Downloaded from UvA-DARE, the institutional repository of the University of Amsterdam (UvA)
http://hdl.handle.net/11245/2.69406

File ID	uvapub:69406
Filename	Chapter 2: The asymptotic and finite sample (un)conditional distributions of OLS and simple IV in simultaneous equations
Version	unknown

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	PhD thesis
Title	Consequences and detection of invalid exogeneity conditions
Author(s)	J. Niemczyk
Faculty	FEB: Amsterdam School of Economics Research Institute (ASE-RI)
Year	2009

FULL BIBLIOGRAPHIC DETAILS:
http://hdl.handle.net/11245/1.312972

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content licence (like Creative Commons).

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
(pagedate: 2014-11-18)

Chapter 2

The asymptotic and finite sample (un)conditional distributions of OLS and simple IV in simultaneous equations

2.1 Introduction

Relatively little attention has been paid in the econometric literature to the limiting distribution of inconsistent estimators. Usually, when developing and rating alternative estimators, consistency has been considered a minimum requirement. This seems very reasonable when actual samples are so large that estimation variance is relatively small. in finite samples, however, it could well be the case that, when the bias of alternative consistent and inconsistent estimators is of similar magnitude whereas the inconsistent one has smaller variance than its consistent rival, the consistent estimator is actually less precise according to reasonable criteria to be operationalized below. An example where this occurs is in estimating dynamic panel data models, where so-called fully efficient GMM estimators may actually have larger mean squared error (MSE) than inconsistent least-squares estimators, see Bun and Kiviet (2006). For a completely specified data generating process any such differences can easily be assessed from Monte Carlo experiments,

but may only persuade practitioners to use inconsistent but actually more precise estimators when at the same time techniques are developed to use them accurately for inference purposes. The present study embarks on this by deriving an explicit characterization of the limiting distribution of an inconsistent estimator and examining its accuracy for actual behavior in finite samples.

We focus on least-squares and instrumental variable estimators in a linear structural equation from a simultaneous system. Goldberger (1964, p. 359) considers a very specific case and derives the asymptotic variance of inconsistent OLS. An early – but incomplete – attempt to obtain the limiting distribution of OLS in a simple specific case can be found in Phillips and Wickens (1978, problem 6.10). A derivation in a more general context for an IV estimator that may contain invalid instruments (note that OLS is thus a special case) can be found in Maasoumi and Phillips (1982), see also Hendry (1979, 1982). However, they do not provide an explicit representation, and they focus on the unconditional limiting distribution in a large dynamic system, whereas we shall obtain an explicit expression for both the conditional and unconditional limiting distribution of inconsistent OLS in particular linear models. Such an approach is also followed in Rothenberg (1972). Our approach differs, because we do not start off from an errors in variables context, but from a more generic parametrization, which covers all kinds of linear contemporaneous and lagged dependence of regressors on disturbances. Joseph and Kiviet (2005) also made an attempt to derive an explicit representation of the limiting distribution of an inconsistent OLS estimator, but we will show here that that result is incomplete. By developing a useful decomposition of the OLS estimation error and by applying a rather standard form of the central limit theorem (CLT), we will derive here a general representation of the limiting distribution of OLS, both unconditional and conditional on predetermined information, in a linear regression model where the regressors are stationary and contemporaneously correlated with the disturbance term. We find this distribution to be normal and centered at the pseudo true value (true coefficient plus inconsistency) with an asymptotic variance that can simply be expressed as a correction to the asymptotic variance of a consistent OLS estimator, where this correction is based on the actual inconsistency and a measure for the simultaneity. It can easily be shown that in general this asymptotic

variance gets smaller (in a matrix sense) when the simultaneity and thus the inconsistency become more severe. However, this is not the case for the first-order asymptotic approximation to the MSE of OLS. We make comparisons with the asymptotic variance of consistent IV implementations in specific simple static simultaneous models. By that we establish areas in the parameter space where OLS beats IV on the basis of asymptotic MSE. In addition, we examine the accuracy of these asymptotic approximations via simulation experiments. In order to ease the presentation, absorption and interpretation of our extensive numerical findings they are all put into colored 2D and 3D diagrams. All these diagrams are in fact single images of animations (3D and 4D diagrams) which, when viewed as a film on a monitor via the web, allow to depict the various most relevant phenomena in more than three dimensions.

In order to limit the size of this Chapter we make actual comparisons between OLS and just identified consistent IV estimation only, i.e. exploiting precisely as many valid instruments as regressors. This implies that we have to take into account the non-existence of moments of IV. At a later stage we also plan to examine overidentified cases and to compare consistent IV and inconsistent IV implementations which exploit some invalid instruments. Then a recent study by Hall and Inoue (2003) will become relevant. They examined generalized method of moments estimators in misspecified models. Loosely formulated they define misspecification as exploiting orthogonality conditions which are in fact false for any possible parameter value, whereas they exclude the case where as many orthogonality conditions as parameters are employed. Hence, they exclude the case of OLS when some of the regressors are in fact invalid instruments, which is precisely the main focus of the present study.

Our major finding is that inconsistent OLS often outperforms consistent IV when the sample size is finite. For a simple specific class of models we find that in samples with a size between 20 and 200 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 to the estimation errors of OLS (both conditional and unconditional) are very

accurate even in relatively small samples, which is not the case for IV when instruments are weak, see also Bound, Jaeger, and Baker (1995). For consistent IV one needs alternative asymptotic sequences when instruments are weak, see for an overview Andrews and Stock (2007), whereas generally speaking standard first-order asymptotic approximations seem to work very well for OLS, which by its very nature always uses the strongest possible, though possibly invalid, instruments. Especially when simultaneity is serious, the actual conditional distribution of OLS is found to be more attractive than its unconditional counterpart. Hence, its asymptotic distribution derived here, which turns out to be highly accurate, could and should be used in future research as a tool for producing inference based on OLS and an assumption on the degree of simultaneity.

The structure of this Chapter is as follows. In Section 2 we introduce the model and some of its particulars, especially the standard asymptotic properties of OLS and IV when the data are stationary. Next in Section 3 we derive the limiting distribution of OLS when the dependent variable is in fact jointly dependent with some of the regressors. We distinguish between the unconditional limiting distribution, and the effects of conditioning on predetermined variables. In Section 4 we discuss the measures that we will use to make comparisons between the performance of different estimators. We address the issues that are relevant when using the limiting behavior of an inconsistent estimator for such a comparison. For representing the actual finite sample performance obtained from Monte Carlo experiments, we develop alternative measures for situations where IV has no finite moments and simply calculating the mean squared error from the simulations would be inappropriate. Next, in Section 5, we present graphical results for a particular simple class of models. In order to make different models from this class comparable over relevant parts of its parameter space, we develop a useful transformation of this parameter space. Section 6 concludes.

2.2 Model, estimators and standard asymptotics

We examine method of moments estimators for the single linear structural model

$$y = X\beta + \varepsilon, \tag{2.1}$$

where y and ε are $n \times 1$ vectors, X is a full column rank $n \times k$ matrix of regressors, which may contain exogenous regressors but also endogenous variables (i.e. jointly dependent with y) and lagged endogenous (i.e. weakly exogenous) variables. The $k \times 1$ vector β contains the unknown coefficients of this relationship between y and X. These are the parameters of primary interest. The relationship must be well-specified, because we assume that the disturbances are white noise (unconditionally), i.e.

$$E(\varepsilon) = 0, \mathrm{Var}(\varepsilon) = \sigma_\varepsilon^2 I_n. \tag{2.2}$$

While the functional relationship of model (2.1) is supposed to be adequately specified, we examine the consequences of misspecification of the chosen set of instrumental variables. We focus on the specific case where all regressors of the full rank matrix X are used as instruments, i.e. OLS is applied and any simultaneity is neglected.

The OLS estimator of model (2.1) is

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y. \tag{2.3}$$

Because we consider here exclusively models with stationary variables, $\hat{\beta}_{OLS}$ will be consistent and asymptotically efficient only if $E(X'\varepsilon) = 0$, and will yield an inconsistent estimator otherwise. Then, consistent estimators could be obtained by exploiting instrumental variables Z for which $E(Z'\varepsilon) = 0$. Here we will only consider as a competitor of OLS the case where Z is a full column rank $n \times k$ matrix, which yields the simple (just identified) IV estimator

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

Matrix Z should be such that $Z'X$ has rank k.

We make standard mild stationarity assumptions yielding

$$X'X = O_p(n),\ X'X = O_p(n),\ Z'X = O_p(n), \tag{2.5}$$

and we define (for $n \to \infty$)

$$\Sigma_{X'X} \equiv \text{plim}\frac{1}{n}X'X,\ \Sigma_{Z'Z} \equiv \text{plim}\frac{1}{n}Z'Z,\ \Sigma_{Z'X} \equiv \text{plim}\frac{1}{n}Z'X, \tag{2.6}$$

which all are supposed to have full rank. This yields standard results on the limiting distributions of the estimators, provided that the instruments actually used are valid, i.e.

$$n^{1/2}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N(0, \sigma_\varepsilon^2 \Sigma_{Z'X}^{-1} \Sigma_{Z'Z} \Sigma_{X'Z}^{-1}), \qquad \text{if } E(Z'\varepsilon) = 0, \tag{2.7}$$

and

$$n^{1/2}(\hat{\beta}_{OLS} - \beta) \xrightarrow{d} N(0, \sigma_\varepsilon^2 \Sigma_{X'X}^{-1}), \qquad \text{if } E(X'\varepsilon) = 0. \tag{2.8}$$

However, when $E(X'\varepsilon) \neq 0$, OLS is inconsistent and its limiting distribution will be different from (2.8).

Below, we restrict ourselves to cases where $E(Z'\varepsilon) = 0$ whereas $E(X'\varepsilon)$ may be non-zero, i.e. the instruments Z are valid and some of the regressors in X may be contemporaneously correlated with the disturbance term. Although we will examine cases where some instruments may be weak (then the columns of $Z'X$ are almost linearly dependent), in this study we will not consider alternative asymptotic sequences, as in (approaches referred to in) Staiger and Stock (1997). We first want to obtain under standard regularity conditions the counterpart of (2.8) when OLS is inconsistent and compare it with (2.7) and with actual behavior of the estimators in finite samples. No doubt these regularity conditions and the specification of our data generating scheme can be relaxed in various ways, as is done in for instance Gallant and White (1988). However, the present strict framework easily yields, after some further specialization of the regularity assumptions, an explicit and calculable characterization of the limiting behavior of inconsistent OLS.

2.3 The asymptotic distribution of inconsistent OLS

We allow for linear contemporaneous dependence of the observations x_i on the disturbances ε_i. We use the $k \times 1$ parameter vector ξ to expresses this dependence, such that matrix X can be decomposed as

$$X = \bar{X} + \varepsilon\xi', \tag{2.9}$$

with

$$E(\bar{X}'\varepsilon) = 0 \text{ and } E(X'\varepsilon) = n\sigma_\varepsilon^2\xi. \tag{2.10}$$

Note that this does not exclude cases where X contains lagged endogenous variables. These could be a part of the component $\bar{X}$ and have a corresponding element in ξ equal to zero. Only current endogenous regressors will have corresponding elements of ξ different from zero.

Decomposition (2.9) with properties (2.10) implies

$$\Sigma_{X'X} = \text{plim}\frac{1}{n}(\bar{X}'\bar{X} + \bar{X}'\varepsilon\xi' + \xi\varepsilon'\bar{X} + \xi\varepsilon'\varepsilon\xi') = \text{plim}\frac{1}{n}\bar{X}'\bar{X} + \sigma_\varepsilon^2\xi\xi'.$$

We define $\Sigma_{\bar{X}'\bar{X}} \equiv \text{plim}\frac{1}{n}\bar{X}'\bar{X}$ and find

$$\Sigma_{\bar{X}'\bar{X}} = \Sigma_{X'X} - \sigma_\varepsilon^2\xi\xi'. \tag{2.11}$$

The probability limit of $\hat{\beta}_{OLS}$ will be denoted as β^*_{OLS}, for which we obtain

$$\beta^*_{OLS} \equiv \text{plim}\hat{\beta}_{OLS} = \beta + \Sigma^{-1}_{X'X}\text{plim}\frac{1}{n}X'\varepsilon = \beta + \sigma_\varepsilon^2\Sigma^{-1}_{X'X}\xi. \tag{2.12}$$

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

$$\ddot{\beta}_{OLS} \equiv \beta^*_{OLS} - \beta = \sigma_\varepsilon^2\Sigma^{-1}_{X'X}\xi, \tag{2.13}$$

which is the inconsistency of the OLS estimator.

For obtaining a characterization of the unconditional limiting distribution of inconsistent OLS, we will assume that the data are in fact IID (independently and identically

distributed). Hence, for obtaining a result on the unconditional limiting distribution of inconsistent OLS we have to exclude occurrence of lagged dependent variables. More in particular, for the transpose of the i-th row ($i = 1, ..., n$) of X, the $k \times 1$ vector x_i, we will assume that we have

$$x_i \sim IID(0, \Sigma_{X'X}), \tag{2.14}$$

where the zero expectation is easily obtained by removing the intercept from the model (if present) by taking the covariance stationary y_i and x_i observations in deviation from their expectation. Then the remaining coefficients are the slopes of the original regression with nonzero but constant unconditional expectation of the regressors. Clearly, the IID assumption excludes most time-series applications. Below, for finding the conditional limiting distribution of inconsistent OLS, the IID assumption is not required, but for obtaining the unconditional limiting distribution it simplifies the derivations considerably.

Like Goldberger (1964, p. 359) we rewrite the model as

$$y = X(\beta^*_{OLS} - \ddot{\beta}_{OLS}) + \varepsilon = X\beta^*_{OLS} + u, \tag{2.15}$$

where $u \equiv \varepsilon - X\ddot{\beta}_{OLS}$. Under assumption (2.14) we find that $E(u) = 0$,

$$\begin{aligned} \sigma_u^2 &\equiv E(u_i^2) = \sigma_\varepsilon^2(1 - 2\ddot{\beta}'_{OLS}\xi) + \ddot{\beta}'_{OLS}\Sigma_{X'X}\ddot{\beta}_{OLS} \\ &= \sigma_\varepsilon^2(1 - \sigma_\varepsilon^2\xi'\Sigma^{-1}_{X'X}\xi), \end{aligned} \tag{2.16}$$

and $E(u_iu_j) = 0$ for $i \neq j$. Moreover, $E(x_iu_i) = E(x_i\varepsilon_i) - E(x_ix'_i)\ddot{\beta}_{OLS} = \sigma_\varepsilon^2\xi - \Sigma_{X'X}\ddot{\beta}_{OLS} = 0$, thus $E(X'u) = 0$. Hence, in the alternative model specification (2.15) OLS will be consistent and the disturbances have a scalar covariance matrix. Therefore, applying OLS to this model yields the limiting distribution

$$n^{1/2}(\hat{\beta}_{OLS} - \beta^*_{OLS}) \xrightarrow{d} N(0, \sigma_\varepsilon^2(1 - \sigma_\varepsilon^2\xi'\Sigma^{-1}_{X'X}\xi)\Sigma^{-1}_{X'X}). \tag{2.17}$$

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

$$\text{plim}\frac{1}{n}\hat{u}'\hat{u} = \text{plim}\frac{1}{n}(\varepsilon - X\ddot{\beta}_{OLS})'(\varepsilon - X\ddot{\beta}_{OLS}) = \sigma_u^2. \tag{2.18}$$

Thus, standard OLS inference in the regression of y on X makes sense and is in fact asymptotically valid when the data are IID, but it concerns unconditional (because it has been built on the stochastic properties of X) inference on the pseudo true value $\beta^*_{OLS} = \beta + \sigma^2_\varepsilon \Sigma^{-1}_{X'X}\xi$, and not on β, unless $\xi = 0$.

Next we shall refine result (2.17) by focussing on the limiting distribution of $\hat{\beta}_{OLS}$ conditional on the predetermined variables $\bar{X}$ (which in practice have not all been observed, usually, because they are conflated with unknown reduced form parameters and reduced form disturbances, but that turns out not to matter), no longer restricting ourselves to (2.14), hence serially correlated regressors and lagged dependent explanatory variables are again allowed. As suggested in Rothenberg (1972), we do not center now at β^*_{OLS}, but at

$$\beta^*_{n,OLS} \equiv \beta + \ddot{\beta}_{n,OLS} = \beta + \sigma^2_\varepsilon(\frac{1}{n}\bar{X}'\bar{X} + \sigma^2_\varepsilon\xi\xi')^{-1}\xi, \tag{2.19}$$

where $\text{plim}\beta^*_{n,OLS} = \beta^*_{OLS}$. Conditioning on $\bar{X}$ (and specializing (2.2) to $Var(\varepsilon|\bar{X}) = \sigma^2_\varepsilon I_n$), we will derive that $n^{1/2}(\hat{\beta}_{OLS} - \beta^*_{n,OLS}) \xrightarrow{d} N(0, V)$, and we establish the variance matrix V of this zero mean limiting distribution. From this result we find $\beta^*_{n,OLS}$ to be a first-order asymptotic approximation to the expectation of $\hat{\beta}_{OLS}$ in finite samples, whereas V_n/n for its variance, provided $\text{plim}V_n = V$. First-order approximations to the quantiles of $\hat{\beta}_{OLS}$ can straightforwardly be obtained from the corresponding normal distribution.

Remark 2.1 *Conditioning on $\bar{X}$ will always be unfeasible, because we cannot observe it. But, it is 'the maximal set of information' on which we can possibly condition, and hence will yield the highest precision, which we will express without using $\bar{X}$ and which we can assess only by making assumptions on ξ (or on $\rho_{X\varepsilon} = \Sigma^{-1/2}_{X'X}\sigma_\varepsilon\xi$ where $\Sigma_{X'X}$ and σ^2_ε can be assessed consistently, as will be show in Section 6 of Chapter 5).*

So, we set out to examine the limiting behavior of

$$\begin{aligned} n^{1/2}(\hat{\beta}_{OLS} - \beta^*_{n,OLS}) &= n^{1/2}[(\frac{1}{n}X'X)^{-1}\frac{1}{n}X'\varepsilon - \ddot{\beta}_{n,OLS}] \\ &= (\frac{1}{n}X'X)^{-1}[\frac{1}{\sqrt{n}}X'\varepsilon - n^{1/2}(\frac{1}{n}X'X)\ddot{\beta}_{n,OLS}]. \end{aligned} \tag{2.20}$$

For the terms between square brackets we find

$$\begin{aligned}
& \frac{1}{\sqrt{n}}X'\varepsilon - n^{1/2}(\frac{1}{n}X'X)\ddot{\beta}_{n,OLS} \\
= & \frac{1}{\sqrt{n}}[\bar{X}'\varepsilon + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi] + n^{1/2}\sigma_\varepsilon^2\xi - n^{1/2}(\frac{1}{n}X'X)\ddot{\beta}_{n,OLS} \\
= & \frac{1}{\sqrt{n}}[\bar{X}'\varepsilon + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi] - n^{1/2}[\frac{1}{n}X'X - (\frac{1}{n}\bar{X}'\bar{X} + \sigma_\varepsilon^2\xi\xi')]\ddot{\beta}_{n,OLS} \\
= & \frac{1}{\sqrt{n}}[\bar{X}'\varepsilon + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi] - \frac{1}{\sqrt{n}}[\bar{X}'\varepsilon\xi' + \xi\varepsilon'\bar{X} + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi\xi']\ddot{\beta}_{n,OLS} \\
= & \frac{1}{\sqrt{n}}[(1 - \xi'\ddot{\beta}_{n,OLS})I_k - \xi\ddot{\beta}'_{n,OLS}]\bar{X}'\varepsilon + \frac{1}{\sqrt{n}}(1 - \xi'\ddot{\beta}_{n,OLS})\xi(\varepsilon'\varepsilon - n\sigma_\varepsilon^2) \\
= & \frac{1}{\sqrt{n}}[A'_n\varepsilon + a_n(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)], \qquad (2.21)
\end{aligned}$$

where A_n is an $n \times k$ matrix and a_n a $k \times 1$ vector, viz.

$$\begin{aligned}
A'_n &\equiv [(1 - \xi'\ddot{\beta}_{n,OLS})I_k - \xi\ddot{\beta}'_{n,OLS}]\bar{X}', \\
a_n &\equiv (1 - \xi'\ddot{\beta}_{n,OLS})\xi.
\end{aligned} \qquad (2.22)$$

Denoting the ith row of A_n as $A'_{n,i}$ we can now write (2.21) as a scaled sample average of n mutually uncorrelated zero mean random vectors $A_{n,i}\varepsilon_i + a_n(\varepsilon_i^2 - \sigma_\varepsilon^2)$ and apply (while conditioning on $\bar{X}$) the standard CLT, giving

$$n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left[A_{n,i}\varepsilon_i + a_n(\varepsilon_i^2 - \sigma_\varepsilon^2)\right]\right) \xrightarrow{d} N\left(0, \lim\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}\left[A_{n,i}\varepsilon_i + a_n(\varepsilon_i^2 - \sigma_\varepsilon^2)\right]\right). \qquad (2.23)$$

Since $\mathrm{Var}[A_{n,i}\varepsilon_i + a_n(\varepsilon_i^2 - \sigma_\varepsilon^2) \mid \bar{x}_i] = \sigma_\varepsilon^2 A_{n,i}A'_{n,i} + \sigma_\varepsilon^3\mu_3(A_{n,i}a'_n + a_nA'_{n,i}) + \sigma_\varepsilon^4(\mu_4 - 1)a_na'_n$, where $\mu_3 \equiv E(\varepsilon_i^3/\sigma_\varepsilon^3)$ and $\mu_4 \equiv E(\varepsilon_i^4/\sigma_\varepsilon^4)$, we find that $n^{1/2}(\hat{\beta}_{OLS} - \beta^*_{n,OLS})$ has a limiting distribution conditional on $\bar{X}$ given by

$$N\left(0, \sigma_\varepsilon^2\Sigma_{X'X}^{-1}\left[\lim\frac{1}{n}\left[A'_nA_n - \sigma_\varepsilon\mu_3(A'_n\iota a'_n + a_n\iota' A_n)\right] + \sigma_\varepsilon^2(\mu_4 - 1)a_na'_n\right]\Sigma_{X'X}^{-1}\right). \qquad (2.24)$$

For the special case with normal disturbances, and exploiting (2.11), the conditional

asymptotic variance specializes to

$$\begin{aligned} & \sigma_\varepsilon^2 \Sigma_{X'X}^{-1}[(1-\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1}\xi)I_k - \sigma_\varepsilon^2 \xi\xi' \Sigma_{X'X}^{-1}]\Sigma_{\bar{X}'\bar{X}} \times \\ & [(1-\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1}\xi)I_k - \sigma_\varepsilon^2 \Sigma_{X'X}^{-1}\xi\xi']\Sigma_{X'X}^{-1} + 2\sigma_\varepsilon^4(1-\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1}\xi)^2 \Sigma_{X'X}^{-1}\xi\xi'\Sigma_{X'X}^{-1} \\ = \; & (1-\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1}\xi)[(1-\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1}\xi)\sigma_\varepsilon^2 \Sigma_{X'X}^{-1} - (1-2\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1}\xi)\sigma_\varepsilon^4 \Sigma_{X'X}^{-1}\xi\xi'\Sigma_{X'X}^{-1}]. \end{aligned} \tag{2.25}$$

Note that when $\xi = 0$, i.e. when OLS is consistent and efficient, the above formula yields $\sigma_\varepsilon^2 \Sigma_{X'X}^{-1}$ for the asymptotic variance, as it should. Also note that $\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1}\xi$ constitutes the population R^2 of the auxiliary regression of ε on X; denoting the OLS estimator of this regression as $\hat{\theta} = (X'X)^{-1}X'\varepsilon$, we find

$$R_{\varepsilon,X}^2 \equiv \text{plim}\frac{\hat{\theta}'X'X\hat{\theta}}{\varepsilon'\varepsilon} = \text{plim}\frac{\varepsilon' X(X'X)^{-1}X'\varepsilon}{\varepsilon'\varepsilon} = \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1}\xi, \tag{2.26}$$

which expresses the seriousness of the simultaneity. Substituting (2.26) and (2.13), result (2.25) implies

$$\mathsf{AVar}_C^N(\hat{\beta}_{OLS}) = \frac{1}{n}(1-R_{\varepsilon,X}^2)[(1-R_{\varepsilon,X}^2)\sigma_\varepsilon^2 \Sigma_{X'X}^{-1} - (1-2R_{\varepsilon,X}^2)\ddot{\beta}_{OLS}\ddot{\beta}_{OLS}'], \tag{2.27}$$

where the superscript N indicates that we assumed that the first four moments of the disturbances conform to normal and the subscript C indicates that this concerns the conditional distribution. Of course, $0 < 1 - R_{\varepsilon,X}^2 \leq 1$. Because $\ddot{\beta}_{OLS}\ddot{\beta}_{OLS}'$ is positive semi-definite, we find that as a rule, and certainly when $R_{\varepsilon,X}^2 < 0.5$, simultaneity has a mitigating effect on the asymptotic variance of the OLS estimator. This is plausible because by the pseudo true value also part of the disturbances is explained, and hence the effective signal-to-noise ratio becomes larger under simultaneity.

For the case with symmetric disturbances ($\mu_3 = 0$) and excess kurtosis ($\mu_4 \neq 3$) the asymptotic variance (2.27) changes to

$$\frac{1}{n}(1-R_{\varepsilon,X}^2)\{(1-R_{\varepsilon,X}^2)\sigma_\varepsilon^2 \Sigma_{X'X}^{-1} - [(4-\mu_4)-(5-\mu_4)R_{\varepsilon,X}^2]\ddot{\beta}_{OLS}\ddot{\beta}_{OLS}'\}. \tag{2.28}$$

Assuming that the first column of X equals ι so that $\Sigma_{X'X}^{-1}\Sigma_{X'\iota} = e_1 = (1,0,...,0)'$ is

a unit vector whereas $\xi' e_1 = 0$, then in case of skewness, the extra contribution to the variance of the limiting distribution is

$$\frac{1}{n}\sigma_\varepsilon^3 \mu_3 (1 - R^2_{\varepsilon,X})^2 [e_1 \xi' \Sigma^{-1}_{X'X} + \Sigma^{-1}_{X'X} \xi e_1']. \tag{2.29}$$

Note that – in agreement with established knowledge – the contributions due to $\mu_3 \neq 0$ or $\mu_4 \neq 3$ are nil when $\xi = 0$.

Returning now to the unconditional limiting distribution given in (2.17) and using (2.26) we find

$$\mathsf{AVar}_U(\hat{\beta}_{OLS}) = \frac{1}{n}(1 - R^2_{\varepsilon,X})\sigma_\varepsilon^2 \Sigma^{-1}_{X'X}, \tag{2.30}$$

which holds irrespective of the distribution of the disturbances. By its very nature it should be larger in a matrix sense than its conditional counterpart.

An expression that can be shown to be similar to (2.27) can be found in Rothenberg (1972). However, his formula (4.7), which is employed in Hausman (1978) and Hahn and Hausman (2003), is more difficult to interpret. It has been obtained from a particular errors in variables model specification in which no allowance has been made for lagged dependent regressor variables, whereas ours stems from a much more general (possibly dynamic) regression specification, which is generic concerning the problem of contemporaneous correlation of regressors and disturbances. By the decomposition (2.9) we avoided an explicit specification of the variance matrix of the disturbances in a reduced form for X, as employed by Rothenberg (1972), and then from (2.25) it is easy to recognize that, apart from $\sigma_\varepsilon^2 \Sigma^{-1}_{X'X}$, the only determining factors of the asymptotic variance are the very meaningful characteristics: (i) the inconsistency $\ddot{\beta}_{OLS} = \beta^*_{OLS} - \beta = \sigma_\varepsilon^2 \Sigma^{-1}_{X'X} \xi$ and (ii) a measure for the simultaneity $R^2_{\varepsilon,X} = \xi' \ddot{\beta}_{OLS}$. The derivations in Joseph and Kiviet (2005) yielded the expression $\frac{1}{n}[\sigma_\varepsilon^2 \Sigma^{-1}_{X'X} + \ddot{\beta}_{OLS} \ddot{\beta}'_{OLS}]$ for $\mathsf{AVar}^N_C(\hat{\beta}_{OLS})$. It can be shown that the difference between this incorrect and the complete formula given above is positive semi-definite. Hence, the area in the parameter space where OLS beats IV on the basis of their limiting distribution is actually even larger than indicated in that earlier study. Kiviet and Niemczyk (2007) presented the complete expression for $\mathsf{AVar}_C(\hat{\beta}_{OLS})$, but it

was inadequately indicated[1] that it concerned the distribution conditional on $\bar{X}$.

2.4 Measures for estimator accuracy

We want to use characteristics of the limiting distributions of OLS and IV estimators in order to express the essentials of their location and spread, so that we can make useful comparisons, which hopefully will also prove to approximate their relative qualities in finite samples reasonably well. Apart from using first-order asymptotic theory to approximate these finite sample characteristics, in addition we shall use simulation to assess them. The asymptotic distributions of OLS and IV in the models to be considered are all normal and have finite moments.

Let for the generic estimator $\hat{\beta}$ of β, with pseudo true value β^*, the limiting distribution be given by

$$n^{1/2}(\hat{\beta} - \beta^*) \xrightarrow{d} N(0, V). \tag{2.31}$$

Under a complete specification of the data generating processes for both y and the variables occurring in X and Z, matrices like $\Sigma_{X'X}$ and $\Sigma_{Z'X}$ and vector ξ are determined just by the model parameters. Then all elements of both β^* and V depend on the parameters only. The first order asymptotic approximation to the variance of $\hat{\beta}$ is given by

$$\mathsf{AVar}(\hat{\beta}) \equiv \frac{1}{n}V, \tag{2.32}$$

and to its bias by $\beta^* - \beta$. Hence, the first-order asymptotic approximation to the MSE (mean squared error) can be defined as

$$\mathsf{AMSE}(\hat{\beta}) \equiv \frac{1}{n}V + (\beta^* - \beta)(\beta^* - \beta)', \tag{2.33}$$

which for a consistent estimator simplifies to $\frac{1}{n}V$.

The simple IV estimators $\hat{\beta}_{IV}$ considered in this study do not have finite moments in

[1] We thank Peter Boswijk for bringing this forward.

finite samples and hence their bias $E(\hat{\beta}-\beta)$, their variance $\mathrm{Var}(\hat{\beta})$, and their MSE, i.e.

$$\mathsf{MSE}(\hat{\beta}) \equiv E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' = \mathrm{Var}(\hat{\beta}) + E(\hat{\beta}-\beta)E(\hat{\beta}-\beta)', \tag{2.34}$$

do not exist. This makes the usual measures of the actual distribution of $\hat{\beta}$, calculated on the basis of Monte Carlo sample moments, unsuitable. Denoting the series of mutually independent simulated realizations of the estimator by $\hat{\beta}^{(1)}, ..., \hat{\beta}^{(R)}$, where R is the number of replications, the habitual Monte Carlo estimator of $E(\hat{\beta})$ is the Monte Carlo sample average

$$\mathsf{ME}(\hat{\beta}) \equiv R^{-1}\sum\nolimits_{r=1}^{R}\hat{\beta}^{(r)}. \tag{2.35}$$

However, $\mathsf{ME}(\hat{\beta})$ will not converge for $R \to \infty$ if $E(\hat{\beta})$ does not exist. Self-evidently, similar problems arise for the Monte Carlo assessment of the variance, i.e.

$$\mathsf{MVar}(\hat{\beta}) \equiv \frac{1}{R-1}\sum\nolimits_{r=1}^{R}(\hat{\beta}^{(r)} - \mathsf{ME}(\hat{\beta}))(\hat{\beta}^{(r)} - \mathsf{ME}(\hat{\beta}))', \tag{2.36}$$

and for the empirical (Monte Carlo) MSE, i.e.

$$\mathsf{MMSE}(\hat{\beta}) \equiv \frac{1}{R}\sum\nolimits_{r=1}^{R}(\hat{\beta}^{(r)} - \beta)(\hat{\beta}^{(r)} - \beta)', \tag{2.37}$$

if the corresponding moments do not exist. Therefore, to find expressions for estimator quality obtained from Monte Carlo results such that they will always summarize location and spread in a meaningful way, we will choose measures here which are based directly on characteristics of the empirical Monte Carlo density or the empirical distribution function $\hat{F}_i$ of the ith element of the vector $\hat{\beta}$, such as the median and other quantiles.

For any real argument value x the empirical distribution function of $\hat{\beta}_i$, obtained from the Monte Carlo experiments, is defined as

$$\hat{F}_i(x) \equiv \frac{1}{R}\sum\nolimits_{r=1}^{R} I(\hat{\beta}_i^{(r)} \leq x), \tag{2.38}$$

where $I(\cdot)$ is the Kronecker indicator function. Then the empirical median or second quartile is $\hat{F}_i^{-1}(0.5)$, and the first and third empirical quartiles are $\hat{F}_i^{-1}(0.25)$ and $\hat{F}_i^{-1}(0.75)$,

respectively. These q^{th} quartiles can easily be obtained after sorting the $\hat{\beta}_i^{(r)}$ in non-decreasing order and then taking (assuming R is a multiple of 100)

$$\hat{F}_i^{-1}(q/4) = 0.5(\hat{\beta}_i^{(qR/4)} + \hat{\beta}_i^{(1+qR/4)}),\ q = 1, 2, 3. \tag{2.39}$$

To mimic the RMSE (root mean squared error) criterion, which is $\sqrt{\sigma_i^2 + b_i^2}$, when σ_i and b_i are the standard deviation and the bias of $\hat{\beta}_i$ respectively, a similar alternative empirical measure, not requiring existence of finite moments, seems the following. We replace σ_i by $q^*_{0.75}[\hat{F}_i^{-1}(0.75) - \hat{F}_i^{-1}(0.25)]/2$, for some real number $q^*_{0.75}$, and b_i by $\hat{F}_i^{-1}(0.5) - \beta_i$. We can choose $q^*_{0.75}$ such that in case an estimator is in fact normally distributed the criterion conforms precisely to RMSE. Indicating the standard normal distribution function by Φ this requires $q^*_{0.75}[\Phi^{-1}(0.75) - \Phi^{-1}(0.25)]/2 = 1$, which results in $q^*_{0.75} = (0.67499)^{-1} = 1.4815$. As an alternative to the RMSE we could then use

$$\sqrt{(q^*_{0.75})^2[\hat{F}_i^{-1}(0.75) - \hat{F}_i^{-1}(0.25)]^2/4 + [\hat{F}_i^{-1}(0.5) - \beta_i]^2}.$$

However, we do not necessarily have to use the quartiles. More generally, for any $0.5 < p < 1$, we may define

$$d(p) \equiv [\Phi^{-1}(p) - \Phi^{-1}(1-p)]/2.$$

Let $\Phi_{\mu,\sigma}$ be the distribution function of $N(\mu, \sigma^2)$, then

$$\Phi^{-1}_{\mu,\sigma}(p) - \Phi^{-1}_{\mu,\sigma}(1-p) = 2\sigma d(p).$$

Now as an assessment $\hat{\sigma}_i(p)$ from an empirical distribution $\hat{F}_i$ that should mimic σ_i (if this exists), we may use

$$\hat{\sigma}_i(p) \equiv \frac{1}{2d(p)}[\hat{F}_i^{-1}(p) - \hat{F}_i^{-1}(1-p)]. \tag{2.40}$$

This will work perfectly well for any $0.5 < p < 1$ if $\hat{F}_i$ is in fact normal. We have experimented with a few values of p, trying Chi-squared (skewed) and Student (fat tailed) distributions, and found especially $p = 0.841345$, for which $d(p) = 1$, to work well.

Therefore, when finite moments do not exist, instead of RMSE, we will use what we call the "empirical quantile error distance", which we define as

$$\mathsf{EQED}(\hat{\beta}_i) \equiv \sqrt{[\hat{F}_i^{-1}(0.841345) - \hat{F}_i^{-1}(1-0.841345)]^2/4 + [\hat{F}_i^{-1}(0.5) - \beta_i]^2}. \tag{2.41}$$

See also Pearson and Tukey (1965)[2] on approximating means by distances between quantiles, where the authors also consider the approximation to the standard deviation (2.40) for p=0.95, 0.975 and 0.99 (Pearson and Tukey's equations (6)-(8)).

Below, we will calculate this for alternative estimators for the same model (and same parameter values and sample size), including the consistent and asymptotically optimal estimator, and then depict the logarithm of the ratio (with the asymptotically optimal in the denominator), so that positive and negative values directly indicate which estimator has more favorable EQED criterion for particular parameter values. Having smaller EQED will be interpreted as being more accurate in finite samples. Hence, negative values for the log of the ratio will indicate that the asymptotically optimal is actually less accurate in finite samples.

To examine the accuracy in finite samples of the precision criteria obtained from the limiting distribution we can calculate the log ratio of $\mathsf{EQED}(\hat{\beta}_i)$ and the asymptotic root mean squared error

$$\mathsf{ARMSE}(\hat{\beta}_i) \equiv \sqrt{\frac{1}{n}V_{ii} + (\beta_i^* - \beta_i)^2}. \tag{2.42}$$

For an estimator with finite moments we can simply take the log ratio of the Monte Carlo root mean squared error

$$\mathsf{MRMSE}(\hat{\beta}_i) \equiv \sqrt{\frac{1}{R}\sum\nolimits_{r=1}^{R}(\hat{\beta}_i^{(r)} - \beta_i)^2} \tag{2.43}$$

and $\mathsf{ARMSE}(\hat{\beta}_i)$.

Note that for an inconsistent estimator, where $\beta_{OLS,i}^* \neq \beta_i$, the ARMSE criterion will converge for $n \to \infty$ to $\mid \beta_{OLS,i}^* - \beta_i \mid \neq 0$, whereas it will converge to zero for any consistent estimator. Hence the criterion follows the logic that, since estimator variance

[2] We thank Prof. J.G. de Gooijer for this reference.

gets smaller in larger samples irrespective of whether the estimator is consistent, the larger the sample size the more pressing it becomes to have a consistent estimator. On the other hand, when sample size is moderate, an inconsistent estimator with possibly a substantial bias in finite samples but a relatively small variance could well be more attractive than a consistent estimator, especially when the latter's distribution has fat tails, and is not median unbiased with possibly a wide spread. In the models to be defined below, we will first examine the log ratios of the ARMSE criterion for OLS and IV, with IV in the denominator, so that positive values of this ratio indicate parameter values for which IV is more accurate on the basis of first-order asymptotic theory. Next we will examine whether the findings from first-order asymptotic theory are vindicated in finite samples by simulation experiments.

2.5 Pictured parametrizations

In this section we specify a class of simple specific models that easily allow to parameterize the asymptotic characteristics of both OLS and IV. Models from this class will be simulated too in order to assess the actual behavior in finite samples and to examine the accuracy of the asymptotic approximations. We restricted our study to cases where disturbances are normally distributed. In all simulations we use the same set of random draws for the various disturbance vectors for all grid-points in the parameter space examined. To further reduce the experimental variance, exploiting the assumed symmetry of the disturbances, we also made use of the simple variance reduction method of re-using vectors of normal random numbers by simply changing their sign. The number of Monte Carlo replications for each parameter combination is 1,000,000 for densities and 50,000 for all grid points in the 3-D pictures. The diagrams presented below are single images from animated versions (available at www.feb.uva.nl/ke/jfk.htm) which allow to inspect the relevant phenomena over a much larger part of the parameter space.

For the simple static models that we examine below some analytic finite sample properties are available; see Woglom (2001) and Hillier (2006) for some recent contributions and further references. We have not made use of these and employed straightforward

Monte Carlo simulation, which as yet seems the only option for assessing finite sample properties for most of the phenomena examined here.

2.5.1 A basic static IID model

We consider a model with one regressor and one valid and either strong or weak instrument. The two variables x and z, together with the dependent variable y, are jointly IID with zero mean and finite second moments. This case may be denoted as

$$y_i = \beta x_i + \varepsilon_i, \tag{2.44}$$

$$x_i = \bar{x}_i + \xi \varepsilon_i, \tag{2.45}$$

where ξ is scalar now. Data for y, x and z can be obtained by the generating scheme

$$\begin{aligned} \varepsilon_i &= \sigma_\varepsilon v_{1,i}, \\ \bar{x}_i &= \alpha_1 v_{2,i}, \\ z_i &= \alpha_2 v_{2,i} + \alpha_3 v_{3,i}, \end{aligned}$$

where $v_i = (v_{1,i}, v_{2,i}, v_{3,i})' \sim \text{IID}(0, I_3)$. Thus

$$\begin{pmatrix} \varepsilon_i \\ x_i \\ z_i \end{pmatrix} = P v_i = \begin{pmatrix} \sigma_\varepsilon & 0 & 0 \\ \sigma_\varepsilon \xi & \alpha_1 & 0 \\ 0 & \alpha_2 & \alpha_3 \end{pmatrix} v_i, \tag{2.46}$$

giving $(\varepsilon_i, x_i, z_i)' \sim \text{IID}(0, PP')$.

We will focus on this model just for the case $\beta = 1$. This is merely a normalization and not a restriction, because we can imagine that we started from a model $y_i = \mathring{\beta} \mathring{x}_i + \varepsilon_i$, with $\mathring{\beta} \neq 0$, and rescaled the explanatory variable such that $x_i = \mathring{x}_i / \mathring{\beta}$. We can impose some further normalizations on the 5 parameters of P, because, without loss of generality,

we may take

$$\sigma_\varepsilon = 1, \tag{2.47}$$

$$\sigma_z^2 = \alpha_2^2 + \alpha_3^2 = 1. \tag{2.48}$$

By (2.47) we normalize all results with respect to σ_ε, and because the IV estimator is invariant to the scale of the instruments (only the space spanned by z is relevant) we may impose (2.48) which will be used to obtain the value

$$\alpha_3^2 = 1 - \alpha_2^2 \geq 0. \tag{2.49}$$

From the above we find the following data variances, covariances and related correlations:

$$\left.\begin{array}{ll} \sigma_x^2 = \xi^2 + \alpha_1^2 & \sigma_y^2 = \xi^2 + 2\xi + 1 + \alpha_1^2 \\ \sigma_{x\varepsilon} = \xi & \rho_{x\varepsilon} = \xi/\sqrt{\xi^2 + \alpha_1^2} \\ \sigma_{z\varepsilon} = 0 & \rho_{z\varepsilon} = 0 \\ \sigma_{xz} = \alpha_1\alpha_2 & \rho_{xz} = \alpha_1\alpha_2/\sqrt{\xi^2 + \alpha_1^2} \end{array}\right\} \tag{2.50}$$

Note that these depend on only 3 remaining free parameters: viz. ξ, α_1 and α_2, and so will the expressions for asymptotic variance (together with μ_3 and μ_4, the 3rd and 4th moments of $v_{1,i}$).

However, instead of designing our results in terms of the three parameters ξ, α_1 and α_2, we prefer another parametrization. We shall use as a base of the design parameter space for this simple model, the three parameters: $\rho_{x\varepsilon}$, ρ_{xz} and SN (signal-noise ratio), where

$$SN = \beta^2\sigma_x^2/\sigma_\varepsilon^2 = \sigma_x^2 \geq 0. \tag{2.51}$$

This reparametrization is useful because the parameters $\rho_{x\varepsilon}$, ρ_{xz} and SN have a direct econometric interpretation, viz. the degree of simultaneity, instrument strength and model fit, respectively. The population fit of the model might be expressed as $PF = \beta^2\sigma_x^2/(\beta^2\sigma_x^2 + \sigma_\varepsilon^2) = SN/(SN+1)$. By varying the three parameters $|\rho_{x\varepsilon}| < 1$, $|\rho_{xz}| < 1$ and $0 < PF < 1$, we can examine the whole parameter space of this model. For

given values of $SN = PF/(1-PF) = \sigma_x^2$ and $\rho_{x\varepsilon}$ one can obtain ξ and α_1, i.e.

$$\xi = \rho_{x\varepsilon}\sigma_x, \tag{2.52}$$

$$\alpha_1 = \left|\sqrt{\sigma_x^2 - \xi^2}\right| = \left|\sigma_x\sqrt{1-\rho_{x\varepsilon}^2}\right|. \tag{2.53}$$

With ρ_{xz} we can now obtain

$$\alpha_2 = \rho_{xz}/\left|\sqrt{1-\rho_{x\varepsilon}^2}\right| \tag{2.54}$$

and, of course,

$$\alpha_3 = \left|\sqrt{1-\alpha_2^2}\right| = \left|\sqrt{(1-\rho_{x\varepsilon}^2-\rho_{xz}^2)/(1-\rho_{x\varepsilon}^2)}\right|, \tag{2.55}$$

so that $\rho_{x\varepsilon}^2 + \rho_{xz}^2 < 1$.

In this simple model we have

$$\left.\begin{array}{l} \beta_{OLS}^* - \beta = \frac{\xi}{\sigma_x^2} = \rho_{x\varepsilon}/\sigma_x \\ R_{\varepsilon,X}^2 = \frac{\xi^2}{\sigma_x^2} = \rho_{x\varepsilon}^2 \\ \Sigma_{X'X} = \sigma_x^2 = SN \\ \Sigma_{Z'X}^{-1}\Sigma_{Z'Z}\Sigma_{X'Z}^{-1} = \sigma_z^2/\sigma_{xz}^2 = 1/\rho_{xz}^2\sigma_x^2 = \rho_{xz}^2/\sigma_x^2. \end{array}\right\} \tag{2.56}$$

In the simulations of the finite sample distributions and the evaluations of the first-order asymptotic approximations, we want to distinguish between the unconditional and the conditional cases. When conditioning on $\bar{X}$ all Monte Carlo replications should use the same drawing, i.e. just one single realization of the series $v_{2,i}$. However, an arbitrary draw of $v_{2,i}$ might give rise to an atypical $\bar{x}$ series, and when one would condition the distribution of $\hat{\beta}_{IV}$ on the exogenous Z as well, the actual strength of the instrument would not be fully under control because the sample correlation between $v_{2,i}$ and $v_{3,i}$ would not be precisely zero. Therefore, when conditioning, we replaced $v_{3,i}$ by its residuals after regressing on $v_{2,i}$ and an intercept, in order to guarantee a sample correlation of zero. And to make sure that sample mean and variance of both $v_{2,i}$ and $v_{3,i}$ are appropriate we standardized them too. An effect of this is that in the simulations $\bar{x}'\bar{x}/n + \sigma_\varepsilon^2\xi^2 = \sigma_x^2$ and thus $\beta_{n,OLS}^* = \beta_{OLS}^*$, which makes it easier to notice the major consequences of conditioning. Another consequence is that the assessment in the Monte Carlo of the

first-order approximation to the variance of $\hat{\beta}_{IV}$ is the same for both the conditional and unconditional case, as is the case for OLS when $\xi = 0$.

So, for the case where all variables are (almost) normally distributed

$$\mathsf{AVar}_C^N(\hat{\beta}_{OLS}) = \frac{1}{n}(1-\rho_{x\varepsilon}^2)(1-2\rho_{x\varepsilon}^2+2\rho_{x\varepsilon}^4)/\sigma_x^2. \tag{2.57}$$

This yields

$$\frac{\partial \mathsf{AVar}_C^N(\hat{\beta}_{OLS})}{\partial \rho_{x\varepsilon}^2} = -\frac{1}{n}(3-8\rho_{x\varepsilon}^2+6\rho_{x\varepsilon}^4)/\sigma_x^2,$$

which is strictly negative, because the polynomial factor between parentheses is strictly positive. Therefore, the asymptotic variance of OLS decreases when the simultaneity aggravates, even when $R_{\varepsilon,X}^2 \geq 0.5$ (compare with the finding below (2.27)).

Result (2.57) implies for the first-order asymptotic approximation to the mean squared error under normality of the disturbances the specific result

$$\mathsf{AMSE}_C^N(\hat{\beta}_{OLS}) = [\frac{1}{n}(1-\rho_{x\varepsilon}^2)(1-2\rho_{x\varepsilon}^2+2\rho_{x\varepsilon}^4)+\rho_{x\varepsilon}^2]/\sigma_x^2, \tag{2.58}$$

from which we find $\frac{\partial}{\partial \rho_{x\varepsilon}^2}\mathsf{AMSE}_C^N(\hat{\beta}_{OLS}) > 0$ for $n > 3$. So, first order asymptotic theory predicts that in all cases of practical interest the reduction in variance due to an increase in simultaneity will be offset by the squared increased inconsistency.

We want to compare expression (2.58) with the corresponding quantity for IV

$$\mathsf{AVar}(\hat{\beta}_{IV}) = 1/(n\sigma_x^2\rho_{xz}^2), \tag{2.59}$$

which holds for both the unconditional and the conditional distribution. Note that, unlike $\mathsf{AVar}_C(\hat{\beta}_{OLS})$ and $\mathsf{AVar}_U(\hat{\beta}_{OLS})$, this is invariant with respect to $\rho_{x\varepsilon}$. According to first order asymptotic criteria, OLS will be more accurate than IV for all combinations of parameter values and n satisfying $\mathsf{AMSE}_C^N(\hat{\beta}_{OLS}) < \mathsf{AMSE}(\hat{\beta}_{IV}) = \mathsf{AVar}(\hat{\beta}_{IV})$, i.e. for

$$\rho_{xz}^2[(1-\rho_{x\varepsilon}^2)(1-2\rho_{x\varepsilon}^2+2\rho_{x\varepsilon}^4)+n\rho_{x\varepsilon}^2] < 1. \tag{2.60}$$

Note that this watershed between IV and OLS as far as AMSE is concerned is invariant

with respect to $SN = \sigma_x^2$, and so is the relative (but not the absolute) difference in AMSE. Self-evidently (2.60) shows that for $\rho_{x\varepsilon} = 0$ OLS will always be more accurate. It is also obvious that IV runs into weak instrument problems when ρ_{xz}^2 gets close to zero. When $\rho_{xz}^2 = 0$ the equation is not identified. For IV this implies an exploding variance but not for OLS, where $\mathsf{AMSE}_C^N(\hat{\beta}_{OLS})$ is not affected by ρ_{xz}^2. So, although obtaining meaningful inference on β from it may seem an illusion, $\hat{\beta}_{OLS}$ has still a well-defined distribution.

Since

$$\begin{aligned} \hat{\beta}_{OLS} - \beta &= \frac{1}{\sigma_x} \frac{\sum_{i=1}^n \left(\rho_{x\varepsilon} v_{i1} + \sqrt{1-\rho_{x\varepsilon}^2} v_{i2}\right) v_{i1}}{\sum_{i=1}^n \left(\rho_{x\varepsilon} v_{i1} + \sqrt{1-\rho_{x\varepsilon}^2} v_{i2}\right)^2}, \\ \hat{\beta}_{IV} - \beta &= \frac{1}{\sigma_x} \frac{\sum_{i=1}^n (\alpha_2 v_{i2} + \alpha_3 v_{i3}) v_{i1}}{\sum_{i=1}^n (\alpha_2 v_{i2} + \alpha_3 v_{i3}) \left(\rho_{x\varepsilon} v_{i1} + \sqrt{1-\rho_{x\varepsilon}^2} v_{i2}\right)}, \end{aligned} \tag{2.61}$$

the finite sample distributions of both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ are determined by $SN = \sigma_x^2$ in a very straightforward way. In fact, the shape of the densities is not affected, but only the scale. This is also the case for the inconsistency, see the first formula in (2.56), and thus carries over to the asymptotic variances (2.27) and (2.59) too. From (2.61) we can also see that due to the symmetry of v_i, the densities of both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ are not affected by the sign of $\rho_{x\varepsilon}$ nor by the sign of ρ_{xz}, so we will examine positive values only.

2.5.2 Actual findings

The actual values of β_{OLS}^* and of (the square root of) $\mathsf{AMSE}^N(\hat{\beta}_{OLS})$ and $\mathsf{AVar}(\hat{\beta}_{IV})$ could be calculated and tabulated now for various values of n, SN, $\rho_{x\varepsilon}$ and ρ_{xz} and then (to find out how accurate these first-order asymptotic approximations are) be compared with simulation estimates for the expectation (or median) and the standard error (or interquartile range). We have chosen, however, for a visual and more informative representation of these phenomena by focussing both on density functions and on graphs of ratios of the performance measures mentioned in section 2.4. We will portray these over the relevant parameter space. From the foregoing it is clear that varying $SN = \sigma_x^2$ will have a rather straightforward and relatively neutral effect, so we focus much more on the effects of $\rho_{x\varepsilon}$, ρ_{xz} and n.

In Figure 2.1 densities are presented, both for OLS and for IV, for the conditional

and the unconditional distribution, both for the actual empirical distribution and for its asymptotic approximation, as indicated in the legend below.

Legend for Figures 2.1 and 2.2

line type:	*density of:*
———	OLS, actual, conditional
□ □ □	OLS, asymptotic, conditional
– – – –	OLS, actual, unconditional
◊ ◊ ◊	OLS, asymptotic, unconditional
· · · · ·	IV, actual, conditional
○ ○ ○ ○	IV, actual, unconditional
+ + +	IV, asymptotic, both conditional and unconditional

For the asymptotic approximations we take

$$\begin{aligned} \hat{\beta}_{OLS} &\overset{a}{\sim} N(\beta^*_{OLS}, \tfrac{1}{n}\mathsf{AVar}^N_C(\hat{\beta}_{OLS})), \\ \hat{\beta}_{OLS} &\overset{a}{\sim} N(\beta^*_{OLS}, \tfrac{1}{n}\mathsf{AVar}_U(\hat{\beta}_{OLS})), \\ \hat{\beta}_{IV} &\overset{a}{\sim} N(\beta, \tfrac{1}{n}\mathsf{AVar}(\hat{\beta}_{IV})). \end{aligned} \tag{2.62}$$

In the simulations we took $v_i \sim \mathrm{IIN}(0, I_3)$. From the results we may expect to get quick insights into issues as the following. For which combinations of the design parameter values are the actual densities of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ close (regarding mean/median, spread, symmetry, unimodality, tail behavior) to their respective normal approximations (2.62)? Is there a qualitative difference between the accuracy of the OLS and the IV asymptotic approximations? What are the effects of conditioning? Do these densities clearly disclose where IV seems to perform better (or worse) than OLS? Hence, we focus on the correspondences and differences in shape, location and spread of the four pairs of asymptotic and empirical distributions.

Figure 2.1 consists of six panels of 2×2 diagrams each. Every panel has a fixed value of n and of ρ_{xz}; the latter is shown in the middle of each panel. The three left-hand panels are for $n = 50$ and the three right-hand panels for $n = 200$. The three rows of panels are for different values of ρ_{xz}. From top to bottom we distinguish a relatively strong instrument ($\rho_{xz} = 0.8$), a much weaker one ($\rho_{xz} = 0.2$) and a very weak instrument ($\rho_{xz} = 0.02$). Hence, in these three rows of panels the OLS results do not change, as they do not depend on ρ_{xz}. However, the scale on both the horizontal and vertical axes differs, so their appearance does differ. The four diagrams in each panel concern very mild simultaneity ($\rho_{x\varepsilon} = 0.1$), slightly stronger simultaneity ($\rho_{x\varepsilon} = 0.2$) and the bottom two diagrams show more severe simultaneity for $\rho_{x\varepsilon} = 0.4$ and $\rho_{x\varepsilon} = 0.6$ respectively. These all fulfill the requirement $\rho_{xz}^2 \leq 1 - \rho_{x\varepsilon}^2$. Note that when $\rho_{x\varepsilon} = 0.6$ and $\rho_{xz} = 0.8$ instrument z_i is a multiple of $\bar{x}_i$ ($\alpha_3 = 0$), according to (2.55), and cannot possibly be made stronger. Each panel contains the seven densities for $SN = \sigma_x^2 = 10$, implying population fit $PF = 10/11 = 0.909$, which value has just a straightforward multiplicative effect and does not affect the qualitative differences between the densities.

From Figure 2.1 we find that for a relatively strong instrument the three densities depicted for IV are extremely close to each other, even for $n = 50$, irrespective of the severity of simultaneity. Obviously, when the instrument is much weaker all three densities become much flatter, but when the instrument is really weak we note a serious discrepancy between the unconditional and conditional actual distribution, where the latter is much more erratic and shows bimodality for the smaller sample size. As has been established in the literature before, the standard asymptotic approximation is clearly inaccurate for a very weak instrument, and is in fact much too pessimistic regarding the spread of the actual distribution. This is seen more clearly in Figure 2.2, where the two panels of the third row of Figure 2.1 for the very weak instrument are depicted again, but now on a different scale, without the OLS densities. Note that conditional IV tends to be bimodal, especially for smaller sample size and more severe simultaneity. With respect to OLS Figure 2.1 shows that even for the smaller sample size the two asymptotic approximations are very accurate for their respective finite sample densities, which both are almost similar for mild simultaneity, but clearly demonstrate for more severe simultaneity the smaller

variance of the conditional distribution. The latter occurs for both sample sizes examined. It is evident that, in case of substantial simultaneity, IV can be more attractive than OLS when the instrument is relatively strong, especially for the larger sample size. However, it is much less obvious when the sample size is small and simultaneity very mild, even for a strong instrument. That OLS may on average have smaller estimation errors than IV when the instrument is weak is also clearly exposed, especially when the simultaneity is mild. Because the bias of OLS is relatively small in comparison to the increased spread of IV, this seems to be the case much more generally. For which particular parameter value combinations OLS beats IV indeed can be learned from the diagrams in Figure 2.3.

To examine more closely for which parameter values the performance measures developed in section 4 show a positive (negative) difference between the precision of OLS and IV in finite samples, we produce here 3D graphs (and 4D graphs on the web) of

$$\log[\mathsf{EQED}_C^N(\hat{\beta}_{OLS})/\mathsf{EQED}_C^N(\hat{\beta}_{IV})], \tag{2.63}$$

for fixed values of SN and n over the $(\rho_{x\varepsilon}, \rho_{xz})$ plane. This log-ratio (2.63) is positive when IV performs better (*yellow/amber surface*) and negative (*light/dark blue surface*) when OLS is more precise. The four panels in Figure 2.3 correspond to $n = 20$, 50, 100 and 200 respectively. We took again $SN = 10$, but ratio (2.63) is invariant with respect to this value, due to (2.61). These graphs have been obtained from simulating the conditional distributions of $\hat{\beta}_{IV}$ and $\hat{\beta}_{OLS}$. They illustrate that IV performs better when both $\rho_{x\varepsilon}$ and ρ_{xz} are large in absolute value, i.e. when both simultaneity is severe and the instrument relatively strong. The (blue) area where OLS performs better diminishes when n increases. Where the ratio equals 2, IV is exp(2) or about 7.5 times as accurate as OLS, whereas where the log-ratio is less than -3 OLS is more than exp(3) (i.e. about 20) times as accurate as IV. We notice that over a substantial area in the parameter space (which obeys $\rho_{x\varepsilon}^2 + \rho_{xz}^2 < 1$) the OLS efficiency gains over IV are much more impressive than its potential losses can ever be.

A measure for the weakness of an instrument is the first-stage population F value (see,

for instance, Staiger and Stock (1997)), which in this model is

$$F \equiv n\frac{\sigma_x^2 - \sigma_x^2(1-\rho_{xz}^2)}{\sigma_x^2(1-\rho_{xz}^2)} = n\frac{\rho_{xz}^2}{1-\rho_{xz}^2}. \tag{2.64}$$

Instrument weakness is associated with small values of F, say $F \leq 10$. The latter implies here $\rho_{xz}^2 \leq 10/(n+10)$ or $|\rho_{xz}| \leq 0.58$ (for $n = 20$) and $|\rho_{xz}| \leq 0.3$ (for $n = 100$). From Figure 2.3 we see that this criterion lacks the influence of $\rho_{x\varepsilon}$ in order to be useful to identify all the cases where IV performs better/worse than OLS.

Figure 2.4 examines for conditional OLS the quality of the asymptotic approximation to represent the actual empirical OLS distribution. Because OLS has finite moments, we simply use the RMSE criterion. The 3D graphs represent

$$\log[\mathsf{ARMSE}_C^N(\hat{\beta}_{OLS})/\mathsf{MRMSE}_C^N(\hat{\beta}_{OLS})], \tag{2.65}$$

hence positive values indicate pessimism of the asymptotic approximation (actual RMSE smaller than first-order asymptotic approximation) and negative values optimism. Self-evidently ρ_{xz} has no effect, neither has $SN = \sigma_x^2$, but $\rho_{x\varepsilon}$ has. We find that the asymptotic approximation of MSE developed in this study may be slightly pessimistic, but is especially accurate when the simultaneity is serious. Even in a very small sample the over assessment by the asymptotic approximation of the actual RMSE is usually below 10 .

The above model can easily be generalized, for instance by including another, possibly serially correlated or a lagged-dependent, explanatory variable for y_i, as we did in Kiviet and Niemczyk (2007), although not yet taking conditioning properly into account. This will be examined in future research. Note that when $\bar{x}_i$ is serially correlated the IID assumption does no longer hold, and the asymptotic approximation to the unconditional distribution of OLS does not apply.

2.6 Conclusions

Econometrics developed as a field separate from statistics, mainly because it focusses on the statistical analysis of observational non-experimental data, whereas standard statistics

generally analyzes data that have been obtained from appropriately designed experiments. This option is often not open in economics, where data are usually not random samples from a well-defined population. Unlike data obtained from experiments, most variables may be jointly dependent. As a consequence the structural relationships become part of a simultaneous system, and their explanatory variables may be contemporaneously correlated with the equation's disturbance term. In that situation the least-squares estimator exhibits bias, not just in finite samples. In simultaneous equations of stationary variables least-squares estimators are inconsistent. Hence, even asymptotically (in infinitely large samples) this estimator produces systematic estimation errors. For that reason its actual distribution has received relatively little attention in the literature, mainly because in an identified (partial-) simultaneous system alternative consistent method of moments estimators are available. However, in finite samples these instrumental variable estimators have systematic estimation errors too, and may even have no finite moments. The fact that they can be very inefficient (even in large samples) has been highlighted recently in the literature on weak instruments; see Dufour (2003) for an overview. In extreme cases these method of moment estimators are no longer consistent either, whereas in less extreme cases, they may still have reasonable location properties, while showing an unfavorable spread.

In this Chapter we provide further evidence on the behavior of inconsistent least-squares and consistent just identified instrumental variable estimators. This evidence enables us to monitor the trade-off options between: (i) the systematic but generally bounded dislocation of the least-squares estimator, and (ii) the vulnerability of the instrumental variable estimator regarding both its location and its scale (we avoid here addressing these as mean and variance, because just identified instrumental variable estimators have no finite moments). To achieve this we first derive the limiting distribution of the least-squares estimator when applied to a simultaneous equation. We consider both the unconditional distribution and the effects of conditioning on predetermined information in static models. We are not aware of any published study that provides an explicit representation for this conditional asymptotic distribution in terms of its inconsistency and the degree of simultaneity as given in Kiviet and Niemczyk (2007). Analyzing it in

a particular simple class of models shows that simultaneity usually has a mitigating effect on the asymptotic variance of OLS, and comparing it with results from Monte Carlo experiments shows that even in very small samples the derived conditional asymptotic variance of least-squares provides a very accurate approximation to the actual variance. The asymptotic distribution of IV is often very informative on its behavior in finite samples, but not in cases of weak instruments due to poor identification. This is natural, because under weak instruments the standard asymptotic results do not apply.

From the limiting distribution of OLS we straightforwardly obtain a first-order asymptotic approximation to its MSE, which we can compare with its counterpart for instrumental variables. We do so over all feasible parameter values of the simple class of models examined. We find that under moderate simultaneity or for moderately weak instruments in samples of a limited size least-squares can perform much better, even substantially so, than instrumental variables. On the other hand, when both simultaneity and instrument strength are extreme, IV estimation is only marginally more (or on a root mean squared error criterion in moderately large samples roughly about twice as) precise than least-squares, although IV is uniformly superior when the sample is really large. These general predictions from first-order asymptotic theory are vindicated in simulation experiments of actual samples of sizes in the range from 20 till 200. To make such comparisons we need an equivalent to the root mean squared error, which is still meaningful when moments do not exist. Therefore we developed what we call the empirical quantile error distance, which proves to work adequately.

In practice, very often least-squares estimators are being used in situations where, according to common text-book knowledge, more sophisticated method of moments estimators seem to be called for. Some of the results in this Chapter can be used to rehabilitate the least-squares estimator for use in linear simultaneous models. However, we should warn that the present study does not provide yet proper accurate inference methods (estimated standard errors, tests, confidence sets) that can be applied to least squares when it is inconsistent. This is on the agenda for future research, that should focus also on methods to modify least-squares, in order to render it consistent, and examining its effects on the resulting efficiency.

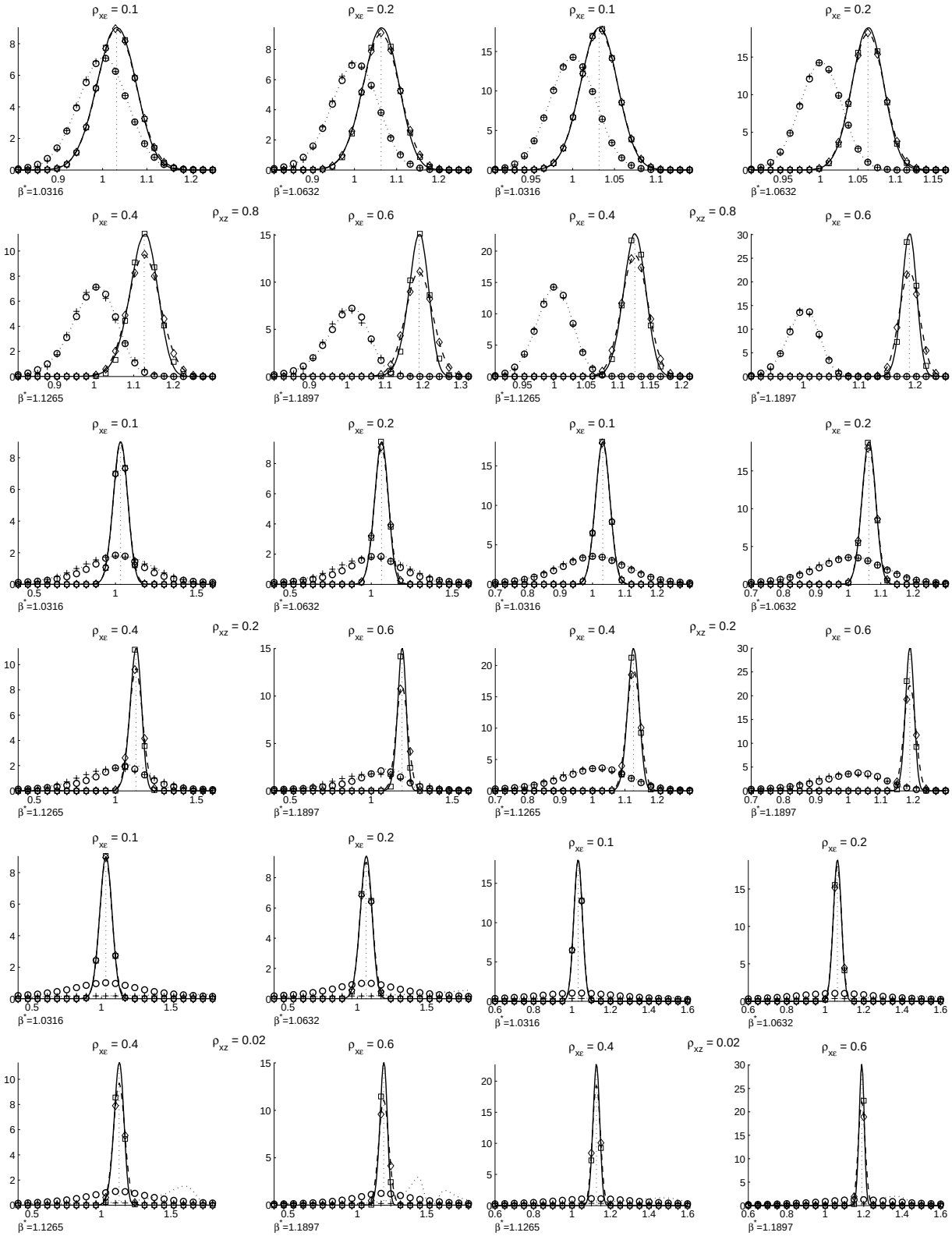

Figure 2.1: Actual/asymptotic (un)conditional densities of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{IV}$ in basic static model; $SN = 10$; $n = 50$ (first two columns), 200 (last two columns); $\rho_{xz} = 0.8$, 0.2, 0.02; $\rho_{x\varepsilon} = 0.1$, 0.2, 0.4, 0.6

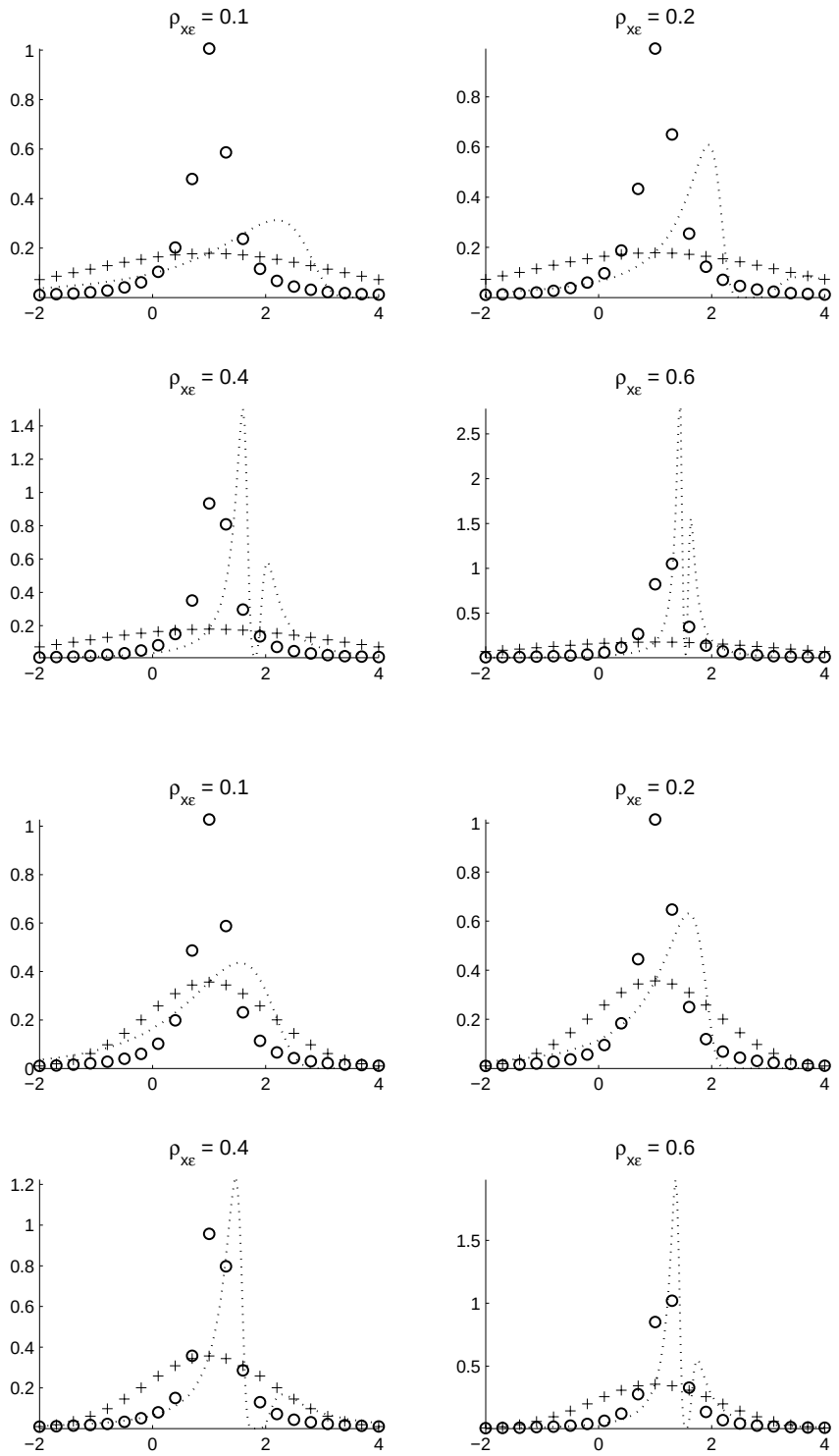

Figure 2.2: Actual/asymptotic (un)conditional densities of $\hat{\beta}_{IV}$ in basic static model; $SN = 10$; the top four diagrams are for $n = 50$ and the bottom four diagrams for 200; $\rho_{xz} = 0.02$; $\rho_{x\varepsilon} = 0.1, 0.2, 0.4, 0.6$

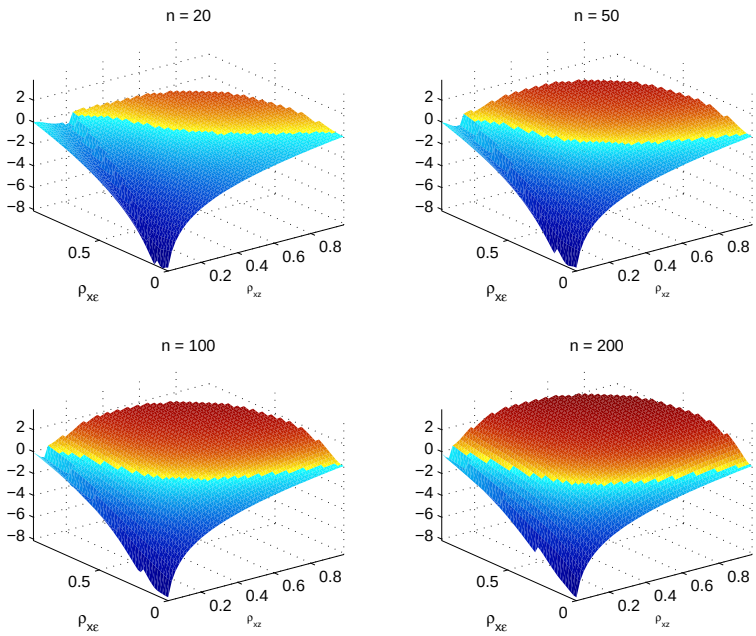

Figure 2.3: Static model, $\log[\mathsf{EQED}_C^N(\hat{\beta}_{OLS})/\mathsf{EQED}_C^N(\hat{\beta}_{IV})]$.

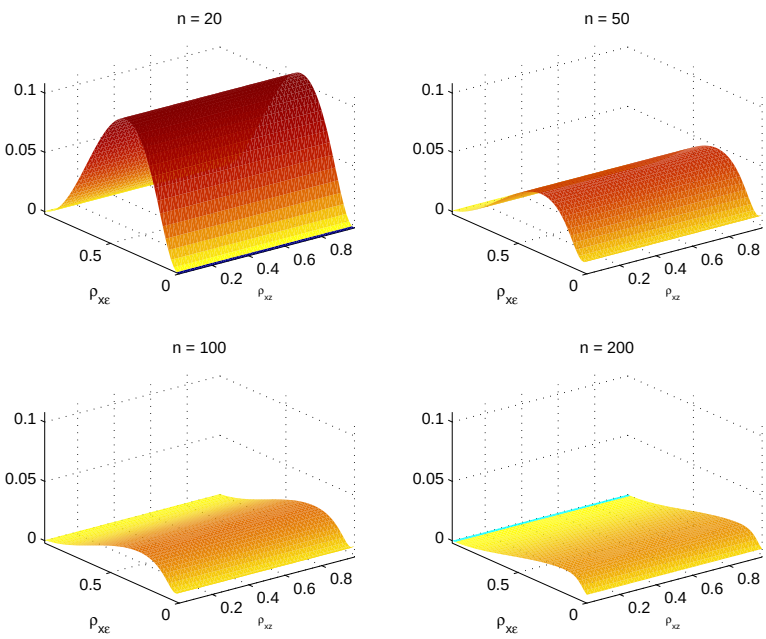

Figure 2.4: Static model, $\log[\mathsf{ARMSE}_C^N(\hat{\beta}_{OLS})/\mathsf{MRMSE}_C^N(\hat{\beta}_{OLS})]$.