# An Alternative Approach for Constructing Small Sample and Limiting Distributions of Maximum Likelihood Estimators

Frank Kleibergen*

Econometric Institute Report 9844/A

**Abstract**

We construct limiting and small sample distributions of maximum likelihood estimators (mle) from the property that they satisfy the first order condition (foc). The foc relates the mle of the analyzed model to the mle of an encompassing model and shows that the mle of the analyzed model is a realization from the limiting or small sample distribution of the mle of the encompassing model given that the foc holds. We can thus use the unique conditional (limiting or small sample) density of the mle of the encompassing model given that the foc holds to construct the limiting or small sample density/distribution of the mle of the analyzed model. To proof the validity of this approach and thus of the concept of an unique conditional density, we use it to construct the small sample and limiting distribution of the limited information mle and show that they are identical to resp. the sampling density and the expression discussed elsewhere in the literature. We analyze the small sample density further and relate it to existing expressions and show its limiting behavior in case of weak and strong instruments.

## 1 Introduction

The statistical properties of the maximum likelihood estimator of the parameters of a specific model are typically obtained from its closed form analytical expression. The statistical properties then result from the different (random) elements of the closed form expression of this estimator. The limiting and small sample distribution of the maximum likelihood estimator are typically constructed in this way, see *e.g.* Phillips (1983). It is, however, also possible to construct the small sample and limiting distribution through an implicit approach that is based on a property of the maximum likelihood estimator instead of the previously referred to explicit approach which uses the analytical expression of the maximum likelihood estimator. That property is the first order condition for a maximum of the likelihood which is satisfied by the maximum likelihood estimator. The first order condition can be specified such that it relates the maximum likelihood estimator of the parameters of the analyzed model to the maximum likelihood estimator of the parameters of an encompassing model. When the limiting or/and small sample distribution of the maximum likelihood estimator of the parameters

*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam. Email: kleibergen@few.eur.nl. The author thanks the Netherlands Foundation for Scientific Research (N.W.O.) for their financial support.

of the encompassing model is known, this maximum likelihood estimator can be considered as a realization from that distribution. The maximum likelihood estimator of the parameters of the analyzed model is related to this estimator through the first order condition. It can therefore be considered as a realization of the random variable characterizing the maximum likelihood estimator of the parameters of the encompassing model under the condition that the first order condition holds. It is thus a drawing from the conditional density of the maximum likelihood estimator of the parameters of the encompassing model given that the first order condition holds. When this conditional density is unique, it can be used to construct the limiting and/or small sample distribution/density of the maximum likelihood estimator of the parameters of the analyzed model.

In Kleibergen (1998), sufficient conditions for the existence of an unique conditional density are given. These sufficient conditions *a.o.* require that the event where we want to condition on can be represented in an unambigious way. The resulting conditional densities avoid the Borel-Kolmogorov paradox, which states that the density of a probability distribution is not defined on sets of measure zero, see *e.g.* Kolmogorov (1950), Billingsley (1986), and Drèze and Richard (1983), and can therefore be used to construct the limiting and small sample distributions of maximum likelihood estimators.

To show the validity of the concept of an unique conditional density and the small sample and limiting distributions resulting from it, we use the developed methodology in a nontrivial example, *i.e.* to construct the limiting and small sample distribution of the limited information maximum likelihood estimator. The limited information maximum likelihood estimator is used in instrumental variable regression models which are nested in a standard linear model. Using the unique conditional density concept we construct the small sample and limiting distribution of the limited information maximum likelihood estimator from the specified small sample and limiting distribution of the maximum likelihood estimator of the parameters of the encompassing linear model, *i.e.* the least squares estimator. This example is nontrivial in the sense that the restriction leading to the instrumental variable regression model when imposed on the encompassing linear model is more complicated than a linear restriction in which case the results follow straightforwardly. The resulting small sample and limiting distributions are identical to resp. the sampling density and the analytical expression discussed elsewhere in the literature. This proofs the validity of the concept of an unique conditional density and the procedure for constructing small sample and limiting distributions that results from it.

The paper is organized as follows. In section 2, we show how small sample and limiting distributions of maximum likelihood estimators result from the small sample and limiting distributions of the maximum likelihood estimator of the parameters of an encompassing model. They result from the property of the maximum likelihood estimator that it satisfies the first order condition and the concept of an unique conditional density. We therefore also give the sufficient conditions for the existence of an unique conditional density, see Kleibergen (1998). In section 3, we show the validity of the approach by using it to construct the small sample density of the limited information maximum likelihood estimator. We analyze the properties of the resulting small sample density, like its relationship to already existing expressions, see *e.g.* Mariano and Sawa (1972) and Phillips (1983), and its limiting behavior when sample size increases in case of weak and strong instruments, see *e.g.* Staiger and Stock (1997) and Nelson and Startz (1990). We also show its equivalence with the sampling density of the limited information maximum likelihood estimator and thus of the validity of construction the small sample density using the conditional density approach. In section 4, we construct the limiting distribution of the limited information maximum likelihood estimator using the conditional density approach and show that it is equivalent to the expression stated elsewhere in the literature, see *e.g.* Hausman (1983). In section 5 we draw some conclusions *a.o.* that the unique conditional density approach offers a novel procedure for constructing small sample and limiting distributions of maximum likelihood estimators that can sometimes be more or less convenient than the traditional approach that uses the explicit expression of the maximum likelihood estimator.

# 2 Maximum Likelihood Inference using Unique Conditional Densities

The maximum likelihood (ml) estimator of the parameters of a specific nested model can be considered as a realization of a random variable. Furthermore, ml estimators satisfy the first order condition (foc) for a maximum of the likelihood. This foc is often such that it relates the ml estimator of the parameters of a nested model to the ml estimator of the parameters of an encompassing model. Since both these estimators are realizations of random variables, the random variable characterizing the ml estimator of the parameters of the nested model can be considered as a realization of the random variable characterizing the ml estimator of the parameters of the encompassing model given that the foc holds. This property can be used to construct the small sample and limiting distributions of the ml estimator of the parameters of the nested model from the small sample and limiting distribution of the ml estimator of the parameters of the encompassing model. It involves the use of conditional densities that show the behavior of the random variable characterizing the ml estimator of the parameters of the encompassing model given that the foc holds. Because of the Borel-Kolmogorov paradox, see *e.g.* Kolmogorov (1950) and Billingsley (1986), these conditional densities need to be defined such that they unambiguously reflect the imposed restriction, here the foc. We therefore first briefly discuss the uniqueness of these conditional densities before we proceed with applying them to ml estimators. For a more elaborate discussion of unique conditional densities we refer to Kleibergen (1998).

## 2.1 Unique Conditional Densities

The Borel-Kolmogorov paradox, see *e.g.* Kolmogorov (1950), Billingsley (1986), Drèze and Richard (1983), Poirier (1995) and Wolpert (1995), states that the density of a probability distribution is not defined on sets of measure zero. This would imply that we cannot specify the conditional density of a random variable given a certain event in an unique way. On the other hand we often specify joint densities of random variables as products of conditional and marginal densities. For example in case of a bivariate normal distribution it is well known that the conditional density of one of the random variables given the other is normal. This, however, contradicts a literal interpretation of the Borel-Kolmogorov paradox as it indicates that conditional densities are in some cases uniquely defined. In Kleibergen (1998) sufficient conditions for the existence of an unique conditional density are therefore defined.

*Sufficient conditions for the existence of an unique conditional density for the continuous random variable* $x : k \times 1$; *whose space, on which it is defined, is unrestricted and has density* $p(x)$; *given the restriction* $g(x) = 0$; $g(x) : m \times 1$, $m < k$, *and* $g(x)$ *is continuous differentiable and is defined on the whole space of* $x$; *are:*

**Condition 1.** *An invertible relationship between* $x$ *and* $(y, z)$ *exists;* $y = h(x) : m \times 1$, $z = f(x) : (k - m) \times 1$, *where* $h$ *is continuous and continuous differentiable for all* *values of $x$ and $f$ is continuous and continuous differentiable except (maybe) for some lower dimensional subspaces of the space of $x$; which is such that given that $z$ does not result from one of the lower dimensional subspaces where $f(x)$ is not continuous, we can uniquely solve $x$ from $(y, z)$ for all values of $y$ including on all the lower dimensional subspaces of the space on which $y$ is defined.*

**Condition 2.** *The restriction $g(x) = 0$ is equivalent with $y = 0$ and imposes no restrictions on $z$.*

When the sufficient conditions are satisfied, the unique expression for the conditional density function of the random variable $x$ given that $g(x) = 0$ reads, see Kleibergen (1998),

$$\begin{aligned} p_r(x) &\propto p(x)|_{g(x)=0} \\ &\propto p(x(y,z))|_{y=0}|J(x,(y,z))|_{y=0}|, \end{aligned} \tag{1}$$

where $J(x,(y,z))$ is the jacobian of the transformation from $x$ to $(y, z)$ and $|_{y=0}$ stands for evaluated in $y = 0$. In classical statistical analysis these unique conditional densities can *a.o.* be used to construct small sample and limiting distributions of ml estimators. The resulting small sample or limiting distributions can then be compared with already known analytical expressions or densities constructed through Monte-Carlo simulation to show the validity of the concept of an unique conditional density. Note that the sufficient conditions for an unique conditional density are stricter than the conditions for transforming random variables, see Kleibergen (1998).

## 2.2 Restrictions imposed by the First Order Condition

Ml estimators satisfy the foc for a maximum of the likelihood. Given the limiting or small sample distribution of a ml estimator of the parameters of an encompassing model, the foc can be seen as a restriction imposed on the unrestricted random variable of which the ml estimator of the parameters of the encompassing model is a realization. When these restrictions satisfy the sufficient conditions for the existence of an unique conditional density, these unique conditional densities can be used to construct the limiting or small sample distribution of the ml estimator of the parameters of the nested model from the limiting or small sample distribution of the ml estimator of the parameters of the encompassing model.

Consider for example the model,

$$y = Xf(\varphi) + \varepsilon, \tag{2}$$

where $y : T \times 1$; $X : T \times k$; $\varepsilon : T \times 1$, $\varepsilon \sim n(0, \sigma^2 I_T)$, $\varphi : m \times 1$; $f(\varphi) : k \times 1$, continuous and continuous differentiable except (maybe) for some lower dimensional subspaces of the space on which $\varphi$ is defined. Examples of model (2) are the simultaneous equation model, the autoregressive moving average model and numerous others.

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

$$\begin{aligned}
\frac{1}{\sigma^2}\left(\frac{\partial f}{\partial \varphi'}|_{\hat\varphi}\right)' X'\left(y - Xf(\hat\varphi)\right) &= 0 \Leftrightarrow \\
\frac{1}{\sigma^2}\left(\frac{\partial f}{\partial \varphi'}|_{\hat\varphi}\right)' X'X\left((X'X)^{-1}X'y - f(\hat\varphi)\right) &= 0 \Leftrightarrow \\
\frac{1}{\sigma^2}\left(\frac{\partial f}{\partial \varphi'}|_{\hat\varphi}\right)' X'X\left(\hat\Phi - f(\hat\varphi)\right) &= 0 \Leftrightarrow \\
\left((X'X)^{\frac{1}{2}}\frac{\partial f}{\partial \varphi'}|_{\hat\varphi}\sigma^{-1}\right)'\left((X'X)^{\frac{1}{2}}\hat\Phi\sigma^{-1} - (X'X)^{\frac{1}{2}}f(\hat\varphi)\sigma^{-1}\right) &= 0 \Leftrightarrow \\
\left(\frac{\partial r}{\partial \varphi}|_{\hat\varphi}\right)'\left(\hat\Theta - r(\hat\psi(\hat\varphi))\right) &= 0,
\end{aligned} \tag{3}$$

where $\hat\Phi = (X'X)^{-1}X'y$; $\hat\Theta = (X'X)^{\frac{1}{2}}\hat\Phi\sigma^{-1}$, $\psi : m \times 1$; $r(\psi) : k \times 1$, $r(\psi) = (X'X)^{\frac{1}{2}} f(\varphi)\sigma^{-1}$, $r(\psi)$ is a continuous differentiable function and an invertible relationship between $\psi$ and $\varphi$ exists such that $\frac{\partial\varphi}{\partial\psi'}$ is invertible for all $\varphi$, $\psi$; $\hat\psi$ and $\hat\varphi$ stands for the ml estimator of the specific parameter.

The foc (3) relates the ml estimator of $\varphi$, $\hat\varphi$, which can be uniquely solved for from $\hat\psi$, to the "$t$-values" $\hat\Theta$ of the linear model,

$$y = X\Phi + \varepsilon. \tag{4}$$

The "$t$-values" $\hat\Theta$ can be considered as a realization of a random variable. Since the foc (3) holds for all realizations of this random variable, we can consider $\hat\psi$, and thus also $\hat\varphi$, as a realization of this random variable given that the restriction imposed by the foc holds. When this restriction satisfies the sufficient conditions for the existence of an unique conditional density and the density function of $\hat\Theta$ is known, we can construct the density of $\hat\psi$ as the unique conditional density of $\hat\Theta$ given that the foc holds. A similar reasoning can be pursued to obtain the limiting distribution of $\hat\varphi$ from the limiting distribution of $\hat\Theta$, when this limiting distribution has a known density function.

### 2.2.1 Small Sample Distributions of Maximum Likelihood Estimators

When the density function of $\hat\Theta$ in (3) is known, say $\hat\Theta \sim n(\Theta_0, I_k)$, $\Theta_0 = (X'X)^{\frac{1}{2}} f(\varphi_0)\sigma^{-1}$, we can construct the small sample distribution of the ml estimator $\hat\varphi$. When $S = X'X$ and has a fixed full rank value, such that $X$ is nonstochastic, and $\frac{\partial f}{\partial\varphi'}$ invertible $\forall\varphi$, the foc (3) essentially imposes the restriction,

$$\hat\Theta - r(\hat\psi) \equiv 0, \tag{5}$$

on the random variable $\hat\Theta$, which has an identity covariance matrix, since (3) holds for all values of the random variable $\hat\Theta$. When (5) satisfies the sufficient conditions for the existence of an unique conditional density, the pdf of the ml estimator $\hat\varphi$ results from the conditional probability density function (pdf) of $\hat\Theta$ given that (5) holds,

$$\begin{aligned}
p_r(\hat\varphi) &\propto p_r(\hat\psi(\hat\varphi))|J(\hat\psi, \hat\varphi)| \\
\text{and } p_r(\hat\psi) &\propto p(\hat\Theta)|_{\hat\Theta = r(\hat\psi)} \\
&\propto p(\hat\Theta(\hat\psi, \hat\lambda))|_{\hat\lambda=0}|J(\hat\Theta, (\hat\psi, \hat\lambda)|_{\hat\lambda=0}|,
\end{aligned} \tag{6}$$

where $\hat{\lambda} : (k-m) \times 1$ represents the restriction (5) (because of condition 2), $r$ stands for restricted to indicate that we donot use the marginal densities directly resulting from $\hat{\Theta}$, and $J(\hat{\psi}, \hat{\varphi})$ is the jacobian of the transformation from $\hat{\psi}$ to $\hat{\varphi}$. In section 3, we apply the above arguments to construct the small sample density of the limited information maximum likelihood (liml) estimator. We show there that the analytical expression of the small sample density is equal to the density obtained by simulation which thus shows the validity of the concept of an unique conditional density and the procedure for constructing small sample distributions that results from it.

The small sample density constructed by using (6) results from an implicit argument as it is obtained from the property of the ml estimator that it satisfies the foc (3). The small sample distributions for maximum likelihood estimators discussed in the literature, see *e.g.* Phillips (1983) and Mariano and Sawa (1972), are all constructed using an explicit expression of the maximum likelihood estimator, *i.e.* the small sample distribution is constructed using the exact closed form analytical expression of the maximum likelihood estimator and the densities of the different elements of it, and do not result from a restriction imposed on a specific random variable. The procedure for constructing the small sample distribution using the unique conditional densities is appealing as it can also be applied in cases where we cannot construct closed form analytical expressions of the maximum likelihood estimator.

We note that for an observed sample, $\hat{\Theta}$ is just a statistic such that the foc (3) does not imply the restriction (5), which is also not possible, and we can just solve for $\hat{\varphi}$ from (3). However, when $\hat{\Theta}$ is a random variable defined on the $R^k$ and (3) holds for all of its realizations, the foc (3) does impose the restriction (5) on the random variable $\hat{\Theta}$.

### 2.2.2 Limiting Distribution of Maximum Likelihood Estimators

The foc (3) can also be used to obtain the limiting behavior of the ml estimator $\hat{\varphi}$ when the limiting behavior of $\hat{\Theta}$ around its true value $\Theta_0$ is known, say $\sqrt{T}(\hat{\Theta} - \Theta_0) \Rightarrow n(0, I_k)$, $\hat{\Theta} = S^{\frac{1}{2}}\hat{\Phi}\sigma^{-1}$, $\Theta_0 = S^{\frac{1}{2}} f(\varphi_0)\sigma^{-1}$, $S = p\lim_{T\to\infty}\left(\frac{X'X}{T}\right)$. This limiting behavior can be constructed using the restriction imposed by the foc on $\hat{\Theta}$,

$$\left((X'X)^{\frac{1}{2}} \frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\sigma^{-1}\right)' \left((X'X)^{\frac{1}{2}}\,\hat{\Phi}\sigma^{-1} - (X'X)^{\frac{1}{2}} f(\hat{\varphi})\sigma^{-1}\right) \equiv 0 \Leftrightarrow \tag{7}$$

$$\left((X'X)^{\frac{1}{2}} \frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\sigma^{-1}\right)' \left((X'X)^{\frac{1}{2}} \left(\hat{\Phi} - f(\varphi_0)\right)\sigma^{-1} - (X'X)^{\frac{1}{2}} \left(f(\hat{\varphi}) - f(\varphi_0)\right)\sigma^{-1}\right) \equiv 0.$$

The limiting behavior of $\hat{\Theta}$ is such that $\sqrt{T}\left(\hat{\Theta} - \Theta_0\right)$ converges to a random variable $x$ with density function,

$$p(x) = (2\pi)^{-\frac{1}{2}k} \exp\left[-\frac{1}{2}x'x\right]. \tag{8}$$

So, since $\sqrt{T}\left(\hat{\Theta}-\Theta_0\right)$ converges to a random variable and the condition (7) holds for every value $\hat{\varphi}$, it also implies that,

$$\begin{aligned}
\lim_{T\to\infty}\left[\sqrt{T}\left(S^{\frac{1}{2}}\frac{\partial f}{\partial\varphi'}\sigma^{-1}\right)'\left(\sqrt{T}S^{\frac{1}{2}}\left(\hat{\Phi}-f(\varphi_0)\right)\sigma^{-1}-\sqrt{T}S^{\frac{1}{2}}\left(f(\hat{\varphi})-f(\varphi_0)\right)\sigma^{-1}\right)\right] &\equiv 0 \Leftrightarrow \\
\lim_{T\to\infty}\left[\left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}}\right)'\left(\sqrt{T}\left(\hat{\Theta}-\Theta_0\right)-\sqrt{T}\left(r(\hat{\psi})-r(\psi_0)\right)\right)\right] &\equiv 0 \Leftrightarrow \\
\lim_{T\to\infty}\left[\left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}}\right)'\left(x-\sqrt{T}\left(r(\hat{\psi})-r(\psi_0)\right)\right)\right] &\equiv 0. \quad (9)
\end{aligned}$$

as $S = p\lim_{T\to\infty}\left(\frac{X'X}{T}\right)$ and $r(\hat{\psi}) = S^{\frac{1}{2}}f(\hat{\varphi})\sigma^{-1}$ is an invertible continuous differentiable function. This results from the continuous mapping theorem, see Billingsley (1986). (9) holds for every value of $x$ which is a random variable defined on $R^k$. This is therefore only possible when

$$x-\sqrt{T}\left(r(\hat{\psi})-r(\psi_0)\right)\equiv 0. \qquad (10)$$

So, when (9) satisfies the sufficient conditions for the existence of an unique conditional density, we can construct the limiting distribution of $\hat{\psi}$, and thus also of $\hat{\varphi}$, from the density of $x$ given that (10) holds,

$$\sqrt{T}\left(r(\hat{\psi})-r(\psi_0)\right)\Rightarrow x|_{g(x)=0}, \qquad (11)$$

where $g(x)=0$ is equivalent with (10). In section 4, we give an example where we construct the limiting distibribution of the liml estimator using the above arguments and show that the resulting limiting distribution is equivalent to the one stated in the literature, see *e.g.* Hausman (1983). This further shows the validity of the above arguments and the concept of an unique conditional density.

We note that we obtain the limiting distribution of the ml estimator of the parameters of the nested model by explicitly using condition (10) which results from the foc. This results since (9) implied (10) because (9) holds for the random variable $x$. When we analyze an observed sample, we have only one realization such that (9) does not imply equality of $f(\hat{\varphi})$ and $\hat{\Phi}$ for the observed sample, which is also not possible in general. Identical to the previous subsection where the foc (3) implied (5), the foc (9) implies (11) because it is defined in terms of a random variable, $x$.

# 3 Small Sample Density of the LIML Estimator

To proof and illustrate the applicability of the unique conditional densities for constructing small sample densities of ml estimators, we use the instrumental variables regression model. An instrumental variables regression model can be considered as a restriction on the parameters of a linear model and we use this property to construct the small sample density of its maximum likelihood estimator, the limited information maximum likelihood (liml) estimator. We use this example because the restriction imposed on the parameters of the linear model by the instrumental variables regression model is not that straightforward which it is for example when we construct the small sample density of the ml estimator of the parameters of a linear model by using that is a restriction on an encompassing linear model. The applicability of the concept of unique conditional densities is obvious in that case.

### 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 Hausman (1983) and Kleibergen and Zivot (1998),

$$\begin{aligned} y_1 &= Y_2\beta + Z\gamma + \varepsilon_1, \\ Y_2 &= X\Pi + Z\Gamma + V_2, \end{aligned} \tag{12}$$

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 $T \times (m-1)$ 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 $Z$, which may contain lagged endogenous variables, are assumed to be of full column rank, uncorrelated with $\varepsilon_1$ and $V_2$ and weakly exogenous for the structural parameter $\beta$. 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

$$\Sigma = var \begin{pmatrix} \varepsilon_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{13}$$

The degree of endogeniety of $Y_2$ in (12) 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 nonlinearly *restricted reduced form* specification

$$Y = X\Pi B + Z\Psi + V, \tag{14}$$

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

$$\Omega = var \begin{pmatrix} v_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} e_1' \\ B \end{pmatrix}' \Sigma \begin{pmatrix} e_1' \\ B \end{pmatrix}, \tag{15}$$

where $e_1 : m \times 1$ is the first $m$ dimensional unity vector. Note that $\Psi$ is an 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

$$Y = X\Phi + Z\Psi + V, \tag{16}$$

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 $k_2 > m-1$. $k-m+1$ is thus the degree of overidentification. $\beta$ 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 maximum likelihood estimator of $\beta$, $\hat{\beta}$, is obtained from the concentrated log-likelihood which results when we have concentrated out $\Pi$ and $\Sigma$ from the log-likelihood of the parameters of the model (12), see *e.g.* Hausman (1983),

$$\begin{aligned} \log(L(\beta|X,Y)) &= \frac{1}{2}T\log\left|\frac{(y_1 - Y_2\beta)'M_X(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'(y_1 - Y_2\beta)}\right| \\ &= \frac{1}{2}T\log\left|1 - \frac{(y_1 - Y_2\beta)'X(X'X)^{-1}X'(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'(y_1 - Y_2\beta)}\right| \\ &= \frac{1}{2}T\log|1-\eta|, \end{aligned} \tag{17}$$

where $M_X = I_T - X(X'X)^{-1}X'$, $\eta = \frac{(y_1 - Y_2\beta)'X(X'X)^{-1}X'(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'(y_1 - Y_2\beta)}$. 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 $\eta$,

$$\eta = \min_{\beta}\left[\frac{(y_1 - Y_2\beta)'X(X'X)^{-1}X'(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'(y_1 - Y_2\beta)}\right], \tag{18}$$

which is identical to solving the eigenvalue problem,

$$\begin{aligned} |\eta Y'Y - Y'X(X'X)^{-1}X'Y| &= 0 \Leftrightarrow \\ |\eta I_m - (Y'Y)^{-1}\hat{\Phi}X'X\hat{\Phi}| &= 0, \end{aligned} \tag{19}$$

where $\hat{\Phi} = (X'X)^{-1}X'Y$, and to use the smallest root of (19), 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 (12). The foc (3), which we use to construct the small sample density, is, however, specified on the linear model (2). We therefore use the restricted reduced form (14) to obtain the foc that allows for constructing the small sample density of the liml estimator. This foc reads,

$$
\begin{aligned}
&\left(\frac{\partial vec(\Pi B)}{(\partial vec(\beta)'\ \partial vec(\Pi)')}|_{(\hat\Omega,\hat\beta,\hat\Pi)}\right)' vec(X'(Y-X\Pi B)\Omega^{-1})|_{(\hat\Omega,\hat\beta,\hat\Pi)}=0 \\
\Leftrightarrow\ &\left(\frac{\partial vec(\Pi B)}{(\partial vec(\beta)'\ \partial vec(\Pi)')}|_{(\hat\Omega,\hat\beta,\hat\Pi)}\right)' vec(X'X(\hat\Phi-\Pi B)\Omega^{-1})|_{(\hat\Omega,\hat\beta,\hat\Pi)}=0 \\
\Leftrightarrow\ &\left(\frac{\partial vec(\Pi B)}{(\partial vec(\beta)'\ \partial vec(\Pi)')}|_{(\hat\Omega,\hat\beta,\hat\Pi)}\right)' vec(X'X(\hat\Phi-\Pi B)\Omega^{-1})|_{(\hat\Omega,\hat\beta,\hat\Pi)}=0 \\
\Leftrightarrow\ &\left(\left(\Omega^{-\frac{1}{2}}\otimes S_0^{\frac{1}{2}}\right)\left(\frac{\partial vec(\Pi B)}{(\partial vec(\beta)'\ \partial vec(\Pi)')}|_{(\hat\Omega,\hat\beta,\hat\Pi)}\right)\right)' vec(\hat\Theta-\hat\Gamma\hat D)|_{(\hat\Omega,\hat\beta,\hat\Pi)}=0 \\
\Leftrightarrow\ &\left(\frac{\partial vec(\Gamma D)}{(\partial vec(\beta)'\ \partial vec(\Pi)')}|_{(\hat\Omega,\hat\beta,\hat\Pi)}\right)' vec(\hat\Theta-\hat\Gamma\hat D)|_{(\hat\Omega,\hat\beta,\hat\Pi)}=0,
\end{aligned}
\tag{20}
$$

where $S_0=X'X$, $\hat\Phi=(X'X)^{-1}X'Y$, $\hat\Theta=S_0^{\frac{1}{2}}\hat\Phi\Omega^{-\frac{1}{2}}$, $\Gamma D=S_0^{\frac{1}{2}}\Pi B\Omega^{-\frac{1}{2}}$, $\Gamma: k\times(m-1)$ and $D:(m-1)\times m$, $D=\begin{pmatrix}\delta & I_{m-1}\end{pmatrix}$, $|_{(\hat\Omega,\hat\beta,\hat\Pi)}$ stands for evaluated in $(\hat\Omega,\hat\beta,\hat\Pi)$ and $\hat\Gamma\hat D$ stands for $\Gamma D$ evaluated at $(\hat\Omega,\hat\beta,\hat\Pi)$. Note that the foc (20) also involves the covariance matrix estimator $\hat\Omega$ as we donot know its value like in the foc (3). Note also that only in case of an unknown value of the covariance matrix are the structural form (12) and the restricted reduced form (14) equivalent.

In case the density of the disturbances is known, we can construct the small sample density of the liml estimator. For example when $V\sim n(0,\Omega_0\otimes I_T)$ and the true value of $\Pi$, $\beta$, is $\Pi_0$, $\beta_0$ respectively. The least squares estimator is then distributed as

$$\hat\Phi\sim n(\Phi_0,\Omega_0\otimes S_0^{-1}), \tag{21}$$

where $\Phi_0=\Pi_0B_0$, $B_0=\begin{pmatrix}\beta_0 & I_{m-1}\end{pmatrix}$. The covariance matrix estimator $\hat\Omega$ $(=\frac{1}{T}Y'M_XY=\frac{1}{T}(Y-X\hat\Phi)'(Y-X\hat\Phi))$ based on the least squares estimator $\hat\Phi$ is distributed as

$$\hat\Omega\sim W(\tfrac{1}{T}\Omega_0,T-k), \tag{22}$$

which stands for a Wishart distribution with scale matrix $\frac{1}{T-k}\Omega_0$ and $T-k$ degrees of freedom. The mean of this distribution is $\frac{T-k}{T}\Omega_0$. For details on the Wishart distribution, see Muirhead (1982). The distribution of $\hat\Theta$ $(=S_0^{\frac{1}{2}}\hat\Phi\Omega_0^{-\frac{1}{2}})$, *i.e.* the "$t$-values" of $\hat\Phi$, directly results from (21),

$$\hat\Theta\sim n(\Theta_0,I_m\otimes I_k), \tag{23}$$

where $\Theta_0=S_0^{\frac{1}{2}}\Phi_0\Omega_0^{-\frac{1}{2}}$. So, the statistic $\hat\Theta$ can be seen as a realization from the distribution (23). When we consider the foc (20) expressed in the random variable $\hat\Theta$, the foc can only be satisfied when $\hat\Theta$ is generated under the condition that

$$\hat\Theta-\hat\Gamma\hat D\equiv 0. \tag{24}$$

This holds as $\hat\Theta$ is a random variable and the foc (20) holds for all of its realizations. When (24) satisfies the sufficient conditions for the existence of an unique conditional density, we can construct the small sample density of the liml estimator from that conditional density.

The restriction (24) is a reduced rank restriction as it implies that the rank of the $k\times m$ (random) matrix $\hat\Theta$ is equal to $m-1$. $\hat\Gamma$ is namely a $k\times(m-1)$ matrix and $\hat D$ a $(m-1)\times m$

matrix such that the rank of $\hat{\Gamma}\hat{D}$ is equal to $m-1$. The rank of a matrix is represented by the number of nonzero singular values, which are generalized eigenvalues of non-symmetric matrices, see Golub and van Loan (1989). The singular values result from the singular value decomposition,

$$\hat{\Theta} = USV', \tag{25}$$

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 nonnegative singular values in decreasing order on its main diagonal $(= (s_{11} ... s_{mm}))$ and is equal to zero elsewhere. The reduced rank restriction (24) imposed by the foc is thus the restriction that the smallest singular value of $\hat{\Theta}$ is equal to zero.

It is convenient to represent the rank restriction on $\hat{\Theta}$ using the specification

$$\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_{\perp}\hat{\lambda}\hat{D}_{\perp}, \tag{26}$$

where $\hat{\Gamma}_{\perp}$ is a $k \times (k-m+1)$ matrix such that $\hat{\Gamma}'\hat{\Gamma}_{\perp} \equiv 0$ and $\hat{\Gamma}'_{\perp}\hat{\Gamma}_{\perp} \equiv I_{k-m+1}$; $\hat{D}_{\perp}$ is a $1 \times m$ vector such that $\hat{D}\hat{D}'_{\perp} \equiv 0$, $\hat{D}_{\perp}\hat{D}'_{\perp} \equiv 1$; and $\hat{\lambda}$ is a $(k-m+1) \times 1$ vector to be specified. $\hat{\Gamma}_{\perp}$ and $\hat{D}_{\perp}$ can 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 representation (26) is an unrestricted specification of $\hat{\Theta}$ and results from the singular value decomposition (25) with

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, \; S = \begin{pmatrix} S_1 & 0 \\ 0 & s_2 \end{pmatrix} \text{ and } V = \begin{pmatrix} v_{11} & v_{12} \\ V_{21} & v_{22} \end{pmatrix}, \tag{27}$$

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. Explicit expressions for $\hat{\delta}$, $\hat{\Gamma}$ and $\hat{\lambda}$ are derived in Kleibergen (1998) and Kleibergen and van Dijk (1998) and are given by

$$\hat{\Gamma} = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V'_{21}, \; \hat{\delta} = V_{21}^{\prime -1} v'_{11}, \; \hat{\lambda} = (U_{22}U'_{22})^{-\frac{1}{2}} U_{22} s_2 v'_{12} (v_{12} v'_{12})^{-\frac{1}{2}}. \tag{28}$$

The specification of $\hat{\lambda}$ in (28) is such that $\hat{\lambda}$ is an orthogonal transformation of the smallest singular value contained in $s_2$. This implies that we can always solve $s_2$ from $\hat{\lambda}$ for all values of $\hat{\lambda}$. This is also reflected in the Jacobian of the transformation from $s_2$ to $\hat{\lambda}$ which is equal to one and independent of the other parameters as well as the data. Restricting the smallest singular value to zero is thus equivalent to restricting $\hat{\lambda}$ to zero and the rank restriction (24) is thus equivalent with $\hat{\lambda} = 0$. The rank restriction (24) resulting from the foc thus satisfies the sufficient conditions for the existence of an unique conditional density and we can use it to construct the small sample density of the liml estimator.

[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 $Q^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P'$.

The conditional density of $\hat{\Theta}$ given that (24) holds then reads,

$$\begin{aligned} p_r(\hat{\Theta}) &\propto p(\hat{\Theta})|_{rank(\hat{\Theta})=m-1} \\ &\propto p(\hat{\Theta}(\hat{\Gamma},\hat{\delta},\hat{\lambda}))|_{\hat{\lambda}=0}|J(\hat{\Theta},(\hat{\Gamma},\hat{\delta},\hat{\lambda}))|_{\hat{\lambda}=0}| \\ &\propto \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D}-\Theta_0\right)'\left(\hat{\Gamma}\hat{D}-\Theta_0\right)\right)\right]\left|\begin{pmatrix} \hat{D}'\otimes I_k & e_1\otimes\hat{\Gamma} & \hat{D}'_\perp\otimes\hat{\Gamma}_\perp \end{pmatrix}\right| \\ &\propto \left|\begin{pmatrix} \hat{D}\hat{D}'\otimes I_k & \hat{\delta}\otimes\hat{\Gamma} \\ \hat{\delta}'\otimes\hat{\Gamma}' & 1\otimes\hat{\Gamma}'\hat{\Gamma}' \end{pmatrix}\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D}-\Theta_0\right)'\left(\hat{\Gamma}\hat{D}-\Theta_0\right)\right)\right] \\ &\propto |\hat{\Gamma}'\hat{\Gamma}|^{\frac{1}{2}}|I_{m-1}+\hat{\delta}\hat{\delta}'|^{\frac{1}{2}(k-m+1)}\exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D}-\Theta_0\right)'\left(\hat{\Gamma}\hat{D}-\Theta_0\right)\right)\right], \end{aligned} \tag{29}$$

where we have used that, see Kleibergen (1998), Kleibergen and van Dijk (1998) and Kleibergen and Zivot (1998),

$$\begin{aligned} |J(\hat{\Theta},(\hat{\Gamma},\hat{\delta},\hat{\lambda}))|_{\hat{\lambda}=0}| &= \left|\begin{pmatrix} \hat{D}'\otimes I_k & e_1\otimes\hat{\Gamma} & \hat{D}'_\perp\otimes\hat{\Gamma}_\perp \end{pmatrix}\right| \\ &= \left|\begin{pmatrix} \hat{D}\hat{D}'\otimes I_k & \hat{\delta}\otimes\hat{\Gamma} \\ \hat{\delta}'\otimes\hat{\Gamma}' & 1\otimes\hat{\Gamma}'\hat{\Gamma}' \end{pmatrix}\right|^{\frac{1}{2}} \\ &= |\hat{\Gamma}'\hat{\Gamma}|^{\frac{1}{2}}|I_{m-1}+\hat{\delta}\hat{\delta}'|^{\frac{1}{2}(k-m+1)}, \end{aligned} \tag{30}$$

where $e_1$ is the first $m$ dimensional unity vector. Since $\hat{\Theta}$ in (29) has a reduced rank value an invertible relationship between $(\hat{\Gamma},\hat{\delta})$, which span the reduced rank space of $\hat{\Theta}$, and $(\hat{\beta},\hat{\Pi})$ exists,

$$\hat{\Gamma}\hat{D} = S_0^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega_0^{-\frac{1}{2}} = S_0^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega_2\begin{pmatrix} \left(\hat{B}\Omega_2\right)^{-1}\hat{B}\omega_1 & I_{m-1} \end{pmatrix}, \tag{31}$$

where $\Omega_0^{-\frac{1}{2}} = \begin{pmatrix} \omega_1 & \Omega_2 \end{pmatrix}$ with $\omega_1$ a $m\times 1$ vector and $\Omega_2$ a $m\times(m-1)$ matrix such that $\hat{\delta} = \left(\hat{B}\Omega_2\right)^{-1}\hat{B}\omega_1$ and $\hat{\Gamma} = S^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega_2$. (31), however, assumes that we know the value of $\Omega_0$ which we donot and we therefore need to replace it by an estimator of it, which also has a density function, in order to solve $(\hat{\beta},\hat{\Pi})$ from $(\hat{\Gamma},\hat{\delta})$. A possible estimator of $\Omega_0$ is $\hat{\Omega}$ and we can use it instead of $\Omega_0$ in (31). $\hat{\Omega}$ is, however, not the only estimator/random variable that can represent $\Omega_0$ and we can use in order to construct the small sample density of the liml estimator. Any estimator/random variable with a mean that is proportional to $\Omega_0$ and is stochastically independent of $\hat{\Theta}$ can be used such that we can choose that estimator/random variable which leads to the most convenient expression of the small sample density. We therefore use $\hat{\Lambda} = \Omega_0\hat{\Omega}^{-1}\Omega_0$ instead of $\hat{\Omega}$ as it leads to a more convenient expression of the small sample density. $\hat{\Lambda}$ is distributed as,

$$\hat{\Lambda} \sim iW(T\Omega_0, T-k),$$

where $iW$ stands for inverted-Wishart, and has a mean equal $\frac{T}{T-k-m-1}\Omega_0$, see Muirhead (1982), such that it can be used as the estimator/random variable representing $\Omega_0$ in (31). We, therefore use instead of (31),

$$\hat{\Gamma}\hat{D} = S_0^{\frac{1}{2}}\hat{\Pi}\hat{B}\hat{\Lambda}^{-\frac{1}{2}} = S_0^{\frac{1}{2}}\hat{\Pi}\hat{B}\Delta_2\begin{pmatrix} \left(\hat{B}\Delta_2\right)^{-1}\hat{B}\Delta_1 & I_{m-1} \end{pmatrix}, \tag{32}$$

where $\hat{\Lambda}^{-\frac{1}{2}} = \begin{pmatrix} \Delta_1 & \Delta_2 \end{pmatrix}$ with $\Delta_1$ a $m \times 1$ vector and $\Delta_2$ a $m \times (m-1)$ matrix such that $\hat{\delta} = \left(\hat{B}\Delta_2\right)^{-1}\hat{B}\Delta_1$ and $\hat{\Gamma} = S^{\frac{1}{2}}\hat{\Pi}\hat{B}\Delta_2$, to obtain $(\hat{\beta}, \hat{\Pi})$ from $(\hat{\Gamma}, \hat{\delta})$. The joint density of $(\hat{\beta}, \hat{\Pi})$ and $\hat{\Lambda}$ then results from the joint density of $(\hat{\Gamma}, \hat{\delta})$ and $\hat{\Lambda}$ as $\hat{\Lambda}$ is stochastically independent of $\hat{\Theta}$,

$$\begin{aligned} p_r(\hat{\Gamma}, \hat{\delta}, \hat{\Lambda}) &\propto p_r(\hat{\Theta})p(\hat{\Lambda}) \\ &\propto |\hat{\Gamma}'\hat{\Gamma}|^{\frac{1}{2}}|I_{m-1} + \hat{\delta}\hat{\delta}'|^{\frac{1}{2}(k-m+1)}|\hat{\Lambda}|^{-\frac{1}{2}(T-k+m+1)} \\ &\quad \exp\left[-\frac{1}{2}tr\left(T\Omega_0\hat{\Lambda}^{-1} + \left(\hat{\Gamma}\hat{D} - \Theta_0\right)'\left(\hat{\Gamma}\hat{D} - \Theta_0\right)\right)\right], \end{aligned} \tag{33}$$

by transforming $(\hat{\Gamma}, \hat{\delta})$ to $(\hat{\beta}, \hat{\Pi})$ we then obtain the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$, see appendix A,

$$\begin{aligned} p_r(\hat{\beta}, \hat{\Pi}, \hat{\Lambda}) &\propto p_r(\hat{\Gamma}(\hat{\beta}, \hat{\Pi}, \hat{\Lambda}), \hat{\delta}(\hat{\beta}, \hat{\Pi}, \hat{\Lambda}), \hat{\Lambda})|J((\hat{\Gamma}, \hat{\delta}), (\hat{\beta}, \hat{\Pi}))| \\ &\propto |\hat{\Lambda}|^{-\frac{1}{2}(T-k+2m)}\left|\hat{\Pi}'S_0\hat{\Pi}\right|^{\frac{1}{2}}\left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)}|S_0|^{\frac{1}{2}(m-1)} \\ &\quad \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(T\Omega_0 + \left(\hat{\Pi}\hat{B} - \Pi_0 B_0\right)' S_0\left(\hat{\Pi}\hat{B} - \Pi_0 B_0\right)\right)\right)\right]. \end{aligned} \tag{34}$$

In case that $m = 2$, we can integrate out $\hat{\Pi}$ from (34) analytically to obtain the joint density of $(\hat{\beta}, \hat{\Lambda})$, see appendix A,

$$\begin{aligned} &p_r(\hat{\beta}, \hat{\Lambda}) \\ \propto\ & p_r(\hat{\beta}|\hat{\Lambda})p_r(\hat{\Lambda}) \\ \propto\ & |\hat{\Lambda}|^{-\frac{1}{2}(T-k+2m)}\exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(T\Omega_0 + B_0'\Pi_0'S_0\Pi_0B_0\right)\right)\right] \\ & \left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ & \left[\sum_{j=0}^{\infty}\left(\left(\frac{\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \beta_0'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^2 \Pi_0'S_0\Pi_0}{2\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|}\right)^j\right.\right. \\ & \left.\left.\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right], \end{aligned} \tag{35}$$

such that

$$\begin{aligned} p_r(\hat{\beta}|\hat{\Lambda}) &\propto \left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ &\quad \left[\sum_{j=0}^{\infty}\left(\left(\frac{\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \beta_0'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^2 \Pi_0'S_0\Pi_0}{2\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|}\right)^j\right.\right. \\ &\quad \left.\left.\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right] \\ p_r(\hat{\Lambda}) &\propto |\hat{\Lambda}|^{-\frac{1}{2}(T-k+2m)}\exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(T\Omega_0 + B_0'\Pi_0'S_0\Pi_0B_0\right)\right)\right], \end{aligned} \tag{36}$$

which shows that $\hat{\Lambda} \sim iW(T\Omega_0 + B_0'\Pi_0'S_0\Pi_0B_0, T-k+m-1)$ and has thus changed because of the rank reduction imposed on $\hat{\Theta}$ and the dependence of $(\hat{\beta}, \hat{\Pi})$ on $\hat{\Lambda}$. Note that the mean of this inverted-Wishart density is $\frac{1}{T-k+m-1}(T\Omega_0 + B_0'\Pi_0'S_0\Pi_0B_0) \approx \Omega_0 + \frac{1}{T}B_0'\Pi_0'S_0\Pi_0B_0$ and that this density is centered close around its mean for reasonably large values of $T$ ($T > 25$). A good approximation of the marginal density of $\hat{\beta}$ is therefore $p_r(\hat{\beta}|\hat{\Lambda} = \Omega_0 + \frac{1}{T}B_0'\Pi_0'S_0\Pi_0B_0)$. This is also the reason why we use $\hat{\Lambda}$ instead of $\hat{\Omega}$ since the resulting marginal density of $\hat{\Omega}$ does not belong to a standard class while the density of $\hat{\Lambda}$ does such that we can use its properties, like the mean as shown above.

## 3.4 Properties of the Small Sample Density

The previous (sub)sections focussed on the construction of the small sample density of the liml estimator. In this section we discuss the properties of the resulting expression of the small sample density (35). We discuss the small sample density itself, its relationship with the already existing expressions in the literature, its convergence properties when the sample size increases, how it relates to the sampling density and its implications for testing hypotheses.

### 3.4.1 The Small Sample Density

The conditional small sample density of $\hat{\beta}$ given $\hat{\Lambda}$ (36) consists of the product of a Cauchy kernel and a single infinite sum. Since the first element of the infinite sum does not 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 moments besides the distribution exist. Furthermore, because the marginal density of $\hat{\Lambda}$ is finite everywhere, as it is an inverted-Wishart density, the tail behavior of the marginal density of $\hat{\beta}$ is thus identical to the tail behavior of the conditional density of $\hat{\beta}$ given $\hat{\Lambda}$. So, also marginally no finite moments of $\hat{\beta}$ besides the distribution exist. When $\Pi_0 = 0$, the only element remaining of the infinite sum is a constant such that in that case the conditional small sample density of $\hat{\beta}$ given $\hat{\Lambda}$ is even equal to a Cauchy density. Another simplification occurs when $\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} = \beta_0'$ since the conditional density is symmetric in that case.

A nice and elegant feature of the joint small density of $(\hat{\beta}, \hat{\Lambda})$ is that it can be decomposed into the product of a marginal density of $\hat{\Lambda}$ that belongs to a known class of density functions, *i.e.* the inverted-Wishart, and a conditional density of $\hat{\beta}$ given $\hat{\Lambda}$. This is also the reason why we use $\hat{\Lambda}$ instead of $\hat{\Omega}$. We can therefore use the properties of the inverted-Wishart density in our analysis. The mean of the inverted-Wishart distribution of $\hat{\Lambda}$ is equal to $\frac{1}{T-k+m-1}(T\Omega_0 + B_0'\Pi_0'S_0\Pi_0B_0) \approx \Omega_0 + \frac{1}{T}B_0'\Pi_0'S_0\Pi_0B_0$ and its variance is proportional to $\frac{1}{T}$, see Muirhead (1982). The mean of $\hat{\Lambda}$ is therefore equal to $\frac{1}{T}$ times the expectation of the quadratic form of the endogenous variables, $\frac{1}{T}E(Y'Y)$. This result is not that surprising since $Y'Y$ is used to construct the liml estimator in (19). Because the variance of $\hat{\Lambda}$ is proportional to $\frac{1}{T}$, the density of $\hat{\Lambda}$ quickly concentrates around its mean when the sample size, $T$, increases. This convergence is quite fast which can be concluded for example from the well-known result that an univariate $t$ density with 25 degrees of freedom is almost identical to the normal density. It implies that already for quite small sample sizes ($T > 25$), the density of $\hat{\Lambda}$ is so concencrated around its mean that we can approximate the marginal density of $\hat{\beta}$ by the conditional density of $\hat{\beta}$ given that $\hat{\Lambda}$ is equal to the mean of its marginal density very well. So, we then use $p_r(\hat{\beta}) \approx p_r(\hat{\beta}|\hat{\Lambda} = \Omega_0 + \frac{1}{T}B_0'\Pi_0'S_0\Pi_0B_0)$. For smaller sample sizes, we can compute the marginal density of $\hat{\beta}$ straightforwardly by sampling $\hat{\Lambda}$ from its marginal density and taking the average of the conditional density of $\hat{\beta}$ given the sampled $\hat{\Lambda}$'s.

The small sample density of $(\hat{\beta}, \hat{\Lambda})$ reads as (35) 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 moment of the determinant of a noncentral-Wishart distributed random matrix. For details on this, see Muirhead (1982). Since we do have the expression of the joint small sample density of $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$ 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 $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$, 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 small sample density (34) results from a different approach then the one 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 small sample density in Mariano and Sawa (1972) which consists of a triple infinite sum while the small sample density (36) consists of a single infinite sum. One reason for this is that the small sample density constructed by Mariano and Sawa is the marginal density while (36) is the conditional density of $\hat{\beta}$ given $\hat{\Lambda}$. Another reason is that these small sample densities are constructed using different approaches. The small sample density (36) results from an unique conditional density that exploits the property of the liml estimator that it satisfies the foc while the traditional approach constructs the small sample density from a closed form expression of the liml estimator. Since we already discussed the construction of the small sample density (36) at length, we now briefly discuss the traditional way of constructing the small sample density of the liml estimator to show the differences and similarities with the unique conditional density approach.

The liml estimator results from the characteristic polynomial (19), see Mariano and Sawa (1972),

$$\begin{aligned} |\eta Y'Y - Y'X(X'X)^{-1}X'Y| &= 0 \Leftrightarrow \\ |\eta \left(Y'M_X Y + Y'X(X'X)^{-1}X'Y\right) - Y'X(X'X)^{-1}X'Y| &= 0, \end{aligned} \tag{37}$$

and is defined such that the eigenvector associated with the smallest root of (37) 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 noncentral Wishart distribution, $Y'M_XY$ has a standard Wishart distribution and these random matrices are stochastically independent. Since the liml estimator results from an eigenvector of (37), it satisfies the relationship,

$$\eta \left(Y'M_X Y + Y'X(X'X)^{-1}X'Y\right) \begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix} = Y'X(X'X)^{-1}X'Y \begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix}, \tag{38}$$

where $\eta$ is the smallest root of (37). The small sample density of the liml estimator $\hat{\beta}$ can thus 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 stochastically 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),

$$\hat{\beta} = (\hat{y}'y_2)^{-1}\hat{y}'y_1, \tag{39}$$

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 (37), 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 small sample 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 $\hat{\Lambda}$, this expression has Cauchy tails such that no finite moments besides the distribution exist. Anderson and Sawa (1979) constructed the small sample 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 (37) by the estimated reduced form covariance matrix $Y'M_XY$. They show that the small sample density of this limlk estimator is less complicated than the small sample density of the liml estimator, as it consists of a double infinite sum, and approximates the small sample density of the liml estimator often quite well, see Anderson *et. al.* (1983).

The liml estimator that results from the characteristic polynomial (37) is closely related to the singular value decomposition used to construct the unique conditional density. This can be shown by specifying (37) as,

$$\begin{aligned}
\left|\eta Y'Y - \hat{\Phi}'X'X\hat{\Phi}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \hat{\Theta}'\hat{\Theta}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix}' \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix}' \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix}' \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix} \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix} \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix} \begin{pmatrix} \hat{D} \\ \hat{D}_\perp \end{pmatrix}' \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix}' \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix}' \begin{pmatrix} \hat{\Gamma} & \hat{\Gamma}_\perp \end{pmatrix} \begin{pmatrix} I_{m-1} & 0 \\ 0 & \hat{\lambda} \end{pmatrix}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \begin{pmatrix} \hat{D}\hat{D}' & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\Gamma}'\hat{\Gamma} & 0 \\ 0 & \hat{\lambda}^2 \end{pmatrix}\right| &= 0 \Leftrightarrow \\
\left|\eta I_m - \begin{pmatrix} \hat{D}\hat{D}'\hat{\Gamma}'\hat{\Gamma} & 0 \\ 0 & \hat{\lambda}^2 \end{pmatrix}\right| &= 0
\end{aligned} \tag{40}$$

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 (26). The liml estimator of $\beta$, $\hat{\beta}$, results from the eigenvector associated with the smallest root of (40) which is $\hat{\lambda}^2$. This eigenvector can be specified by $a\hat{D}'_\perp$, where $a$ is a nonzero scalar. Because of the specification of $\hat{D}_\perp$ $(= (1 + \hat{\delta}'\hat{\delta})^{-\frac{1}{2}}(1 \ -\hat{\delta}'))$, see (26), $\hat{\beta}$ and $\hat{\delta}$ coincide. To construct the unique conditional density, $\hat{\lambda}$ is set equal to zero and $\hat{\Theta}$ is therefore set equal to $\hat{\Gamma}\hat{D}$. $(\hat{\Pi}, \hat{\beta})$ is then solved from $\hat{\Gamma}\hat{D}$ in (32). So, to obtain the density of $(\hat{\Pi}, \hat{\beta})$ from the conditional density of $\hat{\Theta}$ given that it has rank $m-1$, we have to account for the variance and cannot directly solve $(\hat{\Pi}, \hat{\beta})$ from $\hat{\Theta}$ as in the construction of the density of $(\hat{\Pi}, \hat{\beta})$ based on (37). This is natural as the $\hat{\Theta}$ used there has an identity covariance matrix and (40) shows that the liml estimator can be considered to result from imposing rank reduction on a devarianced $\hat{\Phi}$. Note also that we solve $(\hat{\Pi}, \hat{\beta})$ from $\hat{\Theta}$ in the conditional density using the covariance matrix $\hat{\Lambda}$ which has a mean, according to its marginal density, equal to $\frac{1}{T}E(Y'Y)$, and that $Y'Y$ is used in (40). The conditional density 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 small sample densities (36) and the one from Mariano and Sawa (1972) are also different in the sense that (36) is the conditional density of $\hat{\beta}$ given $\hat{\Lambda}$ while the small sample density in Mariano and Sawa is the marginal density. For two reasons is this difference not so important as it seems. First, the marginal density of $\hat{\Lambda}$ is known and belongs to a standard class of density functions such that we for example know that it concentrates quickly around its mean when sample size increases and how to sample from it. Second, $\hat{\Lambda}$ is a nuisance parameter and the identification of $\hat{\beta}$ does not depend on it. The conditional density is therefore not dramatically different for various values of $\hat{\Lambda}$ like for example the conditional density of $\hat{\beta}$ given $\hat{\Pi}$ that strongly differs over the value of $\hat{\Pi}$ as the identification of $\beta$ depends on $\Pi$.

A nice feature of the unique conditional density approach is that we straightforwardly obtain the analytical expression of the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$ (34) without the involvement of the noncentral Wishart density. This density is only involved in the integration over $\hat{\Pi}$ to obtain the joint density of $(\hat{\beta}, \hat{\Lambda})$. As the traditional approach for constructing the small sample density uses the noncentral Wishart density from the outset, see previous discussion, the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$ is more complicated to construct using that approach. Furthermore, given that we have the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$ 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 thus enable us to compute the small sample density of the liml estimator also in case of more than two endogenous variables. We therefore donot 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 $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$ (34) is identical 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 Small Sample Distribution to Limiting Distribution

The small sample density of $\hat{\beta}$ given $\hat{\Lambda}$,

$$\begin{aligned} p_r(\hat{\beta}|\hat{\Lambda}) \propto & \left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\ & \left[\sum_{j=0}^{\infty}\left(\left(\frac{\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \beta_0'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^2 \Pi_0' S_0 \Pi_0}{2\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|}\right)^j \right.\right. \\ & \left.\left. \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right], \end{aligned} \tag{41}$$

can be used to analyze the limiting distributions of $\hat{\beta}$ for different values of $\Pi_0$. Below we discuss three different cases, $\Pi_0 = 0$, $\Pi_0 = \Psi_0/\sqrt{T}$, where $\Psi_0$ is a fixed full rank matrix, and $\Pi_0$ is a fixed nonzero 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 (41). 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 small sample density, they show the convergence issues at specific values of $\Pi_0$ in a rather illustrative way.

$\Pi_0 = 0$ : is known as the case of total nonidentification. It implies that the conditional density of $\hat{\beta}$ given $\hat{\Lambda}$ (41) 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).

$\Pi_0 = \Psi_0/\sqrt{T}$: 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 $\Pi_0$ 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 and small sample distribution are identical and as the small sample distribution is nonnormal, it can easily generate "$t$-values" which seem significant when one mistakenly uses normal critical values but are nonsignificant when one uses the correct ones. The similarity of the limiting and small sample distribution results because $S_0 = X'X$ and $p\lim_{T\to\infty}\left(\frac{X'X}{T}\right) = Q_0$ is a fixed full rank matrix, such that,

$$\begin{aligned} p\lim_{T\to\infty}(\Pi_0' S_0 \Pi_0) &= p\lim_{T\to\infty}(\Pi_0' X'X \Pi_0) = p\lim_{T\to\infty}\left(\left(\frac{\Psi_0}{\sqrt{T}}\right)' X'X \left(\frac{\Psi_0}{\sqrt{T}}\right)\right) \\ &= p\lim_{T\to\infty}\left(\Psi_0'\left(\frac{X'X}{T}\right)\Psi_0\right) \\ &= \Psi_0' Q_0 \Psi_0. \end{aligned} \tag{42}$$

$\Pi_0' S_0 \Pi_0$ thus remains a finite constant when the sample size goes to infinity. The continuous mapping theoren, see Billingsley (1986), then implies that

$$
\begin{aligned}
& p \lim_{T\to\infty} p_r(\hat\beta|\hat\Lambda) \qquad (43) \\
\propto\ & p \lim_{T\to\infty} \Bigg[ \left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|^{-\frac{1}{2}m} \\
& \left[ \sum_{j=0}^{\infty} \left( \left( \frac{\left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \beta_0'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|^2 \Pi_0' S_0 \Pi_0}{2 \left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|} \right)^j \right. \right. \\
& \left. \left. \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \right) \right] \Bigg] \\
\propto\ & \left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|^{-\frac{1}{2}m} \\
& \left[ \sum_{j=0}^{\infty} \left( \left( \frac{\left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \beta_0'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|^2 p\lim_{T\to\infty}\left(\Pi_0' S_0 \Pi_0\right)}{2 \left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|} \right)^j \right. \right. \\
& \left. \left. \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \right) \right] \Bigg] \\
\propto\ & \left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|^{-\frac{1}{2}m} \\
& \left[ \sum_{j=0}^{\infty} \left( \left( \frac{\left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \beta_0'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|^2 \Pi_0' Q_0 \Pi_0}{2 \left| \hat\Lambda_{22}^{-1} + \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right)' \hat\Lambda_{11.2}^{-1} \left(\hat\Lambda_{12}\hat\Lambda_{22}^{-1} - \hat\beta'\right) \right|} \right)^j \right. \right. \\
& \left. \left. \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))} \right) \right] \Bigg],
\end{aligned}
$$

which shows that the small sample and limiting distributions are identical, what also holds for the previous case. So, $\hat\beta$ remains a random variable when sample size increases and does thus not converge to the fixed constant $\beta_0$. 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.

$\Pi_0$ **fixed full rank:** implies that $\Pi_0' S_0 \Pi_0$ converges to infinity when sample size increases. To illustrate the convergence of the small sample distribution we now use the joint small sample density of $(\hat\beta, \hat\Pi, \hat\Lambda)$ (34),

$$
\begin{aligned}
p_r(\hat\beta, \hat\Pi, \hat\Lambda) \ \propto\ & |\hat\Lambda|^{-\frac{1}{2}(T-k+2m)} \left|\hat\Pi' S_0 \hat\Pi\right|^{\frac{1}{2}} \left|\hat B \hat\Lambda^{-1} \hat B'\right|^{\frac{1}{2}(k-m+1)} |S_0|^{\frac{1}{2}(m-1)} \\
& \exp\left[ -\frac{1}{2} tr\left( \hat\Lambda^{-1} \left( T\Omega_0 + \left(\hat\Pi\hat B - \Pi_0 B_0\right)' S_0 \left(\hat\Pi\hat B - \Pi_0 B_0\right) \right) \right) \right].
\end{aligned}
$$

Since $S_0 = X'X$ and $p\lim_{T\to\infty}\left(\frac{X'X}{T}\right) = Q_0$ is a fixed full rank matrix, it follows from the continuous mapping theorem, see Billingsley (1986), that,

$$p\lim_{T\to\infty} p_r(\hat{\beta},\hat{\Pi},\hat{\Lambda}) \propto p\lim_{T\to\infty}\left[|\hat{\Lambda}|^{-\frac{1}{2}(T-k+2m)}\left|\hat{\Pi}'S_0\hat{\Pi}\right|^{\frac{1}{2}}\left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)}|S_0|^{\frac{1}{2}(m-1)} \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(T\Omega_0+\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)'S_0\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right)\right)\right]\right] \quad (44)$$

We can now divide $S_0$ by $T$ and multiply $\hat{\Pi}\hat{B}-\Pi_0B_0$ by $\sqrt{T}$ without affecting the joint density,

$$\begin{aligned}&p\lim_{T\to\infty} p_r(\hat{\beta},\hat{\Pi},\hat{\Lambda}) \\ \propto\ & p\lim_{T\to\infty}\left[|\hat{\Lambda}|^{-\frac{1}{2}(T-k+2m)}\left|\hat{\Pi}'X'X\hat{\Pi}\right|^{\frac{1}{2}}\left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)}|X'X|^{\frac{1}{2}(m-1)}\right. \\ & \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}T\Omega_0+\hat{\Lambda}^{-1}\left(\sqrt{T}\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right)'\left(\frac{X'X}{T}\right)\right.\right. \\ & \left.\left.\left.\left(\sqrt{T}\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right)\right)\right]\right].\end{aligned} \quad (45)$$

Since $\hat{\Pi}\hat{B} = \begin{pmatrix}\hat{\Pi}\hat{\beta} & \hat{\Pi}\end{pmatrix}$ and $\Pi_0B_0 = \begin{pmatrix}\Pi_0\beta_0 & \Pi_0\end{pmatrix}$, $\hat{\Pi}\Rightarrow\Pi_0$ and $\hat{\beta}\Rightarrow\beta_0$ ($\hat{B}\Rightarrow B_0$) such that, because of the continuous mapping theorem,

$$\begin{aligned}&p\lim_{T\to\infty} p_r(\hat{\beta},\hat{\Pi},\hat{\Lambda}) \\ \propto\ & p\lim_{T\to\infty}\left[|\hat{\Lambda}|^{-\frac{1}{2}(T-k+2m)}|\Pi_0'X'X\Pi_0|^{\frac{1}{2}}\left|B_0\hat{\Lambda}^{-1}B_0'\right|^{\frac{1}{2}(k-m+1)}|X'X|^{\frac{1}{2}(m-1)}\right. \\ & \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}T\Omega_0+\hat{\Lambda}^{-1}\left(\sqrt{T}\begin{pmatrix}\Pi_0(\hat{\beta}-\beta_0) & \hat{\Pi}-\Pi_0\end{pmatrix}\right)'\left(\frac{X'X}{T}\right)\right.\right. \\ & \left.\left.\left.\left(\sqrt{T}\begin{pmatrix}\Pi_0(\hat{\beta}-\beta_0) & \hat{\Pi}-\Pi_0\end{pmatrix}\right)\right)\right]\right].\end{aligned} \quad (46)$$

By performing a transformation of the random variables from $\hat{\beta}$ to $\sqrt{T}(\hat{\beta}-\beta_0)$, and $\hat{\Pi}$ to $\sqrt{T}(\hat{\Pi}-\Pi_0)$, with jacobian $|J((\hat{\beta},\hat{\Pi}),(\sqrt{T}(\hat{\beta}-\beta_0),\sqrt{T}(\hat{\Pi}-\Pi_0)))| = T^{-\frac{1}{2}((k+1)(m-1))}$, we then obtain that

$$\begin{aligned}&p\lim_{T\to\infty} p(\sqrt{T}(\hat{\beta}-\beta_0),\sqrt{T}(\hat{\Pi}-\Pi_0),\hat{\Lambda}) \\ \propto\ & p\lim_{T\to\infty} p_r((\hat{\beta},\hat{\Pi})(\sqrt{T}(\hat{\beta}-\beta_0),\sqrt{T}(\hat{\Pi}-\Pi_0)),\hat{\Lambda}) \\ & |J((\hat{\beta},\hat{\Pi}),(\sqrt{T}(\hat{\beta}-\beta_0),\sqrt{T}(\hat{\Pi}-\Pi_0)))| \\ \propto\ & p\lim_{T\to\infty}\left[|\hat{\Lambda}|^{-\frac{1}{2}(T-k+2m)}\left|\Pi_0'\left(\frac{X'X}{T}\right)\Pi_0\right|^{\frac{1}{2}}\left|B_0\hat{\Lambda}^{-1}B_0'\right|^{\frac{1}{2}(k-m+1)}\left|\frac{X'X}{T}\right|^{\frac{1}{2}(m-1)}\right. \\ & \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}T\Omega_0+\hat{\Lambda}^{-1}\left(\sqrt{T}\begin{pmatrix}\Pi_0(\hat{\beta}-\beta_0) & \hat{\Pi}-\Pi_0\end{pmatrix}\right)'\left(\frac{X'X}{T}\right)\right.\right. \\ & \left.\left.\left.\left(\sqrt{T}\begin{pmatrix}\Pi_0(\hat{\beta}-\beta_0) & \hat{\Pi}-\Pi_0\end{pmatrix}\right)\right)\right]\right],\end{aligned} \quad (47)$$

which shows that

$$\begin{aligned}\sqrt{T}(\hat{\beta}-\beta_0) &\Rightarrow n(0,\omega_{11.2}(\Pi_0'Q_0\Pi_0)^{-1}), \\ \sqrt{T}(\hat{\Pi}-\Pi_0) &\Rightarrow n(0,\Omega_{22}\otimes Q_0^{-1}), \\ \hat{\Lambda} &\Rightarrow \Omega_0.\end{aligned} \quad (48)$$

The limiting distributions in (48) are constructed under the assumption of normality of the disturbances. In the next section, we derive the same result also using the concept of an unique conditional density under the less stringent assumption that the least squares estimator of the encompassing unrestricted reduced form has a normal limiting distribution. The limiting distributions in (48) also 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 the small sample density. The small sample density is thus a convenient tool using which all convergence issues can be illustrated.

#### 3.4.4 Small Sample Density versus Sampling Density

To show the validity of the concept of an unique conditional density and its applicability for constructing the small sample density, we compare the resulting small sample density with the sampling density for specific parameter values. We therefore sampled one million datasets from the model,

$$\begin{aligned} y_1 &= \beta y_2 + \varepsilon_1, \\ y_2 &= X\pi + v_2, \end{aligned} \tag{49}$$

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 conditional density of $\hat{\beta}$ given $\hat{\Lambda}$ (36) with $\hat{\Lambda} = \Omega + \frac{1}{T} B'\pi' X' X \pi B$, $B = \begin{pmatrix} \beta & 1 \end{pmatrix}$, $\Omega = \begin{pmatrix} e_1 & B' \end{pmatrix}' \Sigma \begin{pmatrix} e_1 & B' \end{pmatrix}$. This value of $\hat{\Lambda}$ equals the mean of the marginal density of $\hat{\Lambda}$ and since $T$ is quite large the marginal density is concentrated around its mean. Note that $X$ is fixed over the datasets, and we also use it in the conditional density $p_r(\hat{\beta}|\hat{\Lambda})$, 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 $\Omega_{22}^{-1}\omega_{21} = 2$. 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 nonnormal small sample and sampling densities. A coinciding small sample and sampling density at these extreme parameter values is therefore a strong indication of the correctness of the small sample density and thus of the appropriateness of the concept of an unique conditional density.

In figure 1, the small sample and sampling densities in case of total nonidentification, $\pi_1 = 0$, are shown and they are indistinguishable. We only show the exact identified case because increasing the degree of overidentification does not affect the small sample or sampling density at all (as was to be expected from (36)). 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, $\pi_1 = 0$, with $k = 1$, 5 (figure 3) and $k = 20$ (figure 4). For all cases, the small sample and sampling densities are hard to distinguish from one another which is, given the extreme values

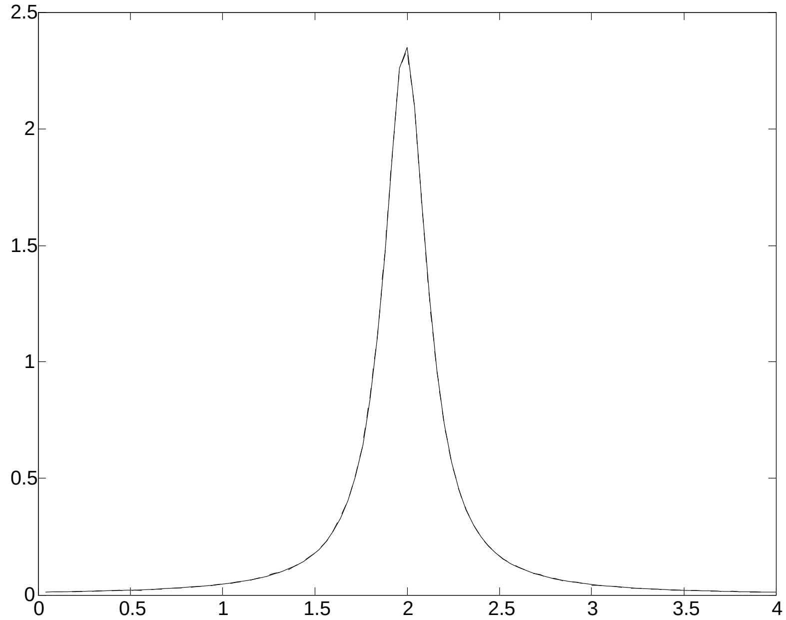


Figure 1: Exact small sample density (-) and sampling density (- -), $\pi_1 = 0$.

of the parameters of the data generating process, strong evidence in support of the construction of small sample densities using the concept of an unique conditional density and of this concept itself. Note also the peculiarity in the densities in case of weak identification which are equal to zero in $\hat{\beta} = 2$ $(= \Omega_{22}^{-1}\omega_{21})$ 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 unique conditional density gives a straightforward explanation is the relative insensitivity of the small sample and sampling densities to adding superfluous instruments. The unique conditional density approach namely shows that the small sample 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 (24). The "$t$-values" of the superfluous instruments are nonsignificant 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 discarding its eigenvector. Since the "$t$-values" of the superfluous instruments are nonsignificant, they will be associated with the smallest singular value and its eigenvector thus has nonzero elements at the positions of the superfluous instruments. When we thus restrict the smallest singular value to zero and discard its eigenvector, we essentially remove the superfluous instruments. As a consequence, the small sample density of the liml estimator is relatively insensitive to adding superfluous instruments. The small sample 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).

### 3.4.5 Small Sample Testing

The small sample density of the liml estimator (36) does not belong to a standard class of densities nor does the joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$ (34). The joint density of $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$ can also

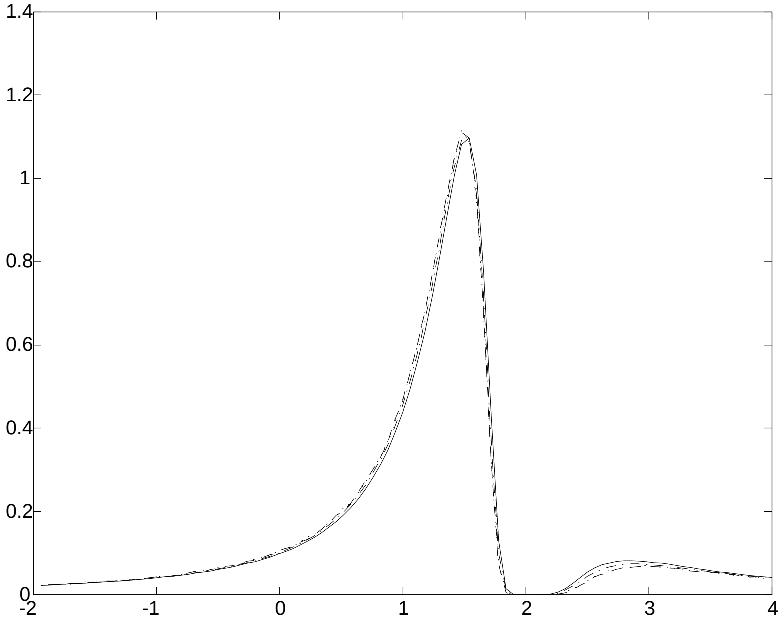


Figure 2: $\pi_1 = 0.1$, $k = 1$: Exact (-) and sampling density (- -); $k = 5$ : exact (-.) and sampling density (..)

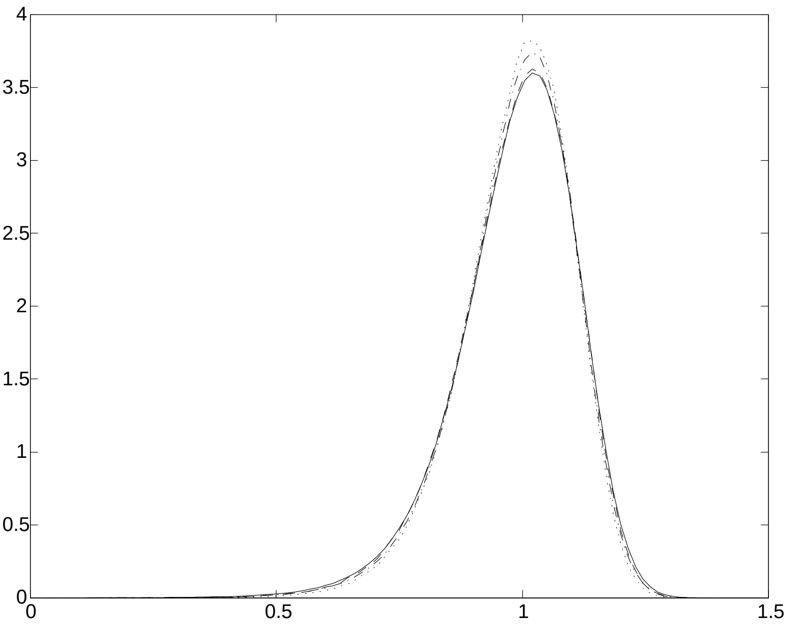


Figure 3: $\pi_1 = 1$, $k = 1$: Exact (-) and sampling density (- -); $k = 5$ : exact (-.) and sampling density (..)

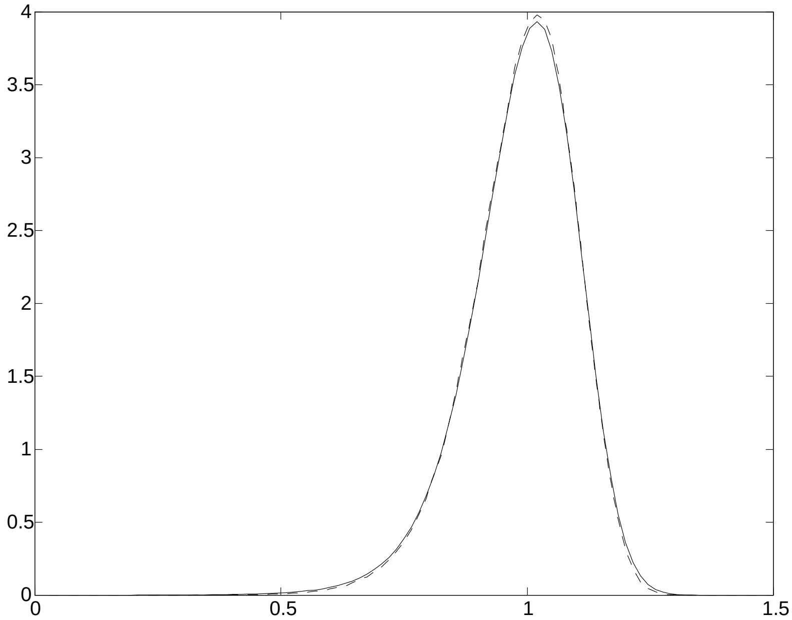


Figure 4: $\pi_1 = 1$, $k = 20$: Exact (-) and sampling density (- -)

not be factorized such that the marginal or conditional density given $\hat{\Lambda}$ of the "$t$-value" of $\hat{\beta}$,

$$\hat{\tau} = \left(\hat{\Pi}' S_0 \hat{\Pi}\right)^{\frac{1}{2}} (\hat{\beta} - \beta_0)\hat{\Lambda}_{11.2}^{-\frac{1}{2}}, \tag{50}$$

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" (50). The factorization is not possible because of the term $|\hat{B}\hat{\Lambda}^{-1}\hat{B}'|^{\frac{1}{2}(k-m+1)}$ in the joint density (34) and consequently the transformation of $(\hat{\beta}, \hat{\Pi}, \hat{\Lambda})$ to $(\hat{\tau}, \hat{\Pi}, \hat{\Lambda})$ leads to a joint density of $(\hat{\tau}, \hat{\Pi}, \hat{\Lambda})$ for which we cannot construct the marginal and conditional density of $\hat{\tau}$ given $\hat{\Lambda}$ analytically. The density of $\hat{\tau}$ can thus only be assessed numerically. The small sample distribution of "$t$" and Wald statistics testing hypotheses on $\beta$ are therefore nonstandard and standard critical values are only asymptotically valid when $\Pi_0$ is a fixed full rank matrix. In case of weak instruments, where $\Pi_0 = \Psi_0/\sqrt{T}$, 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 thus also nonstandard, see *e.g.* Staiger and Stock (1997), Wang and Zivot (1998) and Zivot *et. al.* (1998).

Instead of testing using the "$t$-statistic" $\hat{\tau}$ (50), we can also conduct inference directly using the small sample density of $\hat{\beta}$. The density of $\hat{\beta}$ is less complicated than the density of $\hat{\tau}$ and for example in case that $m = 2$, an analytical expression of the joint density of $(\hat{\beta}, \hat{\Lambda})$ exists which can be used to compute the marginal density, see sections 3.4.1-3.4.4. So, to conduct inference on $\hat{\beta}$ it can be more convenient to use the small sample density of $\hat{\beta}$ itself than to use the small sample density of its "$t$-value". The small sample density of $\hat{\beta}$ (36) involves unobserved parameters, which are such that when we replace these by estimates from the data, *i.e.* $\Omega + \frac{1}{T}B'\pi'X'X\pi B$ by $\frac{1}{T}Y'Y$ and $\Pi_0 B_0$ by $\hat{\Phi}$, that the small sample density is equivalent to the posterior of $\beta$ using the Jeffreys' prior, see Kleibergen and Zivot (1998). Note that when $T$ is large, the approximation of $\Omega + \frac{1}{T}B'\pi'X'X\pi B$ by $\frac{1}{T}Y'Y$ is quite accurate and harmless but one has to be careful in the specification of $\Pi_0 B_0$.

Likelihood ratio and score statistics can also be used to test hypotheses. The small sample distribution of the score statistic can also be constructed using the concept of an unique conditional density, see Kleibergen (1998). For reasons of space we donot elaborate on that here.

# 4 Limiting Distribution of LIML

When we donot know the distribution of the disturbances $V$, we can still use the foc (3) to conduct statistical inference on the liml estimators by using the limiting distribution of the least squares estimator $\hat{\Phi}$,

$$\sqrt{T}\left(\hat{\Phi}-\Phi_0\right) \Rightarrow n\left(0,\Omega_0\otimes S_0^{-1}\right), \tag{51}$$

where $S_0 = p\lim_{T\to\infty}\left(\frac{X'X}{T}\right)$, $\Phi_0=\Pi_0B_0$, such that

$$\sqrt{T}\left(\hat{\Theta}-\Theta_0\right) \Rightarrow n\left(0,I_m\otimes I_k\right), \tag{52}$$

with $\hat{\Theta} = S_0^{\frac{1}{2}}\hat{\Phi}\Omega_0^{-\frac{1}{2}}$, $\Theta_0 = S_0^{\frac{1}{2}}\Phi_0\Omega_0^{-\frac{1}{2}} = S_0^{\frac{1}{2}}\Pi_0B_0\Omega_0^{-\frac{1}{2}}$, and the foc (20). (20) can also be expressed as

$$\lim_{T\to\infty}\left[\sqrt{T}\left(\frac{\partial vec(\Gamma D)}{(\partial vec(\beta)'\ \partial vec(\Pi)')}|_{(\hat{\Omega},\hat{\beta},\hat{\Pi})}\right)' vec\left(\sqrt{T}\left(\hat{\Theta}-\Theta_0\right)-\sqrt{T}\left(\hat{\Gamma}\hat{D}-\Theta_0\right)\right)|_{(\hat{\Omega},\hat{\beta},\hat{\Pi})}\right]=0, \tag{53}$$

such that since $\sqrt{T}\left(\hat{\Theta}-\Theta_0\right)$ converges to a random variable $x$, which is distributed as $n(0,I_m\otimes I_k)$, and (53) holds for all realizations of this random variable $x$, it imposes the restriction,

$$\begin{aligned}\lim_{T\to\infty}\left[\sqrt{T}\left(\hat{\Theta}-\Theta_0\right)-\sqrt{T}\left(\hat{\Gamma}\hat{D}-\Theta_0\right)\right] &\equiv 0 \Leftrightarrow \\ \lim_{T\to\infty}\sqrt{T}\left(\hat{\Gamma}\hat{D}-\Theta_0\right) &\equiv x,\end{aligned} \tag{54}$$

on $x$. Just identical to (24), (54) is a reduced rank restriction as it implies that the rank of the $k\times m$ (random) matrix $x$ is $m-1$. Using the conditional density of $x$ given that $rank(x)=m-1$ and assuming that $\Pi_0$ has full rank, we show in appendix B that the limiting distributions of the liml estimators $(\hat{\beta},\hat{\Pi})$ that result from the conditional density are

$$\begin{aligned}\sqrt{T}\left(\hat{\Pi}-\Pi_0\right) &\Rightarrow n(0,\Omega_{22}\otimes S_0^{-1}), \\ \sqrt{T}\left(\hat{\beta}-\beta_0\right) &\Rightarrow n(0,\sigma_{11.2}\left(\Pi_0'S_0\Pi_0\right)^{-1}),\end{aligned} \tag{55}$$

where $\Omega_0=\begin{pmatrix}\omega_{11} & \omega_{12}\\ \omega_{21} & \Omega_{22}\end{pmatrix}$, $\omega_{11}:1\times1$, $\omega_{12},\omega_{21}':1\times(m-1)$, $\Omega_{22}:(m-1)\times(m-1)$, $\omega_{11.2}=\omega_{11}-\omega_{12}\Omega_{22}^{-1}\omega_{21}=\sigma_{11.2}=\sigma_{11}-\sigma_{12}\sigma_{22}^{-1}\sigma_{21}$, $\Sigma_0=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \Sigma_{22}\end{pmatrix}$, $\sigma_{11}:1\times1$, $\sigma_{12},\sigma_{21}':1\times(m-1)$, $\Sigma_{22}:(m-1)\times(m-1)$, and are thus identical to the limiting distributions discussed elsewhere in the literature, see *e.g.* Hausman (1983), which are, however, derived using a completely different approach. This again shows the applicability of the unique conditional densities.

# 5 Conclusions

We developed a novel approach for constructing small sample and limiting distributions of maximum likelihood estimators that is based on the property of the maximum likelihood estimator that it satisfies the first order condition and the concept of an unique conditional density. We applied this approach to construct the small sample and limiting distribution of the limited information maximum likelihood estimator. The resulting expressions are identical to resp. the sampling density and the expression stated elsewhere in the literature and thus proof the validity of the concept of an unique conditional density and the limiting and small sample distributions that result from it.

As our novel approach differs from the traditional one it can sometimes be more or less convenient to apply than the traditional one. The construction of the small sample and limiting distribution of the limited information maximum likelihood estimator already show this as it is fairly complicated to obtain the limiting distribution using the novel approach compared to the traditional one but the construction of the small sample distribution could be considered as less complicated. An advantage is there that the joint density of all elements of the maximum likelihood estimator is obtained straightforwardly and doesnot involve the noncentral Wishart density which is only needed to integrate out specific parameters. It is therefore interesting to see for which models the unique conditional density approach allows us to construct the small sample distribution of the maximum likelihood estimator and also how it can be used to construct exact test statistics. Both areas are topics for further research.

# Appendix

### A. Small Sample Distribution LIML estimator

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

**1.** Construct joint small sample density of $t-$values least squares estimator and 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 conditional density of $(\hat{\beta}, \hat{\Pi})$ given the covariance matrix estimator.

**4.** Integrate out $\hat{\Pi}$ to obtain small sample density of the liml estimator $\hat{\beta}$ given covariance matrix estimator.

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

**1.** To construct the small sample 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,

$$\hat{\Phi} \sim n(\Phi_0, \Omega_0 \otimes S_0^{-1}),$$

where $\Phi_0 = \Pi_0 B_{0,}$ $B_0 = \begin{pmatrix} \beta_0 & I_{m-1} \end{pmatrix}$, $S_0 = X'X$.

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

$$\hat{\Omega} \sim W(\tfrac{1}{T}\Omega_0, T-k),$$

and is stochastically independent of $\hat{\Phi}$. The expectation of this random variable is $\frac{T-k}{T}\Omega_0$. Because of the expressions of the conditional densities it is convenient to define the random variable $\hat{\Lambda} = \Omega_0 \hat{\Omega}^{-1} \Omega_0$,

$$\hat{\Lambda} \sim iW(T\Omega_0, T-k),$$

since $\hat{\Omega}^{-1} \sim iW(T\Omega_0^{-1}, T-k)$, which has expectation $\frac{T}{T-k-m-1}\Omega_0$, and the density function of $\hat{\Lambda}$ reads,

$$p(\hat{\Lambda}) \propto |\hat{\Lambda}|^{-\frac{1}{2}(T-k+m+1)} \exp\left[-\frac{1}{2} tr\left(\hat{\Lambda}^{-1} T\Omega_0\right)\right],$$

and $\hat{\Lambda}$ is also stochastically independent of $\hat{\Phi}$. The $t$-values of $\hat{\Phi}$ are defined by $\hat{\Theta} = S_0^{\frac{1}{2}} \hat{\Phi} \Omega_0^{-\frac{1}{2}}$, and are distributed as,

$$\hat{\Theta} \sim n(\Theta_0, I_m \otimes I_k),$$

where $\Theta_0 = S_0^{\frac{1}{2}} \Phi_0 \Omega_0^{-\frac{1}{2}}$. The density function of these $t$-values therefore reads,

$$p(\hat{\Theta}) \propto \exp\left[-\frac{1}{2} tr\left(\left(\hat{\Theta} - \Theta_0\right)' \left(\hat{\Theta} - \Theta_0\right)\right)\right].$$

**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 $\hat{\lambda} = 0$, which is proportional to the conditional density of the "$t$-values" given that they have reduced rank,

$$\begin{aligned}
p_r(\hat{\Gamma}, \hat{\delta}|\Omega) &\propto p(\hat{\Theta}|\Omega)|_{rank(\hat{\Theta})=m-1} \\
&\propto p(\hat{\Theta}(\hat{\Gamma}, \hat{\delta}, \hat{\lambda})|\Omega)|_{\hat{\lambda}=0}|J(\hat{\Theta}, (\hat{\Gamma}, \hat{\delta}, \hat{\lambda}))|_{\hat{\lambda}=0}| \\
&\propto \left|\begin{pmatrix} \hat{D}\hat{D}' \otimes I_k & \hat{\delta} \otimes \hat{\Gamma} \\ \hat{\delta}' \otimes \hat{\Gamma}' & \hat{\Gamma}'\hat{\Gamma} \end{pmatrix}\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D} - \Theta_0\right)'\left(\hat{\Gamma}\hat{D} - \Theta_0\right)\right)\right] \\
&\propto \left|\hat{\Gamma}'\hat{\Gamma}\right|^{\frac{1}{2}} \left|(I_{m-1} \otimes I_k) - (\hat{\delta}\hat{\delta}' \otimes M_{\hat{\Gamma}})\right|^{\frac{1}{2}} \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D} - \Phi_0\right)'\left(\hat{\Gamma}\hat{D} - \Phi_0\right)\right)\right] \\
&\propto \left|\hat{\Gamma}'\hat{\Gamma}\right|^{\frac{1}{2}} \left|I_{m-1} + \hat{\delta}\hat{\delta}'\right|^{\frac{1}{2}(k-m+1)} \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Gamma}\hat{D} - \Phi_0\right)'\left(\hat{\Gamma}\hat{D} - \Phi_0\right)\right)\right],
\end{aligned}$$

where $r$ indicates that the density is not the marginal density but the conditional density given that $\hat{\lambda} = 0$.

**3.** The liml estimators can be solved from $\hat{\Gamma}\hat{D}$ by using an estimator for the unknown covariance matrix $\Omega_0$. This estimator is also a random variable and needs to have a mean proportional to $\Omega_0$ to be suitable. Instead of $\hat{\Omega}$ we use $\hat{\Lambda}$ as estimator/random variable to represent $\Omega_0$ as it leads to a more convenient expression of the small sample density of the liml estimator. Because of the rank reduction imposed on $\hat{\Theta}$, we can exactly solve for the liml estimators from $\hat{\Gamma}\hat{D}$ ,

$$\hat{\Gamma}\hat{D} = S_0^{\frac{1}{2}}\hat{\Pi}\hat{B}\hat{\Lambda}^{-\frac{1}{2}} = S_0^{\frac{1}{2}}\hat{\Pi}\hat{B}\Delta_2\begin{pmatrix} \left(\hat{B}\Delta_2\right)^{-1}\hat{B}\Delta_1 & I_{m-1} \end{pmatrix},$$

where $\hat{\Lambda}^{-\frac{1}{2}} = \begin{pmatrix} \Delta_1 & \Delta_2 \end{pmatrix}$ with $\Delta_1$ a $m \times 1$ vector and $\Delta_2$ a $m \times (m-1)$ matrix such that $\hat{\delta} = \left(\hat{B}\Delta_2\right)^{-1}\hat{B}\Delta_1$ and $\hat{\Gamma} = S^{\frac{1}{2}}\hat{\Pi}\hat{B}\Delta_2$.

To construct the Jacobian of the transformation from $(\hat{\Gamma}, \hat{\delta})$ to $(\hat{\Pi}, \hat{\beta})$, $J\left((\hat{\Gamma}, \hat{\delta}), \left(\hat{\Pi}, \hat{\beta}\right)\right)$, $\hat{\delta} = (B\Delta_2)^{-1}B\Delta_1$, $\hat{\Gamma} = S_0^{\frac{1}{2}}\hat{\Pi}\hat{B}\Delta_2$, we use the following results:

$$\begin{aligned}
\frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\beta})'} &= \left(\Delta_1' \otimes \left(\hat{B}\Delta_2\right)^{-1}\right)\frac{\partial vec(\hat{B})}{\partial vec(\hat{\beta})'} \\
&\quad - \left(\Delta_1'\hat{B}' \otimes I_{m-1}\right)\left(\left(\hat{B}\Delta_2\right)^{-1\prime} \otimes \left(\hat{B}\Delta_2\right)^{-1}\right)(\Delta_2' \otimes I_{m-1})\frac{\partial vec(\hat{B})}{\partial vec(\hat{\beta})'} \\
&= \left(\Delta_1'e_1 \otimes \left(\hat{B}\Delta_2\right)^{-1}\right) - \left(\Delta_1'\hat{B}'\left(\hat{B}\Delta_2\right)^{-1\prime}\Delta_2'e_1 \otimes \left(\hat{B}\Delta_2\right)^{-1}\right) \\
&= \left(\Delta_1'\left(I_m - \hat{B}'\left(\hat{B}\Delta_2\right)^{-1\prime}\Delta_2'\right)e_1 \otimes \left(\hat{B}\Delta_2\right)^{-1}\right), \\
\frac{\partial vec(\hat{\Gamma})}{\partial vec(\hat{\Pi})'} &= \left(\Delta_2'\hat{B}' \otimes S_0^{\frac{1}{2}}\right),
\end{aligned}$$

where $e_1$ is the first $m$ dimensional unity vector. Because $\frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\Pi})'} = 0$, the Jacobian $\left| J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi},\hat{\beta}\right)\right)\right|$ then becomes

$$\begin{aligned}
\left| J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi},\hat{\beta}\right)\right)\right| &= \left|\frac{\partial vec(\hat{\delta})}{\partial vec(\hat{\beta})'}\right| \left|\frac{\partial vec(\hat{\Gamma})}{\partial vec(\hat{\Pi})'}\right| \\
&= \left|\Delta_1'\left(I_m - \hat{B}'\left(\hat{B}\Delta_2\right)^{-1\prime}\Delta_2'\right)e_1 \otimes \left(\hat{B}\Delta_2\right)^{-1}\right| \left|\left(\Delta_2'\hat{B}'\otimes S_0^{\frac{1}{2}}\right)\right| \\
&= \left|\hat{B}\Delta_2\right|^{k-1} |S_0|^{\frac{1}{2}(m-1)} \left|\Delta_1'\left(I_m - \hat{B}'\left(\hat{B}\Delta_2\right)^{-1\prime}\Delta_2'\right)e_1\right|^{(m-1)}.
\end{aligned}$$

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

$$\begin{aligned}
p_r(\hat{\Pi},\hat{\beta}|\hat{\Lambda}) &\propto p_r(\hat{\Gamma}(\hat{\Pi},\hat{\beta},\hat{\Lambda}),\hat{\delta}(\hat{\Pi},\hat{\beta},\hat{\Lambda})) \left| J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi},\hat{\beta}\right)\right)\right| \\
&\propto \left|\Delta_2'B'\hat{\Pi}'S_0\hat{\Pi}B\Delta_2\right|^{\frac{1}{2}} \left| I_{m-1} + \left(\hat{B}\Delta_2\right)^{-1}\hat{B}\Delta_1\Delta_1'\hat{B}'\left(\hat{B}\Delta_2\right)^{-1\prime}\right|^{\frac{1}{2}(k-m+1)} \\
&\quad \left| J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi},\hat{\beta}\right)\right)\right| \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)' S_0\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right)\right] \\
&\propto |B\Delta_2|^{-(k-m)} \left|\hat{\Pi}'S_0\hat{\Pi}\right|^{\frac{1}{2}} \left|\hat{B}\Delta_2\Delta_2'\hat{B}' + \hat{B}\Delta_1\Delta_1'\hat{B}'\right|^{\frac{1}{2}(k-m+1)} \\
&\quad \left| J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi},\hat{\beta}\right)\right)\right| \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)' S_0\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right)\right] \\
&\propto |B\Delta_2|^{-(k-m)} \left|\hat{\Pi}'S_0\hat{\Pi}\right|^{\frac{1}{2}} \left|B\hat{\Lambda}^{-1}B'\right|^{\frac{1}{2}(k-m+1)} \\
&\quad \left| J\left((\hat{\Gamma},\hat{\delta}),\left(\hat{\Pi},\hat{\beta}\right)\right)\right| \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)' S_0\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right)\right],
\end{aligned}$$

since $\Delta_2\Delta_2' + \Delta_1\Delta_1' = \hat{\Lambda}^{-1}$ and $\hat{B}\Delta_2$ is a square matrix.

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

$$\begin{aligned}
\left|\Delta_2'\hat{B}'\right| \left|\Delta_1'\left(I_m - B'\left(\Delta_2'B'\right)^{-1}\Delta_2'\right)e_1\right| &= \left|\begin{pmatrix} \Delta_1'e_1 & \Delta_1'B' \\ \Delta_2'e_1 & \Delta_2'B' \end{pmatrix}\right| \\
= \left|\begin{pmatrix} \Delta_1 & \Delta_2 \end{pmatrix}'\begin{pmatrix} e_1 & \hat{B}' \end{pmatrix}\right| &= \left|\begin{pmatrix} \Delta_1 & \Delta_2 \end{pmatrix}\right| = |\hat{\Lambda}|^{-\frac{1}{2}},
\end{aligned}$$

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

The density $p_r(\hat{\Pi},\hat{\beta}|\hat{\Lambda})$ then becomes,

$$\begin{aligned} p_r(\hat{\Pi},\hat{\beta}|\hat{\Lambda}) &\propto \left|\hat{B}\Delta_2\right|^{-(k-m)} \left|\hat{\Pi}'S_0\hat{\Pi}\right|^{\frac{1}{2}} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)} \\ &\quad \left|\hat{B}\Delta_2\right|^{k-1} |S_0|^{\frac{1}{2}(m-1)} \left|\Delta_1'\left(I_m - \hat{B}'\left(\hat{B}\Delta_2\right)^{-1\prime}\Delta_2'\right)e_1\right|^{(m-1)} \\ &\quad \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)'S_0\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right)\right] \\ &\propto |\hat{\Lambda}|^{-\frac{1}{2}(m-1)} \left|\hat{\Pi}'S_0\hat{\Pi}\right|^{\frac{1}{2}} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)} |S_0|^{\frac{1}{2}(m-1)} \\ &\quad \exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)'S_0\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right)\right]. \end{aligned}$$

**4.** To construct the small sample density of $\hat{\beta}$ given $\hat{\Lambda}$, we first decompose the trace component of the conditional density $p_r(\hat{\Pi},\hat{\beta}|\hat{\Lambda})$ as,

$$\begin{aligned} & tr\left(\hat{\Lambda}^{-1}\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)'S_0\left(\hat{\Pi}\hat{B}-\Pi_0B_0\right)\right) \\ = \; & tr\left(\hat{\Lambda}^{-1}\left(\hat{B}'\hat{\Pi}'S_0\hat{\Pi}\hat{B}-\left(\hat{B}'\hat{\Pi}'S_0\Pi_0B_0+B_0'\Pi_0'S_0\hat{\Pi}\hat{B}\right)+B_0'\Pi_0'S_0\Pi_0B_0\right)\right) \\ = \; & tr\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\left(\hat{\Pi}-\Psi_0\right)'S_0\left(\hat{\Pi}-\Psi_0\right)\right) \\ & +tr\left(\left(\hat{\Lambda}^{-1}-\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}\right)B_0'\Pi_0'S_0\Pi_0B_0\right), \end{aligned}$$

where $\Psi_0 = \Pi_0B_0\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}$. To obtain the conditional density of $\hat{\beta}$ given $\hat{\Lambda}$, we construct the integral of the joint density of $\hat{\Pi}$ and $\hat{\beta}$ over $\hat{\Pi}$,

$$\begin{aligned} p_r(\hat{\beta}|\hat{\Lambda}) &\propto |\hat{\Lambda}|^{-\frac{1}{2}(m-1)} \left|\hat{\Pi}'S_0\hat{\Pi}\right|^{\frac{1}{2}} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)} |S_0|^{\frac{1}{2}(m-1)} \\ &\quad \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Lambda}^{-1}-\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}\right)B_0'\Pi_0'S_0\Pi_0B_0\right)\right] \\ &\quad \int\left|\hat{\Pi}'S_0\hat{\Pi}\right|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\left(\hat{\Pi}-\Psi_0\right)'S_0\left(\hat{\Pi}-\Psi_0\right)\right)\right]d\hat{\Pi} \\ &\propto |\hat{\Lambda}|^{-\frac{1}{2}(m-1)} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{\frac{1}{2}(k-m+1)} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{-\frac{1}{2}(k+1)} \\ &\quad \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Lambda}^{-1}-\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}\right)B_0'\Pi_0'S_0\Pi_0B_0\right)\right] \\ &\quad \int\left|\hat{\Upsilon}'\hat{\Upsilon}\right|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr\left(\left(\hat{\Upsilon}-\Upsilon_0\right)'\left(\hat{\Upsilon}-\Upsilon_0\right)\right)\right]d\hat{\Upsilon} \\ &\propto |\hat{\Lambda}|^{-\frac{1}{2}(m-1)} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{-\frac{1}{2}m} \\ &\quad \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Lambda}^{-1}-\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}\right)B_0'\Pi_0'S_0\Pi_0B_0\right)\right] \\ &\quad \int\left|\hat{\Upsilon}'\hat{\Upsilon}\right|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr\left(\left(\hat{\Upsilon}-\Upsilon_0\right)'\left(\hat{\Upsilon}-\Upsilon_0\right)\right)\right]d\hat{\Upsilon} \end{aligned}$$

where $\hat{\Upsilon} = S_0^{\frac{1}{2}} \hat{\Pi} \left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{\frac{1}{2}}$, $\Upsilon_0 = S_0^{\frac{1}{2}} \Psi_0 \left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{\frac{1}{2}} = S_0^{\frac{1}{2}} \Pi_0 B_0 \hat{\Lambda}^{-1} \hat{B}' \left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-\frac{1}{2}}$, since $|J(\hat{\Pi}, \hat{\Upsilon})| = |S_0|^{-\frac{1}{2}(m-1)} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{-\frac{1}{2}k}$ and $\left|\hat{\Pi}' S_0 \hat{\Pi}\right|^{\frac{1}{2}} = \left|\hat{\Upsilon}'\hat{\Upsilon}\right|^{\frac{1}{2}} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{-\frac{1}{2}}$. The integral in the above expression is a noncentral 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 $\hat{\Upsilon} \sim n(\Upsilon_0, I_k)$, it holds that $w = \hat{\Upsilon}'\hat{\Upsilon} \sim \chi^2(k, \mu)$, where $\mu = \Upsilon_0'\Upsilon_0$ is the noncentrality parameter of the noncentral $\chi^2$ distribution and $k$ the degrees of freedom parameter. The density function of a noncentral $\chi^2$ 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 $w$ is $\chi^2(k+2j)$ distributed 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 noncentral $\chi^2$ distribution therefore reads,

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

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

$$\begin{aligned} & \int |\hat{\Upsilon}'\hat{\Upsilon}|^{\frac{1}{2}} \exp\left[-\frac{1}{2} tr\left(\left(\hat{\Upsilon} - \Upsilon_0\right)'\left(\hat{\Upsilon} - \Upsilon_0\right)\right)\right] d\hat{\Upsilon} \\ \propto\ & E_{\chi^2(k,\mu)}\left[w^{\frac{1}{2}}\right] \\ \propto\ & \sum_{j=0}^{\infty} \left( \frac{\left(\frac{1}{2}\Upsilon_0'\Upsilon_0\right)^j}{j!} \exp\left[-\frac{1}{2}\Upsilon_0'\Upsilon_0\right] 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \right) \\ \propto\ & \sum_{j=0}^{\infty} \left( \frac{\left(\frac{1}{2}\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}B_0'\Pi_0'S_0\Pi_0B_0\right)^j}{j!} \right. \\ & \left. \exp\left[-\frac{1}{2}\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}B_0'\Pi_0'S_0\Pi_0B_0\right] 2^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))} \right), \end{aligned}$$

such that the conditional density of $\hat{\beta}$ given $\hat{\Lambda}$ reads,

$$\begin{aligned}
& p_r(\hat{\beta}|\hat{\Lambda}) \\
\propto\ & \left|\hat{\Lambda}\right|^{-\frac{1}{2}(m-1)} \left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{-\frac{1}{2}m} \\
& \exp\left[-\frac{1}{2}tr\left(\left(\hat{\Lambda}^{-1}-\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}\right)B_0'\Pi_0'S_0\Pi_0B_0\right)\right] \\
& \left[\sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}B_0'\Pi_0'S_0\Pi_0B_0\right)^j}{j!}\right.\right. \\
& \left.\left.\exp\left[-\frac{1}{2}\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}B_0'\Pi_0'S_0\Pi_0B_0\right]2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right)\right] \\
\propto\ & \left|\hat{\Lambda}\right|^{-\frac{1}{2}(m-1)}\exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}B_0'\Pi_0'S_0\Pi_0B_0\right)\right]\left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right|^{-\frac{1}{2}m} \\
& \left[\sum_{j=0}^{\infty}\left(\frac{\left(\frac{1}{2}\hat{\Lambda}^{-1}\hat{B}'\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)^{-1}\hat{B}\hat{\Lambda}^{-1}B_0'\Pi_0'S_0\Pi_0B_0\right)^j}{j!}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right)\right] \\
\propto\ & \left|\hat{\Lambda}\right|^{-\frac{1}{2}(m-1)}\exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}B_0'\Pi_0'S_0\Pi_0B_0\right)\right]\left|\hat{\Lambda}_{22}^{-1}+\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\
& \left[\sum_{j=0}^{\infty}\left(\left(\frac{\left(B_0\hat{\Lambda}^{-1}\hat{B}'\right)^2\Pi_0'S_0\Pi_0}{2\left(\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right)}\right)^j\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right]
\end{aligned}$$

$$\begin{aligned}
\propto\ & \left|\hat{\Lambda}\right|^{-\frac{1}{2}(m-1)}\exp\left[-\frac{1}{2}tr\left(\hat{\Lambda}^{-1}B_0'\Pi_0'S_0\Pi_0B_0\right)\right]\left|\hat{\Lambda}_{22}^{-1}+\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\
& \left[\sum_{j=0}^{\infty}\left(\left(\frac{\left|\hat{\Lambda}_{22}^{-1}+\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\beta_0'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}'\right)\right|^2\Pi_0'S_0\Pi_0}{2\left|\hat{\Lambda}_{22}^{-1}+\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}'\right)'\hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}'\right)\right|}\right)^j\right.\right. \\
& \left.\left.\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right]
\end{aligned}$$

since

$$\begin{aligned}
\left|\hat{B}\hat{\Lambda}^{-1}\hat{B}'\right| &= |\hat{\Lambda}_{22}^{-1}+(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}')'\hat{\Lambda}_{11.2}^{-1}(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}')|, \\
\text{and } \left|B_0\hat{\Lambda}^{-1}B'\right| &= \left|\hat{\Lambda}_{22}^{-1}+\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\beta_0'\right)'\ \hat{\Lambda}_{11.2}^{-1}\left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1}-\hat{\beta}'\right)\right|.
\end{aligned}$$

As $\hat{\Lambda}$ is stochastically independent of $\hat{\Theta}$, the joint density of $(\hat{\beta}, \hat{\Lambda})$ thus becomes,

$$\begin{aligned}
& p_r(\hat{\beta}, \hat{\Lambda}) \\
\propto\ & p(\hat{\Lambda}) p_r(\hat{\beta}|\hat{\Lambda}) \\
\propto\ & |\hat{\Lambda}|^{-\frac{1}{2}(T-k+m+1)} \exp\left[-\frac{1}{2} tr\left(\hat{\Lambda}^{-1} T\Omega_0\right)\right] \\
& \left|\hat{\Lambda}\right|^{-\frac{1}{2}(m-1)} \exp\left[-\frac{1}{2} tr\left(\hat{\Lambda}^{-1} B_0' \Pi_0' S_0 \Pi_0 B_0\right)\right] \left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\
& \left[\sum_{j=0}^{\infty} \left(\left(\frac{\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \beta_0'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^2 \Pi_0' S_0 \Pi_0}{2\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|}\right)^j \right.\right. \\
& \left.\left. \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right],
\end{aligned}$$

such that

$$\begin{aligned}
p_r(\hat{\beta}|\hat{\Lambda}) \ \propto\ & \left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^{-\frac{1}{2}m} \\
& \left[\sum_{j=0}^{\infty} \left(\left(\frac{\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \beta_0'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|^2 \Pi_0' S_0 \Pi_0}{2\left|\hat{\Lambda}_{22}^{-1} + \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)' \hat{\Lambda}_{11.2}^{-1} \left(\hat{\Lambda}_{12}\hat{\Lambda}_{22}^{-1} - \hat{\beta}'\right)\right|}\right)^j \right.\right. \\
& \left.\left. \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right)\right] \\
p_r(\hat{\Lambda}) \ \propto\ & |\hat{\Lambda}|^{-\frac{1}{2}(T-k+2m)} \exp\left[-\frac{1}{2} tr\left(\hat{\Lambda}^{-1}\left(T\Omega_0 + B_0'\Pi_0' S_0 \Pi_0 B_0\right)\right)\right],
\end{aligned}$$

which shows that $\hat{\Lambda} \sim iW(T\Omega_0 + B_0'\Pi_0' S_0 \Pi_0 B_0, T-k+m-1)$ and has thus changed because of the rank reduction imposed on $\hat{\Theta}$ and the dependence of $(\hat{\beta}, \hat{\Pi})$ on $\hat{\Lambda}$. Note that the mean of this inverted-Wishart density is $\frac{1}{T-k+m-1}(T\Omega_0 + B_0'\Pi_0' S_0 \Pi_0 B_0) \approx \Omega_0 + \frac{1}{T} B_0'\Pi_0' S_0 \Pi_0 B_0$ and that this density is centered close around its mean for reasonably large values of $T$ ($T > 25$). A good approximation of the marginal density of $\hat{\beta}$ is therefore $p_r(\hat{\beta}|\hat{\Lambda} = \Omega_0 + \frac{1}{T} B_0'\Pi_0' S_0 \Pi_0 B_0)$. This is also the reason why we use $\hat{\Lambda}$ instead of $\hat{\Omega}$ since the resulting marginal density of $\hat{\Omega}$ does not belong to a standard class while the density of $\hat{\Lambda}$ does such that we can use its properties, like the mean as shown above.

When $\Pi_0 = 0$, the conditional density of $\hat{\beta}$ simplifies to,

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

which is a Cauchy density. Another simplification occurs when $\beta_0 = \hat{\Lambda}_{22}^{-1}\hat{\Lambda}_{21}$ as in that case the term $B_0\hat{\Lambda}^{-1}\hat{B}'$ is equal to $\hat{\Lambda}_{22}^{-1}$ and $p_r(\hat{\beta}|\hat{\Lambda})$ is a symmetric density then.

We note that the density $p_r(\hat{\beta}|\hat{\Lambda})$ has a simpler functional form than the density derived in Mariano and Sawa (1971), which involves a triplicate infinite series whereas the $p_r(\hat{\beta}|\hat{\Lambda})$

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 $\hat{\Lambda}$. We do know the marginal density of $\hat{\Lambda}$, however, such that we can use its properties, like the fact that it quickly concentrates around its (known) mean when the number of observations increases or how to simulate from it, to conduct marginal inference on $\hat{\beta}$.

### B. Limiting Distributions LIML estimator

The limiting distributions of the liml estimator result from,

$$\sqrt{T}\left(\hat{\Gamma}\hat{D}-\Theta_0\right)\Rightarrow x|_{rank(x)=m-1},$$

where $x\sim n(0,I_m\otimes I_k)$, $\hat{\Gamma}\hat{D}=S_0^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega_0^{-\frac{1}{2}}$, $\Theta_0=S_0^{\frac{1}{2}}\Pi_0 B_0\Omega_0^{-\frac{1}{2}}$, $S_0=p\lim_{T\to\infty}\left(\frac{X'X}{T}\right)$. The conditional density function of $x$ given that $rank(x)=m-1$, is obtained by specifying $x$ as,

$$x=zw+z_\perp\lambda w_\perp,$$

where $z:k\times(m-1)$, $z_\perp:k\times(k-m+1)$, $z_\perp'z_\perp=I_{k-m+1}$, $z_\perp'z=0$, $w:(m-1)\times m$, $w_\perp:1\times m$, $w=\begin{pmatrix}A & I_{m-1}\end{pmatrix}$, $A:(m-1)\times 1$, $w_\perp w_\perp'=1$, $ww_\perp'=0$, such that $x|_{rank(x)=m-1}=zw$. This conditional density then reads,

$$\begin{aligned}
p_r(z,A) &\propto p(x)|_{rank(x)=m-1}\\
&\propto p(x(z,A,\lambda))|_{\lambda=0}|J(x,(z,A,\lambda))|_{\lambda=0}|\\
&\propto \left|\begin{pmatrix} ww'\otimes I_k & A\otimes z\\ A'\otimes z' & z'z\end{pmatrix}\right|^{\frac{1}{2}}\exp\left[-\frac{1}{2}tr\left(\Omega_0^{-1}w'z'zw\right)\right]\\
&\propto |z'z|^{\frac{1}{2}}\left|I_{m-1}+AA'\right|^{\frac{1}{2}(k-m+1)}\exp\left[-\frac{1}{2}tr\left(\Omega_0^{-1}w'z'zw\right)\right].
\end{aligned}$$

Since,

$$\begin{aligned}
\sqrt{T}\left(\hat{\Theta}-\Theta_0\right) &=\\
\sqrt{T}S_0^{\frac{1}{2}}\left(\hat{\Pi}B-\Pi_0B_0\right)\Omega_0^{-\frac{1}{2}} &=\\
\sqrt{T}S_0^{\frac{1}{2}}\begin{pmatrix}\left(\hat{\Pi}\hat{\beta}-\Pi_0\beta_0\right) & \left(\hat{\Pi}-\Pi_0\right)\end{pmatrix}\Omega_0^{-\frac{1}{2}} &\Rightarrow \begin{pmatrix} zA & z\end{pmatrix}\Leftrightarrow\\
\sqrt{T}\begin{pmatrix}\left(\hat{\Pi}\hat{\beta}-\Pi_0\beta_0\right) & \left(\hat{\Pi}-\Pi_0\right)\end{pmatrix} &\Rightarrow S_0^{-\frac{1}{2}}zw\Omega_0^{\frac{1}{2}}\Leftrightarrow\\
\sqrt{T}\begin{pmatrix}\left(\hat{\Pi}\hat{\beta}-\Pi_0\beta_0\right) & \left(\hat{\Pi}-\Pi_0\right)\end{pmatrix} &\Rightarrow S_0^{-\frac{1}{2}}zw\Omega_2\begin{pmatrix}(w\Omega_2)^{-1}w\omega_1 & I_{m-1}\end{pmatrix},
\end{aligned}$$

where $\Omega_0^{\frac{1}{2}}=\begin{pmatrix}\omega_1 & \Omega_2\end{pmatrix}$, $\omega_1:m\times 1$, $\Omega_2:m\times(m-1)$. It thus holds that,

$$\sqrt{T}\left(\hat{\Pi}-\Pi_0\right)\Rightarrow S_0^{-\frac{1}{2}}zw\Omega_2,$$

and

$$\begin{aligned}
\sqrt{T}\left(\hat{\Pi}\hat{\beta}-\Pi_0\beta_0\right) &\Rightarrow S_0^{-\frac{1}{2}} zw\Omega_2\left(w\Omega_2\right)^{-1} w\omega_1 \Leftrightarrow \\
\sqrt{T}\left(\hat{\Pi}\left(\hat{\beta}-\beta_0\right)+\left(\hat{\Pi}-\Pi_0\right)\beta_0\right) &\Rightarrow S_0^{-\frac{1}{2}} zw\Omega_2\left(w\Omega_2\right)^{-1} w\omega_1 \Leftrightarrow \\
\sqrt{T}\hat{\Pi}\left(\hat{\beta}-\beta_0\right) &\Rightarrow S_0^{-\frac{1}{2}} zw\Omega_2\left(\left(w\Omega_2\right)^{-1} w\omega_1-\beta_0\right) \Leftrightarrow \\
\sqrt{T}\Pi_0\left(\hat{\beta}-\beta_0\right) &\Rightarrow S_0^{-\frac{1}{2}} zw\Omega_2\left(\left(w\Omega_2\right)^{-1} w\omega_1-\beta_0\right) \Leftrightarrow \\
\sqrt{T}\left(\hat{\beta}-\beta_0\right) &\Rightarrow \left(\Pi_0' S\Pi_0\right)^{-1}\Pi_0' S^{\frac{1}{2}} zw\Omega_2\left(\left(w\Omega_2\right)^{-1} w\omega_1-\beta_0\right).
\end{aligned}$$

Note that we use the property that $\Pi_0$ has full rank here. Similar to appendix A, we now define $\Gamma = zw\Omega_2$ and $\psi = (w\Omega_2)^{-1} w\omega_1 - \beta_0$ such that

$$\begin{aligned}
\sqrt{T}\left(\hat{\Pi}-\Pi_0\right) &\Rightarrow S_0^{-\frac{1}{2}}\Gamma, \\
\sqrt{T}\left(\hat{\beta}-\beta_0\right) &\Rightarrow \left(\Pi_0' S\Pi_0\right)^{-1}\Pi_0 S^{\frac{1}{2}}\Gamma\psi,
\end{aligned}$$

and

$$zw = \Gamma\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}\Omega_0^{-\frac{1}{2}}.$$

The derivatives of this transformation read (we construct the inverse of the jacobian here),

$$\begin{aligned}
\frac{\partial vec(\psi)}{\partial vec(A)'} &= \left(\omega_1' \otimes (w\Omega_2)^{-1}\right)\frac{\partial vec(w)}{\partial vec(A)'} \\
&\quad -\left(\omega_1' w' \otimes I_{m-1}\right)\left((w\Omega_2)^{-1\prime} \otimes (w\Omega_2)^{-1}\right)\left(\Omega_2' \otimes I_{m-1}\right)\frac{\partial vec(w)}{\partial vec(A)'} \\
&= \left(\omega_1' e_1 \otimes (w\Omega_2)^{-1}\right) - \left(\omega_1' z' (w\Omega_2)^{-1\prime}\Omega_2' e_1 \otimes (w\Omega_2)^{-1}\right) \\
&= \left(\omega_1'\left(I_m - w'(w\Omega_2)^{-1\prime}\Omega_2'\right)e_1 \otimes (w\Omega_2)^{-1}\right) \\
\frac{\partial vec(\Gamma)}{\partial vec(z)'} &= \left((w\Omega_2)' \otimes I_k\right),
\end{aligned}$$

and we note that

$$\begin{aligned}
\left|(w\Omega_2)'\right|\left|\omega_1'\left(I_m - w'(w\Omega_2)^{-1\prime}\Omega_2'\right)e_1\right| &= \left|\begin{pmatrix} \omega_1' e_1 & \omega_1' w' \\ \Omega_2' e_1 & \Omega_2' w' \end{pmatrix}\right| \\
&= \left|\begin{pmatrix} \omega_1 & \Omega_2 \end{pmatrix}'\begin{pmatrix} e_1 & w' \end{pmatrix}\right| \\
&= \left|\begin{pmatrix} \omega_1 & \Omega_2 \end{pmatrix}\right| = |\Omega_0|^{\frac{1}{2}},
\end{aligned}$$

such that

$$\begin{aligned}
|J((\Gamma,\psi),(z,A))| &= \left|\left(\omega_1'\left(I_m - w'(w\Omega_2)^{-1\prime}\Omega_2'\right)e_1 \otimes (w\Omega_2)^{-1}\right)\right| \\
&\quad \left|\left((w\Omega_2)' \otimes I_k\right)\right| \\
&= |w\Omega_2|^{k-m}|\Omega_0|^{\frac{1}{2}(m-1)}.
\end{aligned}$$

The joint density of $(\Gamma, \psi)$ then reads,

$$\begin{aligned}
p_r(\Gamma,\psi) &\propto p_r(z(\Gamma,\psi), A(\Gamma,\psi)) \left|J\left((\Gamma,\psi),(z,A)\right)\right|^{-1} \\
&\propto |w\Omega_2|^{-(k-m)} |z'z|^{\frac{1}{2}} |ww'|^{\frac{1}{2}(k-m+1)} \exp\left[-\frac{1}{2}tr\left(\Omega_0^{-1}w'z'zw\right)\right] \\
&\propto |w\Omega_2|^{-(k-m)} \left|(w\Omega_2)^{-1\prime}\,\Gamma'\Gamma\,(w\Omega_2)^{-1}\right|^{\frac{1}{2}} \\
&\quad \left|w\Omega_2 \begin{pmatrix} (w\Omega_2)^{-1}w\omega_1 & I_{m-1} \end{pmatrix} \Omega_0^{-1} \begin{pmatrix} (w\Omega_2)^{-1}w\omega_1 & I_{m-1} \end{pmatrix}' w\Omega_2\right|^{\frac{1}{2}(k-m+1)} \\
&\quad \exp\left[-\frac{1}{2}tr\left(\Omega^{-1}\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}'\Gamma'\Gamma\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}\right)\right] \\
&\propto |\Gamma'\Gamma|^{\frac{1}{2}} \left|\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}\Omega_0^{-1}\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}'\right|^{\frac{1}{2}(k-m+1)} \\
&\quad \exp\left[-\frac{1}{2}tr\left(\Omega_0^{-1}\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}'\Gamma'\Gamma\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}\right)\right]
\end{aligned}$$

We now use that $\Omega_0 = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix}$, such that

$$\begin{aligned}
& \begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}\Omega_0^{-1}\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}' \\
=\;& \Omega_{22}^{-1} + \left(\psi - \left(\Omega_{22}^{-1}\omega_{21} - \beta_0\right)\right)\omega_{11.2}^{-1}\left(\psi - \left(\Omega_{22}^{-1}\omega_{21} - \beta_0\right)\right)' \\
=\;& \Omega_{22}^{-\frac{1}{2}}(I_{m-1} + (\delta-\delta_0)(\delta-\delta_0)')\Omega_{22}^{-\frac{1}{2}},
\end{aligned}$$

and that

$$\begin{aligned}
& tr\left(\Omega_0^{-1}\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}'\Gamma'\Gamma\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}\right) \\
=\;& tr\left(\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}\Omega_0^{-1}\begin{pmatrix} \psi+\beta_0 & I_{m-1} \end{pmatrix}'\Gamma'\Gamma\right) \\
=\;& tr(\Omega_{22}^{-\frac{1}{2}}(I_{m-1} + (\delta-\delta_0)'(\delta-\delta_0))\Omega_{22}^{-\frac{1}{2}}\Gamma'\Gamma) \\
=\;& tr((I_{m-1} + (\delta-\delta_0)'(\delta-\delta_0))\Upsilon'\Upsilon) \\
=\;& tr\left(\Upsilon'\Upsilon\right) + tr\left((\delta-\delta_0)'\,\Upsilon'\Upsilon\,(\delta-\delta_0)\right),
\end{aligned}$$

by performing another transformation of the variables to $\Upsilon = \Gamma\Omega_{22}^{-\frac{1}{2}}$, $\delta = \Omega_{22}^{\frac{1}{2}}\psi\omega_{11.2}^{-\frac{1}{2}}$, $\delta_0 = \Omega_{22}^{\frac{1}{2}}\left(\Omega_{22}^{-1}\omega_{21} - \beta_0\right)\omega_{11.2}^{-\frac{1}{2}}$, such that

$$\begin{aligned}
p_r(\delta,\Upsilon) &\propto p_r(\Gamma(\delta,\Upsilon), \psi(\delta,\Upsilon)) \left|J\left((\Gamma,\psi),(\delta,\Upsilon)\right)\right| \\
&\propto |\Upsilon'\Upsilon|^{\frac{1}{2}} \left|I_{m-1} + (\delta-\delta_0)(\delta-\delta_0)'\right|^{\frac{1}{2}(k-m+1)} \\
&\quad \exp\left[-\frac{1}{2}\left(tr\left(\Upsilon'\Upsilon\right) + tr\left((\delta-\delta_0)'\,\Upsilon'\Upsilon\,(\delta-\delta_0)\right)\right)\right].
\end{aligned}$$

Using the results from section A, it then follows that

$$p_r(\delta) \propto \left|1 + (\delta-\delta_0)'(\delta-\delta_0)\right|^{-\frac{1}{2}m},$$

because $\Upsilon'\Upsilon$ has a Wishart distribution with $k$ degrees of freedom such that, using the Wishart integration step,

$$\begin{aligned}
& \int |\Upsilon'\Upsilon|^{\frac{1}{2}} \exp\left[-\frac{1}{2}tr(\Upsilon'\Upsilon(I_{m-1}+(\delta-\delta_0)(\delta-\delta_0)')\right] d\Upsilon \\
\propto & \int |\Upsilon'\Upsilon|^{\frac{1}{2}(k+1-(m-1)-1)} \exp\left[-\frac{1}{2}tr(\Upsilon'\Upsilon(I_{m-1}+(\delta-\delta_0)(\delta-\delta_0)')\right] d\Upsilon'\Upsilon \\
\propto & \left|I_{m-1}+(\delta-\delta_0)(\delta-\delta_0)'\right|^{-\frac{1}{2}(k+1)} \\
\propto & \left|1+(\delta-\delta_0)'(\delta-\delta_0)\right|^{-\frac{1}{2}(k+1)},
\end{aligned}$$

which is a Cauchy density with location $\delta_0$ and identity scale matrices. A Cauchy random variable $x_c : (m-1)\times 1$, of this type can be defined as, see Zellner (1971),

$$\begin{aligned}
x_c &= x_n(y_n^2)^{-\frac{1}{2}}+\delta_0 \\
\text{or } x_c &= (Y_n'Y_n)^{-\frac{1}{2}}x_n+\delta_0,
\end{aligned}$$

where $x_n : (m-1)\times 1$, $Y_n : (m-1)\times(m-1)$, $y_n : 1\times 1$, are all independent standard normal random variables with mean zero and identity covariance matrices. In case of the latter definition, the joint density of $Y_n$ and $x_c$ reads,

$$p(x_c,Y_n) \propto |Y_n'Y_n|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left[tr\left(Y_n'Y_n\right)+tr\left((x_c-\delta_0)'Y_n'Y_n(x_c-\delta_0)\right)\right]\right].$$

This definition is closely related to the joint density of $(\delta,\Upsilon)$ shown above. This shows that a Cauchy random variable can also be defined as,

$$x_c=(Z_n'Z_n)^{-\frac{1}{2}}w+\delta_0,$$

where $w : (m-1)\times 1$, $Z_n : k\times(m-1)$ is a standard normal random matrix with mean zero and identity covariance matrix and the joint density of $(w,Z_n)$ reads,

$$p(w,Z_n) \propto \left|1+w'\left(Z_n'Z_n\right)^{-1}w\right|^{\frac{1}{2}(k-m+1)} \exp\left[-\frac{1}{2}\left[tr\left(Z_n'Z_n\right)+tr\left(w'w\right)\right]\right].$$

The joint density of $x_c$ and $Z_n$ then reads,

$$\begin{aligned}
p(x_c,Z_n) \propto\ & |Z_n'Z_n|^{\frac{1}{2}}\left|1+(x_c-\delta_0)'(x_c-\delta_0)\right|^{\frac{1}{2}(k-m+1)} \\
& \exp\left[-\frac{1}{2}\left[tr\left(Z_n'Z_n\right)+tr\left((x_c-\delta_0)'Z_n'Z_n(x_c-\delta_0)\right)\right]\right],
\end{aligned}$$

which is identical to the joint density of $(\delta,\Upsilon)$ shown above and thus leads to a marginal density of $x_c$ which is Cauchy. So, when

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

and

$$\begin{aligned}
p(x_c|Z_n) \propto\ & |Z_n'Z_n|^{\frac{1}{2}}\left|1+(x_c-\delta_0)'(x_c-\delta_0)\right|^{\frac{1}{2}(k-m+1)} \\
& \exp\left[-\frac{1}{2}tr\left((x_c-\delta_0)'Z_n'Z_n(x_c-\delta_0)\right)\right],
\end{aligned}$$

then

$$p(x_c) \propto \left|1 + (x_c - \delta_0)' (x_c - \delta_0)\right|^{-\frac{1}{2}m},$$

and vice versa. The marginal distribution of $\Upsilon$ is as a consequence standard normal with mean zero and identity covariance matrix. The limiting behavior of the estimators now results as

$$\begin{aligned} \sqrt{T}\left(\hat{\Pi} - \Pi_0\right) &\Rightarrow S_0^{-\frac{1}{2}} \Upsilon \Omega_{22}^{\frac{1}{2}} \sim n(0, \Omega_{22} \otimes S_0^{-1}), \\ \sqrt{T}\left(\hat{\beta} - \beta_0\right) &\Rightarrow (\Pi_0' S_0 \Pi_0)^{-1} \Pi_0' S_0^{\frac{1}{2}} \Upsilon \delta \omega_{11.2}^{\frac{1}{2}}. \end{aligned}$$

The random variable $\Pi_0' S_0^{\frac{1}{2}} \Upsilon \Omega_{22}^{\frac{1}{2}} \sim n(0, \Omega_{22} \otimes \Pi_0' S_0 \Pi_0)$ and $Y_n = (\Pi_0' S_0 \Pi_0)^{-\frac{1}{2}} \Pi_0' S_0^{\frac{1}{2}} \Upsilon = (\Pi_0' S_0 \Pi_0)^{-\frac{1}{2}} \Pi_0' S_0^{\frac{1}{2}} \Upsilon \Omega_{22}^{\frac{1}{2}} \Omega_{22}^{-\frac{1}{2}} \sim n(0, I_{m-1} \otimes I_{m-1})$. We can now define $\delta$ as,

$$\delta = (Y_n' Y_n)^{-\frac{1}{2}} x_n + \delta_0,$$

where $x_n \sim n(0, I_{m-1})$, such that

$$\Pi_0' S^{\frac{1}{2}} \Upsilon \delta \omega_{11.2}^{\frac{1}{2}} = (\Pi_0' S_0 \Pi_0)^{\frac{1}{2}} Y_n \left((Y_n' Y_n)^{-\frac{1}{2}} x_n + \delta_0\right) \omega_{11.2}^{\frac{1}{2}}.$$

So, conditional on $Y_n$, $\delta$ has a normal distribution with mean $\delta_0$ and covariance matrix $(Y_n' Y_n)^{-1}$, $\delta | Y_n) \sim n(\delta_0, (Y_n' Y_n)^{-1})$. Given $Y_n$, the limiting behavior of $\hat{\beta}$ can thus be characterized by,

$$\sqrt{T}(\hat{\beta} - \beta_0 | Y_n) \Rightarrow (\Pi_0' S_0 \Pi_0)^{-\frac{1}{2}} Y_n \left[n(\delta_0, (Y_n' Y_n)^{-1})\right] \omega_{11.2}^{\frac{1}{2}}.$$

The limiting behavior marginal with respect to $Y_n$ is obtained by taking the expectation with respect to the standard normal variable $Y_n$,

$$\begin{aligned} \sqrt{T}(\hat{\beta} - \beta_0) &\Rightarrow E_{Y_n}((\Pi_0' S_0 \Pi_0)^{-\frac{1}{2}} Y_n \left[n(\delta_0, (Y_n' Y_n)^{-1})\right] \omega_{11.2}^{\frac{1}{2}}) \\ &\Rightarrow n(E_{Y_n}((\Pi_0' S_0 \Pi_0)^{-\frac{1}{2}} Y_n \delta_0 \omega_{11.2}^{\frac{1}{2}}, \\ &\quad \omega_{11.2} E_{Y_n}((\Pi_0' S_0 \Pi_0)^{-\frac{1}{2}} Y_n Y_n^{-1} Y_n^{-1\prime} Y_n' (\Pi_0' S_0 \Pi_0)^{-\frac{1}{2}})), \\ &\Rightarrow n(0, \omega_{11.2} (\Pi_0' S_0 \Pi_0)^{-1}), \end{aligned}$$

since $Y_n$ is a square matrix, and because $\omega_{11.2} = \sigma_{11.2}$, this expression can also be specified as

$$\sqrt{T}(\hat{\beta} - \beta_0) \Rightarrow n(0, \sigma_{11.2} (\Pi_0' S_0 \Pi_0)^{-1}),$$

which is the well known expression for the limiting behavior of the liml estimator. Note that when $\Pi_0 = 0$, $\hat{\beta}$ converges to a random variable, see section 3.4.3 and Phillips (1989).

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

[7] Drèze, J.H. and J.F. Richard. Bayesian Analysis of Simultaneous Equations systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).

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

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

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

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

[12] 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).

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

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

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

[16] Kleibergen, F. Conditional Densities in Econometrics. Unpublished manuscript, Econometric Institute, Erasmus University Rotterdam, 1998.

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

[18] Kleibergen, F. and H.K. van Dijk. Bayesian Simultaneous Equation Analysis using Reduced Rank Structures. *Econometric Theory*, **14**, 1998. forthcoming.

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

[20] 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).

[21] Kolmogorov, A.N. *Foundations of the Theory of Probability.* Chelsea (New York), (1950).

[22] 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.

[23] 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).

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

[25] 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.

[26] 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.

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

[28] Poirier, D.J. *Intermediate Statistics and Econometrics: A Comparative Approach.* MIT Press, (Cambridge, MA), (1995).

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

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

[31] Wang, J. and E. Zivot. Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments. *Econometrica*, Forthcoming, 1998.

[32] Wolpert, R.J. Comment on: Inference for a Deterministic Population Dynamics Model for Bowhead Whales, by, A.E. Raftery, G.H. Rivens and J.E. Zeh. *Journal of the American Statistical Association*, **90**:426–427, (1995).

[33] Zellner, A. *An Introduction to Bayesian Inference in Econometrics.* Wiley (New York), (1971).

[34] Zivot, E., R. Startz, and C. R. Nelson. Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review*, Forthcoming, 1998.