Instrument-free inference under confined regressor endogeneity; derivations and applications

Kiviet, Jan

University of Amsterdam

6 November 2019

Instrument-free inference under confined regressor endogeneity; derivations and applications

Jan F. Kiviet*

Version of 6 November 2019

JEL-Classifications: C12, C13, C21, C22, C26.

Keywords: *endogeneity robust inference, instrument validity tests, replication studies, sensitivity analysis, two-stage least-squares.*

Abstract

A fully-fledged alternative to Two-Stage Least-Squares (TSLS) inference is developed for general linear models with endogenous regressors. This alternative approach does not require the adoption of external instrumental variables. It generalizes earlier results which basically assumed all variables in the model to be normally distributed and their observational units to be stochastically independent. Now the chosen underlying framework corresponds completely to that of most empirical cross-section or time-series studies using TSLS. This enables revealing empirically relevant replication studies, also because the new technique allows testing the earlier untestable exclusion restrictions adopted when applying TSLS. For three illustrative case studies a new perspective on their empirical findings results. The new technique is computationally not very demanding. It involves scanning least-squares-based results over all compatible values of the nuisance parameters established by the correlations between regressors and disturbances.

1. Introduction

For rather specific models with endogenous regressors Kiviet (2019) develops an alternative approach that does not require the use of instrumental variables. Instead of strict orthogonality assumptions on instrumental variables and disturbances, it requires bounds on the possible nonorthogonality of regressors and disturbances. Then, as long as the actual endogeneity respects the specified bounds, asymptotically valid instrument-free inference on coefficients can be produced. In its derivations, however, it has been assumed that the sample observations on all variables involved are: (i) i.i.d. (independently and identically distributed), which excludes most time-series applications, and (ii) either normally distributed or have at least no excess kurtosis. In this study we

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

start-off from the much more general framework usually adopted when TSLS (two-stage least-squares) or IV (instrumental variables) are applied to either cross-section or time-series data. In that context we derive the limiting distribution of a least-squares-based consistent instrument-free estimator, which is however explicit in a nuisance parameter characterizing any endogeneity. Also for possibly time-dependent and nonnormal data this enables feasible and asymptotically valid –and, as is demonstrated, in finite samples remarkably accurate and efficient– instrument-free inference in static or dynamic linear regression models under mild regularity conditions similar to those justifying practitioners when applying TSLS.

Besides producing inference in its own right, these new techniques also allow a sensitivity analysis of standard (and non-standard weak-instrument) IV or TSLS inference, including a comprehensive check on the validity of instruments. This check is more stringent than previously available, because –in addition to validity of any over-identification restrictions– the just-identification restrictions can be verified too. In three replication studies the new techniques are applied to the data used in earlier instrument-based publications. This reveals that some of the assumptions on which these studies have been built seem doubtful. For all case studies examined a new perspective regarding their empirical findings results.

When regressors are endogenous they are correlated with the model errors, which may lead to serious bias of least-squares estimators, irrespective of the size of the sample. In such situations one usually reverts to applying method of moment estimators, which are built on the assumed orthogonality of so-called instrumental variables and the model errors. Such estimators may have two serious impediments, which are associated with the proclaimed validity and relevance of the employed instrumental variables; see, for instance, Bound, Jaeger and Baker (1995), Murray (2006, 2017), Kiviet and Niemczyk (2012), Andrews, Marmer and Yu (2019), Andrews, Stock and Sun (2019) and many of the further references in those studies. The validity or orthogonality of instruments and errors can only very partially be vindicated on the basis of statistical evidence; the major justification of instrument validity depends as a rule just on subject matter specific rhetoric verbal persuasion. Whereas external instruments can only be valid if they do not have a direct effect on the dependent variable, so their exclusion from the regression relationship should be true, at the same time in order to be relevant they should have a relatively strong indirect effect on the dependent variable through their association with the endogenous regressors. If this association is weak then method of moment estimators may be as seriously biased as least-squares estimators are, and they will also be harmed by having an unattractive large dispersion and possibly a highly nonnormal distribution.

For the alternative instrument-free inference methods validity and relevance of instruments are not an issue, self-evidently. Their primary impediment is actual credibility regarding the chosen range of likely values of the degree of endogeneity of the individual regressors. A narrow range may yield seemingly more efficient but also unmistakably less credible inference; wide ranges will be more credible but will as a rule result in less pronounced statistical conclusions, as our applications will illustrate. These applications indicate that the new techniques provide a useful sensitivity analysis of instrument-based findings, revealing any vulnerability regarding presumptuous orthogonality conditions. Always, however, they will also provide very attractive autonomous alternative inferences, because, even for rather wide intervals for the endogeneity correlations, resulting instrument-free confidence intervals are usually narrower. Especially narrower than those constructed by weak-instrument robust methods, whereas their confidence coefficients are much more trustworthy than those of the standard instrument-based methods.

Identification of the parameters of single equations, or of the complete system to which they belong, has usually been obtained by exploiting normalization and exclusion restrictions or more general coefficient restrictions, see Koopmans, Rubin and Leipnik (1950) and Fisher (1959). Achieving identification by exploiting restrictions on the covariances of the disturbances has been introduced by Fisher (1963) and extended by Wegge (1965); more recently it has been specialized to exploiting heteroskedasticity for identification, see Lewbel (2012). In the approach developed here identification of a single structural equation is based on restricting yet other parameters, namely the correlations between regressors and disturbances. At first sight this may seem odd, because in current practice the actual sign and magnitude of these correlations are usually disregarded, except for the case of them being zero or not. Simulations in Kiviet and Niemczyk (2012) show, however, that these correlations are nuisance parameters which may seriously distort the finite sample distribution of TSLS based estimators and test statistics. Therefore, and because TSLS estimators are built on statistically unverifiable preconditions, statistical inference on the actual value of these endogeneity correlations seems mostly highly unreliable. However, below we will indicate that in many practical situations the theoretical arguments used to suggest a particular model specification implicitly entail assumptions on the sign and likely magnitude of endogeneity correlations. Moreover, our procedures do not require assumptions on the true values of these correlations, but just to specify intervals which should enclose these true values.

In the next section we first review three basic empirically relevant situations which may give rise to endogeneity of regressors. For all three it is also shown that, in the rather general context of relationships that can be parameterized linearly, endogenous regressors can be decomposed into two mutually uncorrelated components, where only one is proportional to the model error. These decompositions facilitate to make credible assumptions on the likely sign and magnitude of any endogeneity of regressors. They are also used in the derivation of the asymptotic validity of the alternative instrument-free inference methods. Because these derivations are rather cumbersome for a model with an arbitrary number of regressors, from which probably more than one is endogenous, Section 3 first considers the simple model with just one regressor for which all matrix algebra can be avoided. This regressor may be endogenous, nonnormal, and also dependent on its own past, as is often the case for regressors in time-series relationships. This oversimplified model provides a helpful stepping stone towards the presentation of the result in Section 4 for single linear multiple regression models with some endogenous explanatories. The technical derivations of the results presented in Sections 3 and 4 can be found in appendices. Section 5 provides some simulation evidence on the accuracy of the proposed methods in finite samples. Those who are primarily interested in the actual practical achievements of the new approach may immediately jump to Section 6 which contains three empirical replication studies, where standard and non-standard instrumental variable based inferences are supplemented with instrument-free results. The latter reveal frailties in and provide alternatives to the earlier findings. Finally, Section 7 concludes.

2. Endogenous regressors in linear regression models

Especially when relationships are modeled under specification uncertainty or on basis of so-called observational data (these do not stem from controlled experiments), as is usually the case in social science and especially in economics and business, explanatory variables may be contemporaneously correlated with the model error (the random disturbance term). Like the dependent or explained variable (regressand), which is unavoidably contemporaneously correlated with the disturbances, such regressors are labeled endogenous. There are three fundamental sources for endogeneity of regressors, namely: (i) simultaneity, (ii) errors in explanatories, and (iii) wrongly omitted explanatories. One may argue that a fourth possibility is joint occurrence of autoregressive disturbances and lagged dependent variable regressors in a time-series regression. However, such endogeneity could be resolved in principle by including in the regression further lags of all regressors. So, in essence, this case is already covered by (iii).

In the context of linear regression models these three separate sources of endogeneity of regressors are characterized by situations which in rather basic form can be represented as indicated below. These expositions serve to demonstrate that endogenous regressors can in theory always be decomposed into two contemporaneously uncorrelated additive components, where one is endogenous and the other predetermined or exogenous. This decomposition will prove to be helpful in the further derivations and also when it comes to making an assessment of the likely sign and magnitude of regressor endogeneity. For the sake of simplicity, we suppose for the moment that all variables have zero mean and have finite and constant (co)variance over the sample observations. The latter are indexed by $i = 1, ..., n$.

In case (i), simultaneity (or reciprocal causality) the equation of primary interest, here assumed to have just one endogenous explanatory variable $x_i^{(1)}$ next to an arbitrary number of exogenous regressors in vector $x_i^{(2)}$, is given by

$$y_i = \beta_1 x_i^{(1)} + \beta_2' x_i^{(2)} + u_i, \tag{2.1}$$

where u_i is the random disturbance term. This equation is assumed to be part of a larger system. Another relationship from this system may be given by, say,

$$x_i^{(1)} = \gamma_0 y_i + \gamma_3' x_i^{(3)} + v_i, \tag{2.2}$$

where vectors $x_i^{(2)}$ and $x_i^{(3)}$ may have some elements in common. The disturbances u_i and v_i could be uncorrelated, but possibly $v_i = \psi u_i + v_i^*$ with $E(v_i^* \mid u_i) = 0$ and $\psi \neq 0$. Substituting (2.1) into (2.2) yields

$$x_i^{(1)} = (1 - \gamma_0\beta_1)^{-1}(\gamma_0\beta_2' x_i^{(2)} + \gamma_3' x_i^{(3)} + \gamma_0 u_i + v_i). \tag{2.3}$$

This shows that regressor $x_i^{(1)}$ is endogenous in (2.1), because generally $E(x_i^{(1)} u_i) \neq 0$. Focussing on the simple case where $x_i^{(3)}$ is exogenous with respect to u_i too, one finds $E(x_i^{(1)} u_i) = (1 - \gamma_0\beta_1)^{-1}(\gamma_0 + \psi)E(u_i^2)$. This is zero only when $\psi = -\gamma_0$. In cases where it can be argued that $1 - \gamma_0\beta_1 > 0$ the sign of the endogeneity (correlation between $x_1^{(1)}$ and u_i) is similar to that of $\gamma_0 + \psi$.

In case (ii), errors in explanatories, the situation may be as follows. Let the true data generating process (DGP) be represented by

$$y_i = \beta_1 x_i^* + \beta_2' x_i^{(2)} + \varepsilon_i, \tag{2.4}$$

where none of the regressors is endogenous, because $E(x_i^* \varepsilon_i) = 0$ and $E(x_i^{(2)} \varepsilon_i) = 0$. However, estimating this equation is unfeasible, because scalar variable x_i^* has only been observed with (measurement) errors. We consider the very simple case where one has observed the proxy $x_i^{(1)}$ for x_i^*, where

$$x_i^{(1)} = x_i^* + \eta_i, \tag{2.5}$$

for which we assume $E(\eta_i \mid x_i^*) = 0$ and $E(\varepsilon_i \mid \eta_i) = 0$. Substitution yields the feasible regression model

$$y_i = \beta_1 x_i^{(1)} + \beta_2' x_i^{(2)} + u_i, \tag{2.6}$$

with disturbance $u_i = \varepsilon_i - \beta_1 \eta_i$. Since $E(x_i^{(1)} u_i) = -\beta_1 E(\eta_i^2) \neq 0$ regressor $x_i^{(1)}$ is endogenous in (2.6), unless $\beta_1 = 0$ or $x_i^{(1)} \equiv x_i^*$ for all i. Sign and magnitude of the endogeneity are determined by $corr(x_i^{(1)}, u_i) = -\beta_1 \left| \sigma_\eta / \sigma_{x^{(1)}} \right|$.

In case (iii), wrongly omitted explanatories, it is assumed that the DGP is now given by

$$y_i = \beta_1' x_i^{(1)} + \beta_2 x_i^{(2)} + \varepsilon_i, \tag{2.7}$$

with $E(x_i^{(1)} \varepsilon_i) = 0$ and $E(x_i^{(2)} \varepsilon_i) = 0$, but that single regressor $x_i^{(2)}$ is not available, or has not been included in the regression for other reasons, so one uses the underspecified model

$$y_i = \beta_*' x_i^{(1)} + u_i \tag{2.8}$$

in an attempt to estimate vector β_1. Let us assume, in line with all the other linearity assumptions made here, that

$$E(x_i^{(2)} \mid x_i^{(1)}) = \phi' x_i^{(1)}. \tag{2.9}$$

Hence,

$$x_i^{(2)} = \phi' x_i^{(1)} + \omega_i, \tag{2.10}$$

with $E(\omega_i \mid x_i^{(1)}) = 0$, thus $E(\omega_i) = 0$ and $E(x_i^{(1)} \omega_i) = E[E(x_i^{(1)} \omega_i \mid x_i^{(1)})] = 0$. Substituting (2.10) into (2.7) yields

$$y_i = (\beta_1' + \beta_2 \phi') x_i^{(1)} + \beta_2 \omega_i + \varepsilon_i,$$

so in terms of (2.8) this suggests $\beta_* = \beta_1 + \beta_2 \phi$ with $u_i = \beta_2 \omega_i + \varepsilon_i$, where $E(x_i^{(1)} u_i) = \beta_2 E(x_i^{(1)} \omega_i) + E(x_i^{(1)} \varepsilon_i) = 0$. Hence, regressor $x_i^{(1)}$ turns out to be not endogenous in (2.8) for estimating $\beta_1 + \beta_2 \phi$.

However, forcing the interpretation $\beta_* = \beta_1$ results in $u_i = \beta_2 \phi' x_i^{(1)} + \beta_2 \omega_i + \varepsilon_i = \beta_2 x_i^{(2)} + \varepsilon_i$, which leads to $E[x_i^{(1)}(\beta_2 x_i^{(2)} + \varepsilon_i)] = \beta_2 \phi E(x_i^{(1)})^2 \neq 0$, unless $\beta_2 = 0$ (then regressor $x_i^{(2)}$ is not wrongly omitted but redundant) or $\phi = 0$ (implying orthogonality of regressors $x_i^{(1)}$ and $x_i^{(2)}$). If $\beta_2 \neq 0$ ($x_i^{(2)}$ is wrongly omitted) and $\phi \neq 0$ (the elements of $x_i^{(1)}$ which correspond to nonzero elements of ϕ are related to $x_i^{(2)}$) then $x_i^{(2)}$ is called a confounder. Sign and magnitude of the resulting endogeneity of elements of $x_i^{(2)}$ are determined by vector $\beta_2\phi\,|\sigma_{x^{(1)}}/\sigma_u|$.

In the majority of practical cases (see Young, 2019) investigators estimate models with just one endogenous regressor. Even then the endogeneity may be due to a combination of the three basic situations sketched above, so that it may not be self-evident what the sign and actual magnitude of the endogeneity may be. Things get certainly more complex when more than one regressor is affected by reciprocal causality or measurement errors and certainly when more than one explanatory has been wrongly omitted, especially when more than one of the included regressors are correlated with the omitted ones. Nevertheless, as we will demonstrate in our illustrations, many applied studies could benefit tremendously by paying more attention to the likely values of the elements of vector ρ_{xu}.

In all three cases that give rise to endogeneity of (elements of) regressor $x_i^{(1)}$ in the model with disturbance u_i we find that $x_i^{(1)}$ can be decomposed into two additive components, namely

$$x_i^{(1)} = \xi_i + \lambda u_i, \tag{2.11}$$

with $E(u_i \mid \xi_i) = 0$. In case of simultaneity (2.3) implies $\xi_i = (1-\gamma_0\beta_1)^{-1}(\gamma_0\beta_2' x_i^{(2)} + \gamma_3' x_i^{(3)} + v_i)$ and $\lambda = (\gamma_0+\psi)/(1-\gamma_0\beta_1)$. Under errors in explanatories we find from $E[(x_i^{(1)} - \lambda u_i)u_i] = 0$ that $\lambda = -\beta_1(\sigma_\eta^2/\sigma_u^2)$ and $\xi_i = x_i^{(1)} + \beta_1(\sigma_\eta^2/\sigma_u^2)u_i$. And under omitted variables we obtain in a similar way $\lambda = E[x_i^{(1)}(\beta_2\phi' x_i^{(1)} + \beta_2\omega_i + \varepsilon_i)]/E(u_i^2) = \beta_2 E(x_i^{(1)} x_i^{(1)\prime})\phi/\sigma_u^2$ and $\xi_i = x_i^{(1)} - \lambda u_i$.

The above suggests (and it can be formally proved) that in a single linear regression model

$$y_i = x_i'\beta + u_i, \tag{2.12}$$

with K possibly endogenous regressors collected in the $K \times 1$ vector x_i, we will have

$$x_i = \xi_i + \lambda u_i, \tag{2.13}$$

where ξ_i and λ are both $K \times 1$ vectors now, with $E(u_i \mid \xi_i) = 0$, hence $E(\xi_i u_i) = 0$ and $E(x_i u_i) = \lambda\sigma_u^2$. Denoting $E(x_{ik}^2) = \sigma_k^2$ and $\rho_k = E(x_{ik}u_i)/(\sigma_k\sigma_u)$, we find

$$\lambda_k = E(x_{ik}u_i)/\sigma_u^2 = \rho_k\,|\sigma_k/\sigma_u| \text{ for } k = 1, ..., K. \tag{2.14}$$

The OLS (ordinary least-squares) estimator for β, given by $\hat\beta_{OLS} = (X'X)^{-1}X'y$, where $X = (x_1, ..., x_n)'$ is an $n \times K$ matrix and $y = (y_1, ..., y_n)'$ an $n \times 1$ vector, has (invoking the law of large numbers) probability limit given by

$$\begin{aligned}\text{plim}\,\hat\beta_{OLS} &= \beta + [\text{plim}\, n^{-1}X'X]^{-1}\,\text{plim}\, n^{-1}X'u \\ &= \beta + \sigma_u\Sigma_{xx}^{-1}\Sigma_x\rho_{xu},\end{aligned} \tag{2.15}$$

where $(\Sigma_{xx})_{jk} = \sigma_{jk} = E(x_{ij}x_{ik})$, $\sigma_k^2 = \sigma_{kk}$, $\Sigma_x = \text{diag}(|\sigma_1|, ..., |\sigma_K|)$ and $\rho_{xu} = (\rho_1, ..., \rho_K)'$, giving $\lambda = \sigma_u^{-1}\Sigma_x\rho_{xu}$. Hence, in general, each element of $\hat\beta_{OLS}$ is inconsistent (thus biased, irrespective of the size of the sample) if any element of ρ_{xu} (or of λ) is nonzero. Such a nonzero element undermines the moment $E(x_i u_i) = \sigma_u^2\lambda = \sigma_u\Sigma_x\rho_{xu}$ to establish a valid orthogonality condition.

3. Instrument-free inference in a very simple model

The approach to and full proof of instrument-free and fairly robust inference in general linear regression models with some possibly endogenous explanatory variables will be introduced here first for a simple relationship with just one zero-mean regressor, denoted as

$$y_i = \beta x_i + u_i. \tag{3.1}$$

Disturbance $u_i \sim (0, \sigma_u^2)$ is assumed homoskedastic and serially uncorrelated, hence $E(u_i u_t) = 0$ for $i \neq t = 1, ..., n$. Earlier we focussed on i.i.d. cross-section samples, see Kiviet (2013, 2016). Now we want to cover time-series regressions with forms of dependence between the sample observations as well. Therefore we assume

$$E(x_i u_t) = 0 \text{ for } 1 \leq i < t \leq n, \tag{3.2}$$

but allow

$$E(x_i u_t) \neq 0 \text{ for } 1 \leq t \leq i \leq n. \tag{3.3}$$

Hence, although the regressor could be exogenous, namely when $E(x_i u_t) = 0 \; \forall i, t$, or predetermined when $E(x_i u_t) = 0 \; \forall i \leq t$, it could also be endogenous. In fact, we will assume that

$$E(u_i \mid x_{i-1}, ..., x_1, u_{i-1}, ..., u_1) = 0 \text{ for } i \geq 2. \tag{3.4}$$

Then u_i is called an innovation with respect to its own past and that of x_i, whereas x_i could depend on current and past u_i and on past x_i. So, the sample observations are not necessarily independent. For the sake of simplicity, though, we will assume them to be identically distributed. Note that assumption (3.4) easily matches with i.i.d. cross-section applications, and also with time-series regressions when the endogeneity stems from simultaneity or from errors in regressors, provided u_i, ε_i and η_i are serially uncorrelated indeed. However, in case of wrongly omitted time-series regressors the assumption that $u_i = \beta_2 x_i^{(2)} + \varepsilon_i$ is serially uncorrelated would require that omitted regressor $x_i^{(2)}$ is serially uncorrelated too, which will not be the case in many empirical time-series applications.

In line with Section 2 we assume that the scalar regressor x_i can be decomposed as

$$x_i = \xi_i + \lambda u_i \sim (0, \sigma_x^2), \tag{3.5}$$

with $\sigma_x > 0$, λ nonrandom, and where $\xi_i \sim (0, \sigma_\xi^2)$. Moreover, $E(u_i \mid \xi_i, ..., \xi_1) = 0$, so component ξ_i is predetermined but could in fact be strictly exogenous.

The endogeneity of the regressor can be expressed by the constant correlation

$$\rho_{xu} = \lambda \sigma_u / \sigma_x. \tag{3.6}$$

Because $\sigma_x^2 = \sigma_\xi^2 + \lambda^2 \sigma_u^2 = \sigma_\xi^2 + \rho_{xu}^2 \sigma_x^2$ we have $\sigma_\xi^2 = (1 - \rho_{xu}^2)\sigma_x^2$.

In this one-regressor model[1]

$$\hat{\beta}_{OLS} = (\Sigma x_i^2)^{-1} \Sigma x_i y_i = \beta + (\Sigma x_i^2)^{-1} \Sigma x_i u_i, \tag{3.7}$$

[1] All summations Σ that follow are over the range $i = 1, ..., n$.

and (2.15) specializes into

$$\hat{\beta}_{OLS} = \beta + n^{-1}\Sigma x_i u_i/(n^{-1}\Sigma x_i^2) \rightarrow \beta + \rho_{xu}\sigma_u/\sigma_x. \tag{3.8}$$

where $\rightarrow$ indicates convergence in probability. Hence, $\hat{\beta}_{OLS}$ is inconsistent when the degree of endogeneity ρ_{xu} is nonzero.

Assuming for the moment that ρ_{xu} were known, then a consistent estimator of β could be obtained, if we can find consistent estimators for σ_u and σ_x as well. Since

$$\hat{\sigma}_x^2 = n^{-1}\Sigma x_i^2 \rightarrow \sigma_x^2 \tag{3.9}$$

we have $\hat{\sigma}_x \rightarrow \sigma_x$. From $\hat{u}_i = y_i - x_i\hat{\beta}_{OLS} = u_i - x_i(\hat{\beta}_{OLS} - \beta) = u_i - x_i\Sigma x_i u_i/\Sigma x_i^2$, we find $\Sigma\hat{u}_i^2 = \Sigma u_i^2 - (\Sigma x_i u_i)^2/\Sigma x_i^2$, thus

$$n^{-1}\Sigma\hat{u}_i^2 = n^{-1}\Sigma u_i^2 - (n^{-1}\Sigma x_i u_i)^2/n^{-1}\Sigma x_i^2 \rightarrow \sigma_u^2 - \rho_{xu}^2\sigma_u^2 = \sigma_u^2(1-\rho_{xu}^2),$$

so

$$\hat{\sigma}_u^2(\rho_{xu}) = (1-\rho_{xu}^2)^{-1}n^{-1}\Sigma\hat{u}_i^2 \rightarrow \sigma_u^2, \tag{3.10}$$

giving $\hat{\sigma}_u(\rho_{xu}) \rightarrow \sigma_u$. From these we obtain what we called in previous studies the kinky least-squares (KLS) estimator

$$\hat{\beta}_{KLS}(\rho_{xu}) = \hat{\beta}_{OLS} - \rho_{xu}\hat{\sigma}_u(\rho_{xu})/\hat{\sigma}_x \rightarrow \beta, \tag{3.11}$$

which is consistent, although unfeasible, unless ρ_{xu} is really known.

If we obtain the limiting distribution of $\hat{\beta}_{KLS}(\rho_{xu})$ as a function of ρ_{xu}, then we can construct for any hypothesis on scalar β its studentized statistic. Next, scanning the p-value of this statistic by taking for ρ_{xu} a dense grid of real values r in the interval $[r^L, r^U]$ such that $-1 < r^L \leq r \leq r^U < 1$, we can produce inference on β which is robust to endogeneity, provided

$$r^L \leq \rho_{xu} \leq r^U \tag{3.12}$$

indeed.

In Appendix A we derive that the distribution of $\hat{\beta}_{KLS}(\rho_{xu})$, in case $E(u_i^4) = \kappa_u\sigma_u^4$ and $E(x_i^4) = \kappa_x\sigma_x^4$, can be approximated by a normal distribution centered at β with a variance that can be estimated consistently by

$$\widehat{Var}[\hat{\beta}_{KLS}(\rho_{xu})] = \theta(\rho_{xu},\kappa_u,\kappa_x)\frac{\hat{\sigma}_u^2(\rho_{xu})}{\Sigma x_i^2}, \tag{3.13}$$

$$\text{with } \theta(\rho_{xu},\kappa_u,\kappa_x) = \frac{4 + (\kappa_u+\kappa_x-14)\rho_{xu}^2 - 2(\kappa_u-5)\rho_{xu}^4}{4(1-\rho_{xu}^2)^2}.$$

Note that this expression is invariant regarding the sign of ρ_{xu}. Skewness of the series x_i and u_i does not have an effect, and neither do their fifth and higher-order moments. When $\kappa_x = \kappa_u = 3$ (and thus also under normality) we find that $\theta(\rho_{xu}, 3, 3) = 1$, giving $\widehat{Var}[\hat{\beta}_{KLS}(\rho_{xu})] = \hat{\sigma}_u^2(\rho_{xu})/\Sigma x_i^2$. When $\rho_{xu} = 0$ the KLS estimator specializes to $\hat{\beta}_{OLS}$, which has limiting distribution $\mathcal{N}(0, \sigma_u^2/\sigma_x^2)$, irrespective of the third and higher-order moments of the data. Since we should find $\widehat{Var}[\hat{\beta}_{KLS}(0)] = \widehat{Var}(\hat{\beta}_{OLS})$, it makes sense to replace in (3.13) $\hat{\sigma}_u^2(\rho_{xu})$ by the asymptotically equivalent though degrees of freedom corrected expression

$$s_u^2(\rho_{xu}) = \frac{1}{(1-\rho_{xu}^2)}\frac{\Sigma(y_i - x_i'\hat{\beta}_{OLS})^2}{n-K}. \tag{3.14}$$

Variance formula (3.13) makes clear that for pretty small absolute values of ρ_{xu} the variance is not very much affected by how much κ_u and κ_x differ from 3 (called their excess kurtosis).[2] For both κ_x and κ_u smaller than 10 and $|\rho_{xu}| \leq 0.3$ factor $\theta(\rho_{xu}, \kappa_u, \kappa_x)$ is smaller than 1.35, and for $|\rho_{xu}| \leq 0.5$ it does not exceed 2.5, giving a multiplicative boost to the KLS standard error under zero excess kurtosis of 1.16 and 1.58 respectively.

To test hypotheses about β, in addition to bounds (3.12), we should either know κ_u and κ_x (which seems unlikely) or use the consistent estimators

$$\left.\begin{aligned} \hat{\kappa}_u(\rho_{xu}) = n^{-1}\Sigma[y_i - x_i\hat{\beta}_{KLS}(\rho_{xu})]^4/\hat{\sigma}_u^4(\rho_{xu}), \\ \hat{\kappa}_x = n^{-1}\Sigma x_i^4/\hat{\sigma}_x^4, \end{aligned}\right\} \tag{3.15}$$

although these may require pretty large samples in order to be reasonably accurate.

When estimating model (3.1) by IV the strongest possible valid though unfeasible instrument would obviously be variable ξ_i. Its strength is given by $Corr(\xi_i, x_i) = \sigma_\xi/\sigma_x = (1-\rho_{xu}^2)^{1/2}$. Hence, the more serious the endogeneity is, the weaker even the strongest possible instrument will be. And on the other hand: when a valid instrument is really very strong, this implies that the endogeneity cannot be very substantial at the same time. The variance of the limiting distribution of $\hat{\beta}_{IV(\xi)} = \Sigma(\xi_i y_i)/\Sigma(\xi_i x_i)$ is $\sigma_u^2/\sigma_\xi^2 = (1-\rho_{xu}^2)^{-1}\sigma_u^2/\sigma_x^2$, whereas for KLS this is the much more attractive σ_u^2/σ_x^2, provided $\kappa_u = \kappa_x = 3$. It can easily be derived that in the simple one-regressor model, only for substantial excess kurtosis and limited endogeneity, unfeasible but most efficient IV can be more efficient than unfeasible KLS, namely when $\kappa_u + \kappa_\xi > 10$ and $\rho_{xu}^2 < 1 - 4/(\kappa_u + \kappa_\xi - 6) < 1$.

4. Instrument-free inference for more general linear models

In this section we present the major result on which instrument-free inference in linear (dynamic) regressions can be based. Its proof can be found in Appendix C, which uses some basic underlying derivations collected in Appendix B. The notation used and the assumptions made regarding the distribution of vector $(x_i', u_i)'$ for general linear regression model (2.12) are as follows.

KLS Assumptions:
(a) First and second moments: The vectors $\{(x_i', u_i)'; i = 1, ..., n\}$ *are identically (but not necessarily independently) distributed with zero mean and the second moments* $E(x_i x_i') = \Sigma_{xx}$, $E(u_i^2) = \sigma_u^2$ *and* $E(x_i u_i) = \sigma_{xu}$ *are all finite. Scalar* σ_{jk} *denotes the typical element of* Σ_{xx} *and* $\sigma_j = \text{abs}(\sigma_{jj}^{1/2})$ *for* $j, k = 1, ..., K$ *with* $\Sigma_x = \text{diag}(\sigma_1, ..., \sigma_K)$, *hence* σ_{xu} has typical element $\rho_j\sigma_j\sigma_u$, where ρ_j is *the typical element of vector* $\rho_{xu} = \Sigma_x^{-1}\sigma_{xu}/\sigma_u$*;*
(b) Fourth moments: $E(u_i^4) = \kappa_u\sigma_u^4$ *and* $E(x_{ik}^4) = \kappa_x\sigma_k^4$ *for* $k = 1, ..., K$, *where* κ_u *and* κ_x *are both finite and not smaller than unity;*
(c) Time dependence: As $E(u_i u_t) = 0$ *and* $E(x_i u_t) = 0$ *for* $t > i = 1, ..., n$ *and arbitrary otherwise, the disturbances are serially uncorrelated and individual regressors may be either exogenous, predetermined or endogenous.*

[2] Some benchmarks: A Student distribution with 5 degrees of freedom has kurtosis 9, a χ^2 distribution with 3 degrees of freedom has kurtosis 7, and the uniform distribution has kurtosis 1.8.

That all regressors have zero mean is helpful in the proof. As argued in Theorem 2 of Kiviet (2019) the findings will also apply to models with nonzero mean regressors that include an intercept. Also the assumption that all K regressors have the same kurtosis parameter κ_x is convenient in the proof. Its consequences, which will be shown to be minor, will be discussed later.

Further notation:
The sample equivalents of Σ_{xx}, Σ_x^2 *and* Σ_x *are given by* $S_{xx} = n^{-1}\Sigma_{i=1}^n x_i x_i'$, *by* S_x^2 *(the matrix just containing the main diagonal of* S_{xx}*), and by the positive definite diagonal matrix* S_x *(for which* $S_x S_x = S_x^2$*) respectively. By* R *we denote the diagonal* $K \times K$ *matrix with the elements of* ρ_{xu} *on its main diagonal.*

All results to follow are in terms of the unknown parameter ρ_{xu}, and some are also in terms of the unknown parameters κ_u and κ_x. Vector ρ_{xu} can only be estimated consistently if at least K valid and reasonably strong instrumental variables are available, and additionally the sample has to be pretty large for such a consistent estimator of ρ_{xu} to be reasonably accurate. The mere fact that instrument validity remains always doubtful from a statistical point of view forms the major motivation for developing the present instrument-free approach. This will be based on producing inference which is robust regarding regressor endogeneity (and also regarding κ_u and κ_x) provided the true ρ_{xu} vector belongs to a particular chosen set. Choosing the latter set is a matter of exploiting expert knowledge on the subject matter, like it is when adopting instrumental variables. The major difference is, however, that the corresponding set associated with the validity of instruments has measure zero, whereas the set regarding ρ_{xu} can be chosen as wide or narrow as one finds credible.

The limiting distribution in the following theorem establishes our major result.

KLS Theorem:
Under the above KLS Assumptions estimator $\hat{\beta}_{KLS}(\rho_{xu}) = \hat{\beta}_{OLS} - \hat{\sigma}_u(\rho_{xu}) S_{xx}^{-1} S_x \rho_{xu}$, *where* $\hat{\sigma}_u^2(\rho_{xu}) = \hat{\sigma}_{u,OLS}^2/(1 - \rho_{xu}' S_x S_{xx}^{-1} S_x \rho_{xu})$ *and* $\hat{\sigma}_{u,OLS}^2 = n^{-1}\Sigma_{i=1}^n (y_i - x_i'\hat{\beta}_{OLS})^2$, *has limiting distribution*

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

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

$$\begin{aligned}\Theta = \Sigma_{xx} &- (\Sigma_{xx}\mathcal{R}^2 + \mathcal{R}^2\Sigma_{xx}) + \theta^{-1}(\Phi - \Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi - \Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx}) \\ &- 0.25(\kappa_u - 1)\theta^{-1}[\mathcal{R}^2\Phi + \Phi\mathcal{R}^2 + \theta^{-1}(1 - 2\rho_{xu}'\mathcal{R}\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\mathcal{R}\rho_{xu})\Phi] \\ &+ 0.25(\kappa_x - 1)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi),\end{aligned}$$

which uses $\theta = 1 - \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\rho_{xu} > 0$ *and* $\Phi = \Sigma_x\rho_{xu}\rho_{xu}'\Sigma_x$.

This theorem for dependent data specializes for $\kappa_u = \kappa_x = 3$ to Theorem 1 of Kiviet (2019), which (apparently superfluously) supposed in its proof i.i.d. data. From

$$\text{plim}\, \hat{\sigma}_{u,OLS}^2 = \text{plim}\, n^{-1}u'[I - X(X'X)^{-1}X']u = \sigma_u^2(1 - \rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\rho_{xu})$$

it follows that $\theta > 0$ and $\rho_{xu}'\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\rho_{xu} < 1$. Hence, the values of the non-zero elements of ρ_{xu} are confined to an ellipsoid included in a unit sphere. Note, though, that $\hat{\beta}_{KLS}(\rho_{xu})$

only exists for ρ_{xu} vectors obeying $\rho_{xu}' S_x S_{xx}^{-1} S_x \rho_{xu} < 1$. For $K = 1$ this simply implies choosing scalar $\rho_{xu}^2 < 1$. However, for $K > 1$ values close to 1 for the absolute value of elements of vector ρ_{xu} may be unfeasible.

In case there is just one endogenous regressor x_1 in the first column of the regressor matrix $X = (x_1, X_2)$, with the first element of ρ_{xu} equal to ρ_1, then it follows from the theorem that its KLS estimate is simply given by

$$\begin{aligned}\hat{\beta}_{1,KLS}(\rho_1) &= \hat{\beta}_{1,OLS} - \rho_1 \hat{\sigma}_{u,OLS}(1 - \rho_1^2 e_1' S_x S_{xx}^{-1} S_x e_1)^{-1/2} e_1' S_{xx}^{-1} e_1 (n^{-1} \Sigma x_{i1}^2)^{1/2} \\ &= \hat{\beta}_{1,OLS} - \rho_1 [f_1/(1 - f_1 \rho_1^2)]^{1/2} n^{1/2} SE(\hat{\beta}_{1,OLS}). \end{aligned} \quad (4.1)$$

Here $SE(\hat{\beta}_{1,OLS}) = \hat{\sigma}_{u,OLS}[e_1'(X'X)^{-1}e_1]^{1/2}$ is the usual (but when $\rho_1 \neq 0$ naive) estimate for the standard deviation of $\hat{\beta}_{1,OLS}$. Factor

$$f_1 = e_1'(X'X)^{-1} e_1 \Sigma x_{i1}^2 \geq 1 \quad (4.2)$$

is also known (when $\rho_1 = 0$) as the 'variance inflation factor': the ratio of $Var(\hat{\beta}_{1,OLS})$ and its hypothetical value if all regressors X_2 happened to be orthogonal to x_1. In deviation from the $K = 1$ case, scalar estimator $\hat{\beta}_{1,KLS}(\rho_1)$ is now only defined for

$$\rho_1^2 < 1/f_1 \leq 1.$$

An asymptotically valid estimator of the variance of (4.1), derived in Appendix D, is

$$\widehat{Var}[\hat{\beta}_{1,KLS}(\rho_1)] = s_u^2(\rho_1) \frac{4 - 8\rho_1^2 + (\kappa_u + \kappa_x - 6)\rho_1^2 f_1 - 2(\kappa_u - 5)\rho_1^4 f_1^2}{4(1 - \rho_1^2 f_1)^2} \frac{f_1}{\Sigma_i x_{i1}^2}. \quad (4.3)$$

This variance increases with κ_x. And because $\rho_1^2 f_1 - 2\rho_1^4 f_1^2 = \rho_1^2 f_1 (1 - 2\rho_1^2 f_1)$ we note that it increases with κ_u only if $\rho_1^2 < 0.5/f_1$ and decreases when $\rho_1^2 > 0.5/f_1$. For the case with an arbitrary number of endogenous regressors the variance of the KLS coefficient estimator can be estimated (asymptotically conservative) by

$$\widehat{Var}[\hat{\beta}_{KLS}(\rho_{xu})] = n s_u^2(\rho_{xu}) S_{xx}^{-1} \hat{\Theta} S_{xx}^{-1}, \quad (4.4)$$

where $\hat{\Theta}$ is obtained by replacing in the expression for Θ given in the theorem Σ_{xx} and Σ_x by S_{xx} and S_x respectively, and κ_u and κ_x by

$$\left.\begin{aligned} \hat{\kappa}_u(\rho_{xu}) &= n^{-1} \Sigma [y_i - x_i' \hat{\beta}_{KLS}(\rho_{xu})]^4 / \hat{\sigma}_u^4(\rho_{xu}), \\ \hat{\kappa}_x &= \max_{j=1,\dots,K} n^{-1} \Sigma x_{ij}^4 / \hat{\sigma}_{x_j}^4. \end{aligned}\right\} \quad (4.5)$$

The Schur Theorem on Hadamard products implies that the contribution to Θ of the term involving κ_x is positive-semidefinite. Hence, by taking for κ_x the maximum of the K individual kurtosis estimates we avoid asymptotically underestimating the variance. In the next sections we will find out that the actual contributions to the KLS variance of the two terms involving kurtosis are in fact fairly insignificant. That we use the degrees of freedom corrected $s_u^2(\rho_{xu})$ in $\widehat{Var}[\hat{\beta}_{KLS}(\rho_{xu})]$ and the uncorrected $\hat{\sigma}_{u,OLS}$ in $\hat{\beta}_{KLS}(\rho_{xu})$ is deliberate, because in simulations these choices proved to be preferable in (very) small samples.

So, for any chosen numerical $K \times 1$ vector r, such that $r'S_xS_{xx}^{-1}S_xr < 1$, estimator $\hat{\beta}_{KLS}(r)$ is a consistent estimator of β, provided $r = \rho_{xu}$ indeed, and its variance can be adequately estimated too. Thus, contingent on using the true value for ρ_{xu}, restrictions on β can be tested and confidence regions constructed by consistent methods which can control significance levels asymptotically (and, as Section 5 will demonstrate, also surprisingly accurate in finite samples). By calculating $p_T(r)$, the p-value of a particular test statistic T, calculated assuming $\rho_{xu} = r$, over a dense grid of r values in some region $\mathcal{C}$ (this region has dimension equal to the number of potentially endogenous regressors in the relationship), instrument-free inference can be produced. Assuming $\rho_{xu} \in \mathcal{C}$, this inference is (asymptotically) conservative (meaning cautious by securing that asymptotically type I error probabilities will never exceed some critical threshold p^{crit}) by rejecting the hypothesis if $\forall r \in \mathcal{C}$ one finds $p(r) < p^{crit}$, and not rejecting the hypothesis if $p(r) > p^{crit}$ for $\forall r \in \mathcal{C}$; otherwise, when some $p(r)$ values exceed and some (for different r values in $\mathcal{C}$) do not exceed p^{crit}, the test is inconclusive over $\mathcal{C}$. Note that it is also possible to construct a range of regions, say $\mathcal{C}_h$ (for $h = 1, 2, ...$), such that the test is always conclusive over each separate subregion.

An intriguing feature of tests based on KLS estimates is that they allow to test validity of instruments by directly testing exclusion restrictions. Since KLS estimates are identified by some non-orthogonality conditions and not just by classic orthogonality conditions as in TSLS, each classic identifying restriction associated with an external instrument (not just the over-identifying ones!) can be tested, either on its own or in groups. Let $y_i = \beta_1'x_{i1} + \beta_2'x_{i2} + u_i$, where $\rho_{x_2u} = 0$, with the variables in $K_1 \times 1$ vector x_{i1} possibly endogenous. For method of moments estimation at least K_1 external but valid instruments are required. Let $K_3 \times 1$ vector x_{i3} contain $K_3 \geq 1$ candidate external instruments. Augmenting the model and estimating $y_i = \beta_1'x_{i1} + \beta_2'x_{i2} + \beta_3'x_{i3} + u_i^*$, by KLS over a credible subspace of likely values for ρ_{x_1u}, and then testing the exclusion restrictions $\beta_3 = 0$ on basis of $\hat{\beta}_{3,KLS}(r)$ may endorse or refute the acceptability of variables x_{i3} as valid external instruments.

In the applications of the KLS instrument validity test to follow, a peculiar and confusing phenomenon emerges. We find that, in particular when $K_1 = K_3 = 1$, the p-value of the exclusion restriction test is 1, or very close to 1, for r close to $\hat{\rho}_1 = n^{-1}x_1'\hat{u}_{TSLS}/(\hat{\sigma}_1\hat{\sigma}_{u,TSLS})$, the TSLS-based estimator of endogeneity correlation ρ_1. At first sight this seems to suggest that instrument x_3 is valid especially for ρ values close to the value suggested by the ρ estimate based on assuming validity of the very same instrument. However, in Appendix E we demonstrate that this is a fallacy. We show that when x_3 is a valid instrument then estimator $\hat{\beta}_{3,KLS}(\rho_1)$, evaluated at the true value ρ_1, tends to zero, as it should. Unfortunately, when x_3 is an invalid instrument, then estimator $\hat{\beta}_{3,KLS}(\hat{\rho}_1)$, evaluated at inconsistent estimator $\hat{\rho}_1$, tends to zero too. Hence, the test lacks power for values r which are deceivingly instigated by endogeneity estimates obtained by an invalid instrument. Therefore, we better just use the test to indicate values for ρ_1 for which the tested external instruments seem invalid, due to low p-values, rather than claiming validity of the instruments in an area around $\hat{\rho}_1$, for which p-values are not small.

Often primary interest is in estimating (or testing a linear restriction on) just a subset of the K coefficients β. Suppose that we can decompose the regressors and corresponding coefficients such that $X = (X_1, X_2)$ and $\beta_1' = (\beta_1', \beta_2')$ and that all endogenous regressors belong to $n \times K_1$ matrix X_1 and possibly some predetermined regressors as well, but all regressors X_2 have zeroes in their corresponding elements of vector ρ_{xu}. It is well known from partitioned regression that vector $\hat{\beta}_{1,OLS} = H\hat{\beta}_{OLS}$, where $H = (I_{K_1}, O)$, can also be obtained by regressing $M_2 y = y^*$ on $M_2 X_1 = X_1^*$, where $M_2 = I - X_2(X_2'X_2)^{-1}X_2'$. Since the sum of squared residuals of the regressions of y on X and of y^* on X_1^* are equivalent, also $\hat{\sigma}^2_{u,OLS} = \hat{\sigma}^2_{u^*,OLS}$, where $u^* = M_2 u$. This is in agreement with $\operatorname{plim} n^{-1}u^{*\prime}u^* = \operatorname{plim} n^{-1}u'M_2 u = \operatorname{plim} n^{-1}u'u$ from which it follows that $\sigma^2_{u^*} = \sigma^2_u$. From

$$\operatorname{plim} n^{-1}X_1^{*\prime}u^* = \operatorname{plim} n^{-1}X_1'[I - X_2(X_2'X_2)^{-1}X_2']u = \operatorname{plim} n^{-1}X_1^{*\prime}u = \operatorname{plim} n^{-1}X_1'u$$

we obtain $S_{x_1^*}\rho_{x_1^*u^*} = S_{x_1^*}\rho_{x_1^*u} = S_{x_1}\rho_{x_1u}$. Therefore, using a well-know result for the inverse of a partitioned symmetric matrix, $HS_{xx}^{-1}S_x\rho_{xu} = S^{-1}_{x_1^*x_1^*}S_{x_1}\rho_{x_1u} = S^{-1}_{x_1^*x_1^*}S_{x_1^*}\rho_{x_1^*u^*}$, and also $\rho_{xu}'S_xS_{xx}^{-1}S_x\rho_{xu} = \rho'_{x_1^*u^*}S_{x_1^*}S^{-1}_{x_1^*x_1^*}S_{x_1^*}\rho_{x_1^*u^*}$, thus $\hat{\sigma}_u(\rho_{xu}) = \hat{\sigma}_{u^*}(\rho_{x_1^*u^*})$. From this we find

$$\begin{aligned}\hat{\beta}_{1,KLS}(\rho_{xu}) &= H\hat{\beta}_{KLS}(\rho_{xu}) = \hat{\beta}_{1,OLS} - \hat{\sigma}_u(\rho_{xu})HS_{xx}^{-1}S_x\rho_{xu} \\ &= \hat{\beta}_{1,OLS} - \hat{\sigma}_{u^*}(\rho_{x_1^*u^*})S^{-1}_{x_1^*x_1^*}S_{x_1^*}\rho_{x_1^*u^*} = \hat{\beta}_{1,KLS}(\rho_{x_1^*u^*}). \qquad (4.6)\end{aligned}$$

Hence, when the focus is just on β_1, the KLS Theorem can also be applied to the regression of y^* on X_1^*, under the understanding that for full correspondence of the KLS coefficient estimates vector ρ_{x_1u} has to be replaced then by $\rho_{x_1^*u^*} = S^{-1}_{x_1^*}S_{x_1}\rho_{x_1u}$. Note that each individual elements of $\rho_{x_1^*u^*}$ cannot be smaller than the corresponding element of ρ_{x_1u}, because $X_1'X_1 - X_1^{*\prime}X_1^* = X_1'X_2(X_2'X_2)^{-1}X_2'X_1$ is positive-semidefinite.

Result (4.6) can be useful to deal slightly more satisfactorily with kurtosis of the regressors. Partialling out as many predetermined regressors (including dummy variables) as possible then requires to make an assessment only of the maximum of the kurtosis of the K_1 variables in X_1^*. In the model with just one endogenous regressor, after partialling out all predetermined regressors, inference on the coefficient of the endogenous regressor can directly be obtained on the basis of the kurtotis of the single variable M_2x_1. Then taking the maximum of all K kurtosis estimates has been avoided, and the "conservativeness" problem circumvented. Note that partialling out endogenous regressors would lead to complications because then $\rho_{x_1^*u^*} \neq \rho_{x_1u}$. The KLS Theorem supposes that the intercept has been partialled out already.

5. The accuracy of KLS estimates assessed by simulation

By executing controlled experiments we want to assess whether the actual distribution of KLS is well approximated by the limiting distribution that we obtained, and whether it behaves favorably in comparison to IV/TSLS estimates. Only when this is the case it seems worthwhile to further examine whether KLS based test procedures have reasonable probability to reject true and false hypotheses on coefficient values. For such simulation analyses we first have to design simple but representative families of DGPs.

In the next subsection we focus on the accuracy of the asymptotic approximation for the very simple model of Section 3, for the special case where it represents i.i.d. data from a cross-section, but where the regressor and disturbances are not necessarily normally distributed. In the second subsection we examine the actual and asymptotic distribution of the KLS estimator in a dynamic time-series model with additional regressors as in Section 4, and compare these also with OLS and TSLS, but now all the time sticking to cases where the variables are normal.

5.1. Results for a simple nonnormal cross-section model

We examine the KLS density for simple model (3.1) when the data are i.i.d., choosing a few different values for κ_u, κ_ξ and ρ_{xu} at a specific finite sample size n. To make all the densities to be obtained comparable we will choose $\sigma_u/\sigma_x = 1$, by taking $\sigma_u = 1$ and $\sigma_x^2 = 1$, which requires $\sigma_\xi^2 = \sigma_x^2 - \lambda^2\sigma_u^2 = 1 - \rho_{xu}^2$. We will focus on the density for the estimation error $\hat{\beta}_{KLS} - \beta$, which is invariant regarding the actual value of β, which we will therefore choose to be zero in the DGP. Next to kurtosis, we also want to examine the effects of skewness of the distributions of u_i and ξ_i. This is achieved by obtaining drawings u_i and ξ_i not only from the standard normal distribution $\mathcal{N}(0,1)$, but also from transformed Student(v) and transformed $\chi^2(v)$ distributions, where v indicates degrees of freedom. When η_i is Student(v) with $v > 4$ then standardized drawings $\eta_i/[v/(v-2)]^{1/2}$, to be indicated by $St^*(v)$, have zero mean, unit variance and kurtosis $3 + 6/(v-4)$. Next to this symmetric distribution we will also consider a skew one. When ψ_i is $\chi^2(v)$ distributed then standardized drawings $(\psi_i - v)/(2v)^{1/2}$, indicated by $Chi^*(v)$, have zero mean, unit variance and kurtosis $3 + 12/v$, whereas its skewness is $(8/v)^{1/2}$. Next to situations where $\kappa_u = 3$ and $\kappa_\xi = 3$ we will consider situations where $\kappa_u = 9$ and/or $\kappa_\xi = 9$ by using drawings from $St^*(5)$ or $Chi^*(2)$. The latter has skewness 2.

In Figure 5.1 we examine and compare (cumulative) densities for $n = 100$ at $\rho_{xu} = 0.2$ (left-hand panels) and $\rho_{xu} = 0.4$ (right-hand panels). All simulated densities (top-row panels) are obtained from 10^6 drawings from the relevant KLS distribution, whereas their asymptotic approximations (mid-row panels) directly represent the density of the $\mathcal{N}(0, n^{-1/2}[1 + \rho_{xu}^2(\kappa_u + \kappa_\xi - 6)/4])$ distributions, see (A.9). The bottom-row panels represent the discrepancies between the cumulative distributions of the simulated distributions and their asymptotic approximations. We note that when both u_i and ξ_i are normal (black uninterrupted line) the asymptotic approximation is rather accurate for both ρ_{xu} values, as had been established already in Kiviet (2013, 2019) just by comparing tail probabilities. From the present results we also see that when $\kappa_u + \kappa_\xi$ increases up to 18, the accuracy of the asymptotic approximation to represent the actual distribution in finite sample gets slightly worse, and nonsymmetry leads to some further reduction of the accuracy of the asymptotic approximation when $n = 100$. However, the discrepancies occur especially close to the central parts of the distribution, and less in the tail areas. From further calculations (not depicted) for these cases we found that at $n = 500$ the symmetric asymptotic approximation is much more satisfactory.

5.2. Results for a simple time-series model under normality

Now we shall examine KLS for a simple stable synthetic dynamic regression relationship in stationary zero-mean variables, given by

$$y_i = \beta_1 x_i + \beta_2 y_{i-1} + u_i, \text{ for } i = 1, ..., n, \tag{5.1}$$

where $|\beta_2| < 1$ and

$$\begin{aligned} u_i &\sim i.i.d.(0, \sigma_u^2), & \varepsilon_i &\sim i.i.d.[0, (1-\pi^2)\sigma_\xi^2], \\ x_i &= \xi_i + \lambda_1 u_i, & \xi_i &= \pi\xi_{i-1} + \varepsilon_i \text{ with } |\pi| < 1. \end{aligned}$$

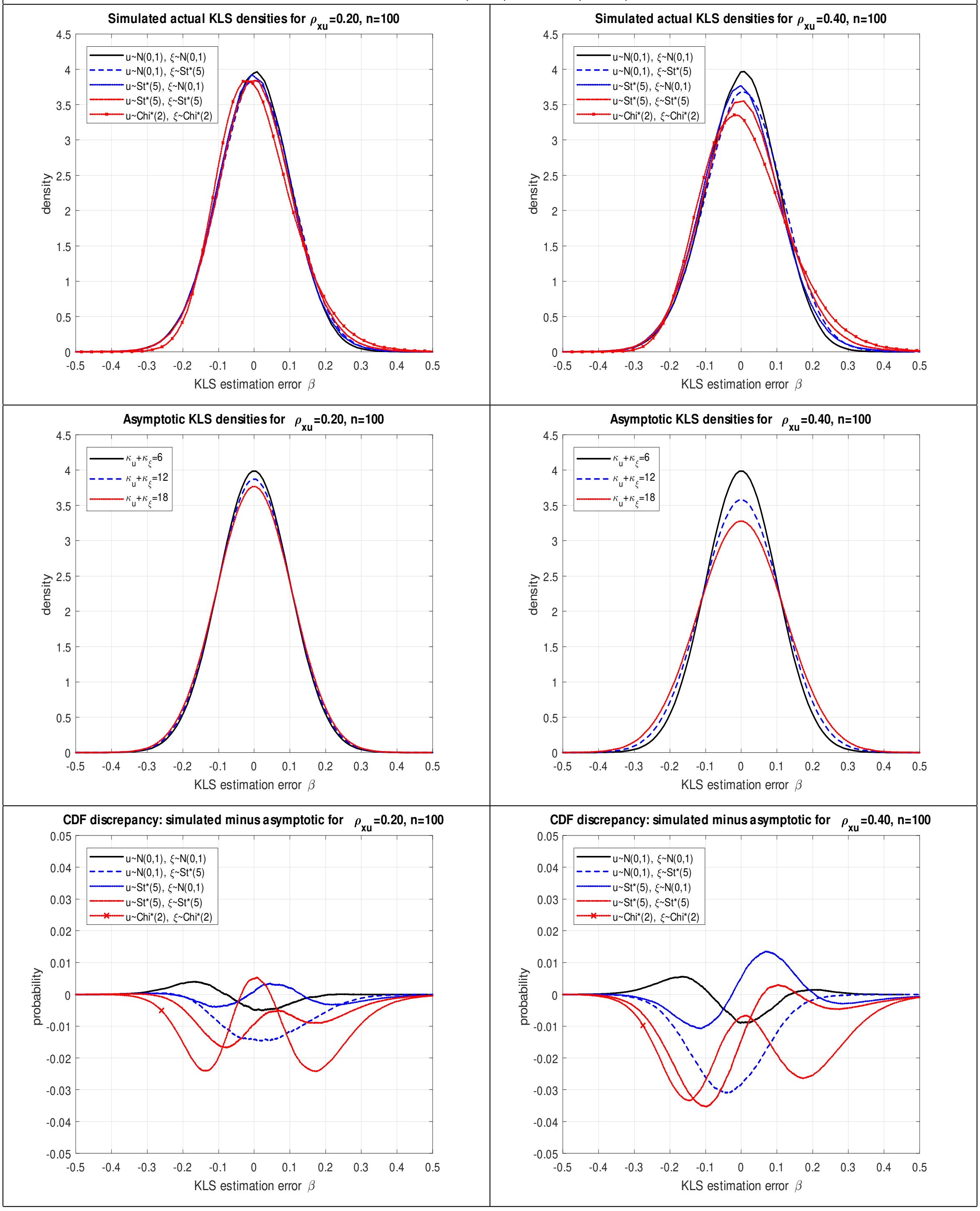
Hence ξ_i is an AR(1) process with $E(\xi_i \xi_{i-l}) = \pi^{|l|}\sigma_\xi^2$. If $\pi \neq 0$ the x_i series will be serially correlated too, whereas x_i is endogenous as well, provided $\lambda_1 \neq 0$. All x_i have variance

$$\sigma_x^2 = \sigma_\xi^2 + \lambda_1^2 \sigma_u^2.$$

The y_i series will have zero-mean too with variance

$$\sigma_y^2 = \beta_1^2 \sigma_x^2 + \beta_2^2 \sigma_y^2 + \sigma_u^2 + 2\beta_1\beta_2 E(x_i y_{i-1}) + 2\beta_1 E(x_i u_i).$$

Figure 5.1 KLS distribution for model (3.1) and (non)normal i.i.d. series u_i and ξ_i



Since

$$E(x_i y_{i-1}) = E[(\pi\xi_{i-1} + \varepsilon_i + \lambda_1 u_i)y_{i-1}] = \pi E(\xi_i y_i),$$

whereas y_i consists of many terms from which just $\beta_1 \Sigma_{l=0}^{\infty} \beta_2^l \xi_{i-l}$ is correlated with ξ_i, we obtain

$$E(x_i y_{i-1}) = \pi\beta_1 \Sigma_{l=0}^{\infty} \beta_2^l E(\xi_i \xi_{i-l}) = \pi\beta_1 \sigma_\xi^2/(1 - \pi\beta_2).$$

Now, using $\lambda_1 = \rho_1\sigma_x/\sigma_u$, $\sigma_x^2 = \sigma_\xi^2 + \rho_1^2\sigma_x^2$ and $\sigma_x^2 = \sigma_\xi^2/(1-\rho_1^2)$, we find

$$\begin{aligned}(1-\beta_2^2)\sigma_y^2 &= \beta_1^2\sigma_x^2 + \sigma_u^2 + 2\beta_1^2\beta_2\pi\sigma_\xi^2/(1-\pi\beta_2) + 2\beta_1\lambda_1\sigma_u^2 \\ &= \sigma_\xi^2\beta_1^2[1/(1-\rho_1^2) + 2\beta_2\pi/(1-\pi\beta_2)] + 2\sigma_\xi\beta_1\rho_1\sigma_u/(1-\rho_1^2)^{1/2} + \sigma_u^2.\end{aligned} \quad (5.2)$$

We shall use the above relationships to choose parameter values for the DGP which seem empirically relevant. Without loss of generality we may choose $\sigma_u^2 = 1$. We are especially interested in moderately nonnegative values of β_2. By choosing $\beta_1 = 1-\beta_2$ the long-run multiplier of y with respect to x will be kept constant at value unity, irrespective of the speed of the dynamic adjustment process determined by β_2. For π we may choose a value like 0.8, so that x_i mimics a smooth time-series, and for ρ_1 values in the interval $[-0.5, +0.5]$ seem most relevant. Given numerical values for σ_u, β_1, β_2, π and ρ_1 we can generate data for model (5.1) as soon as we have also chosen a relevant value for σ_ξ. For this we use the fact that relationships like (5.1) usually have a rather high coefficient of determination (low noise/signal ratio). Therefore, we will impose

$$1 - \sigma_u^2/\sigma_y^2 = R^*, \quad (5.3)$$

with R^* equal to, say, 0.9. Substituting all chosen characteristics in (5.3) and (5.2) yields a polynomial equation in σ_ξ of second order from which we can obtain real positive solutions for σ_ξ, provided we have chosen compatible values for the other parameters. Defining scalars a, b and c such that (5.2) can be rewritten as $\sigma_y^2 = a\sigma_\xi^2 + b\sigma_\xi + c$, then (5.3) yields

$$\frac{a\sigma_\xi^2 + b\sigma_\xi + c - \sigma_u^2}{a\sigma_\xi^2 + b\sigma_\xi + c} = R^*.$$

Next, from the polynomial equation

$$a(1-R^*)\sigma_\xi^2 + b(1-R^*)\sigma_\xi + c(1-R^*) - 1 = 0$$

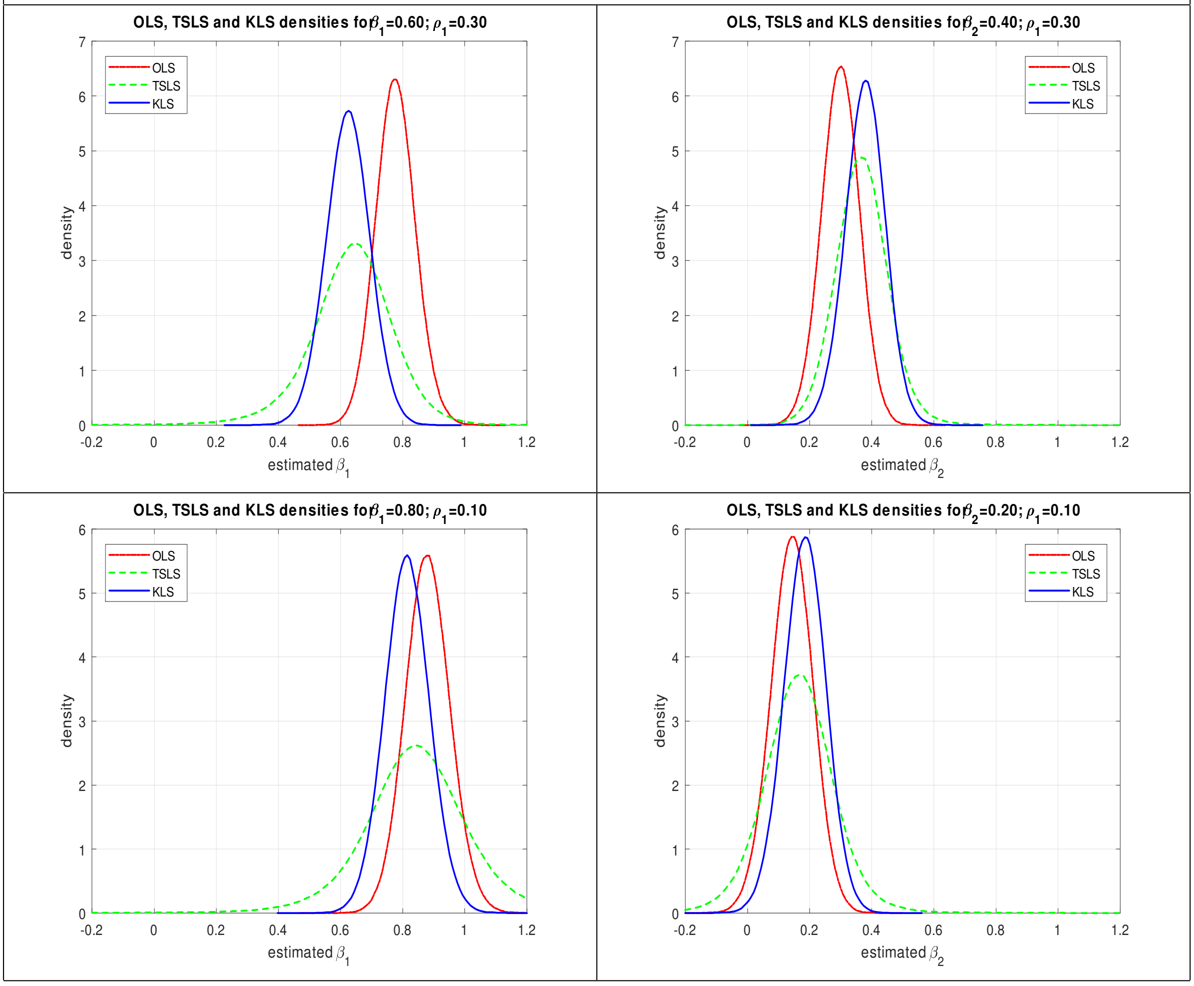
we will consider its positive solutions

$$\sigma_\xi = \{-b(1-R^*) \pm [b^2(1-R^*)^2 - 4a(1-R^*)(c(1-R^*)-1)]^{1/2}\}/[2a(1-R^*)].$$

Then, series $\{x_i, y_i\}$ can be drawn on the basis of generated series $\{\varepsilon_i, u_i\}$, for which we may choose alternative values for skewness and kurtosis.

That consistent estimates of dynamic models like (5.1) may show substantial bias in samples of finite size has aroused a rather massive literature. The magnitude of this bias has been assessed under normality, both by simulation and by analytical methods (higher-order asymptotic approximations), both for models in which x_i is exogenous and OLS has been examined, see Kiviet and Phillips (2012) and its references, and for models in which x_i is endogenous and TSLS has been examined, see Phillips and Liu-Evans (2016) and its references. Here our primary aim is simply to investigate the finite sample density of KLS estimators as obtained from simulation experiments, and compare these with competitors.

Figure 5.2.1 Simulated densities for $n = 50$, $\pi = 0.8$, $R^* = 0.9$, $\kappa_\varepsilon = 3$, $\kappa_u = 3$



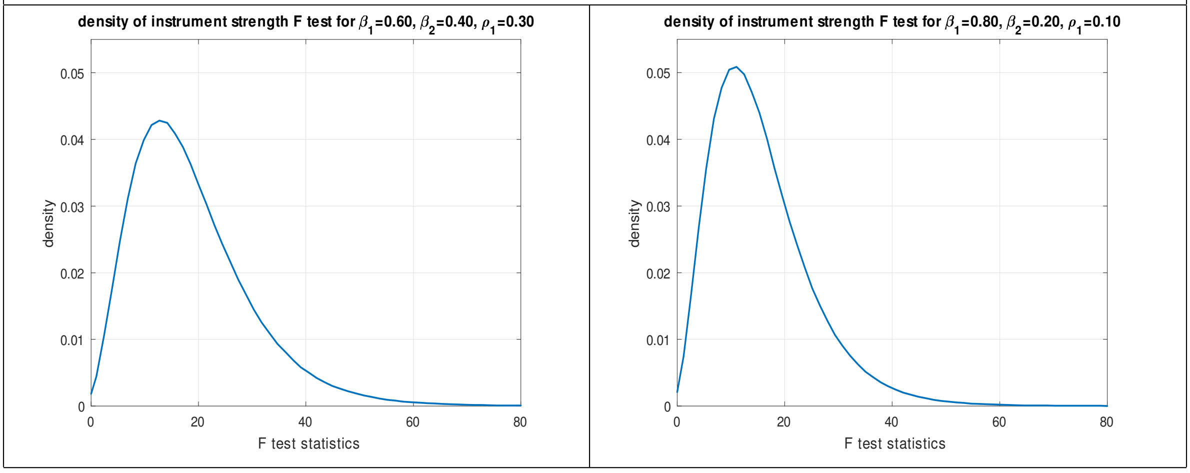
In Figure 5.2.1 we present densities for the OLS, TSLS and KLS estimators of β_1 (left-hand panels) and $\beta_2 = 1 - \beta_1$ (right-hand panels) for $\beta_1 = 0.6$ and $\rho_{xu} = 0.3$ (top-row panels) and for $\beta_1 = 0.8$ and $\rho_{xu} = 0.1$ (bottom-row panels), whereas $\pi = 0.8$, $R^* = 0.9$, $\kappa_\varepsilon = 3$, $\kappa_u = 3$ (in fact ε_i and u_i were drawn from the normal distribution) and $n = 50$. By choosing $\beta_1 + \beta_2 = 1$ (not imposed when estimating) the long-run multiplier of x with respect to y is unity. All these density estimates are based on 10^6 simulation replications. By skipping n initial observations we made sure that the generated series are stationary indeed. The TSLS estimator is actually a simple IV estimator, where x_{i-1} and y_{i-1} have been used as instruments, so the degree of overidentification is zero, implying that formally the estimator has no finite moments thus its distribution may be fat tailed.

We note that the inconsistent OLS estimator is rather severely biased, especially β_1, even for the relatively small value $\rho_1 = 0.3$. The consistent TSLS estimator is better centered around the true value, but it has substantially larger dispersion and it is also slightly skew for β_1. The KLS estimator clearly outperforms both alternatives, showing no substantial bias nor skewness, and visibly having the smallest mean squared error. Even for minor simultaneity ($\rho_1 = 0.1$) and much slower dynamic adjustment ($\beta_2 =$

0.2) the same patterns show up. We also examined larger samples where $n = 250$ (not depicted). Then all densities are more peaked, but still show the same distinctive properties.

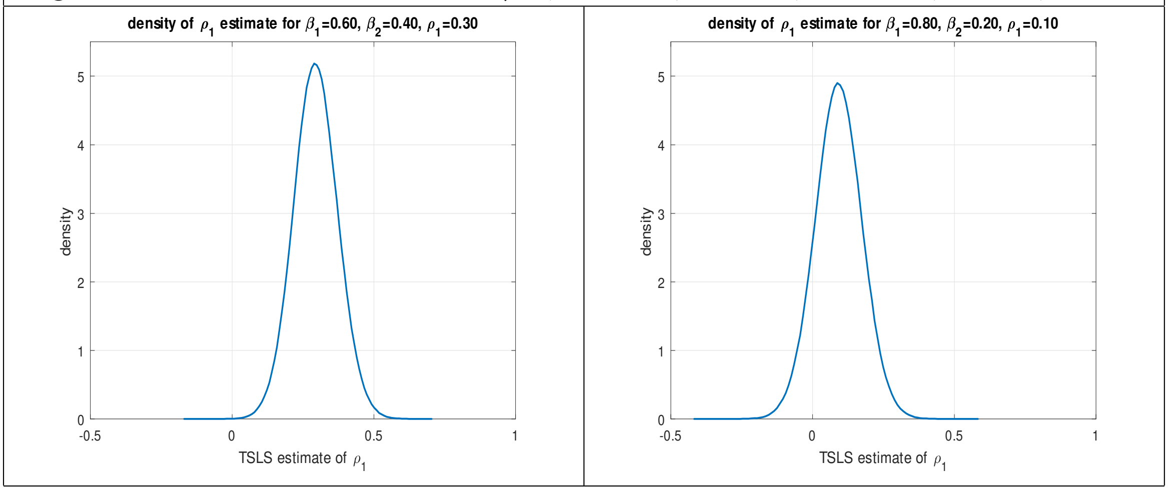
Of course, one should keep in mind that the KLS estimator as presented is unfeasible, because it uses, next to $E(y_{i-1}u_i) = 0$, the true value of ρ_1 which is unknown in practice. Likewise, however, the TSLS estimator is unfeasible, because it exploits the two moment conditions $E(y_{i-1}u_i) = E(x_{i-1}u_i) = 0$, which in practice cannot be vindicated statistically either. In Figure 5.2.2 we depict for both parameterizations examined in Figure 5.2.1 the strength of instrument x_{i-1} by presenting the density of its two-sided significance test in the first-stage regression. We note that in a nonnegligable number of replications this instrument turned out to be weak, but on average the relevant F test statistic has been well above 10 in both cases.

Figure 5.2.2 Simulated densities of F test; $n = 50$, $\pi = 0.8$, $R^* = 0.9$, $\kappa_\varepsilon = 3$, $\kappa_u = 3$



In cases where validity of these moment conditions seems beyond dispute, one might (keeping the result proved in Appendix E in mind) attempt to exploit the KLS properties by substituting for the unknown ρ_1 the estimated correlation $\hat{\rho}_1$ between the TSLS residuals and the endogenous regressor. From simulations for $n = 50$ we found that the distribution of such an estimator $\hat{\rho}$, although consistent and reasonably well centered around its true value, is very imprecise.

Figure 5.2.3 Simulated densities of $\hat{\rho}_{xu}$; $n = 250$, $\pi = 0.8$, $R^* = 0.9$, $\kappa_\varepsilon = 3$, $\kappa_u = 3$



In Figure 5.2.3 we present its simulated density for both parameterizations examined in the Figure 5.2.1, but we took $n = 250$ now. From this it seems clear that TSLS estimates of the degree of endogeneity are not very trustworthy (although we found the reasonable values 0.294 and 0.091 for their estimated expectation respectively), due to their wide dispersion, let alone because of their always questionable consistency.

The few graphs in this section just provide a very superficial impression on the general properties of KLS. Choosing different parameter values may alter the relative performance, especially if they imply the use of weaker/stronger instruments. For the construction of one-sided test statistics, which in both directions have acceptable size control, it is required that the KLS distribution is almost symmetric and well centered around its true value and that the actual variance of the KLS estimator is reasonably well approximated by its asymptotic expression. Verification of that is deferred to future work.

6. Three empirical illustrative replication studies

Below we will illustrate how the techniques discussed here can be used in practice and can place earlier obtained results using instrumental variables into a new revealing perspective, either positively or negatively. The first is an international-macro application in which we re-analyze a country cross-section data set from which the causal effect of international trade on income per capita has been assessed. In this illustration we focus in particular on the effect of nonzero excess kurtosis. In the second example we re-analyze a cross-sectional data set on Vietnamese individuals examining the causal effect of personal income on risk aversion. Here we consider a specification with one and also one with two endogenous regressors and compare KLS findings with weak-instrument inference. In the third illustration we re-analyze a time-series data set and pay extra attention to the fact that this study demonstrates that our formulas also apply to data that are temporally dependent and to dynamic models.

6.1. Effect of trade on growth

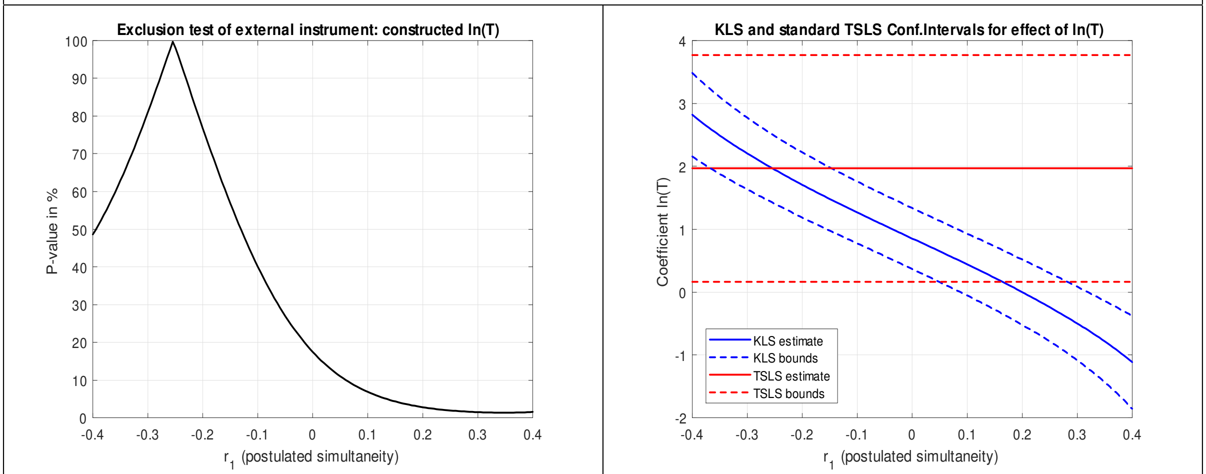
In a much referenced study (over 6000 citations according to Google) by Frankel and Romer (1999), below referred to as F&R, data for the year 1985 on 150 countries have been analyzed from which it has been concluded (F&R, Table 3, column 2) that a 1 percent-point raise of the trade share T (defined as the sum of exports and imports divided by GDP) leads to an increase of about 2% in income per person Y (the coefficient estimate in a regression of $\ln Y$ on T is 1.97 with standard error 0.99). This is the result of a linear IV analysis where trade share is the one and only endogenous regressor supplemented by an intercept and two exogenous covariates, namely log of population $\ln N$ and of area $\ln A$. The study uses one instrument, called the constructed trade share $\hat{T}$, which has been obtained by regressing trade share on a series of geographic characteristics. So, actually TSLS has been employed. However, by not providing the original set of instruments used in the first stage, neither the Sargan test can be re-established from the provided second stage data in F&R's Table A1, nor the Anderson-Rubin confidence set on the coefficient of the endogenous regressor can be obtained.

The constructed single instrument $\hat{T}$ is not weak, but not very strong either, because the relevant first-stage F test statistic of $\hat{T}$ (in a regression of T on an intercept, $\ln A$, $\ln N$

and $\hat{T}$) is 13.1, which may explain the relatively large standard error and consequently large confidence interval (0.03, 3.91) for the parameter of major interest. This interval has nominal confidence coefficient 95% but may in fact be highly unreliable. The OLS estimate of the coefficient of interest is only 0.85 with much smaller standard error 0.25. Of course, both coefficient estimates may be severely biased. However, by a Hausman test F&R establish that endogeneity of trade share is not significant (its t-value is 1.2), and therefore the substantial difference between the IV and OLS coefficient estimates is actually not a significant difference. In the end F&R conclude that the IV coefficient estimate being more than twice the size of the OLS estimate is simply due to random variation, also because they argue at length that they in fact expect that the IV estimate should be smaller than the OLS estimate, simply due to $\rho_1 > 0$.

Our procedures can put some extra statistical evidence on the table. Making assumptions on the value of ρ_1 we can verify whether the constructed trade share seems a valid instrument indeed, by testing whether it has been validly omitted from the specified relationship. This may either reinforce or oust our trust in the TSLS findings. Especially in the latter case our direct instrument-free inference on the parameter of interest, which is not infested with weak or invalid instrument problems, may be of more use here. From the IV results we obtain an estimate for the degree of endogeneity of -0.25, which we know may be very imprecise (as Section 5 indicated), or even inconsistent, if the instrument is invalid, possibly due to a poorly specified structural equation. So, in line with F&R's belief, we better also examine what the consequences would be of more credible values $\rho_1 > 0$.

Figure 6.1 KLS inference regarding the effect of trade on growth



KLS defined over $|r_1| \leq 0.69$ | KLS defined over $|r_1| \leq 0.66$

From Figure 6.1 we see from its left-hand panel that if $\rho_1 > 0$ it does not seem likely that the instrument is valid, due to low p-values of the exclusion restriction test. We also note the high p-value around values for ρ_1 close to its IV estimate. This is the phenomenon proved in Appendix E: Accepting an instrument as valid, yields estimates for the endogeneity correlation which approve the exclusion restriction. Note the circularity here; this does not provide evidence that ρ_1 is negative and neither that $\hat{T}$ is a valid instrument.

From the right-hand panel of Figure 6.1 we see that for any value of $\rho_1 > -0.25$ KLS yields (much) lower values for the estimated effect of trade-share than IV produces, and also yields much narrower confidence intervals. Assuming $-0.14 \leq \rho_1 \leq 0.05$ we conclude with a probability exceeding 95% that the effect of trade share is in between 0 and 2, assuming that the model is adequately specified. Under the same proviso, KLS suggests that if $\rho_1 > 0.08$ then the coefficient of $\ln(T)$ is smaller than 1.

Table 6.1 Skewness and kurtosis of some F&R variables

	$\ln Y$	T	$\ln N$	$\ln A$	$\hat{T}$	IV-resid.
skewness	-0.19	1.64	-0.66	-0.11	5.14	-0.25
kurtosis	2.00	7.90	3.15	2.99	43.36	2.56

Table 6.1 presents estimates of skewness and kurtosis of the relevant variables. Note the extremely large kurtosis of the external instrument $\hat{T}$. The maximized value of κ_x used in the variance estimation for the KLS exclusion restriction test has been 43.36 and 7.9 for the construction of the KLS confidence set. We have also constructed the graphs of Figure 6.1 after partialling out the variables $\ln N$ and $\ln A$, and also after taking $\kappa_x = 3$, but this has hardly a visible effect, so we do not present these here.

From the above we conclude that it seems that the employed structural equation has been rather poorly specified. Therefore it seems most likely that also the regressors $\ln N$ and $\ln A$ are correlated with the disturbance term. Rather than taking that into account with our methods, we think that attempts should be undertaken first to formulate a better explanatory model of income in terms of trade share and further controls not available in the present data set.

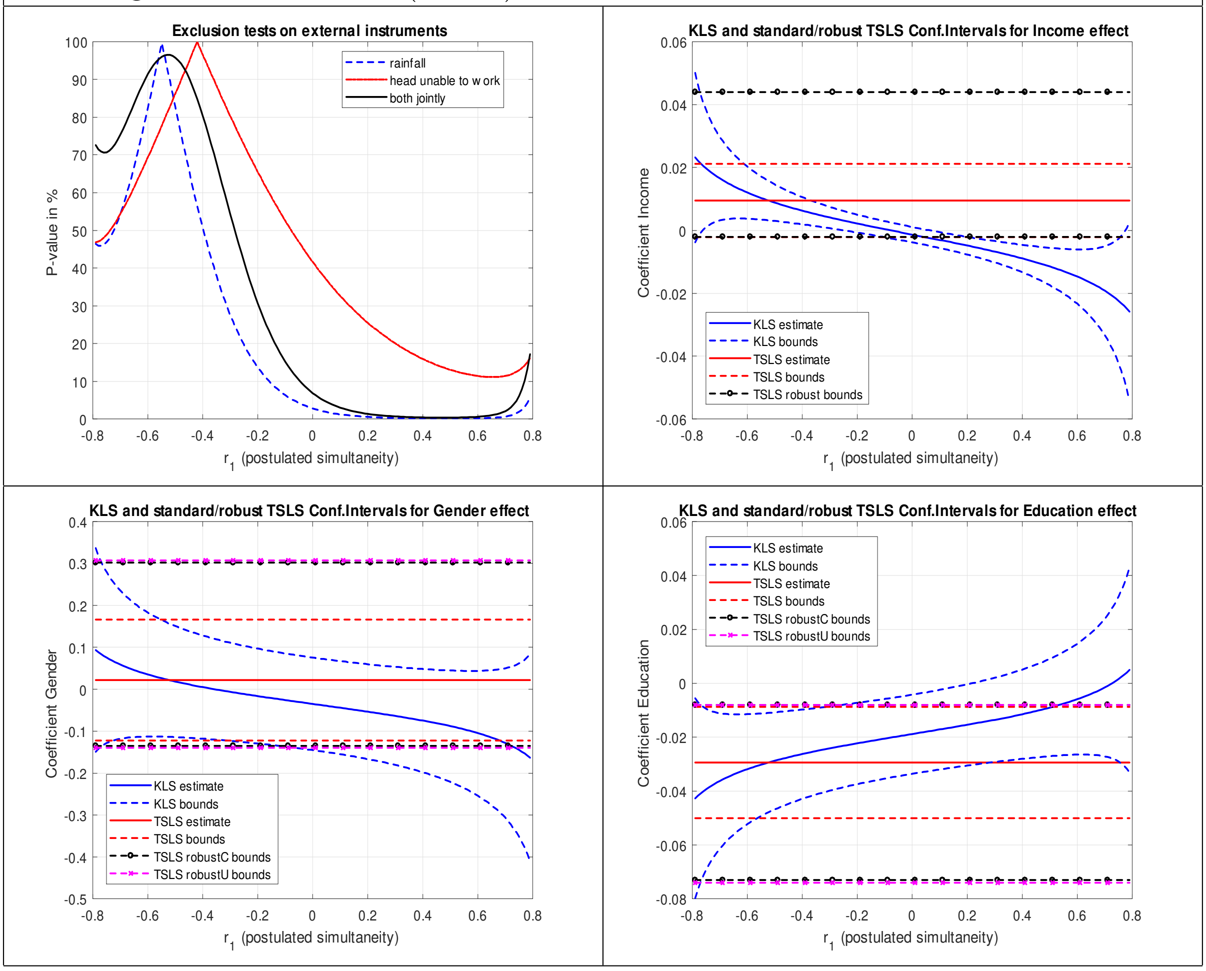
6.2. Risk preferences in Vietnam

Tanaka, Camerer and Nguyen (2010) analyze cross-section data obtained by combining living standard survey data on individuals collected in 2002 with additional experiment-based direct measurements on risk and time preferences of the very same individuals living in nine particular communities. In their Section II the authors (below indicated as TCN) estimate relationships for two different dimensions of risk aversion. Below we will just focus on the results for the dependent variable "concavity of the value function" (σ in the TCN notation), for which they examined two different specifications. In addition to a range of exogenous demographic control variables, they include as endogenous regressor(s) either just household income, or both relative income and mean income within the community. These two relationships are estimated by TSLS employing two external instrumental variables: rainfall and the dummy variable household head's ability to work. Hence, in one equation the degree of over-identification is one, and in the other it is zero. Although TCN use two variants of tests to verify the significance of the endogeneity of the income variables, no test of over-identification restrictions has been employed.

With respect to the validity of the two external instruments they just state that these are: "... unlikely to be correlated with preferences, as instrumental variables for income". This, however, is a most confusing remark, originating in the fact that TCN wrongly use the word correlation for equation coefficient (see also the caption of their Table 4). A relevant instrument should be (preferably substantially) correlated with the endogenous regressors, and will therefore (assuming the endogenous regressor has nonzero coefficient) also be correlated with the dependent variable. However, to be a valid instrument, it should be correctly excluded from the structural equation for the dependent variable and thus have coefficient zero in this equation. Therefore, its impact on the dependent variable should be solely indirect, via the endogenous regressor. Correct exclusion seems perhaps self-evident for rainfall, but is much less straight-forward regarding the dummy instrument (ability to work), because one's preferences regarding risk taking may certainly be structurally affected in case of permanent disability. In their footnote 8 TCN remark that they "... tested several instrumental variables ...". However, here they refer to F-tests in (reduced form) first-stage regressions, so testing the relevance (strength or weakness) of the instruments, but not their validity.

Let us now turn to the empirical findings obtained from the sample of 181 observations. Presupposing validity of the instruments and adequacy of the two specified equations, as TCN persistently do, the tests on endogeneity of the income variables reject consistency of the OLS results. TCN assume that the endogeneity is due to reciprocal causality. Given our findings in Section 2 this would imply the endogeneity correlations to be positive, unless $\gamma_0\beta_1 > 1$. In fact, for the equation with just one endogenous regressor the estimated correlation of the endogeneity is -0.52. Of course, apart from their substantial standard errors, these correlation estimates will only be consistent if the instruments are valid indeed.

Figure 6.2.1 KLS and (robust) TSLS inference for the model with $K_1 = 1$



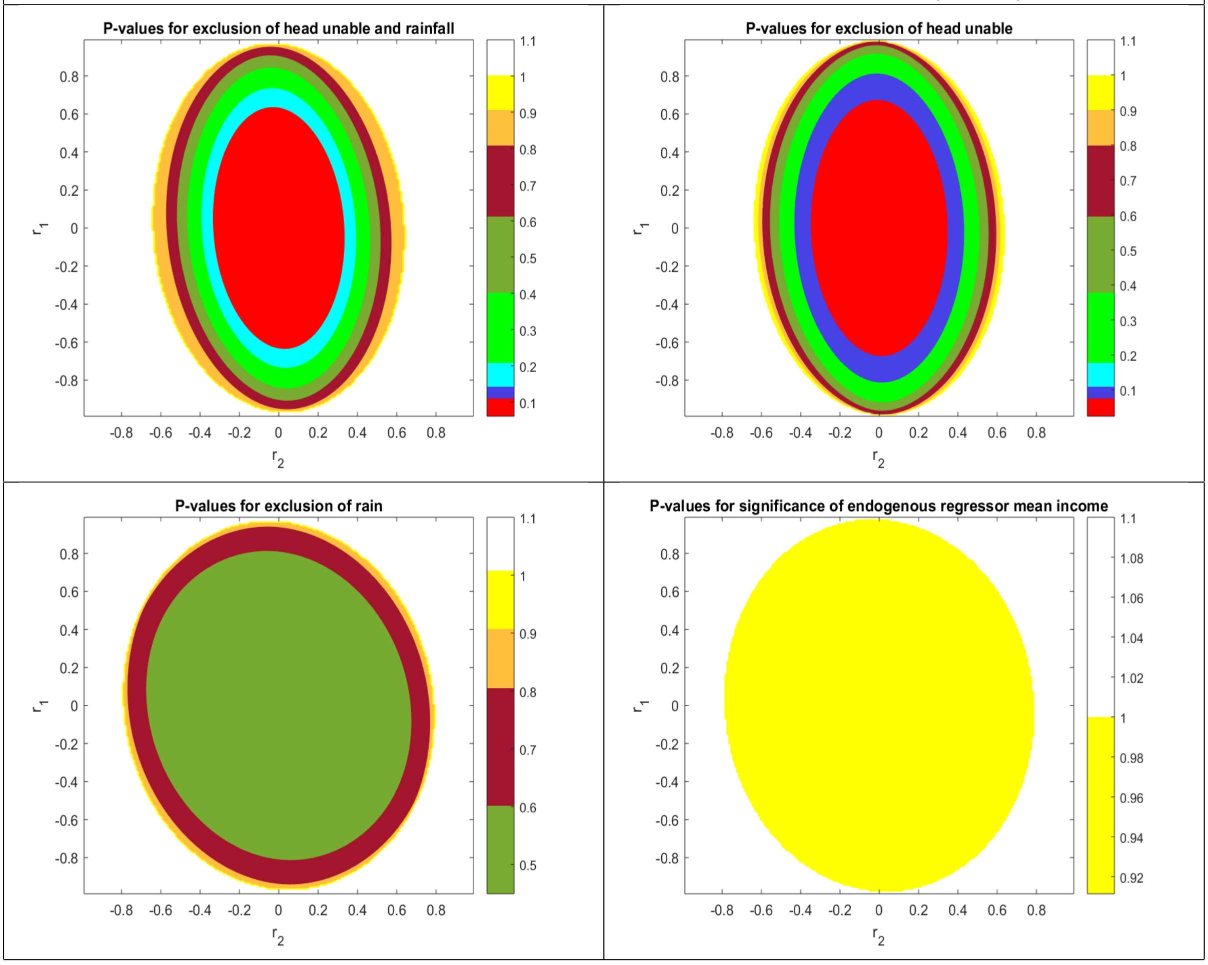
TCN apply TSLS (see their Table 5) and find for the model with one endogenous regressor ($K_1 = 1$) for the coefficient of income 0.010 (standard error 0.006). Using their data set we find that the Sargan test for the overidentification restriction has p-value 0.79, whereas testing the joint strength of the instruments yields F-value 5.96, so on basis of these usual diagnostics the instruments seem valid (implicitly supposing that one instrument is valid anyhow) but seriously weak.

Figure 6.2.1 presents KLS results for the relationship with $K_1 = 1$. The top-left panel (for the three curves KLS is defined for r_1 in absolute value smaller than 0.92, 0.94 and 0.92 respectively) shows that especially the "rainfall" dummy instrument has low p-values for $\rho_1 > -0.2$ and the joint exclusion of the two instruments finds certainly no support for positive ρ_1. Observing the peaks of the three p-value curves, it will not surprise that the TSLS based estimators for ρ_1 obtained when just using the instrument "rainfall" is -0.55, when just using the instrument "head unable to work" it is -0.42, and when using both instruments it is -0.52. Although the endogenous regressor has an estimated kurtosis of 31.7 this high value has little effect on the results in Figure 6.2.1. Substituting (incorrectly) $\kappa_x = 3$ in the variance formula was again found to have just minor effects.

The other panels (defined for $|r_1| \leq 0.95$) produce (asymptotic) 95% confidence intervals for the endogenous regressor Income, and for the exogenous regressors Gender and Education. Next to KLS and standard TSLS intervals they also present the weak-instrument robust intervals constructed by Guggenberger, Kleibergen and Mavroeidis (2019). From the KLS findings on Income we may conclude that its effect is positive provided $\rho_1 < -0.2$. And, assuming that $\rho_1 > -0.6$, one can also conclude that the Income coefficient value does not exceed 0.02. For this coefficient, the robust –and thus much more trustworthy interval than the standard one– highlights that weak-instrument robust TSLS inference is actually not very efficient in comparison to KLS. For the coefficients of exogenous regressors this difference in width of the confidence intervals is less spectacular.

For the model with two endogenous regressors we find that the correlation estimates of the endogeneity of relative and mean income are -0.16 and -0.18 respectively. Figure 6.2.2 presents 2D contour plots of p-values for four different KLS-based tests over all feasible postulated values (r_1, r_2) for (ρ_1, ρ_2). Its North-West panel shows that for moderate absolute values of both endogeneity correlations the joint validity of the two instruments has to be rejected. From the North-East and South-West panels we see that it is again the "head unable to work" instrument which is to blame for this. Therefore, producing further TSLS inference on the coefficients of this relationship seems in vain. Assuming validity of both instruments, TCN and Guggenberger et al conclude that the coefficient of relative income is insignificant, whereas that of mean income is –or is close to– significantly positive. Not using any external instruments, the South-East panel of Figure 6.2.2 presents the p-values (all larger than 0.9) for a KLS-based one-sided t-test. This shows that the hypothesis of a zero coefficient for mean income should not be rejected in favor of a positive value, irrespective of the values of the endogeneity correlations.

Figure 6.2.2 KLS inference for the model with $K_1 = 2$ ($\kappa = 3$)

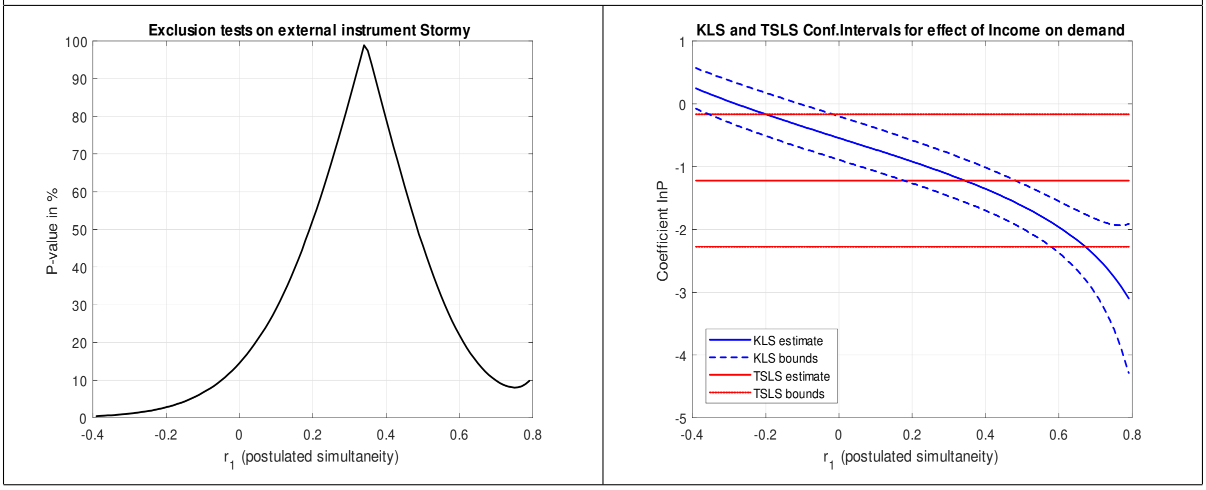


6.3. Demand at the Fulton fish market

To illustrate our procedures when applied to time-series observations, we focus on some studies that examined the market for whiting based on 111 consecutive trading days at the Fulton fish market. These data originate from Graddy (1995). Angrist, Graddy and Imbens (2000) produce in one of the columns of their Table 4 just-identified TSLS results for a base-line static linear in logs demand equation. These TSLS results represent the preferred specification in Graddy (2006, Table 4, column 2) and in Graddy and Kennedy (2006, Table 2, column 2). The latter study argues that supply is not just determined by the previous night's catch, but also by inventory changes, so that at this market quantity traded and its price are simultaneously determined indeed. Using this fish market example, Imbens (2014) provides a thorough explanation to convince especially statisticians that endogeneity of explanatory variables is often a reality in observational studies, despite the confusing fact that simultaneous equations cannot directly be simulated on a computer, without generating endogenous variables by their reduced form equations, or possibly (in case of the fish market) by integrating all relevant aspects of each individual transaction during the day.

In the static demand equation examined in the just mentioned studies the regressand is $\log Q$ (quantity) and we expect endogenous regressor $\log P$ (price) to be positively correlated with the error term. In line with that OLS yields a larger (less negative) price coefficient estimate than IV. This equation also includes day of the week dummies and two variables "cold" and "rainy", characterizing the weather on shore. It is just-identified, because weather variable at sea "stormy" is supposed to be a determinant of supply, but not of demand.

Figure 6.3.1 KLS inference for the static demand equation



Further calculations reveal the following about the static specification. It yields $\hat{\rho} = 0.34$, and for $0.18 \leq \rho \leq 0.48$ the exclusion restriction test has a p-value exceeding 0.5, as can be seen from the left panel of Figure 6.3.1. So, for moderately positive values of ρ, the application of IV finds support. However, the first-stage F test for the external instrument is 5.85, so it is pretty weak. Therefore, we expect the IV estimate of the price elasticity of demand to be biased in the direction of OLS, which would imply that the true value might be closer to -2 than to the obtained -1.22. However, KLS inference, which is not plagued by weak instrument problems and finite sample inaccuracies, indicates that the elasticity seems between -0.2 and -1.7, provided this static specification of the demand equation is appropriate and $0 \leq \rho_1 \leq 0.4$.

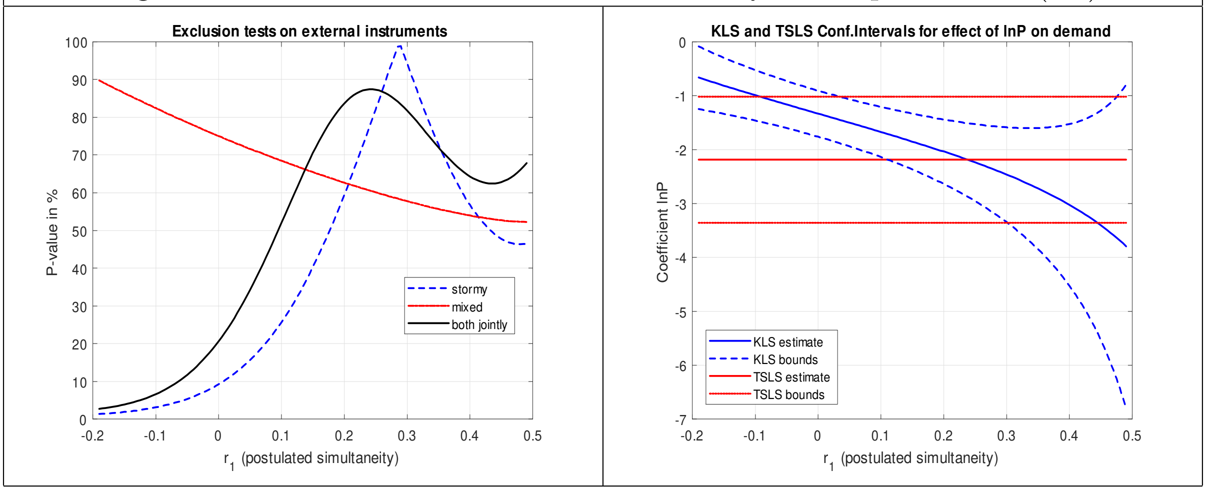
None of the studies presenting the investigated specification report any checks on its adequacy, apparently bedazzled by the ostensible correct signs of the coefficients despite their substantial standard errors. However, a test for 1-st order serial correlation produces a p-value of 0.00. This finding disqualifies all substantive inferences regarding this statistic specification.

Hendry and Nielsen (2006) analyze the same Fulton fish market data, adhering to a methodology which aims to fully respect the temporal-dependence in the observed variables. First they find single descriptive "congruent" time-series models for both $\log Q_t$ and $\log P_t$, and next they combine these in a VAR (vector autoregressive) model of order 1, which also includes two weather dummy variables, namely the earlier "stormy" and another called "mixed" (also used in further investigations by Angrist et al, 2000), together with a dummy "hol" which is unity at three particular dates close to holidays. This VAR specification passes an extensive mis-specification analysis. Next, using decomposition methods to the joint log-likelihood, these initial descriptive exercises inspire to formulate dynamic structural simultaneous equations for supply and demand. For the over-identified demand equation their maximum likelihood estimates are almost similar to the corresponding single equation TSLS estimates (see their Tables 15.5 and 15.9)

$$\underset{(0.07)}{\log Q_t = 8.52} - \underset{(0.59)}{2.19} \log P_t + \underset{(0.47)}{1.83} \log P_{t-1} - \underset{(0.36)}{1.89} hol_t. \quad (6.1)$$

Here "stormy" and "mixed" are used as external instruments. The Sargan test for the single over-identification restriction has p-value 0.60, and a test for 1st order serial correlation has p-value 0.71. The joint first stage F statistic of the two external instruments is 9.45, so although not strong, they are not very seriously weak either.

Figure 6.3.2 KLS inference for over-identified dynamic specification (6.1)



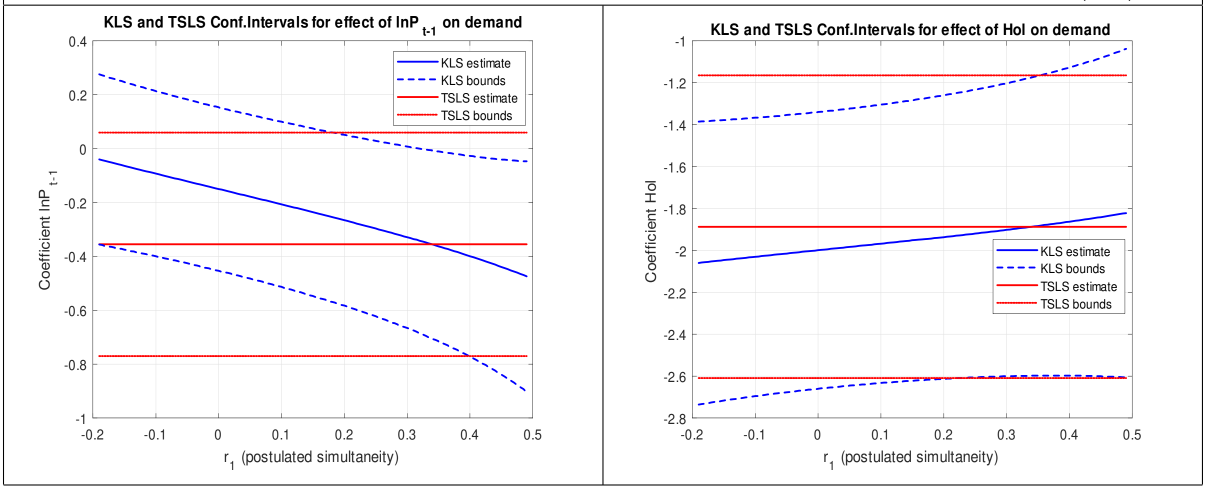
Application of KLS yields the following. The left-hand panel of Figure 6.3.2 supports the validity assumption regarding both instruments, assuming $\rho_1 > 0$ (the $\hat{\rho}_1$ estimates are 0.28, -0.31 and 0.24 for using stormy, mixed or both respectively as instruments). The right-hand panel produces (as we saw before) for moderate r_1 values narrower confidence bands for the coefficient of the endogenous regressor ($\log P_t$), but not as dramatically as for cases where instruments are really weak.

This dynamic specification of demand implies a price elasticity which reacts immediately very strongly to price changes, but with a huge correction to that the next trading day. So, the long-run elasticity (which is attained already after one period) has TSLS estimate of only -0.36. By reformulating the model we can easily obtain its TSLS standard error, giving

$$\underset{(0.07)}{\log Q_t = 8.52} - \underset{(0.59)}{2.19} \Delta \log P_t - \underset{(0.21)}{0.36} \log P_{t-1} - \underset{(0.36)}{1.89} hol_t. \quad (6.2)$$

So, the elasticity seems much smaller in absolute value than inferred before, and it does not even seem significantly negative. Figure 6.3.3 presents KLS inference on the coefficients of the predetermined regressors $\log P_{t-1}$ and hol_t. Note that the KLS results imply a significantly negative long-run demand elasticity provided $\rho_1 > 0.3$.

Figure 6.3.3 Further KLS inference regarding over-identified specification (6.2)



However, all the above inferences on the demand for whiting seem rash, because a TSLS-based Chow breakpoint test half-way the sample applied to model (6.2) yields a p-value of 0.00!

7. Conclusion

Since the 1950's TSLS estimation has been the prominent workhorse in empirical econometrics for causal analysis of single linear relationships involving endogenous regressors. It is based on assuming uncorrelatedness between the error terms of the model and at least as many variables, not occurring in the model equation and therefore called external instruments, as there are endogenous explanatory variables in the model. The latter are or seem correlated with the error term for some reason or another. If there are $K_1 > 0$ endogenous regressors in the model, then, assuming that a subset of $L_2 \geq K_1$ of the instruments is valid (uncorrelated with the errors indeed), the validity of any further candidate instruments can be tested by including them as (non-endogenous) regressor into the model, applying TSLS, and testing whether their exclusion seems statistically acceptable. Thus, for K_1 instruments, testing their validity by TSLS is impossible, whereas their validity is a necessary requirement for testing the validity of any further candidate instruments. This is of course a very sad state of affairs.

In addition, for several decades it is widely known now, that when instruments are weak, meaning that when using them to fit the endogenous regressors this fit is actually rather poor, this has three serious consequences, namely: (a) the TSLS estimators are seriously biased, even in large samples; (b) their variance will be large; (c) moreover, the usual estimate of their variance is much too optimistic.

Only for the last mentioned problem the so-called weak-instrument asymptotic approaches have provided a cure. So, when instruments are weak, weak-instrument robust TSLS inference still suffers from bias and large variance, and often faces hard to refute criticism due to the untested proviso that all instruments are valid. Hence, a fundamentally different approach, based on out of the box thinking, seems to be called for.

To explain the variation in the dependent variable, TSLS uses just the variation in the regressors which can be expressed in terms of the available candidate instruments, in an attempt to get rid of the variation in the regressors which is correlated with the error term. Of course, this does not establish all variation in the regressors that is uncorrelated with the errors when the instruments are weak, and it completely falls through if some instruments are invalid (correlated with the disturbances themselves).

On the other hand, the KLS technique, further developed in the present study, does not need any instruments, because it aims to decompose the variation in the regressors directly into two components, one uncorrelated with the errors and the other proportional to the errors. The latter, not the former, causes all the trouble due to endogeneity of regressors, so there the focus should be. A consistent assessment of the component of the regressors that is infected by the error term is used to correct the inconsistent OLS estimator. This is only possible by making the far-fetched assumption that the investigator knows the degree of endogeneity of all regressors. Although this may be the case for some obviously predetermined or exogenous regressors, where it is zero, usually this is not the case for all explanatories. However, by varying the numerical assumptions regarding endogeneity over intervals thought reasonable, it is nevertheless possible to generate inference that is not suffering from the problems faced by TSLS: there is no need to use instruments, so their validity and strength are not an issue.

KLS estimators prove to be virtually unbiased in cross-section samples, and their finite sample bias in dynamic models seems primarily due just to presence of any predetermined regressors, whereas their variance estimates and the speed of their convergence towards normality are such that inference can claim high levels of accuracy. So, the historical overview by Epstein (1989), which preceded the weak-instrument era and is titled "The fall of OLS in structural estimation", may soon be followed up by a new gospel: The resurrection of Least-Squares in structural estimation. Afterall, KLS is a bias-corrected least-squares estimator, with the pecularity that the least-squares-based estimate of the bias (inconsistency) is of the same stochastic order as the least-squares coefficient estimator. Therefore, the KLS estimator has a different asymptotic variance than the uncorrected estimator, unlike estimators corrected for an O(n^{-1}) bias.

It is striking how little attention users of the TSLS technique have mostly given to the degree of endogeneity of the regressors in their models. When testing the significance of endogeneity of regressors usually two-sided test statistics are being used, because interest has always just been on absence or presence of endogenous regressors. From the illustrations in this paper it is clear that the authors of some of the articles that we used for replication studies should have concluded from their estimates of ρ_{xu}, or from the sign of the difference between TSLS and OLS estimates, that their empirical findings refute the theories and assumptions that they had started with. If they had only noticed these incompatibilities, they should have felt a need to reformulate their assumptions, theories or specifications. Moreover, although the asymptotic distribution of TSLS is invariant regarding ρ_{xu}, its finite sample distribution is certainly not. So, the focus on ρ_{xu}, which is at the heart of KLS, should also inspire users of TSLS to pay more attention to it.

As in Andrews, Stock and Sun (2019), who urge that weak instrument robust methods should be robustified regarding heteroskedasticity as well, likewise the instrument-free procedures presented here, which at present still presuppose homoskedasticity, should be robustified regarding heteroskedasticity; we are working on that. We do not suppose that a robustification with respect to serial correlation is called for too, because any predetermined regressors in an econometric time-series model would become endogenous under general forms of serial correlation. Suggesting reasonable intervals for their endogeneity coefficients would require knowledge which usually does not seem to be available. Hence, KLS should preferably be applied to models which are so complete that serial correlation has been avoided. Of course, though, serial correlation tests and also more general omitted variables tests, including coefficient constancy tests, should as a rule be employed to investigate and eventually overcome specification problems. All empirical applications that we considered were found to suffer from specification problems; Young (2019) too concludes that their occurrence seems ubiquitous in current practice. Presently, KLS has the potential to overcome only those specification problems which may render included regressors endogenous, while leaving the error terms i.i.d.

Acknowledgments

I'm grateful for enjoying the stimulating research environment provided by the UvA-Econometrics group and in particular for comments made on the present study by Jan Willem Gunning and Sebastian Kripfganz.

References

Andrews, D.W.K., Marmer, V., Yu, Z., 2019. On optimal inference in the linear IV model. *Quantitative Economics* 10, 457–485.

Andrews, I., Stock, J., Sun, L., 2019. Weak instruments in IV regression: Theory and practice. *Annual Review of Economics*. Forthcoming.

Angrist, J.D., Graddy, K., Imbens, G.W., 2000. The interpretation of Instrumental Variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies* 76, 499-527.

Bound, J., Jaeger, D., Baker, R., 1995. Problems with Instrumental Variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association* 90, 443-450.

Epstein, R.J., 1989. The fall of OLS in structural estimation. *Oxford Economic Papers* 41, 94-107.

Fisher, F.M., 1959. Generalisation of the rank and order conditions for identifiability. *Econometrica* 27, 431-447.

Fisher, F.M., 1963. Uncorrelated disturbances and identifiability criteria. *International Economic Review* 4, 134-152.

Frankel, J., Romer, D., 1999. Does trade cause growth? *American Economic Review* 89, 379-399.

Graddy, K., 1995. Testing for imperfect competition at the Fulton fish market. *RAND Journal of Economics* 26, 75-92.

Graddy, K., 2006. Markets: The Fulton fish market. *Journal of Economic Perspectives* 20, 207-220.

Graddy, K., Kennedy, P., 2010. When are supply and demand determined recursively rather than simultaneously? *Eastern Economic Journal* 36, 188-197.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., 2019. A more powerful subvector Anderson Rubin test in linear instrumental variables regression. *Quantitative Economics* 10, 487-526.

Hendry, D.F., Nielsen, B., 2006. *Econometric Modeling; A Likelihood Approach*. Princeton University Press, Princeton, USA.

Imbens, G.W., 2014. Instrumental Variables: An econometrician's perspective. *Statistical Science* 29, 323-358.

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., 2019. Testing the impossible: identifying exclusion restrictions. To appear in *The Journal of Econometrics.*

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., Phillips, G.D.A., 2012. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. *Journal of Computational Statistics & Data Analysis* 56, 3705-3729.

Koopmans, T. C., Rubin, H., Leipnik, R. B., 1950. Measuring the equation systems of dynamic economics. Chapter II in *Statistical Inference in Dynamic Economic Models* (T.C. Koopmans, ed.), New York, John Wiley and Sons: Cowles Commission Monograph No. 10.

Lewbel, A., 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics* 30, 67-80.

Murray, M.P., 2006. Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives* 20, 111-132.

Murray, M.P., 2017. Linear model IV estimation when instruments are many or weak. *Journal of Econometric Methods* 6, 1-22.

Phillips, G.D.A., Liu-Evans, G., 2016. Approximating and reducing bias in 2SLS estimation of dynamic simultaneous equation models. *Journal of Computational Statistics & Data Analysis* 100, 734-762.

Tanaka, T., Camerer, C.F., Nguyen, Q., 2010. Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review* 100, 557-571.

Wegge, L.L., 1965. Identifiability criteria for a system of equations as a whole. *Australian Journal of Statistics* 7, 67-77.

Young, A., 2019. Consistency without inference: Instrumental Variables in practical application. Unpublished manuscript.

Appendices

A. KLS asymptotic variance in the simple model

To find an asymptotic approximation to the distribution of $\hat{\beta}_{KLS}(\rho_{xu})$ of (3.11) we consider $n^{1/2}[\hat{\beta}_{KLS}(\rho_{xu}) - \beta]$. Using (3.7) we obtain

$$\begin{aligned} n^{1/2}[\hat{\beta}_{KLS}(\rho_{xu}) - \beta] &= n^{1/2}[n^{-1}\Sigma x_i u_i / n^{-1}\Sigma x_i^2 - \rho_{xu}\hat{\sigma}_u(\rho_{xu})/\hat{\sigma}_x] \\ &= \hat{\sigma}_x^{-2} n^{1/2}[n^{-1}\Sigma x_i u_i - \rho_{xu}\hat{\sigma}_x\hat{\sigma}_u(\rho_{xu})]. \end{aligned} \qquad \text{(A.1)}$$

To find its limiting distribution, we first have to derive the limiting distribution of $n^{1/2}[n^{-1}\Sigma x_i u_i - \rho_{xu}\hat{\sigma}_x\hat{\sigma}_u(\rho_{xu})]$. This is just determined by its largest components. These are $O_p(1)$, which expresses that they have a finite distribution. We may remove all smaller contributions, which are $o_p(1)$. We proceed as follows.

Because

$$n^{-1}\Sigma x_i u_i = \sigma_{xu} + (n^{-1}\Sigma x_i u_i - \sigma_{xu}), \qquad \text{(A.2)}$$

where $n^{-1}\Sigma x_i u_i - \sigma_{xu}$ has expectation zero and variance $n^{-2}\Sigma Var(x_i u_i) = O(n^{-1})$, we find $n^{-1}\Sigma x_i u_i - \sigma_{xu} = O_p(n^{-1/2})$, whereas $\sigma_{xu} = O(1)$. Similarly

$$\hat{\sigma}_x^2 = n^{-1}\Sigma x_i^2 = \sigma_x^2 + (n^{-1}\Sigma x_i^2 - \sigma_x^2) \qquad \text{(A.3)}$$

with $n^{-1}\Sigma x_i^2 - \sigma_x^2 = O_p(n^{-1/2})$. This yields the expansion

$$\hat{\sigma}_x = (n^{-1}\Sigma x_i^2)^{1/2} = \sigma_x + 2^{-1}\sigma_x^{-1}(n^{-1}\Sigma x_i^2 - \sigma_x^2) + o_p(n^{-1/2}), \qquad \text{(A.4)}$$

which is easily verified by squaring both sides. Moreover,

$$\hat{\sigma}_x^{-2} = (n^{-1}\Sigma x_i^2)^{-1} = \sigma_x^{-2} - \sigma_x^{-4}(n^{-1}\Sigma x_i^2 - \sigma_x^2) + o_p(n^{-1/2}). \qquad \text{(A.5)}$$

Its validity can be checked by multiplying by the decomposition (A.3) of $\hat{\sigma}_x^2$. Using (A.2) and (A.5) we find for (3.10)

$$\begin{aligned} \hat{\sigma}_u^2(\rho_{xu}) &= (1-\rho_{xu}^2)^{-1} n^{-1}\Sigma \hat{u}_i^2 \\ &= (1-\rho_{xu}^2)^{-1}[n^{-1}\Sigma u_i^2 - (n^{-1}\Sigma x_i u_i)^2/n^{-1}\Sigma x_i^2] \\ &= (1-\rho_{xu}^2)^{-1}(n^{-1}\Sigma u_i^2 - \sigma_u^2) + (1-\rho_{xu}^2)^{-1}\sigma_u^2 - (1-\rho_{xu}^2)^{-1} \\ &\qquad \times [\sigma_{xu}^2 + 2\sigma_{xu}(n^{-1}\Sigma x_i u_i - \sigma_{xu})][\sigma_x^{-2} - \sigma_x^{-4}(n^{-1}\Sigma x_i^2 - \sigma_x^2)] + o_p(n^{-1/2}) \\ &= \sigma_u^2 + (1-\rho_{xu}^2)^{-1}[(n^{-1}\Sigma u_i^2 - \sigma_u^2) - 2\rho_{xu}\sigma_u\sigma_x^{-1}(n^{-1}\Sigma x_i u_i - \sigma_{xu}) \\ &\qquad + \rho_{xu}^2\sigma_u^2\sigma_x^{-2}(n^{-1}\Sigma x_i^2 - \sigma_x^2)] + o_p(n^{-1/2}). \end{aligned}$$

This yields

$$\begin{aligned} \hat{\sigma}_u(\rho_{xu}) = \sigma_u + 0.5(1-\rho_{xu}^2)^{-1}[&\sigma_u^{-1}(n^{-1}\Sigma u_i^2 - \sigma_u^2) - 2\rho_{xu}\sigma_x^{-1}(n^{-1}\Sigma x_i u_i - \sigma_{xu}) \\ &+ \rho_{xu}^2\sigma_u\sigma_x^{-2}(n^{-1}\Sigma x_i^2 - \sigma_x^2)] + o_p(n^{-1/2}). \end{aligned} \qquad \text{(A.6)}$$

Now we can establish from the factor in square brackets of (A.1) its leading $O_p(n^{-1/2})$ terms, since

$$
\begin{aligned}
&n^{-1}\Sigma x_i u_i - \rho_{xu}\hat{\sigma}_u(\rho_{xu})\hat{\sigma}_x \\
&= (n^{-1}\Sigma x_i u_i - \sigma_{xu}) - 0.5\rho_{xu}\sigma_u\sigma_x^{-1}(n^{-1}\Sigma x_i^2 - \sigma_x^2) \\
&\quad - 0.5(1-\rho_{xu}^2)^{-1}\rho_{xu}\sigma_x[\sigma_u^{-1}(n^{-1}\Sigma u_i^2 - \sigma_u^2) - 2\rho_{xu}\sigma_x^{-1}(n^{-1}\Sigma x_i u_i - \sigma_{xu}) \\
&\qquad\qquad + \rho_{xu}^2\sigma_u\sigma_x^{-2}(n^{-1}\Sigma x_i^2 - \sigma_x^2)] + o_p(n^{-1/2}) \\
&= -0.5\rho_{xu}(1-\rho_{xu}^2)^{-1}[\sigma_u^{-1}\sigma_x(n^{-1}\Sigma u_i^2 - \sigma_u^2) + \sigma_u\sigma_x^{-1}(n^{-1}\Sigma x_i^2 - \sigma_x^2)] \\
&\quad + (1-\rho_{xu}^2)^{-1}(n^{-1}\Sigma x_i u_i - \sigma_{xu}) + o_p(n^{-1/2}).
\end{aligned}
$$

Substituting $x_i = \xi_i + \rho_{xu}\sigma_u^{-1}\sigma_x u_i$, this yields

$$
\begin{aligned}
n^{1/2}[n^{-1}\Sigma x_i u_i - \rho_{xu}\hat{\sigma}_u(\rho_{xu})\hat{\sigma}_x] &= 0.5\rho_{xu}\sigma_x\sigma_u^{-1}n^{1/2}(n^{-1}\Sigma u_i^2 - \sigma_u^2) + n^{1/2}(n^{-1}\Sigma\xi_i u_i) \\
&\quad - 0.5\rho_{xu}(1-\rho_{xu}^2)^{-1}\sigma_u\sigma_x^{-1}n^{1/2}(n^{-1}\Sigma\xi_i^2 - \sigma_\xi^2) \\
&\quad + o_p(1). \qquad (A.7)
\end{aligned}
$$

Its three leading terms establish a rescaled sample average of zero mean random elements, to which a central limit theorem applies yielding its limiting normal distribution.

To obtain the variance of this limiting distribution we have to derive the variance of the sum of these three leading terms. To find this we use that for $i = 1, ..., n$

$$Var(u_i^2 - \sigma_u^2) = (\kappa_u - 1)\sigma_u^4,\ Var(\xi_i u_i) = \sigma_\xi^2\sigma_u^2,\ Var(\xi_i^2 - \sigma_\xi^2) = (\kappa_\xi - 1)\sigma_\xi^4,$$

upon denoting $E(u_i^4) = \kappa_u\sigma_u^4$, $E(\xi_i^4) = \kappa_\xi\sigma_\xi^4$ and $E(x_i^4) = \kappa_x\sigma_x^4$. We also use that for $w_{1i} = u_i^2 - \sigma_u^2$, $w_{2i} = \xi_i u_i$ and $w_{3i} = \xi_i^2 - \sigma_\xi^2$, we have, for $j \neq k = 1, 2, 3$ and $i \neq t = 1, ..., n$, that

$$E(w_{ji}w_{ki}) = 0,\ E(w_{ji}w_{jt}) = 0 \text{ and } E(w_{ji}w_{kt}) = 0,$$

because, for instance, $E[(u_i^2 - \sigma_u^2)\xi_i u_i] = E(u_i^3\xi_i) = E[E(u_i^3\xi_i \mid u_i)] = E[u_i^3 E(\xi_i \mid u_i)] = E(0) = 0$, and, taking $i > t$, $E[(u_i^2 - \sigma_u^2)(u_t^2 - \sigma_u^2)] = E\{E[(u_i^2 - \sigma_u^2)(u_t^2 - \sigma_u^2) \mid u_t]\} = E\{(u_t^2 - \sigma_u^2)E[(u_i^2 - \sigma_u^2) \mid u_t]\} = 0$. Similar zero covariances are found for all further cross-terms.

So, we find for the variance of the sum of the three $O_p(1)$ terms of (A.7)

$$
\begin{aligned}
&0.25\rho_{xu}^2\sigma_x^2\sigma_u^{-2}(\kappa_u - 1)\sigma_u^4 + \sigma_\xi^2\sigma_u^2 + 0.25\rho_{xu}^2(1-\rho_{xu}^2)^{-2}\sigma_u^2\sigma_x^{-2}(\kappa_\xi - 1)\sigma_\xi^4 \\
&= \sigma_u^2[1 + \rho_{xu}^2(\kappa_\xi + \kappa_u - 6)/4]\sigma_x^2. \qquad (A.8)
\end{aligned}
$$

Respecting (A.1) this has to be multiplied by the square of plim $\hat{\sigma}_x^2$ in order to obtain the KLS limiting variance, giving

$$n^{1/2}[\hat{\beta}_{KLS}(\rho_{xu}) - \beta] \xrightarrow{d} \mathcal{N}\{0, [1 + \rho_{xu}^2(\kappa_\xi + \kappa_u - 6)/4]\sigma_u^2/\sigma_x^2\}. \qquad (A.9)$$

Since

$$
\begin{aligned}
E(x_i^4) &= E(\xi_i^4 + 4\lambda\xi_i^3 u_i + 6\lambda^2\xi_i^2 u_i^2 + 4\lambda^3\xi_i u_i^3 + \lambda^4 u_i^4) \\
&= [\kappa_\xi(1-\rho_{xu}^2)^2 + 6\rho_{xu}^2(1-\rho_{xu}^2) + \rho_{xu}^4\kappa_u]\sigma_x^4 = \kappa_x\sigma_x^4
\end{aligned}
$$

we find

$$\kappa_\xi = [\kappa_x - 6\rho_{xu}^2(1-\rho_{xu}^2) - \rho_{xu}^4\kappa_u]/(1-\rho_{xu}^2)^2, \tag{A.10}$$

so we can express (A.9) equivalently as

$$n^{1/2}[\hat{\beta}_{KLS}(\rho_{xu}) - \beta] \to \mathcal{N}\left(0, \frac{4 + (\kappa_u + \kappa_x - 14)\rho_{xu}^2 - 2(\kappa_u - 5)\rho_{xu}^4}{4(1-\rho_{xu}^2)^2}\frac{\sigma_u^2}{\sigma_x^2}\right), \tag{A.11}$$

which avoids using the kurtosis of the latent variable ξ_i.

For $\kappa_\xi = \kappa_u = 3$, implying $\kappa_x = 3$, which covers the case that $(x_i\ u_i)'$ is multivariate normal, this result specializes to that of Theorem 4.1 in Kiviet (2013, p.S44), also given as formula (2.11) in Kiviet (2016, p.194)[3] and as Corollary 1.2 in Kiviet (2019). These earlier results, however, had been derived just for i.i.d. cross-section observations. The derivation above shows that it is also valid for dependent data. Moreover, the results in this study indicate how the limiting variance of KLS changes for data with excess kurtosis.

B. Some auxiliary results

We prove here, all the time invoking a standard version of the central limit theorem, that the results (A.1) through (A.7) of Kiviet (2019) still hold under possible time dependence of the regressors.

The assumptions on u_i^2 did not change, so we still have

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

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

That we still have

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

giving $X'u/n - \sigma_{xu} = O_p(n^{-1/2})$, is again found by decomposing x_i into $x_i = \xi_i + \sigma_{xu}\sigma_u^{-2}u_i$, where $E(u_i \mid \xi_i) = 0$ and $E(u_i^2 \mid \xi_i) = \sigma_u^2$. Of course, $E(\xi_i) = 0$ and $E(\xi_iu_i) = 0$, so $E(x_iu_i) = \sigma_{xu}$ indeed. Since $\Sigma_{xx} = Var(\xi_i) + \sigma_u^{-2}\sigma_{xu}\sigma_{xu}'$, result (B.2) follows from $Var(x_iu_i - \sigma_{xu}) = E(u_i^2x_ix_i') - \sigma_{xu}\sigma_{xu}' = E(u_i^2\xi_i\xi_i') + \sigma_{xu}\sigma_{xu}'\sigma_u^{-4}E(u_i^4) - \sigma_{xu}\sigma_{xu}' = \sigma_u^2Var(\xi_i) + (\kappa_u - 1)\sigma_{xu}\sigma_{xu}' = \sigma_u^2\Sigma_{xx} + (\kappa_u - 2)\sigma_{xu}\sigma_{xu}'$, whereas under the adopted time dependence for $t < i$ we still have $E[(x_iu_i - \sigma_{xu})(x_tu_t - \sigma_{xu})'] = E(u_iu_tx_ix_t') - \sigma_{xu}\sigma_{xu}' = E(u_iu_t\xi_i\xi_t') + \sigma_u^{-4}E(u_i^2u_t^2)\sigma_{xu}\sigma_{xu}' - \sigma_{xu}\sigma_{xu}' = EE(u_iu_t\xi_i\xi_t' \mid u_t, \xi_t) = 0$.

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

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

This is proved by decomposing $x_{ik} = ax_{ij} + \eta_{ij}$, where $E(\eta_{ij} \mid x_{ij}) = 0$ and $E(\eta_{ij}^2 \mid x_{ij}) = \sigma_{\eta_j}^2$, thus η_{ij} has zero mean and unconditional variance $\sigma_{\eta_j}^2$. Because $E(x_{ik}^2) = \sigma_k^2 = a^2\sigma_j^2 + \sigma_{\eta_j}^2$ and $E(x_{ik}x_{ij}) = \sigma_{kj} = a\sigma_j^2$ we have $a = \sigma_{kj}\sigma_j^{-2}$ and $\sigma_{\eta_j}^2 = \sigma_k^2 - \sigma_{kj}^2\sigma_j^{-2}$. Now we obtain $E(x_{ij}^2x_{ik}^2) = E[x_{ij}^2(a^2x_{ij}^2 + 2ax_{ij}\eta_{ij} + \eta_{ij}^2)] = \kappa_x\sigma_{jk}^2 + \sigma_j^2(\sigma_k^2 - \sigma_{kj}^2\sigma_j^{-2}) =$

[3] In the derivations leading to this result there are two awkward typos in the result just above formula (2.7). This should read: $Var(\sigma_x^{-2}\omega_{x\varepsilon} - \sigma_{x\varepsilon}\sigma_x^{-4}\omega_{xx}) = n^{-1}[1 + (\kappa_x + \kappa_\varepsilon - 7)\rho_{x\varepsilon}^2 - 2(\kappa_\varepsilon - 3)\rho_{x\varepsilon}^4]\sigma_\varepsilon^2\sigma_x^{-2}$.

$\sigma_j^2\sigma_k^2+(\kappa_x-1)\sigma_{jk}^2$, thus $Var(x_{ij}x_{ik}-\sigma_{jk})=\sigma_j^2\sigma_k^2+(\kappa_x-2)\sigma_{jk}^2$, from which (B.3) follows, because $E[(x_{ij}x_{ik}-\sigma_{jk})(x_{tj}x_{tk}-\sigma_{jk})]=0$ for $i\neq t$. 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

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

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 $E(x_{ij}^2x_{ik}^2)$ just derived, we find $E[(x_{ij}^2-\sigma_j^2)(x_{ik}^2-\sigma_k^2)]=(\kappa_x-1)\sigma_{jk}^2$ and $E[(x_{ij}^2-\sigma_j^2)(x_{tk}^2-\sigma_k^2)]=0$ for $i\neq t$. Thus $E[(x_{ij}^2-\sigma_j^2)\rho_j(x_{ik}^2-\sigma_k^2)\rho_k]=(\kappa_x-1)\rho_j\sigma_{jk}^2\rho_k$ is the typical element of the limiting variance matrix of (B.4).

We also need the mutual covariances of scalar (B.1) and vectors (B.2) and (B.4). We find $E[(u_i^2-\sigma_u^2)(x_iu_i-\sigma_{xu})]=E[(u_i^2-\sigma_u^2)(\xi_iu_i+\sigma_{xu}\sigma_u^{-2}u_i^2-\sigma_{xu})]=(\kappa_u-1)\sigma_u^2\sigma_{xu}$ and $E[(u_i^2-\sigma_u^2)(x_tu_t-\sigma_{xu})]=0$ for $i\neq t$. Hence,

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

Using $x_{ij}=\xi_{ij}+\rho_j\sigma_j\sigma_u^{-1}u_i$, from $E[(u_i^2-\sigma_u^2)(x_{ij}^2-\sigma_j^2)\rho_j]=\rho_jE[(u_i^2-\sigma_u^2)(\xi_{ij}^2+2\sigma_j\sigma_u^{-1}u_i\xi_{ij}+\rho_j^2\sigma_j^2\sigma_u^{-2}u_i^2)]=(\kappa_u-1)\sigma_u^2\rho_j^3\sigma_j^2$ and $E[(u_i^2-\sigma_u^2)(x_{tj}^2-\sigma_j^2)\rho_j]=0$ for $i\neq t$, we find

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

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

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

C. Proof of the KLS Theorem

To find the limiting distribution of the inconsistency corrected OLS estimator $\hat{\beta}_{KLS}(\rho_{xu})=\hat{\beta}_{OLS}-n\cdot\hat{\sigma}_u(\rho_{xu})(X'X)^{-1}S_x\rho_{xu}$ we have to examine

$$n^{1/2}[\hat{\beta}_{KLS}(\rho_{xu})-\beta]=(n^{-1}X'X)^{-1}[n^{-1/2}X'u-n^{1/2}\hat{\sigma}_u(\rho_{xu})S_x\rho_{xu}].\tag{C.1}$$

In Appendix B of Kiviet (2019) it is shown that the factor in square brackets can be rewritten as

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

where $\theta = 1 - \rho'_{xu}\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\rho_{xu} > 0$. Its three terms with finite distributions are all sample averages of zero-mean random vectors, so a central limit theorem applies warranting its limiting normal distribution. Denoting its limiting variance matrix as $\sigma_u^2\Theta$ we have

$$n^{1/2}[\hat{\beta}_{KLS}(\rho_{xu}) - \beta] \xrightarrow{d} \mathcal{N}(0, \sigma_u^2\Sigma_{xx}^{-1}\Theta\Sigma_{xx}^{-1}). \tag{C.3}$$

Employing the asymptotic variances and covariances derived in Appendix B, and making use of the fact that diagonal matrices commute, we find for Θ the expression

$$\begin{aligned}
&(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})[\Sigma_{xx} + (\kappa_u - 2)\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x](I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)\\
&+ 0.25(\kappa_x - 1)(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)\\
&+ 0.25(\kappa_u - 1)\theta^{-2}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\\
&- (I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\Sigma_{xx}\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)\\
&- 0.5(\kappa_u - 3)(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)\\
&- (I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\mathcal{R}^2\Sigma_{xx}(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)\\
&- 0.5(\kappa_u - 3)(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\mathcal{R}^2\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)\\
&- 0.5\theta^{-1}(\kappa_u - 1)(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\\
&- 0.5\theta^{-1}(\kappa_u - 1)\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x)\\
&+ 0.25(\kappa_u - 1)\theta^{-1}(I + \theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\Sigma_{xx}^{-1})\mathcal{R}^2\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\\
&+ 0.25(\kappa_u - 1)\theta^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x).
\end{aligned}$$

To simplify this we denote $\Phi = \Sigma_x\rho_{xu}\rho'_{xu}\Sigma_x$ and use $\Phi\Sigma_{xx}^{-1}\Phi = (1-\theta)\Phi$, and find

$$\begin{aligned}
\Theta = \Sigma_{xx} &+ \theta^{-1}\Phi + (\kappa_u - 1)\theta^{-2}\Phi\\
&+ 0.25(\kappa_x - 1)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi)\\
&+ 0.25(\kappa_u - 1)\theta^{-2}\Phi\\
&- (I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_{xx}\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi)\\
&- 0.5(\kappa_u - 3)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Phi\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi)\\
&- (I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}^2\Sigma_{xx}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi)\\
&- 0.5(\kappa_u - 3)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}^2\Phi(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi)\\
&- 0.5\theta^{-1}(\kappa_u - 1)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Phi\\
&- 0.5\theta^{-1}(\kappa_u - 1)\Phi(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi)\\
&+ 0.25(\kappa_u - 1)\theta^{-1}(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}^2\Phi\\
&+ 0.25(\kappa_u - 1)\theta^{-1}\Phi\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi).
\end{aligned}$$

Making use of $(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Phi = \theta^{-1}\Phi$ this gives

$$
\begin{aligned}
\Theta = \Sigma_{xx} &+ \theta^{-1}\Phi + 1.25(\kappa_u - 1)\theta^{-2}\Phi \\
&+ 0.25(\kappa_x - 1)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi) \\
&- (I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_{xx}\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi) \\
&- 0.5(\kappa_u - 3)\theta^{-1}\Phi\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi) \\
&- (I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}^2\Sigma_{xx}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi) \\
&- 0.5(\kappa_u - 3)\theta^{-1}(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}^2\Phi \\
&- \theta^{-2}(\kappa_u - 1)\Phi \\
&+ 0.25(\kappa_u - 1)\theta^{-1}(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\mathcal{R}^2\Phi \\
&+ 0.25(\kappa_u - 1)\theta^{-1}\Phi\mathcal{R}^2(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi).
\end{aligned}
$$

$$
\begin{aligned}
= \Sigma_{xx} &+ \theta^{-1}\Phi + 0.25(\kappa_u - 1)\theta^{-2}\Phi \\
&+ 0.25(\kappa_x - 1)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi) \\
&- \Sigma_{xx}\mathcal{R}^2 - \theta^{-1}\Phi\mathcal{R}^2 - \theta^{-1}\Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi - \theta^{-2}\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi \\
&- 0.5(\kappa_u - 3)\theta^{-1}\Phi\mathcal{R}^2 - 0.5(\kappa_u - 3)\theta^{-2}\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi \\
&- \mathcal{R}^2\Sigma_{xx} - \theta^{-1}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx} - \theta^{-1}\mathcal{R}^2\Phi - \theta^{-2}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi \\
&- 0.5(\kappa_u - 3)\theta^{-1}\mathcal{R}^2\Phi - 0.5(\kappa_u - 3)\theta^{-2}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi \\
&+ 0.25(\kappa_u - 1)\theta^{-1}\mathcal{R}^2\Phi + 0.25(\kappa_u - 1)\theta^{-2}\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi \\
&+ 0.25(\kappa_u - 1)\theta^{-1}\Phi\mathcal{R}^2 + 0.25(\kappa_u - 1)\theta^{-2}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi,
\end{aligned}
$$

$$
\begin{aligned}
= \Sigma_{xx} &+ \theta^{-1}\Phi + 0.25(\kappa_u - 1)\theta^{-2}\Phi \\
&- \Sigma_{xx}\mathcal{R}^2 - \mathcal{R}^2\Sigma_{xx} \\
&- \theta^{-1}\mathcal{R}^2\Phi - 0.5(\kappa_u - 3)\theta^{-1}\mathcal{R}^2\Phi + 0.25(\kappa_u - 1)\theta^{-1}\mathcal{R}^2\Phi \\
&- \theta^{-1}\Phi\mathcal{R}^2 - 0.5(\kappa_u - 3)\theta^{-1}\Phi\mathcal{R}^2 + 0.25(\kappa_u - 1)\theta^{-1}\Phi\mathcal{R}^2 \\
&- \theta^{-1}\Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi - \theta^{-1}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx} \\
&- \theta^{-2}\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi - 0.5(\kappa_u - 3)\theta^{-2}\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi + 0.25(\kappa_u - 1)\theta^{-2}\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi \\
&- \theta^{-2}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi - 0.5(\kappa_u - 3)\theta^{-2}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi + 0.25(\kappa_u - 1)\theta^{-2}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi \\
&+ 0.25(\kappa_x - 1)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi),
\end{aligned}
$$

$$
\begin{aligned}
= \Sigma_{xx} &+ \theta^{-1}[1 + 0.25(\kappa_u - 1)\theta^{-1}]\Phi \\
&- \Sigma_{xx}\mathcal{R}^2 - \mathcal{R}^2\Sigma_{xx} - 0.25(\kappa_u - 1)\theta^{-1}(\mathcal{R}^2\Phi + \Phi\mathcal{R}^2) \\
&- \theta^{-1}\Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi - \theta^{-1}\Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx} \\
&- 0.25(\kappa_u - 1)\theta^{-2}\Phi(\mathcal{R}^2\Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}\mathcal{R}^2)\Phi \\
&+ 0.25(\kappa_x - 1)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi).
\end{aligned}
$$

Further simplification, and employing

$$
\begin{aligned}
\Phi\mathcal{R}^2\Sigma_{xx}^{-1}\Phi &= \Sigma_x\rho_{xu}\rho_{xu}'\Sigma_x\mathcal{R}^2\Sigma_{xx}^{-1}\Sigma_x\rho_{xu}\rho_{xu}'\Sigma_x \\
&= (\rho_{xu}'\Sigma_x\mathcal{R}^2\Sigma_{xx}^{-1}\Sigma_x\rho_{xu})\Sigma_x\rho_{xu}\rho_{xu}'\Sigma_x \\
&= (\rho_{xu}'\mathcal{R}\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\mathcal{R}\rho_{xu})\Phi = \Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Phi,
\end{aligned}
$$

yields

$$\begin{aligned}\Theta &= \Sigma_{xx} - (\Sigma_{xx}\mathcal{R}^2 + \mathcal{R}^2\Sigma_{xx}) \\ &\quad + [1 + 0.25(\kappa_u - 1)\theta^{-1}(1 - 2\rho'_{xu}\mathcal{R}\Sigma_x\Sigma_{xx}^{-1}\Sigma_x\mathcal{R}\rho_{xu})]\theta^{-1}\Phi \\ &\quad - 0.25(\kappa_u - 1)\theta^{-1}(\mathcal{R}^2\Phi + \Phi\mathcal{R}^2) \\ &\quad - \theta^{-1}(\Sigma_{xx}\mathcal{R}^2\Sigma_{xx}^{-1}\Phi + \Phi\Sigma_{xx}^{-1}\mathcal{R}^2\Sigma_{xx}) \\ &\quad + 0.25(\kappa_x - 1)(I + \theta^{-1}\Phi\Sigma_{xx}^{-1})\Sigma_x^{-1}\mathcal{R}(\Sigma_{xx} \circ \Sigma_{xx})\mathcal{R}\Sigma_x^{-1}(I + \theta^{-1}\Sigma_{xx}^{-1}\Phi).\end{aligned}$$

D. KLS variance estimation in the model with $K_1 = 1$

When the first column of X contains the one and only endogenous regressor, whereas its correlation with the disturbance is ρ_1, then, using e_1 to denote a $K \times 1$ vector with a unit element in the first position and all others equal to zero, we have $R = \rho_1 e_1 e_1'$, $\Phi = e_1 e_1' \rho_1^2 \sigma_1^2$, $\theta = 1 - \rho_1^2\phi_1$ and $\phi_1 = \sigma_1^2 e_1' \Sigma_{xx}^{-1} e_1$. Using these when evaluating $e_1'\Sigma_{xx}^{-1}\Theta\Sigma_{xx}^{-1}e_1$ one finds this to be proportional to $1/\sigma_1^2$ with factor

$$\begin{aligned}&(1 - \rho_1^2)\phi_1 - 2\theta^{-1}\rho_1^4\phi_1^2 + \theta^{-1}\rho_1^2\phi_1^2 - 0.5(\kappa_u - 1)\theta^{-1}\rho_1^4\phi_1^2 \\ &- 0.25(\kappa_u - 1)\theta^{-2}(1 - 2\rho_1^4\phi_1)\rho_1^2\phi_1^2 + 0.25(\kappa_x - 1)\rho_1^2(\phi_1 + \theta^{-1}\rho_1^2\phi_1^2)^2,\end{aligned}$$

and this is found to be proportional to $\phi_1/(4\theta^2)$ with factor

$$-2 + 2\rho_1^2 + \theta(6 - \rho_1^2) + \kappa_u(1 - \theta)(1 - 2\rho_1^2) + \kappa_x\phi_1(\theta + 1).$$

Hence,

$$e_1'\Sigma_{xx}^{-1}\Theta\Sigma_{xx}^{-1}e_1 = \frac{\phi_1[4 - 8\rho_1^2 - 6\phi_1\rho_1^2 + 10\rho_1^4\phi_1^2 + \kappa_u(\rho_1^2\phi_1 - 2\rho_1^4\phi_1^2) + \kappa_x\rho_1^2\phi_1]}{4\theta^2\sigma_1^2},$$

from which result (4.3) follows.

E. On the power of the KLS instrument validity test

We focus on the situation sketched at the end of Section 4 for the case $K_1 = K_3 = 1$ and $K_2 = 0$. The latter assumption is not restrictive because we may assume that further relevant predetermined regressors have been partialled out. From the general formula for $\hat{\beta}_{KLS}(\rho_{xu})$ we find that in this special augmented regression we have

$$\begin{aligned}\hat{\beta}_{3,KLS}(\rho_1) &= \hat{\beta}_{3,OLS} - \begin{pmatrix} 0 & 1 \end{pmatrix}\hat{\sigma}_{u^*}(\rho_1)\begin{pmatrix} s_{11} & s_{13} \\ s_{31} & s_{33} \end{pmatrix}^{-1}\begin{pmatrix} s_1 & 0 \\ 0 & s_3 \end{pmatrix}\begin{pmatrix} \rho_1 \\ 0 \end{pmatrix} \\ &= \hat{\beta}_{3,OLS} - \hat{\sigma}_{u^*}(\rho_1)s_1\rho_1[-s_{13}/(s_{11}s_{33} - s_{13}^2)].\end{aligned}$$

We will derive the probability limits of $\hat{\beta}_{3,KLS}(\rho_1)$ and of $\hat{\beta}_{3,KLS}(\hat{\rho}_1)$, where $\hat{\rho}_1$ results from IV estimation. Consider first

$$\hat{\beta}_{3,OLS} + \hat{\sigma}_{u^*}(\hat{\rho}_1)(n^{-1}x_1'x_1)^{1/2}\hat{\rho}_1\{n^{-1}x_1'x_3/[(n^{-1}x_1'x_1)(n^{-1}x_3'x_3) - (n^{-1}x_1'x_3)^2]\}.$$

Using $M_1 = I - x_1(x_1'x_1)^{-1}x_1'$, $M_{1,3} = M_1 - M_1x_3(x_3'M_1x_3)^{-1}x_3'M_1$ and denoting $E(x_{i3}u_i) = \rho_3\sigma_3\sigma_u$, $E(x_{i1}x_{i3}) = \rho_{13}\sigma_1\sigma_3$, we find, assuming $\rho_{13} \neq 0$ (variable x_3 is a relevant instrument for x_1):

$$\begin{aligned}
\hat{\beta}_{3,OLS} &= (x_3'M_1x_3)^{-1}x_3'M_1y = (x_3'M_1x_3)^{-1}x_3'M_1(x_1\beta_1 + u) \\
&= (n^{-1}x_3'M_1x_3)^{-1}n^{-1}x_3'M_1u \to \frac{\rho_3 - \rho_{13}\rho_1}{1-\rho_{13}^2}\frac{\sigma_u}{\sigma_3}, \\
\hat{u}_{IV} &= y - x_1\hat{\beta}_{IV} = y - (x_3'y/x_3'x_1)x_1 = u - (x_3'u/x_3'x_1)x_1 \to u - \frac{\sigma_u}{\sigma_1}\frac{\rho_3}{\rho_{13}}x_1, \\
n^{-1}\hat{u}_{IV}'\hat{u}_{IV} &= n^{-1}u'u - 2\frac{\sigma_u}{\sigma_1}\frac{\rho_3}{\rho_{13}}n^{-1}u'x_1 + \frac{\sigma_u^2}{\sigma_1^2}\frac{\rho_3^2}{\rho_{13}^2}n^{-1}x_1'x_1 \\
&\to \sigma_u^2\rho_{13}^{-2}(\rho_{13}^2 - 2\rho_1\rho_3\rho_{13} + \rho_3^2), \\
n^{-1}x_1'\hat{u}_{IV} &= n^{-1}x_1'u - \frac{\sigma_u}{\sigma_1}\frac{\rho_3}{\rho_{13}}n^{-1}x_1'x_1 \to \sigma_u\sigma_1(\rho_1 - \rho_3/\rho_{13}), \\
\hat{\rho}_1 &= (n^{-1}x_1'\hat{u}_{IV})/[(n^{-1}x_1'x_1)(n^{-1}\hat{u}_{IV}'\hat{u}_{IV})]^{1/2} \\
&\to (\rho_1\rho_{13} - \rho_3)[\rho_{13}^2 - 2\rho_1\rho_3\rho_{13} + \rho_3^2]^{-1/2}, \\
n^{-1}x_1'M_3x_1 &\to \sigma_1^2(1-\rho_{13}^2), \\
n^{-1}y'M_{1,3}y &\to \sigma_u^2[1 - \rho_1^2 - (\rho_3 - \rho_1\rho_{13})^2/(1-\rho_{13}^2)], \\
\hat{\sigma}_{u^*}^2(\hat{\rho}_1) &= (n^{-1}y'M_{1,3}y)/[1 - \hat{\rho}_1^2(n^{-1}x_1'x_1)/(n^{-1}x_1'M_3x_1)^{-1}] \\
&\to \sigma_u^2(1 - 2\rho_1\rho_3/\rho_{13} + \rho_3^2/\rho_{13}^2).
\end{aligned}$$

Hence, because $\hat{\sigma}_{u^*}(\hat{\rho}_1)\hat{\rho}_1 \to \sigma_u(\rho_1 - \rho_3/\rho_{13})$ and

$$n^{-1}x_1'x_3/[(n^{-1}x_1'x_1)(n^{-1}x_3'x_3) - (n^{-1}x_1'x_3)^2] \to \frac{1}{\sigma_1\sigma_3}\frac{\rho_{13}}{1-\rho_{13}^2},$$

we obtain

$$\hat{\beta}_{3,KLS}(\hat{\rho}_1) \to \frac{\sigma_u}{\sigma_3}\left\{\frac{\rho_3 - \rho_{13}\rho_1}{1-\rho_{13}^2} + (\rho_1 - \rho_3/\rho_{13})\frac{\rho_{13}}{1-\rho_{13}^2}\right\} = 0,$$

irrespective of the actual value of ρ_3. Whereas, because $\hat{\sigma}_{u^*}^2(\rho_1) \to \sigma_u^2$,

$$\hat{\beta}_{3,KLS}(\rho_1) \to \frac{\sigma_u}{\sigma_3}\left\{\frac{\rho_3 - \rho_{13}\rho_1}{1-\rho_{13}^2} + \frac{\rho_1\rho_{13}}{1-\rho_{13}^2}\right\} = \frac{\sigma_u}{\sigma_3}\frac{\rho_3}{1-\rho_{13}^2},$$

which is zero just when $\rho_3 = 0$, which makes x_3 a valid instrument. Note that ρ_{13} close to zero (x_3 a weak instrument for x_1) does not upset these results.