Division of Economics, EGC
School of Humanities and Social Sciences
Nanyang Technological University
14 Nanyang Drive
Singapore 637332

On the Limiting and Empirical Distributions of IV Estimators When Some of the Instruments are Actually Endogenous

Jan F. KIVIET and Jerzy NIEMCZYK

29 December 2013

EGC Report No: 2013/11

HSS-04-88
Tel: +65 67905689
Email: D-EGC@ntu.edu.sg
http://egc.hss.ntu.edu.sg

On the limiting and empirical distributions of IV estimators when some of the instruments are actually endogenous

Jan F. Kiviet* and Jerzy Niemczyk†

29 December 2013

JEL-classification: *C13, C15, C30*

Keywords: *empirical density, inconsistent estimators, invalid instruments, (un)conditional asymptotic distribution, weak instruments*

Abstract

IV estimation is examined when some instruments may be invalid. This is relevant because the initial just-identifying orthogonality conditions are untestable, whereas their validity is required when testing the orthogonality of additional instruments by so-called over-identification restriction tests. Moreover, these tests have limited power when samples are small, especially when instruments are weak. Distinguishing between conditional and unconditional settings, we analyze the limiting distribution of inconsistent IV and examine normal first-order asymptotic approximations to its density in finite samples. For simple classes of models we compare these approximations with their simulated empirical counterparts over almost the full parameter space. The latter 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. Also, 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.

* *corresponding author:* Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332 (email: jfkiviet@ntu.edu.sg); Emeritus Professor of Econometrics, University of Amsterdam, The Netherlands (email: j.f.kiviet@uva.nl).

† European Central Bank, Frankfurt, Germany (email: jerzy.niemczyk@ecb.int).

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 difficult to assess whether an explanatory or an external variable excluded from the explanatory variables is uncorrelated with the disturbance term indeed, since such an orthogonality test for the validity of candidate instruments is viable only when just identification or overidentification by valid instruments has been assumed already. Thus, econometric methodology is handicapped due to the unfortunate prerequisite that some non-econometric methodology should provide first some truly valid instrumental variables, in number at least as large as the number of unknown coefficients (k) to be estimated. The validity of the initial k instruments is statistically untestable. Moreover, orthogonality tests of additional candidate instrumental variables 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. Then the IV estimator for the structural parameters will be inconsistent, even when the structural equation itself is correctly specified regarding the parameters of interest; moreover, standard and weak-instrument robust coefficient restriction tests will be unreliable.[1] Here we examine the magnitude of some consequences of using invalid instruments and analyze the determining factors of these consequences both in large and in finite samples.

In this paper we consider general (and in much more detail some simple specific) forms of static linear structural equations and corresponding partial reduced form systems in covariance stationary variables and examine the IV estimator when actually some of its exploited orthogonality conditions 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 a possibly invalid IV estimator for a single structural equation that itself has been specified correctly in the sense that its implied vector of error terms is IID (independent and identically distributed) with unconditional expectation equal to zero.[2] In addition to the expression for the inconsistency of the IV estimator in terms of parameters and data moments, we also examine its limiting distribution under standard large sample asymptotics. Unlike for consistent estimators this limiting distribution depends on whether or not one conditions on exogenous variables. The resulting analytic first-order asymptotic approximations to the actual distribution of inconsistent IV estimators in finite samples are compared by simulation experiments with the actual estimator distribution for widely varying model parameter values.

[1] See Murray (2006), Small (2007), Berkowitz et al. (2008, 2012) and Guggenberger (2012).

[2] Also the conditional distribution is assumed IID with zero expectation provided no endogenous regressors nor any invalid instruments are included in the conditioning set. An alternative point of departure is chosen in Hale et al. (1980), Rhodes and Westbrook (1981), Knight (1982) and Skeels (1995) where instruments are invalid due to omitted regressors.

Despite its obvious relevance, the analysis of IV estimators employing invalid instruments has not received very much attention in the literature. Yet, the special case of inconsistent OLS has been addressed in various studies. Goldberger (1964, p.359) examines for IIN (independent and identically normally distributed) regressors the unconditional limiting distribution and obtains the inconsistency and asymptotic variance. In an unpublished paper, Rothenberg (1972), conditioning on the unobserved means of the regressors, derives the limiting distribution of OLS when some regressors are affected by measurement errors. From this result Hausman (1978, formula 2.11) obtains the limiting distribution of OLS in a simultaneous equation with $k = 1$, while conditioning on the explanatory part of the reduced form equation. The same case can be found as a problem exercise in Phillips and Wickens (1978, Question 6.10). In their solution they indicate (but do not solve) the complexities of obtaining the variance of the limiting distribution. The analysis of the much more general problem of obtaining the limiting distribution of an inconsistent instrumental variables estimator has received attention too. For one equation from a normal linear dynamic simultaneous system with serial correlation this was taken up by Hendry (1979), and further examined in Maasumi and Phillips (1982) and Hendry (1982). These three studies seek a so-called control variate, which has the same limiting distribution, and examine its potential to enhance the effectiveness of Monte Carlo simulation findings by response surface analysis. The focus in these three studies is on the unconditional distribution of inconsistent IV estimation assuming normality of all variables involved, whereas ours will primarily be on the conditional distribution rendering distributional assumptions on predetermined variables redundant. Much more recently Bound et al. (1995, p.444) and Hahn and Hausman (2005) have examined effects of (locally) invalid instruments in simple linear one regressor models, i.e. $k = 1$, and Hall and Inoue (2003) the consequences of misspecification in a general nonlinear GMM context. Ashley (2009) derives the counterpart of Golberger's unconditional result for inconsistent 2SLS applied to an alleged just identified model ($l = k$) and demonstrates how it can be used in practice to produce inference on structural parameters by exposing its sensitivity to varying assumptions on the correlation of instruments and disturbances.

Our major analytic result will be an explicit formula for the asymptotic variance of inconsistent 2SLS in case $l \geq k \geq 1$, which does not require an assumption on nor estimates of higher-order population moments of genuinely exogenous components of the model, because we condition on their sample realizations. In principle, this conditional variance is smaller (in a matrix sense) than the corresponding unconditional variance, and hence more attractive, provided one is interested in inference that refers more directly to the actually observed data and less to a putative multivariate data generating system on all the variables involved. This conditional variance is obtained here by extending the approach introduced by Rothenberg (1972) for inconsistent OLS estimators, also followed for OLS estimation of a simultaneous equation in Kiviet and Niemczyk (2007, 2012) and Kiviet (2013), which complete some initial results obtained in Joseph and Kiviet (2005). For some simple cases we examine any discrepancies between conditional and unconditional asymptotic distributions and their simulated finite sample counterparts. We do not resort to the response surface technique, but to (dynamic) graphical techniques. These allow to cover almost the entire parameter space, in which we do not put the DGP parameters on the axes but transformations of these which much more directly convey the basic model characteristics, being: degree of simultaneity, instrument strength, instrument invalidity, model fit and sample size. 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 standard coefficient estimators when instruments may be both weak and invalid.

The analysis of the exact finite sample properties of consistent IV estimators has a long history. Early contributions are Sawa (1969), Mariano (1973) and Phillips (1980, 1983). More recent contributions (and further references) can be found in, for instance, Phillips (2006) and Hillier (2006). Some of our numerical 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. Here, we supplement these findings with illustrations for the case of invalid instruments, also addressing the yet underexposed effects of conditioning. Apart from providing general insights into the consequences of employing invalid instruments, the results might also provide a stepping stone towards the development of alternative operational empirical conditional inference techniques in the same vein as is provided in Ashley (2009) on the basis of inconsistent unconditional just-identified IV and in Kiviet (2013) on the basis of inconsistent unconditional OLS.

From our simulations we establish that invalid but not very weak 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 when inconsistent it has its probability mass centered around the pseudo-true-value instead of the true value. Even when instruments are weak and sample size is not too small we establish that the accuracy of the normal conditional asymptotic distribution is not all that bad, whereas standard large-sample asymptotics of the unconditional distribution can be 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 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 skewness, bimodality and a median away from the pseudo-true-value may occur for much more moderate weakness and simultaneity when instruments are invalid. Note, however, that in practice one can easily avoid using weak instruments, since weakness (unlike validity) can straightforwardly 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; see also Kiviet (2013). 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, this consistent weak instruments based IV estimator may well have a less attractive actual distribution than an inconsistent one based on strong but invalid instruments.

The structure of this paper is as follows. In Section 2 we introduce the model and the statistical assumptions on all instrumental and explanatory variables involved. Focussing on the alleged overidentified case we consider the generalized IV or 2SLS estimator and derive its inconsistency and conditional limiting distribution (proofs in appendices) for the case where all variables are covariance stationary. Next, these results are specialized for the just identified case, for which we make comparisons with the unconditional limiting distribution. These explicit limiting distributions are then compared with the more implicit results of Maasoumi and Phillips (1982), Hendry (1982) and Hall and Inoue (2003). Section 3 contains graphic illustrations of both the asymptotic and finite sample distributions in specific simple ($k = 1$) models, and we illustrate the consequences and some advantages of conditioning. Moreover, we compare the performance of valid and invalid IV with OLS. The latter uses always extremely strong but possibly invalid instruments. In separate subsections we examine models with $l = 1$ and with $l = 2$. Substantial attention is paid to obtaining a transparent design for simulating the data for these models and instruments and to establishing any invariance and symmetry properties of the estimators over the parameter space. Instead of presenting extensive tables, we present a few series of 2D and 3D graphs in print, and we provide dynamic multi-dimensional visualization techniques to present our findings more effectively via links to the web through animations on a monitor. Section 4 concludes and conjectures on how to use the obtained results in practice.

2 Model, assumptions and theorems

Before we define our framework in full detail, we will first sketch a broad outline. We consider data generating processes for variables for which n observations have been collected in the rows of $y = (y_1, ..., y_n)'$, $X = (x_1, ..., x_n)'$ and $Z = (z_1, ..., z_n)'$. The full column rank matrices X and Z have k and l columns respectively, with $l \geq k$; also $Z'X$ has full column rank. For $i = 1, ..., n$ column vector x_i contains the explanatory variables for y_i in a linear structural model, which is well-specified in a sense that implies $\varepsilon_i = y_i - x_i'\beta \sim IID(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon > 0$ for a true and unique β. The l variables collected in Z will be used as instrumental variables for estimating the k structural parameters of interest β of model

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

However, not all these instruments are necessarily valid. Some elements of z_i may be correlated with ε_i. Moreover, some of the z_i may only be weakly correlated with some or all elements of x_i, whereas some correspond to (or are spanned by) columns of Z and hence are extremely strongly correlated with elements of x_i. All variables are stationary and we denote their non-centered second moment matrices as $\Sigma_{X'X}$, $\Sigma_{Z'X}$ and $\Sigma_{Z'Z}$ respectively, which all three are assumed to have full column rank.

In this framework the generalized instrumental variable (GIV) or two-stage least-squares estimator of β exists and is given by

$$\hat{\beta}_{GIV} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y = (\hat{X}'\hat{X})^{-1}\hat{X}'y, \tag{2}$$

where we introduced the notation

$$\hat{X} \equiv Z\hat{\Pi} = Z(Z'Z)^{-1}Z'X. \tag{3}$$

Here $\hat{\Pi} = (Z'Z)^{-1}Z'X$ contains the OLS coefficient estimates of the first-stage regressions. The probability limit for $n \to \infty$ of $\hat{\beta}_{GIV}$ exists and can be expressed as

$$\begin{aligned}
\beta^*_{GIV} &\equiv \operatorname{plim} \hat{\beta}_{GIV} \\
&= \beta + \operatorname{plim}[X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon \\
&= \beta + \sigma^2_\varepsilon[\Sigma_{X'Z}\Sigma^{-1}_{Z'Z}\Sigma_{Z'X}]^{-1}\Sigma_{X'Z}\Sigma^{-1}_{Z'Z}\zeta,
\end{aligned} \tag{4}$$

where $\zeta \equiv \sigma^{-2}_\varepsilon \operatorname{plim} n^{-1}Z'\varepsilon = \sigma^{-2}_\varepsilon E(z'_i\varepsilon_i)$, $\forall i$. Vector β^*_{GIV} is also known as the pseudo-true-value of $\hat{\beta}_{GIV}$. We shall denote the inconsistency of $\hat{\beta}_{GIV}$ as

$$\ddot{\beta}_{GIV} \equiv \beta^*_{GIV} - \beta = \sigma^2_\varepsilon \Sigma^{-1}_{\hat{X}'\hat{X}}\Pi'\zeta, \tag{5}$$

where we used $\Sigma_{\hat{X}'\hat{X}} \equiv \Sigma_{X'Z}\Sigma^{-1}_{Z'Z}\Sigma_{Z'X}$ and $\Pi \equiv \operatorname{plim}(Z'Z)^{-1}Z'X = \Sigma^{-1}_{Z'Z}\Sigma_{Z'X}$. Note that the GIV estimator is consistent ($\beta^*_{GIV} = \beta$) if and only if $\Pi'\zeta = 0$, so in particular when $\zeta = 0$ or in the pathological case that $\zeta \neq 0$ but is in the $l - k$ dimensional subspace orthogonal to the space spanned by the columns of Π.

Since $\Pi = [E(Z'Z)]^{-1}E(Z'X)$, we may write

$$X = Z\Pi + V, \tag{6}$$

where $E(Z'V) = O$. This establishes (without imposing any restrictions) an orthogonal (in a stochastic sense) decomposition of X. Only when $\zeta = 0$ (i.e. when all instruments are valid) and when Z includes all the predetermined variables from the (here not specified) complete simultaneous system then (6) would contain all the reduced form equations for all the endogenous regressors in X. In our set-up, though, where Z may contain invalid instruments, these will often just be pseudo reduced form equations.

In a similar way as we decomposed X 'orthogonally' with respect to the matrix Z, we may decompose both Z and V with respect to the vector ε. Since $E(Z'\varepsilon) = n\sigma^2_\varepsilon\zeta$ this yields

$$Z = \bar{Z} + \varepsilon\zeta', \text{ with } E(\bar{Z}'\varepsilon) = 0, \tag{7}$$

and

$$V = \bar{V} + \varepsilon\nu', \text{ with } E(\bar{V}'\nu) = 0, \tag{8}$$

where $\nu \equiv n^{-1}\sigma^{-2}_\varepsilon E(V'\varepsilon)$. Substituting these in (6) yields for the regressor matrix X another 'orthogonal' decomposition (now not with respect to Z but with respect to ε)

$$X = (\bar{Z} + \varepsilon\zeta')\Pi + \bar{V} + \varepsilon\nu' = \bar{Z}\Pi + \bar{V} + \varepsilon(\zeta'\Pi + \nu') = \bar{X} + \varepsilon\xi', \tag{9}$$

where $\bar{X} \equiv \bar{Z}\Pi + \bar{V}$ and $\xi \equiv \Pi\zeta + \nu$, with $E(\bar{X}'\varepsilon) = 0$.

Note that for the observations on subject (or at time period) i, i.e. (y_i, x'_i, z'_i), all information not conflated with ε_i is embodied in $\bar{z}'_i$ (the components of z'_i which are uncorrelated with ε_i) and $\bar{x}'_i = \bar{z}_i\Pi + \bar{\nu}'_i$ (where $\bar{\nu}'_i$ contains those components of the pseudo reduced form disturbances which are uncorrelated with ε_i). The variables in the columns of $\bar{Z}$ and $\bar{X}$ are exogenous (with respect to ε). Apart from their columns corresponding with known zero elements in ζ and ξ they are unobservable, and thus $\bar{Z}$ and $\bar{X}$ do not establish a feasible information set. Nevertheless, we will demonstrate that it is viable to derive statistical – albeit unfeasible – results when conditioning on $\bar{Z}$ and $\bar{X}$. This is useful too, because it represents the benchmark for the situation where inference is not meant to refer to the population from which $\bar{Z}$ and $\bar{X}$ are drawn, but just to their actual realizations in the sample.

The above parametrizations and regularity conditions can now be summarized and formalized in the following framework definition.

Framework A. *We have for observations $i = 1, ..., n$: (i) the structural equation $y_i = x_i'\beta + \varepsilon_i$ and the alleged instruments z_i', with (ii) $E(x_i\varepsilon_i) = \sigma_\varepsilon^2\xi$ and $E(z_i\varepsilon_i) = \sigma_\varepsilon^2\zeta$, where the $k \times 1$ vector ξ parametrizes simultaneity and the $l \times 1$ vector ζ parametrizes instrument invalidity, whereas (iii) the disturbances are mutually uncorrelated and have the finite conditional moments $E(\varepsilon_i \mid \bar{X}, \bar{Z}) = 0$, $E(\varepsilon_i^2 \mid \bar{X}, \bar{Z}) = \sigma_\varepsilon^2$, $E(\varepsilon_i^3 \mid \bar{X}, \bar{Z}) = \mu_3\sigma_\varepsilon^3$ and $E(\varepsilon_i^4 \mid \bar{X}, \bar{Z}) = \mu_4\sigma_\varepsilon^4$, where $\bar{X} \equiv X - \varepsilon\xi'$ and $\bar{Z} \equiv Z - \varepsilon\zeta'$. Moreover, (iv) $\forall i$ $E(x_ix_i') = \Sigma_{X'X} \equiv \text{plim}_{n\to\infty} n^{-1}X'X$, $E(z_iz_i') = \Sigma_{Z'Z} \equiv \text{plim}_{n\to\infty} n^{-1}Z'Z$ and $E(z_ix_i') = \Sigma_{Z'X} \equiv \text{plim}_{n\to\infty} n^{-1}Z'X$ all have full column rank, and also (v) $X'X$, $Z'Z$ and $Z'X$ all have full column rank with probability one.*

Note that if $\xi_j = 0$ for some $j \in \{1, ...k\}$ then the j-th regressor in X is exogenous or 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 an invalid instrument otherwise. Note that this framework does not allow that either X or Z or both include lagged dependent variables. Our major result concerns the asymptotic variance of inconsistent instrumental variables estimation of β, conditional on the exogenous variables $\bar{X}$ and $\bar{Z}$ (hence, not requiring a distributional assumption on $\bar{X}$ and $\bar{Z}$). The unconditional asymptotic variance will only be examined for the case $l = k$ under normality.

In the special case $l = k$ of (alleged) just identification the above GIV results specialize to simple IV, i.e.

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y, \;\beta_{IV}^* = \beta + \sigma_\varepsilon^2\Sigma_{Z'X}^{-1}\zeta, \;\ddot{\beta}_{IV} = \sigma_\varepsilon^2\Sigma_{Z'X}^{-1}\zeta. \tag{10}$$

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

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y, \;\beta_{OLS}^* = \beta + \sigma_\varepsilon^2\Sigma_{X'X}^{-1}\xi, \;\ddot{\beta}_{OLS} = \sigma_\varepsilon^2\Sigma_{X'X}^{-1}\xi. \tag{11}$$

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

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

With increasing sample size the estimator $\hat{\beta}_{GIV}$ tends to β_{GIV}^*, instead of β. Therefore, 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 β_{GIV}^* too, see Rothenberg (1972). The derivations are relatively smooth if we center at

$$\beta_{n,GIV}^* \equiv \sigma_\varepsilon^2 S_{\hat{X}'\hat{X}}^{-1}P', \tag{12}$$

where $S_{\hat{X}'\hat{X}} \equiv S_{Z'X}'S_{Z'Z}^{-1}S_{Z'X}$, $P \equiv S_{Z'Z}^{-1}S_{Z'X}$, with

$$S_{Z'Z} \equiv n^{-1}E(Z'Z) = n^{-1}\bar{Z}'\bar{Z} + \sigma_\varepsilon^2\zeta\zeta' \text{ and } S_{Z'X} \equiv n^{-1}E(Z'X) = n^{-1}\bar{Z}'\bar{X} + \sigma_\varepsilon^2\zeta\xi'. \tag{13}$$

Note that $\text{plim}\, \beta^*_{n,GIV} = \beta^*_{GIV}$. For GIV estimators we obtain the following result on its convergence in distribution (proof in Appendix A).

Theorem 1. *In Framework B and conditional on the exogenous variables $\bar{X}$ and $\bar{Z}$ we have $n^{1/2}(\hat{\beta}_{GIV} - \beta^*_{n,GIV}) \xrightarrow{d} \mathcal{N}(0, V^{CN}_{GIV})$, with*

$$\begin{aligned} V^{CN}_{GIV} \;=\; & \sigma^2_\varepsilon c_5(1-c_3+c_4)\Sigma^{-1}_{\hat{X}'\hat{X}} + \sigma^2_\varepsilon c_4^2 \Sigma^{-1}_{\hat{X}'\hat{X}}\Sigma_{X'X}\Sigma^{-1}_{\hat{X}'\hat{X}} \\ & - c_4[\Sigma^{-1}_{\hat{X}'\hat{X}}\Sigma_{X'X}\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV} + \ddot{\beta}_{GIV}\ddot{\beta}'_{GIV}\Sigma_{X'X}\Sigma^{-1}_{\hat{X}'\hat{X}}] \\ & + \sigma^4_\varepsilon c_4(1-2c_4)\Sigma^{-1}_{\hat{X}'\hat{X}}\xi\xi'\Sigma^{-1}_{\hat{X}'\hat{X}} \\ & + \sigma^2_\varepsilon c_4(1-2c_5)[\Sigma^{-1}_{\hat{X}'\hat{X}}\xi\ddot{\beta}'_{GIV} + \ddot{\beta}_{GIV}\xi'\Sigma^{-1}_{\hat{X}'\hat{X}}] \\ & + [c_5(1-2c_5) - c_3 + \sigma^{-2}_\varepsilon(\ddot{\beta}'_{GIV}\Sigma_{X'X}\ddot{\beta}_{GIV})]\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV}, \end{aligned}$$

where $c_1 \equiv \sigma^2_\varepsilon\zeta'\Sigma^{-1}_{Z'Z}\zeta$, $c_2 \equiv \zeta'\Pi\ddot{\beta}_{GIV}$, $c_3 \equiv \xi'\ddot{\beta}_{GIV}$, $c_4 \equiv c_1 - c_2$ and $c_5 \equiv 1 - c_3 - c_4$.

The superindex CN of V^{CN}_{GIV} indicates that it refers to the conditional limiting distribution and to the case where the disturbances are "almost normal", because $\mu_3 = 0$ and $\mu_4 = 3$. The limiting distribution of $\hat{\beta}_{GIV}$ is still genuinely normal when instruments are invalid, although no longer centered at β but at the pseudo-true-value β^*_{GIV}. When all instruments are valid ($\zeta = 0$) then $\beta^*_{GIV} = \beta$, $\ddot{\beta}_{GIV} = 0$ 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 $n^{1/2}(\hat{\beta}_{GIV} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2_\varepsilon\Sigma^{-1}_{\hat{X}'\hat{X}})$. Note that when $\zeta = 0$ the asymptotic variance of $\hat{\beta}_{GIV}$ is not determined by the simultaneity ξ, because $\Sigma_{\hat{X}'\hat{X}} = \Sigma_{X'Z}\Sigma^{-1}_{Z'Z}\Sigma_{Z'X} = \Sigma_{\bar{X}'\bar{Z}}\Sigma^{-1}_{\bar{Z}'\bar{Z}}\Sigma_{\bar{Z}'\bar{X}}$. However, when instruments are invalid ($\zeta \neq 0$) then $\Sigma_{Z'X} = \Sigma_{\bar{Z}'\bar{X}} + \sigma^2_\varepsilon\zeta\xi'$ and $\Sigma_{Z'Z} = \Sigma_{\bar{Z}'\bar{Z}} + \sigma^2_\varepsilon\zeta\zeta'$, thus $\Sigma_{\hat{X}'\hat{X}}$ is now determined by both ξ and ζ. When fitting X to Z, while the $\varepsilon\xi'$ part of X is not (asymptotically) orthogonal to Z due to its component $\varepsilon\zeta'$, this does not only lead to the inconsistency, but also to the many extra terms in the asymptotic conditional variance.

When $l = k$ and GIV specializes to IV we have $\zeta'\Pi\ddot{\beta}_{IV} = \sigma^2_\varepsilon\zeta'\Sigma^{-1}_{Z'Z}\Sigma_{Z'X}\Sigma^{-1}_{Z'X}\zeta = \sigma^2_\varepsilon\zeta'\Sigma^{-1}_{Z'Z}\zeta$, so $c_1 = c_2$, $c_4 = 0$ and $c_5 = 1 - c_3$. This leads to substantial simplification, giving:

Corollary 1. *In Framework B for the special case $l = k$ we have that conditional on $\bar{X}$ and $\bar{Z}$ we have $n^{1/2}(\hat{\beta}_{IV} - \beta^*_{n,IV}) \xrightarrow{d} \mathcal{N}(0, V^{CN}_{IV})$, with*

$$V^{CN}_{IV} = \sigma^2_\varepsilon(1-c_3)^2\Sigma^{-1}_{Z'X}\Sigma_{Z'Z}\Sigma^{-1}_{X'Z} - [(1-c_3)^2 - \sigma^2_\varepsilon\zeta'\Sigma^{-1}_{Z'X}\Sigma_{\bar{X}'\bar{X}}\Sigma^{-1}_{X'Z}\zeta]\ddot{\beta}_{IV}\ddot{\beta}'_{IV},$$

where $\beta^*_{n,IV} \equiv \beta + \sigma^2_\varepsilon S^{-1}_{Z'X}\zeta$ and $c_3 \equiv \xi'\ddot{\beta}_{IV} = \sigma^2_\varepsilon\xi'\Sigma^{-1}_{Z'X}\zeta$.

When all instruments are valid ($\zeta = 0$) this result specializes to the standard result $n^{1/2}(\hat{\beta}_{IV} - \beta) \xrightarrow{d} N\left(0, \sigma^2_\varepsilon\Sigma^{-1}_{Z'X}\Sigma_{Z'Z}\Sigma^{-1}_{X'Z}\right)$, which is also true unconditionally and for non-normal disturbances. Since for arbitrary ζ and ξ the scalar $\sigma^2_\varepsilon\xi'\Sigma^{-1}_{Z'X}\zeta$ can either be positive or negative, no general conclusions can be drawn on the behavior of V^{CN}_{IV} in comparison to the reference case $\sigma^2_\varepsilon\Sigma^{-1}_{Z'X}\Sigma_{Z'Z}\Sigma^{-1}_{X'Z}$. Depending on the particular parametrization and data moment matrices the conditional 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 $\hat{\beta}_{IV} = \hat{\beta}_{OLS}$, the resulting $V^{CN}_{IV} = V^{CN}_{OLS}$ is the same as the formula for the conditional asymptotic variance found for an inconsistent OLS estimator when the disturbances are (almost) normal, as earlier derived in Kiviet and Niemczyk (2007, 2012).

Next, we look at the case where the disturbances may have general 3^{rd} and 4^{th} moment (proof in Appendix B). Let ι be a $n \times 1$ vector of unit elements. Upon defining

$$\begin{aligned}\Sigma_{Z'\iota} &\equiv \operatorname{plim} n^{-1}Z'\iota = \operatorname{plim} n^{-1}\bar{Z}'\iota \equiv \Sigma_{\bar{Z}'\iota},\\ \Sigma_{X'\iota} &\equiv \operatorname{plim} n^{-1}X'\iota = \operatorname{plim} n^{-1}\bar{X}'\iota \equiv \Sigma_{\bar{X}'\iota},\end{aligned}$$

we find (superindex CP indicates conditional on $\bar{X}$ and $\bar{Z}$ with IID disturbances following a Pearson distribution):

Theorem 2. *In Framework A and conditional on $\bar{X}$ and $\bar{Z}$ we have $n^{1/2}(\hat{\beta}_{GIV} - \beta^*_{n,GIV}) \xrightarrow{d} \mathcal{N}(0, V^{CP}_{GIV})$, where V^{CP}_{GIV} is equal to V^{CN}_{GIV}, given in Theorem 1, plus two additional terms. When $\mu_4 \neq 3$ an additional term is*

$$(\mu_4 - 3)\{\sigma_\varepsilon^4 c_4^2 \Sigma^{-1}_{\hat{X}'\hat{X}}\xi\xi'\Sigma^{-1}_{\hat{X}'\hat{X}} + \sigma_\varepsilon^2 c_4 c_5[\ddot{\beta}_{GIV}\xi'\Sigma^{-1}_{\hat{X}'\hat{X}} + \Sigma^{-1}_{\hat{X}'\hat{X}}\xi\ddot{\beta}'_{GIV}] + c_5^2\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV}\},$$

where $c_4 \equiv \sigma_\varepsilon^2\zeta'\Sigma^{-1}_{Z'Z}\zeta - \zeta'\Pi\ddot{\beta}_{GIV}$ and *$c_5 \equiv 1 - \xi'\ddot{\beta}_{GIV} - c_4$. When $\mu_3 \neq 0$ another additional term is*

$$\begin{aligned}\mu_3\{&c_4[\sigma_\varepsilon^3 c_4 \Sigma^{-1}_{\hat{X}'\hat{X}}\Sigma_{X'\iota}\xi'\Sigma^{-1}_{\hat{X}'\hat{X}} - \sigma_\varepsilon\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\xi'\Sigma^{-1}_{\hat{X}'\hat{X}} + \sigma_\varepsilon^3 c_5\Sigma^{-1}_{\hat{X}'\hat{X}}\Pi'\Sigma_{Z'\iota}\xi'\Sigma^{-1}_{\hat{X}'\hat{X}}\\ &\qquad + (\sigma_\varepsilon^3\Sigma^{-1}_{\hat{X}'\hat{X}}\xi - \sigma_\varepsilon\ddot{\beta}_{GIV})(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma^{-1}_{Z'Z}\Sigma_{Z'\iota}\xi'\Sigma^{-1}_{\hat{X}'\hat{X}}]\\ &+c_5[\sigma_\varepsilon c_4\Sigma^{-1}_{\hat{X}'\hat{X}}\Sigma_{X'\iota}\ddot{\beta}'_{GIV} - \sigma_\varepsilon^{-1}\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV} + \sigma_\varepsilon c_5\Sigma^{-1}_{\hat{X}'\hat{X}}\Pi'\Sigma_{Z'\iota}\ddot{\beta}'_{GIV}\\ &\qquad + (\sigma_\varepsilon\Sigma^{-1}_{\hat{X}'\hat{X}}\xi - \sigma_\varepsilon^{-1}\ddot{\beta}_{GIV})(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma^{-1}_{Z'Z}\Sigma_{Z'\iota}\ddot{\beta}'_{GIV}]\\ &+c_4[\sigma_\varepsilon^3 c_4\Sigma^{-1}_{\hat{X}'\hat{X}}\xi\Sigma'_{X'\iota}\Sigma^{-1}_{\hat{X}'\hat{X}} - \sigma_\varepsilon\Sigma^{-1}_{\hat{X}'\hat{X}}\xi\Sigma'_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV} + \sigma_\varepsilon^3 c_5\Sigma^{-1}_{\hat{X}'\hat{X}}\xi\Sigma'_{Z'\iota}\Pi\Sigma^{-1}_{\hat{X}'\hat{X}}\\ &\qquad + \Sigma^{-1}_{\hat{X}'\hat{X}}\xi\Sigma'_{Z'\iota}\Sigma^{-1}_{Z'Z}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\sigma_\varepsilon^3\xi'\Sigma^{-1}_{\hat{X}'\hat{X}} - \sigma_\varepsilon\ddot{\beta}'_{GIV})]\\ &+c_5[\sigma_\varepsilon c_4\ddot{\beta}_{GIV}\Sigma'_{X'\iota}\Sigma^{-1}_{\hat{X}'\hat{X}} - \sigma_\varepsilon^{-1}\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV} + \sigma_\varepsilon c_5\ddot{\beta}_{GIV}\Sigma'_{Z'\iota}\Pi\Sigma^{-1}_{\hat{X}'\hat{X}}\\ &\qquad + \ddot{\beta}_{GIV}\Sigma'_{Z'\iota}\Sigma^{-1}_{Z'Z}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\sigma_\varepsilon\xi'\Sigma^{-1}_{\hat{X}'\hat{X}} - \sigma_\varepsilon^{-1}\ddot{\beta}'_{GIV})]\}.\end{aligned}$$

When all the instruments are valid ($\zeta = 0$) this result again collapses to the standard one, i.e. $V^{CP}_{GIV} = \sigma_\varepsilon^2\Sigma^{-1}_{\hat{X}'\hat{X}}$, 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 that conditional on $\bar{X}$ and $\bar{Z}$ we have $n^{1/2}(\hat{\beta}_{IV} - \beta^*_{n,IV}) \xrightarrow{d} \mathcal{N}(0, V^{CP}_{IV})$, with*

$$\begin{aligned}V^{CP}_{IV} &= \sigma_\varepsilon^2 c_5^2\Sigma^{-1}_{Z'X}\Sigma_{Z'Z}\Sigma^{-1}_{X'Z} + \mu_3\sigma_\varepsilon c_5^2[\Sigma^{-1}_{Z'X}\Sigma_{Z'\iota}\ddot{\beta}'_{IV} + \ddot{\beta}_{IV}\Sigma'_{Z'\iota}\Sigma^{-1}_{X'Z}]\\ &\quad -[(5-\mu_4)c_5^2 + 2c_5(\mu_3\sigma_\varepsilon^{-1}\ddot{\beta}'_{IV}\Sigma_{X'\iota} - 1) + 1 - \sigma_\varepsilon^{-2}\ddot{\beta}'_{IV}\Sigma_{X'X}\ddot{\beta}_{IV}]\ddot{\beta}_{IV}\ddot{\beta}'_{IV}.\end{aligned}$$

where $c_5 \equiv 1 - \xi'\ddot{\beta}_{IV} = 1 - \sigma_\varepsilon^2\xi'\Sigma^{-1}_{Z'X}\zeta$.

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

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 $E(v_i v_h') = O$ for $i \neq h = 1, ..., n$, then $n^{1/2}\bar{v} \xrightarrow{d} \mathcal{N}(0, \lim_{n\to\infty} \bar{V})$, where $\bar{v} = n^{-1}\sum_{i=1}^n v_i$ and $\bar{V} = n^{-1}\sum_{i=1}^n V_i$. We employ the following generalized version:

Lemma. *Let ω be a $k \times 1$ nonrandom vector and $W = (w_1, ..., w_n)'$ a $n \times m$ random matrix with w_i being stationary $\forall i$, whereas the $n \times 1$ vector $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$ has mutually uncorrelated elements for which $E(\varepsilon_i \mid W) = 0$, $E(\varepsilon_i^2 \mid W) = \sigma_\varepsilon^2$, $E(\varepsilon_i^3 \mid W) = \mu_3\sigma_\varepsilon^3$ and $E(\varepsilon_i^4 \mid W) = \mu_4\sigma_\varepsilon^4$. Then for the $k \times 1$ vector $v_i = w_i\varepsilon_i + \omega(\varepsilon_i^2 - \sigma_\varepsilon^2)$ we have $E(v_i \mid W) = 0$, $E(v_i v_i' \mid W) = V_i = \sigma_\varepsilon^2 w_i w_i' + \mu_3\sigma_\varepsilon^3(w_i\omega' + \omega w_i') + (\mu_4 - 1)\sigma_\varepsilon^4\omega\omega'$, whereas $E(v_i v_h' \mid W) = O$ for $i \neq h$, so that for $n^{-1/2}\sum_{i=1}^n v_i = n^{-1/2}[W'\varepsilon + \omega(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)]$ a CLT implies that conditional on W*

$$n^{1/2}\bar{v} \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2\Sigma_{W'W} + \mu_3\sigma_\varepsilon^3(\Sigma_{W'\iota}\omega' + \omega\Sigma_{\iota'W}) + (\mu_4 - 1)\sigma_\varepsilon^4\omega\omega'),$$

where $\Sigma_{W'W} \equiv \operatorname{plim} n^{-1}W'W = E(w_i w_i')$ and $\Sigma_{W'\iota} \equiv \Sigma_{\iota'W}' \equiv \operatorname{plim} n^{-1}W'\iota = E(w_i)$ $\forall i$, with ι a $n \times 1$ vector of unit elements.

The principle difference between the conditional approach followed above and the results obtained in Maasoumi and Phillips (1982) is the following. Instead of simplifying (by conditioning) formula (52) of Appendix A into an expression to which the above Lemma applies and from which the asymptotic variance directly follows, they substitute (52) in (49) and then rewrite it as $n^{1/2}Hvec[n^{-1}F'F - E(n^{-1}F'F)]$, in which H is a $k \times nm$ transformation matrix determined by selection and data moment matrices and parameters, whereas the random zero mean vectorized object is obtained from the $n \times m$ matrix F, which contains all the n observations on all the m variables (both endogenous and predetermined) from a multi-equation dynamic simultaneous system with serial correlation. This expression constitutes the control variate that has the same unconditional limiting distribution as a single equation GIV estimator in which invalid instruments are being used. Next they sketch how under normality of all the m variables (including any exogenous and predetermined variables) and exploiting a spectral density matrix its variance could be established. Hence, they do not provide an explicit formula and only consider the unconditional asymptotic variance under normality. This also applies to the result in Hendry (1982), who provides algebraic instructions on how the unconditional asymptotic variance can be computed when all variables are normal. Note, though, that evaluation of the unconditional asymptotic variance, like the conditional one, is unfeasible in the sense that it requires substitution of the values of unobserved parameters.

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 Ω. This matrix, which is left unspecified, is the variance of the limiting distribution of a $k + 2l$ vector, say v^*, which in our particular setting specializes to

$$v^* \equiv \begin{pmatrix} n^{-1/2}\left[Z'(y - X\beta_{GIV}^*) - n \operatorname{plim} n^{-1}Z'(y - X\beta_{GIV}^*)\right] \\ n^{1/2}\left(n^{-1}X'Z - \Sigma_{X'Z}\right)\Sigma_{Z'Z}^{-1} \operatorname{plim} n^{-1}Z'(y - X\beta_{GIV}^*) \\ n^{1/2}\left[(n^{-1}Z'Z)^{-1} - \Sigma_{Z'Z}^{-1}\right] \operatorname{plim} n^{-1}Z'(y - X\beta_{GIV}^*) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Omega).$$

How Ω should be obtained and whether or not it is supposed to represent the unconditional or the conditional variance they do not mention. It can be shown, though, that in our model

$$n^{1/2}(\hat{\beta}_{GIV} - \beta_{n,GIV}^*) = \Sigma_{\hat{X}'\hat{X}}^{-1}\begin{pmatrix} \Pi' & I_k & \Sigma_{X'Z} \end{pmatrix} v^* + o_p(1), \tag{14}$$

and thus the conditional asymptotic variance of GIV that we obtained can indeed be expressed in $\Sigma_{Z'X}$, $\Sigma_{Z'Z}$ and Ω (which we did express for our linear model explicitly in $\Sigma_{Z'X}$, $\Sigma_{Z'Z}$, ξ, ζ, μ_3, μ_4 and σ_ε^2). The added value 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 and, because they are explicit in parameters and data moments, they allow to depict the numerical effects of those various characteristics, as we will show in the next section.

When $l = k$ and the data are IID, it is relatively straightforward to generalize the result by Goldberger (1964, p.359) on inconsistent OLS and obtain the unconditional limiting distribution of the IV estimator when some instruments are invalid. Subtracting from all components in the model their unconditional expectation, and thus removing the intercept (if present) from both the model and the set of instruments, the remaining regressors all have expectation zero. Still using the same notation for the model (although it may now have one regressor less), we consider also the alternative model specification

$$y = X\beta + \varepsilon = X\beta_{IV}^* + u, \tag{15}$$

where $u \equiv \varepsilon - X\ddot{\beta}_{IV}$. Because now $E(x_i) = 0$, we have $E(u_i) = E(\varepsilon_i) - E(x_i')\ddot{\beta}_{IV} = 0$, and also $\sigma_u^2 \equiv E(u_i^2) = \sigma_\varepsilon^2(1 - 2\ddot{\beta}_{IV}'\xi) + \ddot{\beta}_{IV}'\Sigma_{X'X}\ddot{\beta}_{IV}$, $E(u_i u_j) = 0$ for $i \neq j$, whereas $E(z_i u_i) = \sigma_\varepsilon^2\zeta - E(z_i x_i')\ddot{\beta}_{IV} = \sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{IV} = 0$. Thus, the instruments Z are valid in this alternative model specification, whereas for $l > k$ this would not generally be true. Ashley (2009) derives the variance of the unconditional limiting distribution of the IV estimator of β_{IV}^*, which in the general IID case involves a great many fourth order moments. He also shows how it simplifies when all regressors, instruments and the disturbances are normal, in particular when $l = k = 1$. This latter case we will examine more closely in the next section.

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 samples. We will also 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 $n^{1/2}(\hat{\beta} - \beta_n^*) \xrightarrow{d} \mathcal{N}(0, V)$ and they imply a first-order approximation to the distribution of $\hat{\beta}$ in finite samples which can be expressed as

$$\hat{\beta} \overset{a}{\sim} \mathcal{N}(\beta + \ddot{\beta}, n^{-1}V). \tag{16}$$

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

$$\text{AMSE}(\hat{\beta}) \equiv n^{-1}V + \ddot{\beta}\ddot{\beta}'. \tag{17}$$

The actual values of $\ddot{\beta}$ and of (the square root of) AMSE($\hat{\beta}$) can be computed for any n and any given values of the model parameters and relevant data moments. To find out how accurate the first-order asymptotic approximation (17) 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 samples, as is the case for GIV 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 β the median error ME($\hat{\beta}$) and the median absolute error MAE($\hat{\beta}$) are defined as

$$\begin{array}{l}\Pr\{(\hat{\beta}-\beta)\leq \mathrm{ME}(\hat{\beta})\}=0.5,\\ \Pr\{\mid \hat{\beta}-\beta \mid\leq \mathrm{MAE}(\hat{\beta})\}=0.5.\end{array} \tag{18}$$

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

$$\begin{aligned}0.5 &= \Pr\{\mid \hat{\beta}-\beta \mid\leq m\}=1-\Pr\{\mid \hat{\beta}-\beta \mid> m\}\\ &= 1-\Pr\{\hat{\beta}-\beta>m\}-\Pr\{\hat{\beta}-\beta<-m\}\\ &= \Pr\{\hat{\beta}-\beta<m\}-\Pr\{\hat{\beta}-\beta<-m\}\\ &\overset{a}{=} \Phi_{\ddot{\beta},\sigma_{\hat{\beta}}}(m)-\Phi_{\ddot{\beta},\sigma_{\hat{\beta}}}(-m),\end{aligned}$$

so that we can solve[3] m from

$$\Phi_{\ddot{\beta},\sigma_{\hat{\beta}}}(m)=0.5+\Phi_{\ddot{\beta},\sigma_{\hat{\beta}}}(-m). \tag{19}$$

Below we will examine the empirical finite sample distribution of scalar $\hat{\beta}_{GIV}$ and compare it with $\hat{\beta}_{GIV}\overset{a}{\sim}\mathcal{N}(\beta+\ddot{\beta}_{GIV},n^{-1}V_{GIV}^{CN})$. In addition, for various estimators $\hat{\beta}_{GIV}$ (including $\hat{\beta}_{IV}$ and $\hat{\beta}_{OLS}$), we examine MAE($\hat{\beta}_{GIV}$) and compare it with AMAE($\hat{\beta}_{GIV}$) 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

[3]Since $m=\Phi^{-1}_{\ddot{\beta},\sigma_{\hat{\beta}}}[0.5+\Phi_{\ddot{\beta},\sigma_{\hat{\beta}}}(-m)]$, we employed the iterative scheme, $m_0=0$, $m_{i+1}=\Phi^{-1}_{\ddot{\beta},\sigma_{\hat{\beta}}}[0.5+\Phi_{\ddot{\beta},\sigma_{\hat{\beta}}}(-m_i)]$ for $i=0,1,...$ until convergence. When $\ddot{\beta}=0$ no iteration is required since $m=\Phi^{-1}_{0,\sigma_{\hat{\beta}}}(0.75)$ conforms to the quartile.

[4]This document contains a few series of graphs only; more complete and animated pictures are available via the web, see `https://sites.google.com/site/homepagejfk/`

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 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 (2005). Though, only the latter two papers consider (locally) invalid instruments.

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

$$\begin{aligned} \ddot{\beta}_{IV} &= \beta_{IV}^* - \beta = \sigma_\varepsilon^2 \Sigma_{Z'X}^{-1} \zeta = \frac{\sigma_{z\varepsilon}}{\sigma_{xz}} = \frac{\rho_{z\varepsilon}}{\rho_{xz}} \frac{\sigma_\varepsilon}{\sigma_x} \\ c_3 &= \sigma_\varepsilon^2 \xi' \Sigma_{Z'X}^{-1} \zeta = \frac{1}{\sigma_\varepsilon^2} \frac{\sigma_{z\varepsilon}\sigma_{x\varepsilon}}{\sigma_{xz}} = \frac{\rho_{z\varepsilon}\rho_{x\varepsilon}}{\rho_{xz}} \\ \ddot{\beta}_{IV}' \Sigma_{X'X} \ddot{\beta}_{IV} &= \sigma_\varepsilon^4 \zeta' \Sigma_{Z'X}^{-1} \Sigma_{X'X} \Sigma_{X'Z}^{-1} \zeta = \sigma_{z\varepsilon}^2 \frac{\sigma_x^2}{\sigma_{xz}^2} = \sigma_\varepsilon^2 \frac{\rho_{z\varepsilon}^2}{\rho_{xz}^2}, \end{aligned} \tag{20}$$

giving in the case where the disturbances are (almost) normally distributed

$$V_{IV}^{CN} = \frac{\sigma_\varepsilon^2}{\sigma_x^2} \frac{(1-\rho_{z\varepsilon}^2)(\rho_{xz} - \rho_{z\varepsilon}\rho_{x\varepsilon})^2 + \rho_{z\varepsilon}^4(1-\rho_{x\varepsilon}^2)}{\rho_{xz}^4}. \tag{21}$$

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

For this simple model we find from Ashley (2009, formula 15) for the unconditional asymptotic variance under normality

$$V_{IV}^{UN} = \frac{\sigma_\varepsilon^2}{\sigma_x^2} \frac{(\rho_{xz} - \rho_{z\varepsilon}\rho_{x\varepsilon})^2 + \rho_{z\varepsilon}^2(1-\rho_{x\varepsilon}^2)}{\rho_{xz}^4}, \tag{22}$$

which makes immediately clear that the conditional asymptotic variance (21) is smaller if $\rho_{z\varepsilon} \neq 0$, as expected.

We now set out to design the Monte Carlo. 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 taken as

$$SN = \frac{\beta^2\sigma_x^2}{\sigma_\varepsilon^2} = \frac{\sigma_x^2}{\sigma_\varepsilon^2}. \tag{23}$$

From (20) and (21) we find that V_{IV}^{CN} and $\ddot{\beta}$ are proportional to (the square root of) the inverse of SN. In fact, after the normalization $\beta = 1$, in this simple model the approximation to the distribution of the IV estimator $\hat{\beta}_{IV} \overset{a}{\sim} \mathcal{N}(\beta + \ddot{\beta}, n^{-1}V_{IV}^{CN})$ is completely determined by n and the four model characteristics ρ_{xz}, $\rho_{x\varepsilon}$, $\rho_{z\varepsilon}$ 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 is given by

$$\left.\begin{array}{r} y_i = \beta x_i + \varepsilon_i \\ x_i = \bar{x}_i + \xi\varepsilon_i \\ z_i = \bar{z}_i + \zeta\varepsilon_i \end{array}\right\} \tag{24}$$

where ξ and ζ are scalars. In order to obtain $(\varepsilon_i, x_i, z_i)' \sim \text{IID}(0, \Omega)$, with appropriate 3×3 covariance matrix Ω, 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:

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

This provides full generality. The coefficient α_1 determines $\sigma^2_{\bar{x}}$, whereas $E(\bar{x}_i\varepsilon_i) = 0$, as it should. Also $E(\bar{z}_i\varepsilon_i) = 0$, and α_2 and α_3 enable any correlation between $\bar{x}_i$ and $\bar{z}_i$ and any value of $\sigma^2_{\bar{z}}$. The above implies

$$\begin{pmatrix} \varepsilon_i \\ x_i \\ z_i \end{pmatrix} = \Omega^{1/2} v_i = \begin{pmatrix} \sigma_\varepsilon & 0 & 0 \\ \sigma_\varepsilon\xi & \alpha_1 & 0 \\ \sigma_\varepsilon\zeta & \alpha_2 & \alpha_3 \end{pmatrix} \begin{pmatrix} v_{i,1} \\ v_{i,2} \\ v_{i,3} \end{pmatrix}. \tag{25}$$

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

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

$$\hat{\beta}_{IV} = \frac{\sum z_i y_i}{\sum z_i x_i} = \beta + \frac{\sum z_i \varepsilon_i}{\sum z_i x_i}, \tag{26}$$

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 $\rho_{z\varepsilon} \geq 0$, because of the following reasoning. The value of ρ_{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 $\rho_{z\varepsilon} \leq 0$ and arbitrary positive or negative value of $\rho_{x\varepsilon}$ is equivalent with the distribution of $-(\hat{\beta}_{IV} - \beta)$ for $-\rho_{z\varepsilon} \geq 0$ and $-\rho_{x\varepsilon}$. It is also obvious that σ_x and σ_ε 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 Ω. For these we choose

$$\sigma_\varepsilon = 1, \tag{27}$$

$$\sigma^2_z = \zeta^2 + \alpha_2^2 + \alpha_3^2 = 1. \tag{28}$$

By (27) we normalize all results with respect to σ_ε, and (23) simplifies to

$$SN = \sigma^2_x. \tag{29}$$

Because any GIV estimator is invariant to the scale of the instruments (only the space spanned by the instruments is relevant) we may impose (28), which will be used to obtain the value

$$\alpha_3 = \left|(1 - \zeta^2 - \alpha_2^2)^{1/2}\right|, \tag{30}$$

where, without loss of generality, we may stick to positive values for α_3 as long as $\upsilon_{i,3}$ is symmetrically distributed. For similar reasons we would get observationally equivalent data realizations if both α_1 and α_2 would be changed in sign. Therefore, below we will restrict ourselves to just positive values for both α_1 and α_3.

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

$$\begin{array}{ll} \sigma_x^2 = \xi^2 + \alpha_1^2 & \sigma_y^2 = \xi^2 + 2\xi + 1 + \alpha_1^2 \\ \sigma_{x\varepsilon} = \xi & \rho_{x\varepsilon} = \xi/\sqrt{\xi^2 + \alpha_1^2} \\ \sigma_{z\varepsilon} = \zeta & \rho_{z\varepsilon} = \zeta \\ \sigma_{xz} = \xi\zeta + \alpha_1\alpha_2 & \rho_{xz} = (\xi\zeta + \alpha_1\alpha_2)/\sqrt{\xi^2 + \alpha_1^2} \end{array} \tag{31}$$

Note that these, after the normalizations $\beta = 1$, $\sigma_\varepsilon = 1$ and $\sigma_z = 1$, depend on only 4 free parameters of the data generating process (DGP), viz. ξ, ζ, α_1 and α_2. As we already established, the expressions for inconsistency in (20) and asymptotic variance (21) evaluated under $\mu_3 = 0$ and $\mu_4 = 3$ (the 3^{rd} and 4^{th} moment of $\upsilon_{i,1}$) depend on just four characteristics too, viz. on $\rho_{x\varepsilon}$, $\rho_{z\varepsilon}$, ρ_{xz} and $SN = \sigma_x^2$. 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

$$\begin{array}{l} \zeta = \rho_{z\varepsilon}, \\ \xi = \rho_{x\varepsilon}\sigma_x, \\ \alpha_1 = \sigma_x \left|(1 - \rho_{x\varepsilon}^2)^{1/2}\right|, \\ \alpha_2 = (\rho_{xz} - \rho_{x\varepsilon}\rho_{z\varepsilon})/\left|(1 - \rho_{x\varepsilon}^2)^{1/2}\right|, \end{array} \tag{32}$$

from which α_3 follows directly via (30). This reparametrization[5] is useful, because the parameters $\rho_{x\varepsilon}$, $\rho_{z\varepsilon}$, ρ_{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 $|\rho_{x\varepsilon}| < 1$, $0 \leq \rho_{z\varepsilon} < 1$, $0 < \rho_{xz} < 1$ and $0 < \sigma_x^2/(\sigma_x^2 + 1) < 1$, we can examine the limiting and finite sample (un)conditional distributions of $\hat{\beta}_{IV}$ over the entire parameter space of this model. Note, however, that not all values of these parameters will be admissible. For example, when $\rho_{x\varepsilon}$ is large and $\rho_{z\varepsilon}$ is small, this cannot be compatible with ρ_{xz} being very large. Moreover, σ_x has just an effect on the scale of $\ddot{\beta}_{IV}$, V_{IV}^{CN}, V_{IV}^{U} and $\hat{\beta}_{IV} - \beta$, so we may choose just one fixed value for σ_x and from these findings the results for any value of σ_x can be obtained simply by rescaling. In our calculations and simulations we will fix $SN = 10$, yielding a population fit of the model of $10/11 = 0.909$.

In the simulations we took $\upsilon_i \sim \text{IIN}(0, I_3)$ and used 2,000,000 replications. From the results we may expect to get quick insights into issues as the following. For which

[5] For more background regarding the role of the simulation design parameter space and its relationship with the data generating process parameter space see Kiviet (2011).

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 β^*_{IV}, (c) the actual tail behavior of $\hat{\beta}_{IV}$ reasonably well represented by that of the asymptotic normal approximation, and (d) how does the magnitude of the actual estimation errors depend on the design parameters. Hence, we will assess the correspondences and differences in shape, location and spread of the asymptotic and the empirical distributions, both for the conditional and the unconditional 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 tails of the finite sample distribution.

For our results on the simulated conditional distribution only one arbitrary draw of the series $\upsilon_{i,2}$ and $\upsilon_{i,3}$ has been used. However, because such a single random realization will not yield two series which have an actual sample correlation of zero and a sample variance of one, this implies that we would loose full control over the actual values of SN and of the strength of the instrument. We overcame this problem by transforming these single draws of two n-element vectors υ_j by replacing one by its residuals after regressing on the other and by normalizing them, such that they obtained zero sample mean, zero sample correlation and inner-product n. Naturally, when establishing the conditional asymptotic distribution by evaluating $\ddot{\beta}_{IV}$ and V^{CN}_{IV}, one has to replace σ_{xz} by $\bar{x}'\bar{z}/n + \sigma^2_\varepsilon\zeta\xi$ and σ^2_x by $\bar{x}'\bar{x}/n + \sigma^2_\varepsilon\xi^2$. Due to our transformation of the υ_j series $\bar{x}'\bar{x} = \alpha^2_1\upsilon'_2\upsilon_2 = n\alpha^2_1$ and likewise $\bar{x}'\bar{z} = n\alpha_1\alpha_2$ yielding exactly the same numerical values for σ_{xz} and σ^2_x as for the unconditional case. Therefore, the assessment of the standard asymptotic distribution when $\rho_{z\varepsilon} = 0$ will be the same for the conditional and unconditional case, which would not occur for untransformed single realizations of $\bar{x}$ and $\bar{z}$; and, for $\rho_{z\varepsilon} \neq 0$, our "standardized" results will underexpose the effects of conditioning.

In order to find out how accurate the first-order asymptotic approximations are, actual values of $\ddot{\beta}_{IV}$ and of AMAE($\hat{\beta}_{IV}$) can now be calculated for any set of compatible values of n, $\rho_{x\varepsilon}$, $\rho_{z\varepsilon}$, ρ_{xz} and σ_x, and they can be compared with corresponding Monte Carlo estimates obtained from $\hat{\beta}_{IV}$ realizations in simulated finite samples. Before we present these summarizing characteristics, we will first examine the actual density functions themselves. The Figures 1.1 through 1.4 consider $\rho_{xz} = 0.8$, 0.4, 0.1, 0.02 respectively. Each figure contains for $n = 50$, 200, 2000 a set of 8 panels, so 24 in total. The 8 panels depict in two rows densities for the four cases $\rho_{x\varepsilon} = -0.3$, 0, 0.3, 0.6, with next to each other results for $\rho_{z\varepsilon} = 0$ and $\rho_{z\varepsilon} = 0.2$ respectively. All these panels present three or four densities, viz. for the actual empirical conditional distribution (*solid line*), for the unconditional finite sample distribution (*dashed line*) and for the available asymptotic approximations. For these, either both the conditional and unconditional distributions are available and equivalent (then we use *circles*), or if different and available the conditional asymptotic distribution (*squares*) and the unconditional (*diamonds*) are presented separately. These asymptotic distributions have been taken as $\hat{\beta}_{IV} \overset{a}{\sim} \mathcal{N}(\beta + \ddot{\beta}_{IV}, n^{-1}V)$, where V is given by (21) or (22) respectively.

In Figure 1.1 $\rho_{xz} = 0.8$, so the instrument is far from weak. The panels in columns 1 and 3 have $\rho_{z\varepsilon} = 0$, i.e. the instrument is valid and the standard asymptotic result applies for both conditional and unconditional distributions. In columns 2 and 4 $\rho_{z\varepsilon} = 0.2$, i.e. the instrument is (mildly) invalid, hence the IV estimator is inconsistent and the asymptotic approximations differ for the conditional and unconditional cases. From Figure 1.1 we find that for a relatively strong instrument the resemblance between corresponding actual and approximated distributions is quite remarkable even for sample sizes as small as $n = 50$, but for the sample sizes considered we already note some differences between conditional and unconditional distributions when $\rho_{z\varepsilon} \neq 0$. In Figure 1.2, where $\rho_{xz} = 0.4$, we note that due to the weaker instrument the actual distributions display some skewness, especially when sample size is small, except for $\rho_{x\varepsilon} = \rho_{z\varepsilon} = 0$. By its very nature, such skewness is completely overlooked by the first-order normal approximations. Hence, although any inconsistency and the dispersion of the actual distributions (which both increased due to the weaker instrument) are quite well approximated, the actual behavior in the tails of the distribution may be over or under stated by standard asymptotic approximations. In Figure 1.3 where the instrument is weak we note more pronounced skewness when both $\rho_{x\varepsilon} \neq 0$ and $\rho_{z\varepsilon} \neq 0$. The actual unconditional distribution is approximated well only for a really large sample. On the other hand, especially when $\rho_{x\varepsilon} \geq 0$ and $\rho_{z\varepsilon} > 0$, the conditional asymptotic approximation is remarkably accurate. For a much weaker instrument, Figure 1.4 shows that even for large sample size the actual distributions may become bimodal, unlike their normal first-order approximations. Only for $\rho_{x\varepsilon} = 0$ and $\rho_{z\varepsilon} = 0$ the standard asymptotic approximation captures the actual conditional distribution very well, even when the sample size is small. When the instrument is invalid and the sample size gets larger the approximation of the conditional distribution is found to be better than that of the unconditional one. When the instrument is very weak, whether valid or not, the actual distributions become so widely dispersed and wrongly located that they compare very unfavorably with that of a much stronger instrument, even if this were not valid.

The magnitude of the (median) bias of consistent IV and of OLS in relation to the effects of weakness of the instrument has been analyzed by many authors, see Sawa (1969) and (further references in) Hillier (2006). From Figure 1.4 we find that when the instrument is both invalid and weak then the finite sample distribution of the inconsistent IV estimator (both conditional and unconditional) is not centered at the pseudo-true-value. Surprisingly, it is often quite close 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 unconditional consistent IV estimator occurs for much more severe simultaneity than examined here, viz. for $\rho_{x\varepsilon} = 0.99$, whereas Phillips (2006) shows that it is omnipresent in the simple Keynesian model where simultaneity is always severe. We find that the conditional distribution may be bimodal for much smaller values of $\rho_{x\varepsilon}$. For moderate degree of simultaneity our findings also suggest that using instruments that are both weak and invalid may lead to bimodality in small samples, irrespective of whether one conditions or not.

From Figures 1.1 through 1.4 we conclude that, irrespective of whether the instruments are valid or not, one should avoid to use standard first-order 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 $\rho_{z\varepsilon} = 0.2$ columns of Figure 1.1. Not only are asymptotic approximations very accurate here, but also the actual distributions are much better concentrated around the true value than that of the consistent estimators depicted in the $\rho_{z\varepsilon} = 0$ columns of Figures 1.3 and 1.4. For $n = 50$ most of the inconsistent IV estimates for $\rho_{z\varepsilon} = 0.2$ and $\rho_{xz} = 0.8$ fall within the (0.9, 1.2) interval for all examined values of $\rho_{x\varepsilon}$, whereas this range is close to (0, 2) for the consistent estimates when $\rho_{xz} = 0.1$.

A wider picture of the major findings from Figures 1.1 through 1.4 on the conditional distribution will be given now by scanning the median absolute error over almost the full parameter space of this simple model. Figure 1.5 provides an overview in 16 diagrams of the (in)accuracy of the asymptotic conditional distribution of IV as an approximation to the actual conditional distribution in finite samples. Each of these diagrams (based on 50,000 replications) cover all compatible positive values of $\rho_{x\varepsilon}$ and ρ_{xz}, and relate to a particular relatively small sample size, viz. $n = 20$, 50, 100, 200, and a particular discrete value of $\rho_{z\varepsilon}$, viz. $\rho_{z\varepsilon} = 0$, 0.1, 0.3 and 0.6. The accuracy is expressed as $\log[\text{MAE}(\hat{\beta}_{IV})/\,\text{AMAE}(\hat{\beta}_{IV})]$. Hence, positive values (yellow-amber, or light grey) indicate larger absolute errors in finite samples than indicated by the asymptotic approximation and negative values (blue, or dark grey) indicate that standard asymptotics is too pessimistic about the absolute errors of $\hat{\beta}_{IV}$ in finite samples. This log-ratio is invariant regarding the value of $SN = \sigma_x^2/\sigma_\varepsilon^2$. We find that the approximation is quite good, even for $n = 20$, as long as the instrument is not very weak.

Figure 1.6 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 samples the absolute estimation errors committed by OLS are larger than those of IV only when $\rho_{x\varepsilon}$ and ρ_{xz} are both large. Of course, the area where IV beats OLS gets smaller for larger $\rho_{z\varepsilon}$. 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 ε_i, so that $\zeta_2 = 0$. This implies that

$$\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'\varepsilon} = \sigma_{\hat{x}\varepsilon} = \sigma_{xz_1}\sigma_{z_1\varepsilon} = \rho_{xz_1}\rho_{z_1\varepsilon}\sigma_x, \tag{33}$$

where, of course, $\rho_{z_1\varepsilon} = \zeta_1$. Now the various entries in the formula of Theorem 1 specialize to

$$\begin{aligned}
\Sigma_{X'X} &= \sigma_x^2 > 0, \\
\Sigma_{\hat{X}'\hat{X}} &= \sigma_{\hat{x}}^2 = \rho_{x\hat{x}}^2\sigma_x^2 > 0, \\
\ddot{\beta}_{GIV} &= \sigma_\varepsilon^2\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'\varepsilon} = \frac{\sigma_{\hat{x}\varepsilon}}{\sigma_{\hat{x}}^2} = \frac{\rho_{xz_1}\rho_{z_1\varepsilon}\sigma_x}{\rho_{x\hat{x}}^2\sigma_x^2} = \frac{\rho_{xz_1}\rho_{z_1\varepsilon}}{\rho_{x\hat{x}}^2\sigma_x}, \\
c_1 &= \sigma_\varepsilon^2\Sigma_{\varepsilon'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'\varepsilon} = \rho_{z_1\varepsilon}^2, \\
c_2 &= \sigma_\varepsilon^2\Sigma_{\varepsilon'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'X}\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'\varepsilon} = \rho_{\hat{x}\varepsilon}^2 = \frac{\rho_{xz_1}^2\rho_{z_1\varepsilon}^2}{\rho_{x\hat{x}}^2}, \\
c_3 &= \Sigma_{\varepsilon'X}\ddot{\beta}_{GIV} = \frac{\rho_{x\varepsilon}\rho_{\hat{x}\varepsilon}\sigma_x}{\sigma_{\hat{x}}} = \frac{\rho_{x\varepsilon}\rho_{\hat{x}\varepsilon}}{\rho_{x\hat{x}}} = \frac{\rho_{x\varepsilon}\rho_{xz_1}\rho_{z_1\varepsilon}}{\rho_{x\hat{x}}^2},
\end{aligned} \tag{34}$$

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: σ_x, $\rho_{x\hat{x}}$,

$\rho_{x\varepsilon}$, $\rho_{z_1\varepsilon}$ and ρ_{xz_1}. However, in the special case $\rho_{z_1\varepsilon} = 0$ the minimal set of parameters is just one dimensional, because $\rho_{x\hat{x}}\sigma_x$ suffices. For the general case we find

$$\begin{aligned} V_{GIV}^{CN} &= \frac{\sigma_\varepsilon^2}{\sigma_x^2\rho_{x\hat{x}}^2}\left[(1-\rho_{z_1\varepsilon}^4) + \rho_{z_1\varepsilon}\frac{\gamma_1}{\rho_{x\hat{x}}^2} + \rho_{z_1\varepsilon}^2\rho_{xz_1}^2\frac{\gamma_2}{\rho_{x\hat{x}}^4} + 2\rho_{z_1\varepsilon}^4\rho_{xz_1}^4\frac{\gamma_3}{\rho_{x\hat{x}}^6}\right] \\ &\quad + \frac{\sigma_\varepsilon^2}{\sigma_x^4\rho_{x\hat{x}}^4}\rho_{z_1\varepsilon}^4\left(1-\frac{\rho_{xz_1}^2}{\rho_{x\hat{x}}^2}\right)\left[1-2\rho_{z_1\varepsilon}^2\left(1-\frac{\rho_{xz_1}^2}{\rho_{x\hat{x}}^2}\right)\right], \end{aligned} \tag{35}$$

where

$$\begin{aligned} \gamma_1 &= \rho_{z_1\varepsilon}^3 - 2\rho_{x\varepsilon}\rho_{xz_1}(1+\rho_{z_1\varepsilon}^2 - 2\rho_{z_1\varepsilon}^4) - \rho_{xz_1}^2(\rho_{z_1\varepsilon} - 5\rho_{z_1\varepsilon}^3 + 2\rho_{z_1\varepsilon}^5), \\ \gamma_2 &= \rho_{x\varepsilon}^2 + 4\rho_{z_1\varepsilon}\rho_{x\varepsilon}\rho_{xz_1} - 4\rho_{z_1\varepsilon}^2[1+3\rho_{z_1\varepsilon}\rho_{x\varepsilon}\rho_{xz_1} - \rho_{x\varepsilon}^2 + \rho_{xz_1}^2(1-\rho_{z_1\varepsilon}^2)], \\ \gamma_3 &= 2 - (\rho_{x\varepsilon} - \rho_{z_1\varepsilon}\rho_{xz_1})(3\rho_{x\varepsilon} - \rho_{z_1\varepsilon}\rho_{xz_1}). \end{aligned}$$

Note that this variance is invariant to sign changes of the correlations as long as the sign of $\rho_{z_1\varepsilon}\rho_{x\varepsilon}\rho_{xz_1}$ is not affected, or when either $\rho_{x\varepsilon}$ or $\rho_{z_1\varepsilon}$ is zero. The sign of the inconsistency $\ddot{\beta}_{GIV}$ is determined by the sign of $\rho_{z_1\varepsilon}\rho_{xz_1}$. For given values of $\rho_{x\hat{x}}$ and $\rho_{z_1\varepsilon}$ the magnitude of $\ddot{\beta}_{GIV}$ is a multiple of ρ_{xz_1}, so it will be large when the invalid instrument is relatively strong. For the special case $\rho_{xz_1} = \rho_{x\hat{x}}$, i.e. the second instrument is orthogonal to x, the variance formula specializes to

$$\frac{\sigma_\varepsilon^2}{\sigma_x^2}\frac{(1-\rho_{z_1\varepsilon}^2)(\rho_{xz_1} - \rho_{z_1\varepsilon}\rho_{x\varepsilon})^2 + \rho_{z_1\varepsilon}^4(1-\rho_{x\varepsilon}^2)}{\rho_{xz_1}^4} \tag{36}$$

which, not surprisingly, corresponds to (21).

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 (24) and (25) can be extended as follows. Let now $\upsilon_i \sim \text{IID}(0, I_4)$. Again we take $\varepsilon_i = \upsilon_{i,1}$ and, again restricting ourselves to positive α_1 for symmetrically distributed υ_i, we have

$$\begin{aligned} x_i &= \xi\upsilon_{i,1} + \alpha_1\upsilon_{i,2} \\ &= \sigma_x[\rho_{x\varepsilon}\upsilon_{i,1} + \left|(1-\rho_{x\varepsilon}^2)^{1/2}\right|\upsilon_{i,2}], \end{aligned} \tag{37}$$

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

$$\begin{aligned} z_{i,1} &= \zeta_1\upsilon_{i,1} + \alpha_2\upsilon_{i,2} + \alpha_3\upsilon_{i,3} + \alpha_4\upsilon_{i,4}, \\ z_{i,2} &= \zeta_2\upsilon_{i,1} + \alpha_5\upsilon_{i,2} + \alpha_6\upsilon_{i,3} + \alpha_7\upsilon_{i,4}. \end{aligned}$$

These entail full generality, because they allow both instruments to have any correlation with the disturbance ε_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 $\upsilon_{i,1}$. This corresponds to taking $\zeta_2 = 0$ and re-interpreting α_5, α_6 and α_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. 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

$$\left.\begin{array}{l} z_{i,1} = \zeta_1 v_{i,1} + \alpha_2 v_{i,2} + \left|(1-\zeta_1^2-\alpha_2^2)^{1/2}\right| v_{i,3}, \\ z_{i,2} = \alpha_5 v_{i,2} + \alpha_6 v_{i,3} + \left|(1-\alpha_5^2-\alpha_6^2)^{1/2}\right| v_{i,4}. \end{array}\right\} \tag{38}$$

Due to the symmetry of v_i generality is maintained when we restrict ourselves to cases where particular coefficients are nonnegative. This extends to α_2 and α_5, because the space spanned by the instruments does not change by multiplying all elements by -1, yielding

$$0 \leq \alpha_2 \leq 1,\ 0 \leq \alpha_5 \leq 1. \tag{39}$$

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

$$\alpha_6 = -\alpha_2\alpha_5(1-\zeta_1^2-\alpha_2^2)^{-1/2}. \tag{40}$$

So, for given values of σ_x, $\rho_{x\varepsilon}$ and $\rho_{z_1\varepsilon} = \zeta_1$, we would be able to generate data according to (37) and (38) if we also knew α_2 and α_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

$$R_{xZ}^2 = \frac{\Sigma_{x'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x}}{\sigma_x^2}. \tag{41}$$

Note that

$$\rho_{x\hat{x}}^2 = \frac{(\Sigma_{x'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x})^2}{\sigma_x^2\sigma_{\hat{x}}^2} = \frac{\sigma_{\hat{x}}^2}{\sigma_x^2} = R_{xZ}^2, \tag{42}$$

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

$$\begin{aligned} \rho_{x\hat{x}}^2 &= \frac{\Sigma_{x'Z}\Sigma_{Z'x}}{\sigma_x^2} = \rho_{xz_1}^2 + (1-\rho_{x\varepsilon}^2)\alpha_5^2, \\ \rho_{xz_1} &= \rho_{x\varepsilon}\rho_{z_1\varepsilon} + \left|(1-\rho_{x\varepsilon}^2)^{1/2}\right|\alpha_2. \end{aligned}$$

From these we can express the (nonnegative) values of α_5 and α_2 as

$$\alpha_5 = \left|\left(\frac{\rho_{x\hat{x}}^2 - \rho_{xz_1}^2}{1-\rho_{x\varepsilon}^2}\right)^{1/2}\right| \tag{43}$$

and

$$\alpha_2 = \left|\frac{\rho_{xz_1} - \rho_{x\varepsilon}\rho_{z_1\varepsilon}}{(1-\rho_{x\varepsilon}^2)^{1/2}}\right|, \tag{44}$$

from which α_6 follows directly by evaluating (40).

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 σ_x, $\rho_{x\hat{x}}$, $\rho_{x\varepsilon}$, ρ_{xz_1} and $\rho_{z_1\varepsilon}$. Here again, these are found to be the data moments that characterize the asymptotic distribution. They determine ξ and α_1 as in (32) and α_2, α_5 and α_6 via (44), (43) and (40), respectively. We may restrict ourselves to cases where $\rho_{x\hat{x}} > 0$ (since the coefficients of the simulation design are just determined by $\rho_{x\hat{x}}^2$). Note that the coefficients of the data generation process, notably α_2, are unaffected (thus yielding the same distribution of $\hat{\beta}_{GIV}$) if both $\rho_{z_1\varepsilon}$ and ρ_{xz_1} are changed in sign. Therefore, we will only examine cases with $\rho_{xz_1} \geq 0$. However, we shall also just examine nonnegative values of $\rho_{z_1\varepsilon}$, because when the distribution of ε is symmetric changing the signs of both $\rho_{z_1\varepsilon}$ and $\rho_{x\varepsilon}$ 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 $\rho_{z_1\varepsilon} \leq 0$ is equivalent with the distribution of $-(\hat{\beta}_{GIV} - \beta)$ for $-\rho_{z_1\varepsilon} \geq 0$ and $-\rho_{x\varepsilon}$, because: If we change the signs of $\rho_{z_1\varepsilon}$, $\rho_{x\varepsilon}$ and all ε_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 $\zeta_1 = \rho_{z_1\varepsilon} = 0$ in (38). Thus full generality is maintained by making $z_{i,2}$ independent of $\upsilon_{i,3}$, 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

$$\begin{aligned} z_{i,1} &= \alpha_2\upsilon_{i,2} + (1-\alpha_2^2)^{1/2}\upsilon_{i,3}, \\ z_{i,2} &= \upsilon_{i,4}. \end{aligned} \tag{45}$$

These imply $\rho_{xz_1} = \rho_{x\hat{x}}$ and $\alpha_2 = \left|\rho_{x\hat{x}}(1-\rho_{x\varepsilon}^2)^{-1/2}\right|$ instead of (44). Hence, when $\zeta = 0$ the finite sample distribution is determined by just 3 parameters (viz. σ_x, $\rho_{x\varepsilon}$ and $\rho_{x\hat{x}}$) instead of 5 (apart from n), whereas the limiting distribution just depends on $\sigma_{\hat{x}}^2 = \rho_{x\hat{x}}^2\sigma_x^2$.

In all calculations we again fixed $SN = 10$ (which here too has only a multiplicative effect, i.e. just affects the scale of the densities). For the conditional distribution we used the same draws for $\upsilon_{i,2}$ and $\upsilon_{i,3}$ as in the previous example, and added a $\upsilon_{i,4}$ and used the same strategy regarding normalizing and orthogonalizing them. Since $l - k = 1$, GIV does have a finite first moment now, and for $l > k$ we do not have an explicit formula for the unconditional asymptotic variance when $\rho_{z_1\varepsilon} \neq 0$. Figures 2.1 through 2.4 present illustrative densities for $\rho_{x\hat{x}} = 0.8$, 0.4, 0.1, 0.02 respectively. Again we examine $n = 50$, 200, 2000 and as before we chose values $\rho_{x\varepsilon} = -0.3$, 0.0, 0.3, 0.6 and $\rho_{z_1\varepsilon} = \zeta_1 = 0.0$, 0.2, whereas we fixed $\rho_{xz_1} = \rho_{x\hat{x}}/2$. Hence, z_1 provides half of the joint strength of the instruments, and we omitted cases where $\rho_{xz_1} = \rho_{x\hat{x}}$ (and $\rho_{xz_2} = 0$). Although we already established that the latter cases yield similar asymptotic results as obtained for $k = l = 1$, from the simulations we found that the finite sample densities do differ slightly from the "no finite moments" case, and more so for a weaker instrument z_1.

Figure 2.1 shows that for strong instruments the available asymptotic approximations are reasonably accurate, but slightly less so for small sample size in Figure 2.2 where the instruments are weaker. For really weak instruments we note from Figures 2.3 and 2.4 that the asymptotic approximations for the conditional distribution are very accurate when $\rho_{x\varepsilon} = 0$, but are much too pessimistic regarding the inter quartile range otherwise. Forchini (2006) suspects bimodality in the overidentified model when the instruments are valid but weak; for $\rho_{x\hat{x}} = 0.02$ we establish that to be the case, especially for the conditional distribution, and also for the unconditional distribution when there is an invalid instrument. When both instruments are valid and $\rho_{x\hat{x}} = 0.02$ the $l = 2$ case produces estimators which are substantially more efficient than the corresponding $l = 1$ estimators in Figure 1.4. 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.

Finally, we look again at median absolute error results. Figure 2.5 gives for this model a more global impression of the remarkable accuracy of the conditional asymptotic approximation even in very small samples. However, the approximation is much too pessimistic about the actual estimation errors when overall instrument strength $\rho_{x\hat{x}}$ is very low and simultaneity $\rho_{x\varepsilon}$ severe, the more so for serious instrument invalidity $\rho_{z_1\varepsilon}$. Figure 2.6 makes comparisons with OLS. 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 (2005) that 2SLS is the preferred estimator when $n \geq 100$ and $\rho^2_{x\hat{x}} \geq 0.1$. They arrive at this conclusion by comparing second-order asymptotic approximations to MSE, whereas ours are obtained from simulating the actual distributions.

4 Conclusions

The asymptotic variance of a consistent method of moments estimator for the coefficients of a linear model does just involve the variance of the disturbances and an expression involving population data moments. For these moments consistent estimates obtained from the sample data are being substituted when estimating this variance. Usually, it is inconsequential whether or not one considers exogenous regressors as fixed or as random realizations. For models where all observations are IID we demonstrated that when the estimator is inconsistent due to exploiting invalid moment conditions, a relevant random vector series (to which the central limit theorem can be applied in order to obtain the limiting distribution) has elements which are no longer linear but quadratic in the disturbance. Therefore the asymptotic variance also involves third and fourth order moments of the disturbances. Moreover, conditioning on exogenous variables generally has consequences for the variance expression of the limiting distribution of inconsistent estimators, which, although still normal, are now centered at the pseudo-true-value (true coefficient plus inconsistency). It remains a challenge to generalize the explicit results obtained here for IID data and IV estimators, to time-series models and GMM estimators, such as for 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 (but low dimensional) 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 in general yield unattractive biased estimators with widely dispersed and often bimodal distributions, is abandoned altogether. We have shown that exclusively exploiting strong instruments, even if these constitute invalid instruments, often 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 very accurate approximation to the finite sample distribution. But, to render this approximation feasible one requires information on the simultaneity parameter ξ and the instrument invalidity parameter ζ. That seems hard to obtain, and if such information was available other estimators than those obtained by minimizing an (in)appropriate method of moments criterion function might be better for producing accurate inference on the coefficient β. However, the now available explicit though unfeasible conditional and unconditional asymptotic distributions of possibly inconsistent estimators can always be used to produce a sensitivity analysis of OLS or IV inference under alternative assumptions regarding the degree of simultaneity and of instrument invalidity.

References

Andrews, D.W.K., Stock, J.H., 2007. Inference with weak instruments. In R. Blundell, W.K. Newey, T. Persson (Eds.), *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Vol. 3, 122-173. Cambridge, UK: Cambridge University Press.

Ashley, R., 2009. Assessing the credibility of instrumental variables inference with imperfect instruments via sesitivity analysis. *Journal of Applied Econometrics* 24, 325-337.

Berkowitz, D., Caner, M., Ying Fang, 2008. Are "nearly exogenous instruments" reliable? *Economics Letters* 101, 20-23.

Berkowitz, D., Caner, M., Ying Fang, 2012. The validity of instruments revisited. *Journal of Econometrics* 166, 255-266.

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variable estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90, 443-450.

Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. *Econometrica* 69, 1161-1191.

Forchini, G., 2006. On the bimodality of the exact distribution of the TSLS estimator. *Econometric Theory* 22, 932-946.

Goldberger, A.S., 1964. *Econometric Theory.* John Wiley & Sons. New York.

Guggenberger, P., 2012. On the asymptotic size distortion of tests when instruments locally violate the exogeneity assumption. *Econometric Theory* 28, 387-421.

Hahn, J., Hausman, J.A., 2005. IV estimation with valid and invalid instruments. *Annales d'Économie et de Statistique* 79/80, 25-57.

Hahn, J., Inoue, A., 2002. A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators. *Econometric Reviews* 21, 309-336.

Hale, C., Mariano, R.S., Ramage, J.G., 1980. Finite sample analysis of misspecification in simultaneous equation models. *Journal of the American Statistical Association* 75, 418-427.

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

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251-1271.

Hendry, D.F., 1979. The behaviour of inconsistent instrumental variables estimators in dynamic systems with autocorrelated errors. *Journal of Econometrics* 9, 295-314.

Hendry, D.F., 1982. A reply to professors Maasoumi and Phillips. *Journal of Econometrics* 19, 203-213.

Hillier, G., 2006. Yet more on the exact properties of IV estimators. *Econometric Theory* 22, 913-931.

Joseph, A.S., Kiviet, J.F., 2005. Viewing the relative efficiency of IV estimators in models with lagged and instantaneous feedbacks. *Journal of Computational Statistics and Data Analysis* 49, 417-444.

Kiviet, J.F., 2011. Monte Carlo Simulation for Econometricians, *Foundations and Trends in Econometrics* 5, 1-181.

Kiviet, J.F., 2013. Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *The Econometrics Journal* 16, S24-S59.

Kiviet, J.F., Niemczyk, J., 2007. The asymptotic and finite sample distributions of OLS and simple IV in simultaneous equations. *Journal of Computational Statistics and Data Analysis* 51, 3296-3318.

Kiviet, J.F., Niemczyk, J., 2012. Comparing the asymptotic and empirical (un)conditional distributions of OLS and IV in a linear static simultaneous equation. *Journal of Computational Statistics & Data Analysis* 56, 3567-3586.

Knight, J.L., 1982. A note on finite sample analysis of misspecifcation in simultaneous equation models. *Economics Letters* 9, 275-279.

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

Maddala, G.S., Jeong, J., 1992. On the exact small sample distribution of the instrumental variable estimator. *Econometrica* 60, 181-183.

Mariano, R.S., 1973. Approximations to the distribution functions of the ordinary least-squares and two-stage least-squares estimators in the case of two included endogenous variables. *Econometrica* 41, 67-77.

Murray, M.P., 2006. Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives* 20, 111-132.

Phillips, P.C.B., 1980. The exact distribution of instrumental variable estimators in an equation containing $n+1$ endogenous variables. *Econometrica* 48, 861-878.

Phillips, P. C. B., 1983. Exact small sample theory in the simultaneous equations model. In Z. Griliches, M. D. Intriligator (Eds.), *Handbook of Econometrics*, Vol. 1, 449-516. Amsterdam: North-Holland.

Phillips, P.C.B., 2006. A remark on bimodality and weak instrumentation in structural equation estimation. *Econometric Theory* 22, 947-960.

Phillips, P.C.B., Wickens, M.R., 1978. *Exercises in Econometrics*. Philip Allen and Ballinger, Cambridge MA.

Rhodes, G.F., Westbrook, M.D., 1981. A study of estimator densities and performance under misspecifcation. *Journal of Econometrics* 16, 311-337.

Rothenberg, T.J., 1972. The asymptotic distribution of the least squares estimator in the errors in variables model. Unpublished mimeo.

Sawa, T., 1969. The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *Journal of the American Statistical Association* 64, 923-937.

Skeels, C. L., 1995. Instrumental variables estimation in misspecified single equations. *Econometric Theory* 11, 498-529.

Small, D.S., 2007. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Society* 102, 1049-1058.

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

A Proof of Theorem 1

Centering around $\beta^*_{n,GIV}$ defined in (12) we obtain

$$\sqrt{n}(\hat{\beta}_{GIV} - \beta^*_{n,GIV}) = \sqrt{n}[(\hat{X}'\hat{X})^{-1}\hat{X}'\varepsilon - \sigma^2_\varepsilon S^{-1}_{\hat{X}'\hat{X}}P'\zeta]. \tag{46}$$

To find the limiting distribution we shall rewrite the right-hand side of (46) such that we can invoke the Lemma given at the end of Section 2. Below we first show that (46) can be rewritten as

$$\sqrt{n}(\hat{\beta}_{GIV} - \beta^*_{n,GIV}) = (\frac{1}{n}\hat{X}'\hat{X})^{-1}\frac{1}{\sqrt{n}}[W'\varepsilon + \omega(\varepsilon'\varepsilon - n\sigma^2_\varepsilon)] + o_p(1), \tag{47}$$

for appropriate $n \times k$ matrix W, with $E(W'\varepsilon) = 0$, and nonrandom $k \times 1$ vector ω. Invoking also a theorem often attributed to Cramér, the Lemma then yields

$$\sqrt{n}(\hat{\beta}_{GIV} - \beta^*_{n,GIV}) \xrightarrow{d} N\left[0, \sigma^2_\varepsilon \Sigma^{-1}_{\hat{X}'\hat{X}}\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}W'W + 2\sigma^2_\varepsilon\omega\omega'\right)\Sigma^{-1}_{\hat{X}'\hat{X}}\right], \tag{48}$$

upon assuming $\mu_3 = 0$ and $\mu_4 = 3$.

We first set out to rewrite (46) in the form (47). We easily obtain

$$\begin{aligned}\sqrt{n}(\hat{\beta}_{GIV} - \beta^*_{n,GIV}) &= \sqrt{n}(\hat{X}'\hat{X})^{-1}[\hat{\Pi}'(Z'\varepsilon - n\sigma^2_\varepsilon\zeta) + n\sigma^2_\varepsilon\hat{\Pi}'\zeta - \sigma^2_\varepsilon\hat{X}'\hat{X}S^{-1}_{\hat{X}'\hat{X}}P'\zeta] \\ &= (\frac{1}{n}\hat{X}'\hat{X})^{-1}\frac{1}{\sqrt{n}}\{\hat{\Pi}'[\bar{Z}'\varepsilon + (\varepsilon'\varepsilon - n\sigma^2_\varepsilon)\zeta] + n\sigma^2_\varepsilon[\hat{\Pi}' - \frac{1}{n}\hat{X}'\hat{X}S^{-1}_{\hat{X}'\hat{X}}P']\zeta\}. \end{aligned} \tag{49}$$

For the second expression between square brackets in the final line of (49) we find

$$\begin{aligned}&\hat{\Pi}' - \frac{1}{n}\hat{X}'\hat{X}S^{-1}_{\hat{X}'\hat{X}}S_{X'Z}S^{-1}_{Z'Z} \\ =\ & \hat{\Pi}'(S_{Z'Z} - \frac{1}{n}Z'Z)S^{-1}_{Z'Z} + (\frac{1}{n}X'Z - \frac{1}{n}\hat{X}'\hat{X}S^{-1}_{\hat{X}'\hat{X}}S_{X'Z})S^{-1}_{Z'Z} \\ =\ & \hat{\Pi}'(S_{Z'Z} - \frac{1}{n}Z'Z)S^{-1}_{Z'Z} + (\frac{1}{n}X'Z - S_{X'Z})S^{-1}_{Z'Z} + (S_{\hat{X}'\hat{X}} - \frac{1}{n}\hat{X}'\hat{X})S^{-1}_{\hat{X}'\hat{X}}S_{X'Z}S^{-1}_{Z'Z}. \end{aligned} \tag{50}$$

Its third term contains a factor which can be decomposed as

$$\begin{aligned}&S_{\hat{X}'\hat{X}} - \frac{1}{n}\hat{X}'\hat{X} \\ =\ & S_{X'Z}S^{-1}_{Z'Z}S_{Z'X} - \hat{\Pi}'\frac{1}{n}Z'X \\ =\ & (S_{X'Z}S^{-1}_{Z'Z} - \hat{\Pi}')S_{Z'X} - \hat{\Pi}'(\frac{1}{n}Z'X - S_{Z'X}) \\ =\ & \{(S_{X'Z} - \frac{1}{n}X'Z)S^{-1}_{Z'Z} - \frac{1}{n}X'Z[(\frac{1}{n}Z'Z)^{-1} - S^{-1}_{Z'Z}]\}S_{Z'X} - \hat{\Pi}'(\frac{1}{n}Z'X - S_{Z'X}) \\ =\ & \{(S_{X'Z} - \frac{1}{n}X'Z)S^{-1}_{Z'Z} - \hat{\Pi}'[(S_{Z'Z} - \frac{1}{n}Z'Z)S^{-1}_{Z'Z}]\}S_{Z'X} - \hat{\Pi}'(\frac{1}{n}Z'X - S_{Z'X}). \end{aligned} \tag{51}$$

Now substituting the decompositions obtained in (50) and (51) into the expression within curly brackets in the final line of (49), and next using $\hat{\Pi} = \Pi + O_p(n^{-1/2})$, $P = \Pi + O_p(n^{-1/2})$ and $\ddot{\beta}_{n,GIV} = \ddot{\beta}_{GIV} +$

$O_p(n^{-1/2})$, we obtain

$$
\begin{aligned}
& \hat{\Pi}'[\bar{Z}'\varepsilon + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\zeta] + n\sigma_\varepsilon^2[\hat{\Pi}' - \frac{1}{n}\hat{X}'\hat{X}S_{\hat{X}'\hat{X}}^{-1}S_{X'Z}S_{Z'Z}^{-1}]\zeta \\
= \; & \hat{\Pi}'[\bar{Z}'\varepsilon + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\zeta] \\
& + n\sigma_\varepsilon^2\hat{\Pi}'(S_{Z'Z} - \frac{1}{n}Z'Z)S_{Z'Z}^{-1}\zeta + n\sigma_\varepsilon^2(\frac{1}{n}X'Z - S_{X'Z})S_{Z'Z}^{-1}\zeta + n(S_{\hat{X}'\hat{X}} - \frac{1}{n}\hat{X}'\hat{X})\ddot{\beta}_{n,GIV}
\end{aligned}
$$

$$
\begin{aligned}
= \; & \Pi'[\bar{Z}'\varepsilon + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\zeta] - n\sigma_\varepsilon^2\Pi'(\frac{1}{n}Z'Z - S_{Z'Z})S_{Z'Z}^{-1}\zeta + n\sigma_\varepsilon^2(\frac{1}{n}X'Z - S_{X'Z})S_{Z'Z}^{-1}\zeta \\
& - n(\frac{1}{n}X'Z - S_{X'Z})\Pi\ddot{\beta}_{GIV} + n\Pi'(\frac{1}{n}Z'Z - S_{Z'Z})\Pi\ddot{\beta}_{GIV} - n\Pi'(\frac{1}{n}Z'X - S_{Z'X})\ddot{\beta}_{GIV} + O_p(1).
\end{aligned} \tag{52}
$$

All the six terms of (52) are $O_p(n^{1/2})$ and will, after premultiplication by $n^{-1/2}(\frac{1}{n}\hat{X}'\hat{X})^{-1}$, affect the limiting distribution of $\hat{\beta}_{GIV}$.

In deriving in a more general context such a limiting distribution Hendry (1979, formula 16) takes into account only terms similar to the first and sixth of (52). Hence, he incorrectly neglects four of the six $O_p(n^{1/2})$ terms of (52), as demonstrated by Maasoumi and Phillips (1982, formula 17), who also show that this only affects the $l > k$ case. From (52) it can easily be seen that in the special case $l = k$ terms 2 and 5, and 3 and 4 cancel. Maasoumi and Phillips (1982) and Hendry (1982) provide algebra by which, exploiting formulas related to (49) and (52), the unconditional limiting distribution of the GIV estimator can be obtained for one equation of a particular simultaneous dynamic system with first order serial correlation involving an arbitrary number of endogenous and autoregressive predetermined variables, which are all assumed to be normally distributed.

We will here obtain an explicit result for the particular situation described by our Frameworks A and B, while conditioning on the (latent) predetermined variables. Hence, we do not make an assumption on the distribution of $\bar{X}$ and $\bar{Z}$, but make the substitutions

$$
\begin{aligned}
\frac{1}{n}Z'Z - S_{Z'Z} &= \frac{1}{n}\bar{Z}'\bar{Z} + \frac{1}{n}\bar{Z}'\varepsilon\zeta' + \frac{1}{n}\zeta\varepsilon'\bar{Z} + \frac{1}{n}\zeta\varepsilon'\varepsilon\zeta' - \frac{1}{n}\bar{Z}'\bar{Z} - \sigma_\varepsilon^2\zeta\zeta' \\
&= \frac{1}{n}\bar{Z}'\varepsilon\zeta' + \frac{1}{n}\zeta\varepsilon'\bar{Z} + \frac{1}{n}(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\zeta\zeta', \\
\frac{1}{n}X'Z - S_{X'Z} &= \frac{1}{n}\bar{X}'\varepsilon\zeta' + \frac{1}{n}\xi\varepsilon'\bar{Z} + \frac{1}{n}(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi\zeta'.
\end{aligned}
$$

in the final expression given for (52), which yields

$$
\begin{aligned}
& \Pi'\bar{Z}'\varepsilon + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\Pi'\zeta - \sigma_\varepsilon^2\Pi'\bar{Z}'\varepsilon\zeta'S_{Z'Z}^{-1}\zeta - \sigma_\varepsilon^2\Pi'\zeta\varepsilon'\bar{Z}S_{Z'Z}^{-1}\zeta - \sigma_\varepsilon^2\Pi'(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\zeta\zeta'S_{Z'Z}^{-1}\zeta \\
& + \sigma_\varepsilon^2\bar{X}'\varepsilon\zeta'S_{Z'Z}^{-1}\zeta + \sigma_\varepsilon^2\xi\varepsilon'\bar{Z}S_{Z'Z}^{-1}\zeta + \sigma_\varepsilon^2(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi\zeta'S_{Z'Z}^{-1}\zeta + \Pi'\bar{Z}'\varepsilon\zeta'\Pi\ddot{\beta}_{GIV} + \Pi'\zeta\varepsilon'\bar{Z}\Pi\ddot{\beta}_{GIV} \\
& + \Pi'(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\zeta\zeta'\Pi\ddot{\beta}_{GIV} - \bar{X}'\varepsilon\zeta'\Pi\ddot{\beta}_{GIV} - \xi\varepsilon'\bar{Z}\Pi\ddot{\beta}_{GIV} - (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi\zeta'\Pi\ddot{\beta}_{GIV} - \Pi'\zeta\varepsilon'\bar{X}\ddot{\beta}_{GIV} \\
& - \Pi'\bar{Z}'\varepsilon\xi'\ddot{\beta}_{GIV} - \Pi'(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\zeta\xi'\ddot{\beta}_{GIV} + O_p(1).
\end{aligned}
$$

This can be further simplified by using

$$
\begin{aligned}
c_1 &\equiv \sigma_\varepsilon^2\zeta'\Sigma_{Z'Z}^{-1}\zeta, \\
c_2 &\equiv \zeta'\Pi\ddot{\beta}_{GIV} = \sigma_\varepsilon^2\zeta'\Sigma_{Z'Z}^{-1}\Sigma_{Z'X}\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\zeta, \\
c_3 &\equiv \xi'\ddot{\beta}_{GIV},
\end{aligned}
$$

giving

$$
\begin{aligned}
& \Pi'\bar{Z}'\varepsilon + (\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\Pi'\zeta - c_1\Pi'\bar{Z}'\varepsilon - \sigma_\varepsilon^2\Pi'\zeta\zeta'\Sigma_{Z'Z}^{-1}\bar{Z}'\varepsilon - c_1(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\Pi'\zeta \\
& + c_1\bar{X}'\varepsilon + \sigma_\varepsilon^2\xi\zeta'\Sigma_{Z'Z}^{-1}\bar{Z}'\varepsilon + c_1(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi + c_2\Pi'\bar{Z}'\varepsilon + \Pi'\zeta\ddot{\beta}_{GIV}'\Pi'\bar{Z}'\varepsilon + c_2(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\Pi'\zeta \\
& - c_2\bar{X}'\varepsilon - \xi\ddot{\beta}_{GIV}'\Pi'\bar{Z}'\varepsilon - c_2(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\xi - \Pi'\zeta\ddot{\beta}_{GIV}'\bar{X}'\varepsilon - c_3\Pi'\bar{Z}'\varepsilon - c_3(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\Pi'\zeta + o_p(n^{1/2})
\end{aligned}
$$

$$
\begin{aligned}
= \; & [c_4 I_k - \Pi'\zeta\ddot{\beta}_{GIV}']\bar{X}'\varepsilon + [c_5\Pi' + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}_{GIV}'\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}]\bar{Z}'\varepsilon \\
& + (c_4\xi + c_5\Pi'\zeta)(\varepsilon'\varepsilon - n\sigma_\varepsilon^2) + O_p(1),
\end{aligned} \tag{53}
$$

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

Note that (53) re-expresses the factor in curly brackets in the final line of (49). We want to derive its limiting distribution after scaling by the factor $1/\sqrt{n}$, so we may neglect the remainder term. Defining the $k \times k$ matrix Θ_1, the $k \times l$ matrix Θ_2 and the $k \times 1$ vector ω, such that

$$\begin{aligned} \Theta_1 &\equiv c_4 I_k - \Pi'\zeta\ddot{\beta}'_{GIV}, \\ \Theta_2 &\equiv c_5\Pi' + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}, \\ \omega &\equiv c_4\xi + c_5\Pi'\zeta, \end{aligned} \tag{54}$$

we can now invoke the Lemma with

$$W' = \Theta_1\bar{X}' + \Theta_2\bar{Z}'. \tag{55}$$

For the case $\mu_3 = 0$ and $\mu_4 = 3$ we then obtain, conditioning on $\bar{X}$ and $\bar{Z}$, the limiting distribution

$$\frac{1}{\sqrt{n}}\left[\Theta_1\bar{X}'\varepsilon + \Theta_2\bar{Z}'\varepsilon + \omega(\varepsilon'\varepsilon - n\sigma_\varepsilon^2)\right] \xrightarrow{d} \mathcal{N}(0, \sigma_\varepsilon^2 V_0),$$

where

$$V_0 = \Theta_1\Sigma_{\bar{X}'\bar{X}}\Theta_1' + \Theta_2\Sigma_{\bar{Z}'\bar{Z}}\Theta_2' + 2\sigma_\varepsilon^2\omega\omega' + \Theta_1\Sigma_{\bar{X}'\bar{Z}}\Theta_2' + \Theta_2\Sigma_{\bar{Z}'\bar{X}}\Theta_1'. \tag{56}$$

In evaluating V_0 we make use of

$$\begin{aligned} \Sigma_{\bar{Z}'\bar{Z}} &\equiv \operatorname{plim} n^{-1}\bar{Z}'\bar{Z} = \Sigma_{Z'Z} - \sigma_\varepsilon^2\zeta\zeta' \\ \Sigma_{\bar{X}'\bar{X}} &\equiv \operatorname{plim} n^{-1}\bar{X}'\bar{X} = \Sigma_{X'X} - \sigma_\varepsilon^2\xi\xi' \\ \Sigma_{\bar{Z}'\bar{X}} &\equiv \operatorname{plim} n^{-1}\bar{Z}'\bar{X} = \Sigma_{Z'X} - \sigma_\varepsilon^2\zeta\xi' \\ \Sigma_{\bar{X}'\bar{Z}} &\equiv \operatorname{plim} n^{-1}\bar{X}'\bar{Z} = \Sigma'_{\bar{Z}'\bar{X}} \end{aligned}$$

and find that V_0 can be expressed as

$$\begin{aligned} &[c_4 I_k - \Pi'\zeta\ddot{\beta}'_{GIV}](\Sigma_{X'X} - \sigma_\varepsilon^2\xi\xi')[c_4 I_k - \ddot{\beta}_{GIV}\zeta'\Pi] \\ &+[c_5\Pi' + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}](\Sigma_{Z'Z} - \sigma_\varepsilon^2\zeta\zeta')[c_5\Pi + \Sigma_{Z'Z}^{-1}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\xi' - \zeta'\Pi)] \\ &+2\sigma_\varepsilon^2[c_4\xi + c_5\Pi'\zeta][c_4\xi' + c_5\zeta'\Pi] \\ &+[c_4 I_k - \Pi'\zeta\ddot{\beta}'_{GIV}](\Sigma_{X'Z} - \sigma_\varepsilon^2\xi\zeta')[c_5\Pi + \Sigma_{Z'Z}^{-1}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\xi' - \zeta'\Pi)] \\ &+[c_5\Pi' + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}](\Sigma_{Z'X} - \sigma_\varepsilon^2\zeta\xi')[c_4 I_k - \ddot{\beta}_{GIV}\zeta'\Pi]. \end{aligned}$$

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

$$\begin{aligned} & c_4^2(\Sigma_{X'X} - \sigma_\varepsilon^2\xi\xi') - c_4(\Sigma_{X'X} - \sigma_\varepsilon^2\xi\xi')\ddot{\beta}_{GIV}\zeta'\Pi \\ & -c_4\Pi'\zeta\ddot{\beta}'_{GIV}(\Sigma_{X'X} - \sigma_\varepsilon^2\xi\xi') + \Pi'\zeta\ddot{\beta}'_{GIV}(\Sigma_{X'X} - \sigma_\varepsilon^2\xi\xi')\ddot{\beta}_{GIV}\zeta'\Pi \\ = \; & c_4^2\Sigma_{X'X} - \sigma_\varepsilon^2 c_4^2\xi\xi' - c_4\Sigma_{X'X}\ddot{\beta}_{GIV}\zeta'\Pi + \sigma_\varepsilon^2 c_3 c_4\xi\zeta'\Pi \\ & -c_4\Pi'\zeta\ddot{\beta}'_{GIV}\Sigma_{X'X} + \sigma_\varepsilon^2 c_3 c_4\Pi'\zeta\xi' + (\ddot{\beta}'_{GIV}\Sigma_{X'X}\ddot{\beta}_{GIV} - \sigma_\varepsilon^2 c_3^2)\Pi'\zeta\zeta'\Pi, \end{aligned}$$

the second is

$$\begin{aligned} & [c_5\Pi'\Sigma_{Z'Z} + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})][c_5\Pi + \Sigma_{Z'Z}^{-1}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\xi' - \zeta'\Pi)] \\ & -\sigma_\varepsilon^2[c_5\Pi'\zeta\zeta' + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\zeta\zeta'][c_5\Pi + \Sigma_{Z'Z}^{-1}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\xi' - \zeta'\Pi)] \end{aligned}$$

$$\begin{aligned} = \; & c_5^2\Pi'\Sigma_{Z'Z}\Pi + c_5(\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta'\Pi - \ddot{\beta}'_{GIV}\Sigma_{X'Z}\Pi) + (\sigma_\varepsilon^2 c_5\Pi'\zeta - c_5\Pi'\Sigma_{Z'X}\ddot{\beta}_{GIV})(\xi' - \zeta'\Pi) \\ & +(\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta'\Sigma_{Z'Z}^{-1} - \ddot{\beta}'_{GIV}\Pi')(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\xi' - \zeta'\Pi) \\ & -\sigma_\varepsilon^2 c_5[c_5\Pi'\zeta\zeta'\Pi + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\zeta\zeta'\Pi] \\ & -\sigma_\varepsilon^2[c_5\Pi'\zeta\zeta' + (\xi - \Pi'\zeta)(c_1\zeta' - c_2\zeta')][(\sigma_\varepsilon^2\Sigma_{Z'Z}^{-1}\zeta - \Pi\ddot{\beta}_{GIV})(\xi' - \zeta'\Pi)] \end{aligned}$$

$$
\begin{aligned}
&= c_5^2\Pi'\Sigma_{Z'Z}\Pi + \sigma_\varepsilon^2 c_5\xi\zeta'\Pi - \sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi - c_5\xi\ddot\beta_{GIV}'\Sigma_{X'Z}\Pi + c_5\Pi'\zeta\ddot\beta_{GIV}'\Sigma_{X'Z}\Pi \\
&\quad +\sigma_\varepsilon^2 c_5\Pi'\zeta\xi' - c_5\Pi'\Sigma_{Z'X}\ddot\beta_{GIV}\xi' - \sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi + c_5\Pi'\Sigma_{Z'X}\ddot\beta_{GIV}\zeta'\Pi \\
&\quad +(\xi-\Pi'\zeta)(\sigma_\varepsilon^2\zeta'\Sigma_{Z'Z}^{-1} - \ddot\beta_{GIV}'\Pi')(\sigma_\varepsilon^2\zeta\xi' - \Sigma_{Z'X}\ddot\beta_{GIV}\xi') \\
&\quad -(\xi-\Pi'\zeta)(\sigma_\varepsilon^2\zeta'\Sigma_{Z'Z}^{-1} - \ddot\beta_{GIV}'\Pi')(\sigma_\varepsilon^2\zeta\zeta'\Pi - \Sigma_{Z'X}\ddot\beta_{GIV}\zeta'\Pi) \\
&\quad -\sigma_\varepsilon^2 c_5[c_5\Pi'\zeta\zeta'\Pi + (\xi-\Pi'\zeta)(c_1\zeta'\Pi - c_2\zeta'\Pi)] \\
&\quad -\sigma_\varepsilon^2[c_5\Pi'\zeta\zeta' + (\xi-\Pi'\zeta)(c_1\zeta' - c_2\zeta')](\sigma_\varepsilon^2\Sigma_{Z'Z}^{-1}\zeta\xi' - \Pi\ddot\beta_{GIV}\xi') \\
&\quad +\sigma_\varepsilon^2[c_5\Pi'\zeta\zeta' + (\xi-\Pi'\zeta)(c_1\zeta' - c_2\zeta')](\sigma_\varepsilon^2\Sigma_{Z'Z}^{-1}\zeta\zeta'\Pi - \Pi\ddot\beta_{GIV}\zeta'\Pi)
\end{aligned}
$$

$$
\begin{aligned}
&= c_5\Pi'\Sigma_{Z'Z}\Pi + \sigma_\varepsilon^2 c_5\xi\zeta'\Pi - \sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi - \sigma_\varepsilon^2 c_5\xi\zeta'\Pi + \sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi \\
&\quad +\sigma_\varepsilon^2 c_5\Pi'\zeta\xi' - \sigma_\varepsilon^2 c_5\Pi'\zeta\xi' - \sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi + \sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi \\
&\quad +(\xi-\Pi'\zeta)(\sigma_\varepsilon^2 c_1\xi' - \sigma_\varepsilon^2 c_2\xi') - (\xi-\Pi'\zeta)(\sigma_\varepsilon^2 c_2\xi' - \sigma_\varepsilon^2\ddot\beta_{GIV}'\Pi'\zeta\xi') \\
&\quad -(\xi-\Pi'\zeta)(\sigma_\varepsilon^2 c_1\zeta'\Pi - \sigma_\varepsilon^2 c_2\zeta'\Pi) + (\xi-\Pi'\zeta)(\sigma_\varepsilon^2 c_2\zeta'\Pi - \sigma_\varepsilon^2\ddot\beta_{GIV}'\Pi'\zeta\zeta'\Pi) \\
&\quad -\sigma_\varepsilon^2 c_5(c_5\Pi'\zeta\zeta'\Pi + c_1\xi\zeta'\Pi - c_1\Pi'\zeta\zeta'\Pi - c_2\xi\zeta'\Pi + c_2\Pi'\zeta\zeta'\Pi) \\
&\quad -\sigma_\varepsilon^2[c_1c_5\Pi'\zeta\xi' + (\xi-\Pi'\zeta)c_1c_4\xi'] + \sigma_\varepsilon^2[c_2c_5\Pi'\zeta\xi' + (\xi-\Pi'\zeta)c_2c_4\xi'] \\
&\quad +\sigma_\varepsilon^2[c_1c_5\Pi'\zeta\zeta'\Pi + c_1c_4(\xi\zeta'\Pi - \Pi'\zeta\zeta'\Pi)] - \sigma_\varepsilon^2[c_2c_5\Pi'\zeta\zeta'\Pi + c_2c_4(\xi\zeta'\Pi - \Pi'\zeta\zeta'\Pi)]
\end{aligned}
$$

$$
\begin{aligned}
&= c_5^2\Sigma_{\hat X'\hat X} + \sigma_\varepsilon^2 c_4\xi\xi' - \sigma_\varepsilon^2 c_4\Pi'\zeta\xi' - \sigma_\varepsilon^2 c_2\xi\xi' + \sigma_\varepsilon^2 c_2\xi\xi' \\
&\quad -\sigma_\varepsilon^2 c_4\xi\zeta'\Pi + \sigma_\varepsilon^2 c_4\Pi'\zeta\zeta'\Pi + \sigma_\varepsilon^2(c_4-c_5)c_5\Pi'\zeta\zeta'\Pi - \sigma_\varepsilon^2 c_4c_5\xi\zeta'\Pi \\
&\quad -\sigma_\varepsilon^2 c_4c_5\Pi'\zeta\xi' - \sigma_\varepsilon^2 c_4^2\xi\xi' + \sigma_\varepsilon^2 c_4^2\Pi'\zeta\xi' + \sigma_\varepsilon^2 c_4c_5\Pi'\zeta\zeta'\Pi + \sigma_\varepsilon^2 c_4^2\xi\zeta'\Pi - \sigma_\varepsilon^2 c_4^2\Pi'\zeta\zeta'\Pi
\end{aligned}
$$

$$
\begin{aligned}
&= c_5^2\Sigma_{\hat X'\hat X} + \sigma_\varepsilon^2 c_4(1-c_4)\xi\xi' + \sigma_\varepsilon^2 c_4(c_4-1-c_5)\Pi'\zeta\xi' + \sigma_\varepsilon^2 c_4(c_4-1-c_5)\xi\zeta'\Pi \\
&\quad -\sigma_\varepsilon^2[c_4(c_4-1-2c_5) + c_5^2]\Pi'\zeta\zeta'\Pi,
\end{aligned}
$$

the third is

$$
2\sigma_\varepsilon^2 c_4^2\xi\xi' + 2\sigma_\varepsilon^2 c_4c_5\Pi'\zeta\xi' + 2\sigma_\varepsilon^2 c_4c_5\xi\zeta'\Pi + 2\sigma_\varepsilon^2 c_5^2\Pi'\zeta\zeta'\Pi,
$$

the fourth is

$$
\begin{aligned}
&[c_4\Sigma_{X'Z} - \Pi'\zeta\ddot\beta_{GIV}'\Sigma_{X'Z}][c_5\Pi + \Sigma_{Z'Z}^{-1}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot\beta_{GIV})(\xi' - \zeta'\Pi)] \\
&-[\sigma_\varepsilon^2 c_4\xi\zeta' - \sigma_\varepsilon^2\Pi'\zeta\ddot\beta_{GIV}'\xi\zeta'][c_5\Pi + \Sigma_{Z'Z}^{-1}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot\beta_{GIV})(\xi' - \zeta'\Pi)]
\end{aligned}
$$

$$
\begin{aligned}
&= c_4c_5\Sigma_{X'Z}\Pi - c_5\Pi'\zeta\ddot\beta_{GIV}'\Sigma_{X'Z}\Pi \\
&\quad +\sigma_\varepsilon^2 c_4\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\zeta\xi' - c_4\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'X}\ddot\beta_{GIV}\xi' - \sigma_\varepsilon^2 c_4\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\zeta\zeta'\Pi + c_4\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'X}\ddot\beta_{GIV}\zeta'\Pi \\
&\quad -\sigma_\varepsilon^2 c_2\Pi'\zeta\xi' + \Pi'\zeta\ddot\beta_{GIV}'\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'X}\ddot\beta_{GIV}\xi' + \sigma_\varepsilon^2 c_2\Pi'\zeta\zeta'\Pi - \Pi'\zeta\ddot\beta_{GIV}'\Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'X}\ddot\beta_{GIV}\zeta'\Pi \\
&\quad -\sigma_\varepsilon^2 c_4c_5\xi\zeta'\Pi + \sigma_\varepsilon^2 c_3c_5\Pi'\zeta\zeta'\Pi - \sigma_\varepsilon^2 c_1c_4\xi\xi' + \sigma_\varepsilon^2 c_1c_3\Pi'\zeta\xi' + \sigma_\varepsilon^2 c_2c_4\xi\xi' - \sigma_\varepsilon^2 c_2c_3\Pi'\zeta\xi' \\
&\quad +\sigma_\varepsilon^2 c_1c_4\xi\zeta'\Pi - \sigma_\varepsilon^2 c_1c_3\Pi'\zeta\zeta'\Pi - \sigma_\varepsilon^2 c_2c_4\xi\zeta'\Pi + \sigma_\varepsilon^2 c_2c_3\Pi'\zeta\zeta'\Pi
\end{aligned}
$$

$$
\begin{aligned}
&= c_4c_5\Sigma_{X'Z}\Pi - \sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi + \sigma_\varepsilon^2 c_4\Pi'\zeta\xi' - \sigma_\varepsilon^2 c_4\Pi'\zeta\xi' - \sigma_\varepsilon^2 c_4\Pi'\zeta\zeta'\Pi + \sigma_\varepsilon^2 c_4\Pi'\zeta\zeta'\Pi \\
&\quad -\sigma_\varepsilon^2 c_2\Pi'\zeta\xi' + \sigma_\varepsilon^2 c_2\Pi'\zeta\xi' + \sigma_\varepsilon^2 c_2\Pi'\zeta\zeta'\Pi - \sigma_\varepsilon^2 c_2\Pi'\zeta\zeta'\Pi \\
&\quad +\sigma_\varepsilon^2 c_4(c_4-c_5)\xi\zeta'\Pi + \sigma_\varepsilon^2 c_3(c_5-c_4)\Pi'\zeta\zeta'\Pi - \sigma_\varepsilon^2 c_4^2\xi\xi' + \sigma_\varepsilon^2 c_3c_4\Pi'\zeta\xi'
\end{aligned}
$$

$$
= c_4c_5\Sigma_{\hat X'\hat X} - \sigma_\varepsilon^2 c_4^2\xi\xi' + \sigma_\varepsilon^2 c_4(c_4-c_5)\xi\zeta'\Pi + \sigma_\varepsilon^2 c_3c_4\Pi'\zeta\xi' + \sigma_\varepsilon^2 c_3(c_5-c_4)\Pi'\zeta\zeta'\Pi - \sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi,
$$

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

$$
\begin{aligned}
V_0 &= c_4^2\Sigma_{X'X} - \sigma_\varepsilon^2 c_4^2\xi\xi' - c_4\Sigma_{X'X}\ddot\beta_{GIV}\zeta'\Pi + \sigma_\varepsilon^2 c_3c_4\xi\zeta'\Pi \\
&\quad -c_4\Pi'\zeta\ddot\beta_{GIV}'\Sigma_{X'X} + \sigma_\varepsilon^2 c_3c_4\Pi'\zeta\xi' + (\ddot\beta_{GIV}'\Sigma_{X'X}\ddot\beta_{GIV} - \sigma_\varepsilon^2 c_3^2)\Pi'\zeta\zeta'\Pi \\
&\quad +(1-c_4-c_3)^2\Sigma_{\hat X'\hat X} + \sigma_\varepsilon^2 c_4(1-c_4)\xi\xi' + \sigma_\varepsilon^2 c_4(2c_4+c_3-2)\Pi'\zeta\xi' + \sigma_\varepsilon^2 c_4(2c_4+c_3-2)\xi\zeta'\Pi \\
&\quad -\sigma_\varepsilon^2[c_4(c_4-1-2c_5) + c_5^2]\Pi'\zeta\zeta'\Pi + 2\sigma_\varepsilon^2 c_4^2\xi\xi' + 2\sigma_\varepsilon^2 c_4c_5\Pi'\zeta\xi' + 2\sigma_\varepsilon^2 c_4c_5\xi\zeta'\Pi + 2\sigma_\varepsilon^2 c_5^4\Pi'\zeta\zeta'\Pi \\
&\quad +2c_4c_5\Sigma_{\hat X'\hat X} - 2\sigma_\varepsilon^2 c_4^2\xi\xi' + \sigma_\varepsilon^2 c_4(c_4+c_3-c_5)\xi\zeta'\Pi + \sigma_\varepsilon^2 c_4(c_4+c_3-c_5)\Pi'\zeta\xi' \\
&\quad +2\sigma_\varepsilon^2 c_3(c_2-c_1+c_5)\Pi'\zeta\zeta'\Pi - 2\sigma_\varepsilon^2 c_5\Pi'\zeta\zeta'\Pi
\end{aligned}
$$

$$
\begin{aligned}
= \quad & c_4^2\Sigma_{X'X} + 2c_4c_5\Sigma_{\hat{X}'\hat{X}} + c_5^2\Sigma_{\hat{X}'\hat{X}} - c_4\Sigma_{X'X}\ddot{\beta}_{GIV}\zeta'\Pi - c_4\Pi'\zeta\ddot{\beta}_{GIV}'\Sigma_{X'X} \\
& -\sigma_\varepsilon^2c_4^2\xi\xi' + \sigma_\varepsilon^2c_4(1-c_4)\xi\xi' + 2\sigma_\varepsilon^2c_4^2\xi\xi' - 2\sigma_\varepsilon^2c_4^2\xi\xi' \\
& +\sigma_\varepsilon^2c_3c_4\xi\zeta'\Pi + \sigma_\varepsilon^2c_4(2c_4+c_3-2)\xi\zeta'\Pi + 2\sigma_\varepsilon^2c_4c_5\xi\zeta'\Pi + \sigma_\varepsilon^2c_4(c_4+c_3-c_5)\xi\zeta'\Pi \\
& +\sigma_\varepsilon^2c_3c_4\Pi'\zeta\xi' + \sigma_\varepsilon^2c_4(2c_4+c_3-2)\Pi'\zeta\xi' + 2\sigma_\varepsilon^2c_4c_5\Pi'\zeta\xi' + \sigma_\varepsilon^2c_4(c_4+c_3-c_5)\Pi'\zeta\xi' \\
& +(\ddot{\beta}_{GIV}'\Sigma_{X'X}\ddot{\beta}_{GIV} - \sigma_\varepsilon^2c_3^2)\Pi'\zeta\zeta'\Pi - \sigma_\varepsilon^2[c_4(c_4-1-2c_5)+c_5^2]\Pi'\zeta\zeta'\Pi \\
& +2\sigma_\varepsilon^2c_5^2\Pi'\zeta\zeta'\Pi + 2\sigma_\varepsilon^2c_3(c_2-c_1+c_5)\Pi'\zeta\zeta'\Pi - 2\sigma_\varepsilon^2c_5\Pi'\zeta\zeta'\Pi
\end{aligned}
$$

$$
\begin{aligned}
= \quad & c_4^2\Sigma_{X'X} + (2c_4+c_5)c_5\Sigma_{\hat{X}'\hat{X}} - c_4\Sigma_{X'X}\ddot{\beta}_{GIV}\zeta'\Pi - c_4\Pi'\zeta\ddot{\beta}_{GIV}'\Sigma_{X'X} \\
& +\sigma_\varepsilon^2c_4(1-2c_4)\xi\xi' + \sigma_\varepsilon^2c_4(c_4+c_3-c_5)\xi\zeta'\Pi + \sigma_\varepsilon^2c_4(c_4+c_3-c_5)\Pi'\zeta\xi' + \ddot{\beta}_{GIV}'\Sigma_{X'X}\ddot{\beta}_{GIV}\Pi'\zeta\zeta'\Pi \\
& +\sigma_\varepsilon^2\{-c_3^2 - [c_4(c_4-1-2c_5)+c_5^2] + 2c_5^2 + 2c_3(c_5-c_4) - 2c_5\}\Pi'\zeta\zeta'\Pi
\end{aligned}
$$

$$
\begin{aligned}
= \quad & c_4^2\Sigma_{X'X} + [(1-c_3)^2 - c_4^2]\Sigma_{\hat{X}'\hat{X}} - c_4\Sigma_{X'X}\ddot{\beta}_{GIV}\zeta'\Pi - c_4\Pi'\zeta\ddot{\beta}_{GIV}'\Sigma_{X'X} + \sigma_\varepsilon^2c_4(1-2c_4)\xi\xi' \\
& +\sigma_\varepsilon^2c_4(1-2c_5)\xi\zeta'\Pi + \sigma_\varepsilon^2c_4(1-2c_5)\Pi'\zeta\xi' + (\ddot{\beta}_{GIV}'\Sigma_{X'X}\ddot{\beta}_{GIV})\Pi'\zeta\zeta'\Pi + \sigma_\varepsilon^2[c_5(1-2c_5)-c_3]\Pi'\zeta\zeta'\Pi.
\end{aligned}
$$

Thus, the asymptotic variance of the (generalized) GIV estimator (48) is

$$
\begin{aligned}
& V_{GIV}^{CN} \\
= \quad & \sigma_\varepsilon^2[(1-c_3)^2 - c_4^2]\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^2c_4^2\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'X}\Sigma_{\hat{X}'\hat{X}}^{-1} - c_4[\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'X}\ddot{\beta}_{GIV}\ddot{\beta}_{GIV}' + \ddot{\beta}_{GIV}\ddot{\beta}_{GIV}'\Sigma_{X'X}\Sigma_{\hat{X}'\hat{X}}^{-1}] \\
& +\sigma_\varepsilon^4c_4[1-2c_4]\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^2c_4(1-2c_5)[\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\ddot{\beta}_{GIV}' + \ddot{\beta}_{GIV}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1}] \\
& +[c_5(1-2c_5) - c_3 + \sigma_\varepsilon^{-2}(\ddot{\beta}_{GIV}'\Sigma_{X'X}\ddot{\beta}_{GIV})]\ddot{\beta}_{GIV}\ddot{\beta}_{GIV}',
\end{aligned}
$$

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 so the variance simplifies to

$$
\begin{aligned}
V_{IV}^{CN} &= \sigma_\varepsilon^2c_5^2\Sigma_{\hat{X}'\hat{X}}^{-1} + [(1-c_3)(2c_3-1) - c_3 + \sigma_\varepsilon^{-2}(\ddot{\beta}_{IV}'\Sigma_{X'X}\ddot{\beta}_{IV})]\ddot{\beta}_{IV}\ddot{\beta}_{IV}' \\
&= \sigma_\varepsilon^2c_5^2\Sigma_{\hat{X}'\hat{X}}^{-1} - [2c_3^2 - 2c_3 + 1 - \sigma_\varepsilon^{-2}(\ddot{\beta}_{IV}'\Sigma_{X'X}\ddot{\beta}_{IV})]\ddot{\beta}_{IV}\ddot{\beta}_{IV}',
\end{aligned}
$$

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 $\sigma_\varepsilon^4\omega\omega'$. In obtaining the third term of (56) we already found

$$
\omega\omega' = c_4^2\xi\xi' + c_4c_5\Pi'\zeta\xi' + c_4c_5\xi\zeta'\Pi + c_5^2\Pi'\zeta\zeta'\Pi,
$$

so

$$
\begin{aligned}
& \sigma_\varepsilon^4\Sigma_{\hat{X}'\hat{X}}^{-1}\omega\omega'\Sigma_{\hat{X}'\hat{X}}^{-1} \\
= \quad & \sigma_\varepsilon^4c_4^2\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^4c_4c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\zeta\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^4c_4c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\zeta'\Pi\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^4c_5^2\Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\zeta\zeta'\Pi\Sigma_{\hat{X}'\hat{X}}^{-1} \\
= \quad & \sigma_\varepsilon^4c_4^2\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^2c_4c_5\ddot{\beta}_{GIV}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^2c_4c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\ddot{\beta}_{GIV}' + c_5^2\ddot{\beta}_{GIV}\ddot{\beta}_{GIV}'.
\end{aligned}
$$

In the overall asymptotic variance this term has factor $\mu_4 - 1$. The expression for V_{GIV}^{CP} 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

$$
\mu_3\sigma_\varepsilon^3(\Sigma_{W'\iota}\omega' + \omega\Sigma_{\iota'W}).
$$

Since

$$
\begin{aligned}
W'\iota &= \Theta_1\bar{X}'\iota + \Theta_2\bar{Z}'\iota \\
&= c_4\bar{X}'\iota - \Pi'\zeta\ddot{\beta}_{GIV}'\bar{X}'\iota + c_5\Pi'\bar{Z}'\iota + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}_{GIV}'\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\bar{Z}'\iota,
\end{aligned}
$$

we find

$$\Sigma_{W'\iota} = c_4\Sigma_{X'\iota} - \Pi'\zeta\ddot{\beta}'_{GIV}\Sigma_{X'\iota} + c_5\Pi'\Sigma_{Z'\iota} + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota},$$

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

$$\begin{aligned}
& \Sigma_{W'\iota}\omega' \\
= \; & c_4[c_4\Sigma_{X'\iota}\xi' - \Pi'\zeta\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\xi' + c_5\Pi'\Sigma_{Z'\iota}\xi' + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota}\xi'] \\
& + c_5[c_4\Sigma_{X'\iota}\zeta'\Pi - \Pi'\zeta\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\zeta'\Pi + c_5\Pi'\Sigma_{Z'\iota}\zeta'\Pi + (\xi - \Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota}\zeta'\Pi].
\end{aligned}$$

Thus,

$$\begin{aligned}
& \sigma_\varepsilon^3\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{W'\iota}\omega'\Sigma_{\hat{X}'\hat{X}}^{-1} \\
= \; & \sigma_\varepsilon^3 c_4[c_4\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} - \Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\zeta\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\Sigma_{Z'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \quad + (\Sigma_{\hat{X}'\hat{X}}^{-1}\xi - \Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1}] \\
& + \sigma_\varepsilon^3 c_5[c_4\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'\iota}\zeta'\Pi\Sigma_{\hat{X}'\hat{X}}^{-1} - \Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\zeta\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\zeta'\Pi\Sigma_{\hat{X}'\hat{X}}^{-1} + c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\Sigma_{Z'\iota}\zeta'\Pi\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \quad + (\Sigma_{\hat{X}'\hat{X}}^{-1}\xi - \Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\zeta)(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota}\zeta'\Pi\Sigma_{\hat{X}'\hat{X}}^{-1}] \\
= \; & c_4[\sigma_\varepsilon^3 c_4\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^3 c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\Sigma_{Z'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \quad + (\sigma_\varepsilon^3\Sigma_{\hat{X}'\hat{X}}^{-1}\xi - \sigma_\varepsilon\ddot{\beta}_{GIV})(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1}] \\
& + c_5[\sigma_\varepsilon c_4\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'\iota}\ddot{\beta}'_{GIV} - \sigma_\varepsilon^{-1}\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV} + \sigma_\varepsilon c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\Sigma_{Z'\iota}\ddot{\beta}'_{GIV} \\
& \quad + (\sigma_\varepsilon\Sigma_{\hat{X}'\hat{X}}^{-1}\xi - \sigma_\varepsilon^{-1}\ddot{\beta}_{GIV})(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota}\ddot{\beta}'_{GIV}]
\end{aligned}$$

and the additional term is then equal to μ_3 multiplied by

$$\begin{aligned}
& \sigma_\varepsilon^3\Sigma_{\hat{X}'\hat{X}}^{-1}[\Sigma_{W'\iota}\omega' + \omega\Sigma_{\iota'W}]\Sigma_{\hat{X}'\hat{X}}^{-1} \\
= \; & c_4\{\sigma_\varepsilon^3 c_4\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^3 c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\Sigma_{Z'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \quad + (\sigma_\varepsilon^3\Sigma_{\hat{X}'\hat{X}}^{-1}\xi - \sigma_\varepsilon\ddot{\beta}_{GIV})(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1}\} \\
& + c_5[\sigma_\varepsilon c_4\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'\iota}\ddot{\beta}'_{GIV} - \sigma_\varepsilon^{-1}\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV} + \sigma_\varepsilon c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\Pi'\Sigma_{Z'\iota}\ddot{\beta}'_{GIV} \\
& \quad + (\sigma_\varepsilon\Sigma_{\hat{X}'\hat{X}}^{-1}\xi - \sigma_\varepsilon^{-1}\ddot{\beta}_{GIV})(\sigma_\varepsilon^2\zeta' - \ddot{\beta}'_{GIV}\Sigma_{X'Z})\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota}\ddot{\beta}'_{GIV}] \\
& + c_4[\sigma_\varepsilon^3 c_4\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\Sigma'_{X'\iota}\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\Sigma'_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV} + \sigma_\varepsilon^3 c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\Sigma'_{Z'\iota}\Pi\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \quad + \Sigma_{\hat{X}'\hat{X}}^{-1}\xi\Sigma'_{Z'\iota}\Sigma_{Z'Z}^{-1}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\sigma_\varepsilon^3\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon\ddot{\beta}'_{GIV})] \\
& + c_5[\sigma_\varepsilon c_4\ddot{\beta}_{GIV}\Sigma'_{X'\iota}\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon^{-1}\ddot{\beta}'_{GIV}\Sigma_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}'_{GIV} + \sigma_\varepsilon c_5\ddot{\beta}_{GIV}\Sigma'_{Z'\iota}\Pi\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \quad + \ddot{\beta}_{GIV}\Sigma'_{Z'\iota}\Sigma_{Z'Z}^{-1}(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\sigma_\varepsilon\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon^{-1}\ddot{\beta}'_{GIV})].
\end{aligned}$$

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

$$\begin{aligned}
\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'\iota} &= e_{k,1}, \\
\Pi'\Sigma_{Z'\iota} &= \Sigma_{X'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'\iota} = \Sigma_{X'Z}e_{l,1},
\end{aligned}$$

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

$$
\begin{aligned}
& \sigma_\varepsilon^3 \Sigma_{\hat{X}'\hat{X}}^{-1}[\Sigma_{W'\iota}\omega' + \omega\Sigma_{\iota'W}]\Sigma_{\hat{X}'\hat{X}}^{-1} \\
= \; & c_4\{\sigma_\varepsilon^3 c_4 e_{k,1}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon \ddot{\beta}_{GIV}\ddot{\beta}_{GIV}'\Sigma_{X'\iota}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} + \sigma_\varepsilon^3 c_5 \Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'Z}e_{l,1}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \qquad + (\sigma_\varepsilon^3\Sigma_{\hat{X}'\hat{X}}^{-1}\xi - \sigma_\varepsilon\ddot{\beta}_{GIV})(\sigma_\varepsilon^2\zeta' - \ddot{\beta}_{GIV}'\Sigma_{X'Z})e_{l,1}\xi'\Sigma_{\hat{X}'\hat{X}}^{-1}\} \\
& + c_5[\sigma_\varepsilon c_4 e_{k,1}\ddot{\beta}_{GIV}' - \sigma_\varepsilon^{-1}\ddot{\beta}_{GIV}'\Sigma_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}_{GIV}' + \sigma_\varepsilon c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\Sigma_{X'Z}e_{l,1}\ddot{\beta}_{GIV}' \\
& \qquad + (\sigma_\varepsilon\Sigma_{\hat{X}'\hat{X}}^{-1}\xi - \sigma_\varepsilon^{-1}\ddot{\beta}_{GIV})(\sigma_\varepsilon^2\zeta' - \ddot{\beta}_{GIV}'\Sigma_{X'Z})e_{l,1}\ddot{\beta}_{GIV}'] \\
& + c_4[\sigma_\varepsilon^3 c_4\Sigma_{\hat{X}'\hat{X}}^{-1}\xi e_{k,1}' - \sigma_\varepsilon\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\Sigma_{X'\iota}'\ddot{\beta}_{GIV}\ddot{\beta}_{GIV}' + \sigma_\varepsilon^3 c_5\Sigma_{\hat{X}'\hat{X}}^{-1}\xi\Sigma_{X'Z}e_{l,1}'\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \qquad + \Sigma_{\hat{X}'\hat{X}}^{-1}\xi e_{l,1}'(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\sigma_\varepsilon^3\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon\ddot{\beta}_{GIV}')] \\
& + c_5[\sigma_\varepsilon c_4\ddot{\beta}_{GIV}e_{k,1}' - \sigma_\varepsilon^{-1}\ddot{\beta}_{GIV}'\Sigma_{X'\iota}\ddot{\beta}_{GIV}\ddot{\beta}_{GIV}' + \sigma_\varepsilon c_5\ddot{\beta}_{GIV}e_{l,1}'\Sigma_{Z'X}\Sigma_{\hat{X}'\hat{X}}^{-1} \\
& \qquad + \ddot{\beta}_{GIV}e_{l,1}'(\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{GIV})(\sigma_\varepsilon\xi'\Sigma_{\hat{X}'\hat{X}}^{-1} - \sigma_\varepsilon^{-1}\ddot{\beta}_{GIV}')].
\end{aligned}
$$

When $k = l$, i.e. $c_1 = c_2$, $c_5 = 1 - c_3$, $\Pi\Sigma_{\hat{X}'\hat{X}}^{-1} = \Sigma_{X'Z}^{-1}$ and $\sigma_\varepsilon^2\zeta - \Sigma_{Z'X}\ddot{\beta}_{IV} = 0$, V_{GIV}^{CP} specializes to the expression

$$
\begin{aligned}
V_{IV}^{CP} = \; & \sigma_\varepsilon^2 c_5^2\Sigma_{Z'X}^{-1}\Sigma_{Z'Z}\Sigma_{X'Z}^{-1} - [2c_5^2 - 2c_5 + 1 - \sigma_\varepsilon^{-2}(\ddot{\beta}_{IV}'\Sigma_{X'X}\ddot{\beta}_{IV})]\ddot{\beta}_{IV}\ddot{\beta}_{IV}' \\
& + (\mu_4 - 1)c_5^2\ddot{\beta}_{IV}\ddot{\beta}_{IV}' - 2\mu_3 c_5\sigma_\varepsilon^{-1}\ddot{\beta}_{IV}'\Sigma_{X'\iota}\ddot{\beta}_{IV}\ddot{\beta}_{IV}' + \mu_3\sigma_\varepsilon c_5^2[\Sigma_{Z'X}^{-1}\Sigma_{Z'\iota}\ddot{\beta}_{IV}' + \ddot{\beta}_{IV}\Sigma_{Z'\iota}'\Sigma_{X'Z}^{-1}] \\
= \; & \sigma_\varepsilon^2 c_5^2\Sigma_{Z'X}^{-1}\Sigma_{Z'Z}\Sigma_{X'Z}^{-1} + \mu_3\sigma_\varepsilon c_5^2[\Sigma_{Z'X}^{-1}\Sigma_{Z'\iota}\ddot{\beta}_{IV}' + \ddot{\beta}_{IV}\Sigma_{Z'\iota}'\Sigma_{X'Z}^{-1}] \\
& - [(3 - \mu_4)c_5^2 + 2c_5(\mu_3\sigma_\varepsilon^{-1}\ddot{\beta}_{IV}'\Sigma_{X'\iota} - 1) + 1 - \sigma_\varepsilon^{-2}(\ddot{\beta}_{IV}'\Sigma_{X'X}\ddot{\beta}_{IV})]\ddot{\beta}_{IV}\ddot{\beta}_{IV}'.
\end{aligned}
$$

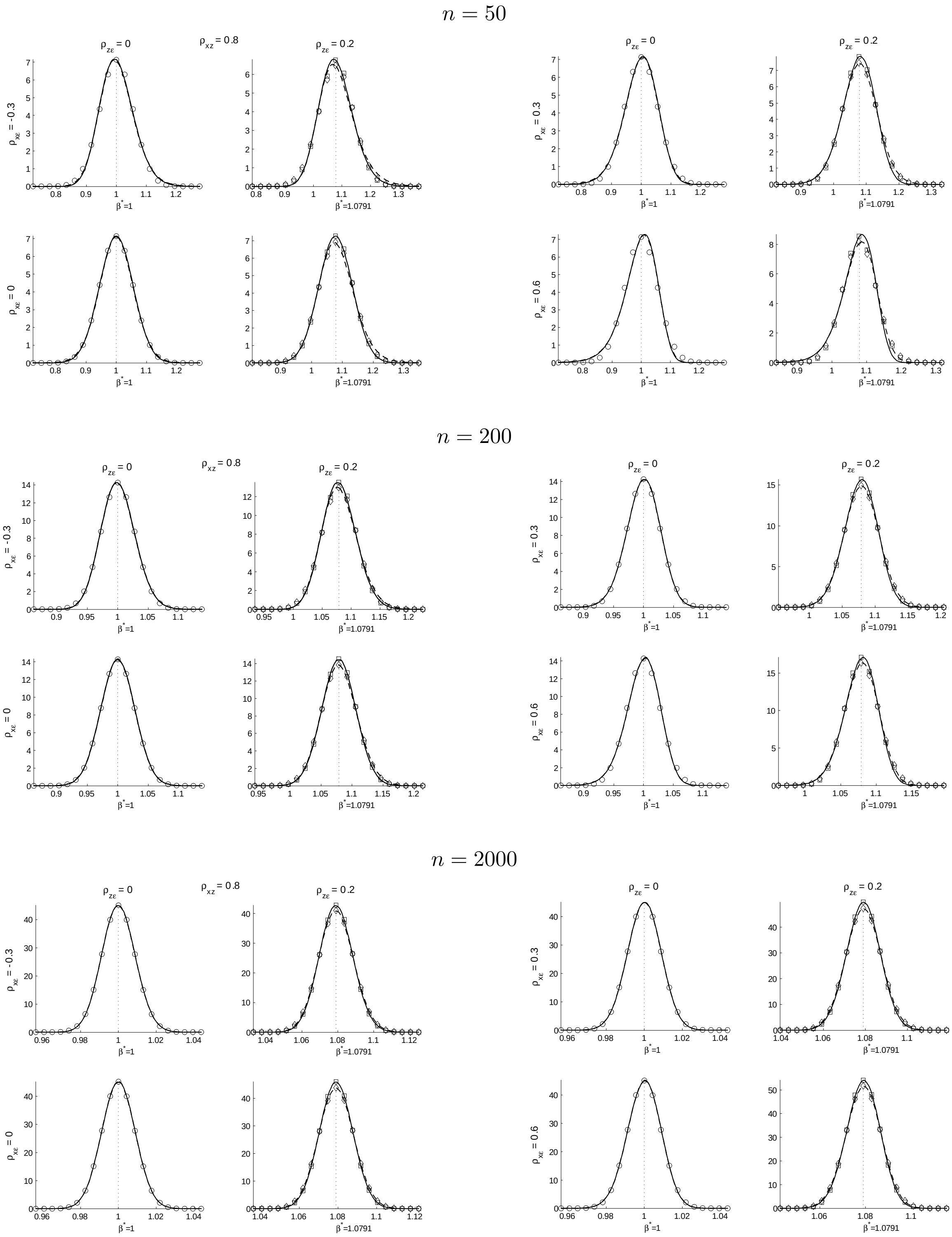
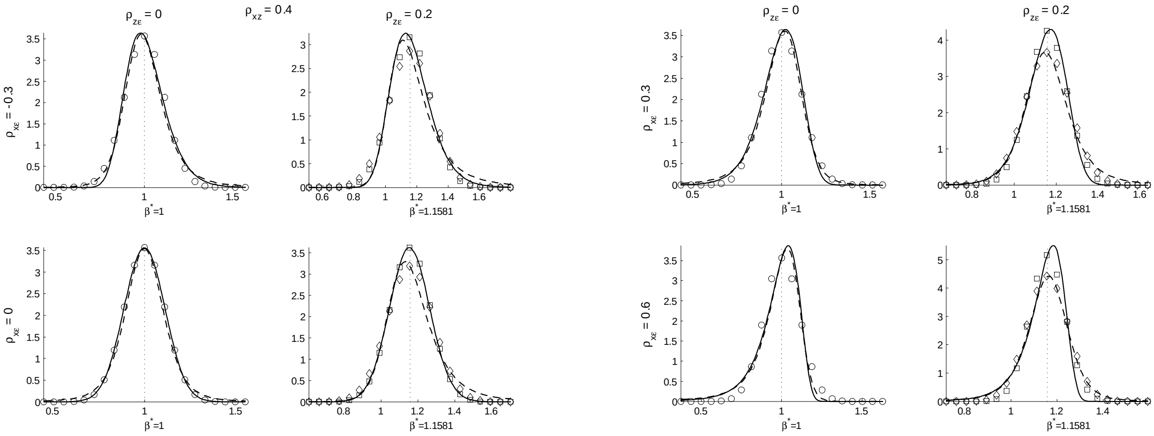
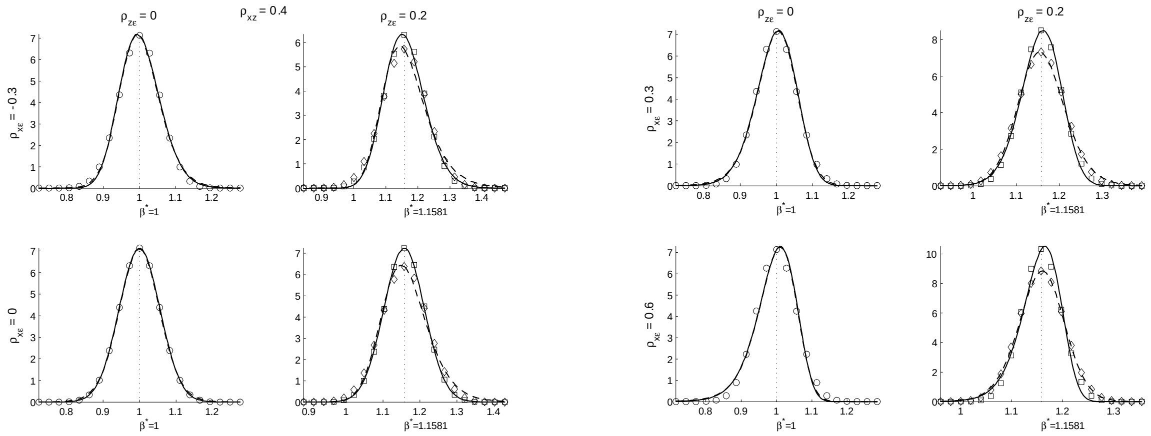


Figure 1.1: Densities for $k = l = 1$; $\beta = 1$; $\sigma_x^2/\sigma_\varepsilon^2 = 10$; $\rho_{xz} = 0.8$
(actual: — = cond., – – = uncond.; asymptotic: □ = cond., ◊ = uncond., O = both)

$n = 50$



$n = 200$



$n = 2000$

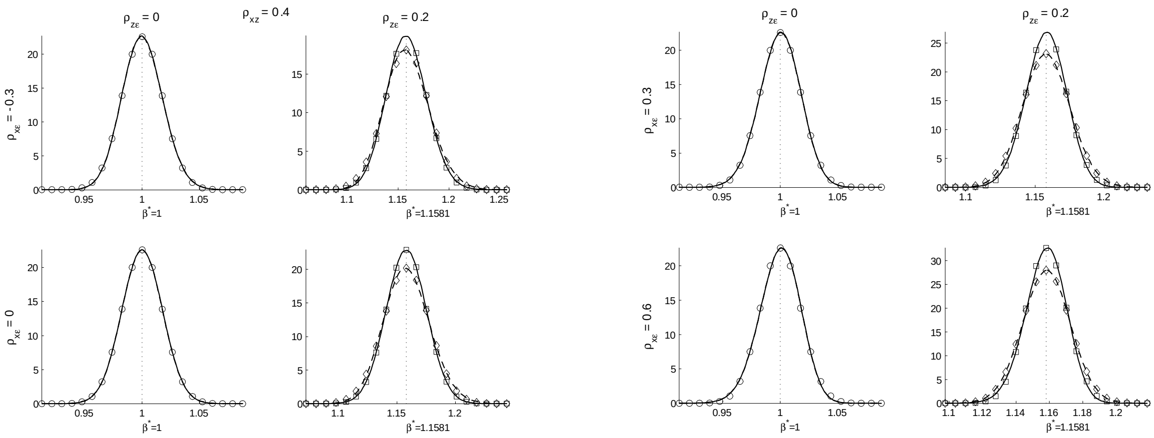


Figure 1.2: Densities for $k = l = 1;\ \beta = 1;\ \sigma_x^2/\sigma_\varepsilon^2 = 10;\ \rho_{xz} = 0.4$
(actual: — = cond., – – = uncond.; asymptotic: □ = cond., ◇ = uncond., O = both)

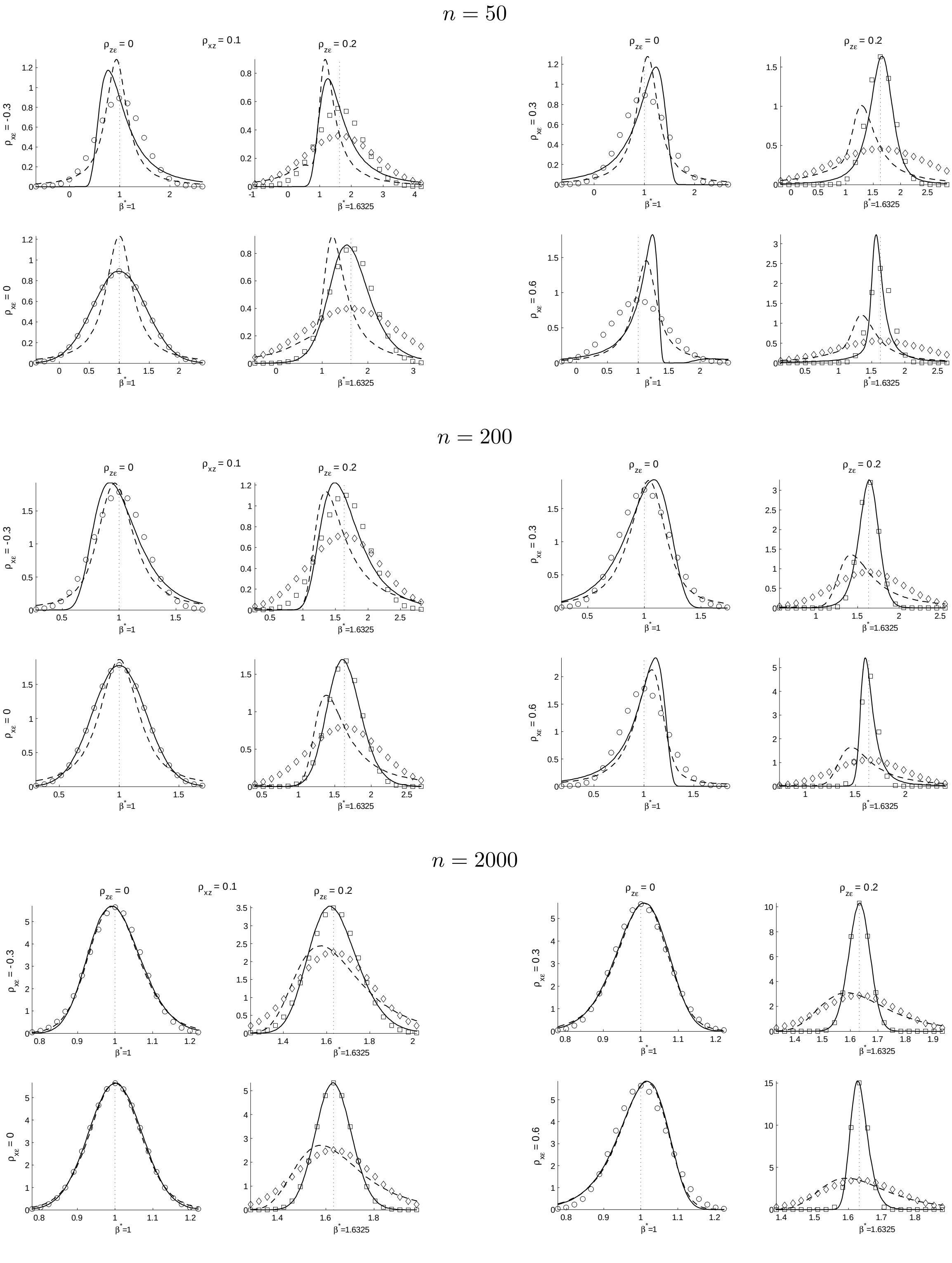


Figure 1.3: Densities for $k = l = 1;\ \beta = 1;\ \sigma_x^2/\sigma_\varepsilon^2 = 10;\ \rho_{xz} = 0.1$
(actual: — = cond., $--$ = uncond.; asymptotic: □ = cond., ◊ = uncond., O = both)

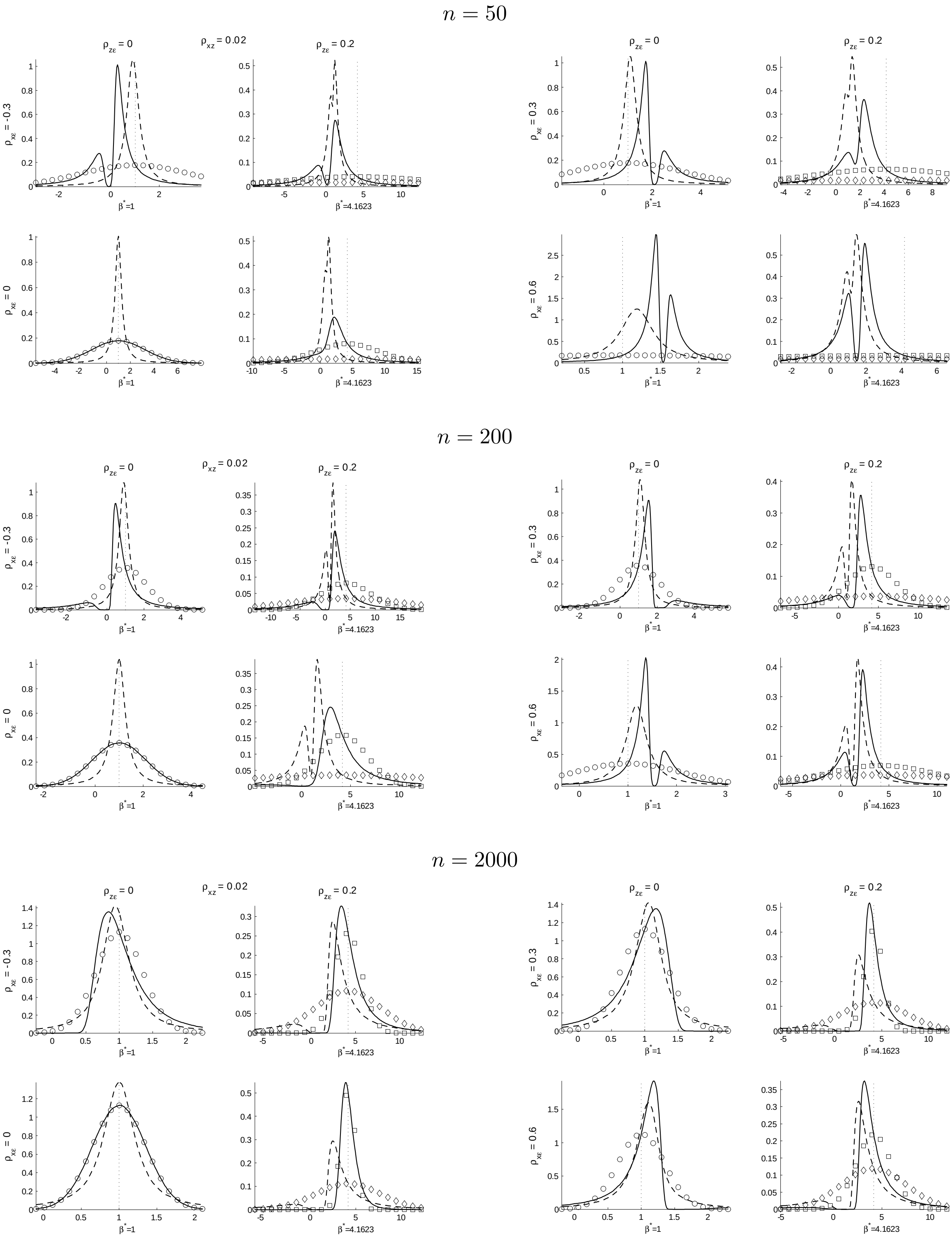


Figure 1.4: Densities for $k = l = 1$; $\beta = 1$; $\sigma_x^2/\sigma_\varepsilon^2 = 10$; $\rho_{xz} = 0.02$
(actual: — = cond., $- - -$ = uncond.; asymptotic: □ = cond., ◊ = uncond., O = both)

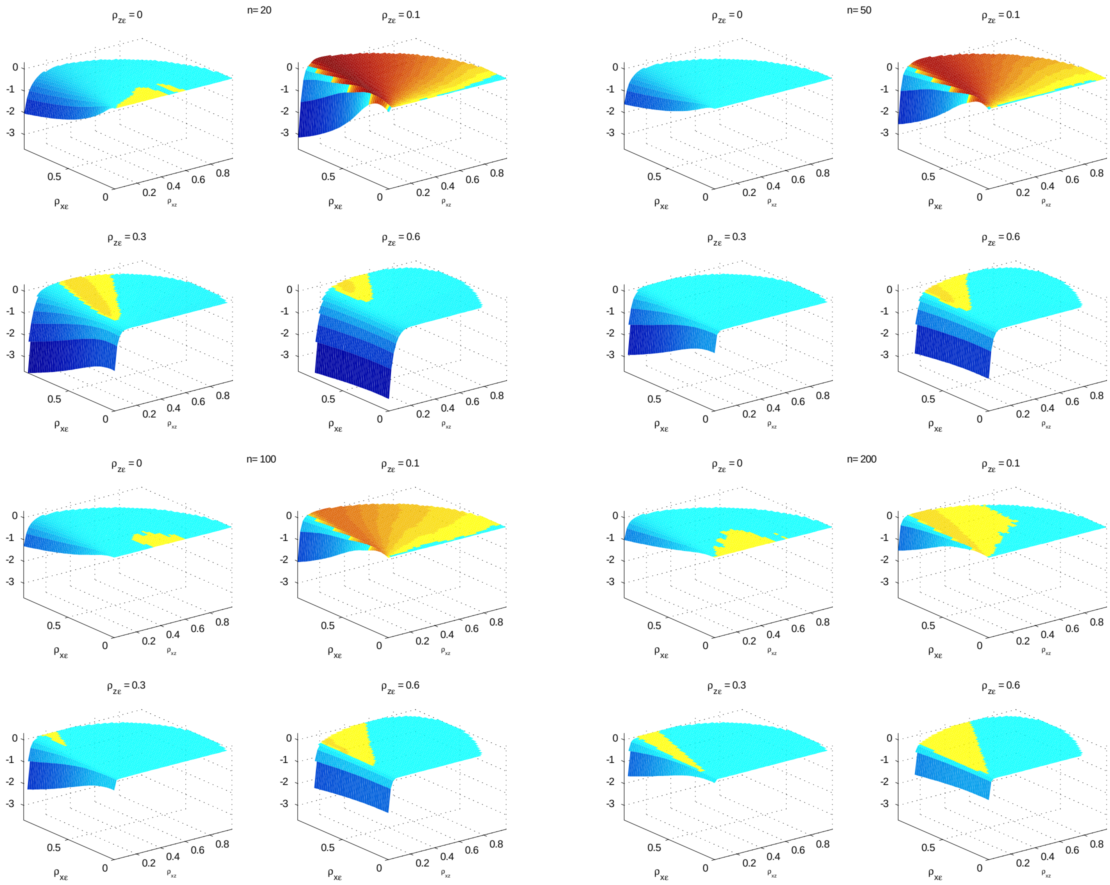


Figure 1.5: Conditioned $\log[MAE(\hat{\beta}_{IV})/AMAE(\hat{\beta}_{IV})]$ for $k = l = 1$; any SN

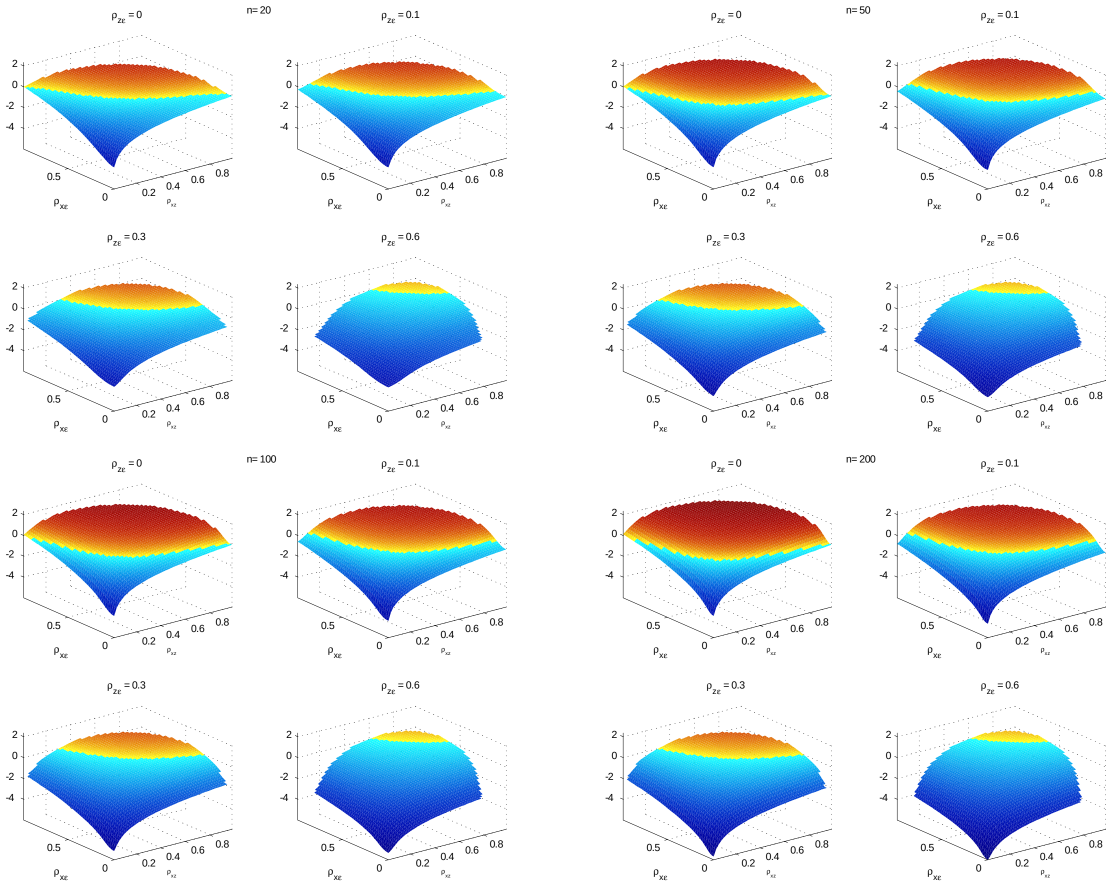


Figure 1.6: Conditioned $\log[MAE(\hat{\beta}_{OLS})/MAE(\hat{\beta}_{IV})]$ for $k = l = 1$; any SN

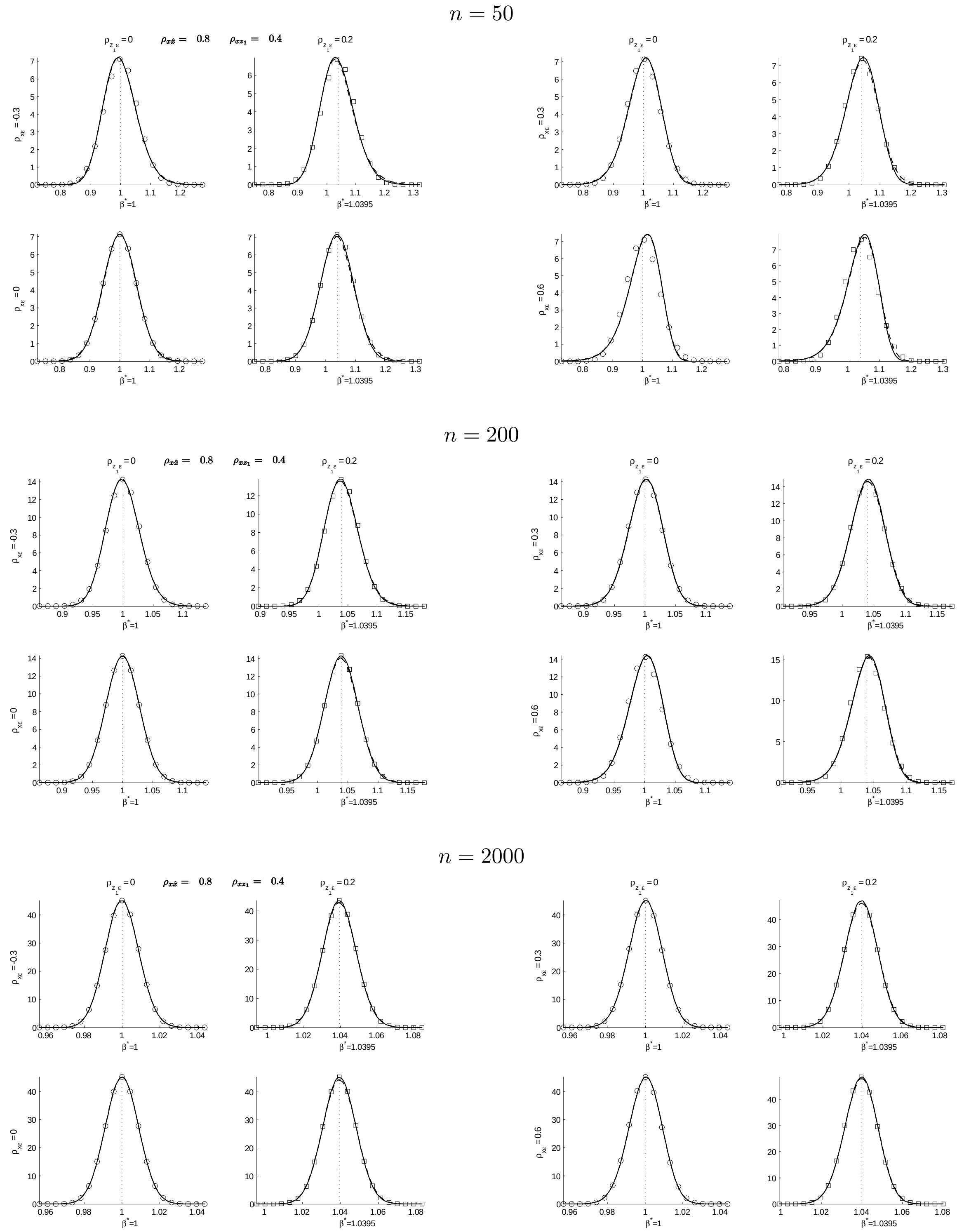


Figure 2.1: Densities for $l-1=k=1;\ \beta=1;\ \sigma_x^2/\sigma_\varepsilon^2=10;\ \rho_{x\hat{x}}=0.8;\ \rho_{xz_1}=0.4$
(actual: — = cond., – – = uncond.; asymptotic: □ = cond., O = both cond. and uncond.)

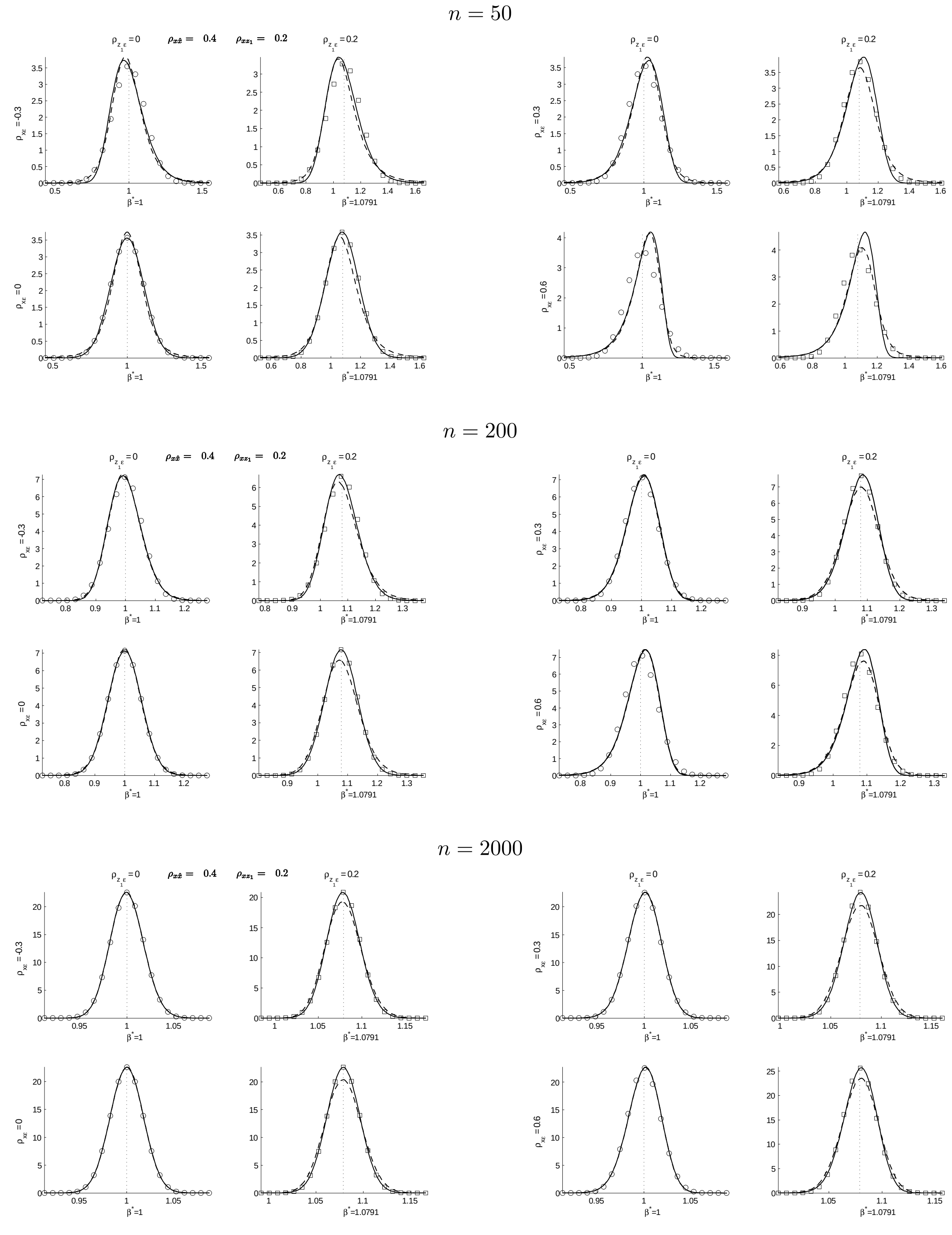
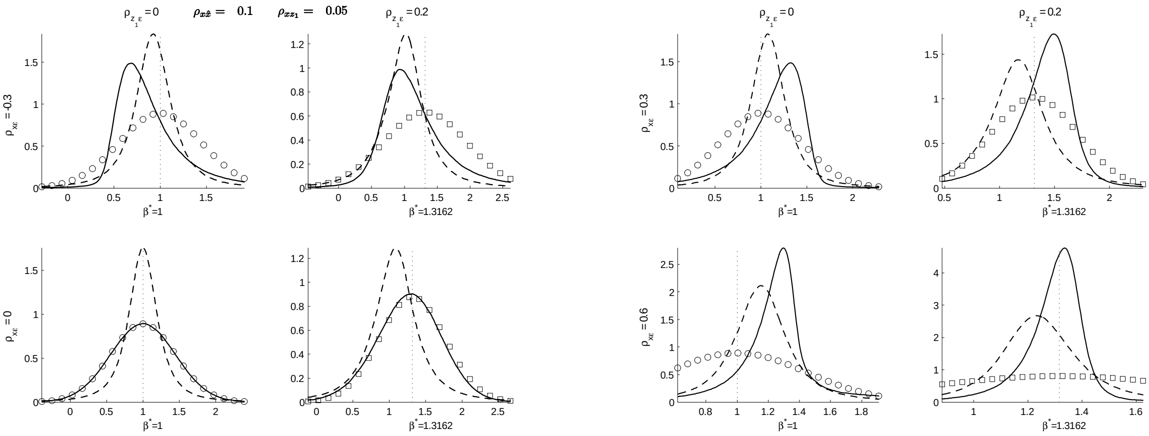
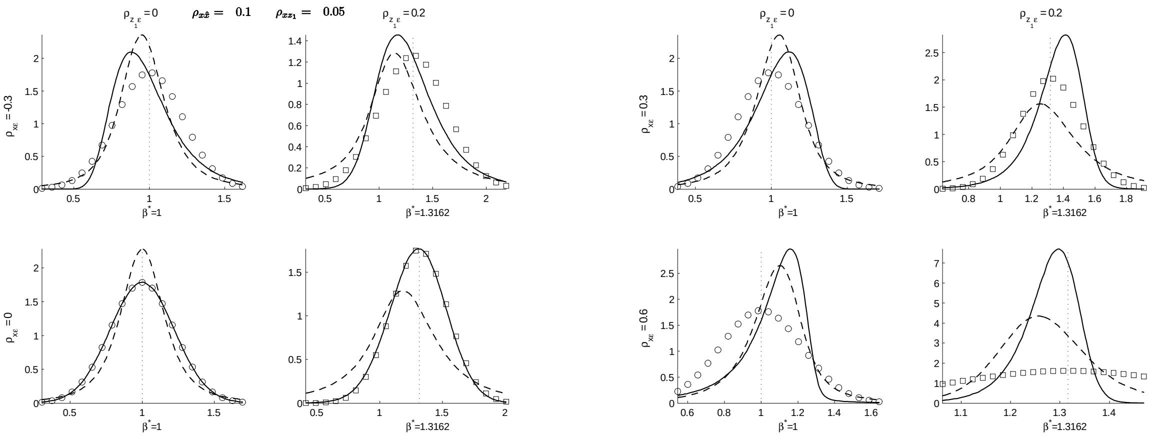


Figure 2.2: Densities for $l-1=k=1;\ \beta=1;\ \sigma_x^2/\sigma_\varepsilon^2=10;\ \rho_{x\hat{x}}=0.4;\ \rho_{xz_1}=0.2$
(actual: — = cond., – – = uncond.; asymptotic: □ = cond., O = both cond. and uncond.)

$n = 50$



$n = 200$



$n = 2000$

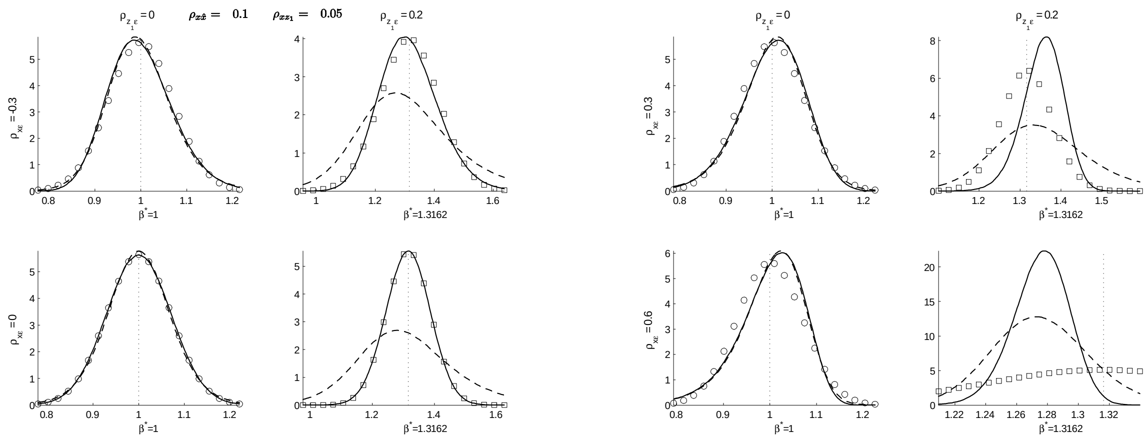


Figure 2.3: Densities for $l-1=k=1$; $\beta=1$; $\sigma_x^2/\sigma_\varepsilon^2=10$; $\rho_{x\hat{x}}=0.1$; $\rho_{xz_1}=0.05$
(actual: — = cond., – – = uncond.; asymptotic: □ = cond., O = both cond. and uncond.)

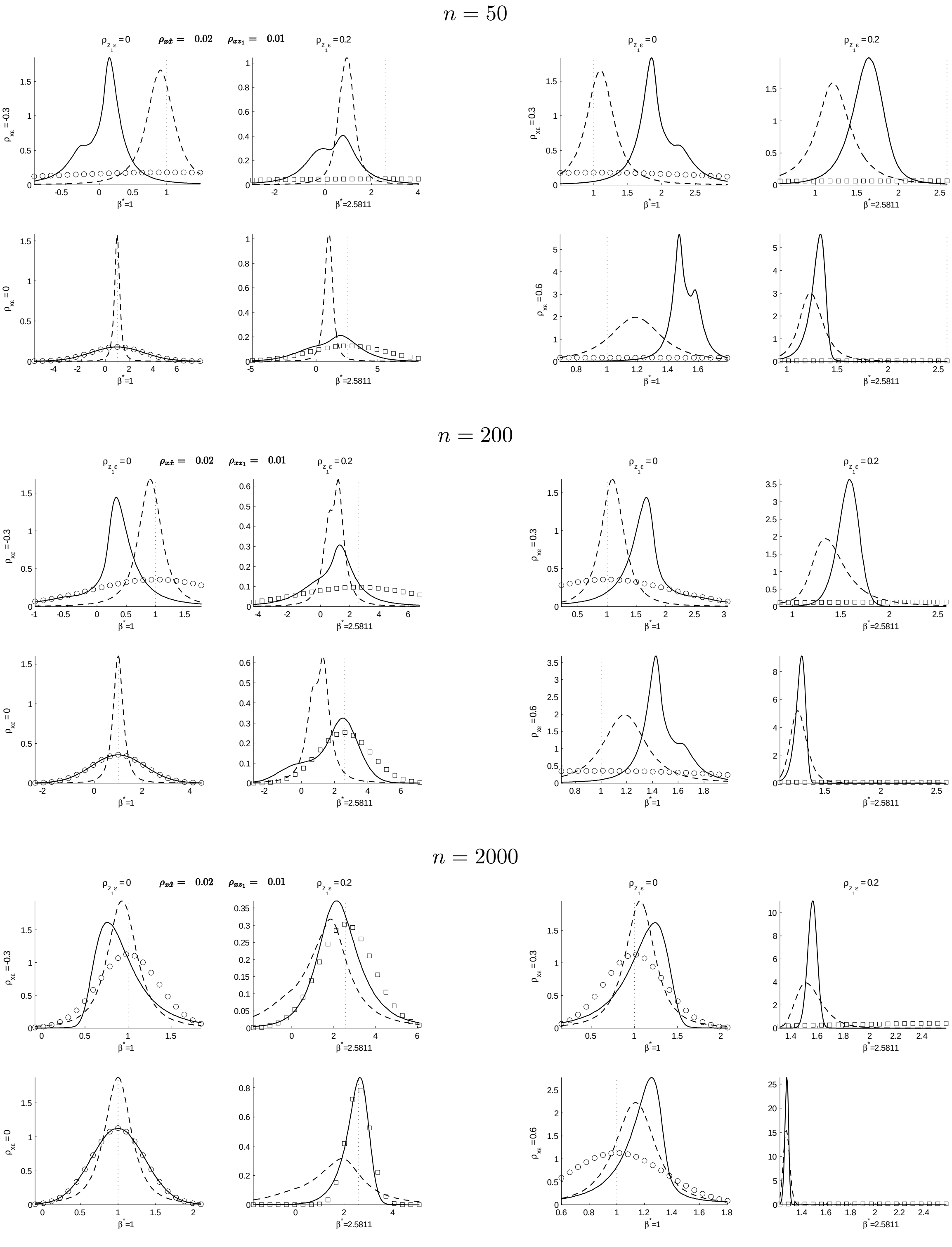


Figure 2.4: Densities for $l-1=k=1$; $\beta=1$; $\sigma_x^2/\sigma_\varepsilon^2=10$; $\rho_{x\hat{x}}=0.02$; $\rho_{xz_1}=0.01$ (actual: — = cond., – – = uncond.; asymptotic: □ = cond., O = both cond. and uncond.)

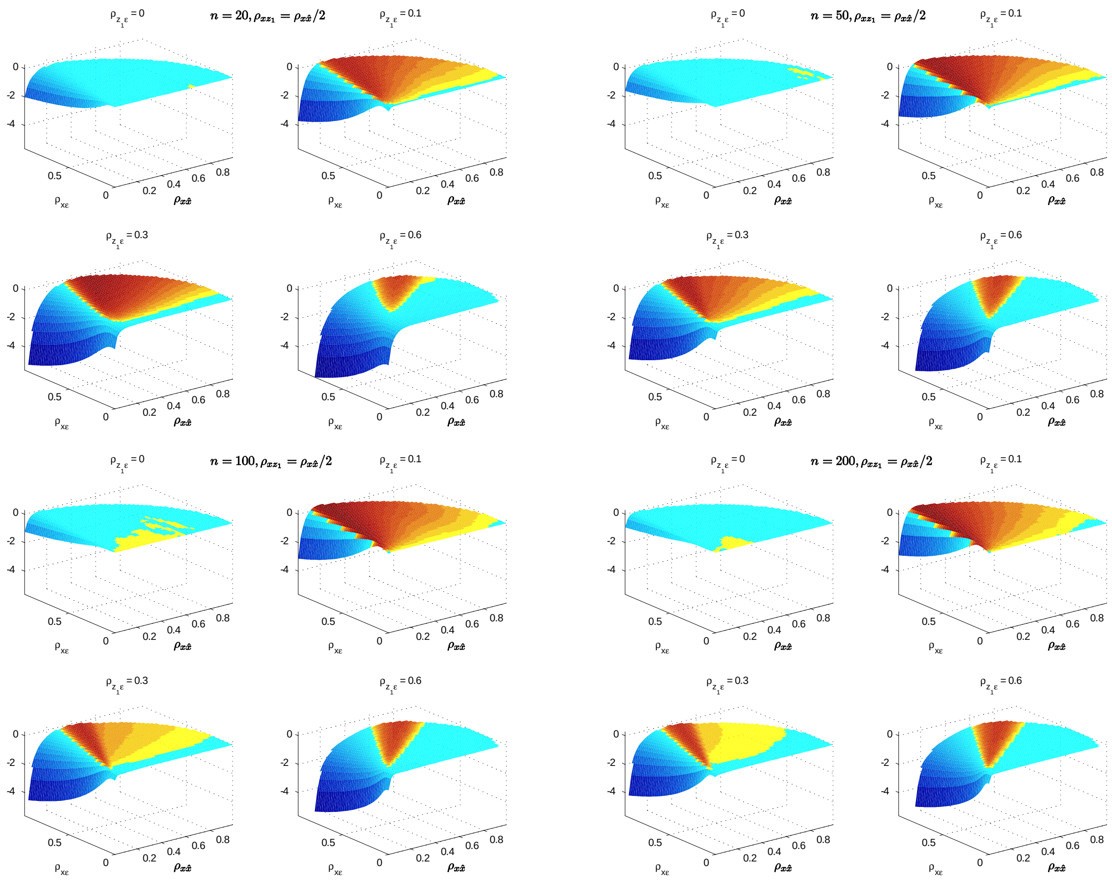


Figure 2.5: Conditioned $\log[MAE(\hat{\beta}_{GIV})/AMAE(\hat{\beta}_{GIV})]$ for $k = l - 1 = 1$; any SN

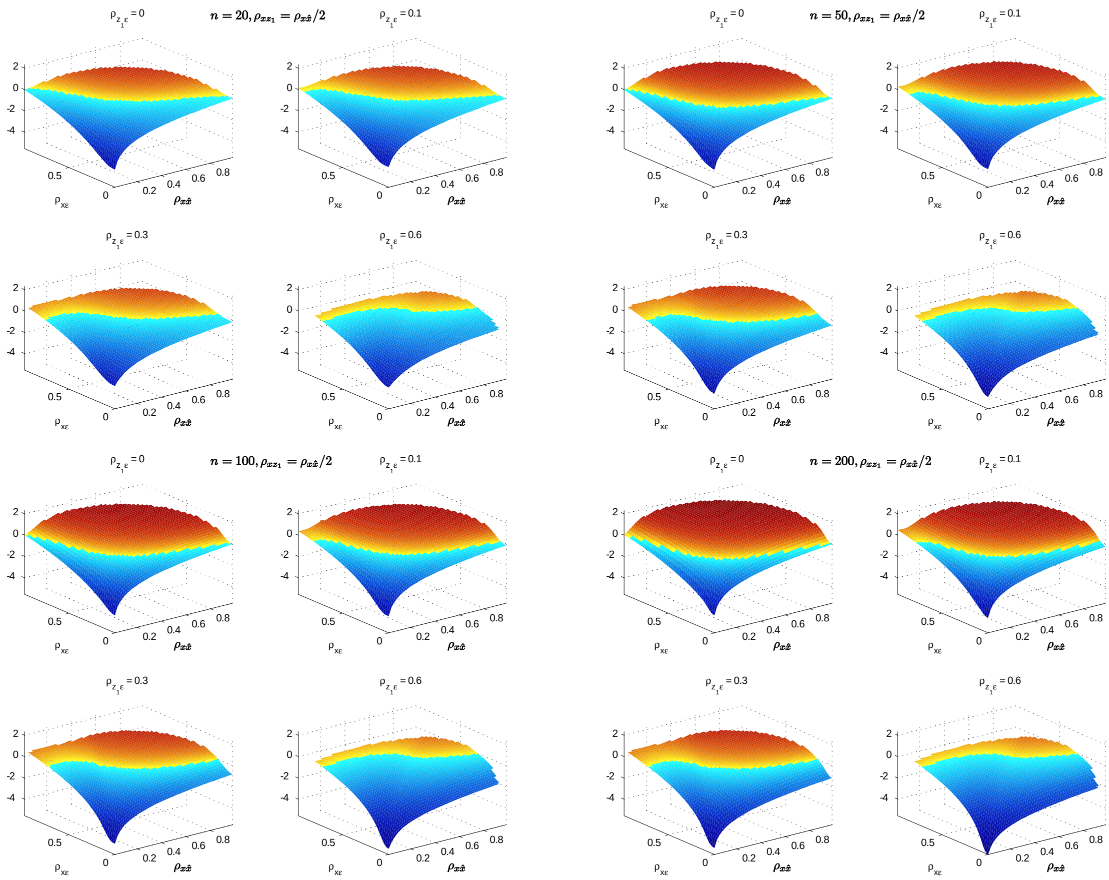


Figure 2.6: Conditioned $\log[MAE(\hat{\beta}_{OLS})/MAE(\hat{\beta}_{GIV})]$ for $k = l - 1 = 1$; any SN