On the limiting and empirical distribution of IV estimators when some of the instruments are invalid

Jan F. Kiviet and Jerzy Niemczyk

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands
On the limiting and empirical distribution of IV estimators when some of the instruments are invalid

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

version of 5 June 2007

JEL-classification: C13, C15, C30

Keywords: empirical density, inconsistent estimators, invalid instruments, limiting distribution, weak instruments

Abstract

We examine IV estimation when instruments may be invalid. This is relevant because validity of the initial just-identification restrictions is untestable. Moreover, tests for the validity of additional instruments, so-called over-identification restriction tests, have limited power when samples are small, especially when instruments are weak. The limiting normal distribution of inconsistent IV is derived and provides a first-order asymptotic approximation to the density in finite sample. In specific simple models we scan this approximation and compare it with the simulated empirical distribution over almost the full parameter space, which is expressed in measures for: model fit, simultaneity, instrument invalidity and instrument weakness. Our major findings are that for the accuracy of large sample asymptotic approximations instrument weakness is much more detrimental than instrument invalidity. IV estimators obtained from strong but possibly invalid instruments are usually much closer to the true parameter values than those obtained from valid but weak instruments.

*Department of Quantitative Economics, Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; phone +31.20.5254217; email J.F.Kiviet@UvA.NL and J.Niemczyk@UvA.NL
1 Introduction

When in a regression model some of the explanatory variables are contemporaneously correlated with the disturbance term, and this correlation is unknown, then one needs further variables in order to find consistent estimators by the method of moments. These instrumental variables should have a known (usually zero) correlation with the disturbances. If this is the case they provide moment conditions (called orthogonality conditions when the correlation is zero) that make it possible to obtain consistent instrumental variable (IV) estimators. In practice, however, it is usually difficult to assess whether an instrumental variable is indeed valid, i.e. is uncorrelated with the disturbance term. Firstly, instrument validity or orthogonality tests are only viable under just identification or overidentification by truly valid instruments. That is, they are based on the prerequisite of having already available a number of undisputed valid instruments, at least as great as the number of coefficients \( k \) to be estimated, whereas the validity of the initial \( k \) instruments is untestable. Moreover, orthogonality tests will have reasonable power only when the instruments employed and those under test are not too weak (are sufficiently correlated with the regressors) and the sample size is substantial. Therefore, it seems very likely that IV estimation will often be employed when some of the instruments are in fact invalid. In this case the IV estimator for the structural parameters is inconsistent, even when the structural equation itself is correctly specified for the parameters of interest.

In this paper we consider general and specific forms of linear structural equations and corresponding partial reduced form systems in stationary variables and examine the IV estimator when some of its exploited orthogonality conditions actually do not hold. We cover the general case where the number of moment conditions exploited \( l \), thus the valid and invalid conditions together, is at least as large as the number of unknown coefficients, i.e. we consider the (alleged) over or just identified case \( l \geq k \). We focus on the distribution of such an invalid IV estimator for a single structural equation that otherwise has been specified correctly\(^1\) in the sense that its implied series of error terms is IID (independent and identically distributed) with unconditional expectation equal to zero. We derive an expression for the inconsistency of the IV estimator in terms of parameters and data moments and also obtain its limiting distribution. These results yield a first-order asymptotic approximation to the actual distribution of inconsistent IV estimators in finite sample. The asymptotic variance proves to be a rather complicated expression, although it can be substantially simplified when specialized for the just identified case \( l = k \). Over the relevant parameter space of simple classes of models we verify by simulation whether these analytic findings are accurate regarding the actual estimator distribution in finite sample and demonstrate how these depend on the various model parameters.

In the illustrations we focus first on a very simple specific type of model, entailing just one explanatory variable and one instrument. Here we show that all our findings, both the analytic asymptotic and the simulated finite sample results, are driven by just four primal econometric model characteristics, in addition to the sample size. These four characteristics are straightforward transformations of the underlying parameters of

\(^1\) An alternative point of departure is chosen in Hale et al. (1980) where instruments are invalid due to omitted regressors.
the data generating process. They are all related to particular correlation coefficients, viz: (i) the model fit, (ii) the degree of simultaneity, (iii) the degree of invalidity of the instrument, and (iv) the degree of instrument weakness. Thus, even in the simple one-regressor one-instrument model, the distributional properties of the IV estimator are functions involving five arguments, which makes it difficult to depict their behavior over all relevant argument values. Instead of presenting extensive tables, we present a few series of 2D and 3D graphs in print, and we use dynamic multi-dimensional visualization techniques to present our findings more elegantly and effectively on screen through animations. We also present some results for an (alleged) overidentified one-regressor model with two instruments. Here we find that both the actual finite sample distribution and its asymptotic approximation can be expressed using just one extra argument, whereas at first sight one might conjecture that both the invalidity and the strength of the extra instrument would matter separately.

The analysis of IV estimators employing invalid instruments has not yet received much attention in the literature. Although its limiting behavior has been examined by Maasumi and Phillips (1982), Hahn and Hausman (2003) and Hall and Inoue (2003), none of these studies provides an explicit formula for the asymptotic variance in the general linear multivariate case, where \( l \geq k \geq 1 \). Such a formula is obtained here by extending an approach\(^2\) that yielded similar results for inconsistent OLS estimators in Kiviet and Niemczyk (2007), which completes some initial results obtained in Joseph and Kiviet (2005). As far as we are aware no simulation evidence has yet been published as presented here on the actual finite sample distribution of invalid IV estimators, covering almost the entire parameter space for some basic models. The analysis of the exact finite sample properties of consistent IV estimators has a long history. Early contributions are Sawa (1969) and Phillips (1980). More recent contributions (and further references) can be found in, for instance, Phillips (2006) and Hillier (2006). Our findings illustrate the effects of instrument weakness on the finite sample density of consistent IV estimators, which have been studied before by Woglom (2001), who focusses on just identified IV estimators, and by Forchini (2006), who gives further theoretical underpinnings in case of overidentification. In addition, however, we supplement these findings with extensive illustrations for the case of invalid instruments. Whereas much of the recent literature on weak instruments focusses on developing appropriate tests and confidence sets when instruments are weak but valid, see for instance Hahn and Inoue (2002) and Andrews and Stock (2007), the present study analyzes and illustrates properties of the distribution of coefficient estimators when instruments may be weak and invalid.

From our simulations we establish that invalid but reasonably strong instruments yield IV estimators which have a distribution in small samples that is rather close to the analytic large-sample asymptotic approximations derived here. Hence, the distribution of these estimators is often close to normal, but has its probability mass centered around the pseudo-true-value instead of the true value. However, when instruments are very weak, we establish that the accuracy of standard large-sample asymptotics is very poor, as had already been established for the valid instrument case. More importantly, though, for both valid and invalid instruments we also find that when the instrument is weak the probability mass of the actual distribution of instrumental variable estimators is generally much closer to the true value of the coefficient than indicated by these

\(^2\)A related approach can be found in an unplished discussion paper by Rothenberg (1972).
much too flat asymptotic approximations. For valid but rather weak instruments it had already been established that the finite sample distribution of IV can be skewed, and that it becomes bimodal for very strong simultaneity, whereas for extreme weakness (i.e. close to underidentification) the dispersion explodes and the median moves away from the true parameter value towards the probability limit of OLS. We find that for invalid weak instruments skewness, bimodality and a median away from the pseudo-true-value may occur for much more moderate weakness and simultaneity. Note, however, that in practice one can easily avoid using weak instruments, since weakness (unlike validity) can straight-forwardly be assessed. Because the invalid IV estimator is reasonably well behaved for reasonably strong instruments, a tentative conclusion is that it seems more promising to attempt to produce accurate inference from IV estimators based on strong (as in OLS) but possibly invalid instruments, than on valid but weak instruments. In the latter case, not only is the standard asymptotic approximation poor, but also the actual behavior of the distribution of the estimator is rather erratic and has much larger estimation errors than invalid but strong instruments produce. Thus, even when its actual behavior could be adequately approximated by alternative weak-instrument asymptotic methods, the valid IV estimator may still have a less attractive actual distribution than that based on strong but possibly invalid instruments.

The structure of this paper is as follows. In Section 2 we introduce the model and the generating schemes for all explanatory and instrumental variables, with their underlying statistical assumptions. Focussing on the alleged overidentified case we consider the generalized IV or 2SLS estimator and derive its inconsistency and limiting distribution (proofs in appendices) for the case where all variables are weakly stationary, i.e. their first and second moments are constant through time. The results are also specialized for the just identified case and are compared with those obtained by Hall and Inoue (2003). Section 3 contains graphic illustrations of both the asymptotic and finite sample distributions in specific simple models. From these we examine the accuracy of the asymptotic approximations and their actual behavior over different values of all the various determining factors. Moreover, we compare the effectiveness of IV with respect to OLS, which uses always extremely strong but possibly invalid instruments. In separate subsections we consider models with $l = k = 1$ and with $l - 1 = k = 1$. Substantial attention is paid to obtaining a transparent design of the parameter space for generating the data for these models and instruments and to establishing any invariance and symmetry properties of the estimators over this parameter space. Section 4 concludes.

2 Model, assumptions and theorems

We consider data generating processes for variables for which $n$ observations have been collected in the rows of $y$, $X$ and $Z$. The matrices $X$ and $Z$ have $k$ and $l$ columns respectively, with $l \geq k$. $X$ contains the explanatory variables for the vector $y$ in a linear structural model with structural disturbance vector $\varepsilon$. The $l$ variables collected in $Z$ will be used as instrumental variables for estimating the $k$ structural parameters of interest $\beta$. Not all these instruments are necessarily valid, some of them may be weak, whereas others may be extremely strong, especially when columns of $X$ correspond to (or are spanned by) columns of $Z$. The basic framework is characterized by the following parametrization and regularity conditions, which involve linearity and stationarity.
Framework A. We have: (i) the structural equation \( y = X\beta + \varepsilon \); (ii) with disturbances having (for \( i \neq h = 1, \ldots, n \)) the (finite) unconditional moments \( \mathbb{E}(\varepsilon_i) = 0, \mathbb{E}(\varepsilon_i\varepsilon_h) = 0, \mathbb{E}(\varepsilon_i^2) = \sigma_\varepsilon^2, \mathbb{E}(\varepsilon_i^3) = \mu_3\sigma_\varepsilon^3 \) and \( \mathbb{E}(\varepsilon_i^4) = \mu_4\sigma_\varepsilon^4 \); (iii) while \( \mathbb{E}(X | \varepsilon) = \varepsilon\xi' \) and \( \mathbb{E}(Z | \varepsilon) = \varepsilon\zeta' \), with \( \xi \) and \( \zeta \) fixed parameter vectors of \( k \) and \( l \) elements respectively. Moreover, (iv) \( \Sigma_{XX} = \text{plim}_{n \to \infty} n^{-1}X'X, \Sigma_{ZZ} = \text{plim}_{n \to \infty} n^{-1}Z'Z \) and \( \Sigma_{ZX} = \text{plim}_{n \to \infty} n^{-1}Z'X \) have all full column rank, and (v) so have \( X'X, Z'Z \) and \( Z'X \) with probability one. Finally, (vi) we have \( \mathbb{E}(n^{-1}Z'X | Z) - \text{plim}_{n \to \infty} n^{-1}Z'X = \sigma_p(n^{-1/2}) \) and \( \mathbb{E}(n^{-1}X'Z | X, Z) - \Sigma_{XX} = \sigma_p(n^{-1/2}) \), where \( X \equiv X - \mathbb{E}(X | \varepsilon) \) and \( Z \equiv Z - \mathbb{E}(Z | \varepsilon). \)

Note that the latter definitions imply the decompositions \( X \equiv \bar{X} + \varepsilon\xi' \) and \( Z \equiv \bar{Z} + \varepsilon\zeta' \). From A(iii) we find \( \mathbb{E}(\bar{X}'\varepsilon) = 0 \) and \( \mathbb{E}(\bar{Z}'\varepsilon) = 0 \), whereas

\[
\mathbb{E}(X'\varepsilon) = n\sigma_\varepsilon^2 \xi \quad \text{and} \quad \mathbb{E}(Z'\varepsilon) = n\sigma_\varepsilon^2 \zeta.
\]  

Hence, if \( \xi_j = 0 \) for some \( j \in \{1, \ldots, k\} \) then the \( j \)-th regressor in \( X \) is predetermined and will establish a valid instrument; otherwise, when \( \xi_j \neq 0 \), the \( j \)-th regressor is endogenous. Likewise, if \( \zeta_g = 0 \) for some \( g \in \{1, \ldots, l\} \) then the \( g \)-th column of \( Z \) establishes a valid instrument, and an invalid instrument otherwise. It can be shown that A(vi) boils down to the mild regularity assumptions \( \frac{1}{n}\bar{Z}'Z - \text{plim}_{n \to \infty} n^{-1}Z'Z = \sigma_p(n^{-1/2}) \) and \( \frac{1}{n}\bar{X}'\bar{Z} - \text{plim}_{n \to \infty} n^{-1}\bar{X}'\bar{Z} = \sigma_p(n^{-1/2}) \).

Since \( l \geq k \) the generalized instrumental variable (GIV) or 2SLS estimator of \( \beta \) exists and is given by

\[
\hat{\beta}_{GIV} = \frac{[X'Z'(Z'Z)^{-1}Z'X]\cdot[X'Z(Z'Z)^{-1}Z'y]}{(X'X)^{-1}X'y},
\]

where we introduced the notation

\[
\hat{X} \equiv \bar{Z}\hat{\Pi} = (Z'Z)^{-1}Z'X,
\]

where \( \hat{\Pi} = (Z'Z)^{-1}Z'X \) contains the (reduced form) coefficient estimates of the first-stage regressions. In Framework A the probability limit of \( \hat{\beta}_{GIV} \) exists. We define

\[
\beta^*_{GIV} \equiv \text{plim} \hat{\beta}_{GIV} = \beta + \sigma_\varepsilon^2[\Sigma_{XX}\Sigma_{ZZ}^{-1}\Sigma_{ZX}]^{-1}\Sigma_{XX}\Sigma_{ZZ}^{-1}\zeta,
\]

where \( \beta^*_{GIV} \) is also known as the pseudo-true-value of \( \hat{\beta}_{GIV} \). We shall denote the inconsistency of \( \hat{\beta}_{GIV} \) as

\[
\hat{\beta}_{GIV} \equiv \beta^*_{GIV} - \beta = \sigma_\varepsilon^2\Sigma_{XX}^{-1} \hat{\Pi}'\zeta,
\]

where we used \( \Sigma_{\hat{X}'\hat{X}} = \Sigma_{XX}\Sigma_{ZZ}^{-1}\Sigma_{ZX} \) and \( \hat{\Pi} = \text{plim}(Z'Z)^{-1}Z'X = \Sigma_{ZZ}^{-1}\Sigma_{ZX} \). Note that in Framework A the GIV estimator is consistent if and only if \( \zeta = 0 \).
Below, we shall also look into the special case \( l = k \) (just identification), where the above GIV results specialize to simple IV, i.e.

\[
\begin{align*}
\hat{\beta}_{IV} &= (Z'X)^{-1}Z'y, \\
\beta^*_{IV} &= \beta + \sigma^2_{\varepsilon} \Sigma^{-1}_{X'X} \xi, \\
\tilde{\beta}_{IV} &= \sigma^2_{\varepsilon} \Sigma^{-1}_{Z'X} \xi.
\end{align*}
\]

When in fact \( Z = X \) (all regressors are used as instruments), i.e. \( \zeta = \xi \), then IV specializes to OLS, i.e.

\[
\begin{align*}
\beta_{OLS} &= (X'X)^{-1}X'y, \\
\beta^*_{OLS} &= \beta + \sigma^2_{\varepsilon} \Sigma^{-1}_{X'X} \xi, \\
\tilde{\beta}_{OLS} &= \sigma^2_{\varepsilon} \Sigma^{-1}_{X'X} \xi.
\end{align*}
\]

For the sake of simplicity, we start with deriving special results for models with disturbances that have 3rd and 4th moments corresponding to those of the normal distribution. Therefore, we state:

**Framework B.** This specializes Framework A to the case: \( \mu_3 = 0 \) and \( \mu_4 = 3 \).

For GIV estimators we obtain the following result (proofs in appendices) on its convergence in distribution.

**Theorem 1.** In Framework B we have \( n^{1/2} (\hat{\beta}_{GIV} - \beta^*_{GIV}) \to N(0, V^N_{GIV}) \), with

\[
V^N_{GIV} = \sigma^2_{\varepsilon} c_3 (1 - c_3 + c_4) \Sigma^{-1}_{X'X} + \sigma^2_{\varepsilon} c_4 \Sigma^{-1}_{Z'X} \Sigma_{X'X} \Sigma^{-1}_{X'X} \\
- c_4 \Sigma^{-1}_{X'X} \Sigma_{X'X} \hat{\beta}'_{GIV} + \hat{\beta}'_{GIV} \Sigma_{X'X} \Sigma^{-1}_{X'X} \\
+ \sigma^2_{\varepsilon} c_4 (1 - 2c_4) \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X} \\
+ \sigma^2_{\varepsilon} c_4 (1 - 2c_5) \Sigma^{-1}_{X'X} \xi' \hat{\beta}'_{GIV} + \hat{\beta}'_{GIV} \Sigma_{X'X} \xi' \Sigma^{-1}_{X'X} \\
+ [c_5 (1 - 2c_5) - c_3 + \sigma^2_{\varepsilon} \Sigma^{-1}_{X'X} \hat{\beta}'_{GIV} \Sigma_{X'X} \hat{\beta}'_{GIV}] \beta_{GIV} \beta'_{GIV},
\]

where \( c_1 \equiv \sigma^2_{\varepsilon} \xi' \Sigma^{-1}_{Z'Z} \xi, c_2 \equiv \xi' \Pi \hat{\beta}_{GIV}, c_3 \equiv \xi' \beta_{GIV}, c_4 \equiv c_1 - c_2 \) and \( c_5 \equiv 1 - c_3 - c_4 \).

The \( N \) in the superindex of \( V^N_{GIV} \) indicates that it refers to the case where the disturbances are "almost normal", because \( \mu_3 = 0 \) and \( \mu_4 = 3 \). We find that the limiting distribution of \( \hat{\beta}_{GIV} \) is still genuinely normal when instruments are invalid, although no longer centered at \( \beta \) but at the pseudo-true-value \( \beta^*_{GIV} \). When all instruments are valid, i.e. \( \zeta = 0 \), then \( \beta^*_{GIV} = \beta, \tilde{\beta}_{GIV} = \beta^*_{GIV} = 0 \) and \( c_1 = c_2 = c_3 = 0 \), giving \( c_4 = 0 \) and \( c_5 = 1 \), so that Theorem 1 specializes to the standard result \( n^{1/2} (\hat{\beta}_{GIV} - \beta) \to N(0, \sigma^2_{\varepsilon} \Sigma^{-1}_{X'X}) \). Note that when all instruments are valid the asymptotic variance of \( \hat{\beta}_{GIV} \) is not determined by the simultaneity \( \xi \), because \( \Sigma_{X'X} = \Sigma_{X'X} \Sigma^{-1}_{Z'Z} \Sigma_{X'X} \Sigma_{Z'Z} \Sigma^{-1}_{Z'X} \). However, when instruments are invalid, i.e. \( \zeta \neq 0 \), then \( \Sigma_{X'X} = \Sigma_{X'X} + \sigma^2_{\varepsilon} \zeta \xi' \) and thus \( \Sigma_{X'X} \) is determined by both \( \xi \) and \( \zeta \). When fitting \( X \) to \( Z \) while the \( \varepsilon' \xi \) part of \( X \) is not (asymptotically) orthogonal to \( Z \), due to the presence of \( \varepsilon' \xi \), this does not only lead to the inconsistency, but also to the many extra terms in the asymptotic variance.
For the special case \( l = k \) we have \( \zeta' \Pi \tilde{\beta}_{GIV} = \sigma^2 \zeta' \Sigma^{-1}_{Z'X} \Sigma^{-1}_{Z'X} \zeta = \sigma^2 \zeta' \Sigma^{-1}_{Z'X} \zeta \), so \( c_1 = c_2, c_4 = 0 \) and \( c_5 = 1 - c_3 \), giving:

**Corollary 1.** In Framework B for the special case \( l = k \) we have \( n^{1/2}(\beta_{IV} - \beta^*_{IV}) \to N(0, V_{IV}^N) \), with

\[
V_{IV}^N = \sigma^2(1 - c_3) \Sigma^{-1}_{Z'X} \Sigma^{-1}_{Z'X} \Sigma^{-1}_{Z'X} - [2c_3 - 2c_3 + 1 - \sigma^2(\beta_{IV}' \Sigma^{-1}_{X'X} \beta_{IV})] \tilde{\beta}_{IV} \tilde{\beta}_{IV}',
\]

where \( c_3 \equiv \xi' \tilde{\beta}_{IV} = \sigma^2 \xi' \Sigma^{-1}_{Z'X} \zeta \).

When all instruments are valid, i.e. \( \zeta = 0 \), this result specializes to the standard result \( n^{1/2}(\beta_{IV} - \beta^*_{IV}) \to N(0, \sigma^2 \Sigma^{-1}_{Z'X} \Sigma^{-1}_{Z'X} \Sigma^{-1}_{Z'X}) \). Since for arbitrary \( \zeta \) and \( \xi \) the scalar \( \sigma^2 \xi' \Sigma^{-1}_{Z'X} \zeta \) can either be positive or negative, no general conclusions can be drawn on the behavior of \( V_{IV}^N \) in comparison to the reference case \( \sigma^2 \Sigma^{-1}_{Z'X} \Sigma^{-1}_{Z'X} \Sigma^{-1}_{Z'X} \). Depending on the particular parametrization and data moment matrices the asymptotic variance of individual coefficient estimates may either increase or decrease, due to \( \xi \neq 0 \) and \( \zeta \neq 0 \).

When \( Z = X \), which gives \( \zeta = \xi \) and \( \beta_{IV} = \beta_{OLS} \), the resulting \( V_{IV}^N = V_{OLS}^N \) is the same as the formula found for an inconsistent OLS estimator when the disturbances are (almost) normal, as derived in Kiviet and Niemczyk (2007).

Next, we look at the case where the disturbances may have general 3rd and 4th moment. Let \( \iota \) be a \( n \times 1 \) vector of unit elements. Upon defining

\[
\Sigma_{Z'i} = \text{plim} n^{-1} Z'i = \text{plim} n^{-1} Z'i \equiv \Sigma_{Z'i},
\]

\[
\Sigma_{X'i} = \text{plim} n^{-1} X'i = \text{plim} n^{-1} X'i \equiv \Sigma_{X'i},
\]

we find (superindex \( NN \) indicates nonnormal disturbances):

**Theorem 2.** In Framework A we have \( n^{1/2}(\beta_{GIV} - \beta^*_{GIV}) \to N(0, V_{GIV}^{NN}) \), where \( V_{GIV}^{NN} \) is equal to \( V_{GIV}^N \), given in Theorem 1, plus two additional terms. When \( \mu_4 \neq 3 \) the additional term is

\[
(\mu_4 - 3)\{\sigma^2 \zeta' \Sigma^{-1}_{X'X} \xi \xi' \Sigma^{-1}_{X'X} + \sigma^2 c_4 c_5 [\beta_{GIV} \xi' \Sigma^{-1}_{X'X} + \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X} + \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}] \}
\]

where \( c_4 \equiv \sigma^2 \zeta' \Sigma^{-1}_{Z'Z} \zeta - \zeta' \Pi \tilde{\beta}_{GIV} \) and \( c_5 \equiv 1 - c_4 - \xi' \tilde{\beta}_{GIV} \). When \( \mu_3 \neq 0 \) the additional term is

\[
\mu_3 \{c_4 [\sigma^2 c_4 \Sigma^{-1}_{X'X} \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} - \sigma^2 \beta_{GIV} \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} + \sigma^2 c_5 \Sigma^{-1}_{X'X} \Pi \Sigma_{Z'Z} \xi' \Sigma^{-1}_{X'X}]
\]

\[
+ (\sigma^2 \Sigma^{-1}_{X'X} \xi - \zeta' \Pi \tilde{\beta}_{GIV}) (\sigma^2 \zeta' - \beta_{GIV}' \Sigma_{X'X}) \Sigma_{Z'Z} \Sigma_{Z'Z} \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}]
\]

\[
+c_5 [\sigma^2 c_4 \Sigma^{-1}_{X'X} \xi \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} - \sigma^2 \beta_{GIV} \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} + \sigma^2 c_5 \Sigma^{-1}_{X'X} \Pi \Sigma_{Z'Z} \xi' \Sigma^{-1}_{X'X}]
\]

\[
+ (\sigma^2 \Sigma^{-1}_{X'X} \xi - \zeta' \Pi \tilde{\beta}_{GIV}) (\sigma^2 \zeta' - \beta_{GIV}' \Sigma_{X'X}) \Sigma_{Z'Z} \Sigma_{Z'Z} \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}]
\]

\[
+c_4 [\sigma^2 c_4 \Sigma^{-1}_{X'X} \xi \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} - \sigma^2 \beta_{GIV} \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} + \sigma^2 c_5 \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}]
\]

\[
+ \Sigma^{-1}_{X'X} \xi' \Sigma_{Z'Z} \Sigma_{Z'Z} \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}
\]

\[
+c_5 [\sigma^2 c_4 \Sigma_{GIV} \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} - \sigma^2 \beta_{GIV} \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} + \sigma^2 c_5 \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}]
\]

\[
+ (\sigma^2 \Sigma^{-1}_{X'X} \xi - \zeta' \Pi \tilde{\beta}_{GIV}) (\sigma^2 \zeta' - \beta_{GIV}' \Sigma_{X'X}) \Sigma_{Z'Z} \Sigma_{Z'Z} \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}]
\]

\[
+c_5 [\sigma^2 c_4 \Sigma_{GIV} \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} - \sigma^2 \beta_{GIV} \Sigma_{X'i} \xi' \Sigma^{-1}_{X'X} + \sigma^2 c_5 \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}]
\]

\[
+ (\sigma^2 \Sigma^{-1}_{X'X} \xi - \zeta' \Pi \tilde{\beta}_{GIV}) (\sigma^2 \zeta' - \beta_{GIV}' \Sigma_{X'X}) \Sigma_{Z'Z} \Sigma_{Z'Z} \Sigma^{-1}_{X'X} \xi' \Sigma^{-1}_{X'X}]
\]
When all the instruments are valid, i.e. \( \zeta = 0 \), this result again collapses to the standard one, i.e. \( V_{GIV}^{NN} = \sigma_{\xi}^2 \Sigma_{X'X}^{-1} \), which highlights that normality of the disturbances is not a requirement for the standard normal limiting distribution of \( \hat{\beta}_{GIV} \).

For the special case \( l = k \) Theorem 2 yields:

**Corollary 2.** In Framework A for the special case \( l = k \) we have \( n^{1/2}(\hat{\beta}_{IV} - \beta_{IV}) \to N(0, V_{IV}^{NN}) \), with

\[
V_{IV}^{NN} = \sigma_{\xi}^2 c_3 \Sigma_{Z'X}^{-1} \Sigma_{Z'Z} \Sigma_{X'X}^{-1} + \mu_3 \sigma_{\xi}^2 (\Sigma_{Z'X}^{-1} \Sigma_{Z'Z} \Sigma_{X'X}^{-1} + \hat{\beta}_{IV} \Sigma_{Z'Z} \Sigma_{X'X}^{-1} \hat{\beta}_{IV}') - [(5 - \mu_4) c_5^2 + 2c_5 (\mu_3 \sigma_{\xi}^2 \hat{\beta}_{IV}' \Sigma_{X'X}^{-1} - 1) + 1 - \sigma_{\xi}^2 \hat{\beta}_{IV}' \Sigma_{X'X} \hat{\beta}_{IV}' \hat{\beta}_{IV}'.
\]

where \( c_5 \equiv 1 - c_4 \xi' \Sigma_{Z'X}^{-1} \xi \).

Of course, for \( \mu_3 = 0 \) and \( \mu_4 = 3 \) this result simplifies to that of Corollary 1. It also shows that an increase (decrease) in the kurtosis leads to a larger (smaller) asymptotic variance.

In the proofs of the above theorems we employ a lemma which is a straightforward extension of the simple CLT (central limit theorem): Let \( v_i \) be a \( k \times 1 \) random vector such that \( E(\v_i) = 0, E(\v_i \v_i') = V_i \) and \( E(\v_i \v_i') \to O \) for \( i \neq h = 1, \ldots, n \), then \( n^{1/2} \tilde{V} \to N(0, \text{lim}_{n \to \infty} V) \), where \( \tilde{V} = n^{-1} \sum_{i=1}^{n} V_i \) and \( V = n^{-1} \sum_{i=1}^{n} V_i \). We employ the following generalized version:

**Lemma.** Let \( W = (w_1, \ldots, w_n)' \) be a \( n \times k \) random matrix and \( \omega \) a \( k \times 1 \) nonrandom vector, whereas the \( n \times 1 \) vector \( \v = (\v_1, \ldots, \v_n)' \) has mutually uncorrelated elements for which \( E(\v_i \mid w_i) = 0, E(\v_i \v_i') = \sigma_i^2, E(\v_i \mid w_i) = \mu_i \sigma_i^2 \) and \( E(\v_i') = \mu_i \sigma_i^2 \). Then the \( k \times 1 \) vector \( v_i = w_i \v_i + \omega (\v_i - \sigma_i^2) \) has zero expectation, conditional variance \( E(\v_i | w_i) = V_i = \sigma_i^2 w_i \omega_i + \mu_i \sigma_i^2 (w_i \omega_i + \omega w_i') + (\mu_i - 1) \sigma_i^4 \omega \omega' \), whereas \( E(\v_i \v_i') = O \) for \( i \neq h \), so that for \( n^{-1/2} \sum_{i=1}^{n} v_i = n^{-1/2} [W \v + \omega (\v - \nabla \v)] \) the CLT implies

\[
n^{1/2} \tilde{V} \to N(0, \sigma_i^2 \Sigma_{W'W} + \mu_i \sigma_i^2 (\Sigma_{W'\omega} + \omega \Sigma_{\v W}) + (\mu_i - 1) \sigma_i^4 \omega \omega'),
\]

where \( \Sigma_{W'W} \equiv \text{plim} \ n^{-1} W'W \) and \( \Sigma_{W'\omega} \equiv \text{plim} \ n^{-1} W' \omega \equiv \Sigma_{\omega W} ' \) with \( \omega \) a \( n \times 1 \) vector of unit elements.

The results in Theorems 1 and 2 in fact address special cases of Theorem 2 in Hall and Inoue (2003, p.369). The latter theorem is more general, but more implicit at the same time. It concerns GMM estimation (both 1-step and 2-step) of a possibly nonlinear misspecified model and expresses its asymptotic variance matrix in a few model characteristics and a matrix \( \Omega \). This matrix, which is not further specified, is the variance of the limiting distribution of a \( 2k + l \) vector, say \( \nu^* \), which in our particular setting specializes to

\[
\nu^* \equiv \begin{pmatrix}
  n^{-1/2} [Z'(y - X \hat{\beta}_{GIV}) - \text{plim} \ n^{-1} Z'(y - X \hat{\beta}_{GIV})]
n^{-1/2} (n^{-1} X'Z - \Sigma_{Z'Z}) \Sigma_{Z'Z}^{-1} \text{plim} \ n^{-1} Z'(y - X \hat{\beta}_{GIV})
n^{-1/2} \left( (n^{-1} Z')^{-1} - \Sigma_{Z'Z}^{-1} \right) \text{plim} \ n^{-1} Z'(y - X \hat{\beta}_{GIV})
\end{pmatrix} \to N(0, \Omega).
\]

It can be shown that

\[
n^{1/2}(\hat{\beta}_{GIV} - \beta_{GIV}) = \Sigma_{X'X}^{-1} \left( \Pi' \ 1_k \ \Sigma_{X'Z} \right) \nu^* + o_p(1),
\]

(8)
and thus the asymptotic variance of GIV can be expressed in $\Sigma_{Z'X}$, $\Sigma_{Z'Z}$ and $\Omega$ (which is itself determined by $\Sigma_{X'X}$ and $\Sigma_{Z'Z}$).

Our approach allows us to evaluate explicitly the entire covariance matrix $\Omega$, and consequently the asymptotic variance of GIV in terms of the data moments $\Sigma_{X'X}$, $\Sigma_{Z'X}$ and $\Sigma_{Z'Z}$ and the further model characteristics, viz: $\xi$, $\zeta$, $\beta$, $\sigma^2_\zeta$, $\mu_3$ and $\mu_4$. Therefore, the advantage of our results is that some general conclusions could be drawn on the joint effects of simultaneity, instrument invalidity and disturbance kurtosis on the limiting distribution of GIV. Moreover, our results allow to depict the numerical effects of those various characteristics, as we will show below.

### 3 Illustrations

To illustrate the analytical asymptotic findings obtained in the foregoing section, we will calculate the various formulas for particular models and show the corresponding normal densities over relevant parts of the parameter space. In addition, we will simulate these models and depict the empirical density of the estimators to check the relevance and accuracy of the first-order asymptotic approximations in finite sample. We also will compare IV and GIV estimators (using possibly invalid and possibly weak instruments) with OLS. The latter estimator always uses extremely strong instruments that at the same time are invalid in case of simultaneity.

The limiting distributions obtained in the foregoing section are all of the generic form $n^{1/2}(\hat{\beta} - \beta_0) \rightarrow N(0, V)$ and they imply a first-order approximation to the distribution of $\hat{\beta}$ in finite sample which can be expressed as

$$\hat{\beta} \overset{a}{\sim} N(\beta + \bar{\beta}, n^{-1}V). \tag{9}$$

This entails a first-order asymptotic approximation to the mean error of $\hat{\beta}$ equal to $\bar{\beta} = \beta^* - \beta$ and to the mean squared error (AMSE) given by

$$\text{AMSE}(\hat{\beta}) \equiv n^{-1}V + \bar{\beta}^2. \tag{10}$$

The actual values of $\bar{\beta}$ and of (the square root of) AMSE($\hat{\beta}$) can be computed for any $n$ and any given values of the model parameters and asymptotic data moments. To find out how accurate the first-order asymptotic approximation (10) is, it should be compared with corresponding Monte Carlo estimates obtained from a series of realizations of $\hat{\beta}$ in simulated finite samples. However, these cannot be achieved in the standard way when $\hat{\beta}$ does not have finite first or second moments in finite sample, as is the case when $l - k \leq 1$. Then, irrespective of the number of Monte Carlo replications employed, the sample moments from Monte Carlo experiments are not informative as they do not converge. Appropriate alternatives for the mean error and for the root mean squared error are then the median error and the median of the absolute error.

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

$$\Pr\{(\hat{\beta} - \beta) \leq \text{ME}(\hat{\beta})\} = 0.5,$$
$$\Pr\{|\hat{\beta} - \beta| \leq \text{MAE}(\hat{\beta})\} = 0.5. \tag{11}$$
From a series of $R$ independent Monte Carlo realizations $\hat{\beta}^{(r)}$ ($r = 1, ..., R$) we estimate $\text{ME}(\hat{\beta})$ by sorting the values $(\hat{\beta}^{(r)} - \beta)$ and taking the median value, and similarly for $\text{MAE}(\hat{\beta})$, taking the median of the sorted $|\hat{\beta}^{(r)} - \beta|$ values. Of course, $\text{AMSE}(\hat{\beta})$ is not the natural asymptotic counterpart of the Monte Carlo estimate of $\text{MAE}(\hat{\beta})$. We assess the (scalar) asymptotic version $\text{AMAE}(\hat{\beta})$ of $\text{MAE}(\hat{\beta})$ in the following way. Let the CDF of the normal approximation to the distribution of $\hat{\beta}$ be indicated by $\Phi_{\hat{\beta}, \sigma_{\hat{\beta}}}(x)$.

Then, for $m \equiv \text{AMAE}(\hat{\beta})$, we have

$$0.5 = \Pr\{|\hat{\beta} - \beta| \leq m\} = 1 - \Pr\{|\hat{\beta} - \beta| > m\} = 1 - \Pr\{\hat{\beta} - \beta > m\} - \Pr\{\hat{\beta} - \beta < -m\} = \Phi_{\hat{\beta}, \sigma_{\hat{\beta}}}(m) - \Phi_{\hat{\beta}, \sigma_{\hat{\beta}}}(-m),$$

so that we can solve\(^3\) for $m$

$$\Phi_{\hat{\beta}, \sigma_{\hat{\beta}}}(m) = 0.5 + \Phi_{\hat{\beta}, \sigma_{\hat{\beta}}}(-m). \tag{12}$$

Below we will examine the empirical finite sample distribution of scalar $\hat{\beta}_{\text{GIV}}$ and compare it with $\hat{\beta}_{\text{GIV}} \sim N(\beta + \hat{\beta}_{\text{GIV}}, n^{-1}V_{\text{GIV}}^{N})$. In addition, for various estimators $\hat{\beta}_{\text{GIV}}$ (including $\hat{\beta}_{\text{IV}}$ and $\hat{\beta}_{\text{OLS}}$), we examine $\text{MAE}(\hat{\beta}_{\text{GIV}})$ and compare it with $\text{AMAE}(\hat{\beta}_{\text{GIV}})$ over almost the entire parameter space of two simple classes of models. Employing normally distributed disturbances\(^4\), we examined these models under Framework B only.

### 3.1 A simple just identified model

We commence by considering the most basic example one can think of, viz. a model with one regressor ($k = 1$) and one instrument ($l = 1$), which is either weak or strong and possibly invalid. The two variables $x$ and $z$, together with the dependent variable $y$, are supposed to be jointly IID with zero mean and finite second moments. Hence, the variables are strongly stationary and our Theorems 1 and 2 apply. This model has often been examined in the past. Recently in Woglom (2001) and Hillier (2006), and for $l \geq 1$ in Bound et al. (1995) and Hahn and Hausman (2003). Though, only the latter paper considers invalid instruments.

We first evaluate the relevant expressions for the asymptotic distribution given in Corollary 1. In the model with $k = l = 1$ we can simplify the notation considerably, by writing $\sigma_{x}^{2}$ for $\Sigma_{X,X}$, $\sigma_{xz}$ or $\rho_{xz}\sigma_{x}\sigma_{z}$ for $\Sigma_{Z,X}$, etc. Using $\xi = \sigma_{xz}/\sigma_{\varepsilon}^{2}$ and $\xi = \sigma_{xz}/\sigma_{\varepsilon}^{2}$ we

\(^3\)Since $m = \Phi^{-1}_{\hat{\beta}, \sigma_{\hat{\beta}}}[0.5 + \Phi_{\hat{\beta}, \sigma_{\hat{\beta}}}(-m)]$, we employed the iterative scheme, $m_{0} = 0$, $m_{i+1} = \Phi^{-1}_{\hat{\beta}, \sigma_{\hat{\beta}}}[0.5 + \Phi_{\hat{\beta}, \sigma_{\hat{\beta}}}(-m_{i})]$ for $i = 0, 1, \ldots$ until convergence. When $\hat{\beta} = 0$ no iteration is required since $m = \Phi^{-1}_{\hat{\beta}, \sigma_{\hat{\beta}}}(0.75)$ conforms to the quartile.

\(^4\)This paper contains a few series of graphs only; more complete and animated pictures are available via the web, see: www.fee.uva.nl/ke/jfk.htm

10
obtain
\[
\hat{\beta}_{IV} = \beta_{IV} - \beta = \sigma_z^2 \Sigma_{XZ}^{-1} \zeta = \frac{\sigma_{xz}}{\sigma_{xz}} = \frac{\rho_{xz}}{\rho_{xz}} \sigma_z
\]  
(13)

\[
c_3 = \sigma_z^2 \Sigma_{XZ}^{-1} \zeta = \frac{1}{\sigma_z^2} \sigma_{xz} \sigma_{xz} = \frac{\rho_{xz}}{\rho_{xz}} \rho_{xz}
\]

\[
\beta_{IV}^2 \Sigma_{X'X} \beta_{IV} = \sigma_z^4 \Sigma_{XZ}^{-1} \Sigma_{X'X} \Sigma_{X'Z} \zeta = \frac{\sigma_z^2}{\sigma_z^2} = \frac{\rho_{xz}^2}{\rho_{xz}^2},
\]

giving in the case where the disturbances are (almost) normally distributed
\[
V_{IV}^N = \frac{\sigma_z^2}{\sigma_z^2} \frac{1 - \rho_{xz}^2}{\rho_{xz}^2} \rho_{xz}^2 (\rho_{xz}^2 - \rho_{xz}^2)^2 + \rho_{xz}^4 (1 - \rho_{xz}^2).
\]  
(14)

The expression for the inconsistency \( \hat{\beta}_{IV} \) shows that its sign is determined by the sign of \( \rho_{ze} / \rho_{xz} \), whereas its magnitude is inversely related to the strength of the instrument, cf. Bound et al. (1995). \( V_{IV}^N \) is unaffected by the signs of \( \rho_{ze} \), \( \rho_{xz} \) and \( \rho_{xz} \) as long as the sign of the product \( \rho_{ze} \rho_{xz} \), remains the same, or when either \( \rho_{ze} \) or \( \rho_{ze} \) is zero. Self-evidently, \( V_{IV}^N \) diverges for \( \rho_{xz} \) approaching zero.

Without loss of generality we may focus in this model on the case \( \beta = 1 \). This is just a normalization and not a restriction, because we can imagine that we started off from a model \( y_i = \beta^* x_i^* + \varepsilon_i \), with \( \beta^* \neq 0 \), and rescaled the explanatory variable such that
\[
x_i = \frac{x_i^*}{\beta^*}.
\]

An important characteristic of the model is the signal-to-noise ratio \( (SN) \), which is here equal to
\[
SN = \frac{\beta^2 \sigma_z^2}{\sigma_z^2} = \frac{\sigma_z^2}{\sigma_z^2}.
\]  
(15)

From (13) and (14) we find that \( V_{IV}^N \) and \( \hat{\beta} \) are proportional to (the square root of) the inverse of \( SN \). In fact, after the normalization \( \hat{\beta} = 1 \), in this simple model the approximation to the distribution of the IV estimator \( \hat{\beta}_{IV} \overset{\approx}{\sim} N(\beta + \hat{\beta}, n^{-1} V_{IV}^N) \) is completely determined by \( n \) and the four model characteristics \( \rho_{xz} \), \( \rho_{ze} \), \( \rho_{xz} \) and \( SN \).

Next, we focus on obtaining an appropriate data generating scheme for this model to be used in the simulations. In the notation of Section 2 it should be given by
\[
\begin{align*}
y_i &= \beta x_i + \varepsilon_i \\
x_i &= \bar{x}_i + \xi \varepsilon_i \\
z_i &= \bar{z}_i + \zeta \varepsilon_i
\end{align*}
\]  
(16)

where \( \xi \) and \( \zeta \) are scalars. In order to obtain \( (\varepsilon_i, x_i, z_i)^T \sim \text{IID}(0, \Omega) \), with appropriate \( 3 \times 3 \) covariance matrix \( \Omega \), we can first generate \( v_i = (v_{i,1}, v_{i,2}, v_{i,3})^T \sim \text{IID}(0, I_3) \) and then parameterize as follows:
\[
\begin{align*}
\varepsilon_i &= \sigma_v v_{i,1} \\
\bar{x}_i &= \alpha_1 v_{i,2} \\
\bar{z}_i &= \alpha_2 v_{i,2} + \alpha_3 v_{i,3}.
\end{align*}
\]

This provides full generality. The coefficient \( \alpha_1 \) determines \( \sigma_v^2 \), whereas \( E(\bar{x}_i \varepsilon_i) = 0 \), as it should. Also \( E(\bar{z}_i \varepsilon_i) = 0 \), and \( \alpha_2 \) and \( \alpha_3 \) enable any correlation between \( \bar{x}_i \) and \( \bar{z}_i \) and
any value of $\sigma_z^2$. The above implies

$$
\begin{pmatrix}
\varepsilon_i \\
x_i \\
z_i
\end{pmatrix} = \Omega^{1/2} v_i =
\begin{pmatrix}
\sigma_\varepsilon & 0 & 0 \\
\sigma_\varepsilon \xi & \alpha_1 & 0 \\
\sigma_\varepsilon \zeta & \alpha_2 & \alpha_3
\end{pmatrix}
\begin{pmatrix}
v_{i,1} \\
v_{i,2} \\
v_{i,3}
\end{pmatrix}.
$$

(17)

Note that the zero elements do not entail restrictions on $\Omega$, because $\Omega^{1/2}$ is non-unique and a lower-triangular form with positive diagonal elements can be found for any positive definite $\Omega$.

In this simple model with $k = l = 1$ we have

$$
\hat{\beta}_{IV} = \frac{\sum z_i y_i}{\sum z_i x_i} = \beta + \frac{\sum z_i \varepsilon_i}{\sum z_i x_i},
$$

(18)

which clarifies that, irrespective of the sample size, the distribution of $\hat{\beta}_{IV}$ is invariant to the scale of $z_i$. We may also change the sign of all the $z_i$ without affecting $\hat{\beta}_{IV}$. Therefore, we may restrict ourselves in the illustrations to cases with $\rho_{xz} > 0$ (the case $\rho_{xz} = 0$ leads to underidentification and was already excluded in the assumptions). Since the distribution of $\hat{\beta}_{IV} - \beta$ becomes just its mirror-image when all $x_i$ are changed in sign, we shall also restrict ourselves to cases where $\rho_{xz} \geq 0$, because of the following reasoning. The value of $\rho_{xz}$ is invariant to changing the signs of all $x_i$ and $z_i$ values. Hence, for any value of $\rho_{xz} > 0$ the distribution of $\hat{\beta}_{IV} - \beta$ for $\rho_{xz} \leq 0$ and arbitrary positive or negative value of $\rho_{xz}$ is equivalent with the distribution of $-(\hat{\beta}_{IV} - \beta)$ for $-\rho_{xz} \geq 0$ and $-\rho_{xz}$. It is also obvious that $\sigma_x$ and $\sigma_\varepsilon$ do not affect the distribution of $\hat{\beta}_{IV}$ separately, but only through their ratio. Hence, without loss of generality, we can impose some genuine equality restrictions on the 6 parameters of $\Omega$. For these we choose

$$
\sigma_\varepsilon = 1,
$$

(19)

$$
\sigma_z^2 = \zeta^2 + \alpha_2^2 + \alpha_3^2 = 1.
$$

(20)

By (19) we normalize all results with respect to $\sigma_\varepsilon$, and (15) simplifies to

$$
SN = \sigma_x^2.
$$

(21)

Because any GIV estimator is invariant to the scale of the instruments (only the space spanned by the instruments is relevant) we may impose (20), which will be used to obtain the value

$$
\alpha_3 = \left| (1 - \zeta^2 - \alpha_2^2)^{1/2} \right|,
$$

(22)

where, without loss of generality, we may stick to positive values for $\alpha_3$ as long as $v_i^{(3)}$ is symmetrically distributed. For similar reasons we would get observationally equivalent data realizations if both $\alpha_1$ and $\alpha_2$ would be changed in sign. Therefore, below we will restrict ourselves to just positive values for both $\alpha_1$ and $\alpha_3$.

The above yields the following data (co)variances and correlations:

$$
\begin{align*}
\sigma_x^2 &= \xi^2 + \alpha_1^2 \\
\sigma_y^2 &= \xi^2 + 2\xi + 1 + \alpha_1^2 \\
\sigma_{xz} &= \xi \\
\rho_{xz} &= \frac{\xi}{\sqrt{\xi^2 + \alpha_1^2}} \\
\sigma_{ze} &= \zeta \\
\rho_{ze} &= \zeta \\
\sigma_{xz} &= \xi \zeta + \alpha_1 \alpha_2 \\
\rho_{xz} &= \frac{(\xi \zeta + \alpha_1 \alpha_2)}{\sqrt{\xi^2 + \alpha_1^2}}
\end{align*}
$$

(23)
Note that these, after the normalizations $\beta = 1$, $\sigma_\epsilon = 1$ and $\sigma_z = 1$, depend on only 4 free parameters of the data generating process (DGP), viz. $\zeta$, $\zeta$, $\alpha_1$ and $\alpha_2$. As we already established, the expressions for inconsistency in (13) and asymptotic variance (14) evaluated under $\mu_3 = 0$ and $\mu_4 = 3$ (the 3rd and 4th moment of $v_{i,t}$) depend on just four characteristics too, viz. on $\rho_{xz}$, $\rho_{zz}$, $\rho_{xx}$ and $SN = \sigma_z^2$. The latter four can be used in this simple model as a base for the Monte Carlo design parameter space, since they determine the parameters of the DGP through the relationships

$$
\begin{align*}
\zeta &= \rho_{xz}, \\
\zeta &= \rho_{xz}\sigma_x, \\
\alpha_1 &= \sigma_x \sqrt{(1 - \rho_{xz}^2)^{1/2}}, \\
\alpha_2 &= (\rho_{xz} - \rho_{xz}\rho_{xx})/(1 - \rho_{xz}^2)^{1/2},
\end{align*}
$$

from which $\alpha_3$ follows directly via (22). This reparametrization is useful, because the parameters $\rho_{xz}$, $\rho_{zz}$, $\rho_{xx}$ and $SN$ have a direct econometric interpretation, viz. the degree of simultaneity, instrument (in)validity, and instrument strength, whereas $SN$ is directly related to the model fit, which can be expressed as $SN/(SN + 1)$. We prefer to avoid to use the ‘concentration parameter’ as one of the relevant characteristics of this model in the present context, because this concept refers exclusively to the case where all instruments are valid.

From the above it follows that by varying the four parameters $|\rho_{xz}| < 1$, $0 \leq \rho_{xz} < 1$, $0 < \rho_{xz} < 1$ and $0 < \sigma_x^2/(\sigma_z^2 + 1) < 1$, we can examine the limiting and finite sample distributions of $\hat{\beta}_{IV}$ over the entire parameter space of this model. Note, however, that not all admissible values of these parameters will be compatible. For example, when $\rho_{xz}$ is large and $\rho_{xz}$ is small, this cannot be compatible with $\rho_{xz}$ being very large. Moreover, $\sigma_x$ has just an effect on the scale of $\hat{\beta}_{IV}$, $V_{IV}^N$ and $\hat{\beta}_{IV} - \beta$, so we may choose just one fixed value for $\sigma_x$ and from these findings the results for any value of $\sigma_x$ can be obtained simply by rescaling. In our calculations and simulations we will fix $\sigma_x^2/\sigma_z^2 = 10$, yielding a population fit of the model of $10/11 = 0.909$.

Actual values of $\hat{\beta}_{IV}$ and of $\text{AMAE}(\hat{\beta}_{IV})$ can now be calculated for any set of compatible values of $n$, $\rho_{xz}$, $\rho_{zz}$, $\rho_{xx}$ and $\sigma_x$, and they can be compared with corresponding Monte Carlo estimates obtained from $\hat{\beta}_{IV}$ realizations in simulated finite samples, in order to find out how accurate the first-order asymptotic approximations are. Before we present these summarizing characteristics, we will first examine the full density functions themselves. The Figures 1 through 4 contain 8 panels each. In all these panels four densities are presented, viz. for $n = 50$ (dark/black lines) and $n = 200$ (grey/red lines), both for the actual empirical distribution (solid lines) and for its asymptotic approximation (dashed lines). The latter has been taken as $\hat{\beta}_{IV} \sim N(\beta + \hat{\beta}_{IV}, n^{-1}V_{IV}^N)$. In the simulations we took $v_i \sim \text{IN}(0, I_3)$ and used 1,000,000 replications. From the results we may expect to get quick insights into issues as the following. For which combinations of the five design parameter values is: (a) the actual density of $\hat{\beta}_{IV}$ close to normal (symmetric, unimodal, etc.), (b) the actual median of $\hat{\beta}_{IV}$ close to $\beta_{IV}$, (c) the actual tail behavior of $\hat{\beta}_{IV}$ reasonably well represented by that of the $N(\beta_{IV}, n^{-1}V_{IV}^N)$ distribution, and (d) how do the actual estimation errors depend on the design parameters. Hence, we focus on the correspondences and differences in shape, location and spread of the asymptotic and the empirical distributions. Since in this just identified model the IV estimator does not have finite moments, we know that even if the instruments are valid, the asymptotic approximation will not capture the fat tail(s) of the finite sample distribution.
In Figure 1 \(\rho_{xz} = 0.8\), so the instrument is not ultra strong, but certainly not weak. In Figure 2 \(\rho_{xz} = 0.3\), in Figure 3 \(\rho_{xz} = 0.1\) and in Figure 4 \(\rho_{xz} = 0.01\). Hence, in the latter figure the instrument is certainly weak and we may expect that standard large sample asymptotics does not provide a very accurate approximation. All four figures contain eight panels for particular combinations of \(\rho_{xz}\) and \(\rho_{ze}\) values. The panels in the left-hand columns have \(\rho_{ze} = 0\), i.e. the instrument is valid and the standard asymptotic result applies. In the right-hand columns \(\rho_{ze} = 0.2\), i.e. the instrument is invalid and the IV estimator is inconsistent. Nevertheless, the asymptotic approximation presented in Corollary 1 applies. The four rows of panels cover the cases \(\rho_{ze} = -0.3, \rho_{ze} = 0\) (no simultaneity; hence, OLS would be more appropriate than IV), \(\rho_{ze} = 0.3\) and \(\rho_{ze} = 0.6\).

From Figure 1 we see in the left-hand column that the standard asymptotic approximation of IV when using a valid and strong instrument is very accurate when the simultaneity is not very serious, but deteriorates when \(\rho_{xz}\) increases, especially when \(n\) is small. We note some skewness and one fat tail, but the asymptotic distribution is never extremely bad for the cases examined. In the right-hand column we see that the new result of Corollary 1 is almost of the same quality but slightly less accurate. Especially for the smaller sample size we note some skewness and at least one fat tail in the empirical distribution, which are not captured by the first-order normal asymptotic approximation. In Figure 2, where the instrument is weaker, we find that when the instrument is valid the distribution is more skewed, and more so for serious simultaneity. In the right-hand column this occurs for different \(\rho_{xz}\) values. In most cases there is a substantial but not a dramatic difference between the actual distribution and its approximation. The discrepancies are more pronounced in Figure 3, and affect both the standard \((\rho_{ze} = 0)\) and the new \((\rho_{ze} \neq 0)\) asymptotic approximations. From Figure 4 it is clear that the asymptotic approximations are useless (at the sample sizes examined) when the instrument is really weak. When the instrument is valid the actual distributions show some median bias, but they are much less dispersed than suggested by \(nV_{IV}^N\).

The magnitude of the bias in relation to the OLS bias and weakness of the instrument has been analyzed by many authors, see Sawa (1969) and (further references in) Hillier (2006). When the instrument is invalid and very weak then the finite sample distribution of the inconsistent IV estimator is not centered at the pseudo-true-value. Surprisingly, it is actually much closer to the true value (also when the instrument is not so weak), whereas the distribution becomes bimodal when the instrument is very weak. Maddala and Jeong (1992), Woglom (2001), Hillier (2006) and Forchini (2006) show that bimodality of the consistent IV estimator occurs for much more severe simultaneity than examined here, viz. for \(\rho_{ze} = 0.99\), whereas Phillips (2006) shows that it is omnipresent in the simple Keynesian model where simultaneity is always severe. Our findings suggest that using instruments that are both weak and invalid leads to bimodality, irrespective of the degree of simultaneity.

From Figures 1 through 4 we conclude that, irrespective of whether the instruments are valid or not, one should avoid to use standard large sample asymptotics when instruments are really weak. If one replaces the weak instrument with a strong one that is invalid (which is always possible by reverting to OLS), one obtains an inconsistent estimator, such as depicted in the right-hand column of Figure 1, which has a distribution that is actually much more concentrated around the true value than that of the consistent estimator depicted in the left-hand column of Figure 4. The general validity
of the findings from Figures 1 through 4 will be illustrated now by scanning the median absolute error over almost the full parameter space of this simple model.

Figure 5 provides an overview of the (in)accuracy of the asymptotic distribution of IV as an approximation to the actual distribution in finite sample for \( n = 20 \) and for \( n = 100 \). These figures (based on 10,000 replications) cover all compatible positive values of \( \rho_{xx} \) and \( \rho_{zx} \), for \( \rho_{xz} = 0, 0.1, 0.3 \) and 0.6. This accuracy is expressed as \( \log [\text{MAE}(\hat{\beta}_{IV}) / \text{MAE}(\hat{\beta}_{IV})] \). Hence, positive values (yellow, amber) indicate larger absolute errors in finite sample than indicated by the asymptotic approximation and negative values (blue) indicate that standard asymptotics is too pessimistic about the absolute errors of \( \hat{\beta}_{IV} \) in finite sample. Note that this log-ratio is invariant regarding the value of \( SN = \sigma_{Z}^{2} / \sigma_{\varepsilon}^{2} \). We find that the degree of simultaneity \( \rho_{xx} \) has little effect, and neither has the (in)validity of the instrument \( \rho_{zx} \). Just instrument weakness (roughly, when \( |\rho_{xz}| < n^{-1/2} \)) seriously deteriorates the accuracy of the large-\( n \) asymptotic approximation.

Figure 6 examines \( \log [\text{MAE}(\hat{\beta}_{OLS}) / \text{MAE}(\hat{\beta}_{IV})] \), which is also invariant with respect to \( SN \). It shows that in finite sample the absolute estimation errors committed by OLS are larger than those of IV only when both \( \rho_{xx} \) and \( \rho_{xz} \) are large. The area where IV beats OLS gets smaller for larger \( \rho_{xx} \). We also note that OLS may beat IV by a much larger margin (when the instrument is weak and the simultaneity not so serious) than IV will ever beat OLS (which happens when the instrument is strong, the simultaneity serious, and the instrument not severely invalid).

### 3.2 A simple overidentified model

The model of the above subsection can be extended such that we have two instruments \( z_{i,1} \) and \( z_{i,2} \), i.e. \( l = 2 \) and \( \zeta = (\zeta_{1}, \zeta_{2})' \). First, we examine by which minimal set of data moments the limiting distribution is determined in this model. Again we assume that all variables in the regression have been scaled such that \( \beta = 1 \) and \( \sigma_{\varepsilon}^{2} = 1 \), whereas the instruments \( Z \) have been transformed such that \( \Sigma_{Z}Z = I \) (while still spanning the original subspace). Such an orthonormal base for this subspace is nonunique, and without loss of generality we may choose one in which only \( z_{i,1} \) is possibly correlated with \( \varepsilon_{i} \), so that \( \zeta_{2} = 0 \). This implies that

\[
\Sigma_{XZ}^{-1}\Sigma_{ZZ}^{-1}\Sigma_{Z\varepsilon} = \sigma_{2\varepsilon} = \sigma_{xz} \sigma_{z1\varepsilon} = \rho_{xz1} \rho_{z1\varepsilon} \sigma_{x},
\]

where, of course, \( \rho_{z1\varepsilon} = \zeta_{1} \). Now the various entries in the formula of Theorem 1 specialize to

\[
\begin{align*}
\Sigma_{XX} &= \sigma_{x}^{2} > 0, \\
\Sigma_{XZ} &= \sigma_{x}^{2} = \rho_{xz}^{2} \sigma_{z}^{2} > 0, \\
\hat{\beta}_{GIV} &= \sigma_{z}^{-2} \Sigma_{Z}Z^{-1} \Sigma_{X} \Sigma_{Z}Z^{-1} \Sigma_{Z} = \sigma_{z\varepsilon}^{2} \sigma_{z\varepsilon}^{2} \sigma_{x}^{2} = \rho_{xz1} \rho_{z1\varepsilon} \sigma_{x} \rho_{xz}^{2} \sigma_{z}^{2}, \\
c_{1} &= \sigma_{z}^{-2} \Sigma_{Z}Z^{-1} \Sigma_{Z}Z = \rho_{z1\varepsilon}^{2}, \\
c_{2} &= \sigma_{z}^{-2} \Sigma_{Z}Z^{-1} \Sigma_{X} \Sigma_{Z}Z = \rho_{z1\varepsilon}^{2}, \\
c_{3} &= \Sigma_{XZ} \hat{\beta}_{GIV} = \rho_{xz} \rho_{z1\varepsilon} \sigma_{x} \sigma_{z\varepsilon} = \rho_{xz} \rho_{z1\varepsilon} \sigma_{x} \rho_{z1\varepsilon} \sigma_{z\varepsilon} \rho_{xz}^{2} \sigma_{z}^{2} \sigma_{z\varepsilon}^{2} \sigma_{x}^{2} = \rho_{xz}^{2} \sigma_{z\varepsilon}^{2} \sigma_{x}^{2}.
\end{align*}
\]
from which $c_4$ and $c_5$ readily follow. From the above we conclude that the limiting
distribution of Theorem 1 is fully determined by (and varies with) the 5 data moments:
$\sigma_x$, $\rho_{xz}$, $\rho_{x\varepsilon}$, $\rho_{z\varepsilon}$ and $\rho_{xz\varepsilon}$. However, in the special case $\rho_{z\varepsilon} = 0$ the minimal set of
parameters is just one dimensional, because $\rho_{xz} \sigma_x$ suffices. For the general case we find

$$V_{GIV}^N = \frac{\sigma_e^2}{\sigma_x^2 \rho_{zx}^2} \left[ (1 - \rho_{z1\varepsilon}^4) + \rho_{z1\varepsilon}^2 \frac{\gamma_1}{\rho_{zx}^2} + \rho_{z1\varepsilon}^2 \rho_{xz}^2 \frac{\gamma_2}{\rho_{zx}^2} + 2\rho_{z1\varepsilon}^2 \rho_{xz}^4 \frac{\gamma_3}{\rho_{zx}^6} \right]$$

$$+ \frac{\sigma_e^2}{\sigma_x^2 \rho_{zx}^2} \rho_{z1\varepsilon}^4 \left[ 1 - 2 \rho_{z1\varepsilon}^2 \left( 1 - \frac{\rho_{z1\varepsilon}^2}{\rho_{zx}^2} \right) \right],$$

where

$$\gamma_1 = \rho_{z1\varepsilon}^2 - 2\rho_{xz} \rho_{z1\varepsilon} (1 + \rho_{z1\varepsilon}^2 - 2\rho_{z1\varepsilon}^4) - \rho_{z1\varepsilon} (\rho_{z1\varepsilon}^2 - 5\rho_{z1\varepsilon}^3 + 2\rho_{z1\varepsilon}^5),$$

$$\gamma_2 = \rho_{zx}^2 + 4\rho_{z1\varepsilon} \rho_{z1\varepsilon} \rho_{z1\varepsilon} - 4\rho_{z1\varepsilon}^2 [1 + 3\rho_{z1\varepsilon} \rho_{zx} \rho_{z1\varepsilon} - \rho_{zx}^2 + \rho_{z1\varepsilon}^2 (1 - \rho_{z1\varepsilon}^2)],$$

$$\gamma_3 = 2 - (\rho_{zx} - \rho_{z1\varepsilon} \rho_{z1\varepsilon})(3\rho_{zx} - \rho_{z1\varepsilon} \rho_{z1\varepsilon}).$$

Note that this variance is invariant to sign changes of the correlations as long as the
sign of $\rho_{z1\varepsilon} \rho_{zx} \rho_{z1\varepsilon}$ is not affected, or when either $\rho_{zx}$ or $\rho_{z1\varepsilon}$ is zero. The sign of the inconsistency
$\hat{\beta}_{GIV}$ is determined by the sign of $\rho_{z1\varepsilon} \rho_{z1\varepsilon}$. For given values of $\rho_{zx}$ and
$\rho_{z1\varepsilon}$ the magnitude of $\hat{\beta}_{GIV}$ is a multiple of $\rho_{z1\varepsilon}$, so it will be large when the invalid
instrument is relatively strong. For the special case $\rho_{z1\varepsilon} = \rho_{zx}$, i.e. the second instrument
is orthogonal to $x$, the variance formula specializes to

$$\frac{\sigma_e^2}{\sigma_x^2} \left( 1 - \rho_{z1\varepsilon}^2 \right) \rho_{z1\varepsilon}^2 \rho_{zx} (\rho_{z1\varepsilon}^2 - \rho_{z1\varepsilon}^4 \rho_{z1\varepsilon}^2 + \rho_{z1\varepsilon}^4 (1 - \rho_{zx}^2))$$

which, not surprisingly, corresponds to (14).

Next we examine whether, apart from $n$, the same number of parameters is required
to obtain in all generality the finite sample distribution of GIV by generating the appro-
priate data processes. For that purpose the schemes (16) and (17) can be extended as
follows. Let now $v_i \sim N(0, I)$ Again we take $\varepsilon_i = \varepsilon_i, \varepsilon_i^2$, and, again restricting ourselves
to positive $\alpha_1$ for symmetrically distributed $v_i$, we have

$$x_i = \xi v_{i,1} + \alpha_1 v_{i,2}$$

$$= \sigma_x [\rho_{zx} v_{i,1} + \sqrt{(1 - \rho_{zx}^2)^2} \, v_{i,2}],$$

with $\sigma_x^2 = SN$ (this is all similar to the earlier example with $l = 1$). Now, however, we
have to compose the $l = 2$ instruments as

$$z_i,1 = \zeta_1 v_{i,1} + \alpha_2 v_{i,2} + \alpha_3 v_{i,3} + \alpha_4 v_{i,4},$$

$$z_i,2 = \zeta_2 v_{i,1} + \alpha_5 v_{i,2} + \alpha_6 v_{i,3} + \alpha_7 v_{i,4}.$$
of one valid and one possibly invalid instrument, as we already argued above from the asymptotic perspective. Now we can perform a similar operation with respect to \( z_{i,1} \), such that we may impose \( \alpha_4 = 0 \). Next, rescaling the instruments such that they have unit variance leads to the generating schemes

\[
\begin{align*}
  z_{i,1} &= \zeta_1 v_{i,1} + \alpha_2 v_{i,2} + \left(1 - \zeta_1^2 - \alpha_2^2\right)^{1/2} v_{i,3}, \\
  z_{i,2} &= \alpha_5 v_{i,1} + \alpha_6 v_{i,3} + \left(1 - \alpha_5^2 - \alpha_6^2\right)^{1/2} v_{i,4}.
\end{align*}
\] (30)

Due to the symmetry of \( v_i \) generality is maintained when we restrict ourselves to cases where particular coefficients are nonnegative. This extends to \( \alpha_2 \) and \( \alpha_5 \), because the space spanned by the instruments does not change by multiplying all elements by \(-1\), yielding

\[
0 \leq \alpha_2 \leq 1, \quad 0 \leq \alpha_5 \leq 1.
\] (31)

We also maintain full generality by imposing zero covariance on the two instruments, which implies \( \alpha_2 \alpha_5 + \alpha_6 \left(1 - \zeta_1^2 - \alpha_2^2\right)^{1/2} = 0 \), from which we find

\[
\alpha_6 = -\alpha_2 \alpha_5 \left(1 - \zeta_1^2 - \alpha_2^2\right)^{-1/2}.
\] (32)

So, for given values of \( \sigma_x, \rho_{xx} \) and \( \rho_{x1} = \zeta_1 \), we would be able to generate data according to (29) and (30) if we also knew \( \alpha_2 \) and \( \alpha_5 \).

The asymptotic overall strength of the two instruments can be controlled by the population \( R^2 \) of the regression of \( x \) on \( Z = (z_1, z_2) \), which is

\[
R^2_{xZ} = \frac{\Sigma x'Z \Sigma^{-1} Z'x}{\sigma_x^2}.
\] (33)

Note that

\[
\rho_{xx}^2 = \frac{\left(\Sigma x'Z \Sigma^{-1} Z'x\right)^2}{\sigma_x^2 \sigma_x^2} = \frac{\sigma_{z1}^2}{\sigma_x^2} = R^2_{xZ},
\] (34)

and, since we imposed \( \Sigma_{Z'Z} = I \), we have

\[
\rho_{xx}^2 = \frac{\Sigma x'Z \Sigma Z'x}{\sigma_x^2} = \rho_{x1}^2 + \left(1 - \rho_{xx}^2\right)^{1/2} \alpha_5^2,
\]

\[
\rho_{x1} = \rho_{xx} \rho_{x1} + \left(1 - \rho_{xx}^2\right)^{1/2} \alpha_2.
\]

From these we can express the (nonnegative) values of \( \alpha_5 \) and \( \alpha_2 \) as

\[
\alpha_5 = \left| \left( \frac{\rho_{xx}^2 - \rho_{x1}^2}{1 - \rho_{xx}^2} \right)^{1/2} \right| \tag{35}
\]

and

\[
\alpha_2 = \frac{\rho_{x1} - \rho_{xx} \rho_{x1}}{\left(1 - \rho_{xx}^2\right)^{1/2}}.
\] (36)

from which \( \alpha_6 \) follows directly by evaluating (32).

Hence, we can scan the finite sample distribution of GIV for this class of model for any \( n \) over its entire parameter space by simulating data for all compatible values of \( \sigma_x, \rho_{xx}, \rho_{xx}, \rho_{x1} \) and \( \rho_{x1} \). Here again, these are found to be those data moments that characterize the asymptotic distribution. They determine \( \xi \) and \( \alpha_1 \) via (29) and \( \alpha_2, \alpha_5 \)
and $\alpha_6$ via (36), (35) and (32), respectively. We may restrict ourselves to cases where $ho_{xx} > 0$ (since the coefficients of the simulation design are just determined by $\rho_{xx}^2$). Note that the coefficients of the data generation process, notably $\alpha_2$, are unaffected (thus yielding the same distribution of $\hat{\beta}_{GIV}$) if both $\rho_{zi}^2$ and $\rho_{zi1}$ are changed in sign. Therefore, we will only examine cases with $\rho_{zi1} \geq 0$. However, we shall also examine nonnegative values of $\rho_{zi}^2$ only, because when the distribution of $\varepsilon$ is symmetric changing the signs of both $\rho_{zi1}$ and $\rho_{xx}$ yields the mirror image of the distribution of $\hat{\beta}_{GIV}$. In line with the just identified model the distribution of $\hat{\beta}_{GIV} - \beta$ for $\rho_{zi1} \leq 0$ is equivalent with the distribution of $-(\hat{\beta}_{GIV} - \beta)$ for $-\rho_{zi1} \geq 0$ and $-\rho_{xx}$, because: If we change the signs of $\rho_{zi1}$, $\rho_{xx}$ and all $\varepsilon_i$ then the variables $x_i$, $z_{i,1}$, $z_{i,2}$ and thus $\hat{x}_i$ remain the same, whereas $\hat{\beta}_{GIV} - \beta = \sum \hat{x}_i \varepsilon_i / \sum \hat{x}_i^2$ changes sign.

In the special case that no instrument is invalid we have $\zeta_1 = \rho_{zi1} = 0$ in (30). Thus full generality is maintained by making $z_{i,2}$ independent of $v_{i,3}$, giving $\alpha_6 = 0$, and zero covariance of the two instruments implies now $\alpha_2 \alpha_5 = 0$. Hence, we may choose $\alpha_5 = 0$, resulting in the simplified generating schemes

$$
\begin{align*}
  z_{i,1} &= \alpha_2 v_{i,2} + (1 - \alpha_2^2)^{1/2} v_{i,3}, \\
  z_{i,2} &= v_{i,4}.
\end{align*}
$$

These imply $\rho_{zz_1} = \rho_{xx}$ and $\alpha_2 = |\rho_{xx}(1 - \rho_{xx}^2)|^{-1/2}$ instead of (36). Hence, when $\zeta = 0$ the finite sample distribution is determined by just 3 parameters (viz. $\sigma_x$, $\rho_{xx}$ and $\rho_{xx}$) instead of 5 (apart from $n$), whereas the limiting distribution just depends on $\sigma_x^2 = \rho_{xx}^2 \sigma_x^2$.

In all calculations we again fixed $\sigma_x^2/\sigma_{\varepsilon}^2 = 10$ (which here too has only a multiplicative effect, i.e. just affects the scale of the densities), as before we chose values $\rho_{xx} = \{0.3, 0.0, 0.3, 0.6\}$, $\rho_{zi1} = \zeta_1 = \{0.0, 0.2\}$ and $\rho_{xx} = \{0.8, 0.3, 0.1, 0.01\}$, whereas $\rho_{zz_1} = \{\rho_{xx}, \rho_{xx}/2, \rho_{xx}/8\}$. The latter values are associated with decreasing relative strength of $z_1$ (complementary, a valid instrument $z_2$ is either uncorrelated with $x$, contributes 50% to the joint strength of the instruments, or is relatively strong). Figures 7 and 8 contain some illustrative densities for $\rho_{xx} < \rho_{xx}$, again for $n = 50$ and $n = 200$. Since $l = k = 1$, GIV does have a finite first moment now. To save space we have put more cases into one figure. Moreover, we have omitted the cases where $\rho_{xx1} = \rho_{xx}$ (and $\rho_{zz_1} = 0$). Although we already established that these yield similar asymptotic results as the $k = l = 1$ case, from the simulations we found that the finite sample densities do differ slightly from the "no finite moments" case, even more so for a weaker instrument $z_1$, especially when $\rho_{xx1} = \rho_{xx} = 0.01$. When both instruments are valid and $\rho_{xx} = 0.01$ the $l = 2$ case produces estimators which are slightly more efficient than the corresponding $l = 1$ estimators. This seems at odds with the findings in Donald and Newey (2001) which suggest that efficiency benefits when weak instruments are discarded. Note, however, that their analysis assumes that the number of instruments grows at a smaller rate than the sample size, whereas in our experiments the number of instruments is fixed. When $\rho_{xx1} = 0.01$, $\rho_{zz_1} = 0$ and $\rho_{zi1} = 0.2$ we find that the bimodality of the GIV estimator is less pronounced than for the IV estimator.

Figure 7 presents densities for $\rho_{xx} = 0.8$ and 0.3, and Figure 8 for the weaker instruments. In Figure 7 the asymptotic approximations are mostly reasonably accurate, but not in Figure 8, especially when $\rho_{xx} = 0.01$. In the latter case we again note that the asymptotic approximations are much too pessimistic. The actual (median) bias of the GIV estimator is much less dramatic as the inconsistency $\hat{\beta}_{GIV}$ suggests, and even
though both instruments are very weak and one of them is also invalid, the actual density of \( \hat{\beta}_{GIV} \) has most of its probability mass remarkably close to the true value 1. Although Forchini (2006) suspects bimodality in the overidentified model when the instruments are valid but weak, we do not find it at \( \rho_{x\hat{x}} = 0.01 \).

Finally, we look again at median absolute error results. Figure 9 gives a more global impression of the accuracy of the asymptotic approximation in this model. We present results for \( n = 20 \) only and establish that the overall instrument strength \( \rho_{x\hat{x}} \) is the major determining factor, although the measures for instrument invalidity and simultaneity have an effect too. Figure 10 makes comparisons with OLS for \( n = 100 \). We note that, especially in the presence of invalid instruments, there is much scope for OLS to produce smaller estimation errors than GIV. Anyhow, our simulation results do not generally support the conclusion by Hahn and Hausman (2003) that 2SLS is the preferred estimator when \( n \geq 100 \) and \( \rho_{x\hat{x}}^2 \geq 0.1 \). They arrive at this conclusion by comparing second-order asymptotic approximations to MSE.

4 Conclusions

In this paper we obtained an explicit formula for the asymptotic variance of the generalized instrumental variable estimator when some of the employed instruments are invalid. We showed that the limiting distribution of such an inconsistent estimator is normal, and is centered at the pseudo-true-value (true coefficient plus inconsistency), whereas its asymptotic variance includes a number of terms and factors additional to the standard result. It can only be expressed when one is willing to make assumptions on the first four moments of the disturbances. To obtain our results we assumed covariance stationarity of all variables, i.e. the dependent, the explanatory and the instrumental variables. It can only be expressed when one is willing to make assumptions on the first four moments of the disturbances. To obtain our results we assumed covariance stationarity of all variables, i.e. the dependent, the explanatory and the instrumental variables. In the simple illustrative models which we used, the data observations are in fact IID, as is often assumed in cross-section applications. Note, however, that our theorems also hold for time-series applications, where independence of the sample observations is unrealistic and the model may include lagged dependent variables. The theorems are also directly applicable in cases involving non-stationary series, provided the model is formulated in error correction form and the long-run multipliers (the coefficients of the cointegrating vector) have been imposed, so that the model and the instruments can all be represented by transformations of the original data that are integrated to order zero. Also note that the theorems can easily be generalized such that they would apply to GMM estimators and to dynamic panel data models.

We examined the accuracy of our analytic large sample results by comparing them with the simulated actual behavior of instrumental variable estimators in finite samples. Through a reparametrization of the structural and reduced form coefficients into parameters that directly express the degree of simultaneity, the degree of (in)validity of the instrument(s), the strength of the instrument(s) and the signal-to-noise ratio of the model, and by condensing the numerical results into graphic displays, it proved possible to produce a rather complete taxonomy of the behavior of the examined instrumental variables estimators over their full parameter space.

There is a quickly expanding literature on the shortcomings of standard large sample asymptotic approximations to the distribution of IV and GMM estimators and tests when the sample size is small or moderate and some of the instruments are weak but
valid, and how alternative and better approximations could be obtained. The present study shows that it is possible to obtain an explicit large sample asymptotic approximation to the distribution of method of moments estimators, also when some of the exploited moment conditions are invalid. Not surprisingly, however, that approximation is found to be vulnerable too, when instruments are weak. One option now would be to replace it by an approximation that aims to cope with weakness of instruments. However, our illustrations also suggest an alternative approach in which the employment of weak instruments, which invariably yields estimators with flat distributions, is abandoned altogether. We have shown that exclusively exploiting strong instruments, even if these constitute invalid instruments, yields much smaller absolute estimation errors in comparison with those obtained on the basis of weak instruments. For that situation we have produced here a reasonably accurate approximation to the finite sample distribution. But, to render this approximation feasible one requires information on the simultaneity parameter $\xi$ and the instrument invalidity parameter $\zeta$. That seems hard to obtain, and if such information was available other estimators than those obtained by minimizing an (in)appropriate GMM criterion function might be better for producing accurate inference on the coefficient $\beta$.

References


and Data Analysis 49, 417-444.


A Proof of Theorem 1

Because with an increase of sample size the estimator \( \hat{\beta}_{GIV} \) tends to \( \beta_{GIV}^* \), instead of \( \beta \), in order to establish its limiting distribution we should not focus on \( \sqrt{n}(\hat{\beta}_{GIV} - \beta) \), but choose a center of the distribution that tends to \( \beta_{GIV}^* \) too, see Rothenberg (1972). For the sake of simplicity we shall center at \( \beta_{GIV}^* \) itself. Note that

\[
\sqrt{n}(\hat{\beta}_{GIV} - \beta_{GIV}^*) = \sqrt{n}[(\hat{X}^\prime \hat{X})^{-1} \hat{X}^\prime \epsilon - \sigma_z^2 \Sigma_{-1}^{-1} \Sigma_{XZ} \Sigma_{Z}^{-1} Z^\prime \epsilon].
\]  

(38)

To obtain the limiting distribution we shall rewrite the right-hand side of (38) such that we can invoke the Lemma given at the end of Section 2. Below we first show that (38) can be rewritten as

\[
\sqrt{n}(\hat{\beta}_{GIV} - \beta_{GIV}^*) = (\frac{1}{n} \hat{X}^\prime \hat{X})^{-1} \frac{1}{\sqrt{n}} [W^\prime \epsilon + \omega(\epsilon^\prime \epsilon - n\sigma_z^2)] + o_p(1),
\]  

(39)

for appropriate \( n \times k \) matrix \( W \), with \( E(W^\prime \epsilon) = 0 \), and nonrandom \( k \times 1 \) vector \( \omega \). Invoking a theorem often attributed to Cramér, the lemma yields

\[
\sqrt{n}(\hat{\beta}_{GIV} - \beta_{GIV}^*) \rightarrow N \left[ 0, \sigma_z^2 \frac{1}{X^\prime \Sigma_{Z}^{-1} Z^\prime} \right],
\]  

(40)

upon assuming \( \mu_3 = 0 \) and \( \mu_4 = 3 \).

We first set out to rewrite (38) in the form (39). Using \( \hat{X} = (Z^\prime Z)^{-1} Z^\prime X \) we get

\[
\sqrt{n}(\hat{\beta}_{GIV} - \beta_{GIV}^*) = \sqrt{n}[\hat{X}^\prime \hat{X}]^{-1} \left[ \hat{W}^\prime [Z^\epsilon - E(Z^\epsilon)] + \frac{1}{\sqrt{n}} \hat{W}^\prime E(Z^\epsilon) - \sigma_z^2 \hat{X}^\prime \hat{X} \Sigma_{XZ} \Sigma_{Z}^{-1} \Sigma_{X} \Sigma_{Z}^{-1} \Sigma_{Z} \Sigma_{Z}^{-1} Z^\prime \epsilon \right],
\]  

(41)

\[
= (\frac{1}{n} \hat{X}^\prime \hat{X})^{-1} \frac{1}{\sqrt{n}} \left[ \hat{W}^\prime [Z^\epsilon - E(Z^\epsilon)] + \sigma_z^2 \frac{1}{\sqrt{n}} \hat{W}^\prime - \frac{1}{n} \hat{X}^\prime \hat{X} \Sigma_{XZ} \Sigma_{XZ} \Sigma_{Z}^{-1} \Sigma_{Z} \Sigma_{Z}^{-1} \Sigma_{Z} \Sigma_{Z}^{-1} Z^\prime \epsilon \right].
\]
For the second expression between square brackets in the final line of (41) we find
\[
\hat{\Pi}' = \frac{1}{n} \hat{X}' \hat{X} \Sigma_{\hat{X}' \hat{X}}^{-1} \Sigma_{\hat{X}' \hat{Z} \Sigma_{\hat{Z}' \Sigma_{\hat{Z}' \hat{X}}}}^{-1} \\
= \hat{\Pi}'(\Sigma_{\hat{Z}' \hat{Z}} - \frac{1}{n} \hat{Z}' \hat{Z}) \Sigma_{\hat{Z}' \hat{Z}}^{-1} + \left(\frac{1}{n} \hat{X}' \hat{X} \Sigma_{\hat{X}' \hat{X}}^{-1} \Sigma_{\hat{X}' \hat{Z} \Sigma_{\hat{Z}' \hat{X}}}ight) \Sigma_{\hat{Z}' \hat{Z}}^{-1} \\
= \hat{\Pi}'(\Sigma_{\hat{Z}' \hat{Z}} - \frac{1}{n} \hat{Z}' \hat{Z}) \Sigma_{\hat{Z}' \hat{Z}}^{-1} + \left(\frac{1}{n} \hat{X}' \hat{X} \Sigma_{\hat{X}' \hat{X}}^{-1} \Sigma_{\hat{X}' \hat{Z} \Sigma_{\hat{Z}' \hat{X}}}ight) \Sigma_{\hat{Z}' \hat{Z}}^{-1},
\]
and this contains a factor which can be rewritten as
\[
\Sigma_{\hat{X}' \hat{X}} = \frac{1}{n} \hat{X}' \hat{X}
\]
\[
\Sigma_{\hat{X}' \hat{Z}} \Sigma_{\hat{Z}' \hat{Z}} \Sigma_{\hat{Z}' \hat{X}} - \hat{\Pi}' \frac{1}{n} \hat{Z}' \hat{X}
\]
\[
= (\Sigma_{\hat{X}' \hat{Z}} \Sigma_{\hat{Z}' \hat{Z}} - \hat{\Pi}' \hat{Z}' \hat{X} - \hat{\Pi}' \frac{1}{n} \hat{Z}' \hat{X} - \Sigma_{\hat{Z}' \hat{X}})
\]
\[
= \{(\Sigma_{\hat{X}' \hat{Z}} - \frac{1}{n} \hat{X}' \hat{Z}) \Sigma_{\hat{Z}' \hat{Z}} - \frac{1}{n} \hat{X}' \hat{Z}((\frac{1}{n} \hat{Z}' \hat{Z})^{-1} - \Sigma_{\hat{Z}' \hat{Z}})\} \Sigma_{\hat{Z}' \hat{X}} - \hat{\Pi}'(\frac{1}{n} \hat{Z}' \hat{X} - \Sigma_{\hat{Z}' \hat{X}})
\]
\[
= \{(\Sigma_{\hat{X}' \hat{Z}} - \frac{1}{n} \hat{X}' \hat{Z}) \Sigma_{\hat{Z}' \hat{Z}} - \hat{\Pi}'(\Sigma_{\hat{Z}' \hat{Z}} - \frac{1}{n} \hat{Z}' \hat{Z}) \Sigma_{\hat{Z}' \hat{Z}})\} \Sigma_{\hat{Z}' \hat{X}} - \hat{\Pi}'(\frac{1}{n} \hat{Z}' \hat{X} - \Sigma_{\hat{Z}' \hat{X}}).
\]
Now substituting the decompositions obtained in (42) and (43) into the expression within curly brackets in the final line of (41) we obtain
\[
\hat{\Pi}'[\hat{Z}' \hat{c} - \hat{E}(\hat{Z}' \hat{c})] + n \sigma^2 \hat{\Pi}' - \frac{1}{n} \hat{X}' \hat{X} \Sigma_{\hat{X}' \hat{X}}^{-1} \Sigma_{\hat{X}' \hat{Z} \Sigma_{\hat{Z}' \hat{X}}}
\]
\[
\hat{\Pi}'[\hat{Z}' \hat{c} - \hat{E}(\hat{Z}' \hat{c})] + n \sigma^2 \hat{\Pi}'(\Sigma_{\hat{Z}' \hat{Z}} - \frac{1}{n} \hat{Z}' \hat{Z}) \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi
\]
\[
+ n \sigma^2(\frac{1}{n} \hat{X}' \hat{Z} - \Sigma_{\hat{X}' \hat{Z}}) \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi + n \sigma^2(\frac{1}{n} \hat{X}' \hat{Z} - \Sigma_{\hat{X}' \hat{Z}}) \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi
\]
\[
= \hat{\Pi}'[\hat{Z}' \hat{c} - \hat{E}(\hat{Z}' \hat{c})] - n \sigma^2 \hat{\Pi}'(\frac{1}{n} \hat{Z}' \hat{X} - \Sigma_{\hat{Z}' \hat{X}}) \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi
\]
\[
+ n \sigma^2((\frac{1}{n} \hat{X}' \hat{Z}) \Pi - \hat{\Pi}'(\Sigma_{\hat{Z}' \hat{Z}} - \frac{1}{n} \hat{Z}' \hat{Z}) \Sigma_{\hat{Z}' \hat{Z}}^{-1} \Sigma_{\hat{Z}' \hat{X}} - \hat{\Pi}'(\frac{1}{n} \hat{Z}' \hat{X} - \Sigma_{\hat{Z}' \hat{X}})) \beta_{GIV},
\]
where we substituted \( \Pi \equiv \Sigma_{\hat{X}' \hat{X}}^{-1} \Sigma_{\hat{X}' \hat{Z} \Sigma_{\hat{Z}' \hat{X}}} \). Now exploiting item (vi) of Framework A, we can employ
\[
\hat{Z}' \hat{c} - \hat{E}(\hat{Z}' \hat{c}) = \hat{Z}' \hat{c} + (\hat{c}' \hat{c} - n \sigma^2) \xi,
\]
\[
\frac{1}{n} \hat{Z}' \hat{Z} - \Sigma_{\hat{Z}' \hat{Z}} = \frac{1}{n} \hat{Z}' \hat{Z} - \frac{1}{n} \hat{E}(\hat{Z}' \hat{Z} | \hat{Z}) + \frac{1}{n} \hat{E}(\hat{Z}' \hat{Z} | \hat{Z}) - \Sigma_{\hat{Z}' \hat{Z}}
\]
\[
\frac{1}{n} \hat{Z}' \hat{c} - \Sigma_{\hat{X}' \hat{Z}} = \frac{1}{n} \hat{Z}' \hat{c} - \frac{1}{n} \hat{E}(\hat{X}' \hat{Z} | \hat{X}, \hat{Z}) + \frac{1}{n} \hat{E}(\hat{X}' \hat{Z} | \hat{X}, \hat{Z}) - \Sigma_{\hat{X}' \hat{Z}}
\]
so that the final expression given for (44) can be written as
\[
\hat{\Pi}' \hat{Z}' \hat{c} + (\hat{c}' \hat{c} - n \sigma^2) \hat{\Pi}' \xi - \sigma^2 \hat{\Pi}' \hat{Z}' \hat{c} \xi \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi - \sigma^2 \hat{\Pi}' \hat{c}' \hat{c} \xi \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi
\]
\[
+ \sigma^2 \hat{X}' \hat{Z} \xi \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi + \sigma^2 \hat{Z}' \hat{c} \xi \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi + \sigma^2 \hat{X}' \hat{c} \xi \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi + \hat{\Pi}' \hat{Z}' \hat{c} \xi \Pi \hat{\beta}_{GIV} + \Pi \hat{\Pi}' \hat{c}' \hat{Z} \Pi \hat{\beta}_{GIV}
\]
\[
+ \Pi'(\hat{c}' \hat{c} - n \sigma^2) \xi \Pi \hat{\beta}_{GIV} - \hat{X}' \hat{c} \xi \Pi \hat{\beta}_{GIV} - \hat{\xi}' \hat{Z} \xi \Pi \hat{\beta}_{GIV} - (\hat{c}' \hat{c} - n \sigma^2) \xi \Pi \hat{\beta}_{GIV} - \Pi \hat{\xi}' \hat{X} \Pi \hat{\beta}_{GIV}
\]
\[
- \hat{\Pi}' \hat{Z}' \hat{c} \xi \Pi \hat{\beta}_{GIV} - \hat{\Pi}' \hat{c}' \hat{c} \xi \Pi \hat{\beta}_{GIV} + o_p(n^{-1/2}).
\]
This can be further simplified by using
\[
c_1 \equiv \sigma^2 \xi \xi \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi,
\]
\[
c_2 \equiv \xi \xi \Pi \hat{\beta}_{GIV} = \sigma^2 \xi \xi \Sigma_{\hat{Z}' \hat{Z}}^{-1} \Sigma_{\hat{X}' \hat{X}}^{-1} \Sigma_{\hat{X}' \hat{Z}} \Sigma_{\hat{Z}' \hat{Z}}^{-1} \xi,
\]
\[
c_3 \equiv \xi \xi \Pi \hat{\beta}_{GIV},
\]
\[
22
\]
giving

\[ \begin{align*}
&\hat{\Pi}'Z'e + (\varepsilon'e - n\sigma^2\varepsilon)\hat{\Pi}'\xi - c_1\hat{\Pi}'Z'e - \sigma^2\varepsilon\hat{\Pi}'\xi^\prime\Sigma_{Z'Z}^{-1}Z'e - c_1(\varepsilon'e - n\sigma^2\varepsilon)\hat{\Pi}'\xi
+ c_1\hat{\Pi}'Z'e + \sigma^2\varepsilon(\xi^\prime\Sigma_{Z'Z}^{-1}Z'e + c_1(\varepsilon'e - n\sigma^2\varepsilon)\hat{\Pi}'\xi
- c_2\hat{X}'\xi - \xi_\beta GIV'\hat{\Pi}'Z'e - c_2(\varepsilon'e - n\sigma^2\varepsilon)\hat{\Pi}'\xi
- c_2\hat{X}'\xi - \xi_\beta GIV'\hat{\Pi}'Z'e - c_2(\varepsilon'e - n\sigma^2\varepsilon)\hat{\Pi}'\xi
+ c_2\xi + c_3\hat{\Pi}'\xi)^2(\varepsilon'e - n\sigma^2\varepsilon) + o_p(n^{1/2}),
\end{align*} \]

\[ \begin{align*}
= [c_4I_k - \Pi'\xi_\beta GIV]\hat{X}'\xi' + [c_5\hat{\Pi}' + (\xi - \Pi'\xi)(\sigma^2\xi' - \beta GIV\Sigma X'Z)\Sigma_{Z'Z}^{-1}\hat{Z}'\xi' + (c_4\xi + c_5\Pi'\xi)^2(\varepsilon'e - n\sigma^2\varepsilon) + o_p(n^{1/2}),
\end{align*} \]

where \( c_4 \equiv c_1 - c_2 \) and \( c_5 \equiv 1 - c_3 - c_4. \)

Note that (45) is equal to the factor in curly brackets in the final line of (41), and we want to derive its limiting distribution after scaling by the factor \( 1/\sqrt{n} \), so we may neglect the remainder term. Cramér's theorem implies that in deriving this limiting distribution we may replace \( \hat{\Pi} \) by its probability limit \( \Pi \). Hence, defining the \( k \times k \) matrix \( \Theta_1 \), the \( k \times l \) matrix \( \Theta_2 \) and the \( k \times 1 \) vector \( \omega \), such that

\[ \begin{align*}
\Theta_1 &= c_4I_k - \Pi'\xi_\beta GIV, \\
\Theta_2 &= c_5\Pi' + (\xi - \Pi'\xi)(\sigma^2\xi' - \beta GIV\Sigma X'Z)\Sigma_{Z'Z}^{-1}, \\
\omega &= c_4\xi + c_5\Pi'\xi,
\end{align*} \]

we can now invoke the Lemma with

\[ \begin{align*}
W' &= \Theta_1\hat{X}' + \Theta_2\hat{Z}'.
\end{align*} \]

For the case \( \mu_3 = 0 \) and \( \mu_4 = 3 \) we then obtain the limiting distribution

\[ \begin{align*}
\frac{1}{\sqrt{n}} [\Theta_1\hat{X}'\varepsilon + \Theta_2\hat{Z}'\varepsilon + \omega(\varepsilon'e - n\sigma^2\varepsilon)] &\rightarrow N(0, \sigma^2_2V_0),
\end{align*} \]

where

\[ \begin{align*}
V_0 &= \Theta_1\Sigma_{X'X}\Theta_1' + \Theta_2\Sigma_{Z'Z}\Theta_2' + 2\sigma^2_2\omega\omega' + \Theta_1\Sigma_{X'Z}\Theta_2' + \Theta_2\Sigma_{Z'X}\Theta_1'.
\end{align*} \]

In evaluating \( V_0 \) we make use of

\[ \begin{align*}
\Sigma_{Z'Z} &= \text{plim } n^{-1}\hat{Z}'\hat{Z} = \Sigma_{Z'Z} - \sigma^2_2\xi'\xi', \\
\Sigma_{X'X} &= \text{plim } n^{-1}\hat{X}'\hat{X} = \Sigma_{X'X} - \sigma^2_2\xi'\xi', \\
\Sigma_{Z'X} &= \text{plim } n^{-1}\hat{Z}'\hat{X} = \Sigma_{Z'X} - \sigma^2_2\xi'\xi', \\
\Sigma_{X'Z} &= \text{plim } n^{-1}\hat{X}'\hat{Z} = \Sigma_{X'Z} - \sigma^2_2\xi'\xi',
\end{align*} \]

and find that \( V_0 \) can be expressed as

\[ \begin{align*}
&[c_4I_k - \Pi'\xi_\beta GIV][\Sigma_{X'X} - \sigma^2_2\xi'\xi'][c_4I_k - \beta GIV\xi_\Pi]
+ [c_5\Pi' + (\xi - \Pi'\xi)(\sigma^2\xi' - \beta GIV\Sigma X'Z)\Sigma_{Z'Z}^{-1}[\Sigma_{Z'Z} - \sigma^2_2\xi'\xi']\Sigma_{Z'Z}^{-1}(\xi - \xi'\Pi)][c_6\Pi + \Sigma_{Z'Z}^{-1}(\sigma^2_2\xi' - \Sigma_{Z'X}\beta GIV)(\xi' - \xi'\Pi)]
+ 2\sigma^2_2[c_4\xi + c_5\Pi'\xi][c_4\xi + c_5\Pi'\xi] \]
\[ \begin{align*}
&[c_4I_k - \Pi'\xi_\beta GIV][\Sigma_{X'Z} - \sigma^2_2\xi'\xi'][c_6\Pi + \Sigma_{Z'Z}^{-1}(\sigma^2_2\xi' - \Sigma_{Z'X}\beta GIV)(\xi' - \xi'\Pi)]
+ [c_5\Pi' + (\xi - \Pi'\xi)(\sigma^2\xi' - \beta GIV\Sigma X'Z)\Sigma_{Z'Z}^{-1}[\Sigma_{Z'X} - \sigma^2_2\xi'\xi']\Sigma_{X'X}^{-1}\hat{X}'\xi' + (\beta GIV\Sigma X'X\beta GIV - \sigma^2_2\xi'\xi')\Pi'\xi_\Pi,
\end{align*} \]

Next, we examine these 5 terms of \( V_0 \) one by one. The first one is

\[ \begin{align*}
&c_2^2(\Sigma_{X'X} - \sigma^2_2\xi'\xi') - c_4(\Sigma_{X'X} - \sigma^2_2\xi'\xi')\beta GIV\xi_\Pi
- c_4\Pi'\beta GIV(\Sigma_{X'X} - \sigma^2_2\xi'\xi')\beta GIV\xi_\Pi
= c_2^2\Sigma_{X'X} - \sigma^2_2\xi'\xi' - c_4\Sigma_{X'X}\beta GIV\xi_\Pi + c_2^2\sigma_2\xi_\Pi
- c_4(\Sigma_{X'X}\beta GIV - \sigma^2_2\xi_\Pi)(\beta GIV\Sigma X'X\beta GIV - \sigma^2_2\xi_\Pi),
\end{align*} \]

the second is

\[ \begin{align*}
&[c_5\Pi'\Sigma Z'Z + (\xi - \Pi'\xi)(\sigma^2\xi' - \beta GIV\Sigma X'Z)]\Sigma_{Z'Z}^{-1}(\sigma^2_2\xi' - \Sigma_{Z'X}\beta GIV)(\xi' - \xi'\Pi)
- \sigma^2_2[c_5\Pi'\xi_\Pi + (\xi - \Pi'\xi)(\sigma^2\xi' - \beta GIV\Sigma X'Z)\Sigma_{Z'Z}^{-1}\xi_\Pi][c_6\Pi + \Sigma_{Z'Z}^{-1}(\sigma^2_2\xi' - \Sigma_{Z'X}\beta GIV)(\xi' - \xi'\Pi)]
\]
\[= c_3^2\Pi'\Sigma Z_z \Pi + c_3(\xi - \Pi'\zeta)(\sigma_2^2\zeta' \Pi - \bar{\beta}_{GIV}^1 \Sigma_r \Sigma x \bar{\beta}_{GIV}) (\xi' - \zeta' \Pi) + (\xi - \Pi' \zeta)(\sigma_2^2\zeta' \Sigma^{-1}_z z - \bar{\beta}_{GIV} \Pi)(\sigma_2^2\zeta - \Sigma x \bar{\beta}_{GIV}^1) (\xi' - \zeta' \Pi) - \sigma_2^2 c_3 \Pi' (\xi' \Pi) + (\xi - \Pi' \zeta) (\sigma_2^2 \zeta' - \bar{\beta}_{GIV} \Sigma x \Sigma x \bar{\beta}_{GIV} \Pi) \]

\[= c_3^2\Pi'\Sigma Z_z \Pi + c_3^2\pi \xi \zeta' \Pi - c_3^2 \xi \zeta' \Pi - c_3^2 \xi \zeta' \Pi + c_3^2 \xi \zeta' \Pi + c_3^2 \xi \zeta' \Pi
+ (\xi - \Pi' \zeta)(\sigma_2^2 \zeta' \Sigma^{-1}_z z - \bar{\beta}_{GIV} \Pi)(\sigma_2^2 \zeta' - \Sigma x \bar{\beta}_{GIV}^1)
- (\xi - \Pi' \zeta)(\sigma_2^2 \zeta' \Sigma^{-1}_z z - \bar{\beta}_{GIV} \Pi)(\sigma_2^2 \zeta' \Pi - \Sigma x \bar{\beta}_{GIV} \Pi)
- \sigma_2^2 c_3 \Pi' \xi \zeta' \Pi + (\xi - \Pi' \zeta) (c_1 \xi' \Pi - c_3 \xi' \Pi)
+ \sigma_2^2 c_3 \Pi' \xi \zeta' \Pi + (\xi - \Pi' \zeta)(c_1 \xi' \Pi - c_3 \xi' \Pi)(\sigma_2^2 \zeta' \Pi - \Pi \bar{\beta}_{GIV} \Pi)
\]

\[= c_3^2 \Sigma X + c_3^2 c_4 \xi \zeta' \Pi - c_3^2 c_4 \xi \zeta' \Pi - c_3^2 c_4 \xi \zeta' \Pi + c_3^2 c_4 \xi \zeta' \Pi
- \sigma_2^2 c_4 \xi \zeta' \Pi + c_3^2 c_4 \xi \zeta' \Pi + c_3^2 c_4 \xi \zeta' \Pi - c_3^2 c_4 \xi \zeta' \Pi
- \sigma_2^2 c_4 \xi \zeta' \Pi - \sigma_2^2 c_4 \xi \zeta' \Pi + c_3^2 c_4 \xi \zeta' \Pi + c_3^2 c_4 \xi \zeta' \Pi + c_3^2 c_4 \xi \zeta' \Pi - c_3^2 c_4 \xi \zeta' \Pi
\]

\[= c_3^2 \Sigma X + c_3^2 c_4 (c_4 - 1 - c_3) \Pi' \zeta' \Pi + c_3^2 c_4 (c_4 - 1 - c_3) \zeta' \Pi
- \sigma_2^2 c_4 \xi \zeta' \Pi + c_3^2 c_4 (c_4 - 1 - c_3) \Pi' \zeta' \Pi + c_3^2 c_4 (c_4 - 1 - c_3) \zeta' \Pi
\]

\[= c_3^2 \Sigma X + c_3^2 c_4 (c_4 - 1 - c_3) \zeta' \Pi + c_3^2 c_4 (c_4 - 1 - c_3) \zeta' \Pi + c_3^2 c_4 (c_4 - 1 - c_3) \zeta' \Pi + c_3^2 c_4 (c_4 - 1 - c_3) \zeta' \Pi + c_3^2 c_4 (c_4 - 1 - c_3) \zeta' \Pi + c_3^2 c_4 (c_4 - 1 - c_3) \zeta' \Pi
\]
and the final one is the transpose of the fourth. Collecting all terms we find

$$V_0 = -\sigma^2_G\Sigma_{X,X} - \sigma^2_G\Sigma_{X,X}\beta_{GIV}\zeta'\Pi + \sigma^2_G c_3 \zeta'\Pi,$$

and the fifth one is the transpose of the fourth. Collecting all terms we find

$$V_0 = -\sigma^2_G\Sigma_{X,X} - \sigma^2_G\Sigma_{X,X}\beta_{GIV}\zeta'\Pi + \sigma^2_G c_3 \zeta'\Pi,$$

so

$$\hat{V}_{GIV} = \frac{1}{\sigma^2_G}\left[\frac{\sigma^2_G\Sigma_{X,X}^2}{(1 - \sigma_G^2)^2} - \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2} + \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2} - \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2}\right].$$

Note that when $c_1 = c_2$, so $c_4 = 0$ and $c_5 = 1 - c_3$, and so the variance simplifies to

$$\hat{V}_{GIV} = \frac{1}{\sigma^2_G}\left[\frac{\sigma^2_G\Sigma_{X,X}^2}{(1 - \sigma_G^2)^2} - \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2} + \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2} - \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2}\right].$$

which is the result of Theorem 1.

Thus, the asymptotic variance of the (generalized) GIV estimator (40) is

$$\hat{V}_{GIV} = \frac{1}{\sigma^2_G}\left[\frac{\sigma^2_G\Sigma_{X,X}^2}{(1 - \sigma_G^2)^2} - \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2} + \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2} - \frac{\sigma^2_G\Sigma_{X,X}^2}{\Sigma_{X,X}^2}\right].$$

which is the result of Corollary 1.

**B Proof of Theorem 2**

When $\mu_4 \neq 3$ then there is an additional contribution to the asymptotic variance for which we have to evaluate $\sigma^2_G\omega'$. In obtaining the third term of (48) we already found

$$\omega' = c_4 c_5\zeta' + c_4 c_5\zeta' + c_4 c_5\zeta' + c_3\zeta'\Pi,$$

so

$$\sigma^2_G\Sigma_{X,X}^{\omega'}/\sigma^2_G\Sigma_{X,X}^{\omega'}.$$
In the overall asymptotic variance this term has factor $\mu_4 - 1$. The expression for $V_{GIV}^{N}$ already contains it with factor 2, so the additional term given in Theorem 4 has factor $\mu_4 - 3$.

When $\mu_3 \neq 0$ there is another additional contribution to the asymptotic variance, for which we have to evaluate

$$\mu_3 \sigma^2 c_3 (\Sigma W) \omega' + \omega \Sigma W.$$ 

Since

$$W' = \Theta_1 X' + \Theta_2 Z' t,$$

$$= c_4 X' t - \Pi' \beta' GIV X' t + c_5 \Pi' Z' t + (\xi - \Pi' \zeta)(\sigma_2^2 \zeta' - \beta' GIV \Sigma X' Z) \Sigma Z' Z' t,$$

we find

$$\Sigma W' = c_4 \Sigma X' t + \Pi' \beta' GIV \Sigma X' t + c_5 \Pi' Z' t + (\xi - \Pi' \zeta)(\sigma_2^2 \zeta' - \beta' GIV \Sigma X' Z) \Sigma Z' Z' t,$$

and with $\omega' = c_4 \xi' + c_5 \xi' \Pi'$, we obtain

$$\Sigma W' \omega' = c_4 [c_4 \Sigma X' t - \Pi' \beta' GIV \Sigma X' t + c_5 \Pi' Z' t + (\xi - \Pi' \zeta)(\sigma_2^2 \zeta' - \beta' GIV \Sigma X' Z) \Sigma Z' Z' t]$$

Thus,

$$\sigma^2 \Sigma_{X' t} \Sigma W' \omega' \Sigma_{t' X}$$

$$= \sigma^2 c_4 [c_4 \Sigma X' t - \Pi' \beta' GIV \Sigma X' t + c_5 \Pi' Z' t + (\xi - \Pi' \zeta)(\sigma_2^2 \zeta' - \beta' GIV \Sigma X' Z) \Sigma Z' Z' t]$$

$$+ (\sigma^2 c_4 \Sigma X' t - \Pi' \beta' GIV \Sigma X' t + c_5 \Pi' Z' t + (\xi - \Pi' \zeta)(\sigma_2^2 \zeta' - \beta' GIV \Sigma X' Z) \Sigma Z' Z' t)$$

and the additional term is then equal to $\mu_3$ multiplied by

$$\sigma^2 \Sigma_{X' t} \Sigma W' \omega' \Sigma_{t' X}$$

$$= c_4 [c_4 \Sigma X' t - \Pi' \beta' GIV \Sigma X' t + c_5 \Pi' Z' t + (\xi - \Pi' \zeta)(\sigma_2^2 \zeta' - \beta' GIV \Sigma X' Z) \Sigma Z' Z' t]$$

This expression can be simplified slightly when we assume that both matrices $X$ and $Z$ have a first column of ones. Then

$$\Sigma_{X' t} \Sigma_{X' t} = e_{k, 1},$$

$$\Pi' Z' t = \Sigma_{X' Z' t} \Sigma_{Z' t} = \Sigma_{X' Z} e_{1, 1},$$
where $e_{f,g}$ denotes a $f \times 1$ unit vector which has all elements equal to zero apart from a unit element in position $g$. This yields

\[
\sigma_e^3 \Sigma_{X'X}^{-1} [\Sigma_{X'X} \omega' + \omega \Sigma_{X'}] \Sigma_{X'X}^{-1} \\
= c_4 \{ \sigma_e^2 e_{k,1} \zeta e_{k,1} \Sigma_{X'X}^{-1} \Sigma_{X'X} - \sigma_e^2 \beta_{GIV} \beta_{GIV} \Sigma_{X'X} \xi' \Sigma_{X'X}^{-1} \Sigma_{X'X} + \sigma_e^2 c_5 \Sigma_{X'X} \Sigma_{X'} e_{k,1} \xi' \Sigma_{X'X}^{-1} \Sigma_{X'X} \\
+ (\sigma_e^2 \Sigma_{X'X}^{-1} \xi - \sigma_e^2 \beta_{GIV})(\sigma_e^2 \xi' - \beta_{GIV} \Sigma_{X'} e_{k,1} \xi' \Sigma_{X'X}^{-1} \Sigma_{X'X}) \} \\
+ c_6 [\sigma_e^2 e_{k,1} \beta_{GIV} - \sigma_e^2 \beta_{GIV} \Sigma_{X'} \beta_{GIV} \beta_{GIV} + \sigma_e^2 c_5 \Sigma_{X'X} \Sigma_{X'} e_{k,1} \beta_{GIV} \\
+ (\sigma_e^2 \Sigma_{X'X}^{-1} \xi - \sigma_e^2 \beta_{GIV})(\sigma_e^2 \xi' - \beta_{GIV} \Sigma_{X'} e_{k,1} \beta_{GIV}) \} \\
+ c_4 [\sigma_e^2 \Sigma_{X'X}^{-1} \xi e_{k,1} \Sigma_{X'X} \beta_{GIV} \beta_{GIV} + \sigma_e^2 c_5 \Sigma_{X'X} \Sigma_{X'} e_{k,1} \beta_{GIV} \\
+ \Sigma_{X'X} \xi e_{k,1} (\sigma_e^2 \xi - \Sigma_{X'} \beta_{GIV})(\sigma_e^2 \xi' - \beta_{GIV} \Sigma_{X'} e_{k,1} \beta_{GIV}) \} \\
+ c_6 [\sigma_e^2 \beta_{GIV} e_{k,1} - \sigma_e^2 \beta_{GIV} \Sigma_{X'} \beta_{GIV} \beta_{GIV} + \sigma_e^2 c_5 \beta_{GIV} e_{k,1} \Sigma_{X'} \Sigma_{X'}^{-1} \Sigma_{X'X} \\
+ \beta_{GIV} e_{k,1} \sigma_e^2 \xi - \Sigma_{X'} \beta_{GIV}) (\sigma_e^2 \xi' - \beta_{GIV} \Sigma_{X'} e_{k,1} \beta_{GIV}) \}.
\]

When $k = l$, i.e. $c_1 = c_2$, $c_5 = 1 - c_3$, $\Pi \Sigma_{X'X}^{-1} = \Sigma_{X'X}^{-1}$ and $\sigma_e^2 \xi - \Sigma_{X'} \beta_{GIV} = 0$, $V^{NN}_{GIV}$ specializes to the expression

\[
V^{NN}_{IV} = \sigma_e^2 \Sigma_{X'X} \Sigma_{Z'Z} \Sigma_{X'X}^{-1} \Sigma_{X'X}^{-1} Z [2c_2 - 2c_5 + 1 - \sigma_e^2 (\beta_{IV} \Sigma_{X'} \beta_{IV})] \beta_{IV} \beta_{IV}^{-1} \\
+ (\mu_4 - 1) \sigma_e^2 \beta_{IV} \beta_{IV}^{-1} - 2\mu_5 \sigma_e^2 \beta_{IV} \Sigma_{X'} \beta_{IV} \beta_{IV} + \mu_5 \sigma_e^2 \Sigma_{X'} \Sigma_{Z'} e_{k,1} \beta_{IV} + \beta_{IV} \Sigma_{X'} \Sigma_{X'}^{-1} \Sigma_{X'X} \\
= \sigma_e^2 \Sigma_{X'X} \Sigma_{Z'Z} \Sigma_{X'X}^{-1} \Sigma_{X'X}^{-1} Z + \mu_5 \sigma_e^2 \Sigma_{X'} \Sigma_{X'} \Sigma_{X'} \Sigma_{X'} \Sigma_{X'} \Sigma_{X'} \Sigma_{X'}^{-1} \Sigma_{X'X} \\
- [(3 - \mu_4) \sigma_e^2 + 2c_5 (\mu_4 \sigma_e^2 \beta_{IV} \Sigma_{X'} - 1) + 1 - \sigma_e^2 (\beta_{IV} \Sigma_{X'} \beta_{IV})] \beta_{IV} \beta_{IV}^{-1}.
\]
Figure 1: Densities for $k = l = 1; \beta = 1; \sigma_x^2/\sigma_z^2 = 10; n = 50, 200.$
Figure 2: Densities for $k = l = 1; \beta = 1; \sigma_z^2/\sigma_x^2 = 10; n = 50, 200.$
Figure 3: Densities for $k = l = 1; \beta = 1; \sigma_l^2/\sigma_z^2 = 10; n = 50, 200.$
Figure 4: Densities for $k = l = 1; \beta = 1; \sigma_z^2/\sigma_x^2 = 10; n = 50, 200$. 

31
Figure 5: $\log[MAE(\hat{\beta}_{IV})/AMAE(\hat{\beta}_{IV})]$ for $k = l = 1$; any $SN$; $n = 20, 100$. 
Figure 6: log\[MAE(\hat{\beta}_{OLS})/MAE(\hat{\beta}_{IV})\] for $k = l = 1$; any $SN$; $n = 20, 100$. 

33
Figure 7: Densities for \( l - 1 = k = 1; \) \( \beta = 1; \) \( \frac{\sigma_x^2}{\sigma_z^2} = 10; \) \( n = 50, 200. \)
Figure 8: Densities for $l - 1 = k = 1; \beta = 1; \sigma_x^2/\sigma_z^2 = 10; n = 50, 200.$
Figure 9: log[MAE(\(\hat{\beta}_{GIV}\))/AMAE(\(\hat{\beta}_{GIV}\))] for \(k = l - 1 = 1\); any SN; \(n = 20\)
Figure 10: \( \log[MAE(\hat{\beta}_{OLS})/MAE(\hat{\beta}_{GIV})] \) for \( k = l - 1 = 1 \); any SN; \( n = 100 \).