array}{c} \alpha_{a,IV}-\alpha_a \\ \alpha_{b,IV}-\alpha_b \end{array}\right] = GF_n^{-1}F_n\left(\alpha_{IV}-\alpha\right),$$

giving

$$F_nG^{-1}\left(\beta_{IV}-\beta\right) = F_n\left(\alpha_{IV}-\alpha\right) \Rightarrow MN\left(0, \sigma_{00}M^{-1}\right). \quad (22)$$

Since $G^{-1} = \left[\begin{array}{cc} I_m & \Pi' \\ 0 & 1 \end{array}\right]$ and $F_n = \operatorname{diag}\left(\sqrt{n}I_{k_a}, \frac{1}{\sqrt{n}}I_{k_b}\right)$, we have the partitioned asymptotics

$$\begin{aligned} & F_nG^{-1}\left(\beta_{IV}-\beta\right) \\ =& \left[\begin{array}{cc} \sqrt{n}I_m & \sqrt{n}\Pi' \\ 0 & 1/\sqrt{n} \end{array}\right]\left[\begin{array}{c} \beta_{a,IV}-\beta_a \\ \beta_{b,IV}-\beta_b \end{array}\right] \\ =& \left[\begin{array}{c} \sqrt{n}\left(\beta_{a,IV}-\beta_a\right)+\sqrt{n}\Pi'\left(\beta_{b,IV}-\beta_b\right) \\ \frac{1}{\sqrt{n}}\left(\beta_{b,IV}-\beta_b\right) \end{array}\right] \\ \Rightarrow& MN\left(0, \sigma_{00}\left[\begin{array}{cc} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xz}\Sigma_{zz}^{-1}A_z \\ A_z'\Sigma_{zz}^{-1}\Sigma_{zx} & A_z'\Sigma_{zz}^{-1}A_z \end{array}\right]^{-1}\right). \end{aligned} \quad (23)$$

Recall that $\alpha = G^{-1}\beta = \left[\begin{array}{c} \beta_a+\Pi'\beta_b \\ \beta_b \end{array}\right] =: \left[\begin{array}{c} \alpha_a \\ \alpha_b \end{array}\right]$ so that $\beta_{a,IV}+\Pi'\beta_{b,IV}$ is consistent for $\beta_a+\Pi'\beta_b$ and $\sqrt{n}\left(\beta_{a,IV}-\beta_a\right)+\sqrt{n}\Pi'\left(\beta_{b,IV}-\beta_b\right)$ has a limiting mixed normal distribution, whereas $\beta_{b,IV}-\beta_b$ diverges at the rate $\sqrt{n}$. More specifically, we have by partitioning the limit covariance matrix in (23) that

$$\sqrt{n}\left(\beta_{a,IV}+\Pi'\beta_{b,IV}-\beta_a-\Pi'\beta_b\right) \Rightarrow MN0, \sigma_{00}H_{\beta\beta}^{-1},$$

where $H_{\beta\beta} = \left[\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}-\Sigma_{xz}\Sigma_{zz}^{-1}A_z\left(A_z'\Sigma_{zz}^{-1}A_z\right)^{-1}A_z'\Sigma_{zz}^{-1}\Sigma_{zx}\right]$.

Part (ii) We next consider the IV error variance estimate $\tilde{\sigma}^2 = \frac{1}{n}\tilde{u}'\tilde{u}$, where $\tilde{u} = y - X\beta_{IV} = y - W\alpha_{IV} = u_0 - W\left(W'P_ZW\right)^{-1}W'P_Zu_0$. The estimate can be expanded as follows

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{1}{n}u_0'u_0 - \frac{2}{n}u_0'P_ZW\left(W'P_ZW\right)^{-1}W'u_0 + \frac{1}{n}u_0'P_ZW\left(W'P_ZW\right)^{-1}W'W\left(W'P_ZW\right)^{-1}W'P_Zu_0 \\ &= \frac{1}{n}u_0'u_0 - \frac{2}{n}u_0'P_ZWF_n^{-1}\left(F_n^{-1}W'P_ZWF_n^{-1}\right)^{-1}F_n^{-1}W'u_0 \\ &\quad + \frac{1}{n}u_0'P_ZWF_n^{-1}\left(F_n^{-1}W'P_ZWF_n^{-1}\right)^{-1}F_n^{-1}W'WF_n^{-1}\left(F_n^{-1}W'P_ZWF_n^{-1}\right)^{-1}F_n^{-1}W'P_Zu_0.\end{aligned}$$

Observe that $n^{-1}u_0'u_0 \to_{a.s.} \sigma_{00}$, $F_n\left(\alpha_{IV} - \alpha\right) \Rightarrow Z_\alpha := MN\left(0, \sigma_{00}M^{-1}\right)$ in view of (19), and

$$F_n^{-1}W'u_0 = \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^n x_t^0 u_{0t} \\ \sqrt{n}\sum_{t=1}^n a_t v_t u_{0t} \end{bmatrix},$$

so that

$$\begin{aligned}&\frac{1}{n}u_0'P_ZW\left(W'P_ZW\right)^{-1}W'u_0 = \frac{1}{n}\left(\alpha_{IV} - \alpha\right)'F_nF_n^{-1}W'u_0 \\ &= \frac{1}{\sqrt{n}}\left(\alpha_{IV} - \alpha\right)'F_n\begin{bmatrix} \frac{1}{n}\sum_{t=1}^n x_t^0 u_{0t} \\ \sum_{t=1}^n a_t v_t u_{0t} \end{bmatrix} = O_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Next, note that

$$\begin{aligned}\frac{1}{n}F_n^{-1}W'WF_n^{-1} &= \frac{1}{n}\begin{bmatrix} \frac{1}{n}\sum_{t=1}^n x_t^0 x_t^{0\prime} & \sum_{t=1}^n a_t x_t^0 v_t' \\ \sum_{t=1}^n a_t v_t x_t^{0\prime} & n\sum_{t=1}^n a_t^2 v_t^0 v_t' \end{bmatrix} = \begin{bmatrix} \frac{1}{n^2}\sum_{t=1}^n x_t^0 x_t^{0\prime} & \frac{1}{n}\sum_{t=1}^n a_t x_t^0 v_t' \\ \frac{1}{n}\sum_{t=1}^n a_t v_t x_t^{0\prime} & \sum_{t=1}^n a_t^2 v_t^0 v_t' \end{bmatrix} \\ &= \begin{bmatrix} O_p\left(n^{-1}\right) & O_p\left(n^{-1}\right) \\ O_p\left(n^{-1}\right) & \sum_{t=1}^n a_t v_t^0 v_t' \end{bmatrix}, \text{ under A}' \text{ and B(ii).}\end{aligned}$$

Alternatively under A′ and B(i), we have

$$\begin{bmatrix} \frac{1}{n^2}\sum_{t=1}^n x_t^0 x_t^{0\prime} & \frac{1}{n}\sum_{t=1}^n a_t x_t^0 v_t' \\ \frac{1}{n}\sum_{t=1}^n a_t v_t x_t^{0\prime} & \sum_{t=1}^n a_t^2 v_t^0 v_t' \end{bmatrix} = \begin{bmatrix} O_p\left(n^{-1}\right) & O_p\left(n^{-1+\eta}\right) \\ O_p\left(n^{-1+\eta}\right) & \sum_{t=1}^n a_t^2 v_t^0 v_t' \end{bmatrix},$$

for some small $\eta > 0$. Using these results we obtain

$$\begin{aligned}&\frac{1}{n}u_0'P_ZWF_n^{-1}\left(F_n^{-1}W'P_ZWF_n^{-1}\right)^{-1}F_n^{-1}W'WF_n^{-1}\left(F_n^{-1}W'P_ZWF_n^{-1}\right)^{-1}F_n^{-1}W'P_Zu_0 \\ &= \left(\alpha_{IV} - \alpha\right)'F_n\left\{\frac{1}{n}F_n^{-1}W'WF_n^{-1}\right\}F_n\left(\alpha_{IV} - \alpha\right) \\ &= \left(\alpha_{IV} - \alpha\right)'F_n\begin{bmatrix} O_p\left(n^{-1}\right) & O_p\left(n^{-1+\eta}\right) \\ O_p\left(n^{-1+\eta}\right) & \sum_{t=1}^n a_t v_t v_t' \end{bmatrix}F_n\left(\alpha_{IV} - \alpha\right) \\ &= \left\{\frac{1}{\sqrt{n}}\left(\alpha_{b,IV} - \alpha_b\right)'\right\}\left(\sum_{t=1}^n a_t^2 v_t v_t'\right)\left\{\frac{1}{\sqrt{n}}\left(\alpha_{b,IV} - \alpha_b\right)\right\} + o_p(1) \\ &\Rightarrow \sigma_{00}\psi_b'H^{-1/2}\left(\sum_{t=1}^\infty a_t^2 v_t v_t'\right)H^{-1/2}\psi_b\end{aligned}$$

where we use the fact that $\frac{1}{\sqrt{n}}(\alpha_{b,IV}-\alpha_b)=\frac{1}{\sqrt{n}}\left(\beta_{b,IV}-\beta_b\right)\Rightarrow MN\left(0,\sigma_{00}H^{-1}\right)=\sigma_{00}^{1/2}H^{-1/2}\psi_b$ with $\psi_b=N(0,I_{k_b})$ and

$$H=\left[A_z'\Sigma_{zz}^{-1}A_z-A_z'\Sigma_{zz}^{-1}\Sigma_{zx}\left(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\right)^{-1}\Sigma_{xz}\Sigma_{zz}^{-1}A_z\right].$$

It follows that

$$\tilde{\sigma}^2=\frac{1}{n}\tilde{u}'\tilde{u}\rightarrow_p\sigma_{00}\left\{1+\omega_{zz}\right\}$$

where $\omega_{zz}=\psi_b'H^{-1/2}\left(\sum_{t=1}^{\infty}a_t^2v_tv_t'\right)H^{-1/2}\psi_b$, as stated.

Part (iii) The block Wald test is

$$\begin{aligned}W_n &= \beta_{IV}'X'P_ZX\beta_{IV}/\tilde{\sigma}^2=\beta_{IV}'G'^{-1}G'X'P_ZXGG^{-1}\beta_{IV}/\tilde{\sigma}^2=\alpha_{IV}'W'P_ZW\alpha_{IV}/\tilde{\sigma}^2\\ &= \alpha_{IV}'F_n\left(F_n^{-1}W'P_ZWF_n^{-1}\right)F_n\alpha_{IV}/\tilde{\sigma}^2.\end{aligned}$$

Under the null hypothesis $\mathbb{H}_0:\beta=\alpha=0$ we have from (22) that $F_n\alpha_{IV}\Rightarrow N\left(0,\sigma_{00}M^{-1}\right)$, and from (18) that $F_n^{-1}W'P_ZWF_n^{-1}\Rightarrow M$. It follows that

$$\left\{F_n^{-1}W'P_ZWF_n^{-1}\right\}^{1/2}F_n\alpha_{IV}/\sigma_{00}^{1/2}\Rightarrow Z\sim N\left(0,I_k\right),$$

so that

$$W_n\Rightarrow\frac{1}{\left\{1+\omega_{zz}\right\}}Z'Z=\chi_k^2/\left\{1+\omega_{zz}\right\},$$

as stated.

Part (iv) The Sargan test for overidentification has the form

$$\begin{aligned}S_n &= \tilde{u}'P_Z\tilde{u}/\tilde{\sigma}^2=u_0'\left\{P_z-P_zX\left(X'P_ZX\right)^{-1}X'P_Z\right\}u_0/\tilde{\sigma}^2\\ &= u_0'\left\{P_z-P_zW\left(W'P_ZW\right)^{-1}W'P_Z\right\}u_0/\tilde{\sigma}^2\\ &= \zeta_n'\left\{I_K-\left(Z'Z\right)^{-1/2}Z'W\left(W'P_ZW\right)^{-1}W'Z\left(Z'Z\right)^{-1/2}\right\}\zeta_n/\tilde{\sigma}^2,\end{aligned}$$

where $\zeta_n=\left(n^{-1}Z'Z\right)^{-1/2}\left(n^{-1/2}Z'u_0\right)\Rightarrow\zeta\equiv MN\left(0,\sigma_{00}I_K\right)\equiv N\left(0,\sigma_{00}I_K\right)$ by the MGCLT in Lemma A. Note that the limit distribution and random vector ζ is independent of (z_t). Use the earlier finding (18) that

$$F_n^{-1}W'P_ZWF_n^{-1}\rightarrow_{a.s}\begin{bmatrix}\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xz}\Sigma_{zz}^{-1}A_z\\ A_z'\Sigma_{zz}^{-1}\Sigma_{zx} & A_z'\Sigma_{zz}^{-1}A_z\end{bmatrix},$$

where $A_z=\sum_{t=1}^{\infty}a_tz_tv_t'$, and $F_n=\mathrm{diag}\left(\sqrt{n}I_m,\frac{1}{\sqrt{n}}\right)$. We further note that

$$\begin{aligned}\left(Z'Z\right)^{-1/2}Z'WF_n^{-1} &= \left(\frac{Z'Z}{n}\right)^{-1/2}\frac{Z'W}{\sqrt{n}}F_n^{-1}\\ &= \left(\frac{Z'Z}{n}\right)^{-1/2}\left[\begin{array}{cc}\sum_{t=1}^n z_tx_t^0 & \sum_{t=1}^n a_tz_tv_t'\end{array}\right]\rightarrow_{a.s}\Sigma_{zz}^{-1/2}\left[\Sigma_{zx},A_z\right],\end{aligned}$$

and, defining $Q = \Sigma_{zz}^{-1/2} \left[\begin{array}{cc} \Sigma_{zx} & A_z \end{array} \right]$, observe that

$$\left[\begin{array}{cc} \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xz}\Sigma_{zz}^{-1}A_z \\ A_z'\Sigma_{zz}^{-1}\Sigma_{zx} & A_z'\Sigma_{zz}^{-1}A_z \end{array} \right] = \left[\begin{array}{c} \Sigma_{xz} \\ A_z' \end{array} \right] \Sigma_{zz}^{-1} \left[\begin{array}{cc} \Sigma_{zx} & A_z \end{array} \right] = Q'Q.$$

We dedice that

$$\begin{aligned} S_n &= \zeta_n' \left\{ I_K - (Z'Z)^{-1/2} Z'W \left(W'P_Z W\right)^{-1} W'Z \left(Z'Z\right)^{-1/2} \right\} \zeta_n / \tilde{\sigma}^2 \\ &= \zeta_n' \left\{ I_K - Q \left(Q'Q\right)^{-1} Q' + o_{a.s.}(1) \right\} \zeta_n / \left\{ \sigma_{00} \left[1 + \omega_{zz}\right] + o_{a.s.}(1) \right\} \\ &\Rightarrow \chi^2_{K-k} / \left\{1 + \omega_{zz}\right\}, \end{aligned}$$

since $P_Q = I_K - Q\left(Q'Q\right)^{-1} Q'$ is symmetric and idempotent of rank $K-k$. Hence, the Sargan overidentification test statistic is distributed in the limit as $\chi^2_{K-k} / \left\{1 + \omega_{zz}\right\}$, as stated.

6 References

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Application.* New York: Academic Press.

Li D., P. C. B. Phillips and J. Gao (2016). "Uniform Consistency of Nonstationary Kernel-Weighted Sample Covariances for Nonparametric Regression", Econometric Theory (forthcoming).

Pagan, A and A. Ullah (1999). *Nonparametric Econonometrics.* Cambridge University Press.

Park, J. Y. and P. C. B. Phillips (1988). "Statistical Inference in Regressions With Integrated Processes: Part 1," *Econometric Theory* 4, 468–497.

Park, J. Y. and P. C. B. Phillips (1989). "Statistical Inference in Regressions With Integrated Processes: Part 2," *Econometric Theory* 5, 95-131.

Park, J.Y. and P.C.B. Phillips (2000). "Nonstationary binary choice," *Econometrica*, 68, 1249-1280.

Phillips, P.C. B. (1988). "Multiple Regression with Integrated Processes," In N. U. Prabhu, (ed), *Statistical Inference Inference from Stochastic Processes, Contemporary Mathematics* 80, 79-106.

Phillips, P. C. B. (1989). "Partially identified econometric models," *Econometric Theory* 5, 181–240.

Phillips, P. C. B. (2007). "Regression with slowly varying regressors and nonlinear trends". *Econometric Theory,* 23, 557-614.

Phillips, P. C. B., D. Li and J. Gao (2014). "Estimating Smooth Structural Changes in Cointegration Models" Cowles Foundation Discussion Paper, Yale University.

Phillips, P. C. B. and T. Magdalinos (2013). "Inconsistent VAR Regression with Common Explosive Roots", *Econometric Theory*, 29, 808-837

Phillips, P. C. B. and D. Sul (2007). "Transition modeling and econometric convergence tests", *Econometrica,* 75, 1771-1855.

Phillips, P. C. B. and D. Sul (2009). "Economic Transition and Growth", *Journal of Applied Econometrics*, 24, 1153-1185.

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

Sargan, J. D. (1983). "Identification and lack of identification". *Econometrica,* 51, 1605-1633.

Sims, C. A., J. H. Stock and M. W. Watson (1990). "Inference in linear time series models with some unit roots," *Econometrica* 58, 113–144.

Staiger, D. and J.H. Stock (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65, 557-586.

Toda, H. and P. C. B. Phillips (1993). "Vector autoregressions and causality, *Econometrica* 61, 1367–1393.

Robinson, P. M. (2012). "Inference on power law spatial trends", *Bernoulli,* 18, 644-677.