Conditional Densities in Econometrics

Frank Kleibergen*

Econometric Institute Report 9853/A

Abstract

Statistical inference in nested linear models that result from linear restrictions on the parameters of encompassing linear models can be considered to result from the conditional distribution under the encompassing model. We extend this reasoning to nested models that result from general (nonlinear) restrictions by defining sufficient conditions that, when satisfied by the random variables and the restrictions, ensure the existence of an unique expression of the conditional density. Statistical inference in these nested models can then be considered to result from such a conditional density. This novel manner of conducting statistical analyzes leads both to some new results and allows one to obtain some already known results in a different manner. In Bayesian statistics, the conditional densities show how to construct specific classes of priors for the parameters of nested models, priors on the parameters of encompassing models that imply an already specified prior on the parameters of the nested model, Bayes factors using (generalized) Savage-Dickey density ratios and Bayesian score statistics. In classical statistical analysis, the conditional densities offer an alternative approach for constructing small sample and limiting distributions of maximum likelihood estimators and classical score statistics.

1. Introduction

In the classical statistical analysis of linear models, the least squares estimator of the regression parameters often has a normal distribution either in small samples or asymptotically. Under the appropriate assumptions, the least squares estimators of regression parameters of other linear models that are nested in the initial encompassing linear model are then also normally distributed. Usually these results are obtained independently by evaluating the closed form expressions of the estimators. We can, however, also obtain this result by considering that the parameters of the encompassing model consist of two disjunct sets, one that contains the parameters of the nested model and one that contains the parameters that show the difference between the encompassing and nested model. The distribution of the least squares estimator of the parameters of the nested model can then be obtained from the conditional distribution of the least squares estimator of the parameters of the first set given that the least squares estimator of the parameters of the second set is equal to zero. The normality of the least squares estimator of the parameters of the nested model then results because the conditional

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

densities of the different elements of a multivariate normal distributed random vector are all normal.

In Bayesian statistical analysis a similar reasoning can be pursued to obtain the posteriors of the parameters of different nested linear models from a normal posterior of the parameters of an encompassing linear model, see Tiao *et. al.* (1977).

The inference in nested linear models, that result from linear restrictions on the parameters of encompassing linear models, can thus in both statistical paradigms be considered to result from conditionalization arguments. For more complicated nested models, that result from nonlinear restrictions on the parameters of encompassing models, thorough statistical analyzes based upon conditionalization arguments have not been developed previously. These analyzes have not been developed as it was thought that they would suffer from the Borel-Kolmogorov paradox, see *e.g.* Kolmogorov (1950) and Billingsley (1986), which states that the density of a probability distribution is not uniquely defined on sets of measure zero. In the case of the nested linear models that result from linear restrictions on the parameters of encompassing linear models, it is, however, possible to base the statistical inference on these kind of conditional densities. This shows that the Borel-Kolmogorov paradox is not as restrictive as it seems and that conditional densities given linear restrictions on the random variables are properly defined. It is even possible to further extend this class of restrictions, for which unique expressions of the conditional densities exist, and to specify sufficient conditions that, when satisfied by the (unrestricted) random variables and the restriction upon which we want to condition, ensure the existence of an unique expression of the conditional density given the restriction. The two crucial elements of these conditions are that the density of the (unrestricted) random variables has to be known and that parameters exist that unambiguously reflect the restrictions and nothing else. In case of the linear restrictions, the latter parameters result in a natural way but in case of nested models that result from nonlinear restrictions it is often less obvious how to construct them and whether they satisfy the sufficient conditions. We show, however, that the restrictions imposed by some commonly used models in econometrics and time series analysis on the parameters of encompassing linear models do satisfy the sufficient conditions and that we thus can construct an unique expression of the conditional density given these restrictions.

Once unique conditional densities can be specified, various inferential procedures in both Bayesian and classical statistical analyzes can be based upon them. Examples where they can be applied in Bayesian analyzes are: (*i.*) To construct general classes of priors for the parameters of nested models that result from restrictions, which satisfy the sufficient conditions for an unique conditional density, on the parameters of encompassing models where the general classes of priors are already defined for. (*ii.*) To construct the prior on the parameters of an encompassing model that implies an already specified prior on the parameters of a nested model. This is especially convenient when the updating from prior to posterior is obvious in the encompassing model but not in the nested model. In that case the prior on the parameters of the encompassing model reveals properties of the posterior of the parameters of the nested model and shows the plausiblity of the specified prior on the parameters of the nested model. Furthermore, it can be used for a sensitivity analysis. (*iii.*) To construct the Bayes factor for model comparison using the (generalized) Savage-Dickey density ratio. (*iv.*) To construct a Bayesian test statistic, *i.e.* the Bayesian score statistic. In classical statistical analyzes, the unique conditional density can for example be used: (I.) To construct the small sample and limiting distributions of the maximum likelihood estimator of the parameters of a nested model from the small sample and limiting distributions of the maximum likelihood estimator of the parameters of an encompassing model and (II.) to construct exact and asymptotic distributions of classical score statistics. In the paper, we briefly discuss each of the above applications and compare them with the traditional analyzes (when these exist).

The contents of the paper are organized as follows. In section 2, we discuss the different issues to construct the conditional density given a specific restriction with respect to the Borel-Kolmogorov paradox. We give sufficient conditions that, when satisfied by the (unrestricted) random variables and the restriction, ensure that an unique expression of the conditional density exists. In section 3, as an illustration, we construct the unique conditional densities that result when we impose reduced rank or exponential restrictions on normal distributed random matrices, vectors. Section 4 shows that commonly used models in econometrics and time series analysis, like cointegration, simultaneous equation and autoregressive moving average models, correspond with the restrictions from section 3 when imposed on the parameters of a linear model. Section 5 discusses the different consequences (*i.*)-(*iv.*) of the unique conditional densities for Bayesian analyzes. In section 6, we discuss the consequences (I.)-(II.) of the unique conditional densities for classical statistical inference on maximum likelihood estimators and classical score statistics. Finally the seventh section concludes.

2. Do unique Conditional Densities Exist?

Many statistical analyzes conducted in econometrics and time series analysis implicitly impose restrictions on (unrestricted) random variables. Distributional results can be obtained in these cases by constructing conditional Probability Density Functions (pdfs) of random variables given that a certain event, *i.e.* the restriction, occurs. These conditional pdfs can then be used for various cases in both Bayesian and classical statistics of which we discuss a few examples. A crucial problem in this kind of analysis is, however, that we, in most cases, cannot characterize the event, where we want to condition on, in an unique way. This phenomenon is known as the Borel-Kolmogorov paradox, see *e.g.* Billingsley (1986), Drèze and Richard (1983), Kolmogorov (1950), Poirier (1995) and Wolpert (1995), which states that the density of a probability distribution is not uniquely defined on sets of measure zero. It implies that we cannot find unique expressions of the conditional pdfs where we are interested in. Thus the answer to the question raised in the title of this section would be no! On the other hand, joint pdfs of several random variables are decomposed into products of conditional and marginal pdfs. It is well-known for example that the different elements of a bivariate normal distributed random vector have conditional normal pdfs given the other element, see for example Zellner (1971). This, however, contradicts a literal interpretation of the Borel-Kolmogorov paradox since it suggests that the conditional pdf of a random variable given a specific event in some cases has an unique expression. So, the answer to the question raised in the title of this section would then be sometimes! We therefore need to determine in which cases we can and cannot construct unique expressions for conditional pdfs and whether it is possible to define necessary and/or sufficient conditions that need to be satisfied in order to be able to do so. We illustrate this point using an example of a bivariate normal distribution.

2.1. Conditional Densities of the Bivariate Normal

Consider the scalar random variables x and y which are bivariate normal distributed with mean 0 and covariance matrix Ω ($= \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}$). The joint pdf of x and y reads,

$$p(x,y) = (2\pi)^{-1}|\Omega|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}' \Omega^{-1} \begin{pmatrix} x \\ y \end{pmatrix}\right], \tag{2.1}$$

such that the conditional pdf of x given y and the marginal pdf of y read,

$$p(x|y) = (2\pi\omega_{11.2})^{-\frac{1}{2}} \exp\left[-\frac{1}{2\omega_{11.2}}(x - \omega_{12}\omega_{22}^{-1}y)^2\right], \tag{2.2}$$

$$p(y) = (2\pi\omega_{22})^{-\frac{1}{2}} \exp\left[-\frac{1}{2\omega_{22}}y^2\right], \tag{2.3}$$

where $\omega_{11.2} = \omega_{11} - \omega_{12}\omega_{22}^{-1}\omega_{21}$. In the following we discuss two different kind of restrictions that we impose on x and y which illustrate the issues involving the Borel-Kolmogorov paradox. The first restriction we impose is $y = 0$.

2.1.1. Conditioning on $y = 0$

Imposing the restriction $y = 0$ on the bivariate normal distributed random variables x and y can be conducted in various ways. The most obvious one is directly using y. The pdf of x under the restriction that $y = 0$ then reads

$$\begin{aligned} p_r(x) &\propto p(x,y)|_{y=0} \\ &= p(x|y=0) \\ &= (2\pi\omega_{11.2})^{-\frac{1}{2}} \exp\left[-\frac{1}{2\omega_{11.2}}x^2\right], \end{aligned} \tag{2.4}$$

where r stands for restricted and $|_{y=0}$ for evaluated in $y = 0$. The pdf under the restriction that $y = 0$ is in this case identical to the conditional density of x (2.2) given $y = 0$. Another way to impose the restriction is using $z = \frac{y}{x} = 0$. The joint density of (x, z) reads

$$p(x,z) = (2\pi)^{-1}|x||\Omega|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}x^2\begin{pmatrix} 1 \\ z \end{pmatrix}' \Omega^{-1} \begin{pmatrix} 1 \\ z \end{pmatrix}\right], \tag{2.5}$$

since $|J((x,y),(x,z))| = |x|$, where $J((x,y),(x,z)) = \frac{\partial(x\ y)'}{\partial(x\ z)}$. When we impose $z = 0$ in (2.5), we obtain

$$\begin{aligned} p_r(x) &\propto p(x,z)|_{z=0} \\ &\propto |x| \exp\left[-\frac{1}{2\omega_{11.2}}x^2\right], \end{aligned} \tag{2.6}$$

which is not identical to (2.4), see Drèze and Richard (1983), Poirier (1995), and Wolpert (1995) for similar examples. This phenomenon shows a feature of the Borel-Kolmogorov paradox. In this case, however, $z = 0$ is not always equivalent to $y = 0$ since $z = 0$ can both imply $y = 0$ and $x = \pm\infty$ such that when $x = \pm\infty$ and $z = 0$, no unique value for y is implied. So, the two restrictions $y = 0$ and $z = 0$ are not identical and in this case, given x, the specification using z does not allow us to uniquely solve y from (x,z) for every value of z. It is therefore natural to use the restriction $y=0$ (or continuous invertible functions thereof) instead of $z=0$ as the latter specification not always represents the analyzed restriction. A direct consequence of this example is that unique expressions of conditional pdfs of (x,y) given for example $z=\frac{y}{x}=1$ or $z=xy=1$, do not exist because we cannot uniquely solve y from every value of (x,z). So, even though an analytical expression of the marginal pdf of z can be constructed, which is a Cauchy density, we cannot represent the restriction in terms of it and this shows that the existence of an analytical expression of the marginal density is another issue then the existence of an unique conditional density.

The example shows the importance of the spaces on which the random variables are defined. When these spaces are finite, $x=\pm\infty$ is outside of the space of x and $z=0$ ($z=\frac{y}{x}$) then corresponds with $y=0$. For a finite space of (x,y) we can thus uniquely solve y from every value of (x,z) and an unique expression of the conditional pdf does therefore not exist in that case (one can have doubt about the value of the jacobian of the transformation from (x,y) to (x,z) in that case though).

Conditional pdfs on events like $y=0$ in this example can be constructed in a quite natural way since variables exist that unambiguously represent the analyzed restriction. For other kind of restrictions, like (the mentioned) $xy=1$, this is, however, not the case.

2.1.2. Conditioning on $x^2+y^2=1$

The restriction $x^2+y^2=1$ can be specified in various ways. Examples are $\lambda=x-\sqrt{1-y^2}=0$, $\mu=y-\sqrt{1-x^2}=0$, $\theta=x^2+y^2=1$ and $r^2f(\varphi)=1$, where $f(\varphi)=\begin{pmatrix}\cos(\varphi)\\ \sin(\varphi)\end{pmatrix}'\Omega\begin{pmatrix}\cos(\varphi)\\ \sin(\varphi)\end{pmatrix}$, $\begin{pmatrix}x\\ y\end{pmatrix}=\Omega^{\frac{1}{2}}\begin{pmatrix}r\cos(\varphi)\\ r\sin(\varphi)\end{pmatrix}$. Of these restrictions only the last one, $r^2f(\varphi)=1$, is valid for specifying the restriction since the transformation from (x,y), $-\infty<x<\infty$, $-\infty<y<\infty$, to (r,φ), $0\leq r<\infty$, $0\leq\varphi<2\pi$, is such that (x,y) can be uniquely solved from (r,φ) for all values of (r,φ) which does not hold for the transformations from (x,y) to (λ,y), (μ,x) and (θ,x). This again shows the importance of the Borel-Kolmogorov paradox and it shows that restrictions should be able to be represented in terms of variables from which we can always solve for the original variables in order to have unique conditional densities, even on all lower dimensional spaces. This condition is stronger than just invertibility between the specification of the random variables containing a variable representing the restriction and the orginal variables as shown in the previous example where the joint density of (x,z) (2.5) is properly defined but representing the restriction using z does not lead to an unique expression of the conditional pdf because we cannot always uniquely solve y from (x,z). In general, restrictions do not allow for such a representation.

The joint pdf of (r,φ) reads

$$p(r,\varphi)=(2\pi)^{-1}|r|\exp\left[-\frac{1}{2}r^2\right], \tag{2.7}$$

such that the pdf of (x,y) under the restriction $x^2+y^2=r^2f(\varphi)=1$ can be represented by

(Note that since r is non-negative, square roots are uniquely defined)

$$\begin{aligned} p_r(x,y) &\propto p_r(\varphi(x,y))|J((r,\varphi),(x,y))|_{x^2+y^2=1}|, \\ p_r(\varphi) &\propto p(r,\varphi)|_{r=(f(\varphi))^{-\frac{1}{2}}} \\ &\propto \left| \begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix}' \Omega \begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix} \right|^{-\frac{1}{2}} \\ & \quad \exp\left[-\frac{1}{2} \left(\begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix}' \Omega \begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix} \right)^{-1} \right]. \end{aligned} \tag{2.8}$$

We can construct the conditional density (2.8) because every line with direction φ beginning in (0,0) only cuts through the circle $x^2 + y^2 = 1$ once such that only one unique value of r is associated with every value of φ.

2.2. Sufficient Conditions for Unique Conditional Densities

The examples of the bivariate normal in the previous subsection show some cases where an unique expression of the conditional pdf given the restriction exists. These are, however, quite exceptional cases as these expressions in general do not exist. The conditional pdfs exist in these cases because the restrictions satisfy two conditions that allow us to obtain unique expressions of the conditional densities. The involved random variables namely have an invertible relationship with another specification of them that consists of two disjunct sets of variables. These two sets are such that: $(i.)$ the variables in the first set only and completely reflect the imposed restriction and $(ii.)$ given that the variables in the second set do not lie in a lower dimensional subspace from which we cannot uniquely solve for the orginal variables, it is possible to uniquely solve for the value of the original variables for all values of the variables in the first set (this is also partly reflected in the word only in (i)). In the example in 2.1.1 the restriction is specified on y and no restriction is imposed on x such that these conditions are satisfied by definition. In the example in 2.1.2, we transform (x, y) to (r, φ). The restriction is then completely captured by r and φ is not affected by it. Furthermore, given φ, we can solve (x, y) from (r, φ) for every value of r such that the conditions are satisfied. So, in both examples we can represent the restrictions in terms of the (transformed) random variables in an unambiguous way.

The functions that represent the previously analyzed restrictions are all continuous and continuous differentiable. The involved random variables are also continuous and the space on which they are defined is not formally restricted. For these cases we can thus define sufficient conditions for the existence of an unique expression of the conditional pdf of the random variables given the restrictions.

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 pdf* $p(x)$; *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.*

Condition 1 implies that there are no lower dimensional subsets of the space on which y is defined where we cannot uniquely obtain x from (y, z), given that z does not result from a value of x that lies in a lower dimensional subspace where $f(x)$ is not continuous. As a result of this and combined with condition 2, y only represents the restriction and nothing else. We use y to obtain an unique expression for the conditional pdf which is unique since y allows us to only condition on the restriction and nothing else as other specifications, like the ones referred to in 2.1, would do because they not only condition on the restriction but also on other things. Condition 1 also ensures that the spaces of y and $g(x)$ have equal measures/dimensions and largely therefore we avoid the Borel-Kolmogorov paradox. The conditions on z correspond with the standard conditions for transforming random variables, see Poirier (1995). It implies for example that when $x = (x_1\ x_2)'$, $z = \frac{x_1}{x_2}$ is allowed for but $y = \frac{x_1}{x_2}$ isnot. The conditions on y are thus stronger than the normal conditions for transforming random variables.

In general, the specification of (y, z) is not unique. For example, when (y, z) satisfies conditions 1 and 2, so does $(p(y), z)$, where p is an invertible and continuous differentiable function. Conditions 1 and 2 still lead to an unique expression of the conditional pdf though because, when (y_1, z_1) and (y_2, z_2) both satisfy conditions 1 and 2, the jacobian of the transformation from x to (y_1, z_1) evaluated in $y_1 = 0$ is proportional to product of the jacobian of the transformation from x to (y_2, z_2) evaluated in $y_2 = 0$ and the jacobian of the transformation from z_2 to z_1. This is stated in lemma 1. Since the latter jacobian does not involve y_1 and y_2, this is not a trivial result and it therefore gives us a convenient tool to verify whether conditions 1 and 2 hold. It also results in a convenient transitivity property since no matter what specification used initially, it shows that we can still obtain the pdf of the desired specification. Later we show examples where we explicitly verify these properties.

Lemma 1. *When (y_1, z_1) and (y_2, z_2), (y_1, z_1) and (y_2, z_2) are not identical, satisfy conditions 1 and 2 then the transitivity property,*

$$|J(x, (y_1, z_1))|_{y_1=0}| \propto |J(x, (y_2(y_1, z_1), z_2(y_1, z_1)))|_{y_2=0}||J(z_2, z_1)|,$$

holds. Furthermore, all specifications (y, z) that have an invertible relationship with x and accord both with condition 2 and the transitivity property also satisfy condition 1.

Proof: see Appendix A.1.

The above conditions are such that the resulting conditional pdf exactly corresponds with the conditional pdf that results from decomposing the joint pdf of (y, z) into a product of conditional and marginal densities; $p(y, z) = p(z|y)p(y)$, such that $p_r(z) \propto p(z|y = 0)$. The conditions rule out any ambiguity regarding the specification of the restrictions in terms of the (transformed) random variables, such that we avoid the Borel-Kolmogorov paradox, and they imply that the random variables representing the restriction only represent the restriction and nothing else. The conditional densities we use are easier to obtain, however, than the complete decomposition into conditionals and marginals because it is not necessary to have an analytical expression of $p(y)$ in order to be able to construct them. We can therefore still obtain these conditional densities even in case it is not possible to construct the marginal density $p(y)$ analytically. So, the sufficient conditions correspond with the existence of a well-defined decomposition into conditionals and marginals of the kind discussed for example for the bivariate normal in 2.1.1 and 2.1.2.

The set of restrictions for which unique conditional pdfs exist may be larger than the one resulting from conditions 1 and 2, that is why we refer to them as sufficient conditions, but these conditions are weak enough to be able to analyze some models commonly used in econometrics and time series analysis and strong enough to imply unique expressions of the conditionals pdfs given the restrictions as stated in the next theorem.

Theorem 2.1. *When x and the restriction $g(x) = 0$ satisfy conditions 1 and 2, an unique expression for the conditional density of x given that $g(x) = 0$,*

$$\begin{aligned} p_r(x) &\propto p(x)|_{g(x)=0} \\ &\propto p_r(z(x))|J((y,z),x)|_{g(x)=0}|, \end{aligned} \tag{2.9}$$

where $p_r(z) \propto p(y,z)|_{y=0} \propto p(x(y,z))|_{y=0}|J(x,(y,z))|_{y=0}|$ and (y,z) result from conditions 1 and 2; exists.

Proof: see Appendix A.2.

Before we proceed with discussing how the above conditions can be applied in classical and Bayesian statistics, we first show the conditional densities that result when we impose specific restrictions on normal distributed random variables. As we will show later, these restrictions correspond with some commonly used models in econometrics and time series analysis.

3. Conditional Density given Restriction for the Normal Distribution

In this section we show two examples of the conditional pdf that results when we impose a specific restriction on a normal distributed random variable. Two other kind of restrictions on normal distributed random variables are already discussed in 2.1.1 and 2.1.2 and direct extensions of these examples, to general linear restrictions and spheres/ellipsoids with the mean of the normal random variables as their center, are straightforward. We therefore donot further discuss extensions of these restrictions. In the following two subsections, we impose reduced rank and exponential restrictions on normal distributed random vectors/matrices.

3.1. Reduced Rank Restriction on Matric-variate Normal

Consider the random matrix, $X : k \times m$, that has a matric-variate normal distribution, see *e.g.* Drèze and Richard (1983) and Zellner (1971), with mean μ and covariance matrix $(\Omega \otimes P^{-1})$, where $\mu : k \times m$, $\Omega : m \times m$ and $P : k \times k$, Ω and P positive definite symmetric matrices. The pdf of X thus reads,

$$p(X) = (2\pi)^{-\frac{1}{2}km}|P|^{\frac{1}{2}m}|\Omega|^{-\frac{1}{2}k}\exp\left[-\frac{1}{2}tr(\Omega^{-1}(X-\mu)'P(X-\mu))\right]. \tag{3.1}$$

We want to analyze the random matrix X under the restriction that its rank is equal to r which is less than $\min(k, m)$ and larger than 0,

$$rank(X) = r < \min(k, m). \tag{3.2}$$

When X satisfies restriction (3.2) and thus has reduced rank r, we can represent X as the product of two matrices, $Z : k \times r$, and $W : r \times m$, that each have rank r, see Golub and van Loan (1989),

$$X = ZW. \tag{3.3}$$

Since the parameters in (3.3) are not identified, as postmultiplying Z by $Q : r \times r$ and premultiplying W by Q^{-1} leads to the same value of X, we need to impose restrictions on Z and W in order to identify them. Two possible identified specifications are:

$$\begin{aligned} \mathbf{1}.\ X &= Z_1W_1;\ W_1 = \begin{pmatrix} A & I_r \end{pmatrix}; A : r \times (m-r); Z_1, A : \text{unrestricted}, & (3.4)\\ \mathbf{2}.\ X &= Z_2W_2;\ Z_2 = \begin{pmatrix} I_r \\ B \end{pmatrix};\ B : (k-r) \times r;\ W_2, B : \text{unrestricted}, & (3.5) \end{aligned}$$

where $Z_1 : k \times r$, $W_2 : r \times m$.

Specifications (3.3) and (3.4), (3.5) are specifications that already assume a reduced rank value of X and therefore donot span the reduced rank restriction itself. We span the reduced rank restriction using the singular values of the unrestricted random matrix X. Singular values are generalized eigenvalues of (nonsymmetric) matrices and are a natural way to represent the rank of a matrix, see Golub and van Loan (1989). Using these singular values we construct the unrestricted (full rank) specification of X,

$$X = Z_iW_i + Z_\perp Y W_\perp, \qquad i = 1, 2, \tag{3.6}$$

where $Z_\perp : k \times (k-r)$, $W_\perp : (m-r) \times m$, $Y : (k-r) \times (m-r)$; and $Z_i'Z_\perp \equiv 0$, $W_iW_\perp' \equiv 0$, $i = 1, 2$; $Z_\perp'Z_\perp \equiv I_{k-r}$, $W_\perp W_\perp' \equiv I_{m-r}$, such that $Z_\perp = (-B\ I_{k-r})'(I_{k-r} + BB')^{-\frac{1}{2}} = (-Z_{12}Z_{11}^{-1}\ I_{k-r})'(I_{k-r} + Z_{12}Z_{11}^{-1}Z_{11}^{-1\prime}Z_{12}')^{-\frac{1}{2}}$, $W_\perp = (I_{m-r} + A'A)^{-\frac{1}{2}}(I_{m-r}\ -A') = (I_{m-r} + W_{21}'W_{22}^{-1\prime}W_{22}^{-1}W_{21})^{-\frac{1}{2}}(I_{m-r}\ -W_{21}'W_{22}^{-1\prime})$, where $Z_1 = (Z_{11}'\ Z_{12}')'$, $Z_{11} : r \times r$, $Z_{12} : (k-r) \times r$; $W_2 = (W_{21}\ W_{22})$, $W_{21} : r \times (m-r)$, $W_{22} : r \times r$. We note that the square root of a positive (semi-)definite symmetric matrix Q is well-defined, $Q = D\Lambda D'$, D orthogonal, Λ diagonal and nonnegative, then $Q^{\frac{1}{2}} = D\Lambda^{\frac{1}{2}}D'$.

Specification (3.6) results from a singular value decompostion (svd) of X, see Golub and van Loan (1989) and Magnus and Neudecker (1988),

$$X = USV', \tag{3.7}$$

where $U : k \times k$, $U'U = I_k$; $V : m \times m$, $V'V = I_m$; and $S : k \times m$ is a rectangular matrix containing the (nonnegative) singular values (in decreasing order) on its main diagonal ($= (s_{11}...s_{nn})$ with $n = \min(k, m)$). If we now write,

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix},\ S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix},\ \text{and } V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}, \tag{3.8}$$

where U_{11}, S_1, $V_{21} : r \times r$; $V_{12} : (m-r) \times (m-r)$; V_{11}', $V_{22} : r \times (m-r)$; U_{12}, $U_{21}' : r \times (k-r)$; $U_{22} : (k-r) \times (k-r)$; $S_2 : (k-r) \times (m-r)$, then the following relationship exists between (Z_1, A, Y), (B, W_2, Y) and (U, S, V), for a proof see appendix B:

$$\begin{aligned} Z_1 &= \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1V_{21}',\ A = V_{21}^{-1\prime}V_{11}', & (3.9)\\ W_2 &= U_{11}S_1 \begin{pmatrix} V_{11} \\ V_{21} \end{pmatrix}',\ B = U_{21}U_{11}^{-1}, \\ Y &= (U_{22}U_{22}')^{-\frac{1}{2}}U_{22}S_2V_{12}'(V_{12}V_{12}')^{-\frac{1}{2}}. \end{aligned}$$

Since $(U_{22}U_{22}')^{-\frac{1}{2}}U_{22}$ and $V_{12}'(V_{12}V_{12}')^{-\frac{1}{2}}$ are orthogonal matrices, Y is a rotation of the singular values contained in S_2. So, restricting Y is equivalent with restricting the smallest $\min(k-r, m-r)$ singular values contained in S_2 since the jacobian of the transformation from Y to S_2 is one. It is also more convenient to impose a restriction on Y since it has no sign and ordering restrictions like the singular values.

Specification (3.6)-(3.9) satisfies conditions 1 and 2. Condition 1 is satisfied since X can only not uniquely be solved from (Z_1, A, Y) or (W_2, B, Y) for lower rank values of V_{21}, U_{11} respectively. This results from the inverse that is taken of these matrices when constructing A and B. When V_{21}, U_{11} have full rank values, we can uniquely solve X from (Z_1, A, Y), (W_2, B, Y) for all values of Y. This results because Y in (3.9) is an orthogonal transformation of the singular values and the singular values are uniquely defined, see Golub and van Loan (1989). So, since the restriction is not specified on A or B, condition 1 is satisfied. Condition 2 is satisfied because the reduced rank restriction (3.2) is equivalent with $Y = 0$ and doesnot involve the other variables. This is shown in (3.6), as $Y = 0$ implies that (3.6) is equivalent with (3.3)-(3.5). To further verify whether condition 1 and 2 are satisfied we check for the transtivity property implied by lemma 1. For this, we use the jacobian of the transformation of the unrestricted X to (Z_1, A, Y) evaluated in $Y = 0$, which is constructed in appendix C,

$$\begin{aligned}
& |J(X,(Z_1,A,Y))|_{Y=0}| \qquad (3.10)\\
= & \left|\begin{pmatrix} W_1'\otimes I_k & \begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix}\otimes Z_1 & W_\perp'\otimes Z_\perp \end{pmatrix}\right| \\
= & \left|\begin{pmatrix} W_1'\otimes I_k & \begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix}\otimes Z_1 \end{pmatrix}'\begin{pmatrix} W_1'\otimes I_k & \begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix}\otimes Z_1 \end{pmatrix}\right|^{\frac{1}{2}} \\
= & \left|\left(\frac{\partial vec(X)}{\partial(vec(Z_1)'\ vec(A)')}\right)'\left(\frac{\partial vec(X)}{\partial(vec(Z_1)'\ vec(A)')}\right)\right|^{\frac{1}{2}},
\end{aligned}$$

and is simplified by using that $Z_1'Z_\perp = 0$, $Z_\perp'Z_\perp = I_{k-r}$, $W_1W_\perp' = 0$, $W_\perp W_\perp' = I_{m-r}$. The transivity property results from (3.10) because when (B, W_2, Y) is another set of variables, where (W, B_2) span X under reduced rank, that satisfy conditions 1 and 2, it holds that

$$\begin{aligned}
& |J(X,(W_2,B,Y))|_{Y=0}| \qquad (3.11)\\
= & \left|\left(\frac{\partial vec(X)}{\partial(vec(W_2)'\ vec(B)')}\right)'\left(\frac{\partial vec(X)}{\partial(vec(W_2)'\ vec(B)')}\right)\right|^{\frac{1}{2}} \\
= & \left|\left(\left(\frac{\partial vec(X)}{\partial(vec(Z_1)'\ vec(A)')}\right)\left(\frac{\partial(vec(Z_1)'\ vec(A)')'}{\partial(vec(W_2)'\ vec(B)')}\right)\right)'\right. \\
& \left.\left(\left(\frac{\partial vec(X)}{\partial(vec(Z_1)'\ vec(A)')}\right)\left(\frac{\partial(vec(Z_1)'\ vec(A)')'}{\partial(vec(W_2)'\ vec(B)')}\right)\right)\right|^{\frac{1}{2}} \\
= & \left|\left(\frac{\partial vec(X)}{\partial(vec(Z_1)'\ vec(A)')}\right)'\left(\frac{\partial vec(X)}{\partial(vec(Z_1)'\ vec(A)')}\right)\right|^{\frac{1}{2}} \\
& \left|\left(\frac{\partial(vec(Z_1)'\ vec(A)')'}{\partial(vec(W_2)'\ vec(B)')}\right)\right| \\
= & |J(X,(Z_1,A,Y))|_{Y=0}||J((Z_1,A),(W_2,B)|,
\end{aligned}$$

because $\frac{\partial(vec(Z_1)'\ vec(A)')'}{\partial(vec(W_2)'\ vec(B)')}$ is a square matrix and it shows that the transivity property holds for specification (3.6)-(3.8) and that it thus satisfies conditions 1 and 2. So, since the two sufficient conditions for the existence of an unique expression of the conditional density are satisfied, we can construct the conditional density of the matric-variate normal distributed random matrix X given that its rank is equal to r,

$$\begin{aligned}
p_r(X) &\propto p(X)|_{rank(X)=r} \\
&\propto p_r(Z_1(X), A(X))|J((Z_1, A, Y), X)|_{rank(X)=r}|, \\
p_r(Z_1, A) &\propto p(X(Z_1, A, Y))|_{Y=0}|J(X, (Z_1, A, Y))|_{Y=0}| \\
&\propto \left| \begin{pmatrix} W_1W_1' \otimes I_k & A \otimes Z_1 \\ A' \otimes Z_1' & I_{m-r} \otimes Z_1'Z_1 \end{pmatrix} \right|^{\frac{1}{2}} \\
&\quad \exp\left[-\frac{1}{2} tr(\Omega^{-1}(Z_1W_1 - \mu)'P(Z_1W_1 - \mu))\right] \\
&\propto q(Y|Z_1, A)|_{Y=0} q(Z_1, A),
\end{aligned} \tag{3.12}$$

where

$$\begin{aligned}
& q(Y|Z_1, A) \\
=\ & (2\pi)^{-\frac{1}{2}(k-r)(m-r)} |W_\perp\Omega^{-1}W_\perp'|^{\frac{1}{2}(k-r)} |Z_\perp'PZ_\perp|^{\frac{1}{2}(m-r)} \\
& \exp\left[-\frac{1}{2} tr(W_\perp\Omega^{-1}W_\perp'(Y - \mu_Y)'Z_\perp'PZ_\perp(Y - \mu_Y))\right],
\end{aligned} \tag{3.13}$$

$$\begin{aligned}
& q(Z_1, A) \\
\propto\ & \left| \begin{pmatrix} W_1W_1' \otimes I_k & W_1' \otimes Z_1 \\ A' \otimes Z_1 & \begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix}' \otimes Z_1'Z_1 \end{pmatrix}' \left(\begin{pmatrix} W_1' \otimes I_k & I_m \otimes Z_1 \end{pmatrix}' (\Omega \otimes P^{-1}) \right. \right. \\
& \left. \left. \begin{pmatrix} W_1' \otimes I_k & I_m \otimes Z_1 \end{pmatrix} \right)^{-1} \begin{pmatrix} W_1W_1' \otimes I_k & W_1' \otimes Z_1 \\ A' \otimes Z_1 & \begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix}' \otimes Z_1'Z_1 \end{pmatrix} \right|^{\frac{1}{2}} \\
& \exp[-\frac{1}{2}(tr(\Omega^{-1}(Z_1W_1 - \mu)'P(Z_1W_1 - \mu)) - tr(\Omega^{-1}W_\perp'(W_\perp\Omega^{-1}W_\perp')^{-1} \\
& W_\perp\Omega^{-1}(Z_1W_1 - \mu)'PZ_\perp(Z_\perp'PZ_\perp)^{-1}Z_\perp'P(Z_1W_1 - \mu)))],
\end{aligned} \tag{3.14}$$

and $\mu_Y = (Z_\perp'PZ_\perp)^{-1}Z_\perp'P(\mu - Z_1W_1)\Omega^{-1}W_\perp'(W_\perp\Omega^{-1}W_\perp')^{-1}$.

In (3.12) we show that $p_r(X)$ is proportional to the product of the conditional density $q(Y|Z_1, A)$ (3.13) evaluated in $Y = 0$ and $q(Z_1, A)$ (3.14). We note that these densities donot represent the conditional density of Y given (Z_1, A) and the marginal density of (Z_1, A). We therefore also indicate these densities by q instead of p. We cannot analytically construct the integrating constant of the true conditional density of Y given (Z_1, A) such that no analytical expression of the marginal density of (Z_1, A) exists. This results because Y is present in the jacobian of the transformation from X to (Z_1, A, Y), $J(X, (Z_1, A, Y))$, see appendix C. The property that $p_r(X)$ can be specified as the product of the conditional and marginal shown in (3.12) allows us, however, to conduct some of the analyzes discussed in later sections. This explains why we explicitly mention this decomposition. We note that these analyzes are based on $q(Y|Z_1, A)$ (3.13) and donot involve $q(Z_1, A)$ (3.14).

The pdf $p_r(X)$ in (3.12) doesnot belong to a known class of densities. Analytical expressions for the moments of X under reduced rank or the moments of the variables spanning its reduced rank space, like Z_1 and A, are therefore unknown. The density is proper though since it is constructed using a proper density of the unrestricted random matrix X. A property of it is also mentioned just above where it is shown that it can be specified as the product of the conditional density $q(Y|Z_1, A)$ (3.13) evaluated in $Y = 0$ and the marginal density $q(Z_1, A)$ (3.14). In Kleibergen and Paap (1998) and Kleibergen and van Dijk (1998), algorithms to sample from (3.12) are constructed.

3.2. Exponential Restrictions on Multivariate Normal

Consider the random vector x; $x : k \times 1$, $k \geq 2$; that has a multi-variate normal distribution with mean μ and covariance matrix P^{-1}, $\mu : k \times 1$, and $P : k \times k$ positive definite symmetric. The pdf of x reads,

$$p(x) = (2\pi)^{-\frac{1}{2}k}|P|^{\frac{1}{2}} \exp\left[-\frac{1}{2}(x-\mu)'P(x-\mu)\right]. \tag{3.15}$$

We impose $k-2$ exponential restrictions on x,

$$x_i = zx_{i-1}, \qquad i = 2, ..., k, \tag{3.16}$$

where $x = (x_1...x_k)'$, and $z : 1 \times 1$, is a random variable that results from the restrictions. We represent these restrictions using,

$$y_i = x_{i+2} - zx_{i+1}, \qquad i = 1, ..., k-2, \tag{3.17}$$

such that x can be specified as,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ x_k \end{pmatrix} = \begin{pmatrix} x_1 \\ zx_1 \\ y_1 + z^2x_1 \\ . \\ y_{k-2} + z^{k-1}x_1 \end{pmatrix}, \tag{3.18}$$

where $y = (y_1...y_{k-2})'$. The specification (3.17)-(3.18) satisfies the sufficient conditions 1 and 2 for the existence of an unique conditional pdf of x given the restrictions. With respect to condition 1, we can uniquely solve for x in (3.18) from (x_1, z, y) for all values of (x_1, z, y) when $x_1 \neq 0$. Since the restriction is not specified on x_1, condition 1 is thus satisfied. Condition 2 is satisfied because (3.18) is equivalent to (3.16) when $y = 0$ in (3.18). To further verify these conditions we check whether the transivity property resulting from lemma 1 holds. For this, it is important to note that allthough z is present in all restrictions, it can only be obtained as the ratio of $\frac{x_{i+1}}{x_i}$, $i = 1, ..., k-1$, since square roots of negative values are not properly defined. Another specification of x in terms of variables representing the restriction (3.16) is therefore

$$\begin{pmatrix} x_1 \\ . \\ x_{i-1} \\ x_i \\ x_{i+1} \\ x_{i+2} \\ . \\ x_k \end{pmatrix} = \begin{pmatrix} w_1 + v^{-i}x_i \\ . \\ w_{i-1} + v^{-1}x_i \\ x_i \\ vx_i \\ w_i + v^2x_i \\ . \\ w_{k-2} + v^{k-i-1}x_i \end{pmatrix}, \tag{3.19}$$

where $1 < i < k$, $v : 1 \times 1$, $w = (w_1...w_{k-2})'$ and when $w = 0$, (3.19) is equivalent to (3.16). The jacobian of the transformation from x to (x_1, z, y), $|J(x, (x_1, z, y))|$, is equal to $|x_1|$. The jacobian of the transformation from x to (x_i, v, w), $|J(x,(x_i,v,w))|$, is equal to $|x_i|$. Because $x_1 = v^{-i}x_i$, $z = v$, when $w = y = 0$; it thus holds that $|J((x_i,v),(x_1,z))| = |z^{-i}|$ such that $|J(x,(x_1,z,y))|_{y=0}| = |J(x,(x_i,v,w))|_{w=0}||J((x_i,v),(x_1,z))| = |z^i x_1||z^{-i}| = |x_1|$, which shows that the transitivity property holds and therefore that conditions 1 and 2 are satisfied. So, since all conditions are satisfied we can construct the unique conditional density of x given (3.16),

$$\begin{aligned}
p_r(x) &\propto p(x)|_{x_{i+1}=zx_i, i=1,\ldots,k-1} \\
&\propto p_r(x_1(x), z(x))|J((x_1,z,y),x)|_{x_{i+1}=zx_i, i=1,\ldots,k-1}|, \\
p_r(x_1,z) &\propto p(x(x_1,z,y))|_{y=0}|J(x,(x_1,z,y))|_{y=0}| \\
&\propto |x_1| \exp\left[-\frac{1}{2}\left(x_1\begin{pmatrix}1\\ z\\ .\\ z^{k-1}\end{pmatrix}-\mu\right)' P \left(x_1\begin{pmatrix}1\\ z\\ .\\ z^{k-1}\end{pmatrix}-\mu\right)\right] \\
&\propto p(y|x_1,z)|_{y=0}p(z,x_1),
\end{aligned} \tag{3.20}$$

where

$$p(y|x_1,z) = (2\pi)^{-\frac{1}{2}(k-2)}|P_{22}|^{\frac{1}{2}}\exp\left[-\frac{1}{2}(y-\mu_y)'P_{22}(y-\mu_y)\right], \tag{3.21}$$

$$\begin{aligned}
p(x_1,z) = {} & (2\pi)^{-1}|x_1||P_{11.2}|^{\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\begin{pmatrix}x_1\\ z\end{pmatrix}-\mu_{xz}\right)'\right. \\
& \left.\begin{pmatrix}1 & 0\\ 0 & x_1\end{pmatrix}P_{11.2}\begin{pmatrix}1 & 0\\ 0 & x_1\end{pmatrix}\left(\begin{pmatrix}x_1\\ z\end{pmatrix}-\mu_{xz}\right)\right],
\end{aligned} \tag{3.22}$$

and $P = \begin{pmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{pmatrix}$; $P_{11} : 2\times 2$; $P_{22} : (k-2)\times(k-2)$; P_{12}', $P_{21} : (k-2)\times 2$; $P_{11.2} = P_{11} - P_{12}P_{22}^{-1}P_{21}$; $\mu = (\mu_1' \ \mu_2')'$; $\mu_1 : 2\times 1$; $\mu_2 : (k-2)\times 1$; $\mu_y = \mu_2 + x_1\begin{pmatrix}z^2\\ .\\ z^{k-1}\end{pmatrix} + P_{22}^{-1}P_{21}\begin{pmatrix}1 & 0\\ 0 & x_1\end{pmatrix}\left(\begin{pmatrix}x_1\\ z\end{pmatrix}-\mu_{xz}\right)$; $\mu_{xz} = \begin{pmatrix}1 & 0\\ 0 & x_1^{-1}\end{pmatrix}\mu_1$.

The conditional density (3.20) shows the functional form of the density of a multivariate normal distributed random vector under the exponential restriction (3.16). This density doesnot belong to a known class of pdfs such that analytical expressions of its moments do not exist. It is a proper density though since it is constructed from a proper normal pdf. The conditional pdf (3.20) has the same property as the conditional pdf (3.12) since it can be specified as the product of a conditional normal pdf evaluated in 0 and another (in this case also normal) marginal pdf. An important difference with (3.12) is, however, that (3.21) and (3.22) are the true conditional and marginal pdfs of y and (x_1,z) respectively while this doesnot hold for (3.13) and (3.14). We therefore also indicate them by p instead of q as in (3.13) and (3.14). We mention these decompositions since they allow us to conduct some of the analyzes discussed in later sections. In Kleibergen and Hoek (1996), algorithms to sample from (3.20) are constructed.

4. Models corresponding with the Conditional Densities

Before we proceed with the classical and Bayesian applications of the unique conditional densities, we first briefly discuss the statistical models in the analyzes of whom the conditional densities from the previous section are used.

4.1. Reduced Rank Parameter Matrices

Reduced rank parameter matrices appear in commonly used statistical models in econometrics and time series analysis. Amongst these are simultaneous equations models, instrumental variable regression models, cointegration models and factor models. In the following, we briefly discuss two of these models, *i.e.* simultaneous equations and cointegration models.

4.1.1. Simultaneous Equations Model

Consider a simultaneous equations model (sem) with one completely specified structural equation while the other equations are in (unrestricted) reduced form. The resulting kind of analysis is also known as limited information analysis or instrumental variable regression, see Hausman (1983),

$$\begin{aligned} y_1 &= Y_2\beta + \xi_1, \\ Y_2 &= X\Pi + \xi_2, \end{aligned} \tag{4.1}$$

where $y_1 : T \times 1$, $Y_2 : T \times (m-1)$, contain the endogenous variables; $X : T \times k$, contains the instruments that are weakly exogenous or lagged endogenous variables, see Engle *et. al.* (1983), $k \geq (m-1)$; $(\xi_1 \ \ \xi_2) \sim n(0, \Sigma \otimes I_T)$; $\Sigma : m \times m$, positive definite symmetric. A more general specification of (4.1) results when we add identical explanatory variables to all of its equations. Since this doesnot change the analysis of (4.1) besides conducting an auxiliary regression beforehand, we use the notationally more convenient specification (4.1). The reduced rank property of the parameters of (4.1) becomes clear when we construct its restricted reduced form by substituting Y_2 in the first equation of (4.1),

$$\begin{pmatrix} y_1 & Y_2 \end{pmatrix} = X\Pi \begin{pmatrix} \beta & I_{m-1} \end{pmatrix} + \varepsilon, \tag{4.2}$$

where $\varepsilon = \begin{pmatrix} \xi_1 + \xi_2\beta & \xi_2 \end{pmatrix}$, $\varepsilon \sim n(0, \Omega \otimes I_T)$, $\Omega = \begin{pmatrix} 1 & 0 \\ \beta & I_{m-1} \end{pmatrix}' \Sigma \begin{pmatrix} 1 & 0 \\ \beta & I_{m-1} \end{pmatrix}$. The reduced rank property results since the rank of the $k \times m$ matrix $\Pi \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$ can never exceed $m-1$. The specification $\Pi \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$ corresponds with the identified representation of a reduced rank matrix (3.4) with $r = m-1$, $A = \beta$, $Z_1 = \Pi$ $(k = k, m = m)$. So, the reduced rank restriction implied by the sem (4.1) on the parameters of a standard linear model,

$$\begin{pmatrix} y_1 & Y_2 \end{pmatrix} = X\Phi + \varepsilon, \tag{4.3}$$

where $\Phi : k \times m$, unrestricted; can be represented by the variables spanning the reduced rank restriction implied by (3.4). When $\lambda : (k-m+1) \times 1$, represents the parameters spanning the reduced rank restriction on Φ and satisfies conditions 1 and 2, it is thus equal to Y in (3.9). Because Φ is identical to X in (3.6), it can therefore be (unrestrictedly) specified as

$$\Phi = \Pi \begin{pmatrix} \beta & I_{m-1} \end{pmatrix} + \Pi_{\perp}\lambda \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_{\perp}, \tag{4.4}$$

where $\Pi_\perp$ and $\begin{pmatrix} \beta & I_{m-1} \end{pmatrix}_\perp$ are equivalent to $Z_\perp$, $W_\perp$ in (3.6). Both conditions 1 and 2 are satisfied by λ, see 3.1, such that when $\lambda = 0$ in (4.4), (4.3) is equivalent to (4.2) and the sem (4.1) therefore corresponds with the linear model (4.3) under the restriction that its parameter Φ is such that $\lambda = 0$.

4.1.2. Error Correction Cointegration Model

The reduced rank property of an error correction cointegration model (eccm) is obvious from its specification, see Engle and Granger (1987) and Johansen (1991),

$$\Delta x_t = \alpha'\beta' x_{t-1} + \varepsilon_t, \qquad t = 1, ..., T, \tag{4.5}$$

where β, $\alpha' : k \times r$, $\beta = (I_r\ \beta_2')'$, $\beta_2 : (k-r) \times r$, $x_t : k \times 1$, $\varepsilon_t : k \times 1$, is i.i.d. normal with mean 0 and covariance matrix Ω; that in matrix notation reads,

$$\Delta X = X_{-1}\beta\alpha + \varepsilon, \tag{4.6}$$

where $\Delta X = (\Delta x_1 ... \Delta x_T)'$, $X_{-1} = (x_0 ... x_{T-1})'$, $\varepsilon = (\varepsilon_1 ... \varepsilon_T)'$. The rank of the $k \times k$ parameter matrix $\beta\alpha$ cannot exceed r. Furthermore, the normalizing restrictions imposed on β are such that $\beta\alpha$ corresponds with the identified representation of a reduced rank parameter matrix (3.5) with $r = r$, $B = \beta_2$, $W_2 = \alpha$ $(k = k, m = m)$. The eccm can therefore be specified as a reduced rank restriction on the parameters of the linear model,

$$\Delta X = X_{-1}\Gamma + \varepsilon, \tag{4.7}$$

where $\Gamma : k \times m$, unrestricted. Because the specification of $\beta\alpha$ is identical to (3.5), Γ is equal to X in (3.6),

$$\Gamma = \beta\alpha + \beta_\perp \lambda \alpha_\perp, \tag{4.8}$$

where $\beta_\perp$, $\alpha_\perp$ are identical to $Z_\perp$, $W_\perp$ in (3.6). λ is identical to Y in (3.9) and since it satisfies both conditions 1 and 2, the eccm (4.6) is equivalent to the linear model (4.7) with its parameter Γ restricted to values for which $\lambda = 0$. We note that (4.5) represents the simplest specification of an eccm and various other variables like lags of Δx_t and deterministic components can be added to it. For notational convenience we only discuss (4.5).

4.1.3. Implications of conditional densities for reduced rank models

The conditional density of a matric-variate normal given rank reduction, that is discussed in 3.1, can be used to analyze the sem, eccm and other reduced rank models in both Bayesian and classical analyzes. In Bayesian analysis, the parameters are considered to be a drawing from a pdf and a Bayesian analysis of the reduced rank parameter models therefore results from the conditional density of a matric-variate normal given rank reduction. We discuss the consequences of this in section 5. In classical analysis, limiting and small sample distributions of the least squares estimator of the parameters of the unrestricted reduced form of a sem are normal, see Hausman (1983) and Phillips (1983). The limited information maximum likelihood (liml) estimator, see Anderson and Rubin (1949), implicitly reduces the rank of the least squares estimator of the unrestricted reduced form. It can therefore be considered as a drawing from the small sample distribution of the least squares estimator given that it has reduced rank. This implies that we can use the conditional density of the matric-variate normal given rank reduction to construct the small sample and limiting distribution of the liml estimator and to perform exact tests. We further discuss this in section 6.

4.2. ARMA(1,1) model

The exponential restriction discussed in 3.2 corresponds with the restriction which an AutoRegressive Moving Average model of order (1,1) (ARMA(1,1)) imposes on an autoregressive model. The ARMA(1,1) model reads,

$$(1-\rho L)x_t = (1-\alpha L)\varepsilon_t, \qquad t=1,...,T, \tag{4.9}$$

where L is the lagoperator, $L^j x_t = x_{t-j}$, ε_t i.i.d. normal with mean 0 and variance σ^2. The AR model implied by (4.9) reads

$$x_t = \sum_{i=1}^{t} c_i x_{t-i} + \varepsilon_t, \tag{4.10}$$

where $c_1 = \rho - \alpha$, $c_i = \alpha c_{i-1}$, $i = 2,...,T$, $c = (c_1...c_T)'$. We note that to construct (4.10), we assumed a nonzero value for x_0 and zero values of $x_{-1},...,x_{-\infty}$. Given these assumptions, the implied AR polynomial can be obtained from the ARMA polynomial even when the invertibility conditions of the MA polynomial are not satisfied, see Fuller (1976). The restrictions on c_i correspond with the restrictions (3.16) with $k = T$. We can thus specify the ARMA(1,1) model (4.9),

$$y = \rho y_{-1} - \alpha\varepsilon_{-1} + \varepsilon, \tag{4.11}$$

where $y = (x_1...x_T)'$, $y_{-1} = (x_0...x_{T-1})'$, $\varepsilon_{-1} = (\varepsilon_0...\varepsilon_T)'$, $\varepsilon = (\varepsilon_1...\varepsilon_T)'$ $(\varepsilon_0 = y_0)$, as the restriction $\lambda = 0$ on the parameter θ of the linear model,

$$y = Z\theta + \varepsilon, \tag{4.12}$$

where $Z = (y_{-1}...y_{-T})$, $y_{-i} = (0...0\ y_0...y_{T-i})' : T \times 1$, and θ, α, ρ, λ are identical to resp. x, z, $x_1 + z$ and y from (3.18). Since λ is equivalent to y in (3.18), it satisfies conditions 1 and 2 and the ARMA(1,1) model is therefore equivalent to the linear model (4.12) with its parameter θ restricted to values for which $\lambda = 0$. We can thus use the conditional density of the normal given the exponential restriction constructed in 3.2 to analyze the ARMA(1,1) model. Since the parameters are considered to be a drawing from a pdf in Bayesian analysis, a Bayesian analysis of ARMA models results from this conditional pdf. We discuss this more elaborately in section 5. In classical analysis and under stationarity, the limiting distribution of the least squares estimator of θ (c) is normal, see Fuller (1976). The maximum likelihood estimator of (ρ, α) implicitly imposes the restriction $\lambda = 0$ (or equivalently $c_i = \alpha c_{i-1}$, $i = 1,...,T$), on the least squares estimator of θ (c). We can therefore construct the conditional density which results when we impose these restrictions on the limiting distribution of θ (c). This then results in the limiting distribution of (α, ρ). We further discuss this in section 6.

5. Bayesian Analysis

The unique conditional densities given the restrictions that result from sections 2 and 3 can be used for different purposes in Bayesian analysis. In the following subsections we discuss four topics where the conditional densities appear. First, we discuss how these densities are used to construct general classes of priors, like noninformative and conjugate, for the parameters of the from the restrictions resulting models. Second, given a specified prior on the parameters of the analyzed nested model, the conditional densities show us the prior on the parameters of the encompassing model that implies it. When the latter model is linear in the parameters, each property of its' parameters prior is directly reflected in the posterior and therefore enables us to verify the plausibility of the specified prior. Third, we show that Bayes factors for model comparison result from the conditional densities and fourth, the conditional densities allow one to construct invariant Bayesian test statistics, Bayesian Score Statistics, to test the validity of the nested model. The next four subsections briefly discuss each of the above topics.

5.1. General Classes of Priors

The sufficient conditions, lemma 1 and theorem 2.1 not only allow us to construct unique conditional densities that are based on the normal density, as in section 3, but can be applied to other continuous densities as well. We can therefore use these conditional densities to construct the general classes of priors, like noninformative and (natural) conjugate, for the parameters of nested models that result from restrictions, which satisfy the conditions for unique conditioning, on the parameters of encompassing models where the general classes of priors are already defined for. Examples of these kind of models are given in the previous section where the sem, eccm and ARMA(1,1) are all restrictions, for which we can construct unique conditional densities, on the parameters of linear models where general classes of priors are already defined for.

Given the parameters of the encompassing model, say Θ, the parameters of the analyzed nested model, say φ, and the parameters spanning the restriction, that result in the nested model when imposed on the parameters of the encompassing linear model and satisfy conditions 1 and 2, say λ, a prior $p(\Theta)$ specified on Θ thus implies the prior

$$\begin{aligned} p_r(\varphi) &\propto p(\varphi,\lambda)|_{\lambda=0} \\ &\propto p(\Theta(\varphi,\lambda))|_{\lambda=0}|J(\Theta,(\varphi,\lambda))|_{\lambda=0}|, \end{aligned} \tag{5.1}$$

where r stands for restricted to indicate that it is not the marginal prior of φ in the encompassing model; on the parameters of the nested model φ. For the parameters of linear models, general classes of priors like noninformative and (natural) conjugate are specified, see *e.g.* Berger (1985), Poirier (1995) and Zellner (1971). (5.1) shows how to extend these classes of priors to the parameters of nested models that result from a restriction which satisfies conditions 1 and 2 on the parameters of an encompassing linear model. Conjugate priors for the parameters of the models discussed in the previous section therefore directly result from the conditional densities from section 3. This results since apart from the covariance parameters, conjugate priors for the parameters of linear models are normal densities. In section 3, we imposed restrictions on these normal densities that exactly correspond with the sem, eccm and ARMA(1,1) model and constructed the conditional densities given these restrictions. These conditional densities are therefore identical to the priors resulting from (5.1) when $p(\Theta)$ is a normal density. Conjugate priors for the parameters of the sem and eccm therefore correspond with (3.12), when we use the appropriate expressions for W_i and Z_i, while a conjugate prior for the parameters of the ARMA(1,1) model corresponds with (3.20), when we use the appropriate expressions for x_1 and z. These conjugate priors result since no restrictions are specified on Ω or σ^2. In Kleibergen and Hoek (1996), Kleibergen and Paap (1998), Kleibergen and van Dijk (1998), and Kleibergen and Zivot (1998), the resulting priors are used in Bayesian analyzes of respectively arma, cointegration, simultaneous equations and instrumental variable regression models.

The posteriors of the parameters are proportional to the product of the prior and the likelihood and therefore have an identical structure as (5.1) since the likelihood is continuous in the parameters,

$$\begin{aligned} p_r(\varphi|D) &\propto p_r(\varphi)L(D|\varphi) \\ &\propto p(\Theta(\varphi,\lambda))|_{\lambda=0}|J(\Theta,(\varphi,\lambda))|_{\lambda=0}|L(D|\Theta(\varphi,\lambda))|_{\lambda=0} \\ &\propto p(\Theta(\varphi,\lambda)|D)|_{\lambda=0}|J(\Theta,(\varphi,\lambda))|_{\lambda=0}|. \end{aligned} \tag{5.2}$$

This also shows that the conjugate priors resulting from (5.1) are posteriors of artificial datasets.

In Kleibergen and Zivot (1998), it is shown that also the Jeffreys' prior in the instrumental variable regression model, which is proportional to the square root of the determinant of information matrix, results from the conditional density approach as it leads to a posterior of the parameters of the instrumental variable regression model that is the unique conditional posterior of the parameters of a linear model given that a specific parameter matrix has reduced rank.

5.2. Implied Prior on Parameters Encompassing Model

The relationship between the prior of the parameters of a nested model and the prior of the parameters of the encompassing model (5.1) also implies the prior on the parameters of the encompassing model when we directly specify the prior on the parameters of the nested model. This results from (2.9) since when $p_r(\varphi)$ is specified

$$\begin{aligned} p(\Theta)|_{g(\Theta)=0} &\propto p(\varphi(\Theta),\lambda(\Theta))|_{g(\Theta)=0}|J((\varphi,\lambda),\Theta)|_{g(\Theta)=0}| \\ &\propto p_r(\varphi(\Theta))|J((\varphi,\lambda),\Theta)|_{g(\Theta)=0}| \\ &\propto p_r(\varphi(\Theta))(|J(\Theta,(\varphi(\Theta),\lambda(\Theta)))|_{\lambda=0}|)^{-1}, \end{aligned} \tag{5.3}$$

where we have used that $\lambda=0$ is equivalent with $g(\Theta)=0$ and that $(\varphi,\lambda=0)$ uniquely defines Θ (conditions 1, 2). Only the prior for values of Θ for which $g(\Theta)=0$ directly results from $p_r(\varphi)$ such that the implied prior for general values of Θ reads,

$$p(\Theta) \propto q(\Theta)p(\Theta)|_{g(\Theta)=0}, \tag{5.4}$$

where $q(\Theta)\equiv 1$ when $g(\Theta)=0$.

Allthough the prior of φ only defines the prior of Θ on a subset of the space on which it is defined, the subset for which $g(\Theta)=0$, such that we donot know its behavior in general, it is nevertheless a convenient tool to verify the plausibility of the prior of φ. This especially holds when the nested model is not linear in its parameters while the encompassing model is. In that case, the properties of the implied prior on the parameters of the encompassing model appear in an identical way in the posterior of the parameters of the nested model. This results because the encompassing model is linear in the parameters such that the properties of the prior are also present in the posterior. The nested model is, however, nonlinear in the parameters such that the updating from the prior specified on the parameters of the nested model to its posterior is less obvious. Furthermore, to analyze the nested model we donot need to know the behavior of $p(\Theta)$ for values of Θ for which $g(\Theta)\neq 0$ since these values donot correspond with the analyzed model which implies that $g(\Theta)=0$.

In Kleibergen and Zivot (1998), the implied prior on the parameters of an encompassing linear model is used to analyze several "diffuse priors" on the parameters of an instrumental variable regression model. As this model is nonlinear in the parameters, it is not directly obvious how the prior is updated to the posterior and therefore how the prior exactly affects the posterior. The encompassing model is linear and, since we know how priors are updated to posteriors for linear models, the implied prior thus exactly shows how the prior affects the posterior. It therefore enables us to analyze the sensitivity of the posterior with respect to the prior without the need to compute the posterior. This is shown in Kleibergen and Zivot (1998) as all stylized facts of the marginal posteriors using the different "diffuse priors" are straightforwardly shown by the implied priors on the parameters of the encompassing linear model. The latter is, however, constructed analytically and is therefore a convenient tool to analyze the plausibility of the specified prior on the parameters of the instrumental variable regression model. So, the properties of priors specified on the parameters of the instrumental variables regression model for the posterior are revealed by the implied prior on the parameters of the encompassing linear model and are not obvious from the specified priors themselves.

5.3. Model Comparison using the Bayes Factor

In Bayesian analysis, model comparison is often conducted using the Bayes factor, see *e.g.* Berger (1985) and Poirier (1995). The conditions for unique conditional densities, conditions 1 and 2, show how the Bayes factor to compare a nested model that results from a restriction that satisfies conditions 1 and 2 on the parameters of an encompassing model, with this encompassing model is constructed. Depending on the specification of the prior, the Bayes factor that results then either corresponds with the Savage-Dickey density ratio, see Dickey (1971), or its generalization by Verdinelli and Wasserman, see Verdinelli and Wasserman (1995).

Consider a nested model with parameter $\varphi : m \times 1$, and an encompassing model with parameter Θ, $\Theta(\varphi, \lambda) : k \times 1$. The nested model corresponds with the restriction $g(\Theta) = 0$ or equivalently $\lambda = 0 : (k-m) \times 1$, that satisfies conditions 1 and 2, on the parameters of the encompassing model. A proper prior $p_r(\varphi)$ is specified on the parameters of the nested model and a proper prior $p(\Theta)$ is specified on the parameters of the encompassing model. As shown in section 5.1, the prior $p(\Theta)$ specified on the parameters of the encompassing model implies the prior $p(\varphi, \lambda) = p(\Theta(\varphi, \lambda))|J(\Theta, (\varphi, \lambda))|$, (5.1), on the specification of the encompassing model in terms of the parameter of the nested model φ and the parameter representing the restriction λ. The Bayes factor for comparing the nested and encompassing model then reads, see Verdinelli and Wasserman (1995),

$$\begin{aligned}
BF(\lambda &= 0) = \frac{\int p_r(\varphi)L(D|\varphi)d\varphi}{\int p(\Theta)L(D|\Theta)d\Theta} = \frac{\int p_r(\varphi)L(D|\varphi)d\varphi}{\iint p(\varphi,\lambda)L(D|\varphi,\lambda)d\lambda d\varphi} \qquad (5.5)\\
&= p(\lambda|D)|_{\lambda=0}\frac{\int\left(\frac{p_r(\varphi)L(D|\varphi,\lambda)|_{\lambda=0}}{p(\lambda|D)|_{\lambda=0}}\right)d\varphi}{\iint p(\varphi,\lambda)L(D|\varphi,\lambda)d\lambda d\varphi}\\
&= p(\lambda|D)|_{\lambda=0}\frac{\int\left(\frac{p_r(\varphi)L(D|\varphi,\lambda)|_{\lambda=0}p(\varphi|\lambda,D)|_{\lambda=0}}{p(\lambda|D)|_{\lambda=0}p(\varphi|\lambda,D)|_{\lambda=0}}\right)d\varphi}{\iint p(\varphi,\lambda)L(D|\varphi,\lambda)d\lambda d\varphi}\\
&= p(\lambda|D)|_{\lambda=0}\frac{\int\left(\frac{p_r(\varphi)L(D|\varphi,\lambda)|_{\lambda=0}p(\varphi|\lambda,D)|_{\lambda=0}}{p(\varphi,\lambda|D)|_{\lambda=0}}\right)d\varphi}{\iint p(\varphi,\lambda)L(D|\varphi,\lambda)d\lambda d\varphi}
\end{aligned}$$

$$
\begin{aligned}
&= p(\lambda|D)|_{\lambda=0}\frac{\int\left(\frac{p_r(\varphi)L(D|\varphi,\lambda)|_{\lambda=0}p(\varphi|\lambda,D)|\lambda=0}{\left(\frac{p(\varphi,\lambda)|_{\lambda=0}L(D|\varphi,\lambda)|_{\lambda=0}}{\iint p(\varphi,\lambda)L(D|\varphi,\lambda)d\lambda d\varphi}\right)}\right)d\varphi}{\iint p(\varphi,\lambda)L(D|\varphi,\lambda)d\lambda d\varphi} \\
&= p(\lambda|D)|_{\lambda=0}\int\left(\frac{p_r(\varphi)}{p(\varphi,\lambda)|_{\lambda=0}}p(\varphi|\lambda,D)|_{\lambda=0}\right)d\varphi \\
&= \int\left(p_r(\varphi)\frac{p(\varphi,\lambda|D)|_{\lambda=0}}{p(\varphi,\lambda)|_{\lambda=0}}\right)d\varphi
\end{aligned}
$$

where we have used that $p(\varphi,\lambda|D)=\frac{p(\varphi,\lambda)L(D|\varphi,\lambda)}{\iint p(\varphi,\lambda)L(D|\varphi,\lambda)d\lambda d\varphi}$, and when

$$
p_r(\varphi)=p(\varphi|\lambda)|_{\lambda=0}=\frac{p(\varphi,\lambda)|_{\lambda=0}}{\int p(\varphi,\lambda)|_{\lambda=0}d\varphi}=\frac{p(\Theta(\varphi,\lambda))|_{\lambda=0}|J(\Theta,(\varphi,\lambda))|_{\lambda=0}|}{\int p(\Theta(\varphi,\lambda))|_{\lambda=0}|J(\Theta,(\varphi,\lambda))|_{\lambda=0}|d\varphi}, \tag{5.6}
$$

this expression simplifies to the Savage-Dickey density ratio, see Dickey (1971), *i.e.* the ratio of the marginal posterior and prior height in the point of the restriction,

$$
\begin{aligned}
SD(\lambda &= 0)=p(\lambda|D)|_{\lambda=0}\int\left(\frac{p_r(\varphi)}{p(\varphi,\lambda)|_{\lambda=0}}p(\varphi|\lambda,D)|_{\lambda=0}\right)d\varphi \qquad (5.7)\\
&= p(\lambda|D)|_{\lambda=0}\int\left(\frac{p_r(\varphi)}{p(\varphi|\lambda=0)p(\lambda)|_{\lambda=0}}p(\varphi|\lambda,D)|_{\lambda=0}\right)d\varphi \\
&= \int\frac{\left(\frac{p(\varphi,\lambda)|_{\lambda=0}}{\int p(\varphi,\lambda)|_{\lambda=0}d\varphi}\right)p(\varphi,\lambda|D)|_{\lambda=0}}{p(\varphi,\lambda)|_{\lambda=0}}d\varphi \\
&= \frac{\int p(\varphi,\lambda|D)|_{\lambda=0}d\varphi}{\int p(\varphi,\lambda)|_{\lambda=0}d\varphi} \\
&= \frac{p(\lambda|D)|_{\lambda=0}}{p(\lambda)|_{\lambda=0}}.
\end{aligned}
$$

The Bayes factors (5.5), (5.7) show that the unique conditional densities are necessary to obtain the Bayes factor for comparing a nested model with an encompassing model using the Savage-Dickey density ratio or its generalization by Verdinelli and Wasserman, see Dickey (1971) and Verdinelli and Wasserman (1995).

The specification of the prior $p_r(\varphi)$ compared to $p(\Theta)$ is important for the coherency of the Bayes Factor. When $p_r(\varphi)$ strongly differs from $p(\varphi|\lambda)|_{\lambda=0}$ in functional form, the resulting Bayes Factors become (weakly) incoherent, see Berger (1985) and Poirier (1995). The ratio of the generalized Savage-Dickey density ratio (5.5) and the Savage-Dickey density ratio (5.7) is therefore important as it reveals that part of the Bayes Factor that can be attributed to not specifying the prior $p_r(\varphi)$ according to the unique conditional density arguments from section 2,

$$
\begin{aligned}
\frac{BF(\lambda=0)}{SD(\lambda=0)} &= \frac{\int p_r(\varphi)L(D|\varphi)d\varphi}{\int p(\varphi|\lambda)|_{\lambda=0}L(D|\varphi,\lambda)|_{\lambda=0}d\varphi} \qquad (5.8)\\
&= \frac{\int\left(p_r(\varphi)\frac{p(\varphi,\lambda|D)|_{\lambda=0}}{p(\varphi,\lambda)|_{\lambda=0}}\right)d\varphi}{\left(\frac{\int p(\varphi,\lambda|D)|_{\lambda=0}d\varphi}{\int p(\varphi,\lambda)|_{\lambda=0}d\varphi}\right)}
\end{aligned}
$$

$$
\begin{aligned}
&= \int \left(p_r(\varphi) \frac{\left(\frac{p(\varphi,\lambda|D)|_{\lambda=0}}{\int p(\varphi,\lambda|D)|_{\lambda=0} d\varphi} \right)}{\left(\frac{p(\varphi,\lambda)|_{\lambda=0}}{\int p(\varphi,\lambda)|_{\lambda=0} d\varphi} \right)} \right) d\varphi \\
&= \int \left(p_r(\varphi) \frac{p(\varphi|\lambda, D)|_{\lambda=0}}{p(\varphi|\lambda)|_{\lambda=0}} \right) d\varphi.
\end{aligned}
$$

Using the implied prior on the parameters Θ of the encompassing model (5.3), (5.8) can also be specified as,

$$
\begin{aligned}
\frac{BF(\lambda = 0)}{SD(\lambda = 0)} &= \int \left(p_r(\varphi) \frac{p(\varphi|\lambda, D)|_{\lambda=0}}{p(\varphi|\lambda)|_{\lambda=0}} \right) d\varphi. \qquad (5.9) \\
&= \int_{g(\Theta)=0} \left(p_r(\Theta) \frac{\left(\frac{p(\Theta|D)}{\int_{g(\Theta)=0} p(\Theta|D) d\Theta} \right)}{\left(\frac{p(\Theta)}{\int_{g(\Theta)=0} p(\Theta) d\Theta} \right)} \right) d\Theta \\
&= \int_{g(\Theta)=0} a(\Theta) \frac{b(\Theta)}{c(\Theta)} d\Theta,
\end{aligned}
$$

where $a(\Theta) = p_r(\Theta)$, $b(\Theta) = \frac{p(\Theta|D)}{\int_{g(\Theta)=0} p(\Theta|D)d\Theta}$, $c(\Theta) = \frac{p(\Theta)}{\int_{g(\Theta)=0} p(\Theta)d\Theta}$ and $\int_{g(\Theta)=0} a(\Theta)d\Theta = \int_{g(\Theta)=0} b(\Theta)d\Theta = \int_{g(\Theta)=0} c(\Theta)d\Theta = 1$, which sometimes reveals the properties of the ratio of the Bayes Factor and the Savage-Dickey density ratio in a more straightforward way.

Especially when the nested model is such that specific parameters are locally nonidentified, as in the sem, eccm and arma models, and the proper prior of the parameters of the nested model doesnot capture this feature, Bayes factors can become incoherent. This results even though we have specified a proper prior and instead of comparing the model with $\lambda = 0$ with the model with $\lambda \neq 0$, the Bayes factor then essentially compares the model with $\lambda = 0$ and the other parameters at their values leading to local nonidentification with the unrestricted model. So, the Bayes factor then doesnot test the specified hypothesis and a fair bet is thus not possible, *i.e.* the Bayes factor is then incoherent.

The (generalized) Savage-Dickey density ratio is only intended to compare nested models with encompassing models and therefore involves unique conditional densities. The Bayes factor for comparing models that are not in someway nested within one another can still be obtained using the (generalized) Savage-Dickey density ratio when both models have an encompassing model in common. When this is not the case, the (generalized) Savage-Dickey density ratio can not be used to obtain the Bayes factor. In econometrics, the (generalized) Savage-Dickey density ratio is closely related to the posterior information criterium of Phillips and Ploberger, see Phillips and Ploberger (1994,1996) and Phillips (1996).

In Kleibergen and Paap (1998), the Savage-Dickey density ratio (5.7) is used to compare eccms with different cointegration ranks. All these models are nested in the same encompassing linear model and one prior is specified on the parameters of that model from which the priors on the parameters of the models with the different cointegration ranks are then constructed from using (5.1). In Kleibergen and Hoek (1996), the Savage-Dickey density ratio (5.7) is used to compare ARMA models with the same summed lag lengths of their AR and MA polynomials. So, an ARMA(2,1) is compared with an AR(3), ARMA(1,2) and MA(3). In both of these applications of the Savage-Dickey density ratio the analyzed models are such that specific parameters are locally nonidentified. By deducing the priors of the nested models from the proper conjugate prior on the parameters of the encompassing linear model, the problem of the local nonidentification is dealt with in a straightforward way.

5.4. Bayesian Score Statistics

The unique conditional densities imply that we can consider the posterior of the parameters of a nested model as proportional to the conditional posterior of the parameters of the encompassing model given that the parameters which represent the restriction of equality of the two models and satisfy conditions 1 and 2 are equal to zero. So, we essentially consider the parameter configuration of the nested model as a drawing from the posterior of the parameters of the encompassing model that is such that the parameters which represent the restriction are exactly equal to zero. In Kleibergen and Paap (1998), test statistics, to which they refer as Bayesian Score Statistics (BSS), are constructed that evaluate whether the drawing from the posterior of the parameters of the encompassing model that leads to the nested model lies in the tails of the posterior of the parameters of the encompassing model. The BSS thus shows the plausibility of the restriction imposed on the parameters of the encompassing model. The BSS only uses the posterior of the parameters of the nested model and is invariant to parameter transformations. It is therefore different from the highest posterior density based test statistics that are discussed in Box and Tiao (1973). It is somewhat related to the Bayesian test-statistic that is constructed in Poirier (1988) but differs in construction from that statistic as it is based on an improper prior.

The derivation of the BSS consists of four steps:

1. The posterior of the parameters of the nested model is specified as the product of a standard conditional density, of the parameter reflecting the restriction evaluated in the point where the restriction holds, and a marginal density of the parameters of the nested model, like in (3.12) and (3.20).

2. In the unrestricted case of no imposed restriction, the conditional density of the parameter reflecting the restriction is used to construct a function of all the parameters that is such that it, *i.e.* the BSS, is stochastically independent of the parameters that donot represent the restriction, *i.e.* the parameters of the nested model, and has a standardized distribution.

3. Since the BSS is stochastically independent of the parameters that donot represent the restriction, it retains the same distribution when we take the expectation of it with respect to these parameters and the marginal density of them that results from the decomposition of the posterior of the parameters of the nested model.

4. The BSS to test for the plausibility of the restriction imposed by the nested model on the parameters of the encompassing model results by taking the expectation of the BSS over the parameters of the nested model with respect to the posterior of the parameters of the nested posterior instead of the marginal density used in the previous step. We then compare the value of the resulting statistic with its distribution in the unrestricted case.

The BSS is thus a test statistic that is computed under the null hypothesis but its distribution results from the alternative hypothesis. In Kleibergen and Paap (1998) a proof of the validity of the BSS and its exact functional form are given. Furthermore, the different specifications of it for several commonly tested hypothezes in linear models, like omitted variables, serial correlation, cointegration, are constructed. As the computation of the BSS only involves the nested model it is a convenient diagnostic to test the plausibility of the implicitly, by the nested model, imposed restrictions.

6. Classical Analysis

The previous section showed Bayesian applications of the unique conditional densities. The unique conditional densities also have consequences for classical statistical analyzes of econometric and time series models. In the following subsections, we briefly discuss two of these consequences. The first consequence involves the limiting and small sample distributions of maximum likelihood estimators (mles). The second one deals with exact tests of hypotheses.

6.1. Distributions of Maximum Likelihood Estimators

The mle satisfies the first order condition (foc) for a maximum of the log-likelihood. When the mle is the unique solution to the foc, we can in principle construct its limiting and/or small sample distribution. The foc can be specified such that it relates the mle of the parameters of the analyzed nested model to the mle of the parameters of an encompassing model. Both these mles can be considered as realizations of random variables. For all realizations of the mle of the parameters of the encompassing model, the uniqueness condition implies that there is exactly one solution to the foc for a maximum of the log-likelihood of the nested model. This is in general not possible and this condition therefore imposes a restriction on the random variable that characterizes the mle of the parameters of the encompassing model, see Kleibergen (1998). When this restriction satisfies the sufficient conditions for the existence of an unique conditional density, the concept of an unique conditional density can be used to construct the limiting or small sample distribution of the mle of the parameters of the nested model from the limiting or small sample distribution of the mle of the parameters of the encompassing model.

Consider, for example, the model,

$$y = Xf(\varphi) + \varepsilon, \tag{6.1}$$

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 φ is defined. Examples of models like (6.1) are the simultaneous equation model, the autoregressive moving average model and numerous others.

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

$$\begin{aligned}
\frac{1}{\sigma^2}\left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)' X'\left(y - Xf(\hat{\varphi})\right) &= 0 \Leftrightarrow \qquad (6.2)\\
\frac{1}{\sigma^2}\left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)' X'X\left((X'X)^{-1}X'y - f(\hat{\varphi})\right) &= 0 \Leftrightarrow \\
\frac{1}{\sigma^2}\left(\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\right)' X'X\left(\hat{\Phi} - f(\hat{\varphi})\right) &= 0 \Leftrightarrow \\
\left((X'X)^{\frac{1}{2}}\frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}}\sigma^{-1}\right)'\left((X'X)^{\frac{1}{2}}\hat{\Phi}\sigma^{-1} - (X'X)^{\frac{1}{2}}f(\hat{\varphi})\sigma^{-1}\right) &= 0 \Leftrightarrow \\
\left(\frac{\partial r}{\partial \varphi'}|_{\hat{\varphi}}\right)'\left(\hat{\Theta} - r(\hat{\psi}(\hat{\varphi}))\right) &= 0 \Leftrightarrow \\
\left(\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}(\hat{\varphi})}\right)\left(\frac{\partial \psi}{\partial \varphi'}|_{\hat{\varphi}}\right)\right)'\left(\hat{\Theta} - r(\hat{\psi}(\hat{\varphi}))\right) &= 0 \Leftrightarrow \\
\left(\frac{\partial \psi}{\partial \varphi'}|_{\hat{\varphi}}\right)'\left(\frac{\partial r}{\partial \psi'}|_{\hat{\psi}(\hat{\varphi})}\right)'\left(\hat{\Theta} - r(\hat{\psi}(\hat{\varphi}))\right) &= 0,
\end{aligned}$$

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 ψ and φ exists such that $\frac{\partial\varphi}{\partial\psi'}$ is nonsingular for all φ, ψ; $\hat{\psi}$ and $\hat{\varphi}$ stand for the mles of the specific parameters; and we have used the chain rule of differentiation, $\frac{\partial r}{\partial\varphi'}|_{\hat{\varphi}} = \left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}(\hat{\varphi})}\right)\left(\frac{\partial\psi}{\partial\varphi'}|_{\hat{\varphi}}\right)$. The foc (6.2) relates the mle of φ, $\hat{\varphi}$, which can be uniquely solved for from $\hat{\psi}$, to the "t-values" $\hat{\Theta}$ of the linear model,

$$y = X\Phi + \varepsilon, \tag{6.3}$$

and we can solve for $\hat{\varphi}$ from $\hat{\Theta}$ using (6.2). Since $\frac{\partial\varphi}{\partial\psi'}$ is nonsingular everywhere, the foc (6.2) is equivalent with

$$\left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}(\hat{\varphi})}\right)'\left(\hat{\Theta} - r(\hat{\psi}(\hat{\varphi}))\right) = 0, \tag{6.4}$$

and we can also solve for $\hat{\psi}$ from $\hat{\Theta}$ using (6.4) and then solve for $\hat{\varphi}$ from $\hat{\psi}$.

We reflect the foc in its most primitive form such that it does not involve any datamatrices like $X'X$. This is achieved by specifying it in terms of the "t-values" $\hat{\Theta}$ which can be considered as a realization of a random variable with an identity covariance matrix. Because of this latter property, the log-likelihood is strictly quadratic in $\hat{\Theta}$ as $\hat{\Theta}$ is not weighted by a specific datadependent matrix in the log-likelihood. Restrictions can therefore directly be imposed on $\hat{\Theta}$ and do not need to be scaled by an additional datadependent matrix, estimator or parameter.

In Kleibergen (1998), it is shown that the uniqueness of the mle of the parameters of the nested model, $\hat{\psi}$, as the solution to the foc for all realizations of $\hat{\Theta}$ imposes a restriction on the random variable $\hat{\Theta}$. When $\frac{\partial r}{\partial\psi'}$ has full rank for all values of ψ, this restriction is,

$$\hat{\Theta} - r(\hat{\psi}) \equiv 0. \tag{6.5}$$

From realizations of $\hat{\Theta}$ that satisfy this restriction, we can uniquely solve for $\hat{\psi}$, and thus also for $\hat{\varphi}$, and construct the densities of these random variables. In order to be able to do this, the restriction (6.5) has to satisfy the sufficient conditions for the existence of an unique conditional density and the density function of $\hat{\Theta}$ has to be known. When these conditions hold, we can construct the density of $\hat{\psi}$ from the unique conditional density of $\hat{\Theta}$ given that (6.5) 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.

For an observed dataset, the realized value of $\hat{\Theta}$ never satisfies the restriction (6.5) which might be considered as an argument against the validity of the previously discussed approach. However, since the unrestricted $\hat{\Theta}$ has more elements than $\hat{\psi}$, it consists of two distinct parts, one part from which we solve $\hat{\psi}$ and one part which we essentially discard when we solve for $\hat{\psi}$. The part of $\hat{\Theta}$ where $\hat{\psi}$ is solved from does satisfy the restriction (6.5) and that part is the only part in which we are interested to determine the density of $\hat{\psi}$. Estimation algorithms that estimate $\hat{\psi}$ from $\hat{\Theta}$ using the foc (6.4) also operate in this way and also (implicitly) discard specific elements of $\hat{\Theta}$. The main difference between such an approach and the one discussed above is that the latter approach explicitly from the outset shows which part of $\hat{\Theta}$ it is going to use to determine $\hat{\psi}$ while the estimation algorithm implicitly discards certain parts of $\hat{\Theta}$ in estimating $\hat{\psi}$.

6.1.1. Small Sample Distributions of Maximum Likelihood Estimators

When the density function of $\hat{\Theta}$ in (6.4) 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 mle $\hat{\varphi}$. When $S = X'X$ and has a fixed full rank value, such that X is nonstochastic, $\frac{\partial r}{\partial \psi'}$ has full rank, $\frac{\partial \varphi}{\partial \psi'}$ invertible $\forall \varphi$; then, as discussed previously, the uniqueness of the mle as solution to the foc (6.2) imposes the restriction,

$$\hat{\Theta} - r(\hat{\psi}) \equiv 0, \tag{6.6}$$

on the random variable $\hat{\Theta}$, which has an identity covariance matrix, as the mle $\hat{\psi}$ is the unique solution to the foc (6.2) for all values of the random variable $\hat{\Theta}$. When (6.6) satisfies the sufficient conditions for the existence of an unique conditional density, the pdf of the mle $\hat{\varphi}$ results from the conditional pdf of $\hat{\Theta}$ given that (6.6) holds,

$$\begin{aligned} p_r(\hat{\varphi}) &\propto p_r(\hat{\psi}(\hat{\varphi}))|J(\hat{\psi}, \hat{\varphi})| \\ \text{and } p_r(\hat{\psi}) &\propto p(\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0} \\ &\propto p(\hat{\Theta}(\hat{\psi}, \hat{\lambda}))|_{\hat{\lambda}=0} |J(\hat{\Theta}, (\hat{\psi}, \hat{\lambda})|_{\hat{\lambda}=0}|, \end{aligned} \tag{6.7}$$

where $\hat{\lambda} : (k-m) \times 1$ represents the restriction (6.6) (because of the sufficient conditions for an unique conditional density), r stands for restricted to indicate that we donot use the marginal densities resulting from $\hat{\Theta}$ directly, and $J(\hat{\psi}, \hat{\varphi})$ is the jacobian of the transformation from $\hat{\psi}$ to $\hat{\varphi}$.

The small sample density constructed by using (6.7) results from an implicit argument as it is obtained from the property of the mle that it is the unique solution to the foc (6.2). The small sample distributions for mles discussed in the literature, see *e.g.* Phillips (1983) and Mariano and Sawa (1972), are all constructed using an explicit expression of the mle, *i.e.* the small sample distribution is constructed using the exact closed form analytical expression of the mle 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 therefore appealing as it can also be applied in cases where we cannot construct a closed form analytical expression of the mle.

We note that for an observed sample, $\hat{\Theta}$ is just a statistic and the foc (6.2) cannot imply the restriction (6.6). We can then just solve for $\hat{\varphi}$ from (6.2). By solving for $\hat{\varphi}$ we then, however, implicitly discard a part of $\hat{\Theta}$ and the part of $\hat{\Theta}$ from which $\hat{\varphi}$ is essentially solved does satisfy (6.6). So, to construct the small sample density of $\hat{\varphi}$, we can as well impose (6.6) from the outset and construct the density of $\hat{\varphi}$ from the conditional density of $\hat{\Theta}$ given (6.6).

6.1.2. Limiting Distribution of Maximum Likelihood Estimators

The foc (6.2) can also be used to obtain the limiting behavior of the mle $\hat{\varphi}$ when the limiting behavior of $\hat{\Theta}$ around its true value Θ_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 $\hat{\Theta}$,

$$\begin{aligned} \left((X'X)^{\frac{1}{2}} \frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}} \sigma^{-1}\right)' \left((X'X)^{\frac{1}{2}} \hat{\Phi} \sigma^{-1} - (X'X)^{\frac{1}{2}} f(\hat{\varphi})\sigma^{-1}\right) &\equiv 0 \quad (6.8) \\ \left((X'X)^{\frac{1}{2}} \frac{\partial f}{\partial \varphi'}|_{\hat{\varphi}} \sigma^{-1}\right)' \left((X'X)^{\frac{1}{2}} \left(\hat{\Phi} - f(\varphi_0)\right)\sigma^{-1} - (X'X)^{\frac{1}{2}} \left(f(\hat{\varphi}) - f(\varphi_0)\right)\sigma^{-1}\right) &\equiv 0. \end{aligned}$$

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,

$$p(x)=(2\pi)^{-\frac{1}{2}k}\exp\left[-\frac{1}{2}x'x\right]. \tag{6.9}$$

So, since $\sqrt{T}\left(\hat{\Theta}-\Theta_0\right)$ converges to a random variable and the condition (6.8) holds for every value $\hat{\varphi}$, it also implies that,

$$\begin{aligned}
\lim_{T\to\infty}\left[\sqrt{T}\left(S^{\frac{1}{2}}\frac{\partial f}{\partial\varphi'}\sigma^{-1}\right)'\left(\sqrt{T}S^{\frac{1}{2}}\left(\hat{\Phi}-f(\varphi_0)\right)\sigma^{-1}-\sqrt{T}S^{\frac{1}{2}}\left(f(\hat{\varphi})-f(\varphi_0)\right)\sigma^{-1}\right)\right] &\equiv 0\Leftrightarrow\\
\lim_{T\to\infty}\left[\sqrt{T}\left(\frac{\partial\psi}{\partial\varphi'}|_{\hat{\varphi}}\right)'\left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}}\right)'\left(\sqrt{T}\left(\hat{\Theta}-\Theta_0\right)-\sqrt{T}\left(r(\hat{\psi})-r(\psi_0)\right)\right)\right] &\equiv 0\Leftrightarrow\\
\lim_{T\to\infty}\left[\left(\frac{\partial r}{\partial\psi'}|_{\hat{\psi}}\right)'\left(x-\sqrt{T}\left(r(\hat{\psi})-r(\psi_0)\right)\right)\right] &\equiv 0
\end{aligned} \tag{6.10}$$

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 a continuous differentiable function of $\hat{\psi}$, $\psi: m\times 1$, ψ is an invertible function of φ. This results from the continuous mapping theorem, see Billingsley (1986). Identical to the case of the small sample distribution, there is an unique solution $\hat{\psi}$ to (6.10) for every possible value of x which is a random variable defined on R^k. Since $\frac{\partial r}{\partial\psi'}$ has full rank for all values of ψ, this imposes the restriction,

$$x-\lim_{T\to\infty}\sqrt{T}\left(r(\hat{\psi})-r(\psi_0)\right)\equiv 0, \tag{6.11}$$

on the random variable x. So, when (6.11) 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 (6.11) holds,

$$\sqrt{T}\left(r(\hat{\psi})-r(\psi_0)\right)\Rightarrow x|_{g(x)=0}, \tag{6.12}$$

where $g(x)=0$ is equivalent with (6.11).

We note that we obtain the limiting distribution of the mle of the parameters of the nested model by explicitly using condition (6.11) which results since the mle is the unique solution to the foc. For statistics resulting from an observed dataset this condition is not satisfied but the elements of these statistics from which the mle $\hat{\psi}$ is essentially solved, do satisfy (6.11).

6.1.3. Application of the unique conditional densities to liml estimator

In Kleibergen (1998), the limiting and small sample distribution of the limited information maximum likelihood (liml) estimator are constructed using the unique conditional densities. The limiting distribution that results is identical to the expressions discussed elsewhere in the literature, see *e.g.* Hausman (1983), and the obtained small sample density is identical to the sampling density. This thus shows the validity of constructing small sample and limiting distributions of mles using unique conditional densities and also of the concept of an unique conditional density itself.

6.2. Classical Score Statistics

The restrictions (6.6) and (6.12) imposed by the foc on random variables that characterize the small sample and limiting behavior of mles of the parameters of encompassing models can also be used to construct exact and asymptotic distributions of classical score statistics, see *e.g.* Engle (1984). These test statistics result quite naturally because the restrictions (6.6) and (6.12) are imposed on random variables with identity covariance matrices.

Consider the null hypothesis, $H_0 : \Phi_0 = f(\varphi)$, that makes (6.3) equal to (6.1). This hypothesis is equivalent with $H_0 : g(\Theta_0) = g(S^{\frac{1}{2}}\Phi_0\sigma^{-1}) = 0 : (k-m)\times 1$, that satisfies conditions 1 and 2, and implies that the null hypothesis amounts to $H_0 : \lambda_0 = 0$, see section 6.1. Note that Θ_0 is the true value of Θ.

In the case of the small sample distribution, the restriction imposed by the foc (6.6) is equivalent with the restriction $g(\hat{\Theta}) = 0$, that can be modelled using the parameter $\hat{\lambda}$, imposed on the random variable $\hat{\Theta}$, $\hat{\Theta} = S^{\frac{1}{2}}\hat{\Phi}\sigma^{-1}$. The small sample distribution of the maximum likelihood estimator $\hat{\varphi}$ then results as, see (6.7),

$$\begin{aligned} p_r(\hat{\varphi}) &\propto p_r(\hat{\psi}(\hat{\varphi}))\left|J(\hat{\psi},\hat{\varphi})\right| \\ \text{and } p_r(\hat{\psi}) &\propto p(\hat{\psi},\hat{\lambda})|_{\hat{\lambda}=0} \\ &\propto p(\hat{\Theta}(\hat{\psi},\hat{\lambda}))|_{\hat{\lambda}=0}\left|J\left(\hat{\Theta},\left(\hat{\psi},\hat{\lambda}\right)\right)\right|_{\hat{\lambda}=0}, \end{aligned} \tag{6.13}$$

where $\hat{\Theta} = S^{\frac{1}{2}}\hat{\Phi}\sigma^{-1}$, $r(\hat{\psi}) = S^{\frac{1}{2}}f(\hat{\varphi})\sigma^{-1}$, $\hat{\psi} : m \times 1$ and $\hat{\psi}$ has an invertible relationship with $\hat{\varphi}$. To test the null hypothesis, $H_0 : \lambda = 0$, we only use $p_r(\hat{\psi})$ and similar to section 5.4, $p_r(\hat{\psi})$ can often be specified as,

$$p_r(\hat{\psi}) \propto q(\hat{\lambda}|\hat{\psi})|_{\hat{\lambda}=0}q(\hat{\psi}), \tag{6.14}$$

where $q(\hat{\lambda}|\hat{\psi})$ and $q(\hat{\psi})$ are proper density functions of which $q(\hat{\lambda}|\hat{\psi})$ belongs to a standard class. The model obtained under the null hypothesis corresponds with $\lambda_0 = 0$. The conditional density $q(\hat{\lambda}|\hat{\psi})$ under the nullhypothesis, $H_0 : \lambda = 0$, can be used to verify whether (the imposed) $\hat{\lambda} = 0$ is a plausible drawing from the density $q(\hat{\lambda}|\hat{\psi})$ under $H_0 : \lambda = 0$. For this we use the classical score statistic that tests $H_0 : \lambda = 0$ against $H_1 : \lambda \neq 0$. The classical score statistic is a function of $(\hat{\lambda},\hat{\psi})$ for which we can construct its pdf under H_0 from $q(\hat{\psi},\hat{\lambda}) = q(\hat{\lambda}|\hat{\psi})q(\hat{\psi})$. The classical score statistic is such a function that it is stochastically independent from $\hat{\psi}$ and has a standardized density. The marginal pdf of the classical score statistic therefore results from $q(\hat{\psi},\hat{\lambda})$ and we can use it to obtain the critical values. We consider the classical score statistic for the observed dataset as a realization from this distribution and can compare it with the critical values to determine its significance. A similar reasoning leads to the asymptotic distribution of the classical score statistic which uses the limiting distributions of the mle that result from the unique conditional densities and are discussed in the previous subsection.

7. Conclusions

We defined sufficient conditions for the existence of an unique conditional density. Several econometric and time series models result from restrictions on encompassing linear models that satisfy these sufficient conditions. Both Bayesian and classical inference in these models then (can be considered to) result from these conditional densities. In Bayesian statistics the unique conditional densities allow us to construct (*i.*) the specific classes of priors for the parameters of these nested models, $(ii.)$ priors on the parameters of the encompassing models that imply an already specified prior on the parameters of the nested model, $(iii.)$ Bayes factors resulting from (generalized) Savage-Dickey density ratios and $(iv.)$ Bayesian score statistics. In classical statistical analysis, the unique conditional densities offer an alternative approach to construct the small sample and limiting distributions of (I.) maximum likelihood estimators and (II.) classical score statistics.

Several applications of the conditional densities in practice exist: Kleibergen and Hoek (1996) and Kleibergen and Paap (1998) use $(i.)$ and $(iii.)$ in Bayesian analyzes of resp. arma and cointegration models; Kleibergen and van Dijk (1998) use $(i.)$ in a Bayesian analysis of simultaneous equations models; Kleibergen and Zivot (1998) use $(i.)$, $(ii.)$ and (I.) in a comparison of classical and Bayesian analyzes of instrumental variable regression models and Kleibergen (1998) constructs limiting and small sample distributions of limited information maximum likelihood estimators using (I.). These analyzes further show the importance of the unique conditional densities and contain both elements that couldnot have been derived without them $((i.) - (iv.))$ and elements that could also have been obtained using another approach ((I.)-(II.)). The latter thus show the validity of the framework for conducting statistical inference using conditional densities and the conditional densities itself as they lead to the same results as obtained before. It is therefore interesting to further investigate what new statistical results can be obtained using the conditional densities.

Appendix

A.1. Proof of Lemma 1.

Consider two specifications (y_1, z_1) and (y_2, z_2) that are not identical and satisfy conditions 1 and 2. We can specify the jacobian of the transformation from x to (y_2, z_2) as

$$|J(x,(y_2,z_2))| = |J(x,(y_1,z_1))J((y_1,z_1),(y_2,z_2))|.$$

$J((y_1, z_1), (y_2, z_2))$ reads,

$$J((y_1,z_1),(y_2,z_2)) = \begin{pmatrix} \frac{\partial y_1}{\partial y_2'} & \frac{\partial y_1}{\partial z_2'} \\ \frac{\partial z_1}{\partial y_2'} & \frac{\partial z_1}{\partial z_2'} \end{pmatrix}.$$

Since (y_1, z_1) and (y_2, z_2) both satisfy conditions 1 and 2, it holds that y_1 and y_2 completely and only represent the restriction and nothing else. This implies that $\frac{\partial y_1}{\partial z_2'} = 0$ since y_1 would otherwise also represent other things besides the analyzed restriction. Evaluating the jacobian in $y_1 = y_2 = 0$ then gives

$$\begin{aligned} J((y_1,z_1),(y_2,z_2)) &= \begin{pmatrix} \frac{\partial y_1}{\partial y_2'} & \frac{\partial y_1}{\partial z_2'} \\ \frac{\partial z_1}{\partial y_2'} & \frac{\partial z_1}{\partial z_2'} \end{pmatrix} |_{y_1=y_2=0} \\ &= \begin{pmatrix} \frac{\partial y_1}{\partial y_2'}|_{y_1=y_2=0} & 0 \\ \frac{\partial z_1}{\partial y_2'}|_{y_1=y_2=0} & \frac{\partial z_1}{\partial z_2'}|_{y_1=y_2=0} \end{pmatrix} \end{aligned}$$

and therefore

$$\begin{aligned} |J((y_1,z_1),(y_2,z_2))|_{y_1=y_2=0}| &= |\frac{\partial y_1}{\partial y_2'}|_{y_1=y_2=0}|\frac{\partial z_1}{\partial z_2'}|_{y_1=y_2=0} \\ &\propto |J(z_1,z_2)|, \end{aligned}$$

since y_1 and y_2 are evaluated in a fixed point, 0, and donot depend on z_1 and z_2 as they completely and only reflect the restriction. So,

$$|J(x,(y_2,z_2)|_{y_2=0}| \propto |J(x,(y_1,z_1))|_{y_1=0}||J(z_1,z_2)|,$$

which proofs lemma 1. The presence of the jacobian $J(z_1, z_2)$ also shows that the conditions on z are weaker than the conditions on y.

To proof the second part of lemma 1, we note that

$$|J(x,(y_2,z_2)|_{y_2=0}| \propto |J(x,(y_1,z_1))|_{y_1=0}||J(z_1,z_2)| \qquad \forall z_2,$$

thus implies that

$$\begin{aligned} |J((y_1,z_1),(y_2,z_2))|_{y_2=0}| &= |\frac{\partial y_1}{\partial y_2'}|_{y_2=0}||\frac{\partial z_1}{\partial z_2'}|_{y_2=0}| \qquad \forall z_2 \\ &\propto |J(z_1,z_2)| \qquad \forall z_2, \end{aligned}$$

which again implies that both $|\frac{\partial y_1}{\partial y_2'}|_{y_2=0}|$ does not depend on z_2 and that either $\frac{\partial z_1}{\partial y_2'}$ or $\frac{\partial y_1}{\partial z_2'}$ is equal to zero $\forall z_2$. As $\frac{\partial y_1}{\partial y_2'}|_{y_2=0}$ doesnot depend on z_2, y_1 can only depend on z_2 when it consists of two additive parts, one only depending on y_2 and one only depending on z_2. Condition 2 implies that the restriction $g(x) = 0$ corresponds with $y_1 = 0$ and imposes no restrictions on z_1. However, when y_1 consists of two additive parts, one of which only depends on y_2 and one which only depends on z_2, $y_1 = 0$ thus imposes a restriction on z_2 and therefore condition 2 is violated. y_1 thus only depends on y_2 and $\frac{\partial y_1}{\partial z_2'} = 0 \ \forall z_2$. Given condition 2 and invertibility between x and (y, z), the transitivity property is thus equivalent to condition 1, since it shows that y only represents the restriction and nothing else, and it is thus a convenient tool to check whether condition 1 holds.

A.2. Proof of theorem 2.1.

Consider two specifications, (y_1, z_1) and (y_2, z_2) that are not identical and satisfy conditions 1 and 2. The specification of the conditional density of x given $g(x) = 0$ reads,

$$\begin{aligned} p_r(x) &\propto p(x)|_{g(x)=0} \\ &\propto p_r(z_2(x))|J((y_2, z_2), x)|_{g(x)=0}| \end{aligned}$$

where $p_r(z_2) \propto p(y_2, z_2)|_{y_2=0} \propto p(x(y_2, z_2))|_{y_2=0}|J(x, (y_2, z_2))|_{y_2=0}|$. Condition 2 implies that y_2 completely reflects the restriction and because (y_1, z_1) and (y_2, z_2) satisfy condition 1, such that $p(y_1, z_1)$ and $p(y_2, z_2)$ are well-defined, $y_1 = 0$ is equivalent with $y_2 = 0$ as x can be uniquely solved from (y_1, z_1), (y_2, z_2) for all values of y_1, y_2 given that z_1, z_2 allow for it. Furthermore, it results from lemma 1 that,

$$\begin{aligned} |J(x, (y_2(y_1, z_1), z_2(y_1, z_1)))|_{y_2=0}||J(z_2, z_1)| &\propto |J(x, (y_1, z_1))|_{y_1=0}| \Leftrightarrow \\ |J(x, (y_1, z_1))|_{y_1=0}|^{-1}|J(z_2, z_1)| &\propto |J(x, (y_2(y_1, z_1), z_2(y_1, z_1)))|_{y_2=0}|^{-1} \Leftrightarrow \\ |J((y_1, z_1), x)|_{g(x)=0}||J(z_2, z_1)| &\propto |J((y_2, z_2), x)|_{g(x)=0}|. \end{aligned}$$

Combining this yields,

$$\begin{aligned} p_r(x) &\propto p(x)|_{g(x)=0} \\ &\propto p_r(z_2(x))|J((y_2, z_2), x)|_{g(x)=0}| \\ &\propto p_r(z_2(x))|J(z_2, z_1)||J((y_1, z_1), x)|_{g(x)=0}| \\ &\propto p_r(z_1(z_2(x)))|J((y_1, z_1), x)|_{g(x)=0}| \\ &\propto p_r(z_1(x))|J((y_1, z_1), x)|_{g(x)=0}|, \end{aligned}$$

where $p_r(z_1) \propto p(y_1, z_1)|_{y_1=0} \propto p(x(y_1, z_1))|_{y_1=0}|J(x, (y_1, z_1))|_{y_1=0}|$. This proofs theorem 2.1. The presence of the jacobian of the transformation from z_2 to z_1 also shows that the conditions on z are weaker than the conditions on y and correspond with the standard conditions for transforming random variables.

B. Proof of Relationship (3.9) between (3.6) and (3.7)

The specification of X reads,

$$\begin{aligned} X &= \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} V_{11}' & V_{21}' \\ V_{12}' & V_{22}' \end{pmatrix} \\ &= Z_i W_i + Z_\perp Y W_\perp, \ i = 1, 2, \end{aligned}$$

such that

$$Z_i W_i = \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 \begin{pmatrix} V_{11}' & V_{21}' \end{pmatrix}, \ i = 1, 2,$$

and

$$Z_\perp Y W_\perp = \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} S_2 \begin{pmatrix} V_{12}' & V_{22}' \end{pmatrix}.$$

Consequently,

$$\begin{aligned} Z_1 &= \begin{pmatrix} U_{11} \\ U_{21} \end{pmatrix} S_1 V_{21}',\ A = V_{21}^{-1\prime} V_{11}', \\ W_2 &= U_{11} S_1 \begin{pmatrix} V_{11} \\ V_{21} \end{pmatrix}',\ B = U_{21} U_{11}^{-1}, \end{aligned}$$

Substituting these expressions in the specification of $W_\perp$ and $Z_\perp$ gives,

$$\begin{aligned} Z_\perp &= \begin{pmatrix} -Z_{11}^{-1\prime} Z_{12}' \\ I_{k-r} \end{pmatrix} (I_{k-r} + Z_{12} Z_{11}^{-1} Z_{11}^{-1\prime} Z_{12}')^{-\frac{1}{2}} \\ &= \begin{pmatrix} -U_{11}'^{-1} S_1^{-1} V_{21}^{-1} V_{21} S_1 U_{21}' \\ I_{k-r} \end{pmatrix} (I_{k-r} + U_{21} U_{11}^{-1} U_{11}'^{-1} U_{21}')^{-\frac{1}{2}} \\ &= \begin{pmatrix} -U_{11}'^{-1} U_{21}' \\ I_{k-r} \end{pmatrix} (I_{k-r} + U_{21} U_{11}^{-1} U_{11}'^{-1} U_{21}')^{-\frac{1}{2}}, \end{aligned}$$

as $U_{11}' U_{12} + U_{21}' U_{22} = 0$ (because of the orthogonality of U), $U_{12} U_{22}^{-1} = -U_{11}'^{-1} U_{21}'$, and $U_{12}' U_{12} + U_{22}' U_{22} = I_{k-r}$, such that

$$\begin{aligned} Z_\perp &= \begin{pmatrix} U_{12} U_{22}^{-1} \\ I_{k-r} \end{pmatrix} (I_{k-r} + U_{22}^{-1\prime} U_{12}' U_{12} U_{22}^{-1})^{-\frac{1}{2}} \\ &= \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1} (U_{22}'^{-1} (U_{22}' U_{22} + U_{12}' U_{12}) U_{22}^{-1})^{-\frac{1}{2}} \\ &= \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1} ((U_{22}'^{-1} U_{22}^{-1})^{-1})^{\frac{1}{2}} \\ &= \begin{pmatrix} U_{12} \\ U_{22} \end{pmatrix} U_{22}^{-1} (U_{22} U_{22}')^{\frac{1}{2}}. \end{aligned}$$

Similarly for $W_\perp$,

$$\begin{aligned} W_\perp &= \begin{pmatrix} A & I_r \end{pmatrix}_\perp = (I_{m-r} + A'A)^{-\frac{1}{2}} \begin{pmatrix} I_{m-r} & -A' \end{pmatrix} \\ &= (I_{m-r} + V_{11} V_{21}^{-1} V_{21}'^{-1} V_{11}')^{-\frac{1}{2}} \begin{pmatrix} I_{m-r} & -V_{11} V_{21}^{-1} \end{pmatrix} \\ &= (I_{m-r} + V_{12}^{-1\prime} V_{22}' V_{22} V_{12}^{-1})^{-\frac{1}{2}} \begin{pmatrix} I_{m-r} & V_{12}^{-1\prime} V_{22}' \end{pmatrix} \\ &= (V_{12}^{-1\prime} (V_{12}' V_{12} + V_{22}' V_{22}) V_{12}^{-1})^{-\frac{1}{2}} V_{12}'^{-1} \begin{pmatrix} V_{12}' & V_{22}' \end{pmatrix} \\ &= (V_{12}^{-1\prime} V_{12}^{-1})^{-\frac{1}{2}} V_{12}'^{-1} \begin{pmatrix} V_{12}' & V_{22}' \end{pmatrix} \\ &= (V_{12} V_{12}')^{\frac{1}{2}} V_{12}'^{-1} \begin{pmatrix} V_{12}' & V_{22}' \end{pmatrix}, \end{aligned}$$

since $V_{11}' V_{12} + V_{21}' V_{22} = 0$, such that $-V_{21}'^{-1} V_{11}' = V_{22} V_{12}^{-1}$, and $V_{12}' V_{12} + V_{22}' V_{22} = I_{m-r}$. Consequently in order to have equivalence,

$$\begin{aligned} Y &= (U_{22} U_{22}')^{-\frac{1}{2}} U_{22} S_2 V_{12}' (V_{12} V_{12}')^{-\frac{1}{2}} \\ &= C S_2 D, \end{aligned}$$

where $C=(U_{22}U_{22}')^{-\frac{1}{2}}U_{22}$, and $D=V_{12}'(V_{12}V_{12}')^{-\frac{1}{2}}$. Both C and D are orthogonal matrices, since $C'C=I_{k-r}$ and $DD'=I_{m-r}$, such that Y equals the smallest singular value pre and postmultipied by orthogonal vectors/matrices or stated differently, Y is a rotation of the singular values around the origin.

C. Jacobian of transformation to Reduced Rank Specification

For the derivation of the jacobian of the transformation from the unrestricted matrix X to the specification (3.6)-(3.9), it is notationally convenient to conduct this transformation in two steps, $(i.)$ from X to (Z_{11},Q_2,A,Y) where $Q_2=Z_{12}Z_{11}^{-1}$, and $(ii.)$ from (Z_{11},Q_2,A,Y) to (Z_{11},Z_{12},A,Y). Note that we only construct the jacobian associated with (3.4) and the jacobian of other specifications can be constructed in an identical way as this jacobian.

We can denote X as,

$$\begin{aligned}
X &= \begin{pmatrix} Q & Q_\perp \end{pmatrix}\begin{pmatrix} Z_{11} & 0 \\ 0 & Y \end{pmatrix}\begin{pmatrix} W_1 \\ W_\perp \end{pmatrix} \\
&= QZ_{11}W_1+Q_\perp YW_\perp,
\end{aligned}$$

where $Q=(I_r\ Q_2')'$, $Q_\perp=(-Q_2\ I_{k-r})'(I_{k-r}+Q_2Q_2')^{-\frac{1}{2}}$, $W_1=(A\ I_r)$, $W_\perp=(I_{m-r}+A'A)^{-\frac{1}{2}}(I_{m-r}\ -A')$. The jacobians of X with respect to Z_{11}, Q_2, A and Y then read,

$$\begin{aligned}
J_1 &= \frac{\partial vec(X)}{\partial vec(Z_{11})'}=(W_1'\otimes Q) \\
J_2 &= \frac{\partial vec(X)}{\partial vec(Q_2)'}=(W_1'Z_{11}'\otimes I_k)\frac{\partial vec(Q)}{\partial vec(Q_2)'}+(W_\perp'Y'\otimes I_k)\frac{\partial vec(Q_\perp)}{\partial vec(Q_2)'} \\
J_3 &= \frac{\partial vec(X)}{\partial vec(A)'}=(I_m\otimes QZ_{11})\,\frac{\partial vec(W)}{\partial vec(A)'}+(I_m\otimes W_\perp Y)\,\frac{\partial vec(W_\perp)}{\partial vec(A)'} \\
J_4 &= \frac{\partial vec(X)}{\partial vec(Y)'}=(W_\perp'\otimes Q_\perp)
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial vec(Q)}{\partial vec(Q_2)'} &= \left(I_r\otimes\begin{pmatrix} 0 \\ I_{k-r}\end{pmatrix}\right), \\
\frac{\partial vec(Q_\perp)}{\partial vec(Q_2)'} &= -(H^{-\frac{1}{2}\prime}\otimes\begin{pmatrix} I_r \\ 0\end{pmatrix})K_{k-r,r}+ \\
&\quad (I_{k-r}\otimes\begin{pmatrix} -Q_2' \\ I_{k-r}\end{pmatrix})\frac{\partial vec((H^{\frac{1}{2}})^{-1})}{\partial vec(H^{\frac{1}{2}})'}\frac{\partial vec(H^{\frac{1}{2}})}{\partial vec(H)'}\frac{\partial vec(H)}{\partial vec(Q_2)'}, \\
\frac{\partial vec((H^{\frac{1}{2}})^{-1})}{\partial vec(H^{\frac{1}{2}})'} &= -(H^{-\frac{1}{2}\prime}\otimes H^{-\frac{1}{2}}), \\
\frac{\partial vec(H^{\frac{1}{2}})}{\partial vec(H)'} &= ((I_{k-r}\otimes H^{\frac{1}{2}})+(H^{\frac{1}{2}\prime}\otimes I_{k-r}))^{-1}, \\
\frac{\partial vec(H)}{\partial vec(Q_2)'} &= (Q_2\otimes I_{k-r})+(I_{k-r}\otimes Q_2)K_{k-r,r},
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial vec(W_1)}{\partial vec(A)'} &= (\begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix} \otimes I_r), \\
\frac{\partial vec(W_\perp)}{\partial vec(A)'} &= -(\begin{pmatrix} 0 \\ I_r \end{pmatrix} \otimes D^{-\frac{1}{2}})K_{r,m-r} + \\
& \quad ((\begin{array}{cc} I_{m-r} & -A' \end{array})' \otimes I_{m-r})\frac{\partial vec(D^{-\frac{1}{2}})}{\partial vec(D^{\frac{1}{2}})'}\frac{\partial vec(D^{\frac{1}{2}})}{\partial vec(D)'}\frac{\partial vec(D)}{\partial vec(A)'}, \\
\frac{\partial vec(D^{-\frac{1}{2}})}{\partial vec(D^{\frac{1}{2}})'} &= -(D^{-\frac{1}{2}\prime} \otimes D^{-\frac{1}{2}}) = -D^{-1}, \\
\frac{\partial vec(D^{\frac{1}{2}})}{\partial vec(D)'} &= ((I_{m-r} \otimes D^{\frac{1}{2}}) + (D^{\frac{1}{2}\prime} \otimes I_{m-r}))^{-1}, \\
\frac{\partial vec(D^{\frac{1}{2}})}{\partial vec(A)'} &= (A' \otimes I_{m-r})K_{r,m-r} + (I_{m-r} \otimes A'),
\end{aligned}
$$

and $H = I_{k-r} + Q_2 Q_2'$, $H^{\frac{1}{2}} H^{\frac{1}{2}} = H$, $D = (I_{m-r} + A'A)$, $D^{\frac{1}{2}} D^{\frac{1}{2}} = D$, $K_{i,j} : ij \times ij$, are socalled commutation matrices such that for any $\Upsilon : i \times j$, $vec(\Upsilon') = K_{i,j} vec(\Upsilon)$, $vec(\Upsilon) = K_{j,i} vec(\Upsilon')$, $K_{j,i} = K_{i,j}'$, see Magnus and Neudecker (1988). Note that when Σ is positive (semi-)definite symmetric, $\Sigma = \Psi \Lambda \Psi'$, where Ψ are orthogonal eigenvectors and Λ is a diagonal matrix containing the eigenvalues, then $\Sigma^{\frac{1}{2}} = \Psi \Lambda^{\frac{1}{2}} \Psi'$ is also symmetric.

The jacobian of the transformation from X to (Z_{11}, Q_2, A, Y) then reads,

$$
\frac{\partial vec(X)}{\partial (vec(Z_{11})' \; vec(Q_2)' \; vec(A)' \; vec(Y)')} = \begin{pmatrix} J_1 & J_2 & J_3 & J_4 \end{pmatrix}.
$$

Since $Q_2 = Z_{12} Z_{11}^{-1}$, the jacobians of the transformations from (Z_{11}, Q_2, A, Y) to Z_{11}, Z_{12}, A, and Y read,

$$
\begin{aligned}
G_1 &= \frac{\partial (vec(Z_{11})' \; vec(Q_2)' \; vec(A)' \; vec(Y)')'}{\partial vec(Z_{11})'} = \begin{pmatrix} I_r \otimes I_r \\ -Z_{11}^{-1\prime} \otimes Z_{12} Z_{11}^{-1} \\ 0 \\ 0 \end{pmatrix} \\
G_2 &= \frac{\partial (vec(Z_{11})' \; vec(Q_2)' \; vec(A)' \; vec(Y)')'}{\partial vec(Z_{12})'} = \begin{pmatrix} 0 \\ Z_{11}^{-1\prime} \otimes I_{k-r} \\ 0 \\ 0 \end{pmatrix} \\
G_3 &= \frac{\partial (vec(Z_{11})' \; vec(Q_2)' \; vec(A)' \; vec(Y)')'}{\partial vec(A)'} = \begin{pmatrix} 0 \\ 0 \\ I_{m-r} \otimes I_r \\ 0 \end{pmatrix} \\
G_3 &= \frac{\partial (vec(Z_{11})' \; vec(Q_2)' \; vec(A)' \; vec(Y)')'}{\partial vec(Y)'} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ I_{m-r} \otimes I_{k-r} \end{pmatrix}
\end{aligned}
$$

The jacobian of the transformation from X to (Z_1, A, Y) then becomes,

$$\begin{aligned}
& |J(X,(Z_1,A,Y))| \\
= & \left|\frac{\partial vec(X)}{\partial (vec(Z_1)'\ vec(A)'\ vec(Y)')}\right| \\
= & \left|\frac{\partial vec(X)}{\partial (vec(Z_{11})'\ vec(Q_2)'\ vec(A)'\ vec(Y)')}\right| \\
& \left|\frac{\partial (vec(Z_{11})'\ vec(Q_2)'\ vec(A)'\ vec(Y)')'}{\partial (vec(Z_1)'\ vec(A)'\ vec(Y)')}\right| \\
= & \left|\begin{pmatrix} J_1 & J_2 & J_3 & J_4 \end{pmatrix}\right| \left|\begin{pmatrix} G_1 & G_2 & G_3 & G_4 \end{pmatrix}\right|.
\end{aligned}$$

So,

$$J(X,(Z_1,A,Y))|_{Y=0} = \begin{pmatrix} W_1' \otimes I_k & \begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix} \otimes Z_1 & W_\perp' \otimes Z_\perp \end{pmatrix}.$$

To prove that $|(J(X,(Z_{11},Q_2,Y,A))| \geq |(J(X,(Z_{11},Q_2,Y,A))|_{Y=0})|$, we use that

$$\begin{aligned}
& J(X,(Z_{11},Y,A)) \\
= & J(X,(Z_{11},Q_2,Y,A))J((Z_{11},Q_2,Y,A),(Z_1,Y,A)).
\end{aligned}$$

As shown previously

$$J((Z_{11},Q_2,Y,A),(Z_1,Y,A))|_{Y=0} = J((Z_{11},Q_2,Y,A),(Z_1,Y,A)).$$

It also holds that

$$J(X,(Z_{11},Q_2,Y,A)) = J(X,(Z_{11},Q_2,Y,A))|_{Y=0} + C,$$

where

$$C = \begin{pmatrix} 0 & (W_\perp' Y' \otimes I_k)\frac{\partial vec(Q_\perp)}{\partial vec(Q_2)'} & (I_m \otimes Q_\perp Y)\frac{\partial vec(Q_\perp)}{\partial vec(A)'} & 0 \end{pmatrix}$$

such that

$$(J(X,(Z_{11},Q_2,Y,A))|_{Y=0})C' = 0.$$

This implies that

$$\begin{aligned}
& |J(X,(Z_{11},Q_2,Y,A))| \\
= & |J(X,(Z_{11},Q_2,Y,A))J(X,(Z_{11},Q_2,Y,A))'|^{\frac{1}{2}} \\
= & |(J(X,(Z_{11},Q_2,Y,A))|_{Y=0})(J(X,(Z_{11},Q_2,Y,A))|_{Y=0})' + CC'|^{\frac{1}{2}} \\
\geq & |(J(X,(Z_{11},Q_2,Y,A))|_{Y=0})(J(X,(Z_{11},Q_2,Y,A))|_{Y=0})'|^{\frac{1}{2}} \\
\geq & |(J(X,(Z_{11},Q_2,Y,A))|_{Y=0})|
\end{aligned}$$

and consequently

$$|J(X,(Z_1,Y,A))| \geq |(J(X,(Z_1,Y,A))|_{Y=0})|.$$

References

[1] 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).

[2] Berger, J.O. *Statistical Decision Theory and Bayesian Inference.* Springer-Verlag (New York), (1985).

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

[4] Box, G.E.P. and G.C. Tiao. *Bayesian Inference in Statistical Analysis.* Addison-Wesley (Reading, MA), (1973).

[5] Dickey, J.M. The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *The Annals of Mathematical Statistics*, **42**:204–223, 1971.

[6] 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).

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

[8] Engle, R.F. and C.W.J. Granger. Co-integration and error correction : Representation, estimation and testing. *Econometrica*, **55**:251–276, 1987.

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

[10] Fuller, W.A. *Introduction to Statistical Time Series.* Wiley (New York), (1976).

[11] Geweke, J. Bayesian Reduced Rank Regression in Econometrics. *Journal of Econometrics*, **75**:121–146, (1996).

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

[13] 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).

[14] Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, **59**:1551–1580, (1991).

[15] Kleibergen, F. An Alternative Approach for Constructing Small Sample and Limiting Distributions of Maximum Likelihood Estimators. Econometric Institute Reports 9844/A, 1998.

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

[17] Kleibergen, F. and H. Hoek. Bayesian Analysis of ARMA Models using Noninformative Priors. Econometric Institute Report 9553/A, Erasmus University Rotterdam, 1996.

[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. Bayesian Score Statistics in Linear Models. Technical report, Econometric Institute, Erasmus University Rotterdam, 1998.

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

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

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

[23] 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.

[24] 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.

[25] Phillips, P.C.B. Econometric Model Determination. *Econometrica*, **64**:763–812, (1996).

[26] Phillips, P.C.B. and W. Ploberger. Posterior Odds Testing for a Unit Root with Data-based Model Selection. *Econometric Theory*, **10**:774–808, (1994).

[27] Phillips, P.C.B. and W. Ploberger. An Asymptotic Theory of Bayesian Inference for Time Series. *Econometrica*, **64**:381–412, (1996).

[28] Poirier, D.J. Bayesian Diagnostic Testing in the General Linear Regression Model. In J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, editor, *Bayesian Statistics 3*, pages 725–732. Oxford University Press, 1988.

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

[30] Tiao, G.C., W.-Y. Tan and Y.-C. Chang. Some Aspects of Bivariate Regression Subject to Linear Constraints. *Journal of Econometrics*, 5:13–35, 1977.

[31] Verdinelli, I. and L. Wasserman. Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association*, **90**:614–618, 1995.

[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, (1971).