On the limiting and empirical distribution of IV estimators when some of the instruments are invalid
Kiviet, J.F.; Niemczyk, J.

Citation for published version (APA):

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

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
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*

preliminary draft of 11 July 2006

JEL-classification: C13, C15, C30

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

Abstract

In practice IV estimation will often be employed when in fact some of the instruments are invalid. This is because moment conditions can only be tested if already a sufficient number of valid but untestable instruments is available. Moreover, tests for the validity of additional instruments, i.e. so-called overidentification restriction tests, will have limited power when instruments are weak and samples are small. We examine the case where the model is treated as over or just identified, although some of the instruments may actually be invalid, whereas all variables are stationary. We derive an expression in terms of the various parameters of the data generating process for the inconsistency of the invalid IV estimator and obtain its limiting normal distribution. In specific simple models we scan this approximation to the finite sample distribution over the relevant parameter space and compare it with the actual empirical distribution obtained by simulation. This parameter space is transformed to measures for: (i) model fit; (ii) simultaneity; (iii) instrument invalidity; and (iv) instrument weakness. We present our findings over this multi-dimensional parameter space by using dynamic visualization techniques, which can best be enjoyed on screen. Our major findings are that: (a) for the accuracy of large sample asymptotic approximations instrument weakness is much more detrimental than instrument invalidity; (b) the realizations of IV estimators obtained from strong but possibly invalid instruments seem 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, whereas this correlation is unknown, then one needs further variables to use as instruments in order to find consistent estimators by the method of moments. These instrumental variables should have known (usually zero) correlation with the disturbances. Then they provide so-called orthogonality conditions that make it possible to obtain consistent instrumental variable (IV) estimators. In practice, however, it is mostly hard to assess whether an instrumental variable is valid indeed, 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 built 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 in fact some of the instruments are 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), i.e. 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 ≥ k). We focus on the distribution of such an invalid IV estimator for a single structural equation that for the rest has been specified correctly in the sense that its implied error term is IID (independent and identically distributed) with unconditional expectation equal to zero. An alternative point of departure is chosen in Hale et al. (1980) where instruments are invalid due to omitted regressors. We derive an expression in terms of parameters and data moments for the inconsistency of the invalid IV estimator in an otherwise correctly specified model and also obtain its limiting distribution. These results yield a first-order asymptotic approximation to the actual distribution in finite sample of IV estimators. 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). By simulation we verify over the relevant parameter space of simple classes of models whether these analytic findings are accurate regarding the actual estimator distribution in finite sample and 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 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 some
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 an-
imations. 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 face one might conjecture that both the invalidity and the strength of the extra
instrument might matter separately.

The analysis of IV estimators employing invalid instruments has not received much
attention in the literature yet. Although its limiting behavior has been examined before,
see Maasumi and Phillips (1982), Hahn and Hausman (2003) and Hall and Inoue (2003),
one of these studies provides a simple 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\(^1\) that yielded similar results for inconsistent OLS estimators.
The latter can be found in Kiviet and Niemczyk (2005), which completes some initial
results obtained in Joseph and Kiviet (2005). As far as we know simulation evidence on
the actual finite sample distribution of valid and invalid IV estimators covering almost
the entire parameter space of some basic models as presented here has not been produced
before. The analysis of the exact finite sample properties of consistent IV estimators has
a long history, see Sawa (1969) and Phillips (1980). Recent contributions and further
references can be found in, for instance, Phillips (2005) and Hillier (2005). Our find-
ings illustrate those on the effects of instrument weakness on the finite sample density
of consistent IV estimators by both Woglom (2001), who focusses on just identified IV
estimators, and Forchini (2006), who gives further theoretical underpinnings in case of
overidentification. They also supplement these findings, because we provide more exten-
sive illustrations and consider invalid instruments as well. 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 (2005), the present study analyzes and illustrates properties of the
distribution of coefficient estimators.

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 distri-
bution 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 find also that when the instrument
is weak the probability mass of the actual distribution of instrumental variable estima-
tors is generally much closer to the true value of the coefficient than indicated by these
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 skew, and
that it becomes bimodal for very strong simultaneity, whereas for extreme weakness (i.e.

\(^1\) A related approach can be found in an unpublished discussion paper by Rothenberg (1972).
close to underidentification) the dispersion explodes, while 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 to use weak instruments if one would lift the ban on invalid 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 the standard asymptotic approximation is 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. So, even when its actual behavior could be adequately approximated by alternative weak-instrument asymptotic methods it still may have an actual distribution that is less attractive than that of an IV estimator based on strong but possibly invalid instruments.

The structure of this paper is as follows. In Section 2 we introduce the model to be estimated 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. Next the results are specialized for the just identified case. 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 the 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 \( l - 1 = k = 1 \). Hence, we focus on the very simple model with just one endogenous explanatory variable and one or two - possibly weak and possibly invalid - instrumental variables. Finally, Section 4 concludes.

## 2 Model, assumptions and theorems

We consider data generating processes for variables for which \( n \) observations have been collected in \( y, X \) and \( Z \), which all have \( n \) rows. The matrices \( X \) and \( Z \) have \( k \) and \( l \) columns respectively, with \( l \geq k \). \( X \) contains the explanatory variables for 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 stationarity and regularity conditions.

**Framework A.** We have: (i) the structural equation \( y = X\beta + \varepsilon \); (ii) with disturbances having (for \( i \neq h = 1, ..., n \)) the (finite) unconditional moments \( \mathbb{E}(\varepsilon_i) = 0, \mathbb{E}(\varepsilon_i\varepsilon_h) = \)
\(0, \quad E(\varepsilon_1^2) = \sigma_\varepsilon^2, \quad E(\varepsilon_2^3) = \mu_3\sigma_\varepsilon^3\) and \(E(\varepsilon_4^4) = \mu_4\sigma_\varepsilon^4;\) (iii) while \(X = \bar{X} + \varepsilon\xi'\) and \(Z = \bar{Z} + \varepsilon\zeta',\) such that \(E(\bar{X}'\varepsilon) = 0\) and \(E(\bar{Z}'\varepsilon) = 0,\) with \(\xi\) and \(\zeta\) fixed parameter vectors of \(k\) and \(l\) elements respectively. Moreover, (iv) \(\Sigma_{X'X} = \text{plim}_{n \to \infty} n^{-1}X'X, \quad \Sigma_{Z'Z} = \text{plim}_{n \to \infty} n^{-1}Z'Z\) and \(\Sigma_{Z'X} = \text{plim}_{n \to \infty} n^{-1}Z'X\) have all full column rank, and (v) so have \(X'X, \quad Z'Z\) and \(Z'X\) with probability one. Finally, (vi) we have \(E(\frac{1}{n}Z'Z) = 0\) and \(E(\frac{1}{n}X'X | \bar{X}, \bar{Z}) = \Sigma_{X'X} = o_p(n^{-1/2}).\)

Note that A(iii) implies
\[
E(X'\varepsilon) = n\sigma_\varepsilon^2 \xi \quad \text{and} \quad E(Z'\varepsilon) = n\sigma_\varepsilon^2 \zeta. \tag*{(1)}
\]

Hence, if \(\xi_j = 0\) for some \(j \in \{1,...,k\}\) 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 \(g \in \{1,...,l\}\) then the \(g\)-th column of \(Z\) establishes a valid instrument, and it can be shown that A(vi) boils down to the mild regularity assumptions \(\frac{1}{n}Z'Z - \text{plim}_{n \to \infty} n^{-1}Z'Z = o_p(n^{-1/2})\) and \(\frac{1}{n}X'X - \text{plim}_{n \to \infty} n^{-1}X'X = o_p(n^{-1/2}).\)

Since \(l \geq k\) the generalized instrumental variable (GIV) or 2SLS estimator of \(\beta\) exists and is given by
\[
\hat{\beta}_{\text{GIV}} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y = (\bar{X}'\bar{X})^{-1}\bar{X}'y, \tag*{(2)}
\]

where we introduced the notation
\[
\bar{X} = ZZ\hat{\Pi} = (Z(Z'Z)^{-1}Z'X, \tag*{(3)}
\]

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 \(\hat{\beta}_{\text{GIV}}\) exists. We define
\[
\beta_{\text{GIV}}^* = \text{plim} \hat{\beta}_{\text{GIV}} = \beta + \text{plim} [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon = \beta + \sigma_\varepsilon^2 \Sigma_{X'Z}^{-1} \Sigma_{Z'Z}^{-1} \Sigma_{X'X} \Sigma_{Z'Z}^{-1} \zeta, \tag*{(4)}
\]

where \(\beta_{\text{GIV}}^*\) is also known as the pseudo-true-value of \(\hat{\beta}_{\text{GIV}}\). We shall denote the inconsistency of \(\hat{\beta}_{\text{GIV}}\) as
\[
\hat{\beta}_{\text{GIV}} = \beta_{\text{GIV}}^* - \beta = \sigma_\varepsilon^2 \Sigma_{X'Z}^{-1} \Sigma_{Z'Z}^{-1} \Sigma_{X'X} \Sigma_{Z'Z}^{-1} \zeta = \sigma_\varepsilon^2 \Sigma_{X'X} \Pi' \zeta, \tag*{(5)}
\]

where we used \(\Sigma_{X'X} = \Sigma_{X'Z} \Sigma_{Z'Z}^{-1} \Sigma_{Z'X}\) and \(\Pi = \text{plim}(Z'Z)^{-1}Z'X = \Sigma_{Z'Z}^{-1} \Sigma_{Z'X}.\) 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.
\[
\hat{\beta}_{\text{IV}} = (Z'X)^{-1}Z'y, \tag*{(6)}
\]
\[
\beta_{\text{IV}}^* = \beta + \sigma_\varepsilon^2 \Sigma_{Z'X}^{-1} \zeta, \tag*{iv}
\]
\[
\hat{\beta}_{\text{IV}} = \sigma_\varepsilon^2 \Sigma_{Z'X}^{-1} \zeta. \tag*{v}
\]
When in fact $Z = X$ (all regressors are used as instruments), i.e. $\zeta = \xi$, then IV specializes to OLS, i.e.

$$\begin{align*}
\beta_{\text{OLS}} &= (X'X)^{-1}X'y, \\
\beta_{\text{OLS}}^* &= \beta + \sigma_\varepsilon^2 \Sigma_{X,X}^{-1} \xi, \\
\hat{\beta}_{\text{OLS}} &= \sigma_\varepsilon^2 \Sigma_{X,X}^{-1} \xi.
\end{align*}$$

(7)

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

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

For GIV and IV estimators we now obtain the following results (proved in appendices) on their convergence in distribution.

**Theorem 1.** In Framework B we have $n^{1/2}(\hat{\beta}_{\text{GIV}} - \beta_{\text{GIV}}^*) \to N\left(0, V_{\text{GIV}}^N\right)$, with

$$V_{\text{GIV}}^N = \sigma_\varepsilon^2 c_1 (1 - c_3 + c_4) \Sigma_{X,X}^{-1} + \sigma_\varepsilon^2 c_4 \Sigma_{X,X}^{-1} \Sigma_{X,X} \Sigma_{X,X}^{-1} X \xi
$$

$$- c_4 [\Sigma_{X,X}^{-1} \Sigma_{X,X} \hat{\beta}_{\text{GIV}}^* \hat{\beta}_{\text{GIV}}^*] + \sigma_\varepsilon^2 c_4 (1 - 2c_4) \Sigma_{X,X}^{-1} \xi \hat{\beta}_{\text{GIV}}^* + \sigma_\varepsilon^2 c_4 (1 - 2c_4) \Sigma_{X,X}^{-1} \xi
$$

$$+ [c_3 (1 - 2c_3) - c_3 + \sigma_\varepsilon^2 (\Sigma_{X,X}^{-1} \Sigma_{X,X} \Sigma_{X,X}^{-1} \Sigma_{X,X}^{-1} X \xi)] \Sigma_{X,X}^{-1} \Sigma_{X,X}^{-1} X \xi,$$

where $c_1 = \sigma_\varepsilon^2 c' \Sigma_{X,X}^{-1} \xi$, $c_2 = c_1 c_3 = c_3$, $c_4 = c_1 - c_2$ and $c_5 = 1 - c_3 - c_4$.

The $N$ in the superindex of $V_{\text{GIV}}^N$ 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 $\hat{\beta}_{\text{GIV}}^*$ is still genuinely normal when instruments are invalid, although no longer centered at $\beta$ but at the pseudo_true-value $\beta_{\text{GIV}}^*$. When all instruments are valid, i.e. $\zeta = 0$, then $\beta_{\text{GIV}}^* = \beta$, $\beta_{\text{GIV}} = 0$ 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 $n^{1/2}(\hat{\beta}_{\text{GIV}} - \beta) \to N\left(0, \sigma_\varepsilon^2 \Sigma_{X,X}^{-1}\right)$. Note that when all instruments are valid the asymptotic variance of $\beta_{\text{GIV}}$ is not determined by the simultaneity $\zeta$, because $\Sigma_{X,X} = \Sigma_{X,X} \Sigma_{X,X} \Sigma_{X,X}^{-1} \Sigma_{X,X} \Sigma_{X,X}^{-1} X \xi$. However, when instruments are invalid, i.e. $\zeta \neq 0$, then $\Sigma_{Z,X} = \Sigma_{Z,X} + \sigma_\varepsilon^2 \zeta \xi$ and thus $\Sigma_{X,X}$ is determined by both $\zeta$ and $\xi$. Then, when fitting $X$ to $Z$, the $\varepsilon \zeta$ part of $X$ is no longer (asymptotically) orthogonal to $Z$, due to the presence of $\varepsilon \zeta$. 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 $\zeta' \Sigma_{X,X} \hat{\beta}_{\text{GIV}} = \sigma_\varepsilon^2 \zeta' \Sigma_{X,X} \Sigma_{X,X} \Sigma_{X,X}^{-1} \xi = \sigma_\varepsilon^2 \zeta' \Sigma_{X,X} \Sigma_{X,X}^{-1} \xi$, 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 $n^{1/2}(\hat{\beta}_{\text{IV}} - \beta_{\text{IV}}^*) \to N\left(0, V_{\text{IV}}^N\right)$, with

$$V_{\text{IV}}^N = \sigma_\varepsilon^2 (1 - c_3)^2 \Sigma_{Z,Z} \Sigma_{Z,Z} \Sigma_{Z,Z} - [2c_3 - 2c_3 + 1 - \sigma_\varepsilon^2 (\beta_{\text{IV}} \Sigma_{X,X} \beta_{\text{IV}})] \beta_{\text{IV}} \beta_{\text{IV}}^*,$$
where \( c_3 \equiv \xi' \tilde{\beta}_{IV} = \sigma^2 \xi' \Sigma_{Z'X}^{-1} \xi \).

When all instruments are valid, i.e. \( \xi = 0 \), this result specializes to the standard result \( n^{1/2}(\hat{\beta}_{IV} - \beta^*_IV) \to N \left( 0, \sigma^2 \Sigma_{X'X}^{-1} \right) \). Since for general \( \zeta \) and \( \xi \) the scalar \( \sigma^2 \xi' \Sigma_{Z'X}^{-1} \xi \) 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 \( \sigma^2 \Sigma_{Z'Z}^{-1} \Sigma_{Z'X} \Sigma_{X'Z}^{-1} \). Depending on the particular parametrization and data moment matrices the asymptotic variance of individual coefficient estimates may either increase or decrease, due to \( \xi \neq 0 \) or \( \xi \neq 0 \).

When \( Z = X \), which gives \( \zeta = 0 \) 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 (2005).

Next we look at the case where the disturbances may have general 3rd and 4th moment. Let \( n \) be a \( n \times 1 \) vector of unit elements. Upon defining
\[
\Sigma_{Z'V} = \text{plim} \, n^{-1} Z'V = \text{plim} \, n^{-1} \tilde{Z}'V \equiv \Sigma_{Z'V},
\]
\[
\Sigma_{Z'X} = \text{plim} \, n^{-1} Z'X = \text{plim} \, n^{-1} \tilde{Z}'X \equiv \Sigma_{Z'X},
\]
we find (superindex \( NN \) indicates nonnormal disturbances):

**Theorem 2.** In Framework A we have \( n^{1/2}(\hat{\beta}_{GIV} - \beta^*_GIV) \to N \left( 0, V_{GIV}^{NN} \right) \), 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
\[
(\mu_4 - 3)\left\{ \sigma^2 \xi' \Sigma_{X'X} \xi' \Sigma_{X'X}^{-1} - \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} \right\} \left\{ \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} \right\}
\]
where \( c_4 = \sigma^2 \xi' \Sigma_{Z'X}^{-1} \xi' \Sigma_{X'X}^{-1} + c_3 - \mu_4 - \xi' \Sigma_{Z'X}^{-1} \xi' \Sigma_{X'X}^{-1} \). When \( \mu_3 \neq 0 \) the additional term is
\[
\mu_3 \left\{ c_4 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} \xi' \Sigma_{X'X}^{-1} - \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} \right\} \left\{ \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} \right\}
\]
\[
+ c_5 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} \xi' \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} + \sigma^2 \xi' \Sigma_{X'X} \Sigma_{X'X}^{-1} \right\}
\]
When all instruments are valid, i.e. \( \zeta = 0 \), then this result again collapses to the standard result, i.e. \( V_{GIV}^{NN} = \sigma^2 \Sigma_{X'X}^{-1} \), which highlights that normality of the disturbances is not a requirement for the standard normal limiting distribution of \( \beta_{GIV} \).

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

**Corollary 2.** In Framework A for the special case \( l = k \) we have \( n^{1/2}(\hat{\beta}_{IV} - \beta^*_IV) \to N \left( 0, V_{IV}^{NN} \right) \), with
\[
V_{IV}^{NN} = \sigma^2 \Sigma_{Z'X} \Sigma_{Z'X} \Sigma_{X'X}^{-1} + \mu_5 \xi_3 \xi_3 \Sigma_{Z'X} \Sigma_{Z'X} \Sigma_{Z'X} \Sigma_{X'X}^{-1} + \tilde{\beta}_{IV} \Sigma_{Z'X} \Sigma_{X'X}^{-1}
\]
\[
- \left\{ (5 - \mu_4) \xi_3 + 2 \xi_3 \mu_3 \Sigma_{Z'X} \Sigma_{X'X}^{-1} \right\} + 1 - \sigma^2 \Sigma_{X'X} \Sigma_{X'X} \Sigma_{X'X}^{-1} \tilde{\beta}_{IV} \Sigma_{Z'X} \Sigma_{X'X}^{-1}
\]
where \( c_3 \equiv 1 - \xi' \tilde{\beta}_I V = 1 - \sigma^2 \xi' \Sigma^{-1}{Z}'X \xi. \)

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 that is a straightforward extension of the following simple CLT (central limit theorem), which says: Let \( v_i \) be a \( k \times 1 \) random vector such that \( E(v_i) = 0, E(v_i v_i') = V_i \) and \( E(v_i v_h') = O \) for \( i \neq h = 1, \ldots, n, \) then \( n^{1/2} \tilde{v} \rightarrow N(0, \lim_{n \rightarrow \infty} V) \), 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 \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \) has mutually uncorrelated elements for which \( E(\varepsilon_i | w_i) = 0, E(\varepsilon_i^2 | w_i) = \sigma^2 \), \( E(\varepsilon_i^3 | w_i) = \mu_3 \sigma^3 \varepsilon \) and \( E(\varepsilon_i^4) = \mu_4 \sigma^4 \). Then the \( k \times 1 \) vector \( v_i = w_i \varepsilon_i + \omega (\varepsilon_i^2 - \sigma^2) \) has zero expectation, conditional variance \( E(v_i v_i' | w_i) = V_i = \sigma^2 w_i w_i' + \mu_3 \varepsilon (w_i \omega' + \omega w_i') + (\mu_4 - 1) \sigma^4 \omega \omega' \), whereas \( E(v_i v_h') = O \) for \( i \neq h \), so that for \( n^{-1/2} \sum_{i=1}^n v_i = n^{-1/2} [W' \varepsilon + \omega (\varepsilon^2 - n \sigma^2)] \) the CLT implies

\[
n^{1/2} \tilde{v} \rightarrow N[0, \sigma^2 \Sigma_{WW} + \mu_3 \sigma^3 \Sigma_W \omega' + \omega \Sigma_{WW}] + (\mu_4 - 1) \sigma^4 \omega \omega',
\]

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

3 Illustrations

To illustrate the analytical asymptotic findings obtained in the foregoing section, we shall 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. Also we 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

\[
n^{1/2}(\hat{\beta} - \beta - \bar{\beta}) \rightarrow N(0, V)
\]

and they imply a first-order approximation to the distribution of \( \hat{\beta} \) in finite sample that can be expressed as

\[
\hat{\beta} \sim N(\beta + \bar{\beta}, n^{-1} V).
\]

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

\[
\text{AMSE}(\hat{\beta}) \equiv n^{-1} V + \tilde{\beta} \tilde{\beta}'.
\]

The actual values of \( \hat{\beta} \) 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 (9) 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 \( l - k \leq 1 \). 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

\[
\Pr\{ (\hat{\beta} - \beta) \leq \text{ME}(\hat{\beta}) \} = 0.5, \\
\Pr\{ |\hat{\beta} - \beta| \leq \text{MAE}(\hat{\beta}) \} = 0.5. \tag{10}
\]

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 likewise for \( \text{MAE}(\hat{\beta}) \) after sorting the values \( |\hat{\beta}^{(r)} - \beta| \). 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

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

so that we can solve\(^2\) \( m \) from

\[
\Phi_{\hat{\beta},\sigma_{\hat{\beta}}}(m) = 0.5 + \Phi_{\hat{\beta},\sigma_{\hat{\beta}}}(-m). \tag{11}
\]

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

### 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 and one possibly invalid and either strong or weak instrument, i.e. \( k = l = 1 \). The two variables \( x \) and \( z \), together with the dependent variable \( y \), are

\(^2\)Since \( m = \Phi_{\hat{\beta},\sigma_{\hat{\beta}}}^{-1}[0.5 + \Phi_{\hat{\beta},\sigma_{\hat{\beta}}}(-m)] \), we employed the iterative scheme, \( m_0 = 0 \), \( m_{i+1} = \Phi_{\hat{\beta},\sigma_{\hat{\beta}}}^{-1}[0.5 + \Phi_{\hat{\beta},\sigma_{\hat{\beta}}}(-m_i)] \) for \( i = 0, 1, \ldots \) until convergence. When \( \hat{\beta} = 0 \) no iteration is required since \( m = \Phi_{0,\sigma_{\hat{\beta}}}^{-1}(0.75) \) conforms to the quartile.
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 been addressed often before, recently in Woglom (2001) and Hillier (2005), and for \( l \geq 1 \) in Bound et al. (1995) and Hahn and Hausman (2003), although only in the latter paper invalid instruments are being considered.

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 \( \zeta = \sigma_{xe}/\sigma_x^2 \) and \( \xi = \sigma_{xe}/\sigma_z^2 \) we obtain

\[
\hat{\beta}_{IV} = \frac{\beta_{IV} - \beta}{1 - \rho_{xe}^2} = \frac{\sigma_x^2 \Sigma_{X'X} \zeta}{\sigma_{xz}} = \frac{\rho_{xe} \sigma_x}{\rho_{xz} \sigma_z} \tag{12}
\]

\[
c_3 = \frac{\sigma_x^2 \Sigma_{X'X} \zeta}{\sigma_{x}^2} = \frac{\rho_{xe} \rho_{xz}}{\rho_{xz}} \tag{13}
\]

\[
\hat{\beta}'_{IV} \Sigma_{X'X} \hat{\beta}_{IV} = \frac{\sigma_x^2 \Sigma_{X'X} \Sigma_{X'Z} \zeta}{\sigma_{xz}^2} = \frac{\sigma_x^2 \rho_{xz}^2}{\sigma_{xz}^2},
\]

giving

\[
V^N_{IV} = \frac{\sigma_x^2}{\sigma_z^2} \frac{(1 - \rho_{xe}^2)(\rho_{xz} - \rho_{xe}\rho_{xz})^2 + \rho_{xe}^4(1 - \rho_{xe}^2)}{\rho_{xz}^2},
\]

in the case where the disturbances are (almost) normally distributed. The expression for the inconsistency \( \hat{\beta}_{IV} \) shows that its sign is determined by the sign of \( \rho_{xe}/\rho_{xz} \), whereas its magnitude is inversely related to the strength of the instrument, cf. Bound et al. (1995). \( V^N_{IV} \) is unaffected by the signs of \( \rho_{xe}, \rho_{xz} \) and \( \rho_{xz} \) as long as the sign of the product \( \rho_{xe}\rho_{xz} \) remains the same, or when either \( \rho_{xe} \) or \( \rho_{xz} \) is zero. Self-evidently, \( V^N_{IV} \) 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 \( x_i = x_i^\# / \beta^\# \).

An important characteristic of the model is the signal-to-noise ratio (SN), which is equal to

\[
SN = \frac{\beta^2 \sigma_x^2}{\sigma_z^2} = \frac{\sigma_x^2}{\sigma_z^2}. \tag{14}
\]

From (12) and (13) we find that \( V^N_{IV} \) and \( \beta \) are proportional to (the square root of) the inverse of SN. In fact, after normalization to \( \beta = 1 \), the approximation to the distribution of the IV estimator in this simple model \( \hat{\beta}_{IV} \sim N(\beta + \beta^\#, n^{-1}V^N_{IV}) \) is completely determined by \( n \) and the four model characteristics \( \rho_{xz}, \rho_{xe}, \rho_{xz} \) and \( SN \).

Next we focus on obtaining an appropriate data generating scheme for this model, which is to be used in the simulations. In the notation of Section 2 it should be given by

\[
y_i = \beta x_i + \varepsilon_i \\
x_i = \tilde{x}_i + \xi \varepsilon_i \\
z_i = \tilde{z}_i + \zeta \varepsilon_i \tag{15}
\]

where \( \xi \) and \( \zeta \) are scalar. In order to obtain \((\varepsilon_i, x_i, z_i)' \sim \text{IID}(0, \Omega)\), with appropriate \( 3 \times 3 \) covariance matrix \( \Omega \), we can first generate \( v_i = (v_{i,1}, v_{i,2}, v_{i,3})' \sim \text{IID}(0, I_3) \) and
then parameterize as follows:

\[ \begin{align*}
\varepsilon_i & = \sigma_\varepsilon v_{i,1}, \\
\bar{x}_i & = \alpha_1 v_{i,2}, \\
\bar{z}_i & = \alpha_2 v_{i,2} + \alpha_3 v_{i,3}. 
\end{align*} \]

This provides full generality. The coefficient \( \alpha_1 \) determines \( \sigma_\varepsilon^2 \), whereas \( E(\varepsilon_i \varepsilon_i) = 0 \), 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 \( \sigma_\varepsilon^2 \). The above implies

\[ \begin{pmatrix} \varepsilon_i \\ x_i \\ z_i \end{pmatrix} = \Omega^{1/2} v_i = \begin{pmatrix} \sigma_\varepsilon & 0 & 0 \\ \sigma_\varepsilon \xi & \alpha_1 & 0 \\ \sigma_\varepsilon \zeta & \alpha_2 & \alpha_3 \end{pmatrix} \begin{pmatrix} v_{i,1} \\ v_{i,2} \\ v_{i,3} \end{pmatrix}. \quad (16) \]

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

\[ \hat{\beta}_{IV} = \frac{\sum z_i y_i}{\sum z_i x_i} = \beta + \frac{\sum z_i \varepsilon_i}{\sum z_i x_i}, \quad (17) \]

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

\[ \begin{align*}
\sigma_\varepsilon & = 1, \\
\sigma_\varepsilon^2 & = \zeta^2 + \alpha_2^2 + \alpha_3^2 = 1. 
\end{align*} \quad (18, 19) \]

By (18) we normalize all results with respect to \( \sigma_\varepsilon \), and (14) simplifies to

\[ SN = \sigma_x^2. \quad (20) \]

Because any GIV estimator is invariant to the scale of the instruments (only the space spanned by the instruments is relevant) we may impose (19), which will be used to obtain the value

\[ \alpha_3 = \left| (1 - \zeta^2 - \alpha_2^2)^{1/2} \right|, \quad (21) \]

where, without loss of generality, we may stick to positive values for \( \alpha_3 \) as long as \( \nu_i^{(3)} \) 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:

\[
\begin{align*}
\sigma_x^2 &= \xi^2 + \alpha_1^2 \\
\sigma_y^2 &= \xi^2 + 2\xi + 1 + \alpha_1^2 \\
\sigma_{xz} &= \xi \\
\rho_{xz} &= \xi / \sqrt{\xi^2 + \alpha_1^2} \\
\rho_{xx} &= \zeta \\
\rho_{xz} &= (\xi \zeta + \alpha_1 \alpha_2) / \sqrt{\xi^2 + \alpha_1^2}
\end{align*}
\] (22)

Note that these, after the normalizations $\beta = 1$, $\sigma_x = 1$ and $\sigma_z = 1$, depend on only 4 free parameters of the data generating process (DGP), viz. $\xi$, $\zeta$, $\alpha_1$ and $\alpha_2$. As we already established, the expressions for inconsistency in (12) and asymptotic variance (13) evaluated under $\mu_3 = 0$ and $\mu_4 = 3$ (the 3rd and 4th moment of $v_{i,1}$) depend on just four characteristics too, viz. on $\rho_{xz}$, $\rho_{ze}$, $\rho_{xz}$ 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

\[
\begin{align*}
\xi &= \rho_{ze}, \\
\xi &= \rho_{ze} \sigma_x, \\
\alpha_1 &= \sigma_x \left(1 - \rho_{xz}^2\right)^{1/2}, \\
\alpha_2 &= \left(\rho_{xz} - \rho_{ze} \rho_{ze}\right) / \left(1 - \rho_{xz}^2\right)^{1/2},
\end{align*}
\] (23)

from which $\alpha_3$ follows directly by evaluating (21). This reparametrization is useful, because the parameters $\rho_{xz}$, $\rho_{ze}$, $\rho_{xz}$ 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 AMAE($\hat{\beta}_{IV}$) can be calculated now for any set of compatible values of $n$, $\rho_{xz}$, $\rho_{ze}$, $\rho_{xz}$ and $\sigma_x$. Next 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 $v_i \sim \text{IIN}(0, I_3)$ and used 1,000,000 replications. From the

12
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 $\beta_{IV}$ close to normal (symmetric, unimodal, etc.); (b) the actual median of $\beta_{IV}$ close to $\beta_{IV}^*$; (c) the actual tail behavior of $\beta_{IV}$ reasonably well represented by that of the $N(\beta_{IV}^*, n^{-1}V_{IV}^N)$ distribution. Hence, we focus on the correspondence 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 do know that even when the instruments are valid, the asymptotic approximation will not capture the fat tail characteristic 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 in 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 $\rho_{xe}$ values. The panels in the left-hand columns have $\rho_{xe} = 0$, i.e. the instrument is valid and the standard asymptotic result applies. In the right-hand columns $\rho_{xe} = 0.2$, 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 $\rho_{xe} = -0.3$, $\rho_{xe} = 0$ (no simultaneity; hence, OLS would be more appropriate than IV), $\rho_{xe} = 0.3$ and $\rho_{xe} = 0.6$.

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 skew, and more so for serious simultaneity. In the right-hand column this occurs for different $\rho_{xe}$ 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 ($\rho_{xe} = 0$) and the new ($\rho_{xe} \neq 0$) 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 amy authors, see Sawa (1969) and (further remences in) Hillier (2005). 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 (2005) and Forchini (2005) show that bimodality of the consistent IV estimator occurs for much more severe simultaneity than examined here, viz. for $\rho_{xz} = 0.99$, whereas Phillips (2005) 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 may be able to produce inference on \( \beta \) by an inconsistent estimator, such as depicted in the right-hand column of Figure 1, which has a distribution that is much more concentrated around the true value than that from 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 \( \rho_{xz} \), for \( \rho_{xx} = 0, 0.1, 0.3 \) and 0.6. This accuracy is expressed as \( \log[\text{MAE(\hat{\beta}_{IV})}/ \text{AMAE(\hat{\beta}_{IV})}] \). 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 \( SN = \sigma_{\varepsilon}^2/\sigma_{\varepsilon}'^2 \). We find that the degree of simultaneity \( \rho_{xz} \) has little effect, and neither has the (in)validity of the instrument \( \rho_{xz} \). 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 \( \rho_{xz} \). 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 strongly invalid).

### 3.2 A simple overidentified model

The model of the above subsection can be extended such that we have two instruments \( z_{1i} \) and \( z_{i2} \), 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. We assume again 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 \( z_{i1} \) is possibly correlated with \( \varepsilon_i \), so that \( \zeta_2 = 0 \). This implies that

\[
\Sigma_{X'Z} \Sigma_{Z'Z}^{-1} \Sigma Z_{\varepsilon} = \Sigma_{Xz} \Sigma_{z1} = \rho_{xz1} \rho_{z1} \sigma_{x},
\]  

(24)
where, of course, \( \rho_{z1\varepsilon} = \zeta_1 \). Now the various entries in the formula of Theorem 1 specialize to

\[
\begin{align*}
\Sigma_{XX'} &= \sigma_x^2 > 0, \\
\Sigma_{XX'X} &= \sigma_x^2 = \rho_{xx}^2 \sigma_x^2 > 0, \\
\left( \beta_{GIV} \right)' &= \sigma_x^2 \Sigma_{X'X}^{-1} \Sigma_{XX'} \Sigma_{X'X}^{-1} \Sigma_{X'z'} = \frac{\sigma_{z\varepsilon}}{\sigma_x} = \frac{\rho_{xz1}\rho_{z1\varepsilon}}{\rho_{xx}^2 \sigma_x} = \frac{\rho_{xz1}\rho_{z1\varepsilon}}{\rho_{xx}^2 \sigma_x}, \\
c_1 &= \sigma_z^2 \Sigma_{zX}^{-1} \Sigma_{z'z} \Sigma_{z'z} = \rho_{z1\varepsilon}, \\
c_2 &= \sigma_z^2 \Sigma_{z'z} \Sigma_{z'z} \Sigma_{z'z} = \rho_{z1\varepsilon}, \\
c_3 &= \Sigma_{z'z} \beta_{GIV} = \frac{\rho_{xz1}\rho_{z1\varepsilon}}{\sigma_x} = \frac{\rho_{xz1}\rho_{z1\varepsilon}}{\sigma_x},
\end{align*}
\]

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: \( \sigma_x, \rho_{xx}, \rho_{xe}, \rho_{z1\varepsilon} \) and \( \rho_{xz1} \). However, in the special case \( \rho_{z1\varepsilon} = 0 \) the minimal set of parameters is just one dimensional, because \( \rho_{xx} \sigma_x \) suffices. For the general case we find

\[
V_{GIV}^N = \frac{\sigma_x^2}{\sigma_x \rho_{xx}^2} \left[ (1 - \rho_{z1\varepsilon}^4) + \rho_{z1\varepsilon} \gamma_1 + \rho_{z1\varepsilon} \rho_{xz1} \gamma_2 + 2 \rho_{z1\varepsilon}^4 \rho_{xz1} \gamma_3 \right] (28)
\]

where

\[
\begin{align*}
\gamma_1 &= \rho_{z1\varepsilon}^3 - 2 \rho_{xz1} (1 + \rho_{z1\varepsilon}^2 - 2 \rho_{z1\varepsilon}^4) - \rho_{xz1}^2 (\rho_{z1\varepsilon}^2 - 5 \rho_{z1\varepsilon}^3 + 2 \rho_{z1\varepsilon}^4), \\
\gamma_2 &= \rho_{z1\varepsilon}^4 + 2 \rho_{z1\varepsilon} (1 + 3 \rho_{z1\varepsilon} \rho_{xx} \rho_{xz1} - \rho_{xz1}^2 (1 - \rho_{z1\varepsilon})], \\
\gamma_3 &= 2 - (\rho_{xz1} - \rho_{z1\varepsilon} \rho_{xz1}) (3 \rho_{xz1} - \rho_{z1\varepsilon} \rho_{xz1}).
\end{align*}
\]

Note that this variance is invariant to sign changes of the correlations as long as the sign of \( \rho_{z1\varepsilon} \rho_{xz1} \rho_{z1\varepsilon} \) is not affected, or when either \( \rho_{xz1} \) or \( \rho_{z1\varepsilon} \) is zero. The sign of the inconsistency is determined by the sign of \( \rho_{z1\varepsilon} \rho_{xz1} \). For given values of \( \rho_{xx} \) and \( \rho_{z1\varepsilon} \) the magnitude of \( \beta_{GIV} \) is a multiple of \( \rho_{xz1} \), so it will be large when the invalid instrument is relatively strong. For the special case \( \rho_{xz1} = \rho_{xx} \), i.e. the second instrument is orthogonal to \( x \), the variance formula specializes to

\[
\frac{\sigma_x^2}{\sigma_x^2} (1 - \rho_{z1\varepsilon}^2) (\rho_{xz1} - \rho_{z1\varepsilon} \rho_{xx})^2 + \rho_{z1\varepsilon}^4 \rho_{z1\varepsilon}^2 (1 - \rho_{xx}^2)
\]

which, not surprisingly, corresponds to (13).

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 appropriate data processes. For that purpose the schemes (15) and (16) can be extended as follows. Let now \( v_i \sim \text{IID}(0, I_1) \). Again we take \( \varepsilon_i = v_{i,1} \) and, again restricting ourselves to positive \( \alpha_1 \) for symmetrically distributed \( v_i \), we have

\[
\begin{align*}
x_i &= \xi v_{i,1} + \alpha_1 v_{i,2} \\
&= \sigma_x [\rho_{xx} v_{i,1} + \frac{1}{\sqrt{1 - \rho_{xx}^2}} v_{i,2}],
\end{align*}
\]
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

\[ z_{i,1} = \zeta_1 v_{i,1} + \alpha_2 v_{i,2} + \alpha_3 v_{i,3} + \alpha_4 v_{i,4}, \]
\[ z_{i,2} = \zeta_2 v_{i,1} + \alpha_5 v_{i,2} + \alpha_6 v_{i,3} + \alpha_7 v_{i,4}. \]

These entail full generality, because they allow for both instruments any correlation with the disturbance $\varepsilon_i$, any correlation with the regressor $x_i$ and any mutual correlation. Since it is only the space spanned by these two instruments that matters for $\hat{\beta}_{GIV}$, we may replace $z_{i,2}$ by a linear combination of $z_{i,1}$ and $z_{i,2}$ such that it no longer depends on $v_{i,1}$. This corresponds to taking $\zeta_2 = 0$ and re-interpreting $\alpha_5$, $\alpha_6$ and $\alpha_7$. Hence, the general case of two possibly invalid instruments can be represented fully by that of one valid and one possibly invalid instrument, as we already argued above from the asymptotic perspective. We can perform a similar operation again, now 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

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

(29)

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

\[ 0 \leq \alpha_2 \leq 1, \quad 0 \leq \alpha_5 \leq 1. \]  

(30)

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

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

(31)

So, for given values of $\sigma_x$, $\rho_{xz}$ and $\rho_{z\varepsilon} = \zeta_1$, we would be able to generate data according to (28) and (29) 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

\[ R^2_{xz} = \frac{\Sigma_{x'z} \Sigma_{z'z}^{-1} \Sigma_{z'x}}{\sigma_z^2}. \]  

(32)

Note that

\[ \rho_{xx}^2 = \frac{(\Sigma_{x'z} \Sigma_{z'z}^{-1} \Sigma_{z'x})^2}{\sigma_z^2 \sigma_z^2} = \frac{\sigma_x^2}{\sigma_z^2} = R^2_{xz}, \]  

and, since we imposed $\Sigma_{z'z} = I$, we have

\[ \rho_{xx}^2 = \frac{\Sigma_{x'z} \Sigma_{z'x}}{\sigma_z^2} = \rho_{xx}^2 + (1 - \rho_{xx}^2)^{1/2} \alpha_5^2, \]
\[ \rho_{xz} = \rho_{xx} \rho_{z\varepsilon} + (1 - \rho_{xx}^2)^{1/2} \alpha_2. \]

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

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

(34)
and

$$\alpha_2 = \frac{\rho_{xz1} - \rho_{xe}\rho_{z1\varepsilon}}{(1 - \rho_{xe}^2)^{1/2}},$$

(35)

from which $\alpha_6$ follows directly by evaluating (31).

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 $\sigma_x$, $\rho_{zx}, \rho_{xe}, \rho_{xz1}$ and $\rho_{z1\varepsilon}$. Here again, these are found to be those data moments that characterize the asymptotic distribution. They determine $\xi$ and $\alpha_1$ via (28) and $\alpha_2, \alpha_5$ and $\alpha_6$ via (35), (34) and (31), respectively. We may restrict ourselves to cases where $\rho_{zx} > 0$ (since the coefficients of the simulation design are just determined by $\rho_{zx}^2$).

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 $\rho_{z1\varepsilon}$ and $\rho_{xz1}$ are changed in sign. Therefore, we will only examine cases with $\rho_{xz1} \geq 0$. However, we shall also examine only nonnegative values of $\rho_{z1\varepsilon}$, because changing the signs of both $\rho_{z1\varepsilon}$ and $\rho_{xe}$ yields the mirror image of the distribution of $\hat{\beta}_{GIV}$ when the distribution of $\varepsilon$ is symmetric. In line with the just identified model the distribution of $\hat{\beta}_{GIV} - \beta$ for $\rho_{z1\varepsilon} \leq 0$ is equivalent with the distribution of $-(\hat{\beta}_{GIV} - \beta)$ for $-\rho_{z1\varepsilon} \geq 0$ and $-\rho_{xe}$, because: If we change the signs of $\rho_{z1\varepsilon}, \rho_{xe}$ 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 $\zeta_1 = \rho_{z1\varepsilon} = 0$ in (29) and 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

$$z_{i,1} = \alpha_2v_{i,2} + (1 - \alpha_2^2)^{1/2}v_{i,3},$$

$$z_{i,2} = v_{i,4}. $$

(36)

These imply $\rho_{xz1} = \rho_{zx}$ and $\alpha_2 = |\rho_{zx}(1 - \rho_{xe}^2)^{-1/2}|$ instead of (35). Hence, when $\zeta = 0$ the finite sample distribution is determined by just 3 parameters (viz. $\sigma_x, \rho_{xe}$ and $\rho_{zx}$) instead of 5 (apart from $n$), whereas the limiting distribution just depends on $\sigma_x^2 = \rho_{zx}^2\sigma_x^2$.

In all calculations we fixed again $\sigma_x^2/\sigma_\varepsilon^2 = 10$ (which here too has only a multiplicative effect, i.e. just affects the scale of the densities), as before we chose values $\rho_{xe} = \{-0.3, 0.0, 0.3, 0.6\}$, $\rho_{z1\varepsilon} = \zeta = \{0.0, 0.2\}$ and $\rho_{zx} = \{0.8, 0.3, 0.1, 0.01\}$, whereas $\rho_{xz1} = \{\rho_{zx}, \rho_{zx}/2, \rho_{zx}/8\}$. The latter values are associated with decreasing relative strongness of $z_1$ (and, complementary, a valid instrument $z_2$ that 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 $\rho_{xz1} < \rho_{zx}$, 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 $\rho_{xz1} = \rho_{zx}$ (and $\rho_{xz2} = 0$). Although we already established that these yield similar asymptotic results as the $k = l = 1$ case, from the simulations we found that in this situation the finite sample densities do differ slightly from the "no finite moments" case, the more so for a weaker instrument $z_1$, especially when $\rho_{xz1} = \rho_{zx} = 0.01$. When both instruments are valid and $\rho_{zx} = 0.01$ 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 $\rho_{x_1x} = 0.01$, $\rho_{x_2x} = 0$ and $\rho_{z_1x} = 0.2$ we find that the bimodality of the $GIV$ estimator is less pronounced than for the $IV$ estimator.

Figure 7 presents densities for $\rho_{x_1x} = 0.8$ 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 $\rho_{x_1x} = 0.01$. In the latter case we note again 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 (2005) suspects bimodality in the overidentified model when the instruments are valid but weak, we do not find it at $\rho_{x_1x} = 0.01$.

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 $\rho_{x_1x}$ 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 more efficient inference 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 $\rho_{x_1}^2 \geq 0.1$. 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 pseuso-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. In the simple illustrative models that 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. They are also directly applicable in case non-stationary series are involved, 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.

We examined the accuracy of our analytic large sample results in small samples by simulating a simple just identified and a simple overidentified model and establishing the actual behavior of instrumental variable estimators. 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 estimators 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 IV estimators when some of the instruments 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 alternative approximation that can cope with weakness of instruments. However, our illustrations suggest that it seems more worthwhile to abandon the employment of weak instruments altogether and just stick to strong instruments, even if they are invalid. For that situation we at least seem to have obtained here a reasonably accurate approximation to its finite sample distribution, whereas at the same time this finite sample distribution is such that it may yield much more accurate inference than that obtained on the basis of weak instruments.

References


and Data Analysis 49, 417-444.


A Proof of Theorem 1

Because the estimator \( \hat{\beta}_{GIV} \) tends for an increasing sample size not to \( \beta \), but to \( \beta^*_GIV \), 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

\[
\sqrt{n}(\hat{\beta}_{GIV} - \beta^*_GIV) = \sqrt{n}[(\hat{X}'\hat{X})^{-1}\hat{X}'e - \sigma^2_e\Sigma^{-1}_{X'X}\Sigma_{X'Z}\Sigma^{-1}_{Z'Z}e]. \tag{37}
\]

To obtain the limiting distribution we shall rewrite the right-hand side of (37) such that we can invoke the Lemma given at the end of Section 2. Below we first show that (37) can be rewritten as

\[
\sqrt{n}(\hat{\beta}_{GIV} - \beta^*_GIV) = \left( \frac{1}{n} \hat{X}'\hat{X} \right)^{-1} \frac{1}{\sqrt{n}} \left[ W'e + \omega(e' - na_e^2) \right] + o_p(1), \tag{38}
\]

for appropriate \( n \times k \) matrix \( W \), with \( E(W'e) = 0 \), and nonrandom \( k \times 1 \) vector \( \omega \). Next, invoking also a theorem often attributed to Cramér, the lemma yields

\[
\sqrt{n}(\hat{\beta}_{GIV} - \beta^*_GIV) \rightarrow N \left[ 0, \sigma^2_e\Sigma^{-1}_{X'X} \lim_{n \rightarrow \infty} \frac{1}{n} W'W + 2\sigma^2_e\omega\omega' \right] \Sigma^{-1}_{X'X} \tag{39}
\]

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

We first set out to rewrite (37) in the form (38). Using \( \hat{\Pi} \equiv (Z'Z)^{-1}Z'X \) we get

\[
\sqrt{n}(\hat{\beta}_{GIV} - \beta^*_GIV) = \sqrt{n}[(\hat{X}'\hat{X})^{-1} \left( \hat{\Pi}'[Z'e - E(Z'e)] + \hat{\Pi}'E(Z'e) - \sigma^2_e\hat{X}'\hat{X} \Sigma^{-1}_{X'X} \Sigma_{X'Z} \Sigma^{-1}_{Z'Z} \zeta \right)] \tag{40}
\]

\[
= \frac{1}{n} \left( \frac{1}{n} \hat{X}'\hat{X} \right)^{-1} \frac{1}{\sqrt{n}} \left[ \hat{\Pi}'[Z'e - E(Z'e)] + na_e^2 \hat{\Pi}' - \frac{1}{n} \hat{X}'\hat{X} \Sigma^{-1}_{X'X} \Sigma_{X'Z} \Sigma^{-1}_{Z'Z} \zeta \right].
\]
For the second expression between square brackets in the final line of (40) we find

\[ \hat{\Pi}' = \frac{1}{n} \hat{X}'\hat{X}\Sigma_{X'}^{-1}\Sigma_{X'}Z \Sigma_{Z'}^{-1} \]

\[ = \hat{\Pi}'(\Sigma_{Z'}^{-1} - \frac{1}{n}Z'Z)\Sigma_{Z'}^{-1} + (\frac{1}{n}X'Z - \frac{1}{n}X'\hat{X}\Sigma_{X'}^{-1}\Sigma_{X'}Z)\Sigma_{Z'}^{-1} \]

\[ = \hat{\Pi}'(\Sigma_{Z'}^{-1} - \frac{1}{n}Z'Z)\Sigma_{Z'}^{-1} + [(\frac{1}{n}X'Z - \Sigma_{X'}Z) + (\Sigma_{X'}\hat{X} - \frac{1}{n}X'\hat{X})\Sigma_{X'}^{-1}\Sigma_{X'}Z] \Sigma_{Z'}^{-1}, \]

and this contains a factor which can be rewritten as

\[ \Sigma_{X'}\hat{X} - \frac{1}{n}X'\hat{X} \]

\[ = \Sigma_{X'}Z_{Z'}^{-1} \Sigma_{Z'}X - \hat{\Pi}' \frac{1}{n}Z'X \]

\[ = (\Sigma_{X'}Z_{Z'}^{-1} - \hat{\Pi}') \Sigma_{Z'}X - \hat{\Pi}'(\frac{1}{n}Z'X - \Sigma_{Z'}X) \]

\[ = ([\Sigma_{X'}X - \frac{1}{n}X'X]Z_{Z'}^{-1} - \frac{1}{n}X'X(\frac{1}{n}X'Z - \Sigma_{X'}Z)] \Sigma_{Z'}X - \hat{\Pi}'(\frac{1}{n}Z'X - \Sigma_{Z'}X) \]

\[ = ([\Sigma_{X'}X - \frac{1}{n}X'X]Z_{Z'}^{-1} - \hat{\Pi}'([\Sigma_{Z'}X - \frac{1}{n}Z'Z]Z_{Z'}^{-1}]) \Sigma_{Z'}X - \hat{\Pi}'(\frac{1}{n}Z'X - \Sigma_{Z'}X). \]

Now substituting the decompositions obtained in (41) and (42) into the expression within curly brackets in the final line of (40) we obtain

\[ \hat{\Pi}'[Z'\varepsilon - E(Z'\varepsilon)] + n\sigma_{\varepsilon}^2[\hat{\Pi}' - \frac{1}{n}X'\hat{X}\Sigma_{X'}^{-1}\Sigma_{X'}Z \Sigma_{Z'}^{-1}]\zeta \]

\[ = \hat{\Pi}'[Z'\varepsilon - E(Z'\varepsilon)] + n\sigma_{\varepsilon}^2[\hat{\Pi}'(\Sigma_{Z'}^{-1} - \frac{1}{n}Z'Z)\Sigma_{Z'}^{-1}]\zeta \]

\[ + n\sigma_{\varepsilon}^2(\frac{1}{n}X'Z - \Sigma_{X'}Z) \Sigma_{Z'}^{-1}\zeta + n\sigma_{\varepsilon}^2(\Sigma_{X'}\hat{X} - \frac{1}{n}X'\hat{X})\Sigma_{X'}^{-1}\Sigma_{X'}Z \Sigma_{Z'}^{-1}\zeta \]

\[ = \hat{\Pi}'[Z'\varepsilon - E(Z'\varepsilon)] - n\sigma_{\varepsilon}^2\hat{\Pi}'(\frac{1}{n}X'Z - \Sigma_{X'}Z) \Sigma_{Z'}^{-1}\zeta + n\sigma_{\varepsilon}^2(\frac{1}{n}X'Z - \Sigma_{X'}Z) \Sigma_{Z'}^{-1}\zeta \]

\[ + n\sigma_{\varepsilon}^2 \left[ (\Sigma_{X'}X - \frac{1}{n}X'X) \Pi - \hat{\Pi}'(\Sigma_{Z'}X - \frac{1}{n}Z'Z) \Sigma_{Z'}^{-1}\Sigma_{Z'}X - \hat{\Pi}'(\frac{1}{n}Z'X - \Sigma_{Z'}X) \right] \bar{\beta}_{GIV}, \]

where we substituted \( \Pi \equiv \Sigma_{Z'}^{-1}\Sigma_{Z'}X \) and \( \bar{\beta}_{GIV} \equiv \sigma_{\varepsilon}^2\Sigma_{X'}^{-1}\Sigma_{X'}Z \Sigma_{Z'}^{-1}\zeta. \) Now exploiting item (vi) of Framework A, we can employ

\[ Z'\varepsilon - E(Z'\varepsilon) = Z'\varepsilon + (\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\zeta, \]

\[ \frac{1}{n}Z'Z - \Sigma_{Z'}X = \frac{1}{n}Z'Z - \frac{1}{n}E(Z'Z | \bar{Z}) + \frac{1}{n}E(Z'Z | \bar{Z}) - \Sigma_{Z'}Z \]

\[ = \frac{1}{n}Z'\zeta' + \frac{1}{n}\zeta'\bar{Z} + \frac{1}{n}(\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\zeta' + o_p(n^{-1/2}), \]

\[ \frac{1}{n}X'Z - \Sigma_{X'}Z = \frac{1}{n}X'Z - \frac{1}{n}E(X'Z | \bar{X}, \bar{Z}) + \frac{1}{n}E(X'Z | \bar{X}, \bar{Z}) - \Sigma_{X'}Z \]

\[ = \frac{1}{n}X'\zeta' + \frac{1}{n}\zeta'\bar{Z} + \frac{1}{n}(\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\zeta' + o_p(n^{-1/2}), \]

so that the final expression given for (43) can be written as

\[ \hat{\Pi}'[Z'\varepsilon + (\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\hat{\Pi}' - \sigma_{\varepsilon}^2\hat{\Pi}'\zeta'\zeta^{-1}\zeta - \sigma_{\varepsilon}^2\hat{\Pi}'\zeta'Z\Sigma_{Z'}^{-1}\zeta - \sigma_{\varepsilon}^2\hat{\Pi}'(\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\zeta'\zeta^{-1}\zeta \]

\[ + \sigma_{\varepsilon}^2X'\zeta'\zeta^{-1}\zeta \Sigma_{Z'}^{-1}\zeta + \sigma_{\varepsilon}^2X'\zeta'\zeta^{-1}\zeta \Sigma_{Z'}^{-1}\zeta + \sigma_{\varepsilon}^2(\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\zeta'\zeta^{-1}\zeta \Sigma_{Z'}^{-1}\zeta + \hat{\Pi}'Z'\varepsilon\Pi\beta_{GIV} + \hat{\Pi}'\zeta'\zeta\Pi\beta_{GIV} \]

\[ + \hat{\Pi}'(\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\zeta'\zeta\Pi\beta_{GIV} - \varepsilon'\zeta\Pi\beta_{GIV} - (\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\zeta'\zeta\Pi\beta_{GIV} - \Pi'\varepsilon'X\beta_{GIV} - \Pi'(\varepsilon'\varepsilon - n\sigma_{\varepsilon}^2)\zeta'\zeta\beta_{GIV} + o_p(n^{1/2}). \]
This can be simplified further by using
\[ c_1 \equiv \sigma_z^2 \zeta \Sigma_{Z'Z}', \]
\[ c_2 \equiv \zeta' \Pi \beta GIV = \sigma_z^2 \zeta' \Sigma_{Z'Z}' \Sigma_{X'X}^{-1} \Sigma_{X'Z} \Sigma_{Z'Z}', \]
\[ c_3 \equiv \zeta' \beta GIV, \]
giving
\[
\hat{\Pi} Z' \varepsilon + (\varepsilon' - n \sigma_z^2) \hat{\Pi} \zeta - c_1 \hat{\Pi} Z' \varepsilon - \sigma_z^2 \hat{\Pi} \zeta' \Sigma_{Z'Z}' Z' \varepsilon - c_1(\varepsilon' - n \sigma_z^2) \hat{\Pi} \zeta \\
+ c_1 \hat{X}' \varepsilon + \sigma_z^2 \hat{\Pi} \zeta' \Sigma_{Z'Z}' Z' \varepsilon + c_2(\varepsilon' - n \sigma_z^2) \hat{\Pi} \zeta + c_3 \hat{\Pi}' \zeta' \beta GIV \Sigma_{X'Z} Z' \varepsilon + c_2(\varepsilon' - n \sigma_z^2) \hat{\Pi} \zeta' \\
- c_2 \hat{X}' \varepsilon - \zeta' \beta GIV \Pi Z' \varepsilon - c_2(\varepsilon' - n \sigma_z^2) \hat{\Pi} \zeta' - c_3 \hat{\Pi} Z' \varepsilon - c_3(\varepsilon' - n \sigma_z^2) \hat{\Pi} \zeta' + o_p(1) \\
= [c_4 I_k - \Pi' \zeta' \beta GIV] X' \varepsilon + [c_5 \hat{\Pi}' + (\xi - \Pi' \zeta)(\sigma_z^2 \zeta' - \beta GIV \Sigma_{X'Z}) \Sigma_{Z'Z}^{-1}] Z' \varepsilon \\
+ (c_4 \xi + c_5 \hat{\Pi} \zeta)(\varepsilon' - n \sigma_z^2) + o_p(1),
\]
where \( c_4 \equiv c_1 - c_3 \) and \( c_5 \equiv 1 - c_1 - c_4 \).

Note that (44) is equal to the factor in curly brackets in the final line of (40), and we want to derive its limiting distribution after scaling by the factor \( 1/\sqrt{n} \), so we may neglect the remainder term. Cramèr’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
\[
\Theta_1 \equiv c_4 I_k - \Pi' \zeta' \beta GIV, \\
\Theta_2 \equiv c_5 \Pi' + (\xi - \Pi' \zeta)(\sigma_z^2 \zeta' - \beta GIV \Sigma_{X'Z}) \Sigma_{Z'Z}^{-1}, \\
\omega \equiv c_4 \xi + c_5 \Pi' \zeta,
\]
we can invoke the Lemma now with
\[
W' = \Theta_1 X' + \Theta_2 Z'.
\]
For the case \( \mu_3 = 0 \) and \( \mu_4 = 3 \) we then obtain the limiting distribution
\[
\frac{1}{\sqrt{n}} [\Theta_1 X' \varepsilon + \Theta_2 Z' \varepsilon + \omega(\varepsilon' - n \sigma_z^2)] \rightarrow N(0, \sigma_z^2 V_0),
\]
where
\[
V_0 = \Theta_1 \Sigma_{X'X} \Theta_1' + \Theta_2 \Sigma_{Z'Z} \Theta_2' + 2 \sigma_z^2 \omega \omega' + \Theta_1 \Sigma_{X'Z} \Theta_2' + \Theta_2 \Sigma_{Z'X} \Theta_1'.
\]
In evaluating \( V_0 \) we can make use of
\[
\Sigma_{Z'Z} = \text{plim} n^{-1} Z' Z = \Sigma_{Z'Z} - \sigma_z^2 \zeta' \\
\Sigma_{X'X} = \text{plim} n^{-1} X' X = \Sigma_{X'X} - \sigma_z^2 \zeta' \\
\Sigma_{Z'X} = \text{plim} n^{-1} Z' X = \Sigma_{Z'X} - \sigma_z^2 \zeta' \\
\Sigma_{X'Z} = \text{plim} n^{-1} X' Z = \Sigma_{X'Z}
\]
and find that \( V_0 \) can be expressed as
\[
[c_4 I_k - \Pi' \zeta' \beta GIV](\Sigma_{X'X} - \sigma_z^2 \xi \xi')[c_4 I_k - \beta GIV \zeta' \Pi] \\
+ [c_5 \Pi' + (\xi - \Pi' \zeta)(\sigma_z^2 \zeta' - \beta GIV \Sigma_{X'Z}) \Sigma_{Z'Z}^{-1}][\Sigma_{Z'Z} - \sigma_z^2 \zeta \zeta'][c_5 \Pi + \Sigma_{Z'Z}^{-1} \Sigma_{X'Z} \beta GIV (\xi' - \zeta' \Pi)] \\
+ 2 \sigma_z^2 [c_4 \xi + c_5 \Pi' \zeta][c_4 \xi' + c_5 \Pi' \zeta] \\
+ [c_4 I_k - \Pi' \beta GIV] (\Sigma_{X'Z} - \sigma_z^2 \xi \xi')[c_5 \Pi + \Sigma_{Z'Z}^{-1} \Sigma_{X'Z} \beta GIV] (\xi' - \zeta' \Pi) \\
+ [c_5 \Pi' + (\xi - \Pi' \zeta)(\sigma_z^2 \zeta' - \beta GIV \Sigma_{X'Z}) \Sigma_{Z'Z}^{-1}][\Sigma_{Z'Z} - \sigma_z^2 \zeta \zeta'][c_4 I_k - \beta GIV \zeta' \Pi].
\]
Next we examine these 5 terms of \( V_0 \) one by one. The first one is
\[
\sigma_z^2 (\Sigma_{X'X} - \sigma_z^2 \xi \xi') - c_4 (\Sigma_{X'X} - \sigma_z^2 \xi \xi') \beta GIV \zeta' \Pi \\
- c_2 \Pi' \zeta' \beta GIV (\Sigma_{X'X} - \sigma_z^2 \xi \xi') + \Pi' \beta GIV (\Sigma_{X'X} - \sigma_z^2 \xi \xi') \beta GIV \zeta' \Pi \\
= \sigma_z^2 (\Sigma_{X'X} - \sigma_z^2 \xi \xi') - c_2 \Sigma_{X'X} \beta GIV \zeta' \Pi + \sigma_z^2 c_4 \xi \xi' \\
- c_2 \Pi' \zeta' \beta GIV (\Sigma_{X'X} + \sigma_z^2 c_4 \Pi' \zeta' \Pi + (\beta GIV \Sigma_{X'X} \beta GIV - \sigma_z^2 c_3) \Pi' \zeta' \Pi, \\
22
the second one is

\[c_3 \Pi' \Sigma_{Z;Z} + (\xi - \Pi' \zeta' (\sigma_2^2 \zeta' - \tilde{\beta}_{GIV} \Sigma_{Z;Z})) [c_3 \Pi + \Sigma_{Z;Z}^{-1} (\sigma_2^2 \zeta - \Sigma_{Z;Z} \tilde{\beta}_{GIV}) (\zeta' - \zeta' \Pi)]
\]

\[-\sigma_2^2 (c_3 \Pi' \zeta' (\sigma_2^2 \zeta' - \tilde{\beta}_{GIV} \Sigma_{Z;Z}) \Sigma_{Z;Z}^{-1} (\zeta') [c_3 \Pi + \Sigma_{Z;Z}^{-1} (\sigma_2^2 \zeta - \Sigma_{Z;Z} \tilde{\beta}_{GIV}) (\zeta' - \zeta' \Pi)]
\]

\[= c_2^2 \Pi' \Sigma_{Z;Z} \Pi + c_3 (\xi - \Pi' \zeta' (\sigma_2^2 \zeta' - \tilde{\beta}_{GIV} \Sigma_{Z;Z}) \Pi) + (\sigma_2^2 c_3 \Pi' \zeta - c_3 \Pi' \Sigma_{Z;Z} \tilde{\beta}_{GIV}) (\zeta' - \zeta' \Pi) + (\xi - \Pi' \zeta') (\sigma_2^2 \zeta' \Pi) (\sigma_2^2 \zeta' - \Sigma_{Z;Z} \tilde{\beta}_{GIV} (\zeta' - \zeta' \Pi))
\]

\[-\sigma_2^2 (c_3 \Pi' \zeta' (\sigma_2^2 \zeta' - \tilde{\beta}_{GIV} \Sigma_{Z;Z}) \Sigma_{Z;Z}^{-1} (\zeta') [c_3 \Pi + \Sigma_{Z;Z}^{-1} (\sigma_2^2 \zeta - \Sigma_{Z;Z} \tilde{\beta}_{GIV}) (\zeta' - \zeta' \Pi)]
\]

the third one is

\[2c_2^2 \sigma_2^2 \epsilon_0 \epsilon' + 2c_2^2 \sigma_2^2 \epsilon_0 \epsilon' + 2c_2^2 \sigma_2^2 \epsilon_0 \epsilon' + 2c_2^2 \sigma_2^2 \epsilon_0 \epsilon'
\]

the fourth one is

\[c_3 \Sigma_{X;Z} - \Pi' \tilde{\beta}_{GIV} \Sigma_{X;Z} [c_3 \Pi + \Sigma_{X;Z}^{-1} (\sigma_2^2 \zeta - \Sigma_{X;Z} \tilde{\beta}_{GIV}) (\zeta' - \zeta' \Pi)]
\]

\[-\sigma_2^2 (c_3 \Pi' \zeta' (\sigma_2^2 \zeta' - \tilde{\beta}_{GIV} \Sigma_{X;Z}) \Sigma_{X;Z}^{-1} (\zeta') [c_3 \Pi + \Sigma_{X;Z}^{-1} (\sigma_2^2 \zeta - \Sigma_{X;Z} \tilde{\beta}_{GIV}) (\zeta' - \zeta' \Pi)]
\]

\[= c_3 \Pi' \Sigma_{X;Z} \Pi + c_3 (\xi - \Pi' \zeta' (\sigma_2^2 \zeta' - \tilde{\beta}_{GIV} \Sigma_{X;Z}) \Pi) + (\sigma_2^2 c_3 \Pi' \zeta - c_3 \Pi' \Sigma_{X;Z} \tilde{\beta}_{GIV} \pi) (\zeta' - \zeta' \Pi) + (\xi - \Pi' \zeta') (\sigma_2^2 \zeta' \Pi) (\sigma_2^2 \zeta' - \Sigma_{X;Z} \tilde{\beta}_{GIV} (\zeta' - \zeta' \Pi))
\]

\[-\sigma_2^2 (c_3 \Pi' \zeta' (\sigma_2^2 \zeta' - \tilde{\beta}_{GIV} \Sigma_{X;Z}) \Sigma_{X;Z}^{-1} (\zeta') [c_3 \Pi + \Sigma_{X;Z}^{-1} (\sigma_2^2 \zeta - \Sigma_{X;Z} \tilde{\beta}_{GIV}) (\zeta' - \zeta' \Pi)]
\]
$$V_0 = c_4 c_5 \Sigma_{X'X} - c_2^2 c_4^2 \zeta' + c_4 \Sigma_{X'X} \beta_{GVG} \zeta' \Pi - c_2 c_4 \Sigma_{X'X} \beta_{GVG} \zeta',$$

$$= c_4 c_5 \Sigma_{X'X} - c_2^2 c_4^2 \zeta' + c_4 \Sigma_{X'X} \beta_{GVG} \zeta' \Pi - c_2 c_4 \Sigma_{X'X} \beta_{GVG} \zeta',$$

$$= c_4 c_5 \Sigma_{X'X} - c_2^2 c_4^2 \zeta' + c_4 \Sigma_{X'X} \beta_{GVG} \zeta' \Pi - c_2 c_4 \Sigma_{X'X} \beta_{GVG} \zeta',$$

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

$$V_0 = c_4^2 \Sigma_{X'X} - c_2^2 c_4^2 \zeta' + c_4 \Sigma_{X'X} \beta_{GVG} \zeta' \Pi - c_2 c_4 \Sigma_{X'X} \beta_{GVG} \zeta',$$

$$= c_4^2 \Sigma_{X'X} - c_2^2 c_4^2 \zeta' + c_4 \Sigma_{X'X} \beta_{GVG} \zeta' \Pi - c_2 c_4 \Sigma_{X'X} \beta_{GVG} \zeta',$$

Thus, the asymptotic variance of the (generalized) GIV estimator (39) is

$$V_{GIV}^N = \sigma_e^2 \beta_{GVG} \Sigma_{X'X} \beta_{GVG},$$

which gives the result of Theorem 1.

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

$$V_{IV}^N = \sigma_e^2 \beta_{GVG} \Sigma_{X'X} \beta_{GVG},$$

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 $\sigma_e^2 \omega'$. In obtaining the third term of (47) we already found

$$\omega' = c_4^2 \xi' + c_4 c_5 \Pi' \zeta' \Pi + c_4 c_5 \xi' \Pi' \zeta' \Pi,$$
\[ \sigma^4_{X'X} \omega \omega' \Sigma^1_{X'X} \]
\[ = \sigma^4 c_4 \Sigma^1_{X'X} \zeta \zeta' \Sigma^1_{X'X} + \sigma^2 c_4 \Sigma^1_{X'X} \Pi' \zeta' \Sigma^1_{X'X} + \sigma^2 c_4 \Sigma^1_{X'X} \zeta' \Pi' \zeta' \Sigma^1_{X'X} + \sigma^2 c_4 \Sigma^1_{X'X} \Pi' \zeta' \Pi' \Sigma^1_{X'X} \]
\[ = \sigma^4 c_4 \Sigma^1_{X'X} \zeta \zeta' \Sigma^1_{X'X} + \sigma^2 c_4 \Sigma^1_{X'X} \beta_{GIV} \zeta \Sigma^1_{X'X} + \sigma^2 c_4 \Sigma^1_{X'X} \zeta \beta_{GIV} + \sigma^2 c_4 \Sigma^1_{X'X} \beta_{GIV} \beta_{GIV}. \]

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
\[ \mu_3 \sigma^3_{X'X} \Sigma^1_{W'W} \omega' \omega. \]

Since
\[ W' = \Theta_1 X' + \Theta_2 Z' \]
\[ c_4 X' - \Pi' \zeta' \Sigma^1_{X'X} + c_5 \Pi' \zeta' \Sigma^1_{X'X} \]
we find
\[ \Sigma_{W'W} = c_4 \Sigma_{X'X} - \Pi' \zeta' \beta_{GIV} \Sigma_{X'X} + c_3 \Pi' \Sigma_{Z'Z} + (\xi - \Pi' \zeta' \beta_{GIV} \Sigma_{X'X}) \Sigma^1_{Z'Z} \Sigma_{Z'Z}, \]
and with \( \omega' = c_4 \zeta' + c_3 \zeta' \Pi \), we obtain
\[ \Sigma_{W'W} \]
\[ = c_4 [c_4 \Sigma_{X'X} \zeta' - \Pi' \zeta' \beta_{GIV} \Sigma_{X'X} + c_3 \Pi' \Sigma_{Z'Z} + (\xi - \Pi' \zeta' \beta_{GIV} \Sigma_{X'X}) \Sigma^1_{Z'Z} \Sigma_{Z'Z}]. \]

Thus
\[ \sigma^3_{X'X} \omega \omega' \Sigma^1_{X'X} \]
\[ = \sigma^3 c_4 [c_4 \Sigma^1_{X'X} \Sigma_{X'X} \zeta \zeta' \Sigma^1_{X'X} - \Sigma^1_{X'X} \Pi' \zeta' \beta_{GIV} \Sigma_{X'X} + c_3 \Sigma^1_{X'X} \Pi' \Sigma_{Z'Z} + (\xi - \Pi' \zeta' \beta_{GIV} \Sigma_{X'X}) \Sigma^1_{Z'Z} \Sigma_{Z'Z}]. \]

and the additional term is then equal to \( \mu_3 \) multiplied by
\[ \sigma^3 \Sigma^1_{X'X} [\Sigma_{W'W} \omega' + \omega \Sigma_{W'W}] \Sigma^1_{X'X} \]
\[ = c_4 [\sigma^3 c_4 \Sigma^1_{X'X} \Sigma_{X'X} \zeta \zeta' \Sigma^1_{X'X} - \sigma^3 \beta_{GIV} \Sigma_{X'X} \zeta \zeta' \Sigma^1_{X'X} + \sigma^3 \Sigma^1_{X'X} \Pi' \Sigma_{Z'Z} + (\xi - \Pi' \zeta' \beta_{GIV} \Sigma_{X'X}) \Sigma^1_{Z'Z} \Sigma_{Z'Z}]. \]
This expression can be simplified slightly when we assume that both matrices $X$ and $Z$ have a first column of ones. Then

$$
\Sigma_{X'}^{-1} \Sigma_{X'k} = e_{k,1},
$$

$$
\Pi' \Sigma_{Z'} = \Sigma_{X'Z} \Sigma_{Z'}^{-1} \Sigma_{Z'k} = \Sigma_{X'Z} e_{1,1},
$$

where $e_{1,1}$ denotes a $f \times 1$ unit vector, which has all elements equal to zero, apart from a unit element in position $g$. This yields

$$
\begin{align*}
\sigma^2 \xi_{X'k}^{-1} \{ \Sigma_{W'} \omega' + \omega \Sigma_{W'} \Sigma_{X'X}^{-1} \\
= c_4 \{ \sigma^2 c_4 e_{k,1} \xi_{X'k}^{-1} - \sigma_\xi \beta_{GIV} \Sigma_{X'} \xi_{X'k}^{-1} + \sigma^2 \xi_{X'k}^{-1} \Sigma_{X'Z} e_{1,1} \xi_{X'k}^{-1} \\
+ (\sigma^2 \xi_{X'k}^{-1} \xi - \sigma_\xi \beta_{GIV}) (\sigma^2 \xi_{X'k}^{-1} - \beta_{GIV} \Sigma_{X'Z} e_{1,1} \xi_{X'k}^{-1}) \} \\
+ c_5 [\sigma c_4 e_{k,1} \beta_{GIV} - \sigma_\xi \beta_{GIV} \Sigma_{X'} \beta_{GIV} + \sigma c_5 \Sigma_{X'X} \xi_{X'k}^{-1} + \Sigma_{X'X} \xi_{X'k}^{-1} ] \\
+ \Sigma_{X'X} \xi_{X'k}^{-1} \xi_{X'k}^{-1} (\sigma^2 \xi_{X'k}^{-1} - \Sigma_{X'X} \beta_{GIV}) (\sigma^2 \xi_{X'k}^{-1} - \sigma_\xi \beta_{GIV}) \\
+ c_6 [\sigma c_4 \beta_{GIV} e_{k,1} - \sigma_\xi \beta_{GIV} \Sigma_{X'} \beta_{GIV} + \sigma c_5 \beta_{GIV} e_{1,1} \Sigma_{X'X}^{-1} ] \\
+ \beta_{GIV} e_{1,1} (\sigma^2 \xi_{X'k}^{-1} - \Sigma_{X'X} \beta_{GIV}) (\sigma_\xi \xi_{X'k}^{-1} - \sigma_\xi \beta_{GIV}).
\end{align*}
$$

When $k = l$, i.e. $c_1 = c_2$, $c_5 = 1 - c_4$, $\Pi \Sigma_{X'X}^{-1} = \Sigma_{X'Z}$ and $\sigma^2 \xi - \Sigma_{X'X} \beta_{IV} = 0$, $V_{GIV}^{NN}$ specializes to the expression

$$
V_{IV}^{NN} = \sigma^2 \xi_{X'k}^{-1} \Sigma_{X'Z} \Sigma_{X'Z}^{-1} - \{ \sigma^2 \xi_{X'k}^{-1} - \sigma_\xi \beta_{IV} \Sigma_{X'X} \beta_{IV} \} \beta_{IV} \beta_{IV} \\
+ (\mu_4 - 1) \sigma^2 \xi_{X'k}^{-1} - 2 \sigma_\xi \beta_{IV} \Sigma_{X'X} \beta_{IV} + \mu_3 \sigma_\xi \xi_{X'k}^{-1} [\Sigma_{X'X} \beta_{IV} + \beta_{IV} \Sigma_{Z'} \Sigma_{X'X}^{-1}] \\
= \sigma^2 \xi_{X'k}^{-1} \Sigma_{X'Z} \Sigma_{X'Z}^{-1} + \sigma_\xi \xi_{X'k}^{-1} [\Sigma_{X'X} \beta_{IV} + \beta_{IV} \Sigma_{Z'} \Sigma_{X'X}^{-1}] \\
- (3 - \mu_4) \sigma^2 \xi_{X'k}^{-1} + 2 \sigma_\xi \xi_{X'k}^{-1} [\Sigma_{X'X} \beta_{IV} + \beta_{IV} \Sigma_{Z'} \Sigma_{X'X}^{-1}] \\
+ 1 - \sigma_\xi \xi_{X'k}^{-1} [\Sigma_{X'X} \beta_{IV} + \beta_{IV} \Sigma_{Z'} \Sigma_{X'X}^{-1}].
$$
Figure 1: Densities for $k = l = 1$; $\beta = 1$; $\sigma_z^2 / \sigma_x^2 = 10$; $n = 50, 200$. 
Figure 2: Densities for $k = l = 1; \beta = 1; \sigma_x^2/\sigma_z^2 = 10; n = 50, 200$. 
Figure 3: Densities for $k = l = 1; \beta = 1; \sigma^2_\epsilon/\sigma^2_\xi = 10; n = 50, 200$. 

29
Figure 4: Densities for $k = l = 1; \beta = 1; \sigma_z^2/\sigma^2 = 10; n = 50, 200.$
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 \( SN \); \( n = 20, 100 \).
Figure 7: Densities for $l - 1 = k = 1; \beta = 1; \sigma_x^2/\sigma_z^2 = 10; n = 50, 200.$
Figure 8: Densities for $l - 1 = k = 1; \beta = 1; \sigma^2_x / \sigma^2_z = 10; n = 50, 200$. 

34
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 $SN$; $n = 100$.  

36