

Discussion Paper: 2010/01

# Comparing the asymptotic and empirical (un)conditional distributions of OLS and IV in a linear static simultaneous equation

Jan F. Kiviet and Jerzy Niemczyk

[www.feb.uva.nl/ke/UvA-Econometrics](http://www.feb.uva.nl/ke/UvA-Econometrics)

## **Amsterdam School of Economics**

Department of Quantitative Economics  
Roetersstraat 11  
1018 WB AMSTERDAM  
The Netherlands

UvA  UNIVERSITEIT VAN AMSTERDAM



# Comparing the asymptotic and empirical (un)conditional distributions of OLS and IV in a linear static simultaneous equation

Jan F. Kiviet\* and Jerzy Niemczyk†

30 June 2010

JEL-classification: *C13, C15, C30*

Keywords:

*conditioning, efficiency comparisons, inconsistent estimation, Monte Carlo design, simultaneity bias, weak instruments*

## Abstract

In designing Monte Carlo simulation studies for analyzing finite sample properties of econometric inference methods, one can use either IID drawings in each replication for any series of exogenous explanatory variables or condition on just one realization of these. The results will usually differ, as do their interpretations. Conditional and unconditional limiting distributions are often equivalent, thus yielding similar asymptotic approximations. However, when an estimator is inconsistent, its limiting distribution may change under conditioning. These phenomena are analyzed and numerically illustrated for OLS (ordinary least-squares) and IV (instrumental variables) estimators in single static linear simultaneous equations. The results obtained supplement – and occasionally correct – earlier results. The findings demonstrate in particular that the asymptotic approximations to the unconditional and a conditional distribution of OLS are very accurate even in small samples, and that the actual absolute estimation errors of inconsistent OLS in finite samples are often much smaller than those of consistent IV, even when the instruments are not extremely weak. It is also shown that conditioning reduces the estimation errors of OLS, whereas it deranges the distribution of IV when instruments are weak. Finally it is indicated how OLS could be modified to produce accurate inference under assumptions regarding the degree of simultaneity.

## 1 Introduction

Classic Monte Carlo simulation is widely used to assess finite sample distributional properties of parameter estimators and associated test procedures when employed to

---

\*corresponding author: Tinbergen Institute and Department of Quantitative Economics, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; phone +31.20.5254217; email j.f.kiviet@uva.nl

†European Central Bank, Frankfurt, Germany; email: Jerzy.Niemczyk@ecb.int

particular classes of models. This involves executing experiments in which data are being generated from a fully specified DGP (data generating process) over a grid of relevant values in its parameter space. The endogenous variable(s) of such a DGP usually depend on some exogenous explanatory variables, and in a time-series context they may also depend on particular initial conditions. These initial observations and exogenous variables are either generated by particular typical synthetic and possibly stochastic processes or they are taken from empirically observed samples. In the latter case, and if in the former case all replications use just one single realization of such exogenous and pre-sample processes, then the simulation yields the conditional distribution of the analyzed inference methods with respect to those particular realizations. The unconditional distribution is obtained when each replication is based on new independent random draws of these variables. In principle, both simulation designs may yield very useful information, which, however, addresses aspects of different underlying populations. For practitioners, it may often be the more specific conditional distribution that will be of primary interest, provided that the conditioning involves – or mimics very well – the actually observed relevant empirical exogenous and pre-sample variables. Note that a much further fine-tuning of the simulation design (such that it may come very close to an empirically observed DGP, possibly by using for the parameter values in the simulated data their actual empirical estimates) may convert a classic Monte Carlo simulation study on general properties in finite samples of particular inference methods into the generation of alternative inference on a particular data set obtained by resampling, popularly known as bootstrapping.

The large sample asymptotic null distribution of test statistics in well-specified models is usually invariant with respect to the exogenous regressors and their coefficient values. In finite samples this is also the case in the classic normal linear regression model and in a few other cases. In many situations, however, the finite sample null distribution is nonpivotal, and in a Monte Carlo study of possible size distortions and of test power, it will generally matter which type of process is chosen for the exogenous variables, and also whether one conditions on just one realization or does not. For consistent estimators under usual regularity conditions, their conditional and unconditional limiting distributions are equivalent, and when translating these into an asymptotic approximation to the finite sample distribution, it does not seem to matter whether one aims at a conditional or an unconditional interpretation. For inconsistent estimators, however, their limiting distributions may be substantially different depending on whether one conditions or not, which naturally induces a difference between conditional and unconditional asymptotic approximations. In this paper, these phenomena are analyzed analytically and they are also implemented in simulation experiments, when applying either OLS (ordinary least-squares) or IV (instrumental variables) estimators in single static linear simultaneous equations. The results obtained extend – and some of them correct – earlier results published in Kiviet and Niemczyk (2007). The major corrections and their direct consequences have been implemented in the (online available) discussion paper Kiviet and Niemczyk (2009a), which conforms to Chapter 2 of Niemczyk (2009). These closely follow the earlier published text of Kiviet and Niemczyk (2007), and hence provide a refurbished version, in which: (a) the main formula (asymptotic variance of inconsistent OLS) is still the same, but its derivation has been corrected; (b) it is shown now to establish a conditional asymptotic variance for static simultaneous models; (c) also an unconditional asymptotic variance of OLS has been obtained; (d) illustrations are provided which enable to compare (both conditional and unconditional) the asymptotic approximations to and the actual empirical distributions of OLS and IV estimators

in finite samples.

In the present study these new results are more systematically presented and at the same time put into a broader context. Conditioning and its implications for both asymptotic analysis and simulation studies are examined, and especially the consequences of conditioning on latent variables are more thoroughly analyzed and illustrated. Our findings demonstrate that inference based on inconsistent OLS, especially when conditioned on all the exogenous components of the relevant partial reduced form system, may often be more attractive than that obtained by consistent IV when the instruments are very or just moderately weak. However, such inference is unfeasible, because some of its components become available only if one makes an assumption on the degree of simultaneity in the model. If one is willing to do so, possibly for a range of likely situations, this may provide useful additional inference (conditional on the assumptions made).

Recent studies on general features which are relevant when designing Monte Carlo studies, such as Doornik (2006) and Kiviet (2007), do not address the issue of whether one should or should not condition on just one realization of the exogenous variables included in the examined DGP. An exception is Edgerton (1996), who argues against conditioning. However, his only argument is that a conditional simulation study, although unbiased, provides an inefficient assessment of the unconditional distribution. This is obviously true, but it is not relevant when recognizing that the conditional distribution may be of interest in its own right. Actually, it is sometimes more and sometimes less attractive than the unconditional distribution for obtaining inference on the parameters of interest, as we will see. Below, we will reconsider these issues. Our illustrations show that both approaches deserve consideration and comparison, especially in cases where they are not just different in finite samples, but differ asymptotically as well. We also show that conditioning on purely arbitrary draws of the exogenous variables leads to results that are hard to interpret, but that this is avoided by stylizing these draws in such a way that comparison with unconditional results does make sense.

As already mentioned, conditioning has consequences asymptotically too when we consider inconsistent estimators. We shall focus on applying OLS to one simultaneous equation from a larger system. Goldberger (1964) did already put forward the unconditional limiting distribution for the special case where all structural and reduced form regressors are IID (independently and identically distributed). We shall critically review the conditions under which this result holds. Phillips and Wickens (1978, Question 6.10c) consider the model with just one explanatory variable which is endogenous and has a reduced form with also just one explanatory variable. Because this exogenous regressor is assumed to be fixed, the variables are not IID here. In their solution to the question, they list the various technical complexities that have to be surpassed in order to find the limiting distribution of the inconsistent OLS coefficient estimator, but they do not provide an explicit answer. Hausman (1978) considers the same type of model and, exploiting an unpublished result for errors in variables models by Rothenberg (1972), presents its limiting distribution, self-evidently conditioning on the fixed reduced form regressors. Kiviet and Niemczyk (2007) aimed at generalizing this result for the model with an arbitrary number of endogenous and exogenous stationary regressors, without explicitly specifying the reduced form. Below, we will demonstrate that the limiting distribution they obtained is correct for the case of conditioning on all exogenous information, but that the proof that they provided has some flaws. These will be repaired here, and at the same time we will further examine the practical consequences of the conditioning. In the illustrations in Kiviet and Niemczyk (2007) the obtained asymp-

otic approximation was compared inappropriately with simulation results in which we did not condition on just one single draw of the exogenous regressors. (We thank Peter Boswijk for bringing this to our attention.) Here, we will provide illustrations which allow to appreciate the effects of conditioning both for the limiting distribution of OLS, and for its distribution in finite samples. Moreover, we make comparisons between the accuracy of inconsistent OLS and consistent IV estimation, both conditional and unconditional. Results for inconsistent IV estimators can be found in Kiviet and Niemczyk (2009b).

Our major findings are that inconsistent OLS often outperforms consistent IV when the sample size is finite, irrespective of whether one conditions or not. For a simple specific class of models we find that in samples with a size between 20 and 500 the actual estimation errors of IV are noticeably smaller than those of OLS only when the degree of simultaneity is substantial and the instruments are far from weak. However, when instruments are weak OLS always wins, even for a substantial degree of simultaneity. We also find that the first-order asymptotic approximations (both conditional and unconditional) to the estimation errors of OLS are very accurate even in relatively small samples. This is not the case for IV when instruments are weak, see also Bound et al. (1995). For consistent IV one needs alternative asymptotic sequences when instruments are weak, see for an overview Andrews and Stock (2007). However, we also find that the problems with IV when instruments are weak are much less serious for the unconditional distribution than for the conditional one, which is inflicted by serious skewness and bimodality, see Woglom (2001). Especially when simultaneity is serious, the conditional distribution of OLS is found to be more efficient than its unconditional counterpart.

The structure of this paper is as follows. Section 2 introduces the single structural equation model from an only partially specified linear static simultaneous system. Next, in separate subsections, two alternative frameworks are defined for obtaining either unconditional or conditional asymptotic approximations to the distribution of estimators, and for generating their finite sample properties from accordingly designed simulation experiments. In Section 3 the unconditional and conditional limiting distributions of IV and OLS coefficient estimators are derived. These are shown to be similar for consistent IV and diverging for inconsistent OLS. Section 4 discusses particulars of the simulation design of the various simple cases that we considered, and addresses in detail how we implemented conditioning in the simulations. Next, graphical results are presented which easily allow to make general and more specific comparisons between IV and OLS estimation and analyze the effects of the particular form of conditioning that we adopted. Finally, we indicate how our findings might be used in practice. Section 5 concludes.

## 2 Model and two alternative frameworks

To examine the consequences for estimators under either a particular unconditional regime or under conditioning on some relevant information set, we will define in separate subsections two alternative frameworks, viz. Framework U and Framework C. For both we will examine in Section 3 how IV and OLS estimators converge under a matching asymptotic sequence. In Section 4 both will also establish the blueprint for two alternative data generating schemes to examine in finite samples by Monte Carlo experiments unconditional and conditional inference respectively. These two frameworks are polar in nature, but intermediary implementations could be considered too. First, we will state what both implementations do have in common.

Both focus on a single standard static linear simultaneous equation

$$y_t = x_t' \beta + \varepsilon_t, \quad (1)$$

for observations  $t = 1, \dots, n$ , where  $x_t$  and  $\beta$  are  $k \times 1$  vectors. Both these vectors can be partitioned correspondingly in  $k_1$  and  $k_2 = k - k_1 \geq 0$  elements respectively, giving  $x_t' \beta = x_{1t}' \beta_1 + x_{2t}' \beta_2$ . Regarding the disturbances we assume that  $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$ , but also that  $E(\varepsilon_t | x_{2t}) \neq 0$ , hence  $x_{2t}$ , if not void, will contain some endogenous explanatory variables. In addition, we have  $l \geq k$  variables collected in an  $l \times 1$  vector  $z_t$ , which can be partitioned in  $k_1$  and  $l - k_1 \geq 0$  elements respectively, i.e.  $z_t' = (z_{1t}' \ z_{2t}')$ , whereas  $z_{1t} = x_{1t}$ . Below, we will distinguish between nonrandom and random  $z_t$ . In the latter case we assume that  $\varepsilon_t | z_1, \dots, z_n \sim IID(0, \sigma_\varepsilon^2)$ . Hence, in both cases the variables  $z_t$  are exogenous and establish valid instruments. If  $k_1 > 0$  then equation (1) contains at least  $k_1$  exogenous regressors  $x_{1t}$ .

All  $n$  observations on the variables and the  $n$  realizations of the random disturbances can be collected, as usual, in vectors  $y$  and  $\varepsilon$  and matrices  $X = (X_1, X_2)$  and  $Z = (Z_1, Z_2)$ , where  $Z_1 = X_1$ . Both  $X$  and  $Z$  have full column rank, and so has  $Z'X$ , thus the necessary and sufficient condition for identification of the coefficients  $\beta$  by the sample are satisfied; i.e. a unique generalized IV estimator exists. Note that we did not specify the structural equations for the variables in  $X_2$ , nor their reduced form equations, so whether the necessary and sufficient rank condition for asymptotic identification holds is not clear at this stage.

## 2.1 Framework U

Under this framework for unconditional analysis we assume that all variables are random, and that after centering they are weakly stationary. So,  $x_t - E(x_t)$  and  $z_t - E(z_t)$  have constant and bounded second moments. Using  $E(y_t) = E(x_t')\beta$  and subtracting it from (1) leads to a model without intercept (if there was one) where all variables have zero mean. Since our primary interest lies in inference on slope parameters we may therefore assume, without loss of generality, that  $y_t$ ,  $x_t$  and  $z_t$  (after the above transformation of the model) all have zero mean. For the second moments we define (all plim's are here for  $n \rightarrow \infty$ )

$$\Sigma_{X'X} \equiv \text{plim } \frac{1}{n} X'X = E(x_t x_t'), \quad \Sigma_{Z'Z} \equiv \text{plim } \frac{1}{n} Z'Z = E(z_t z_t'), \quad (2)$$

$$\Sigma_{Z'X} \equiv \text{plim } \frac{1}{n} Z'X = E(z_t x_t'), \quad \forall t.$$

We also assume that  $\Sigma_{X'X}$ ,  $\Sigma_{Z'Z}$  and  $\Sigma_{Z'X}$  all have full column rank, which guarantees the asymptotic identification of  $\beta$  by these instruments. Note that, although we assume that  $(z_t' \ x_{2t}')$  has  $\forall t$  identical second moments, (2) does not imply that  $(z_t' \ x_{2t}')$  and  $(z_s' \ x_{2s}')$  are independent for  $t \neq s$ , but any dependence should disappear for  $|t - s|$  large.

Using  $\Sigma_{Z'X} = (\Sigma_{Z'X_1}, \Sigma_{Z'X_2})$  and defining

$$\Pi \equiv \Sigma_{Z'Z}^{-1} \Sigma_{Z'X} = \Sigma_{Z'Z}^{-1} (\Sigma_{Z'X_1}, \Sigma_{Z'X_2}) = ((I_{k_1}, O_{k_1, k_2})', \Pi_2), \quad (3)$$

where  $I_p$  is a  $p \times p$  identity matrix and  $O_{p,q}$  a  $p \times q$  zero matrix, we can easily characterize implied linear reduced form equations for  $x_{2t}$  as follows. Decomposing  $x_{2t}$  into two components, where one is linear in  $z_t$ , we obtain

$$x_{2t}' = z_t' \Pi_2 + v_{2t}', \quad (4)$$

where  $E(v'_{2t}) = 0'$  and  $E(z_t v'_{2t}) = E[z_t(x'_{2t} - z'_t \Pi_2)] = \Sigma_{Z'X_2} - \Sigma_{Z'Z} \Pi_2 = O_{l, k_2}$ . Equations (4) correspond with the genuine reduced form equations only if  $z_t$  contains all exogenous variables from the complete simultaneous system, which we leave unspecified.

The endogeneity of  $x_{2t}$  implies nonzero covariance between  $v_{2t}$  and  $\varepsilon_t$ . We may denote (i.e. parametrize) this covariance as

$$E(\varepsilon_t x'_{2t}) = E(\varepsilon_t v'_{2t}) \equiv \sigma_\varepsilon^2 \xi'_2. \quad (5)$$

This enables to decompose  $v_{2t}$  as

$$v'_{2t} = \bar{v}'_{2t} + \varepsilon_t \xi'_2, \quad (6)$$

where  $E(\bar{v}'_{2t}) = 0'$  and  $E(\varepsilon_t \bar{v}'_{2t}) = 0'$ . Now another decomposition for  $x'_{2t}$  is

$$x'_{2t} = \bar{x}'_{2t} + \varepsilon_t \xi'_2, \quad (7)$$

where

$$\bar{x}'_{2t} = z'_t \Pi_2 + \bar{v}'_{2t}. \quad (8)$$

This establishes a different decomposition of the endogenous regressors as the implied partial reduced form equations (4) do. The latter have an exogenous component that is a linear combination of just the instruments  $z'_t$  and the former have an exogenous component that also contains  $\bar{v}'_{2t}$ , which establishes the implied reduced form disturbances in as far as uncorrelated with  $\varepsilon_t$ . These could be interpreted as the effects on  $x'_{2t}$  of all exogenous variables yet omitted from the implied reduced form (4).

Decomposition (7) implies  $\forall t$

$$x'_t = \bar{x}'_t + \varepsilon_t \xi'_2, \quad (9)$$

where  $\xi \equiv (0', \xi'_2)'$ , with

$$E(x_t \varepsilon_t) = \sigma_\varepsilon^2 \xi. \quad (10)$$

Hence,

$$X = \bar{X} + \varepsilon \xi', \quad (11)$$

with

$$\text{plim} \frac{1}{n} X' \varepsilon = \sigma_\varepsilon^2 \xi, \quad \text{plim} \frac{1}{n} \bar{X}' \bar{X} = \Sigma_{X'X} - \sigma_\varepsilon^2 \xi \xi' \quad \text{and} \quad \text{plim} \frac{1}{n} Z' \bar{X} = \Sigma_{Z'X}. \quad (12)$$

Decomposition (11) will be relevant too when we consider conditioning, as we shall see.

## 2.2 Framework C

In this framework the variables  $z_t$ , and hence  $x_{1t}$ , are all (treated as) fixed for  $t = 1, \dots, n$ . Like in Framework U, the structural equation (now in matrix notation) is

$$y = X_1 \beta_1 + X_2 \beta_2 + \varepsilon, \quad (13)$$

and  $\varepsilon_t$  is correlated with  $x_{2t}$ . This correlation can again be parametrized, like in (5), so that

$$E(X'_2 \varepsilon) \equiv n \sigma_\varepsilon^2 \xi_2. \quad (14)$$

Indicating the "genuine" partial reduced form disturbances for  $X_2$  as  $V_2^* \equiv X_2 - E(X_2)$ , and decomposing  $V_2^* = \bar{V}_2^* + \varepsilon\xi_2'$  with  $E(\bar{V}_2^*\varepsilon) = 0$ , we find a decomposition of  $X$  which can be expressed (again) as

$$X = \bar{X} + \varepsilon\xi', \quad \text{with } \bar{X} = (\bar{X}_1, \bar{X}_2) \text{ and } \xi' = (0', \xi_2'), \quad (15)$$

where now

$$X_2 = \bar{X}_2 + \varepsilon\xi_2' = E(X_2) + \bar{V}_2^* + \varepsilon\xi_2'. \quad (16)$$

Here  $E(X_2)$  contains the deterministic part of the genuine partial reduced form (a linear combination of all exogenous regressors from the unspecified system), whereas component  $\bar{V}_2^*$  is random with zero mean; its  $t^{\text{th}}$  row  $\bar{v}_{2t}^*$  consists of components of the disturbances from the genuine but unspecified reduced form, in as far as uncorrelated with  $\varepsilon_t$ .

We can use this framework to analyze the consequences of conditioning on the obtained realizations of  $z_t = (x'_{1t}, z'_{2t})'$ . However, in the practical situation in which an investigator realizes that the variables  $z_t$  will most probably contain only a subset of the regressors of the reduced form, one might also contemplate conditioning on an extended information set, not only containing  $z_t$ , but also  $E(x'_{2t})$ , and even  $\bar{v}_{2t}^*$ , although both  $E(x'_{2t})$  and  $\bar{v}_{2t}^*$  are unobserved. That they are in practice unobserved is no limitation in a Monte Carlo simulation study, where these components of the DGP – like the in practice unobserved parameter values – will always be available. Also in practice, though, one may have the ambition to condition inference on all the specific circumstances (both observed and unobserved) which are exogenous with respect to the disturbances  $\varepsilon_t$ . Below we will examine whether it is worthwhile to use for conditioning the widest possible set, which is provided under Framework C by  $(z'_t, \bar{x}'_t)$ .

For an asymptotic analysis in large samples under Framework C we will resort to the "constant in repeated samples" concept, see Theil (1971, p.364). Thus, we consider samples of size  $mn$  in which  $Z_m$  is an  $mn \times l$  matrix in which the  $n \times l$  matrix  $Z$  has been stacked  $m$  times. Then we obtain (now all plim's are for  $m \rightarrow \infty$ )

$$\Sigma_{Z'Z} \equiv \text{plim} \frac{1}{mn} Z'_m Z_m = \text{plim} \frac{1}{mn} \sum_{j=1}^m Z'_j Z_j = \frac{1}{n} Z' Z, \quad (17)$$

implying  $\Sigma_{X'_1 X_1} = \frac{1}{n} X'_1 X_1$  and  $\Sigma_{Z' X_1} = \frac{1}{n} Z' X_1$ , which are all finite, self-evidently. However, one does not keep  $\varepsilon$  constant in these (imaginary) enlarged samples. All the components of the  $mn \times 1$  vector  $\varepsilon_m$  are IID( $0, \sigma_\varepsilon^2$ ), and because  $E(Z'_m \varepsilon_m) = Z'_m E(\varepsilon_m) = 0$ , also  $E(Z'_m \varepsilon_m) \xi_2' = O_{l, k_2}$ . Thus,  $\Sigma_{Z' X_2} = \text{plim} \frac{1}{mn} Z'_m \bar{X}_{2m} = \frac{1}{n} Z' \bar{X}_2$  and  $\Sigma_{X'_1 X_2} = \frac{1}{n} X'_1 \bar{X}_2$ , whereas  $\Sigma_{X'_2 X_2} = \frac{1}{n} \bar{X}'_2 \bar{X}_2 + \sigma_\varepsilon^2 \xi_2 \xi_2'$ , thus

$$\Sigma_{X'X} = \frac{1}{n} \bar{X}' \bar{X} + \sigma_\varepsilon^2 \xi \xi', \quad \Sigma_{Z'X} = \frac{1}{n} (Z' X_1, Z' \bar{X}_2). \quad (18)$$

Note that the above implementation of the "constant in repeated samples" concept excludes the possibility that some of the instruments (or variables in  $x_{1t}$ ) are actually weakly-exogenous, because that would require to incorporate lags of  $\varepsilon_t$  in  $z_t$ .

In both frameworks U and C, the asymptotic sequence leads to finite second data moments, but these are being assembled in different ways. In both frameworks  $X$  can be decomposed as in (11). But, under U the matrix  $\bar{X}$  is random and

$$\Sigma_{X'X} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \bar{X}' \bar{X} + \sigma_\varepsilon^2 \xi \xi' = E \bar{x}_t \bar{x}'_t + \sigma_\varepsilon^2 \xi \xi', \quad \forall t, \quad (19)$$

whereas under C the matrix  $\bar{X}$  is nonrandom and  $\Sigma_{X'X}$  is given by (18). In the next sections we will examine the respective consequences for estimation.



### 3 Limiting distributions for IV and OLS

We shall now derive the limiting distributions of the IV and OLS estimators of  $\beta$  under both frameworks, and from these investigate the analytical consequences regarding the first-order asymptotic approximations in finite samples to the unconditional and conditional distributions.

#### 3.1 IV estimation

The model introduced above is in practice typically estimated by the methods of moments technique, in which a surplus of  $l - k$  moment conditions is optimally exploited by the (generalized) IV estimator

$$\hat{\beta}_{GIV} = (X'P_ZX)^{-1}X'P_Zy, \quad (20)$$

where  $P_Z \equiv Z(Z'Z)^{-1}Z'$ . When  $l = k$ , thus  $Z'X$  is a square and invertible matrix, this simplifies to  $\hat{\beta}_{IV} = (Z'X)^{-1}Z'y$ . For  $l \geq k$ , and under the regularity conditions adopted in either Framework U or Framework C, it can be shown in the usual way that GIV is consistent and asymptotically normal with limiting distribution

$$n^{1/2}(\hat{\beta}_{GIV} - \beta) \xrightarrow{d} N\left(0, \mathbf{AVar}(\hat{\beta}_{GIV})\right), \quad (21)$$

where

$$\mathbf{AVar}(\hat{\beta}_{GIV}) = \sigma_\varepsilon^2(\Sigma'_{Z'X}\Sigma_{Z'Z}^{-1}\Sigma_{Z'X})^{-1}. \quad (22)$$

The estimator for  $\sigma_\varepsilon^2$  is based on the GIV residuals  $\hat{\varepsilon}_{GIV} = y - X\hat{\beta}_{GIV}$ . It is not obvious in what way its finite sample properties could be improved by employing a degrees of freedom correction and therefore one usually employs simply the consistent estimator

$$\hat{\sigma}_{\varepsilon,GIV}^2 = \frac{1}{n}\hat{\varepsilon}'_{GIV}\hat{\varepsilon}_{GIV}. \quad (23)$$

Hence, in practice, one uses under both frameworks

$$\widehat{Var}(\hat{\beta}_{GIV}) = \hat{\sigma}_{\varepsilon,GIV}^2(X'P_ZX)^{-1} \quad (24)$$

as an estimator of  $Var(\hat{\beta}_{GIV})$ , because  $n\widehat{Var}(\hat{\beta}_{GIV})$  is a consistent estimator of (22). This easily follows from the consistency of  $\hat{\sigma}_{\varepsilon,GIV}^2$ , and because under both frameworks  $n(X'P_ZX)^{-1} = [\frac{1}{n}X'Z(\frac{1}{n}Z'Z)^{-1}\frac{1}{n}Z'X]^{-1}$  has probability limit  $(\Sigma'_{Z'X}\Sigma_{Z'Z}^{-1}\Sigma_{Z'X})^{-1}$ .

Hence, irrespective of whether one adopts Framework U or C, there are no material differences between the consequences of standard first-order asymptotic analysis for consistent IV estimation. In both cases the consistent estimators  $\hat{\beta}_{GIV}$ ,  $\hat{\sigma}_{\varepsilon,GIV}^2$  and  $\widehat{Var}(\hat{\beta}_{GIV})$  are all obtained from the very same expressions in actually observed sample data moments. How well they serve to approximate the characteristics of the actual unconditional and conditional distributions in finite samples will be examined by simulations under the two respective frameworks. A point of special concern here is that in finite samples  $\hat{\beta}_{GIV}$  has no finite moments from order  $l - k + 1$  onwards, which is not reflected by the Gaussian approximation. As a consequence,  $\widehat{Var}(\hat{\beta}_{GIV})$  approximates a non-existing quantity when  $l = k$  or  $l = k + 1$ . Therefore, in the illustrations in Section 4, we will only present density functions and particular quantiles.

## 3.2 OLS estimation

When one neglects the simultaneity in model (1) and employs the OLS estimator

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y, \quad (25)$$

then under both frameworks its probability limit is

$$\beta_{OLS}^* \equiv \text{plim } \hat{\beta}_{OLS} = \beta + \Sigma_{X'X}^{-1} \text{plim } n^{-1}X'\varepsilon = \beta + \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} \xi. \quad (26)$$

This is the so-called pseudo true value of  $\hat{\beta}_{OLS}$ . We may also define

$$\ddot{\beta}_{OLS} \equiv \beta_{OLS}^* - \beta = \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} \xi, \quad (27)$$

which is the inconsistency of the OLS estimator.

Under both frameworks we will next derive the limiting distribution  $n^{1/2}(\hat{\beta}_{OLS} - \beta_{OLS}^*) \xrightarrow{d} N(0, V)$ . Note that this is not centered at  $\beta$ , but at  $\beta_{OLS}^*$ , and that  $V$  will be different under the two frameworks. For the variance matrix  $V$  of this zero mean limiting distribution we will find a different expression under Framework U than under C. In Section 4 we shall use  $\beta_{OLS}^* - \beta = \ddot{\beta}_{OLS} = \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} \xi$  as a first-order asymptotic approximation to the bias of  $\hat{\beta}_{OLS}$  in finite samples, and  $V/n$  for the variance of  $\hat{\beta}_{OLS}$ . Or, rather than  $\ddot{\beta}_{OLS}$  and  $V/n$ , similar expressions in which the matrix of population data moments  $\Sigma_{X'X}$  has been replaced by the corresponding sample data moments. Like  $\ddot{\beta}_{OLS}$ , both expressions for  $V/n$  will also appear to depend on the parameters  $\sigma_\varepsilon^2$  and  $\xi$ . That is not problematic when we evaluate these first-order asymptotic approximations in the designs that we use for our simulation study in order to examine the qualities of the asymptotic approximations, but of course it precludes that these can be used simply and directly for inference in practice.

### 3.2.1 Unconditional limiting distribution of OLS

For obtaining a characterization of the unconditional limiting distribution of inconsistent OLS, like Goldberger (1964, p.359), we rewrite the model as

$$y = X(\beta_{OLS}^* - \ddot{\beta}_{OLS}) + \varepsilon = X\beta_{OLS}^* + u, \quad (28)$$

where  $u \equiv \varepsilon - X\ddot{\beta}_{OLS}$ . Under Framework U we have (after employing the transformation that removed the intercept)  $E(X) = O_{n,k}$ , and hence  $E(u) = 0$ . From  $Var(x_t) = \Sigma_{X'X}$  and (10) we find for  $u_t \equiv \varepsilon_t - x_t'\ddot{\beta}_{OLS}$  that

$$\begin{aligned} \sigma_u^2 &\equiv E(u_t^2) = \sigma_\varepsilon^2(1 - 2\ddot{\beta}_{OLS}'\xi) + \ddot{\beta}_{OLS}'\Sigma_{X'X}\ddot{\beta}_{OLS} \\ &= \sigma_\varepsilon^2(1 - \xi'\ddot{\beta}_{OLS}). \end{aligned} \quad (29)$$

Moreover,  $E(x_t u_t) = E(x_t \varepsilon_t) - E(x_t x_t')\ddot{\beta}_{OLS} = \sigma_\varepsilon^2 \xi - \Sigma_{X'X}\ddot{\beta}_{OLS} = 0$ . Thus, in the alternative model specification (28) OLS will yield a consistent estimator for  $\beta_{OLS}^*$ .

To obtain its limiting distribution, one has to evaluate  $Var(x_t u_t) = E(u_t^2 x_t x_t')$  and  $E(u_t u_s x_t x_s')$  for  $t \neq s$ . These depend on characteristics of the joint distribution of  $\varepsilon_t$  and  $x_t$  that have not yet been specified in Framework U. Here we will just examine the consequences of a further specialization of Framework U by assuming that  $\forall t$

$$\varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \text{ and } x_t \sim NID(0, \Sigma_{X'X}). \quad (30)$$

Note that by assuming independence of  $x_t$  and  $x_s$  for  $t \neq s$  typically most time-series applications are excluded.

From (30) we obtain  $u_t \sim NID(0, \sigma_u^2)$ , so that  $E(x_t u_t) = 0$  now implies independence of  $x_t$  and  $u_t$ . Then we find  $E(u_t^2 x_t x_t') = \sigma_u^2 \Sigma_{X'X}$  and also  $E(u_t u_s x_t x_s') = O_{k,k}$  for  $t \neq s$ , so that a standard central limit theorem can be invoked, yielding the limiting distribution

$$n^{1/2}(\hat{\beta}_{OLS} - \beta_{OLS}^*) \xrightarrow{d} N\left(0, \mathbf{AVar}_U^{NID}(\hat{\beta}_{OLS})\right), \quad (31)$$

with

$$\mathbf{AVar}_U^{NID}(\hat{\beta}_{OLS}) \equiv (1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi) \sigma_\varepsilon^2 \Sigma_{X'X}^{-1}, \quad (32)$$

where self-evidently the indices  $U$  and  $NID$  refer to the adopted framework, specialized with (30). Goldberger (1964, p.359) presents a similar result without adopting normality of  $\varepsilon_t$  and  $x_t$ , which does not seem right. The same remark is made by Rothenberg (1972, p.16), but he condemns result (31) anyhow, simply because he finds the assumption  $E(x_t) = 0$  unrealistic in general. We claim, however, that this can be justified in Framework U by interpreting this limiting distribution as just referring to the slope coefficients after centering the relationship.

For the OLS residuals  $\hat{u} = y - X\hat{\beta}_{OLS}$  one easily obtains

$$\text{plim} \frac{1}{n} \hat{u}' \hat{u} = \text{plim} \frac{1}{n} (\varepsilon - X\ddot{\beta}_{OLS})' (\varepsilon - X\ddot{\beta}_{OLS}) = \sigma_u^2. \quad (33)$$

Thus, when the data are Gaussian and IID, standard OLS inference in the regression of  $y$  on  $X$ , and estimating  $\text{Var}(\hat{\beta}_{OLS})$  by  $\frac{\hat{u}' \hat{u}}{n} (X'X)^{-1}$ , makes sense and is in fact asymptotically valid, but it concerns unconditional (note that it has really been built on the stochastic properties of  $X$ ) inference on the pseudo true value  $\beta_{OLS}^* = \beta + \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} \xi$ , and not on  $\beta$ , unless  $\xi = 0$ .

### 3.2.2 Conditional limiting distribution of OLS

Next we shall focus on the limiting distribution while conditioning on the exogenous variables  $\bar{X}$  for which Framework C is suitable, because it treats  $\bar{X}$  as fixed. So, we do no longer restrict ourselves to (30), hence nonnormal disturbances and serially correlated regressors are allowed again. However, as will become clear below, we have to extend Framework C with the assumption  $\text{Var}(\varepsilon \mid \bar{X}) = \sigma_\varepsilon^2 I_n$ , and hence exclude particular forms of conditional heteroskedasticity.

The conditional limiting distribution is obtained as follows. Obvious substitutions yield

$$\begin{aligned} n^{1/2}(\hat{\beta}_{OLS} - \beta_{OLS}^*) &= n^{1/2} \left[ \left( \frac{1}{n} X'X \right)^{-1} n^{-1} X' \varepsilon - \ddot{\beta}_{OLS} \right] \\ &= \left( \frac{1}{n} X'X \right)^{-1} [n^{-1/2} X' \varepsilon - n^{-1/2} X' X \ddot{\beta}_{OLS}]. \end{aligned} \quad (34)$$

To examine the terms in square brackets, we substitute the decomposition (15), and for the second term we also use that from (18) and (27) it follows that

$$\bar{X}' \bar{X} \ddot{\beta}_{OLS} = n \Sigma_{X'X} \ddot{\beta}_{OLS} - n \sigma_\varepsilon^2 \xi \xi' \ddot{\beta}_{OLS} = n \sigma_\varepsilon^2 (1 - \xi' \ddot{\beta}_{OLS}) \xi.$$

Then we obtain

$$\begin{aligned}
& n^{-1/2} X' \varepsilon - n^{-1/2} X' X \ddot{\beta}_{OLS} \\
= & n^{-1/2} [(\bar{X}' \varepsilon + \varepsilon' \varepsilon \xi) - (\bar{X}' \bar{X} + \bar{X}' \varepsilon \xi' + \xi \varepsilon' \bar{X} + \varepsilon' \varepsilon \xi \xi') \ddot{\beta}_{OLS}] \\
= & n^{-1/2} [(\bar{X}' \varepsilon + \varepsilon' \varepsilon \xi) - n \sigma_\varepsilon^2 (1 - \xi' \ddot{\beta}_{OLS}) \xi - (\bar{X}' \varepsilon \xi' + \xi \varepsilon' \bar{X}) \ddot{\beta}_{OLS} - (\xi' \ddot{\beta}_{OLS}) \varepsilon' \varepsilon \xi] \\
= & n^{-1/2} \{[(1 - \xi' \ddot{\beta}_{OLS}) I_k - \xi \ddot{\beta}'_{OLS}] \bar{X}' \varepsilon + (1 - \xi' \ddot{\beta}_{OLS}) \xi (\varepsilon' \varepsilon - n \sigma_\varepsilon^2)\} \\
= & n^{-1/2} [A' \varepsilon + a(\varepsilon' \varepsilon - n \sigma_\varepsilon^2)], \tag{35}
\end{aligned}$$

with

$$\begin{aligned}
A' & \equiv [(1 - \xi' \ddot{\beta}_{OLS}) I_k - \xi \ddot{\beta}'_{OLS}] \bar{X}', \\
a & \equiv (1 - \xi' \ddot{\beta}_{OLS}) \xi, \tag{36}
\end{aligned}$$

which, when conditioning on  $\bar{X}$ , are both deterministic.

This conforms to equations (20)-(22) in Kiviet and Niemczyk (2007), which were obtained under the extra assumption that the expression in their equation (19) has to be zero. By putting the derivations now into Framework C, thus fully recognizing that we condition on  $\bar{X}$ , that specific expression is zero by definition. Also note that the equation at the bottom of Kiviet and Niemczyk (2007, p.3300) only holds when  $\bar{X}$  is nonrandom, which was not respected in the simulations presented in that paper.

The conditional limiting distribution now follows exactly as in the derivations in Kiviet and Niemczyk (2007, p.3301) and, when one assumes  $E(\varepsilon_i^3) = 0$  and  $E(\varepsilon_i^4) = 3\sigma_\varepsilon^4$ , that leads to

$$n^{1/2} (\hat{\beta}_{OLS} - \beta_{OLS}^*) \xrightarrow{d} N\left(0, \mathbf{AVar}_C^N(\hat{\beta}_{OLS})\right), \tag{37}$$

with

$$\begin{aligned}
& \mathbf{AVar}_C^N(\hat{\beta}_{OLS}) \tag{38} \\
\equiv & (1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi) [(1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi) \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} - (1 - 2\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi) \sigma_\varepsilon^4 \Sigma_{X'X}^{-1} \xi \xi' \Sigma_{X'X}^{-1}] \\
= & (1 - \xi' \ddot{\beta}_{OLS}) [(1 - \xi' \ddot{\beta}_{OLS}) \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} - (1 - 2\xi' \ddot{\beta}_{OLS}) \ddot{\beta}_{OLS} \ddot{\beta}'_{OLS}],
\end{aligned}$$

where now the superindex N refers to the assumed almost normality of just the disturbances. For the additional terms that would follow when the disturbances have 3rd and 4th moment different from the normal we refer to the earlier paper.

In the illustrations to follow we will compare (38) with the variance of the unconditional limiting distribution given in (32), and both will also be compared with the actual finite sample variance obtained from simulating models under the respective frameworks. These comparisons are made by depicting the respective densities.

Rothenberg (1972) examined the limiting distribution of inconsistent OLS in a linear regression model with measurement errors. It has been used by Schneeweiss and Srivastava (1994) to analyze in such a model the MSE (mean squared error) of OLS up to the order of  $1/n$ . By reinterpreting his results Rothenberg obtains also the asymptotic variance of OLS (his equation 4.7) in a structural equation where for all endogenous regressors the deterministic part of their reduced form equations is given and fixed. Hausman (1978, p.1257) and Hahn and Hausman (2003, p.124) used Rothenberg's result to express the asymptotic variance of OLS (conditioned on all exogenous regressors) in the structural equation model for the case  $k = 1$ . We will return to their result in Section 4, where we also specialize to the case  $k = 1$ . Our result (38) is directly obtained for the general ( $k \geq 1$ ) linear structural equation model, and by the decomposition (15) we also avoided an explicit specification of the reduced form and of the variance matrix of the disturbances in the structural equation and the partial reduced form for  $X$ ,

as is required when employing Rothenberg's result. From formula (38) it can be seen directly that  $AVar_C^N(\hat{\beta}_{OLS})$  is a modification of the asymptotic variance  $\sigma_\varepsilon^2 \Sigma_{X'X}^{-1}$  of the standard consistent OLS case. The only determining factor of this modification is the parameter regarding the simultaneity  $\xi$ , and then more in particular how  $\xi$  (transformed by standard asymptotic variance  $\sigma_\varepsilon^2 \Sigma_{X'X}^{-1}$ ) affects the inconsistency  $\ddot{\beta}_{OLS} = \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} \xi$ , which (in the innerproduct  $\xi' \ddot{\beta}_{OLS} = \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi$ ) is prominent in the modification. Note that the factor  $1 - \xi' \ddot{\beta}_{OLS}$ , also occurring in (32), is equal to  $\text{plim } \varepsilon' M_X \varepsilon / \varepsilon' \varepsilon$ , where  $M_X = I - X(X'X)^{-1}X'$ , and is therefore nonnegative and not exceeding 1, thus simultaneity mitigates the asymptotic (un)conditional variance of OLS.

## 4 Actual qualities of the approximations

The relevance and accuracy of our various results will now be investigated numerically. The actual densities of the various estimators  $\hat{\beta}$  will be assessed by simulating them by generating finite samples from particular DGP's. These densities will be graphically compared with their first order asymptotic approximations of the generic form  $N(\beta^*, n^{-1}AVar(\hat{\beta}))$ . To summarize such findings it is also useful to consider and compare a one-dimensional measure for the magnitude of the estimation errors. For this we do not use the (root) MSE because we will consider models for which  $l = k = 1$ , where IV does not have finite moments. Therefore we will use the median absolute error  $MAE(\hat{\beta})$  instead, which can be estimated from the Monte Carlo results, and compared with the asymptotic MAE, or  $AMA E(\hat{\beta})$ , for the relevant normal limiting distributions.

For a scalar estimator  $\hat{\beta}$  of  $\beta$  the median absolute error  $MAE(\hat{\beta})$  is defined as

$$\Pr\{|\hat{\beta} - \beta| \leq MAE(\hat{\beta})\} = 0.5. \quad (39)$$

From a series of  $R$  independent Monte Carlo realizations  $\hat{\beta}^{(r)}$  ( $r = 1, \dots, R$ ) we estimate  $MAE(\hat{\beta})$  by sorting the values  $|\hat{\beta}^{(r)} - \beta|$  and taking the median value. The natural asymptotic counterpart of the Monte Carlo estimate of  $MAE(\hat{\beta})$  is the (scalar) asymptotic version  $AMA E(\hat{\beta})$ , which we define as follows. Let the CDF of the normal approximation to the distribution of  $\hat{\beta} - \beta$  be indicated by  $\Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}(x)$ , where  $\text{plim } \hat{\beta} = \beta + \ddot{\beta}$ . Then, for  $m \equiv AMA E(\hat{\beta})$ , we have

$$\begin{aligned} 0.5 &= \Pr\{|\hat{\beta} - \beta| \leq m\} = 1 - \Pr\{|\hat{\beta} - \beta| > m\} \\ &= 1 - \Pr\{\hat{\beta} - \beta > m\} - \Pr\{\hat{\beta} - \beta < -m\} \\ &= \Pr\{\hat{\beta} - \beta < m\} - \Pr\{\hat{\beta} - \beta < -m\} \\ &\stackrel{a}{=} \Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}(m) - \Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}(-m), \end{aligned}$$

so that we can solve  $AMA E(\hat{\beta})$  from

$$\Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}(m) = 0.5 + \Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}(-m). \quad (40)$$

Since  $m = \Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}^{-1}[0.5 + \Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}(-m)]$ , we employed the iterative scheme,  $m_0 = 0$ ,  $m_{i+1} = \Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}^{-1}[0.5 + \Phi_{\ddot{\beta}, \sigma_{\ddot{\beta}}}(-m_i)]$  for  $i = 0, 1, \dots$  until convergence. When  $\ddot{\beta} = 0$  no iteration is required since  $AMA E(\hat{\beta}) = \Phi_{0, \sigma_{\ddot{\beta}}}^{-1}(0.75)$  conforms to the quartile.

We will re-examine here only the basic static model, described in the next subsection, that was earlier examined in Kiviet and Niemczyk (2007). In that paper the

conditional asymptotic approximation has been compared (inappropriately) with simulation results obtained under Framework U. Here, in the second subsection, we will supplement these results with simulations under Framework C and asymptotic approximations for the unconditional case, and then appropriate comparisons can be made. The diagrams presented below are single images from animated versions (available via <http://www.feb.uva.nl/ke/jfk.htm>), which allow to inspect the relevant phenomena over a much larger part of the parameter space. In the final subsection we sketch how our results on inconsistent OLS could be used for inference in practice.

## 4.1 The basic static IID model

We consider a model with one regressor and one valid and either strong or weak instrument, i.e.  $k = 1$  and  $l = 1$ . The two variables  $x$  and  $z$ , together with the dependent variable  $y$ , are jointly Gaussian IID with zero mean and finite second moments. For the data under Framework U we generated them exactly as in the earlier study, where we used as a base for the parameter space of the simulation design the three parameters:  $\rho_{x\varepsilon}$ ,  $\rho_{xz}$  and  $PF$  or population fit, where  $PF = SN/(SN + 1)$ , with  $SN$  (the signal-noise ratio) given by

$$SN = \beta^2 \sigma_x^2 / \sigma_\varepsilon^2 = \sigma_x^2 \geq 0, \quad (41)$$

because both  $\sigma_\varepsilon^2$  and  $\beta$  were standardized and taken equal to unity. This implies that

$$\sigma_x = \sqrt{PF/(1 - PF)}. \quad (42)$$

By varying the three parameters  $|\rho_{x\varepsilon}| < 1$  (simultaneity),  $|\rho_{xz}| < 1$  (instrument strength) and  $0 < PF < 1$  (model fit), we can examine the whole parameter space of this model, where  $\xi$  is now scalar and in fact equals  $\rho_{x\varepsilon} \sigma_x$ . The data for  $\varepsilon_i$ ,  $x_i$  and  $z_i$ , where the latter without loss of generality can be standardized such that  $\sigma_z = 1$ , can now be generated by transforming a  $3 \times 1$  vector  $v_i \sim NID(0, I_3)$  as follows:

$$\begin{pmatrix} \varepsilon_i \\ x_i \\ z_i \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{x\varepsilon} \sigma_x & \sigma_x \sqrt{1 - \rho_{x\varepsilon}^2} & 0 \\ 0 & \rho_{xz} / \sqrt{1 - \rho_{x\varepsilon}^2} & \sqrt{1 - \rho_{x\varepsilon}^2 - \rho_{xz}^2} / \sqrt{1 - \rho_{x\varepsilon}^2} \end{pmatrix} \begin{pmatrix} v_{1,i} \\ v_{2,i} \\ v_{3,i} \end{pmatrix}. \quad (43)$$

It is easy to check that this yields the appropriate data (co)variances and correlation coefficients indeed, with  $\rho_{z\varepsilon} = 0$ . After calculating  $y_i = x_i + \varepsilon_i$  one can straightforwardly obtain  $\hat{\beta}_{IV} = \Sigma z_i y_i / \Sigma z_i x_i$  and  $\hat{\beta}_{OLS} = \Sigma x_i y_i / \Sigma x_i^2$  in order to compare many independent replications of these estimators (in fact, we calculated the estimators as appropriate in models with an intercept, although this was actually zero in the DGP.) with their (pseudo-)true values  $\beta = 1$  and  $\beta^* = 1 + \rho_{x\varepsilon} / \sigma_x$  respectively. Likewise, one can calculate  $n\widehat{Var}(\hat{\beta}_{IV})$  and  $n\widehat{Var}(\hat{\beta}_{OLS})$  to compare these with (22), which here specializes to

$$AVar(\hat{\beta}_{IV}) = \rho_{xz}^{-2} / \sigma_x^2, \quad (44)$$

and with (32) or (38), which here specialize to

$$AVar_U^{NID}(\hat{\beta}_{OLS}) = (1 - \rho_{x\varepsilon}^2) / \sigma_x^2, \quad (45)$$

or

$$AVar_C^N(\hat{\beta}_{OLS}) = (1 - \rho_{x\varepsilon}^2)(1 - 2\rho_{x\varepsilon}^2 + 2\rho_{x\varepsilon}^4) / \sigma_x^2 \quad (46)$$

respectively. It can be shown that Rothenberg's formula (as used by Hausman), in which the conditioning is on the instruments, simplifies in this model to  $[1 - \rho_{x\varepsilon}^2(1 + 2\rho_{xz}^4)]/\sigma_x^2$ .

Since  $0 < \rho_{xz}^2 < 1$  and  $0 \leq \rho_{x\varepsilon}^2 < 1$  we have

$$\text{AVar}(\hat{\beta}_{IV}) > \text{AVar}_U^{NID}(\hat{\beta}_{OLS}) \geq \text{AVar}_C^N(\hat{\beta}_{OLS}). \quad (47)$$

From the simulations we will investigate how much these systematic asymptotic differences, jointly with the inconsistency of OLS, will affect the accuracy of these estimators in finite samples for particular values of  $\rho_{x\varepsilon}$ ,  $\rho_{xz}$  and  $n$ , and also how much conditioning does matter.

When simulating under Framework C, i.e. conditioning on  $\bar{x}_i = \sigma_x(1 - \rho_{x\varepsilon}^2)^{1/2}v_{2,i}$  and  $z_i = \rho_{xz}(1 - \rho_{x\varepsilon}^2)^{-1/2}v_{2,i} + (1 - \rho_{x\varepsilon}^2 - \rho_{xz}^2)^{1/2}(1 - \rho_{x\varepsilon}^2)^{-1/2}v_{3,i}$ , all Monte Carlo replications should use the same drawings of  $\bar{x}_i$  and  $z_i$ , i.e. be based on just one single realization of the series  $v_{2,i}$  and  $v_{3,i}$ . However, an arbitrary draw of  $v_{2,i}$  would generally give rise to an atypical  $\bar{x}$  series, in the sense that the resulting sample mean and sample variance of  $x$  may deviate from the values that they are supposed to have in the population. For the same reason the sample correlation of  $z_i$  and  $\bar{x}_i$  would differ from  $\rho_{xz}$ , and hence we would lose full control over the strength of the instrument. Therefore, when conditioning, although we used just one arbitrary draw of the series  $v_{2,i}$  and  $v_{3,i}$ , we did replace  $v_{3,i}$  by its residual after regressing on  $v_{2,i}$  and an intercept, in order to guarantee a sample correlation of zero between them. And next, to make sure that sample mean and variance of both  $v_{2,i}$  and  $v_{3,i}$  are appropriate too, we standardized them so that they have zero sample mean and unit sample variance. By stylizing  $\bar{x}_i$  and  $z_i$  in the simulations under Framework C, we realize that  $\bar{x}'\bar{x}/n + \sigma_\varepsilon^2\xi^2 = \sigma_x^2$  and  $z'\bar{x}/n = \rho_{xz}\sigma_x = \sigma_{xz}$ , as required by (18). In this way comparing the results of both frameworks becomes meaningful.

It is easily seen that the estimation errors (difference between estimate and true value  $\beta$ ) of both OLS and IV are a multiple of  $\sigma_x$ . Therefore, we do not have to vary  $\sigma_x$  in our simulations. We will just consider the case  $\sigma_x^2 = 10$ , which implies  $SN = 10$  and  $PF = 10/11 = .909$ . Results for different values of  $\sigma_x$  can directly be obtained from these by rescaling. Hence, we will only have to vary  $n$ ,  $\rho_{x\varepsilon}$  and  $\rho_{xz}$ , where the latter self-evidently has no effects on OLS estimation. In the present model we have to restrict the grid of values to the circle  $\rho_{x\varepsilon}^2 + \rho_{xz}^2 \leq 1$ . We just consider nonnegative values of  $\rho_{x\varepsilon}$  and  $\rho_{xz}$  because the effects of changing their sign follow rather simple symmetry rules.

## 4.2 Comparisons of actual and asymptotic distributions

In the Figures 1 through 4 below general characteristics of the (un)conditional distributions of IV and OLS are analyzed by comparing their MAE's, both in actual finite samples and as approximated by standard first-order asymptotics. From (43) it is easily seen that  $\hat{\beta}_{OLS} - \beta = \Sigma x_i \varepsilon_i / \Sigma x_i^2$  and  $\hat{\beta}_{IV} - \beta = \Sigma z_i \varepsilon_i / \Sigma z_i x_i$  are both multiples of  $\sigma_x^{-1}$ , thus so are their MAE's. Hence, ratios of MAE's are invariant with respect to PF (i.e. to  $\sigma_x$ ), and the only determining factors for the IV results are  $\rho_{x\varepsilon}$ ,  $\rho_{xz}$  and  $n$ , and for OLS just  $\rho_{x\varepsilon}$  and  $n$ . These figures are based on  $10^6$  replications in the Monte Carlo simulations. The Figures 5 through 9 present densities for specific DGP's. There we used  $2 \times 10^6$  replications. The results on the conditional distribution have all been obtained for the same stylized series of arbitrary  $v_{2,i}$  and  $v_{3,i}$  series. We also tried a few different stylized series, but the results did not differ visibly.

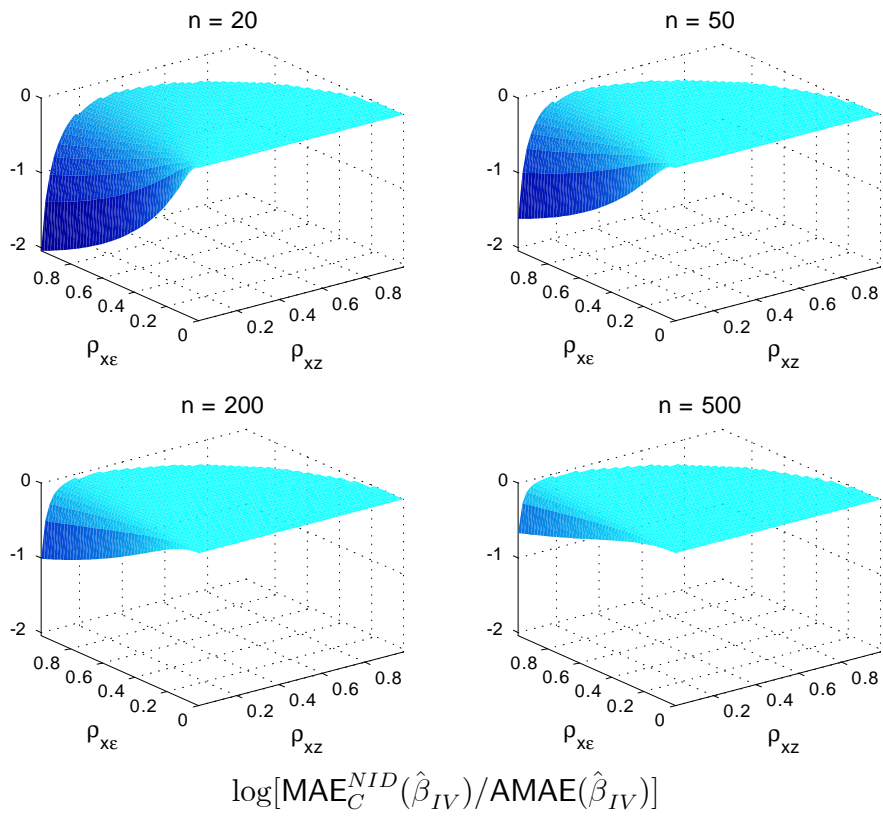
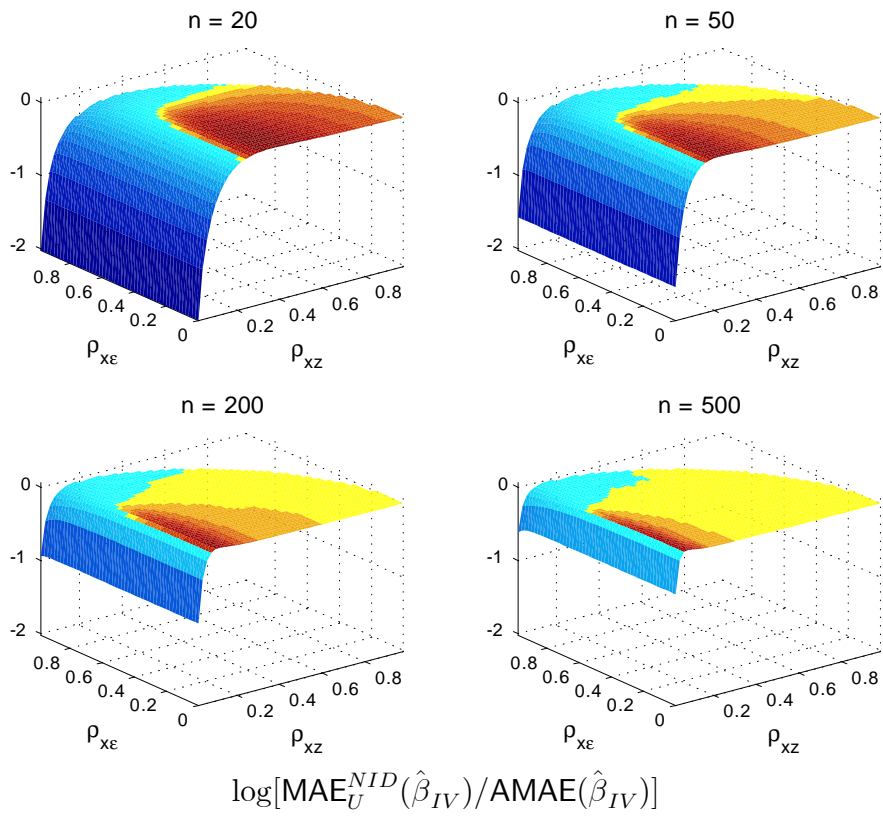


Figure 1: Accuracy of asymptotic approximations for IV



Figure 1 depicts for different values of  $n$  the accuracy of the asymptotic approximations for IV, over all compatible positive values of  $\rho_{x\varepsilon}$  and  $\rho_{xz}$ . We see from the 3D graphs on  $\log[\text{MAE}_U^{NID}(\hat{\beta}_{IV})/\text{AMAe}(\hat{\beta}_{IV})]$  and  $\log[\text{MAE}_C^{NID}(\hat{\beta}_{IV})/\text{AMAe}(\hat{\beta}_{IV})]$  that for this model with NID observations the asymptotic approximation seems reasonably accurate when the instrument is not weak, even when the sample size is quite small. But for small values of  $\rho_{xz}$ , as is well known, the approximation errors by standard asymptotics are huge and much too pessimistic. However, we also establish here that for the conditional distribution they are less severe when the simultaneity is mild. Note that when these ratio's are  $-1$  this means that the asymptotic approximation overstates the actual MAE's by a factor  $\exp(1) \approx 2.7$ . Hence, we find that the asymptotic approximation for the unconditional distribution is rather unsatisfactory for  $|\rho_{xz}| < 0.1$ , especially when  $n$  is small, irrespective of  $\rho_{x\varepsilon}$ , whereas only for large  $\rho_{x\varepsilon}$  the same holds for the conditional distribution. Note, however, that these graphs show that the actual distribution of IV when instruments are weak is not as bad as the asymptotic distribution suggests.

Figure 2 presents similar results for OLS. We note a remarkable difference with IV. Here (for  $n \geq 20$ ) the first-order asymptotic approximations never break down, because no weak instrument problem exists. The accuracy varies nonmonotonically with the degree of simultaneity. For  $n$  only 20 the discrepancy does not exceed 2.1% (for the unconditional distribution) or 3.9% (for the conditional distribution). Asymptotics has a tendency to understate the accuracy of the unconditional distribution and to overstate the accuracy under conditioning.

In Figure 3 we focus on the effects on estimator accuracy of conditioning in finite samples. In the 3D graphs on IV we note a substantial difference in MAE (especially for small  $n$ ) when both  $\rho_{x\varepsilon}$  and  $\rho_{xz}$  are small, with the unconditional distribution more tightly centered around the true value of  $\beta$  than the conditional distribution. However, especially when the sample size is small the conditional distribution is somewhat more attractive when the instrument is not very weak. The two panels with graphs on OLS show that conditioning has moderate positive effects on OLS accuracy for intermediate values of  $\rho_{x\varepsilon}$  and especially when the sample size is small. The pattern of this phenomenon is predicted by the asymptotic approximations, but not without approximation errors.

Figure 4 provides a general impression of the actual qualities of IV and OLS in finite samples in terms of relative MAE. The top panel compares unconditional OLS and IV. We note that IV performs better than OLS when both  $\rho_{x\varepsilon}$  and  $\rho_{xz}$  are substantial in absolute value, i.e. when both simultaneity is serious and the instrument relatively strong. Of course, the area where OLS performs better diminishes when  $n$  increases. Where the ratio equals 2, IV is  $\exp(2) \times 100\%$  or about 7.5 times as accurate as OLS, whereas where the log-ratio is -3 OLS is  $\exp(3)$  (i.e. about 20) times as accurate as IV. We notice that over a substantial area in the parameter space the OLS efficiency gains over IV are much more impressive than its maximal losses can ever be. OLS seems to perform worst when  $|\rho_{x\varepsilon}| = |\rho_{xz}| = 0.5$ . We did not include similar 3D graphs comparing conditional OLS and IV, but found that for the smaller sample size the OLS gains over IV are even more substantial than under conditioning when the instrument is weak, especially when the simultaneity is moderate. The effects in this respect of conditioning can directly be observed from the bottom panel of Figure 4 in which the pattern of the difference between the two relevant log MAE ratios is shown.

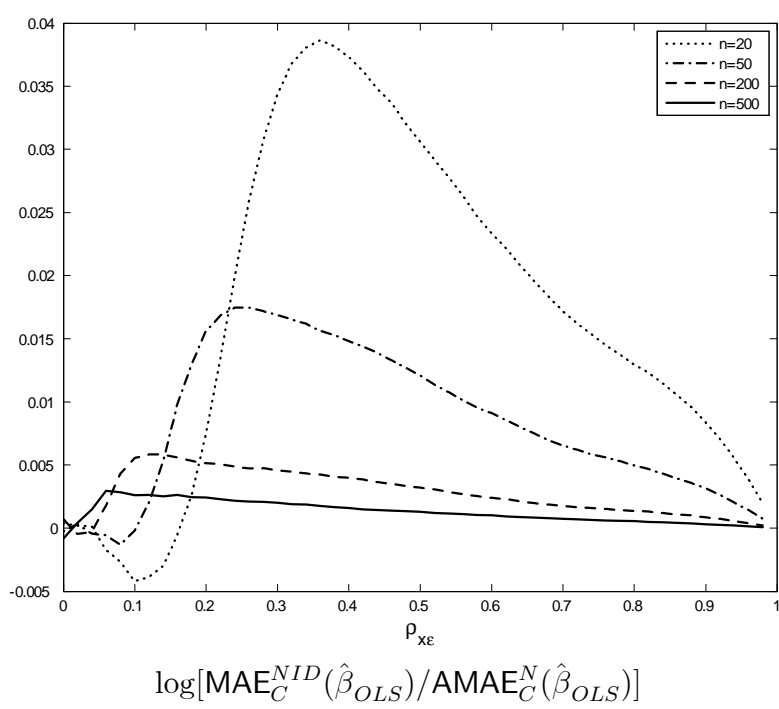
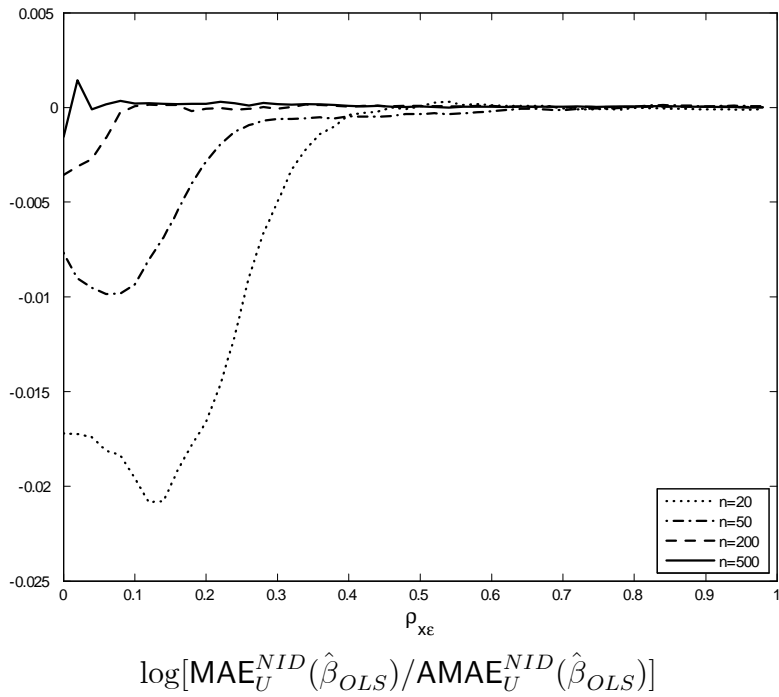


Figure 2: Accuracy of asymptotic approximations for OLS

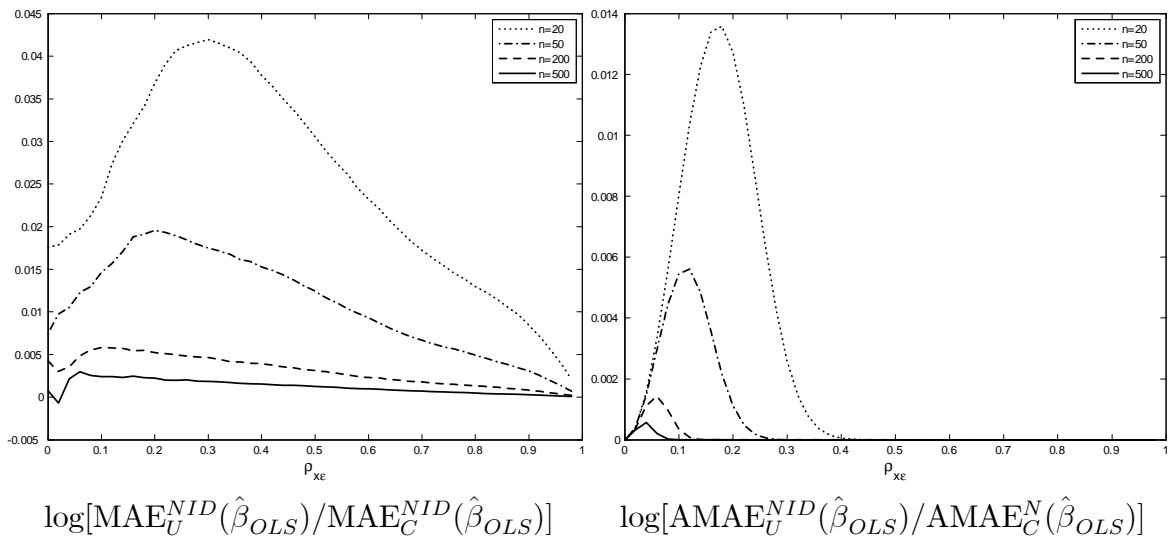
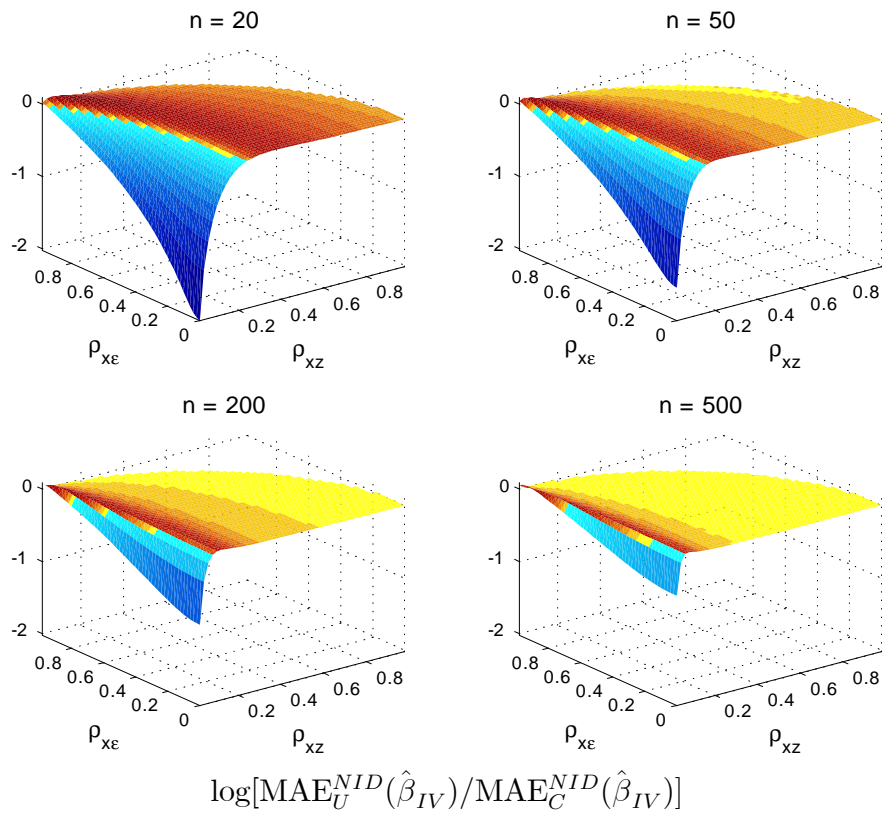


Figure 3: Effect of conditioning on efficiency for IV and for OLS

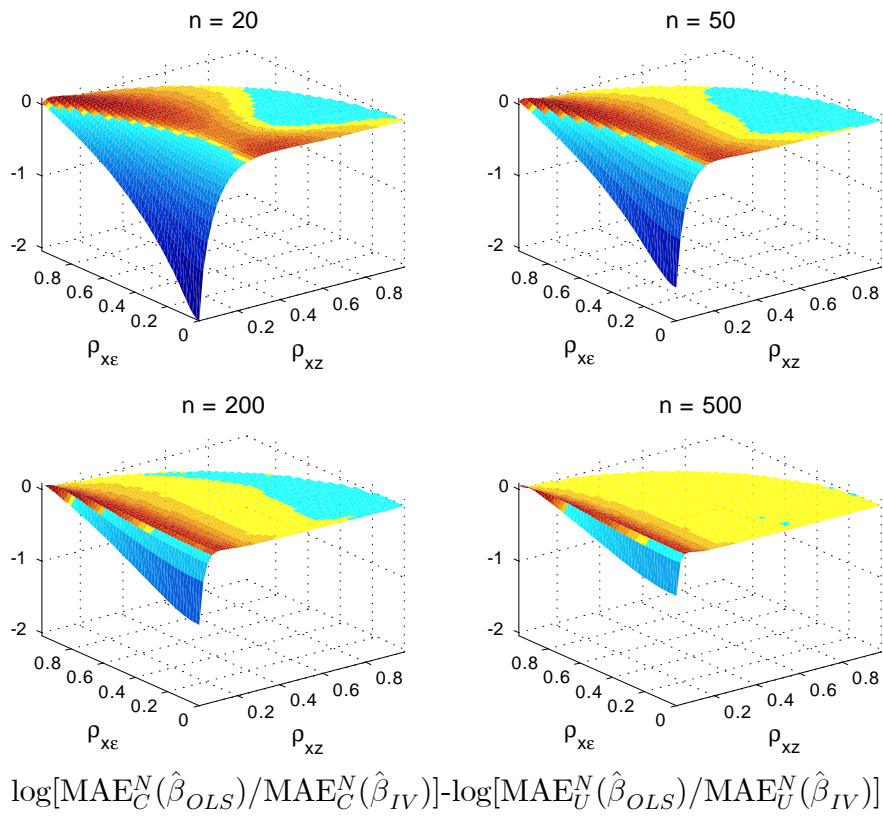
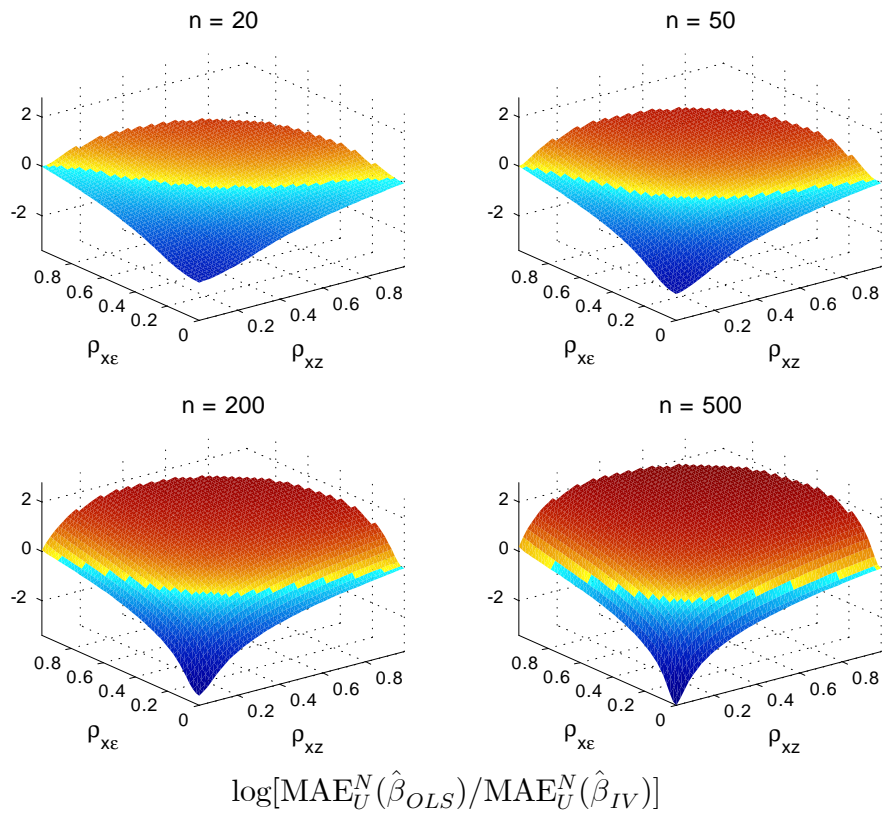


Figure 4: Relative actual estimator efficiency, OLS versus IV, C versus U

The remaining figures contain the actual densities of IV and OLS for particular values of  $n$ ,  $\rho_{xz}$  and  $\rho_{x\varepsilon}$  and their first-order asymptotic approximation. From these one can see more subtle differences than by the unidimensional MAE criterion, because they expose separately any differences in location and in scale, and also deviations from normality like skewness or bimodality. All these figures consist of two panels, where each panel contains densities for the four cases  $\rho_{x\varepsilon} = 0.1, 0.2, 0.4$  and  $0.6$  respectively. Note that within each of these Figures the scales of the vertical and the horizontal axes are kept constant, but that these differ between most of the Figures.

Figures 5 and 6 present the same cases for  $n = 50$  and  $n = 200$  respectively. In Figure 5 we see both OLS and IV for a strong instrument where  $\rho_{xz} = 0.8$ . For OLS we note the inconsistency and also the smaller variance of the conditional distribution and the great accuracy of the asymptotic approximations. For IV with a strong instrument the distribution is well centered and the asymptotic approximation is not bad either, but for serious simultaneity we already note some skewness of the actual distributions which self-evidently is not a characteristic of the Gaussian approximation. In Figure 6, due to the larger sample size, the approximations are more accurate of course. Differences between the conditional and unconditional distributions become apparent only for OLS when the simultaneity is serious.

Figures 7, 8 and 9 are all about IV. In Figure 7 the instrument is weak, since  $\rho_{xz} = 0.2$ , but not as weak as in Figure 8, where  $\rho_{xz} = 0.02$ . The upper panels are for  $n = 50$  and the lower panels (using the same scale) are for  $n = 200$ . Hence, any problems in the upper panels become milder in the lower panels, but we note that they are still massive for the very weak instrument when  $n = 200$ . All these panels show that the unconditional IV distribution is more attractive than the conditional one, as we already learned from the MAE figures. The conditional distribution is more skew, and shows bimodality when the instrument is very weak and the simultaneity substantial. The asymptotic approximation is still reasonable when  $\rho_{xz} = 0.2$ , but useless and much too pessimistic when  $\rho_{xz} = 0.02$ , also when  $n = 200$ . Figure 9 examines the very weak instrument case for larger samples, and shows that even at  $n = 1000$  the approximation is very poor, and the unconditional distribution is better behaved than the conditional one. At  $n = 5000$  the approximation is reasonable, provided the simultaneity is mild. Though note, that the IV estimator at  $n = 5000$  varies over a domain which is much wider than that of OLS at  $n = 50$ , which highlights that employing a strong invalid instrument is preferable to using a valid but weak one.

### 4.3 Adapted OLS inference for a simple simultaneous equation

Without knowing the degree of simultaneity  $\rho_{x\varepsilon}$ , it is impossible to provide a measure for the bias and the variance of OLS. Nevertheless, our approximations to the unconditional and to the more attractive conditional distributions of inconsistent OLS should allow to produce an indication of the magnitude of the OLS bias and its standard error under a range of likely values of  $\rho_{x\varepsilon}$ . Next, it should be feasible to produce a range of confidence sets for  $\beta$ , conditional on these  $\rho_{x\varepsilon}$  values. Their union should in principle provide a confidence set for  $\beta$  which is asymptotically exact (but probably conservative) for, say,  $\rho_{x\varepsilon} \in [-0.2, 0.6]$ . This approach has been followed for one empirical data set in Niemczyk (2009, p.166). In a very small simulation study we will examine here whether such an approach works in the simple model with  $k = 1$ .

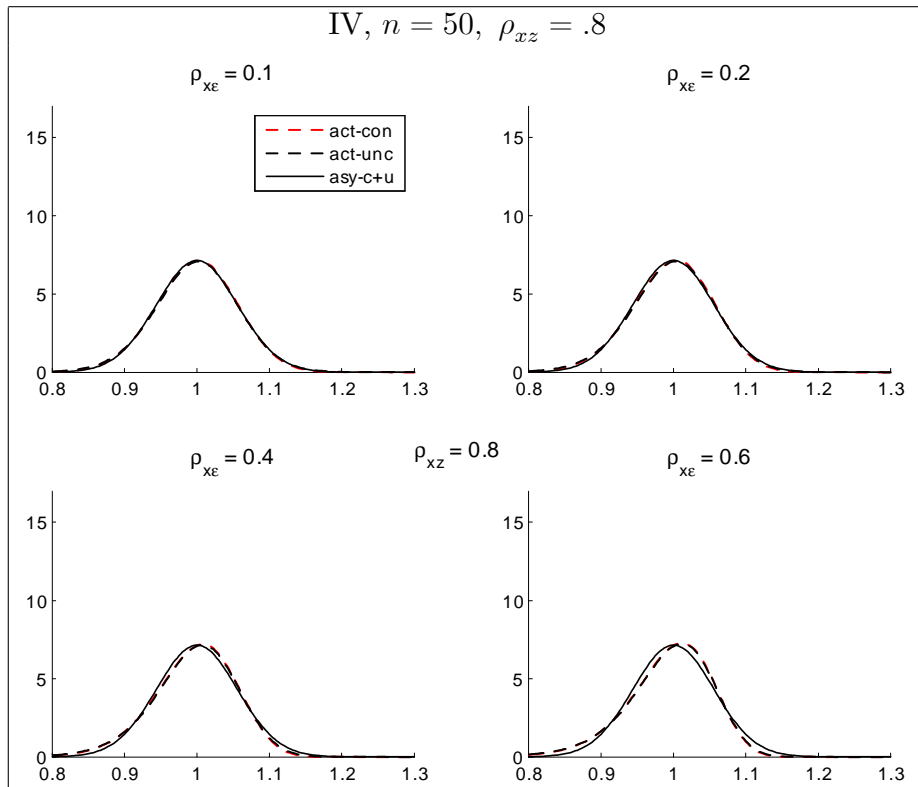
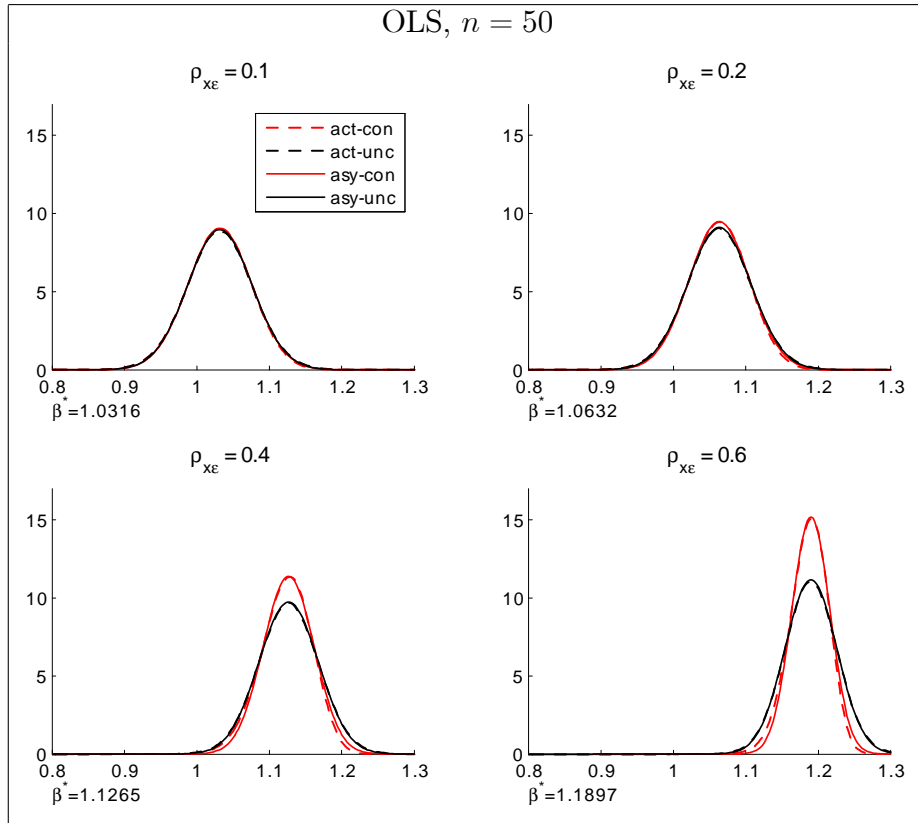


Figure 5: OLS and IV (strong) for  $n = 50$

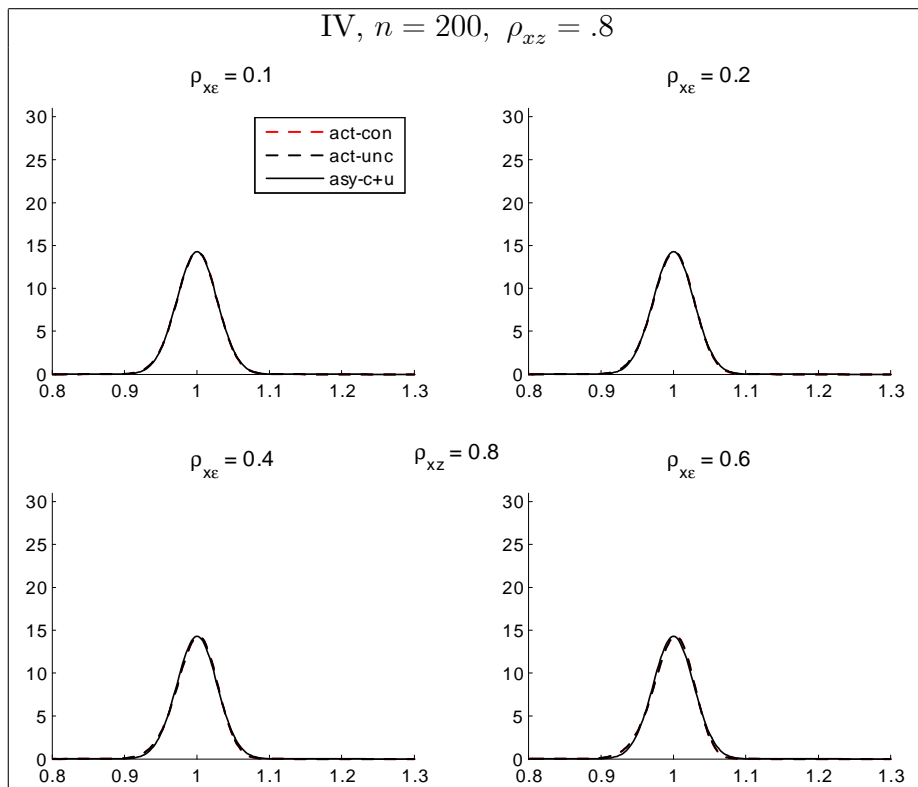
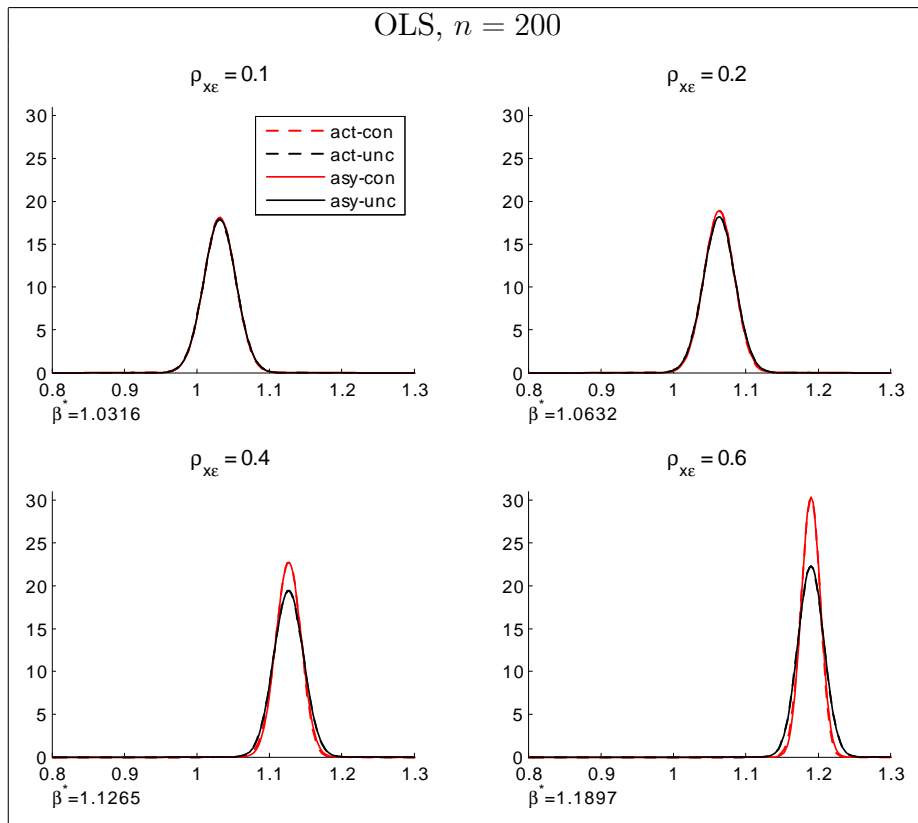


Figure 6: OLS and IV (strong) for  $n = 200$

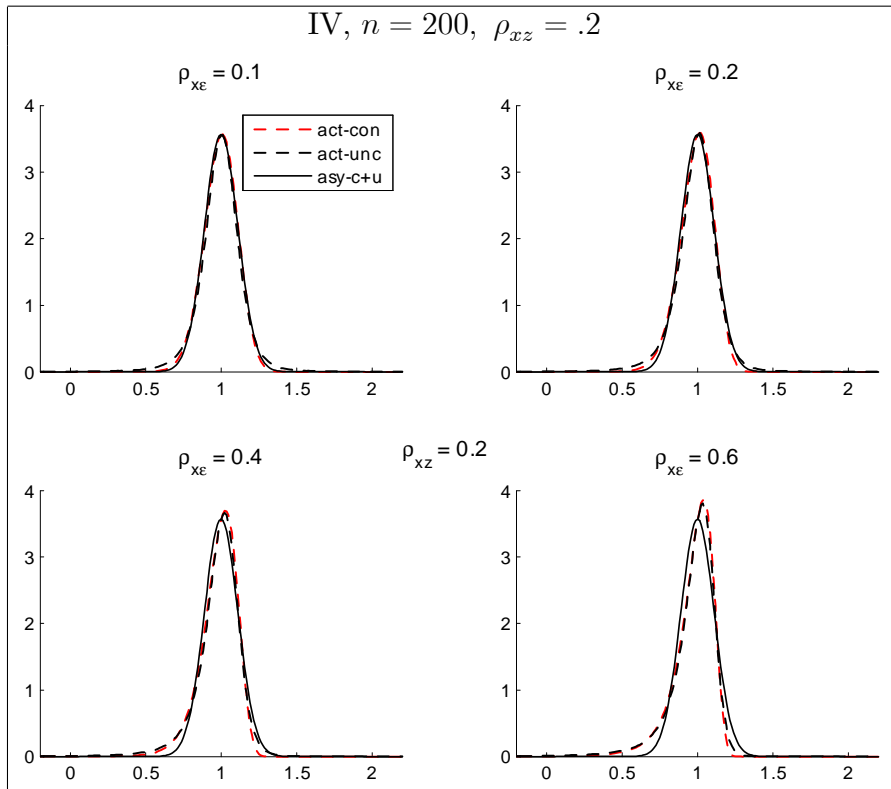
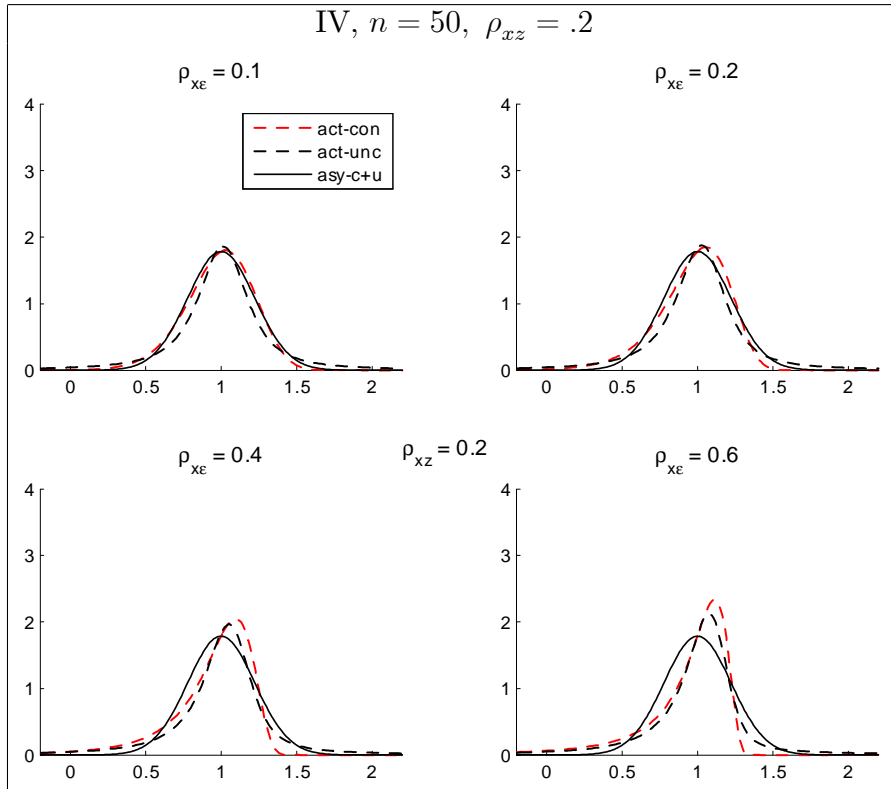


Figure 7: IV (weak) for  $n = 50, 200$



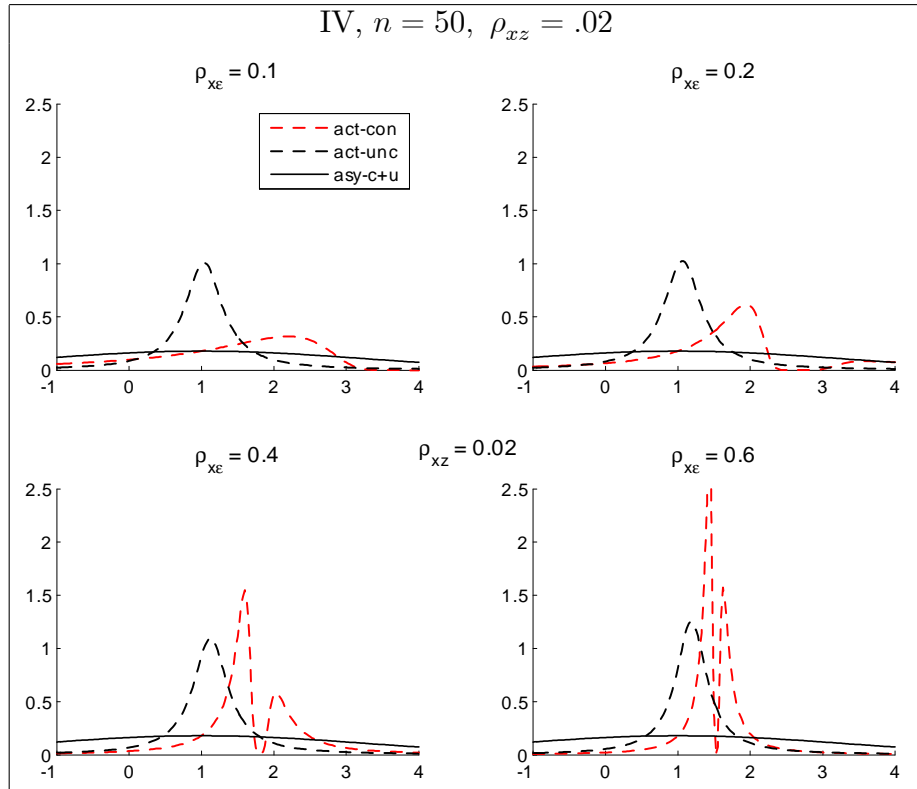


Figure 8: IV (very weak) for  $n = 50, 200$

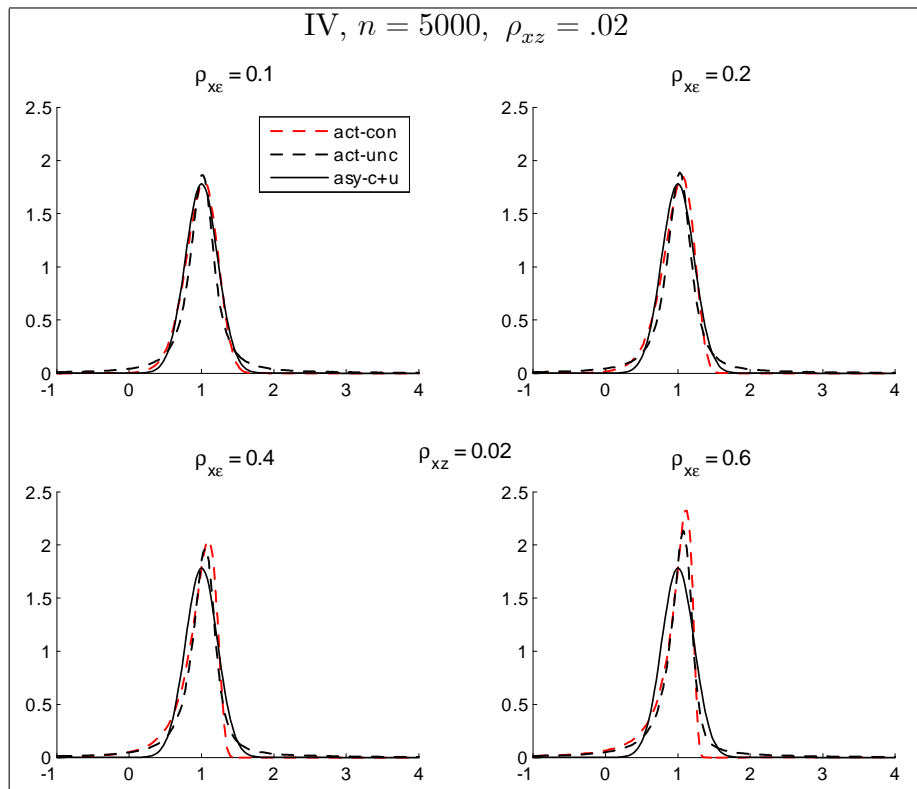
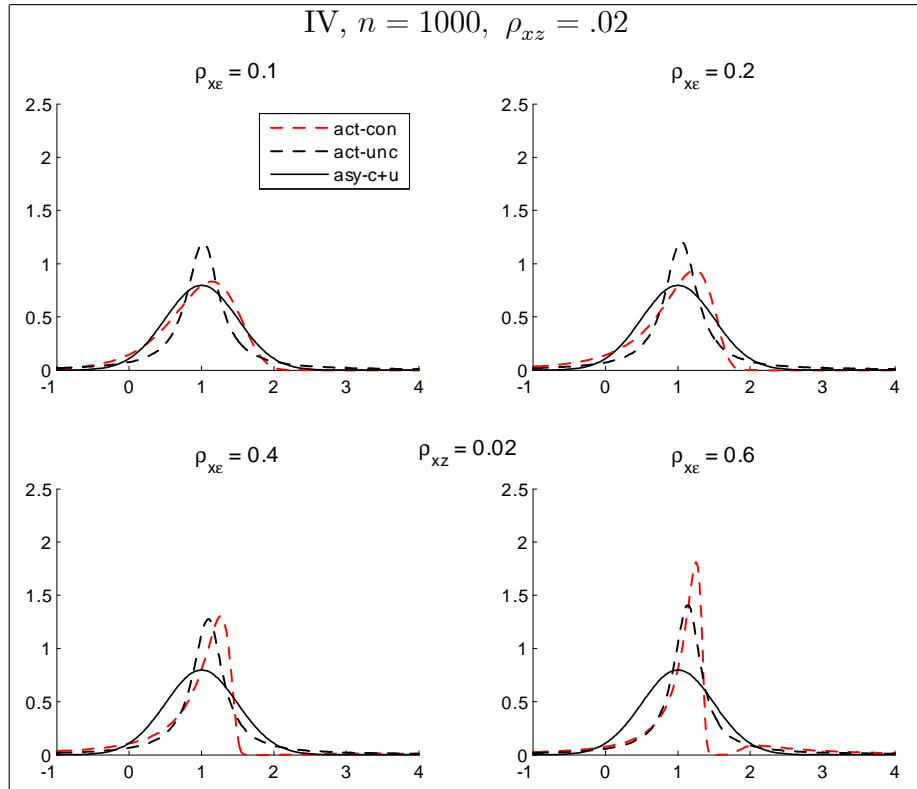


Figure 9: IV (very weak) for  $n = 1000, 5000$

Applying OLS we obtain  $\hat{\beta}_{OLS}$  and  $\widehat{Var}(\hat{\beta}_{OLS}) = s^2 / \sum_{i=1}^n x_i^2$ , where  $s^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2 / (n-1)$  with  $\hat{\varepsilon}_i = y - \hat{\beta}_{OLS}x_i$ . If all model characteristics were known, one might use in the model with Gaussian disturbances, while conditioning on  $\bar{x}$ , the modified though totally unfeasible estimators

$$\begin{cases} \check{\beta}_{OLS} \equiv \hat{\beta}_{OLS} - \rho_{x\varepsilon}\sigma_\varepsilon/\sigma_x, \\ \widehat{Var}_C^N(\check{\beta}_{OLS}) = n^{-1}\mathbf{AVar}_C^N(\hat{\beta}_{OLS}) \equiv (1 - \rho_{x\varepsilon}^2)(1 - 2\rho_{x\varepsilon}^2 + 2\rho_{x\varepsilon}^4)\sigma_\varepsilon^2/(n\sigma_x^2). \end{cases} \quad (48)$$

Exploiting  $\text{plim } s^2 = (1 - \rho_{x\varepsilon}^2)\sigma_\varepsilon^2$  and  $\text{plim } n\widehat{Var}(\hat{\beta}_{OLS}) = (1 - \rho_{x\varepsilon}^2)\sigma_\varepsilon^2/\sigma_x^2$ , we can adapt these such that they just contain OLS estimates and the parameter  $\rho_{x\varepsilon}$ , giving

$$\begin{cases} \check{\beta} \equiv \hat{\beta}_{OLS} - \left[ \frac{\rho_{x\varepsilon}^2}{1 - \rho_{x\varepsilon}^2} n\widehat{Var}(\hat{\beta}_{OLS}) \right]^{1/2}, \\ \widehat{Var}_C^N(\check{\beta}) \equiv (1 - 2\rho_{x\varepsilon}^2 + 2\rho_{x\varepsilon}^4)\widehat{Var}(\hat{\beta}_{OLS}). \end{cases} \quad (49)$$

Note that, conditional on making an assumption on the value of  $\rho_{x\varepsilon}$ ,  $\check{\beta}$  is consistent for  $\beta$  and  $n\widehat{Var}_C^N(\check{\beta})$  for  $\mathbf{AVar}_C^N(\check{\beta}) = \mathbf{AVar}_C^N(\hat{\beta}_{OLS})$ , in the same way as standard OLS is consistent provided the assumption  $\rho_{x\varepsilon} = 0$  is correct.

Table 1: Qualities of adapted OLS inference

<i>SNR</i>	<i>n</i>	$\rho_{x\varepsilon}$	<i>Bias</i> ( $\hat{\beta}_{OLS}$ )	<i>Bias</i> ( $\check{\beta}$ )	<i>Var</i> ( $\check{\beta}$ )	$E\left(\widehat{Var}_C^N(\check{\beta})\right)$	<i>Cov.Prob.</i>	
3	40	.2	.113	.001 [-.003]	.0084 [.0077]	.0076 [.0084]	.929 [.945]	
		.5	.286	.066 [-.003]	.0065 [.0041]	.0041 [.0047]	.714 [.944]	
	100	.2	.114	.003 [-.001]	.0033 [.0030]	.0030 [.0031]	.934 [.949]	
		.5	.288	.070 [-.001]	.0025 [.0016]	.0016 [.0018]	.540 [.947]	
	10	40	.2	.062	.001 [-.001]	.0025 [.0023]	.0023 [.0025]	.929 [.945]
			.5	.157	.036 [-.002]	.0020 [.0012]	.0012 [.0014]	.714 [.944]
100		.2	.063	.002 [-.001]	.0010 [.0009]	.0009 [.0009]	.934 [.949]	
		.5	.157	.038 [-.001]	.0008 [.0005]	.0005 [.0005]	.540 [.947]	

Table 1 contains some results on  $\hat{\beta}_{OLS}$  and the estimators of (49) based on 100,000 Monte Carlo replications for  $\sigma_x^2 = 3$  and  $\sigma_x^2 = 10$ ,  $n = 40$  and  $n = 100$ , and  $\rho_{x\varepsilon} = 0.2$  and  $\rho_{x\varepsilon} = 0.5$ . In square brackets the corresponding results are presented for the strongly unfeasible estimators (48). The latter demonstrate again that despite the simultaneity the asymptotic OLS approximations are very accurate in finite samples and yield confidence intervals (symmetric around  $\tilde{\beta}$  and based on the  $\alpha = 0.025$  critical value of the normal distribution) with actual coverage probabilities very close to the nominal level. Asymptotically valid inference based on  $\tilde{\beta}$ , which is only unfeasible regarding  $\rho_{x\varepsilon}$ , is quite accurate for low values of  $\rho_{x\varepsilon}$ , but especially for higher  $\rho_{x\varepsilon}$  it clearly requires some finite sample corrections. It is obvious that  $\widehat{Var}_C^N(\tilde{\beta})$  of (49) only takes the asymptotic variance of the first term of  $\tilde{\beta}$  into account, and not of the second one which is random too for  $n$  finite. Moreover, the bias of  $\tilde{\beta}$  shows that its present definition would benefit from some further refinements. These could be based either on higher-order expansions or on bootstrapping, or on both. This requires substantial further study, and when successful, the length of the resulting confidence intervals (based on an assumed interval for  $\rho_{x\varepsilon}$ ) should be compared with the length of intervals obtained by IV (for a range of values of  $\rho_{xz}$ ), where in case of weak instruments these intervals should be made robust, see Mikusheva (2010).

## 5 Conclusions

In this paper we examined the effects of conditioning for rather standard econometric models and estimators. We analyzed the analytic and numerical effects of conditioning on first-order asymptotic approximations, as well as its consequences in finite samples by running appropriately designed Monte Carlo experiments. For many published results on simulation studies it is not always clear whether or not they have been obtained by keeping exogenous variables fixed or by redrawing them every replication, whereas knowing this is crucial when interpreting the results. From our simulations it seems, that many of the complexities that have been studied recently on the consequences of weak instruments for the IV distribution, such as bimodality, are simply the result of conditioning; see, for instance, Hillier (2006) and the references therein. We find that the unconditional IV distribution may be quite well behaved (it is much closer to normal and less dispersed). Although it is still not very accurately approximated by standard asymptotic methods, it is most probably much easier to find a good approximation for it, than for the deranged conditional distribution.

Our major finding is that a consistent estimator does not necessarily outperform an inconsistent one in finite samples, and our major contribution is that we provide an accurate asymptotic approximation to the distribution of inconsistent OLS and show how this is affected by conditioning. It may be illuminating to highlight the similarities and differences with results obtained recently in Kiefer and Vogelsang (2005). They consider just one technique, viz. HAC estimation, but examine asymptotic approximations based on two different parameter sequences, the standard one which yields consistency and an alternative one which is inconsistent. Here the inconsistent one yields better approximations to the actual finite sample distribution of test statistics incorporating HAC estimation.

We find that the dispersion of both unconditional and conditional IV when the instrument is weak is such that inconsistent OLS in general establishes a much more accurate estimator. From our Figures we find that for  $n \leq 200$  less than 100% of the

(un)conditional IV estimates of  $\beta = 1$  in the simulation were (when  $\rho_{xz} = .02$ ) in the interval  $[0, 2]$ , whereas all OLS estimates were in the much narrower interval  $[.9, 1.3]$ . For  $\rho_{xz} = .2$  this IV interval is  $[.5, 1.5]$ , underscoring that OLS estimates are often much and much more accurate than IV estimates. We have also indicated how OLS, which by its very nature always uses the strongest – though possibly invalid – instruments, might still be used in practice for inference purposes, when it has been assessed that some of the available valid instruments are too weak to put one’s trust fully in extremely inefficient standard IV inference.

## Acknowledgements

We are especially grateful for the advice from the Editors Erricos Kontoghiorghes and Lynda Khalaf and for the constructive comments by two anonymous referees.

## References

- Andrews, D.W.K., Stock, J.H., 2007. Inference with Weak Instruments, Chapter 6 in: Blundell, R., Newey, W.K., Persson, T. (eds.), *Advances in Economics and Econometrics, Theory and Applications*, 9th Congress of the Econometric Society, Vol. 3. Cambridge, UK: Cambridge University Press.
- Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variable estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90, 443-450.
- Doornik, J.A., 2006. The Role of Simulation in Econometrics. Chapter 22 (pp. 787-811) in: Mills, T.C., Patterson, K. (eds.). *Palgrave Handbooks of Econometrics* (Volume 1, Econometric Theory). Basingstoke, Palgrave MacMillan.
- Edgerton, D.L., 1996. Should stochastic or non-stochastic exogenous variables be used in Monte Carlo experiments? *Economics Letters* 53, 153-159.
- Goldberger, A.S., 1964. *Econometric Theory*. John Wiley & Sons. New York.
- Hahn, J., Hausman, J.A., 2003. Weak instruments: Diagnosis and cures in empirical econometrics. *American Economic Review* 93, 181-125.
- Hausman, J.A., 1978. Specification tests in econometrics. *Econometrica* 46, 1251-1271.
- Hillier, G., 2006. Yet more on the exact properties of IV estimators. *Econometric Theory* 22, 913-931.
- Joseph, A.S., Kiviet, J.F., 2005. Viewing the relative efficiency of IV estimators in models with lagged and instantaneous feedbacks. *Journal of Computational Statistics and Data Analysis* 49, 417-444.
- Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* 21, 1130-1164.
- Kiviet, J.F., 2007. Judging contending estimators by simulation: Tournaments in dynamic panel data models. Chapter 11 (pp.282-318) in Phillips, G.D.A., Tzavalis, E. (Eds.). *The refinement of econometric estimation and test procedures; finite Sample and asymptotic analysis*. Cambridge University Press.
- Kiviet, J.F., Niemczyk, J., 2007. The asymptotic and finite sample distribution of OLS and simple IV in simultaneous equations. *Journal of Computational Statistics and Data Analysis* 51, 3296-3318.
- Kiviet, J.F., Niemczyk, J., 2009a. The asymptotic and finite sample (un)conditional

distributions of OLS and simple IV in simultaneous equations. *UvA-Econometrics Discussion paper 2009/01*.

Kiviet, J.F., Niemczyk, J., 2009b. On the limiting and empirical distribution of IV estimators when some of the instruments are invalid. *UvA-Econometrics Discussion paper 2006/02 (revised September 2009)*.

Mikusheva, A., 2010. Robust confidence sets in the presence of weak instruments. *Journal of Econometrics* 157, 236-247.

Niemczyk, J., 2009. *Consequences and detection of invalid exogeneity conditions*. PhD-thesis, Tinbergen Institute Research Series no. 462, Amsterdam.

Phillips, P.C.B., Wickens, M.R., 1978. *Exercises in Econometrics*. Philip Allen and Ballinger, Cambridge MA.

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

Schneeweiss, H., Srivastava, V.K., 1994. Bias and mean squared error of the slope estimator in a regression with not necessarily normal errors in both variables. *Statistical Papers* 35, 329-335.

Theil, H., 1971. *Principles of Econometrics*. John Wiley and Sons, New York.

Woglom, G., 2001. More results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 69, 1381-1389.