# UvA ECONOMETRICS

Discussion Paper: 2011/03

# Mauling the Method of Moments into Kinky Least Squares

## Jan F. Kiviet

UvA UNIVERSITEIT VAN AMSTERDAM

# Mauling the Method of Moments
# into Kinky Least Squares

Jan F. Kiviet

University of Amsterdam[*]

21 June 2011

The major challenge of econometrics is assessing the essentials of relationships between empirical phenomena, where this has to be based on data which could not be collected from controlled experiments. This calls for inference procedures which can handle both exogenous and endogenous explanatory variables. For proper interpretation of econometric inference various assumptions of a technical statistical nature should hold, whereas for some of these conditions their validity cannot be corroborated. Therefore, they simply have to be adopted, either on the basis of conventions or other often highly subjective convictions. In this note we demonstrate that some of these crucial but statistically unverifiable assumptions can be replaced by others, in order to make inferences not only more credible, but by the same stroke more robust, efficient and accurate as well.

## 1  A simple though unidentified model

Consider the two linear equations

$$y_i^{(1)} = \beta y_i^{(2)} + \gamma_1 z_i^{(1)} + \gamma_2 z_i^{(2)} + \varepsilon_i, \tag{1}$$

$$y_i^{(2)} = \pi_1 z_i^{(1)} + \pi_2 z_i^{(2)} + v_i, \tag{2}$$

which model two endogenous variables $y_i^{(1)}$ and $y_i^{(2)}$ for observations $i = 1, ..., n$. Let the random disturbances be such that $\varepsilon_i \sim NID(0, \sigma_\varepsilon^2)$ and $v_i \sim NID(0, \sigma_v^2)$, with $E(\varepsilon_i v_i) = \sigma_{\varepsilon v} = \rho_{\varepsilon v} \sigma_\varepsilon \sigma_v$, where $|\rho_{\varepsilon v}| < 1$, $\sigma_\varepsilon > 0$ and $\sigma_v > 0$. We suppose that

$$E(\varepsilon_i z_i^{(j)}) = E(v_i z_i^{(j)}) = 0 \text{ for } j = 1, 2. \tag{3}$$

So, in both equations the two variables $z_i^{(j)}$, for which we may assume $z_i^{(j)} \sim NID(0, \sigma_{z^{(j)}}^2)$, are predetermined, or even exogenous. However, we have $E(y_i^{(2)} \varepsilon_i) \neq 0$ when $\rho_{\varepsilon v} \neq 0$. Then the two variables $y_i^{(j)}$ are jointly dependent. The first equation is the structural (causal) relationship for $y_i^{(1)}$ and the second is the reduced form equation for $y_i^{(2)}$. The three coefficients of the

---

[*]Address: Department of Quantitative Economics, Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands; phone +31.20.5254224; email j.f.kiviet@uva.nl. From 1 July 2011 onwards my major affiliation will be: Nanyang Technological University, Singapore.

structural equation (1) are not identified by the two orthogonality conditions $E(\varepsilon_i z_i^{(j)}) = 0$. We will clarify this and examine by what alternative further restricting assumptions their identification can be achieved.

Continuing to use obvious notation for variances and covariances we easily obtain the following relationships between second moments

$$\sigma_{y^{(1)}y^{(2)}} = \beta\sigma_{y^{(2)}}^2 + \gamma_1\sigma_{y^{(2)}z^{(1)}} + \gamma_2\sigma_{y^{(2)}z^{(2)}} + \sigma_{y^{(2)}\varepsilon}, \tag{4}$$

$$\sigma_{y^{(1)}z^{(1)}} = \beta\sigma_{y^{(2)}z^{(1)}} + \gamma_1\sigma_{z^{(1)}}^2 + \gamma_2\sigma_{z^{(1)}z^{(2)}}, \tag{5}$$

$$\sigma_{y^{(1)}z^{(2)}} = \beta\sigma_{y^{(2)}z^{(2)}} + \gamma_1\sigma_{z^{(1)}z^{(2)}} + \gamma_2\sigma_{z^{(2)}}^2, \tag{6}$$

$$\sigma_{y^{(2)}z^{(1)}} = \pi_1\sigma_{z^{(1)}}^2 + \pi_2\sigma_{z^{(1)}z^{(2)}}, \tag{7}$$

$$\sigma_{y^{(2)}z^{(2)}} = \pi_1\sigma_{z^{(1)}z^{(2)}} + \pi_2\sigma_{z^{(2)}}^2. \tag{8}$$

Note that by virtue of the Law of Large Numbers such second moments regarding observed variables can directly be estimated consistently by their sample equivalent. This is not the case for $\sigma_{y^{(2)}\varepsilon}$, simply because $\varepsilon$ has not been observed. The equations (5) through (8) do no longer contain such a covariance involving an unobserved disturbance term, because we did already substitute the four orthogonality conditions (3).

From the two equations (7) and (8) consistent estimators for the two reduced form coefficients $\pi_1$ and $\pi_2$ directly follow. Denoting sample second moments by putting a hat on $\sigma$, we find from

$$\hat\sigma_{y^{(2)}z^{(1)}} = \hat\pi_1\hat\sigma_{z^{(1)}}^2 + \hat\pi_2\hat\sigma_{z^{(1)}z^{(2)}},$$

$$\hat\sigma_{y^{(2)}z^{(2)}} = \hat\pi_1\hat\sigma_{z^{(1)}z^{(2)}} + \hat\pi_2\hat\sigma_{z^{(2)}}^2,$$

that

$$\begin{pmatrix} \hat\pi_1 \\ \hat\pi_2 \end{pmatrix} = \begin{pmatrix} \hat\sigma_{z^{(1)}}^2 & \hat\sigma_{z^{(1)}z^{(2)}} \\ \hat\sigma_{z^{(1)}z^{(2)}} & \hat\sigma_{z^{(2)}}^2 \end{pmatrix}^{-1} \begin{pmatrix} \hat\sigma_{y^{(2)}z^{(1)}} \\ \hat\sigma_{y^{(2)}z^{(2)}} \end{pmatrix} = (Z'Z)^{-1}Z'y^{(2)}, \tag{9}$$

where the $n \times 2$ matrix $Z$ and $n \times 1$ vector $y^{(2)}$ are defined in the obvious way. Hence, as is well-known, reduced form coefficients are identified and can be estimated consistently by the method of moments (MM), which in this case is equivalent to least-squares (LS). From yet another equation in second moments

$$\sigma_{y^{(2)}}^2 = \pi_1^2\sigma_{z^{(1)}}^2 + 2\pi_1\pi_2\sigma_{z^{(1)}z^{(2)}} + \pi_2^2\sigma_{z^{(2)}}^2 + \sigma_v^2, \tag{10}$$

we easily find a consistent MM estimator for $\sigma_v^2$, namely

$$\hat\sigma_v^2 = \sigma_{y^{(2)}}^2 - \hat\pi_1^2\hat\sigma_{z^{(1)}}^2 - 2\hat\pi_1\hat\pi_2\hat\sigma_{z^{(1)}z^{(2)}} - \hat\pi_2^2\hat\sigma_{z^{(2)}}^2 = \frac{1}{n}y^{(2)\prime}[I - Z(Z'Z)^{-1}Z']y^{(2)}. \tag{11}$$

Due to the presence of the unknown term $\sigma_{y^{(2)}\varepsilon}$, identifying $\beta$, $\gamma_1$ and $\gamma_2$ and obtaining consistent estimators for them from the three equations (4), (5) and (6) is only possible by incorporating extra information. We will first review three options for this that have been discussed in the econometric literature for over half a century already, and next two further options, which have been suggested only very recently. We shall also try to indicate the advantages and disadvantages of these various approaches with respect to what in our opinion should be the major aspirations in econometrics.

2

# 2 Inference aspirations

In econometrics we want our inferences to be: credible, robust, efficient and accurate. These four qualities refer to the following. Although we have to accept that econometric inferences will be based on particular assumptions that simply have to be adopted, which neither can be imposed by brute force nor be verified empirically, we nevertheless often have a choice regarding which assumptions to adopt, or whether to formulate them very strictly or less restrictively. While adopting purely hypothetical and rather abstract assumptions may provide – as it seems – a fertile starting point in producing inference in economic theory, it cannot serve empirical econometrics in the same way, simply because in that field the assumptions made should not be incompatible with reality. Although employing sound econometric theory built on premises which do not hold for the data under study might still be labelled conditionally valid, it is at the same time misleading and of no practical use.

So, in econometrics both illusory and elusive assumptions should best be avoided as much as possible. Nevertheless, the extent of the credibility of the adopted restrictions will be determined largely by subjective preferences. It is obvious, though, that less binding assumptions are usually more credible. Designing inferences such that they apply to less binding assumptions also boosts robustness. Robustness as such refers to appropriateness of the inference technique under a wide set of assumptions. However, robustness usually comes with a price in terms of reduced efficiency. The asymptotic efficiency of inferences and their precision in finite samples is often expressed by measures such as (asymptotic) root mean squared error, (local) power of tests, or length of confidence intervals. Naturally, we will prefer the inference technique which achieves higher precision than its competitors, but self-evidently here we will as a rule face trade-off dilemmas between efficiency and robustness. Usually these cannot be enhanced both at the same time. Finally, we want our inferences to be accurate, by which we mean that they should fulfill their claimed precision. Hence, estimated standard errors of estimators should concur closely with the actual corresponding standard deviations, and the actual type I error probability of tests should be close to the nominal significance level aimed at, whereas the actual coverage probability of confidence sets should be close to their claimed confidence coefficient.

Regarding the classic simultaneous equations model, which we did set out above in simplified rudimentary form, and which has been studied intensively since World War II, it may seem almost impossible that a new and useful technique can still be added to the existing econometricians toolbox. Nonetheless, we will suggest one below, and claim that what we nicknamed kinky least squares (KLS) has a very high CREA-factor. Very often it creates jointly more Credibility, Robustness, Efficiency and Accuracy than the existing techniques.

# 3 Five routes towards identification

First we discuss three strategies that have been very well documented in the literature[1] already, and next two that have been suggested only very recently. They all claim to allow consistent estimation of the coefficients of the structural equation for $y_i^{(1)}$. Note that in practice what we denote here as the separate vectors of explanatory variables $y_i^{(2)}$, $z_i^{(1)}$, $z_i^{(2)}$ might actually be vectors including more than one variable. However, allowing for that would complicate

---

[1] For a concise and easily accessible though rather complete overview of the merits and pitfalls of the standard approaches, see Larcker and Rusticus (2010).

the notation substantially, but not lead to qualitatively very different results, although having an unequal number of elements in $\gamma_1$ and $\gamma_2$ would require some special attention. The five distinct routes towards identification that we are aware of are the following.

**A. Assuming $\rho_{\varepsilon v} = 0$.** This restriction implies $\rho_{y^{(2)}\varepsilon} = 0$, so that $y_i^{(2)}$ is predetermined. Then the structural equation can be estimated consistently by LS. Of course, if in fact $\rho_{\varepsilon v} \neq 0$, then simultaneity is neglected and the LS estimator is inconsistent[2], which jeopardizes the accuracy of standard LS inference. Because the LS residuals are orthogonal to all the regressors by construction, it is impossible to use them to test the validity of the restriction $\rho_{y^{(2)}\varepsilon} = 0$. Such a test is only feasible when the structural equation is identified. This requires that it obeys at least one exclusion restriction. However, according to established wisdom, such a restriction cannot be tested either, unless the equation is identified. Hence, for this option A it does not seem that its credibility is easily substantiated. Thus, its accuracy is often doubtful.

**B. Assuming $\gamma_2 = 0$.** Note that our model is symmetric in $z_i^{(1)}$ and $z_i^{(2)}$, so assuming $\gamma_1 = 0$ would have similar consequences. If this exclusion restriction is valid then the remaining coefficients are identified. The two equations (5) and (6) contain two unknowns now and direct application of MM yields

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_{y^{(2)}z^{(1)}} & \hat{\sigma}^2_{z^{(1)}} \\ \hat{\sigma}_{y^{(2)}z^{(2)}} & \hat{\sigma}_{z^{(1)}z^{(2)}} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\sigma}_{y^{(1)}z^{(1)}} \\ \hat{\sigma}_{y^{(1)}z^{(2)}} \end{pmatrix} = (Z'X)^{-1}Z'y^{(1)}, \tag{12}$$

where $X$ is the $n \times 2$ matrix with $(y_i^{(2)} \; z_i^{(1)})$ on the $i^{th}$ row. This is the well-known instrumental variables (IV) estimator (and in a more general setting we would have found the 2SLS estimator). It is consistent. However, when at least one of the instruments is weak, it will suffer from finite sample bias almost as severe as inconsistent LS and have very poor efficiency too. These problems aggravate, and the IV estimator becomes inconsistent, when the exclusion restriction $\gamma_2 = 0$ is in fact invalid. Then $\gamma_2 z_i^{(2)}$ has to be accommodated by the disturbance term with the effect that orthogonality of this implied disturbance and $z_i^{(2)}$ ceases to be a possibility. Because the equations (4) through (6) do not identify $\gamma_2$ the exclusion restriction cannot be tested from them. Hence, any inference on $\beta$ and $\gamma_1$ based on this restriction is not credible without strong supplementary (from external sources) evidence on its validity.

The exclusion restriction issue refers to the classic problem of identifying the direct (causal) effect of $y^{(2)}$ on $y^{(1)}$ when $y^{(2)}$ itself depends on $y^{(1)}$. To measure this direct effect of endogenous explanatory variable $y^{(2)}$ on $y^{(1)}$ variable $z^{(2)}$ is an effective instrument if: (a) it has no direct effect on $y^{(1)}$ ($\gamma_2 = 0$); (b) but only an indirect effect via $y^{(2)}$ ($\pi_2 \neq 0$). If the effect of $z^{(2)}$ on $y^{(2)}$ is not very strong ($\pi_2 \sigma_{z^{(2)}}$ is relatively small), then the instrument is weak and both $\beta$ and $\gamma_1$ are only weakly identified, which results in poor efficiency plus poor accuracy. If $\pi_2 = 0$ then $\beta$ and $\gamma_1$ are not identified, irrespective whether or not the exclusion restriction $\gamma_2 = 0$ holds. The strength of instruments can always be assessed by estimating the reduced form, but statistical testing of the exclusion restriction is in the present context impossible.

**C. Extending the reduced form equation.** Instead of imposing restrictions, one can also try to extend the information, by disentangling disturbance $v_i$ and exploiting (alleged?) explanatory variables earlier omitted from the reduced form equation. Then this may lead to

---

[2]More consequences for LS of neglected simultaneity are analyzed in Kiviet and Niemczyk (2010).

an extra component $\pi_3 z_i^{(3)}$ in (2) with also an extra corresponding orthogonality condition. Assuming now that $\pi_3 \neq 0$, whereas $z_i^{(3)}$ has coefficient zero in the structural equation, we can find consistent IV estimators for $\beta$, $\gamma_1$ and $\gamma_2$ from the three equations (5), (6) and

$$\sigma_{y^{(1)}z^{(3)}} = \beta\sigma_{y^{(2)}z^{(3)}} + \gamma_1\sigma_{z^{(1)}z^{(3)}} + \gamma_2\sigma_{z^{(3)}}^2. \tag{13}$$

Whether this approach is effective hinges upon the strength of instrument $z_i^{(3)}$ and the credibility of its zero coefficient in the structural equation. As always, the strength can be measured, but it is likely to be poor, given that omission of this variable was first seen as acceptable. Again, the exclusion restriction is untestable (because in the present case it does not concern an overidentification restriction). Hence, faith in its credibility gained from other than statistical sources is again crucial. Note that in the end cases C and B have much in common. Where B achieves identification by imposing a restriction on the structural form, C does by no longer imposing an earlier implicit restriction on the reduced form.

**D. Assuming $\gamma_2 = \gamma_{20}$.** If the exclusion restriction seems inappropriate, and one does no longer aim at estimating $\gamma_2$ because information on its value can be obtained from other sources, then estimates for $\beta$ and $\gamma_1$ can be obtained directly from (5) and (6). This yields

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_{y^{(2)}z^{(1)}} & \hat{\sigma}_{z^{(1)}}^2 \\ \hat{\sigma}_{y^{(2)}z^{(2)}} & \hat{\sigma}_{z^{(1)}z^{(2)}} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\sigma}_{y^{(1)}z^{(1)}} + \gamma_{20}\hat{\sigma}_{z^{(1)}z^{(2)}} \\ \hat{\sigma}_{y^{(1)}z^{(2)}} + \gamma_{20}\hat{\sigma}_{z^{(2)}}^2 \end{pmatrix}$$
$$= (Z'X)^{-1}Z'y^{(1)} + (Z'X)^{-1}Z'Z\begin{pmatrix} 0 \\ \gamma_{20} \end{pmatrix}. \tag{14}$$

Of course, this modified IV estimator is inconsistent when the assumption on $\gamma_2$ is incorrect. However, such an assumption can be made more flexible, for instance, by assuming $\gamma_2$ to be the realization of a random drawing. This approach, and some Bayesian extensions, have been put forward recently by Kraay (2011) and Timothy *et al.* (2011). Although they replace the often very incredible exclusion restriction by a possibly much less restrictive and therefore also more credible assumption on $\gamma_2$, the resulting efficiency and accuracy will still be highly dependent on the validity of the instruments (3) and on their strength.

**E. Assuming $\rho_{y^{(2)}\varepsilon} = \rho_0$.** Yet another option is to make an assumption on actual the degree of simultaneity. This leads to a procedure which does not require to impose any untestable exclusion restrictions, and neither is its efficiency and accuracy dependent on the strength of instruments. More details are given in Kiviet (2011). The essentials are as follows. Defining the $t^{th}$ row of $X$ now to contain $(y_i^{(2)} \; z_i^{(1)} \; z_i^{(2)})$ and substituting $\sigma_{y^{(2)}\varepsilon} = \rho_0\sigma_{y^{(2)}}\sigma_\varepsilon$, the equations (4) through (6) yield

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = (X'X)^{-1}X'y^{(1)} + (X'X)^{-1}\begin{pmatrix} n\rho_0\hat{\sigma}_{y^{(2)}}\hat{\sigma}_\varepsilon \\ 0 \\ 0 \end{pmatrix}. \tag{15}$$

Here $\hat{\sigma}_\varepsilon$ is the square root of the consistent estimator for $\sigma_\varepsilon^2$ given by

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{1 - \rho_0^2[1 - y^{(1)\prime}Z(Z'Z)^{-1}Z'y^{(1)}/y^{(1)\prime}y^{(1)}]} \frac{y^{(1)\prime}[I - X(X'X)^{-1}X']y^{(1)}}{n}. \tag{16}$$

This estimator involves application of LS, which is next modified in such a way that the various moment equations are being respected, securing consistency. At first sight, it seems unfeasible, because it requires the value of the unobservable $\rho_{y^{(2)}\varepsilon}$. But then any MM estimator, including IV, is actually unfeasible, because they are obtained by substituting $\rho_{z^{(j)}\varepsilon} = 0$, whereas $\rho_{z^{(j)}\varepsilon}$ is unobservable too (and, in addition, the exclusion restrictions it requires for just identification are untestable). So, estimator (15) is certainly not necessarily less credible than other MM estimators. Moreover, its credibility can be enhanced by extending the underlying adopted assumption to the interval $\rho_{y^{(2)}\varepsilon} \in [\rho_{y^{(2)}\varepsilon}^L, \rho_{y^{(2)}\varepsilon}^U]$. Of course, the wider this interval is, the more robust (although less efficient) the resulting inference will be. We do admit that (15) is an odd estimator, and it is curious that LS after a modification produces a useful estimator of coefficients which under the usual notions are not even identified. However, making an interval assumption on the simultaneity has apparently sufficient identifying power. All in all, Kinky LS seems an appropriate name for this uncommon estimator.

# 4 Where does KLS lead to?

KLS based inference can only be used effectively in practice by exploiting the limiting distribution of estimator (15), provided this is reasonably accurate for its properties in finite samples. In Kiviet (2011) initial results can be found which demonstrate that under normality of all the regressor variables and the structural disturbance, the expression for the limiting distribution of KLS is remarkably simple. Moreover, in simulation experiments inference in finite samples proves to be highly accurate when built on the true value of the simultaneity coefficient. Already for moderately strong instruments it is more accurate and efficient than IV based inference, and especially so when the instruments are weak. KLS inference can be made more credible and more robust by extending the interval $[\rho_{y^{(2)}\varepsilon}^L, \rho_{y^{(2)}\varepsilon}^U]$. If the true value is in the interval this means that KLS inference will be too prudent (the coverage probability of confidence intervals is much higher than required) but can still often be more efficient than IV inference. Only when the true value of the simultaneity coefficient is not in the adopted interval KLS inference becomes inaccurate and may be worse than IV inference. Presently, work is underway to provide full proofs for cases with an arbitrary number of endogenous and exogenous regressors, supplemented with applications which will throw new light on established empirical results which are based on IV inference exploiting often weak and possibly even invalid instruments. In principle, it will then be possible to test the allegedly untestable exclusion restrictions by KLS!

## References

Conley, G.C., Hansen, C.B., Rossi, P.E., 2011. Plausibly exogenous. *Review of Economics and Statistics*, forthcoming.

Kiviet, J.F., 2011. Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *UvA-Econometrics Discussion Paper 2011/02.*

Kiviet, J.F., Niemczyk, J., 2011. The asymptotic and finite sample (un)conditional distributions of OLS and simple IV in simultaneous equations. *Journal of Computational Statistics and Data Analysis*, forthcoming. (*UvA-Econometrics Discussion Paper 2010/01*).

Kraay, A., 2011. Instrumental variables regressions with uncertain exclusion restrictions: A Bayesian approach. *Journal of Applied Econometrics*, forthcoming.

Larcker, D.F., Rusticus, T.O., 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49, 186-205.