

Discussion Paper: 2015/05

When is it really justifiable to ignore explanatory variable endogeneity in a regression model?

Jan F. Kiviet

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics

Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA  UNIVERSITEIT VAN AMSTERDAM



When is it really justifiable to ignore explanatory variable endogeneity in a regression model?

Jan F. Kiviet*

Version of 5 December 2015

JEL Classifications: *C2, C13, O5*

Keywords:

asymptotics, endogeneity, exogeneity, growth, robust OLS inference

Abstract

A conversion of standard ordinary least-squares results into inference which is robust under endogeneity of some regressors has been put forward in Ashley and Parmeter, *Economics Letters*, 137 (2015) 70-74. However, their conversion is based on an incorrect (though by accident conservative) asymptotic approximation and entails a neglected but avoidable randomness. By a very basic example it is illustrated why a much more sophisticated asymptotic expansion under a stricter set of assumptions is required than used by these authors. Next, particular aspects of their consequently flawed sensitivity analysis for an empirical growth model are replaced by results based on a proper limiting distribution for a feasible inconsistency corrected least-squares estimator. Finally we provide references to literature where relevant asymptotic approximations have been derived which should enable to produce similar endogeneity robust inference for more general models and hypotheses than currently available.

1. Introduction

An attempt is made in Ashley and Parmeter (2015a), henceforth APLS, to convert ordinary least-squares (OLS) inference such that it becomes robust in some sense with respect to simultaneity. The methods used in APLS are basically a specialization of an approach put forward in Ashley and Parmeter (2015b), henceforth APIV, which aims to robustify instrumental variables (IV) inference for the situation where instruments are

*Emeritus Professor of Econometrics, Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 NJ Amsterdam, The Netherlands (j.f.kiviet@uva.nl) and Visting Professor at the Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332 (jfkiviet@ntu.edu.sg).

in fact endogenous. The APLS results are obtained by simply adopting in the APIV approach the regressors as instruments.

In Kiviet (2013), henceforth KLS, we pursued similar goals as APLS by developing an asymptotically valid inference method on the basis of inconsistent OLS results, through achieving identification by making an assumption on a nonzero moment condition, instead of the habitual zero moment conditions exploited by consistent IV. APLS use a different conversion of OLS inference than put forward in KLS, but both approaches depart from the same starting point, namely an assessment of the limiting distribution of an unfeasible estimator which corrects OLS for its inconsistency. In this paper we will focus on the results in APLS and show that its asymptotic analysis is much too naive, and that their approach unnecessarily leads to an assessment of robustness over a set of possible degrees of endogeneity which is random instead of deterministic as in KLS.

By analyzing in all detail a very simple case we will show in Section 2 that obtaining the required asymptotic results is much more involved than suggested in APLS, and that feasible asymptotically valid OLS-based inference can be produced over an arbitrary chosen range of degrees of simultaneity. However, a valid operational technique has yet only been achieved in KLS for the case where just one regressor is endogenous. For that particular situation we produce in Section 3 robust and asymptotically valid inference for the growth data analyzed also in APLS. In the concluding Section 4 we put the foregoing into perspective, and indicate some literature that seems useful for achieving the goals of robustifying OLS and IV with respect to invalid orthogonality conditions for models with possibly more than just one endogenous regressor.

2. Examination of a simple case

Consider the very simple linear regression model with just one regressor and zero mean serially uncorrelated and homoskedastic disturbances, hence

$$y = x\beta + \varepsilon \text{ with } \varepsilon \sim (0, \sigma_\varepsilon^2 I), \quad (2.1)$$

where both y and x are $n \times 1$ observed data vectors. Regressor x could be endogenous, therefore we suppose that

$$x = \mu_x + \lambda\varepsilon, \quad (2.2)$$

where $\mu_x = E(x)$ is an arbitrary nonrandom $n \times 1$ vector. For $\lambda = 0$ regressor x is exogenous and for $\lambda \neq 0$ endogenous. We make standard regularity assumptions, involving that the unknown scalars β , λ and σ_ε are finite, and so are all elements of μ_x . Supposing $\mu_x \neq 0$, we have $x'x > 0$, and the ordinary least-squares estimator for β is

$$b = x'y/x'x = \beta + x'\varepsilon/x'x. \quad (2.3)$$

We shall examine its limiting behavior for $n \rightarrow \infty$.

2.1. Preparations

From $E(x'x) = \mu'_x \mu_x + \lambda^2 n \sigma_\varepsilon^2$ and by denoting $v = \varepsilon' \varepsilon - n \sigma_\varepsilon^2$ we obtain

$$\begin{aligned} n^{1/2}(b - \beta) &= (x'x/n)^{-1} x' \varepsilon / n^{1/2} \\ &= (\mu'_x \mu_x / n + 2\lambda \mu'_x \varepsilon / n + \lambda^2 \varepsilon' \varepsilon / n)^{-1} (\mu'_x \varepsilon / n^{1/2} + \lambda \varepsilon' \varepsilon / n^{1/2}) \\ &= [E(x'x)/n + 2\lambda \mu'_x \varepsilon / n + \lambda^2 v / n]^{-1} [\mu'_x \varepsilon / n^{1/2} + \lambda v / n^{1/2} + n^{1/2} \lambda \sigma_\varepsilon^2]. \end{aligned} \quad (2.4)$$

This is the ratio of two factors and each factor has three scalar terms. Some of these are nonrandom and the others are random with zero mean. We can determine their order (of probability). For $E(x'x)/n = \mu'_x \mu_x / n + \lambda^2 \sigma_\varepsilon^2$ we find that it will be deterministic and finite, irrespective of the magnitude of n . This is indicated as $E(x'x)/n = O(1)$ or $O(n^0)$. From $\mu'_x \varepsilon \sim (0, \sigma_\varepsilon^2 \mu'_x \mu_x)$ where $\sigma_\varepsilon^2 \mu'_x \mu_x = O(n)$ we find $\mu'_x \varepsilon = O_p(n^{1/2})$ thus $2\lambda \mu'_x \varepsilon / n = O_p(n^{-1/2})$ whereas $\mu'_x \varepsilon / n^{1/2} = O_p(1)$ has a so-called finite distribution. From $v = \varepsilon' \varepsilon - n \sigma_\varepsilon^2 \sim (0, \kappa n \sigma_\varepsilon^4)$, where the kurtosis κ would be 2 if ε were multivariate normal, we obtain $v = O_p(n^{1/2})$, assuming that the regularity also includes that κ is finite. Thus, $\lambda^2 v / n = O_p(n^{-1/2})$ and $\lambda v / n^{1/2} = O_p(1)$, and of course $n^{1/2} \lambda \sigma_\varepsilon^2 = O(n^{1/2})$.

Note that in the factor of (2.4) that has to be inverted we have one $O(1)$ term and two $O_p(n^{-1/2})$ terms. Let us indicate them by $c = E(x'x)/n = O(1)$ and $d = 2\lambda \mu'_x \varepsilon / n + \lambda^2 v / n = O_p(n^{-1/2})$ and consider the following expansion and approximation (which uses $c > 0$)

$$\begin{aligned} (c + d)^{-1} &= c^{-1} (1 + d/c)^{-1} \\ &= c^{-1} (1 - d/c + d^2/c^2 - d^3/c^3 + \dots) \\ &= c^{-1} - dc^{-2} + O_p(n^{-1}). \end{aligned} \quad (2.5)$$

Hence, this takes for granted that all the omitted terms, which are of decreasing order, have a sum of the same order as the first and largest omitted term, which is $d^2/c^3 = O_p(n^{-1})$.

The other factor of (2.4) has two $O_p(1)$ terms and one $O(n^{1/2})$. Let us indicate them by $f = \mu'_x \varepsilon / n^{1/2} + \lambda v / n^{1/2} = O_p(1)$ and $g = n^{1/2} \lambda \sigma_\varepsilon^2 = O(n^{1/2})$. Now employing (2.5) we obtain for (2.4)

$$\begin{aligned} (c + d)^{-1} (f + g) &= (c^{-1} - dc^{-2}) (f + g) + O_p(n^{-1/2}) \\ &= g/c + f/c - dg/c^2 + O_p(n^{-1/2}). \end{aligned} \quad (2.6)$$

Since $df/c^2 = O_p(n^{-1/2})$ it can be absorbed into the remainder term which is of the same order, because it follows from multiplying the remainder term of (2.5) by the largest of f and g , which is g . Note that $g/c = O(n^{1/2})$ is deterministic, whereas the two terms $f/c = O_p(1)$ and $dg/c^2 = O_p(1)$ are of similar stochastic order. It is this latter term which is not respected by APLS in their formula (7).

Substituting (2.6) into (2.4) yields $n^{1/2}(b - \beta) - g/c = f/c - dg/c^2 + O_p(n^{-1/2})$. When writing $E(x'x/n) = q_{xx} > 0$ this gives

$$\begin{aligned} n^{1/2}(b - \beta - \lambda\sigma_\varepsilon^2/q_{xx}) &= [\mu'_x\varepsilon/n^{1/2} + \lambda v/n^{1/2}]/q_{xx} \\ &\quad - \lambda\sigma_\varepsilon^2[2\lambda\mu'_x\varepsilon/n^{1/2} + \lambda^2 v/n^{1/2}]/q_{xx}^2 + O_p(n^{-1/2}). \end{aligned}$$

Also writing $E(x'\varepsilon/n) = q_{x\varepsilon} = \lambda\sigma_\varepsilon^2$ and $\rho = q_{x\varepsilon}/(\sigma_\varepsilon^2 q_{xx})^{1/2}$ this can be expressed as

$$\begin{aligned} n^{1/2}(b - \beta - q_{x\varepsilon}/q_{xx}) &= (1 - 2\rho^2)q_{xx}^{-1}\mu'_x\varepsilon/n^{1/2} \\ &\quad + \rho(1 - \rho^2)(\sigma_\varepsilon^2 q_{xx})^{-1/2}v/n^{1/2} + O_p(n^{-1/2}). \end{aligned} \quad (2.7)$$

2.2. Limiting distributions

Obviously, the expectation of the two $O_p(1)$ terms of (2.7) is zero. To find the variance of their sum the third and fourth moments of the disturbances are required. Assuming that the disturbances are normal, this gives $E(v\varepsilon) = 0$ and $E(v^2) = 2\sigma_\varepsilon^4$. Substituting $\mu'_x\mu_x/n = q_{xx} - \lambda^2\sigma_\varepsilon^2 = q_{xx} - q_{x\varepsilon}^2/\sigma_\varepsilon^2 = (1 - \rho^2)q_{xx}$, a neat expression for this variance can be found and, under all the assumptions made, a Central Limit Theorem can be invoked yielding the limiting distribution

$$n^{1/2}(b - \beta - q_{x\varepsilon}/q_{xx}) \xrightarrow{d} \mathcal{N}(0, (1 - \rho^2)(1 - 2\rho^2 + 2\rho^4)\sigma_\varepsilon^2 q_{xx}^{-1}). \quad (2.8)$$

APLS use in their formula (7) the same infeasible inconsistency expression $q_{x\varepsilon}/q_{xx}$ for the correction of the OLS estimator. However, the naively chosen (standard) expression $\sigma_\varepsilon^2 q_{xx}^{-1}$ for the variance of its limiting distribution is too conservative, as $(1 - \rho^2)(1 - 2\rho^2 + 2\rho^4) \leq 1$ for $|\rho| < 1$.

In KLS it is highlighted that limiting distribution (2.8) is obtained by conditioning on μ_x . In an unconditional setting (assuming μ_x , like ε , to be random and normal, and also independent from ε) the limiting distribution of the same infeasible estimator has a larger variance and is given by

$$n^{1/2}(b - \beta - q_{x\varepsilon}/q_{xx}) \xrightarrow{d} \mathcal{N}(0, (1 - \rho^2)\sigma_\varepsilon^2 q_{xx}^{-1}). \quad (2.9)$$

For that case it has also been derived in KLS that for the feasible consistent estimator b_ρ^* , which is obtained by correcting b using an assumed value of ρ , one has

$$n^{1/2}(b_\rho^* - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2 q_{xx}^{-1}), \quad (2.10)$$

where

$$\begin{aligned} b_\rho^* &= b - n^{1/2}\rho(1 - \rho^2)^{-1/2}se(b), \text{ with} \\ se(b) &= s/(x'x)^{1/2} \text{ and } s^2 = (y - xb)'(y - xb)/(n - k). \end{aligned} \quad (2.11)$$

Here $se(b)$ is just the expression for the usual OLS standard error estimate of b as produced under assumed exogeneity of x . In the present special model $k = 1$ but for the asymptotic result to hold a degrees of freedom correction is irrelevant, of course. Estimator b_ρ^* is consistent for β , because $q_{x\varepsilon}/q_{xx} = \rho(\sigma_\varepsilon^2/q_{xx})^{1/2}$ whereas $x'x/n$ is consistent for q_{xx} and $s^2/(1 - \rho^2)$ is a consistent estimator of σ_ε^2 . Note that $n^{1/2}se(b) = O_p(1)$ and therefore the inconsistency correction term in b_ρ^* is finite too and vanishes only for $\rho = 0$.

One may perhaps find it inappropriate that estimator b_ρ^* is addressed as a feasible estimator, whereas in practice the value of ρ is commonly unknown. A not unreasonable response to that is: OLS/IV estimators are usually not labelled as unfeasible either, whereas their underlying just identifying orthogonality conditions simply adopt the value zero for ρ , for which it is equally difficult or sheer impossible to find statistical evidence, see Kiviet (2015).

Result (2.10) is quite remarkable, not because consistent feasible estimator b_ρ^* is found to have (due to estimating $q_{x\varepsilon}/q_{xx}$) a larger asymptotic variance than the unfeasible estimator $b - q_{x\varepsilon}/q_{xx}$, but because this increment exactly leads to the same asymptotic variance as b has when it is consistent. However, this equivalence only holds (unconditionally) under joint normality of x and ε ; it can be shown that the asymptotic variance of b_ρ^* will be larger in case ε has excess kurtosis.

2.3. Endogeneity robust OLS inference

To produce asymptotically valid inference based on (2.10) we use the asymptotic approximation $b_\rho^* \stackrel{a}{\sim} \mathcal{N}(\beta, s^2/[(1 - \rho^2)x'x])$. Given the value of ρ , a confidence interval for β with asymptotic confidence level $(1 - \alpha)100\%$ is given by

$$\{b_\rho^* + z_{\alpha/2}s[(1 - \rho^2)x'x]^{-1/2}, b_\rho^* + z_{1-\alpha/2}s[(1 - \rho^2)x'x]^{-1/2}\},$$

where z_p denotes for $0 < p < 1$ the p^{th} quantile of the standard normal distribution. After substitution of (2.11) this can also be expressed as

$$[b + se(b)(z_{\alpha/2} - n^{1/2}\rho)(1 - \rho^2)^{-1/2}, b + se(b)(z_{1-\alpha/2} - n^{1/2}\rho)(1 - \rho^2)^{-1/2}]. \quad (2.12)$$

Note that the interval is not symmetric around b , and that its width is equal to $(z_{1-\alpha/2} - z_{\alpha/2})(1 - \rho^2)^{-1/2}se(b)$. Hence, it will not only be wider for smaller α (as is always the case) and for smaller n (because $se(b) = O_p(n^{-1/2})$), but also for larger $|\rho|$, and even be infinitely wide for $|\rho| \rightarrow 1$. The latter sheds doubts on the conclusion in APLS that they did establish OLS inferences which were found to be robust to any degree of simultaneity.

The essential elements of the APLS approach in the context of the above simple one coefficient model involves the following. Its focus is on hypothesis testing, say $\mathcal{H}_0 : \beta = \beta_0$ against $\mathcal{H}_1 : \beta \neq \beta_0$ for known numerical value β_0 . Suppose for

a particular data set one has found $|(b - \beta_0)/se(b)| > z_{1-\alpha/2}$, so \mathcal{H}_0 is rejected at level α , presupposing exogeneity. Now the method seeks to establish the values ρ for which \mathcal{H}_0 would still be rejected. This is pursued as follows. The set of values for $q_{x\varepsilon}$, say $\mathcal{S}(q_{x\varepsilon})$, is assessed (by a random search procedure) for which $\mathcal{S}(q_{x\varepsilon}) = \{q_{x\varepsilon} \in \mathbb{R} : z_{\alpha/2} < (b_{q_{x\varepsilon}}^* - \beta_0)/se(b_{q_{x\varepsilon}}^*) < z_{1-\alpha/2}\}$, where

$$\begin{aligned} b_{q_{x\varepsilon}}^* &= b - q_{x\varepsilon}/(x'x/n) \text{ with} \\ se(b_{q_{x\varepsilon}}^*) &= \hat{\sigma}_{q_{x\varepsilon}}/(x'x)^{1/2} \text{ and } \hat{\sigma}_{q_{x\varepsilon}}^2 = (y - xb_{q_{x\varepsilon}}^*)'(y - xb_{q_{x\varepsilon}}^*)/n. \end{aligned} \quad (2.13)$$

When the numerical problem to assess $\mathcal{S}(q_{x\varepsilon})$ has been solved, the corresponding set of values for $\hat{\rho}$ given by $\mathcal{S}(\hat{\rho}) = \{\hat{\rho} = q_{x\varepsilon}/[\hat{\sigma}_{q_{x\varepsilon}}^2(x'x/n)]^{1/2} : q_{x\varepsilon} \in \mathcal{S}(q_{x\varepsilon})\}$ is assessed, and the conclusion is drawn that rejection of \mathcal{H}_0 is robust with respect to endogeneity for all values $\rho \in \mathcal{S}(\hat{\rho})$.

Hence, a crucial difference with KLS is that a choice is made regarding $q_{x\varepsilon}$ and not with respect to ρ directly. The price for that is that APLS find a random set for ρ , and omit to discuss the consequences of this randomness (which is not due to the random search, but to the dependence of $\hat{\rho}$ on $\hat{\sigma}_{q_{x\varepsilon}}^2$ and on $x'x$). Also, the relevant limiting distribution for APLS is not that of b_ρ^* (which they employ), but that of $b_{q_{x\varepsilon}}^* = b - q_{x\varepsilon}/(x'x/n)$. However, comparing with (2.6), we find

$$\begin{aligned} n^{1/2}(b_{q_{x\varepsilon}}^* - \beta) &= n^{1/2}[b - q_{x\varepsilon}/(x'x/n) - \beta] \\ &= n^{1/2}(x'\varepsilon/n - q_{x\varepsilon})/(x'x/n) \\ &= f/c - dg/c^2 + O_p(n^{-1/2}), \end{aligned}$$

thus it conforms to the limiting distribution of $b - q_{x\varepsilon}/q_{xx}$ given in (2.8) or (unconditionally) in (2.9). Since these have smaller asymptotic variance than b_ρ^* more powerful APLS inference would be obtained by replacing $se(b_{q_{x\varepsilon}}^*)$ given in (2.13) by

$$se(b_{q_{x\varepsilon}}^*) = \hat{\sigma}_{q_{x\varepsilon}} \{1 - q_{x\varepsilon}/[\hat{\sigma}_{q_{x\varepsilon}}^2(x'x/n)]^{1/2}\}^{1/2}/(x'x)^{1/2}$$

when invoking (2.9).

3. Empirical illustration

The methods developed in APLS have been applied to a particular empirical growth model presented in Mankiw et al. (1992) where possible endogeneity of regressors has not been taken into account. This model for 98 countries can be represented as $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta x_{i4} + \varepsilon_i$, where y_i is per capita output, x_{i2} is the rate of human capital, x_{i3} is investment in physical capital and x_{i4} is the logarithm of the sum of the population growth rate, the growth rate in technology and the depreciation rate. The

focus in the APLS analysis is on testing $\beta_2 = 0$ and $\beta_2 + \beta_3 + \beta_4 = 1$ (constant returns to scale). APLS find OLS results as given in their equation (8), which are slightly different from those in Mankiw et al. (1992, p.420, Table II, first column). APLS perform a sensitivity analysis for four different scenarios, which allow either x_{i2} , or x_{i3} or x_{i4} to be endogenous, or both x_{i2} and x_{i3} could be endogenous.

The KLS results for regression models with just one explanatory variable that may be endogenous can also serve the situation where the model has some further exogenous regressors, which have all been partialled out. This means that only for the first scenario of APLS (x_{i2} endogenous) we can produce KLS results relevant for testing β_2 . For these access to the observations on the regressand and regressors is not required. The OLS coefficient estimates and standard errors presented in APLS equation (8) suffice. Just for illustrative purposes we also present in Table 1 results relevant on testing β_3 allowing x_{i3} to be endogenous and on β_4 allowing x_{i4} to be endogenous.

Table 1: Asymptotic 95% KLS confidence intervals

ρ	β_2		β_3		β_4	
-0.3	0.72	1.01	0.83	1.36	-1.32	0.37
-0.2	0.65	0.93	0.69	1.21	-1.74	-0.10
-0.1	0.58	0.86	0.56	1.08	-2.15	-0.53
0.0	0.51	0.79	0.44	0.94	-2.55	-0.95
0.1	0.44	0.72	0.30	0.82	-2.97	-1.35
0.2	0.37	0.65	0.17	0.69	-3.40	-1.76
0.3	0.29	0.58	0.02	0.55	-3.87	-2.18
0.4	0.20	0.50	-0.15	0.41	-4.40	-2.64
0.5	0.09	0.41	-0.35	0.24	-5.02	-3.17
0.6	-0.04	0.30	-0.59	0.04	-5.80	-3.79

From the confidence intervals for particular assumed values of ρ we see that for $\rho = 0.6$ coefficient β_2 is no longer significantly different from zero. However, the statement that β_2 is significantly positive at level 2.5% is robust with respect to simultaneity provided $\rho \leq 0.5$ (further calculations revealed that it is actually for $\rho \leq 0.57$). The parallel finding in APLS is: $r_{\min} = 0.425$ at 5%. One can also conclude from the β_2 results that the interval $0.09 \leq \beta_2 \leq 0.86$ has a confidence coefficient of at least 95% (asymptotically) assuming $-0.1 \leq \rho \leq 0.5$, and so on. Also note that under possible endogeneity of x_{i3} its coefficient is significantly positive provided $\rho \leq 0.3$, whereas under endogeneity of x_{i4} its coefficient is significantly negative provided $\rho \geq -0.2$.

4. How to tackle more general cases?

From the above it should be clear that the KLS approach has yet been developed for only very few special simple cases, whereas the APLS approach has flaws in its asymptotic

underpinnings. Though, even when employing sound asymptotics, the latter will lead to findings which are hard to interpret because the resulting r_{\min} vector is random and obtaining an asymptotic approximation to its distribution to assess its accuracy seems far from easy. On the other hand, further development of the KLS approach seems possible, but requires an asymptotic analysis which certainly cannot be characterized as in APLS (just above their section 2.2) as "easy" and "straightforward", because of the following three reasons.

First, although the unfeasible estimator (6) in APLS is clearly consistent¹, it is not self-evident that it is asymptotically normal unless one has verified whether the conditions for invoking a central limit theorem are satisfied. Doing so reveals that extra conditions are required and that the resulting asymptotic variance involves extra terms. Second, when a consistent estimator is biased, its bias usually being $O(n^{-1})$, it can be corrected by subtracting an $O_p(n^{-1})$ assessment of this bias, while this corrected estimator retains the same limiting distribution as the uncorrected estimator, because the correction just affects higher-order asymptotic aspects. However, an inconsistent estimator has a bias which is generally $O(1)$, and hence any useful random assessment of it will be $O_p(1)$ as well, and thus as a rule employing such a correction will affect the limiting distribution. Third, a further complicating issue is that whereas the limiting distribution of standard consistent estimators is similar whether or not one conditions on exogenous variables, this situation apparently changes when consistency is achieved by correcting an inconsistent estimator.

Many aspects of these complications for regression models with an arbitrary number of endogenous and exogenous regressors have already been addressed in Kiviet and Niemczyk (2012) with respect to OLS estimation and in Kiviet and Niemczyk (2014) with respect to IV estimation when invalid instruments may have been used. A next step should be to obtain for these more general settings the limiting distributions of inconsistency corrected estimators which are feasible in the sense as used in KLS, and next exploit these to produce inference which is robust over a credible set of values for relevant endogeneity correlations. Only after this has been achieved it seems justifiable to provide an answer to the question when explanatory variable endogeneity may be ignored in a regression model.

References

Ashley, R.A., Parmeter, C.F., 2015a. When is it justifiable to ignore explanatory variable endogeneity in a regression model? *Economics Letters* 137, 70-74.

¹Note that when $\Sigma_{X\varepsilon}$ is introduced in (6) of APLS there is a confusing typo: $E(X'_i\varepsilon_i)$ should read $E(X'\varepsilon)$.

Ashley, R.A., Parmeter, C.F., 2015b. Sensitivity analysis for inference in 2SLS estimation with possibly-flawed instruments. *Empirical Economics* forthcoming.

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., 2015. Discriminating between (in)valid external instruments and (in)valid exclusion restrictions. UvA-Econometrics working paper 2015/04.

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., Niemczyk, J., 2014. On the limiting and empirical distribution of IV estimators when some of the instruments are actually endogenous, pp.425-490 in *Advances in Econometrics, Volume 33; Essays in Honor of Peter C.B. Phillips* (Eds. Yoosoon Chang, Thomas B. Fomby, Joon Y. Park), Emerald Group Publishing Limited, Bingley, UK.

Mankiw, N.G., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. *Quarterly Journal of Economics* 107, 407-437.