TI 2000-039/4
Tinbergen Institute Discussion Paper

# Exact Test Statistics and Distributions of Maximum Likelihood Estimators that result from Orthogonal Parameters

*Frank R. Kleibergen*

**Tinbergen Institute**

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

**Tinbergen Institute Amsterdam**
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

**Tinbergen Institute Rotterdam**
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

# Exact Test Statistics and Distributions of Maximum Likelihood Estimators that result from Orthogonal Parameters

with applications to the Instrumental Variables Regression Model

Frank Kleibergen*

May 4, 2000

**Abstract**

We show that three convenient statistical properties that are known to hold for the linear model with normal distributed errors that: (*i.*) when the variance is known, the likelihood based test statistics, Wald, Likelihood Ratio and Score or Lagrange Multiplier, coincide, (*ii.*) when the variance is unknown, exact test statistics exist, (*iii.*) the density of the maximum likelihood estimator (mle) of the parameters of a nested model equals the conditional density of the mle of the parameters of an encompassing model, also apply to a larger class of models. This class contains models that are nested in a linear model and allow for orthogonal parameters to span the difference with the encompassing linear model. Next to linear models, an important set of models that belongs to this class are the reduced rank regression models. An example of a reduced rank regression model is the instrumental variables regression model. We use the three convenient statistical properties to conduct exact inference in the instrumental variables regression model and use them to construct both the density of the limited information maximum likelihood estimator and novel exact statistics to test instrument validity, over-identification and hypothezes on all or subsets of the structural form parameters.

# 1 Introduction

Maximum likelihood estimators (mles) in standard linear models with exogenous regressors and (mixture of) normal disturbances have a number of appealing statistical properties. Two well-known examples of these properties are that the likelihood based test statistics, *i.e.* the Wald, Likelihood Ratio and Lagrange Multiplier or Score statistic, are identical when the variance is a priori known, see *e.g.* Engle (1984), and that test statistics with an exact distribution exist when the variance is unknown, *i.e.* the $F$-statistics. Another convenient property concerns the density of the mle of the parameters of the linear model, which is the least squares estimator. This density is equal to the conditional density of the mle of the parameters

*Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Email:kleiberg@fee.uva.nl.

of an encompassing linear model given that the part of this mle that represents the difference between the nested and encompassing model is equal to zero. The latter property results from the normal distribution since the conditional densities of the elements of a bivariate normal distributed random vector are also normal. As a consequence of these three properties, we can both conduct tests on the parameters and construct the density of the mle in a straightforward way.

We show that the aforementioned properties hold for a more general class of statistical models than pure linear models which are also linear in the parameters. The convenient statistical properties namely apply to any model that is nested in a pure linear model and for which a set of orthogonal parameters, which concept we define lateron and implies amongst others the global orthogonality as defined by Cox and Reid (1987), spans the difference with the encompassing linear model.

Next to standard linear models, an important class of models that allow for the construction of orthogonal parameters are reduced rank regression models. Reduced rank regression models are commonly used in econometrics and some well-known representants of this class of models are, for example, the instrumental variables regression model, the error correction cointegration model, the factor model and the simultaneous equations model. We show that indeed the convenient statistical properties apply to these models, when they also satisfy the other assumptions, by analyzing the instrumental variables regression model. For the instrumental variables regression model, we construct the density of the mle, which is the limited information maximum likelihood (liml) estimator, and exact test statistics by using the convenient statistical properties that result from the orthogonal parameters.

The density of the liml estimator that we construct results from a different approach then the one traditionally pursued in the literature, see *e.g.* Mariano and Sawa (1973), Phillips (1983) and Anderson (1982). That approach constructs the density from a closed-form analytical expression of the liml estimator while our approach is based on a property which the liml estimator obeys, *i.e.* that it satisfies the first order condition. As a consequence, we obtain a different expression for the density then the one obtained previously. We compared our expression of the density with the sampling density of the liml estimator for a data generating process with strong endogeneity and for which we varied the quality of the instruments and the degree of over-identification. For all of the simulated data generating processes, the sampling and theoretical density of the liml estimator coincide. We also use the density of the liml estimator to illustrate the case of weak instruments, see *e.g.* Staiger and Stock (1997), for which we show that it leads to coinciding small sample and limiting distributions. In case of strong instruments, the density of the liml estimator converges to the density that corresponds with the limiting distribution which is known from the literature, see *e.g.* Hausman (1983).

Next to the marginal density of the liml estimator, we also obtain the marginal density of the mle of the orthogonal parameters. We use this mle to construct exact statistics with distributions that do not depend on nuisance parameters and that can be used to test hypothezes on the parameters of the instrumental variables regression model. This is remarkable as the distribution of the liml estimator is non-standard and depends on unobserved nuisance parameters. Because of these latter properties, the distributions of many statistics that test hypothezes on the parameters of the instrumental variables regression model depend on unobserved nuisance parameters, see *e.g.* Staiger and Stock (1997), Nelson and Startz (1990) and Zivot, Nelson and Startz (1998). The hypothezes that can be tested using the exact test statistics are: *i.* the validity of the instruments, *ii.* over-identification, *iii.* value of all the structural form parameters, *iv.* value of some of the structural form parameters. The Anderson-Rubin statistic, see Anderson and Rubin (1949), is one of the statistics that results (*i.*) but we also obtain novel exact statistics to conduct tests on (subsets of) the structural form parameters (*iii.* and *iv.*), whose distributions do not depend on the degree of over-identification. This is a well-known problem when one uses the Anderson-Rubin statistic to conduct tests on the structural form parameters. We also obtain a novel exact statistic to test for over-identification (*ii.*).

The paper is organized as follows. In section 2, we discuss the framework for constructing densities of mles and exact test statistics that results from the orthogonal parameters. Sections 3 and 4 apply the framework to the instrumental variables regression model. Section 3 constructs and discusses the distribution of the liml estimator while section 4 constructs and discusses the exact test statistics. Finally, the fifth section concludes.

Throughout the whole paper, the *vec* operator stands for the stacked columns of a matrix such that when $A = (a_1 \cdots a_N)$, $vec(A) = (a_1' \cdots a_N')'$. Most of the time, we consider the estimator of a parameter $a$, indicated by $\hat{a}$, as a random variable and not as a function of the data. For reasons of space, we do not explicitly distinguish these two cases. $a$ then represents the "true value".

# 2 Distribution Maximum Likelihood Estimator and Test Statistics

## 2.1 The First Order Condition

We construct the distribution of the maximum likelihood estimator (mle) for linear models, that can be non-linear in the parameters, with normal disturbances and construct it from the first order condition (foc) for a maximum of the likelihood. We thus consider the model

$$y = Xf(\varphi) + \varepsilon, \tag{1}$$

where $y$ is a $T \times 1$ vector that contains the $T$ observations of the dependent variable, $X$ is a $T \times k$ matrix of (weakly) exogenous regressors, $\varepsilon$ is a $T \times 1$ vector that contains the disturbances and is assumed to be distributed as $\varepsilon \sim N(0, \sigma^2 I_T)$, $I_T$ is the $T \times T$ dimensional identity matrix, $f(\varphi)$ is a $k \times 1$ vector function in the $m \times 1$ vector $\varphi$ that is continuous differentiable except (maybe) for some lower dimensional manifolds of the space on which $\varphi$ is defined, *i.e.* the $\mathbb{R}^m$, $k \geq m$. Examples of model (1) are the standard linear model and, in case of more than one equation, the instrumental variables regression model.

The first order condition for a maximum of the (log-) likelihood reads

$$\begin{aligned}
\tfrac{1}{\sigma^2}\left(\tfrac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)' X'\left(y - Xf(\hat{\varphi})\right) &= 0 \Leftrightarrow \\
\tfrac{1}{\sigma^2}\left(\tfrac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)' X'X\left((X'X)^{-1}X'y - f(\hat{\varphi})\right) &= 0 \Leftrightarrow \\
\left((X'X)^{\frac{1}{2}}\tfrac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\sigma^{-1}\right)'\left((X'X)^{\frac{1}{2}}\hat{\Phi}\sigma^{-1} - (X'X)^{\frac{1}{2}}f(\hat{\varphi})\sigma^{-1}\right) &= 0 \Leftrightarrow \\
\left(\tfrac{\partial r}{\partial \varphi'}|_{\hat{\varphi}}\right)'\left(\hat{\Theta} - r(\hat{\psi}(\hat{\varphi}))\right) &= 0 \Leftrightarrow \\
\left(\left(\tfrac{\partial r}{\partial \psi'}|_{\hat{\psi}(\hat{\varphi})}\right)\left(\tfrac{\partial \psi}{\partial \varphi'}|_{\hat{\varphi}}\right)\right)'\left(\hat{\Theta} - r(\hat{\psi}(\hat{\varphi}))\right) &= 0 \Leftrightarrow \\
\left(\tfrac{\partial \psi}{\partial \varphi'}|_{\hat{\varphi}}\right)'\left(\tfrac{\partial r}{\partial \psi'}|_{\hat{\psi}(\hat{\varphi})}\right)'\left(\hat{\Theta} - r(\hat{\psi}(\hat{\varphi}))\right) &= 0,
\end{aligned} \tag{2}$$

where $\hat{\psi}$ and $\hat{\varphi}$ are the mles of the respective parameters, $\hat{\Phi} = (X'X)^{-1}X'y$; $\hat{\Theta} = (X'X)^{\frac{1}{2}}\hat{\Phi}\sigma^{-1}$, $r(\hat{\psi})$ is a $k \times 1$ vector function in the $m \times 1$ (random) variable $\hat{\psi}$ which is such that $r(\hat{\psi}) =$

$(X'X)^{\frac{1}{2}} f(\hat{\varphi})\sigma^{-1}$ and an invertible relationship between $\hat{\psi}$ and $\hat{\varphi}$ exists such that $\frac{\partial\hat{\varphi}}{\partial\hat{\psi}'}$ is non-singular for all $\hat{\varphi}$, $\hat{\psi}$. We have also used the chain rule of differentiation, $\frac{\partial r}{\partial\varphi'}|_{\hat{\varphi}} = \left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}(\hat{\varphi})}\right)\left(\frac{\partial\psi}{\partial\varphi'}|_{\hat{\varphi}}\right)$. We make the assumption.

**Assumption 1:** $\left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}}\right)$ *has full rank for all values of* $\hat{\psi}$ *except maybe on lower dimensional manifolds of the space of* $\hat{\psi}$, *i.e. the* $\mathbb{R}^m$.

Assumption 1 is equivalent with assuming that $\frac{\partial f}{\partial\varphi'}|_{\varphi}$ has full rank for all values of $\varphi$ except maybe for lower dimensional manifolds of the space of $\varphi$.

## 2.2 Orthogonal Parameters

The foc (2) can also directly be specified in terms of $\hat{\psi}$,

$$\left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}}\right)'\left(\hat{\Theta} - r(\hat{\psi})\right) = 0, \tag{3}$$

such that we can solve for $\hat{\psi}$ from the foc (3) and then obtain $\hat{\varphi}$ from $\hat{\psi}$ as $\hat{\varphi}$ and $\hat{\psi}$ entertain an invertible relationship. In foc (3), $\hat{\Theta}$ is a $k \times 1$ vector while $\hat{\psi}$ is a $m \times 1$ vector. $\hat{\psi}$ is consequently over-identified and no unique solution of $\hat{\Theta}$ given $r(\hat{\psi})$ to the foc (3) exists. To express the solutions of $\hat{\Theta}$ given $r(\hat{\psi})$ to the foc (3) in a convenient way we make the following assumption about the specification of $\hat{\Theta}$.

**Assumption 2:** *An invertible relationship between the* $k \times 1$ *vector* $\hat{\Theta}$ *and the* $m \times 1$ *and* $(k-m) \times 1$ *vectors* $\hat{\psi}$ *and* $\hat{\lambda}$ *exists which reads*

$$\hat{\Theta} = r(\hat{\psi}) + q(\hat{\psi})\hat{\lambda} \tag{4}$$

*and where* $q(\hat{\psi})$ *results from definition 1.*

**Definition 1** $q(\hat{\psi})$ *is a* $k \times (k-m)$ *matrix function of the* $m \times 1$ *vector* $\hat{\psi}$ *which satisfies*

$$\begin{array}{llr} \textit{i.} & r(\hat{\psi})'q(\hat{\psi}) & \equiv 0 \\ \textit{ii.} & \left(\frac{\partial r}{\partial\psi'}\right)' q(\hat{\psi}) & \equiv 0 \\ \textit{iii.} & q(\hat{\psi})'q(\hat{\psi}) & \equiv I_{k-m}. \end{array} \tag{5}$$

Assumption 2 essentially contains two conditions as it both implies that an invertible relationship between $\hat{\Theta}$ and $(\hat{\psi}, \hat{\lambda})$ exists and that $\hat{\Theta}$ is a linear function of $\hat{\lambda}$ as specified in (4). Definition 1 also consists of more than one condition. Its second condition partly results from the first condition such that models exist for which $q(\hat{\psi})$ satisfies all three conditions jointly although the number of restrictions in (5) $(k(k-m+1))$ exceeds the total number of elements of $q(\hat{\psi})$ $(k(k-m))$. Examples of these models are linear models and reduced rank regression models. For many other models it is not possible to construct $q(\hat{\psi})$ such that it accords with all of the conditions in definition 1. These models then do not allow for the kind of analysis that we conduct lateron.

$q(\hat{\psi})$ spans the space of solutions to the foc (3) since we can add $q(\hat{\psi})\hat{\lambda}$ to every solution $\hat{\Theta} = r(\hat{\psi})$, for every $\hat{\lambda}$ that is an element of the $\mathbb{R}^{k-m}$, and still the foc (3) is satisfied. The solutions to the foc (3) thus span a $k-m$ dimensional manifold in the space of $\hat{\Theta}$, which is the $\mathbb{R}^k$, see Hillier and Armstrong (1998) and Tjur (1980). Values of $\hat{\Theta}$, like $\hat{\Theta} = r(\hat{\psi}) + q(\hat{\psi})\hat{\lambda}$, thus exist that satisfy the foc (3) but which essentially do not correspond with the analyzed model (1) because no $f(\hat{\varphi})$ exists such that $f(\hat{\varphi}) = (X'X)^{-\frac{1}{2}}\hat{\Theta}\sigma$.

Since an invertible relationship exists between $\hat{\Theta}$ and $(\hat{\psi}, \hat{\lambda})$, we can construct the density of $(\hat{\psi}, \hat{\lambda})$ from the density of $\hat{\Theta}$. $\hat{\Theta}$ results from the least squares estimator $\hat{\Phi}$, *i.e.* it consists of the "$t$-values of $\hat{\Phi}$", which has a known distribution

$$\hat{\Phi} \sim N(\Phi, \sigma^2(X'X)^{-1}), \tag{6}$$

where $\Phi = f(\varphi)$, such that the distribution of $\hat{\Theta}$ reads

$$\hat{\Theta} \sim N(\Theta, I_k), \tag{7}$$

where $\Theta = r(\psi) = (X'X)^{\frac{1}{2}} f(\varphi)\sigma^{-1}$, with density

$$p(\hat{\Theta}) \propto \exp\left[-\frac{1}{2}\left(\hat{\Theta} - \Theta\right)'\left(\hat{\Theta} - \Theta\right)\right]. \tag{8}$$

The density of $(\hat{\psi}, \hat{\lambda})$ then results after an appropriate transformation of random variables,

$$p(\hat{\psi}, \hat{\lambda}) \propto p(\hat{\Theta}(\hat{\psi}, \hat{\lambda}))\left|J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))\right|, \tag{9}$$

where $J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))$ is the jacobian of the transformation from $\hat{\Theta}$ to $(\hat{\psi}, \hat{\lambda})$, and we obtain the density of the mle by integrating (9) over $\hat{\lambda}$, see Hillier and Armstrong (1999) and Tjur (1980),

$$p(\hat{\psi}) = \int_{\mathbb{R}^{k-m}} p(\hat{\psi}, \hat{\lambda})d\hat{\lambda}. \tag{10}$$

The density of the mle can, however, also be obtained by using the specification of $\hat{\Theta}$ (4) and $q(\hat{\psi})$ (5). This specification implies *a.o.* a special structure for the jacobian that is stated in theorem 2.

**Theorem 2** *The Jacobian of the transformation from $\hat{\Theta}$ to $(\hat{\psi}, \hat{\lambda})$ is characterized by*

$$\begin{aligned} &\left|J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))\right| \\ &= \left|\begin{pmatrix} \left(\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right) + \left(\hat{\lambda}' \otimes I_k\right)\left(\frac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right)\right)'\left(\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right) + \left(\hat{\lambda}' \otimes I_k\right)\left(\frac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right)\right) & 0 \\ 0 & I_{k-m} \end{pmatrix}\right|^{\frac{1}{2}} \\ &= \left|\left(\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right) + \left(\hat{\lambda}' \otimes I_k\right)\left(\frac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right)\right)'\left(\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right) + \left(\hat{\lambda}' \otimes I_k\right)\left(\frac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right)\right)\right|^{\frac{1}{2}}. \end{aligned} \tag{11}$$

**Proof.**

$$\begin{aligned} \left|J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))\right| &= \left|\begin{pmatrix} \frac{\partial\hat{\Theta}}{\partial\hat{\psi}'} & \frac{\partial\hat{\Theta}}{\partial\hat{\lambda}'} \end{pmatrix}\right| \\ &= \left|\begin{pmatrix} \left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right) + \left(\hat{\lambda}' \otimes I_k\right)\left(\frac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right) & q(\hat{\psi}) \end{pmatrix}\right| \\ &= \left|J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))'J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))\right|^{\frac{1}{2}}. \end{aligned}$$

Definition 1 implies that $q(\hat{\psi})'q(\hat{\psi}) \equiv I_{k-m}$, such that

$$\begin{aligned}\left[\left(I_{k-m} \otimes q(\hat{\psi})'\right) I_{k(k-m)} + \left(q(\hat{\psi})' \otimes I_{k-m}\right) K_{k,(k-m)}\right] \left(\tfrac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right) &= 0 \Leftrightarrow \\ \left[K_{(k-m),(k-m)} + I_{(k-m)^2}\right] \left(I_{k-m} \otimes q(\hat{\psi})'\right) \left(\tfrac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right) &= 0,\end{aligned}$$

where $K_{k,(k-m)}$ is a $k(k-m)\times k(k-m)$ dimensional commutation matrix such that $K_{k,(k-m)}vec(A') = vec(A)$ with $A$ a $k(k-m) \times k(k-m)$ matrix, see Magnus and Neudecker (1988), and which implies that

$$\left(I_{k-m} \otimes q(\hat{\psi})'\right) \left(\frac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right) = 0.$$

Combining this with the definition, $\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)' q(\hat{\psi}) \equiv 0$, we obtain

$$\left(\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right) + \left(\hat{\lambda}' \otimes I_k\right) \left(\frac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}}\right)\right)' q(\hat{\psi}) = 0.$$

These results imply that $\left|J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))'J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))\right|^{\frac{1}{2}}$ is equal to (11). ■

Theorem 2 shows that the random variables $\hat{\psi}$ and $\hat{\lambda}$ are globally orthogonal as defined by Cox and Reid (1987). This results as the information matrix of $(\hat{\psi}, \hat{\lambda})$, which equals $J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))'J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))$, is block diagonal. As a consequence, the conditional density of $\hat{\psi}$ given $\hat{\lambda}$ only minorly depends on $\hat{\lambda}$. The global orthogonality results only from conditions (*ii.*) and (*iii.*) from definition 1 and condition (*i.*) is redundant for it. The orthogonality that results from definition 1 is therefore stronger, and thus more restrictive, than the global orthogonality defined by Cox and Reid.

## 2.3 The Density of the MLE and the Score Vector

In appendix A, we construct the marginal densities of the mle $\hat{\psi}$ and the random variable $\hat{\lambda}$ which, is both orthogonal to $\hat{\psi}$ and spans the difference with the encompassing linear model and, is equal to the score vector. Theorem 3 states these marginal densities.

**Theorem 3** *When $r(\hat{\psi})$ and $\hat{\Theta}$ satisfy assumptions 1, 2, the conditions from definition 1 and $\Theta = r(\psi)$, the marginal density of $\hat{\psi}$ reads*

$$\begin{aligned} p(\hat{\psi}) &= \textstyle\int_{\mathbb{R}^{k-m}} p(\hat{\psi}, \hat{\lambda}) d\hat{\lambda} \\ &= p(\hat{\psi}|\hat{\lambda} = 0) \\ &\propto \left|\left(\tfrac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)' \left(\tfrac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)\right|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}\left(r(\hat{\psi}) - r(\psi)\right)' \left(r(\hat{\psi}) - r(\psi)\right)\right], \end{aligned} \tag{12}$$

*and the marginal density of the score vector $\hat{\lambda}$ that results from assumption 2 and definition 1 reads*

$$p(\hat{\lambda}) \propto \exp\left[-\frac{1}{2}\hat{\lambda}'\hat{\lambda}\right]. \tag{13}$$

**Proof.** see Appendix A. ■

The marginal density of $\hat{\psi}$ (12) is proportional to the conditional density of $(\hat{\psi}, \hat{\lambda})$ given that $\hat{\lambda} = 0$,

$$\begin{aligned} p(\hat{\psi}) & = p(\hat{\psi}|\hat{\lambda} = 0) \\ & \propto p(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0} \\ & \propto p(\hat{\Theta}(\hat{\psi}, \hat{\lambda}))|_{\hat{\lambda}=0} \left| J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))|_{\hat{\lambda}=0} \right|, \end{aligned} \tag{14}$$

and corresponds with the conditional densities given a conditioning statistic that are constructed in, *e.g.*, Barndorf-Nielsen (1980,1983). Equation (14) therefore shows that, when the analyzed model (1) allows for the orthogonal parameters stated in definition 1, the marginal density of the mle is equal to the conditional density of the mle given the orthogonal conditioning statistic evaluated at the parameter point where the orthogonal conditioning statistic is equal to zero. Cox and Reid (1987) argue that in case $\hat{\psi}$ and $\hat{\lambda}$ are globally orthogonal that the conditional density of $\hat{\psi}$ given $\hat{\lambda}$ only minorly depends on $\hat{\lambda}$. As the orthogonality that results from definition 1 is stronger than global orthogonality, since we added condition (*i.*), we are also able to make a stronger statement, that the marginal density of $\hat{\psi}$ is equal to the conditional density given that $\hat{\lambda} = 0$. This results as, because of condition (*i.*), not only the information matrix is block-diagonal, which resulted from conditions (*ii.*)-(*iii.*), but also no products of $r(\hat{\psi})$ and $\hat{\lambda}$ in the exponent term of the density of $\hat{\Theta}$ appear. The only element of the exponent term where $\hat{\psi}$ and $\hat{\lambda}$ appear jointly is $\Theta' q(\hat{\psi})\hat{\lambda}$, as $\hat{\lambda}' q(\hat{\psi})' q(\hat{\psi})\hat{\lambda} = \hat{\lambda}'\hat{\lambda}$, that solely operates in the space orthogonal to $r(\hat{\psi})$. It therefore does not interfere with $r(\hat{\psi})$ and the marginal density of $\hat{\psi}$. The orthogonality of $\hat{\psi}$ and $\hat{\lambda}$ comes close to stochastic independence which implies that the marginal density of $\hat{\psi}$ is equal to the conditional density given any value of $\hat{\lambda}$ while it is only equal to the conditional density given that $\hat{\lambda}$ is equal to zero in case of orthogonality.

The specification of $(\hat{\psi}, \hat{\lambda})$ from definition 1 satisfies the sufficient conditions for a unique conditional density of $\hat{\Theta}$ given that $\hat{\Theta} = r(\hat{\psi})$, see Kleibergen (2000). The density (14) results from this unique conditional density

$$p(\hat{\Theta})|_{\hat{\Theta}=r(\hat{\psi})} \propto p(\hat{\psi}(\hat{\Theta}), \hat{\lambda}(\hat{\Theta}))|_{\hat{\Theta}=r(\hat{\psi})} \left| J((\hat{\psi}, \hat{\lambda}), \hat{\Theta})|_{\hat{\Theta}=r(\hat{\psi})} \right|, \tag{15}$$

where $p(\hat{\psi}(\hat{\Theta}), \hat{\lambda}(\hat{\Theta}))|_{\hat{\Theta}=r(\hat{\psi})} \propto p(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}$. So, although values of $\hat{\Theta}$ exist that satisfy the foc (3) but which essentially do not correspond with the analyzed model (1), when we integrate over these values to obtain the density of the mle $\hat{\psi}$, we obtain a density of the mle that can be considered to consist of just those solutions to the foc that exactly correspond with the analyzed model. Thus the solution of $\hat{\Theta}$ to the foc (3) given $r(\hat{\psi})$ is not uniquely defined but we can construct the density of $\hat{\psi}$ from the unique conditional density of $\hat{\Theta}$ given that $\hat{\Theta} = r(\hat{\psi})$.

The marginal density (13) of $\hat{\lambda}$ can be used to construct statistics that test the restriction imposed on $\hat{\Theta}$, $\hat{\Theta} = r(\hat{\psi})$, that leads to the marginal density of $\hat{\psi}$ (14). As a result of definition 1, $q(\hat{\psi})$ is both orthogonal to $r(\hat{\psi})$ and $\frac{\partial r}{\partial \psi'}$. As a consequence, in case of a known variance, $\hat{\lambda}'\hat{\lambda}$ is equal to all three of the likelihood based test statistics, *i.e.* the Score, Wald and Likelihood Ratio statistic. The equality of these three test statistics, in case of a known value of the variance, thus holds when we can represent the difference between the null and alternative hypothesis with orthogonal parameters, which occurs, for example, when we test in linear models.

## 2.4 The Density of the MLE with known variance

The marginal density of the mle (12) is defined in terms of the parameters of the orthogonal specification from definition 1. This density can be transformed into the density of the mle of the parameters of the orginal model (1) which then becomes

$$p(\hat{\varphi}) \propto \left| \sigma^{-2} \left( \frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}} \right)' X'X \left( \frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}} \right) \right|^{\frac{1}{2}} \exp \left[ -\frac{1}{2\sigma^2} \left( f(\hat{\varphi}) - f(\varphi) \right)' X'X \left( f(\hat{\varphi}) - f(\varphi) \right) \right]. \quad (16)$$

The density (16) consists of the square root of the determinant of the observed information matrix and the exponent term of a normal density. All these elements can directly be constructed once we have specified the model (1) as the orthogonal parameters, that are defined in definition 1, are not directly present in (16). These orthogonal parameters are, however, crucial for constructing (16) and (16) is not a valid reflection of the density of the mle when we can not construct orthogonal parameters that satisfy *all* the conditions from assumption 2 and definition 1. Especially the first two conditions in definition 1 are important in this respect as they both need to be satisfied. These two conditions imply that the number of restrictions on $q(\hat{\psi})$ exceeds the number of elements of $q(\hat{\psi})$. For many specifications of $f(\varphi)$ we are therefore only able to construct $q(\hat{\psi})$ such that just the second and third conditions are satisfied. This shows that we explicitly have to verify the conditions from definition 1 and that we can not apply (16) mechanically.

## 2.5 The Density of the MLE using a Variance Estimator

When $\sigma^2$ is unknown, we use the joint density of $(\hat{\psi}, \hat{\lambda})$ and an unbiased variance estimator that is stochastic independent from $(\hat{\psi}, \hat{\lambda})$ to construct the density of the mle. A convenient estimator of $\sigma^2$ for this purpose is based on the estimator that results from the least squares regression of $X$ on $y$,

$$s^2 = \frac{1}{T-k} y' M_X y = \frac{1}{T-k} (y - X\hat{\Phi})'(y - X\hat{\Phi}), \quad (17)$$

which is distributed as

$$(T-k)\frac{s^2}{\sigma^2} \sim \chi^2(T-k), \quad (18)$$

and stochastic independent from $\hat{\Phi}$. The convenient estimator $\hat{\sigma}^2$ then results as

$$\hat{\sigma}^2 = \sigma^2 \left( s^2 \right)^{-1} \sigma^2 \quad (19)$$

and is distributed as

$$\hat{\sigma}^2 \sim iW((T-k)\sigma^2, T-k), \quad (20)$$

where $iW$ stands for inverted-Wishart, with density

$$p(\hat{\sigma}^2) \propto \left| \hat{\sigma}^2 \right|^{-\frac{1}{2}(T-k+2)} \exp \left[ -\frac{(T-k)\sigma^2}{2\hat{\sigma}^2} \right]. \quad (21)$$

Note that the estimator $\hat{\sigma}^2$ defined in (19) depends on the unknown variance $\sigma^2$ and can therefore not be constructed as such. We, however, only use $\hat{\sigma}^2$ to construct the density of the mle $\hat{\varphi}$ as it leads to a more convenient expression of this density than $s^2$. Since $s^2$ is stochastic independent from the least squares estimator $\hat{\Phi}$ also $\hat{\sigma}^2$ is stochastic independent of $\hat{\Phi}$. The resulting density of $(\hat{\psi}, \hat{\lambda}, \hat{\sigma}^2)$ then becomes

$$p(\hat{\psi}, \hat{\lambda}, \hat{\sigma}^2) \propto p(\hat{\Theta}(\hat{\psi}, \hat{\lambda}))p(\hat{\sigma}^2) \left| J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda})) \right|, \tag{22}$$

where $\hat{\Theta} = (X'X)^{\frac{1}{2}}\hat{\Phi}\sigma^{-1}$ and is stochastic independent from $\hat{\sigma}^2$. Identical to (14), we obtain the joint density of the mle $\hat{\psi}$ and $\hat{\sigma}^2$ as the conditional density of $(\hat{\psi}, \hat{\sigma}^2)$ given that $\hat{\lambda} = 0$

$$p(\hat{\psi}, \hat{\sigma}^2) \propto p(\hat{\psi}, \hat{\lambda}, \hat{\sigma}^2)|_{\hat{\lambda}=0}. \tag{23}$$

Instead of $\sigma^2$, we now use $\hat{\sigma}^2$ to solve for $\hat{\varphi}$ from $\hat{\psi}$ to obtain the joint density of $(\hat{\varphi}, \hat{\sigma}^2)$,

$$\begin{aligned} p(\hat{\varphi}, \hat{\sigma}^2) &\propto p(\hat{\psi}(\hat{\varphi}), \hat{\sigma}^2) \left| J(\hat{\psi}, \hat{\varphi}) \right| \\ &\propto \left| \hat{\sigma}^{-2} \left(\frac{\partial f}{\partial \varphi'}\right)' X'X \left(\frac{\partial f}{\partial \varphi'}\right) \right|^{\frac{1}{2}} \exp\left[-\frac{1}{2\hat{\sigma}^2} \left(f(\hat{\varphi}) - f(\varphi)\right)' X'X \left(f(\hat{\varphi}) - f(\varphi)\right)\right] \\ &\quad \left|\hat{\sigma}^2\right|^{-\frac{1}{2}(T-k+2)} \exp\left[-\frac{(T-k)\sigma^2}{2\hat{\sigma}^2}\right]. \end{aligned} \tag{24}$$

Although $\hat{\sigma}^2$ is stochastic independent from $\hat{\psi}$, $\hat{\sigma}^2$ and $\hat{\varphi}$ are not stochastic independent since we use $\hat{\sigma}^2$ to solve $\hat{\varphi}$ from $\hat{\psi}$. As a consequence, the marginal density of $\hat{\sigma}^2$ that results from (24) differs from (21). We show an example of this in section 3.3.

The joint density (24) shows why we use $\hat{\sigma}^2$ instead of $s^2$, as the estimator of $\sigma^2$, since the exponent term of the joint density now only has $\hat{\sigma}^2$ in the denominator. This allows us to obtain some further analytical results as we show in section 3.3 for the limited information maximum likelihood estimator.

## 2.6 The Distribution of Score, Wald and Likelihood Ratio Statistics

The score vector, who's marginal density is (13), can be obtained from the second and third condition from definition 1

$$\begin{aligned} q(\hat{\psi}) &= (X'X)^{-\frac{1}{2}} \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp \sigma \left(\sigma^2 \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp' (X'X)^{-1} \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp\right)^{-\frac{1}{2}}, \\ \hat{\lambda} &= q(\hat{\psi})'\hat{\Theta} = \left(\sigma^2 \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp' (X'X)^{-1} \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp\right)^{-\frac{1}{2}} \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp' (X'X)^{-1}X'y, \end{aligned} \tag{25}$$

where we use the least squares estimator $(q(\hat{\psi})'q(\hat{\psi}))^{-1}q(\hat{\psi})'\hat{\Theta} = q(\hat{\psi})'\hat{\Theta}$ to obtain $\hat{\lambda}$ from $\hat{\Theta}$ as $q(\hat{\psi})$ is orthogonal to $r(\hat{\psi})$, such that

$$\begin{aligned} \hat{\lambda}'\hat{\lambda} &= y'X(X'X)^{-1} \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp \left(\sigma^2 \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp' (X'X)^{-1} \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp\right)^{-1} \left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)_\perp' (X'X)^{-1}X'y \\ &= \frac{1}{\sigma^2} y' \left[M_{X\left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)} - M_X\right] y \\ &\sim \chi^2(k-m), \end{aligned} \tag{26}$$

where $M_V = I_T - V(V'V)^{-1}V'$, $V = X\left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)$, $V = X$, and which tests the null-hypothesis $H_0 : \Phi = f(\varphi)$ against $H_1 : \Phi \neq f(\varphi)$.

We use the expression of the score vector to construct the statistic $\hat{\lambda}'\hat{\lambda}$. Because of the orthogonality of $q(\hat{\psi})$ to both $(X'X)^{-\frac{1}{2}}\left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)$ and $(X'X)^{-\frac{1}{2}}f(\hat{\varphi})$, which results from definition 1, $\hat{\lambda}'\hat{\lambda}$ is equal to all three of the likelihood based test statistics, *i.e.* Score, Wald and Likelihood Ratio, when the variance is known. We only use the second and third condition of definition 1 to construct the score in (25) but also the first condition has to hold for (26) to be valid. This first condition is not automatically satisfied when the second and third condition hold such that we have to verify it explicitly.

When the value of $\sigma^2$ is unknown, we use the estimator $s^2$ (17) in (26) and divide (26) by both $\frac{s^2}{\sigma^2}$ and $k-m$ such that it becomes,

$$\begin{aligned} F(H_0|H_1) &= \tfrac{1}{(k-m)s^2} y' \left[ M_{X\left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)} - M_X \right] y \\ &= \tfrac{1}{k-m} \tfrac{\hat{\lambda}'\hat{\lambda}}{s^2/\sigma^2} \\ &= \frac{\left(\frac{1}{\sigma^2} y' \left[ M_{X\left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)} - M_X \right] y\right) \Big/ (k-m)}{\left(\frac{(T-k)s^2}{\sigma^2}\right) \Big/ (T-k)} \\ &\sim \tfrac{\chi^2(k-m)/(k-m)}{\chi^2(T-k)/(T-k)} \\ &\sim F(k-m, T-k), \end{aligned} \tag{27}$$

since $(\hat{\psi}, \hat{\lambda})$ are stochastic independent from $s^2$. Equation (27) is related to both the Likelihood Ratio statistic, because it uses the mle under $H_0$, $\hat{\varphi}$, and the Wald statistic, since the covariance matrix $s^2$ is computed under $H_1$. It shows that we can construct an exact statistic to test the null hypothesis of model (1) against the alternative hypothesis of an encompassing linear model when we are able to construct a set of orthogonal parameters that satisfy all conditions from definition 1.

The statistic in (27) tests the analyzed model (1) against an encompassing linear model. The degrees of freedom of the $F$ distribution of the statistic, $k-m$, can then be quite large which affects the power of the statistic. We can often also construct exact statistics that test a nested model against an encompassing model, that both allow for the construction of orthogonal parameters which satisfy the conditions from assumption 2 and definition 1, where the encompassing model doesnot have to be linear. Examples of this in case of the instrumental variables regression model are discussed in section 4.

In the next sections, we apply the above results to conduct exact inference in the instrumental variables regression model. We both construct the density of the limited information maximum likelihood estimator, for which we show that the resulting density accurately represents the sampling density, and exact test statistics.

# 3 Density of the LIML Estimator

Next to standard linear models, an important class of models that allow for the construction of the orthogonal parameters, that result from assumption 2 and definition 1, are the so-called reduced rank regression models. Examples of reduced rank regression models in econometrics and time series analysis are cointegration, factor and simultaneous equation models. We use the methodology developed in the previous section to construct the density of the mle of the parameters of the instrumental variables regression model, *i.e.* the limited information maximum likelihood (liml) estimator, which belongs to the latter class of reduced rank regression models. In the next section, we also construct exact statistics to test hypothezes on the parameters of this model.

## 3.1 Instrumental Variable Regression Model

The instrumental variables regression model in *structural form* can be represented as a limited information simultaneous equation model, see *e.g.* Hausman (1983) and Kleibergen and Zivot (1998),

$$\begin{aligned} y_1 &= Y_2\beta + Z\gamma + \varepsilon_1 \\ Y_2 &= X\Pi + Z\Gamma + V_2, \end{aligned} \tag{28}$$

where $y_1$ and $Y_2$ are a $T \times 1$ and $T \times (m-1)$ matrix of endogenous variables, respectively, $Z$ is a $T \times k_1$ matrix of included exogenous variables, $X$ is a $T \times k_2$ matrix of excluded exogenous variables (or instruments), $\varepsilon_1$ is a $T \times 1$ vector of structural errors and $V_2$ is a $T \times (m-1)$ matrix of reduced form errors. The $(m-1) \times 1$ and $k_1 \times 1$ parameter vectors $\beta$ and $\gamma$ contain the structural parameters. The variables in $X$ and $Z$ are assumed to be of full column rank, uncorrelated with $\varepsilon_1$ and $V_2$, and weakly exogenous for the parameters $\beta$ and $\Pi$, see Engle *et. al.* (1983). The error terms $\varepsilon_{1t}$ and $V_{2t}$, where $\varepsilon_{1t}$ denotes the $t$-th observation on $\varepsilon_1$ and $V_{2t}$ is a column vector denoting the $t$-th row of $V_2$, are assumed to be normally distributed with zero mean, and to be serially uncorrelated and homoskedastic with $m \times m$ covariance matrix

$$\Sigma = var \begin{pmatrix} \varepsilon_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{29}$$

which is assumed to be unknown. The degree of endogeniety of $Y_2$ in (28) is determined by the vector of correlation coefficients defined by $\rho = \Sigma_{22}^{-1/2\prime}\Sigma_{21}\sigma_{11}^{-1/2}$ and the quality of the instruments is captured by $\Pi$.

Substituting the reduced form equation for $Y_2$ into the structural equation for $y_1$ gives the non-linearly *restricted reduced form* specification

$$Y = X\Pi B + Z\Psi + V, \tag{30}$$

where $Y = \begin{pmatrix} y_1 & Y_2 \end{pmatrix}$, $B = \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$, $\Psi = \Gamma B + \begin{pmatrix} \gamma & 0 \end{pmatrix}$, $V = \begin{pmatrix} v_1 & V_2 \end{pmatrix}$, $v_1 = \varepsilon_1 + V_2\beta$ and, hence, $(v_{1t}\ V_{2t}')'$ has covariance matrix

$$\Omega = var \begin{pmatrix} v_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} e_1' \\ B \end{pmatrix}' \Sigma \begin{pmatrix} e_1' \\ B \end{pmatrix}, \tag{31}$$

where $e_1 : m \times 1$ is the first $m$ dimensional unity vector. Note that $\Psi$ is a unrestricted $k_1 \times m$ matrix.

The *unrestricted reduced form* of the model expresses each endogenous variable as a linear function of the exogenous variables and is given by

$$Y = X\Phi + Z\Psi + V, \tag{32}$$

where $\Phi : k_2 \times m$. Since the unrestricted reduced form is a multivariate linear regression model, all of the reduced form parameters are identified. It is assumed that $k_2 \geq m-1$ so that the structural parameter vector $\beta$ is "apparently" identified by the order condition. We call the model just-identified when $k_2 = m-1$ and the model over-identified when $k_2 > m-1$. $k_2 - m + 1$

is therefore the degree of over-identification. $\beta$ is identified if and only if $rank(\Pi) = m - 1$. The extreme case in which $\beta$ is totally unidentified occurs when $\Pi = 0$ and, hence, $rank(\Pi) = 0$, see Phillips (1989). The case of "weak instruments", as discussed by Nelson and Startz (1990), Staiger and Stock (1997), Wang and Zivot (1998), and Zivot, Nelson and Startz (1998), occurs when $\Pi$ is close to zero or, as discussed by Kitamura (1994), Dufour and Khalaf (1997) and Shea (1997) when $\Pi$ is close to having reduced rank.

The parameter $\beta$ is typically the focus of the analysis. We can therefore simplify the presentation of the results without changing their implications by setting $\gamma = 0$ and $\Gamma = 0$ ($\Psi = 0$) so that $Z$ drops out of the model. In what follows, let $k = k_2$ denote the number of instruments. We note that the form of the analytical results for $\beta$ in this simplified case carry over to the more general case where $\gamma \neq 0$ and $\Gamma \neq 0$ by interpreting all data matrices as residuals from the projection on $Z$.

## 3.2 LIML estimator

The mle of $\beta$, $\hat{\beta}$, is obtained from the concentrated log-likelihood that results when we have concentrated out $\Pi$ and $\Sigma$ from the log-likelihood of the parameters of the model (28), see *e.g.* Hausman (1983),

$$\begin{aligned} \log(L(\beta|X,Y)) &= \tfrac{1}{2}T\log\left|\tfrac{(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)}{(y_1-Y_2\beta)'(y_1-Y_2\beta)}\right| \\ &= \tfrac{1}{2}T\log\left|1-\tfrac{(y_1-Y_2\beta)'X(X'X)^{-1}X'(y_1-Y_2\beta)}{(y_1-Y_2\beta)'(y_1-Y_2\beta)}\right| \\ &= \tfrac{1}{2}T\log\left|1-\eta\right|, \end{aligned} \tag{33}$$

where $\eta = \frac{(y_1-Y_2\beta)'X(X'X)^{-1}X'(y_1-Y_2\beta)}{(y_1-Y_2\beta)'(y_1-Y_2\beta)}$. Since the concentrated log-likelihood of $\beta$ is a monotonic decreasing function of $\eta$, maximizing with respect to $\beta$ is identical to finding the minimal value of $\eta$

$$\eta = \min_{\beta}\left[\frac{(y_1-Y_2\beta)'X(X'X)^{-1}X'(y_1-Y_2\beta)}{(y_1-Y_2\beta)'(y_1-Y_2\beta)}\right], \tag{34}$$

which is identical to solving the eigenvalue problem

$$\begin{aligned} |\eta Y'Y - Y'X(X'X)^{-1}X'Y| &= 0 \Leftrightarrow \\ |\eta I_m - (Y'Y)^{-1}\hat{\Phi}X'X\hat{\Phi}| &= 0, \end{aligned} \tag{35}$$

where $\hat{\Phi} = (X'X)^{-1}X'Y$, and to use the smallest root of (35), see Anderson and Rubin (1949) and Hood and Koopmans (1953). The liml estimator of $\beta$, $\hat{\beta}$, is then contructed such that the eigenvector associated with $\eta$ equals $a(1\ -\hat{\beta}')'$, where $a$ is the first element of the eigenvector associated with $\eta$.

## 3.3 Reduced Rank Restriction on Random Matrix

The liml estimator constructed previously uses the likelihood of the structural form (28). The foc (2), that we use to construct the density of the mle, is, however, specified on the linear model (1). We therefore use the unrestricted and restricted reduced forms, (32) and (30), to construct the density of the liml estimator. The distribution of the least squares estimator of the unrestricted reduced form results from the assumption of normality of the disturbances and the weak exogeneity of $X$ for $(\beta, \Pi)$

$$\hat{\Phi} \sim N(\Phi, \Omega \otimes (X'X)^{-1}), \tag{36}$$

where $\Phi = \Pi B$ contains the "true values" of the parameters $\beta$ and $\Pi$. The density of the mle in the previous section is constructed using the density of the "$t$-values" of the least squares estimator

$$\hat{\Theta} = (X'X)^{\frac{1}{2}} \hat{\Phi} \Omega^{-\frac{1}{2}}, \tag{37}$$

whose distribution directly results from (36)

$$\hat{\Theta} \sim N(\Theta, I_m \otimes I_k), \tag{38}$$

where $\Theta = (X'X)^{\frac{1}{2}} \Phi \Omega^{-\frac{1}{2}}$, with density

$$p(\hat{\Theta}) \propto \exp\left[ -\frac{1}{2} tr \left\{ \left( \hat{\Theta} - \Theta \right)' \left( \hat{\Theta} - \Theta \right) \right\} \right]. \tag{39}$$

The crucial element for the construction of the density of the mle is the specification of the parameters that are orthogonal to the parameters of the restricted reduced form. An elegant representation for these orthogonal parameters results from specifying $\hat{\Theta}$ as, see (4),

$$\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_\perp \hat{\lambda} \hat{D}_\perp, \tag{40}$$

with $\hat{\Gamma}\hat{D} = (X'X)^{\frac{1}{2}} \hat{\Pi} \hat{B} \Omega^{\frac{1}{2}}$, $\hat{B} = \begin{pmatrix} \hat{\beta} & I_{m-1} \end{pmatrix}$, such that we, for example, can specify $\hat{\Gamma} : k \times (m-1)$, as $\hat{\Gamma} = (X'X)^{\frac{1}{2}} \hat{\Pi} \hat{B} \Omega_2$, and $\hat{D} : (m-1) \times m$, with $\hat{D} = \begin{pmatrix} \hat{\delta} & I_{m-1} \end{pmatrix}$, $\hat{\delta} = (\hat{B}\Omega_2)^{-1} \hat{B} \omega_1$, where $\Omega^{-\frac{1}{2}} = \begin{pmatrix} \omega_1 & \Omega_2 \end{pmatrix}$ with $\omega_1$ a $m \times 1$ vector and $\Omega_2$ a $m \times (m-1)$ matrix. The $(k-m+1) \times 1$ vector $\hat{\lambda}$ is such that when $\hat{\lambda} = 0$, (40) corresponds with the restricted reduced form. $\hat{\Gamma}_\perp$ and $\hat{D}_\perp$ in (40) are specified such that $\hat{\Gamma}_\perp$ is a $k \times (k-m+1)$ matrix and $\hat{\Gamma}'\hat{\Gamma}_\perp \equiv 0$, $\hat{\Gamma}'_\perp \hat{\Gamma}_\perp \equiv I_{k-m+1}$; and $\hat{D}_\perp$ is a $1 \times m$ vector and $\hat{D}\hat{D}'_\perp \equiv 0$, $\hat{D}_\perp \hat{D}'_\perp \equiv 1$. $\hat{\Gamma}_\perp$ and $\hat{D}_\perp$ can thus be constructed from the elements of $\hat{\Gamma}$ and $\hat{D}$ as $\hat{\Gamma}_\perp = \begin{pmatrix} -\hat{\Gamma}_2 \hat{\Gamma}_1^{-1} & I_{k-m+1} \end{pmatrix}' (I_{k-m+1} + \hat{\Gamma}_2 \hat{\Gamma}_1^{-1} \hat{\Gamma}_1^{-1\prime} \hat{\Gamma}'_2)^{-\frac{1}{2}}$, where $\hat{\Gamma} = \begin{pmatrix} \hat{\Gamma}'_1 & \hat{\Gamma}'_2 \end{pmatrix}'$ with $\hat{\Gamma}_1 : (m-1) \times (m-1)$, $\hat{\Gamma}_2 : (k-m+1) \times (m-1)$; and $\hat{D}_\perp = (1 + \hat{\delta}'\hat{\delta})^{-\frac{1}{2}} \begin{pmatrix} 1 & -\hat{\delta}' \end{pmatrix}$[1].

The specification of $\hat{\Gamma}_\perp$ and $\hat{D}_\perp$ satisfies all the conditions from definition 1. The first condition is satisfied as

$$vec(\hat{\Theta}) = vec(\hat{\Gamma}\hat{D}) + (\hat{D}'_\perp \otimes \hat{\Gamma}_\perp) vec(\hat{\lambda}), \tag{41}$$

and

$$(\hat{D}'_\perp \otimes \hat{\Gamma}_\perp)' vec(\hat{\Gamma}\hat{D}) = vec(\hat{\Gamma}'_\perp \hat{\Gamma} \hat{D} \hat{D}'_\perp) \equiv 0. \tag{42}$$

The second condition is satisfied as the derivative of $vec(\hat{\Gamma}\hat{D})$ with respect to $\hat{\delta}$ and $\hat{\Gamma}$

$$\begin{aligned} \frac{\partial vec(\hat{\Gamma}\hat{D})}{(\partial vec(\hat{\delta})' \;\; \partial vec(\hat{\Gamma})')} &= \begin{pmatrix} e_1 \otimes \hat{\Gamma} & \hat{D}' \otimes I_k \end{pmatrix} \\ &= \begin{pmatrix} e_1 \otimes \hat{\Gamma} & \begin{pmatrix} \hat{\delta} & I_{m-1} \end{pmatrix}' \otimes I_k \end{pmatrix} \end{aligned} \tag{43}$$

[1] Let $Q$ be an $n \times n$ symmetric matrix with spectral decomposition $Q = P \Lambda P'$ where $P$ is an $n \times n$ orthogonal matrix of eigenvectors and $\Lambda$ is an $n \times n$ diagonal matrix of eigenvalues. The square root of $Q$ is then defined as $Q^{\frac{1}{2}} = P \Lambda^{\frac{1}{2}} P'$.

is orthogonal to $(\hat{D}'_\perp \otimes \hat{\Gamma}_\perp)$,

$$(\hat{D}'_\perp \otimes \hat{\Gamma}_\perp)'\left(\frac{\partial vec(\hat{\Gamma}\hat{D})}{(\partial vec(\hat{\delta})'\ \partial vec(\hat{\Gamma})')}\right) = (\hat{D}'_\perp \otimes \hat{\Gamma}_\perp)'\left(\begin{array}{cc} e_1 \otimes \hat{\Gamma} & \hat{D}' \otimes I_k \end{array}\right) \equiv 0. \tag{44}$$

The third condition is satisfied by construction since

$$(\hat{D}'_\perp \otimes \hat{\Gamma}_\perp)'(\hat{D}'_\perp \otimes \hat{\Gamma}_\perp) \equiv (1 \otimes I_{k-m+1}). \tag{45}$$

All conditions of definition 1 are thus satisfied.

The transformation from $\hat{\Theta}$ to $(\hat{\delta}, \hat{\Gamma}, \hat{\lambda})$ is a proper transformation of random variables which can be shown using a Singular Value Decomposition (SVD) of $\hat{\Theta}$, see Golub and van Loan (1989),

$$\hat{\Theta} = USV', \tag{46}$$

where $U$ and $V$ are $k \times k$ and $m \times m$ matrices such that $U'U \equiv I_k$ and $V'V \equiv I_m$, and $S$ is a $k \times m$ rectangular matrix which contains the non-negative singular values in decreasing order on its main diagonal $(= (s_{11} \ldots s_{mm}))$ and is equal to zero elsewhere. The representation (40) of $\hat{\Theta}$ can be shown to result from the singular value decomposition (46) when we specify $U$, $S$ and $V$ as,

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix},\ S = \begin{pmatrix} S_1 & 0 \\ 0 & s_2 \end{pmatrix} \text{ and } V = \begin{pmatrix} v_{11} & v_{12} \\ V_{21} & v_{22} \end{pmatrix}, \tag{47}$$

where $U_{11}$, $S_1$, $V_{21}$ are $(m-1) \times (m-1)$ matrices; $v_{12}$ is $1 \times 1$; $v'_{11}$, $v_{22}$ are $(m-1) \times 1$ vectors, $U_{12}$, $U_{21}$, and $U_{22}$ are $(m-1) \times (k-m+1)$, $(k-m+1) \times (m-1)$ and $(k-m+1) \times (k-m+1)$ matrices and $s_2$ is a $(k-m+1) \times 1$ vector. The explicit expressions for $\hat{\delta}$, $\hat{\Gamma}$ and $\hat{\lambda}$ in terms of the elements of $U$, $S$, and $V$ then read, for the construction of these we refer to Kleibergen (2000) and Kleibergen and van Dijk (1998),

$$\hat{\Gamma} = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V'_{21},\ \hat{\delta} = V_{21}'^{-1} v'_{11},\ \hat{\lambda} = (U_{22}U'_{22})^{-\frac{1}{2}} U_{22} s_2 v'_{12} (v_{12} v'_{12})^{-\frac{1}{2}}. \tag{48}$$

The specification of $\hat{\lambda}$ in (48) is such that $\hat{\lambda}$ is an orthogonal transformation of the smallest singular value contained in $s_2$ as $(U_{22}U'_{22})^{-\frac{1}{2}}U_{22}$ and $v'_{12}(v_{12}v'_{12})^{-\frac{1}{2}}$ are orthonormal matrices, *i.e.* $\left((U_{22}U'_{22})^{-\frac{1}{2}}U_{22}\right)\left((U_{22}U'_{22})^{-\frac{1}{2}}U_{22}\right)' = I_{k-m+1}$ and $\left(v'_{12}(v_{12}v'_{12})^{-\frac{1}{2}}\right)'\left(v'_{12}(v_{12}v'_{12})^{-\frac{1}{2}}\right) = 1$. The singular values are generalized eigenvalues of non-symmetric matrices and represent the rank of a matrix. As $\hat{\lambda}$ is an orthogonal transformation of the smallest singular value it thus directly represents the rank of $\hat{\Theta}$, *i.e.* $rank(\hat{\Theta}) = m-1 \Leftrightarrow \hat{\lambda} = 0$.

Both conditions of assumption 2 are thus satisfied as an invertible relationship between $\hat{\Theta}$ and $(\hat{\delta}, \hat{\Gamma}, \hat{\lambda})$ exists and (40) is identical to (4). Also the conditions from definition 1 are satisfied. Theorem 3 thus holds and we can consider the marginal density of $(\hat{\delta}, \hat{\Gamma})$ as proportional to the conditional density of $(\hat{\delta}, \hat{\Gamma})$ given that $\hat{\lambda} = 0$, see (14),

$$\begin{aligned} p(\hat{\delta}, \hat{\Gamma}) &\propto p(\hat{\delta}, \hat{\Gamma}, \hat{\lambda})|_{\hat{\lambda}=0} \\ &\propto p(\hat{\Theta}(\hat{\delta}, \hat{\Gamma}, \hat{\lambda}))|_{\hat{\lambda}=0} \left| J(\hat{\Theta}, (\hat{\delta}, \hat{\Gamma}, \hat{\lambda}))|_{\hat{\lambda}=0} \right|. \end{aligned} \tag{49}$$

The jacobian of the transformation from $\hat{\Theta}$ to $(\hat{\delta}, \hat{\Gamma}, \hat{\lambda})$ evaluated in $\hat{\lambda} = 0$ reads, see Kleibergen (2000), Kleibergen and van Dijk (1998) and Kleibergen and Zivot (1998),

$$\begin{aligned}
|J(\hat{\Theta}, (\hat{\Gamma}, \hat{\delta}, \hat{\lambda}))|_{\hat{\lambda}=0}| &= \left| \begin{pmatrix} \frac{\partial vec(\hat{\Theta})}{\partial vec(\hat{\Gamma})'} & \frac{\partial vec(\hat{\Theta})}{\partial vec(\hat{\delta})'} & \frac{\partial vec(\hat{\Theta})}{\partial vec(\hat{\lambda})'} \end{pmatrix} \Big|_{\hat{\lambda}=0} \right| \\
&= \left| \begin{pmatrix} \hat{D}' \otimes I_k & e_1 \otimes \hat{\Gamma} & \hat{D}'_{\perp} \otimes \hat{\Gamma}_{\perp} \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} \hat{D}\hat{D}' \otimes I_k & \hat{\delta} \otimes \hat{\Gamma} \\ \hat{\delta}' \otimes \hat{\Gamma}' & 1 \otimes \hat{\Gamma}'\hat{\Gamma}' \end{pmatrix} \right|^{\frac{1}{2}} \\
&= |\hat{\Gamma}'\hat{\Gamma}|^{\frac{1}{2}} |I_{m-1} + \hat{\delta}\hat{\delta}'|^{\frac{1}{2}(k-m+1)},
\end{aligned} \tag{50}$$

and combining this with (39), the density (49) becomes

$$\begin{aligned}
p(\hat{\Gamma}, \hat{\delta}) &\propto p(\hat{\Gamma}, \hat{\delta}, \hat{\lambda})|_{\hat{\lambda}=0} \\
&\propto p(\hat{\Theta}(\hat{\Gamma}, \hat{\delta}, \hat{\lambda}))|_{\hat{\lambda}=0} |J(\hat{\Theta}, (\hat{\Gamma}, \hat{\delta}, \hat{\lambda}))|_{\hat{\lambda}=0}| \\
&\propto |\hat{\Gamma}'\hat{\Gamma}|^{\frac{1}{2}} |I_{m-1} + \hat{\delta}\hat{\delta}'|^{\frac{1}{2}(k-m+1)} \exp\left[ -\tfrac{1}{2} tr\left( \left( \hat{\Gamma}\hat{D} - \Theta \right)' \left( \hat{\Gamma}\hat{D} - \Theta \right) \right) \right].
\end{aligned} \tag{51}$$

To obtain the density of the liml estimators, we transform $(\hat{\Gamma}, \hat{\delta})$ to $(\hat{\Pi}, \hat{\beta})$ using the expression given below (40). The (invertible) functional dependence of $(\hat{\Pi}, \hat{\beta})$ on $(\hat{\Gamma}, \hat{\delta})$ depends on the unknown covariance matrix $\Omega$. As discussed in the previous section, we can replace $\Omega$ in these expressions by an estimator of it that is stochastic independent from $\hat{\Theta}$.

The covariance matrix estimator that is based on a least squares regression of $X$ on $Y$,

$$S = \frac{1}{T-k} Y' M_X Y = \frac{1}{T-k} (Y - X\hat{\Phi})'(Y - X\hat{\Phi}), \tag{52}$$

is distributed as, see Muirhead (1982),

$$S \sim W\left( \frac{1}{T-k} \Omega, T-k \right), \tag{53}$$

where $W$ stands for the Wishart distribution, with density

$$p(S) \propto |S|^{\frac{1}{2}(T-k-m-1)} \exp\left[ -\frac{1}{2} tr\left( (T-k) S \Omega^{-1} \right) \right]. \tag{54}$$

The covariance matrix estimator (52) is stochastic independent from the least squares estimator $\hat{\Phi}$ and therefore also stochastic independent from $\hat{\Theta}$. Instead of $S$, we use the covariance matrix estimator

$$\hat{\Omega} = \Omega S^{-1} \Omega, \tag{55}$$

which is distributed as

$$\hat{\Omega} \sim iW((T-k)\Omega, T-k), \tag{56}$$

with density

$$p(\hat{\Omega}) \propto |\hat{\Omega}|^{-\frac{1}{2}(T-k+m+1)} \exp\left[ -\frac{1}{2} tr\left( (T-K) \hat{\Omega}^{-1} \Omega \right) \right]. \tag{57}$$

Since $S$ is stochastic independent from $\hat{\Theta}$, so is $\hat{\Omega}$. The joint density of $(\hat{\delta}, \hat{\Gamma}, \hat{\Omega})$ then reads

$$\begin{aligned} p(\hat{\delta}, \hat{\Gamma}, \hat{\Omega}) & \propto p(\hat{\delta}, \hat{\Gamma}, \hat{\lambda})|_{\hat{\lambda}=0} p(\hat{\Omega}) \\ & \propto |\hat{\Gamma}'\hat{\Gamma}|^{\frac{1}{2}} |I_{m-1} + \hat{\delta}\hat{\delta}'|^{\frac{1}{2}(k-m+1)} |\hat{\Omega}|^{-\frac{1}{2}(T-k+m+1)} \\ & \exp\left[-\tfrac{1}{2} tr \left\{ \left(\hat{\Gamma}\hat{D} - \Theta\right)' \left(\hat{\Gamma}\hat{D} - \Theta\right) + (T-K)\hat{\Omega}^{-1}\Omega \right\}\right]. \end{aligned} \tag{58}$$

We solve for $(\hat{\beta}, \hat{\Pi})$ from $(\hat{\delta}, \hat{\Gamma})$ using the estimator $\hat{\Omega}$ (Note that since $\hat{\Omega}$ (55) depends on the unknown $\Omega$, we cannot construct $\hat{\Omega}$ in practice but we use $\hat{\Omega}$ instead of $S$ (52) as it leads to a more convenient expression of the density of the liml estimator)

$$\hat{\Gamma}\hat{D} = (X'X)^{\frac{1}{2}} \hat{\Pi}\hat{B}\hat{\Omega}^{-\frac{1}{2}} = (X'X)^{\frac{1}{2}} \hat{\Pi}\hat{B}\hat{\Omega}_2 \begin{pmatrix} (\hat{B}\hat{\Omega}_2)^{-1}\hat{B}\hat{\omega}_1 & I_{m-1} \end{pmatrix}, \tag{59}$$

where $\hat{\Omega}^{-\frac{1}{2}} = (\hat{\omega}_1 \ \hat{\Omega}_2)$, $\hat{\omega}_1 : m \times 1$, $\hat{\Omega}_2 : m \times (m-1)$. The joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$ is then obtained by transforming $(\hat{\Gamma}, \hat{\delta})$ to $(\hat{\beta}, \hat{\Pi})$, see appendix B,

$$\begin{aligned} p(\hat{\beta}, \hat{\Pi}, \hat{\Omega}) & \propto p(\hat{\Gamma}(\hat{\beta}, \hat{\Pi}, \hat{\Omega}), \hat{\delta}(\hat{\beta}, \hat{\Pi}, \hat{\Omega}), \hat{\Omega}) |J((\hat{\Gamma}, \hat{\delta}), (\hat{\beta}, \hat{\Pi}))| \\ & \propto |\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)} |\hat{\Pi}'X'X\hat{\Pi}|^{\frac{1}{2}} |\hat{B}\hat{\Omega}^{-1}\hat{B}'|^{\frac{1}{2}(k-m+1)} |X'X|^{\frac{1}{2}(m-1)} \\ & \exp\left[-\tfrac{1}{2} tr \left(\hat{\Omega}^{-1}\left((T-k)\Omega + \left(\hat{\Pi}\hat{B} - \Pi B\right)' X'X \left(\hat{\Pi}\hat{B} - \Pi B\right)\right)\right)\right]. \end{aligned} \tag{60}$$

In case that $m = 2$, we can analytically integrate out $\hat{\Pi}$ from (60) to obtain the joint density of $(\hat{\beta}, \hat{\Omega})$, see appendix B,

$$\begin{aligned} p(\hat{\beta}, \hat{\Omega}) & \propto p(\hat{\beta}|\hat{\Omega}) q(\hat{\Omega}) \\ & \propto |\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)} \exp\left[-\tfrac{1}{2} tr \left(\hat{\Omega}^{-1}\left((T-k)\Omega + B'\Pi'X'X\Pi B\right)\right)\right] \\ & \left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ & \left\{ \sum_{j=0}^{\infty} \left[ \left( \frac{\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^2 \Pi'X'X\Pi}{2\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|} \right)^j \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \right] \right\}, \end{aligned} \tag{61}$$

such that

$$\begin{aligned} p(\hat{\beta}|\hat{\Omega}) \propto & \left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ & \left\{ \sum_{j=0}^{\infty} \left[ \left( \frac{\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^2 \Pi'X'X\Pi}{2\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|} \right)^j \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \right] \right\}, \\ q(\hat{\Omega}) \propto & |\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)} \exp\left[-\tfrac{1}{2} tr \left(\hat{\Omega}^{-1}\left((T-k)\Omega + B'\Pi'X'X\Pi B\right)\right)\right]. \end{aligned} \tag{62}$$

The density $q(\hat{\Omega})$ is the density of the inverted-Wishart random matrix $\Delta$, $\Delta \sim iW((T-k)\Omega + B'\Pi'X'X\Pi B, T-k+m-1)$. This inverted-Wishart random matrix has a mean equal to $\frac{1}{T-k+m-1}((T-k)\Omega + B'\Pi'X'X\Pi B) \approx \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B$ and its variance is proportional to $\frac{1}{T}$. The inverted-Wishart density $q(\hat{\Omega})$, which is not the marginal density of $\hat{\Omega}$, is therefore centered close around its mean for reasonably large values of $T$ ($T > 25$). Hence, for moderate values of $T$, we can consider $q(\hat{\Omega})$ as a point mass at $\Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B$ such that $p(\hat{\beta}) =$

$p(\hat{\beta}|\hat{\Omega} = \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B)$, see appendix B for more details. This is the reason why we use $\hat{\Omega}$ instead of $S$ since we can not decompose the joint density of $(\hat{\beta}, S)$ as the product of the conditional density of $\hat{\beta}$ given $S$ and a standard function of $S$ that has convenient convergence properties. Because $q(\hat{\Omega})$ becomes a point mass at $\hat{\Omega} = \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B$, it shows that the marginal density of $\hat{\Omega}$ has changed compared to (57). This results because of the dependence of $(\hat{\beta}, \hat{\Pi})$ on $\hat{\Omega}$ while $(\hat{\Gamma}, \hat{\delta})$ are stochastic independent from $\hat{\Omega}$. The marginal density of $\hat{\Omega}$ therefore changes when we transform $(\hat{\Gamma}, \hat{\delta})$ to $(\hat{\beta}, \hat{\Pi})$.

## 3.4 Properties of the Density of the LIML estimator

The previous (sub)sections focussed on constructing the density of the liml estimator. In this section we discuss the properties of the resulting expression of the density (61). We discuss the density itself, its relationship with the already existing expressions in the literature, its convergence properties when the sample size increases and how it relates to the sampling density.

### 3.4.1 The Small Sample Density

The conditional density of $\hat{\beta}$ given $\hat{\Omega}$ (62) consists of the product of a Cauchy kernel and a single infinite sum. Since the first element of the infinite sum doesnot depend on $\hat{\beta}$, the tail behavior of the conditional density is identical to the tail behavior of the Cauchy density, *i.e.* no finite integer moments besides the distribution exist. Furthermore, because, already for moderate values of $T$, the marginal density of $\hat{\beta}$ is equal to the conditional density of $\hat{\beta}$ given that $\hat{\Omega} = \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B)$, the tail behavior of the marginal density of $\hat{\beta}$ is identical to the tail behavior of the conditional density of $\hat{\beta}$ given $\hat{\Omega}$. Hence, also marginally no finite integer moments of $\hat{\beta}$ besides the distribution exist. When $\Pi = 0$, the only element that remains of the infinite sum is a constant such that in that case the conditional density of $\hat{\beta}$ given $\hat{\Omega}$ is even equal to a Cauchy density. Another simplification occurs when $\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} = \beta'$ since the conditional density is symmetric in that case.

A convenient and elegant feature of the joint density of $(\hat{\beta}, \hat{\Omega})$ is that it can be decomposed into a product of the conditional density of $\hat{\beta}$ given $\hat{\Omega}$ and a function of $\hat{\Omega}$ that has convenient convergence properties. This function of $\hat{\Omega}$ is identical to the density of an inverted-Wishart random matrix. We can therefore use the properties of the inverted-Wishart distribution in our analysis. The mean of the inverted-Wishart random matrix is equal to $\frac{1}{T-k+m-1}((T-k)\Omega + B'\Pi'X'X\Pi B) \approx \Omega + \frac{1}{T}B'\Pi'X'X\Pi B$ and its variance is proportional to $\frac{1}{T}$, see Muirhead (1982). The mean is also equal to $\frac{1}{T}$ times the expectation of the quadratic form of the endogenous variables, $E(Y'Y)$. This result is not that surprising since $Y'Y$ is used to construct the liml estimator in (35). Because the variance of the inverted-Wishart random matrix is proportional to $\frac{1}{T}$, the function of $\hat{\Omega}$ quickly concentrates around the mean of the inverted-Wishart random matrix when the sample size, $T$, increases. This convergence is quite fast which can be concluded, for example, from the well-known result that a univariate $t$ density with 25 degrees of freedom is almost identical to the normal density. It implies that already for moderate sample sizes ($T > 25$), the joint density of $(\hat{\beta}, \hat{\Omega})$ consists of a ridge that depends on $\hat{\beta}$ at $\hat{\Omega} = \Omega + \frac{1}{T}B'\Pi'X'X\Pi B$ and is equal to zero elsewhere. Hence, we can then use that $p(\hat{\beta}) = p(\hat{\beta}|\hat{\Omega} = \Omega + \frac{1}{T}B'\Pi'X'X\Pi B)$.

The density of $(\hat{\beta}, \hat{\Omega})$ reads as (61) when $m = 2$. For larger values of $m$ there is no straightforward analytical expression of the conditional density. It is possible though to construct such an expression but we will not pursue such kind of an analysis here largely because the resulting expressions are quite complicated in nature. This results as it involves the moments of the determinant of a non-central Wishart distributed random matrix. For details on this, see Muirhead (1982). Since we do have the expression of the joint small sample density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$, however, we can sample from that density using Sampling Algorithms like Importance Sampling, see *e.g.* Kloek and van Dijk (1978) and Geweke (1989), or Metropolis-Hastings Sampling, see *e.g.* Metropolis *et. al.* (1953) and Hastings (1970). In Kleibergen and Paap (1998) and Kleibergen and van Dijk (1998), these algorithms are used in Bayesian analyzes of cointegration and simultaneous equation models where the posteriors are closely related to the joint small sample density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$, see Kleibergen and Zivot (1998). The computed densities can then be compared with the sampling density to show the validity of the approach.

### 3.4.2 Relationship with Existing Analytical Expressions

The density (60)-(62) results from a different approach then the one that is traditionally pursued in the literature, see *e.g.* Mariano and Sawa (1972), Phillips (1983) and Anderson (1982). It also has a different functional form then, for example, the density in Mariano and Sawa (1972), which consists of a triple infinite sum while the density (62) consists of a single infinite sum. One reason for this is that the density constructed by Mariano and Sawa is the marginal density while (62) is the conditional density of $\hat{\beta}$ given $\hat{\Omega}$. Another reason is that these densities are constructed using different approaches. The density (62) results from the use of orthogonal parameters such that the density of the mle is the conditional density of the mle given that the orthogonal conditioning statistics are equal to zero. The traditional approach constructs the density from a closed form expression of the liml estimator. Since we already discussed the construction of the density (62) at length, we now briefly discuss the traditional way of constructing the density of the liml estimator to show the differences and similarities with the approach involving the orthogonal parameters.

The liml estimator results from the characteristic polynomial (35), see Mariano and Sawa (1972),

$$\begin{aligned} |\eta Y'Y - Y'X(X'X)^{-1}X'Y| &= 0 \Leftrightarrow \\ |\eta\,(Y'M_XY + Y'X(X'X)^{-1}X'Y) - Y'X(X'X)^{-1}X'Y| &= 0, \end{aligned} \tag{63}$$

and is defined such that the eigenvector associated with the smallest root of (63) is equal to $a(1\ -\hat{\beta}')'$. When we assume independently normal distributed disturbances with mean zero and identical covariance matrices, $Y'X(X'X)^{-1}X'Y$ has a non-central Wishart distribution, $Y'M_XY$ has a standard Wishart distribution and these random matrices are stochastic independent. Since the liml estimator results from an eigenvector of (63), it satisfies the relationship

$$\eta\left(Y'M_XY + Y'X(X'X)^{-1}X'Y\right)\begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix} = Y'X(X'X)^{-1}X'Y\begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix}, \tag{64}$$

where $\eta$ is the smallest root of (63). The density of the liml estimator $\hat{\beta}$ can therefore be constructed from the joint density of $(Y'X(X'X)^{-1}X'Y, Y'M_XY)$, which is the product of their marginal densities since they are stochastic independent, as it is a function of these (random) matrices, see Mariano and Sawa (1972). The notation of the liml estimator as a

$k$-class estimator, see *e.g.* Hausman (1983) and Phillips (1983),

$$\hat{\beta} = (\hat{y}'y_2)^{-1}\,\hat{y}'y_1, \tag{65}$$

where $\hat{y} = \frac{\eta}{1-\eta}y_1 + \frac{1}{1-\eta}X(X'X)^{-1}X'y_1 = \frac{\eta}{1-\eta}M_X y_1 + X(X'X)^{-1}X'y_1$, $\eta$ is the smallest root of (63), directly shows the functional relationship between $\hat{\beta}$ and $(Y'X(X'X)^{-1}X'Y, Y'M_XY)$. Note that $\eta$ is also a function of $(Y'X(X'X)^{-1}X'Y, Y'M_XY)$. The density of $\hat{\beta}$ can now be constructed by performing a transformation of the random variables and integrating out the remaining random variables besides $\hat{\beta}$, see Mariano and Sawa (1972) for details. In Mariano and Sawa (1972), the resulting expression is given and it consists of a triple infinite sum. Identical to the conditional density of $\hat{\beta}$ given $\hat{\Omega}$, this expression has Cauchy tails such that no finite integer moments besides the distribution exist. Anderson and Sawa (1979) constructed the density of an estimator, to which they refer as the limlk estimator, that is closely related to the liml estimator and results by replacing $Y'Y$ in the characteristic polynomial (63) by the estimated reduced form covariance matrix $Y'M_XY$. They show that the density of this limlk estimator is less complicated than the density of the liml estimator, as it consists of a double infinite sum, and approximates the density of the liml estimator often quite well, see Anderson *et. al.* (1983).

The liml estimator that results from the characteristic polynomial (63) is closely related to the singular value decomposition that we used to construct the density (61). This can be shown by specifying (63) as

$$\begin{aligned}
\left|\eta Y'Y - \hat{\Phi}'X'X\hat{\Phi}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \hat{\Theta}'\hat{\Theta}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix}' \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix}' \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix}' \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix} \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix} \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix} \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix}' \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix}' \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix}' \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix} \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \begin{pmatrix} \hat{D}\hat{D}' & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\Gamma}'\hat{\Gamma} & 0 \\ 0 & \hat{\lambda}^2 \end{pmatrix}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \begin{pmatrix} \hat{D}\hat{D}'\hat{\Gamma}'\hat{\Gamma} & 0 \\ 0 & \hat{\lambda}^2 \end{pmatrix}\right| &= 0
\end{aligned} \tag{66}$$

where $\hat{\Phi} = (X'X)^{-1}X'Y$, $\hat{\Theta} = (X'X)^{\frac{1}{2}}\hat{\Phi}(Y'Y)^{-\frac{1}{2}}$, $\hat{\Theta} = \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix} \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix} \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix}$, see (40). The liml estimator of $\beta$, $\hat{\beta}$, results from the eigenvector associated with the smallest root of (66) which is $\hat{\lambda}^2$. This eigenvector can be specified by $a\hat{D}'_\perp$, where $a$ is a non-zero scalar. Because of the specification of $\hat{D}_\perp$ $(= (1+\hat{\delta}'\hat{\delta})^{-\frac{1}{2}}(1\ -\hat{\delta}'))$, see (40), $\hat{\beta}$ and $\hat{\delta}$ then coincide. When we then conditional on $\hat{\beta}$ estimate $\hat{\Pi}$

$$\hat{\Pi} = (X'X)^{-1}X'Y(Y'Y)^{-1}\hat{B}'(\hat{B}(Y'Y)^{-1}\hat{B}')^{-1}, \tag{67}$$

where $\hat{B} = \begin{pmatrix} \hat{\beta} & I_{m-1} \end{pmatrix}$, the estimated $\hat{\Pi}\hat{B}$ that results is identical to the one that is obtained when we specify $\hat{\Theta}$ according to (40) and discard $\hat{\lambda}$. The liml estimation procedure therefore amounts to imposing rank reduction and discards that part of $\hat{\Theta}$ associated with its smallest singular value. Hence, although $\hat{\Theta}$ itself doesnot have a reduced rank value, the part of $\hat{\Theta}$ where $(\hat{\beta}, \hat{\Pi})$ are essentially solved from has a reduced rank value. The main difference between the two approaches for constructing the density of the liml estimator is thus the stage in which they impose the rank reduction. The approach using the orthogonal parameters imposes rank reduction from the outset and then solves for the liml estimator while the approach using the closed form expression of the liml estimator works exactly the other way around and first constructs the liml estimator and then implicitly imposes the rank reduction. Note also that we solve $(\hat{\Pi}, \hat{\beta})$ from $\hat{\Gamma}\hat{D}$ and use the covariance matrix $\hat{\Omega}$ which has a mean, according to its marginal density, equal to $\frac{1}{T}E(Y'Y)$, and that $Y'Y$ is used in (66). Our approach therefore does lead to the density of the liml estimator and not of the limlk estimator.

Besides being different from a constructional point of view, the densities (62) and the one from Mariano and Sawa (1972) are also different in the sense that (62) is the conditional density of $\hat{\beta}$ given $\hat{\Omega}$ while the density in Mariano and Sawa is the marginal density. Already for moderate values of $T$, however, the marginal density of $\hat{\beta}$ is equal to the conditional density of $\hat{\beta}$ given that $\hat{\Omega} = \Omega + \frac{1}{T}B'\Pi'X'X\Pi B$.

A convenient feature of our approach for constructing the density of the liml estimator is that we straightforwardly obtain the analytical expression of the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$ (60) without the involvement of the non-central Wishart density. This density is only involved in the integration over $\hat{\Pi}$ to obtain the joint density of $(\hat{\beta}, \hat{\Omega})$. The traditional approach for constructing the density of the liml estimator uses the non-central Wishart density from the outset, see previous discussion, and the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$ is therefore more complicated to construct using that approach. Furthermore, given that we have an analytical expression for the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$, we can analyze its properties directly or by sampling from it. In this way, we can compute and analyze the marginal densities also in case of more than two endogenous variables. To sample from the joint density, we can use Sampling Algorithms like Metropolis-Hastings sampling, see *e.g.* Metropolis *et. al.* (1953) and Hastings (1970), and Importance Sampling, see *e.g.* Kloek and van Dijk (1978) and Geweke (1989), which therefore enable us to compute the density of the liml estimator also in case of more than two endogenous variables. Hence, we do not have to rely on complicated analytical integration procedures in order to construct these densities. These simulation algorithms are primarily used in Bayesian statistics but since the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$ (60) is identical in functional form to the posterior of the parameters of an instrumental variable regression model using a Jeffreys' prior, see Kleibergen and Zivot (1998), these simulation techniques can as well be used to compute and analyze the marginal densities of the liml estimator. In Kleibergen and van Dijk (1998) and Kleibergen and Paap (1998), these simulation algorithms are used to simulate from these kind of posteriors to obtain the marginal posteriors of the parameters of instrumental variable regression and cointegration models.

### 3.4.3 Convergence of Distribution LIML estimator to Limiting Distribution

The density of $\hat{\beta}$ given $\hat{\Omega}$,

$$\begin{aligned} p(\hat{\beta}|\hat{\Omega}) \;\propto\; & \left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ & \left[\sum_{j=0}^{\infty} \left( \left( \frac{\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \beta'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{2} \Pi' X' X \Pi}{2\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|} \right)^{j} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \right)\right]. \end{aligned} \tag{68}$$

can be used to analyze the limiting distribution of $\hat{\beta}$ for different values of $\Pi$. Below we discuss three different cases, $\Pi = 0$, $\Pi = \Psi/\sqrt{T}$ where $\Psi$ is a fixed full rank matrix, and $\Pi$ is a fixed non-zero full rank matrix, that cover the main possibilities. Note that most other cases can be considered as combinations of these and that the assumption of normally distributed disturbances is made in order to construct (68). The results are therefore less general then the ones obtained elsewhere in the literature, see *e.g.* Phillips (1989) and Staiger and Stock (1997), but since the convergence properties result straightforwardly from the density of the mle, they show the convergence issues for specific values of $\Pi$ in a rather illustrative way.

$\Pi = 0$ : is known as the case of total non-identification. It implies that the conditional density of $\hat{\beta}$ given $\hat{\Omega}$ (68) is a Cauchy density and remains that regardless of the sample size. The liml estimator $\hat{\beta}$ thus has a Cauchy distribution regardless of the sample size and converges to a random variable with a Cauchy distribution when the sample size $T$ goes to infinity, see also Phillips (1989).

$\Pi = \Psi/\sqrt{T}$: is known as the case of weak identification, see *e.g.* Nelson and Startz (1990), Staiger and Stock (1997), and Zivot, Nelson and Startz (1998), and implies that the value of $\Pi$ decreases with sample size. It functionalizes the, in practice, often observed combination of a large sample size and small but significant "$t$-values" of $\beta$, see, for example, Angrist and Krueger (1991). This results since, similar to the previous case, the limiting distribution is identical to the distribution of the liml estimator and as the distribution of the liml estimator is not a normal distribution, it can easily generate "$t$-values" which seem significant when one mistakenly uses normal critical values but are non-significant when one uses the correct ones. The similarity results because when $\lim_{T\to\infty}\left(\frac{X'X}{T}\right) = Q$ is a fixed full rank matrix,

$$\begin{aligned} \lim_{T\to\infty}\left(\Pi'X'X\Pi\right) &= \lim_{T\to\infty}\left(\left(\tfrac{\Psi}{\sqrt{T}}\right)'X'X\left(\tfrac{\Psi}{\sqrt{T}}\right)\right) \\ &= \lim_{T\to\infty}\left(\Psi'\left(\tfrac{X'X}{T}\right)\Psi\right) = \Psi'Q\Psi, \end{aligned} \tag{69}$$

and $\Pi'X'X\Pi$ remains a finite constant when the sample size goes to infinity. The mapping theorem, see *e.g.* Billingsley (1986), then implies that

$$\begin{aligned} \lim_{T\to\infty} p(\hat{\beta}|\hat{\Omega}) \propto\; & \lim_{T\to\infty}\left[\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m}\right. \\ & \left.\left[\sum_{j=0}^{\infty}\left(\frac{\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|^{2}\Pi'X'X\Pi}{2\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|}\right)^{j}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right]\right] \\ \propto\; & \left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ & \left[\sum_{j=0}^{\infty}\left(\frac{\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|^{2}p\lim[\Pi'X'X\Pi]}{2\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|}\right)^{j}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right] \\ \propto\; & \left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ & \left[\sum_{j=0}^{\infty}\left(\frac{\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|^{2}\Psi'Q\Psi}{2\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|}\right)^{j}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right] \end{aligned} \tag{70}$$

which shows that the distribution and limiting distribution are identical, what also held for the previous case. Hence, $\hat{\beta}$ remains a random variable when sample size increases and does thus not converge to the fixed constant $\beta$. Staiger and Stock (1997) analyzed this case without the normality assumption on the disturbances. Their focus is also esspecially on testing and we therefore discuss the testing implications in a later section.

**$\Pi$ fixed full rank:** implies that $\Pi' X' X \Pi$ converges to infinity when the sample size increases. To illustrate the convergence of the distribution of the liml estimator we now use the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$ (60). Since $\lim_{T\to\infty}\left(\frac{X'X}{T}\right) = Q$ is a fixed full rank matrix, it follows from the mapping theorem, see *e.g.* Billingsley (1986), that,

$$\begin{aligned}\lim_{T\to\infty} p(\hat{\beta}, \hat{\Pi}, \hat{\Omega}) \propto\ & \lim_{T\to\infty}\left[|\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)}\left|\hat{\Pi}' X' X \hat{\Pi}\right|^{\frac{1}{2}}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)}|X'X|^{\frac{1}{2}(m-1)}\right. \\ & \left.\exp\left[-\tfrac{1}{2}tr\left(\hat{\Omega}^{-1}\left((T-k)\Omega + \left(\hat{\Pi}\hat{B} - \Pi B\right)' X'X\left(\hat{\Pi}\hat{B} - \Pi B\right)\right)\right)\right]\right]\end{aligned} \tag{71}$$

We can now divide $X'X$ by $T$ and multiply $\hat{\Pi}\hat{B} - \Pi B$ by $\sqrt{T}$ without affecting the joint density,

$$\begin{aligned}\lim_{T\to\infty} p(\hat{\beta}, \hat{\Pi}, \hat{\Omega}) \propto\ & \lim_{T\to\infty}\left[|\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)}\left|\hat{\Pi}' X' X \hat{\Pi}\right|^{\frac{1}{2}}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)}|X'X|^{\frac{1}{2}(m-1)}\right. \\ & \exp\left[-\tfrac{1}{2}tr\left(\hat{\Omega}^{-1}(T-k)\Omega + \hat{\Omega}^{-1}\left(\sqrt{T}\left(\hat{\Pi}\hat{B} - \Pi B\right)\right)'\left(\tfrac{X'X}{T}\right)\right.\right. \\ & \left.\left.\left.\left(\sqrt{T}\left(\hat{\Pi}\hat{B} - \Pi B\right)\right)\right)\right]\right].\end{aligned} \tag{72}$$

Since $\hat{\Pi} \Rightarrow \Pi$, $\hat{\beta} \Rightarrow \beta$, and $\hat{\Pi}\hat{B} = \begin{pmatrix} \hat{\Pi}\hat{\beta} & \hat{\Pi} \end{pmatrix}$, $\Pi B = \begin{pmatrix} \Pi\beta & \Pi \end{pmatrix}$, the mapping theorem implies,

$$\begin{aligned}\lim_{T\to\infty} p(\hat{\beta}, \hat{\Pi}, \hat{\Omega}) \propto\ & \lim_{T\to\infty}\left[|\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)}|\Pi' X' X \Pi|^{\frac{1}{2}}\left|B\hat{\Omega}^{-1}B'\right|^{\frac{1}{2}(k-m+1)}|X'X|^{\frac{1}{2}(m-1)}\right. \\ & \exp\left[-\tfrac{1}{2}tr\left(\hat{\Omega}^{-1}(T-k)\Omega + \hat{\Omega}^{-1}\left(\sqrt{T}\begin{pmatrix} \Pi(\hat{\beta}-\beta) & \hat{\Pi}-\Pi \end{pmatrix}\right.\right.\right. \\ & \left.\left.\left.\left.\begin{pmatrix} 1 & 0 \\ \beta & I_{m-1} \end{pmatrix}\right)'\left(\tfrac{X'X}{T}\right)\left(\sqrt{T}\begin{pmatrix} \Pi(\hat{\beta}-\beta) & \hat{\Pi}-\Pi \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \beta & I_{m-1} \end{pmatrix}\right)\right)\right]\right].\end{aligned}$$

By performing a transformation of the random variables from $\hat{\beta}$ to $\sqrt{T}(\hat{\beta}-\beta)$, and $\hat{\Pi}$ to $\sqrt{T}(\hat{\Pi}-\Pi)$, with jacobian $|J((\hat{\beta},\hat{\Pi}),(\sqrt{T}(\hat{\beta}-\beta),\sqrt{T}(\hat{\Pi}-\Pi)))| = T^{-\frac{1}{2}((k+1)(m-1))}$, we then obtain that

$$
\begin{aligned}
&\lim_{T\to\infty} p(\sqrt{T}(\hat{\beta}-\beta),\sqrt{T}(\hat{\Pi}-\Pi),\hat{\Omega}) \\
&\propto \lim_{T\to\infty} p((\hat{\beta},\hat{\Pi})(\sqrt{T}(\hat{\beta}-\beta),\sqrt{T}(\hat{\Pi}-\Pi)),\hat{\Omega})|J((\hat{\beta},\hat{\Pi}),(\sqrt{T}(\hat{\beta}-\beta),\sqrt{T}(\hat{\Pi}-\Pi)))| \\
&\propto \lim_{T\to\infty}\left[|\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)}\left|\Pi'\left(\tfrac{X'X}{T}\right)\Pi\right|^{\frac{1}{2}}\left|B\hat{\Omega}^{-1}B'\right|^{\frac{1}{2}(k-m+1)}\left|\tfrac{X'X}{T}\right|^{\frac{1}{2}(m-1)}\right. \\
&\quad \exp\left[-\tfrac{1}{2}tr\left(\hat{\Omega}^{-1}(T-k)\Omega+\hat{\Omega}^{-1}\left(\sqrt{T}\begin{pmatrix}\Pi(\hat{\beta}-\beta) & \hat{\Pi}-\Pi\end{pmatrix}\begin{pmatrix}1 & 0\\ \beta & I_{m-1}\end{pmatrix}\right)'\left(\tfrac{X'X}{T}\right)\right.\right. \\
&\quad \left.\left.\left.\left(\sqrt{T}\begin{pmatrix}\Pi(\hat{\beta}-\beta) & \hat{\Pi}-\Pi\end{pmatrix}\begin{pmatrix}1 & 0\\ \beta & I_{m-1}\end{pmatrix}\right)\right)\right]\right] \\
&\propto \lim_{T\to\infty}\left[|\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)}\left|\Pi'\left(\tfrac{X'X}{T}\right)\Pi\right|^{\frac{1}{2}}|\Sigma|^{\frac{1}{2}(k-m+1)}\left|\tfrac{X'X}{T}\right|^{\frac{1}{2}(m-1)},\right. \\
&\quad \exp\left[-\tfrac{1}{2}tr\left(\hat{\Omega}^{-1}(T-k)\Omega+\Sigma^{-1}\left(\sqrt{T}\begin{pmatrix}\Pi(\hat{\beta}-\beta) & \hat{\Pi}-\Pi\end{pmatrix}\right)'\left(\tfrac{X'X}{T}\right)\right.\right. \\
&\quad \left.\left.\left.\left(\sqrt{T}\begin{pmatrix}\Pi(\hat{\beta}-\beta) & \hat{\Pi}-\Pi\end{pmatrix}\right)\right)\right]\right]
\end{aligned} \tag{73}
$$

which shows that

$$
\begin{aligned}
\sqrt{T}(\hat{\beta}-\beta) &\Rightarrow N(0,\sigma_{11}(\Pi'Q\Pi)^{-1}),\\
\sqrt{T}(\hat{\Pi}-\Pi) &\Rightarrow N(0,\Sigma_{22}\otimes Q^{-1}),\\
\hat{\Omega} &\Rightarrow \Omega.
\end{aligned} \tag{74}
$$

The limiting distributions in (74) are constructed under the assumption of normality of the disturbances. The limiting distributions in (74) accord with the ones discussed in the literature, see *e.g.* Hausman (1983).

The above results show that all convergence issues of the liml estimator can directly be shown using its density. Hence, the density of the mle is a convenient tool using which all convergence issues can be illustrated.

### 3.4.4 Theoretical Density versus Sampling Density

To show the validity of constructing the density of the liml estimator as the conditional density of the liml estimator given that an orthogonal conditioning statistic is equal to zero, we compare the resulting density with the sampling density for specific parameter values. We therefore sampled one million datasets from the model

$$
\begin{aligned}
y_1 &= \beta y_2+\varepsilon_1\\
y_2 &= X\pi+v_2,
\end{aligned} \tag{75}
$$

where $y_1$, $y_2 : T\times 1$, $X : T\times k$, $(\varepsilon_1\ v_2)\sim N(0,\Sigma\otimes I_T)$; $X\sim N(0,I_k\otimes I_T)$, $T=100$, $\pi : k\times 1$, $\pi=(\pi_1...\pi_k)'$, $\pi_2=...=\pi_k=0$, $\beta=1$, $\Sigma=\begin{pmatrix}1 & 0.99\\ 0.99 & 1\end{pmatrix}$; for a few different values of $(k,\pi_1)$ and compared the obtained sampling density with the marginal density of $\hat{\beta}$ that is equal to the conditional density of $\hat{\beta}$ given $\hat{\Omega}$ (62) with $\hat{\Omega}=\Omega+\frac{1}{T}B'\pi'X'X\pi B$, $B=\begin{pmatrix}\beta & 1\end{pmatrix}$, $\Omega=\begin{pmatrix}e_1 & B'\end{pmatrix}'\Sigma\begin{pmatrix}e_1 & B'\end{pmatrix}$. We can use the conditional density as a marginal density as $T$ is sufficiently large ($T=100$). Note that $X$ is fixed over the datasets, and we also use it in the conditional density $p(\hat{\beta}|\hat{\Omega})$, such that we only sample $(\varepsilon_1\ v_2)$ one million times (it is not necessary to perform so many simulations but in this way we obtain an accurate and smooth sampling density).

The model from which we simulate has strong endogeneity as $\rho = 0.99$ and $\Omega_{22}^{-1}\omega_{21} = 1.99$. Furthermore, when we increase $k$, we only add superfluous instruments to the model because the elements of $\pi$ associated with these additional instruments are equal to zero. In this way we can analyze the sensitivity with respect to including too many instruments. We selected these parameter values to have highly non-normal densities. A coinciding theoretical and sampling density at these extreme parameter values is then a strong indication of the correctness of the expression of the theoretical density and thus of the appropriateness of the concept of using the conditional density of the mle given that an orthogonal conditioning statistic is equal to zero as marginal density of the mle.

In figure 1, the theoretical and sampling densities in case of total non-identification, $\pi_1 = 0$, are shown and they are indistinguishable. We only show the exact identified case because increasing the degree of overidentification doesnot affect the small sample or sampling density at all (as was to be expected from (62)). Figure 2 shows the case of weak identification, $\pi_1 = 0.1$, for $k = 1$ (exactly identified) and $k = 5$ (4 degrees of overidentification). The densities are again very similar and it is hard to distinguish them. The same holds for figures 3 and 4 where we show small sample and sampling densities for the properly identified case, $\pi_1 = 1$, with $k = 1$, 5 (figure 3) and $k = 20$ (figure 4). For all cases, the theoretical and sampling densities are hard to distinguish from one another which is, given the extreme values of the parameters of the data generating process, strong evidence in support of constructing the theoretical density as a conditional density. Note also the peculiarity in the densities in case of weak identification which are equal to zero in $\hat{\beta} = 1.99$ $(= \Omega_{22}^{-1}\omega_{21})$ at which point the mode in case of no identification is located.

An interesting phenomenon, that is apparent from all figures, for which the approach using the conditional density gives a straightforward explanation is the relative insensitivity of the density of the liml estimator to adding superfluous instruments. The conditional density approach namely shows that the density of the liml estimator results from imposing rank reduction on the "$t$-values" of the least squares estimator of the encompassing linear model, see (40). The "$t$-values" of the superfluous instruments are non-significant and close to zero. The rank reduction is imposed by restricting the smallest singular value of the "$t$-values" parameter matrix to zero and thus discards the eigenvector associated with this smallest singular value. Since the "$t$-values" of the superfluous instruments are non-significant, they are associated with the smallest singular value and the eigenvector associated with this singular value therefore has non-zero elements at the positions of the superfluous instruments. Hence, when we restrict the smallest singular value to zero and discard its eigenvector, we essentially remove the superfluous instruments. As a consequence, the density of the liml estimator is relatively insensitive to adding superfluous instruments. The densities of other instrumental variable estimators, like for example two stage least squares, are quite sensitive to adding superfluous instruments though, see *e.g.* Phillips (1983) and Kleibergen and Zivot (1998).

The figures of the densities of the liml estimators, figures 1-4, show that the shape of the density changes quite strongly when the instrument quality deteriorates. First, in case of good instruments the densities are unimodal, then they become bimodal when instruments are weak, to become unimodal again when the instruments are invalid. Of course we have specified the data generating processes such that the degree of endogeneity $\rho$ is maximal and the bimodality is therefore more pronounced. The density is, however, essentially always bimodal except for the non-identified case. In the good instrument case, the local modes lie that far apart, though, and differ that strongly in size that only one mode is visible. This can,

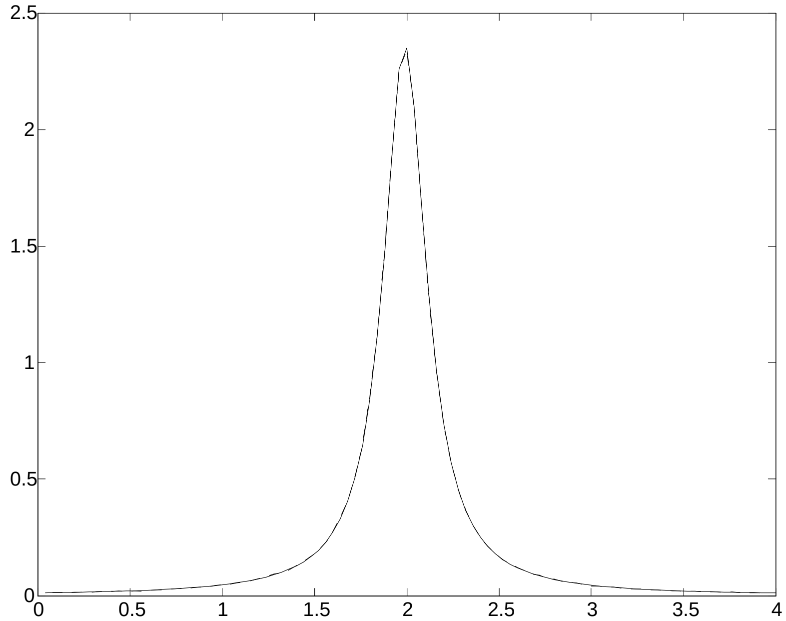


Figure 1: Exact small sample density (-) and sampling density (- -), $\pi_1 = 0$.

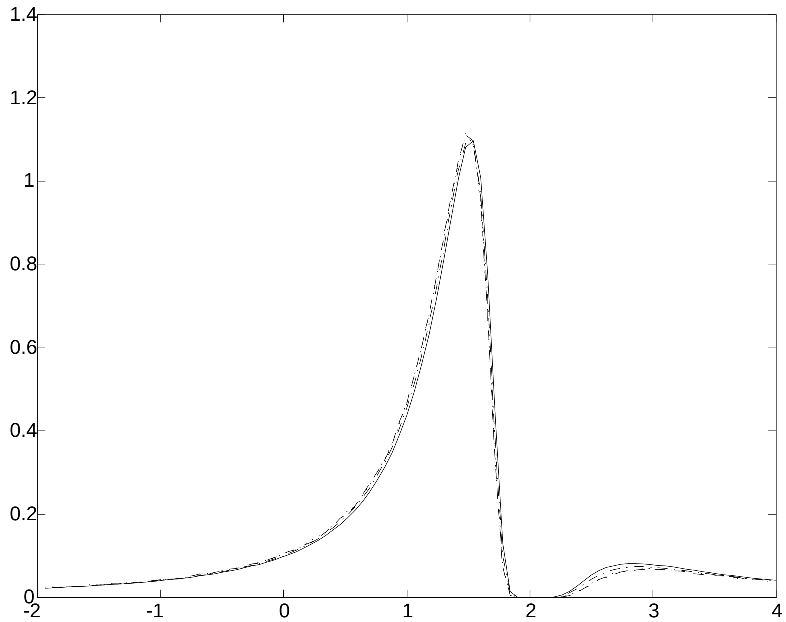


Figure 2: $\pi_1 = 0.1$, $k = 1$: Exact (-) and sampling density (- -); $k = 5$ : exact (-.) and sampling density (..)

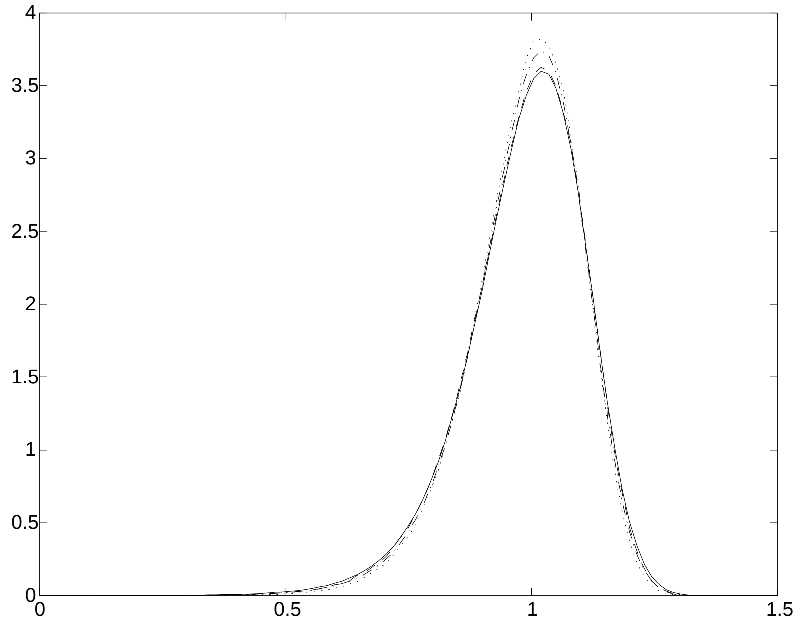


Figure 3: $\pi_1 = 1$, $k = 1$: Exact (-) and sampling density (- -); $k = 5$ : exact (-.) and sampling density (..)

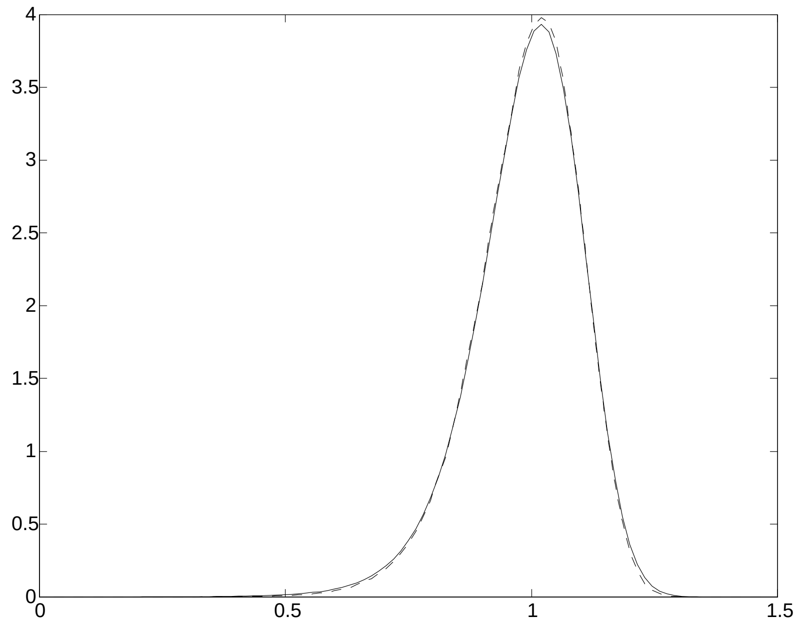


Figure 4: $\pi_1 = 1$, $k = 20$: Exact (-) and sampling density (- -)

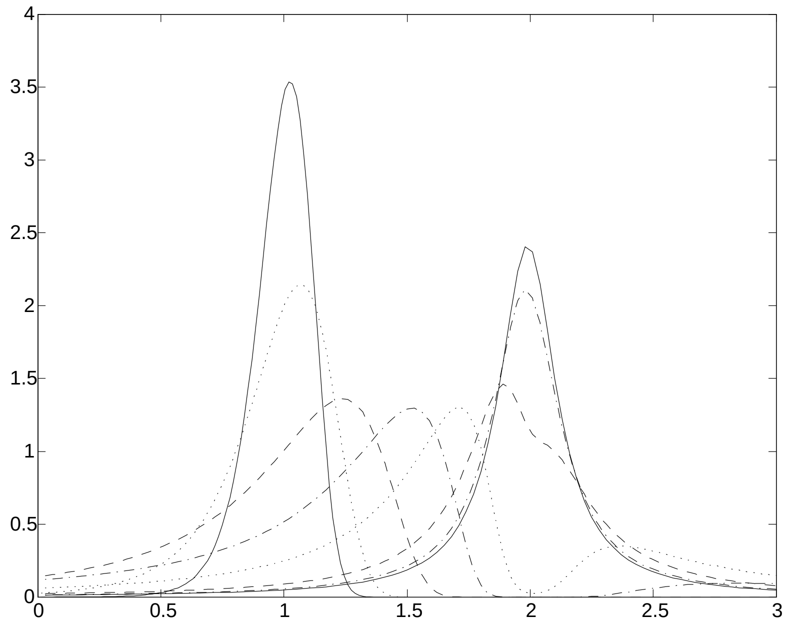


Figure 5: Density $\hat{\beta}$, $\pi_1 = 1$ (-), 0.5 (..), 0.25 (-), 0.1 (-.), 0.05 (..), 0.02 (-), 0.01 (-.), 0 (-).

for example, be concluded from the analytical expression of the conditional density of $\hat{\beta}$ given $\hat{\Omega}$ (62). When we use $\hat{\Omega} = \Omega + \frac{1}{T}B'\Pi'X'X\Pi B$, this density is equal to the marginal density of $\hat{\beta}$. The density has local modes at $\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$, which results from the $t$-kernel in (62), and at the location of the mode of $\frac{B\Omega^{-1}\Pi'X'X\Pi\Omega^{-1}\hat{B}'}{2\hat{B}\hat{\Omega}^{-1}\hat{B}'}$. In the good instrument case, the latter mode strongly dominates the first mode and the modes also lie far apart. As the quality of the instruments deteriorates, however, two phenomena occur. First, the locations of both modes converge to one another and second the difference in importance of the two modes decreases. Thus two things occur in our simulation experiment when the quality of the instruments deteriorates, *i.e.* if $\pi_1$ converges to zero:

$(i)$. $\Omega + \frac{1}{T}B'\Pi'X'X\Pi B$ converges to $\Omega$ and therefore $\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}$ converges to $(\Omega)_{12}\,(\Omega)_{22}^{-1} = 1.99$.

$(ii)$. $\Pi'X'X\Pi$ converges to zero.

The above two phenomena imply that the two local modes of $p(\hat{\beta}|\hat{\Omega})$, one at $\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}$ and one at the mode of the infinite sum, converge to one another and become equally important. This can be shown by computing the density of $\hat{\beta}$ for different values of $\pi_1$. In figure 5, we use the same data generating process as used previously with a fixed degree of over-identification equal to 2. We then visualize the influence of a deterioration of the quality of the instruments by letting $\pi_1$ converge to zero in different steps. This is done by assigning values to $\pi_1$ equal to 1, 0.5, 0.25, 0.1, 0.05, 0.02, 0.01 and 0. The resulting densities nicely show that when the quality of the instruments deteriorates that the location of both local modes converges to 1.99 and that both modes become equally important.

# 4 Testing in Instrumental Variable Regression Models

## 4.1 Test Statistics based on the Density of the LIML estimator

The density of the liml estimator (62) doesnot belong to a standard class of densities nor does the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$ (60). The joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$ can also not be factorized such that the marginal or conditional density given $\hat{\Omega}$ of the "$t$-value" of $\hat{\beta}$

$$\hat{\tau} = (\hat{\Pi}'X'X\hat{\Pi})^{\frac{1}{2}}(\hat{\beta} - \beta)\hat{\Omega}_{11.2}^{-\frac{1}{2}}, \tag{76}$$

where $\hat{\Omega}_{11.2} = \hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{12}$, can be constructed analytically. This results as the dependence of $\hat{\beta}$ on $\hat{\Pi}$ is more complicated than presumed by the relationship underlying the "$t$-value" (76). The factorization is not possible because of the term $|\hat{B}\hat{\Omega}^{-1}\hat{B}'|^{\frac{1}{2}(k-m+1)}$ in the joint density (60) and consequently the transformation of $(\hat{\beta}, \hat{\Pi}, \hat{\Omega})$ to $(\hat{\tau}, \hat{\Pi}, \hat{\Omega})$ leads to a joint density of $(\hat{\tau}, \hat{\Pi}, \hat{\Omega})$ for which we can not construct the marginal and conditional density of $\hat{\tau}$ given $\hat{\Omega}$ analytically. The density of $\hat{\tau}$ can therefore only be assessed numerically and depends on nuisance parameters. The small sample distributions of "$t$" and Wald statistics testing hypothezes on $\beta$ are as a consequence non-standard and standard critical values are only asymptotically valid when $\Pi$ is a fixed full rank matrix.

In case of weak instruments, where $\Pi = \Psi/\sqrt{T}$, see section 3.4.3, the asymptotic distribution of $\hat{\beta}$ is identical to the small sample distribution and in that case the asymptotic distribution of the "$t$-statistic" is then also non-standard, see *e.g.* Staiger and Stock (1997), Wang and Zivot (1998) and Zivot *et. al.* (1998).

So, test statistics that are based on the density of the mle have inconvenient properties. This also holds for cases where $m > 2$, for which the analytical expression of the density of the mle is even unknown.

## 4.2 Testing hypothezes using Orthogonal Parameters

Instead of using the density of the liml estimator to construct test statistics, we can also, like in section 2.7, construct test statistics that are based on the specification of the reduced rank regression model and its orthogonal parameters in (40)-(48). Using this approach we construct four different statistics that can be used to test hypothezes on the parameters of the instrumental variables regression model. The four hypothezes for which we construct (exact!) test statistics are: the validity of all instruments, over-identification, the value of all elements of the structural form parameter, and the value of some elements of the structural form parameter.

### 4.2.1 Validity of all Instruments: Anderson-Rubin Statistic

The first statistic that we construct using orthogonal parameters tests the hypothesis that all the instruments are invalid for the structural relationship. This hypothesis is tested in the unrestricted reduced form (32). It is specified as, $H_{01} : \varphi_1 = 0$, where $\Phi = \begin{pmatrix} \varphi_1 & \Phi_2 \end{pmatrix}$, $\varphi_1 : k \times 1$, $\Phi_2 : k \times (m-1)$, and is tested against the alternative hypothesis $H_1 : \varphi_1 \neq 0$. To reflect the parameters of hypothesis $H_{01}$ in terms of the orthogonal parameters of definition 1, we specify $\hat{\Theta}$ $(= (X'X)^{\frac{1}{2}}\hat{\Phi}\Omega^{-\frac{1}{2}})$ as,

$$\hat{\Theta} = \hat{\psi}\begin{pmatrix} 0 & I_{m-1} \end{pmatrix} + \hat{\lambda}_1 e_1', \tag{77}$$

where $e_1$ is the first $m$ dimensional unity vector and $\hat{\psi} : k \times (m-1)$, $\hat{\lambda}_1 : k \times 1$. As $e_1'$ is orthogonal to $\begin{pmatrix} 0 & I_{m-1} \end{pmatrix}$, *i.e.* $\begin{pmatrix} 0 & I_{m-1} \end{pmatrix} e_1 = 0$, $\hat{\psi}$ and $\hat{\lambda}_1$ satisfy the conditions from definition 1. To show that we obtain the specification under $H_{01}$ when $\hat{\lambda}_1 = 0$, we specify $\Omega^{-1}$ as,

$$\begin{array}{rl} \Omega^{-1} & = \begin{pmatrix} \omega_{11}^{-1} + \omega_{11}^{-1}\omega_{12}\Omega_{22.1}^{-1}\omega_{21}\omega_{11}^{-1} & -\omega_{11}^{-1}\omega_{12}\Omega_{22.1}^{-1} \\ -\Omega_{22.1}^{-1}\omega_{21}\omega_{11}^{-1} & \Omega_{22.1}^{-1} \end{pmatrix} \\ & = \begin{pmatrix} \omega_{11}^{-\frac{1}{2}} & -\omega_{11}^{-1}\omega_{12}\Omega_{22.1}^{-\frac{1}{2}} \\ 0 & \Omega_{22.1}^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \omega_{11}^{-\frac{1}{2}} & -\omega_{11}^{-1}\omega_{12}\Omega_{22.1}^{-\frac{1}{2}} \\ 0 & \Omega_{22.1}^{-\frac{1}{2}} \end{pmatrix}' \\ & = \Omega^{-\frac{1}{2}}\Omega^{-\frac{1}{2}\prime}, \end{array} \tag{78}$$

where $\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix}$, $\omega_{11} : 1 \times 1$, $\omega_{12}, \omega_{21}' : 1 \times (m-1)$, $\Omega_{22} : (m-1) \times (m-1)$, $\Omega_{22.1} = \Omega_{22} - \omega_{21}\omega_{11}^{-1}\omega_{12}$, such that

$$\Omega^{-\frac{1}{2}} = \begin{pmatrix} \omega_{11}^{-\frac{1}{2}} & -\omega_{11}^{-1}\omega_{12}\Omega_{22.1}^{-\frac{1}{2}} \\ 0 & \Omega_{22.1}^{-\frac{1}{2}} \end{pmatrix}. \tag{79}$$

Using the specification of $\Omega^{-\frac{1}{2}}$, we obtain that,

$$\begin{array}{c} \hat{\lambda}_1 = (X'X)^{\frac{1}{2}}\hat{\varphi}_1\omega_{11}^{-\frac{1}{2}}, \ \hat{\psi} = (X'X)^{\frac{1}{2}}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12}\right)\Omega_{22.1}^{-\frac{1}{2}}, \\ \text{and } \Theta = (X'X)^{\frac{1}{2}}\Phi\Omega^{-\frac{1}{2}} = \begin{pmatrix} 0 & \Theta_2 \end{pmatrix}, \end{array} \tag{80}$$

where $\Theta_2 = (X'X)^{\frac{1}{2}}\Phi_2\Omega_{22.1}^{-\frac{1}{2}}$, and which shows that $\hat{\lambda} = 0$ corresponds with the model under $H_0$. Equation (80) also shows that an invertible relationship between $(\hat{\psi}, \hat{\lambda})$ and $(\hat{\varphi}_1, \hat{\Phi}_2)$ exists. $\hat{\Theta}$ has a standard normal distribution, see (7), such that under $H_{01}$

$$\hat{\lambda}_1 \sim N(0, I_k), \ \hat{\psi} \sim N(\Theta_2, I_{m-1} \otimes I_k), \tag{81}$$

and $\hat{\lambda}_1$ and $\hat{\psi}$ are stochastic independent. For a known value of $\Omega$, $\hat{\lambda}_1'\hat{\lambda}_1$ equals the likelihood based test statistics, Wald, Likelihood ratio and Score, and is under $H_{01}$ distributed as

$$\begin{array}{rl} \hat{\lambda}_1'\hat{\lambda}_1 & = \omega_{11}^{-\frac{1}{2}}\hat{\varphi}_1'(X'X)^{-\frac{1}{2}\prime}(X'X)^{-\frac{1}{2}}\hat{\varphi}_1\omega_{11}^{-\frac{1}{2}} \\ & = \frac{y_1'X(X'X)^{-1}X'y_1}{\omega_{11}} \sim \chi^2(k). \end{array} \tag{82}$$

Similar to section 2.7, when the value of $\Omega$ is unknown, we use an estimator of $\Omega$ that is stochastic independent from $\hat{\Theta}$, *i.e.* $S$ (52), and then we obtain the exact test statistic,

$$\begin{array}{rl} F(H_{01}|H_1) & = \frac{1}{k}\frac{y_1'X(X'X)^{-1}X'y_1}{s_{11}} = \frac{1}{k}\frac{\hat{\lambda}_1'\hat{\lambda}_1}{s_{11}/\omega_{11}} = \dfrac{\frac{y_1'X(X'X)^{-1}X'y_1}{\omega_{11}}\Big/k}{\frac{(T-k)s_{11}}{\omega_{11}}\Big/(T-k)} \\ & \sim \frac{\chi^2(k)/k}{\chi^2(T-k)/(T-k)} \sim F(k, T-k), \end{array} \tag{83}$$

where $s_{11}$ results from $S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & S_{22} \end{pmatrix}$ (52)-(54), with $s_{11} : 1 \times 1$, $s_{12}, s_{21}' : 1 \times (m-1)$, $S_{22} : (m-1) \times (m-1)$. Equation (83) is the Anderson-Rubin statistic, see Anderson and Rubin (1949). The Anderson-Rubin statistic is known to have an exact distribution and the above shows how we can derive this result using the orthogonal parameters. It therefore also shows how we in general can construct $F$ statistics in linear models and that linear models belong to the class of models that allow for orthogonal parameters. The exact distribution of the Anderson-Rubin statistic is confirmed by the simulations, that we conducted, which show that the empirical distribution of the Anderson-Rubin statistic is identical to the theoretical distribution.

### 4.2.2 Over-identification

The hypothesis of over-identification equates the unrestricted reduced form model (32) to the restricted reduced form (30). It can therefore be reflected in terms of the parameters as $H_{02}: \Phi = \Pi B$, and is tested against the alternative hypothesis $H_1 : \Phi \neq \Pi B$. To reflect the hypothesis $H_{02}$ in terms of orthogonal parameters, we use (40)

$$\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_{\perp}\hat{\lambda}_2\hat{D}_{\perp}, \tag{84}$$

where $\hat{\lambda}_2 : (k - m + 1) \times 1$, and the other elements are defined in (40). Over-identification corresponds with $H_{02} : \lambda_2 = 0$ and $\hat{\lambda}_2$ is orthogonal to $(\hat{\delta}, \hat{\Gamma})$, which is shown in section 3.3. Under $H_{02}$, $\hat{\lambda}_2$ is then standard normal distributed, see (13),

$$\hat{\lambda}_2 \sim N(0, I_{k-m+1}). \tag{85}$$

We use the orthonormality of $\hat{\Gamma}_{\perp}$, $\hat{D}_{\perp}$, and the orthogonality of these matrices to $\hat{\Gamma}$ and $\hat{D}$, respectively, to construct $\hat{\lambda}_2$ from $\hat{\Theta}$,

$$\begin{aligned} \hat{\lambda}_2 &= \hat{\Gamma}_{\perp}'\hat{\Theta}\hat{D}_{\perp}' = \hat{\Gamma}_{\perp}'(X'X)^{\frac{1}{2}}\hat{\Phi}\Omega^{-\frac{1}{2}}\hat{D}_{\perp}' \\ &= \hat{\Gamma}_{\perp}'(X'X)^{-\frac{1}{2}}X'Y\Omega^{-\frac{1}{2}}\hat{D}_{\perp}'. \end{aligned} \tag{86}$$

In the previous section, we constructed the density of the liml estimator from the conditional density of $(\hat{\delta}, \hat{\Gamma})$ given that $\hat{\lambda}_2 = 0$. Using $\hat{\Theta}$, we can obtain the specification of $\hat{\Gamma}_{\perp}$ and $\hat{D}_{\perp}$ in terms of the liml estimators contained in $\hat{\Pi}$ and $\hat{B}$ as

$$\begin{aligned} \hat{\Gamma}_{\perp} &= (X'X)^{-\frac{1}{2}}\hat{\Pi}_{\perp}(\hat{\Pi}_{\perp}'(X'X)^{-1}\hat{\Pi}_{\perp})^{-\frac{1}{2}} \\ \hat{D}_{\perp} &= (\hat{B}_{\perp}\Omega\hat{B}_{\perp}')^{-\frac{1}{2}}\hat{B}_{\perp}\Omega^{\frac{1}{2}} \end{aligned} \tag{87}$$

such that the expression for $\hat{\lambda}_2$ reads

$$\hat{\lambda}_2 = (\hat{\Pi}_{\perp}'(X'X)^{-1}\hat{\Pi}_{\perp})^{-\frac{1}{2}}\hat{\Pi}_{\perp}'(X'X)^{-1}X'Y\hat{B}_{\perp}'(\hat{B}_{\perp}\Omega\hat{B}_{\perp}')^{-\frac{1}{2}}. \tag{88}$$

The likelihood based test statistics for testing $H_{02}$ against $H_1$ then become

$$\begin{aligned} \hat{\lambda}_2'\hat{\lambda}_2 &= (\hat{B}_{\perp}\Omega\hat{B}_{\perp}')^{-1}\hat{B}_{\perp}Y'X(X'X)^{-1}\hat{\Pi}_{\perp}(\hat{\Pi}_{\perp}'(X'X)^{-1}\hat{\Pi}_{\perp})^{-1}\hat{\Pi}_{\perp}'(X'X)^{-1}X'Y\hat{B}_{\perp}' \\ &= (\hat{B}_{\perp}\Omega\hat{B}_{\perp}')^{-1}\hat{B}_{\perp}Y'(M_{X\hat{\Pi}} - M_X)Y\hat{B}_{\perp}' \sim \chi^2(k - m + 1). \end{aligned} \tag{89}$$

For a known value of $\Omega$, (89) is equal to the Wald, Likelihood ratio and Score statistic for testing $H_{02}$ against $H_1$. In practice $\Omega$ is typically unknown and we then need to use an estimator of it that is stochastic independent from the score vector. A convenient estimator for this purpose is $S$ (52). $\hat{B}_{\perp}$ is then also stochastic independent from $S$ such that

$$(T - K)\frac{\hat{B}_{\perp}S\hat{B}_{\perp}'}{\hat{B}_{\perp}\Omega\hat{B}_{\perp}'} \sim \chi^2(T - k) \tag{90}$$

which results as $S$ has a Wishart distribution with scale matrix $\Omega$ and $T-k$ degrees of freedom, see (53), and $\hat{B}_\perp$ is stochastic independent from $S$, see Muirhead (1982, theorem 3.2.8). We can then construct the test statistic

$$\begin{aligned}
F(H_{02}|H_1) &= \tfrac{1}{k-m+1} tr\left(\left(S^{-1} - S^{-1}\hat{B}'(\hat{B}S^{-1}\hat{B}')^{-1}\hat{B}S^{-1}\right)Y'(M_{X\hat{\Pi}} - M_X)Y\right) \\
&= \tfrac{1}{k-m+1}\tfrac{\hat{B}_\perp Y'X\hat{\Pi}_\perp(\hat{\Pi}'_\perp(X'X)^{-1}\hat{\Pi}_\perp)^{-1}\hat{\Pi}'_\perp X'Y\hat{B}'_\perp}{\hat{B}_\perp S\hat{B}'_\perp} \\
&= \tfrac{1}{k-m+1}\tfrac{\hat{\lambda}'_2\hat{\lambda}_2}{\hat{B}_\perp S\hat{B}'_\perp/\hat{B}_\perp\Omega\hat{B}'_\perp} \\
&= \frac{\frac{\hat{B}_\perp Y'X\hat{\Pi}_\perp(\hat{\Pi}'_\perp(X'X)^{-1}\hat{\Pi}_\perp)^{-1}\hat{\Pi}'_\perp X'Y\hat{B}'_\perp}{\hat{B}_\perp\Omega\hat{B}'_\perp}\Big/(k-m+1)}{(T-k)\frac{\hat{B}_\perp S\hat{B}'_\perp}{\hat{B}_\perp\Omega\hat{B}'_\perp}\Big/(T-k)} \\
&\sim \tfrac{\chi^2(k-m+1)/(k-m+1)}{\chi^2(T-k)/(T-k)} \\
&\sim F(k-m+1, T-k).
\end{aligned} \tag{91}$$

We simulated the test statistic (91) using the data generating process from section 3.4.4 and we compared the empirical distribution function with the theoretical distribution function for various values of $\pi_1$ and $k$. Figures 6 and 7 contain the empirical and theoretical distribution functions of the over-identification statistic (91) for values of $\pi_1$ and $k$ equal to 0.1 and 1, for $\pi_1$, and 5 and 20, for $k$. All other parameters in the data generating process are identical to the ones used in section 3.4.4. The empirical and theoretical distribution functions are in all cases identical which is quite surprising given the large differences of the densities of the mles for the different values of $\pi_1$ which are shown in figures 2, 3 and 4. This results as the over-identification statistic (91) is an exact test statistic such that its distribution doesnot depend on unobserved nuisance parameters.

Figure 8 shows the empirical and (asymptotic) theoretical distributions of the over-identification statistic (91) and the likelihood ratio statistic divided by $k-2$ for the data generating process used for figure 6 with $k = 20$ and $\pi_1 = 0.1$. The empirical distribution of the likelihood ratio statistic differs from its asymptotic theoretical distribution which indicates that the likelihood ratio statistic is not an exact test statistic and that its distribution in small samples depends on nuisance parameters. The over-identification statistic (91) is an exact statistic, however, and, as the figures show, its distribution doesnot depend on nuisance parameters.

### 4.2.3 Value of all elements of the structural form parameter

**Likelihood Ratio Based Statistic** Under the hypothesis $H_{03} : \beta = 0$, the reduced rank regression model is identical to the model under the hypothesis $H_{01}$. Testing $H_{03}$ against $H_{13} : \beta \neq 0$ is thus identical to testing $H_{01}$ against $H_{02}$. When $\Omega$ is known, the likelihood ratio statistic for testing $H_{01}$ against $H_{02}$, and therefore also the likelihood ratio statistic for testing $H_{03}$ against $H_{13}$, is equal to, see (82) and (89),

$$\begin{aligned}
LR(H_{03}|H_{13}) &= LR(H_{01}|H_{02}) = LR(H_{01}|H_1) - LR(H_{02}|H_1) \\
&= \hat{\lambda}'_1\hat{\lambda}_1 - \hat{\lambda}'_2\hat{\lambda}_2 \\
&= \tfrac{y'_1X(X'X)^{-1}X'y_1}{\omega_{11}} - \tfrac{\hat{B}_\perp Y'(M_{X\hat{\Pi}} - M_X)Y\hat{B}'_\perp}{\hat{B}_\perp\Omega\hat{B}'_\perp}.
\end{aligned} \tag{92}$$

The statistic (92) is under $H_{03}$ distributed as a $\chi^2(m-1)$ random variable. Because of the orthogonality, (92) is also equal to the Score and Wald statistic for testing $H_{03}$ against $H_{13}$. When we instead of $\Omega$ use the estimator $S$, which is stochastic independent from the other

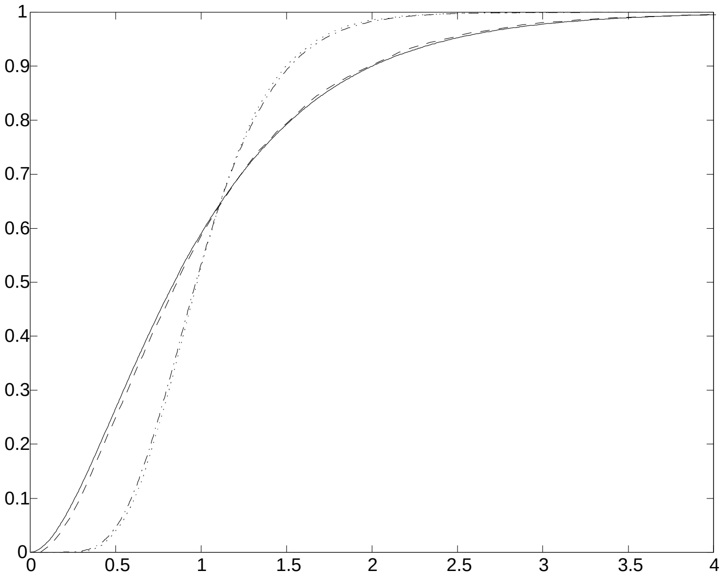


Figure 6: Distribution over-identification statistic, $k = 5$ : empirical (- -), theoretical (-), $k = 20$ : empirical (..), theoretical (-.), $\pi_1 = 0.1$.

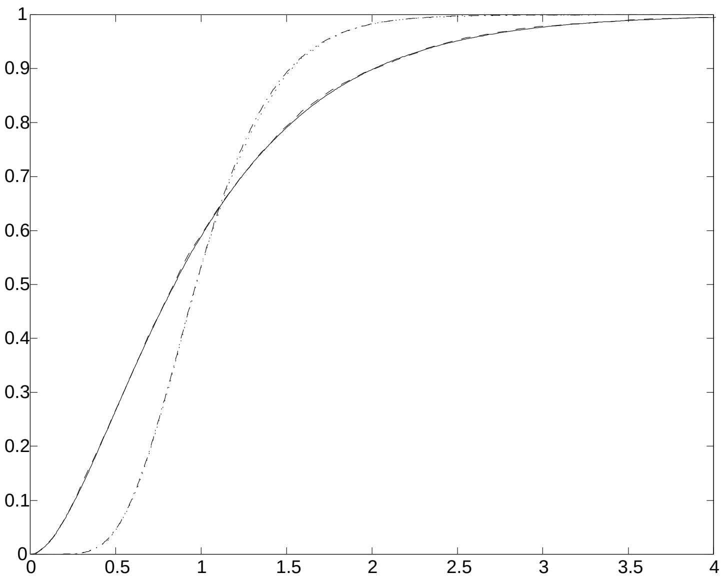


Figure 7: Distribution over-identification statistic, $k = 5$ : empirical (- -), theoretical (-), $k = 20$ : empirical (..), theoretical (-.), $\pi_1 = 1$.

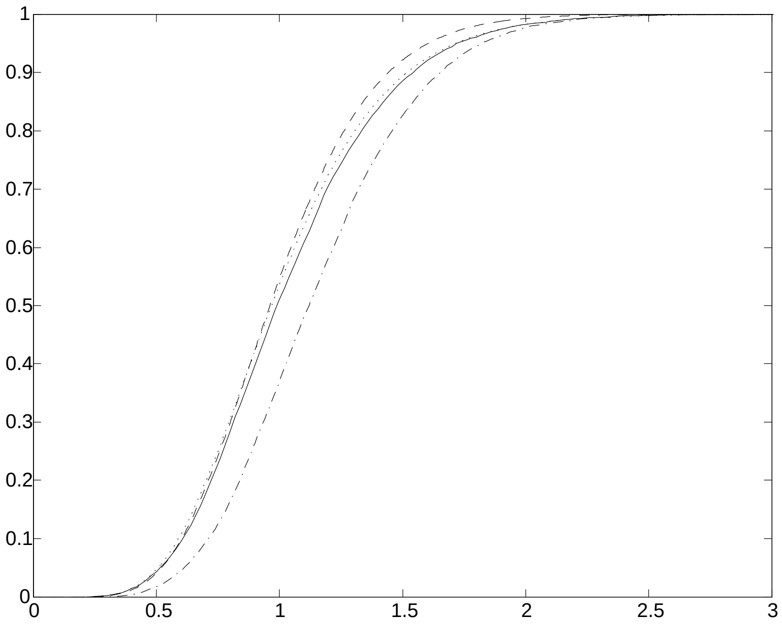


Figure 8: Distribution over-identification statistic, $k = 20$, $\pi_1 = 0.1$, empirical (-), theoretical (..), LR$/(k-2)$ : empirical (-.-), asymptotic theoretical (- -).

elements of (92), and divide (92) by $m-1$, we obtain,

$$\begin{aligned} F(H_{03}|H_{13}) &= \frac{1}{m-1}\left[\frac{y_1'X(X'X)^{-1}X'y_1}{s_{11}} - tr\left(\left(S^{-1} - S^{-1}\hat{B}'(\hat{B}S^{-1}\hat{B}')^{-1}\hat{B}S^{-1}\right)Y'(M_{X\hat{\Pi}} - M_X)Y\right)\right] \\ &= \frac{1}{m-1}\left[\frac{\hat{\lambda}_1'\hat{\lambda}_1}{s_{11}/\omega_{11}} - \frac{\hat{\lambda}_2'\hat{\lambda}_2}{\hat{B}_\perp S\hat{B}_\perp'/\hat{B}_\perp\Omega\hat{B}_\perp'}\right] \\ &= \frac{1}{m-1}\left[\frac{\frac{y_1'X(X'X)^{-1}X'y_1}{\omega_{11}}}{(T-k)\frac{s_{11}}{\omega_{11}}\Big/(T-k)} - \frac{\frac{\hat{B}_\perp Y'X\hat{\Pi}_\perp(\hat{\Pi}_\perp'(X'X)^{-1}\hat{\Pi}_\perp)^{-1}\hat{\Pi}_\perp'X'Y\hat{B}_\perp'}{\hat{B}_\perp\Omega\hat{B}_\perp'}}{(T-k)\frac{\hat{B}_\perp S\hat{B}_\perp'}{\hat{B}_\perp\Omega\hat{B}_\perp'}\Big/(T-k)}\right]. \end{aligned} \tag{93}$$

Under $H_{03}$, the elements in the denominator of (93) are all $\chi^2(T-k)$ distributed random variables that are divided by their degrees of freedom parameter, $T-k$. For reasonable large $T$, the elements in the denominator are therefore approximately equal to one. We are then left with the elements in the numerator of (93) which are, according to (92), $\chi^2(m-1)$ distributed. So, for reasonable large $T$, we obtain that under $H_{03}$,

$$\begin{aligned} F(H_{03}|H_{13}) &= \frac{1}{m-1}\left[\frac{y_1'X(X'X)^{-1}X'y_1}{s_{11}} - tr\left(\left(S^{-1} - S^{-1}\hat{B}'(\hat{B}S^{-1}\hat{B}')^{-1}\hat{B}S^{-1}\right)Y'(M_{X\hat{\Pi}} - M_X)Y\right)\right] \\ &= \frac{1}{m-1}\left[\frac{\frac{y_1'X(X'X)^{-1}X'y_1}{\omega_{11}}}{(T-k)\frac{s_{11}}{\omega_{11}}\Big/(T-k)} - \frac{\frac{\hat{B}_\perp Y'X\hat{\Pi}_\perp(\hat{\Pi}_\perp'(X'X)^{-1}\hat{\Pi}_\perp)^{-1}\hat{\Pi}_\perp'X'Y\hat{B}_\perp'}{\hat{B}_\perp\Omega\hat{B}_\perp'}}{(T-k)\frac{\hat{B}_\perp S\hat{B}_\perp'}{\hat{B}_\perp\Omega\hat{B}_\perp'}\Big/(T-k)}\right] \\ &\sim F(m-1, T-k). \end{aligned} \tag{94}$$

For the data generating process from section 3.4.4 with a value of $\beta$ equal to zero, so under $H_{03}$, we compared the empirical and (asymptotic) theoretical distribution of the test statistic (94) for different values of $\pi_{01}$ and $k$. In figure 9, for a value of $\pi_1$ equal to 0.1, the (asymptotic) theoretical and empirical distribution function for values of $k$ equal to 5 and 20 are shown. For $k = 5$, the empirical and (asymptotic) theoretical distribution function of the

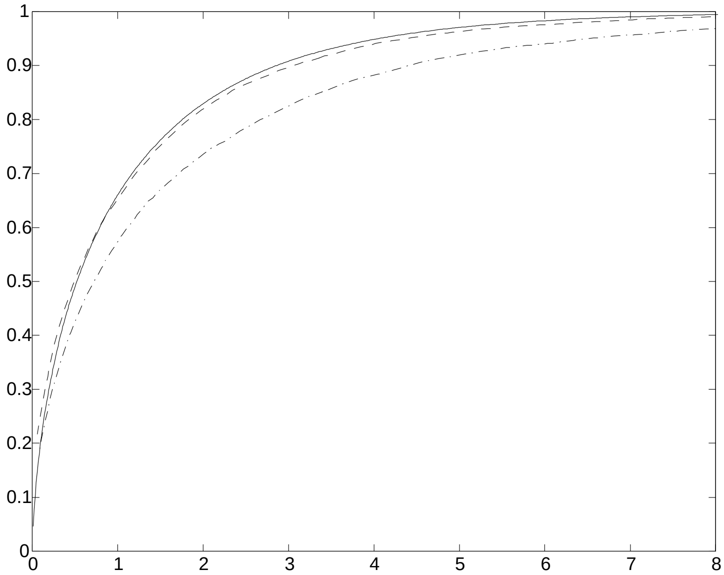


Figure 9: Distribution test statistic $\beta = 0$, (asymptotic) theoretical (-), empirical: $k = 5$ (- -), $k = 20$ (-.), $\pi_1 = 0.1$.

test statistic (94) are identical. For $k = 20$, which implies a large degree of (nonsense) over-identification, there is a small difference between the empirical and (asymptotic) theoretical distribution function of (94) which disappears when the number of observations increases. This is shown in figure 10 where the empirical and (asymptotic) theoretical distribution function are shown in case of 500 observations. $\pi_1$ equal to 0.1 implies a very weak instrument and a highly non-standard density of the mle of $\beta$, see figure 2. Figure 11 contains the empirical and (asymptotic) theoretical distribution function of (94) in case that $\pi_1 = 1$, for values of $k$ equal to 5 and 20. For both values of $k$, the empirical and (asymptotic) theoretical distribution function coincide.

The statistic (94) can directly be used to conduct tests on $\beta$ when it is a vector instead of a scalar, so when $m$ exceeds 2. In that case the analytical expression of the density of the mle is even unknown. Simulation experiments show that the quality of the approximation of the empirical distribution by the asymptotic theoretical distribution is less accurate in this case compared to the univariate setting where $m$ is equal to 2.

**Exact Test Statistic** The statistic (94) is not an exact test statistic such that its distribution depends on nuisance parameters. We can also construct an exact test statistic for conducting tests on the value of the structural form parameter. To construct this statistic we use the specification of $\hat{\Theta}$ $(= (X'X)^{\frac{1}{2}}\hat{\Phi}\Omega^{-\frac{1}{2}})$ (77),

$$\hat{\Theta} = \hat{\psi} \begin{pmatrix} 0 & I_{m-1} \end{pmatrix} + \hat{\lambda}_1 e_1', \tag{95}$$

that we also used to construct the Anderson-Rubin statistic (83). $H_{03} : \beta = 0$ corresponds with $H_{01} : \varphi_1 = 0$ such that, see (80),

$$\Theta = (X'X)^{\frac{1}{2}}\Phi\Omega^{-\frac{1}{2}} = \begin{pmatrix} 0 & \Theta_2 \end{pmatrix}, \tag{96}$$

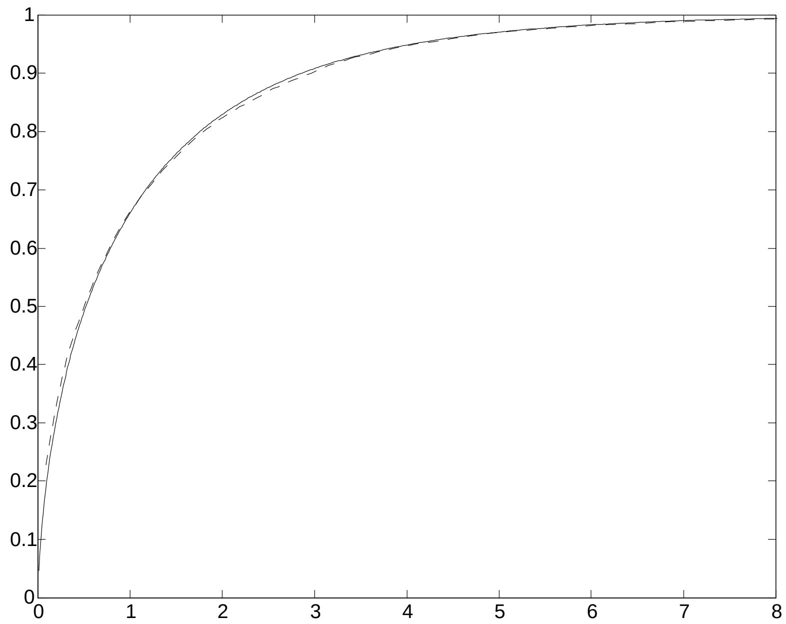


Figure 10: Distribution test statistic $\beta = 0$, (asymptotic) theoretical (-), empirical $k = 20$ (- -), $\pi_1 = 0.1$, $T = 500$.

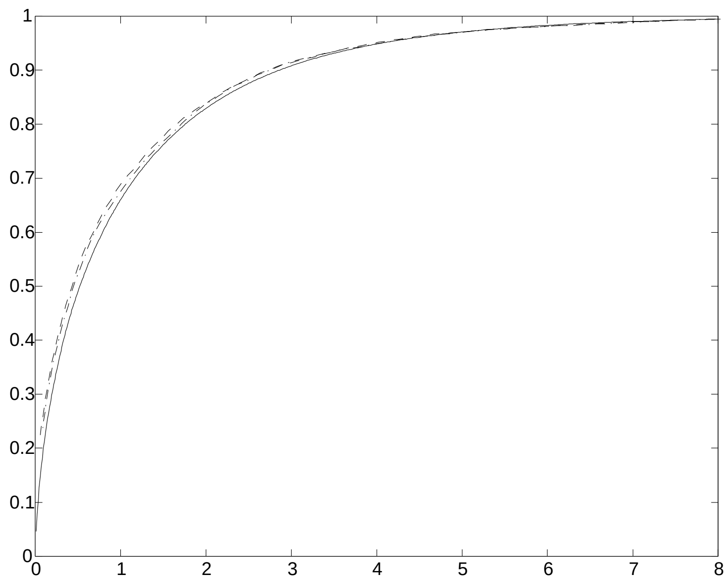


Figure 11: Distribution test statistic $\beta = 0$, (asymptotic) theoretical (-), empirical: $k = 5$ (- -), $k = 20$ (-.), $\pi_1 = 1$.

where $\Theta_2 = (X'X)^{\frac{1}{2}}\Phi_2\Omega_{22.1}^{-\frac{1}{2}}$. For the test of the hypothesis of over-identification $H_{02} : \Phi = \Pi B$, which is identical to $H_{13} : \beta \neq 0$, $\hat{\Theta}$ is specified as (84)

$$\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_\perp \hat{\lambda}_2 \hat{D}_\perp. \tag{97}$$

As $\hat{D} = \begin{pmatrix} \hat{\delta} & I_{m-1} \end{pmatrix}$, $\hat{\psi}$ corresponds with $\hat{\Gamma}$ and $\hat{\lambda}_1 e_1'$ corresponds with $\hat{\Gamma}\begin{pmatrix} \hat{\delta} & 0 \end{pmatrix} + \hat{\Gamma}_\perp \hat{\lambda}_2 \hat{D}_\perp$. $\hat{\lambda}_2$ is normally distributed, see (85), and reflects over-identification. We can then remove the part of $\hat{\lambda}_1 e_1'$ ($\leftrightarrow \hat{\Gamma}\begin{pmatrix} \hat{\delta} & 0 \end{pmatrix} + \hat{\Gamma}_\perp \hat{\lambda}_2 \hat{D}_\perp$) that is due to the over-identification by pre-multiplying $\hat{\lambda}_1 e_1'$ by $(\hat{\psi}'\hat{\psi})^{-\frac{1}{2}}\hat{\psi}'$ and post-multiplying by $e_1$

$$\begin{aligned} (\hat{\psi}'\hat{\psi})^{-\frac{1}{2}}\hat{\psi}'\hat{\lambda}_1 e_1' e_1 & \leftrightarrow (\hat{\Gamma}'\hat{\Gamma})^{-\frac{1}{2}}\hat{\Gamma}'\left[\hat{\Gamma}\begin{pmatrix} \hat{\delta} & 0 \end{pmatrix} + \hat{\Gamma}_\perp \hat{\lambda}_2 \hat{D}_\perp\right] e_1 \\ & = \begin{pmatrix} (\hat{\Gamma}'\hat{\Gamma})^{\frac{1}{2}}\hat{\delta} & 0 \end{pmatrix} e_1 = (\hat{\Gamma}'\hat{\Gamma})^{\frac{1}{2}}\hat{\delta} \end{aligned} \tag{98}$$

such that $(\hat{\psi}'\hat{\psi})^{-\frac{1}{2}}\hat{\psi}'\hat{\lambda}_1$ corresponds with $(\hat{\Gamma}'\hat{\Gamma})^{\frac{1}{2}}\hat{\delta}$ and we have used that $e_1'e_1 = 1$ and that $\hat{\psi}$ corresponds with $\hat{\Gamma}$. We can therefore use $(\hat{\psi}'\hat{\psi})^{-\frac{1}{2}}\hat{\psi}'\hat{\lambda}_1$ to conduct inference on $\hat{\delta}$ and thus also $\hat{\beta}$. As $\hat{\psi}$ $(= \hat{\Gamma})$ is stochastic independent from $\hat{\lambda}_1$, $(\hat{\psi}'\hat{\psi})^{-\frac{1}{2}}\hat{\psi}'\hat{\lambda}_1$ has under $H_{03}$ a standard normal distribution

$$\hat{\lambda}_3 = (\hat{\psi}'\hat{\psi})^{-\frac{1}{2}}\hat{\psi}'\hat{\lambda}_1 \sim N(0, I_{m-1}), \tag{99}$$

and contains the elements of $\hat{\lambda}_1$ that reflect the significance of the structural parameters. We can now construct an exact test statistic to test hypothezes on the structural parameters using (99) as

$$\hat{\lambda}_3'\hat{\lambda}_3 \sim \chi^2(m-1). \tag{100}$$

To construct this statistic, we use the specification of $\hat{\lambda}_1$ and $\hat{\psi}$ (80)

$$\hat{\lambda}_1 = (X'X)^{\frac{1}{2}}\hat{\varphi}_1\omega_{11}^{-\frac{1}{2}}, \quad \hat{\psi} = (X'X)^{\frac{1}{2}}(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12})\Omega_{22.1}^{-\frac{1}{2}}, \tag{101}$$

and that $\hat{\lambda}_1$ and $\hat{\psi}$ are stochastic independent. Both $\hat{\lambda}_1$ and $\hat{\psi}$ contain unknown parameters. We replace these unknown parameters by estimators that are stochastic independent from $\hat{\Phi}_2$ and $\hat{\varphi}_1$. $\hat{\varphi}_1$ and $(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12})$ are stochastic independent and distributed as

$$\begin{aligned} \hat{\varphi}_1 & \sim N(0, \omega_{11} \otimes (X'X)^{-1}), \\ \hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12} & \sim N(\Phi_2, \Omega_{22.1} \otimes (X'X)^{-1}). \end{aligned} \tag{102}$$

Furthermore, see Muirhead (1982),

$$s_{11}^{-1}s_{12}|s_{11}) \sim N(\omega_{11}^{-1}\omega_{12}, \Omega_{22.1} \otimes s_{11}^{-1}), \tag{103}$$

such that we can replace $\omega_{11}^{-1}\omega_{12}$ by $s_{11}^{-1}s_{12}$ in (102) and obtain that

$$\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\,|s_{11}\right) \sim N(\Phi_2, \Omega_{22.1} \otimes (X'X)^{-1}) + a, \tag{104}$$

where $a = \hat{\varphi}_1(s_{11}^{-1}s_{12} - \omega_{11}^{-1}\omega_{12})$ and $a$ is the product of two independent normal random variables that both have mean zero as $\hat{\varphi}_1 \sim N(0, \omega_{11}(X'X)^{-1})$, $(s_{11}^{-1}s_{12} - \omega_{11}^{-1}\omega_{12}|s_{11}) \sim$

$N(0, \Omega_{22.1} \otimes s_{11}^{-1})$. We can simplify $a$ as $a = (X'X)^{-\frac{1}{2}} bc' \Omega_{22.1}^{\frac{1}{2}}$, where $b \sim N(0, I_k)$ $(b = \hat{\lambda}_1)$, $c|s_{11}) \sim N(0, \frac{\omega_{11}}{s_{11}} I_{m-1})$ and $b$ and $c$ are stochastic independent.

It also holds that

$$S_{22.1} = S_{22} - s_{21} s_{11}^{-1} s_{12} \sim W(\Omega_{22.1}, T-k-1), \tag{105}$$

where $W$ stands for Wishart distributed and because of which we can specify $S_{22.1}^{-\frac{1}{2}}$ as

$$S_{22.1}^{-\frac{1}{2}} = \Omega_{22.1}^{-\frac{1}{2}} Q^{-\frac{1}{2}}, \tag{106}$$

where $Q \sim W(I_{m-1}, T-k-1)$, see Muirhead (1982). We can therefore use $S_{22.1}$ to correct for the unknown variance $\Omega_{22.1}$ in (104),

$$\left[(X'X)^{\frac{1}{2}} (\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1} s_{12}) S_{22.1}^{-\frac{1}{2}} | s_{11}, Q\right] \sim N(\Theta_2 Q^{-\frac{1}{2}}, Q^{-1} \otimes I_k) + bc' Q^{-\frac{1}{2}}. \tag{107}$$

Both the conditioning random variables and the two random elements of (107) are stochastic independent from $\hat{\lambda}_1$. The first element is stochastic independent as it is equal to $\hat{\psi} Q^{-\frac{1}{2}}$ and $\hat{\psi}$ and $\hat{\lambda}_1$ are stochastic independent. The second element is stochastic independent from $\hat{\lambda}_1$ $(= b)$ as $c$ is stochastic independent from $\hat{\lambda}_1$. When we construct $\hat{\psi}$ as equal to (107), its conditional mean given $s_{11}$ and $Q$ is equal to $\Theta_2 Q^{-\frac{1}{2}}$. The resulting specification of $\hat{\psi}$,

$$\hat{\psi} = (X'X)^{\frac{1}{2}} \left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1} s_{12}\right) S_{22.1}^{-\frac{1}{2}}, \tag{108}$$

is then such that $S_{22.1}$ cancels out of the expression $\hat{\psi}(\hat{\psi}'\hat{\psi})^{-1}\hat{\psi}'$. The random variable $Q$ does then not affect $\hat{\psi}(\hat{\psi}'\hat{\psi})^{-1}\hat{\psi}'$ such that we can consider the mean of $\hat{\psi}$ as equal to $\Theta_2$. This specification of $\hat{\psi}$ thus satisfies the two necessary conditions to be used in (99) and (100) as it is a random variable that is stochastic independent from $\hat{\lambda}_1$ with a mean equal to $\Theta_2$. The exact test statistic for testing $H_{03}$ against $H_{13}$ then results as

$$\begin{aligned} F(H_{03}|H_{13}) &= \tfrac{1}{(m-1)s_{11}} y_1' X(X'X)^{-1} X'(Y_2 - y_1 s_{11}^{-1} s_{12}) \left[(Y_2 - y_1 s_{11}^{-1} s_{12})' X(X'X)^{-1} X' \right. \\ &\quad \left. (Y_2 - y_1 s_{11}^{-1} s_{12})\right]^{-1} (Y_2 - y_1 s_{11}^{-1} s_{12})' X(X'X)^{-1} X' y_1 \\ &= \tfrac{1}{m-1} \tfrac{\hat{\varphi}_1' X'X \left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1} s_{12}\right)\left(\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1} s_{12}\right)' X'X \left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1} s_{12}\right)\right)^{-1} \left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1} s_{12}\right)' X'X \hat{\varphi}_1}{s_{11}} \\ &= \tfrac{1}{m-1} \tfrac{\hat{\varphi}_1' (X'X)^{\frac{1}{2}} \hat{\psi}(\hat{\psi}'\hat{\psi})^{-1}\hat{\psi}' (X'X)^{\frac{1}{2}} \hat{\varphi}_1}{s_{11}} = \tfrac{\hat{\lambda}_1' \hat{\psi}(\hat{\psi}'\hat{\psi})^{-1}\hat{\psi}' \hat{\lambda}_1 / (m-1)}{(T-k)\frac{s_{11}}{\omega_{11}} / (T-k)} = \tfrac{\hat{\lambda}_3' \hat{\lambda}_3 / (m-1)}{(T-k)\frac{s_{11}}{\omega_{11}} / (T-k)} \\ &\sim \tfrac{\chi^2(m-1)/(m-1)}{\chi^2(T-k)/(T-k)} \sim F(m-1, T-k). \end{aligned} \tag{109}$$

To illustrate that (109) is an exact test statistic, we computed its empirical distribution for the data generating process for which the empirical distribution of the likelihood ratio based statistic (94) was different from its (asymptotic) theoretical distribution, *i.e.* the data generating process used for figure 9 with $k = 20$. Figure 12 contains this distribution jointly with the theoretical distribution and the empirical distributions of the likelihood ratio and likelihood ratio based statistics. As (109) is an exact test statistic, its empirical distribution coincides with its theoretical distribution. Figure 12 also shows that the distributions of the likelihood ratio and likelihood ratio based statistic (94) are identical but differ from their (asymptotic) theoretical distribution as they depend on nuisance parameters.

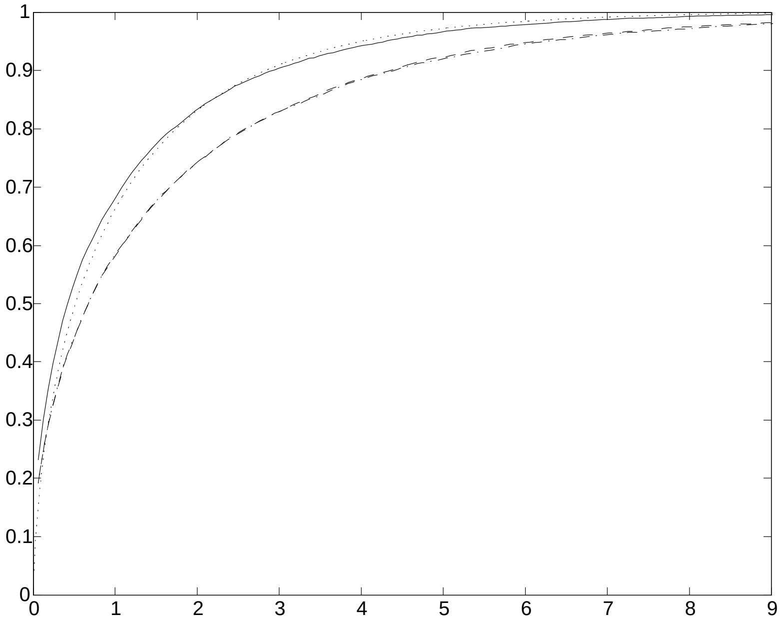


Figure 12: Distribution test statistic $\beta = 0$, $\pi_1 = 0.1$, $k = 20$, $F(1, 80)$ theo. dis. (..), exact stat. (109) (-), LR based stat. (94) (- -), LR stat (-.-).

The statistic (109) can also be used to conduct tests on $\beta$ when it has more than one element. Also for these values of $m$ it is an exact test statistic. To illustrate this, we computed its empirical distribution for a data generating process with weak instruments, $m = 3$ and $k = 20$ with 18 superfluous instruments. Figure 13 shows that the theoretical and empirical distribution of the statistic (109) coincide as it is an exact statistic. Figure 13 also shows the empirical distribution of the likelihood ratio based statistic (94) which differs from its (asymptotic) theoretical distribution as it is not an exact statistic.

#### 4.2.4 Value of some elements of the structural form parameter

The statistics (94) and (109) conduct a joint test on all of the elements of $\beta$. We can also use the orthogonal parameters to construct statistics that can be used to conduct tests on subsets of the elements of $\beta$. Consider, for example, the model,

$$\begin{aligned} y_1 &= Y_2\beta_1 + Y_3\beta_2 + \varepsilon_1, \\ Y_2 &= X\Pi_1 + V_2, \\ Y_3 &= X\Pi_2 + V_3, \end{aligned} \tag{110}$$

where $y_1,\ v_1 : T \times 1$; $Y_2,\ V_2 : T \times m_2$; $Y_3,\ V_3 : T \times m_3$, $m = m_2 + m_3 + 1$, $\beta_1 : m_2 \times 1$, $\beta_2 : m_3 \times 1$, $\Pi_1 : k \times m_2$, $\Pi_2 : k \times m_3$, $Y = (y_1\ Y_2\ Y_3)$, where we want to test the null hypothesis, $H_{04} : \beta_2 = 0$, against the alternative hypothesis $H_{14} : \beta_2 \neq 0$. The same assumptions hold for (110) as for the instrumental variables regression model (28). We specify the covariance matrix $\Omega$ and its estimator $S$ $(= \frac{1}{T-k}Y'M_XY)$ of the reduced form as, $\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$, $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$ where $\Omega_{11},\ S_{11} : (m_2 + 1) \times (m_2 + 1)$; $\Omega_{12},\ \Omega_{21}',\ S_{12},\ S_{21}' : (m_2 + 1) \times m_3$; $\Omega_{22},\ S_{22} : m_3 \times m_3$.

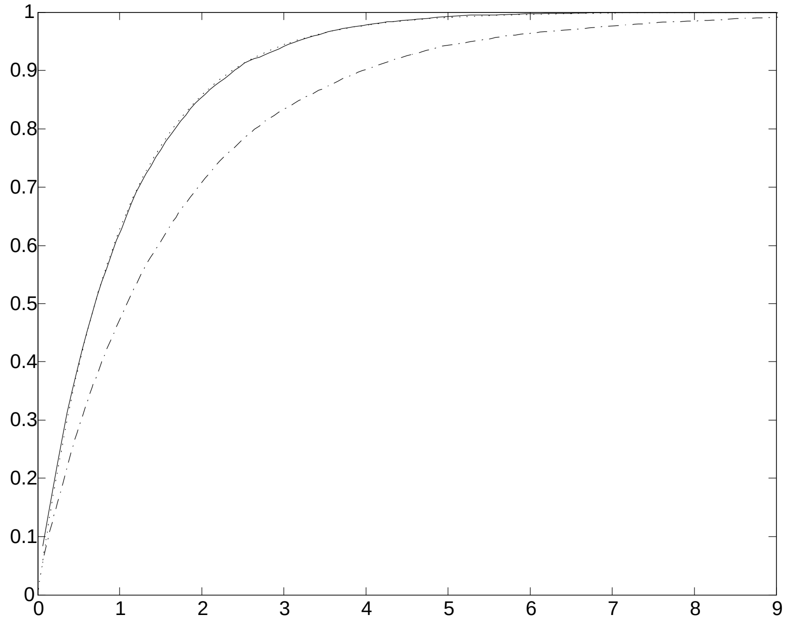


Figure 13: Distribution test statistic $\beta = 0$, $m = 3$, weak instruments $k = 20$, $F(2, 80)$ theo. dis. (..), exact stat. (109) (-), LR based stat. (94) (-.-).

**Likelihood Ratio Based Statistic** Under $H_{04}$, the degree of over-identification of (110) is equal to $k - m_2 + 1$. When $\Omega$ is known, the likelihood based test statistics for testing $H_{04}$ against the unrestricted reduced form result from (89)

$$LR(H_{04}|H_1) = \hat{\lambda}_{04}'\hat{\lambda}_{04} = (\hat{B}_{04\perp}\Omega_{11}\hat{B}_{04\perp}')^{-1}\hat{B}_{04\perp}Y_1'(M_{X\hat{\Pi}_1} - M_X)Y_1\hat{B}_{04\perp}' \sim \chi^2(k - m_2 + 1). \tag{111}$$

where $\hat{B}_{04} = \begin{pmatrix} \hat{\beta}_1 & I_{m_2} \end{pmatrix}$, $Y_1 = \begin{pmatrix} y_1 & Y_2 \end{pmatrix}$ and $(\hat{\beta}_1, \hat{\Pi}_1)$ are the liml estimators which are computed with $H_{04}$ imposed, *i.e.* the liml estimators of the instrumental variables regression model that only consists of the first $(m_2 + 1)$ equations of (110). Similarly, for testing the model under $H_{14}$ against the unrestricted reduced form, the likelihood based test statistics read

$$LR(H_{14}|H_1) = \hat{\lambda}_{14}'\hat{\lambda}_{14} = (\hat{B}_\perp\Omega\hat{B}_\perp')^{-1}\hat{B}_\perp Y'(M_{X\hat{\Pi}} - M_X)Y\hat{B}_\perp' \sim \chi^2(k - m + 1), \tag{112}$$

where $\hat{B} = \begin{pmatrix} \hat{\beta} & I_{m-1} \end{pmatrix}$, $\beta = (\beta_1' \; \beta_2')'$. When $\Omega$ is known, the likelihood ratio statistic for testing $H_{04}$ against $H_{14}$ is equal to the difference between (111) and (112)

$$\begin{aligned} LR(H_{04}|H_{14}) &= LR(H_{04}|H_1) - LR(H_{14}|H_1) \\ &= \hat{\lambda}_{04}'\hat{\lambda}_{04} - \hat{\lambda}_{14}'\hat{\lambda}_{14} \\ &= (\hat{B}_{04\perp}\Omega_{11}\hat{B}_{04\perp}')^{-1}\hat{B}_{04\perp}Y_1'(M_{X\hat{\Pi}_1} - M_X)Y_1\hat{B}_{04\perp}' \\ &\quad -(\hat{B}_\perp\Omega\hat{B}_\perp')^{-1}\hat{B}_\perp Y'(M_{X\hat{\Pi}} - M_X)Y\hat{B}_\perp', \end{aligned} \tag{113}$$

and is under $H_{04}$ distributed as a $\chi^2(m_3)$ random variable. Similar to (90), we have that under $H_{04}$

$$(T - k)\frac{\hat{B}_{03\perp}S_{11}\hat{B}_{03\perp}'}{\hat{B}_{03\perp}\Omega_{11}\hat{B}_{03\perp}'} \sim \chi^2(T - k), \tag{114}$$

where $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$ (52), with $S_{11} : (m_2+1) \times (m_2+1)$; $S_{12}$, $S'_{21} : (m_2+1) \times m_3$; $S_{22} : m_3 \times m_3$, such that

$$\begin{aligned} F(H_{04}|H_{14}) &= \tfrac{1}{m_3}\left[tr\left(\left(S_{11}^{-1} - S_{11}^{-1}\hat{B}'_{04}(\hat{B}_{04}S_{11}^{-1}\hat{B}'_{04})^{-1}\hat{B}_{04}S_{11}^{-1}\right)Y_1'(M_{X\hat{\Pi}_1} - M_X)Y_1\right)\right. \\ &\quad \left. -tr\left(\left(S^{-1} - S^{-1}\hat{B}'(\hat{B}S^{-1}\hat{B}')^{-1}\hat{B}S^{-1}\right)Y'(M_{X\hat{\Pi}} - M_X)Y\right)\right] \\ &= \tfrac{1}{m_3}\left[\frac{\hat{B}_{04\perp}Y_1'(M_{X\hat{\Pi}_1} - M_X)Y_1\hat{B}'_{04\perp}}{\hat{B}_{04\perp}S_{11}\hat{B}'_{04\perp}} - \frac{\hat{B}_{\perp}Y'(M_{X\hat{\Pi}} - M_X)Y\hat{B}'_{\perp}}{\hat{B}_{\perp}S\hat{B}'_{\perp}}\right] \\ &= \left[\frac{\frac{\hat{B}_{04\perp}Y_1'(M_{X\hat{\Pi}_1} - M_X)Y_1\hat{B}'_{04\perp}}{\hat{B}_{04\perp}\Omega_{11}\hat{B}'_{04\perp}}\Big/ m_3}{(T-k)\frac{\hat{B}_{04\perp}S_{11}\hat{B}'_{04\perp}}{\hat{B}_{04\perp}\Omega_{11}\hat{B}'_{04\perp}}\Big/(T-k)} - \frac{\frac{\hat{B}_{\perp}Y'(M_{X\hat{\Pi}} - M_X)Y\hat{B}'_{\perp}}{\hat{B}_{\perp}\Omega\hat{B}'_{\perp}}\Big/ m_3}{(T-k)\frac{\hat{B}_{\perp}S\hat{B}'_{\perp}}{\hat{B}_{\perp}\Omega\hat{B}'_{\perp}}\Big/(T-k)}\right] \\ &\sim F(m_3, T-k), \end{aligned} \tag{115}$$

when $T$ is sufficiently large. This results since both $(T-k)\frac{\hat{B}_{04\perp}S_{11}\hat{B}'_{04\perp}}{\hat{B}_{04\perp}\Omega_{11}\hat{B}'_{04\perp}}\Big/(T-k)$ and $(T-k)\frac{\hat{B}_{\perp}S\hat{B}'_{\perp}}{\hat{B}_{\perp}\Omega\hat{B}'_{\perp}}\Big/(T-k)$ are $\chi^2(T-k)$ random variables divided by their degrees of freedom parameters. When $T$ is sufficiently large they are therefore approximately equal to one.

**Exact Test Statistic** The statistic (115) is not an exact test statistic such that its distribution depends on nuisance parameters. In a similar way as for the joint test statistic on all the structural form parameters, we can also construct an exact test statistic to test hypothezes on subsets of the structural form parameters such as $H_{04} : \beta_2 = 0$ against $H_{14} : \beta_2 \neq 0$. To construct the exact test statistic for this hypothesis we use the specifications of $\hat{\Theta}$ $(= (X'X)^{\frac{1}{2}}\hat{\Phi}\Omega^{-\frac{1}{2}})$,

$$\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_{\perp}\hat{\lambda}_4\hat{D}_{\perp}, \tag{116}$$

where $\hat{\Gamma} = \begin{pmatrix} \hat{\Gamma}_1 & \hat{\Gamma}_2 \end{pmatrix}$, $\hat{\Gamma}_1 : k \times m_2$, $\hat{\Gamma}_2 : k \times m_3$, $\hat{D} = \begin{pmatrix} \hat{\delta}_1 & I_{m_2} & 0 \\ \hat{\delta}_2 & 0 & I_{m_3} \end{pmatrix} = \begin{pmatrix} \hat{D}_1 \\ \hat{D}_2 \end{pmatrix}$, $\hat{\delta}_1 : m_2 \times 1$, $\hat{\delta}_2 : m_3 \times 1$, $\hat{\lambda}_4 : (k - m_2 - m_3) \times 1$, $(k - m_2 - m_3) > 0$; and

$$\hat{\Theta} = \begin{pmatrix} \hat{\Theta}_1 & \hat{\Theta}_2 \end{pmatrix}, \ \hat{\Theta}_1 = \hat{G}\hat{E} + \hat{G}_{\perp}\hat{\lambda}_5\hat{E}_{\perp}, \tag{117}$$

where $\hat{\Theta}_1 : k \times (m_2+1)$, $\hat{\Theta}_2 : k \times m_3$, $\hat{E} = \begin{pmatrix} \hat{\xi}_1 & I_{m_2} \end{pmatrix}$, $\hat{\xi}_1 : m_2 \times 1$, $\hat{G} : k \times m_2$, and $\hat{\lambda}_5 : (k - m_2) \times 1$. As $\hat{\Theta}$ is identical in both specifications, $\hat{\Gamma}_1$, $\hat{\Gamma}_2$, and $\hat{\delta}_1$ from (116) correspond with $\hat{G}$, $\hat{\Theta}_2$ and $\hat{\xi}_1$ from (117) resp.. This then implies that

$$\begin{pmatrix} \hat{\Gamma}_1 & \hat{\Gamma}_2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ \hat{\delta}_2 & 0 & 0 \end{pmatrix} + \hat{\Gamma}_{\perp}\hat{\lambda}_4\hat{D}_{\perp} \tag{118}$$

from (116) corresponds with $\begin{pmatrix} \hat{G}_{\perp}\hat{\lambda}_5\hat{E}_{\perp} & 0 \end{pmatrix}$ from (117). We can pre-multiply (118) by $\hat{\Gamma}'_2$ and post-multiply by $\left(\hat{D}_1\right)'_{\perp}$, which corresponds with $\begin{pmatrix} \hat{E}'_{\perp} & 0 \end{pmatrix}$, and obtain that $\hat{\Gamma}'_2 \begin{pmatrix} \hat{\delta}_2 & 0 & 0 \end{pmatrix}$ $\left(\hat{D}_1\right)'_{\perp}$ corresponds with $\hat{\Theta}'_2\hat{G}_{\perp}\hat{\lambda}_5$. This shows that we can use $\left(\hat{\Theta}'_2\hat{G}_{\perp}\hat{G}'_{\perp}\hat{\Theta}_2\right)^{-\frac{1}{2}}\hat{\Theta}'_2\hat{G}_{\perp}\hat{\lambda}_5$ to conduct inference on $\hat{\delta}_2$ and thus also $\hat{\beta}_2$. We therefore analyze the different elements of this expression to obtain an exact test statistic for conducting tests on $\beta_2$ only.

$\hat{\lambda}_5$ is normally distributed

$$\hat{\lambda}_5 \sim N(0, I_{k-m_2}), \tag{119}$$

from which it results that $\hat{\lambda}_6 = \left(\hat{\Theta}_2'\hat{G}_\perp\hat{G}_\perp'\hat{\Theta}_2\right)^{-\frac{1}{2}}\hat{\Theta}_2'\hat{G}_\perp\hat{\lambda}_5$ is distributed as

$$\hat{\lambda}_6 \sim N(0, I_{m_3}), \tag{120}$$

since both $\hat{G}_\perp$ and $\hat{\Theta}_2$ are stochastic independent from $\hat{\lambda}_5$ and $k - m_2 - m_3 > 0$. We can use (120) to construct an exact test statistic to test $H_{04}$ against $H_{14}$ as

$$\hat{\lambda}_6'\hat{\lambda}_6 \sim \chi^2(m_3). \tag{121}$$

To construct the exact test statistic for testing $H_{04}$ against $H_{14}$, we first construct the different elements of $\hat{\lambda}_6$. $\hat{\Theta}_2$ results from (80)

$$\begin{aligned} \hat{\Theta}_2 &= (X'X)^{\frac{1}{2}}\left(\hat{\Phi}_2 - \hat{\Phi}_1\Omega_{11}^{-1}\Omega_{12}\right)\Omega_{22.1}^{-\frac{1}{2}} \\ &= (X'X)^{-\frac{1}{2}}X'\left(Y_3 - \begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)\Omega_{22.1}^{-\frac{1}{2}}, \end{aligned} \tag{122}$$

where $\hat{\Phi}_1 = (X'X)^{-1}X'\begin{pmatrix} y_1 & Y_2 \end{pmatrix}$, $\hat{\Phi}_2 = (X'X)^{-1}X'Y_3$, while $\hat{\lambda}_5$ and $\hat{G}_\perp$ result from the decomposition of $\hat{\Theta}_1$ (117) and (86)-(88)

$$\begin{aligned} \hat{\lambda}_5 &= (\hat{\Pi}_{1\perp}'(X'X)^{-1}\hat{\Pi}_{1\perp})^{-\frac{1}{2}}\hat{\Pi}_{1\perp}'(X'X)^{-1}X'\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\hat{B}_{1\perp}'(\hat{B}_{1\perp}\Omega_{11}\hat{B}_{1\perp}')^{-\frac{1}{2}} \\ \hat{G}_\perp &= (X'X)^{-\frac{1}{2}}\hat{\Pi}_{1\perp}(\hat{\Pi}_{1\perp}'(X'X)^{-1}\hat{\Pi}_{1\perp})^{-\frac{1}{2}}, \end{aligned} \tag{123}$$

where $\hat{\Pi}_1$ and $\hat{\beta}_1$, which is used in $\hat{B}_1 = \begin{pmatrix} \hat{\beta}_1 & I_{m_2} \end{pmatrix}$, are the liml estimators of the parameters of the instrumental variables regression model that only consists of the first $(m_2+1)$ equations of (110). Combining these expressions, we obtain that

$$\hat{\Theta}_2'\hat{G}_\perp = \Omega_{22.1}^{-\frac{1}{2}\prime}\left(\hat{\Phi}_2 - \hat{\Phi}_1\Omega_{11}^{-1}\Omega_{12}\right)'\hat{\Pi}_{1\perp}(\hat{\Pi}_{1\perp}'(X'X)^{-1}\hat{\Pi}_{1\perp})^{-\frac{1}{2}}, \tag{124}$$

such that

$$\begin{aligned} \hat{\Theta}_2'\hat{G}_\perp\hat{G}_\perp'\hat{\Theta}_2 &= \Omega_{22.1}^{-\frac{1}{2}\prime}\left(\hat{\Phi}_2 - \hat{\Phi}_1\Omega_{11}^{-1}\Omega_{12}\right)'\hat{\Pi}_{1\perp}(\hat{\Pi}_{1\perp}'(X'X)^{-1}\hat{\Pi}_{1\perp})^{-1}\hat{\Pi}_{1\perp}'\left(\hat{\Phi}_2 - \hat{\Phi}_1\Omega_{11}^{-1}\Omega_{12}\right)\Omega_{22.1}^{-\frac{1}{2}} \\ &= \Omega_{22.1}^{-\frac{1}{2}\prime}\left(Y_3 - \begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)'\left[M_{X\hat{\Pi}_1} - M_X\right]\left(Y_3 - \begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)\Omega_{22.1}^{-\frac{1}{2}}, \end{aligned} \tag{125}$$

and

$$\begin{aligned} \hat{\Theta}_2'\hat{G}_\perp\hat{\lambda}_5 &= \Omega_{22.1}^{-\frac{1}{2}\prime}\left(\hat{\Phi}_2 - \hat{\Phi}_1\Omega_{11}^{-1}\Omega_{12}\right)'\hat{\Pi}_{1\perp}(\hat{\Pi}_{1\perp}'(X'X)^{-1}\hat{\Pi}_{1\perp})^{-1}\hat{\Pi}_{1\perp}' \\ &\quad (X'X)^{-1}X'\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\hat{B}_{1\perp}'(\hat{B}_{1\perp}\Omega_{11}\hat{B}_{1\perp}')^{-\frac{1}{2}} \\ &= \Omega_{22.1}^{-\frac{1}{2}\prime}\left(Y_3 - \begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)\left[M_{X\hat{\Pi}_1} - M_X\right]\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\hat{B}_{1\perp}'(\hat{B}_{1\perp}\Omega_{11}\hat{B}_{1\perp}')^{-\frac{1}{2}}. \end{aligned} \tag{126}$$

As a consequence, we can express $\hat{\lambda}_6'\hat{\lambda}_6$ as

$$\begin{aligned}
\hat{\lambda}_6'\hat{\lambda}_6 &= \tfrac{1}{\hat{B}_{1\perp}\Omega_{11}\hat{B}_{1\perp}'}\hat{B}_{1\perp}\begin{pmatrix} y_1 & Y_2 \end{pmatrix}'\left[M_{X\hat{\Pi}_1}-M_X\right]\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right) \\
&\left(\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)'\left[M_{X\hat{\Pi}_1}-M_X\right]\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)\right)^{-1} \\
&\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)'\left[M_{X\hat{\Pi}_1}-M_X\right]\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\hat{B}_{1\perp}' \\
&= tr\left\{\left(\Omega_{11}^{-1}-\Omega_{11}^{-1}\hat{B}_1'\left(\hat{B}_1\Omega_{11}^{-1}\hat{B}_1'\right)^{-1}\hat{B}_1\Omega_{11}^{-1}\right)\begin{pmatrix} y_1 & Y_2 \end{pmatrix}'\left[M_{X\hat{\Pi}_1}-M_X\right]\right. \\
&\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)\left[\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)'\left[M_{X\hat{\Pi}_1}-M_X\right]\right. \\
&\left.\left.\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)\right]^{-1}\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)'\left[M_{X\hat{\Pi}_1}-M_X\right]\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\right\}.
\end{aligned} \tag{127}$$

The expression of $\hat{\lambda}_6'\hat{\lambda}_6$ in (127) contains the parameters $\Omega_{11}$ and $\Omega_{12}$ which are typically unobserved. We replace these parameters by estimators that are stochastic independent from $\hat{\Theta}$, to obtain the exact test statistic. Using (90), it results that

$$(T-K)\frac{\hat{B}_{1\perp}S_{11}\hat{B}_{1\perp}'}{\hat{B}_{1\perp}\Omega_{11}\hat{B}_{1\perp}'}\sim\chi^2(T-k), \tag{128}$$

as $S_{11}$ has a Wishart distribution with scale matrix $\Omega_{11}$ and $\hat{B}_{1\perp}$ is stochastic independent from $S_{11}$. As explained for the exact test statistic for testing all the elements of $\beta$, we can also replace $\Omega_{11}^{-1}\Omega_{12}$ by $S_{11}^{-1}S_{12}$. The resulting expression of the exact statistic for testing $H_{04}:\beta_2=0$ against $H_{14}:\beta_2\neq 0$ that results from $\hat{\lambda}_6'\hat{\lambda}_6$ then becomes,

$$\begin{aligned}
&F(H_{04}|H_{14}) \\
&= \tfrac{1}{m_3}tr\left\{\left(S_{11}^{-1}-S_{11}^{-1}\hat{B}_1'\left(\hat{B}_1S_{11}^{-1}\hat{B}_1'\right)^{-1}\hat{B}_1S_{11}^{-1}\right)\begin{pmatrix} y_1 & Y_2 \end{pmatrix}'\left[M_{X\hat{\Pi}_1}-M_X\right]\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}S_{11}^{-1}S_{12}\right)\right. \\
&\quad\left[\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}S_{11}^{-1}S_{12}\right)'\left[M_{X\hat{\Pi}_1}-M_X\right]\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}S_{11}^{-1}S_{12}\right)\right]^{-1} \\
&\quad\left.\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}S_{11}^{-1}S_{12}\right)'\left[M_{X\hat{\Pi}_1}-M_X\right]\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\right\} \\
&= \tfrac{1}{\left((T-k)\hat{B}_{1\perp}S_{11}\hat{B}_{1\perp}'/\hat{B}_{1\perp}\Omega_{11}\hat{B}_{1\perp}'\right)/(T-k)}\left\{\tfrac{1}{m_3}\tfrac{1}{\hat{B}_{1\perp}\Omega_{11}\hat{B}_{1\perp}'}\hat{B}_{1\perp}\begin{pmatrix} y_1 & Y_2 \end{pmatrix}'\left[M_{X\hat{\Pi}_1}-M_X\right]\right. \\
&\quad\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)\left[\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)'\left[M_{X\hat{\Pi}_1}-M_X\right]\right. \\
&\quad\left.\left.\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)\right]^{-1}\left(Y_3-\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\Omega_{11}^{-1}\Omega_{12}\right)'\left[M_{X\hat{\Pi}_1}-M_X\right]\begin{pmatrix} y_1 & Y_2 \end{pmatrix}\hat{B}_{1\perp}'\right\} \\
&\sim \tfrac{\chi^2(m_3)/m_3}{\chi^2(T-k)/(T-k)}\sim F(m_3,T-k).
\end{aligned} \tag{129}$$

For a data generating process with $T=100$, $m=3$, $m_3=1$, weak instruments, strong endogeneity and a large degree of over-identification (18), we computed the empirical distributions of the exact test statistic (129) and the likelihood ratio based statistic (115). Both of these empirical distributions and the theoretical distribution of the exact statistic (129), which is a $F(1,80)$ distribution, are shown in figure 14. The theoretical distribution of (129) can be considered as the asymptotic distribution of (115). Figure 14 shows that the test statistic (129) is indeed an exact statistic as its empirical and theoretical distributions are identical. The distribution of the likelihood ratio based statistic (115) differs from its asymptotic distribution which shows that it is not an exact statistic and that its distribution in small samples depends on nuisance parameters.

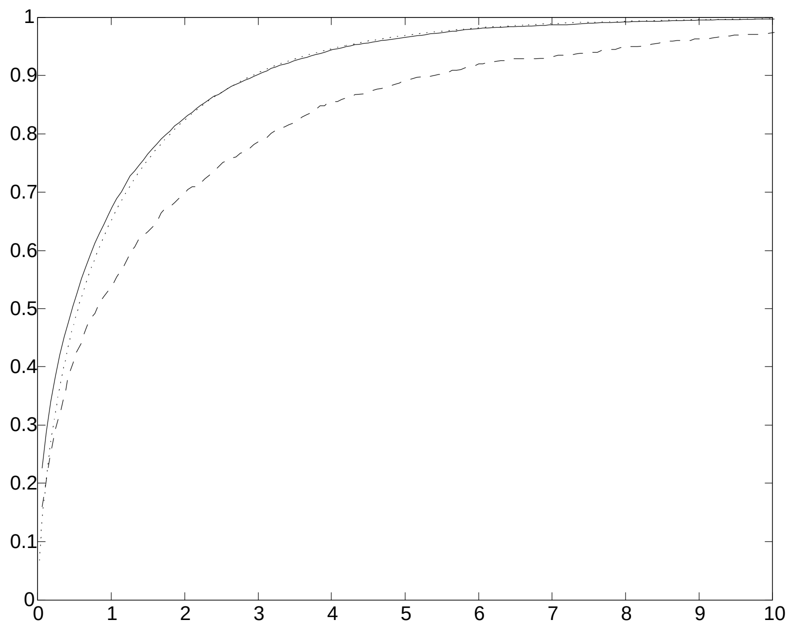


Figure 14: Empirical distribution exact stat. (129) (-), lr based stat. (115) (- -) and $F(1,80)$ theoretical distribution (..).

## 4.3 Confidence Regions

Instead of testing for zero values of the structural form parameters the statistics in section 4.2 can also be used to test for other values. In that case we replace the endogenous variable $y_1$ by $y_1^* = y_1 - Y_2\beta_0$ when the hypothesis of interest is $H_{01}$ or $H_{03}$ and the hypothesized value for $\beta$ is $\beta_0$, or $y_1^* = y_1 - Y_3\beta_{20}$ in case $H_{04}$ is the hypothesis of interest and the hypothesized value for $\beta_2$ is $\beta_{20}$. Except for $y_1$ the whole analysis conducted in the previous sections then remains unchanged.

We can invert an observed test statistic using its distribution under the hypothesis of interest to obtain the $p$-value of the hypothesis. Using this procedure for different hypothesized values of $\beta_0$ or $\beta_{20}$ enables us to construct a $(100-\alpha)\%$ confidence region for $\beta$ or $\beta_2$. Since the statistics from the previous section are exact we can thus construct exact confidence regions even though the density of the mle depends on unobserved nuisance parameters. Note that the confidence regions of the parameters of the instrumental variables regression model can be unbounded, discontinuous or empty, see *e.g.* Dufour (1997) and Zivot *et. al.* (1998).

# 5 Conclusions

We showed that the convenient statistical properties that hold in linear models also apply to a more general class of models. This allows us to analyze these models in a novel manner. We therefore conducted such an analysis of the instrumental variables regression model and in this way obtained new insights into the statistical properties of this model. We constructed, for example, a novel expression for the density of the liml estimator and exact test statistics. Especially the latter are important for practitioners given the common appearance of weak instruments in applied work, see *e.g.* Angrist and Krueger (1991), and the robustness of these test statistics to this phenomenon. An important area for future research is therefore to apply our test statistics to real data-sets and to study their power properties. This will also show how important our initial assumptions, like, for example, the normality of the disturbances, are. The assumption of normality of the disturbances can at least be relaxed to normality of the least squares estimator but some of our results probably also apply for mixtures of normal disturbances.

Another area of future research is to apply the analysis to other models that satisfy the orthogonality condition.

# Appendix

**A. Proof of Theorem 3: The marginal densities of the MLE and the Score Vector**

The joint density of $(\hat{\psi}, \hat{\lambda})$ can be specified as

$$\begin{aligned} p(\hat{\psi}, \hat{\lambda}) \propto\ & p(\hat{\Theta}(\hat{\psi}, \hat{\lambda}))|J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda}))| \\ \propto\ & \left| \left( \left( \tfrac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right) + \left( \hat{\lambda}' \otimes I_k \right) \left( \tfrac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}} \right) \right)' \left( \left( \tfrac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right) + \left( \hat{\lambda}' \otimes I_k \right) \left( \tfrac{\partial vec(q)}{\partial \psi'}|_{\hat{\psi}} \right) \right) \right|^{\frac{1}{2}} \\ & \exp \left[ -\tfrac{1}{2} \left( r(\hat{\psi}) + q(\hat{\psi})\hat{\lambda} - \Theta \right)' \left( r(\hat{\psi}) + q(\hat{\psi})\hat{\lambda} - \Theta \right) \right]. \end{aligned}$$

The orthogonality conditions (5) imply that we can also directly solve $\hat{\psi}$ from $\hat{\Theta}$ without the involvement of $\hat{\lambda}$ since

$$\left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)' \hat{\Theta} = \left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)' (r(\hat{\psi}) + q(\hat{\psi})\hat{\lambda}) = \left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)' r(\hat{\psi}) = \left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)' \hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}.$$

As a consequence, we can solve for $\hat{\psi}$ from $\hat{\Theta}$ by using (3)

$$\left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)' \hat{\Theta} = \left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)' r(\hat{\psi}),$$

which, given a value of $\hat{\Theta}$, are $m$ equations with the $m$ elements of $\hat{\psi}$ as the only unknown elements such that $\hat{\psi}$ is exactly identified. We can solve for $\hat{\psi}$ as it results from the first two orthogonality conditions from (5) that $r(\hat{\varphi})$ is spanned by $\frac{\partial r}{\partial \psi'}$, *i.e.* $r(\hat{\psi}) = \left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)' g(\hat{\psi})$ with $g(\hat{\psi})$ a $m$-dimensional continuous differentiable function of $\hat{\psi}$, such that a unique solution $\hat{\psi}$ exists. When we solve for $\hat{\psi}$ from $\hat{\Theta}(\hat{\psi}, \hat{\lambda})$, we first map $\hat{\Theta}(\hat{\psi}, \hat{\lambda})$ onto $\hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}$, which is equal to $r(\hat{\psi})$, and then solve for $\hat{\psi}$ from $\hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}$. The projection of $\hat{\Theta}(\hat{\psi}, \hat{\lambda})$ onto $\hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}$ is an orthogonal projection as the difference between $\hat{\Theta}(\hat{\psi}, \hat{\lambda})$ and $\hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}$, *i.e.* $q(\hat{\varphi})\hat{\lambda}$, is orthogonal to $\hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}$. Only the projection onto $\hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}$ is an orthogonal projection as projections onto other values of $\hat{\lambda}$, say $\lambda_0 \neq 0$, do not have the property that the difference between the original value and the projected value is orthogonal to the projected value, $(\hat{\Theta}(\hat{\psi}, \hat{\lambda}) - \hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=\lambda_0})'\hat{\Theta}(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=\lambda_0} \neq 0$. When we have obtained the value of $\hat{\psi}$ from $\hat{\Theta}$, we can construct $\hat{\lambda}$ as

$$\hat{\lambda} = q(\hat{\psi})'\hat{\Theta}.$$

Since $\hat{\lambda}$ is not involved when we solve $\hat{\psi}$ from $\hat{\Theta}$, we can construct the marginal density of $\hat{\psi}$ directly from the marginal density of $\hat{\Theta}$. This is also reflected in the equation for $\hat{\lambda}$ as that equation shows that, by construction, $\hat{\lambda}$ is stochastic independent of $\left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)' \hat{\Theta}$, since $\left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)'$ and $q(\hat{\psi})$ are orthogonal and $\hat{\Theta}$ has a normal distribution with an identity covariance matrix, while $\left( \frac{\partial r}{\partial \psi'} \right)' \hat{\Theta}$ is the random variable from which we obtain $\hat{\psi}$. All implicit values of $\hat{\lambda}$ in $\hat{\Theta}(\hat{\psi}, \hat{\lambda})$ thus lead to the same value of $\hat{\psi}$ when we solve for $\hat{\psi}$. This shows that $\hat{\lambda}$ operates in the space orthogonal to $\left( \frac{\partial r}{\partial \psi'}|_{\hat{\psi}} \right)$ and doesnot influence the solution of $\hat{\psi}$. When we solve for $\hat{\psi}$

we therefore (implicitly) conduct an orthogonal projection of $\hat{\Theta}(\hat{\psi},\hat{\lambda})$ onto $\hat{\Theta}(\hat{\psi},\hat{\lambda})|_{\hat{\lambda}=0}=r(\hat{\psi})$ for all values of $\hat{\lambda}$ since $(\hat{\Theta}(\hat{\psi},\hat{\lambda})-\hat{\Theta}(\hat{\psi},\hat{\lambda})|_{\hat{\lambda}=0})'\hat{\Theta}(\hat{\psi},\hat{\lambda})|_{\hat{\lambda}=0}=0$. Afterwards, given $\hat{\psi}$, we can then obtain the value of $\hat{\lambda}$ from $\hat{\Theta}(\hat{\psi},\hat{\lambda})$. Integrating the joint density of $(\hat{\psi},\hat{\lambda})$ over $\hat{\lambda}$ to obtain the marginal density of $\hat{\psi}$ is then identical to conditioning on the value of $\hat{\lambda}$ where all values of $\hat{\lambda}$ are mapped on using the orthogonal projection that we use to solve for $\hat{\psi}$, *i.e.* $\hat{\lambda}=0$. The marginal density of $\hat{\psi}$ is therefore equal to the conditional density of $\hat{\psi}$ given that $\hat{\lambda}$ is equal to zero

$$\begin{array}{rl} p(\hat{\psi}) \propto & \int_{\mathbb{R}^{k-m}} p(\hat{\psi},\hat{\lambda})d\hat{\lambda} \\ \propto & \int_{\mathbb{R}^{k-m}} p(\hat{\Theta}(\hat{\psi},\hat{\lambda}))|J(\hat{\Theta},(\hat{\psi},\hat{\lambda}))|d\hat{\lambda} \\ \propto & \int_{\mathbb{R}^{k-m}} \left[p(\hat{\Theta}(\hat{\psi},\hat{\lambda}))|_{\hat{\lambda}=0}|J(\hat{\Theta},(\hat{\psi},\hat{\lambda}))|_{\hat{\lambda}=0}|\right] p(\hat{\lambda}|\hat{\psi})d\hat{\lambda} \\ \propto & \left[p(\hat{\Theta}(\hat{\psi},\hat{\lambda}))|_{\hat{\lambda}=0}|J(\hat{\Theta},(\hat{\psi},\hat{\lambda}))|_{\hat{\lambda}=0}|\right] \int_{\mathbb{R}^{k-m}} p(\hat{\lambda}|\hat{\psi})d\hat{\lambda} \\ \propto & p(\hat{\Theta}(\hat{\psi},\hat{\lambda}))|_{\hat{\lambda}=0}|J(\hat{\Theta},(\hat{\psi},\hat{\lambda}))|_{\hat{\lambda}=0}| \\ \propto & p(\hat{\psi},\hat{\lambda})|_{\hat{\lambda}=0} \\ = & p(\hat{\psi}|\hat{\lambda}=0). \end{array}$$

Since we can solve for $\hat{\psi}$ from $\hat{\Theta}$ in a way that doesnot involve $\hat{\lambda}$, and therefore obtain the marginal density of $\hat{\psi}$, we can also construct the conditional density of $\hat{\lambda}$ given $\hat{\psi}$. Because $\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)'\hat{\Theta}$ is stochastic independent from $q(\hat{\psi})'\hat{\Theta}$, the conditional density of $\hat{\lambda}$ given $\hat{\psi}$ then results as

$$p(\hat{\lambda}|\hat{\psi}) \propto \exp\left[-\frac{1}{2}(\hat{\lambda}-\lambda(\hat{\psi}))'(\hat{\lambda}-\lambda(\hat{\psi}))\right],$$

where $\lambda(\hat{\psi})=q(\hat{\psi})'\Theta$, and is a normal density

$$\hat{\lambda}|\hat{\psi}) \sim N(q(\hat{\psi})'\Theta, I_{k-m}).$$

The random variable $\hat{\lambda}$ is essentially stochastic independent from $\hat{\psi}$ as $\hat{\psi}$ results from $\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)'\hat{\Theta}$ which is stochastic independent of $\hat{\lambda}$. $\hat{\psi}$ is therefore only reflected in the mean of the conditional (normal) density of $\hat{\lambda}$ and not in the variance. To construct the marginal density of $\hat{\lambda}$, we have to take the expectation of the mean of the conditional density of $\hat{\lambda}$ with respect to $\hat{\psi}$. In order to construct this expectation, we first construct the moment generating function of $r(\hat{\psi})$. The marginal density of $\hat{\psi}$ reads

$$\begin{array}{rl} p(\hat{\psi}) & = p(\hat{\psi}|\hat{\lambda}=0) \\ & \propto \left|\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)'\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(r(\hat{\psi})-r(\psi)\right)'\left(r(\hat{\psi})-r(\psi)\right)\right], \end{array}$$

where we have assumed that $E_{\hat{\Theta}}(\hat{\Theta})=r(\psi)$. We construct the moment generating function of $r(\hat{\psi})$ using

$$M(u)=E_{\hat{\psi}}\left[\exp\left(u'r(\hat{\psi})\right)\right]$$

such that

$$E_{\hat{\psi}}(r(\hat{\psi})) = \left.\frac{\partial M(u)}{\partial u}\right|_{u=0}.$$

The moment generating function of $r(\hat{\psi})$ then results as

$$\begin{aligned}
M(u) &= E_{\hat{\psi}}\left[\exp\left(u'r(\hat{\psi})\right)\right] \\
&= \int_{\mathbb{R}^m} \exp\left(u'r(\hat{\psi})\right) p(\hat{\psi}) d\hat{\psi} \\
&= \int_{\mathbb{R}^m} c\left|\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)'\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)\right|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}\left(r(\hat{\psi})-r(\psi)\right)'\left(r(\hat{\psi})-r(\psi)\right)+u'r(\hat{\psi})\right] d\hat{\psi} \\
&= \exp\left[r(\psi)'u+\tfrac{1}{2}u'u\right] \\
&\quad \int_{\mathbb{R}^m} c\left|\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)'\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)\right|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}\left(r(\hat{\psi})-(r(\psi)+u)\right)'\left(r(\hat{\psi})-(r(\psi)+u)\right)\right] d\hat{\psi} \\
&= \exp\left[r(\psi)'u+\tfrac{1}{2}u'u\right]
\end{aligned}$$

where $c^{-1}=\int_{\mathbb{R}^m}\left|\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)'\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}}\right)\right|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}\left(r(\hat{\psi})-r(\psi)\right)'\left(r(\hat{\psi})-r(\psi)\right)\right] d\hat{\psi}$, such that

$$E_{\hat{\psi}}(r(\hat{\psi})) = \left.\frac{\partial M(u)}{\partial ru}\right|_{u=0} = r(\psi).$$

Because $E_{\hat{\Theta}}(\hat{\Theta}) = r(\psi)$ and $E_{\hat{\psi}}(r(\hat{\psi})) = r(\psi)$,

$$\begin{aligned}
E_{\hat{\Theta}}(\hat{\Theta}) =&\ r(\psi) \Leftrightarrow \\
E_{(\hat{\lambda},\hat{\psi})}(r(\hat{\psi})+q(\hat{\psi})\hat{\lambda}) =&\ r(\psi),
\end{aligned}$$

it results that

$$E_{(\hat{\lambda},\hat{\psi})}(q(\hat{\psi})\hat{\lambda}) = \ 0.$$

As explained before, $\hat{\lambda}$ results from $q(\hat{\psi})'\hat{\Theta}$ and $\hat{\psi}$ results from $\left(\frac{\partial r}{\partial \psi'}\right)'\hat{\Theta}$ which are stochastic independent. Furthermore, $\hat{\psi}$ is independently obtained of $\hat{\lambda}$. As a consequence,

$$E_{(\hat{\lambda},\hat{\psi})}(q(\hat{\psi})\hat{\lambda}) = E_{\hat{\lambda}}E_{\hat{\psi}|\hat{\lambda}}(q(\hat{\psi})\hat{\lambda}) = E_{\hat{\lambda}}E_{\hat{\psi}}(q(\hat{\psi})\hat{\lambda}) = E_{\hat{\psi}}(q(\hat{\psi}))E_{\hat{\lambda}}(\hat{\lambda})$$

and combining this with $E_{(\hat{\lambda},\hat{\psi})}(q(\hat{\psi})\hat{\lambda}) = 0$ implies that $E_{\hat{\lambda}}(\hat{\lambda}) = 0$.

The marginal density of $\hat{\lambda}$ results by integrating the product of the conditional density of $\hat{\lambda}$ and the marginal density of $\hat{\psi}$ over $\hat{\psi}$. $\hat{\psi}$ is only present in the mean of the conditional density of $\hat{\lambda}$ such that we only need to consider the expectation of the mean of the conditional density with respect to $\hat{\psi}$. As $E_{\hat{\lambda}}(\hat{\lambda}) = 0$, it then results that

$$0 = E_{\hat{\lambda}}(\hat{\lambda}) = E_{\hat{\psi}}\left(E_{\hat{\lambda}|\hat{\psi}}(\hat{\lambda})\right) = E_{\hat{\psi}}\left(q(\hat{\psi})'r(\psi)\right) = E_{\hat{\psi}}\left(q(\hat{\psi})\right)'r(\psi)$$

and the marginal density of $\hat{\lambda}$ is therefore standard normal

$$\hat{\lambda} \sim N(0, I_{k-m}).$$

### B. Density LIML estimator

The density of the liml estimator $\hat{\beta}$ is constructed in four steps:

1. Construct the densities of the "$t$-values" of the least squares estimator and the covariance matrix estimator.
2. Construct conditional density of "$t-$values" given that they have reduced rank.
3. Solve for the liml estimator from the "$t-$values" under reduced rank and construct the joint density of $(\hat{\beta}, \hat{\Pi})$ and the covariance matrix estimator.
4. Integrate out $\hat{\Pi}$ and obtain density of the liml estimator $\hat{\beta}$ given the covariance matrix estimator.

In the following we discuss each of the four different steps:

**1.** To construct the density of the LIML estimator of $\beta$, $\hat{\beta}$, we use that the OLS estimator, $\hat{\Phi} = (X'X)^{-1}X'Y$, is distributed as,

$$\hat{\Phi} \sim N(\Phi, \Omega \otimes (X'X)^{-1}),$$

where $\Phi = \Pi B$ $B = \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$. The "$t-$values" of $\hat{\Phi}$ are defined by $\hat{\Theta} = (X'X)^{\frac{1}{2}}\hat{\Phi}\Omega^{-\frac{1}{2}}$, and are distributed as,

$$\hat{\Theta} \sim N(\Theta, I_m \otimes I_k),$$

where $\Theta = (X'X)^{\frac{1}{2}}\Phi\Omega^{-\frac{1}{2}}$. The density function of these "$t-$values" therefore reads,

$$p(\hat{\Theta}) \propto \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Theta} - \Theta\right)'\left(\hat{\Theta} - \Theta\right)\right)\right].$$

The covariance matrix estimator $S = \frac{1}{T-k}Y'M_XY$ is distributed as,

$$S \sim W(\tfrac{1}{T-k}\Omega, T-k),$$

and is stochastically independent of $\hat{\Phi}$. The expectation of this random variable is $\Omega$. Instead of the covariance matrix estimator $S$, we use the covariance matrix estimator $\hat{\Omega} = \Omega S^{-1}\Omega$ which is distributed as,

$$\hat{\Omega} \sim iW((T-k)\Omega, T-k),$$

since $\hat{\Omega}^{-1} \sim iW((T-k)\Omega^{-1}, T-k)$, and has expectation $\frac{T-k}{T-k-m-1}\Omega$. The density function of $\hat{\Omega}$ reads,

$$p(\hat{\Omega}) \propto |\hat{\Omega}|^{-\frac{1}{2}(T-k+m+1)} \exp\left[-\frac{1}{2}tr\left((T-k)\hat{\Omega}^{-1}\Omega\right)\right],$$

and $\hat{\Omega}$ is also stochastically independent of $\hat{\Phi}$ and $\hat{\Theta}$.

**2.** To construct the conditional density of the "$t-$values" given that they have reduced rank, we specify $\hat{\Theta}$ as,

$$\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_{\perp}\hat{\lambda}\hat{D}_{\perp},$$

where $\hat{\Gamma} : k \times (m-1)$, $\hat{D} : (m-1) \times m$, $\hat{D} = \begin{pmatrix} \hat{\delta} & I_{m-1} \end{pmatrix}$, $\hat{\delta} : (k-m+1) \times 1$, and $\hat{\Gamma}_{\perp}'\hat{\Gamma} \equiv 0$, $\hat{\Gamma}_{\perp}'\hat{\Gamma}_{\perp} \equiv I_{k-m+1}$, $\hat{D}_{\perp}\hat{D}' \equiv 0$, $\hat{D}_{\perp}\hat{D}_{\perp}' \equiv 1$, which results from a singular value decomposition of $\hat{\Theta}$. The density of the liml estimators results from the density of $(\hat{\Gamma}, \hat{\delta})$ given that $\hat{\lambda} = 0$,

$$\begin{aligned}
p(\hat{\Gamma}, \hat{\delta}) & \propto p(\hat{\Gamma}, \hat{\delta}, \hat{\lambda})|_{\hat{\lambda}=0} \\
& \propto p(\hat{\Theta}(\hat{\Gamma}, \hat{\delta}, \hat{\lambda}))|_{\hat{\lambda}=0} |J(\hat{\Theta}, (\hat{\Gamma}, \hat{\delta}, \hat{\lambda}))|_{\hat{\lambda}=0}| \\
& \propto \left| \begin{pmatrix} \hat{D}\hat{D}' \otimes I_k & \hat{\delta} \otimes \hat{\Gamma} \\ \hat{\delta}' \otimes \hat{\Gamma}' & \hat{\Gamma}'\hat{\Gamma} \end{pmatrix} \right|^{\frac{1}{2}} \exp\left[ -\tfrac{1}{2} tr \left( \left( \hat{\Gamma}\hat{D} - \Theta \right)' \left( \hat{\Gamma}\hat{D} - \Theta \right) \right) \right] \\
& \propto \left| \hat{\Gamma}'\hat{\Gamma} \right|^{\frac{1}{2}} \left| (I_{m-1} \otimes I_k) - (\hat{\delta}\hat{\delta}' \otimes M_{\hat{\Gamma}}) \right|^{\frac{1}{2}} \exp\left[ -\tfrac{1}{2} tr \left( \left( \hat{\Gamma}\hat{D} - \Phi \right)' \left( \hat{\Gamma}\hat{D} - \Phi \right) \right) \right] \\
& \propto \left| \hat{\Gamma}'\hat{\Gamma} \right|^{\frac{1}{2}} \left| I_{m-1} + \hat{\delta}\hat{\delta}' \right|^{\frac{1}{2}(k-m+1)} \exp\left[ -\tfrac{1}{2} tr \left( \left( \hat{\Gamma}\hat{D} - \Phi \right)' \left( \hat{\Gamma}\hat{D} - \Phi \right) \right) \right].
\end{aligned}$$

**3.** The liml estimators can be solved from $\hat{\Gamma}\hat{D}$ by using an estimator for the unknown covariance matrix $\Omega$. This estimator is also a random variable and needs to have a mean proportional to $\Omega$ and to be stochastically independent from $\hat{\Theta}$. Instead of $S$ we use $\hat{\Omega}$ as estimator/random variable to represent $\Omega$ as it leads to a more convenient expression of the density of the liml estimator. Because of the rank reduction imposed on $\hat{\Theta}$, we can exactly solve for the liml estimators from $\hat{\Gamma}\hat{D}$,

$$\hat{\Gamma}\hat{D} = (X'X)^{\frac{1}{2}}\hat{\Pi}\hat{B}\hat{\Omega}^{-\frac{1}{2}} = (X'X)^{\frac{1}{2}}\hat{\Pi}\hat{B}\hat{\Omega}_2 \begin{pmatrix} \left( \hat{B}\hat{\Omega}_2 \right)^{-1} \hat{B}\hat{\omega}_1 & I_{m-1} \end{pmatrix},$$

where $\hat{\Omega}^{-\frac{1}{2}} = \begin{pmatrix} \hat{\omega}_1 & \hat{\Omega}_2 \end{pmatrix}$ with $\hat{\omega}_1$ a $m \times 1$ vector and $\hat{\Omega}_2$ a $m \times (m-1)$ matrix such that $\hat{\delta} = (\hat{B}\hat{\Omega}_2)^{-1}\hat{B}\hat{\omega}_1$ and $\hat{\Gamma} = (X'X)^{\frac{1}{2}}\hat{\Pi}\hat{B}\hat{\Omega}_2$.

To construct the Jacobian of the transformation from $(\hat{\Gamma}, \hat{\delta})$ to $(\hat{\Pi}, \hat{\beta})$, $J((\hat{\Gamma}, \hat{\delta}), (\hat{\Pi}, \hat{\beta}))$, $\hat{\delta} = (B\hat{\Omega}_2)^{-1}B\hat{\omega}_1$, $\hat{\Gamma} = (X'X)^{\frac{1}{2}}\hat{\Pi}\hat{B}\hat{\Omega}_2$, we use the following results:

$$\begin{aligned}
\frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\beta})'} & = \left( \hat{\omega}_1 \otimes (\hat{B}\hat{\Omega}_2)^{-1} \right) \frac{\partial vec(\hat{B})}{\partial vec(\hat{\beta})'} \\
& \quad - \left( \hat{\omega}_1 \hat{B}' \otimes I_{m-1} \right) \left( (\hat{B}\hat{\Omega}_2)^{-1\prime} \otimes (\hat{B}\hat{\Omega}_2)^{-1} \right) \left( \hat{\Omega}_2 \otimes I_{m-1} \right) \frac{\partial vec(\hat{B})}{\partial vec(\hat{\beta})'} \\
& = \left( \hat{\omega}_1 e_1 \otimes (\hat{B}\hat{\Omega}_2)^{-1} \right) - \left( \hat{\omega}_1 \hat{B}'(\hat{B}\hat{\Omega}_2)^{-1\prime}\hat{\Omega}_2 e_1 \otimes (\hat{B}\hat{\Omega}_2)^{-1} \right) \\
& = \left( \hat{\omega}_1 \left( I_m - \hat{B}'(\hat{B}\hat{\Omega}_2)^{-1\prime}\hat{\Omega}_2 \right) e_1 \otimes (\hat{B}\hat{\Omega}_2)^{-1} \right), \\
\frac{\partial vec(\hat{\Gamma})}{\partial vec(\hat{\Pi})'} & = \left( \hat{\Omega}_2 \hat{B}' \otimes (X'X)^{\frac{1}{2}} \right),
\end{aligned}$$

where $e_1$ is the first $m$ dimensional unity vector. Because $\frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\Pi})'} = 0$, the jacobian then becomes

$$\begin{aligned}
\left| J((\hat{\Gamma}, \hat{\delta}), (\hat{\Pi}, \hat{\beta})) \right| & = \left| \frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\beta})'} \right| \left| \frac{\partial vec(\hat{\Gamma})}{\partial vec(\hat{\Pi})'} \right| \\
& = \left| \hat{\omega}_1 \left( I_m - \hat{B}'(\hat{B}\hat{\Omega}_2)^{-1\prime}\hat{\Omega}_2 \right) e_1 \otimes (\hat{B}\hat{\Omega}_2)^{-1} \right| \left| (\hat{\Omega}_2 \hat{B}' \otimes (X'X)^{\frac{1}{2}}) \right| \\
& = \left| \hat{B}\hat{\Omega}_2 \right|^{k-1} |X'X|^{\frac{1}{2}(m-1)} \left| \hat{\omega}_1 \left( I_m - \hat{B}'(\hat{B}\hat{\Omega}_2)^{-1\prime}\hat{\Omega}_2 \right) e_1 \right|^{(m-1)}.
\end{aligned}$$

The joint density of $(\hat{\Pi}, \hat{\beta})$ and $\hat{\Omega}$ then reads,

$$
\begin{aligned}
&p(\hat{\Pi}, \hat{\beta}, \hat{\Omega}) \\
&\propto p(\hat{\Gamma}(\hat{\Pi}, \hat{\beta}, \hat{\Omega}), \hat{\delta}(\hat{\Pi}, \hat{\beta}, \hat{\Omega})) \left| J(\hat{\Gamma}, \hat{\delta}), (\hat{\Pi}, \hat{\beta}) \right| p(\hat{\Omega}) \\
&\propto \left| \hat{\Omega}_2 B' \hat{\Pi}' X' X \hat{\Pi} B \hat{\Omega}_2 \right|^{\frac{1}{2}} \left| I_{m-1} + \left(\hat{B}\hat{\Omega}_2\right)^{-1} \hat{B}\hat{\omega}_1\hat{\omega}_1\hat{B}' \left(\hat{B}\hat{\Omega}_2\right)^{-1\prime} \right|^{\frac{1}{2}(k-m+1)} |\hat{\Omega}|^{-\frac{1}{2}(T-k+m+1)} \\
&\quad \left| J\left((\hat{\Gamma}, \hat{\delta}), \left(\hat{\Pi}, \hat{\beta}\right)\right) \right| \exp\left[ -\tfrac{1}{2} tr \left( \hat{\Omega}^{-1} \left\{ \left(\hat{\Pi}\hat{B} - \Pi B\right)' X'X \left(\hat{\Pi}\hat{B} - \Pi B\right) + (T-k)\Omega \right\} \right) \right] \\
&\propto \left| B\hat{\Omega}_2 \right|^{-(k-m)} \left| \hat{\Pi}' X' X \hat{\Pi} \right|^{\frac{1}{2}} \left| \hat{B}\hat{\Omega}_2\hat{\Omega}_2\hat{B}' + \hat{B}\hat{\omega}_1\hat{\omega}_1\hat{B}' \right|^{\frac{1}{2}(k-m+1)} |\hat{\Omega}|^{-\frac{1}{2}(T-k+m+1)} \\
&\quad \left| J\left((\hat{\Gamma}, \hat{\delta}), \left(\hat{\Pi}, \hat{\beta}\right)\right) \right| \exp\left[ -\tfrac{1}{2} tr \left( \hat{\Omega}^{-1} \left\{ \left(\hat{\Pi}\hat{B} - \Pi B\right)' X'X \left(\hat{\Pi}\hat{B} - \Pi B\right) + (T-k)\Omega \right\} \right) \right] \\
&\propto \left| B\hat{\Omega}_2 \right|^{-(k-m)} \left| \hat{\Pi}' X' X \hat{\Pi} \right|^{\frac{1}{2}} \left| B\hat{\Omega}^{-1} B' \right|^{\frac{1}{2}(k-m+1)} |\hat{\Omega}|^{-\frac{1}{2}(T-k+m+1)} \\
&\quad \left| J\left((\hat{\Gamma}, \hat{\delta}), \left(\hat{\Pi}, \hat{\beta}\right)\right) \right| \exp\left[ -\tfrac{1}{2} tr \left( \hat{\Omega}^{-1} \left\{ \left(\hat{\Pi}\hat{B} - \Pi B\right)' X'X \left(\hat{\Pi}\hat{B} - \Pi B\right) + (T-k)\Omega \right\} \right) \right],
\end{aligned}
$$

since $\hat{\Omega}_2\hat{\Omega}_2 + \hat{\omega}_1\hat{\omega}_1 = \hat{\Omega}^{-1}$ and $\hat{B}\hat{\Omega}_2$ is a square matrix.

In the following we use that $\hat{\Omega} = \begin{pmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} \\ \hat{\Omega}_{21} & \hat{\Omega}_{22} \end{pmatrix}$, $\hat{\Omega}_{11} : 1 \times 1$; $\hat{\Omega}_{21}$, $\hat{\Omega}_{12}' : (m-1) \times 1$; $\hat{\Omega}_{22} : (m-1) \times (m-1)$, $\hat{\Omega}_{11.2} = \hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$, and that

$$
\begin{aligned}
\left| \hat{\Omega}_2 \hat{B}' \right| \left| \hat{\omega}_1 \left( I_m - B' \left(\hat{\Omega}_2 B'\right)^{-1} \hat{\Omega}_2 \right) e_1 \right| &= \left| \begin{pmatrix} \hat{\omega}_1 e_1 & \hat{\omega}_1 \hat{B}' \\ \hat{\Omega}_2 e_1 & \hat{\Omega}_2 \hat{B}' \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} \hat{\omega}_1 & \hat{\Omega}_2 \end{pmatrix} \begin{pmatrix} e_1 & \hat{B} \end{pmatrix}' \right| \\
&= \left| \begin{pmatrix} \hat{\omega}_1 & \hat{\Omega}_2 \end{pmatrix} \right| = |\hat{\Omega}|^{-\frac{1}{2}},
\end{aligned}
$$

since $\hat{B} = \begin{pmatrix} \hat{\beta} & I_{m-1} \end{pmatrix}$.

The density $p(\hat{\Pi}, \hat{\beta}, \hat{\Omega})$ then becomes,

$$
\begin{aligned}
p(\hat{\Pi}, \hat{\beta}, \hat{\Omega}) \;\propto\; & \left| \hat{B}\hat{\Omega}_2 \right|^{-(k-m)} \left| \hat{\Pi}' X' X \hat{\Pi} \right|^{\frac{1}{2}} \left| \hat{B}\hat{\Omega}^{-1}\hat{B}' \right|^{\frac{1}{2}(k-m+1)} |\hat{\Omega}|^{-\frac{1}{2}(T-k+m+1)} \\
& \left| \hat{B}\hat{\Omega}_2 \right|^{k-1} |X'X|^{\frac{1}{2}(m-1)} \left| \hat{\omega}_1 \left( I_m - \hat{B}' \left(\hat{B}\hat{\Omega}_2\right)^{-1\prime} \hat{\Omega}_2 \right) e_1 \right|^{(m-1)} \\
& \exp\left[ -\tfrac{1}{2} tr \left( \hat{\Omega}^{-1} \left\{ \left(\hat{\Pi}\hat{B} - \Pi B\right)' X'X \left(\hat{\Pi}\hat{B} - \Pi B\right) + (T-k)\Omega \right\} \right) \right] \\
\propto\; & |\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)} \left| \hat{\Pi}' X' X \hat{\Pi} \right|^{\frac{1}{2}} \left| \hat{B}\hat{\Omega}^{-1}\hat{B}' \right|^{\frac{1}{2}(k-m+1)} |X'X|^{\frac{1}{2}(m-1)} \\
& \exp\left[ -\tfrac{1}{2} tr \left( \hat{\Omega}^{-1} \left\{ \left(\hat{\Pi}\hat{B} - \Pi B\right)' X'X \left(\hat{\Pi}\hat{B} - \Pi B\right) + (T-k)\Omega \right\} \right) \right].
\end{aligned}
$$

**4.** To construct the conditional density of $\hat{\beta}$ given $\hat{\Omega}$, we first decompose the trace component of the density $p(\hat{\Pi}, \hat{\beta}, \hat{\Omega})$ as,

$$tr\left(\hat{\Omega}^{-1}\left(\hat{\Pi}\hat{B}-\Pi B\right)' X'X\left(\hat{\Pi}\hat{B}-\Pi B\right)\right) =$$
$$tr\left(\hat{\Omega}^{-1}\left(\hat{B}'\hat{\Pi}'X'X\hat{\Pi}\hat{B}-\left(\hat{B}'\hat{\Pi}'X'X\Pi B+B'\Pi'X'X\hat{\Pi}\hat{B}\right)+B'\Pi'X'X\Pi B\right)\right) =$$
$$tr\left(\left(\hat{\Omega}^{-1}-\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}\right)B'\Pi'X'X\Pi B\right)+tr\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\left(\hat{\Pi}-\Psi\right)' X'X\left(\hat{\Pi}-\Psi\right)\right)$$

where $\Psi = \Pi B\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}$. To obtain the conditional density of $\hat{\beta}$ given $\hat{\Omega}$, we construct the integral of the joint density of $(\hat{\Pi}, \hat{\beta}, \hat{\Omega})$ over $\hat{\Pi}$,

$$\begin{aligned}
&p(\hat{\beta}, \hat{\Omega}) \\
&\propto |\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)}\left|\hat{\Pi}'X'X\hat{\Pi}\right|^{\frac{1}{2}}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)}|X'X|^{\frac{1}{2}(m-1)} \\
&\quad \exp\left[-\tfrac{1}{2}tr\left(\left(\hat{\Omega}^{-1}-\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}\right)B'\Pi'X'X\Pi B\right)+(T-k)\hat{\Omega}^{-1}\Omega\right] \\
&\quad \int\left|\hat{\Pi}'X'X\hat{\Pi}\right|^{\frac{1}{2}}\exp\left[-\tfrac{1}{2}tr\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\left(\hat{\Pi}-\Psi\right)'X'X\left(\hat{\Pi}-\Psi\right)\right)\right]d\hat{\Pi} \\
&\propto |\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{-\frac{1}{2}(k+1)} \\
&\quad \exp\left[-\tfrac{1}{2}tr\left(\left(\hat{\Omega}^{-1}-\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}\right)B'\Pi'X'X\Pi B+(T-k)\hat{\Omega}^{-1}\Omega\right)\right] \\
&\quad \int\left|\hat{\Upsilon}'\hat{\Upsilon}\right|^{\frac{1}{2}}\exp\left[-\tfrac{1}{2}tr\left(\left(\hat{\Upsilon}-\Upsilon\right)'\left(\hat{\Upsilon}-\Upsilon\right)\right)\right]d\hat{\Upsilon} \\
&\propto |\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{-\frac{1}{2}m}\exp\left[-\tfrac{1}{2}tr\left(\left(\hat{\Omega}^{-1}-\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}\right)B'\Pi'X'X\Pi B\right)\right] \\
&\quad \exp\left[-\tfrac{1}{2}tr\left((T-k)\hat{\Omega}^{-1}\Omega\right)\right]\int\left\{\left|\hat{\Upsilon}'\hat{\Upsilon}\right|^{\frac{1}{2}}\exp\left[-\tfrac{1}{2}tr\left(\left(\hat{\Upsilon}-\Upsilon\right)'\left(\hat{\Upsilon}-\Upsilon\right)\right)\right]\right\}d\hat{\Upsilon},
\end{aligned}$$

where $\hat{\Upsilon} = (X'X)^{\frac{1}{2}}\hat{\Pi}\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{\frac{1}{2}}$, $\Upsilon = (X'X)^{\frac{1}{2}}\Psi\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{\frac{1}{2}} = (X'X)^{\frac{1}{2}}\Pi B\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-\frac{1}{2}}$, since $|J(\hat{\Pi}, \hat{\Upsilon})| = |X'X|^{-\frac{1}{2}(m-1)}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{-\frac{1}{2}k}$ and $\left|\hat{\Pi}'X'X\hat{\Pi}\right|^{\frac{1}{2}} = \left|\hat{\Upsilon}'\hat{\Upsilon}\right|^{\frac{1}{2}}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{-\frac{1}{2}}$. The integral in the above expression is a non-central moment of a matrix normal random matrix. We construct this expression for the case that $\hat{\Upsilon}$ is a vector which implies that $m = 2$.

When $\hat{\Upsilon} \sim n(\Upsilon, I_k)$, it holds that $w = \hat{\Upsilon}'\hat{\Upsilon} \sim \chi^2(k, \mu)$, where $\mu = \Upsilon'\Upsilon$ is the non-centrality parameter of the non-central $\chi^2$ distribution and $k$ the degrees of freedom parameter. The density function of a non-central $\chi^2$ reads, see Johnson and Kotz (1970) and Muirhead (1982),

$$p_{\chi^2(k,\mu)}(w) = \sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\mu\right)^j}{j!}\exp\left[-\frac{1}{2}\mu\right]\right)p_{\chi^2(k+2j)}(w),$$

where $p_{\chi^2(k+2j)}(w)$ is the density function of a standard $\chi^2$ random variable with $k+2j$ degrees of freedom. Note that the weights, which correspond with a Poisson density, sum to one. The expectation of $w^{\frac{1}{2}}$ when $w \sim \chi^2(k+2j)$ reads,

$$E_{\chi^2(k+2j)}\left[w^{\frac{1}{2}}\right] = 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}.$$

The expectation of $w^{\frac{1}{2}}$ over the non-central $\chi^2$ distribution therefore reads,

$$\begin{aligned} E_{\chi^2(k,\mu)}\left[w^{\frac{1}{2}}\right] &= \sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\mu\right)^j}{j!}\exp\left[-\tfrac{1}{2}\mu\right]\right)E_{\chi^2(k+2j)}\left[w^{\frac{1}{2}}\right] \\ &= \sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\mu\right)^j}{j!}\exp\left[-\tfrac{1}{2}\mu\right]\right)2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}. \end{aligned}$$

The integral needed to obtain the conditional density of $\hat{\beta}$ given $\hat{\Omega}$ thus reads,

$$\begin{aligned} &\int|\hat{\Upsilon}'\hat{\Upsilon}|^{\frac{1}{2}}\exp\left[-\tfrac{1}{2}tr\left(\left(\hat{\Upsilon}-\Upsilon\right)'\left(\hat{\Upsilon}-\Upsilon\right)\right)\right]d\hat{\Upsilon} \\ &\propto E_{\chi^2(k,\mu)}\left[w^{\frac{1}{2}}\right] \\ &\propto \sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\Upsilon'\Upsilon\right)^j}{j!}\exp\left[-\tfrac{1}{2}\Upsilon'\Upsilon\right]2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right) \\ &\propto \sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}B'\Pi'X'X\Pi B\right)^j}{j!}\right. \\ &\qquad \left.\exp\left[-\tfrac{1}{2}\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}B'\Pi'X'X\Pi B\right]2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right) \end{aligned}$$

such that the joint density of $\hat{\beta}$ and $\hat{\Omega}$ becomes,

$$\begin{aligned} &p(\hat{\beta},\hat{\Omega}) \\ &\propto \left|\hat{\Omega}\right|^{-\frac{1}{2}(T-k+2m)}\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{-\frac{1}{2}m}\exp\left[-\tfrac{1}{2}\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}B'\Pi'X'X\Pi B\right] \\ &\quad \exp\left[-\tfrac{1}{2}tr\left((T-k)\hat{\Omega}^{-1}\Omega\right)\right]\left[\sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}B'\Pi'X'X\Pi B\right)^j}{j!}\right.\right. \\ &\quad \left.\left.2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right)\right]\exp\left[-\tfrac{1}{2}tr\left(\left(\hat{\Omega}^{-1}-\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}\right)B'\Pi'X'X\Pi B\right)\right] \\ &\propto \left|\hat{\Omega}\right|^{-\frac{1}{2}(T-k+2m)}\exp\left[-\tfrac{1}{2}tr\left(\hat{\Omega}^{-1}\left\{(T-k)\Omega+B'\Pi'X'X\Pi B\right\}\right)\right]\left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right|^{-\frac{1}{2}m} \\ &\quad \left[\sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\hat{\Omega}^{-1}\hat{B}'\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Omega}^{-1}B'\Pi'X'X\Pi B\right)^j}{j!}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right)\right] \\ &\propto \left|\hat{\Omega}\right|^{-\frac{1}{2}(T-k+2m)}\exp\left[-\tfrac{1}{2}tr\left(\hat{\Omega}^{-1}\left\{(T-k)\Omega+B'\Pi'X'X\Pi B\right\}\right)\right]\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\right. \\ &\quad \left.\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|^{-\frac{1}{2}m}\left[\sum_{j=0}^{\infty}\left(\left(\frac{\left(B\hat{\Omega}^{-1}\hat{B}'\right)^2\Pi'X'X\Pi}{2\left(\hat{B}\hat{\Omega}^{-1}\hat{B}'\right)}\right)^j\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right] \\ &\propto \left|\hat{\Omega}\right|^{-\frac{1}{2}(T-k+2m)}\exp\left[-\tfrac{1}{2}tr\left(\hat{\Omega}^{-1}\left\{(T-k)\Omega+B'\Pi'X'X\Pi B\right\}\right)\right] \\ &\quad \left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ &\quad \sum_{j=0}^{\infty}\left[\left(\frac{\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\beta'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|^2\Pi'X'X\Pi}{2\left|\hat{\Omega}_{22}^{-1}+\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Omega}_{11.2}^{-1}\left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}'\right)\right|}\right)^j\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right] \end{aligned}$$

since

$$\begin{aligned} \left|\hat{B}\hat{\Omega}^{-1}\hat{B}'\right| &= |\hat{\Omega}_{22}^{-1}+(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}')'\hat{\Omega}_{11.2}^{-1}(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}')|, \\ \text{and } \left|B\hat{\Omega}^{-1}B'\right| &= \left|\hat{\Omega}_{22}^{-1}+(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\beta')'\hat{\Omega}_{11.2}^{-1}(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}-\hat{\beta}')\right|. \end{aligned}$$

and

$$p(\hat{\beta}, \hat{\Omega}) \propto p(\hat{\beta}|\hat{\Omega})q(\hat{\Omega})$$

such that

$$\begin{array}{rl} p(\hat{\beta}|\hat{\Omega}) & \propto \left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ & \sum_{j=0}^{\infty} \left[ \left( \frac{\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^2 \Pi' X' X \Pi}{2\left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|} \right)^j \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \right] \\ q(\hat{\Omega}) & \propto |\hat{\Omega}|^{-\frac{1}{2}(T-k+2m)} \exp\left[-\frac{1}{2} tr\left(\hat{\Omega}^{-1}\left((T-k)\Omega + B'\Pi' X' X \Pi B\right)\right)\right]. \end{array}$$

The function $q(\hat{\Omega})$ is the density that belongs to an inverted-Wishart distributed random matrix $\Delta$, $\Delta \sim iW((T-k)\Omega + B'\Pi'X'X\Pi B, T-k+m-2)$. This inverted-Wishart random matrix has a mean equal to $\frac{1}{T-k+m-1}((T-k)\Omega + B'\Pi'X'X\Pi B) \approx \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B$ and its variance is proportional to $\frac{1}{T}$, see Muirhead (1982). The inverted-Wishart density $q(\hat{\Omega})$, which is not the marginal density of $\hat{\Omega}$, is therefore centered close around its mean for reasonably large values of $T$ $(T > 25)$. Hence, already for moderate values of $T$, we can consider $q(\hat{\Omega})$ as a point mass at $\Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B$. The marginal density of $\hat{\beta}$ then results as

$$\begin{array}{rl} p(\hat{\beta}) & \propto \int p(\hat{\beta}, \hat{\Omega}) d\hat{\Omega} \\ & \propto \int p(\hat{\beta}, \hat{\Omega}) I(\hat{\Omega}, \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B) d\hat{\Omega} \\ & \propto p(\hat{\beta}, \hat{\Omega})|_{\hat{\Omega}=\Omega+B'\Pi'\left(\frac{X'X}{T}\right)\Pi B} \\ & = p(\hat{\beta}|\hat{\Omega} = \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B). \end{array}$$

where $I(\hat{\Omega}, \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B) = 1$ when $\hat{\Omega} = \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B$ and is equal to zero elsewhere. This is the reason why we use $\hat{\Omega}$ instead of $S$ since we can not decompose the joint density of $(\hat{\beta}, S)$ as the product of the conditional density of $\hat{\beta}$ given $S$ and a standard function of $S$ that has convenient convergence properties.

When $\Pi = 0$, the conditional density of $\hat{\beta}$ simplifies to,

$$p(\hat{\beta}|\hat{\Omega}) \propto \left|\hat{\Omega}_{22}^{-1} + \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Omega}_{11.2}^{-1} \left(\hat{\Omega}_{12}\hat{\Omega}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m},$$

which is a Cauchy density. Another simplification occurs when $\beta = \hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$ as in that case the term $B\hat{\Omega}^{-1}B'$ is equal to $\hat{\Omega}_{22}^{-1}$ and $p(\hat{\beta}|\hat{\Omega})$ is a symmetric density then.

We note that the density $p(\hat{\beta}|\hat{\Omega})$ has a simpler functional form than the density derived in Mariano and Sawa (1971), which involves a triplicate infinite series whereas $p(\hat{\beta}|\hat{\Omega})$ constructed above only involves a single infinite series. The density constructed by Mariano and Sawa is the marginal density though while the density constructed above is the conditional density given $\hat{\Omega}$. As the joint density $p(\hat{\beta}, \hat{\Omega})$ quickly converges, when $T$ increases, to the density $p(\hat{\beta}, \hat{\Omega}) I(\hat{\Omega}, \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B) \propto p(\hat{\beta}|\hat{\Omega} = \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B) I(\hat{\Omega}, \Omega + B'\Pi'\left(\frac{X'X}{T}\right)\Pi B)$, we can use the latter conditional density as the marginal density of $\hat{\beta}$.

# References


[1] Anderson, T.W. Some recent developments on the distributions of single-equation estimators. In Werner Hildebrand, editor, *Advances in Econometrics: Invited Papers for the*

*Fourth World Congress of the Econometric Society at Aix-En-Provence, September 1980*, pages 109–122. Cambridge University Press, 1982.

[2] Anderson, T.W. and H. Rubin. Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, **21**:570–582, (1949).

[3] Anderson, T.W. and T. Sawa. Evaluation of the Distribution Function of the Two Stage Least Squares Estimate. *Econometrica*, **47**:163–182, 1979.

[4] Anderson, T.W., N. Kunitomo, and T. Sawa. Comparison of the Densities of the TSLS and LIMLK estimators for Simultaneous Equations. In F.G. Adams and B.G. Hickman, editor, *Global Econometrics: Essays in Honor of Lawrence R. Klein*, pages 102–124. MIT Press, 1983.

[5] Angrist, J.D. and A.B. Krueger. Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, **106**:979–1014, 1991.

[6] Barndorf-Nielsen, O. Conditionality Resolutions. *Biometrika*, **67**:293–310, 1980.

[7] Barndorf-Nielsen, O. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**:343–365, 1983.

[8] Billingsley, P. *Probability and Measure*. Wiley (New York), (1986).

[9] Cox, D.R, and N. Reid. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **49**:1–39, 1987.

[10] Dufour, J.-M. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, **65**:1365–388, 1997.

[11] Dufour, J.-M. and L. Khalaf. Simulation Based Finite and Large Sample Inference Methods in Simultaneous Equations. Working paper, CRDE, Universite de Montreal, 1997.

[12] Engle, R.F. Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, Volume 2*. Elsevier Science (Amsterdam), 1984.

[13] Engle, R.F., Hendry D.F. and Richard J.-F. Exogeneity. *Econometrica*, **51**:277–304, 1983.

[14] Geweke, J. Bayesian Inference in Econometric Models using Monte Carlo Integration. *Econometrica*, **57**:1317–1339, (1989).

[15] Golub, G.H. and C.F. van Loan. *Matrix Computations*. The John Hopkins University Press (Baltimore), (1989).

[16] Hastings, W.K. Monte Carlo Sampling using Markov Chains and their Applications. *Biometrika*, **57**:97–10, (1970).

[17] Hausman, J.A. Specification and Estimation of Simultaneous Equations Systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).

[18] Hillier, G. and M. Armstrong. The Density of the Maximum Likelihood Estimator. *Econometrica*, **67**:1459–1470, 1999.

[19] Hood, W.C. and T.C. Koopmans. *Studies in Econometric Method*, volume 14 of *Cowles Foundation Monograph*. Wiley (New York), 1953.

[20] Johnson, N.L. and S. Kotz. *Continuous Univariate Distributions*, volume 2. John Wiley (New York), 1970.

[21] Kitamura, Y. Specification Tests with Instrumental Variables and Rank Deficiency. Working paper, Department of Economics, University of Minnesota, 1994.

[22] Kleibergen, F. Conditional Densities in Statistical Modelling. Working paper, University of Amsterdam, 2000.

[23] Kleibergen F. and E. Zivot. Bayesian and Classical Approaches to Instrumental Variable Regression. Econometric Institute Report 9835/A, 1998.

[24] Kleibergen, F. and H.K. van Dijk. Bayesian Simultaneous Equation Analysis using Reduced Rank Structures. *Econometric Theory*, **14**:701–743, 1998.

[25] Kleibergen, F. and R. Paap. Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration. Econometric Institute Report 9821/A, 1998.

[26] Kloek, T. and H.K. van Dijk. Bayesian Estimates of Equation System Parameters : An Application of Integration by Monte-Carlo. *Econometrica*, **44**:345–351, (1978).

[27] Magnus, J.R. and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley (Chichester), (1988).

[28] Mariano, R.S. and T. Sawa. The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables. *Journal of the American Statistical Association*, **67**:159–163, 1972.

[29] Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of State Calculations by fast Computing Machines. *Journal of Chemical Physics*, **21**:1087–1092, (1953).

[30] Muirhead, R.J. *Aspects of Multivariate Statistical Theory.* John Wiley (New York), 1982.

[31] Nelson, C.R. and R. Startz. Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator. *Econometrica*, **58**:967–976, 1990.

[32] Phillips, P.C.B. Exact Small Sample Theory in the Simultaneous Equations Model. In Z. Griliches and M.D. Intrilligator, editors, *Handbook of Econometrics, Vol.1.* North-Holland Publishing Co., Amsterdam, 1983.

[33] Phillips, P.C.B. Partially Identified Econometric Models. *Econometric Theory*, **5**:181–240, (1989).

[34] Shea, J. Instrument relevance in multivariate linear models: A simple measure. *The Review of Economics and Statistics*, **79**:348–352, 1997.

[35] Staiger, D. and J.H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, **65**:557–586, 1997.

[36] Tjur, T. *Probability based on Radon Measures.* Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, New York, 1980.

[37] Wang, J. and E. Zivot. Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments. *Econometrica*, **66**:1389–1405, 1998.

[38] Zivot, E., R. Startz, and C. R. Nelson. Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review*, **39**:1119–1144, 1998.