Testing the impossible: identifying exclusion restrictions

Jan F. Kiviet*

First version: 29 December 2016
Second version: 6 November 2017
Third version: of 23 October 2018
This version: 2 November 2018
JEL Classifications: *C12, C13, C18, C21, C26, I12, I26*
Keywords: *endogeneity robust inference; exclusion restrictions test; identification analysis; invalid instruments; sensitivity analysis.*

Abstract

Method of moment estimators are generally obtained by adopting orthogonality conditions, in which particular functions in terms of the observed data and unknown parameters are supposed to have zero expectation. For regression models this implies exploiting presumed uncorrelatedness of the model disturbances and identifying instrumental variables. Here, utilizing non-orthogonality conditions is examined for linear cross-section multiple simultaneous regression models. Employing flexible bounds on the correlations between disturbances and regressors one avoids: (i) adoption of often incredible and unverifiable strictly zero correlation assumptions, and (ii) imprecise inference due to possibly weak or invalid instruments. The asymptotic validity of the suggested alternative form of inference is proved and its finite sample accuracy is demonstrated by simulation. It enables to produce inference on coefficient values that within constraints is endogeneity robust. Also a sensitivity analysis of standard least-squares and instrument-based inference is possible, but even tests on exclusion restrictions regarding external instruments. In the standard approach their adoption is unavoidable though non-testable. The practical relevance is illustrated in a few applications borrowed from the textbook literature.

1. Introduction

The standard quasi-experimental approach in applied econometric research requires the adoption of so-called orthogonality conditions. An initial set of such conditions has to be justified on the basis of persuasive common sense or economic-theoretical arguments.

*Professor of Econometrics, Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 NJ Amsterdam, The Netherlands (j.f.kiviet@uva.nl).

Contenders, however, may easily disqualify these arguments as opportunistic subjective beliefs, since these conditions cannot be vindicated by empirical statistical evidence without adopting further non-testable conditions. Although formal testing of any overidentification restrictions is feasible, its unambiguous interpretation is contingent on the validity of the initial just-identifying set of non-testable orthogonality conditions. For the analysis of single regression equations this implies that at least as many excluded variables must be uncorrelated with the unobservable random disturbances of the equation as there are unknown coefficients of the endogenous explanatory variables, which are those that could be correlated with the disturbances. So these excluded variables, also addressed as the external instruments, should have no direct effect on the dependent variable of the equation of interest. Moreover, for yielding effective inference, they should at the same time have a substantial indirect effect through considerable correlation with the endogenous regressors of the relationship.

Here it will be shown that there is an alternative route towards identification by adopting non-orthogonal moment conditions, which may in fact be more credible than the just-identifying orthogonality conditions, because the moments concerned do not have to be strictly zero but may vary over an interval. A simple implementation of this yields a mutation of the ordinary least-squares (OLS) estimator. It is consistent if the correlation between all regressors and the disturbance is known. From the limiting distribution of this infeasible estimator feasible asymptotic test procedures for regression coefficients can be constructed. In addition to some standard orthogonality conditions with respect to any exogenous regressors, these exploit also bounds on the degree of non-orthogonality or simultaneity of the endogenous regressors. In simulations it is demonstrated that even in quite small samples these tests have appropriate level control and impressive power. They can be converted into feasible asymptotic confidence intervals with conservative coverage, which are often more informative than intervals obtained by instrumental variables (IV) methods, especially when some instruments are weak.

Moreover, this robustified OLS-based test can be implemented to verify exclusion restrictions in the following way: It can produce the set of values of simultaneity correlations which endorses the exclusion restrictions, as well as the set under which these should be rejected at a chosen significance level. Thus, depending on the span and the location of these two sets, the credibility of exclusion restrictions may either be supported, wiped away, or remain uncertain.

The methods developed here are based on different assumptions than those made under the standard approach. Fewer assumptions because no external instruments and corresponding exogeneity assumptions are required. On the other hand, some extra assumptions have to be adopted, namely on the fourth moments of the regressors and disturbances which jointly determine the variance of the adapted least-squares estimator. And, when it comes to making decisions on the basis of the produced inference, this has to be confronted with an opinion of the researcher on the likely degree of endogeneity. So, basically, strict exogeneity assumptions on external instruments are exchanged for interval assumptions with respect to the endogeneity of the regressors.

The techniques developed here are a generalization of some basic initial results already published in Kiviet (2013, Section 4) and further justified in Kiviet (2016). Now they allow for an arbitrary number of endogenous and exogenous regressors and for tests on sets of linear restrictions on the coefficients of both endogenous and exogenous regressors. This approach, addressed as kinky least-squares (KLS), was triggered by some rudimentary findings dating back to Goldberger (1964, p.359) and Rothenberg (1972). That even the standard OLS estimator may beat IV or two-stage least-squares (TSLS) in very small samples has been argued already in Kadane (1971). In Kiviet and Niemczyk (2012) and Kiviet (2013) differences between the asymptotic and simulated finite sample distributions of OLS and IV have been mapped with respect to instrument strength and degree of simultaneity. That the here presented generalized KLS procedures are preferable to standard OLS, and in many cases also to IV, is because they produce accurate statistical inference in models with endogenous regressors, while avoiding the hazards of weakness or invalidity of instruments.

For an overview of complicating issues undermining the accuracy of statistical inference in models with endogenous regressors due to employment of weak or invalid external instruments see, for instance, Dufour (2003). In essence these complications are fourfold: (i) under weak though valid instruments standard asymptotic IV inference is inaccurate (poorly approximates seriously biased non-normal coefficient estimates and their standard errors, resulting in bad level control of tests); (ii) employing more sophisticated weak-instrument techniques may result in improved level control, but yields confidence sets which are often very wide or even unbounded; (iii) the use of invalid instruments produces as a rule highly inaccurate inference; (iv) testing the validity of particular instruments seems only possible when a sufficient number of valid instruments is already available. During the last decades (i) and (ii) received a lot of attention in the literature. This study addresses the two more fundamental problems (iii) and (iv). It escapes from problem (iii) by developing a formal frequentist approach to produce accurate inference in simultaneous models not employing any external instrumental variables at all, by incorporating into the analysis an interval assumption on the degree of simultaneity. When implemented as an exclusion restrictions test this approach also allows to break out of the vicious circle of problem (iv) through testing the validity of instruments without requiring any untested orthogonality conditions.

Various other studies have addressed problem (iii). The degree of invalidity of instruments is incorporated into a frequentist analysis by Ashley (2009)[1] and by Bayesian methods in Kraay (2012). Nevo and Rosen (2012) derive set estimates under assumptions on the signs and relative magnitudes of the simultaneity and instrument invalidity. Conley et al. (2012) augment the model with the instruments and make assumptions on its coefficients (which would be zero under correct exclusion) which next allow frequentist or Bayesian methods to obtain inference allowing for instrument invalidity. However, unlike ours, all these approaches still employ IV methods and so do not escape from problem (i) nor (ii). To our knowledge feasible tests for (iv) have not been developed before, apart from various informal procedures, such as suggested in, for instance, Bound and Jaeger (2000) and Altonji et al. (2005).

In Section 2, after having reviewed how in a linear multiple regression equation with some endogenous regressors consistent estimators can be obtained by exploiting classic identifying orthogonality conditions, we demonstrate how this can also be achieved by adopting non-orthogonality conditions. This yields an adapted least-squares estimator which is a function of the nuisance parameter vector containing the correlations between all the regressors and the disturbances. The limiting distribution of this infeasible es-

[1] Kiviet (2016) addresses flaws in the asymptotic derivations in offsprings of this study.

timator is presented in Section 3 (and derived in Appendices). From this a feasible test procedure for a set of general restrictions on the coefficient values readily follows. Section 4 demonstrates how this procedure can be employed for testing any exclusion restrictions relevant within the context of a classic instrumental variables based analysis. Section 5 provides simulation results on size and power of exclusion restriction tests in simultaneous models with one or two endogenous regressors. Section 6 demonstrates how the various techniques can be employed in practice by analyzing three empirical data sets used for illustrative purposes in well-known textbooks. Section 7 concludes.

2. Two distinct approaches towards identification

Consider a sample of n independent and identically distributed (i.i.d.) zero-mean observations $\{y_i, x_i'; i = 1, ..., n\}$ on a linear causal relationship given by

$$y_i = x_i'\beta + u_i, \text{ with } x_i \sim (0, \Sigma_{xx}) \text{ and } u_i \sim (0, \sigma_u^2), \tag{2.1}$$

where β is an unknown constant $K \times 1$ coefficient vector, $K \times K$ matrix Σ_{xx} is positive definite with all its elements finite and $0 < \sigma_u^2 < \infty$. Vector $x_i' = (x_i^{(1)\prime}, x_i^{(2)\prime})$ has $K_1 + K_2 = K$ elements, such that

$$E(x_i^{(1)}u_i) = \sigma_{x^{(1)}u} \text{ and } E(x_i^{(2)}u_i) = 0, \tag{2.2}$$

with $\sigma_{x^{(1)}u} \in \mathbb{R}^{K_1}$ which in practice is generally unknown.

Model (2.1) has zero-mean regressors and no intercept. However, all results to be derived for its slope coefficients will also apply to the more usual model with an arbitrary intercept and with the same disturbances and same K slope coefficients, though corresponding to i.i.d. regressors with an arbitrary observation-constant mean. Taking all observations in deviation from their sample average in this more general model annihilates the intercept and results in a model with zero-mean regressors. Although the observations are no longer independent in this transformed model, we will prove that all results still apply.

A non-zero but constant correlation between elements of the $K_1 \times 1$ vector of regressors x_{1i} and the disturbance u_i may be due either to simultaneity (elements of $x_i^{(1)}$, while constituting causes for y_i, are causally dependent on y_i themselves too), or to measurement errors in $x_i^{(1)}$, or perhaps to particular omissions in the regression specification. Generally, such non-zero correlations render all elements of the OLS estimator $\hat{\beta}_{OLS} = (\Sigma_{i=1}^n x_i x_i')^{-1}\Sigma_{i=1}^n x_i y_i$ biased for β and inconsistent under common regularity conditions. A standard method to achieve consistent estimators is reviewed in the next subsection, followed in a second subsection by the development of an alternative and non-standard procedure which in a sense repairs the inconsistency of least-squares.

2.1. Exploiting orthogonality conditions

The standard approach to achieve identification and consistent estimation of β is to find a $L_2 \times 1$ vector of observations $z_i^{(2)}$ such that, for the $L \times 1$ vector $z_i' = (x_i^{(2)\prime}, z_i^{(2)\prime})$ with $L = K_2 + L_2 \geq K$, one is willing to assume validity of the orthogonality conditions

$$E(z_i u_i) = 0, \ \forall i. \tag{2.3}$$

These imply

$$E(z_i y_i) = E(z_i x_i')\beta, \ \forall i.$$

Next, the "analogy principle" of the method of moments, by which expectations are replaced by corresponding sample averages, suggests as an estimator for β the "best" solution $\hat{\beta}$ of $n^{-1}\Sigma_{i=1}^{n} z_i y_i = (n^{-1}\Sigma_{i=1}^{n} z_i x_i')\,\hat{\beta}$, hence of

$$Z'y = Z'X\hat{\beta},$$

where $X = (x_1, ..., x_n)'$, $y = (y_1, ..., y_n)'$ and $Z = (z_1, ..., z_n)'$. In case $L = K$ and $Z'X$ has full rank, there is a unique solution, namely

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y. \tag{2.4}$$

This realizes $Z'(y - X\hat{\beta}_{IV}) = 0$, thus achieving orthogonality of the residuals $\hat{u}_{IV} = y - X\hat{\beta}_{IV}$ and the instruments in the sample, similar to the zero moments $E(Z'u) = 0$.

If $L > K$, while $X'Z$ has rank K and $Z'Z$ has rank L, then orthogonality of all individual instruments and the residuals cannot be achieved, but a unique solution is found by minimizing a quadratic form in the vector $Z'(y - X\hat{\beta})$, namely $(Z'y - Z'X\hat{\beta})'W(Z'y - Z'X\hat{\beta})$, where W is some symmetric positive definite weighing matrix. This yields the estimator $\hat{\beta}_{WIV} = (X'ZWZ'X)^{-1}X'ZWZ'y$. Under standard regularity conditions $\hat{\beta}_{WIV}$ (like $\hat{\beta}_{IV}$) is consistent and has a limiting normal distribution. When $L > K$ the efficient Generalized Method of Moment (GMM) estimator is obtained by choosing W proportional to $[Z'Var(u)Z]^{-1}$, where $u = (u_1, ..., u_n)'$. Because we have here $Var(u) = \sigma_u^2 I$ this simplifies to the TSLS estimator

$$\hat{\beta}_{TSLS} = (X'P_Z X)^{-1}X'P_Z y, \tag{2.5}$$

where $P_Z = Z(Z'Z)^{-1}Z'$. If $L > K$ it does not realize in the sample orthogonality of Z and $\hat{u}_{TSLS} = y - X\hat{\beta}_{TSLS}$, but it does realize the orthogonality relationships $\hat{X}'\hat{u}_{TSLS} = 0$. Here $\hat{X} = P_Z X$ is the orthogonal projection of the K regressors X on the L dimensional sub-space spanned by the instrumental variables Z.

In the above approach, asymptotic validity of all inference is based on validity of the $L \geq K$ orthogonality conditions (2.3). It seems obvious that in case $L = K$ no statistical evidence can be produced on this validity from the sample under study, simply because then $Z'(y - X\hat{\beta}_{IV})$ equals zero by construction. When $L > K$ then, due to $\hat{X}'\hat{u}_{TSLS} = X'Z(Z'Z)^{-1}Z'\hat{u}_{TSLS} = CZ'\hat{u}_{TSLS} = 0$ with C a $K \times L$ matrix of rank K, K linearly independent combinations of the instruments Z will be orthogonal to $\hat{u}_{TSLS}$ by construction. This leaves room for the Sargan test[2] to verify the orthogonality of the complimentary subspace of Z with dimension $L - K$ by a quadratic form in $Z'\hat{u}_{TSLS}$. Asymptotically this has a χ^2_{L-K} distribution under (2.3). Put differently, orthogonality of the external instruments $z_i^{(2)}$ would require validity of the $L_2 = L - K_2$ zero restrictions $\beta_z^{(2)} = 0$ when model (2.1) is extended with $z_i^{(2)\prime}\beta_z^{(2)}$, because $\beta_z^{(2)} \neq 0$ would jeopardize $E(z_i^{(2)}u_i) = 0$. However, in order to cope with the endogeneity of $x_i^{(1)}$ while adopting exogeneity of $x_i^{(2)}$, testing these zero or exclusion restrictions seems only possible in two cases: (i) Either, K_1 further valid external instruments should be found. However, to test the validity of these extra instruments, yet another K_1 valid instruments should be

[2]See Sargan (1958) and Hansen (1982) for a GMM generalization.

available, and so on. Thus, in the end, this requires the adoption of $K_2+K_1 = K$ untested instruments. (ii) Or, we should directly adopt validity of K_1 of the external instruments in $z_i^{(2)}$ (or K_1 linearly independent linear combinations of them) and test the significance when augmenting the model with just the $L_2 - K_1 = L - K$ remaining elements of $z_i^{(2)}$. In both cases, when the model with $L \geq K$ candidate internal and external instruments forms the starting point, only when K instruments are presupposed to be valid, the validity of $L-K$ exclusion restrictions, establishing $L-K$ overidentification restrictions, can be tested. This is equivalent with testing the validity of $L-K$ instruments in addition to K valid –though untested– instruments by the Sargan test, see Kiviet (2017).

If one is not willing to adopt validity of K instruments then overidentification restrictions tests are mute about instrument validity, as is argued in Deaton (2010) and in Parente and Santos Silva (2012), because significance of an overidentification test can occur when a subset of the instruments is valid and insignificance may occur when all instruments are invalid. In fact, as Deaton (2010, p.431) remarks and Windmeijer (2018) formally proves: "Such tests can tell us whether estimates change when we select different subsets from a set of possible instruments". However, as argued in Kiviet (2017), such tests are much more informative if one is willing to incorporate validity of at least K instruments (internal and external together) in the maintained hypothesis, which should also incorporate adequacy of the model specification. Then the Sargan test is consistent and optimal for validity of the remaining $L - K$ instruments, see Newey (1985). Resistance to plainly adopt validity of K instruments originates directly from the widely and firmly held belief that statistical testing of the validity of the initial just-identifying set of K instruments is simply impossible.[3]

Because the consistent interpretation of the outcome of overidentification or instrument validity tests is conditional on the legitimacy of adopting K zero correlation assumptions, nonexistence of a statistical test for the latter is highly uncomfortable. It embodies the Achilles heel of many applied econometric studies. As yet, this vulnerability can only be concealed by non-statistical often highly speculative rhetoric arguments, which usually provide just meager protection against dissident views. As we shall demonstrate, the formal test developed in Section 4 could settle such disputes.

2.2. Exploiting some non-orthogonality conditions as well

Next we consider an alternative to the standard approach regarding achieving identification. Instead of adopting $L \geq K$ orthogonality conditions $E(z_i u_i) = 0$, which imply zero correlation between each instrument and the disturbance, consider adopting a numerical assumption concerning the K elements of the correlation vector $\rho_{xu} = (\rho_{x_1 u}, ..., \rho_{x_K u})'$, where

$$\rho_{x_j u} = E(x_{ji} u_i)/(\sigma_u \sigma_j), \; j = 1, ..., K,$$

[3] On the impossibility to test the exogeneity of all the instruments, see for instance: Stock and Watson (2015, p.491) on the case of an exactly identified model: "Then it is impossible to develop a statistical test of the hypothesis that the instruments are in fact exogenous. That is, empirical evidence cannot be brought to bear on the question of whether these instruments satisfy the exogeneity restriction. In this case, the only way to assess whether the instruments are exogenous is to draw on expert opinion and your personal knowledge of the empirical problem at hand". And Carter Hill et al. (2012, p.421): "Unfortunately, not every instrument can be tested for validity".

with σ_j^2 equal to the j-th diagonal element of matrix Σ_{xx}. Hence, suppose we replace the assumption $\rho_{zu} = (\rho_{z_1u}, ..., \rho_{z_Lu})' = 0$ by

$$\rho_{xu} = r, \tag{2.6}$$

where $r = (r_1, ..., r_K)'$ with scalar r_j the adopted value of the correlation between x_{ij} and u_i, so $|r_j| < 1$, $\forall j$. This implies adopting the K moment conditions

$$E(x_{ij}u_i) = \sigma_{x_ju} = r_j\sigma_u\sigma_j, \;\; j = 1, ..., K. \tag{2.7}$$

If we are still convinced of the exogeneity of the regressors $x_i^{(2)}$ we could have

$$r_j = 0 \text{ for } j = K_1 + 1, ..., K \text{ and } r_j \neq 0 \text{ otherwise.}$$

Then we adopt K_2 standard orthogonality conditions and K_1 non-orthogonality conditions.

One may object that in practice one generally would not know the true values of the elements of ρ_{xu}, so r will generally differ from ρ_{xu}. Although true, this will turn out to be of moderate concern, because in the analysis to follow r_j will not necessarily be kept fixed, but may vary within the interval $(-1, +1)$. Moreover, in the classic approach the adopted strictly zero values for the L elements of ρ_{zu} may be false too, raising far more serious credibility issues, because this approach does not allow for non-zero correlations between instruments and disturbances.

Using

$$\Sigma_x = diag(\sigma_1, ..., \sigma_K) \tag{2.8}$$

(2.7) implies

$$E(x_iu_i) = \sigma_{xu} = E[x_i(y_i - x_i'\beta)] = E(x_iy_i) - E(x_ix_i')\beta = \sigma_u\Sigma_x r.$$

Invoking again the "analogy principle" this suggests for the method of moments estimator the solution $\hat{\beta}$, where

$$n^{-1}X'y - n^{-1}X'X\hat{\beta} = \sigma_u S_x r.$$

Here S_x is the sample equivalent of Σ_x. The j-th diagonal element of S_x could either be taken as the square root of $n^{-1}\Sigma_{i=1}^n x_{ij}^2$ (since the regressors have zero expectation) or $(n-1)^{-1}\Sigma_{i=1}^n(x_{ij} - \bar{x}_j)^2$ with $\bar{x}_j = n^{-1}\Sigma_{i=1}^n x_{ij}$ (which may be beneficial in small samples where $\bar{x}_j$ may deviate seriously from zero). This yields solution

$$\begin{aligned}\hat{\beta}(r, \sigma_u) &= (X'X)^{-1}X'y - \sigma_u(n^{-1}X'X)^{-1}S_x r \\ &= \hat{\beta}_{OLS} - \sigma_u(n^{-1}X'X)^{-1}S_x r.\end{aligned} \tag{2.9}$$

This estimator involves a correction to the OLS estimator, aiming to correct its inconsistency when $\rho_{xu} \neq 0$.

Estimator $\hat{\beta}(r, \sigma_u)$ is infeasible as long as σ_u has not been replaced by a sample equivalent. Of course, the standard OLS estimator of σ_u^2, which is given by $\hat{\sigma}_{u,OLS}^2 =$

$\hat{u}'_{OLS}\hat{u}_{OLS}/(n-K)$, where $\hat{u}_{OLS}=y-X\hat{\beta}_{OLS}$, is inconsistent, like $\hat{\beta}_{OLS}$, when $\rho_{xu}\neq 0$, since

$$\begin{aligned}\text{plim}\,\hat{\sigma}^2_{u,OLS} &= \text{plim}\,u'[I-X(X'X)^{-1}X']u/n \\ &= \sigma_u^2 - \text{plim}\,n^{-1}u'X(\text{plim}\,n^{-1}X'X)^{-1}\,\text{plim}\,n^{-1}X'u \\ &= \sigma_u^2(1-\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}). \end{aligned} \tag{2.10}$$

Thus, a feasible though r-based estimator, which attempts to correct $\hat{\sigma}^2_{u,OLS}$ for its inconsistency, is

$$\hat{\sigma}^2_u(r)=\hat{\sigma}^2_{u,OLS}/(1-r'S_xS_{xx}^{-1}S_xr), \tag{2.11}$$

where element j,k of S_{xx} equals either $n^{-1}\Sigma_{i=1}^n x_{ij}x_{ik}$ or $(n-1)^{-1}\Sigma_{i=1}^n(x_{ij}-\bar{x}_j)(x_{ik}-\bar{x}_k)$. Now a feasible estimator which aims to correct $\hat{\beta}_{OLS}$ for its inconsistency is

$$\hat{\beta}(r)=\hat{\beta}_{OLS}-\hat{\sigma}_u(r)S_{xx}^{-1}S_xr. \tag{2.12}$$

It is obvious that the correction factor of (2.11) and correction term of (2.12) only really succeed in discarding the OLS estimators from their inconsistency if $r=\rho_{xu}$.

3. KLS inference for multiple regressions

In order to examine the expedience of estimator (2.12) for producing inference on β we shall first examine its limiting distribution under the (unrealistic) assumption that the K values r equal the true correlations ρ_{xu}. Under the assumptions made above, it is found that consistent (though infeasible) least-squares estimator $\hat{\beta}(\rho_{xu})$ has limiting normal distribution with in general a rather involved variance matrix. Its asymptotic variance appears to be affected by the kurtosis of the distributions of u_i and x_i. Substantial simplifications occur when the fourth moments correspond to those of the normal distribution and especially in models with just one endogenous explanatory variable a remarkably neat result emerges.

For the infeasible estimator $\hat{\beta}(\rho_{xu})$, which generalizes for multiple structural regression models the KLS estimator of Kiviet (2013, Section 4), we find the following result (proof in Appendix B) for models with an arbitrary number of endogenous regressors, where all regressors and disturbances are identically distributed and have no excess kurtosis.

Theorem 1: *In zero mean i.i.d. cross-section model (2.1), where* $E(x_iu_i)=\sigma_u\Sigma_x\rho_{xu}$, $E(u_i^4)=3\sigma_u^4$ *and* $E(x_{ij}^4)=3\sigma_j^4$ *for* $j=1,...,K$ *and* $i=1,...,n$, *we find for* $\hat{\beta}(\rho_{xu})=\hat{\beta}_{OLS}-\hat{\sigma}_u(\rho_{xu})S_{xx}^{-1}S_x\rho_{xu}$, *with* $\hat{\sigma}^2_u(\rho_{xu})=\hat{\sigma}^2_{u,OLS}/(1-\rho'_{xu}S_xS_{xx}^{-1}S_x\rho_{xu})$, *the limiting distribution*

$$n^{1/2}[\hat{\beta}(\rho_{xu})-\beta]\xrightarrow{d}\mathcal{N}[0,\sigma_u^2V(\rho_{xu})],$$

with $V(\rho_{xu})=\Sigma_{xx}^{-1}\Theta\Sigma_{xx}^{-1}$, *where*

$$\begin{aligned}\Theta = {}&\Sigma_{xx}-\Sigma_{xx}\mathcal{R}^2-\mathcal{R}^2\Sigma_{xx}-\theta^{-1}(\Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi+\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx}) \\ &-0.5\theta^{-1}(\Phi\mathcal{R}^2+\mathcal{R}^2\Phi)+[\theta^{-1}+\theta^{-2}(0.5-\rho'_{xu}\mathcal{R}\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\mathcal{R}\rho_{xu})]\Phi \\ &+0.5(I+\theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}\Sigma_x^{-1}(\Sigma_{xx}\circ\Sigma_{xx})\Sigma_x^{-1}\mathcal{R}(I+\theta^{-1}\Sigma_{xx}^{-1}\Phi).\end{aligned}$$

Here $\mathcal{R} = \mathrm{diag}(\rho_{xu})$ *is the diagonal matrix with the elements of* ρ_{xu} *on its main diagonal,* $\theta = 1 - \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}$ *and* $\Phi = \Sigma_x\rho_{xu}\rho_{xu}'\Sigma_x$.

The following corollaries readily follow.

Corollary 1.1: *If in the situation of Theorem 1 one has* $K = 1$, *thus* $\Sigma_{xx} = \sigma_x^2$ *and* $\mathcal{R} = \rho_{xu}$ *are scalar, then* $\Theta = \sigma_x^2$ *so that* $V(\rho_{xu}) = \sigma_x^{-2}$ *is invariant with respect to* ρ_{xu}, *and* $n^{1/2}[\hat{\beta}(\rho_{xu}) - \beta] \xrightarrow{d} \mathcal{N}(0, \sigma_u^2/\sigma_x^2)$.

A direct proof of Corollary 1.1 can already be found in Kiviet (2013, Theorem 4.1).

Another interesting special case of Theorem 1 considers the situation where just one regressor (say the first one) is endogenous ($K_1 = 1$) and all further regressors are exogenous. This leads to the following (proof in Appendix C).

Corollary 1.2: *If in the situation of Theorem 1* $\rho_{xu} = (\rho_1, 0, ..., 0)'$ *then, denoting* $\hat{\beta}(\rho_{xu})$ *now as* $\hat{\beta}(\rho_1)$, *we have* $n^{1/2}[\hat{\beta}(\rho_1) - \beta] \xrightarrow{d} \mathcal{N}[0, \sigma_u^2 V(\rho_1)]$ *with*

$$V(\rho_1) = \Sigma_{xx}^{-1} - \theta^{-1}\rho_1^2\{e_1e_1'\Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}e_1e_1' - [1 + \theta^{-1}(1 - \rho_1^2)]\sigma_1^2\Sigma_{xx}^{-1}e_1e_1'\Sigma_{xx}^{-1}\},$$

where $\theta = 1 - \rho_1^2\sigma_1^2\sigma^{11}$ *with* $\sigma^{11} = (\Sigma_{xx}^{-1})_{1,1}$. *Moreover, for the first element of vector* $\hat{\beta}(\rho_1)$, *denoted* $\hat{\beta}_1(\rho_1)$, *this yields*

$$n^{1/2}[\hat{\beta}_1(\rho_1) - \beta_1] \xrightarrow{d} \mathcal{N}(0, \sigma_u^2\sigma^{11}),$$

which is invariant with respect to ρ_1.

Surprisingly, in a model with just one endogenous regressor, the limiting distribution of the KLS estimator of the coefficient of the endogenous regressor is equivalent to that of OLS in case all regressors are exogenous. This is no longer the case when $K_1 > 1$, or under excess kurtosis. Note that all findings are invariant with respect to skewness.

The zero-mean assumption regarding the regressors simplifies the proof of Theorem 1 considerably, but does not restrain the applicability of its findings to more general linear regression models. This follows from the next result (proved in Appendix D).

Theorem 2: *In model* $y_i^* = \beta_0 + x_i^{*\prime}\beta + u_i$ *with* $x_i^* \sim (\mu, \Sigma_{xx})$ *and* $u_i \sim (0, \sigma_u^2)$ *both i.i.d. over the observations, where* $E(x_i^*u_i) = \sigma_u\Sigma_x\rho_{xu}$ *and* $E(x_{ij} - \mu_j)^4 = 3\sigma_j^4$ *for* $j = 1, ..., K$ *and* $i = 1, ..., n$, *the results of Theorem 1 still hold upon taking* $x_i = x_i^* - n^{-1}\Sigma_{i=1}^n x_i^*$ *and* $y_i = y_i^* - n^{-1}\Sigma_{i=1}^n y_i^*$.

To produce feasible inference the result of Theorem 1 can be exploited as follows. Suppose we are interested in testing jointly h ($\leq K$) linear restrictions on the coefficients β of model (2.1), given by $\mathcal{H}_0 : Q\beta = q$, where Q is a known $h \times K$ matrix of rank h and q a $h \times 1$ known vector. Now consider the feasible test statistic

$$W(Q, q, r) = [Q\hat{\beta}(r) - q]'[n^{-1}Q\hat{V}(r)Q']^{-1}[Q\hat{\beta}(r) - q]/\hat{\sigma}_u^2(r), \quad (3.1)$$

where $\hat{V}(r) = S_{xx}^{-1}\hat{\Theta}S_{xx}^{-1}$, with

$$\begin{aligned}\hat{\Theta} = {} & S_{xx} - S_{xx}R^2 - R^2S_{xx} - \hat{\theta}^{-1}(S_{xx}R^2S_{xx}^{-1}\hat{\Phi} + \hat{\Phi}S_{xx}^{-1}R^2S_{xx}) \\ & - 0.5\hat{\theta}^{-1}(\hat{\Phi}R^2 + R^2\hat{\Phi}) + [\hat{\theta}^{-1} + \hat{\theta}^{-2}(0.5 - r'RS_xS_{xx}^{-1}S_xRr)]\hat{\Phi} \\ & + 0.5(I + \hat{\theta}^{-1}\hat{\Phi}S_{xx}^{-1})RS_x^{-1}(S_{xx} \circ S_{xx})S_x^{-1}R(I + \hat{\theta}^{-1}\Sigma_{xx}^{-1}\hat{\Phi}),\end{aligned}$$

and $\hat{\theta} = 1 - r'S_xS_{xx}^{-1}S_xr$, $\hat{\Phi} = S_xrr'S_x$ and $R = \text{diag}(r)$. Under $\mathcal{H}_0$ and the conditions of Theorem 1 we have

$$W(Q, q, \rho_{xu}) \xrightarrow{d} \chi^2(h). \tag{3.2}$$

From (2.10) it follows that $\theta = 1 - \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\rho_{xu} \geq 0$. Obviously, KLS will only make sense in practice when $\hat{\theta} = 1 - r'S_xS_{xx}^{-1}S_xr > 0$ too, so that the corrected estimator for σ_u^2 will always be positive. Now for $r_\gamma = (\gamma', 0')'$ and $\chi^2_{1-\alpha}(h)$ the $(1-\alpha) \times 100\%$ quantile of the chi-squared distribution with h degrees of freedom, the set

$$\mathcal{C}_0(Q, q, \alpha) = \left\{\gamma \in \mathbb{R}^{K_1} : r_\gamma' S_xS_{xx}^{-1}S_xr_\gamma < 1 \text{ and } W(Q, q, r_\gamma) < \chi^2_{1-\alpha}(h)\right\} \tag{3.3}$$

represents all possible values of $\rho_{x^{(1)}u}$ for which $\mathcal{H}_0$ does not have to be rejected at asymptotic significance level α. Likewise, the complement of $\mathcal{C}_0(Q, q, \alpha)$ in $\mathbb{R}^{K_1}$, given by $\mathcal{C}_1(Q, q, \alpha) \equiv \mathbb{R}^{K_1} \backslash \mathcal{C}_0(Q, q, \alpha)$, represents all possible values of $\rho_{x^{(1)}u}$ for which $\mathcal{H}_0$ should be rejected at asymptotic significance level α.

The sets $\mathcal{C}_0(Q, q, \alpha)$ and $\mathcal{C}_1(Q, q, \alpha)$ enable to supplement standard OLS inference on $\mathcal{H}_0$ with an indication of its robustness or sensitivity regarding simultaneity. Suppose that the $K_1 \times 1$ zero vector is in set $\mathcal{C}_0(Q, q, \alpha)$, then this set represents also all non-zero values of $\rho_{x^{(1)}u}$ which corroborate under simultaneity the non-rejection of $\mathcal{H}_0$ established under full exogeneity. The OLS decision not to reject $\mathcal{H}_0$ is robust regarding simultaneity as long as it obeys the restrictions set by $\mathcal{C}_0(Q, q, \alpha)$, whereas for values of $\rho_{x^{(1)}u}$ in $\mathcal{C}_1(Q, q, \alpha)$ $\mathcal{H}_0$ should be rejected. When the zero vector is in set $\mathcal{C}_1(Q, q, \alpha)$, thus standard OLS inference rejects $\mathcal{H}_0$, this decision can be extended under simultaneity represented by all vectors $\rho_{x^{(1)}u}$ in $\mathcal{C}_1(Q, q, \alpha)$, but should be reversed for values of $\rho_{x^{(1)}u}$ in $\mathcal{C}_0(Q, q, \alpha)$. So, the asymptotic validity of standard OLS inference for the case $\rho_{x^{(1)}u} = 0$ has been extended here to asymptotic validity for either $\rho_{x^{(1)}u} \in \mathcal{C}_0(Q, q, \alpha)$ or $\rho_{x^{(1)}u} \in \mathcal{C}_1(Q, q, \alpha)$. This could be labelled constraint robustness.

It is obvious that KLS-based inference as just described could be fully robust with respect to simultaneity only if $\mathcal{C}_0(Q, q, \alpha)$ is either empty or contains all γ such that $r_\gamma' S_xS_{xx}^{-1}S_xr_\gamma < 1$. This seems highly unlikely to ever happen in practice, as becomes clear when we examine the case $K_1 = K = 1$. The Studentized version of test (3.1) is given in equation (2.13) of Kiviet (2016). For any single hypothesis test for which the OLS test statistic is found to be t_{OLS}, it specializes set (3.3) to the region

$$\mathcal{C}_0^1(\alpha) = \{\gamma \in \mathbb{R}^1 : |\gamma| < 1 \text{ and } z_{\alpha/2} < (1-\gamma^2)^{1/2}t_{OLS} - n^{1/2}\gamma < z_{1-\alpha/2}\}. \tag{3.4}$$

Since the asymptotic critical values are finite for $0 < \alpha < 1$, with $z_{\alpha/2} < 0 < z_{1-\alpha/2}$, it is obvious that, given n and for any real value of t_{OLS}, as a rule values for γ can be found that belong to $\mathcal{C}_0^1(\alpha)$ and other values which do not belong to $\mathcal{C}_0^1(\alpha)$.

When the tested restriction just concerns the coefficient of a single endogenous regressor it is actually quite simple to produce KLS inference, as already exposed in Kiviet (2013, 2016). However, when K_1 is (much) larger than 1, the actual numerical assessment and representation of the above sets regarding $\rho_{x^{(1)}u}$ may seem quite complicated in practice. In the illustrations of Section 6 we will demonstrate how one can deal with general restrictions ($h \leq K$) in models with $K_1 \leq 2$ and discuss options for dealing with $K_1 > 2$.

4. Testing exclusion restrictions

One of the paradigms of classic econometric theory is that in a just identified ($L = K$) model (2.1) the $K_1 = L_2$ exclusion restrictions cannot be tested, and that in overidentified models ($L_2 - K_1 > 0$) one cannot test all L_2 exclusion restrictions but just $L_2 - K_1 = L - K$ (the degree of overidentification) of them. Hence, always K_1 exclusion restrictions are non-testable.

By the methodology exposed above, however, it is possible to a certain extent to test any subset of L_2 exclusion restrictions. It enables to assess, at a chosen nominal significance level, the set of all possible ρ_{xu} values for which any arbitrary subset of exclusion restrictions should be rejected. If this set seems to cover the area in which the true value of ρ_{xu} may reside, then one should reject validity of the variables associated with these exclusion restrictions as instruments. On the other hand, when it seems likely that the true value of ρ_{xu} will not be in the assessed set (or when this set is empty) rejection of validity of the instruments under test is not indicated (which, of course, does not directly imply their validity). Hence, at the stage of deciding whether or not the true value of ρ_{xu} seems covered by a particular non-empty set, expert knowledge is required to decide on the invalidity of instruments, as in the case regarding adopting $\rho_{zu} = 0$. However, the assessed set regarding ρ_{xu} may turn out to be so wide (or so narrow) that the decision becomes relatively easy. By calculating P-values of relevant $W(Q, q, r)$ tests we will show in the illustrations below how evidence on the (in)validity of external instruments can be produced which in many cases may be much more convincing than evidence just based on pure rhetorical arguments.

Before we consider cases where $K_1 \geq 1$ and $L \geq K$ we will first work out in detail the test for just-identifying exclusion restrictions for the model introduced in section 2.1 focussing on the special case $K_1 = 1$, whereas $L = K$. Hence, the structural multiple regression model is just identified, x_{i1} is the one and only endogenous regressor, but the question is whether the single external scalar variable z_{i2} is exogenous indeed and thus can be used as an instrument next to the $K - 1$ regressor variables $x_i^{(2)}$, which are maintained to be exogenous. The assumption $E(z_{i2}u_i) = 0$ is untestable by the Sargan-Hansen approach, because this requires $L > K$.

Attempting to test $E(z_{i2}u_i) = 0$ could be done by including z_{i2} in the regression and test by an appropriate method whether its coefficient is significant. Its insignificance would endorse (although certainly not guarantee) its valid exclusion from the regression and use as an external instrument. To test by the established methods in model[4]

$$y_i = x_{i1}\beta_1 + x_i^{(2)\prime}\beta_2 + z_{i2}\beta_z + u_i, \tag{4.1}$$

the exclusion hypothesis $\mathcal{H}_e : \beta_z = 0$, while respecting at the same time the simultaneity of regressor x_{i1}, would require yet another valid external instrument, which would bring their number to $L + 1 = K + 1$, whereas we assumed that, apart from the $K - 1$ exogenous regressors $x_i^{(2)}$, the only further candidate instrument is z_{i2}. So, testing the exclusion restriction $\beta_z = 0$ seems impossible indeed.

Though, in the present situation Corollary 1.2 applies, after translating it from model (2.1) into the context of augmented model (4.1). The latter we will denote as $y = X^*\beta^* +$

[4] In correspondence with our set-up in Section 2, we may assume that the arbitrary intercept in the relationship has been partialled out, from y_i, x_{i1}, $x_i^{(2)}$ and z_{i2}, by taking these sample values in deviation from their sample mean.

u, where $X^* = (X, z_2)$ and $\beta^* = (\beta', \beta_z)'$. With $1\times(K+1)$ matrix $Q = (0, ..., 0, 1) = e'_{K+1}$ and $q = 0$, and r_1 still indicating the assumed value for ρ_{x_1u}, we may use for this single hypothesis the test statistic

$$t_{K+1}(r_1) = \hat{\beta}_z(r_1)/[n^{-1}\hat{\sigma}_u^{*2}(r_1)e'_{K+1}\hat{V}^*(r_1)e_{K+1}]^{1/2}. \quad (4.2)$$

Here

$$\hat{\beta}_z(r_1) = e'_{K+1}[\hat{\beta}^* - \hat{\sigma}_u^*(r_1)S^{-1}_{x^*x^*}e_1s_1r_1] = \hat{\beta}_z - \hat{\sigma}_u^*(r_1)s_1r_1(e'_{K+1}S^{-1}_{x^*x^*}e_1), \quad (4.3)$$

$$e'_{K+1}\hat{V}^*(r_1)e_{K+1} = e'_{K+1}S^{-1}_{x^*x^*}e_{K+1}+r_1^2[1+(1-r_1^2)/\hat{\theta}^*(r_1)]s_1^2(e'_1S^{-1}_{x^*x^*}e_{K+1})^2/\hat{\theta}^*(r_1), \quad (4.4)$$

where $S^{-1}_{x^*x^*} = (n^{-1}X^{*\prime}X^*)^{-1}$, e_1 is now the unit vector with $K + 1$ elements, and

$$\hat{\beta}_z = [z'_2(I - P_X)z_2]^{-1}z'_2(I - P_X)y, \quad (4.5)$$

$$\hat{\theta}^*(r_1) = 1 - r_1^2s_1^2(e'_1S^{-1}_{x^*x^*}e_1), \quad (4.6)$$

$$\hat{\sigma}_u^{*2}(r_1) = \tilde{n}^{-1}y'(I - P_{X^*})y/\hat{\theta}^*(r_1). \quad (4.7)$$

For $\tilde{n}$ one may simply take n, or if one wants to employ a small sample adjustment it could be taken $n - K - 1$ or, when all variables have been taken in deviation from their sample average, $n - K - 2$. To achieve $\hat{\theta}^*(r_1) > 1$ we should not vary r_1 over the whole $(-1, +1)$ interval, but just examine $r_1^2 < (e'_1S^{-1}_{x^*x^*}e_1)/s_1^2$.

Extending the assumptions of Theorem 1 to model (4.1) and evaluating (4.2) in ρ_1, we have $t_{K+1}(\rho_1) \xrightarrow{d} \mathcal{N}(0, 1)$ under $\mathcal{H}_e$. Hence, if ρ_1 were known an asymptotically exact test would be available. If ρ_1 is unknown, and when testing two-sided, we should seek the set of all r_1 values for which $[t_{K+1}(r_1)]^2 > \chi^2_{1-\alpha}(1)$ or

$$[\hat{\beta}_z(r_1)]^2 - [n^{-1}\hat{\sigma}_u^{*2}(r_1)e'_{K+1}\hat{V}^*(r_1)e_{K+1}] \times \chi^2_{1-\alpha}(1) > 0. \quad (4.8)$$

The left hand side of this inequality is non-linear in the scalar r_1. Assuming that finding the roots of $[\hat{\beta}_z(r_1)]^2 - [n^{-1}\hat{\sigma}_u^{*2}(r_1)e'_{K+1}\hat{V}^*(r_1)e_{K+1}] \times \chi^2_{1-\alpha}(1) = 0$ while $r_1^2 < (e'_1S^{-1}_{x^*x^*}e_1)/s_1^2$ is feasible, finding the set of r_1 values for which inequality (4.8) holds will be feasible too. Under the assumption that the true value of ρ_1 is contained in this set, the hypothesis $\rho_{z^{(2)}u} = 0$ should be rejected. This procedure has an asymptotic significance level not exceeding α. Small sample performance may improve upon replacing $\chi^2_{1-\alpha}(1)$ by $F_{1-\alpha}(1, \tilde{n})$.

Instead of finding the roots of (4.8) at a particular α a much easier and more informative approach is to construct a graph over all relevant values of r_1, satisfying $r_1^2 < (e'_1S^{-1}_{x^*x^*}e_1)/s_1^2$, of the P-values of $t^2_{K+1}(r_1)$ with respect to the $F(1, \tilde{n})$ distribution. For any α this immediately shows the range of values for ρ_{x_1u} where the test statistic rejects the exclusion restriction (or not).

In models where $K_1 \geq 1$ and $L \geq K$, for any $L_2^* \times 1$ subset z_i^* from $L_2 \times 1$ vector $z_i^{(2)}$ its valid exclusion from model (2.1) can be tested in a similar way. This involves a special implementation of test (3.1). Now let $X^* = (X, Z^*)$ and $\beta^* = (\beta', \beta'_{z^*})'$ with $\hat{\beta}_{z^*} = (Z^{*\prime}M_XZ^*)^{-1}Z^{*\prime}M_Xy$ whereas $Q^* = (O, I_{L_2^*})$. Consider test statistic

$$W^*(Q^*, r) = \hat{\beta}'_{z^*}[n^{-1}Q^*\hat{V}^*(r)Q^{*\prime}]^{-1}\hat{\beta}_{z^*}/[L_2^* \times \hat{\sigma}_u^2(r)], \quad (4.9)$$

where $\hat{V}^*(r) = S^{-1}_{x^*x^*}\hat{\Theta}^* S^{-1}_{x^*x^*}$, with $\hat{\Theta}^*$ the appropriate adaptation of $\hat{\Theta}$ below (3.1) to the present extended model. When $E(z_i^* u_i) = 0$ then $L_2^* \times W^*(Q^*, \rho_{xu}) \xrightarrow{d} \chi^2(L_2^*)$. Calculating the P-value of $W^*(Q^*, r)$ with respect to the $F(L_2^*, \tilde{n})$ distribution over all relevant values r indicates for which values ρ_{xu} validity of the instruments z_i^* seems (un)likely. When $K_1 = 2$ these can be visualized easily by a contour plot.

Note that the procedure just sketched is not an alternative to the Sargan test for overidentifying restrictions, which presupposes next to exogeneity of some regressors (the internal instruments) validity of a number of external instruments equal to the number of endogenous regressors in the model. The procedure discussed here can be implemented such that it produces inference on the validity of any subset of external instruments. Also for the set on which standard (and incremental or difference) overidentifying restrictions tests do build without any prior statistical verification.

5. Simulation results

In this section we want to produce simulation evidence on the finite sample behavior of the inference techniques suggested in this study. As always such a study is only feasible when one strongly restrains the number of parameters of the simulation design. For practical reasons one also has to constrain the grid of parameter values from the design parameter space for which the techniques are actually examined. Therefore simulated models do not often fully mimic all aspects of empirically relevant models, but just their major basic characteristics. However, sometimes one can prove that the phenomenon of interest is in fact invariant with respect to particular parameters, which implies that relatively few calculations for a discrete choice of parameter values can represent the relevant properties for a whole subspace of the design parameter space. The model introduced in Section 2 is primarily characterized by the values n, K_1, K_2 and L_2, where $L_2 \geq K_1 \geq 1$, $L_1 = K_2 \geq 0$ and $n \gg L_1 + L_2 \geq K_1 + K_2$, and also by β, Σ_{xx}, Σ_{zz}, Σ_{zx}, σ_u^2, ρ_{xu} and ρ_{zu}. In Kiviet (2013) very favorable results have been produced on the finite sample accuracy of KLS inference on β in the very simple case $L_2 = K_1 = 1$ with $L_1 = K_2 = 0$ when ρ_{xu} were known. But, also for the more realistic situation where ρ_{xu} is unknown and an (in)correct interval $[\rho_{xu}^L, \rho_{xu}^U]$ is adopted which is supposed to contain the true value, it is shown that KLS inference is often much more useful than standard or Anderson-Rubin instrument-based inference. Because one may suppose that the simulated data in these experiments have been obtained after partialling out any exogenous regressors, the results are invariant regarding the chosen value for K_2, so they represent the situation for any $K \geq K_1 = 1$. It has been shown also that the actually chosen values for β, Σ_{xx}, Σ_{zz} and σ_u^2 do not affect the KLS based test.

Below, in the first subsection, we consider the same simulation design as used in Kiviet (2013), but will examine now the finite sample behavior of the exclusion restriction test under situations where ρ_{xu} is either known or unknown. Next we will present simulation results on KLS inference regarding β and exclusion restrictions tests for models where $K \geq K_1 = L_2 = 2$. All presented results are based on 250,000 replications.

5.1. The simplest possible implementation of the exclusion restriction test

For the very simple model with $K = K_1 = L = L_2 = 1$ we will examine here some of the small sample qualities of test (4.2) on a single just-identifying exclusion restriction. Due to invariance, the results will also represent cases where $L_1 = K_2 > 0$ and $K = L > 1$. The Monte Carlo design is constructed as follows. Let ε_i, ξ_i and ζ_i be three mutually independent series ($i = 1, ..., n$) of standard normal drawings. From these we construct the three series

$$u_i = \sigma_u \varepsilon_i \sim \mathcal{N}(0, \sigma_u^2), \tag{5.1}$$

$$x_i = \sigma_x[(1 - \rho_{xu}^2)^{1/2}\xi_i + \rho_{xu}\varepsilon_i] \sim \mathcal{N}(0, \sigma_x^2), \tag{5.2}$$

$$z_i = \sigma_z(\rho_{z\zeta}\zeta_i + \rho_{z\xi}\xi_i + \rho_{zu}\varepsilon_i) \sim \mathcal{N}(0, \sigma_z^2), \tag{5.3}$$

where all ρ coefficients do not exceed 1 in absolute value; moreover,

$$\rho_{z\zeta}^2 + \rho_{z\xi}^2 + \rho_{zu}^2 = 1. \tag{5.4}$$

Obviously, $\sigma_{xu} = \rho_{xu}\sigma_x\sigma_u$, $\sigma_{zu} = \rho_{zu}\sigma_z\sigma_u$ and $\sigma_{zx} = \sigma_z\sigma_x[\rho_{z\xi}(1 - \rho_{xu}^2)^{1/2} + \rho_{zu}\rho_{xu}]$, hence $\rho_{zx} = \rho_{z\xi}(1 - \rho_{xu}^2)^{1/2} + \rho_{zu}\rho_{xu}$, which yields

$$\rho_{z\xi} = (\rho_{zx} - \rho_{zu}\rho_{xu})(1 - \rho_{xu}^2)^{-1/2}, \tag{5.5}$$

for $\rho_{xu}^2 < 1$. From (5.4) we also have

$$\rho_{z\zeta} = (1 - \rho_{z\xi}^2 - \rho_{zu}^2)^{1/2}. \tag{5.6}$$

Hence, when values for $\sigma_u > 0$, $\sigma_x > 0$, $\sigma_z > 0$, $|\rho_{xu}| < 1$, $|\rho_{zx}| \leq 1$ and $|\rho_{zu}| \leq 1$ are chosen, we can generate the series for u_i and x_i and find matching values for $\rho_{z\xi}$ from (5.5) and for $\rho_{z\zeta}$ from (5.6) so that series z_i can be generated as well. However, the three chosen correlations should obey

$$(\rho_{zx} - \rho_{zu}\rho_{xu})^2 \leq (1 - \rho_{xu}^2)(1 - \rho_{zu}^2), \tag{5.7}$$

in order to ensure that $0 \leq \rho_{z\xi}^2 \leq 1$ and $0 \leq \rho_{z\zeta}^2 \leq 1$.

For each realization of the series u_i, x_i and z_i in the simulation replications, we may first subtract their respective sample average from each observation. In that way an arbitrary intercept of an underlying model with one further regressor and one external potential instrument (each distributed with a possibly non-zero arbitrary mean) has been partialled out.

The dependent variable is generated by the model

$$y_i = x_i\beta + z_i\beta_z + u_i, \tag{5.8}$$

where coefficient β_z has true value zero. Its standard least-squares estimator (4.5) simplifies to

$$\hat{\beta}_z = \frac{z'u - z'x(x'x)^{-1}x'u}{z'z - z'x(x'x)^{-1}z'x} = \frac{r_{zu} - r_{zx}r_{xu}}{1 - r_{zx}^2}\frac{s_u}{s_z}, \tag{5.9}$$

where we define the sample statistics as $r_{z'u} = z'u/(z'zu'u)^{1/2}$, $s_u =| \sqrt{s_u^2} |$ with $s_u^2 = u'u/n_1$ (and similar for r_{zx}, r_{xu}, s_x and s_z), where n_1 is either $n - 1$ or n, depending on whether deviations from sample average have been taken or not.

For this special model we further have from (4.6), taking r_1 as appraisal of $\rho_1 = \rho_{xu}$,

$$\hat{\theta}^*(r_1) = 1 - \frac{r_1^2 s_x^2 s_z^2}{s_x^2 s_z^2 (1 - r_{zx}^2)} = \frac{1 - r_1^2 - r_{zx}^2}{1 - r_{zx}^2}, \tag{5.10}$$

and, because $y'(I - P_{X^*})y \geq 0$ specializes here to $u'u[1 - (r_{xu}^2 - 2r_{xu}r_{zx}r_{zu} + r_{zu}^2)/(1 - r_{zx}^2)]$, we find for (4.7) the expression

$$\hat{\sigma}_u^{*2}(r_1) = \frac{u'u}{\tilde{n}} \frac{1 - r_{zx}^2 - r_{zu}^2 + 2r_{zx}r_{zu}r_1 - r_1^2}{1 - r_1^2 - r_{zx}^2}, \tag{5.11}$$

which will be positive (as a variance estimate should) provided $r_1^2 + r_{zx}^2 < 1$ or $\hat{\theta}^*(r_1) > 0$. Clearly, practical problems may emerge in cases where r_1 is chosen large in absolute value and r_{zx}^2 happens to be larger than ρ_{zx}^2. In the simulations we will monitor the occurrence of $\hat{\theta}^*(r_1) \leq 0$ (which may be frequent, especially when n is small and the variance of r_{zx} large) but will skip such replications, because $\hat{\beta}_z(r_1)$ is only defined when $\hat{\theta}^*(r_1) > 0$. In this simple model it specializes to

$$\begin{aligned}
\hat{\beta}_z(r_1) &= \hat{\beta}_z + \hat{\sigma}_u^*(r_1) \frac{r_1 r_{zx}}{1 - r_{zx}^2} \frac{1}{s_z} \\
&= \frac{r_{zu} - r_{zx} r_1 \left[1 - \left(\frac{n_1(1 - r_{zx}^2 - r_{zu}^2 + 2r_{zx}r_{zu}r_1 - r_1^2)}{\tilde{n}(1 - r_1^2 - r_{zx}^2)} \right)^{1/2} \right]}{1 - r_{zx}^2} \frac{s_u}{s_z}.
\end{aligned} \tag{5.12}$$

For its estimated asymptotic variance, assuming $r_1 = \rho_1$, we find

$$\frac{n_1}{\tilde{n} n} \frac{1 - r_{zx}^2 - r_{zu}^2 + 2r_{zx}r_{zu}r_1 - r_1^2}{(1 - r_{zx}^2)^2 \hat{\theta}^*(r_1)} \left[1 + r_1^2 r_{zx}^2 [1 + (1 - r_1^2)/\hat{\theta}^*(r_1)]/\hat{\theta}^*(r_1) \right] \frac{s_u^2}{s_z^2}. \tag{5.13}$$

From the expressions (5.12), (5.13) and (5.10) we observe that in this special model both $\hat{\beta}_z(r_1)$ and its asymptotic standard error are invariant with respect to β and to s_x, whereas both are a multiple of s_u/s_z. Hence, in this simple model the exclusion restriction test statistic (4.2) will be invariant to β, σ_u^2, σ_x^2 and σ_z^2. Therefore, without loss of generality, we may set in the simulation: $\beta = 0$ and $\sigma_u = \sigma_x = \sigma_z = 1$. Another invariance result is the following. If $r_1 = \rho_{xu}$ and from two of the three correlations ρ_{xu}, ρ_{zx} and ρ_{zu} we change their sign, then the square of the exclusion test statistic does not change. Hence, considering in the simulations only cases where these three correlations are nonnegative (as we will) is not as restrictive as it seems at first sight.

We also find

$$\text{plim}\, \hat{\beta}_z(r_1) = \frac{\rho_{zu} - \rho_{zx} r_1 \left[1 - \left(1 + \rho_{zu} \frac{2\rho_{zx} r_1 - \rho_{zu}}{1 - r_1^2 - \rho_{zx}^2} \right)^{1/2} \right]}{1 - \rho_{zx}^2} \frac{\sigma_u}{\sigma_z},$$

which is zero when $\rho_{zu} = 0$. Because it seems to be mostly non-zero for $\rho_{zu} \neq 0$, we are hopeful that a test based on KLS estimator $\hat{\beta}_z(r_1)$ may have power for testing the invalidity of instrument z_i for the regression of y_i on x_i. From (5.9) we find that the pseudo-true-value of $\hat{\beta}_z$ is

$$\text{plim}\, \hat{\beta}_z = \frac{\rho_{zu} - \rho_1 \rho_{zx}}{1 - \rho_{zx}^2} \frac{\sigma_u}{\sigma_z},$$

which is non-zero in general, unless $\rho_{zu} = \rho_1 \rho_{zx}$. Hence, even when $\rho_{zu} = 0$ it may be non-zero, unless also $\rho_1 = 0$ or $\rho_{zx} = 0$. So obviously, the exclusion restriction should not be tested on the basis of the standard OLS estimator $\hat{\beta}_z$.

Table 1: Rejection frequency (in %) of the infeasible exclusion restriction test ($n = 500; \alpha = 0.05$)

	$\rho_{zu} = 0$			$\rho_{zu} = 0.05$			$\rho_{zu} = 0.1$			$\rho_{zu} = 0.2$		
ρ_{xu}	ρ_{zx}			ρ_{zx}			ρ_{zx}			ρ_{zx}		
	0.0	0.4	0.8	0.0	0.4	0.8	0.0	0.4	0.8	0.0	0.4	0.8
0.0	5.09	5.09	5.09	20.1	23.0	46.2	61.0	68.8	96.3	99.4	99.9	100
0.2	5.09	5.07	5.01	20.1	22.7	43.9	61.0	67.9	95.1	99.4	99.8	100
0.4	5.05	4.99	4.68	20.1	22.1	37.2	61.1	66.8	92.7	99.5	99.8	100
0.6	5.02	4.86	100#	20.0	20.4	0.15#	61.2	64.5	0.00#	99.5	99.7	6.09#
0.8	4.84	4.14	-	19.9	10.7	-	61.7	53.4	-	99.6	99.5	-

In Table 1 ($n = 500$) and Table 2 ($n = 50$) we report the rejection frequency of the two-sided infeasible test based on the square of statistic (4.2), where we substituted the true value of ρ_1 for r_1. Since we took the generated data series in deviation from their sample mean we used $n_1 = n - 1$ and, employing the 5% nominal critical value of the F-distribution, we took it at 1 and $n - 3$ degrees of freedom.

In the block of results for $\rho_{zu} = 0$ we observe in Table 1 that the asymptotic test using the true value of ρ_1 demonstrates very good size control[5] when $n = 500$. According to inequality (5.7) the model is not defined for cases where ρ_{zu} is moderate and both ρ_{xu} and ρ_{zx} are large in absolute value (indicated by "-" in the tables). For cases where $\rho_{xu}^2 + \rho_{zx}^2$ is close to unity we observe deterioration of the performance of the test. Settings for which some experiments did produce negative $\hat{\theta}^*(\rho_1)$ realizations are indicated by a hashtag. Such realizations of the variables, where one would not apply KLS in practice, occurred for those cases in about 50% of the replications. At this large sample size the power of the test is already remarkable for $\rho_{zu} = 0.05$, impressive for $\rho_{zu} = 0.1$ and outright splendid for $\rho_{zu} \geq 0.2$. Apart from close to the non existence region, the rejection probability is found to be almost invariant with respect to the degree of simultaneity ρ_{xu}. For $\rho_{zu} \neq 0$ the rejection probability increases with the absolute value of ρ_{zx}, but is also very good for $\rho_{zx} = 0$. So the KLS exclusion restriction test does not suffer in any way from weak instrument problems.

Table 2: Rejection frequency (in %) of the infeasible exclusion restriction test ($n = 50; \alpha = 0.05$)

	$\rho_{zu} = 0$			$\rho_{zu} = 0.1$			$\rho_{zu} = 0.2$			$\rho_{zu} = 0.3$		
ρ_{xu}	ρ_{zx}			ρ_{zx}			ρ_{zx}			ρ_{zx}		
	0.0	0.4	0.8	0.0	0.4	0.8	0.0	0.4	0.8	0.0	0.4	0.8
0.0	5.16	5.16	5.16	10.8	11.8	21.3	28.8	33.4	66.7	56.9	65.0	96.6
0.2	5.12	5.08	4.86	10.8	11.4	17.7	28.8	31.8	57.4	57.1	62.3	91.7
0.4	4.97	4.80	4.20*	10.6	10.2	1.98*	28.9	29.2	22.5*	57.5	58.9	67.1*
0.6	4.47	3.75	100#	10.2	5.54	1.54#	29.0	21.1	0.02#	58.8	50.6	0.03#
0.8	2.17	5.56*	-	7.87	0.64*	-	29.0	0.23*	-	63.7	4.41*	5.11#

Table 2 presents similar findings for sample size $n = 50$. Note that all results marked by an asterisk or hashtag have been obtained from fewer than 250,000 replications, because those where $\hat{\theta}^*(\rho_1)$ turned out to be non-positive have been skipped (this occurred with frequency less than 5% for cases indicated by an asterisk and over 50% for cases indicated by a hashtag). Because in smaller samples r_{zx} may deviate much more from ρ_{zx} we note deterioration of the test qualities over a larger band of cases approaching

[5] Given the number of replications used a probability of 5% will be estimated here with an error that could exceed ± 0.1 with a probability of about 2%.

the non existence area. Otherwise, however, the size properties of the test are still appropriate and power improves with the absolute value of ρ_{zu}, but self-evidently not as sharply as for larger samples.

In practice r_1 will usually deviate from ρ_1. Therefore, as in Kiviet (2013) for inference on β, we will now examine the merits of a feasible exclusion restriction test in this simple model when employed on the basis of an interval $[r_1^L, r_1^U]$ which is supposed to contain ρ_1. We investigate the three cases $r_1^L = \rho_1 - 0.1$, $r_1^U = \rho_1 + 0.1$; $r_1^L = \rho_1 - 0.2$, $r_1^U = \rho_1 + 0.2$; and $r_1^L = 0$, $r_1^U = 0.3$. From Table 3 where $n = 100$ we see that when $\rho_{xu} \in [r_1^L, r_1^U]$ the test is undersized, and still has remarkable power away from the non existence region. The bottom two rows, where $\rho_{xu} \notin [r_1^L, r_1^U]$, show that the test can be either conservative or liberal. Far away from the non existence region the test may still help to produce useful inference on instrument (in)validity, but otherwise its results become uninterpretable. In this table the asterisk stands for a frequency to obtain an undefined test not exceeding 5% and a hashtag for a frequency exceeding 45%.

Table 3: Rejection frequency (in %) of the feasible exclusion restriction test ($n = 100; \alpha = 0.05$)

			$\rho_{zu} = 0$			$\rho_{zu} = 0.2$			$\rho_{zu} = 0.4$		
ρ_{xu}	r_{xu}^L	r_{xu}^U	ρ_{zx}			ρ_{zx}			ρ_{zx}		
			0.0	0.4	0.8	0.0	0.4	0.8	0.0	0.4	0.8
0.0	-0.1	0.1	4.14	1.72	0.10	48.1	41.5	53.4	98.3	98.5	100
0.2	0.1	0.3	4.10	1.63	0.09	47.8	41.5	49.6	98.2	98.4	100
0.4	0.3	0.5	3.99	1.39	0.05	46.6	39.8	21.1*	97.8	98.1	83.7*
0.6	0.5	0.7	3.66	0.80	100#	43.5	43.6	0.00*	96.9	97.7	0.02#
0.0	-0.2	0.2	3.19	0.50	0.00	43.7	24.7	9.30	97.7	95.0	99.4
0.2	0.0	0.4	3.17	0.44	0.00	42.9	26.2	9.88	97.3	95.7	98.9
0.4	0.2	0.6	3.00	0.31	0.00#	40.1	25.5	0.10*	95.7	95.9	21.5#
0.6	0.4	0.8	2.48	0.06*	100#	32.2	15.6*	0.00*	90.0	88.1*	0.00#
0.0	0.0	0.3	3.66	2.60	3.04	44.8	59.1	92.8	97.6	99.6	100
0.2	0.0	0.3	3.64	1.06	0.05	46.4	26.2	9.88	98.1	95.7	99.0
0.4	0.0	0.3	3.63	7.67	42.2	52.2	6.78	5.38	99.2	82.7	30.3
0.6	0.0	0.3	3.63	46.9	100	64.8	0.96	42.9	99.9	59.3	4.46

5.2. Findings for a model with two endogenous regressors

Designing a just identified simultaneous model with two endogenous variables such that one can easily control the degree of simultaneity, the strength of the instruments and the multicollinearity between the regressors is not self-evident, as Kiviet and Pleus (2016, Section 3) shows. For the present purpose the situation is even more complex, because we will have to allow for possible invalidity of the instruments as well when analyzing the power of exclusion restriction tests. We proceed as follows.

Let the 5×1 vectors η_i contain (for $i = 1, ..., n$) independent drawings from a five element multivariate standard normal distribution. Now consider the linear transformation

$$d_i = (x_i^{(1)}, x_i^{(2)}, z_i^{(1)}, z_i^{(2)}, u_i)' = A\eta_i, \qquad (5.14)$$

with $A = (a_{jl})$ a 5×5 upper-diagonal real valued matrix. To realize that all elements of d_i have unit variance, the five rows of matrix A should all have inner-product unity. This directly implies $a_{55} = 1$ and $u_i = \eta_{i5}$. Note that the final column of A actually represents $(\rho_{x^{(1)}u}, \rho_{x^{(2)}u}, \rho_{z^{(1)}u}, \rho_{z^{(2)}u}, 1)'$. In the simulation we will control these four correlation parameters by choosing empirically relevant values for them, as well as for six other relevant correlations, all in the $(-1, +1)$ interval. The 10 yet unknown elements of A will follow from these 4+6 correlations, using the first four equations of (5.14), which are

$$x_i^{(1)} = a_{11}\eta_{i1} + a_{12}\eta_{i2} + a_{13}\eta_{i3} + a_{14}\eta_{i4} + \rho_{x^{(1)}u}\eta_{i5} \tag{5.15}$$

$$x_i^{(2)} = a_{22}\eta_{i2} + a_{23}\eta_{i3} + a_{24}\eta_{i4} + \rho_{x^{(2)}u}\eta_{i5} \tag{5.16}$$

$$z_i^{(1)} = a_{33}\eta_{i3} + a_{34}\eta_{i4} + \rho_{z^{(1)}u}\eta_{i5} \tag{5.17}$$

$$z_i^{(2)} = a_{44}\eta_{i4} + \rho_{z^{(2)}u}\eta_{i5}, \tag{5.18}$$

and the imposed unit variance of all five elements of d_i. The unit variance of (5.18) implies

$$a_{44} = (1 - \rho^2_{z^{(2)}u})^{1/2}. \tag{5.19}$$

By controlling the value of $\rho_{z^{(1)}z^{(2)}}$, which follows from (5.17) and (5.18) to be $\rho_{z^{(1)}u}\rho_{z^{(2)}u} + a_{34}a_{44}$, we find

$$a_{34} = (\rho_{z^{(1)}z^{(2)}} - \rho_{z^{(1)}u}\rho_{z^{(2)}u})/a_{44}. \tag{5.20}$$

Correlating (5.18) and (5.16) we find $\rho_{z^{(2)}x^{(2)}} = a_{44}a_{24} + \rho_{z^{(2)}u}\rho_{x^{(2)}u}$, so

$$a_{24} = (\rho_{z^{(2)}x^{(2)}} - \rho_{z^{(2)}u}\rho_{x^{(2)}u})/a_{44}, \tag{5.21}$$

and correlating (5.18) and (5.15) gives $\rho_{z^{(2)}x^{(1)}} = a_{44}a_{14} + \rho_{z^{(2)}u}\rho_{x^{(1)}u}$, hence

$$a_{14} = (\rho_{z^{(2)}x^{(1)}} - \rho_{z^{(2)}u}\rho_{x^{(1)}u})/a_{44}. \tag{5.22}$$

Due to the unit variance of (5.17) we have

$$a_{33} = (1 - a_{34}^2 - \rho^2_{z^{(1)}u})^{1/2}. \tag{5.23}$$

Then from $\rho_{z^{(1)}x^{(2)}} = a_{33}a_{23} + a_{34}a_{24} + \rho_{z^{(1)}u}\rho_{x^{(2)}u}$ we obtain

$$a_{23} = (\rho_{z^{(1)}x^{(2)}} - a_{34}a_{24} - \rho_{z^{(1)}u}\rho_{x^{(2)}u})/a_{33}, \tag{5.24}$$

and from $\rho_{z^{(1)}x^{(1)}} = a_{33}a_{13} + a_{34}a_{14} + \rho_{z^{(1)}u}\rho_{x^{(1)}u}$ we find

$$a_{13} = (\rho_{z^{(1)}x^{(1)}} - a_{34}a_{14} - \rho_{z^{(1)}u}\rho_{x^{(1)}u})/a_{33}. \tag{5.25}$$

The unit variance of (5.16) yields

$$a_{22} = (1 - a_{23}^2 - a_{24}^2 - \rho^2_{x^{(2)}u})^{1/2}, \tag{5.26}$$

and from $\rho_{x^{(1)}x^{(2)}} = a_{12}a_{22} + a_{13}a_{23} + a_{14}a_{24} + \rho_{x^{(1)}u}\rho_{x^{(2)}u}$ we get

$$a_{12} = (\rho_{x^{(1)}x^{(2)}} - a_{13}a_{23} - a_{14}a_{24} - \rho_{x^{(1)}u}\rho_{x^{(2)}u})/a_{22}, \tag{5.27}$$

and at long last

$$a_{11} = (1 - a_{12}^2 - a_{13}^2 - a_{14}^2 - \rho^2_{x^{(1)}u})^{1/2}. \tag{5.28}$$

So, all elements of matrix A can be expressed in the 10 correlations $\rho_{x^{(1)}u}$, $\rho_{x^{(2)}u}$, $\rho_{z^{(1)}u}$, $\rho_{z^{(2)}u}$, $\rho_{x^{(1)}x^{(2)}}$, $\rho_{z^{(1)}x^{(1)}}$, $\rho_{z^{(1)}x^{(2)}}$, $\rho_{z^{(2)}x^{(1)}}$, $\rho_{z^{(2)}x^{(2)}}$ and $\rho_{z^{(1)}z^{(2)}}$. Not all combinations of values for these correlations in the $(-1, 1)$ interval will be compatible though. Obvious requirements are

$$\left.\begin{array}{r} a_{34}^2 + \rho_{z^{(1)}u}^2 < 1, \\ a_{23}^2 + a_{24}^2 + \rho_{x^{(2)}u}^2 < 1, \\ a_{12}^2 + a_{13}^2 + a_{14}^2 + \rho_{x^{(1)}u}^2 \leq 1. \end{array}\right\} \tag{5.29}$$

We examined just a few compatible combinations of the ten correlations which seem relevant. Also we just considered solutions on the basis of the positive square roots for the diagonal elements of A.

That all elements of d_i have unit variance does not lead to loss of generality. The values to be chosen for β_1 and β_2 can compensate for the unit variance of $x_i^{(1)}$ and $x_i^{(2)}$ in model $y_i = x_i^{(1)}\beta_1 + x_i^{(2)}\beta_2 + u_i$. The KLS based test of joint restrictions on β_1 and β_2 can be shown to be invariant with respect to β_1 and β_2 when the null is true, and so is the KLS based test on the joint significance of $z_i^{(1)}$ and $z_i^{(2)}$ when added to this model, both under the null and alternatives, and it is also invariant to the scale of all regressors. So, when investigating the size of the KLS test of joint restrictions on β_1 and β_2 and the rejection probability both under the null and under alternatives for the joint exclusion restrictions test, we may without loss of generality set $\beta_1 = \beta_2 = 0$. In the simulations we took all vectors d_i in deviation from sample averages, so the results are actually about models that include an intercept as well, whereas they in fact also hold for models which yield similar vectors d_i after partialling out any number of arbitrary further exogenous regressors.

Table 4: Rejection frequencies (in %) of infeasible KLS tests ($K_1 = L_2 = 2; n = 100; \alpha = 0.05$)

$\rho_{z^{(1)}z^{(2)}} = 0.0$; $\rho_{z^{(2)}x^{(2)}} = 0.2$; $\rho_{z^{(2)}x^{(1)}} = 0.1$; $\rho_{z^{(1)}x^{(2)}} = 0.0$; $\rho_{z^{(1)}x^{(1)}} = 0.3$;							$\rho_{z^{(1)}z^{(2)}} = 0.3$; $\rho_{z^{(2)}x^{(2)}} = 0.5$; $\rho_{z^{(2)}x^{(1)}} = 0.3$; $\rho_{z^{(1)}x^{(2)}} = 0.3$; $\rho_{z^{(1)}x^{(1)}} = 0.6$;						
$\rho_{x^{(1)}x^{(2)}}$	$\rho_{z^{(2)}u}$	$\rho_{z^{(1)}u}$	$\rho_{x^{(2)}u}$	$\rho_{x^{(1)}u}$	R_x	R_z	$\rho_{x^{(1)}x^{(2)}}$	$\rho_{z^{(2)}u}$	$\rho_{z^{(1)}u}$	$\rho_{x^{(2)}u}$	$\rho_{x^{(1)}u}$	R_x	R_z
0.2	0.0	0.0	0.0	0.2	5.13	5.15	0.2	0.0	0.0	0.0	0.2	5.14	5.15
				0.5	5.03	4.73					0.5	5.03	4.12
			0.3	0.2	5.18	5.09				0.3	0.2	5.09	4.89
				0.5	5.07	4.65					0.5	5.00	4.05*
		0.4	0.0	0.2	5.12	97.7			0.4	0.0	0.2	5.15	99.8
				0.5	5.03	96.9					0.5	5.02	99.2
			0.3	0.2	5.16	97.9				0.3	0.2	5.08	99.8
				0.5	5.06	97.1					0.5	5.01	97.4*
	0.2	0.0	0.0	0.2	5.14	42.9		0.2	0.0	0.0	0.2	5.15	55.0
				0.5	5.04	42.1					0.5	5.02	51.1
			0.3	0.2	5.18	41.7				0.3	0.2	5.10	50.3
				0.5	5.07	40.3					0.5	5.02	36.0*
		0.4	0.0	0.2	5.14	99.5			0.4	0.0	0.2	5.16	100
				0.5	5.03	99.3					0.5	5.04	99.8
			0.3	0.2	5.16	99.5				0.3	0.2	5.09	99.9
				0.5	5.07	99.2					0.5	5.05	97.3*
0.6	0.0	0.0	0.0	0.2	5.14	5.11	0.6	0.0	0.0	0.0	0.2	5.11	5.11
				0.5	4.72	3.81					0.5	4.74	4.02*
			0.3	0.2	5.20	5.09				0.3	0.2	5.10	5.06
				0.5	5.09	4.71					0.5	5.06	4.11
		0.4	0.0	0.2	5.14	98.1			0.4	0.0	0.2	5.12	99.8
				0.5	4.71	96.0					0.5	4.71	91.6*
			0.3	0.2	5.17	98.7				0.3	0.2	5.11	99.9
				0.5	5.12	97.5					0.5	5.06	99.2
	0.2	0.0	0.0	0.2	5.15	43.1		0.2	0.0	0.0	0.2	5.11	56.5
				0.5	4.74	43.0					0.5	4.73	59.9*
			0.3	0.2	5.19	41.8				0.3	0.2	5.11	51.4
				0.5	5.10	41.6					0.5	5.07	50.7
		0.4	0.0	0.2	5.15	99.5			0.4	0.0	0.2	5.11	99.9
				0.5	4.72	99.0					0.5	4.72	95.2*
			0.3	0.2	5.18	99.7				0.3	0.2	5.12	99.8
				0.5	5.13	99.3					0.5	5.05	99.1

Table 4 presents some results for the two types of KLS tests for the ideal (but unrealistic) situation that $r = \rho_{xu}$ (true value of the degree of simultaneity is known). We examined all 1024 combinations of $\rho_{x^{(1)}u} \in \{0.2, 0.5\}$, $\rho_{x^{(2)}u} \in \{0.0, 0.3\}$, $\rho_{z^{(1)}u} \in \{0.0, 0.4\}$,

$\rho_{z^{(2)}u} \in \{0.0, 0.2\}$, $\rho_{x^{(1)}x^{(2)}} \in \{0.2, 0.6\}$, $\rho_{z^{(1)}x^{(1)}} \in \{0.3, 0.6\}$, $\rho_{z^{(1)}x^{(2)}} \in \{0.0, 0.3\}$, $\rho_{z^{(2)}x^{(1)}} \in \{0.1, 0.3\}$, $\rho_{z^{(2)}x^{(2)}} \in \{0.2, 0.5\}$ and $\rho_{z^{(1)}z^{(2)}} \in \{0.0, 0.3\}$, but present only 64 of them in Table 4. The table has two panels. In the left one we present all 32 results for the lower values of $\rho_{z^{(1)}x^{(1)}}$, $\rho_{z^{(1)}x^{(2)}}$, $\rho_{z^{(2)}x^{(1)}}$, $\rho_{z^{(2)}x^{(2)}}$ and $\rho_{z^{(1)}z^{(2)}}$, and in the right-hand panel those for their higher values. In all experiments the chosen correlation coefficients obeyed the compatibility criteria (5.29). The R_x column contains the rejection frequency (the estimated actual significance level) of the joint significance test on $x_i^{(1)}$ and $x_i^{(2)}$. This test existed in all experiments because we always found $\hat{\theta}(r) > 0$. R_z is the rejection frequency of the joint exclusion restriction test on $z_i^{(1)}$ and $z_i^{(2)}$, which represents its estimated actual significance level for cases where $\rho_{z^{(1)}u} = \rho_{z^{(2)}u} = 0$. For R_z results marked with an asterisk we found $\hat{\theta}(r) \leq 0$ in less than 0.1% of the replications. However, in a few of the cases not included in the table (and closer to the boundary of the admissible parameter space) we found the exclusion test to be undefined much more frequently.

From Table 4 we observe that the size properties of both infeasible KLS test procedures are also very reasonable in the more general model, whereas the power of the exclusion restrictions test seems fine. This being the case for the ideal situation in which $\rho_{x^{(1)}u}$ and $\rho_{x^{(2)}u}$ are supposed to be known provides the appropriate starting point for reasonably successful implementations under more realistic assumptions, as we saw in the preceding subsection.

6. Empirical illustrations

We will reproduce empirical results for three different cross-section data sets which have been used in well-known text books illustrating the standard use of OLS, TSLS and the testing of the strength and validity of instruments. Building on the same assumptions regarding the model specification (the included regressors and adopted homoskedasticity of the disturbances), we supplement these results with KLS inference on the coefficients as well as on the validity of the instruments, in particular the just-identifying restrictions.

6.1. A wage equation for employed women

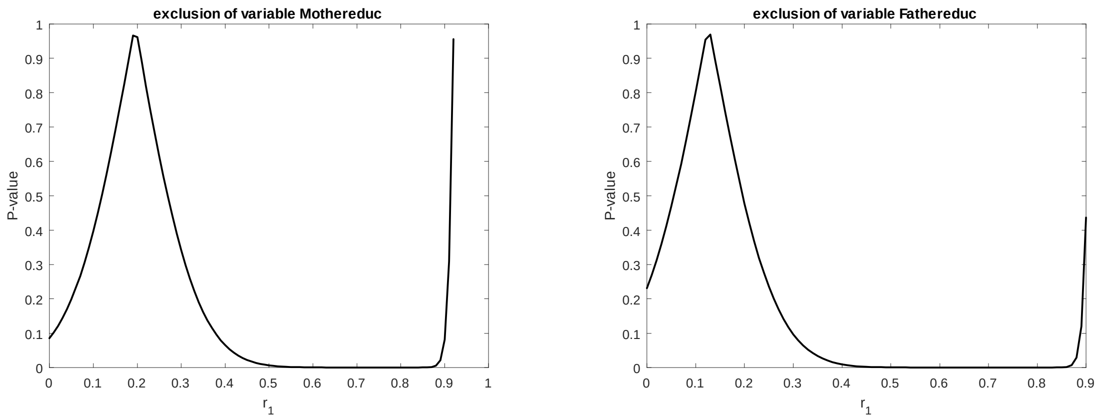
As a first illustration we closely follow a textbook example on IV/TSLS estimation given in Carter Hill et al. (2012, p.415). It concerns a subset of data originating from Mroz, namely a few variables on a sample of $n = 428$ employed women. After taking all these variables in deviation from their mean, the relationship considered is a special case of model (2.1), namely

$$y_i = \beta_1 x_i^{(1)} + x_i^{(2)\prime}\beta_2 + u_i, \tag{6.1}$$

with $K_1 = 1$ and $K_2 = 2$, where y_i is the log of wage, $x_i^{(1)}$ is education in years and vector $x_i^{(2)}$ contains the variables experience in years and its square. The parameter of primary interest is the return to schooling β_1. However, we suppose that its least-squares estimator will be positively biased, due to omission of the unavailable explanatory "ability", which will be positively correlated with $x_i^{(1)}$ while having a positive effect on y_i. Hence, due to $E(x_i^{(1)}u_i) > 0$ least-squares will be inconsistent. Assuming $E(x_i^{(2)}u_i) = 0$, the use of $L_2 = 2$ external instrumental variables $z_i^{(2)} = (z_{i1}^{(2)}, z_{i2}^{(2)})$ is being considered, where $z_{i1}^{(2)}$ is the education in years of the mother and $z_{i2}^{(2)}$ the education of the father of the woman concerned. In reduced form regressions for $x_i^{(1)}$, where next to $x_i^{(2)}$ just $z_{i1}^{(2)}$ is added, its F-test value is 73.95, whereas this is 87.74 for $z_{i2}^{(2)}$, so both instruments are pretty strong. Jointly they have an F-value of 55.40. The Sargan test for the single overidentification restriction when using both instruments has P-value 0.54. So, according to standard practice methods, acceptance of the TSLS results seems vindicated, although these are built on the untested suppositions that $E(x_i^{(2)} u_i) = 0$ and at least one of the instruments in $z_i^{(2)}$ is valid. Using the results of Section 4, we can produce further insights into the tenability of this latter supposition.

Figure 1 shows for all positive r_1 values which yield a positive value of $\hat{\theta}^*(r_1)$ of (4.6) the P-values of the single just-identifying exclusion restriction tests for the variables $z_i^{(1)}$ and $z_i^{(2)}$ respectively. Validity of both instruments seems quite likely only when ρ_1 is pretty close to 0.2; if $\rho_1 > 0.4$ it seems highly unlikely. Testing joint exclusion of both instruments (results not presented) leads to the same conclusion. Thus, without more specific information about the genuine value of ρ_1 testing the just-identification restriction is inconclusive.

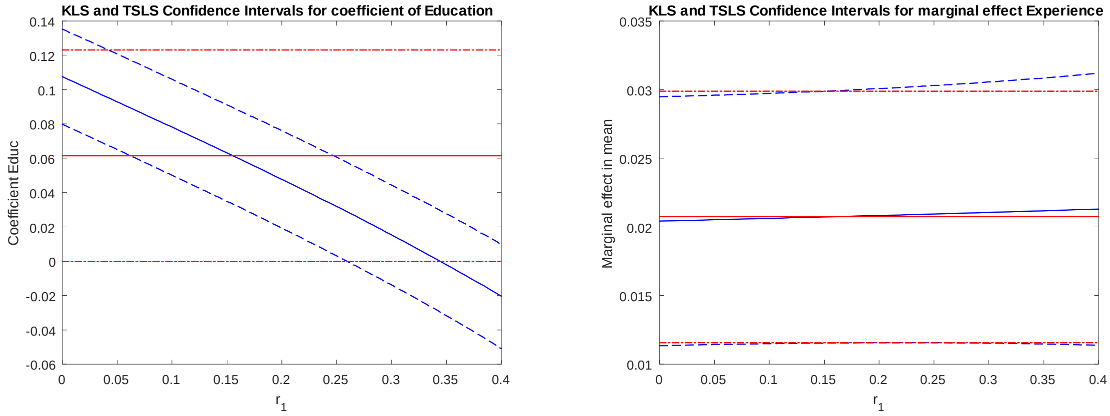
Figure 1: P-values of single just-identifying exclusion restriction tests



In its left graph Figure 2 shows the TSLS asymptotic 95% confidence interval for β_1 (red dash-dotted lines), which is invariant regarding ρ_1, and is centered at the TSLS estimate 0.0614 (red solid line). It also shows the KLS estimator (blue solid line), which varies with r_1 and the KLS asymptotic 95% confidence interval (blue dashed lines). The standard OLS nominal 95% confidence interval is indicated at $r_1 = 0$, centered around 0.108. The graph also shows that the 95% TSLS confidence interval, which is contingent on validity of four instruments, conforms in width to a conservative KLS-based interval contingent on the supposition $0.055 < \rho_1 < 0.265$ and on exogeneity of $x_i^{(2)}$.

In the right-hand graph of Figure 2 we analyze the marginal effect of experience $\partial x_i^{(2)\prime} \beta_2 / \partial x_{i1}^{(2)} = \beta_{21} + 2\beta_{22} x_{i1}^{(2)}$ evaluated in mean experience $\bar{x}_1^{(2)} = 13.04$. We see that here TSLS and KLS inference are almost similar. This seems due to the fact that the external instruments are pretty strong, whereas experience and its square are hardly correlated with the endogenous variable education. That differences between TSLS and KLS can also be huge can be seen from the next illustration.

Figure 2: Inference based on (non-)orthogonality conditions



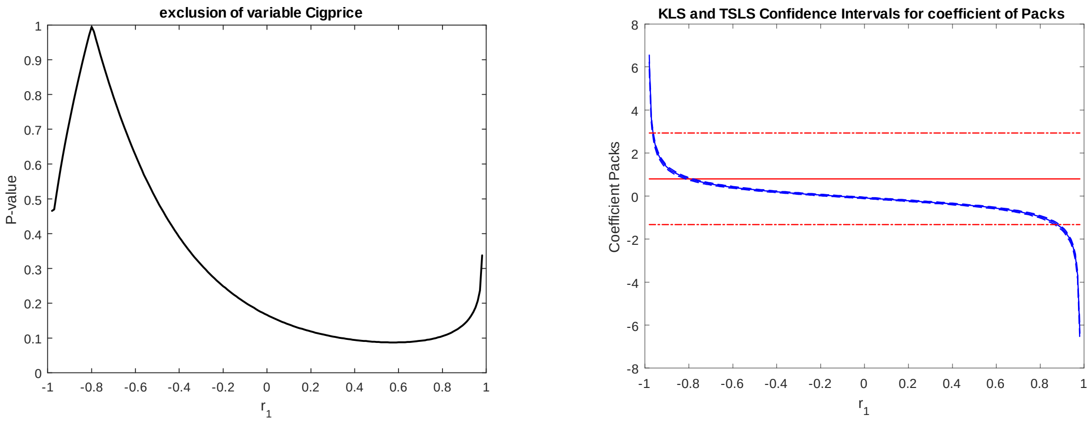
6.2. An analysis of the weight of newborns

In Wooldridge (2010, p.116) an exercise is presented in which it is analyzed for $n = 1388$ newborns whether smoking by their mother during pregnancy affects birth weight. A model like (6.1) is analyzed with $K_1 = 1$ but now $K_2 = 3$, where y_i is the log of birth weight, $x_i^{(1)}$ is the average number of packs of cigarettes smoked per day during pregnancy and vector $x_i^{(2)}$ contains a dummy for the gender of the baby, a variable parity, which is the birth order of the child, and the log of family income. It is assumed that $x_i^{(1)}$ is correlated with the disturbance term, because various further possible determinants of birth weight have been omitted from the model, such as use of alcohol, drugs, unhealthy food, physical inactivity, stress etc. Their effect seems positively correlated with smoking behavior, so we expect again a positive bias of least-squares and $\rho_1 > 0$. It is suggested to use the price of cigarettes as an instrumental variable, because economic theory predicts that it is negatively correlated with packs smoked, whereas it does not seem likely that this price has a direct effect on birth weight.

OLS yields a coefficient estimate for packs of -0.084 with standard error 0.017, which (when consistent) would suggest that each extra cigarette smoked per day (each package containing 20 cigarettes) reduces birth weight by about 0.4%. TSLS yields an outrageous coefficient of 0.651 (positive!) with standard error 0.854. These are clearly affected by weakness of the instrument since the relevant F-value in the first-stage regression is only 1.00. This makes sense, because price elasticity will be moderate, as smoking is addictive.

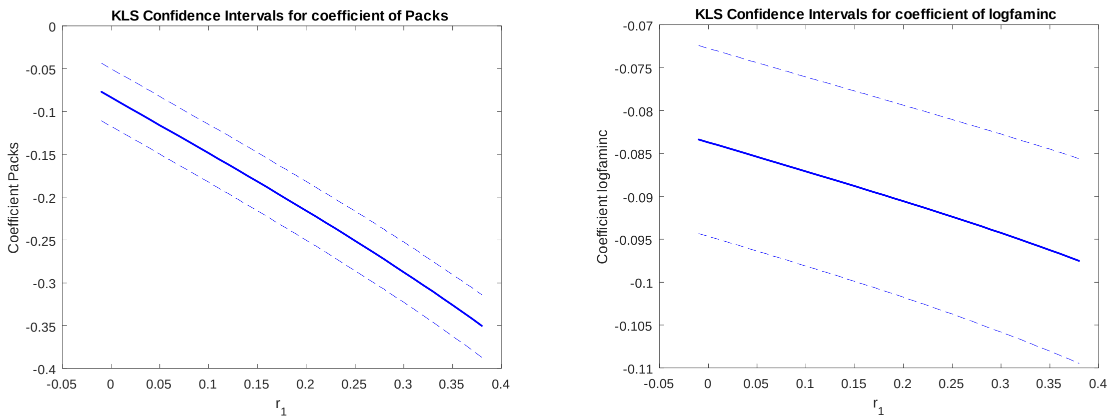
To this standard evidence the procedures developed in this study can add the following. The suspected positive bias of the OLS estimator of β_1 suggests that an extra cigarette per day may reduce birth weight by even more than 0.4%. From the left-hand side of Figure 3 we can see that for $r_1 > 0$ the validity of the instrument lacks strong support. Although the exclusion test does not force to reject at a significance level smaller than 10%, in order to justify the use of the instrument a P-value much larger, say exceeding 50%, would provide much more comfort. Moreover, these relatively low P-values do not encourage to move on by applying weak instrument techniques.

Figure 3: Inference on birth weight data based on (non-)orthogonality conditions



Assuming that the conditions to apply KLS do hold, the right-hand-side graph in Figure 3 shows that if we knew ρ_1 we would be able to produce highly accurate KLS inference on β_1 (blue lines; the confidence interval is so narrow that the figure barely shows it). For instance, it enables to infer rejection of the hypothesis $\beta_1 > 0$, provided $\rho_1 > -0.05$. KLS also allows a sensitivity analysis of TSLS: It shows that the extremely wide (and hence pretty useless) TSLS confidence interval is conservative at the nominal 95% level, provided $-0.95 < \rho_1 < 0.9$. Zooming in on this figure yields the left-hand side of Figure 4, from which we can deduce that for $0 \leq \rho_1 \leq 0.35$ (which does not seem unrealistic) the confidence set $-0.36 \leq \beta_1 \leq -0.05$ has asymptotic confidence coefficient 0.95. In the right-hand side of Figure 4 we produce KLS inference on one of the coefficients of the exogenous regressors, namely the log of family income. The TSLS estimate of this coefficient is 0.064 and its 95% confidence interval is $(-0.048, 0.175)$, but KLS learns that for realistic values of ρ_1 this coefficient is much smaller and significantly negative.

Figure 4: KLS inference on coefficients of birth weight data



6.3. A wage equation for young men

The above illustrations just required using the simple Corollary 1.2, because they concern models with only one endogenous regressor. Next we will exploit Theorem 1 in its full complexity in an empirical model where $K_1 = 2$ which is based on a classic data set originating from work by Griliches and also used for illustrative purposes on a subset ($n = 758$) of these data on young men in Hayashi (2000, p.251). In Kiviet and Pleus (2017, p.18) these same data have been used to illustrate tests on establishing the endogeneity of subsets of regressors. These tests are built on assuming validity of two identifying orthogonality conditions, which are untestable according to the classic approach. Like in the first illustration log wage is the dependent variable, but next to schooling also an iq test score is a possibly endogenous regressor in addition to a range of exogenous controls ($K_2 = 11$), including age and experience. The external instruments used are ($L_2 = 4$): age2 (age squared), expr2 (experience squared), kww (another test score) and kww2 (kww squared). The overall Sargan test (2 degrees of freedom) has satisfying P-value 0.89, but it leaves two underlying just-identifying restrictions untested.

Figure 5 shows (colored) contour plots for the P-values of the KLS based exclusion restrictions tests ($L_2 = 2$) of age2 and expr2 (left-hand) and kww and kww2 (right-hand) respectively. These plots have been obtained by calculating test statistic W^* of (4.9) over a range of values for the simultaneity correlations, where r_1 refers to schooling and r_2 to iq score. The grid values -0.99:0.01:0.99 have been examined for both. For cases where $r'S_xS_{xx}^{-1}S_xr > 0.99$ we did set the P-value at 1.1, to be interpreted as "not defined". Both plots show that the statistic is defined over an ellipse indeed. The exclusion restriction regarding the squares of both included regressors age and expr does not have to be rejected whatever the true values of ρ_1 and ρ_2 will be, since all P-values exceed 0.75. This is pretty hard evidence (although not irrefutable) on the possible validity of these two instruments. For score test variable kww and its square the situation is different. Over a substantial area of (ρ_1, ρ_2) combinations their exclusion test has P-values well below 0.1, whereas the area where it exceeds 0.7 forms just a narrow shell, covering cases where $\rho_1^2 + \rho_2^2$ is relatively large. Especially when the simultaneity is nonexistent or mild the validity of kww and kww2 as instruments seems doubtful. Assuming that both schooling and iq are positively related to "ability", we expect both ρ_1 and ρ_2 to be mildly positive.

Figure 5: P-values of two just-identifying exclusion restriction tests

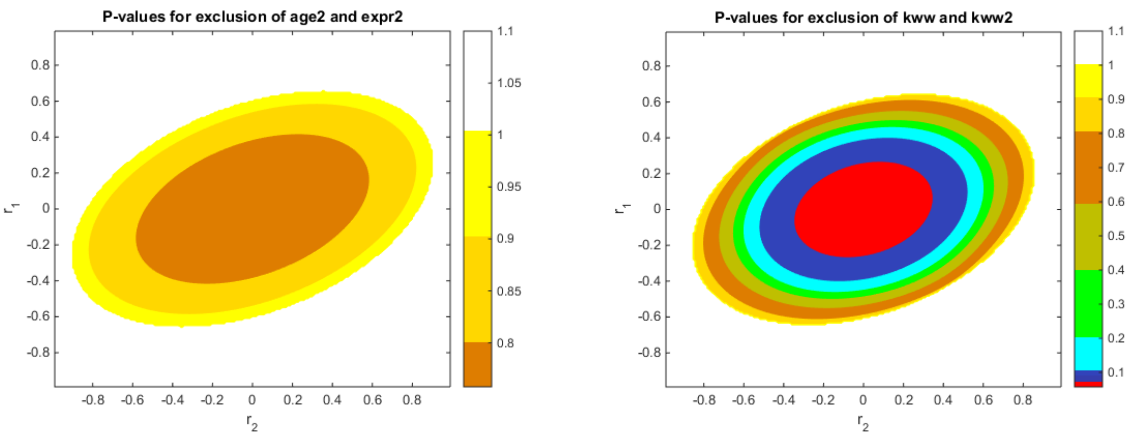
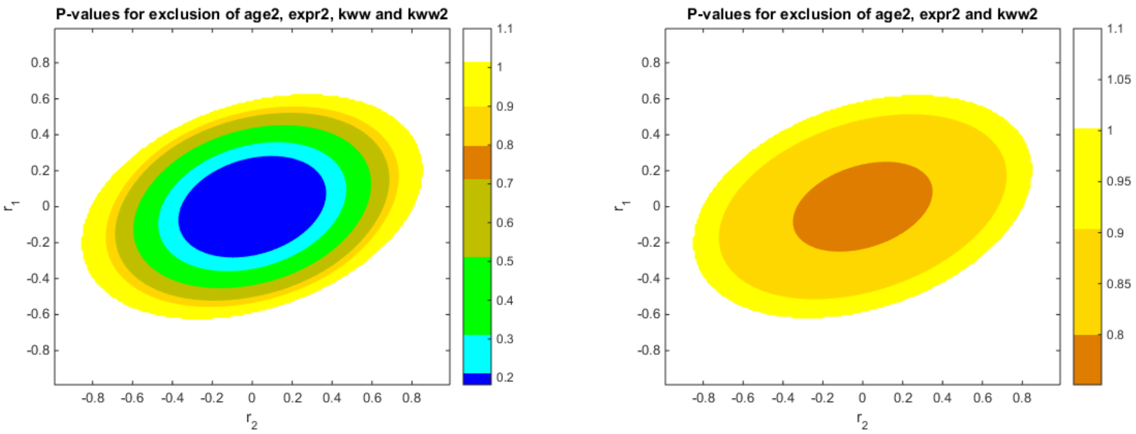


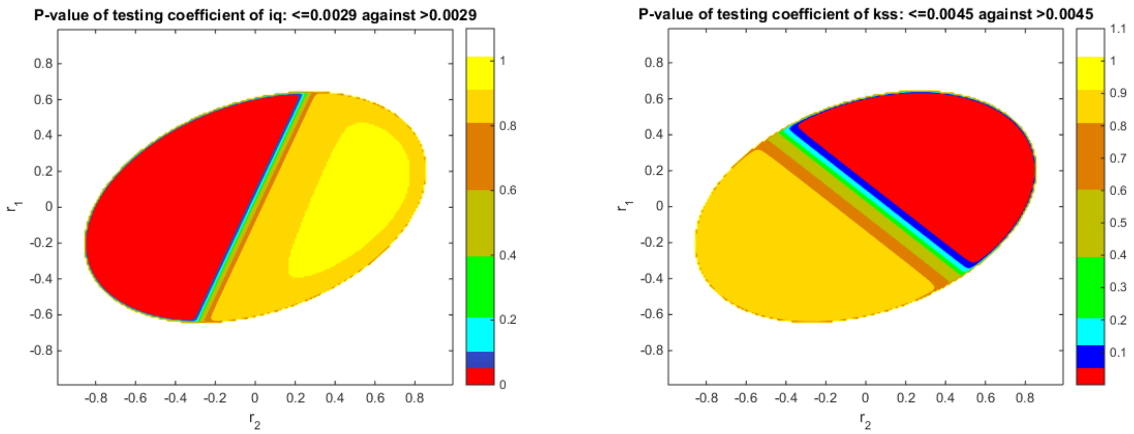
Figure 6: P-values of exclusion restriction tests in two different models



In the left-hand side contour plot of Figure 6 we test the exclusion of the $L_2 = 4$ variables jointly. No P-values are obtained below 0.18 now. This demonstrates that the exclusion test may have limited power when some rightly (age2 and expr2) and some wrongly (kww and kww2) excluded regressors are tested jointly. In the right-hand side of Figure 6 we test the model in which kww has been included as an exogenous regressor ($K_2 = 12$), and it is tested whether the $L_2 = 3$ squared variables seem valid external instruments for the two endogenous regressors schooling and iq. Figure 6 highlights that the inference on endogeneity of these two regressors as presented in Kiviet and Pleus (2017), which uses the model and instruments of the left-hand contour plot, although supported by a large P-value of the Sargan test, should better have been executed for the model and instruments as used in the right-hand contour plot. The latter plot does not discourage the use of the three squared external instruments, irrespective of the actual values of ρ_1 and ρ_2.

The type of results just presented can be used in a TSLS analysis to avoid the wrong exclusion of regressors and employment of invalid instruments. But they can also be used to find an adequate model specification and next –avoiding the use of possibly weak or invalid instruments– analyze its coefficients on the basis of KLS inference. We will illustrate the latter here, just focussing on the coefficients of iq and kss, in the model which includes kww as an extra regressor and excludes the three squared variables ($K_2 = 12$), while treating both schooling and iq as endogenous ($K_1 = 2$). We perform one sided tests on the single hypothesis that these coefficients exceed some particular value. For this value we chose their estimated values as obtained by OLS, which are 0.0029 and 0.0045 respectively. From the contours of Figure 7 one can see that, roughly, this hypothesis is rejected for iq when $\rho_1 > 2\rho_2$ and for kss when $\rho_1 > -0.9\rho_2$. Assuming both ρ_1 and ρ_2 to be about 0.3 it seems likely that the coefficient of iq is smaller than 0.0029 and that of kss larger than 0.0045.

Figure 7: KLS inference on particular coefficients



To depict similar inference results when we would allow kss to be endogenous too becomes more complicated. However, it is not impossible. One could produce, next to the above results where kss is supposed to be exogenous (which could be expressed as $\rho_3 = 0$), also contours for, say, $\rho_3 = 0.1 : 0.1 : 0.5$. So, even in rather general models such series of contour plots allow to produce KLS inference, and examine the sensitivity of OLS and TSLS coefficient estimates and the validity of instruments, conditional on assumptions regarding simultaneity. To a limited degree such inference is robust regarding simultaneity.

7. Conclusion

By standard and by incremental or difference Sargan-Hansen tests for overidentifying restrictions the validity of a subgroup of instruments can be tested, provided a sufficient number of valid and sufficiently strong just- or overidentifying instruments are already available. If one wants to verify whether just-identifying instruments are valid indeed, the only route provided by the standard approach is: first adopt another non-testable set of valid identifying instruments. So, providing full-fledged statistical evidence on the validity of all instruments is simply impossible by these tools. It mimics the miserable situation where for a proof by mathematical induction one can prove the induction step, whereas proof of the truth of a base case is yet missing, and its proof seems even completely beyond reach.

This study presents an approach by which, without exploiting any external instruments, general linear coefficient restrictions can be tested in a multiple regression model with an arbitrary number of endogenous regressors. Instead, it requires a flexible assumption on the degree of endogeneity of all regressors. This approach –which is of substantial practical relevance by itself– can be implemented such that it also allows to generate statistical evidence on the tenability of exclusion restrictions. When this yields an acceptable just-identifying or overidentifying set of instruments, it provides the essential underlying building block for a standard or for a series of incremental Sargan-Hansen tests which was missing so far. This supporting building block seems mandatory if one wants to employ and prop up inference based on using instrumental variables. However, since the general tools developed in this paper also allow to produce inference on coefficient values while avoiding the use of external instruments altogether, one may prefer to avoid the ensuing problems such as sacrificing credibility, accuracy and power due to possible weakness or invalidity of instruments.

Whether the instrument-free approach yields clear-cut inferences depends from case to case on the width of the area which is supposed to contain the actual values of correlation coefficients between regressor variables and disturbances, and on whether P-values of relevant test statistics are relatively constant over this area. Computationally the tools developed here are not very demanding, and they can also be used to provide a sensitivity analysis of least-squares and of instrumental variables based inferences with respect to less strict assumptions regarding their unavoidable orthogonality assumptions.

Of course, as always, deeper insights and further generalizations are called for. Preceding the usual Sargan-Hansen tests by a just-identifying exclusion restrictions test exacerbates the pre-test problems. Our present results presuppose homoskedasticity of both disturbances and regressors. So, if this is not the case yet one should manage to first weigh all observations such that this is achieved as closely as possible. Developing inference methods which are robust regarding both simultaneity and heteroskedasticity, also for dependent data, and at the same time do control size and boost power over the whole model building process, remain a challenge for future efforts.

Acknowledgments

Constructive comments by two referees and a guest coeditor are gratefully acknowledged, and so is the overall guidance by the four guest coeditors. I thank Frank Windmeijer for alerting me that in an earlier version I had wrongly treated and interpreted the first two illustrations as cases of measurement error in stead of omitted variables.

References

Altonji, J.G., Elder, T.E., Taber, C.R., 2005. An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human Resources* 40, 791-821.

Ashley, R., 2009. Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis. *Journal of Applied Econometrics* 24, 325-337.

Bound, J., Jaeger, D.A., 2000. Do compulsory school attendance laws alone explain the association between quarter of birth and earnings? *Research in Labor Economics*, Vol. 19, 83-108

Carter Hill, R., Griffiths, W.E., Lim, G.C., 2012. *Principles of Econometrics*, fourth edition. John Wiley and Sons.

Conley, T.G., Hansen, C.B., Rossi, P.E., 2012. Plausibly exogenous. *The Review of Economics and Statistics* 94, 260-272.

Davidson, R., MacKinnon, J. G., 2004. *Econometric Theory and Methods.* Oxford University Press, Oxford.

Deaton, A., 2010. Instruments, randomization, and learning about development. *Journal of Economic Literature* 48, 424–455.

Dufour, J.-M., 2003. Identification, weak instruments, and statistical inference in econometrics. *Canadian Journal of Economics* 36, 767-808.

Goldberger, A.S., 1964. *Econometric Theory.* John Wiley, New York, NY.

Hansen, P.L., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054.

Hayashi, F. , 2000. Econometrics. Princeton University Press.

Kadane, J.B., 1971. Comparison of k-class estimators when the disturbances are small. *Econometrica* 39, 723-737.

Kiviet, J.F., 2013. Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *The Econometrics Journal* 16, S24-S59.

Kiviet, J.F., 2016. When is it really justifiable to ignore explanatory variable endogeneity in a regression model? *Economics Letters* 145, 192-195.

Kiviet, J.F., 2017. Discriminating between (in)valid external instruments and (in)valid exclusion restrictions. *Journal of Econometric Methods* 6, 1-9.

Kiviet, J.F., Niemczyk, J., 2012. The asymptotic and finite sample (un)conditional distributions of OLS and simple IV in simultaneous equations. *Journal of Computational Statistics and Data Analysis* 56, 3567-3586.

Kiviet, J.F., Pleus, M., 2017. The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation. *Econometrics and Statistics* 2 (2017) 1-21.

Kraay, A., 2012. Instrumental variables regressions with uncertain exclusion restrictions: A Bayesian approach. *Journal of Applied Econometrics* 27, 108-128.

Nevo, A., Rosen, A.M., 2012. Identification with imperfect instruments. *The Review of Economics and Statistics* 94, 659-671.

Newey, W.K., 1985. Generalized method of moment specifiction testing. *Journal of Econometrics* 29, 229-256.

Parente, P.M.D.C, Santos Silva, J.M.C., 2012. A cautionary note on tests of overidentifying restrictions. *Economics Letters* 115, 314-317.

Rothenberg, T.J., 1972. The asymptotic distribution of the least squares estimator in the errors in variables model. *Unpublished mimeo.*

Sargan, J.D., 1958. Estimation of economic relationships using instrumental variables. *Econometrica* 26, 393-514.

Stock, J.H., Watson, M.W., 2015. *Introduction to Econometrics*, updated third edition. Pearson, Boston.

Windmeijer, F., 2018. Two-stage least-squares as minimum distance. Forthcoming in *Econometrics Journal.*

Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*, second edition. MIT Press, Cambridge, Massachusetts.

Appendices

A. Some basic derivations

We have assumed that $\{(x_i', u_i)'; i = 1, ..., n\}$ are independently and identically distributed with zero mean and

$$Var\begin{pmatrix} x_i \\ u_i \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \sigma_{xu} \\ \sigma_{xu}' & \sigma_u^2 \end{pmatrix}.$$

All elements of the latter matrix are assumed to be finite. In addition, we assume that $E(u_i^4) = \kappa_u \sigma_u^4$ and $E(x_{ij}^4) = \kappa_x \sigma_j^4$, with $1 \leq \kappa_u < \infty$ and $1 \leq \kappa_x < \infty$. We denote the typical element of Σ_{xx} by σ_{jk} $(j, k = 1, ..., K)$, but for its diagonal elements we will sometimes use $\sigma_j^2 = \sigma_{jj}$. The typical element of vector σ_{xu} can be denoted $\rho_j \sigma_j \sigma_u$, because $\rho_j = \sigma_{x_j u}/(\sigma_j \sigma_u)$. Next to Σ_{xx} and its sample equivalent S_{xx}, where for the latter we defined two options at the end of Section 2, we will also use the matrices $\Sigma_x^2 = \text{diag}(\sigma_1^2, ..., \sigma_K^2)$ and $\Sigma_x = \text{diag}(\sigma_1, ..., \sigma_K)$, as well as the diagonal matrices S_x and S_x^2. The latter has the same main diagonal as S_{xx}, and $S_x S_x = S_x^2$.

Invoking a standard version of the central limit theorem, we can now obtain the following results, which will be exploited later. We have

$$n^{1/2}(u'u/n - \sigma_u^2) = n^{-1/2}\Sigma_{i=1}^n (u_i^2 - \sigma_u^2) \xrightarrow{d} \mathcal{N}[0, (\kappa_u - 1)\sigma_u^4], \tag{A.1}$$

because $Var(u_i^2 - \sigma_u^2) = E(u_i^4) - \sigma_u^4 = (\kappa_u - 1)\sigma_u^4$. So, $u'u/n - \sigma_u^2 = O_p(n^{-1/2})$. Also

$$n^{1/2}(X'u/n - \sigma_{xu}) = n^{-1/2}\Sigma_{i=1}^n (x_i u_i - \sigma_{xu}) \xrightarrow{d} \mathcal{N}[0, \sigma_u^2 \Sigma_{xx} + (\kappa_u - 2)\sigma_{xu}\sigma_{xu}'], \tag{A.2}$$

hence $X'u/n - \sigma_{xu} = O_p(n^{-1/2})$. This is found by decomposing x_i into two components, $x_i = \xi_i + \sigma_{xu}\sigma_u^{-2} u_i$, where ξ_i is independent of u_i. Of course, $E(\xi_i) = 0$ and $E(\xi_i u_i) = 0$, so $E(x_i u_i) = \sigma_{xu}$ indeed. Since $Var(x_i) = \Sigma_{xx} = Var(\xi_i) + \sigma_u^{-2}\sigma_{xu}\sigma_{xu}'$, result (A.2) follows from $Var(x_i u_i - \sigma_{xu}) = E(u_i^2 x_i x_i') - \sigma_{xu}\sigma_{xu}' = E(u_i^2 \xi_i \xi_i') + \sigma_{xu}\sigma_{xu}'\sigma_u^{-4}E(u_i^4) - \sigma_{xu}\sigma_{xu}' = \sigma_u^2 Var(\xi_i) + (\kappa_u - 1)\sigma_{xu}\sigma_{xu}' = \sigma_u^2 \Sigma_{xx} + (\kappa_u - 2)\sigma_{xu}\sigma_{xu}'$.

For $j, k = 1, ..., K$ we have

$$n^{1/2}(X'X/n - \Sigma_{xx})_{j,k} = n^{-1/2}\Sigma_{i=1}^n (x_{ij}x_{ik} - \sigma_{jk}) \xrightarrow{d} \mathcal{N}[0, \sigma_j^2\sigma_k^2 + (\kappa_x - 2)\sigma_{jk}^2]. \tag{A.3}$$

This is proved by decomposing $x_{ik} = a_1 x_{ij} + a_2 \eta_{ik}$, where x_{ij} and η_{ik} are independent and η_{ik} has zero mean and unit variance. Because $E(x_{ik}^2) = \sigma_k^2 = a_1^2\sigma_j^2 + a_2^2$ and $E(x_{ik}x_{ij}) = \sigma_{kj} = a_1\sigma_j^2$ we have $a_1 = \sigma_{kj}\sigma_j^{-2}$ and $a_2^2 = \sigma_k^2 - \sigma_{kj}^2\sigma_j^{-2}$. Now we obtain $E(x_{ij}^2 x_{ik}^2) = E[x_{ij}^2(a_1^2 x_{ij}^2 + 2a_1 a_2 x_{ij}\eta_{ik} + a_2^2\eta_{ik}^2)] = \kappa_x\sigma_{jk}^2 + \sigma_j^2(\sigma_k^2 - \sigma_{kj}^2\sigma_j^{-2}) = \sigma_j^2\sigma_k^2 + (\kappa_x - 1)\sigma_{jk}^2$, thus $Var(x_{ij}x_{ik} - \sigma_{jk}) = \sigma_j^2\sigma_k^2 + (\kappa_x - 2)\sigma_{jk}^2$, from which (A.3) follows. So, $n^{-1}X'X - \Sigma_{xx} = O_p(n^{-1/2})$.

A result involving the Hadamard (element by element) matrix product (denoted $\circ$) to be exploited later is

$$n^{1/2}(S_x^2 - \Sigma_x^2)\rho_{xu} \xrightarrow{d} \mathcal{N}[0, (\kappa_x - 1)\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}], \tag{A.4}$$

where $\mathcal{R} = \text{diag}(\rho_1, ..., \rho_K)$. For S_x^2, we have $n^{1/2}(S_x^2 - \Sigma_x^2)\rho_{xu} = n^{-1/2}\Sigma_{i=1}^n v_i$ with $v_i' = ((x_{i1}^2 - \sigma_1^2)\rho_1, ..., (x_{iK}^2 - \sigma_K^2)\rho_K)'$. Using the expression for $E(x_{ij}^2 x_{ik}^2)$ just derived, we find $E[(x_{ij}^2 - \sigma_j^2)(x_{ik}^2 - \sigma_k^2)] = (\kappa_x - 1)\sigma_{jk}^2$. Thus $E[(x_{ij}^2 - \sigma_j^2)\rho_j(x_{ik}^2 - \sigma_k^2)\rho_k] = (\kappa_x - 1)\rho_j\sigma_{jk}^2\rho_k$, which is the typical element of the limiting variance matrix of (A.4).

We also need the mutual covariances of scalar (A.1) and vectors (A.2) and (A.4). We find $E[(u_i^2-\sigma_u^2)(x_iu_i-\sigma_{xu})] = E[(u_i^2-\sigma_u^2)(\xi_iu_i+\sigma_{xu}\sigma_u^{-2}u_i^2-\sigma_{xu})] = (\kappa_u-1)\sigma_u^2\sigma_{xu}$, hence

$$nE[(u'u/n-\sigma_u^2)(X'u/n-\sigma_{xu})] = (\kappa_u-1)\sigma_u^2\sigma_{xu}. \tag{A.5}$$

Using $x_{ij} = \xi_{ij}+\rho_j\sigma_j\sigma_u^{-1}u_i$, from $E[(u_i^2-\sigma_u^2)(x_{ij}^2-\sigma_j^2)\rho_j] = \rho_jE[(u_i^2-\sigma_u^2)(\xi_{ij}^2+2\sigma_j\sigma_u^{-1}u_i\xi_{ij}+\rho_j^2\sigma_j^2\sigma_u^{-2}u_i^2)] = (\kappa_u-1)\sigma_u^2\rho_j^3\sigma_j^2$ we find

$$nE[(u'u/n-\sigma_u^2)(S_x^2-\Sigma_x^2)\rho_{xu}] = (\kappa_u-1)\sigma_u^2\Sigma_x^2\mathcal{R}^2\rho_{xu}. \tag{A.6}$$

And, using $Var(\xi_i) = \Sigma_{xx}-\sigma_u^{-2}\sigma_{xu}\sigma_{xu}'$, from

$$\begin{aligned}
&E[(\xi_{ij}u_i+\rho_j\sigma_j\sigma_u^{-1}u_i^2-\rho_j\sigma_j\sigma_u)(\xi_{ik}^2+2\rho_k\sigma_k\sigma_u^{-1}\xi_{ik}u_i+\rho_k^2\sigma_k^2\sigma_u^{-2}u_i^2-\sigma_k^2)\rho_k] \\
&= 2\rho_k^2\sigma_k\sigma_u(\Sigma_{xx}-\sigma_u^{-2}\sigma_{xu}\sigma_{xu}')_{jk}+(\kappa_u-1)\rho_j\sigma_j\sigma_u\rho_k^3\sigma_k^2 \\
&= 2\sigma_u\rho_k^2\sigma_k(\Sigma_{xx})_{jk}+(\kappa_u-3)\sigma_u\sigma_j\rho_j\rho_k^3\sigma_k^2
\end{aligned}$$

we obtain

$$nE[(X'u/n-\sigma_{xu})\rho_{xu}'(S_x^2-\Sigma_x^2)] = 2\sigma_u\Sigma_{xx}\Sigma_x\mathcal{R}^2+(\kappa_u-3)\sigma_u\Sigma_x\rho_{xu}\rho_{xu}'\Sigma_x^2\mathcal{R}^2. \tag{A.7}$$

B. Proof of Theorem 1

To find the limiting distribution of the inconsistency corrected OLS estimator $\hat\beta(\rho_{xu}) = \hat\beta_{OLS}-n\cdot\hat\sigma_u(\rho_{xu})(X'X)^{-1}S_x\rho_{xu}$ we examine

$$n^{1/2}[\hat\beta(\rho_{xu})-\beta] = (n^{-1}X'X)^{-1}[n^{-1/2}X'u-n^{1/2}\hat\sigma_u(\rho_{xu})S_x\rho_{xu}]. \tag{B.1}$$

First, we have to separate from the right-hand side expression the leading $O_p(1)$ terms from $o_p(1)$ terms. Matrix $n^{-1}X'X = O_p(1)$ can be decomposed as

$$n^{-1}X'X = \Sigma_{xx}+(n^{-1}X'X-\Sigma_{xx}), \tag{B.2}$$

where the first component is deterministic and finite, denoted as $\Sigma_{xx} = O(1)$, and the second component is $n^{-1}X'X-\Sigma_{xx} = O_p(n^{-1/2})$, see derivation below (A.3). Exploiting the smaller order of this second component we find

$$\begin{aligned}
(n^{-1}X'X)^{-1} &= (\Sigma_{xx}+n^{-1}X'X-\Sigma_{xx})^{-1} = \Sigma_{xx}^{-1}[I+(n^{-1}X'X-\Sigma_{xx})\Sigma_{xx}^{-1}]^{-1} \\
&= \Sigma_{xx}^{-1}[I-(n^{-1}X'X-\Sigma_{xx})\Sigma_{xx}^{-1} \\
&\qquad +(n^{-1}X'X-\Sigma_{xx})\Sigma_{xx}^{-1}(n^{-1}X'X-\Sigma_{xx})\Sigma_{xx}^{-1}-...] \\
&= \Sigma_{xx}^{-1}-\Sigma_{xx}^{-1}(n^{-1}X'X-\Sigma_{xx})\Sigma_{xx}^{-1}+o_p(n^{-1/2}).
\end{aligned} \tag{B.3}$$

Hence, this inverse has a leading $O(1)$ term, a second term of order $O_p(n^{-1/2})$ plus a remainder of smaller order.

The first term in the factor between square brackets in (B.1) is

$$n^{-1/2}X'u = n^{1/2}\sigma_{xu}+n^{1/2}(n^{-1}X'u-\sigma_{xu}), \tag{B.4}$$

so it can be decomposed in a deterministic $O(n^{1/2})$ and a random $O_p(1)$ component, see (A.2). To find the leading components in decreasing order of the second term in the factor between square brackets in (B.1) we start by considering $S_x^2 = \Sigma_x^2 + (S_x^2 - \Sigma_x^2)$, where $S_x^2 - \Sigma_x^2 = O_p(n^{-1/2})$, as shown in (A.4). Hence, using S_x and Σ_x as defined in Appendix A,

$$S_x = \Sigma_x + 0.5\Sigma_x^{-1}(S_x^2 - \Sigma_x^2) + o_p(n^{-1/2}), \tag{B.5}$$

which is proved by evaluating S_xS_x, as this yields $\Sigma_x^2 + S_x^2 - \Sigma_x^2 + o_p(n^{-1/2}) = S_x^2 + o_p(n^{-1/2})$.

Next we consider (2.11) evaluated in ρ_{xu}, which is

$$\hat{\sigma}_u^2(\rho_{xu}) = [1 - \rho_{xu}' S_x(n^{-1}X'X)^{-1}S_x\rho_{xu}]^{-1}[n^{-1}u'u - (n^{-1}X'u)'(n^{-1}X'X)^{-1}(n^{-1}X'u)]. \tag{B.6}$$

We first decompose $S_x(n^{-1}X'X)^{-1}S_x$ by substituting (B.3) and (B.5), which yields

$$\begin{aligned}
&S_x(n^{-1}X'X)^{-1}S_x \\
&= [\Sigma_x + 0.5\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)][\Sigma_{xx}^{-1} - \Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}] \\
&\qquad \times [\Sigma_x + 0.5\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)] + o_p(n^{-1/2}) \\
&= \Sigma_x\Sigma_{xx}^{-1}\Sigma_x + 0.5\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\Sigma_{xx}^{-1}\Sigma_x - \Sigma_x\Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}\Sigma_x \\
&\quad + 0.5\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2) + o_p(n^{-1/2}).
\end{aligned}$$

As diagonal matrices commute, we have $\Sigma_x^{-1}(S_x^2 - \Sigma_x^2) = (S_x^2 - \Sigma_x^2)\Sigma_x^{-1}$. Using $\theta = 1 - \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}$, we now obtain

$$\begin{aligned}
1 - \rho_{xu}'S_x(n^{-1}X'X)^{-1}S_x\rho_{xu} &= \theta - \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu} \\
&\quad + \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}\Sigma_x\rho_{xu} + o_p(n^{-1/2}).
\end{aligned}$$

Next the first factor of (B.6) can be decomposed as

$$\begin{aligned}
[1 - \rho_{xu}'S_x(n^{-1}X'X)^{-1}S_x\rho_{xu}]^{-1} &= \theta^{-1} + \theta^{-2}\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu} \\
&\quad - \theta^{-2}\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}\Sigma_x\rho_{xu} \\
&\quad + o_p(n^{-1/2}),
\end{aligned}$$

and for the second factor of (B.6) we find, using (B.4), (B.3) and $\sigma_{xu} = \sigma_u\Sigma_x\rho_{xu}$,

$$\begin{aligned}
&n^{-1}u'u - (n^{-1}X'u)'(n^{-1}X'X)^{-1}(n^{-1}X'u) \\
&= \sigma_u^2 + (n^{-1}u'u - \sigma_u^2) \\
&\quad - [\sigma_{xu} + (n^{-1}X'u - \sigma_{xu})]'[\Sigma_{xx}^{-1} - \Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}] \\
&\qquad \times [\sigma_{xu} + (n^{-1}X'u - \sigma_{xu})] + o_p(n^{-1/2}) \\
&= \sigma_u^2\theta + (n^{-1}u'u - \sigma_u^2) - 2\sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'u - \sigma_{xu}) \\
&\quad + \sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}\sigma_{xu} + o_p(n^{-1/2}).
\end{aligned}$$

The above yields for (B.6)

$$
\begin{aligned}
&\hat{\sigma}_u^2(\rho_{xu}) \\
&= \{\theta^{-1} + \theta^{-2}[\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu} - \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}]\}\times \\
&\quad [\sigma_u^2\theta + (n^{-1}u'u - \sigma_u^2) - 2\sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'u - \sigma_{xu}) + \sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}\sigma_{xu}] \\
&\quad + o_p(n^{-1/2}) \\
&= \sigma_u^2 + \theta^{-1}(n^{-1}u'u - \sigma_u^2) - 2\theta^{-1}\sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'u - \sigma_{xu}) + \theta^{-1}\sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}\sigma_{xu} \\
&\quad + \sigma_u^2\theta^{-1}[\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu} - \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}(n^{-1}X'X - \Sigma_{xx})\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}] + o_p(n^{-1/2}) \\
&= \sigma_u^2 + \theta^{-1}(n^{-1}u'u - \sigma_u^2) - 2\theta^{-1}\sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'u - \sigma_{xu}) + \sigma_u^2\theta^{-1}\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu} \\
&\quad + o_p(n^{-1/2}),
\end{aligned}
$$

from which we obtain

$$
\begin{aligned}
\hat{\sigma}_u(\rho_{xu}) = \sigma_u &+ 0.5\sigma_u^{-1}\theta^{-1}(n^{-1}u'u - \sigma_u^2) - \sigma_u^{-1}\theta^{-1}\sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'u - \sigma_{xu}) \\
&+ 0.5\sigma_u\theta^{-1}\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu} + o_p(n^{-1/2}).
\end{aligned}
$$

For the factor between square brackets at the right-hand side of (B.1) we now find from the above, upon collecting all $o_p(1)$ terms in a remainder term,

$$
\begin{aligned}
&n^{-1/2}X'u - n^{1/2}\hat{\sigma}_u(\rho_{xu})S_x\rho_{xu} \\
&= n^{1/2}\sigma_{xu} + n^{1/2}(n^{-1}X'u - \sigma_{xu}) \\
&\quad - n^{1/2}[\sigma_u + 0.5\sigma_u^{-1}\theta^{-1}(n^{-1}u'u - \sigma_u^2) - \sigma_u^{-1}\theta^{-1}\sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'u - \sigma_{xu}) \\
&\qquad + 0.5\sigma_u\theta^{-1}\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu}][\Sigma_x + 0.5\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)]\rho_{xu} + o_p(1) \\
&= n^{1/2}[\sigma_{xu} + (n^{-1}X'u - \sigma_{xu}) - \sigma_u\Sigma_x\rho_{xu} - 0.5\sigma_u^{-1}\theta^{-1}(n^{-1}u'u - \sigma_u^2)\Sigma_x\rho_{xu} \\
&\qquad + \sigma_u^{-1}\theta^{-1}\sigma_{xu}'\Sigma_{xx}^{-1}(n^{-1}X'u - \sigma_{xu})\Sigma_x\rho_{xu} \\
&\qquad - 0.5\sigma_u\theta^{-1}\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu}\Sigma_x\rho_{xu} - 0.5\sigma_u\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu}] + o_p(1) \\
&= n^{1/2}[(I + \theta^{-1}\Sigma_x\rho_{xu}\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1})(n^{-1}X'u - \sigma_{xu}) \\
&\qquad - 0.5\sigma_u(I + \theta^{-1}\Sigma_x\rho_{xu}\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1})\Sigma_x^{-1}(S_x^2 - \Sigma_x^2)\rho_{xu} \\
&\qquad - 0.5\sigma_u^{-1}\theta^{-1}\Sigma_x\rho_{xu}(n^{-1}u'u - \sigma_u^2)] + o_p(1).
\end{aligned} \tag{B.7}
$$

Note that the three explicit terms of (B.7) all have zero mean, are $O_p(1)$ and have a normal limiting distribution, according to our results of Appendix A. Hence, this expression has a limiting normal distribution too, say $[n^{-1/2}X'u - n^{1/2}\hat{\sigma}_u(\rho_{xu})S_x\rho_{xu}] \xrightarrow{d} \mathcal{N}(0, \sigma_u^2\Theta)$. Then, given (B.1), the limiting distribution of the inconsistency corrected OLS estimator is

$$
n^{1/2}[\hat{\beta}(\rho_{xu}) - \beta] \xrightarrow{d} \mathcal{N}(0, \sigma_u^2\Sigma_{xx}^{-1}\Theta\Sigma_{xx}^{-1}). \tag{B.8}
$$

So, in order to establish this we should obtain Θ, the variance of the sum of all $O_p(1)$ terms of the vector (B.7).

Employing the asymptotic variances and covariances derived in Appendix A we find, when specializing for the case $\kappa_u = 3$ and $\kappa_x = 3$,

$$
\begin{aligned}
\Theta = {} & (I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})(\Sigma_{xx} + \Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x) \\
& + 0.5(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x) \\
& + 0.5\theta^{-2}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x \\
& - (I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\Sigma_{xx}\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x) \\
& - (I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\mathcal{R}^2\Sigma_{xx}(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x) \\
& - \theta^{-1}(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x \\
& - \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x) \\
& + 0.5\theta^{-1}(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\mathcal{R}^2\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x \\
& + 0.5\theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x).
\end{aligned}
$$

First we shall evaluate this for the scalar case $K = 1$, where $\rho_{xu}^2 = 1 - \theta$. This yields $1 + \rho_{xu}^2/\theta = \theta^{-1}$, giving

$$
\begin{aligned}
\sigma_x^{-2}\Theta &= (1 + \rho_{xu}^2/\theta)^2(1 + \rho_{xu}^2) + 0.5[(1 + \rho_{xu}^2/\theta)^2\rho_{xu}^2 + \rho_{xu}^2/\theta^2] \\
&\qquad - 2[(1 + \rho_{xu}^2/\theta)^2\rho_{xu}^2 + \rho_{xu}^2(1 + \rho_{xu}^2/\theta)/\theta] + \rho_{xu}^4(1 + \rho_{xu}^2/\theta)/\theta \\
&= \theta^{-2}(1 + \rho_{xu}^2 + \rho_{xu}^2 - 2\rho_{xu}^2 - 2\rho_{xu}^2 + \rho_{xu}^4) \\
&= \theta^{-2}(1 - \rho_{xu}^2)^2 \\
&= 1,
\end{aligned}
$$

which establishes the proof of Corollary 1.1.

For the case $K > 1$ such an elegant result proves to be an illusion, generally speaking. Denoting $\Phi = \Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x$ and using $\Phi\Sigma_{xx}^{-1}\Phi = (1 - \theta)\Phi$, we find for the first term of Θ the expression

$$
\begin{aligned}
& (I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})(\Sigma_{xx} + \Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x) \\
& = (I + \theta^{-1}\Phi\Sigma_{xx}^{-1})(\Sigma_{xx} + \Phi)(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi) \\
& = \Sigma_{xx} + (\theta^{-1} + 2\theta^{-2})\Phi,
\end{aligned}
$$

and for the full expression

$$
\begin{aligned}
\Theta &= \Sigma_{xx} + (\theta^{-1} + 2\theta^{-2})\Phi \\
&\quad + 0.5(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi) + 0.5\theta^{-2}\Phi \\
&\quad - \Sigma_{xx}\mathcal{R}^2 - \theta^{-1}\Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi - \theta^{-1}\Phi\mathcal{R}^2 - \theta^{-2}\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi \\
&\quad - \mathcal{R}^2\Sigma_{xx} - \theta^{-1}\mathcal{R}^2\Phi - \theta^{-1}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx} - \theta^{-2}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi \\
&\quad - 2\theta^{-1}\Phi - 2\theta^{-2}\Phi\Sigma_{xx}^{-1}\Phi \\
&\quad + 0.5\theta^{-1}\mathcal{R}^2\Phi + 0.5\theta^{-2}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi + 0.5\theta^{-1}\Phi\mathcal{R}^2 + 0.5\theta^{-2}\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi \\
&= \Sigma_{xx} + [\theta^{-1} + 2\theta^{-2} + 0.5\theta^{-2} - 2\theta^{-1} - 2\theta^{-2}(1 - \theta)]\Phi + \\
&\quad + 0.5(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}\Sigma_x^{-1}(\Sigma_{xx} \circ \Sigma_{xx})\Sigma_x^{-1}\mathcal{R}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi) \\
&\quad - \Sigma_{xx}\mathcal{R}^2 - \theta^{-1}\Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi - 0.5\theta^{-1}\Phi\mathcal{R}^2 \\
&\quad - \mathcal{R}^2\Sigma_{xx} - 0.5\theta^{-1}\mathcal{R}^2\Phi - \theta^{-1}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx} - \theta^{-2}(\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\mathcal{R}^2\rho_{xu})\Phi \\
&= \Sigma_{xx} - \Sigma_{xx}\mathcal{R}^2 - \mathcal{R}^2\Sigma_{xx} - \theta^{-1}(\Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi + \Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx}) \\
&\quad - 0.5\theta^{-1}(\Phi\mathcal{R}^2 + \mathcal{R}^2\Phi) + [\theta^{-1} + \theta^{-2}(0.5 - \rho'_{xu}\mathcal{R}\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\mathcal{R}\rho_{xu})]\Phi \\
&\quad + 0.5(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}\Sigma_x^{-1}(\Sigma_{xx} \circ \Sigma_{xx})\Sigma_x^{-1}\mathcal{R}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi).
\end{aligned}
$$

C. Proof of Corollary 1.2

Evaluating Θ for the special case where $\rho_{xu} = (\rho_1, 0, ..., 0)'$ a simpler expression for the limiting variance is found, and again a very elegant solution for the coefficient of the endogenous regressor. Let e_1 be the $K \times 1$ vector with all its elements zero apart from the first one which is unity. Now, given this special case of ρ_{xu}, $\mathcal{R} = \rho_1 e_1 e_1'$ and $\Phi = \rho_1^2 \sigma_1^2 e_1 e_1'$. Denoting $e_1' \Sigma_{xx}^{-1} e_1 = \sigma^{11}$ we have $\theta = 1 - \rho_1^2 \sigma_1^2 \sigma^{11}$. Then we find $\Sigma_{xx} \mathcal{R}^2 \Sigma_{xx}^{-1} \Phi = \Sigma_{xx} \rho_1^2 e_1 e_1' \Sigma_{xx}^{-1} \rho_1^2 \sigma_1^2 e_1 e_1' = \rho_1^4 \sigma_1^2 \sigma^{11} \Sigma_{xx} e_1 e_1'$, $\Phi \mathcal{R}^2 = \rho_1^2 \sigma_1^2 e_1 e_1' \rho_1^2 e_1 e_1' = \rho_1^4 \sigma_1^2 e_1 e_1'$, $\rho_{xu}' \Sigma_x \Sigma_{xx}^{-1} \Sigma_x \mathcal{R}^2 \rho_{xu} = \rho_1^4 \sigma_1^2 \sigma^{11}$ and $(I + \theta^{-1} \Phi \Sigma_{xx}^{-1}) \mathcal{R} = (I + \theta^{-1} \rho_1^2 \sigma_1^2 e_1 e_1' \Sigma_{xx}^{-1}) \rho_1 e_1 e_1' = \theta^{-1}(\rho_1 \theta + \rho_1^3 \sigma_1^2 \sigma^{11}) e_1 e_1' = \theta^{-1} \rho_1 e_1 e_1'$. Substituting all these, we obtain

$$\begin{aligned}\Theta &= \Sigma_{xx} - \rho_1^2(\Sigma_{xx} e_1 e_1' + e_1 e_1' \Sigma_{xx}) - \theta^{-1} \rho_1^4 \sigma_1^2 \sigma^{11} (\Sigma_{xx} e_1 e_1' + e_1 e_1' \Sigma_{xx}) - \theta^{-1} \rho_1^4 \sigma_1^2 e_1 e_1' \\ &\quad + \theta^{-1}(1 + 0.5\theta^{-1}) \rho_1^2 \sigma_1^2 e_1 e_1' - \theta^{-2} \rho_1^6 \sigma_1^4 \sigma^{11} e_1 e_1' + 0.5 \theta^{-2} \rho_1^2 \sigma_1^2 e_1 e_1' \\ &= \Sigma_{xx} - \theta^{-1}(\rho_1^2 \theta + \rho_1^4 \sigma_1^2 \sigma^{11})(\Sigma_{xx} e_1 e_1' + e_1 e_1' \Sigma_{xx}) \\ &\quad + \theta^{-2}(-\theta \rho_1^2 + \theta + 1 - \rho_1^4 \sigma_1^2 \sigma^{11}) \rho_1^2 \sigma_1^2 e_1 e_1' \\ &= \Sigma_{xx} - \theta^{-1} \rho_1^2 (\Sigma_{xx} e_1 e_1' + e_1 e_1' \Sigma_{xx}) + \theta^{-2}(\theta + 1 - \rho_1^2) \rho_1^2 \sigma_1^2 e_1 e_1'.\end{aligned}$$

Hence, in this special case

$$\begin{aligned}V(\rho_{xu}) &= \Sigma_{xx}^{-1} \Theta \Sigma_{xx}^{-1} \\ &= \Sigma_{xx}^{-1} - \theta^{-1} \rho_1^2 (e_1 e_1' \Sigma_{xx}^{-1} + \Sigma_{xx}^{-1} e_1 e_1') + [\theta^{-1} + \theta^{-2}(1 - \rho_1^2)] \rho_1^2 \sigma_1^2 \Sigma_{xx}^{-1} e_1 e_1' \Sigma_{xx}^{-1}.\end{aligned} \tag{C.1}$$

We will now focus on the first element of estimator $\hat{\beta}(\rho_{xu})$. The OLS results $\hat{\beta}_1, \hat{u}$ and $\hat{\sigma}_u^2$ regarding β_1, u and σ_u^2 are all invariant under the model transformation given by $y = x_1 \beta_1 + X_2 \beta_2 + u = (M_2 + P_2) x_1 \beta_1 + X_2 \beta_2 + u = M_2 x_1 \beta_1 + X_2[(X_2' X_2)^{-1} X_2' \beta_1 + \beta_2] + u = x_1^* \beta_1 + X_2 \beta_2^* + u$, where $P_2 = X_2 (X_2' X_2)^{-1} X_2'$, $M_2 = I - P_2$ and $X_2' x_1^* = 0$. Therefore, in the model with regressors (x_1^*, X_2), $\sigma_1^2 = n^{-1} x_1^{*\prime} x_1^*$ and $\sigma^{11} = \sigma_1^{-2}$, giving $\theta = 1 - \rho_1^2 \sigma_1^2 \sigma^{11} = 1 - \rho_1^2$, we find for $V(\rho_{xu})$ of (C.1) that

$$\begin{aligned}e_1' V(\rho_{xu}) e_1 &= \sigma^{11} - 2\theta^{-1} \rho_1^2 \sigma^{11} + [\theta^{-1} + \theta^{-2}(1 - \rho_1^2)] \rho_1^2 \sigma_1^2 (\sigma^{11})^2 \\ &= \sigma^{11} - 2\theta^{-1} \rho_1^2 \sigma^{11} + 2\theta^{-1} \rho_1^2 \sigma^{11} = \sigma^{11},\end{aligned}$$

which is invariant with respect to ρ_{xu}.

D. Proof of Theorem 2

After taking the variables of the more general model in deviation from sample averages the resulting sample values of x_i are no longer independent and neither those of u_i. However, this does not affect the results obtained in Appendix A, as we shall show below, and therefore the proof of Theorem 1 in Appendix B is still valid.

We find $\bar{u} = n^{-1} \Sigma_{i=1}^n u_i \sim (0, n^{-1} \sigma_u^2)$, so $\bar{u} = O_p(n^{-1/2})$. Also $E(\bar{u}^2) = n^{-1} \sigma_u^2$ and with minor effort one finds $Var(\bar{u}^2) = O(n^{-2})$, hence $\bar{u}^2 - n^{-1} \sigma_u^2 = O_p(n^{-1})$, but also $\bar{u}^2 = O_p(n^{-1})$. Thus, replacing u_i by $u_i - \bar{u}$ in the left-hand side of (A.1) we find

$$n^{-1/2} \Sigma_{i=1}^n [(u_i - \bar{u})^2 - \sigma_u^2] = n^{-1/2} \Sigma_{i=1}^n (u_i^2 - \sigma_u^2) - n^{1/2} \bar{u}^2,$$

where $n^{-1/2}\Sigma_{i=1}^n(u_i^2 - \sigma_u^2) = O_p(1)$ and $n^{1/2}\bar{u}^2 = O_p(n^{-1/2})$. So the effect of taking deviations from the sample average is of smaller order and does not affect the limiting distribution, giving

$$n^{-1/2}\Sigma_{i=1}^n[(u_i - \bar{u})^2 - \sigma_u^2] \xrightarrow{d} \mathcal{N}[0, (\kappa_u - 1)\sigma_u^4].$$

The same holds regarding (A.2). From $\bar{x}^* = n^{-1}\Sigma_{i=1}^n x_i^* \sim (\mu, n^{-1}\Sigma_{xx})$ we find $\bar{x}^* - \mu = n^{-1}\Sigma_{i=1}^n(x_i^* - \mu) = O_p(n^{-1/2})$, thus

$$\begin{aligned} n^{-1/2}\Sigma_{i=1}^n[(x_i^* - \bar{x}^*)(u_i - \bar{u}) - \sigma_{xu}] &= n^{-1/2}\Sigma_{i=1}^n[(x_i^* - \mu)u_i - \sigma_{xu}] - \bar{u}n^{-1/2}\Sigma_{i=1}^n(x_i^* - \mu) \\ &\quad - (\bar{x}^* - \mu)n^{-1/2}\Sigma_{i=1}^n u_i + n^{1/2}(\bar{x}^* - \mu)\bar{u} \\ &= n^{-1/2}\Sigma_{i=1}^n[(x_i^* - \mu)u_i - \sigma_{xu}] + O_p(n^{-1/2}). \end{aligned}$$

This has the same limiting distribution as obtained earlier. The same can easily be established for (A.3) through (A.7). Hence, the findings of Theorem 1 are unaffected when X and u have actually been obtained by taking regressors X^* and distubances u in deviation form.