Viewing the Relative Efficiency of IV Estimators in Models with Lagged and Instantaneous Feedbacks

Agnes S. Joseph
Jan F. Kiviet*

Faculty of Economics and Econometrics, Universiteit van Amsterdam.

* Tinbergen Institute.

Viewing the Relative Efficiency of IV Estimators in Models with Lagged and Instantaneous Feedbacks

Agnes S. Joseph   and   **Jan F. Kiviet**[*]
Tinbergen Institute, University of Amsterdam

27 February 2004

JEL-codes: C13,C32,C51

keywords: *asymptotic efficiency comparisons, computational visualization, computer algebra, dynamic simultaneous models, instrument weakness*

Abstract

The asymptotic efficiency of OLS and IV estimators is examined in a simple dynamic structural model with a constant and two explanatory variables: the lagged dependent variable and another autoregressive variable, which may also include lagged or instantaneous feedbacks from the dependent variable. The parameter values are such that all variables are stationary. The asymptotic efficiency of OLS and various IV estimators is expressed via the moments of the data series in the model parameters. Various computational and graphical aids are employed to examine and illustrate the relationships between parameter values, instrument weakness, and estimator efficiency. Symbolic computer algebra and image sequences are used in animations to identify regions in the parameter space where consistent IV estimators may be less efficient than inconsistent OLS estimators, upon comparing the asymptotic approximation to their mean squared errors.

1. Introduction

Recent attention has been paid to the effects of the correlation between endogenous explanatory variables and the instruments on instrumental-variable estimation. When this correlation is weak, the finite-sample bias of consistent instrumental-variable (IV) estimators may be as serious as the finite-sample bias of inconsistent ordinary least squares (OLS) estimators. Particular forms of instrument weakness may even jeopardize the standard results on the limiting distribution of IV estimators and their associated test

---

[*] corresponding author: Department of Quantitative Economics, Faculty of Economics and Econometrics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; tel +31.20.5254224, fax +31.20.5254349, email J.F.Kiviet@UvA.NL

statistics. Renewed attention to these problems has been triggered especially by Bound et al. (1995). An overview of many recent contributions to this research is given in Dufour (2003). While this literature applies mostly to simple linear static simultaneous models, we examine here particular aspects in the context of stable dynamic and possibly simultaneous models for stationary variables. Although such models, and non-stationary generalizations, are widely applied in econometrics, the kind of knowledge on the relative efficiency of OLS and method of moments (MM) estimators as provided here is not available.

In a classic static simultaneous-equations model, one can distinguish parametric identification, which is completely determined by the specification of the model irrespective of the actual sample observations, and the effective identification (see Belsley, 1992). For the coefficients of a particular equation, the effective identification is not only determined by the actual values of the parameters, including the relative magnitude of the disturbance variances in the reduced form, but also by the data. Especially the collinearity of the instruments, i.e. both the included and the excluded exogenous variables, is very important. In a static model, the degree of instrument collinearity is not related to the actual values of the structural parameters. However, in the context of the dynamic model under study in this paper, it is quite common that instruments are, or are dependent on, lagged endogenous variables. Collinearity and fit in the reduced form, and hence weakness or strength of the instruments, are then directly determined by the structure and actual parameter values of the full dynamic system. Hence, in dynamic simultaneous models all aspects of effective identification are very much intertwined, which complicates their analysis. We shall use modern computational techniques, such as symbolic algebra and graphical animations, to produce a view on the regions in the parameter space where weak instrument and identification problems appear to occur in simple variants of a particular dynamic simultaneous model. We do that by exploiting and comparing the computational, graphical, and symbolic algebra powers of various widely used software packages, such as Scientific Work Place, Gauss, and Mathematica.

To obtain efficiency measures for OLS and MM estimators, we focus here exclusively on asymptotic results, although in the literature on weak instruments various efforts have been made to develop measures on instrument adequacy that are supposed to be informative of bias and efficiency of IV estimators in finite samples. These include alternative (partial) R^2 measures in reduced-form regressions and the concentration parameter, see for instance, Bound et al. (1995), Shea (1997), Godfrey (1999) and Poskitt and Skeels (2002). Of course, in the end it is the actual behavior of the estimators and tests in finite samples that matters. Therefore, we do not focus here primarily on possible causes of poor efficiency such as instrument weakness, ineffective identification and collinearity, but examine directly the location and dispersion of the estimators. Due to the dynamics in our basic model, the instrumental variables are generally not strongly, but only weakly, exogenous. Therefore their complex stochastic nature makes it very difficult to derive exact finite-sample results or approximations of greater accuracy than those obtained by the present first-order large-sample asymptotic analysis.

When samples are not too small, first-order asymptotic results usually convey very useful general information on properties of estimators. In particular, estimators with relatively poor asymptotic efficiency probably behave poorly in finite samples too. Hence, the present analysis of asymptotic efficiency allows initial screening of various estimation methods for particular types of models. This may provide clues on how to design more sophisticated asymptotic approaches for dynamic models, such as those developed for

static simultaneous models by the local-to-zero approach of Staiger and Stock (1997) and Wang and Zivot (1998). By assuming stationarity, the asymptotic distributions of the estimators are relatively simple and, additionally, the initial conditions (start-up values) do not affect the asymptotic efficiency of the estimators. Since in a fully specified dynamic system estimator efficiency can – in principle – fully be expressed in the model parameters, its evaluation straightforwardly discloses the regions in the parameter space that yield poor inference. In future work we will perform a similar analysis for finite samples using simulation methods to check the inaccuracies of the present asymptotic analysis in relation to actual sample size.

The structure of this study is as follows. In Section 2 we review some well-known results on the asymptotic efficiency of consistent method of moment estimators. Moreover, we develop a first-order asymptotic approximation to the mean squared error (AMSE) of possibly inconsistent estimators. Next in Section 3 the general form of the particular dynamic system to be examined is introduced. The efficiency of its estimators will be determined by six parameters, and we discuss some empirically relevant constraints on the parameter space. Section 4 produces results for relatively simple reduced-form type equations. In a series of subsections, we consider simple autoregressive models, autoregressive models with an additional strongly exogenous regressor, and a dynamic reduced-form equation with weakly exogenous regressors. Here the complexity is such that results on the loss in efficiency when IV estimators are used instead of the optimal OLS estimator can still be obtained by combining simple algebraic analytical derivations with traditional numerical computation. In Section 5, we consider the estimation of an equation from a stylized system involving both instantaneous feedbacks (simultaneity) and lagged feedbacks (dynamics). Now comparisons of the AMSE of consistent IV and inconsistent OLS can only be obtained by using computer algebra, and for interpreting the high-dimensional numerical results, animated graphics proves to be very effective. Section 6 concludes.

2. Asymptotic efficiency of LS and MM estimators

We consider the efficiency of various method of moments estimators of the parameters of interest ϕ in the linear model for the $T \times 1$ dependent variable y given by

$$y = W\phi + \varepsilon, \tag{2.1}$$

where $\varepsilon_t \sim \mathsf{iid}(0, \sigma_\varepsilon^2)$ and the $T \times K$ matrix $W = (w_1, ..., w_T)'$ contains the explanatory variables. The OLS estimator of ϕ is $\hat{\phi}_{\mathrm{OLS}} = (W'W)^{-1}W'y$. Assuming that

$$\operatorname*{plim}_{T \to \infty} \frac{1}{T} W'W = \Sigma_{WW} \tag{2.2}$$

exists (the variables are stationary) with $\operatorname{rank}(\Sigma_{WW}) = K$, and also assuming $\mathsf{E}(w_t \varepsilon_t) = 0$ for $t = 1, ..., T$, estimator $\hat{\phi}_{\mathrm{OLS}}$ is consistent, and under mild regularity $\sqrt{T}(\hat{\phi}_{\mathrm{OLS}} - \phi)$ has a limiting normal distribution with asymptotic variance

$$\mathsf{AV}(\hat{\phi}_{\mathrm{OLS}}) = \sigma_\varepsilon^2 \Sigma_{WW}^{-1}. \tag{2.3}$$

If $\mathsf{E}(w_t \varepsilon_t) \neq 0$ then $\hat{\phi}_{\mathrm{OLS}}$ is inconsistent and (2.3) invalid.

Suppose $Z = (z_1, ..., z_T)'$ is a $T \times L$ matrix of valid instrumental variables, i.e., $\mathsf{E}(z_t \varepsilon_t) = 0$ for all t, and when

$$\operatorname*{plim}_{T \to \infty} \frac{1}{T} Z'W = \Sigma_{ZW}, \ \operatorname*{plim}_{T \to \infty} \frac{1}{T} Z'Z = \Sigma_{ZZ} \tag{2.4}$$

exist with $\operatorname{rank}(\Sigma_{ZW}) = K$ and $\operatorname{rank}(\Sigma_{ZZ}) = L$, then consistent and asymptotically normal MM estimators of ϕ are as follows: if $L = K$, $\hat{\phi}_{\mathrm{IV}} = (Z'W)^{-1}Z'y$ with asymptotic variance

$$\mathsf{AV}(\hat{\phi}_{\mathrm{IV}}) = \sigma_\varepsilon^2 \Sigma_{ZW}^{-1} \Sigma_{ZZ} (\Sigma_{ZW}^{-1})', \tag{2.5}$$

and if $L > K$, $\hat{\phi}_{\mathrm{GIV}} = [W'Z(Z'Z)^{-1}Z'W]^{-1}W'Z(Z'Z)^{-1}Z'y$ with asymptotic variance

$$\mathsf{AV}(\hat{\phi}_{\mathrm{GIV}}) = \sigma_\varepsilon^2 (\Sigma_{ZW}' \Sigma_{ZZ}^{-1} \Sigma_{ZW})^{-1}. \tag{2.6}$$

If $\mathsf{E}(z_t \varepsilon_t) \neq 0$ some of the instruments are invalid, consequently $\hat{\phi}_{\mathrm{IV}}$ and $\hat{\phi}_{\mathrm{GIV}}$ are inconsistent and (2.5) and (2.6) invalid. If $\operatorname{rank}(Z'W) < K$ then the instrument set is deficient and $\hat{\phi}_{\mathrm{IV}}$ and $\hat{\phi}_{\mathrm{GIV}}$ do not exist.

2.1. Efficiency comparisons of consistent estimators

When the instruments are all valid and $\operatorname{rank}(Z'W) = K$ some of the instruments may nevertheless be very weak. As a consequence, some of the diagonal elements of $\mathsf{AV}(\hat{\phi}_{\mathrm{IV}})$ or $\mathsf{AV}(\hat{\phi}_{\mathrm{GIV}})$ may become relatively large. In what follows, we examine and compare these diagonal elements for particular models and, provided $\hat{\phi}_{\mathrm{OLS}}$ is consistent, also make comparisons with $\mathsf{AV}(\hat{\phi}_{\mathrm{OLS}})$.

Occasionally it will be useful to depict the intrinsic efficiency of a particular estimator. A scale-free measure of variability is the coefficient of variation, which equals the ratio of the standard deviation to the actual coefficient. In the same vein, we consider such a ratio here. Let $\hat{\theta}$ be some estimator of a scalar parameter θ. Upon replacing its standard deviation by $\sqrt{\mathsf{AV}(\hat{\theta})/T}$, i.e., the asymptotic approximation to its standard error, we obtain the asymptotic variation coefficient

$$\tau_T^{-1}(\hat{\theta}) = \frac{1}{\theta}\sqrt{\frac{\mathsf{AV}(\hat{\theta})}{T}}. \tag{2.7}$$

Note that $\tau_T^{-1}(\hat{\theta})$ is the inverse of the asymptotic t-ratio for testing $\theta = 0$ evaluated at the true value θ, with omission of a degrees of freedom correction. When $\left|\tau_T^{-1}(\hat{\theta})\right| > 1$ the inefficiency is so serious that the approximated standard deviation is larger than the actual coefficient value. When $\left|\tau_T^{-1}(\hat{\theta})\right| < 0.5$ the absolute value of the true parameter value is larger than twice its standard deviation. Self-evidently, $\tau_T^{-1}(\hat{\theta})$ will be large for θ close to zero, and estimator inefficiency is more serious for smaller sample sizes, although note that in this analysis we neglect (again) any possible deviations of the actual variance $\mathsf{V}(\hat{\theta})$ in finite sample from its first-order asymptotic approximation $\mathsf{AV}(\hat{\theta})/T$. For models where $\mathsf{V}(\hat{\theta})$ can be derived in finite sample and $\theta \neq 0$ Belsley (1982) develops a test for the parameter $\tau_T(\hat{\theta}) = \theta/\mathsf{V}(\hat{\theta})$ in order to detect 'data weakness', which hinges upon both collinearity and high disturbance variance. In Belsley (1992) a generalization of such a test is discussed for static simultaneous models to assess 'the degree of effective identification'.

2.2. Efficiency of inconsistent OLS

The first-order asymptotic efficiency of inconsistent OLS can be expressed as follows. Let

$$W = \bar{W} + \varepsilon\psi', \tag{2.8}$$

with

$$\mathsf{E}(\bar{W}'\varepsilon) = 0 \tag{2.9}$$

and ψ a K-vector which parameterizes the joint-dependence of regressors and disturbances, since

$$\mathsf{E}(W'\varepsilon) = \mathsf{E}(\psi\varepsilon'\varepsilon) = \sigma_\varepsilon^2 T\psi. \tag{2.10}$$

We define

$$\left.\begin{array}{l} \Sigma_{W\varepsilon} \equiv \mathrm{plim}_{T\to\infty} \frac{1}{T}W'\varepsilon = \sigma_\varepsilon^2\psi \\[2mm] \Sigma_{\bar{W}\bar{W}} \equiv \mathrm{plim}_{T\to\infty} \frac{1}{T}\bar{W}'\bar{W} \\[2mm] \Sigma_{WW} \equiv \mathrm{plim}_{T\to\infty} \frac{1}{T}W'W = \Sigma_{\bar{W}\bar{W}} + \sigma_\varepsilon^2\psi\psi'. \end{array}\right\} \tag{2.11}$$

Upon assuming that the skewness coefficient of the disturbances is zero and that the kurtosis is three (as it is under normality), i.e.

$$\mathsf{E}(\varepsilon_t^3) = 0 \quad \text{and} \quad \mathsf{E}(\varepsilon_t^4) = 3\sigma_\varepsilon^4, \quad t = 1, ..., T \tag{2.12}$$

we obtain for the variance of $W'\varepsilon$

$$\begin{aligned} \mathsf{V}(W'\varepsilon) &= \mathsf{E}(W'\varepsilon\varepsilon'W) - \mathsf{E}(W'\varepsilon)[\mathsf{E}(W'\varepsilon)]' \\ &= \mathsf{E}(\bar{W}'\varepsilon\varepsilon'\bar{W}) + \mathsf{E}(\psi\varepsilon'\varepsilon\varepsilon'\varepsilon\psi') - \sigma_\varepsilon^4 T^2\psi\psi' \\ &= \sigma_\varepsilon^2\mathsf{E}(\bar{W}'\bar{W}) + 2\sigma_\varepsilon^4 T\psi\psi', \end{aligned} \tag{2.13}$$

so that we have

$$\frac{1}{\sqrt{T}}[W'\varepsilon - \mathsf{E}(W'\varepsilon)] \to \mathsf{N}(0, \sigma_\varepsilon^2\Sigma_{\bar{W}\bar{W}} + 2\sigma_\varepsilon^4\psi\psi').$$

Note that the variance of this limiting distribution can be written as $\sigma_\varepsilon^2\Sigma_{WW} + \sigma_\varepsilon^4\psi\psi'$. Thus,

$$\sqrt{T}(W'W)^{-1}[W'\varepsilon - \mathsf{E}(W'\varepsilon)] \to \mathsf{N}(0, \sigma_\varepsilon^2\Sigma_{WW}^{-1} + \sigma_\varepsilon^4\Sigma_{WW}^{-1}\psi\psi'\Sigma_{WW}^{-1}), \tag{2.14}$$

which we will employ as follows: For the inconsistency of the OLS estimator we have

$$\begin{aligned} \mathrm{plim}_{T\to\infty}(\hat{\phi}_{\mathrm{OLS}} - \phi) &= \mathrm{plim}_{T\to\infty}[(W'W)^{-1}W'y - \phi] \\ &= \mathrm{plim}_{T\to\infty}[(W'W)^{-1}W'\varepsilon] \\ &= \Sigma_{WW}^{-1}\Sigma_{W\varepsilon} \\ &= \sigma_\varepsilon^2\Sigma_{WW}^{-1}\psi. \end{aligned} \tag{2.15}$$

Denoting this inconsistency by ϕ^*, defined as

$$\phi^* \equiv \sigma_\varepsilon^2\Sigma_{WW}^{-1}\psi, \tag{2.16}$$

we have

$$\begin{aligned}
\sqrt{T}(\hat{\phi}_{\text{OLS}} - \phi - \phi^*) &= \sqrt{T}[(W'W)^{-1}W'y - \phi - \phi^*] \\
&= (\frac{1}{T}W'W)^{-1}\frac{1}{\sqrt{T}}\left[W'\varepsilon - \sigma_\varepsilon^2 T\psi\right] \\
&= (\frac{1}{T}W'W)^{-1}\frac{1}{\sqrt{T}}[W'\varepsilon - \mathsf{E}(W'\varepsilon)].
\end{aligned}$$

Thus

$$\sqrt{T}(\hat{\phi}_{\text{OLS}} - \phi - \phi^*) \to \mathsf{N}(0, \sigma_\varepsilon^2\Sigma_{WW}^{-1} + \sigma_\varepsilon^4\Sigma_{WW}^{-1}\psi\psi'\Sigma_{WW}^{-1}). \tag{2.17}$$

Assuming that it is possible to remove the inconsistency of the OLS estimator, which we will call COLS (Consistent or Corrected OLS), i.e., $\hat{\phi}_{\text{COLS}} \equiv \hat{\phi}_{\text{OLS}} - \phi^*$, this would yield asymptotic variance

$$\begin{aligned}
\mathsf{AV}(\hat{\phi}_{\text{COLS}}) &= \sigma_\varepsilon^2\Sigma_{WW}^{-1} + \sigma_\varepsilon^4\Sigma_{WW}^{-1}\psi\psi'\Sigma_{WW}^{-1} \tag{2.18}\\
&= \sigma_\varepsilon^2\Sigma_{WW}^{-1} + \phi^*\phi^{*\prime}.
\end{aligned}$$

We can use $\mathsf{AV}(\hat{\phi}_{\text{COLS}})$ as a yardstick to measure the relative (in)efficiency of IV estimators, although a drawback is that $\hat{\phi}_{\text{COLS}}$ is not an operational estimator. However, the above analysis can also be used to obtain an efficiency measure for the inconsistent OLS estimator. Above, the asymptotic variance of an estimator was taken to be the variance of its limiting distribution. Let us now also consider the usual first-order asymptotic approximation of the variance in finite samples, denoted here as AAV. We have

$$\mathsf{AAV}(\hat{\phi}_{\text{IV}}) = T^{-1}\sigma_\varepsilon^2\Sigma_{ZW}^{-1}\Sigma_{ZZ}(\Sigma_{ZW}^{-1})', \tag{2.19}$$

and

$$\mathsf{AAV}(\hat{\phi}_{\text{COLS}}) = T^{-1}(\sigma_\varepsilon^2\Sigma_{WW}^{-1} + \phi^*\phi^{*\prime}). \tag{2.20}$$

Then, following the usual concept of mean-squared-error (MSE), which is variance plus outer-product of bias, we may define the first-order asymptotic approximation to the efficiency (denoted here as AMSE) of the inconsistent OLS estimator in a sample of size T as the sum of the first-order approximation of its variance, equal to $\mathsf{AAV}(\hat{\phi}_{\text{COLS}})$, plus the outer-product of the first-order asymptotic approximation to its bias. The latter is given by the inconsistency ϕ^*. This yields

$$\begin{aligned}
\mathsf{AMSE}(\hat{\phi}_{\text{OLS}}) &= T^{-1}(\sigma_\varepsilon^2\Sigma_{WW}^{-1} + \phi^*\phi^{*\prime}) + \phi^*\phi^{*\prime} \tag{2.21}\\
&= T^{-1}\sigma_\varepsilon^2\Sigma_{WW}^{-1} + \frac{T+1}{T}\phi^*\phi^{*\prime}.
\end{aligned}$$

Note that for a consistent estimator AAV equals AMSE. Also note that for comparing the efficiencies of IV and COLS, which are both consistent, it does not matter whether we compare the respective AV or AAV expressions. However, when a comparison is made involving the inconsistent OLS estimator, then the sample size T matters and we should use the ratio of diagonal elements of $\mathsf{AAV}(\hat{\phi}_{\text{IV}}) = \mathsf{AMSE}(\hat{\phi}_{\text{IV}})$ and $\mathsf{AMSE}(\hat{\phi}_{\text{OLS}})$. Note that $\mathsf{AMSE}(\hat{\phi}_{\text{IV}}) = O(T^{-1})$ and $\mathsf{AMSE}(\hat{\phi}_{\text{OLS}}) = O(1)$, provided $\phi^* \neq 0$. Therefore, for large enough T inconsistent OLS cannot be more efficient than consistent IV .

3. A simple model with lagged and instantaneous feedbacks

The particular model we focus on is, in its most general form, given by

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \varepsilon_t, \qquad (3.1)$$

where

$$x_t = \delta + \kappa x_{t-1} + \lambda y_{t-1} + \xi_t, \qquad (3.2)$$

with

$$\left. \begin{array}{l} \mathsf{E}(\varepsilon_t) = 0, \qquad \mathsf{E}(\xi_t) = 0, \\[2mm] \mathsf{V}(\varepsilon_t) = \sigma_\varepsilon^2, \quad \mathsf{V}(\xi_t) = \sigma_\xi^2, \quad \mathsf{E}(\varepsilon_t \xi_t) = \rho \sigma_\varepsilon \sigma_\xi, \\[2mm] \mathsf{E}(\varepsilon_t \varepsilon_s) = 0, \quad \mathsf{E}(\xi_t \xi_s) = 0, \quad \mathsf{E}(\varepsilon_t \xi_s) = 0. \end{array} \right\} \qquad t \neq s = 1, ..., T. \qquad (3.3)$$

We have two observed time-series y_t and x_t, which are serially interdependent and also jointly dependent when $\rho \neq 0$. The two parameters of primary interest are β and γ, but there are nine parameters in total.

To ensure stationarity we examine the VAR representation of the model, which is

$$\left(\begin{array}{c} y_t \\ x_t \end{array} \right) = \left(\begin{array}{c} \alpha + \beta\delta \\ \delta \end{array} \right) + \left(\begin{array}{cc} \gamma + \beta\lambda & \beta\kappa \\ \lambda & \kappa \end{array} \right) \left(\begin{array}{c} y_{t-1} \\ x_{t-1} \end{array} \right) + \left(\begin{array}{c} \beta\xi_t + \varepsilon_t \\ \xi_t \end{array} \right). \qquad (3.4)$$

From this, and upon denoting the lag-operator as L, we establish the final form

$$[1 - (\gamma + \kappa + \beta\lambda)L + \kappa\gamma L^2] \left(\begin{array}{c} y_t \\ x_t \end{array} \right) = \left(\begin{array}{c} (1 - \kappa)\alpha + \beta\delta \\ \lambda\alpha + (1 - \gamma)\delta \end{array} \right) + \left(\begin{array}{cc} 1 - \kappa L & \beta \\ \lambda L & 1 - \gamma L \end{array} \right) \left(\begin{array}{c} \varepsilon_t \\ \xi_t \end{array} \right). \qquad (3.5)$$

Stationarity of the AR(2) parameters implies the restrictions

$$\left. \begin{array}{l} |\kappa\gamma| < 1, \\[3mm] |\gamma + \kappa + \beta\lambda| < 1 + \kappa\gamma. \end{array} \right\} \qquad (3.6)$$

The domain of the nine parameters can be constrained further, since:

1) Obvious restrictions on the parameter space are $\sigma_\varepsilon^2 > 0$, $\sigma_\xi^2 > 0$ and $|\rho| < 1$. In fact, we can make σ_ε the numeraire, and take $\sigma_\varepsilon = 1$;

2) The values of α and δ are of no concern, because all efficiency measures of the estimators for β and γ are found to be invariant with respect to them;

3) Limiting the examination to models where the series $\{y_t\}$ and $\{x_t\}$ are of the same order of integration and, in the present paper, stationary, yields (3.6). Moreover, when $\lambda = 0$ we have $|\gamma| < 1$ and $|\kappa| < 1$;

4) The parameter β is unbounded in principle; without loss of generality, though, we can restrict ourselves to $\beta \geq 0$, because we will find that only the sign of $\beta\lambda$ matters. In practice y_t and x_t are often measured in logs of levels (or perhaps of first differences) and then β and $\beta/(1 - \gamma)$ are the impact multiplier and the long run elasticity, respectively, and will, for most empirically relevant situations, not exceed one digit numbers.

Hence, we have to deal with the six constrained parameters γ, β, κ, λ, ρ and σ_ξ^2. However, not all feasible combinations of parameter values are equally relevant. Therefore, we shall often impose further restrictions. From an empirical point of view it seems

reasonable to restrict our attention to parametrizations such that model (3.1) fits reasonably well. In practice, a time-series regression like (3.1) will usually yield an R^2 of 0.9 or higher. Therefore, we shall often focus on parameter combinations which yield a population coefficient of determination ($\Re^2$) equal to values in the range, say 0.7 through 0.99, where

$$\Re^2 = \frac{\mathsf{V}(\beta x_t + \gamma y_{t-1})}{\mathsf{V}(y_t)}. \tag{3.7}$$

Interesting special cases of the above model from an econometric-theory point of view are (combinations of) $\rho = 0$ (no instantaneous feedback of ε_t on x_t), $\lambda = 0$ (no lagged feedback of y_{t-1} on x_t), $\beta = 0$ (y_t is a pure autoregression). If $\rho = 0$, (3.1) and (3.2) establish a two-dimensional triangular system in which simultaneity of y_t and x_t is ruled out. In that situation, x_t is weakly-exogenous for inference on β and γ and in fact strongly-exogenous if $\lambda = 0$. Hence, when $\rho = 0$, efficient inference on β and γ obtains from applying OLS to (3.1). But, if $\rho \neq 0$, y_t and x_t are jointly dependent and OLS estimates of (3.1) are inconsistent. However, because $\mathsf{E}(y_{t-i}\varepsilon_t) = 0$ and $\mathsf{E}(x_{t-i}\varepsilon_t) = 0$ for $i > 0$, irrespective of the value of ρ, valid instrumental variables in the form of lagged variables are available. Thus, consistent estimators can be obtained by using MM estimators. Their efficiency will depend on the collinearity of x_t and y_{t-1} and on the quality of the instruments. When using (x_{t-1}, y_{t-1}) as instruments for model (3.1) the instrument quality is determined by their explanatory power for x_t. From (3.4) it is obvious that γ and κ both close to zero will lead to a weak instrument problem. In this study, we will not just focus on instrument weakness, but on the absolute and relative efficiency of various estimators.

Using only current exogenous and lagged endogenous and exogenous variables as instruments implies that the asymptotic efficiency is fully determined by the structure and actual parameters of the full dynamic system under study. A similar situation, where only 'internal' instruments are exploited, occurs in the estimation by GMM of dynamic panel-data models with unobserved individual effects. Blundell and Bond (1998) discusses the incidence of instrument weakness for certain numerical parameter values of a particular dynamic panel data model. In a pure time-series context there does not seem to be general knowledge about the relative efficiency of least-squares and method-of-moments estimators for models of the partial-adjustment type given above expressed in terms of their parameter values, not, for instance, in Bowden and Turkington (1984) or in more recent monographs or textbooks (see also West and Wilcox, 1996), although generalizations of this type of specification are widely applied in econometrics. Many macroeconomic systems are of this type, see for instance Hendry (1995) and Favero (2001). Also financial econometrics applications often require specifications of this type, see for instance Ferson and Foerster (1994) and Evans and Lyons (2002). In the latter application, parameter values of γ, β and κ seem relevant that differ widely from those of, say, a macro consumption or wage equation.

Below we make comparisons between asymptotic efficiency of OLS and IV in a series of models of increasing complexity. As will become clear, to obtain results on the critical parameter value combinations that lead to efficiency problems, computational aids of increasing sophistication are called for.

4. Models with lagged but no instantaneous feedbacks

In this section we consider variants of model (3.1) with (3.2) where $\rho = 0$. In the various subsections we focus on the cases $\beta = 0$, $\lambda = 0$, and the unrestricted case, respectively.

4.1. The simple $\mathrm{AR}(1)$ model

When β is known to be zero the relationship (3.1) is a simple AR(1) model. Under full stationarity, which requires $|\gamma| < 1$, we have

$$\left.\begin{array}{l} \mathsf{E}(y_t) = \frac{\alpha}{1-\gamma}, \\[2mm] \mathsf{V}(y_t) = \frac{1}{1-\gamma^2}\sigma_\varepsilon^2, \\[2mm] \mathsf{C}(y_t, y_s) = \mathsf{E}\{[(y_t - \mathsf{E}(y_t)][(y_s - \mathsf{E}(y_s)]\} = \frac{\gamma^{|t-s|}}{1-\gamma^2}\sigma_\varepsilon^2. \end{array}\right\} \tag{4.1}$$

Hence, in this particular case, we find for (2.2)

$$\Sigma_{WW} = \left(\begin{array}{cc} 1 & \mathsf{E}(y_t) \\ \mathsf{E}(y_t) & \mathsf{V}(y_t) + [\mathsf{E}(y_t)]^2 \end{array} \right), \tag{4.2}$$

and, since $\mathsf{E}(y_{t-1}\varepsilon_t) = 0$, it follows from (2.3) that

$$\begin{array}{rcl} \mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}}) & = & \dfrac{\sigma_\varepsilon^2}{\mathsf{V}(y_t)} \\[4mm] & = & 1 - \gamma^2. \end{array} \tag{4.3}$$

This indicates that the OLS efficiency improves for γ further away from zero.

If this model is estimated using not y_{t-1} itself (as in OLS) but y_{t-2} as an instrumental variable (next to the constant) then Σ_{ZZ} equals (4.2), whereas

$$\Sigma_{ZW} = \left(\begin{array}{cc} 1 & \mathsf{E}(y_t) \\ \mathsf{E}(y_t) & \mathsf{C}(y_t, y_{t-1}) + [\mathsf{E}(y_t)]^2 \end{array} \right). \tag{4.4}$$
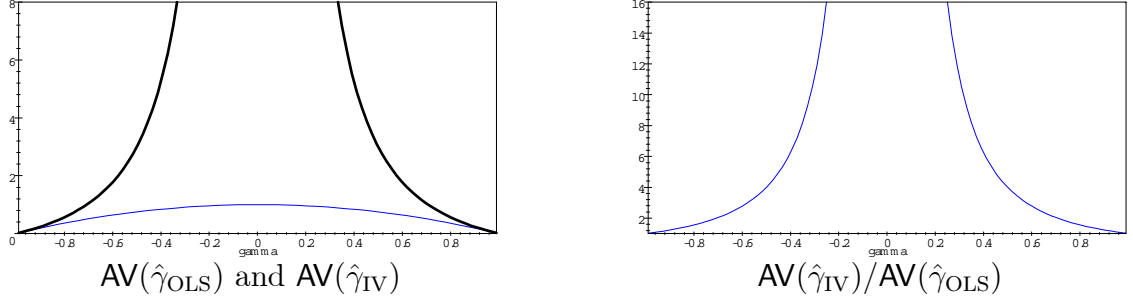
It follows from (2.5) that

$$\begin{array}{rcl} \mathsf{AV}(\hat{\gamma}_{\mathrm{IV}}) & = & \sigma_\varepsilon^2 \dfrac{\mathsf{V}(y_t)}{[\mathsf{C}(y_t, y_{t-1})]^2} \\[4mm] & = & \dfrac{1 - \gamma^2}{\gamma^2}. \end{array} \tag{4.5}$$

This indicates an estimation problem at $\gamma = 0$ caused by lack of correlation between regressor and instrument. For an AR(1) model with $\gamma = 0$, we have $\mathsf{C}(y_t, y_{t-1}) = 0$; then the matrix Σ_{ZW} of (4.4) does not have full rank. Although the instrument is still valid, since $\mathsf{E}(y_{t-2}\varepsilon_t) = 0$, it is unfit when $\gamma = 0$ and weak for γ close to zero.

Note that for any $|\gamma| < 1$, this IV estimator suffers an efficiency loss with respect to OLS, because $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}}) > \mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})$. For both OLS and IV, we find that the absolute efficiency improves for γ further away from zero and in fact becomes the same for $|\gamma| \uparrow 1$. Note that the efficiencies of both $\hat{\gamma}_{\mathrm{OLS}}$ and $\hat{\gamma}_{\mathrm{IV}}$ in the AR(1) model are invariant with respect to the intercept α and the error variance σ_ε^2. These results are all apparent from

Figure 4.1, which clearly shows that, in this model, the efficiency of IV is relatively poor over a large portion of the parameter space (a portion much wider than just the point $\gamma = 0$ where the instrument is unfit).

Figure 4.1: Asymptotic variance in AR(1) model

$\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})$ and $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}})$ $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}})/\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})$

The relative difference in asymptotic efficiency between these two estimators is best characterized by the ratio

$$\frac{\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}})}{\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})} = \gamma^{-2}. \tag{4.6}$$

This is also shown in Figure 4.1. For this model the asymptotic variance of IV is $100\gamma^{-2}$ per cent of the asymptotic variance of OLS. Hence, the IV asymptotic variance is at least double that of OLS if $|\gamma| \leq 0.707$. If we instrument y_{t-1} not by y_{t-2} but by y_{t-3}, efficiency deteriorates by another factor γ^{-2}. When generalized instrumental variables are used (more instruments than regressors) in the AR(1) model, for instance instrumenting y_{t-1} by both y_{t-2} and y_{t-3}, we find no improvement of $\mathsf{AV}(\hat{\gamma}_{\mathrm{GIV}})$ upon $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}})$. This is intuitively obvious, because in fitting the regressor y_{t-1} by the instruments y_{t-2} and y_{t-3}, the latter is a redundant variable in the AR(1) model.

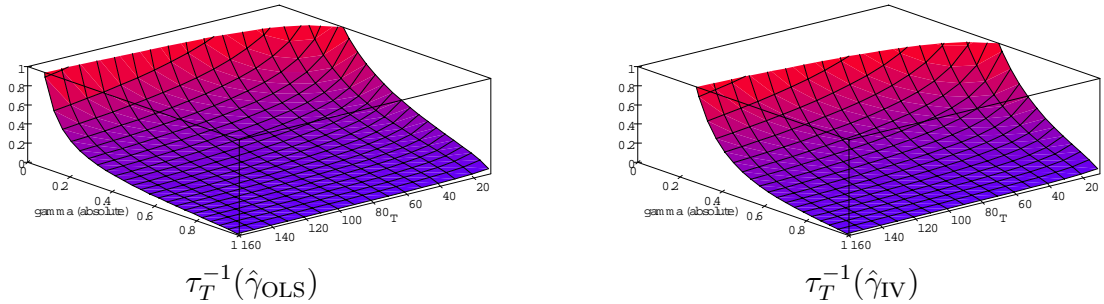Figure 4.2: Asymptotic variation coefficient in AR(1) model

$\tau_T^{-1}(\hat{\gamma}_{\mathrm{OLS}})$ $\tau_T^{-1}(\hat{\gamma}_{\mathrm{IV}})$

Figure 4.2 shows $\tau_T^{-1}(\hat{\gamma}_{\mathrm{IV}})$ and $\tau_T^{-1}(\hat{\gamma}_{\mathrm{IV}})$, defined in (2.7), in the AR(1) model. Note that estimator inefficiency is more serious for smaller sample sizes. At $T = 100$ the asymptotic variation coefficient is larger than one for OLS at $|\gamma| < 0.101$ and for IV when $|\gamma| < 0.308$. It is smaller than 0.5 for OLS when $0.204 < |\gamma| < 1$ and for IV when $0.425 < |\gamma| < 1$.

Figures 4.1.and 4.2 depict a function for which we have an explicit analytical expression available. The formulas enable to exploit directly the graphics facilities of Scientific Work Place for producing the 2-D and 3-D diagrams.

4.2. ARX(1) models with strong exogeneity

The situation gets more interesting but also more complicated when β is unknown. Let us first consider the ARX(1) model $y_t = \alpha + \beta x_t + \gamma y_{t-1} + \varepsilon_t$, where x_t is strongly-exogenous because $\lambda = 0$. Thus $x_t = \delta + \kappa x_{t-1} + \xi_t$ is AR(1), and $\mathsf{E}(\varepsilon_t \xi_t) = 0$ because $\rho = 0$. We now have the four relationships

$$
\left.
\begin{aligned}
\mathsf{V}(y_t) &= \beta^2 \mathsf{V}(x_t) + \gamma^2 \mathsf{V}(y_{t-1}) + 2\beta\gamma \mathsf{C}(x_t, y_{t-1}) + \sigma_\varepsilon^2, \\
\mathsf{V}(x_t) &= (1 - \kappa^2)^{-1} \sigma_\xi^2, \\
\mathsf{C}(x_t, y_{t-1}) &= \kappa \mathsf{C}(x_t, y_t) \\
\mathsf{C}(x_t, y_t) &= \beta \mathsf{V}(x_t) + \gamma \mathsf{C}(x_t, y_{t-1}).
\end{aligned}
\right\}
\tag{4.7}
$$

From the latter two we find

$$
\mathsf{C}(x_t, y_{t-1}) = \frac{\beta\kappa}{1 - \gamma\kappa} \mathsf{V}(x_t),
\tag{4.8}
$$

which, when substituted into the first (and exploiting again the assumed stationarity) gives

$$
(1 - \gamma^2)\mathsf{V}(y_t) = \beta^2 \frac{1 + \gamma\kappa}{1 - \gamma\kappa} \mathsf{V}(x_t) + \sigma_\varepsilon^2.
$$

With the variance of the AR(1) process x_t this yields

$$
\mathsf{V}(y_t) = \frac{1}{1 - \gamma^2} \frac{1 + \gamma\kappa}{1 - \gamma\kappa} \frac{\beta^2}{1 - \kappa^2} \sigma_\xi^2 + \frac{1}{1 - \gamma^2} \sigma_\varepsilon^2.
\tag{4.9}
$$

4.2.1. OLS inference in the simple ARX(1) model

We can now find expressions for the asymptotic variance of the OLS estimators of model (3.1). We obtain from (2.3) and the above

$$
\begin{aligned}
\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}}) &= \sigma_\varepsilon^2 \frac{\mathsf{V}(x_t)}{\mathsf{V}(x_t)\mathsf{V}(y_{t-1}) - [\mathsf{C}(x_t, y_{t-1})]^2} \\
&= \frac{1 - \gamma^2}{1 + \frac{\beta^2}{(1 - \gamma\kappa)^2} \frac{\sigma_\xi^2}{\sigma_\varepsilon^2}}.
\end{aligned}
\tag{4.10}
$$

Note that if $\beta = 0$, i.e., x_t is a redundant exogenous regressor, the asymptotic efficiency (4.10) is similar to that of the estimator in the parsimonious AR(1) model. When $\beta \neq 0$ the presence of the exogenous regressor x_t in the model increases the efficiency of $\hat{\gamma}_{\mathrm{OLS}}$ compared to the AR(1) efficiency (2.3). The asymptotic variance (4.10) depends on the three coefficients γ, β, and κ and on the ratio $\sigma_\xi^2/\sigma_\varepsilon^2$; in fact $\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})$ has only three independent determining factors: γ, κ, and $\beta^2\sigma_\xi^2/\sigma_\varepsilon^2$, because model (3.1) remains observationally equivalent when β and σ_ξ are rescaled inversely.

To make 3-D pictures of the asymptotic variance (4.10), we have to fix one of these three determining factors, preferably one that characterizes a particularly interesting class of model. For this purpose, we find neither γ nor κ or $\beta^2\sigma_\xi^2/\sigma_\varepsilon^2$ very suitable, and would rather fix an empirically more relevant characteristic. For that we chose the population

coefficient of determination $\Re^2$ as defined in (3.7). Note that because $\mathsf{C}(x_t, \varepsilon_t) = 0$ we simply have

$$\Re^2 = 1 - \frac{\sigma_\varepsilon^2}{\mathsf{V}(y_t)}. \tag{4.11}$$

Substituting (4.9), we find

$$\frac{\beta^2 \sigma_\xi^2}{\sigma_\varepsilon^2} = \frac{(1 - \kappa^2)(1 - \gamma\kappa)(\Re^2 - \gamma^2)}{(1 + \gamma\kappa)(1 - \Re^2)}, \tag{4.12}$$

which can only be meaningful when
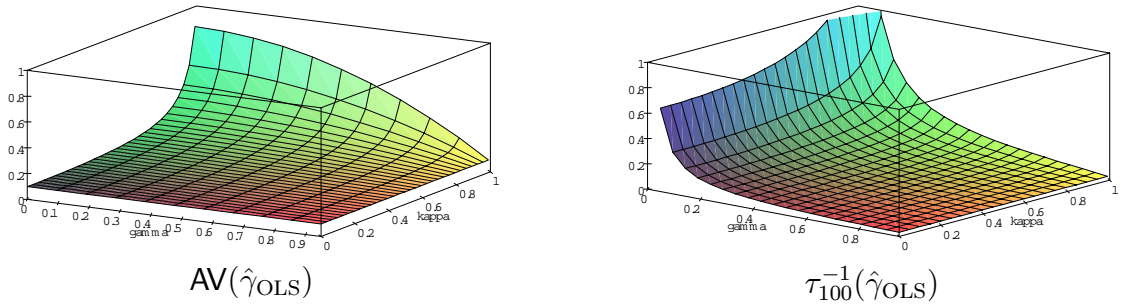
$$0 \leq \gamma^2 \leq \Re^2 < 1. \tag{4.13}$$

Hence, fixing $\Re^2$ constrains the domain of γ. Substituting (4.12) in (4.10) yields

$$\mathsf{AV}(\hat\gamma_{\mathrm{OLS}}) = (1 - \gamma^2\kappa^2)\frac{1 - \Re^2}{1 - \kappa^2\Re^2}. \tag{4.14}$$

This formula shows how $\mathsf{AV}(\hat\gamma_{\mathrm{OLS}})$ depends on γ and κ for fixed $\Re^2$ over (4.13) and $|\kappa| < 1$. Note that for fixed $\Re^2$, the asymptotic efficiency of $\hat\gamma_{\mathrm{OLS}}$ is invariant with respect to the signs of γ and κ, so we have to examine positive values of these only.
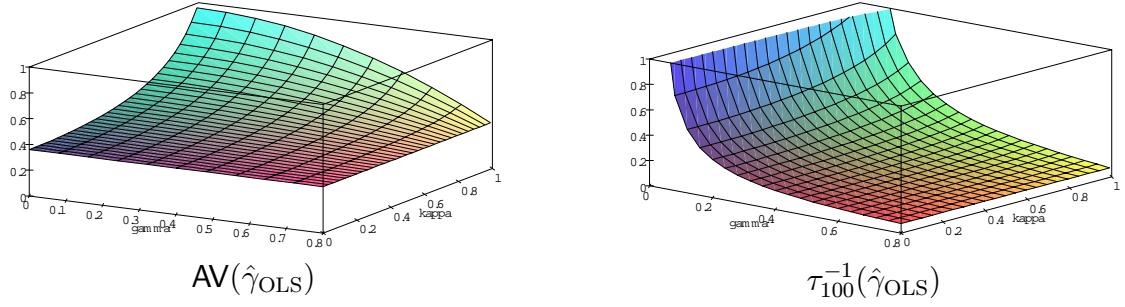
The efficiency of $\hat\gamma_{\mathrm{OLS}}$ is depicted in Figure 4.3 for $\Re^2 = 0.9025$, which implies $|\gamma| \leq 0.95$. Note that the minimal variance is $1 - \Re^2$, and, keeping $\Re^2$ constant, the efficiency is minimal and invariant with respect to γ if $\kappa = 0$, and minimal and invariant with respect to κ if $\gamma^2 = \Re^2$; otherwise the asymptotic variance deteriorates for γ^2 closer to zero and for κ^2 closer to one. Whether the efficiency is reasonable in an absolute sense can be examined again by the asymptotic variation coefficient at particular sample sizes. Figure 4.3 also shows the behaviour of this characteristic for $\Re^2 = 0.9025$ and $T = 100$, and it roughly indicates that useful inference on γ can be obtained if $|\gamma| > 0.2$.

Figure 4.3: ARX(1) model for $\Re^2 = 0.9025$, $|\gamma| \leq 0.95$, $\lambda = 0$

$\mathsf{AV}(\hat\gamma_{\mathrm{OLS}})$ $\tau_{100}^{-1}(\hat\gamma_{\mathrm{OLS}})$

From Figure 4.4, where $\Re^2 = 0.64$, which seems very low for this kind of model, we see that in this type of partial adjustment model accurate inference on γ seems possible for almost all empirically relevant cases, if γ is not too small, irrespective of the value of κ.

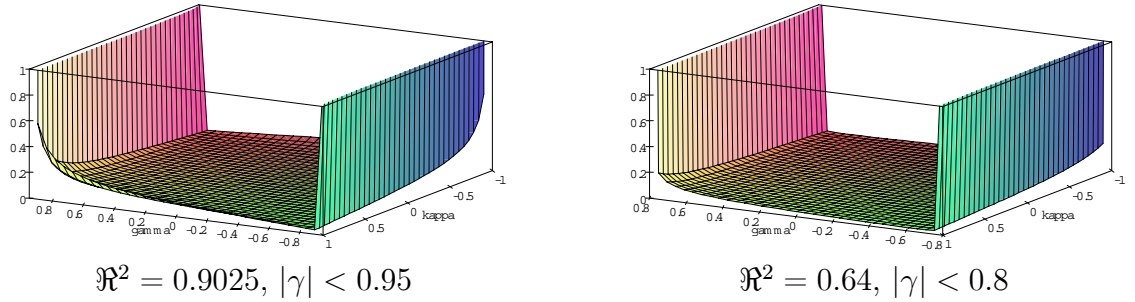Figure 4.4: ARX(1) model for $\Re^2 = 0.64$, $|\gamma| \leq 0.8$, $\lambda = 0$

$\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})$ $\tau_{100}^{-1}(\hat{\gamma}_{\mathrm{OLS}})$

For $\hat{\beta}_{\mathrm{OLS}}$ in the simple ARX(1) model we find

$$
\begin{aligned}
\mathsf{AV}(\hat{\beta}_{\mathrm{OLS}}) &= \sigma_\varepsilon^2 \frac{\mathsf{V}(y_t)}{\mathsf{V}(x_t)\mathsf{V}(y_{t-1}) - [\mathsf{C}(x_t, y_{t-1})]^2} \qquad (4.15) \\
&= \frac{\sigma_\varepsilon^2}{\sigma_\xi^2} \frac{1 - \kappa^2 + \frac{1+\gamma\kappa}{1-\gamma\kappa}\frac{\beta^2\sigma_\xi^2}{\sigma_\varepsilon^2}}{1 + \frac{1}{(1-\gamma\kappa)^2}\frac{\beta^2\sigma_\xi^2}{\sigma_\varepsilon^2}} \\
&= \beta^2 \frac{(1+\gamma\kappa)^2}{\Re^2 - \gamma^2} \frac{1 - \Re^2}{1 - \kappa^2\Re^2}.
\end{aligned}
$$

The sign of $\gamma\kappa$ is relevant now, and there are not three but four determining factors, though again only three for $\tau_T^{-1}(\hat{\beta}_{\mathrm{OLS}})$. Figure 4.5 shows for relevant values of $\Re^2$ that the characteristic $\tau_{100}^{-1}(\hat{\beta}_{\mathrm{OLS}})$ has very moderate values in general, although poor efficiency occurs for γ^2 very close to $\Re^2$, which implies β close to zero and that clarifies that the asymptotic variation coefficient is close to zero here.

Figure 4.5: $\tau_{100}^{-1}(\hat{\beta}_{\mathrm{OLS}})$ in ARX(1) model with $\lambda = 0$

$\Re^2 = 0.9025$, $|\gamma| < 0.95$ $\Re^2 = 0.64$, $|\gamma| < 0.8$

Thus, we find that the asymptotic efficiency of OLS in this model seems pretty good over the relevant part of the parameter space. It is only natural that inference on γ will not be sharp when γ is close to zero (and similarly for β).

4.2.2. IV inference in the simple ARX(1) model using x_t, x_{t-1} as instruments

Next we examine IV estimators in this model. Although we know that the IV efficiency does not exceed that of OLS, because $(\Sigma'_{ZW}\Sigma_{ZZ}^{-1}\Sigma_{ZW})^{-1} - \Sigma_{WW}^{-1}$ is semi positive definite, we want to examine the magnitude of this efficiency loss. There are various options for constructing an instrument set for model (3.1). First we examine using x_t and x_{t-1} as

instruments and denote the resulting estimators by IV_X. From (2.5) we have

$$
\begin{aligned}
\text{AV}(\hat{\gamma}_{\text{IV}_X}) &= \sigma_\varepsilon^2 \frac{[\mathsf{V}(x_t)]^3 - \mathsf{V}(x_t)[\mathsf{C}(x_t, x_{t-1})]^2}{[\mathsf{V}(x_t)\mathsf{C}(x_t, y_t) - \mathsf{C}(x_t, x_{t-1})\mathsf{C}(x_t, y_{t-1})]^2} \\
&= \sigma_\varepsilon^2 \frac{(1 - \gamma\kappa)^2}{\beta^2 \sigma_\xi^2} \\
&= \frac{1 - \gamma^2\kappa^2}{1 - \kappa^2} \frac{1 - \Re^2}{\Re^2 - \gamma^2}.
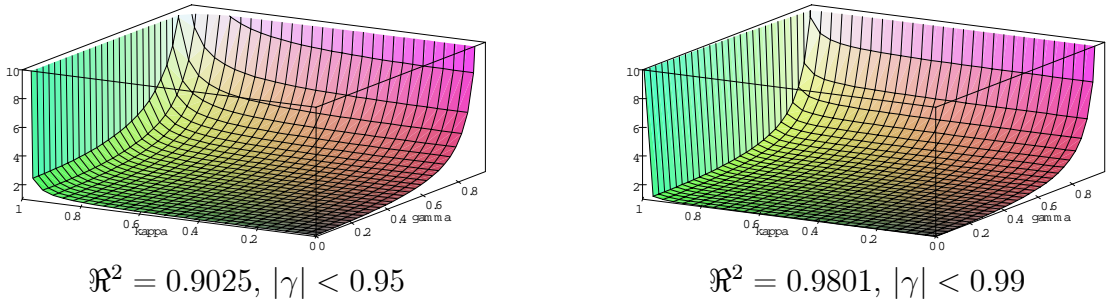\end{aligned}
\tag{4.16}
$$

When the regressor x_t is redundant, $\beta = 0$ which implies $\Re^2 = \gamma^2$. Then Σ_{ZW} does not have full rank, and the instruments, though individually valid, are unfit as a set. Thus, the instruments are weak when β is close to zero.

The (asymptotic) variance inflation factor with respect to OLS due to using the instruments x_t and x_{t-1} is

$$
\frac{\text{AV}(\hat{\gamma}_{\text{IV}_X})}{\text{AV}(\hat{\gamma}_{\text{OLS}})} = \frac{1}{1 - \kappa^2} \frac{1 - \Re^2\kappa^2}{\Re^2 - \gamma^2}.
\tag{4.17}
$$

This is invariant with respect to the signs of γ and κ. From formula (4.17) and Figure 4.6 we see that, for given $\Re^2$, the efficiency loss may be very large for κ^2 very close to one and when $\Re^2 - \gamma^2$ is small. The former leads to an x_t series which, due to its stationarity, tends to a constant, which in the limit would reduce the column rank of the instrument matrix. The latter happens when β^2 is small. This is the weak-instrument situation already indicated, where the processes y_t and x_t tend to independent AR(1) series, and hence x_{t-1} is a valid, but not a very effective, instrument for y_{t-1}. The figures show that a substantial loss of efficiency occurs in models where $\Re^2$ is substantial, not due to the explanatory power of x_t, but primarily due to a high value of γ^2, say, $|\gamma| > 0.8$.

Figure 4.6: $\text{AV}(\hat{\gamma}_{\text{IV}_X})/\text{AV}(\hat{\gamma}_{\text{OLS}})$ in ARX(1) model with $\lambda = 0$

$\Re^2 = 0.9025$, $|\gamma| < 0.95$ \qquad $\Re^2 = 0.9801$, $|\gamma| < 0.99$

With respect to β, we obtain (assuming $\beta \neq 0$)

$$
\begin{aligned}
\text{AV}(\hat{\beta}_{\text{IV}_X}) &= \sigma_\varepsilon^2 \frac{\mathsf{V}(x_t)\,[\mathsf{C}(x_t, y_t)]^2 - 2\mathsf{C}(x_t, x_{t-1})\mathsf{C}(x_t, y_{t-1})\mathsf{C}(x_t, y_t) + \mathsf{V}(x_t)[\mathsf{C}(x_t, y_{t-1})]^2}{[\mathsf{V}(x_t)\mathsf{C}(x_t, y_t) - \mathsf{C}(x_t, x_{t-1})\mathsf{C}(x_t, y_{t-1})]^2} \\
&= \frac{\sigma_\varepsilon^2}{\sigma_\xi^2} \\
&= \frac{\beta^2}{1 - \kappa^2} \frac{1 + \gamma\kappa}{1 - \gamma\kappa} \frac{1 - \Re^2}{\Re^2 - \gamma^2}.
\end{aligned}
\tag{4.18}
$$

In general, this is invariant with respect to γ, β and κ. The intuition behind this at first sight surprising result is as follows: Exploiting results for the inverse of a partitioned
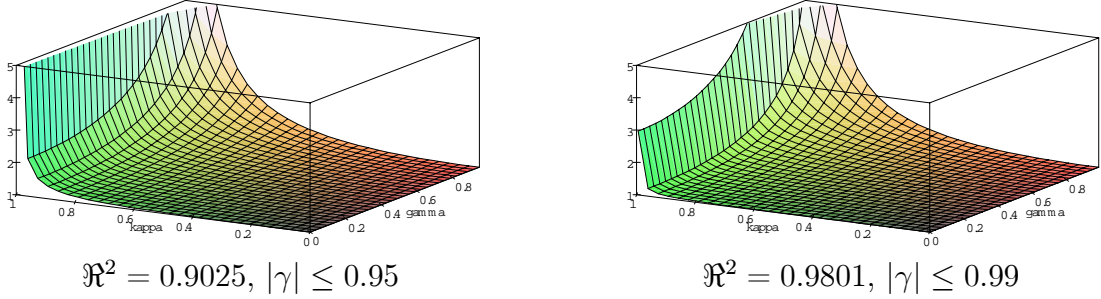
matrix, and taking all variables in deviation from their mean, this asymptotic variance equals the inverse of the sum of squared residuals of the regression of variable $\hat{x}_t$ on $\hat{y}_{t-1}$, multiplied by σ_ε^2 and divided by T. The variables $\hat{x}_t$ and $\hat{y}_{t-1}$ are obtained by fitting the original variables to the instruments x_t and x_{t-1}. Hence, $\hat{x}_t \equiv x_t$ and, due to the AR(1) nature of x_t, $\hat{y}_{t-1}$ is just a multiple of x_{t-1}, asymptotically. Thus, the required residual sum of squares equals simply $\sigma_\xi^2 T$.

However, $\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_X})$ is no longer invariant with respect to γ, β and κ when we fix $\Re^2$. We obtain

$$\frac{\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_X})}{\mathsf{AV}(\hat{\beta}_{\mathrm{OLS}})} = \frac{1}{1-\kappa^2}\frac{1-\Re^2\kappa^2}{1-\gamma^2\kappa^2}, \tag{4.19}$$

which immediately shows that there is no efficiency loss for β with respect to OLS when $\kappa = 0$. From Figure 4.7 we see that the loss seems moderate as long as $|\kappa| < 0.8$, provided that the value of $|\gamma|$ is moderate for large $|\kappa|$.

Figure 4.7: $\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_X})/\mathsf{AV}(\hat{\beta}_{\mathrm{OLS}})$ in ARX(1) model with $\lambda = 0$

$\Re^2 = 0.9025,\ |\gamma| \le 0.95$ $\Re^2 = 0.9801,\ |\gamma| \le 0.99$

We conclude that IV_X can lead to substantial efficiency loss with respect to OLS . This happens for inference on either γ or β when κ^2 is large. Inference on γ also suffers when β^2 is small, which implies γ^2 so large that it is close to $\Re^2$. However, in sharp contrast to the situation found for AR(1) models, in the ARX(1) with strongly exogenous x_t variable the efficiency loss of IV_X with respect to OLS is moderate for moderate values of γ and κ. Note, though, that IV_X uses instruments of a completely different nature than we used in the AR(1).

4.2.3. IV inference in the simple ARX(1) model using x_{t-1}, y_{t-1} as instruments

Next we examine IV estimation when only the lagged variables y_{t-1} and x_{t-1} are used as instruments. We indicate this as IV_L. This is the most obvious choice when $\rho \ne 0$. We first examine the efficiency loss when OLS still remains consistent, because $\rho = 0$. We find
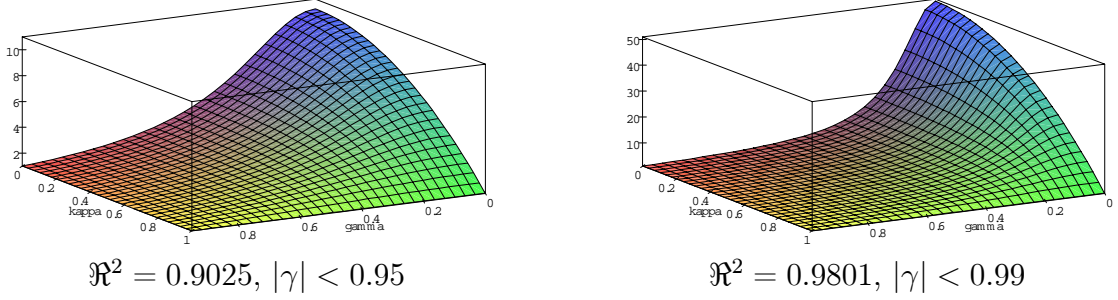
$$
\begin{aligned}
\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L}) &= \sigma_\varepsilon^2 \frac{\mathsf{V}(x_t)[\mathsf{C}(x_t,y_{t-1})]^2 - 2\mathsf{C}(x_t,x_{t-1})\mathsf{C}(x_t,y_{t-1})\mathsf{C}(x_t,y_t) + \mathsf{V}(y_t)[\mathsf{C}(x_t,x_{t-1})]^2}{[\mathsf{V}(y_t)\mathsf{C}(x_t,x_{t-1}) - \mathsf{C}(x_t,y_t)\mathsf{C}(x_t,y_{t-1})]^2} \\[2mm]
&= \frac{1-\gamma^2}{1 + \frac{\beta^2\sigma_\xi^2}{\sigma_\varepsilon^2}\frac{\gamma^2}{(1-\gamma\kappa)^2}} \tag{4.20} \\[2mm]
&= \frac{1-\gamma^2\kappa^2}{1+\gamma^2\frac{1-\kappa^2}{1-\Re^2}}.
\end{aligned}
$$

15

In the first line we have assumed that the denominator is strictly positive, which requires both $\beta \neq 0$ and $\kappa \neq 0$. The ratio

$$\frac{\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L})}{\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})} = \frac{1 - \Re^2 \kappa^2}{1 - \Re^2 + \gamma^2 \left(1 - \kappa^2\right)} \tag{4.21}$$

has a maximum of $(1 - \Re^2)^{-1}$ at $\gamma = \kappa = 0$, which can be seen from Figure 4.8, but note that $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L})$ is not defined for $\kappa = 0$.

Figure 4.8: $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L})/\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})$ in ARX(1), $\kappa \neq 0$, $\lambda = 0$

$\Re^2 = 0.9025$, $|\gamma| < 0.95$ $\Re^2 = 0.9801$, $|\gamma| < 0.99$

Especially when γ^2 is large, the efficiency loss with respect to OLS seems only marginal, although, for both $\Re^2$ values the IV_L asymptotic variance is already about twice the OLS variance at $\gamma = 0.7$ for κ not too close to unity. For smaller values of γ the efficiency loss with respect to OLS is very substantial, as was found in the AR(1) model, although the loss is still finite for $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L})$ when $\gamma = 0$ and $\kappa \neq 0$.

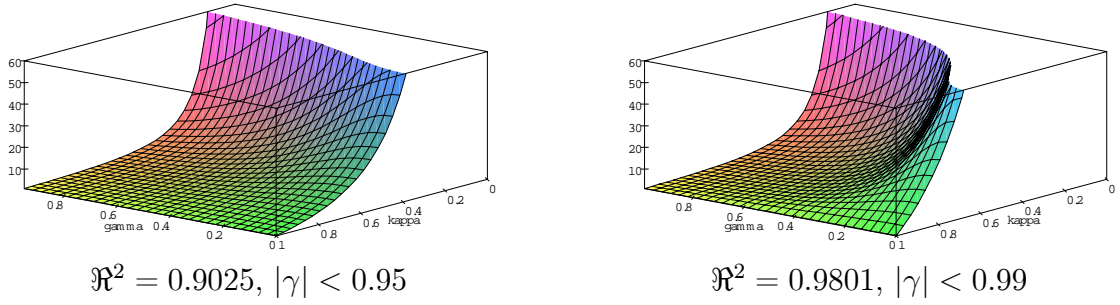With respect to β we find ($\beta \neq 0$ and $\kappa \neq 0$)

$$\begin{aligned}
\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L}) &= \sigma_\varepsilon^2 \frac{\mathsf{V}(x_t)[\mathsf{V}(y_t)]^2 - \mathsf{V}(y_t)[\mathsf{C}(x_t, y_t)]^2}{[\mathsf{V}(y_t)\mathsf{C}(x_t, x_{t-1}) - \mathsf{C}(x_t, y_t)\mathsf{C}(x_t, y_{t-1})]^2} \\[2mm]
&= \frac{\sigma_\varepsilon^2}{\sigma_\xi^2} \frac{1}{\kappa^2} \frac{1 - \kappa^2 + \frac{1+\gamma\kappa}{1-\gamma\kappa} \frac{\beta^2 \sigma_\xi^2}{\sigma_\varepsilon^2}}{1 + \frac{\gamma^2}{(1-\gamma\kappa)^2} \frac{\beta^2 \sigma_\xi^2}{\sigma_\varepsilon^2}} \\[2mm]
&= \frac{\beta^2}{\kappa^2} \frac{1 - \Re^2}{\Re^2 - \gamma^2} \frac{(1 + \gamma\kappa)^2}{1 - \Re^2 + \gamma^2 \left(1 - \kappa^2\right)},
\end{aligned} \tag{4.22}$$

giving the ratio

$$\frac{\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L})}{\mathsf{AV}(\hat{\beta}_{\mathrm{OLS}})} = \frac{1}{\kappa^2} \frac{1 - \kappa^2 \Re^2}{1 - \Re^2 + \gamma^2 \left(1 - \kappa^2\right)}. \tag{4.23}$$
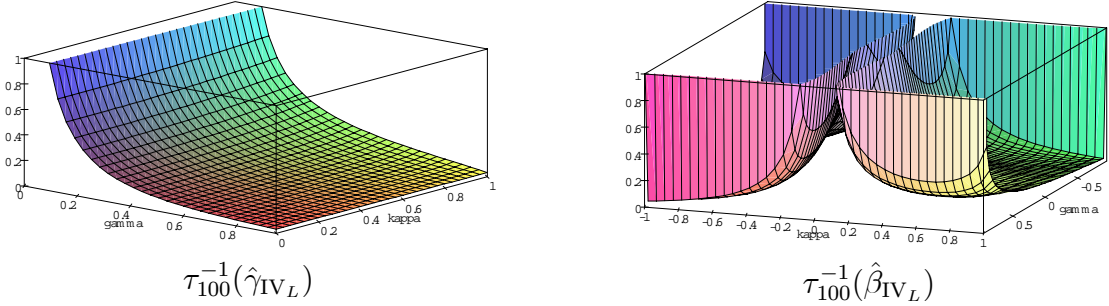
This yields Figure 4.9, which diagnoses a very serious efficiency loss, especially for low κ^2 values but also for low γ^2 values (if κ^2 is not very close to 1). Irrespective of the value of γ, the IV_L variance is here more than double that of OLS for any $|\kappa| < 0.8$.

Figure 4.9: $\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L})/\mathsf{AV}(\hat{\beta}_{\mathrm{OLS}})$ in ARX(1), $\kappa \neq 0$, $\lambda = 0$

$\Re^2 = 0.9025$, $|\gamma| < 0.95$ $\qquad\qquad$ $\Re^2 = 0.9801$, $|\gamma| < 0.99$

We conclude that the efficiency loss of IV_L is much more serious over relevant parts of the parameter space than for IV_X. Inference on γ by IV_L is affected if γ^2 is not very close to $\Re^2$ (except when $\kappa^2 \to 1$) and inference on β by IV_L is affected if κ^2 is not very close to 1, whereas there is obviously a weak-instrument problem here when κ is close to zero. So, we find that estimator efficiency for IV_L is generally much lower than for OLS. Whether this implies that the actual magnitude of the asymptotic variance precludes sharp inference in areas other than those where γ or β are small can be examined again from τ_T^{-1}. Figure 4.10 shows that for $T = 100$ and $\Re^2 = 0.9025$ useful inference by IV_L on γ is possible for $|\gamma| > 0.2$. This conclusion is similar to that for OLS, although we find that the τ_T^{-1} values for OLS at $\Re^2 = 0.64$ correspond more or less to those of IV_L at a much higher $\Re^2$; so here the only effect of using these instruments is loss of efficiency. Figure 4.10 also shows that IV_L inference on β is not only hampered when β is small (which implies γ large, due to the fixed $\Re^2$), but also when κ is small, which, self-evidently, makes x_{t-1} a poor instrument for x_t.

Figure 4.10: Results for ARX(1); $\Re^2 = 0.9025$, $|\gamma| < 0.95$, $\kappa \neq 0$, $\lambda = 0$

$\tau_{100}^{-1}(\hat{\gamma}_{\mathrm{IV}_L})$ $\qquad\qquad\qquad$ $\tau_{100}^{-1}(\hat{\beta}_{\mathrm{IV}_L})$

4.3. ARX(1) models with weak exogeneity

Now we relax the assumption that $\lambda = 0$, but maintain $\rho = 0$. Here we have weak exogeneity, because there is Granger causality from y to x. Under the stationarity conditions (3.6), the OLS estimators for the parameters γ and β are known to be consistent and asymptotically efficient even if $\lambda \neq 0$.

Regarding the second moments of the variables, we find from the expressions in (4.7) that only the one for $\mathsf{V}(y_t)$ is not affected by the introduction of the feedback of y_{t-1} in

x_t. We now have

$$\left.\begin{aligned}
\mathsf{V}(y_t) &= \beta^2\mathsf{V}(x_t) + \gamma^2\mathsf{V}(y_{t-1}) + 2\beta\gamma\mathsf{C}(x_t, y_{t-1}) + \sigma_\varepsilon^2, \\[2mm]
\mathsf{V}(x_t) &= \kappa^2\mathsf{V}(x_t) + \lambda^2\mathsf{V}(y_t) + 2\kappa\lambda\mathsf{C}(x_t, y_t) + \sigma_\xi^2, \\[2mm]
\mathsf{C}(x_t, y_{t-1}) &= \kappa\mathsf{C}(x_t, y_t) + \lambda\mathsf{V}(y_t) \\[2mm]
\mathsf{C}(x_t, y_t) &= \beta\mathsf{V}(x_t) + \gamma\mathsf{C}(x_t, y_{t-1}) \\[2mm]
\mathsf{C}(x_t, x_{t-1}) &= \kappa\mathsf{V}(x_t) + \lambda\mathsf{C}(x_t, y_t).
\end{aligned}\right\} \tag{4.24}$$

These yield

$$\mathsf{V}(x_t) = \frac{(1+\gamma\kappa)\lambda^2\sigma_\varepsilon^2 + [(1-\gamma^2)(1-\gamma\kappa) - 2\gamma\lambda\beta]\sigma_\xi^2}{(1-\gamma\kappa)[(1-\gamma^2)(1-\kappa^2) - 2(\gamma+\kappa)\lambda\beta - \lambda^2\beta^2]}, \tag{4.25}$$

$$\mathsf{V}(y_t) = \frac{[(1-\gamma\kappa)(1-\kappa^2) - 2\kappa\lambda\beta]\sigma_\varepsilon^2 + (1+\gamma\kappa)\beta^2\sigma_\xi^2}{(1-\gamma\kappa)[(1-\gamma^2)(1-\kappa^2) - 2(\gamma+\kappa)\lambda\beta - \lambda^2\beta^2]}, \tag{4.26}$$

$$\mathsf{C}(x_t, y_{t-1}) = \frac{(1-\kappa^2-\kappa\lambda\beta)\lambda\sigma_\varepsilon^2 + [(1-\gamma^2)\kappa + \lambda\beta]\beta\sigma_\xi^2}{(1-\gamma\kappa)[(1-\gamma^2)(1-\kappa^2) - 2(\gamma+\kappa)\lambda\beta - \lambda^2\beta^2]}, \tag{4.27}$$

from which $\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})$ and $\mathsf{AV}(\hat{\beta}_{\mathrm{OLS}})$ can be established using the earlier expressions in moments of the data. This results in expressions in β, γ, κ, λ, σ_ξ^2 and σ_ε^2 that are much too complex to be useful for further direct interpretation or manipulation. Therefore, we can no longer use the SWP facilities to produce 3-D diagrams from explicit analytic expressions.

However, we can proceed by evaluating the data moments and the asymptotic variance expressions numerically over a relevant grid of parameter values, and use these to produce 3-D diagrams. We did so employing Gauss code for the data moments and for the diagonal elements of the relevant Σ_{WW}^{-1} and $\Sigma_{ZW}^{-1}\Sigma_{ZZ}(\Sigma_{ZW}^{-1})^{-1}$ matrices. To fix $\Re^2$ we use (4.11) again, which now yields

$$\beta^2\frac{\sigma_\xi^2}{\sigma_\varepsilon^2} = \frac{(1-\gamma\kappa)[(1-\kappa^2)(\Re^2 - \gamma^2) - 2\gamma\lambda\beta - \lambda^2\beta^2] + 2\kappa(\gamma\kappa - \Re^2)\lambda\beta}{(1+\gamma\kappa)(1-\Re^2)}. \tag{4.28}$$

Note that this simplifies to (4.12) for $\lambda = 0$ indeed. Substitution of (4.28) into (4.24) through (4.27) leads to expressions for $\mathsf{AV}(\hat{\gamma}_{\mathrm{OLS}})/\gamma^2$ and $\mathsf{AV}(\hat{\beta}_{\mathrm{OLS}})/\beta^2$ in terms of γ, κ, $\Re^2$ and $\lambda\beta$. The 'extra' determining factor $\lambda\beta$ complicates designing informative diagrams, because next to fixing the value of $\Re^2$, we have to impose other restrictions. For instance, $\theta = \beta/(1-\gamma) = 1$ (i.e., the long-run multiplier is unity or any other relevant value) and $\lambda \in \{-0.1, 0, +0.1\}$ (i.e., the feedback parameter is zero or relatively mild). Note that for positive γ, κ and β, which are cases of great interest in practice, the second of the inequalities (3.6) would not always be met when λ is positive as well. It can be inferred from (3.5) that we may confine ourselves to examination of positive values of β or θ only, provided that we examine both positive and negative values of λ, in case ε_t and ξ_t are symmetrically distributed around zero.

We present a few results here focussing on $\Re^2 = 0.9025$ and $\theta = 1$, examining $0.01 \leq \gamma \leq 0.9$ and $0.01 \leq \kappa \leq 0.85$. Then the stationarity conditions are fulfilled for $|\lambda| \leq 0.1$.

For $\lambda = 0$, the restriction $\theta = 1$, i.e., $\beta = 1 - \gamma$, becomes inconsequential, because the formulas simplify to those of the former subsection when $\lambda\beta = 0$. We did find that for $\lambda = 0$ our Gauss program produces similar graphs as the Maple figures obtained in the foregoing subsection using SWP. Figure 4.11 gives the efficiency ratio's for $\lambda = -0.1$. For technical reasons in graphs produced by Gauss programs, we have set any data points beyond the range of the vertical axes at the maximum (minimum) value of the plot. Compared to the results in Figures 4.8 and 4.9, respectively, where the same ratio's can be found under strong exogeneity (i.e., for $\lambda = 0$), we see that the efficiency loss of $\hat{\gamma}_{\mathrm{IV}_L}$ with respect to OLS is more moderate here, although we also note more severe problems for κ very close to zero. For $\hat{\beta}_{\mathrm{IV}_L}$ we find that the results for $\lambda = 0$ and $\lambda = -0.1$ do not differ very much. GIV_L employing $x_{t-1}, y_{t-1}, x_{t-2}, y_{t-2}$, does not yield an efficiency improvement asymptotically.

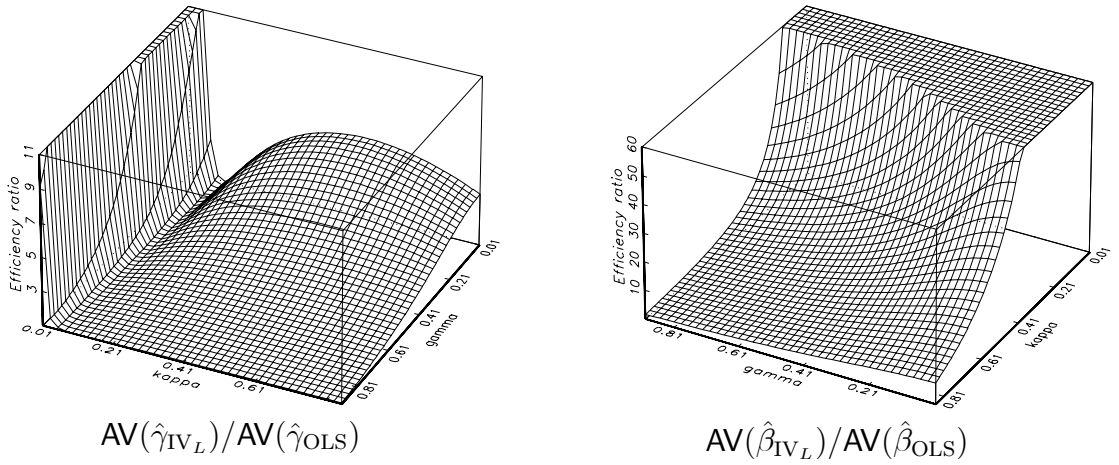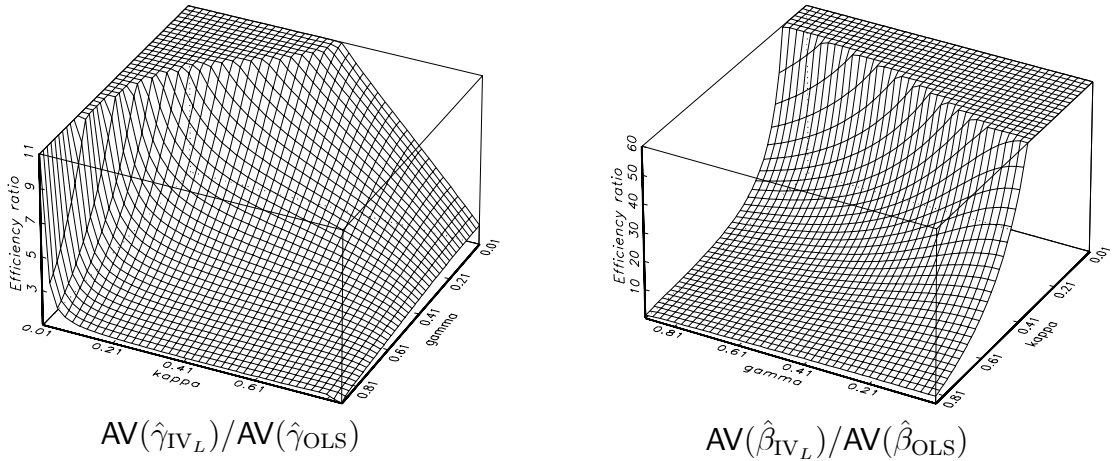Figure 4.11: Results for ARX(1) with $\Re^2 = 0.9025$, $\lambda = -0.1$, $\theta = 1$

$\mathrm{AV}(\hat{\gamma}_{\mathrm{IV}_L})/\mathrm{AV}(\hat{\gamma}_{\mathrm{OLS}})$ $\mathrm{AV}(\hat{\beta}_{\mathrm{IV}_L})/\mathrm{AV}(\hat{\beta}_{\mathrm{OLS}})$

Figure 4.12: Results for ARX(1) with $\Re^2 = 0.9025$, $\lambda = +0.1$, $\theta = 1$

$\mathrm{AV}(\hat{\gamma}_{\mathrm{IV}_L})/\mathrm{AV}(\hat{\gamma}_{\mathrm{OLS}})$ $\mathrm{AV}(\hat{\beta}_{\mathrm{IV}_L})/\mathrm{AV}(\hat{\beta}_{\mathrm{OLS}})$

In Figure 4.12 we present results for $\lambda = 0.1$ and establish a large region of moderate γ and κ values where substantial efficiency losses occur for $\hat{\gamma}_{\mathrm{IV}_L}$ - which must be due to weak instruments. Again we notice little difference regarding $\hat{\beta}_{\mathrm{IV}_L}$. We conclude that it is very important to know whether or not there are instantaneous feedbacks, because the efficiency losses of IV with respect to OLS are moderate only over a very small region

in the (γ, κ) space. Especially the location of the region of great inefficiency of $\hat{\gamma}_{\mathrm{IV}_L}$ is heavily dependent on the occurrence and actual magnitude of any lagged feedbacks.

5. Models with lagged and instantaneous feedbacks

Next we consider the situation where $\mathsf{E}(\varepsilon_t \xi_t) = \rho \sigma_\varepsilon \sigma_\xi$ may be different from zero. If $\rho \neq 0$, the regressor x_t and the disturbance term ε_t are correlated and hence y_t and x_t are jointly dependent (irrespective of the values of κ and β). Note that in a simultaneous model it is not obvious what we mean by the population $\Re^2$. Because $\mathsf{C}(x_t, \varepsilon_t) \neq 0$, our definition (3.7) no longer implies (4.11). From

$$\Re^2 = \frac{\mathsf{V}(\beta x_t + \gamma y_{t-1})}{\mathsf{V}(y_t)} = \frac{\beta^2 \mathsf{V}(x_t) + \gamma^2 \mathsf{V}(y_t) + 2\beta\gamma \mathsf{C}(x_t, y_{t-1})}{\mathsf{V}(y_t)}, \qquad (5.1)$$

this is seen to be smaller than $1 - \sigma_\varepsilon^2 / \mathsf{V}(y_t)$ if $\rho > 0$ and greater otherwise.

The simultaneity also affects some of the data moments of (4.24). The expressions for $\mathsf{V}(x_t)$ and $\mathsf{C}(x_t, y_{t-1})$ remain the same, but we have to replace $\mathsf{V}(y_t)$ and $\mathsf{C}(x_t, y_t)$ by

$$\mathsf{V}(y_t) = \beta^2 \mathsf{V}(x_t) + \gamma^2 \mathsf{V}(y_t) + 2\beta\gamma \mathsf{C}(x_t, y_{t-1}) + \sigma_\varepsilon^2 + 2\beta \mathsf{C}(x_t, \varepsilon_t), \qquad (5.2)$$

and

$$\mathsf{C}(x_t, y_t) = \beta \mathsf{V}(x_t) + \gamma \mathsf{C}(x_t, y_{t-1}) + \mathsf{C}(x_t, \varepsilon_t), \qquad (5.3)$$
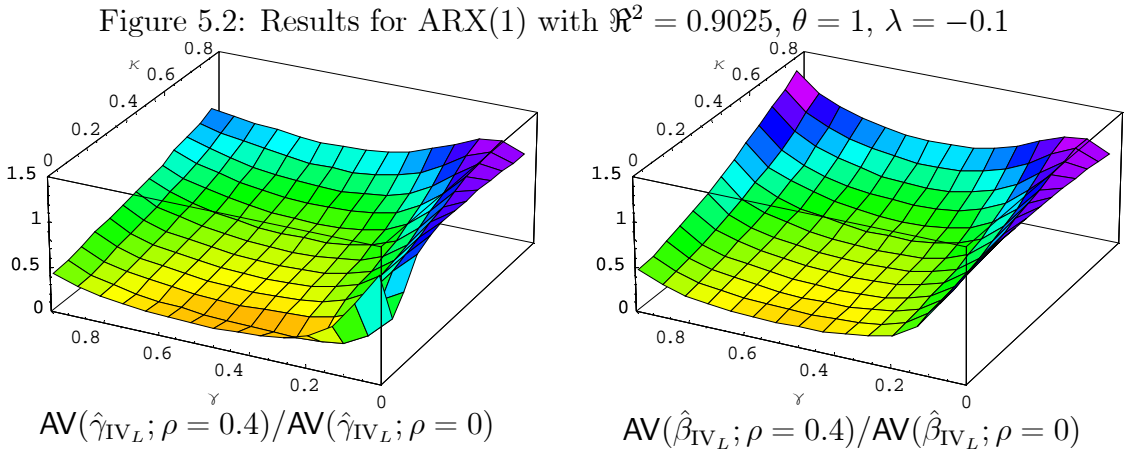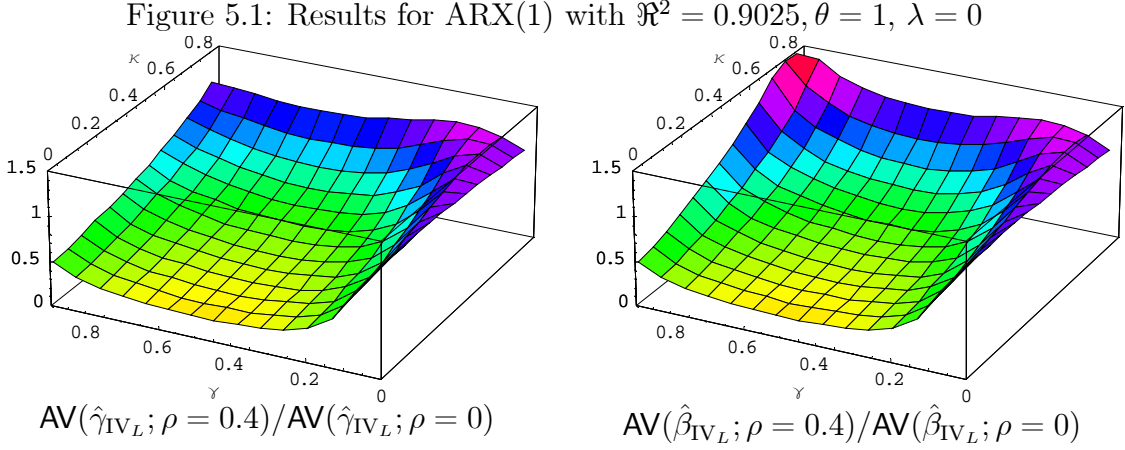
where

$$\mathsf{C}(x_t, \varepsilon_t) = \rho \sigma_\varepsilon \sigma_\xi. \qquad (5.4)$$

These extensions have a profound effect on the complexity of the analytic expressions for the asymptotic efficiency of the estimators, and also explicit expressions for the data moments in parameter values are unruly. This complicates the creation of Gauss code for finding numerical results for the efficiency measures. Another complication for designing graphical results is that the efficiency measures depend on six parameters now, apart from the numeraire σ_ε. The dimensionality can be reduced in a sensible way again by fixing $\Re^2$, the long run multiplier $\theta = \beta/(1-\gamma)$, and perhaps either the simultaneity ρ or the lagged feedback λ. However, that leaves another three free parameters, whereas we can only handle two of them in a standard 3-D plot. It seems a challenge now to exploit capabilities of Mathematica, by which we can both solve through symbolic algebra the asymptotic efficiencies in terms of the transformed model parameters $(\gamma, \kappa, \rho, \lambda, \theta, \Re^2)$, make sure that the stationarity restrictions (3.6) are imposed, and then make animations of 3-D plots, so that we can actually see how the (relative) asymptotic efficiency changes over the γ and κ domain when either ρ, λ, θ or $\Re^2$ varies. We could even make animations where all the latter four parameters vary jointly in a controlled way. At `http://www1.fee.uva.nl/ke/Joseph_Kiviet.htm` a web-site is available where some of the animated versions of graphs from this study can be seen in action. From Mathematica's symbolic manipulations it appeared that, upon fixing $\Re^2$, there are positive and negative roots for σ_ξ. Naturally we examine positive values only.

5.1. ARX(1) models with and without simultaneity

We present some standard 3-D graphs here in which we fix $\theta = 1$ and $\Re^2 = 0.9025$, so that γ, κ, λ and ρ are the remaining free parameters. To examine how much efficiency is lost

due to a non-zero value of ρ, we examine the efficiency ratio's of $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L})$ and $\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L})$ for $\rho = 0.4$ and for $\rho = 0$, respectively. This gives the inflation factor of the efficiency due to simultaneity. Figure 5.1 shows results for $\lambda = 0$. We find that simultaneity may both increase or decrease the efficiency of IV, but the effect is limited (although it may be about $\pm 50\%$). Figure 5.2 shows for $\lambda = -0.1$ how IV_L efficiency changes over the domain of γ and κ when ρ jumps from 0 to $\rho = 0.4$. This seems to mitigate the effects slightly.

Figure 5.1: Results for ARX(1) with $\Re^2 = 0.9025, \theta = 1, \lambda = 0$

$\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L}; \rho = 0.4)/\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L}; \rho = 0)$ $\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L}; \rho = 0.4)/\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L}; \rho = 0)$

Figure 5.2: Results for ARX(1) with $\Re^2 = 0.9025, \theta = 1, \lambda = -0.1$

$\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L}; \rho = 0.4)/\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L}; \rho = 0)$ $\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L}; \rho = 0.4)/\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L}; \rho = 0)$

5.2. IV versus OLS in ARX(1) models with simultaneity

Next to the above comparisons of the same technique for different ρ values, we would like to produce comparisons for different estimators, given a particular value of ρ. We compare IV with COLS and inconsistent OLS, making use of the theoretical analysis of Section 2 which require that the disturbances ε_t are (close to) normally distributed. First we compare the efficiency of two consistent estimators. In Figure 5.3 we examine the ratio of $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L})$ and the non-operational $\mathsf{AV}(\hat{\gamma}_{\mathrm{COLS}})$, both at $\rho = 0.4, \theta = 1$, for $\lambda = 0$ and -0.1. We find that $\hat{\gamma}_{\mathrm{IV}_L}$ is much less efficient, especially for γ moderate or small. When $\lambda \neq 0$ this has substantial impact at particular κ values.

Figure 5.3: $\mathsf{AV}(\hat{\gamma}_{\mathrm{IV}_L})/\mathsf{AV}(\hat{\gamma}_{\mathrm{COLS}})$ for ARX(1) with $\Re^2 = 0.9025, \theta = 1, \rho = 0.4$

$\lambda = 0.0$ $\qquad\qquad\qquad\qquad\qquad\qquad$ $\lambda = -0.1$

Figure 5.4 shows that $\hat{\beta}_{\mathrm{IV}_L}$ is much less efficient than $\hat{\beta}_{\mathrm{COLS}}$, except when κ and γ are both very large. Note that the efficiency difference between these two estimators is often much more substantial than the change in IV efficiency due to the presence of simultaneity as established in the Figures 5.1 and 5.2.

Figure 5.4: $\mathsf{AV}(\hat{\beta}_{\mathrm{IV}_L})/\mathsf{AV}(\hat{\beta}_{\mathrm{COLS}})$ for ARX(1) with $\Re^2 = 0.9025, \theta = 1, \rho = 0.4$

$\lambda = 0.0$ $\qquad\qquad\qquad\qquad\qquad\qquad$ $\lambda = -0.1$

Figures 5.5 through 5.8 show that the difference in efficiency between IV_L and OLS can be substantial too at $T = 100$ (assuming that the asymptotic approximations are reasonably accurate at such a sample size). We have taken the natural logarithm of the efficiency ratio now, so that positive values indicate the superiority of OLS. From Figure 5.5 we find for inference on γ that, when $\lambda = 0$, IV_L is preferable only when ρ, γ and κ are all large. Over a large region, also when $\rho = 0.4$, OLS has a smaller asymptotic mean squared error, despite its bias and inconsistency. For moderate values of γ and κ, the relative OLS efficiency improves with increasing ρ. Hence, its inconsistency seems less of a worry than is the weakness of the instruments for IV. Of course, this evidence is only indicative, because we should check whether, at a sample size of 100, the asymptotic approximations are sufficiently accurate. Additionally, it is not obvious how to exploit the OLS estimator for inference purposes, because of its inconsistency. Its proper use would require an assessment of its bias. Such an assessment, though, if at all possible, would enable to make the (even more efficient) COLS estimator operational.

Figure 5.5: $\ln[\mathbf{AMSE}(\hat{\gamma}_{\mathrm{IV}_L})/\mathbf{AMSE}(\hat{\gamma}_{\mathrm{OLS}})]$ for $\Re^2 = 0.9025$, $\theta = 1$, $T = 100$, $\lambda = 0$

$\rho = -0.4$

$\rho = 0.0$

$\rho = 0.2$

$\rho = 0.4$

Figure 5.6: $\ln[\mathbf{AMSE}(\hat{\beta}_{\mathrm{IV}_L})/\mathbf{AMSE}(\hat{\beta}_{\mathrm{OLS}})]$ for $\Re^2 = 0.9025$, $\theta = 1$, $T = 100$, $\lambda = 0$

$\rho = -0.4$

$\rho = 0.0$

$\rho = 0.2$

$\rho = 0.4$

Figure 5.6 shows that γ has little effect on this IV_L over OLS efficiency ratio regarding estimation of β. When $\rho \neq 0$ IV can beat inconsistent OLS for large $|\kappa|$ only.

When $\lambda \neq 0$ the patterns change drastically with respect to estimating γ (Figure 5.7), but less so for β (Figure 5.8). However, there still exist large regions where OLS is much more efficient than IV.

Figure 5.7: $\ln[\text{AMSE}(\hat{\gamma}_{\text{IV}_L})/\text{AMSE}(\hat{\gamma}_{\text{OLS}})]$ for $\Re^2 = 0.9025$, $\theta = 1$, $T = 100$, $\rho = 0.4$

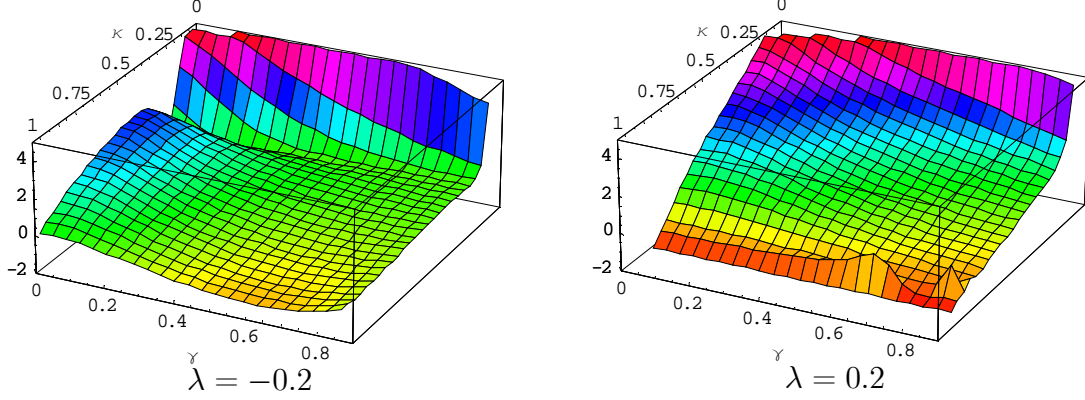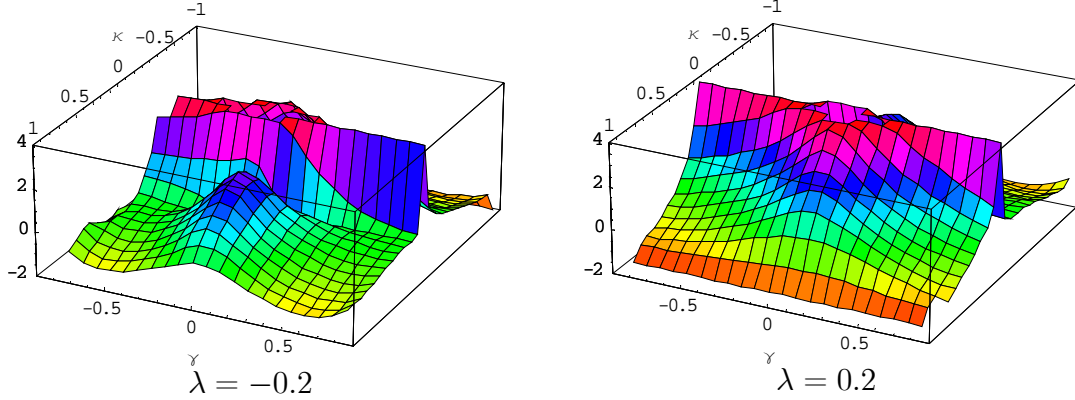

$\lambda = -0.2$ $\lambda = 0.2$

Figure 5.8: $\ln[\text{AMSE}(\hat{\beta}_{\text{IV}_L})/\text{AMSE}(\hat{\beta}_{\text{OLS}})]$ for $\Re^2 = 0.9025$, $\theta = 1$, $T = 100$, $\rho = 0.4$

$\lambda = -0.2$ $\lambda = 0.2$

6. Conclusions

We examine the asymptotic efficiency of various estimators for simple but commonly employed specifications of dynamic regression models with iid disturbances. These specifications incorporate both dynamic adjustments and simultaneity, and they may also entail weak or strong exogeneity. We focus primarily on OLS and IV implementations that exploit lagged values of the variables occurring in the model under study as instruments. Therefore, the asymptotic estimator efficiency is completely characterized by the parameter values of the system, since we assume stationarity, and hence initial conditions do not affect the asymptotic efficiency.

Obviously, if OLS is consistent its asymptotic efficiency makes it preferable to IV. If there is simultaneity in the model OLS yields inconsistent estimators, whereas the IV estimator is consistent provided that the instruments are valid. However, if the instruments employed are weak then problems with respect to scale (i.e., a large second

moment) occur, and, especially in finite samples, also with respect to location (i.e., bias), whereas ultimately, due to either invalid instruments or non-identification, the location problems constitute inconsistent or even totally uninterpretable estimators.

After expressing for the vraious types of models examined the asymptotic second moments of estimators in the model parameters, it is generally rather easy to detect and to rationalize some extremely hazardous situations, usually occurring when particular coefficients approach zero or their boundary values. By scanning the whole parameter space of a model with six parameters, however, we find more areas in the relevant parameter space with serious efficiency problems. We also examine whether and where efficiency problems of consistent estimators originating from instrument weakness turn out to be more detrimental than the efficiency problems of inconsistent estimators caused by instrument invalidity. This is all achieved in close interplay between analytical derivations, graphical methods, computer algebra and animated computerized visualizations.

Comparing the relative efficiency of a particular consistent estimator, when either lagged feedbacks (weak exogeneity) or instantaneous feedbacks (simultaneity) are present (while keeping the respective models comparable otherwise by controlling their signal-to-noise ratio by the coefficient of determination), we find that the effects of these feedbacks are relatively moderate in comparison to efficiency differences due to using alternative estimation techniques. For cases where OLS is consistent, we illustrate that efficiency losses are not marginal but very substantial over large regions of the parameter space when IV is used with lagged variables as valid (but often weak) instruments. That phenomenon also helps to explain why we find that inconsistent OLS is often much more efficient than consistent IV estimation. Upon comparing a particular finite-sample approximation to their mean squared errors (which presupposes that the disturbances are close to normally distributed), we show that enormous gains in efficiency can be achieved by employing invalid but non-weak instruments (as OLS does in the simultaneous model) instead of valid but weak instruments in implementations yielding consistent IV estimators. The reason is that often IV is severely dispersed around a correct location, whereas OLS is surprisingly concentrated around a moderately incorrect location.

To what degree the present findings are relevant for inference in actual finite samples remains to be verified either by Monte Carlo experiments or by more sophisticated possibly higher-order or local-to-zero asymptotic analytical derivations.

References

Belsley, D.A., 1982. Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise. *Journal of Econometrics* 20, 211-253.

Belsley, D.A., 1992. The degree of effective identification and a diagnostic measure for assigning it. In: Amman et al. (eds). *Computational Economics and Econometrics.* Kluwer Academic Publishers, 55-62.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.

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.

Bowden, R.J., Turkington, D.A., 1984. *Instrumental Variables.* Cambridge University Press, Cambridge.

Dufour, J-M., 2003. Identification, weak instruments and statistical inference in

econometrics. *Canadian Journal of Economics* 36, 767-808.

Evans, M.D.D., Lyons, R.K., 2002. Time-varying liquidity in foreign exchange. *Journal of Monetary Economics* 49, 1025-1051.

Favero, C.A., 2001. *Applied Macroeconometrics.* Oxford University Press, Oxford.

Ferson, W.E., Foerster, S.R., 1994. Finite sample properties of the Generalized Method of Moments in tests of conditional asset pricing models. *Journal of Financial Economics* 36, 29-55.

Godfrey, L.G., 1999. Instrument relevance in multivariate linear models. *Review of Economics and Statistics* LXXXI, 550-552.

Hendry, D.F., 1995. *Dynamic Econometrics.* Oxford University Press, Oxford.

Poskitt, D.S., Skeels, C.L., 2002. Assessing instrumental variable relevance: An alternative measure and some exact finite sample theory. Mimeo.

Shea, J., 1997. Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics*, LXXIX 348-352.

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

Wang,J., Zivot, E., 1998. Inference on structural paramters in instrumental variables regression with weak instruments. Econometrica 66, 1389-1404.

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.