UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Orthogonal statistics and the density of the liml estimator

Kleibergen, F.

Publication date
2003

Link to publication

Citation for published version (APA):
Kleibergen, F. (2003). *Orthogonal statistics and the density of the liml estimator*. (UvA Econometrics Discussion Paper; No. 2003/11). Department of Quantitative Economics. http://www1.feb.uva.nl/pp/bin/477fulltext.pdf

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

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:06 Sept 2026

UvA Econometrics

Discussion Paper: 2003/11

Orthogonal statistics and the density of the liml estimator

Frank Kleibergen

www.fee.uva.nl/ke/UvA-Econometrics

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA Universiteit van Amsterdam

Orthogonal Statistics and the Density of the LIML estimator

Frank Kleibergen[‡]

Brown University and University of Amsterdam

Abstract

We show that orthogonalization is helpful for constructing densities of maximum likelihood estimators. We therefore use an orthogonal specification of the reduced form of the instrumental variables regression model to obtain an approximation of the density of the limited information maximum likelihood estimator. The approximation consists of a single infinite sum and is less involved than the expression of the true density. In comparisons with the sampling density the approximation is shown to be accurate indicating the validity of its construction.

1 Introduction

Orthogonal parameters have information matrices that are block-diagonal. Orthogonalizing parameters therefore removes dependence between their estimators, see *e.g.* Cox and Reid (1987). The resulting orthogonal estimators converge more rapidly and under less restrictive assumptions to their limiting distributions. Despite its appealing properties, orthogonalization is not commonly applied since it is often difficult to obtain an orthogonal specification of the econometric model of interest. Orthogonal specifications do, however, exist for some commonly used econometric models. Lancaster (2002) develops orthogonal specifications for panel data models. Kleibergen (2002) and Moreira (2003) obtain orthogonal statistics for the linear instrumental variables (IV) regression model. Their statistics converge to a pre-specified limiting distribution under a much wider set of conditions than the statistics that have been used traditionally. This further illustrates the power of orthogonalization.

Alongside its importance for the convergence of statistics that test hypothezes, as in *e.g.* Kleibergen (2002) and Moreira (2003), we show that orthogonalization can be helpful for constructing densities of maximum likelihood estimators (mles). We therefore use an orthogonal

*I thank the editor, the co-editor and two anonymous referees for helpful comments and suggestions. The research documented in this article is partly funded by the NWO Vernieuwingsimpuls research grant "Empirical Comparison of Economic Models".

†Department of Economics, Brown University, Box B, Providence, RI 02912, United States, Email: Frank_Kleibergen@brown.edu, and Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Homepage: http://www.fee.uva.nl/ke/kleibergen.htm

specification of the instrumental variables (IV) regression model to obtain an expression for the density of the limited information maximum likelihood (liml) estimator. Mariano and Sawa (1972) construct the marginal density of the liml estimator by integrating over the mles of the remaining (covariance) parameters. The orthogonal statistics simplify the integration because integrating over an orthogonal statistic is identical to conditioning on a zero value of it with respect to the construction of the density of the liml estimator. Since conditioning is less involved then integration, we obtain a less complicated expression for the density of the liml estimator than Mariano and Sawa (1972). The orthogonal specification is, however, for the reduced form specification of the IV regression model. The extension to the structural form makes the density an approximation of the density of the liml estimator. The approximation is exact in case of a reasonably large number of observations. The approximate expression of the density of the liml estimator is therefore comparable to that of the limlk estimator of Anderson *et. al.* (1982). It does have a less involved expression since it consists of a single infinite sum while the density of the limlk estimator consists of a double infinite sum. We show that our expression for the density of the liml estimator is an accurate approximation of the sampling density of the liml estimator for a number of data generating processes that cover the cases of invalid, weak and strong instruments. We also use the expression of the density of the liml estimator to illustrate its sensitivity to the quality of the instruments.

The paper is organized as follows. In the second section, we discuss the IV regression model and its orthogonal statistics. In the third section, we show that the information matrix of the orthogonal parameters is block-diagonal and obtain the approximation of the density of the liml estimator. In the fourth section, we illustrate the sensitivity to the quality of the instruments. The fifth section concludes.

Throughout the paper: $\text{vec}(A)$ stands for the column vectorization of the matrix A such that, when $A = (a_1 \cdots a_N)$, $\text{vec}(A) = (a_1' \cdots a_N')'$, I_T is the $T \times T$ identity matrix and $e_{i,m}$ is the i-th column of I_m, $P_X = X(X'X)^{-1}X'$ and $M_X = I_T - P_X$ for a full rank $T \times k$ dimensional matrix X.

2 Instrumental Variables Regression Model

The IV regression model in structural form can be represented as a limited information simultaneous equation model, see *e.g.* Hausman (1983),

$$\begin{array}{rcl} y = & X\beta + \varepsilon \\ X = & Z\Pi + V, \end{array} \tag{1}$$

where y is a $T \times 1$ vector of endogenous variables and Z is a $T \times k$ matrix of pre-determined exogenous variables (or instruments). The $T \times m$ matrix X may contain endogenous as well as exogenous explanatory variables. The latter are assumed to be columns of Z as well and $m \leq k$. If interest is restricted to those elements of the parameter vector β that relate to the endogenous explanatory variables, the instruments in Z might be separated into two parts. For expository purposes we only consider the case that X just consists of endogenous variables. We assume the rows of $(\varepsilon \; V)$ to be independently normally distributed with zero mean; we denote

the $(m+1)\times(m+1)$ covariance matrix by Σ,

$$\Sigma = \text{var}\begin{pmatrix} \varepsilon_t \\ V_t \end{pmatrix} = \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon V} \\ \Sigma_{V\varepsilon} & \Sigma_{VV} \end{pmatrix}. \tag{2}$$

Substituting the reduced form equation for X into the structural equation for y gives the non-linearly restricted reduced form specification

$$(Y\ X) = Z\Pi B + (u\ V), \tag{3}$$

where $B = (\beta\ I_m)$, $u = \varepsilon + V\beta$ and, hence, $(u\ V)$ has covariance matrix

$$\Omega = \text{var}\begin{pmatrix} u_t \\ V_t \end{pmatrix} = \begin{pmatrix} \omega_{uu} & \Omega_{uV} \\ \Omega_{Vu} & \Omega_{VV} \end{pmatrix} = \begin{pmatrix} e'_{1,m+1} \\ B \end{pmatrix}' \Sigma \begin{pmatrix} e'_{1,m+1} \\ B \end{pmatrix}. \tag{4}$$

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

$$(y\ X) = Z\Phi + (u\ V), \tag{5}$$

where $\Phi : k\times(m+1)$. It is assumed that $k \geq m$ so that the structural parameter vector β is identified by the order condition. The degree of over-identification is equal to $k-m$. The parameter β is identified if and only if $\text{rank}(\Pi) = m$.

Orthogonal Specification The parameter matrix ΠB of the restricted reduced form has a rank value equal to m while the rank value of Φ equals $m+1$. The IV regression model is therefore a reduced rank regression model, see *e.g.* Reinsel and Velu (1998). The rank value of a matrix can be reflected by a singular value decomposition (SVD), see *e.g.* Golub and van Loan (1989). We use the SVD to decompose the matrix of "t-values", $\hat{\Theta} = (Z'Z)^{\frac{1}{2}}\hat{\Phi}\Omega^{-\frac{1}{2}}$, of the least squares estimator of Φ, $\hat{\Phi} = (Z'Z)^{-1}Z'(y\ X)$, into two orthogonal components of which one is just a function of the mles of β and Π.

Theorem 1 *Given Ω, the matrix of t-values $\hat{\Theta}$ has the orthogonal specification*

$$\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_\perp\hat{\lambda}\hat{D}_\perp, \tag{6}$$

where $\hat{\Gamma}$ is a unrestricted $k\times m$ matrix, $\hat{\lambda}$ is a unrestricted $(k-m)\times 1$ vector and $\hat{D}$ is a $m\times(m+1)$ matrix that is such that $\hat{D} = (\hat{\delta}\ I_m)$ for a $m\times 1$ vector $\hat{\delta}$. The $k\times(k-m)$ and $1\times(m+1)$ matrices $\hat{\Gamma}_\perp$ and $\hat{D}_\perp$ are such that $\hat{\Gamma}'\hat{\Gamma}_\perp \equiv 0$, $\hat{\Gamma}'_\perp\hat{\Gamma}_\perp \equiv I_{k-m}$ and $\hat{D}\hat{D}'_\perp \equiv 0$, $\hat{D}_\perp\hat{D}'_\perp \equiv 1$. The relationship between $\hat{\Theta}$ and $(\hat{\Gamma},\hat{\delta},\hat{\lambda})$ is invertible because $\hat{\Gamma}_\perp$ and $\hat{D}_\perp$ are functions of just $\hat{\Gamma}$ and $\hat{\delta}$ resp..

When $\hat{\lambda}$ equals zero, we obtain the mles of the parameters of the restricted reduced form so

$$\hat{\Gamma}\hat{D} = (Z'Z)^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega^{-\frac{1}{2}} \tag{7}$$

with $\hat{B} = (\hat{\beta}\ I_{m-1})$ and $\hat{\beta}$, $\hat{\Pi}$ are the mles of β, Π resp.. Hence, $\hat{\Gamma} = (Z'Z)^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega_V$ and $\hat{\delta} = (\hat{B}\Omega_V)^{-1}\hat{B}\omega_u$, with $\Omega^{-\frac{1}{2}} = (\omega_u\ \Omega_V)$ and ω_u and Ω_V a $(m+1)\times 1$ vector and $(m+1)\times m$ matrix.

Proof. see Appendix A. ■

The orthogonal specification from Theorem 1 is motivated by the manner in which the liml estimator is obtained from the concentrated log-likelihood, see *e.g.* Hausman (1983),

$$\ln \mathcal{L}(\beta|y, X, Z) = \tfrac{1}{2} T \log |1 - \eta| , \tag{8}$$

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

$$\hat{\eta} = \min_{\beta} \left[\frac{(y-X\beta)' P_Z (y-X\beta)}{(y-X\beta)'(y-X\beta)} \right], \tag{9}$$

which is identical to solving the eigenvalue problem,

$$\begin{array}{l} |\eta (y\ X)'(y\ X) - (y\ X)' P_Z (y\ X)| = 0 \\ |\eta (\hat{\Omega} + \frac{1}{T-k}(y\ X)' P_Z (y\ X)) - \frac{1}{T-k}(y\ X)' P_Z (y\ X)| = 0. \end{array} \quad \Leftrightarrow \tag{10}$$

with $\hat{\Omega} = \frac{1}{T-k}(y\ X)' M_Z (y\ X)$. The liml estimator results from the eigenvector associated with the smallest eigenvalue since this (orthonormal) eigenvector equals $(1 \vdots -\hat{\beta}')'(1 + \hat{\beta}'\hat{\beta})^{-\frac{1}{2}}$. Anderson *et. al.* (1983) analyze the distribution of the limlk estimator that is obtained when we replace $\hat{\Omega} + \frac{1}{T-k}(y\ X)' P_Z (y\ X)$ by its expectation $\Omega + B'\Pi' \left(\frac{Z'Z}{T-k}\right) \Pi B$. The specification of $\hat{\Theta}$ in Theorem 1 results when use Ω instead of $\hat{\Omega} + \frac{1}{T-k}(y\ X)' P_Z (y\ X)$ since $\hat{\Theta}'\hat{\Theta} = \Omega^{-\frac{1}{2}\prime}(y\ X)' P_Z (y\ X)\Omega^{-\frac{1}{2}}$. The involved covariance matrices therefore differ. This results since the restricted reduced form and structural form are only transformations of one another in case of an unknown covariance matrix. Equation (10) is based on the structural form while Theorem 1 is based on the restricted reduced form with an observed covariance matrix. We show lateron that the density of the mle that results from Theorem 1 approximates the density of the liml estimator when we use a covariance matrix estimator for Ω in Theorem 1. For expository purposes, we refrain from doing so from the outset.

3 Density of the MLE

We use the orthogonal specification from Theorem 1 to obtain the density of the mle of β. Initially we construct the conditional density of the mle given a known value of Ω. To determine the appropriate value of Ω that we should use in this conditional density, we introduce a covariance matrix estimator for Ω. With this value of Ω, the resulting density function accurately approximates the sampling density of the liml estimator. An extensive literature exists about the density of the liml estimator, see *e.g.* Mariano and Sawa (1972), Phillips (1983) and Anderson *et. al.* (1982). We obtain an expression for the density of the liml estimator in an alternative manner by using properties that result from the orthogonal specification in Theorem 1. The expression of the density of the liml estimator that we obtain is less complicated than the expression of the density of the liml estimator obtained by Mariano and Sawa (1972) and that of the density of the limlk estimator from Anderson *et. al.* (1982). Our construction of the density of the liml estimator exploits the orthogonality of $(\hat{\beta}, \hat{\Pi})$ and $\hat{\lambda}$ which is reflected in their block diagonal information matrix.

Theorem 2 *The information matrix of* (β, Π) *and* λ *given* Ω *evaluated at* $(\hat{\beta}, \hat{\Pi}, \hat{\lambda})$ *is block-diagonal for every value of* $(\hat{\beta}, \hat{\Pi})$ *and* $\hat{\lambda}$.

Proof. see Appendix B. ■

Theorem 2 shows that $(\hat{\beta}, \hat{\Pi})$ and $\hat{\lambda}$ are globally orthogonal as defined in *e.g.* Cox and Reid (1987). The global orthogonality results because $(\frac{\partial \text{vec}(\hat{\Theta})}{\partial \hat{\delta}'} \vdots \frac{\partial \text{vec}(\hat{\Theta})}{\partial \text{vec}(\hat{\Gamma})'})$ is orthogonal to $\frac{\partial \text{vec}(\hat{\Theta})}{\partial \hat{\lambda}'}$, where $(\hat{\delta}, \hat{\Gamma})$ is, given Ω, an invertible transformation of $(\hat{\beta}, \hat{\Pi})$. Cox and Reid (1987) show that the marginal density of a statistic is accurately approximated by its conditional density given the globally orthogonal statistic. This reasoning remains valid here because we can solve for $(\hat{\delta}, \hat{\Gamma})$ from $\hat{\Theta}$ without the involvement of $\hat{\lambda}$. Since

$$\begin{aligned}&\left(\begin{array}{cc} \frac{\partial \text{vec}(\hat{\Theta})}{\partial \hat{\delta}'} & \frac{\partial \text{vec}(\hat{\Theta})}{\partial \text{vec}(\hat{\Gamma})'} \end{array} \right)' \text{vec}(\hat{\Theta}) = \Bigg((e_{1,m+1} \otimes \hat{\Gamma}) + (I_{m+1} \otimes \hat{\Gamma}_\perp \hat{\lambda}) \frac{\partial \text{vec}(\hat{D}_\perp)}{\partial \text{vec}(\hat{\delta})'} \vdots \\ &(\hat{D}' \otimes I_k) + (\hat{D}_\perp' \hat{\lambda}' \otimes I_k) \frac{\partial \text{vec}(\hat{\Gamma}_\perp)}{\partial \text{vec}(\hat{\Gamma})'} \Bigg)' \left[\left(\hat{D}' \otimes \hat{\Gamma} \right) \text{vec}(I_m) + \left(\hat{D}_\perp' \otimes \hat{\Gamma}_\perp \right) \hat{\lambda} \right] = \left(\begin{array}{c} \hat{\Gamma}' \hat{\Gamma} \hat{\delta} \\ \text{vec}(\hat{\Gamma} \hat{D} \hat{D}') \end{array} \right),\end{aligned} \tag{11}$$

we obtain $(\hat{\delta}, \hat{\Gamma})$ from $\hat{\Theta}$ using the $m + km$ equations

$$\left(\begin{array}{c} (e_{1,m+1} \otimes \hat{\Gamma})' \text{vec}(\hat{\Theta}) \\ (\hat{D}' \otimes I_k)' \text{vec}(\hat{\Theta}) \end{array} \right) = \left(\begin{array}{c} \hat{\Gamma}' \hat{\Gamma} \hat{\delta} \\ \text{vec}(\hat{\Gamma} \hat{D} \hat{D}') \end{array} \right) \Leftrightarrow \hat{\delta} = (\hat{\Gamma}' \hat{\Gamma})^{-1} \hat{\Gamma}' \hat{\Theta} e_{1,m+1} \text{ and } \hat{\Gamma} = \hat{\Theta} \hat{D}' (\hat{D} \hat{D}')^{-1}. \tag{12}$$

Hence, because all elements that contain $\hat{\lambda}$ cancel out because of the prevailing orthogonality relationships, $(\hat{\delta}, \hat{\Gamma})$ results from those values of $\hat{\Theta}$ for which the implicit value of $\hat{\lambda}$ is equal to zero $\hat{\Theta}|_{\hat{\lambda}=0}$. The set of values of $\hat{\Theta}$ for which $\hat{\lambda} = 0$ spans a $(k+1)m$-dimensional manifold in the space of $\hat{\Theta}$ which is the $\mathbb{R}^{k(m+1)}$. Every value of $(\hat{\delta}, \hat{\Gamma})$ implies a unique value on this $(k+1)m$-dimensional manifold and so the relationship between $(\hat{\delta}, \hat{\Gamma})$ and $\hat{\Theta}|_{\hat{\lambda}=0}$ is invertible. The marginal density of $(\hat{\delta}, \hat{\Gamma})$ therefore not only results from integrating over $\hat{\lambda}$, see *e.g.* Hillier and Armstrong (1998) and Tjur (1980), but also from the conditional density of $(\hat{\delta}, \hat{\Gamma})$ given that $\hat{\lambda} = 0$.

Theorem 3 *The marginal density of* $(\hat{\delta}, \hat{\Gamma})$ *equals the conditional density of* $(\hat{\delta}, \hat{\Gamma})$ *given that* $\hat{\lambda} = 0$ *and reads*

$$\begin{aligned} p(\hat{\delta}, \hat{\Gamma}) \propto & \; p(\hat{\delta}, \hat{\Gamma}, \hat{\lambda})|_{\hat{\lambda}=0} \\ \propto & \; |\hat{\Gamma}' \hat{\Gamma}|^{\frac{1}{2}} |1 + \hat{\delta}' \hat{\delta}|^{\frac{1}{2}(k-m)} \exp \left[-\tfrac{1}{2} tr \left((\hat{\Gamma} \hat{D} - \Gamma D)' (\hat{\Gamma} \hat{D} - \Gamma D) \right) \right]. \end{aligned} \tag{13}$$

Proof. see Appendix C. ■

When we transform $(\hat{\delta}, \hat{\Gamma})$ to $(\hat{\beta}, \hat{\Pi})$ according to the specification in Theorem 1, we obtain the marginal density of $(\hat{\beta}, \hat{\Pi})$ for a known value of Ω which we state in Theorem 4. In case β is a scalar, we can obtain a closed form expression for the marginal density of $\hat{\beta}$ for a known value of Ω which we also state in Theorem 4.

Theorem 4 ***a.*** *The marginal density of* $(\hat{\beta}, \hat{\Pi})$ *given a known value of* Ω *reads*

$$p(\hat{\Pi}, \hat{\beta}) \propto \; |\hat{B} \Omega^{-1} \hat{B}|^{\frac{1}{2}(k-m)} |\hat{\Pi}' Z' Z \hat{\Pi}|^{\frac{1}{2}} \exp[-\tfrac{1}{2} tr(\Omega^{-1} \{ (\hat{\Pi} \hat{B} - \Pi B)' Z' Z (\hat{\Pi} \hat{B} - \Pi B) \})]. \tag{14}$$

b. *When $m = 1$, the marginal density of $\hat{\beta}$ given a known value of Ω reads*

$$\begin{aligned} p(\hat{\beta}) \propto & \exp\left[-\tfrac{1}{2}tr(\Omega^{-1}B'\Pi'X'X\Pi B)\right]\left|\hat{B}\Omega^{-1}\hat{B}'\right|^{-\frac{1}{2}(m+1)} \\ & \textstyle\sum_{j=0}^{\infty}\left(\frac{1}{j!}\left(\frac{(B\Omega^{-1}\hat{B}')^2\Pi'X'X\Pi}{2\hat{B}\Omega^{-1}\hat{B}'}\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right) \\ \propto & \exp\left[-\tfrac{1}{2}tr\left(\Omega^{-1}B'\Pi'X'X\Pi B\right)\right]\left|\Omega_{VV}^{-1}+\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)'\Omega_{uu.V}^{-1}\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)\right|^{-\frac{1}{2}(m+1)} \\ & \textstyle\sum_{j=0}^{\infty}\left[\left(\frac{\left|\Omega_{VV}^{-1}+\left(\Omega_{uV}\Omega_{VV}^{-1}-\beta'\right)'\Omega_{uu.V}^{-1}\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)\right|^2\Pi'X'X\Pi}{2\left|\Omega_{VV}^{-1}+\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)'\Omega_{uu.V}^{-1}\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)\right|}\right)^j\frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right] \end{aligned} \tag{15}$$

with $\Omega_{uu.V} = \Omega_{uu} - \Omega_{uV}\Omega_{VV}^{-1}\Omega_{Vu}$.

Proof. see Appendix D. ■

The joint density of $\hat{\beta}$ and $\hat{\Pi}$ is such that neither of their conditional densities belongs to a standard class of densities. We can therefore only construct a closed form expression for the marginal density of $\hat{\beta}$ when $m = 2$. For higher dimensional cases, we can just express the marginal density of $\hat{\beta}$ as a function of zonal polynomials, see Muirhead (1982). The marginal density of $\hat{\beta}$ (15) is such that it has, in line with the densities of the liml and limlk estimators, Cauchy tails and therefore infinite mean and variance. The functional form of the density of $\hat{\beta}$ consists of a single infinite sum so it is less complicated than the expressions of the liml estimator which consists of a triple infinite sum, see Mariano and Sawa (1972), and that of the limlk estimator which consists of a double infinite sum, see Anderson *et. al.* (1982). The densities in Theorem 4 closely resemble the posteriors of (β, Π) and β in Bayesian analysis using the Jeffreys' prior, see Chao and Phillips (2003) and Kleibergen and Zivot (2002).

The marginal density of $\hat{\beta}$ (15) is based on a known value of the reduced form covariance matrix Ω. The reduced form is equivalent to the structural form when the covariance matrix is unknown. In order to compare the marginal density of $\hat{\beta}$ with that of liml and limlk, that are based on the structural form, we therefore have to use an estimator of the unknown covariance matrix. When we analyze the joint distribution of $\hat{\beta}$ and the covariance matrix estimator, we can determine the value of Ω that we should use in the marginal density of $\hat{\beta}$ (15) in order to compare it with the density functions of liml and limlk. This value of Ω is stated in Theorem 5.

Theorem 5 *When m=1 and we use a covariance matrix estimator for Ω, the marginal density of $\hat{\beta}$ is, for large values of T, $T > 25$, equal to*

$$p(\hat{\beta}) \propto \left|\hat{B}\hat{\Delta}^{-1}\hat{B}'\right|^{-\frac{1}{2}(m+1)}\textstyle\sum_{j=0}^{\infty}\left(\frac{1}{j!}\left(\frac{(B\hat{\Delta}^{-1}\hat{B}')^2\Pi'X'X\Pi}{2\hat{B}\hat{\Delta}^{-1}\hat{B}'}\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right), \tag{16}$$

with $\hat{\Delta} = \Omega + B'\Pi'\frac{X'X}{T-k}\Pi B$.

Proof. see Appendix E. ■

The marginal density of $\hat{\beta}$ in Theorem 5 is, identical to the density of the limlk estimator, an approximation of the density of the liml estimator. The approximation errors converge to zero

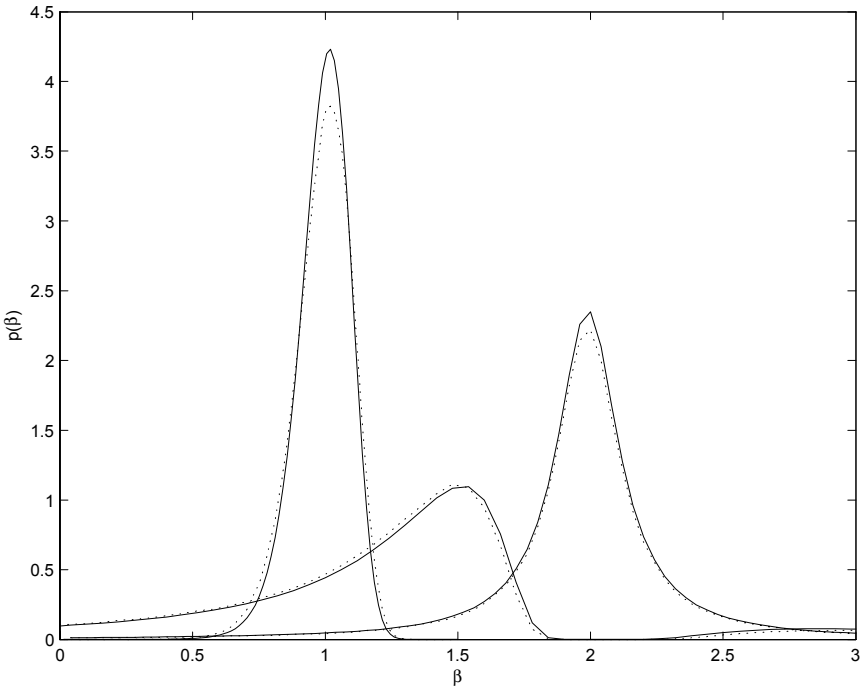

Figure 1: Sampling density of liml estimator (dotted line) and its approximation by (16) (solid line) for $\pi_{11} = 1$ (left), 0.1 (middle) and 0 (right), $\Sigma = \left(\begin{smallmatrix} 1 & 0.99 \\ 0.99 & 1 \end{smallmatrix}\right)$, $T = 100$, $\beta = 1$, $k = 5$ and $Z \sim N(0, I_k \otimes I_T)$.

for large values of T regardless of the value of the other parameters, *i.e.* Π and Ω. The marginal density of $\hat{\beta}$ (16) is a Cauchy density when Π equals zero. The tail behavior of (16) remains as that of a Cauchy density for nonzero values of Π so it has no finite mean or variance.

To illustrate the adequacy of the approximation of the density of the liml estimator by (16) and therefore the validity of the manner in which it is constructed, we compare it with the sampling density that results from one million realizations of the liml estimator. Figures 1 and 2 show the sampling density of the liml estimator and its approximation by (16) for different values of Π and Ω. The densities in both figures are for values of T, k and β equal to 100, 5 and 1 resp.. All densities use the same $T \times k$ dimensional matrix of instruments Z that is generated once from a $N(0, I_k \otimes I_T)$ distribution. The structural form covariance matrix Σ equals $\left(\begin{smallmatrix} 1 & 0.99 \\ 0.99 & 1 \end{smallmatrix}\right)$ in Figure 1 and $\left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$ in Figure 2. We specify Π such that only the first element of Π, π_{11}, can be different from zero. We vary the value of π_{11} to determine if the quality of the approximation depends on the value of Π. Figures 1 and 2 show that the approximation of the density of the liml estimator by (16) is accurate. The approximation reflects the bimodality of the sampling density of the liml estimator and also captures the change of location of the density when the value of π_{11} decreases.

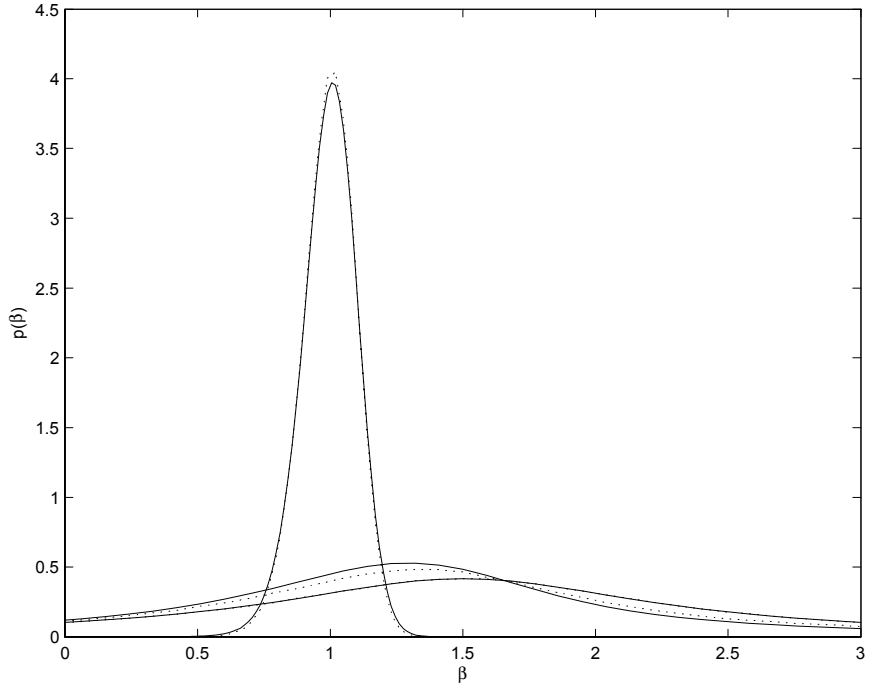

Figure 2: Sampling density of liml estimator (dotted line) and its approximation by (16) (solid line) for $\pi_{11} = 1$ (left), 0.1 (middle) and 0 (right), $\Sigma = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$, $T = 100$, $\beta = 1$, $k = 5$ and $Z \sim N(0, I_k \otimes I_T)$.

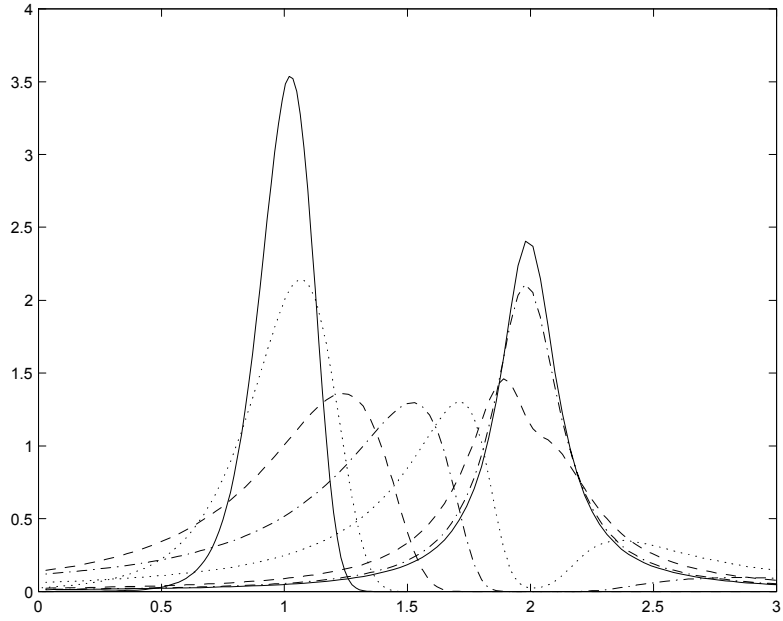

Figure 3: Density $\hat{\beta}$, $\pi_1 = 1$ (-), 0.5 (..), 0.25 (-), 0.1 (-.), 0.05 (..), 0.02 (-), 0.01 (-.), 0 (-).

4 Deteriorating quality of the instruments

We use the approximation of the density of the liml estimator (16) to illustrate the consequences of weak instruments, see *e.g.* Staiger and Stock (1997). We therefore compute the approximation of the density of the liml estimator under a deteriorating quality of the instruments. For this purpose, we use the model

$$\begin{array}{c} y = X\beta + \varepsilon \\ X = Z\pi + v \end{array} \tag{17}$$

where y, $X : T \times 1$, $Z : T \times k$, $(\varepsilon\ v) \sim N(0, \Sigma \otimes I_T)$; $Z \sim N(0, I_k \otimes I_T)$, $T = 100$, $k = 4$, $\pi : k \times 1$, $\pi = (\pi_1 ... \pi_k)'$, $\pi_2 = ... = \pi_k = 0$, $\beta = 1$, $\Sigma = \left(\begin{smallmatrix} 1 & 0.99 \\ 0.99 & 1 \end{smallmatrix}\right)$; for a sequence of values of π_1. For fixed values of β, Z, Σ, $\pi_2, ..., \pi_k$, and a sequence of values of π_1, we compute the approximation of the density of $\hat{\beta}$ (16). The previous section shows that the approximation of the density of the liml estimator by (16) is an accurate one.

Figure 3 shows the densities of the liml estimator $\hat{\beta}$ under a deteriorating quality of the instruments. We achieve this by letting π_1 converge to zero in a sequence of steps, π_1 =1, 0.5, 0.25, 0.1, 0.05, 0.02, 0.01 and 0. The resulting densities show that the shape of the density changes when the instrument quality deteriorates. First, in case of valid instruments the densities are unimodal, then they become bimodal when instruments are weak, to become unimodal again when the instruments are invalid. The expression of (16) indicates that when the quality of the instruments deteriorates:

1. $\Omega + \frac{1}{T} B'\Pi' X'X\Pi B$ converges to Ω such that $\hat{\Delta}_{uV}\hat{\Delta}_{VV}^{-1}$ converges to $(\Omega)_{uV}(\Omega)_{VV}^{-1} = 1.99$.
2. $\Pi' X'X\Pi$ converges to zero so $\frac{(B\hat{\Delta}^{-1}\hat{B}')^2\Pi' X'X\Pi}{\hat{B}\hat{\Delta}^{-1}\hat{B}'}$ becomes independent of $\hat{\beta}$ and converges to zero.

The above two phenomena imply that for a sequence of decreasing values of π_{11} eventually the density of the liml estimator becomes bimodal. For still smaller values of π_1, the location of the two modes both converge to the covariance of the reduced form errors.

5 Conclusions

We show that, alongside its importance for the convergence of tests statistics, orthogonalization can be convenient for the construction of densities of maximum likelihood estimators. For this purpose we propose an orthogonal specification of the instrumental variables regression model. This orthogonal specification allows us to obtain an approximation of the density of the limited information maximum likelihood estimator via a conditioning argument instead of integration. Since conditioning is less involved than integration the resulting approximation is also less involved than the expression of the density of the limited information maximum likelihood estimator of Mariano and Sawa (1972). A comparison of the approximation with the sampling density of the maximum likelihood estimator over a range of data generating processes shows that the approximation is accurate indicating the appropriateness of its derivation.

In order to obtain the density of the maximum likelihood estimator in the proposed manner the statistical model has to satisfy an orthogonal structure. This structure is not present in

most statistical models so the density of the maximum likelihood estimator can then not be constructed in the proposed manner. The instrumental variables regression model is an reduced rank regression model and other reduced rank regression models therefore also have orthogonal specifications. The proposed manner of constructing the density of the maximum likelihood estimators could therefore also be applicable in these reduced rank regression models, like cointegration and factor models.

Appendix

A. Proof of Theorem 1: The orthogonal specification of $\hat{\Theta}$ and its relationship to the mles of $\hat{\beta}$ and $\hat{\Pi}$. A SVD of $\hat{\Theta}$ reads

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

where U is a $k \times k$ orthonormal matrix ($U'U = I_k$), V is a $(m+1) \times (m+1)$ orthonormal matrix ($V'V = I_{m+1}$), and S is a $k \times (m+1)$ matrix that contains the singular values of $\hat{\Theta}$ in decreasing order on its main diagonal and is equal to zero elsewhere. If $k = (m+1)$, S is a $k \times k$ diagonal matrix with the singular values in decreasing order on its main diagonal. If $k \leq m$, S consists of a $k \times k$ diagonal matrix with the k singular values on its main diagonal (in decreasing order) extended on the right hand side with a $k \times (m+1-k)$ matrix of zeros. If $k > m+1$, S consists of a $(m+1) \times (m+1)$ diagonal matrix with the $m+1$ singular values on its main diagonal (in decreasing order) on top of a $(k-m-1) \times (m+1)$ matrix of zeros, see Golub and van Loan (1989) for details.

The expressions for $\hat{\Gamma}$, $\hat{\delta}$, and $\hat{\lambda}$ in terms of the elements of the SVD of $\hat{\Theta}$ follow from the relation

$$\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} v_{11}' & V_{21}' \\ v_{12}' & v_{22}' \end{pmatrix} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_\perp \hat{\lambda} \hat{D}_\perp,$$

where U_{11}, S_1, and V_{21} are $m \times m$ matrices; v_{12} is a scalar; v_{11}' and v_{22} are $m \times 1$ vectors; U_{12} and U_{21}' are $m \times (k-m)$ matrices; and U_{22} is a $(k-m) \times (k-m)$ matrix and s_2 is a $(k-m) \times 1$ vector. The above equation implies that

$$\hat{\Gamma}\hat{D} = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 [v_{11}' \vdots V_{21}'] \text{ and } \hat{\Gamma}_\perp \hat{\lambda} \hat{D}_\perp = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} s_2 [v_{12}' \vdots v_{22}'].$$

Although $\hat{\lambda}$ is initially only identified up to an orthonormal transformation, since $\hat{\Gamma}_\perp^* \hat{\lambda}^* \hat{D}_\perp^* = \hat{\Gamma}_\perp \hat{\lambda} \hat{D}_\perp$ with $\hat{\Gamma}_\perp^* = \hat{\Gamma}_\perp R$, $\hat{\lambda}^* = R' \hat{\lambda} Q'$, $\hat{D}_\perp^* = Q \hat{D}_\perp$ and R and Q are $(k-m) \times (k-m)$ and 1×1 orthonormal matrices, a unique specification of $\hat{\lambda}$ results when we use an appropriate normalization of $\hat{\Gamma}$ and $\hat{D}$ as proposed in Theorem 1: $\hat{\Gamma}$ unrestricted and $\hat{D} = (\hat{\delta} \vdots I_m)$. This normalization allows us to express $\hat{\Gamma}_\perp$ and $\hat{D}_\perp$ as functions of the unrestricted elements of $\hat{\Gamma}$ and $\hat{D}$.

To obtain an invertible relation between the singular value decomposition and $\hat{\Gamma}$, $\hat{\delta}$, and $\hat{\lambda}$, we need to express $\hat{\Gamma}_\perp$ and $\hat{D}_\perp$ in terms of the elements of $\hat{\Gamma}$ and $\hat{D}$ and hence in terms of U_{11},

U_{21}, S_1, V_{11}, and v_{21}. As we show below, this condition is fulfilled if[1]

$$\hat{\Gamma} = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V_{21}', \ \hat{\delta} = (V_{21}')^{-1} v_{11}' \text{ and } \hat{\lambda} = (U_{22}U_{22}')^{-\frac{1}{2}} U_{22} s_2 v_{12}' (v_{12}v_{12}')^{-\frac{1}{2}}.$$

which implies that

$$\hat{\Gamma}_\perp = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1} (U_{22}U_{22}')^{\frac{1}{2}} \text{ and } \hat{D}_\perp = (v_{12}v_{12}')^{\frac{1}{2}} (v_{12}')^{-1} [v_{12}' \vdots v_{22}'].$$

The matrices $(U_{22}U_{22}')^{-\frac{1}{2}} U_{22}$ and $v_{12}'(v_{12}v_{12}')^{-\frac{1}{2}}$ are both orthonormal matrices. Hence, $\hat{\lambda}$ is equal to the vector s_2, that contains the smallest singular value, pre- and postmultipied by orthonormal matrices or stated differently, $\hat{\lambda}$ is just a rotation of the smallest singular value around the origin.

We can now express $\hat{\Gamma}_\perp$ in terms of $\hat{\Gamma} = [\hat{\Gamma}_1' \vdots \hat{\Gamma}_2']'$ with $\hat{\Gamma}_1$ a $m \times m$ matrix and $\hat{\Gamma}_2$ a $(k-m) \times m$ matrix as

$$\begin{aligned}
\hat{\Gamma}_\perp &= \begin{pmatrix} U_{12}U_{22}^{-1} \\ I_{k-m} \end{pmatrix} ((U_{22}')^{-1} U_{22}^{-1})^{-\frac{1}{2}} \\
&= \begin{pmatrix} U_{12}U_{22}^{-1} \\ I_{k-m} \end{pmatrix} ((U_{22}')^{-1} (U_{22}'U_{22} + U_{12}'U_{12}) U_{22}^{-1})^{-\frac{1}{2}} \\
&= \begin{pmatrix} U_{12}U_{22}^{-1} \\ I_{k-m} \end{pmatrix} (I_{k-m} + (U_{22}')^{-1} U_{12}' U_{12} U_{22}^{-1})^{-\frac{1}{2}} \\
&= \begin{pmatrix} -(U_{11}')^{-1} U_{21}' \\ I_{k-m} \end{pmatrix} (I_{k-m} + U_{21} U_{11}^{-1} (U_{11}')^{-1} U_{21}')^{-\frac{1}{2}} \\
&= \begin{pmatrix} -(U_{11}')^{-1} (S_1')^{-1} V_{21}'^{-1} V_{21}' S_1' U_{21}' \\ I_{k-m} \end{pmatrix} (I_{k-m} + U_{21} S_1 V_{21}' (V_{21}')^{-1} S_1^{-1} U_{11}^{-1} (U_{11}')^{-1} (S_1')^{-1} V_{21}^{-1} V_{21} S_1' U_{21}')^{-\frac{1}{2}} \\
&= \begin{pmatrix} -(\hat{\Gamma}_1')^{-1} \hat{\Gamma}_2' \\ I_{k-m} \end{pmatrix} (I_{k-m} + \hat{\Gamma}_2 \hat{\Gamma}_1^{-1} (\hat{\Gamma}_1')^{-1} \hat{\Gamma}_2')^{-\frac{1}{2}},
\end{aligned}$$

where we used that $U_{12}'U_{11} + U_{22}'U_{12} \equiv 0 \Leftrightarrow (U_{22}')^{-1} U_{12}' \equiv -U_{12}U_{11}^{-1}$. Likewise, we specify $\hat{D}_\perp$ in terms of $\hat{\delta}$

$$\begin{aligned}
\hat{D}_\perp &= ((v_{12}')^{-1} v_{12}^{-1})^{-\frac{1}{2}} [1 \vdots (v_{12}')^{-1} v_{22}'] \\
&= ((v_{12}')^{-1} (v_{12}'v_{12} + v_{22}'v_{22}) v_{12}^{-1})^{-\frac{1}{2}} [1 \vdots \ (v_{12}')^{-1} v_{22}'] \\
&= (1 + (v_{12}')^{-1} v_{22}' v_{22} v_{12}^{-1})^{-\frac{1}{2}} [1 \vdots \ (v_{12}')^{-1} v_{22}'] \\
&= (1 + v_{11} V_{21}^{-1} (V_{21}')^{-1} v_{11}')^{-\frac{1}{2}} [1 : -v_{11} V_{21}^{-1}] \\
&= (1 + \hat{\delta}'\hat{\delta})^{-\frac{1}{2}} [1 \vdots -\hat{\delta}'],
\end{aligned}$$

since $v_{12}'v_{11} + v_{22}'V_{21} \equiv 0 \Leftrightarrow (v_{12}')^{-1} v_{22}' = -v_{11}V_{21}^{-1}$. Because $\hat{\Gamma}_\perp$ and $\hat{D}_\perp$ are functions of $\hat{\Gamma}$ and $\hat{\delta}$ only, there is an invertible relationship between $\hat{\Theta}$ and $(\hat{\Gamma}, \hat{\delta}, \hat{\lambda})$.

[1] If C is a positive definite real symmetric matrix, then $C^{\frac{1}{2}} = ELE'$, where L is a diagonal matrix containing the square roots of the eigenvalues of C, E contains the orthonormal eigenvectors of C, and $C^{-\frac{1}{2}} = EL^{-1}E'$, see for example Johansen (1995, p. 222).

The mles $\hat{\Pi}$ and $\hat{\beta}$ satisfy the first order condition for a maximum of the log-likelihood:

$$\begin{array}{rl} \frac{\partial}{\partial \beta'} Ln\mathcal{L}(\beta,\Pi|\Omega,Y,X,Z)|_{\hat{\beta},\hat{\Pi}} = 0 \Leftrightarrow & \text{vec}\{\hat{\Pi}'Z'[(y\ X) - Z\hat{\Pi}\hat{B}]\Omega^{-1}e_{1,m+1}\}' = 0, \\ \frac{\partial}{\partial \text{vec}(\Pi)'} Ln\mathcal{L}(\beta,\Pi|\Omega,Y,X,Z)|_{\hat{\beta},\hat{\Pi}} = 0 \Leftrightarrow & \text{vec}\{Z'[(y\ X) - Z\hat{\Pi}\hat{B}]\Omega^{-1}\hat{B}'\}' = 0. \end{array}$$

When $\hat{\Gamma}\hat{D} = (Z'Z)^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega^{\frac{1}{2}}$, so $\hat{\Gamma} = (Z'Z)^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega_V$ and $\hat{\delta} = (\hat{B}\Omega_V)^{-1}\hat{B}\omega_u$, the first equation of the first order condition can be specified as

$$\begin{array}{rl} \hat{\Pi}'Z'[(y\ X) - Z\hat{\Pi}\hat{B}]\Omega^{-1}e_{1,m+1} = 0 & \Leftrightarrow \\ \hat{\Pi}'Z'(y\ X)\Omega^{-1}e_{1,m+1} - \hat{\Pi}'Z'Z\hat{\Pi}\hat{B}\Omega^{-1}e_{1,m+1} = 0 & \Leftrightarrow \\ \hat{\Pi}'Z'(y\ X)\Omega^{-1}e_{1,m+1} - \hat{\Pi}'(Z'Z)^{\frac{1}{2}}\hat{\Gamma}\hat{D}\hat{\omega}_u = 0 & \Leftrightarrow \\ \hat{\Pi}'(Z'Z)^{\frac{1}{2}'}[\hat{\Theta} - \hat{\Gamma}\hat{D}]\hat{\omega}_u = 0 & \Leftrightarrow \\ (\hat{B}\Omega_V)^{-1\prime}\hat{\Gamma}'[\hat{\Theta} - \hat{\Gamma}\hat{D}]\hat{\omega}_u = 0 & , \end{array}$$

while the second equation can be specified as

$$\begin{array}{rl} Z'[(y\ X) - Z\hat{\Pi}\hat{B}]\Omega^{-1}\hat{B}' = 0 & \Leftrightarrow \\ Z'(y\ X)\Omega^{-1}\hat{B}' - Z'Z\hat{\Pi}\hat{B}\Omega^{-1}\hat{B}' = 0 & \Leftrightarrow \\ Z'(y\ X)\Omega^{-1}\hat{B}' - (Z'Z)^{\frac{1}{2}}\hat{\Gamma}\hat{D}\Omega^{-\frac{1}{2}'}\hat{B}' = 0 & \Leftrightarrow \\ (Z'Z)^{\frac{1}{2}}[\hat{\Theta} - \hat{\Gamma}\hat{D}][\omega_u \vdots \Omega_V]'\hat{B}' = 0 & \Leftrightarrow \\ (Z'Z)^{\frac{1}{2}}[\hat{\Theta} - \hat{\Gamma}\hat{D}]\hat{D}'(\hat{B}\Omega_V)' = 0 & . \end{array}$$

With $\hat{\Theta} = \hat{\Gamma}\hat{D} + \hat{\Gamma}_\perp\hat{\lambda}\hat{D}_\perp$ both equations hold and the value for $\hat{\Theta}$ implied by the mles $\hat{\beta}$ and $\hat{\Pi}$ thus result from a zero value of $\hat{\lambda}$.

B. Proof of Theorem 2: Block diagonal information matrix. The information matrix of $\Theta = (Z'Z)^{-\frac{1}{2}}Z'Y\Omega^{-\frac{1}{2}}$ given Ω equals an identity matrix:

$$\mathcal{I}(\Theta|\Omega) = (I_{m+1} \otimes I_k).$$

The information matrix of (β, Π, λ) results from the quadratic form of the information matrix of Θ and the Jacobian of the transformation from Θ to (β, Π, λ). We construct this information matrix in two steps. First, we construct the information matrix of $(\delta, \Gamma, \lambda)$, with $\Gamma = (Z'Z)^{\frac{1}{2}}\Pi B\Omega_V$ and $\delta = (B\Omega_V)^{-1}B\omega_u$, $\Omega = (\omega_u\ \Omega_V)$, $\omega_u : (m+1) \times 1$ and $\Omega_V : (m+1) \times m$. We show that this information matrix consists of two diagonal blocks, one for (δ, Γ) and one for λ. Because there is an invertible relationship between (δ, Γ) and (β, Π), the diagonal structure carries over to the information matrix of (β, Π, λ). To obtain the information matrix of $(\delta, \Gamma, \lambda)$, we construct the Jacobian of the transformation from $\Theta = \Gamma D + \Gamma_\perp\lambda D_\perp$ to $(\delta, \Gamma, \lambda)$:

$$\begin{array}{rl} \frac{\partial \text{vec}(\Theta)}{\partial \delta'} = & (e_{1,m+1} \otimes \Gamma) + (I_{m+1} \otimes \Gamma_\perp\lambda)\frac{\partial \text{vec}(D_\perp)}{\partial \text{vec}(\delta)'} \\ \frac{\partial \text{vec}(\Theta)}{\partial \text{vec}(\Gamma)'} = & (D' \otimes I_k) + (D'_\perp\lambda' \otimes I_k)\frac{\partial \text{vec}(\Gamma_\perp)}{\partial \text{vec}(\Gamma)'} \\ \frac{\partial \text{vec}(\Theta)}{\partial \lambda'} = & (D'_\perp \otimes \Gamma_\perp), \end{array}$$

where $D = (\delta \vdots I_m)$ and $D'D_\perp \equiv 0$, $D_\perp D'_\perp \equiv 1$, $\Gamma'_\perp \Gamma \equiv 0$, $\Gamma'_\perp \Gamma_\perp \equiv I_{k-m}$. The orthonormality of $\Gamma_\perp$ and $D_\perp$ implies that

$$\begin{aligned} \tfrac{\partial \mathrm{vec}(\Gamma'_\perp \Gamma_\perp)}{\partial \mathrm{vec}(\Gamma)'} \equiv 0 \quad &\Leftrightarrow \\ (I_{k-m} \otimes \Gamma'_\perp)\tfrac{\partial \mathrm{vec}(\Gamma_\perp)}{\partial \mathrm{vec}(\Gamma)'} + (\Gamma'_\perp \otimes I_{k-m})\tfrac{\partial \mathrm{vec}(\Gamma'_\perp)}{\partial \mathrm{vec}(\Gamma)'} \equiv 0 \quad &\Leftrightarrow \\ (I_{(k-m)(k-m)} + K_{k,k-m})(I_{k-m} \otimes \Gamma'_\perp)\tfrac{\partial \mathrm{vec}(\Gamma_\perp)}{\partial \mathrm{vec}(\Gamma)'} \equiv 0 \quad &\Leftrightarrow \\ (I_{k-m} \otimes \Gamma'_\perp)\tfrac{\partial \mathrm{vec}(\Gamma_\perp)}{\partial \mathrm{vec}(\Gamma)'} \equiv 0, & \end{aligned}$$

where $K_{(k-m),(k-m)}$ is the $(k-m)^2 \times (k-m)^2$ dimensional commutation matrix, see *e.g.* Magnus and Neudecker (1988). Similarly,

$$D'_\perp \tfrac{\partial \mathrm{vec}(D_\perp)}{\partial \mathrm{vec}(\delta)'} \equiv 0.$$

As a consequence,

$$\left(\tfrac{\partial \mathrm{vec}(\Theta)}{\partial \lambda'}\right)' \tfrac{\partial \mathrm{vec}(\Theta)}{\partial \delta'} = 0 \text{ and } \quad \left(\tfrac{\partial \mathrm{vec}(\Theta)}{\partial \lambda'}\right)' \tfrac{\partial \mathrm{vec}(\Theta)}{\partial \mathrm{vec}(\Gamma)'} = 0,$$

which implies a diagonal structure of the information matrix of $(\delta, \Gamma, \lambda)$. Since (β, Π) do not depend on λ, this diagonal structure is also present in the information matrix of (β, Π, λ) evaluated at their mles.

C. Proof of Theorem 3: Marginal density of $(\hat{\delta}, \hat{\Gamma})$. Because of the orthogonality of $(\hat{\delta}, \hat{\Gamma})$ and $\hat{\lambda}$, see (11),

$$\begin{aligned} &\left(\begin{array}{cc} \frac{\partial \mathrm{vec}(\hat{\Theta})}{\partial \hat{\delta}'} & \frac{\partial \mathrm{vec}(\hat{\Theta})}{\partial \mathrm{vec}(\hat{\Gamma})'} \end{array} \right)' \mathrm{vec}(\hat{\Theta}) = \Bigg((e_{1,m+1} \otimes \hat{\Gamma}) + (I_{m+1} \otimes \hat{\Gamma}_\perp \hat{\lambda}) \tfrac{\partial \mathrm{vec}(\hat{D}_\perp)}{\partial \mathrm{vec}(\hat{\delta})'} \vdots \\ &(\hat{D}' \otimes I_k) + (\hat{D}'_\perp \hat{\lambda}' \otimes I_k) \tfrac{\partial \mathrm{vec}(\hat{\Gamma}_\perp)}{\partial \mathrm{vec}(\hat{\Gamma})'} \Bigg)' \left[\left(\hat{D}' \otimes \hat{\Gamma} \right) \mathrm{vec}(I_m) + \left(\hat{D}'_\perp \otimes \hat{\Gamma}_\perp \right) \hat{\lambda} \right] = \left(\begin{array}{c} \hat{\Gamma}' \hat{\Gamma} \hat{\delta} \\ \mathrm{vec}(\hat{\Gamma} \hat{D} \hat{D}') \end{array} \right). \end{aligned}$$

We therefore obtain $(\hat{\delta}, \hat{\Gamma})$ from the $(m+1)k$ equations, see (12),

$$\left(\begin{array}{c} (e_{1,m+1} \otimes \hat{\Gamma})' \mathrm{vec}(\hat{\Theta}) \\ (\hat{D}' \otimes I_k)' \mathrm{vec}(\hat{\Theta}) \end{array} \right) = \left(\begin{array}{c} \hat{\Gamma}' \hat{\Gamma} \hat{\delta} \\ \mathrm{vec}(\hat{\Gamma} \hat{D} \hat{D}') \end{array} \right) \Leftrightarrow \hat{\delta} = (\hat{\Gamma}' \hat{\Gamma})^{-1} \hat{\Gamma}' \hat{\Theta} e_{1,m+1} \text{ and } \hat{\Gamma} = \hat{\Theta} \hat{D}' (\hat{D} \hat{D}')^{-1}.$$

These $(m+1)k$ equations show that $(\hat{\delta}, \hat{\Gamma})$ has an invertible relationship with those values of $\hat{\Theta}$ for which the implicit value of $\hat{\lambda}$ equals 0. For values of $\hat{\Theta}$ that correspond with non-zero values of $\hat{\lambda}$, $\hat{\lambda}$ does not affect the solution of $(\hat{\delta}, \hat{\Gamma})$ because all elements that contain $\hat{\lambda}$ cancel out as a result of the prevailing orthogonality relationships. We therefore obtain the marginal density of $(\hat{\delta}, \hat{\Gamma})$ by conducting a transformation of random variables from $\hat{\Theta}|_{\hat{\lambda}=0}$ towards $(\hat{\delta}, \hat{\Gamma})$ because all elements that depend on $\hat{\lambda}$ cancel out when we obtain $(\hat{\delta}, \hat{\Gamma})$ from $\hat{\Theta}$:

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

since $p(\hat{\Theta}) = (2\pi)^{-\frac{1}{2}k(m+1)} \exp\left[-\frac{1}{2}tr(\hat{\Theta} - \Gamma D)'(\hat{\Theta} - \Gamma D)\right]$ and we used that

$$\begin{aligned}
\left|\begin{pmatrix} \hat{D}\hat{D}' \otimes I_k & \hat{\delta} \otimes \hat{\Gamma} \\ \hat{\delta}' \otimes \hat{\Gamma}' & \hat{\Gamma}'\hat{\Gamma} \end{pmatrix}\right| &= |\hat{\Gamma}'\hat{\Gamma}||(\hat{D}\hat{D}' \otimes I_k) - (\hat{\delta} \otimes \hat{\Gamma})(\hat{\Gamma}'\hat{\Gamma})^{-1}(\hat{\delta}' \otimes \hat{\Gamma}')| \\
&= |\hat{\Gamma}'\hat{\Gamma}||((\hat{\delta}\hat{\delta}' + I_m) \otimes I_k) - \hat{\delta}\hat{\delta}' \otimes P_{\hat{\Gamma}}| \\
&= \left|\hat{\Gamma}'\hat{\Gamma}\right| \left|(I_m \otimes I_k) + (\hat{\delta}\hat{\delta}' \otimes M_{\hat{\Gamma}})\right| \\
&= \left|\hat{\Gamma}'\hat{\Gamma}\right| \left|I_m + \hat{\delta}\hat{\delta}'\right|^{(k-m)} = \left|\hat{\Gamma}'\hat{\Gamma}\right| \left|1 + \hat{\delta}'\hat{\delta}\right|^{(k-m)}
\end{aligned}$$

which involves

$$\begin{aligned}
\left|(I_m \otimes I_k) + (\hat{\delta}\hat{\delta}' \otimes M_{\hat{\Gamma}})\right| &= \left|[I_m \otimes (\hat{\Gamma}(\hat{\Gamma}'\hat{\Gamma})^{-\frac{1}{2}} \vdots \hat{\Gamma}_\perp)]'[(I_m \otimes I_k) + (\hat{\delta}\hat{\delta}' \otimes M_{\hat{\Gamma}})][I_m \otimes (\hat{\Gamma}(\hat{\Gamma}'\hat{\Gamma})^{-\frac{1}{2}} \vdots \hat{\Gamma}_\perp)]\right| \\
&= \left|\begin{pmatrix} I_m \otimes (\hat{\Gamma}'\hat{\Gamma})^{-\frac{1}{2}}\hat{\Gamma}'\hat{\Gamma}(\hat{\Gamma}'\hat{\Gamma})^{-\frac{1}{2}} & I_m \otimes (\hat{\Gamma}'\hat{\Gamma})^{-\frac{1}{2}}\hat{\Gamma}'\hat{\Gamma}_\perp \\ I_m \otimes \hat{\Gamma}_\perp\hat{\Gamma}(\hat{\Gamma}'\hat{\Gamma})^{-\frac{1}{2}} & I_m \otimes \hat{\Gamma}'_\perp\hat{\Gamma}_\perp + \hat{\delta}\hat{\delta}' \otimes \hat{\Gamma}'_\perp P_{\hat{\Gamma}'_\perp}\hat{\Gamma}'_\perp \end{pmatrix}\right| \\
&= \left|\begin{pmatrix} I_m \otimes I_m & 0 \\ 0 & (I_m + \hat{\delta}\hat{\delta}') \otimes I_{k-m} \end{pmatrix}\right| \\
&= \left|I_m + \hat{\delta}\hat{\delta}'\right|^{(k-m)}
\end{aligned}$$

since $M_{\hat{\Gamma}} = P_{\hat{\Gamma}_\perp}$. The marginal density of $(\hat{\delta}, \hat{\Gamma})$ is then equal to the conditional density of $(\hat{\delta}, \hat{\Gamma})$ given that $\hat{\lambda} = 0$.

D1. Proof of Theorem 4.a: Density of $(\hat{\beta}, \hat{\Pi})$. We obtain $(\hat{\beta}, \hat{\Pi})$ from $(\hat{\delta}, \hat{\Gamma})$ by using

$$\hat{\Gamma}\hat{D} = (Z'Z)^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega^{-\frac{1}{2}} = (Z'Z)^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega_V \begin{pmatrix} (\hat{B}\Omega_V)^{-1}\hat{B}\omega_u & I_m \end{pmatrix},$$

where $\Omega^{-\frac{1}{2}} = (\omega_u \ \Omega_V)$ with ω_u a $(m+1) \times 1$ vector and Ω_V a $(m+1) \times m$ matrix such that $\hat{\delta} = (\hat{B}\Omega_V)^{-1}\hat{B}\omega_u$ and $\hat{\Gamma} = (Z'Z)^{\frac{1}{2}}\hat{\Pi}\hat{B}\Omega_V$. To construct the Jacobian of the transformation from $(\hat{\Gamma}, \hat{\delta})$ to $(\hat{\Pi}, \hat{\beta})$, we use:

$$\frac{\partial \text{vec}(\hat{\delta})}{\partial \text{vec}(\hat{\beta})'} = (\omega_u'(I_{m+1} - \hat{B}'(\hat{B}\Omega_V)^{-1\prime}\Omega_V)e_{1,(m+1)} \otimes (\hat{B}\Omega_V)^{-1}), \quad \frac{\partial \text{vec}(\hat{\Gamma})}{\partial \text{vec}(\hat{\Pi})'} = (\Omega_V'\hat{B}' \otimes (Z'Z)^{\frac{1}{2}}).$$

Because $\frac{\partial \text{vec}(\hat{\delta})}{\partial \text{vec}(\hat{\Pi})'} = 0$, the Jacobian reads

$$|J((\hat{\Gamma}, \hat{\delta}), (\hat{\Pi}, \hat{\beta}))| = |\hat{B}\Omega_V|^{k-1} |Z'Z|^{\frac{1}{2}m} |\omega_u'(I_{m+1} - \hat{B}'(\hat{B}\Omega_V)^{-1\prime}\Omega_V)e_{1,m+1}|^m,$$

which implies the joint density of $(\hat{\beta}, \hat{\Pi})$:

$$\begin{aligned}
p(\hat{\Pi}, \hat{\beta}) \propto & \ p(\hat{\Gamma}(\hat{\Pi}, \hat{\beta}), \hat{\delta}(\hat{\Pi}, \hat{\beta}))|J((\hat{\Gamma}, \hat{\delta}), (\hat{\Pi}, \hat{\beta}))| \\
\propto & \ \left|\Omega_V'\hat{B}'\hat{\Pi}'Z'Z\hat{\Pi}\hat{B}\Omega_V\right|^{\frac{1}{2}} \left|I_m + (\hat{B}\Omega_V)^{-1}\hat{B}\omega_u\omega_u'\hat{B}'(\hat{B}\Omega_V)^{-1\prime}\right|^{\frac{1}{2}(k-m)} |\hat{B}\Omega_V|^{k-1} |Z'Z|^{\frac{1}{2}m} \\
& |\omega_u'(I_{m+1} - \hat{B}'(\hat{B}\Omega_V)^{-1\prime}\Omega_V)e_{1,m+1}|^m \exp[-\tfrac{1}{2}tr(\Omega^{-1}\{(\hat{\Pi}\hat{B} - \Pi B)'Z'Z(\hat{\Pi}\hat{B} - \Pi B)\})] \\
\propto & \ |\hat{B}\Omega^{-1}\hat{B}|^{\frac{1}{2}(k-m)}|\hat{\Pi}'Z'Z\hat{\Pi}|^{\frac{1}{2}}[|\hat{B}\Omega_V||\omega_u'(I_{m+1} - \hat{B}'(\hat{B}\Omega_V)^{-1\prime}\Omega_V)e_{1,m+1}|]^m \\
& \exp[-\tfrac{1}{2}tr(\Omega^{-1}\{(\hat{\Pi}\hat{B} - \Pi B)'Z'Z(\hat{\Pi}\hat{B} - \Pi B)\})] \\
\propto & \ |\hat{B}\Omega^{-1}\hat{B}|^{\frac{1}{2}(k-m)}|\hat{\Pi}'Z'Z\hat{\Pi}|^{\frac{1}{2}} \exp[-\tfrac{1}{2}tr(\Omega^{-1}\{(\hat{\Pi}\hat{B} - \Pi B)'Z'Z(\hat{\Pi}\hat{B} - \Pi B)\})]
\end{aligned}$$

since $\omega_u\omega_u' + \Omega_V\Omega_V' = \Omega^{-1}$, $\hat{B}\Omega_V$ is a square matrix and, because $\hat{B} = (\hat{\beta}\ I_{m-1})$, it holds that

$$|\Omega_V\hat{B}'||\omega_u'(I_m - B'(\Omega_V'B')^{-1}\Omega_V)e_{1,m+1}| = |\Omega|^{-\frac{1}{2}}.$$

D2: Proof of Theorem 4.b. Marginal density of $\hat{\beta}$. To construct the marginal density of $\hat{\beta}$, the trace component of the density $p(\hat{\Pi}, \hat{\beta})$ is specified as,

$$tr\{[\Omega^{-1} - \Omega^{-1}\hat{B}'(\hat{B}\Omega^{-1}\hat{B}')^{-1}\hat{B}\Omega^{-1}B']\Pi'X'X\Pi B\} + tr\{\hat{B}\Omega^{-1}\hat{B}'(\hat{\Pi} - \Psi)'X'X(\hat{\Pi} - \Psi)\},$$

with $\Psi = \Pi B\Omega^{-1}\hat{B}'(\hat{B}\Omega^{-1}\hat{B}')^{-1}$. The marginal density of $\hat{\beta}$ results from the integral of the joint density of $(\hat{\Pi}, \hat{\beta})$ over $\hat{\Pi}$,

$$\begin{array}{rl} p(\hat{\beta}) \propto & |\hat{B}\Omega^{-1}\hat{B}|^{\frac{1}{2}(k-m)} \exp[-\frac{1}{2}tr\{[\Omega^{-1} - \Omega^{-1}\hat{B}'(\hat{B}\Omega^{-1}\hat{B}')^{-1}\hat{B}\Omega^{-1}]B'\Pi'X'X\Pi B\}] \\ & \int_{\mathbb{R}^{km}} |\hat{\Pi}'Z'Z\hat{\Pi}|^{\frac{1}{2}} \exp[-\frac{1}{2}tr(\Omega^{-1}\{(\hat{\Pi}\hat{B} - \Pi B)'Z'Z(\hat{\Pi}\hat{B} - \Pi B)\})]d\hat{\Pi} \\ \propto & |\hat{B}\Omega^{-1}\hat{B}|^{-\frac{1}{2}(m+1)} \exp[\frac{1}{2}tr\{[\Omega^{-1} - \Omega^{-1}\hat{B}'(\hat{B}\Omega^{-1}\hat{B}')^{-1}\hat{B}\Omega^{-1}]B'\Pi'X'X\Pi B\}] \\ & \int_{\mathbb{R}^{km}} |\hat{\Upsilon}'\hat{\Upsilon}|^{\frac{1}{2}} \exp[-\frac{1}{2}tr\{(\hat{\Upsilon} - \Upsilon)'(\hat{\Upsilon} - \Upsilon)\}]d\hat{\Upsilon} \end{array}$$

with $\hat{\Upsilon} = (X'X)^{\frac{1}{2}}\hat{\Pi}(\hat{B}\Omega^{-1}\hat{B}')^{\frac{1}{2}}$, $\Upsilon = (X'X)^{\frac{1}{2}}\Psi(\hat{B}\Omega^{-1}\hat{B}')^{\frac{1}{2}}$ so $|J(\hat{\Pi}, \hat{\Upsilon})| = |X'X|^{-\frac{1}{2}(m-1)}\,|\hat{B}\Omega^{-1}\hat{B}'|^{-\frac{1}{2}k}$ and $|\hat{\Pi}'X'X\hat{\Pi}|^{\frac{1}{2}} = |\hat{\Upsilon}'\hat{\Upsilon}|^{\frac{1}{2}}|\hat{B}\Omega^{-1}\hat{B}'|^{-\frac{1}{2}}$. The above integral is a non-central moment of a matrix normal random matrix. We construct its closed form expression for the case that $\hat{\Upsilon}$ is a vector which implies that $m = 2$. For higher dimensional cases only expressions in terms of zonal polynomials can be obtained, see Muirhead (1982).

When $\hat{\Upsilon} \sim N(\Upsilon, I_k)$, it holds that $w = \hat{\Upsilon}'\hat{\Upsilon} \sim \chi^2(k, \mu)$, where $\mu = \Upsilon'\Upsilon$ is the non-centrality parameter of the non-central χ^2 distribution and k the degrees of freedom parameter. The density function of a non-central χ^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 χ^2 random variable with $k + 2j$ degrees of freedom. The expectation of $w^{\frac{1}{2}}$ when $w \sim \chi^2(k + 2j)$ reads,

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

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

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

so the expression of the integral involved in the density of $\hat{\beta}$ is

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

such that the density of $\hat{\beta}$ becomes

$$\begin{aligned}
p(\hat{\beta}) \propto\ & \left|\hat{B}\Omega^{-1}\hat{B}'\right|^{-\frac{1}{2}(m+1)} \exp\left[-\tfrac{1}{2}(\Omega^{-1}B'\Pi'X'X\Pi B)\right] \\
& \sum_{j=0}^{\infty}\left(\frac{1}{j!}\left(\frac{(B\Omega^{-1}\hat{B}')^2\Pi'X'X\Pi}{2\hat{B}\Omega^{-1}\hat{B}'}\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right) \\
\propto\ & \exp\left[-\tfrac{1}{2}tr\left(\Omega^{-1}B'\Pi'X'X\Pi B\right)\right]\left|\Omega_{VV}^{-1}+\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)'\Omega_{uu.V}^{-1}\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)\right|^{-\frac{1}{2}(m+1)} \\
& \sum_{j=0}^{\infty}\left[\left(\frac{\left|\Omega_{VV}^{-1}+\left(\Omega_{uV}\Omega_{VV}^{-1}-\beta'\right)'\Omega_{uu.V}^{-1}\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)\right|^2\Pi'X'X\Pi}{2\left|\Omega_{VV}^{-1}+\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)'\Omega_{uu.V}^{-1}\left(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}'\right)\right|}\right)^j \frac{\Gamma(\frac{1}{2}(k+2j+1))}{j!\Gamma(\frac{1}{2}(k+2j))}\right]
\end{aligned}$$

since

$$\begin{aligned}
&\left|\hat{B}\Omega^{-1}\hat{B}'\right| = |\Omega_{VV}^{-1}+(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}')'\Omega_{uu.V}^{-1}(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}')| \\
&\text{and } \left|B\Omega^{-1}\hat{B}'\right| = \left|\Omega_{VV}^{-1}+(\Omega_{uV}\Omega_{VV}^{-1}-\beta')'\Omega_{uu.V}^{-1}(\Omega_{uV}\Omega_{VV}^{-1}-\hat{\beta}')\right|,
\end{aligned}$$

with $\Omega_{uu.V} = \Omega_{uu} - \Omega_{uV}\Omega_{VV}^{-1}\Omega_{Vu}$.

E. Proof of Theorem 5: Appropriate value of Ω for the density of $\hat{\beta}$. To determine the appropriate value of Ω that we need to use for the density of $\hat{\beta}$, we use an estimator of Ω and analyze the joint density of $\hat{\beta}$ and this estimator. Usage of an estimator $\hat{\Delta}$ of Ω in the marginal density of $\hat{\beta}$ (15) implies that it becomes a conditional density given $\hat{\Delta}$ and is given by

$$\begin{aligned}
p(\hat{\beta}|\hat{\Delta}) \propto\ & \exp\left[-\tfrac{1}{2}tr(\hat{\Delta}^{-1}B'\Pi'X'X\Pi B)\right]\left|\hat{B}\hat{\Delta}^{-1}\hat{B}'\right|^{-\frac{1}{2}(m+1)} \\
& \sum_{j=0}^{\infty}\left(\frac{1}{j!}\left(\frac{(B\hat{\Delta}^{-1}\hat{B}')^2\Pi'X'X\Pi}{2\hat{B}\hat{\Delta}^{-1}\hat{B}'}\right)^j 2^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right).
\end{aligned}$$

The joint density of $\hat{\Delta}$ and $\hat{\beta}$ results from multiplying the marginal density of $\hat{\Delta}$ by the conditional density of $\hat{\beta}$ given $\hat{\Delta}$,

$$p(\hat{\beta},\hat{\Delta}) = p(\hat{\beta}|\hat{\Delta})p(\hat{\Delta}).$$

The presence of $\exp\left[-\frac{1}{2}tr(\hat{\Delta}^{-1}B'\Pi'X'X\Pi B)\right]$ in the conditional density of $\hat{\beta}$ given $\hat{\Delta}$ implies that the location of the density of $\hat{\Delta}$ changes. This results since we use $\hat{\Delta}$ to solve $(\hat{\beta},\hat{\Pi})$ from $(\hat{\delta},\hat{\Gamma})$. Hence, $(\hat{\beta},\hat{\Pi})$ is not independent of $\hat{\Delta}$ while $(\hat{\delta},\hat{\Gamma})$ is independent of $\hat{\Delta}$. The influence

on the density of $\hat{\Delta}$ is most straightforward to show when: 1. $p(\hat{\Delta})$ is such that it has $\hat{\Delta}^{-1}$ in the exponent term, 2. $\hat{\Delta}$ is independent of $\hat{\Phi}$ and 3. $p(\hat{\Delta})$ has a mean that is proportional to Ω such that $\hat{\Delta}$ can be used to solve $(\hat{\beta}, \hat{\Pi})$ from $(\hat{\delta}, \hat{\Gamma})$. The covariance matrix estimator $\hat{\Omega} = \frac{1}{T-k}(y\ X)'M_Z(y\ X)$, that is independent of $\hat{\Phi}$ and has a Wishart distribution with $T-k$ degrees of freedom, does not satisfy the first condition because its density function reads

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

We therefore use the infeasible estimator of Ω, $\hat{\Delta} = \Omega\hat{\Omega}^{-1}\Omega$. The estimator $\hat{\Delta}$ is infeasible since it depends on the unknown covariance matrix Ω. Since we only use it to analyze the change of the location of the density of $\hat{\Delta}$, the infeasibility is not a problem for our analysis. The distribution of $\hat{\Delta}$ is an inverted Wishart distribution with density function

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

and mean $\frac{T-k}{T-k-m-1}\Omega_0$. The joint density of $(\hat{\beta}, \hat{\Delta})$ reads

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

For large values of T $(T > 25)$,

$$|\hat{\Delta}|^{-\frac{1}{2}(T-k+m+1)} \exp\left[-\tfrac{1}{2} tr(\hat{\Delta}^{-1}(T-k)\{\Omega + B'\Pi'\tfrac{X'X}{T-k}\Pi B\})\right],$$

becomes a point mass at $\hat{\Delta} = \Omega + B'\Pi'\frac{X'X}{T-k}\Pi B$ so the marginal density of $\hat{\beta}$ is for large values of T,

$$p(\hat{\beta}) \propto \left|\hat{B}\hat{\Delta}^{-1}\hat{B}'\right|^{-\frac{1}{2}(m+1)} \sum_{j=0}^{\infty} \left(\tfrac{1}{j!}\left(\tfrac{(B\hat{\Delta}^{-1}\hat{B}')^2\Pi'X'X\Pi}{2\hat{B}\hat{\Delta}^{-1}\hat{B}'}\right)^j 2^{\frac{1}{2}} \tfrac{\Gamma(\frac{1}{2}(k+2j+1))}{\Gamma(\frac{1}{2}(k+2j))}\right),$$

with $\hat{\Delta} = \Omega + B'\Pi'\frac{X'X}{T-k}\Pi B$. This shows the value of Ω that we should use for the marginal density of $\hat{\beta}$.

References

[1] 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.

[2] Chao, J.C. and P.C.B. Phillips. Jeffreys prior analysis of the simultaneous equations model in the case with n+1 endogenous variables. *Journal of Econometrics*, **111**:251–284, 2002.

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

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

[5] 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).

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

[7] Johansen, S. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 1995.

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

[9] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, **70**:1781–1803, 2002.

[10] Kleibergen F. and E. Zivot. Bayesian and Classical Approaches to Instrumental Variable Regression. *Journal of Econometrics*, **114**:29–72, 2003.

[11] Lancaster, T. Orthogonal Parameters and Panel Data. *Review of Economic Studies*, **69**:647–666, 2002.

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

[13] 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.

[14] Moreira, M.J.,. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, **71**:1027–1048, 2003.

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

[16] 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.

[17] Reinsel, G.C. and R.P. Velu. *Multivariate Reduced Rank Regression: Theory and Applications*. Springer-Verlag, New York, 1998.

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