



TI 2006-078/4

Tinbergen Institute Discussion Paper

# The Asymptotic and Finite Sample Distributions of OLS and Simple IV in Simultaneous Equations

*Jan F. Kiviet*

*Jerzy Niemczyk*

*University of Amsterdam, and Tinbergen Institute.*

**Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

**Tinbergen Institute Amsterdam**

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

**Tinbergen Institute Rotterdam**

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at  
<http://www.tinbergen.nl>.

# The asymptotic and finite sample distributions of OLS and simple IV in simultaneous equations

Jan F. Kiviet and Jerzy Niemczyk  
Tinbergen Institute, University of Amsterdam\*

15 September 2006

JEL-classification: C13, C15, C30

Keywords: *efficiency of an inconsistent estimator, invalid instruments, simultaneity bias, weak instruments, 4D diagrams*

## Abstract

In practice structural equations are often estimated by least-squares, thus neglecting any simultaneity. This paper reveals why this may often be justifiable and when. Assuming data stationarity and existence of the first four moments of the disturbances we find the limiting distribution of the ordinary least-squares (OLS) estimator in a linear simultaneous equations model. In simple static and dynamic models we compare the asymptotic efficiency of this inconsistent estimator with that of consistent simple instrumental variable (IV) estimators and depict cases where – due to relative weakness of the instruments or mildness of the simultaneity – the inconsistent estimator is more precise. In addition, we examine by simulation to what extent these first-order asymptotic findings are reflected in finite sample, taking into account non-existence of moments of the IV estimator. By dynamic visualization techniques we enable to appreciate any differences in efficiency over a parameter space of a much higher dimension than just two, viz. in colored animated image sequences (which are not very effective in print, but much more so in live-on-screen projection).

## 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 to establish a minimum requirement. This seems very reasonable when actual samples are so large that estimation variance is

---

\*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 and J.Niemczyk@UvA.NL. Animated graphs (4D-diagrams) are available via <http://www.fee.uva.nl/ke/jfk.htm>. We want to thank two anonymous referees and co-editor Jan Magnus for helpful comments.

relatively small. In finite sample, 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 the asymptotic variance of an inconsistent estimator. We establish the limiting distribution of such estimators and examine its relevance for actual finite sample behavior.

We focus on least-squares and instrumental variable estimators in a simple linear structural equation from a simultaneous system. 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 Maasumi and Phillips (1982). However, they do not provide an explicit representation. Joseph and Kiviet (2005) make an attempt to derive such an explicit representation for an inconsistent OLS estimator, but we will show here that this 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 in a linear regression model where the regressors are stationary and may be 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 function of the asymptotic variance of a consistent OLS estimator, 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 and dynamic 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 paper 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 some simple specific models we find that in samples with a size as large as hundred observations the actual estimation errors of IV are noticeably smaller than those of OLS only when the degree of simultaneity is very substantial and the instruments are far from weak. We also find that the first-order asymptotic approximations to the error margins of IV and OLS are often very accurate in finite sample, except in those two cases where it has been shown recently that standard asymptotics does not apply, viz. when instruments are very weak and when in dynamic models roots are very close to unity; see, for instance, Bound et al. (1995) and Elliott and Stock (2003), respectively. More generally, we re-establish that first-order asymptotic approximations are often (but not always) reasonably accurate in static stationary models, whereas in dynamic models there is usually room for substantial improvement by higher-order asymptotic approximations, see Kiviet and Phillips (2003).

The structure of this paper 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 regressand of the model is in fact jointly dependent with some of the regressors. 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 various particular simple models which are of great practical importance. In order to make models from specific classes comparable over relevant parts of their parameter space, we impose particular restrictions, such as regarding long-run multipliers and signal-to-noise ratios. Section 6 concludes.

## 2 Model, estimators and standard asymptotics

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

$$y = X\beta + \varepsilon, \tag{1}$$

where  $y$  and  $\varepsilon$  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  $\beta$  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.

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

While the functional relationship of model (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 the regressors  $X$  are used as instruments, i.e. OLS is applied and any simultaneity is neglected.

The OLS estimator of model (1) is

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y. \quad (3)$$

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

$$\hat{\beta}_{IV} = (W'X)^{-1}W'y. \quad (4)$$

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

We make standard mild stationarity assumptions yielding

$$X'X = O_p(n), \quad W'W = O_p(n), \quad W'X = O_p(n), \quad (5)$$

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

$$\Sigma_{X'X} \equiv \text{plim } n^{-1}X'X, \quad \Sigma_{W'W} \equiv \text{plim } n^{-1}W'W, \quad \Sigma_{W'X} \equiv \text{plim } n^{-1}W'X, \quad (6)$$

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

$$n^{1/2}(\hat{\beta}_{IV} - \beta) \rightarrow \mathbf{N}(0, \sigma_\varepsilon^2 \Sigma_{W'X}^{-1} \Sigma_{W'W} \Sigma_{X'W}^{-1}), \quad \text{if } \mathbf{E}(W'\varepsilon) = 0, \quad (7)$$

and

$$n^{1/2}(\hat{\beta}_{OLS} - \beta) \rightarrow \mathbf{N}(0, \sigma_\varepsilon^2 \Sigma_{X'X}^{-1}), \quad \text{if } \mathbf{E}(X'\varepsilon) = 0. \quad (8)$$

However, when  $\mathbf{E}(X'\varepsilon) \neq 0$ , OLS is inconsistent and its limiting distribution will be different from (8).

Below, we restrict ourselves to cases where  $\mathbf{E}(W'\varepsilon) = 0$  whereas  $\mathbf{E}(X'\varepsilon)$  may be non-zero, i.e. the instruments  $W$  are valid and some of the regressors may be correlated with the disturbance term. Although we will examine cases where some instruments may be weak (then the columns of  $W'X$  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 (8) when OLS is inconsistent and compare it with (7) and with actual behavior of the estimators in finite sample. 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 an explicit and calculable characterization of the limiting distribution of inconsistent OLS.

### 3 The asymptotic distribution of inconsistent OLS

We assume that the  $k \times 1$  vector  $\xi$  expresses the dependence of the current observations in the regressor matrix  $X$  on the corresponding disturbances  $\varepsilon$ , such that matrix  $X$  can be decomposed as

$$X = \bar{X} + \varepsilon\xi', \quad (9)$$

with

$$\mathbf{E}(\bar{X}'\varepsilon) = 0 \text{ and } \mathbf{E}(X'\varepsilon) = n\sigma_\varepsilon^2\xi. \quad (10)$$

Note that this does not exclude cases where  $X$  contains lagged endogenous variables. These will be a part of the component  $\bar{X}$  and have a corresponding element in  $\xi$  equal to zero. Only current endogenous regressors will have corresponding elements of  $\xi$  different from zero. Decomposition (9) with properties (10) implies

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

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

$$\Sigma_{\bar{X}'\bar{X}} = \Sigma_{X'X} - \sigma_\varepsilon^2\xi\xi'. \quad (11)$$

Below, we will often condition on (the rows of)  $\bar{X}$ .

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

$$\beta_{OLS}^* \equiv \text{plim } \hat{\beta}_{OLS} = \beta + \Sigma_{X'X}^{-1} \text{plim } n^{-1}X'\varepsilon = \beta + \sigma_\varepsilon^2\Sigma_{X'X}^{-1}\xi. \quad (12)$$

This is the pseudo true value of  $\hat{\beta}_{OLS}$ . Now, exploiting (1), (12) and (10), we obtain

$$\begin{aligned} \hat{\beta}_{OLS} - \beta_{OLS}^* &= (X'X)^{-1}X'\varepsilon - \sigma_\varepsilon^2\Sigma_{X'X}^{-1}\xi \\ &= (X'X)^{-1}[X'\varepsilon - \mathbf{E}(X'\varepsilon)] + \sigma_\varepsilon^2[(n^{-1}X'X)^{-1} - \Sigma_{X'X}^{-1}]\xi. \end{aligned} \quad (13)$$

We examine the limiting behavior of the two terms of this expression. The first term of the final expression of (13) has factor

$$X'\varepsilon - \mathbf{E}(X'\varepsilon) = \bar{X}'\varepsilon + \xi(\varepsilon'\varepsilon - n\sigma_\varepsilon^2), \quad (14)$$

which has two components. Writing  $\bar{X}'_i$  for the  $i^{\text{th}}$  row of  $\bar{X}$ , and making use of the uncorrelatedness of the elements of the disturbance vector  $\varepsilon$ , the Central Limit Theorem (CLT) yields for the first component of (14)

$$n^{-1/2}\bar{X}'\varepsilon = n^{1/2} \left( \frac{1}{n} \sum_{i=1}^n \bar{X}_i\varepsilon_i \right) \rightarrow \mathbf{N} \left( 0, \text{plim } \frac{1}{n} \sum_{i=1}^n \text{Var}(\bar{X}_i\varepsilon_i) \right), \quad (15)$$

where, conditioning on  $\bar{X}_i$ ,  $\text{Var}(\bar{X}_i\varepsilon_i) = \sigma_\varepsilon^2\bar{X}_i\bar{X}'_i$ , so that the asymptotic variance equals  $\sigma_\varepsilon^2\Sigma_{\bar{X}'\bar{X}}$ .

Upon assuming  $\mathbf{E}(\varepsilon_i^4) = \mu_4\sigma_\varepsilon^4$ , which yields variance  $(\mu_4 - 1)\sigma_\varepsilon^4$  for the mutually uncorrelated zero mean scalars  $(\varepsilon_i^2 - \sigma_\varepsilon^2)$ , the CLT also leads to

$$n^{-1/2}(\varepsilon'\varepsilon - n\sigma_\varepsilon^2) = n^{1/2} \left( \frac{1}{n} \sum_{i=1}^n (\varepsilon_i^2 - \sigma_\varepsilon^2) \right) \rightarrow \mathbf{N}[0, (\mu_4 - 1)\sigma_\varepsilon^4] \quad (16)$$

for the second component of (14).

However, to derive the asymptotic distribution of (14), we should apply the CLT to the two components jointly. After appropriate scaling, we obtain

$$\begin{aligned} n^{-1/2} [X'\varepsilon - \mathbf{E}(X'\varepsilon)] &= n^{1/2} \left[ \frac{1}{n} \sum_{i=1}^n [\bar{X}_i \varepsilon_i + \xi(\varepsilon_i^2 - \sigma_\varepsilon^2)] \right] \\ &\rightarrow \mathbf{N} \left( 0, \text{plim} \frac{1}{n} \sum_{i=1}^n \text{Var} [\bar{X}_i \varepsilon_i + \xi(\varepsilon_i^2 - \sigma_\varepsilon^2)] \right), \end{aligned}$$

because the vectors of which we consider the sample mean are mutually uncorrelated. When we also assume  $\mathbf{E}(\varepsilon_i^3) = \mu_3 \sigma_\varepsilon^3$ , we obtain

$$\text{Var} [\bar{X}_i \varepsilon_i + \xi(\varepsilon_i^2 - \sigma_\varepsilon^2)] = \sigma_\varepsilon^2 \bar{X}_i \bar{X}_i' + \sigma_\varepsilon^3 \mu_3 (\bar{X}_i \xi' + \xi \bar{X}_i') + \sigma_\varepsilon^4 (\mu_4 - 1) \xi \xi'.$$

Hence, using  $\Sigma_{\bar{X}'\iota} \equiv \text{plim} \frac{1}{n} \sum_{i=1}^n \bar{X}_i = \text{plim} \frac{1}{n} \bar{X}'\iota$ , where  $\iota$  is an  $n \times 1$  vector with all elements unity, we find

$$n^{-1/2} [X'\varepsilon - \mathbf{E}(X'\varepsilon)] \rightarrow \mathbf{N}[0, \sigma_\varepsilon^2 \Sigma_{\bar{X}'\bar{X}} + \sigma_\varepsilon^3 \mu_3 (\Sigma_{\bar{X}'\iota} \xi' + \xi \Sigma_{\iota'\bar{X}}) + \sigma_\varepsilon^4 (\mu_4 - 1) \xi \xi'].$$

Note that  $\Sigma_{\bar{X}'\iota} = \Sigma_{X'\iota}$ . Hence, when the first column of  $X$  – and thus of  $\bar{X}$  – equals  $\iota$  then  $\Sigma_{\bar{X}'\iota}$  is equal to the first columns of  $\Sigma_{\bar{X}'\bar{X}}$  and  $\Sigma_{X'X}$ . So, for the appropriately scaled first component of (13), i.e.

$$(n^{-1} X'X)^{-1} n^{-1/2} [X'\varepsilon - \mathbf{E}(X'\varepsilon)],$$

we find that it has asymptotic distribution

$$\mathbf{N}[0, \sigma_\varepsilon^2 \Sigma_{\bar{X}'\bar{X}}^{-1} + \sigma_\varepsilon^3 \mu_3 \Sigma_{\bar{X}'\bar{X}}^{-1} (\Sigma_{\bar{X}'\iota} \xi' + \xi \Sigma_{\iota'\bar{X}}) \Sigma_{\bar{X}'\bar{X}}^{-1} + \sigma_\varepsilon^4 (\mu_4 - 1) \Sigma_{\bar{X}'\bar{X}}^{-1} \xi \xi' \Sigma_{\bar{X}'\bar{X}}^{-1}]. \quad (17)$$

Under normality of the disturbances  $\varepsilon$ , which implies  $\mu_3 = 0$  and  $\mu_4 = 3$ , this specializes to formula (17) of Joseph and Kiviet (2005), which – as emerges here – incorrectly omits to take the second component of the final expression of (13) into account.

In order to obtain an improved and complete result for the asymptotic distribution of  $\hat{\beta}_{OLS}$  (which is also quite general, as it does not impose normality), we should not proceed now by deriving the asymptotic distribution of the second component of (13) separately, since we have to apply the (standard) CLT to both components jointly. Therefore, we start off again from  $\hat{\beta}_{OLS} - \beta_{OLS}^*$ , which we scale and decompose now, using (10), as follows

$$\begin{aligned} n^{1/2} (\hat{\beta}_{OLS} - \beta_{OLS}^*) &= n^{1/2} (X'X)^{-1} X'\varepsilon - \sigma_\varepsilon^2 n^{1/2} \Sigma_{X'X}^{-1} \xi \\ &= (n^{-1} X'X)^{-1} n^{-1/2} [X'\varepsilon - \mathbf{E}(X'\varepsilon)] + \sigma_\varepsilon^2 n^{1/2} [(n^{-1} X'X)^{-1} - \Sigma_{X'X}^{-1}] \xi \\ &= (n^{-1} X'X)^{-1} \{ n^{-1/2} [X'\varepsilon - \mathbf{E}(X'\varepsilon)] - \sigma_\varepsilon^2 n^{1/2} [(n^{-1} X'X) - \mathbf{E}(n^{-1} X'X)] \Sigma_{X'X}^{-1} \xi \\ &\quad - \sigma_\varepsilon^2 n^{1/2} [\mathbf{E}(n^{-1} X'X) - \Sigma_{X'X}] \Sigma_{X'X}^{-1} \xi \}. \end{aligned} \quad (18)$$

In the final expression the factor in curly brackets has three terms. The third term, which is non-random, contains a factor that can be simplified by using (9), (10) and (11), viz.

$$n^{1/2} [\mathbf{E}(n^{-1} X'X) - \Sigma_{X'X}] = n^{1/2} [\mathbf{E}(n^{-1} \bar{X}'\bar{X}) - \Sigma_{\bar{X}'\bar{X}}]. \quad (19)$$

In what follows we will neglect the third term because of factor (19). Hence, we assume that it is  $o(1)$ . This certainly holds under weak stationarity of the regressors, which implies that  $\mathbf{E}(\bar{X}_i \bar{X}'_i) = Q$  is constant, so that both  $\mathbf{E}(n^{-1} \bar{X}' \bar{X}) = Q$  and  $\Sigma_{\bar{X}' \bar{X}} = \text{plim } n^{-1} \sum_{i=1}^n \bar{X}_i \bar{X}'_i = \mathbf{E}(\bar{X}_i \bar{X}'_i) = Q$ , giving  $\mathbf{E}(n^{-1} \bar{X}' \bar{X}) - \Sigma_{\bar{X}' \bar{X}} = O$  for any positive  $n$ , so that  $n^{1/2}[\mathbf{E}(n^{-1} \bar{X}' \bar{X}) - \Sigma_{\bar{X}' \bar{X}}] = O$ , also in the limit. We want to remark that assumption (5) as such is not sufficient<sup>1</sup> for (19) to be  $o(1)$ .

The remaining two terms within curly brackets in (18) are of finite order in probability and, both separately (we showed that already for the first one) and jointly, they have a limiting normal distribution, as we shall prove now. Using (9) and (10) we find

$$X'X - \mathbf{E}(X'X) = \bar{X}'\varepsilon\xi' + \xi\varepsilon'\bar{X} + \xi(\varepsilon'\varepsilon - \sigma_\varepsilon^2 n)\xi',$$

and, using (14), we may write the two first terms between curly brackets in (??) as

$$\begin{aligned} & n^{-1/2}[X'\varepsilon - \mathbf{E}(X'\varepsilon)] - \sigma_\varepsilon^2 n^{1/2}[(n^{-1}X'X) - \mathbf{E}(n^{-1}X'X)]\Sigma_{X'X}^{-1}\xi \\ &= n^{-1/2}[\bar{X}'\varepsilon + \xi(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)] - \sigma_\varepsilon^2 n^{1/2}[\bar{X}'\varepsilon\xi' + \xi\varepsilon'\bar{X} + \xi(\varepsilon'\varepsilon - \sigma_\varepsilon^2 n)\xi']\Sigma_{X'X}^{-1}\xi \\ &= n^{-1/2}[A'\varepsilon + a(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)], \end{aligned} \quad (20)$$

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

$$A' \equiv [(1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi)I_k - \sigma_\varepsilon^2 \xi \xi' \Sigma_{X'X}^{-1}] \bar{X}', \quad (21)$$

$$a \equiv (1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi)\xi. \quad (22)$$

Denoting the  $i^{\text{th}}$  row of  $A$  as  $A'_i$  we can now write (20) as a scaled sample mean of uncorrelated random vectors  $A_i \varepsilon_i + a(\varepsilon_i^2 - \sigma_\varepsilon^2)$  and apply the standard CLT, giving

$$n^{1/2} \left( \frac{1}{n} \sum_{i=1}^n [A_i \varepsilon_i + a(\varepsilon_i^2 - \sigma_\varepsilon^2)] \right) \rightarrow \mathbf{N} \left( 0, \text{plim } \frac{1}{n} \sum_{i=1}^n \text{Var} [A_i \varepsilon_i + a(\varepsilon_i^2 - \sigma_\varepsilon^2)] \right). \quad (23)$$

Since  $\text{Var}[A_i \varepsilon_i + a(\varepsilon_i^2 - \sigma_\varepsilon^2)] = \sigma_\varepsilon^2 A_i A'_i + \sigma_\varepsilon^3 \mu_3 (A_i a' + a A'_i) + \sigma_\varepsilon^4 (\mu_4 - 1) a a'$ , we find that  $n^{1/2}(\hat{\beta}_{OLS} - \beta_{OLS}^*)$  has limiting distribution

$$\mathbf{N} \left( 0, \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} \left[ \text{plim } \frac{1}{n} [A'A - \sigma_\varepsilon \mu_3 (A' a' + a' A)] + \sigma_\varepsilon^2 (\mu_4 - 1) a a' \right] \Sigma_{X'X}^{-1} \right). \quad (24)$$

For the special case with normal disturbances, and exploiting (11), the asymptotic variance specializes to

$$\begin{aligned} & \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} [(1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi)I_k - \sigma_\varepsilon^2 \xi \xi' \Sigma_{X'X}^{-1}] \Sigma_{\bar{X}' \bar{X}} \times \\ & [(1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi)I_k - \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} \xi \xi' \Sigma_{X'X}^{-1} + 2\sigma_\varepsilon^4 (1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi)^2 \Sigma_{X'X}^{-1} \xi \xi' \Sigma_{X'X}^{-1}] \\ &= (1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi) [(1 - \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi) \sigma_\varepsilon^2 \Sigma_{X'X}^{-1} - (1 - 2\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi) \sigma_\varepsilon^4 \Sigma_{X'X}^{-1} \xi \xi' \Sigma_{X'X}^{-1}]. \end{aligned} \quad (25)$$

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  $\sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi$

<sup>1</sup>This is illustrated by the following simple, but rather pathologic, example. Let  $\bar{X}$  contain just one variable, such that  $\bar{X}_i^2 = 1 + \sqrt{i} - \sqrt{i-1}$ . Then  $n^{-1} \sum_{i=1}^n \bar{X}_i^2 = 1 + n^{-1/2}$ . Hence,  $\Sigma_{\bar{X}' \bar{X}} = 1$ , but  $n^{1/2}(n^{-1} \sum_{i=1}^n \bar{X}_i^2 - \Sigma_{\bar{X}' \bar{X}}) = 1$ , for any  $n$ . Thus, we require slightly faster convergence of  $n^{-1} \bar{X}' \bar{X}$  to  $\Sigma_{\bar{X}' \bar{X}}$ , such that their discrepancy is  $o(n^{-1/2})$ .

constitutes the population  $R^2$  of the auxiliary regression of  $\varepsilon$  on  $X$ ; denoting the OLS estimator of this regression as  $\hat{\theta} = (X'X)^{-1}X'\varepsilon$ , we find

$$R_{\varepsilon,X}^2 \equiv \text{plim} \frac{\hat{\theta}'X'X\hat{\theta}}{\varepsilon'\varepsilon} = \text{plim} \frac{\varepsilon'X(X'X)^{-1}X'\varepsilon}{\varepsilon'\varepsilon} = \sigma_\varepsilon^2 \xi' \Sigma_{X'X}^{-1} \xi, \quad (26)$$

which expresses the seriousness of the simultaneity. Substituting (26) and (12) result (25) implies

$$\text{AVar}^N(\hat{\beta}_{OLS}) = n^{-1}(1-R_{\varepsilon,X}^2)[(1-R_{\varepsilon,X}^2)\sigma_\varepsilon^2 \Sigma_{X'X}^{-1} - (1-2R_{\varepsilon,X}^2)(\beta_{OLS}^* - \beta)(\beta_{OLS}^* - \beta)'], \quad (27)$$

where the superscript  $N$  indicates that we assumed that the first four moments of the disturbances conform to normal, and where  $0 < 1 - R_{\varepsilon,X}^2 \leq 1$ . Because  $(\beta_{OLS}^* - \beta)(\beta_{OLS}^* - \beta)'$  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 (27) changes to

$$n^{-1}(1 - R_{\varepsilon,X}^2)\{ (1 - R_{\varepsilon,X}^2)\sigma_\varepsilon^2 \Sigma_{X'X}^{-1} - [(4 - \mu_4) - (5 - \mu_4)R_{\varepsilon,X}^2](\beta_{OLS}^* - \beta)(\beta_{OLS}^* - \beta)' \}. \quad (28)$$

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

$$n^{-1}\sigma_\varepsilon^3\mu_3(1 - R_{\varepsilon,X}^2)^2[e_1\xi'\Sigma_{X'X}^{-1} + \Sigma_{X'X}^{-1}\xi e_1']. \quad (29)$$

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

An expression that can be shown to be similar to (27) can be found in Rothenberg (1972). However, his formula (4.7), which is employed in Hahn and Hausman (2003), is much more involved and therefore hard to interpret. By the decomposition (9) we avoided an explicit specification of the variance matrix of the disturbances in the reduced form for  $X$ , as employed by Rothenberg (1972), and then from (25) it is easy to recognize that, apart from  $\sigma_\varepsilon^2 \Sigma_{X'X}^{-1}$ , the only determining factors of the asymptotic variance are the very meaningful characteristics: (i) the inconsistency  $\beta_{OLS}^* - \beta$  and (ii) a measure for the simultaneity  $R_{\varepsilon,X}^2$ . The incorrect result in Joseph and Kiviet (2005) yielded the expression  $n^{-1}[\sigma_\varepsilon^2 \Sigma_{X'X}^{-1} + (\beta_{OLS}^* - \beta)(\beta_{OLS}^* - \beta)']$  for  $\text{AVar}(\hat{\beta}_{OLS})$ . It can be shown that the difference between the incorrect and the correct formula 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.

## 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  $\beta$ , with pseudo true value  $\beta^*$ , the asymptotic distribution be given by

$$n^{1/2}(\hat{\beta} - \beta^*) \rightarrow \mathbf{N}(0, V). \quad (30)$$

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

$$\mathbf{AVar}(\hat{\beta}) \equiv n^{-1}V, \quad (31)$$

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

$$\mathbf{AMSE}(\hat{\beta}) \equiv n^{-1}V + (\beta^* - \beta)(\beta^* - \beta)', \quad (32)$$

which for a consistent estimator simplifies to  $n^{-1}V$ .

The simple IV estimators  $\hat{\beta}_{IV}$  considered in this study do not have finite moments in finite sample and hence their bias  $\mathbf{E}(\hat{\beta} - \beta)$ , their variance  $\mathbf{Var}(\hat{\beta})$ , and their MSE, i.e.

$$\mathbf{MSE}(\hat{\beta}) \equiv \mathbf{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \mathbf{Var}(\hat{\beta}) + \mathbf{E}(\hat{\beta} - \beta)\mathbf{E}(\hat{\beta} - \beta)', \quad (33)$$

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  $\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(R)}$ , where  $R$  is the number of replications, the habitual Monte Carlo estimator of  $\mathbf{E}(\hat{\beta})$  is the Monte Carlo sample average

$$\mathbf{ME}(\hat{\beta}) \equiv R^{-1} \sum_{r=1}^R \hat{\beta}^{(r)}. \quad (34)$$

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

$$\mathbf{MVar}(\hat{\beta}) \equiv \frac{1}{R-1} \sum_{r=1}^R (\hat{\beta}^{(r)} - \mathbf{ME}(\hat{\beta}))(\hat{\beta}^{(r)} - \mathbf{ME}(\hat{\beta}))', \quad (35)$$

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

$$\mathbf{MMSE}(\hat{\beta}) \equiv \frac{1}{R} \sum_{r=1}^R (\hat{\beta}^{(r)} - \beta)(\hat{\beta}^{(r)} - \beta)', \quad (36)$$

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  $i^{\text{th}}$  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

$$\hat{F}_i(x) \equiv \frac{1}{R} \sum_{r=1}^R I(\hat{\beta}_i^{(r)} \leq x), \quad (37)$$

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  $q^{\text{th}}$  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)

$$\hat{F}_i^{-1}(q/4) = 0.5(\hat{\beta}_i^{(qR/4)} + \hat{\beta}_i^{(1+qR/4)}), \quad q = 1, 2, 3. \quad (38)$$

To mimic the RMSE (root mean squared error) criterion, which is  $\sqrt{\sigma_i^2 + b_i^2}$ , when  $\sigma_i$  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  $\sigma_i$  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  $\Phi$  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  $\mathbf{N}(\mu, \sigma^2)$ , then

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

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

$$\hat{\sigma}_i(p) \equiv \frac{1}{2d(p)}[\hat{F}_i^{-1}(p) - \hat{F}_i^{-1}(1-p)]. \quad (39)$$

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

$$\text{EQED}(\hat{\beta}_i) \equiv \sqrt{[\hat{F}_i^{-1}(0.841345) - \hat{F}_i^{-1}(1 - 0.841345)]^2/4 + [\hat{F}_i^{-1}(0.5) - \beta_i]^2}. \quad (40)$$

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 sample. Hence, negative values for the log of the ratio will indicate that the asymptotically optimal is actually less accurate in finite sample.

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

$$\text{ARMSE}(\hat{\beta}_i) \equiv \sqrt{n^{-1}V_{ii} + (\beta_i^* - \beta_i)^2}. \quad (41)$$

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

$$\text{MRMSE}(\hat{\beta}_i) \equiv \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\beta}_i^{(r)} - \beta_i)^2} \quad (42)$$

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

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

## 5 Pictured parametrizations

In this section we specify a few very simple specific models that allow to parametrize the asymptotic characteristics of both OLS and IV. These models will be simulated too in order to assess the actual behavior in finite sample 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 drawings for the various disturbance vectors for all grid-points in the graphs. 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. In dynamic models, where we need initial values, we generated the start-up observations by drawing from the stationary distribution. The number of Monte Carlo replications for each parameter combination is 1,000,000 (for densities at  $n = 100$ ), 100,000 (for densities at  $n = 1000$ ) and 10,000 for the 3-D pictures. The diagrams presented below are single images from animated

versions, which are available via the world-wide-web and allow to inspect the relevant phenomena over a much larger part of the parameter space.

For the simplest of the static models that we examine below some analytic finite sample properties are available; see Woglom (2001) and Hillier (2005) 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 in the more complex, though more relevant, dynamic models.

## 5.1 A basic static model

We commence by considering the most basic example we can think of, viz. a model with one regressor and one valid and either strong or weak instrument. The two variables  $x$  and  $w$ , together with the dependent variable  $y$ , are jointly IID (independent and identically distributed) with zero mean and finite second moments. This case may be denoted as

$$y_i = \beta x_i + \varepsilon_i, \quad (43)$$

$$x_i = \bar{x}_i + \xi \varepsilon_i, \quad (44)$$

where  $\xi$  is scalar now. Data for  $y$ ,  $x$  and  $w$  can be obtained by the generating scheme

$$\varepsilon_i = \sigma_\varepsilon v_{1i},$$

$$\bar{x}_i = \alpha_1 v_{2i},$$

$$w_i = \alpha_2 v_{2i} + \alpha_3 v_{3i},$$

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

$$\begin{pmatrix} \varepsilon_i \\ x_i \\ w_i \end{pmatrix} = P v_i = \begin{pmatrix} \sigma_\varepsilon & 0 & 0 \\ \sigma_\varepsilon \xi & \alpha_1 & 0 \\ 0 & \alpha_2 & \alpha_3 \end{pmatrix} v_i, \quad (45)$$

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

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  $y_i = \hat{\beta} \hat{x}_i + \varepsilon_i$ , with  $\hat{\beta} \neq 0$ , and rescaled the explanatory variable such that  $x_i = \hat{x}_i / \hat{\beta}$ . We can impose some further normalizations on the 5 parameters of  $P$ , because, without loss of generality, we may take

$$\sigma_\varepsilon = 1, \quad (46)$$

$$\sigma_w^2 = \alpha_2^2 + \alpha_3^2 = 1. \quad (47)$$

By (46) we normalize all results with respect to  $\sigma_\varepsilon$ , and because the IV estimator is invariant to the scale of the instruments (only the space spanned by  $w$  is relevant) we may impose (47) which will be used to obtain the value

$$\alpha_3^2 = 1 - \alpha_2^2 \geq 0. \quad (48)$$

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

$$\left. \begin{aligned} \sigma_x^2 &= \xi^2 + \alpha_1^2 & \sigma_y^2 &= \xi^2 + 2\xi + 1 + \alpha_1^2 \\ \sigma_{x\varepsilon} &= \xi & \rho_{x\varepsilon} &= \xi/\sqrt{\xi^2 + \alpha_1^2} \\ \sigma_{w\varepsilon} &= 0 & \rho_{w\varepsilon} &= 0 \\ \sigma_{xw} &= \alpha_1\alpha_2 & \rho_{xw} &= \alpha_1\alpha_2/\sqrt{\xi^2 + \alpha_1^2} \end{aligned} \right\} \quad (49)$$

Note that these depend on only 3 remaining free parameters: viz.  $\xi$ ,  $\alpha_1$  and  $\alpha_2$ , and so will the expressions for asymptotic variance (together with  $\mu_3$  and  $\mu_4$ , the 3rd and 4th moments of  $v_{1i}$ ).

However, instead of designing our results in terms of the three parameters  $\xi$ ,  $\alpha_1$  and  $\alpha_2$ , we prefer another parametrization. We shall use as a base of the design parameter space for this simple model, the three parameters:  $\rho_{x\varepsilon}$ ,  $\rho_{xw}$  and  $PF$ , where

$$PF = \sigma_x^2/(\sigma_x^2 + 1) \geq 0. \quad (50)$$

The latter parameter expresses the population fit of the regression of interest; it is equal to the variance of the explanatory part as a fraction of that variance plus the disturbance variance. Note that the denominator differs from  $\sigma_y^2$ . By fixing  $PF$ , we basically fix the ratio  $\beta^2\sigma_x^2/\sigma_\varepsilon^2 = \sigma_x^2 = PF/(1 - PF)$ , which is the signal-noise ratio.

This reparametrization is useful because the parameters  $\rho_{x\varepsilon}$ ,  $\rho_{xw}$  and  $PF$  have a direct econometric interpretation, viz. the degree of simultaneity, instrument strength and model fit, respectively. By varying the three parameters  $|\rho_{x\varepsilon}| < 1$ ,  $|\rho_{xw}| < 1$  and  $0 < PF < 1$ , we can examine the whole parameter space of this model. For given values of  $PF$  and  $\rho_{x\varepsilon}$  one can obtain  $\xi$  and  $\alpha_1$ , i.e.

$$\xi = \rho_{x\varepsilon} \left| \sqrt{PF/(1 - PF)} \right|, \quad (51)$$

$$\alpha_1 = \left| \sqrt{PF/(1 - PF) - \xi^2} \right| = \left| \sqrt{PF(1 - \rho_{x\varepsilon}^2)/(1 - PF)} \right|. \quad (52)$$

With  $\rho_{xw}$  we can now obtain

$$\alpha_2 = \rho_{xw} / \left| \sqrt{1 - \rho_{x\varepsilon}^2} \right| \quad (53)$$

and, of course,

$$\alpha_3 = \left| \sqrt{1 - \alpha_2^2} \right| = \left| \sqrt{(1 - \rho_{x\varepsilon}^2 - \rho_{xw}^2)/(1 - \rho_{x\varepsilon}^2)} \right|, \quad (54)$$

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

In this simple model we have

$$\left. \begin{aligned} \beta_{OLS}^* - \beta &= \frac{\xi}{\sigma_x^2} = \rho_{x\varepsilon} \left| \sqrt{(1 - PF)/PF} \right| \\ R_{\varepsilon, X}^2 &= \frac{\xi^2}{\sigma_x^2} = \rho_{x\varepsilon}^2 \\ \Sigma_{X'X} &= \sigma_x^2 = PF/(1 - PF) \\ \Sigma_{W'X}^{-1} \Sigma_{W'W} \Sigma_{X'W}^{-1} &= \sigma_w^2/\sigma_{xw}^2 = 1/\rho_{xw}^2 \sigma_x^2 = (1 - PF)/(PF \times \rho_{xw}^2) \end{aligned} \right\} \quad (55)$$

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

$$\text{AVar}^N(\hat{\beta}_{OLS}) = n^{-1}(1 - \rho_{x\varepsilon}^2)(1 - 2\rho_{x\varepsilon}^2 + 2\rho_{x\varepsilon}^4)(1 - PF)/PF. \quad (56)$$

This yields

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

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 (27)).

Result (56) implies for the first-order asymptotic approximation to the mean squared error under normality of the disturbances the specific result

$$\text{AMSE}^N(\hat{\beta}_{OLS}) = [n^{-1}(1 - \rho_{x\varepsilon}^2)(1 - 2\rho_{x\varepsilon}^2 + 2\rho_{x\varepsilon}^4) + \rho_{x\varepsilon}^2](1 - PF)/PF, \quad (57)$$

from which we find  $\frac{\partial}{\partial \rho_{x\varepsilon}^2} \text{AMSE}(\hat{\beta}_{OLS}) > 0$  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 (57) with the corresponding quantity for IV

$$\text{AVar}(\hat{\beta}_{IV}) = (1 - PF)/(n \times PF \times \rho_{xw}^2). \quad (58)$$

Note that, unlike  $\text{AVar}(\hat{\beta}_{OLS})$ , 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  $\text{AMSE}^N(\hat{\beta}_{OLS}) < \text{AMSE}(\hat{\beta}_{IV}) = \text{AVar}(\hat{\beta}_{IV})$ , i.e. for

$$\rho_{xw}^2[(1 - \rho_{x\varepsilon}^2)(1 - 2\rho_{x\varepsilon}^2 + 2\rho_{x\varepsilon}^4) + n\rho_{x\varepsilon}^2] < 1. \quad (59)$$

Note that this watershed between IV and OLS as far as AMSE is concerned is invariant with respect to  $PF$ , and so is the relative (but not the absolute) difference in AMSE. Self-evidently (59) 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  $\rho_{xw}^2$  gets close to zero. When  $\rho_{xw}^2 = 0$  the equation is not identified. For IV this implies an exploding variance but not for OLS, where  $\text{AMSE}^N(\hat{\beta}_{OLS})$  is not affected by  $\rho_{xw}^2$ . So, although obtaining meaningful inference on  $\beta$  will be an illusion,  $\hat{\beta}_{OLS}$  has still a well-defined distribution.

Since

$$\hat{\beta}_{OLS} - \beta = \sqrt{\frac{1-PF}{PF}} \frac{\sum_{i=1}^n (\rho_{x\varepsilon} v_{i1} + \sqrt{1-\rho_{x\varepsilon}^2} v_{i2}) v_{i1}}{\sum_{i=1}^n (\rho_{x\varepsilon} v_{i1} + \sqrt{1-\rho_{x\varepsilon}^2} v_{i2})^2}, \quad (60)$$

$$\hat{\beta}_{IV} - \beta = \sqrt{\frac{1-PF}{PF}} \frac{\sum_{i=1}^n (\alpha_2 v_{i2} + \alpha_3 v_{i3}) v_{i1}}{\sum_{i=1}^n (\alpha_2 v_{i2} + \alpha_3 v_{i3}) (\rho_{x\varepsilon} v_{i1} + \sqrt{1-\rho_{x\varepsilon}^2} v_{i2})},$$

the finite sample distributions of both  $\hat{\beta}_{OLS}$  and  $\hat{\beta}_{IV}$  are determined by  $PF$  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 (55), and thus carries over to the asymptotic variances (27) and (58) too. From (60) 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  $\rho_{xw}$ , so we will examine positive values only.

The actual values of  $\beta_{OLS}^*$  and of (the square root of)  $\text{AMSE}^N(\hat{\beta}_{OLS})$  and  $\text{AVar}(\hat{\beta}_{IV})$  could be calculated and tabulated now for various values of  $n$ ,  $PF$ ,  $\rho_{x\varepsilon}$  and  $\rho_{xw}$  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 4. We will portray these over the relevant parameter space. From the foregoing it is clear that varying  $PF$  will have a rather straightforward and relatively neutral effect, so we focus much more on the effects of  $\rho_{x\varepsilon}$ ,  $\rho_{xw}$  and  $n$ .

In the Figures 1 and 2 densities are presented, both for OLS (*red or grey lines*) and IV (*black or dark lines*), both for the actual empirical distribution (*solid lines*) and for its asymptotic approximation (*dashed lines*). For the latter we take

$$\begin{aligned}\hat{\beta}_{OLS} &\stackrel{a}{\sim} N\left(\beta_{OLS}^*, n^{-1}\mathbf{AVar}^N(\hat{\beta}_{OLS})\right), \\ \hat{\beta}_{IV} &\stackrel{a}{\sim} N\left(\beta, n^{-1}\mathbf{AVar}(\hat{\beta}_{IV})\right).\end{aligned}\tag{61}$$

In the simulations we took  $v_i \sim \text{IIN}(0, I_3)$ . 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 normal approximations (61)? Is there a qualitative difference between the accuracy of the OLS and the IV asymptotic approximations? Do these densities already 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 two pairs of asymptotic and empirical distributions.

Both Figures 1 and 2 consist of eight panels. The top four concern mild simultaneity ( $\rho_{x\varepsilon} = 0.2$ ) and the bottom four more severe simultaneity ( $\rho_{x\varepsilon} = 0.5$ ). In each block of four the panels concern the cases  $\rho_{xw} = 0.1, 0.5, 0.75$  and  $0.85$  respectively. Each panel contains the four densities for the case  $PF = 0.5$ . Figure 1 has  $n = 100$  and Figures 2 has  $n = 1000$ . We find that for a relatively strong instrument, i.e.  $\rho_{xw} \geq 0.5$ , and relatively strong simultaneity, i.e.  $\rho_{x\varepsilon} \geq 0.5$ , the IV estimator is clearly more attractive than the OLS estimator, when  $n \geq 100$ . However, for  $\rho_{x\varepsilon} = 0.2$  this is less clear-cut. For a rather weak instrument ( $\rho_{xw} = 0.1$ ) the density of IV is so flat that it is obvious that OLS is more attractive. Then the bias and inconsistency of OLS do not seem to disqualify the OLS estimator in comparison to IV, because OLS has a relatively moderate variance. The quality of the asymptotic approximation of IV is very bad (as is well known) when the instrument is extremely weak. Self-evidently it improves with the sample size. Especially at  $n = 100$  it is noticeable that the asymptotic approximation of IV does not represent the asymmetry of the actual empirical distribution nor the fatness of at least one of its tails. The asymptotic approximation to the actual distribution of OLS is much better when  $\rho_{x\varepsilon} = 0.2$  than for  $\rho_{x\varepsilon} = 0.5$ , where, even for  $n = 1000$ , the actual and asymptotic densities show substantial discrepancy. Note that each block of four panels contains the same two OLS densities (because they are not affected by  $\rho_{xw}$ ), but just on a different scale.

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 sample we produce here 3D graphs (and 4D graphs on the web) of

$$\log\left(\text{EQED}(\hat{\beta}_{OLS})/\text{EQED}(\hat{\beta}_{IV})\right),\tag{62}$$

for fixed values of  $PF$  and  $n$  over the  $(\rho_{x\varepsilon}, \rho_{xw})$  plane. This log-ratio (62) 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 3 correspond to  $n = 20, 50, 100$  and  $200$  respectively. We took  $PF = 0.9$  but this ratio is invariant with respect to  $PF$ . These graphs illustrate that IV performs better when both  $\rho_{x\varepsilon}$  and  $\rho_{xw}$  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  $\exp(2) \times 100\%$  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  $\rho_{x\varepsilon}^2 + \rho_{xw}^2 < 1$ ) 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

$$F \equiv n \frac{\sigma_x^2 - \sigma_x^2(1 - \rho_{xw}^2)}{\sigma_x^2(1 - \rho_{xw}^2)} = n \frac{\rho_{xw}^2}{1 - \rho_{xw}^2}. \quad (63)$$

Instrument weakness is associated with small values of  $F$ , say  $F \leq 10$ . The latter implies here  $\rho_{xw}^2 \leq 10/(n + 10)$  or  $|\rho_{xw}| \leq 0.58$  (for  $n = 20$ ) and  $|\rho_{xw}| \leq 0.3$  (for  $n = 100$ ). From Figure 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 4 examines the quality of the asymptotic approximation of the empirical OLS distribution on a RMSE criterion. The 3D graphs represent

$$\log \left( \text{ARMSE}(\hat{\beta}_{OLS}) / \text{MRMSE}(\hat{\beta}_{OLS}) \right), \quad (64)$$

hence positive values indicate pessimism of the asymptotic approximation (actual RMSE smaller than first-order asymptotic approximation) and negative values optimism. Self-evidently  $\rho_{x\varepsilon}$  has no effect, but apart from the closeness of the densities as represented in the earlier figures, the relative size of  $\beta_{OLS}^* - \beta$  has. We find that the asymptotic approximation of MSE developed in this study is especially accurate when the simultaneity is serious.

The above model can easily be generalized, for instance by including another explanatory variable for  $y_t$  or by adding further exogenous regressors in the generating scheme for  $\bar{x}_t$ , so that more valid instruments are available. Or we can generalize the generating processes for  $z_t$  or  $\bar{x}_t$  into AR(1) processes. Then  $z_{t-1}$  and  $x_{t-1}$  establish extra valid instruments. Note, however, that any extra valid instruments will be effective and improve the asymptotic performance only if they are also incorporated with a non-zero coefficient in the generating scheme for  $x_t$ . Below we introduce dynamic aspects into the above model in two steps.

### 5.1.1 A basic semi-dynamic model

Below we stick to the simple static model for  $y_t$ , but make the reduced form equation for its stationary explanatory variable  $x_t$  dynamic by choosing an AR(1) scheme with autoregressive coefficient  $|\kappa| < 1$ . Then the model can be written (we index the observations now by  $t = 1, \dots, n$ )

$$\left. \begin{aligned} y_t &= \beta x_t + \varepsilon_t, \\ x_t &= \kappa x_{t-1} + \eta_t + \xi \varepsilon_t, \end{aligned} \right\} \quad (65)$$

where  $\varepsilon_t \sim \text{IID}(0,1)$  is independent of  $\eta_t \sim \text{IID}(0, \sigma_\eta^2)$ . Generating these for  $t = 0, \dots, n$  we can obtain a starting value

$$x_0 = (\eta_0 + \xi\varepsilon_0)/\sqrt{1 - \kappa^2}. \quad (66)$$

Taking  $w_t = x_{t-1}$ , this yields (again fixing  $\beta = 1$ )

$$\left. \begin{aligned} \sigma_x^2 &= (\xi^2 + \sigma_\eta^2)/(1 - \kappa^2) & \sigma_y^2 &= \sigma_x^2 + 2\xi + 1 \\ \sigma_{x\varepsilon} &= \xi & \rho_{x\varepsilon} &= \xi/\sigma_x \\ \sigma_{w\varepsilon} &= 0 & \rho_{w\varepsilon} &= 0 \\ \sigma_{xw} &= \kappa\sigma_x^2 & \rho_{xw} &= \kappa \end{aligned} \right\} \quad (67)$$

Using again the same base for the parametrization, i.e.  $\rho_{x\varepsilon}$ ,  $\rho_{xw}$  and  $PF$ , where now  $\sigma_x^2 = \sigma_w^2 = PF/(1 - PF)$ , we find the coefficients of the generating scheme from

$$\left. \begin{aligned} \kappa &= \rho_{xw} \\ \xi &= \rho_{x\varepsilon} \sqrt{PF/(1 - PF)} \\ \sigma_\eta^2 &= (1 - \rho_{x\varepsilon}^2 - \rho_{xw}^2)PF/(1 - PF) \end{aligned} \right\} \quad (68)$$

The latter result highlights that in this model we again have  $\rho_{x\varepsilon}^2 + \rho_{xw}^2 < 1$  and regarding asymptotic performance we find again the results of (56), (57), (58) and (59), because (55) still holds.

We observed that the finite sample distributions too are little affected by the serial dependence in the stationary series for  $y_t$  and  $x_t$  in this model, because they were found to be virtually similar to those of the static model, especially when  $\rho_{xw}$  is small. Hence, also for this model, see Figure 5, we find a very substantial area in the parameter space where OLS beats IV, and again the population  $F$  statistic of the reduced form cannot properly identify that area. Figure 6 indicates that changing the characteristics of the instrument  $x_{t-1}$  through  $\rho_{xw}$  also affects the OLS estimator now, because it has a direct effect on the regressor  $x_t$ . Note that AMSE is reasonably accurate (but always too optimistic) when  $n$  is not too small and  $\kappa = \rho_{xw}$  is not very large.

### 5.1.2 A simple fully dynamic model

We can make the semi-dynamic model fully dynamic by sticking to the same reduced form for  $x_t$ , and adding a lagged-dependent explanatory variable to the equation of interest, giving

$$\left. \begin{aligned} y_t &= \beta x_t + \gamma y_{t-1} + \varepsilon_t, \\ x_t &= \kappa x_{t-1} + \eta_t + \xi\varepsilon_t, \end{aligned} \right\} \quad (69)$$

with  $|\gamma| < 1$  to ensure stationarity. We can still normalize  $\sigma_\varepsilon^2 = 1$ , without loss of generality. Instead of normalizing again with respect to  $\beta$ , now we prefer to normalize the long-run multiplier of  $y$  with respect to  $x$ , i.e. we take

$$\beta = 1 - \gamma. \quad (70)$$

To establish the asymptotic results on OLS and IV, using as instruments  $x_{t-1}$  and  $y_{t-1}$ , we have to find expressions in terms of the parameters for the elements of  $\Sigma_{X'X}$ ,  $\Sigma_{W'X}$  and  $\Sigma_{W'W}$ , i.e. for  $\text{Var}(x_t)$ ,  $\text{Var}(y_t)$ ,  $\text{Cov}(x_t, y_{t-1})$ ,  $\text{Cov}(x_t, x_{t-1})$  and  $\text{Cov}(x_t, y_t)$ .

Exploiting the assumed stationarity and the normalizations, these five data moments obey the five equations

$$\left. \begin{aligned} (1 - \kappa^2) \text{Var}(x_t) &= \xi^2 + \sigma_\eta^2 \\ (1 - \gamma^2) \text{Var}(y_t) &= (1 - \gamma)^2 \text{Var}(x_t) + 2\gamma(1 - \gamma) \text{Cov}(x_t, y_{t-1}) + 2(1 - \gamma)\xi + 1 \\ \text{Cov}(x_t, y_t) &= (1 - \gamma) \text{Var}(x_t) + \gamma \text{Cov}(x_t, y_{t-1}) + \xi \\ \text{Cov}(x_t, y_{t-1}) &= \kappa \text{Cov}(x_t, y_t) \\ \text{Cov}(x_t, x_{t-1}) &= \kappa \text{Var}(x_t) \end{aligned} \right\} \quad (71)$$

These yield

$$\left. \begin{aligned} \text{Cov}(x_t, x_{t-1}) &= \kappa \text{Var}(x_t) \\ \text{Cov}(x_t, y_t) &= [(1 - \gamma) \text{Var}(x_t) + \xi]/(1 - \gamma\kappa) \\ \text{Cov}(x_t, y_{t-1}) &= \kappa[(1 - \gamma) \text{Var}(x_t) + \xi]/(1 - \gamma\kappa) \\ \text{Var}(y_t) &= \frac{(1-\gamma)(1+\gamma\kappa)}{(1+\gamma)(1-\gamma\kappa)} \text{Var}(x_t) + \frac{2}{(1+\gamma)(1-\gamma\kappa)} \xi + \frac{1}{1-\gamma^2} \\ \text{Var}(x_t) &= (\xi^2 + \sigma_\eta^2)/(1 - \kappa^2) \end{aligned} \right\} \quad (72)$$

Note that all data moments are determined by four parameters, viz.  $\gamma$ ,  $\kappa$ ,  $\xi$  and  $\sigma_\eta^2$ .

A set of meaningful design parameters for this more complex model is obtained as follows. To control for the strength of the instrument we use the population fit of the reduced form regression for  $x_t$ , which we define as

$$PFR \equiv \kappa^2 \sigma_x^2 / \sigma_x^2 = \kappa^2. \quad (73)$$

Furthermore, we have

$$\rho_{x\varepsilon} = \xi / \sigma_x = \xi \sqrt{(1 - \kappa^2) / (\xi^2 + \sigma_\eta^2)}. \quad (74)$$

For the population fit of the equation of primary interest we now have

$$PF = \text{Var}(\beta x_t + \gamma y_{t-1}) / [\text{Var}(\beta x_t + \gamma y_{t-1}) + 1].$$

Using (72) this yields

$$\begin{aligned} PF/(1 - PF) &= (1 - \gamma)^2 \text{Var}(x_t) + \gamma^2 \text{Var}(y_t) + 2\gamma(1 - \gamma) \text{Cov}(x_t, y_{t-1}) \\ &= g(\gamma, \kappa, \xi, \sigma_\eta^2), \end{aligned} \quad (75)$$

where  $g(\cdot)$  is a non-linear function. For chosen values of the three design parameters  $PF$ ,  $PFR$  and  $\rho_{x\varepsilon}$ , however, we cannot solve the four parameters  $\gamma$ ,  $\kappa$ ,  $\xi$ ,  $\sigma_\eta^2$  from the three non-linear equations (75), (73) and (74). Therefore, we shall also use the characterization of the dynamics  $\gamma$  as a design parameter. Provided the three non-linear equations can be solved (for which we use Mathematica) for chosen values of  $PF$ ,  $PFR$ ,  $\rho_{x\varepsilon}$  and  $\gamma$ , the components of  $\text{AVar}^N(\hat{\beta}_{OLS})$  and  $\text{AVar}(\hat{\beta}_{IV})$  can be calculated and for chosen  $n$  be compared with their simulated counterparts. In the simulations we generate  $x_0$  again according to (66) and  $y_0$  as follows. We took

$$y_0 = \alpha_1^* \eta_0 + \alpha_2^* \varepsilon_0, \quad (76)$$

where  $\alpha_1^*$  and  $\alpha_2^*$  are chosen such that  $\text{Var}(y_0) = \alpha_1^{*2} + \alpha_2^{*2}$  and  $\text{Cov}(x_0, y_0) = (\alpha_1^* + \alpha_2^* \xi) / \sqrt{1 - \kappa^2}$  obey the solutions of (72).

We present the results just in the form of densities, for the specific values  $n = 100$ ,  $PF = 0.95$  and  $\gamma = 0.5$ . For  $\rho_{x\varepsilon}$  and  $PFR = \kappa^2$  we choose similar values as before. Figures 7 gives densities for estimates of  $\beta = 0.5$ . Note that the asymptotic approximations are not very far off the corresponding empirical distributions. The IV distributions are very flat, and therefore most of the time OLS seems to be better than IV. Figure 9 presents the densities for estimates of  $\gamma = 0.5$ . The IV distributions are less flat here, but they show some bias in the same direction as OLS. Again the estimation errors made by OLS seem usually less substantial than those of IV.

## 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 not random samples from a well-defined population usually. 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 a recent 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 paper 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 are not aware of any published study that provides an explicit representation for this asymptotic distribution in terms of its inconsistency and the degree of simultaneity. Analyzing it in a few particular 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 in many cases (and in static models especially) the asymptotic variance of least-squares provides a reasonable 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 for various specific types of models over all feasible parameter values for particular classes of these models, where the latter are limited by particular chosen values of long-run multipliers and signal-to-noise ratios. We find that least-squares can perform much better, even substantially so, than instrumental variables under moderate simultaneity or for moderately weak instruments in samples of a limited size. 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 paper can be used to rehabilitate the least-squares estimator for use in linear (dynamic) 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.

## References

- 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.
- Bun, M.J.G., Kiviet, J.F., 2006. The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics* 132, 409-444.
- Dufour, J-M., 2003. Identification, weak instruments and statistical inference in econometrics. *Canadian Journal of Economics* 36, 767-808.
- Elliott, G., Stock, J.H., 2001. Confidence intervals for autoregressive coefficients near one. *Journal of Econometrics* 103, 155-181.
- Gallant, A.R., White, H., 1988. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, Oxford.
- Hahn, J., Hausman, J.A., 2003. IV estimation with valid and invalid instruments: application to the returns of education. mimeo. To appear in *Les Annales d'Economie et de Statistique*.
- Hahn, J., Inoue, A., 2002. A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators. *Econometric Reviews*

21, 309-336.

Hall, A.R., Inoue, A., 2003. The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* 114, 361-394.

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.

Kiviet, J.F., Phillips, G.D.A., 2003. Improved Coefficient and Variance Estimation in Stable First-Order Dynamic Regression Models. UvA-Econometrics discussion paper 2002/02.

Maasumi, E., Phillips, P.C.B., 1982. On the behavior of inconsistent instrumental variable estimators. *Journal of Econometrics* 19, 183-201.

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.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

West, K.D., Wilcox, D.W., 1996. A comparison of alternative instrumental variables estimators of a dynamic linear model. *Journal of Business & Economic Statistics* 14, 281-293.

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

Figure 1:  $\hat{\beta}_{OLS}$  and  $\hat{\beta}_{IV}$  in static model,  $n = 100$ ,  $PF = 0.5$

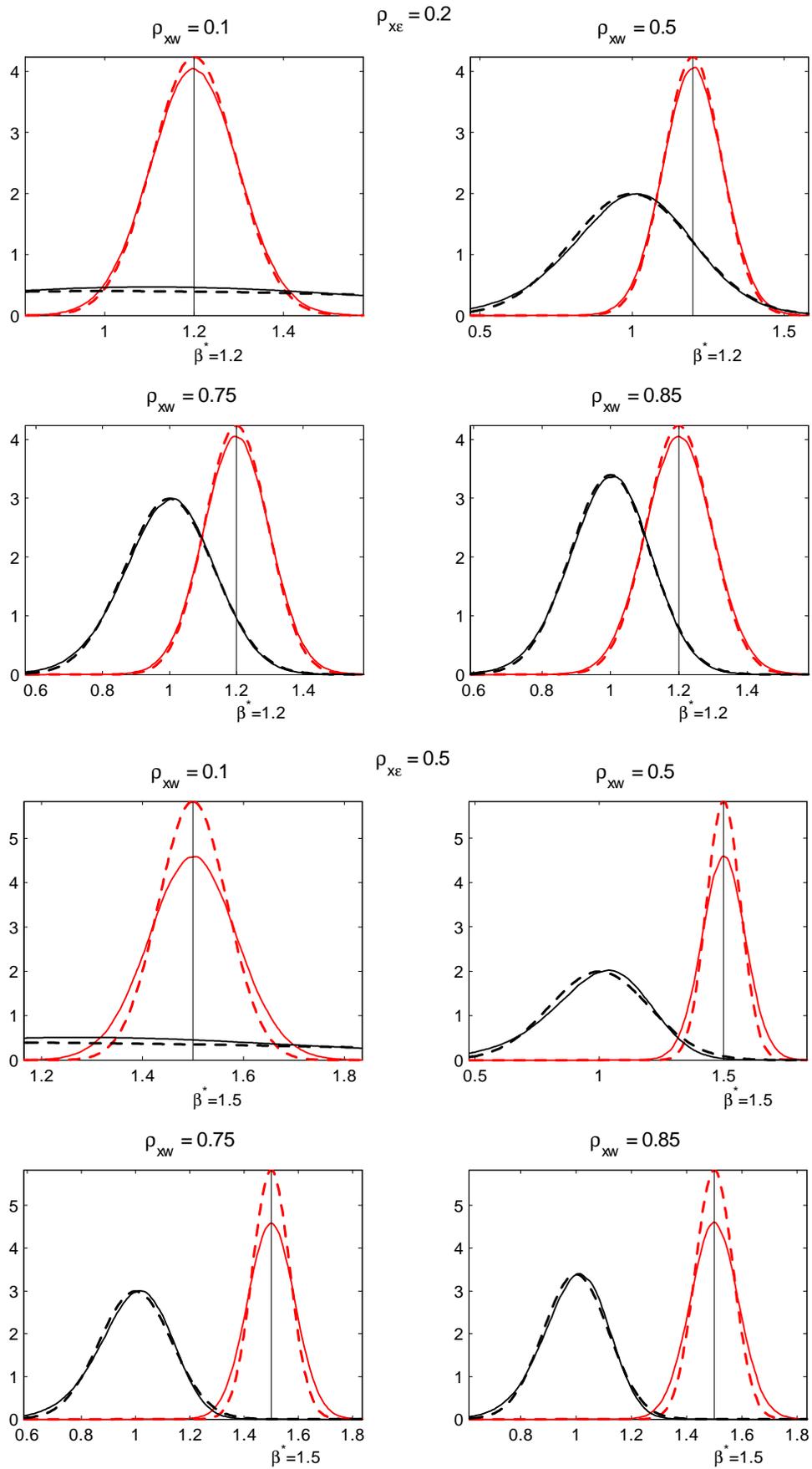


Figure 2:  $\hat{\beta}_{OLS}$  and  $\hat{\beta}_{IV}$  in static model,  $n = 1000$ ,  $PF = 0.5$

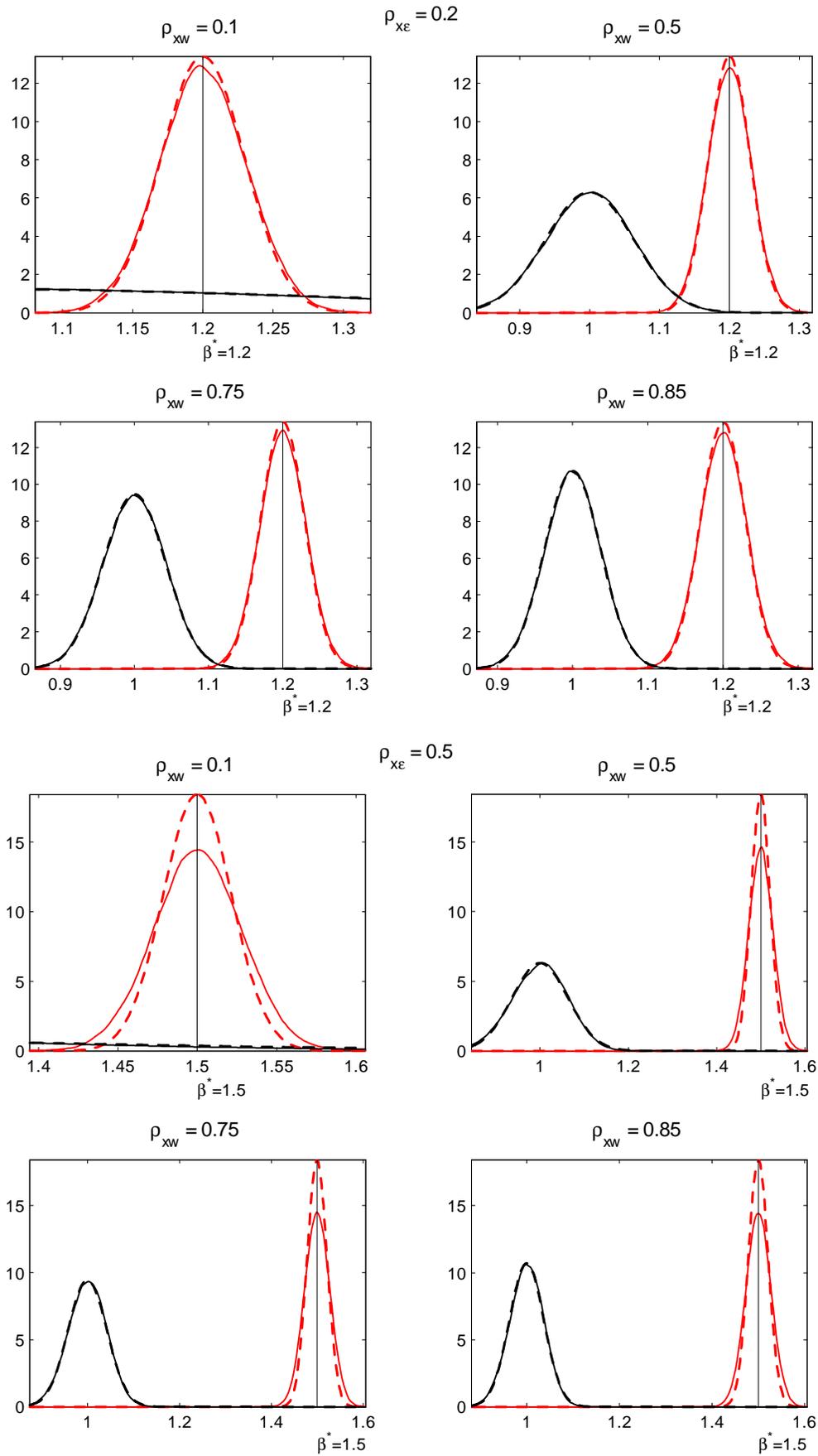


Figure 3: static model,  $\log[\text{EQED}(\hat{\beta}_{OLS})/\text{EQED}(\hat{\beta}_{IV})]$

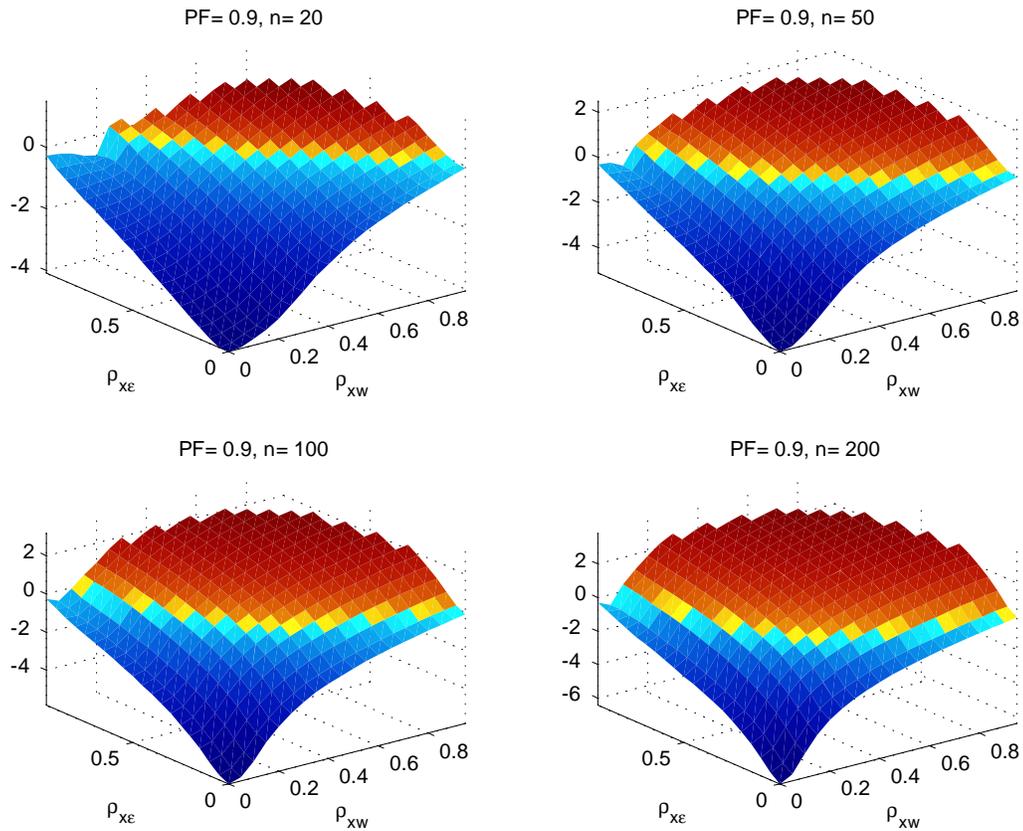


Figure 4: static model,  $\log \text{ARMSE}[(\hat{\beta}_{OLS}), \text{MRMSE}(\hat{\beta}_{OLS})]$

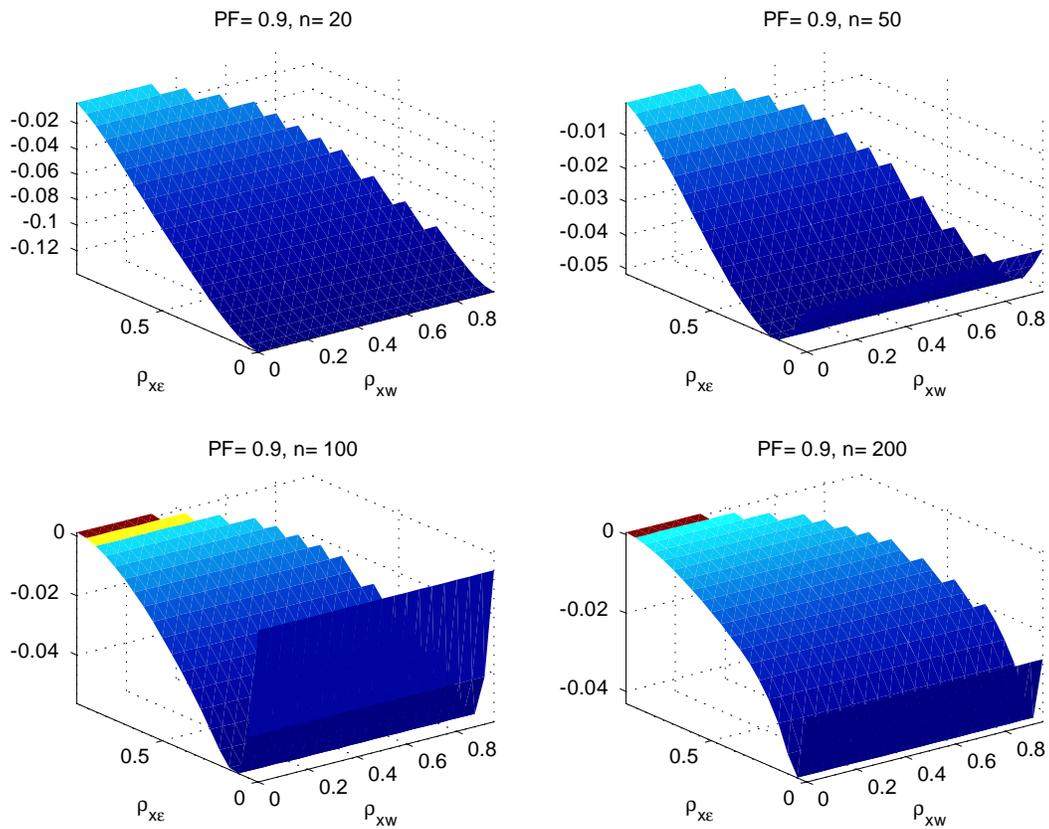


Figure 5: semi-dynamic model,  $\log[\text{EQED}(\hat{\beta}_{OLS})/\text{EQED}(\hat{\beta}_{IV})]$

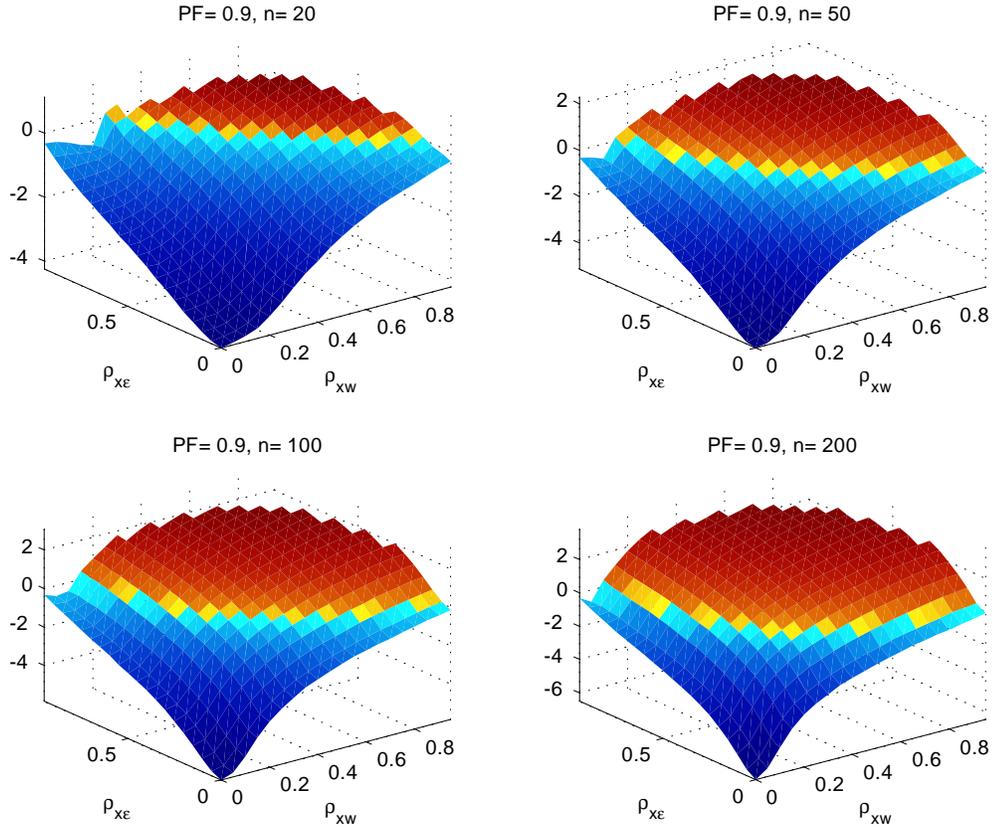


Figure 6: semi-dynamic model,  $\log \text{ARMSE}[(\hat{\beta}_{OLS}), \text{MRMSE}(\hat{\beta}_{OLS})]$

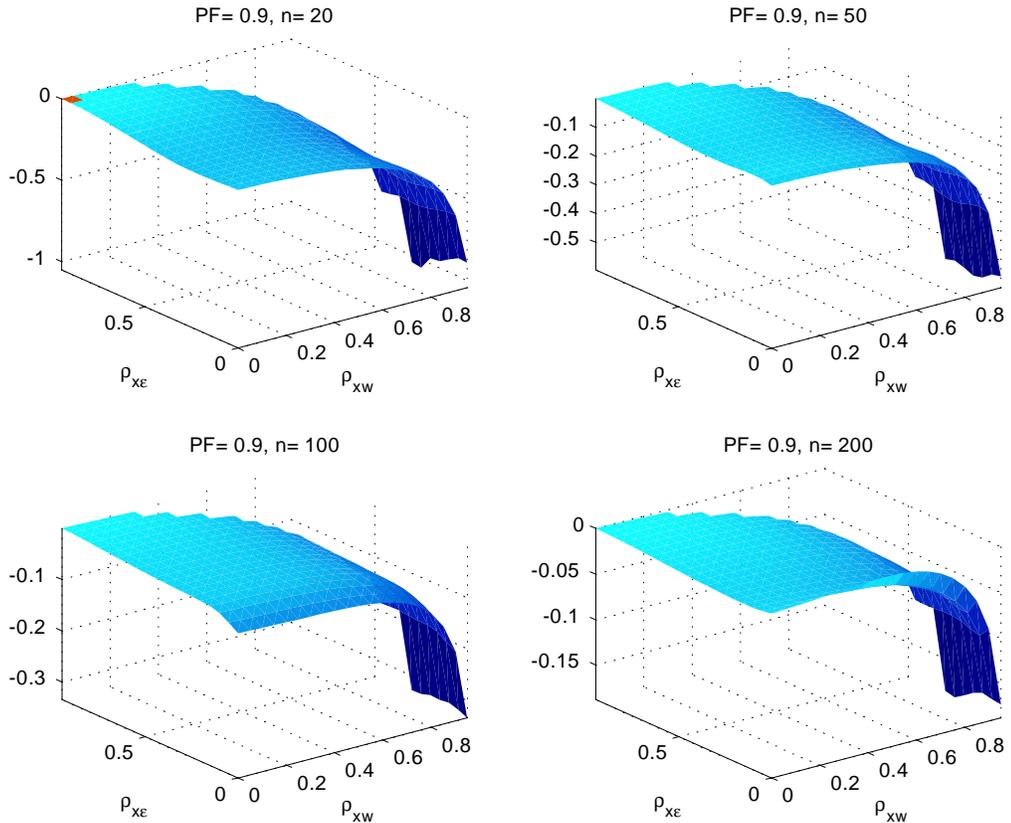


Figure 7:  $\hat{\beta}_{OLS}$  and  $\hat{\beta}_{IV}$  in dynamic model,  $n = 100$ ,  $PF = 0.95$ ,  $\gamma = 0.5$

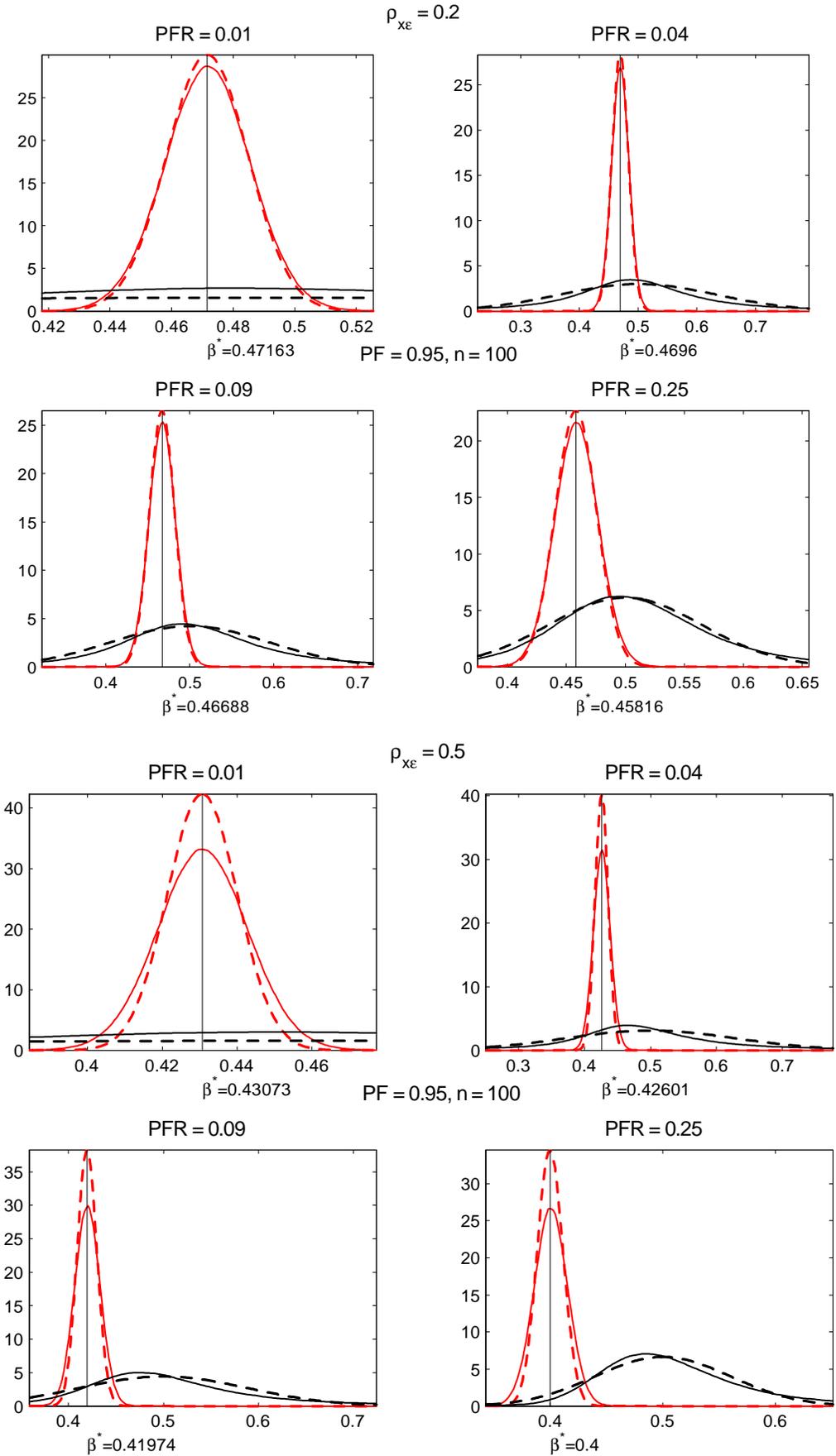


Figure 8:  $\hat{\gamma}_{OLS}$  and  $\hat{\gamma}_{IV}$  in dynamic model,  $n = 100$ ,  $PF = 0.95$ ,  $\gamma = 0.5$

