DO YOU KNOW HOW MANY STRUCTURAL SHOCKS YOU HAVE IN YOUR MODEL? A BAYESIAN FRAMEWORK FOR TESTING ECONOMIC MODELS

Andrzej Kocięcki*
National Bank of Poland
e-mail: andrzej.kociecki@nbp.pl

Abstract: The paper provides with a Bayesian framework for testing macroeconomic models. As many of them imply that the number of economic shocks is less than the number of economic variables under consideration this makes the covariance matrix singular. The paper confronts a common view that such singular models can not be satisfactorily treated with a formal statistical analysis, and in particular that there is no way to convince oneself how many shocks one has in the model. Although there are fundamental problems which make the naïve Bayesian testing inconsistent we designed a more insightful Bayesian framework that can cope with this. Such a framework opens up the way to compare any economic models which may or may not be singular and leads to closer interaction between econometrics and economic theory.

first version: February 2007
this version: March 2008

I. INTRODUCTION

Since the seminal paper by Kydland and Prescott (1982) we have witnessed a large body of macroeconomic literature on Real Business Cycle (RBC) models. Early contributions almost entirely adopted calibration technique instead of the statistical estimation. Needless to say, the calibration, as a method to obtain unknown coefficients, was not uncontroversial from the very beginning. The issue was discussed

* The author thanks Professors Herman van Dijk and Jean–Marie Dufour for some comments and suggestions.

extensively by the profession – see e.g. a special issue of the Journal of Economic Perspectives in 1996. Since that time, RBC models have been gradually replaced by more complicated but also more realistic Dynamic Stochastic General Equilibrium (DSGE) models. In particular, Smets and Wouters (2003) showed that estimated economic model can fit the data quite well i.e. comparably to the statistical model e.g. VAR model. Now, more and more researchers try to estimate their models instead of calibrating them. Unfortunately, very often such models contain less economically interpretable structural shocks than the considered number of economic variables. This introduces nontrivial complication: models become stochastically singular i.e. covariance matrix of the observables is singular. In order to estimate these models one faces the fundamental problem in formulating the likelihood function. In general, the literature adopted two ways to deal successfully with this impediment. One strategy is to consider only a subset of variables so as they are equal in a number to the structural shocks, see e.g. Ruge–Murcia (2003) and references therein. The other way, which has become more popular recently, is to introduce more shocks to the system and give them certain interpretation. These are mainly shocks attributed to the measurement error, see e.g. Ireland (2004) and references therein, and shocks in parameters or unobservable states, see e.g. Smets and Wouters (2003). However, adding additional shocks (whatever they may be rationalized) makes the system estimation feasible but also interpretational problems. As a result, we can no longer be sure whether e.g. technology shock estimated together with added artificial shocks preserves its macroeconomic interpretation. In particular, introducing shocks in different ways e.g. as measurement error or shocks in parameters (or unobservable states), may have far reaching consequences for economic interpretation of model dynamic characteristics. For example, variance decompositions from these models may suggest drastically different contributions of the technology shock, see e.g. Alvarez–Lois et al. (2005). Moreover, as emphasized by Bierens (2007), the singularity of DSGE models is a very general feature that can not be mitigated by adding more shocks to state variables or adding economically interpretable shock which entails appending new observable to the system, see e.g. Ruge–Murcia (2003). Similar point is raised in King and Watson (1996). On the other hand, as stressed e.g. by Landon–Lane (2002), adding measurement errors to all variables implies that there are more shocks than variables which needs incorporating additional restrictions on covariance to achieve the identification of shocks. Alternatively, one should decide which variables are measured with error and which are not. For this strategy see e.g. McGrattan et al. (1997). Therefore it seems there is a demand from macroeconomists to have a formal tool to deal with stochastically singular models in its pure theoretic form i.e. preserving the singularity, which we hope to provide with. This opens up a new perspective to compare singular economic models as they stand (i.e. without incorporating of the additional shocks that make the covariance matrix nonsingular).

As delivering such a framework introduces many technical complications we had to tackle other problems that run parallel with this. For these reasons, we shall put in a systematic way the results concerning singular normal distributions which shed new light on a character, proper statistical treatment and interpretation of singular models in economics. Secondly, we point to remarkable disability of the naïve Bayesian procedure in testing the singularity of covariance matrix and propose the method that overcomes this problem.

Even if the data do not confirm singularity, the macroeconomist would insist that this proves nothing. The reason is that the rejection of the singularity is in fact the rejection of the number of restrictions one of which is singularity. Indeed, since the model entails several constraints which constitute its structure, the rejection of the particular model only means (using basic facts from the Boolean algebra) that at least one of its characteristics does not fit the facts. For example, the assumption of no measurement error, functional model specification or too radical lag truncation in the case of VAR model may not be valid. As rightly noticed by Braithwaite (1953), p. 20, "*in almost every system it is possible to maintain any one hypothesis in the face of apparently contrary evidence at the expense of modifying the others*"[1]. He also emphasized that which restriction has to be modified is a matter of "hunch" of the scientist. In our case, since adding additional shocks may destroy coherent economic reasoning behind the model he or she may be very reluctant to dismiss the assumption that there may be a small number of economically interpreted shocks that are responsible for the total macroeconomic dynamics. He (she) would rather question the other model assumptions. We offer such a macroeconomist the framework to compare existing theories which may induce stochastic singularities. We are in a total agreement with the statement in Eichenbaum (1996) that "*the role of econometrics ought to be the advancement of empirically plausible economic theory*".

[1] This is in fact Duhem's critique of Popper's falsification criterion, see Duhem (1962), pp. 185–187

Indeed, we provide with an econometric tool that helps macroeconomist to specify his or her models.

II. Stochastically Singular Models: Facts and Myths

This section documents some common views about stochastically singular models which will be confronted in the subsequent sections. Our intention is to make readers realize how much misunderstanding prevails among researchers, which calls for a systematic treatment of the subject. We shall illustrate this section with a plenty of citations which articulate the need for clarification. We will also make a few comments regarding the origins of econometric practice that are useful for building the rationale for our model framework. We begin with a short compilation of quotations:

1) Altug (1989) "*since the singularity of these processes is sure to be rejected by actual data*"

2) King (1995) "*the probability that the model is true is zero for stochastically singular models*"

3) Schorfheide (2000) "*in practice these singularities are rejected with a short span of time series data*"

4) An and Schorfheide (2007) "*Any DSGE model that generates a rank-deficient covariance matrix (…) is clearly at odds with the data*"

5) Canova (2007) "*the covariance matrix is singular (…), a restriction unlikely to hold in the data*"

6) Chari, et al. (2005) "*note that in the data, estimated covariance matrices … are never singular*"

7) Sims (2006) "*because the data are clearly not singular, any model that fits multiple time series must necessarily have at least as many sources of stochastic disturbance as there are variables being explained*"

8) Boivin and Giannoni (2006) "*Given that the RBC model considered here has only one source of exogenous fluctuations, using more than one observable series would result in the model rejection (…) the model is said to be stochastically singular in this case. As this is not true in the data, the model is sure to be rejected*"

All these citations have at least two elements in common. First of all, those authors are confident that singularity can not be present. In addition an implicit, common element behind these statements is that the authors did not explain on what methodological ground they reached such a conclusion (except that they do not see it in the data). In fact they all did not provide with any formal results. Concerning the first point, we note that essentially their argument is as follows: If the theory predicts linear restriction among the variables and you observe no such a restriction in the data then the theory is rejected. In the opinion of the author this scientific reasoning is unacceptable as it backs us to the world where the term "statistical inference" had no meaning. Essentially the suggestion is that if you estimate an univariate random walk process and the autoregressive coefficient turns out to be not equal to 1 (which is more than sure) then the unit root hypothesis is rejected (and by the way all enormous literature on the unit root testing). In fact we shall clarify later what the meaning of linear constraints inherent in the data under singularity is and argue forcefully that whether it holds it must be decided using statistical test (which is proposed in this paper).

Furthermore, Bierens (2007) claims that the Bayesian analysis is impossible in the case of singular data distribution. Rust (1988) takes a similar view and interestingly mentions this argument as a general objection to Bayesian approach in economics[2]. We shall confront this claim later and will show that intelligent Bayesian can tackle all these problems.

Contrary to opinions expressed in Bierens and Swanson (2000) it is not difficult to write down the density of singular multivariate normal distribution. In particular it is not identically equal to 0 as argued by Schorfheide (2000), which will be shown. Interestingly some researchers were even skeptical about whether such a function exists e.g. Landon–Lane (2002).

We also mention that the Maximum Likelihood estimation of stochastically singular models is commonly not well treated, but since this argument requires some notation we postpone its discussion to remark 5 in section VII.

As emphasized by Watson (1993), since the economic model need not be enriched by ad hoc random components it is more reliable as an approximation of the economic phenomenon than any statistical model. He, however, concludes that as an economic model "*does not provide a complete probability structure, inference procedures lack statistical foundations and are necessarily ad hoc*". We claim that this conclusion is not valid. Actually our basic aim is to introduce a formal

[2] Actually, he lists 5 arguments against Bayesian approach in economics. All arguments are very weak as pointed out in a rejoinder by prof. Poirier and at least 2 of them are no longer "up–to–date" (i.e. computation and possible inconsistency of nonparametric Bayesian analysis).

framework within which we may choose between economic models that are stochastically singular. Therefore we reject the common view that it is impossible to conduct the formal statistical testing when one faces singular models (e.g. Schorfheide (2000) writes that posterior odds in this situation are not defined).

It is instructive to find out how the creators of the foundations of the econometric methodology addressed the issue of what we call now, stochastic singularity[3]. In particular, Koopmans at al. (1950) p. 58, consider the assumption of nonsingularity of covariance matrix merely as a useful simplification. Interestingly they rationalized this assumption by arguing that essentially the only source of singularity is the functional dependence of the structural shocks which they considered as irrational. See also Koopmans and Hood (1953), p. 121. We agree in this point. Our claim is similar to that of Koopmans at al. (1950) i.e. we assume structural shocks hit independently, but we allow for a possibility that the number of shocks differs from the number of variables. After all it is not a priori unbelievable that in 10–variables model there may be less than 10 structural shocks (e.g. 9)[4]. Indeed, in the recent paper by Fernández–Villaverde et al. (2005), they refer to the case when the number of structural shocks is equal to the number of variables, as "*the lucky situation*". Interestingly, Haavelmo (1944) p. 71, gives his most general construction of (what is now called) simultaneous equation system in which the number of shocks (his m) is distinct from the number of variables in the system (his n). Anyway, in our case the singularity arises in a natural way which was not taken explicitly into account by Koopmans at al. (1950). Moreover those authors argued that singularity is a direct consequence of identities among the variables which however are known prior to estimation so as they may be used to get rid of the singularity in the system of equations. Our problem is more delicate. We do not have a knowledge of any identities that are present in a model although we suspect that they are in (they are suggested by an economic model). This does not automatically mean that we do observe some deterministic relationships among observables but only they are more probable than they are not – "*It would indeed be strange if it were*

[3] The main reason is that they shaped the way we perceive the econometric business and to a large extent we are (maybe unconsciously) stuck in the schematic frame imposed by members of the Cowles commission and its collaborators.

[4] As will become clear after the reading this paper, the fact that e.g. in 10–variables VAR model with 5 lags and constant we have 9 structural shocks induces, in total, 61 restrictions on the model! Thus taking into account the tradeoff between fit and parsimony which is present explicitly in all information criteria and implicitly in Bayes Factor, one may not reject a prior hypothesis that the number of shocks is smaller than the number of variables.

otherwise, since economists would then find themselves in a more favorable position than any other research workers, including the astronomers", Haavelmo (1944) p. 40 (see also Jeffreys (1961) p. 13). The only way to convince oneself whether there is "something like identities" that follow from our construction of the economic model is to conduct the statistical test. We use the term "something like identities" because we aware that the point null hypothesis in the context of economic models may only be perceived as a useful simplification.

A critique (relevant to our discussion) concerning the Haavelmo's approach was made by Tintner (1946). He noted that the assumption that there are as many economic relationships as variables is very strong. He suggested that prior to analysis it is important to "estimate" the number of economic relationships present in the data i.e. "true dimensionality". He warned the Haavelmo's approach followers that they "*may endeavor to accomplish too much, i.e., to determine a greater number of equations than actually exist in the data. This would by necessity lead to nonsensical results*", p. 14. This remark is clearly influenced by the Frisch's confluence analysis in which choosing the number of independent economic equations occupied the central position. What does it have in common with our main objective? In section IV, we will show that the valid representation for our model is the one in which the number of economic equations is less than the number of variables, see equation (4). Thus we seriously take into account the postulate of Tintner. In fact, our whole analysis aims at testing the true dimensionality of the model i.e. the number of independent relationships. Interestingly in a classic paper, Tinbergen (1940), poses the question "*In how many well–known essays on the cycle has it been stated carefully what* [the number of equations] *is?*". He considered it ridiculous to collect the relationships and stop "*after having found that there are as many equations as unknowns*". We provide with a formal tool to avoid this conundrum.

III. Singular Multivariate Normal Distribution

It is important for our further analysis to have a formal definition of a singular normal distribution. Since this topic is to a large extent irrelevant in statistical modeling[5], the contributions to the theory of this distribution are scattered in the

[5] Loosely speaking if the distribution of the vector x is singular there exists a functional dependence among the elements of x. Therefore, at least in principle, we can derive the law of probability for a subset of x and the remaining elements will be deterministic function of this subset. On the other hand in most statistical applications the singularity of the distribution is an event with zero measure which is a consequence of the underlying

literature. Among early developments are Lukomski (1939) and Cramér (1946) – see also Khatri (1968) and Rao (1973). On the other hand the well–known reference book on multivariate distributions i.e. Kotz et al. (2000), confines to the statement that the normal distribution is singular if the covariance matrix is only positive semidefinite and contains no single section devoted to this problem. For these reasons we shall give the precise meaning of the multivariate singular normal distribution. We mention that, to the best of author's knowledge, our results as a whole are not available elsewhere. We begin with the following definition:

Definition 1: *A $m-$dimensional random vector x is said to have a singular multivariate normal distribution of rank k if it can be represented as $x = \mu + C_1 u_1$, where μ is $m \times 1$, C_1 is $m \times k$ with* $\operatorname{rank}(C_1) = k < m$ *and $u_1 \sim N(\mathbf{0}, \mathbf{I}_k)$.*

To derive the density with respect to the Lebesgue measure we have to choose k elements from x, say x_1, to which we attribute all randomness and the remaining elements will be functionally dependent on x_1. Alternatively we can choose k linearly independent combinations of x which may have some virtue in real applications. For simplicity we shall assume that the first k elements of x are random whereas the remaining $m-k$ elements of x are functionally dependent on x_1. Let us denote those elements by x_2. Taking this into account, we rewrite the representation from definition 1 as:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} C_{11} \\ C_{21} \end{bmatrix} u_1 \tag{1}$$

where $x_1 : (k \times 1)$, $x_2 : (m-k) \times 1$, $\mu_1 : (k \times 1)$, $\mu_2 : (m-k) \times 1$, $C_{11} : (k \times k)$, $C_{21} : (m-k) \times k$. Furthermore we assume C_{11} is nonsingular. Additionally we need the following notation:

$$\Sigma = C_1 C_1' = \begin{bmatrix} C_{11}C_{11}' & C_{11}C_{21}' \\ C_{21}C_{11}' & C_{21}C_{21}' \end{bmatrix} \equiv \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{2}$$

Bearing in mind the representation (1) we have the lemma:

assumption of the absolute continuity of the probability measures. Concerning the last point, see however an illuminating critique in Pollard (2002), especially p. 57.

Lemma 1: *If a random vector* $x:(m\times 1)$ *has a representation as in the definition 1, provided the random elements of x are grouped as the first k i.e.* x_1, *we have*:

a) *marginal distribution of* $x_1 \sim N(\mu_1, \Sigma_{11})$

b) *conditional distribution of* x_2 *given* x_1 *is degenerated i.e.*:

$x_2 = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$ *with probability* 1.

Proof: see appendix 1.

Such a distribution will be labeled as a singular normal pdf of rank *k*, or in short, singular normal pdf. For our purposes we shall introduce the concept of an almost singular distribution. It will prove to be a key result for our approach aiming at testing the number of structural shocks in a model. To this end we have the next definition:

Definition 2: *A* $m-$ *dimensional random vector* x *is said to have an almost singular multivariate normal distribution of rank* k *if it can be represented as:*

$x = \mu + [C_1 \vdots C_2]\begin{bmatrix}\mathbf{I}_k & \mathbf{0}\\ \mathbf{0} & \alpha\mathbf{I}_{m-k}\end{bmatrix}\begin{bmatrix}u_1\\ u_2\end{bmatrix}$, *where* $[C_1 \vdots C_2]: m\times m$ *with* $\operatorname{rank}([C_1 \vdots C_2]) = m$ $[u_1' \vdots u_2']' \sim N(\mathbf{0}, \mathbf{I}_m)$ *and* $\alpha \to 0^+$.

Condition $\alpha \to 0^+$ means that α is a constant which is arbitrarily close to 0 but not equal to 0. For if $\alpha = 0$, we arrive at the definition 1. Using notation consistent with (1) and (2) we then have:

Corollary 1: *If a random vector* $x:(m\times 1)$ *has a representation as in definition 2, we have*:

a) *marginal distribution of* $x_1 \sim N(\mu_1, \Sigma_{11})$

b) *conditional distribution of* x_2 *given* x_1 *is*:

$x_2 \mid x_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$ with $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \to \mathbf{0}$

Proof: see appendix 1.

Remark 1: Lemma 1 suggests that the standard formula for the marginal distribution of the subvector of normal multivariate distribution still holds in the case of singular normal pdf of rank k. On the other hand the corollary 1, without a condition $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \to \mathbf{0}$, is simply the standard decomposition of a multivariate normal distribution into the marginal and conditional distributions. We emphasize that the distribution induced by the representation from the definition 2 converges weakly (as $\alpha \to 0^+$) to singular distribution function. This is a formal statement which was proved in appendix 1 by showing that the characteristic function of the almost singular distribution converges to characteristic function of singular distribution.

Remark 2: One may argue that the "soft" restriction $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \approx \mathbf{0}$ is more realistic than $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \mathbf{0}$. This is because the latter restriction expresses the information that $x_2 = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$ precisely. In fact, due to the intangible concept of the Data Generating Process (DGP), μ_2 and μ_1 will not be actual conditional means. For example, in macroeconometrics we often approximate the true unknown DGP by VAR(p) model, where p is relatively small. In this case it is not reasonable to assume the estimated conditional means are exactly equal to the true conditional means of the process. There are also other, more fundamental reasons to rely on corollary 1 rather than on lemma 1, as far as the covariance singularity testing is concerned – see section VII.

IV. The Model and Basic Notation

Having a formal definition of the singular and almost singular normal pdf, we are in a position to introduce the basic model framework that we shall employ in the paper. As (log)linearized approximation of DSGE models around the steady state may be, in general, written in the VARMA framework, there is a link between statistical and economic model – see e.g. Fernández–Villaverde et al. (2005). For simplicity we assume that the VARMA model may be approximated by VAR model with a finite lag – though see Ravenna (2007) for interesting discussion on this topic. Doing so, we implicitly assume that the MA component is invertible. Of course one may proceed directly with VARMA modeling but such a development would be more demanding (though possible extension to our framework). Then, our starting point is exactly the same VAR model as in Bierens (2007):

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + A_0 \varepsilon_{1t} \tag{3}$$

where y_t is $m \times 1$ vector of the observations at time t, c is $m \times 1$ vector of constants, $A_1, A_2, \ldots, A_p$ are $m \times m$ matrices of coefficients on lagged data vectors. The important feature of the model that makes it distinct from the traditional VAR model is that the vector of structural shocks ε_{1t} is $k \times 1$, where k need not to be equal to m (it follows that A_0 is $m \times k$). In particular we will be interested in the case $k \leq m$, which accommodates singularity of the covariance matrix. In what follows we do not consider the case $k > m$, but we note this is much easier to analyze because then the covariance is nonsingular and the key issue is only the structural shocks identification. We assume that structural shocks are normally distributed with $\mathbf{E}(\varepsilon_{1t} \mid y_{t-s}, s > 0) = \mathbf{0}$, $\mathbf{E}(\varepsilon_{1t}\varepsilon_{1t}' \mid y_{t-s}, s > 0) = \mathbf{I}_k$ and importantly $\operatorname{rank}(A_0) = k \leq m$. Note that as the latter condition (i.e. $\operatorname{rank}(A_0) = k$) holds, we can rewrite the basic model (3) as:

$$A_0^* y_t = c^* + A_1^* y_{t-1} + A_2^* y_{t-2} + \ldots + A_p^* y_{t-p} + \varepsilon_{1t} \tag{4}$$

where $A_0^* = (A_0' A_0)^{-1} A_0'$, $A_i^* = A_0^* A_i$ for $i = 1, \ldots p$ and $c^* = A_0^* c$. The model (4) may be seen as a subset of the first k equations from the $m-$variate Structural VAR (SVAR) model.

Next, we find it useful to introduce some notation: $B = [c\ A_1\ A_2\ \ldots\ A_p]$ is $m \times l$ where $l = mp + 1$, $Y' = [y_1\ y_2\ \ldots\ y_T]$ where T denotes the sample size and:

$$X' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_0 & y_1 & \cdots & y_{T-1} \\ y_{-1} & y_0 & \cdots & y_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{-p+1} & y_{-p+2} & \cdots & y_{T-p} \end{bmatrix}$$

Furthermore let $\mathbf{Y}$ denote the data. We will use extensively multivariate gamma function defined as $\Gamma_k(a) = \pi^{\frac{1}{4}k(k-1)} \prod_{i=1}^{k} \Gamma(a - \frac{(i-1)}{2})$. $N(a,b)$ denotes Normal (multivariate or matricvariate) probability density function (pdf) with mean a and covariance b. The symbol "$\xrightarrow{\mathrm{p}}$" means convergence in probability and "$\Rightarrow$" week convergence of distribution function. In addition by the symbol "$\sim$" we mean the asymptotic equivalence i.e. $A \sim B$ implies $\frac{A}{B} \approx 1$ asymptotically. Note however that this symbol will be also used in the context of distribution law e.g. $x \sim N(a,b)$. Lastly, we adopt the usual $O_{\mathrm{p}}(\cdot)$ and $o_{\mathrm{p}}(\cdot)$ to denote the terms magnitudes in probability. Some relevant notation will also be given when it goes naturally with our development.

V. THE LIKELIHOOD FUNCTION

The fundamental object in our analysis is the data density function which we will call interchangeably the likelihood function. Let us first assume that the number of structural shocks is equal to the number of variables i.e. $k = m$. Then A_0 is $m \times m$ and nonsingular. The model (3) may be written as:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \vartheta_{1t} \tag{5}$$

where $\vartheta_{1t} \sim N(\mathbf{0}, A_0 A_0') \equiv N(\mathbf{0}, \Omega)$ and $\Omega = A_0 A_0'$. From now on we build our notation on the covariance Ω rather than A_0 (which appears in the basic specification (3)). The reason is that in order to work with A_0 we have to introduce some identifying restrictions and since our goal is to test the singularity we are satisfied with a reduced form specification (5). For further development we need the following notation:

$$\Omega = \begin{bmatrix} \Omega_{k11} & \Omega_{k12} \\ \Omega_{k21} & \Omega_{k22} \end{bmatrix} \tag{6}$$

$\Omega_{k11} : k \times k$, $\Omega_{k12} : k \times (m-k)$, $\Omega_{k21} = \Omega_{k12}'$, $\Omega_{k22} : (m-k) \times (m-k)$ and:

$$\Omega_{k22.1} = \Omega_{k22} - \Omega_{k21}\Omega_{k11}^{-1}\Omega_{k12} \tag{7}$$

$$\Omega_{k21.1} = \Omega_{k21}\Omega_{k11}^{-1} \tag{8}$$

subscript "k" is introduced to indicate dimensions of the blocks Ω_{kij} which becomes important in the testing procedure. The data density function is given by:

$$L(\mathbf{Y} \mid B, \Omega) = (2\pi)^{-\frac{1}{2}mT} |\Omega|^{-\frac{1}{2}T} \times \operatorname{etr}\left\{-\tfrac{1}{2}\Omega^{-1}(Y' - BX')(Y' - BX')'\right\} \tag{9}$$

For our purposes it is important to rewrite the above function as:

$$L(\mathbf{Y} \mid B, \Omega) = L(\mathbf{Y}_k \mid B, \Omega) \times L(\mathbf{Y}_{m-k} \mid \mathbf{Y}_k, B, \Omega) \tag{10}$$

where $\mathbf{Y}_k$ denotes all observations on the first k variables, $\mathbf{Y}_{m-k}$ all observations on the remaining $m-k$ variables and:

$$L(\mathbf{Y}_k \mid B, \Omega) \equiv L(\mathbf{Y}_k \mid B_k, \Omega_{k11}) =$$
$$= (2\pi)^{-\frac{1}{2}kT} |\Omega_{k11}|^{-\frac{1}{2}T} \times \operatorname{etr}\left\{-\tfrac{1}{2}\Omega_{k11}^{-1}(Y_k' - B_k X')(Y_k' - B_k X')'\right\} \tag{11}$$

$$L(\mathbf{Y}_{m-k} \mid \mathbf{Y}_k, B, \Omega) \equiv L(\mathbf{Y}_{m-k} \mid \mathbf{Y}_k, B_k, B_{m-k}, \Omega_{k22.1}, \Omega_{k21.1}) = (2\pi)^{-\frac{1}{2}(m-k)T} |\Omega_{k22.1}|^{-\frac{1}{2}T} \times$$
$$\operatorname{etr}\left\{-\tfrac{1}{2}\Omega_{k22.1}^{-1}(Y_{m-k}' - B_{m-k}X' - \Omega_{k21.1}(Y_k' - B_k X'))(Y_{m-k}' - B_{m-k}X' - \Omega_{k21.1}(Y_k' - B_k X'))'\right\} \tag{12}$$

where $B_k : k \times l$ (recall $l = mp + 1$) is a submatrix of B comprising its first k rows and $B_{m-k} : (m-k) \times l$ is a submatrix of B comprising its last $m-k$ rows. Analogously $Y_k' : k \times T$ is a submatrix of Y' that contains its first k rows, $Y_{m-k}' : (m-k) \times T$ comprises the remaining rows and X' were defined in section IV.

According to results from section III, if A_0 is $m \times k$, $k < m$ with $\operatorname{rank}(A_0) = k$ then $\Omega_{k22.1} = \mathbf{0}$ and our model is consistent with a singular representation from definition 1 and lemma 1 applies[6]. It means that

$$L(\mathbf{Y} \mid B, \Omega) = L(\mathbf{Y}_k \mid B, \Omega) \times \mathbf{1}_{Y_{m-k}' = B_{m-k}X' + \Omega_{k21.1}(Y_k' - B_k X')}(\mathbf{Y}_{m-k}) \tag{13}$$

[6] Actually for the results to be valid it is extremely important that the upper $k \times k$ block of A_0 is nonsingular, otherwise $\Omega_{k22.1} = \Omega_{k22} - \Omega_{k21}\Omega_{k11}^{-1}\Omega_{k12}$ does not exist since Ω_{k11} is singular. This was assumed in deriving the lemma 1. Of course there is a possibility to rewrite the lemma 1 so as it permits singularity of Ω_{k11} by working with its generalized inverse, but then the marginal density for the first k variables is also singular and the testing procedure would become more cumbersome. Needless to say such a generalization is analytically straightforward.

where $\mathbf{1}_{Y'_{m-k}=B_{m-k}X'+\Omega_{k21.1}(Y'_k-B_kX')}(\mathbf{Y}_{m-k})$ denotes the indicator function equal to 1 if the observations $\mathbf{Y}_{m-k}$ belong to the set $Y'_{m-k} = B_{m-k}X' + \Omega_{k21.1}(Y'_k - B_kX')$ and 0 otherwise. Note also that $Y'_{m-k} = B_{m-k}X' + \Omega_{k21.1}(Y'_k - B_kX')$ if and only if $\Omega_{k22.1} = \mathbf{0}$.

On the other hand if we assume almost singular representation then:

$$L(\mathbf{Y} \mid B,\Omega) = L(\mathbf{Y}_k \mid B,\Omega)L(\mathbf{Y}_{m-k} \mid \mathbf{Y}_k,B,\Omega) \equiv L(\mathbf{Y}_k \mid B_k,\Omega_{k11})L(\mathbf{Y}_{m-k} \mid \mathbf{Y}_k,B,\Omega_{k22.1},\Omega_{k21.1}) \quad (14)$$

where in the latter data density function we force $\Omega_{k22.1} \to \mathbf{0}$ (i.e. $\Omega_{k22.1} \approx \mathbf{0}$) thus $Y'_{m-k} \approx B_{m-k}X' + \Omega_{k21.1}(Y'_k - B_kX')$.

Remark 3: The existence of (exactly) k structural shocks imposes the restriction $Y'_{m-k} = B_{m-k}X' + \Omega_{k21.1}(Y'_k - B_kX')$. It clarifies formally in what sense, if the covariance is singular, there are linear restrictions among the variables to be hold. For such a statement see e.g. Schorfheide (2000), Landon–Lane (2002), Ruge–Murcia (2003), Ireland (2004), Boivin and Giannoni (2006). Those authors, however, did not formalize it. Note that this restriction involves true parameter values and not its estimators or posterior distributions. Thus in small samples (and contrarily to views expressed by many researchers, see e.g. Geweke (1999) and works cited in our sections I and II), although the true covariance matrix is singular, we do not have to necessarily observe that there are linear restrictions inherent in the data. Whether it is probable it must be tested with a statistical tool (e.g. such as proposed in this paper). See also related Remark 2.

Remark 4: The restriction $Y'_{m-k} = B_{m-k}X' + \Omega_{k21.1}(Y'_k - B_kX')$ is quite severe for constant coefficients framework but not for time–varying coefficients and/or covariance framework. Thus if the underlying DGP is singular we will better capture its dynamics using time–varying coefficients and/or covariance models. It is a conjecture but perhaps the success of time–varying regression models in empirical macroeconomics follows from the fact that the situation in which there are less structural shocks than variables in models is quite frequent.

VI. The Prior

In our analysis we adopt the standard conjugate Normal–Wishart prior. This prior setup will permit us to derive many formulas e.g. Prior Predictive Density, in an analytical form. Actually this is the main reason why this prior setup is so popular in empirical studies. The joint Normal–Wishart prior is given by:

$$p(B,\Omega) = p(B \mid \Omega)p(\Omega) \quad (15)$$

where:

$$p(B \mid \Omega) = (2\pi)^{-\frac{1}{2}ml} |\Omega|^{-\frac{1}{2}l} \left|\overline{Z}\right|^{-\frac{1}{2}m} \operatorname{etr}\left\{-\tfrac{1}{2}\Omega^{-1}(B-\overline{B})\overline{Z}^{-1}(B-\overline{B})'\right\} \tag{16}$$

$$p(\Omega) = 2^{-\frac{1}{2}\nu m}[\Gamma_m(\tfrac{\nu}{2})]^{-1} \left|\overline{Q}\right|^{\frac{1}{2}\nu} |\Omega|^{-\frac{1}{2}(\nu+m+1)} \operatorname{etr}\{-\tfrac{1}{2}\Omega^{-1}\overline{Q}\} \tag{17}$$

where $\overline{Z}: l \times l$, $\overline{Q}: m \times m$, $\overline{B}: m \times l$ are hyperparameters (i.e. parameters to be set by the researcher)[7]. In particular for many steps in the analytical integration process we need the following decomposition of the above prior pdf's:

$$p(B \mid \Omega) = p(B_k \mid \Omega) p(B_{m-k} \mid B_k, \Omega) \tag{18}$$

with:

$$p(B_{m-k} \mid B_k, \Omega) = (2\pi)^{-\frac{1}{2}(m-k)l} \left|\Omega_{k22.1}\right|^{-\frac{1}{2}l} \left|\overline{Z}\right|^{-\frac{1}{2}(m-k)} \times$$

$$\times \operatorname{etr}\left\{-\tfrac{1}{2}\overline{Z}^{-1}(B_{m-k} - \overline{B}_{m-k} - \Omega_{k21.1}(B_k - \overline{B}_k))'\Omega_{k22.1}^{-1}(B_{m-k} - \overline{B}_{m-k} - \Omega_{k21.1}(B_k - \overline{B}_k))\right\} \tag{19}$$

$$p(B_k \mid \Omega) = (2\pi)^{-\frac{1}{2}kl} \left|\Omega_{k11}\right|^{-\frac{1}{2}l} \left|\overline{Z}\right|^{-\frac{1}{2}k} \operatorname{etr}\left\{-\tfrac{1}{2}\overline{Z}^{-1}(B_k - \overline{B}_k)'\Omega_{k11}^{-1}(B_k - \overline{B}_k)\right\} \tag{20}$$

where $\overline{B}_k : k \times l$ are prior means for the first k rows of B, $\overline{B}_{m-k} : (m-k) \times l$ are prior means for the remaining $m-k$ rows. Furthermore we decompose the inverted Wishart pdf (17) as:

$$p(\Omega) = p(\Omega_{k21.1} \mid \Omega_{k22.1}) p(\Omega_{k22.1}) p(\Omega_{k11}) \tag{21}$$

with:

$$p(\Omega_{k21.1} \mid \Omega_{k22.1}) =$$

$$= (2\pi)^{-\frac{1}{2}k(m-k)} \left|\Omega_{k22.1}\right|^{-\frac{1}{2}k} \left|\overline{Q}_{k11}\right|^{\frac{1}{2}(m-k)} \operatorname{etr}\left\{-\tfrac{1}{2}\Omega_{k22.1}^{-1}(\Omega_{k21.1} - \overline{Q}_{k21.1})\overline{Q}_{k11}(\Omega_{k21.1} - \overline{Q}_{k21.1})'\right\} \tag{22}$$

$$p(\Omega_{k22.1}) = 2^{-\frac{1}{2}\nu(m-k)}[\Gamma_{m-k}(\tfrac{\nu}{2})]^{-1} \left|\overline{Q}_{k22.1}\right|^{\frac{1}{2}\nu} \left|\Omega_{k22.1}\right|^{-\frac{1}{2}(\nu+m-k+1)} \operatorname{etr}\{-\tfrac{1}{2}\Omega_{k22.1}^{-1}\overline{Q}_{k22.1}\} \tag{23}$$

$$p(\Omega_{k11}) = 2^{-\frac{1}{2}(\nu-m+k)k}[\Gamma_k(\tfrac{\nu-m+k}{2})]^{-1} \left|\overline{Q}_{k11}\right|^{\frac{1}{2}(\nu-m+k)} \left|\Omega_{k11}\right|^{-\frac{1}{2}(\nu-(m-k)+k+1)} \operatorname{etr}\{-\tfrac{1}{2}\Omega_{k11}^{-1}\overline{Q}_{k11}\} \tag{24}$$

with:

$$\overline{Q} = \begin{bmatrix} \overline{Q}_{k11} & \overline{Q}_{k12} \\ \overline{Q}_{k21} & \overline{Q}_{k22} \end{bmatrix}; \qquad \overline{Q}_{k22.1} = \overline{Q}_{k22} - \overline{Q}_{k21}\overline{Q}_{k11}^{-1}\overline{Q}_{k12}; \qquad \overline{Q}_{k21.1} = \overline{Q}_{k21}\overline{Q}_{k11}^{-1}$$

$\overline{Q}_{k11} : k \times k$, $\overline{Q}_{k12} : k \times (m-k)$, $\overline{Q}_{k21} = \overline{Q}_{k12}'$, $\overline{Q}_{k22} : (m-k) \times (m-k)$

Note that the prior (22), (23) and (24) is obtained by the usual change of variables from Ω in the prior (17), to $\Omega_{k21.1}, \Omega_{k22.1}, \Omega_{k11}$ which allows us for such a prior decomposition i.e. (21).

VII. The Problem with Testing

The main purpose of the present paper is to develop the testing procedure for singularity of the multivariate model. Unfortunately such a singularity introduces much complication connected with the measure theoretic issues. As we operate in the Bayesian framework these show up when one tries to obtain the Prior Predictive

[7] In general, all symbols with a bar above throughout the paper indicate the prior hyperparameters.

Density (*PPD*), which is a basis for the Bayes Factor (*BF*) or Posterior Odds Ratio (*POR*). Accordingly we will demonstrate that the null hypothesis that k is equal to some integer between 1 and m, or what is equivalent, the null hypothesis $\Omega_{k22.1} = \mathbf{0}$ for some k, leads to serious pathology of the testing procedure, namely: we always reject the null hypothesis even if it is true. Let us explain this point in more details. If the null hypothesis is $k = i$ ($i = 1,\ldots,m$) or $\Omega_{k22.1} = \mathbf{0}$ for some k, we effectively have to derive *PPD* $m(\mathbf{Y} \mid \mathrm{rank}(\Omega) = k)$ ($m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$). The following proposition shows its form:

Proposition 1: *Given any rational prior one has*:

$$m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) \equiv m(\mathbf{Y} \mid \mathrm{rank}(\Omega) = k) =$$
$$= \int_{Y'_{m-k}[\mathbf{I}_T - \mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}'] = \mathbf{0}} L(\mathbf{Y}_k \mid B_k, \Omega_{k11}) p(\Omega_{k11}, B_k)(d\Omega_{k11})(dB_k)$$

where $\mathrm{F} = [X \vdots (Y_k - XB'_k)]$.

Proof: see appendix 2.

The rationality of the prior means in our context that if $\Omega_{k22.1} = \mathbf{0}$ we necessarily must have that $Y'_{m-k} = B_{m-k}X' + \Omega_{k21.1}(Y'_k - B_k X')$ since these are equivalent statements (i.e. "$\Leftrightarrow$" relation holds). Thus the rational prior under the null should also indicate that the latter constraint is fulfilled. On the other hand the proposition does not restrict in any way the form of the prior (except that the marginal prior for Ω_{k11}, B_k is absolutely continuous so it possesses a density function $p(\Omega_{k11}, B_k)$). In particular $p(\Omega_{k11}, B_k)$ may be any pdf. Note that the region of integration with respect to B_k (as it enters F) is constrained so as Y_{m-k} lies in the null space of $\mathbf{I}_T - \mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}'$. In general such a nonlinear constraint on the integration support can make analytical (or numerical) evaluation of the integral $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$ impossible or very difficult in practice. However for our further reasoning the specific form of $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$ is of no importance since we shall prove that using the latter in the testing leads to serious pathology.

Remark 5: There is a widespread misunderstanding in the literature concerning the maximum likelihood (ML) estimation of stochastically singular models. When one has less structural shocks than variables, the common view is that in order to estimate the model by ML method we should choose as many variables as there are structural shocks and there is no way to estimate all coefficients of the full model – see e.g. Ruge–Murcia (2003). This is only partially true. In fact we employ all variables in estimation process although this automatically imposes restrictions on many coefficients. To be precise what we do is to estimate the subsystem comprising e.g. the first k equations (which is equal to the number of shocks and makes the covariance nonsingular), and the remaining parameters that do not show up in the subsystem i.e. B_{m-k}, $\Omega_{k21.1}$ and $\Omega_{k22.1}$, are restricted by model construction (i.e. singularity). Specifically, in addition to the obvious restriction $\Omega_{k22.1} = \mathbf{0}$, it may be shown that

$$B_{m-k} = Y_{m-k}'XM^{-1} - Y_{m-k}'\Delta'(\Delta\Delta')^{-1}\Delta XM^{-1}$$

$$\Omega_{k21.1} = Y_{m-k}'\Delta'[(\Delta\Delta')^{-1} + (\Delta\Delta')^{-1}\Delta XM^{-1}X'\Delta'(\Delta\Delta')^{-1}] - Y_{m-k}'XM^{-1}X'\Delta'(\Delta\Delta')^{-1}$$

where $M = X'[\mathbf{I}_T - \Delta'(\Delta\Delta')^{-1}\Delta]X$ and $\Delta \equiv Y_k' - B_kX'$. Therefore both B_{m-k} and $\Omega_{k21.1}$ are deterministic functions of observations but also B_k (which is estimated within first k equations subsystem).

As is known from Bayes' formula, the posterior probability of the hypothesis $\Omega_{i22.1} = \mathbf{0}$ for some $i = 1,2,\ldots,m$ (that is $\text{rank}(\Omega) = i$) is given by:

$$P(\text{rank}(\Omega) = i \mid \mathbf{Y}) = \frac{\pi_i m(\mathbf{Y} \mid \text{rank}(\Omega) = i)}{\sum_{j=1}^{m} \pi_j m(\mathbf{Y} \mid \text{rank}(\Omega) = j)} \tag{25}$$

where $\pi_j \equiv p(\text{rank}(\Omega) = j)$ i.e. prior probability for the hypothesis and $\sum_{j=1}^{m}\pi_j = 1$. In order to demonstrate an inconsistency we should show that though the null hypothesis is true i.e. $\Omega_{k22.1} = \mathbf{0}$ for some $k = i$, $P(\text{rank}(\Omega) = i \mid \mathbf{Y})$ converges in probability to 0 as the number of observations tends to infinity. The formal and general proof of the above statement seems to be complicated by two facts. First of all, when the true value of the subset of parameters (i.e. $\Omega_{k22.1}$ for some k) lies on the boundary this would require to deal with directional or one–hand derivatives of the likelihood in the Taylor expansion. For the latter approach see e.g. Andrews (1999). But the most difficult problem to be solved is that under the null $\Omega_{k22.1} = \mathbf{0}$, the likelihood and the prior (integrands in *PPD*) are not defined at all. This makes it impossible to follow standard steps of the proof in which, after stating some technical conditions, one obtains the Taylor expansion of the likelihood function at the ML estimate, retains its quadratic terms and integrates it out. Evidently to prove inconsistency in a formal way there is a necessity to adopt quite nonstandard mathematical apparatus which would accommodate such a nonstandard situation. Unfortunately we could not propose any solution but we hope to cope with this in the future research. What we managed to show is that under the standard conjugate (Normal–Wishart) prior framework, posterior probability of accepting the true hypothesis converges in probability to 0. Of course one may use other specific prior setup but, as the Normal–Wishart prior dominates almost all empirical studies, we had a major candidate to work with. On the other hand since our goal is to show a clear defect of the Bayes procedure in this case (testing precise hypothesis), any prior will be as good as the other one. We adopt the following assumptions:

A1. $B_T^{-1}X'XB_T^{-1} = O_{\mathbf{p}}(1) \equiv \mathbf{Q} > 0$
where $B_T = diag(T^{\alpha_1}, T^{\alpha_2}, \ldots, T^{\alpha_l})$ and $\alpha_i \geq \frac{1}{2}$
A2. ML estimators of all parameters converge in probability to their (pseudo) true values.
A3. The joint prior is Normal–Wishart, as given in section VI.

Assumption A1 states that there is a scaling matrix B_T that ensures $B_T^{-1}X'XB_T^{-1}$ converges in distribution to (degenerated or nondegenerated) random matrix $\mathbf{Q}$ which is positive definite. This is quite general assumption that is valid for both stationary and nonstationary data (including deterministic trends, integrated and exploding processes, but also fractionally integrated processes).

Assumption A2 is necessary for making the arguments in the proof. The validity of this assumption for all parameters except those lying on the boundary is obvious. For those on the boundary i.e. $\Omega_{k22.1}$, the following reasoning will realize that the situation is not so hopeless as it would appear. First of all, as ML estimators $\hat{B}_{m-k}, \hat{B}_k$ and $\hat{\Omega}_{k21.1}$ converge to their (pseudo) true values we conclude $Y'_{m-k} - \hat{B}_{m-k}X' - \hat{\Omega}_{k21.1}(Y'_k - \hat{B}_kX') \xrightarrow{\mathbf{P}} Y'_{m-k} - B_{m-k}X' - \Omega_{k21.1}(Y'_k - B_kX')$. But as long as the covariance is singular we have $\Omega_{k22.1} = \mathbf{0} \Leftrightarrow Y'_{m-k} - B_{m-k}X' - \Omega_{k21.1}(Y'_k - B_kX') = \mathbf{0}$. Therefore consistency of $\hat{B}_{m-k}, \hat{B}_k$ and $\hat{\Omega}_{k21.1}$ implies that $\hat{\Omega}_{k22.1} \xrightarrow{\mathbf{P}} \mathbf{0}$, as it follows from a construction of ML estimator (see right below). Actually the speed in which the ML estimator of $\Omega_{k22.1}$ converges to 0, when $\Omega_{k22.1}$ is indeed equal to 0 (i.e. $\operatorname{rank}(\Omega) = k < m$) is impressively high. To see this note that the ML estimator $\hat{\Omega}_{k22.1}$ of $\Omega_{k22.1}$ takes the following form:

$$\hat{\Omega}_{k22.1} = \tfrac{1}{T}E'_{m-k}E_{m-k} - \tfrac{1}{T}E'_{m-k}E_k(\tfrac{1}{T}E'_kE_k)^{-1}\tfrac{1}{T}E'_kE_{m-k} \tag{26}$$

where $E'_i \equiv Y'_i - \hat{B}_iX'$ and $\hat{B}_i = Y'_iX(X'X)^{-1}$, $i = \{k, m-k\}$. We may rewrite this as:

$$T\hat{\Omega}_{k22.1} = (Y'_{m-k} - \hat{B}_{m-k}X' - \hat{\Omega}_{k21.1}(Y'_k - \hat{B}_kX'))(Y'_{m-k} - \hat{B}_{m-k}X' - \hat{\Omega}_{k21.1}(Y'_k - \hat{B}_kX'))'$$

where $\hat{\Omega}_{k21.1} = \frac{1}{T}E'_{m-k}E_k(\frac{1}{T}E'_kE_k)^{-1}$ is the ML estimator of $\Omega_{k21.1}$. But given that $Y'_{m-k} - \hat{B}_{m-k}X' - \hat{\Omega}_{k21.1}(Y'_k - \hat{B}_kX') \xrightarrow{\mathbf{P}} \mathbf{0}$, we also have $\hat{\Omega}_{k22.1} \xrightarrow{\mathbf{P}} \mathbf{0}$. Thus we obtain $\hat{\Omega}_{k22.1} = \frac{1}{T} \times o_{\mathbf{p}}(1) \times o_{\mathbf{p}}(1)$, and we conclude that $\hat{\Omega}_{k22.1}$ converges to 0 very fast.

Taking the above assumptions we have the following result which establishes the defect of BF as far as singularity testing is concerned and shows that even if the null hypothesis is true, BF will certainly favor the alternative (false) hypothesis in large samples:

Proposition 2: *Provided that assumptions* A1, A2, A3 *hold and if the null singularity hypothesis is true i.e.* $\Omega_{k22.1} = \mathbf{0}$ *for some* $k = 1,\ldots,m-1$, *BF testing the hypothesis* $\Omega_{k22.1} = \mathbf{0}$ *against* Ω *is nonsingular* (*i.e.* $\mathrm{rank}(\Omega) = m$) *results in*:

$$m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) / m(\mathbf{Y} \mid \mathrm{rank}(\Omega) = m) \xrightarrow{\mathbf{P}} 0$$

Proof: see appendix 3.

Of course it also means that for any combination of prior probabilities of hypotheses $\pi_j \equiv p(\mathrm{rank}(\Omega) = j)$, posterior probability of the true hypothesis also converges in probability to 0, since:

$$P(\mathrm{rank}(\Omega) = i \mid \mathbf{Y}) \leq \frac{\pi_i m(\mathbf{Y}|\mathrm{rank}(\Omega)=i)}{\pi_m m(\mathbf{Y}|\mathrm{rank}(\Omega)=m)} \equiv \frac{\pi_i}{\pi_m} \times \frac{m(\mathbf{Y}|\Omega_{i22.1}=\mathbf{0})}{m(\mathbf{Y}|\mathrm{rank}(\Omega)=m)} \xrightarrow{\mathbf{P}} 0 \qquad (27)$$

Thus we have the result:

Corollary 2: *Provided that assumptions* A1, A2, A3 *hold and even if the null hypothesis* $\mathrm{rank}(\Omega) = i$ *for some* $i = 1,\ldots,m-1$ (*i.e.* $\Omega_{i22.1} = \mathbf{0}$) *is true we have*:

$$P(\mathrm{rank}(\Omega) = i \mid \mathbf{Y}) \xrightarrow{\mathbf{P}} 0$$

VIII. The Framework for Testing The Number of Structural Shocks

In the previous section we pointed out to some pathology when adopting standard Bayesian toolkit for testing precise hypotheses. This was a consequence of the partial degeneracy of Prior Predictive Density (*PPD*), whose source lies in the singularity of the covariance. In the present section we lay down the methodology that circumvents the above problems and works well in practice. Our idea of testing the number of structural shocks in a model relies on an almost singular representation of the data density which will be a basis for a modified *PPD*.

In order to derive *PPD* related to the almost singular data density we shall provide with a prior which restricts the integration support so as it implies the data density is almost singular. To this end, it is very intuitive to introduce the prior that assigns almost all probability to the region $\Omega_{k22.1} \approx \mathbf{0}$, which may be called a "soft" restriction. Indeed, from corollary 1 we know that as $\Omega_{k22.1} \to \mathbf{0}$, the likelihood function converges weakly to the singular normal distribution of rank k. Of course this "soft" restriction will be imposed by the appropriate hyperparameters setting in the marginal prior $p(\Omega_{k22.1})$. One caveat is that this marginal prior i.e. $p(\Omega_{k22.1})$, was derived in section VI from the inverted Wishart prior for Ω i.e. $p(\Omega)$ (by the usual changing in variables technique). Thus, by manipulating the hyperparameters of the prior $p(\Omega_{k22.1})$, we effectively do the same for $p(\Omega)$. However we can assume that the priors $p(\Omega_{k21.1} \mid \Omega_{k22.1})$, $p(\Omega_{k22.1})$ and $p(\Omega_{k11})$ enter autonomously and interpret them as separate priors (of course preserving the conjugate setup). Taking this perspective we are legitimate to set the hyperparameters connected with $p(\Omega_{k22.1})$ that will indicate $\Omega_{k22.1} \approx \mathbf{0}$. In particular we can restate the prior for $\Omega_{k22.1}$, i.e. (23), as:

$$p_*(\Omega_{k22.1} \mid \nu_*, \bar{Q}^*_{k22.1}) = 2^{-\frac{1}{2}\nu_*(m-k)}[\Gamma_{m-k}(\tfrac{\nu_*}{2})]^{-1} \left|\bar{Q}^*_{k22.1}\right|^{\frac{1}{2}\nu_*} \left|\Omega_{k22.1}\right|^{-\frac{1}{2}(\nu_*+m-k+1)} \operatorname{etr}\{-\tfrac{1}{2}\Omega^{-1}_{k22.1}\bar{Q}^*_{k22.1}\} \quad (28)$$

where ν_* is a new "degree of freedom" hyperparameter. Setting $\nu_* \to \infty$ and/or $\bar{Q}^*_{k22.1} \to \mathbf{0}$, the prior $p_*(\Omega_{k22.1} \mid \nu_*, \bar{Q}^*_{k22.1})$ becomes more and more concentrated around $\mathbf{0}$. For further development we need the following lemma:

Lemma 2: *We have*:

1) $m(\mathbf{Y} \mid \nu_*, \bar{Q}^*_{k22.1}) = \int m(\mathbf{Y} \mid \Omega_{k22.1}) p_*(\Omega_{k22.1} \mid \nu_*, \bar{Q}^*_{k22.1})(d\Omega_{k22.1}) =$

$$= \pi^{-\frac{1}{2}mT}\Gamma_k(\tfrac{T+\nu-m+k}{2})[\Gamma_k(\tfrac{\nu-m+k}{2})]^{-1}\Gamma_{m-k}(\tfrac{T+\nu_*}{2})[\Gamma_{m-k}(\tfrac{\nu_*}{2})]^{-1} \times$$
$$\times\left|\bar{Z}\right|^{-\frac{1}{2}m}\left|X'X+\bar{Z}^{-1}\right|^{-\frac{1}{2}m}\left|\bar{Q}_{k11}\right|^{\frac{1}{2}\nu}\left|\bar{Q}^*_{k22.1}\right|^{\frac{1}{2}\nu_*}\left|U_k+\bar{Q}^*_{k22.1}\right|^{-\frac{1}{2}(T+\nu_*)} \times$$
$$\times\left|\bar{Q}_{k11}+(\hat{B}_k-\bar{B}_k)((X'X)^{-1}+\bar{Z})^{-1}(\hat{B}_k-\bar{B}_k)'+(Y_k'-\hat{B}_kX')(Y_k'-\hat{B}_kX')'\right|^{-\frac{1}{2}(T+\nu)}$$

2) $m(\mathbf{Y} \mid \operatorname{rank}(\Omega) = m) = \int m(\mathbf{Y} \mid \Omega_{k22.1}) p(\Omega_{k22.1})(d\Omega_{k22.1}) =$

$$= \pi^{-\frac{1}{2}mT}\Gamma_k(\tfrac{T+\nu-m+k}{2})[\Gamma_k(\tfrac{\nu-m+k}{2})]^{-1}\Gamma_{m-k}(\tfrac{T+\nu}{2})[\Gamma_{m-k}(\tfrac{\nu}{2})]^{-1} \times$$
$$\times\left|\bar{Z}\right|^{-\frac{1}{2}m}\left|X'X+\bar{Z}^{-1}\right|^{-\frac{1}{2}m}\left|\bar{Q}_{k11}\right|^{\frac{1}{2}\nu}\left|\bar{Q}_{k22.1}\right|^{\frac{1}{2}\nu}\left|U_k+\bar{Q}_{k22.1}\right|^{-\frac{1}{2}(T+\nu)} \times$$
$$\times\left|\bar{Q}_{k11}+(\hat{B}_k-\bar{B}_k)((X'X)^{-1}+\bar{Z})^{-1}(\hat{B}_k-\bar{B}_k)'+(Y_k'-\hat{B}_kX')(Y_k'-\hat{B}_kX')'\right|^{-\frac{1}{2}(T+\nu)}$$

where:

$$m(\mathbf{Y} \mid \Omega_{k22.1}) = \int L(\mathbf{Y} \mid B_k, B_{m-k}, \Omega_{k22.1}, \Omega_{k21.1}, \Omega_{k11}) p(B_{m-k} \mid B_k, \Omega_{k22.1}, \Omega_{k21.1}) p(B_k \mid \Omega_{k11}) \times$$
$$\times p(\Omega_{k21.1} \mid \Omega_{k22.1}) p(\Omega_{k11})(dB_{m-k})(dB_k)(d\Omega_{k21.1})(d\Omega_{k11});$$

$L(\mathbf{Y} \mid B_k, B_{m-k}, \Omega_{k22.1}, \Omega_{k21.1}, \Omega_{k11}) \equiv L(\mathbf{Y} \mid B, \Omega)$ *is given in* (9), $p(\Omega_{k22.1})$ *and all prior pdf's in the integrand of* $m(\mathbf{Y} \mid \Omega_{k22.1})$ *were defined in section* VI. *Furthermore*:

$$\mathrm{U}_k = (E'_{m-k} - \hat{\Omega}_{k21.1}E'_k)(E'_{m-k} - \hat{\Omega}_{k21.1}E'_k)' + (D_{m-k} - \Omega^*_{k21.1}D_k)((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}(D_{m-k} - \Omega^*_{k21.1}D_k)' +$$
$$+E'_{m-k}E_k(E'_kE_k)^+ E'_kE_{m-k} + \Omega^*_{k21.1}D_k((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}D'_k\Omega'^*_{k21.1} + \overline{Q}_{k21}\overline{Q}^{-1}_{k11}\overline{Q}'_{k21} -$$
$$-(E'_kE_{m-k} + D_k((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}D'_k\Omega'^*_{k21.1} + \overline{Q}'_{k21})' \times$$
$$\times(E'_kE_k + D_k((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}D'_k + \overline{Q}_{k11})^{-1} \times$$
$$\times(E'_kE_{m-k} + D_k((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}D'_k\Omega'^*_{k21.1} + \overline{Q}'_{k21})$$

where:

$D_i = \hat{B}_i - \overline{B}_i$; $\quad E'_i = Y'_i - \hat{B}_iX'$; $\quad \hat{B}_i = Y'_iX(X'X)^{-1}$ $\quad$ for $i = \{k, m-k\}$

$\hat{\Omega}_{k21.1} = \frac{1}{T}E'_{m-k}E_k(\frac{1}{T}E'_kE_k)^+$, where $(\cdot)^+$ denotes Moore–Penrose (M–P) inverse

$\Omega^*_{k21.1} = D_{m-k}((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}D'_k(D_k((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}D'_k)^{-1}$

Proof: straightforward but tedious, hence omitted.

In what follows we assume $\overline{Q}^*_{k22.1} = \overline{Q}_{k22.1}$ i.e. this hyperparameter in $p_*(\Omega_{k22.1} \mid \nu_*, \overline{Q}^*_{k22.1})$ is equal to the counterpart hyperparameter in $p(\Omega_{k22.1})$. This particular case greatly simplifies an asymptotic reasoning, yet it turns out to be sufficient for desirable behavior of the testing procedure. Of course the remaining hyperparameter from $p_*(\Omega_{k22.1} \mid \nu_*, \overline{Q}^*_{k22.1})$, that is the degree of freedom ν_*, will be retained to be distinct from the degree of freedom hyperparameter in $p(\Omega_{k22.1})$ i.e. ν, and will be responsible for expressing the concentration. Actually to keep matters simple we consider the case $\nu_* = \alpha T$, where α is positive constant. Such a setting is sufficiently general for incorporating a belief that the prior for $\Omega_{k22.1}$ becomes increasingly concentrated around 0 as $T \to \infty$ (all moments of this pdf will become 0). This also explains why we consider only the case $\overline{Q}^*_{k22.1} = \overline{Q}_{k22.1}$: because it does not matter asymptotically (unless $\overline{Q}^*_{k22.1}$ is a function of T). Thus in order to control the shrinkage of the support in $p_*(\Omega_{k22.1} \mid \nu_*, \overline{Q}^*_{k22.1})$ we are justified in focusing only on ν_*. In fact, by denoting prior probability measure induced by the density $p_*(\Omega_{k22.1} \mid \nu_*, \overline{Q}^*_{k22.1})$ as $\Pi_0(\Omega_{k22.1} \mid \nu_*)$, we easily find that, as $\nu_* \to \infty$, $\Pi_0(\Omega_{k22.1} \mid \nu_*)$ converges weakly to the degenerated probability measure having a single saltus at the point $\Omega_{k22.1} = \mathbf{0}$. Note however that convergence takes place at all points except the most interesting point i.e. $\Omega_{k22.1} = \mathbf{0}$. Of course, there is nothing unusual in this fact since the point $\Omega_{k22.1} = \mathbf{0}$ is a discontinuity point of the degenerated measure which is a limiting probability measure for $\Pi_0(\Omega_{k22.1} \mid \nu_*)$ (as $\nu_* \to \infty$). Actually, $p_*(\Omega_{k22.1} \mid \nu_*, \overline{Q}^*_{k22.1})$ becomes (as $\nu_* \to \infty$) a Dirac delta function which is not a function in mathematical sense but a distribution. Anyway, taking into account lemma 2, let us define:

$$m(\mathbf{Y}\mid\Omega_{k22.1}\approx\mathbf{0})\equiv\lim_{\nu_*\to\infty}m(\mathbf{Y}\mid\nu_*,\bar{Q}^*_{k22.1})=$$
$$=\lim_{\nu_*\to\infty}\int m(\mathbf{Y}\mid\Omega_{k22.1})p_*(\Omega_{k22.1}\mid\nu_*,\bar{Q}^*_{k22.1})(d\Omega_{k22.1}) \tag{29}$$

Given lemma 2 and provided that $\bar{Q}^*_{k22.1}=\bar{Q}_{k22.1}$ it is easy to check:

$$BF^*_k\equiv m(\mathbf{Y}\mid\Omega_{k22.1}\approx\mathbf{0})\,/\,m(\mathbf{Y}\mid \mathrm{rank}(\Omega)=m)\,|_{\bar{Q}^*_{k22.1}=\bar{Q}_{k22.1}}=$$
$$=\lim_{\nu_*\to\infty}\Gamma_{m-k}(\tfrac{T+\nu_*}{2})[\Gamma_{m-k}(\tfrac{\nu_*}{2})]^{-1}\Gamma_{m-k}(\tfrac{\nu}{2})[\Gamma_{m-k}(\tfrac{T+\nu}{2})]^{-1}\left|\bar{Q}_{k22.1}\right|^{\frac{1}{2}(\nu_*-\nu)}\left|\mathrm{U}_k+\bar{Q}_{k22.1}\right|^{-\frac{1}{2}(\nu_*-\nu)} \tag{30}$$

where "$|_{\bar{Q}^*_{k22.1}=\bar{Q}_{k22.1}}$" means "under the condition $\bar{Q}^*_{k22.1}=\bar{Q}_{k22.1}$". We denote the above BF as BF^*_k to emphasize that this is our key testing tool and the most important practical result in the paper follows:

Proposition 3: *Given the assumptions* A1, A2, A3, *setting* $\bar{Q}^*_{k22.1}=\bar{Q}_{k22.1}$ *and* $\nu_*=\alpha T$, *where* α *is positive constant, we have*:

a) *if* $\mathrm{rank}(\Omega)=k<m$ *i.e.* $\Omega_{k22.1}=\mathbf{0}$ *and* α *fulfills* $\frac{\nu}{T}<\alpha<(\frac{|\mathrm{U}_k+\bar{Q}_{k22.1}|}{|\bar{Q}_{k22.1}|}-1)^{-1}$, *then*:

$\ln BF^*_k\to\infty$ as $T\to\infty$

b) *if* Ω *is nonsingular, and* α *fulfills* $\frac{(\alpha T-\nu)^{\alpha(m-k)}}{(1+\alpha)^{(1+\alpha)(m-k)}}(\frac{|\Omega_{k22.1}|}{|\bar{Q}_{k22.1}|})^{\alpha}>1$ *where* $\Omega_{k22.1}$ *is true value of the parameter, then*:

$\ln BF^*_k\to-\infty$ as $T\to\infty$

c) *if* $\mathrm{rank}(\Omega)=i$, *i.e.* $\Omega_{i22.1}=\mathbf{0}$, *for some* $i=2,\ldots,m-1$ *and* $i>j$, *provided that* $\alpha>\frac{2m}{T}$, *then*:

$\ln[m(\mathbf{Y}\mid\Omega_{i22.1}\approx\mathbf{0})/m(\mathbf{Y}\mid\Omega_{j22.1}\approx\mathbf{0})]=\ln BF^*_i-\ln BF^*_j\to\infty$ as $T\to\infty$

d) *if* $\mathrm{rank}(\Omega)=j$, *i.e.* $\Omega_{j22.1}=\mathbf{0}$, *for some* $j=1,\ldots,m-2$ *and* $i>j$, *provided that* $\frac{\nu}{T}<\alpha<(\frac{|\mathrm{U}_j+\bar{Q}_{j22.1}|}{|\mathrm{U}_i+\bar{Q}_{i22.1}|}\frac{|\bar{Q}_{i22.1}|}{|\bar{Q}_{j22.1}|}-1)^{-1}$, *then*:

$\ln[m(\mathbf{Y}\mid\Omega_{i22.1}\approx\mathbf{0})/m(\mathbf{Y}\mid\Omega_{j22.1}\approx\mathbf{0})]=\ln BF^*_i-\ln BF^*_j\to-\infty$ as $T\to\infty$

and the formula for BF^*_i *is given in* (30).

Proof: see appendix 4.

The result a) states that if indeed $\Omega_{k22.1}=\mathbf{0}$ and we set $\nu<\nu_*<(\frac{|\mathrm{U}_k+\bar{Q}_{k22.1}|}{|\bar{Q}_{k22.1}|}-1)^{-1}T$ then, as $T\to\infty$, BF^*_k, which is essentially BF testing the approximation to the hypothesis $\Omega_{k22.1}=\mathbf{0}$ against the hypothesis that covariance is nonsingular, converges to infinity. Note however that the requirements of the condition $\frac{\nu}{T}<\alpha<(\frac{|\mathrm{U}_k+\bar{Q}_{k22.1}|}{|\bar{Q}_{k22.1}|}-1)^{-1}$ differ drastically between the singular and nonsingular covariance case. If $\Omega_{k22.1}=\mathbf{0}$, $\mathrm{U}_k+\bar{Q}_{k22.1}=O_{\mathbf{p}}(1)$ (see appendix 3). On the other hand if the covariance is nonsingular then $\mathrm{U}_k+\bar{Q}_{k22.1}=O_{\mathbf{p}}(T)$ (see appendix 3).

Thus in the latter case the upper bound for ν_* becomes only bounded in probability (i.e. $O_p(1)$), whereas under singularity the upper bound for ν_* is $O_p(T)$. But this result is reasonable as it says that if Ω is nonsingular then there is no ν_* which grows linearly with T that makes us accept the null (false) hypothesis that Ω is singular.

The result b) says that if Ω is nonsingular then setting $\nu_* = \alpha T$, where α meets the condition $\frac{(\alpha T-\nu)^{\alpha(m-k)}}{(1+\alpha)^{(1+\alpha)(m-k)}}(\frac{|\Omega_{k22.1}|}{|\bar{Q}_{k22.1}|})^{\alpha} > 1$, will make BF_k^* behave consistently. As a matter of fact we could state that under b), for T large enough, we may set any $\alpha > 0$ since the left hand of this condition diverges to infinity as $T \to \infty$.

The results c) and d) consider two other cases, which taken together with a) and b), exhaust the total set of possibilities. The condition for c) does not require any comments. On the other hand in the case d), in order to ensure consistency, α must be set so as $\frac{\nu}{T} < \alpha < (\frac{|\mathrm{U}_j+\bar{Q}_{j22.1}|}{|\mathrm{U}_i+\bar{Q}_{i22.1}|}\frac{|\bar{Q}_{i22.1}|}{|\bar{Q}_{j22.1}|}-1)^{-1}$. If the true null is $\mathrm{rank}(\Omega) = j$, then from results in appendix 4 we conclude that the optimal α is bounded in probability. Unfortunately we can offer no additional insights.

The proposition has direct and immediate consequence, namely: if we use the testing procedure from this section which builds on almost singular representation, the posterior probability of the true hypothesis converges in probability to 1 (provided that the conditions from proposition 3 hold). The main practical question is what values we should assign to α. Unfortunately, the conditions from proposition 3 depend on the data at hand, hyperparameters or even on the true values of parameters as in the case b). It appears that putting α which is slightly but distinctively more that 0 seems to be a good practice. Of course the conditions in proposition 3 are only sufficient and as we found from simulations, the values for α in the whole range $0 < \alpha \leq 0.5$ work surprisingly well.

Proposition 3 establishes the appropriate behavior of the formula (30) which is our proposal for testing singularity of the covariance matrix. In particular the testing procedure is valid for stationary and nonstationary data (assumption A1) but was derived under the particular prior setup i.e. Normal–Wishart. However taking into account that much (almost all) of the empirical work uses the standard Normal–Wishart prior our result may be just what is needed, but we agree that it is useful to have a result for arbitrary prior setup. Moreover, as BF_k^* takes a very simple form and what is more important, is derived analytically i.e. without any resort to MCMC methods, our testing procedure may be routinely used in all empirical studies where the issue of singularity is of great economic importance e.g. DSGE modeling.

IX. The Relationship Between $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$ AND $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$

One may wonder why $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$ derived in proposition 1 leads to inconsistency of the testing procedure whereas $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$, which may seem to be an approximation to $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$, forms a basis for a well behaved testing procedure. The logical answer is that $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$ is not an approximation to $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$ at all. If this is the case then what is the meaning and rationale behind $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$? To clarify this point, consider the following Lebesgue–Stieltjes integral:

$$\int L(\mathbf{Y} \mid \Omega_{k11}, \Omega_{k21.1}, B_k, B_{m-k}, \Omega_{k22.1}) d(\lim_{\nu_* \to \infty} \Pi_0(\Omega_{k21.1}, \Omega_{k11}, B_{m-k}, B_k, \Omega_{k22.1} \mid \nu_*)) \tag{31}$$

in which $\Pi_0(\cdot \mid \nu_*)$ denotes the prior probability measure induced by the prior densities used earlier. In particular the conditioning on ν_* emphasizes that the underlying marginal density for $\Omega_{k22.1}$ is $p_*(\Omega_{k22.1} \mid \nu_*, \bar{Q}^*_{k22.1})$ as defined in (28). Assume that the following factorization of the prior measure holds:

$$\Pi_0(\Omega_{k21.1}, \Omega_{k11}, B_{m-k}, B_k, \Omega_{k22.1} \mid \nu_*) = \Pi_0(\Omega_{k21.1}, \Omega_{k11}, B_{m-k}, B_k \mid \Omega_{k22.1}) \Pi_0(\Omega_{k22.1} \mid \nu_*) \tag{32}$$

in what follows $\Rightarrow$ denotes the weak convergence of measures and it should be understood that it refers to the case when $\nu_* \to \infty$. We have:

$$\Pi_0(\Omega_{k22.1} \mid \nu_* \to \infty) \Rightarrow \delta(\Omega_{k22.1} = \mathbf{0}) \tag{33}$$

where $\delta(\Omega_{k22.1} = \mathbf{0})$ is a degenerated measure having a single saltus at $\Omega_{k22.1} = \mathbf{0}$. Of course this implies:

$$\begin{aligned} &\Pi_0(\Omega_{k21.1}, \Omega_{k11}, B_{m-k}, B_k \mid \Omega_{k22.1}) \Pi_0(\Omega_{k22.1} \mid \nu_*) \Rightarrow \\ &\qquad \Rightarrow \Pi_0(\Omega_{k21.1}, \Omega_{k11}, B_{m-k}, B_k \mid \Omega_{k22.1} = \mathbf{0}) \delta(\Omega_{k22.1} = \mathbf{0}) \end{aligned} \tag{34}$$

Thus the limit of the integral (31), as $\nu_* \to \infty$, is:

$$\int L(\mathbf{Y} \mid \Omega_{k11}, \Omega_{k21.1}, B_k, B_{m-k}, \Omega_{k22.1} = \mathbf{0})(d\Pi_0(\Omega_{k21.1}, \Omega_{k11}, B_{m-k}, B_k \mid \Omega_{k22.1} = \mathbf{0})) \tag{35}$$

but this is simply a definition of $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$. On the other hand $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$ is defined as:

$$\lim_{\nu_* \to \infty} \int L(\mathbf{Y} \mid \Omega_{k11}, \Omega_{k21.1}, B_k, B_{m-k}, \Omega_{k22.1})(d\Pi_0(\Omega_{k21.1}, \Omega_{k11}, B_{m-k}, B_k, \Omega_{k22.1} \mid \nu_*)) \tag{36}$$

Clearly the very reason why $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$ and $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$ differ is that in our case the integral and the limit operator can not be interchanged. It makes our problem peculiar and distinct from most real applications in which the limit operator can be taken indifferently. What prevents us from doing this is the fact that the data density function $L(\mathbf{Y} \mid \Omega_{k11}, \Omega_{k21.1}, B_k, B_{m-k}, \Omega_{k22.1})$ is discontinuous (i.e. not defined) at

$\Omega_{k22.1} = \mathbf{0}$ and importantly the limiting measure of $\Pi_0(\Omega_{k22.1} \mid \nu_* \to \infty)$ (i.e. degenerated measure) gives probability 1 to that point. This violates the condition for weak convergence of the mixtures (i.e. *PPD*'s) as given by theorem 1 in Teicher (1960). This explains why we do not have $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0}) \Rightarrow m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$.

It is natural to ask the question why we should prefer $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$ to $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$. The main reason for such a choice was given in a previous section, namely: the former leads to the consistent testing whereas the latter suffers from serious pathology. In fact as we have only two variants of possible choice we definitely choose the one on which we can build a consistent testing framework. Thus in our case the fact that $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$ does not converge weakly to $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$ is quite fortunate. On the other hand the more basic construction seems to be $m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0})$ since it refers to limiting values of *PPD* itself (hence indirectly to probabilities of hypotheses) and not to limits of integrands comprising *PPD*.

X. The Simulation Study

In order to test our procedure we simulated an artificial data set. To keep things simple we restrict ourselves to 4–dimensional VAR(1) process with a constant: $y_t = c + A_1 y_{t-1} + A_0 \varepsilon_t$, where $c = (0.5, 6, 7, 0.9)'$, $A_1 = diag(1, 0.99, 0.97, 0.999)$. Initializing values for the process were $y_0 = (200, 300, 240, 300)'$. We provide with a simulation for a model with $k = 1,2,3,4$ structural shocks such that $\varepsilon_t \sim \mathbf{IID}N(\mathbf{0}, \mathbf{I}_k)$. Of course for any given k, A_0 is $4 \times k$ matrix of rank k. For simplicity we take it as a submatrix consisting of the first k columns of the general matrix which reads:

$$\begin{bmatrix} -1.005 & 2.9 & 0.91 & -0.7 \\ 1.7 & -2.1 & -1.4 & 1 \\ -0.9 & 0.93 & -1.9 & 1.1 \\ 0.6 & -0.9 & 1.4 & -1.5 \end{bmatrix} \tag{37}$$

thus e.g. for $k = 1$, $A_0 = (-1.005, 1.7, -0.9, 0.6)'$. As far as the prior hyperparameters are concerned we set $\nu_* = 0.4T$ i.e. $\alpha = 0.4$, $\nu = 6$, $\overline{B} = [\mathbf{0} \vdots \mathbf{I}_4]$ (with $\mathbf{0}$, 4×1 vector that stands for prior mean for constants), $\overline{\mathrm{Z}} = diag(10, \frac{0.01}{\mathrm{var}(y_1)}, \frac{0.01}{\mathrm{var}(y_2)}, \frac{0.01}{\mathrm{var}(y_3)}, \frac{0.01}{\mathrm{var}(y_4)})$ where $\mathrm{var}(y_i)$ is estimated error variance from fitting univariate AR(1) regression on the $i-$th variable i.e. $\mathrm{var}(y_i) = \frac{1}{n}\sum_{t=1}^{n}(y_{it} - \hat{\beta}_0 - \hat{\beta}_1 y_{it-1})^2$ ($\hat{\beta}_0, \hat{\beta}_1$ are OLS estimators), $\overline{Q} = diag(\mathrm{var}(y_1), \ldots, \mathrm{var}(y_4))$ (note that by assumption $\overline{Q}^*_{k22.1} = \overline{Q}_{k22.1}$). For each k, we simulated 10000 series of length $T = 100$, and assuming uniform prior for rank of the covariance matrix $\Omega = A_0 A_0'$ i.e. $p(\mathrm{rank}(\Omega) = j) = \frac{1}{4}$ for $j = 1,2,3,4$, we computed the posterior probability for rank of Ω. Testing procedure is based on BF^*_k as explained in section VIII. The results in Table 1 report the frequency of choice (acceptance rate) of each hypothesis[8] $(k=1,2,3,4)$ in 10000 simulations. The true hypothesis is indicated in the first column. The test turns out to behave remarkably well. Only when the covariance is nonsingular (i.e. the last row), there is a minor share of cases when we inconsistently choose the false hypothesis $k=3$.

Table 1 Relative frequency of choosing each hypothesis for *k* = 1,2,3,4

true rank	Accept. Rate $k=1$	Accept. Rate $k=2$	Accept. Rate $k=3$	Accept. Rate $k=4$
$k=1$	1	0	0	0
$k=2$	0	1	0	0
$k=3$	0	0	1	0
$k=4$	0	0	0.024	0.976

Apparently, there is one flaw in our testing procedure. One may argue that the world without the measurement error does not exist. Thus any attempt to test the precise null hypothesis about the singularity of covariance is pointless. There are a few comments on this accusation that are worth mentioning. First of all, even in standard situation the testing of the precise null hypotheses is controversial and many people are very skeptical about its goal – see e.g. Berger and Delampady (1987) with discussion and further references therein. In most cases this hypothesis should be thought as a statement that the probability in the vicinity of this point is unusually large – see e.g. Jeffreys (1961), p. 367. That is we should read it as an approximation to the point hypothesis. This interpretation is particularly well suited for our case, since due to pathology of the standard BF we employed its limiting version. This explicitly considers not the point null (i.e. covariance singularity), but "almost singularity". In fact, almost singular representation of the normal distribution makes it evident that we are testing that some structural shocks have extraordinarily small variation which may be well captured by the singular covariance, which is however only a useful simplification and not the fact itself. Thus our interpretation of the testing procedure is that it checks whether measurement errors for variables in some economic relations are very small and at the same time, those empirical relations reflect underlying true economic relationships quite well. Clearly, if the economic

[8] The hypothesis is accepted for which the posterior probability is the largest.

theory that presupposes less structural shocks than variables is true, then even in the presence of some measurement error the variation of some shocks will be very much smaller than that of the remaining ones, and our limiting version of *BF* should detect it. To demonstrate that this is the case, we modified the above simulation study by introducing some measurement error. To this end we employed the following procedure. For a given k, A_0 was formed from (37) by retaining not only its first k columns, but additionally all the remaining ones with a multiplication them by a factor 0.01. For example, for $k = 2$, A_0 takes the form:

$$\begin{bmatrix} -1.005 & 2.9 & 0.01\times 0.91 & 0.01\times -0.7 \\ 1.7 & -2.1 & 0.01\times -1.4 & 0.01\times 1 \\ -0.9 & 0.93 & 0.01\times -1.9 & 0.01\times 1.1 \\ 0.6 & -0.9 & 0.01\times 1.4 & 0.01\times -1.5 \end{bmatrix}$$

Of course such a design makes the covariance nonsingular, but what is more important, it suggests that there are less dominant shocks than the variables. Those "unimportant" shocks are interpreted as measurement errors. We repeated the same exercise as before (including the same hyperparameters setup), and the results are practically indistinguishable from the earlier ones – see Table 2.

Table 2 Relative frequency of choosing each hypothesis for *k* = 1,2,3,4 (model with measurement error)

true rank	Accept. Rate $k=1$	Accept. Rate $k=2$	Accept. Rate $k=3$	Accept. Rate $k=4$
$k=1$	1	0	0	0
$k=2$	0	1	0	0
$k=3$	0	0	1	0
$k=4$	0	0	0.027	0.973

Of course this is a consequence of the fact that our testing procedure relies on an almost singular representation (i.e. limiting version of *BF*), and literally speaking, it tests whether there are some shocks with variance close to 0 but not exactly so. Needless to say, it works well even when there is no measurement error – as documented in Table 1.

XI. The Case Study

For empirical illustration we use the data set that appeared earlier in Ireland (2004) (the data are available at http://www2.bc.edu/~irelandp/programs.html). These are 3 series of quarterly data: output, consumption and hours worked. The data are from 1948:I to 2002:II and were seasonally adjusted. All series were in logarithms. For detailed description of the construction of the above series we refer to Ireland (2004). What is important, these 3 macroeconomic series (together with investment which however is subject to constraint tying output, consumption and investment hence omitted in estimation) comprise the basis for standard macroeconomic model e.g. Hansen's (1985) real business cycle model. The important economic question is how many structural shocks are driving forces for such a simplified model of economy. In their seminal work, Kydland and Prescott (1982) opt for only one structural (technology) shock i.e. $k = 1$. Using our approach we shall test whether there is any evidence for the presence of only one shock (i.e. technology shock). As we are mainly concerned about the number of shocks we are satisfied with working with the reduced form models. However we emphasize that in general our framework is perfectly suited for comparing economic (i.e. highly structural) models which may or may not be stochastically singular. Needless to say, they could be characterized by distinct number of structural shocks. Indeed, our application serves only as an illustration.

We fitted a 3–variables VAR model with a constant and 5 lags. We chose the number of lags subjectively, inspired by the quarterly character of the data, but the testing procedure turned out to be robust with respect to the lag number. We used a specification of prior hyperparameters which is in the spirit of Minnesota–like prior. In particular $\nu_* = 0.3T$ i.e. $\alpha = 0.3$, $\nu = 5$, $\overline{B} = [\mathbf{0}_{3\times1} \vdots \mathbf{I}_3 \vdots \mathbf{0}_{3\times12}]$, $\overline{\mathrm{Z}}$ is block diagonal matrix with the upper left (1,1) entry equal to 10 and the remaining 5 blocks defined as $\overline{\mathrm{Z}}_i = diag(\frac{0.01}{i\times\mathrm{var}(y_1)}, \frac{0.01}{i\times\mathrm{var}(y_2)}, \frac{0.01}{i\times\mathrm{var}(y_3)})$ for $i = 1,\ldots,5$. $\mathrm{var}(y_i)$ is estimated error variance from fitting univariate AR(5) regression on the $i-$th variable i.e. $\mathrm{var}(y_i) = \frac{1}{n}\sum_{t=1}^{n}(y_{it} - \hat{\beta}_0 - \hat{\beta}_1 y_{it-1} - \ldots - \hat{\beta}_5 y_{it-5})^2$ ($\hat{\beta}_j$ for $j = 0,1,\ldots,5$ are OLS estimators), $\overline{Q} = diag(\mathrm{var}(y_1), \mathrm{var}(y_2), \mathrm{var}(y_3))$. As argued in previous sections we take $\overline{Q}^*_{k22.1} = \overline{Q}_{k22.1}$. We assume equal prior probability for 3 variants of the model i.e. $k = 1,2,3$, so as they do not show up in POR.

Using the framework developed in this paper we checked a possibility of singularity of the model. Using BF^*_k from section VIII, we conducted the test. In particular, $\ln BF^*_1 \approx -147.6$ and $\ln BF^*_2 \approx -75.8$ which according to Jeffreys' "rule of thumb" is decisive (actually overwhelming) evidence against both $k = 1$ and $k = 2$. Therefore in our 3–variables VAR, we found no evidence that there may be less structural shocks than variables. Of course, bearing in mind our discussion in the end of section I, an economist may still maintain the hypothesis of the singularity for his (her) theoretical purposes. Such an economist is at least left with a formal tool that is able to compare hypotheses within the singular covariance framework (e.g. testing particular coefficient in stochastically singular economic model).

XII. The Summary and Conclusion

In the paper, we tried to formally deal with stochastically singular models. Doing so, we confronted many, not necessarily true, common views about the stochastic singularity. As it turns out, the singularity introduces many complications of the measure–theoretic nature which we explained in detail and tried to get around. Using the Normal–Wishart prior we showed that the naïve Bayesian testing tool is inconsistent i.e. even asymptotically chooses the false hypothesis. Recognizing this fact, we proposed a framework to successfully deal with singular models.

Using small simulation study we convinced ourselves that the approach works remarkably well. On the other hand, applying our framework to 3 macroeconomic variables we found no evidence for stochastic singularity of the underlying model.

There is much work to be done in order to complete our research. First of all in order to establish an inconsistency of the naïve BF we used standard conjugate prior since we had no mathematical apparatus to deal with more general prior structures. One would be more satisfied if we went beyond the Normal–Wishart prior.

The same conjugate prior i.e. Normal–Wishart, was adopted for the general framework in the paper. Although the conjugate prior has many virtues, most of all it enables for analytical results which avoid MCMC treatment, it looses some appeal when one draws theoretical conclusions. It would be useful to have the results for more general prior setup.

As it is more natural to analyze the linear approximations of DSGE models using the state–space representation, one may try to generalize the development in this paper to accommodate such a case. This however does not seem to introduce any serious complications, but may be worth working it out explicitly.

APPENDIX 1:

In this appendix we shall derive the density of the singular normal distribution. Throughout this appendix we use the notation from section III. First, we will obtain the characteristic function of $x = \mu + C_1 u_1$:

$$\mathbf{E}(\exp\{it'x\}) = \mathbf{E}(\exp\{it'(\mu + C_1 u_1)\}) = \mathbf{E}(\exp\{it'\mu + it'C_1 u_1\})$$

by defining $t_* = C_1' t$ and using result on characteristic function of multivariate normal distribution (note $u_1 \sim N(\mathbf{0}, \mathbf{I}_k)$) we find:

$$\mathbf{E}(\exp\{it'C_1 u_1\}) \equiv \mathbf{E}(\exp\{it_*' u_1\}) = \exp\{-\tfrac{1}{2} t_*' t_*\} \equiv \exp\{-\tfrac{1}{2} t' C_1 C_1' t\}$$

thus:

$$\mathbf{E}(\exp\{it'x\}) = \exp\{it'\mu - \tfrac{1}{2} t' C_1 C_1' t\}$$

It is easily verified this is characteristic function of the $m-$variate normal distribution with mean μ and covariance $\Sigma \equiv C_1 C_1'$. However $C_1 C_1'$ is singular, $\operatorname{rank}(C_1 C_1') = k$ and $\det(\Sigma) = 0$. Consequently there corresponds no density with respect to Lebesgue measure in $\mathbb{R}^m$. Let us augment C_1 with $C_2 : m \times (m-k)$ and write $C = [C_1 \vdots C_2]$ and assume $\operatorname{rank}(C) = m$. We introduce the following representation which stems from definition 2 in the paper:

$$x^* = \mu + C \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \lambda^{\frac{1}{2}} \mathbf{I}_{m-k} \end{bmatrix} u$$

where $u = [u_1' \vdots u_2']'$, $u \sim N(\mathbf{0}, \mathbf{I}_m)$ and $\lambda \geq 0$ is a constant. We note that $\lambda^{\frac{1}{2}}$ corresponds to α in definition 2. The relabeling was introduced only for practical reasons. Clearly, when $\lambda = 0$ we arrive at the original (singular) representation i.e. $x^* \equiv x$. For $\lambda > 0$ we face the nonsingular representation from which we have:

$$\mathbf{E}(\exp\{it'x^*\}) = \exp\{it'\mu - \tfrac{1}{2} t' \Sigma^* t\}$$

$$\Sigma^* = C \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \lambda \mathbf{I}_{m-k} \end{bmatrix} C'$$

The reason for introducing the parameter λ is to allow for investigating the limiting case as $\lambda \to 0$. In particular since $\exp\{it'\mu - \frac{1}{2} t' \Sigma^* t\} \to \exp\{it'\mu - \frac{1}{2} t' \Sigma t\}$ as $\lambda \to 0$, and the limiting characteristic function is continuous at $t = 0$, by the continuity theorem for characteristic function – see e.g. Cramér (1946) p. 102, the distribution function of x^* converges to singular distribution function of x. Thus a singular normal distribution may be regarded as a limit of a nonsingular normal distribution of x^* as $\lambda \to 0$. This observation together with some technicalities below is essentially the proof of corollary 1.

Since C is nonsingular we can apply the QR decomposition, and without loss of generality (but with some abuse to notation) we can choose C to be lower triangular. Thus we can write:

$$C = \begin{bmatrix} C_{11} & \mathbf{0} \\ C_{21} & C_{22} \end{bmatrix}$$

where $C_{11} : k \times k$, $C_{22} : (m-k) \times (m-k)$ are lower triangular and $C_{21} : (m-k) \times k$. We will also need a block partition of inverse of C:

$$C^{-1} = \begin{bmatrix} C^{11} & \mathbf{0} \\ C^{21} & C^{22} \end{bmatrix} \equiv \begin{bmatrix} C^{1} \\ C^{2} \end{bmatrix}$$

where C^{11} and C^{22} are also lower triangular and $C^1 = [C^{11}\mathbf{0}]$, $C^2 = [C^{21}C^{22}]$. Let us rewrite the characteristic function of x^* as:

$$\mathbf{E}(\exp\{it'CC^{-1}x^*\}) = \exp\{it'CC^{-1}\mu - \tfrac{1}{2}t'C\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \lambda\mathbf{I}_{m-k} \end{bmatrix}C't\}$$

by denoting $\tilde{t}' = t'C$ we get:

$$\mathbf{E}(\exp\{i\tilde{t}'C^{-1}x^*\}) = \exp\{i\tilde{t}'C^{-1}\mu - \tfrac{1}{2}\tilde{t}'\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \lambda\mathbf{I}_{m-k} \end{bmatrix}\tilde{t}\}$$

partitioning $\tilde{t}' = [\tilde{t}_1' \vdots \tilde{t}_2']$, $\tilde{t}_1' : (1 \times k)$; $\tilde{t}_2' : 1 \times (m-k)$:

$$\begin{aligned} \mathbf{E}(\exp\{i\tilde{t}'C^{-1}x^*\}) &= \mathbf{E}(\exp\{i\tilde{t}_1'C^1x^*\}\exp\{i\tilde{t}_2'C^2x^*\}) = \\ &= \exp\{i\tilde{t}_1'C^1\mu + i\tilde{t}_2'C^2\mu - \tfrac{1}{2}\tilde{t}_1'\tilde{t}_1 - \tfrac{1}{2}\lambda\tilde{t}_2'\tilde{t}_2\} \end{aligned}$$

we conclude C^1x^* and C^2x^* are independently distributed with $C^1x^* \sim N(C^1\mu, \mathbf{I}_k)$ and $C^2x^* \sim N(C^2\mu, \lambda\mathbf{I}_{m-k})$. Since $C^1 = [C^{11}\mathbf{0}]$, $C^{11}x_1^* \equiv (C_{11})^{-1}x_1^* \sim N((C_{11})^{-1}\mu_1, \mathbf{I}_k)$ thus $x_1^* \sim N(\mu_1, C_{11}C_{11}')$ where we partitioned $x = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix}$ $x_1^* : k \times 1$, $x_2^* : (m-k) \times 1$ and $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ conformably. Note that the marginal distribution of x_1^* does not depend on the parameter λ, hence whatever λ might be we still have $x_1^* \sim N(\mu_1, C_{11}C_{11}')$. In particular if $\lambda = 0$ we have $x_1^* \equiv x_1 \sim N(\mu_1, C_{11}C_{11}')$. Thus even if covariance Σ is singular and $\mathrm{rank}(\Sigma) = k$, the marginal pdf of x_1 is $N(\mu_1, \Sigma_{11})$ where Σ_{11} is an upper-left (principal) $k \times k$ submatrix of Σ. Thus standard result concerning the marginal distribution of a subvector x_1 when x has nonsingular normal distribution is still valid in the singular case.

Our next goal is to establish the conditional distribution of x_2 (given x_1). Since $C^2x^* \sim N(C^2\mu, \lambda\mathbf{I}_{m-k})$ we may indifferently write:

$$C^{21}x_1^* + C^{22}x_2^* \sim N(C^{21}\mu_1 + C^{22}\mu_2, \lambda\mathbf{I}_{m-k})$$

hence:

$$C_{22}C^{21}x_1^* + x_2^* \sim N(C_{22}C^{21}\mu_1 + \mu_2, \lambda C_{22}C_{22}')$$

and:

$x_2^* \sim N(\mu_2 - C_{22}C^{21}(x_1^* - \mu_1), \lambda C_{22}C_{22}')$
if we partition

$$\Sigma^* = \begin{bmatrix} C_{11} & \mathbf{0} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \lambda \mathbf{I}_{m-k} \end{bmatrix} \begin{bmatrix} C_{11} & \mathbf{0} \\ C_{21} & C_{22} \end{bmatrix}' = \begin{bmatrix} C_{11}C_{11}' & C_{11}C_{21}' \\ C_{21}C_{11}' & C_{21}C_{21}' + \lambda C_{22}C_{22}' \end{bmatrix} \equiv \begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{bmatrix}$$

we get:

$$\lambda C_{22}C_{22}' = \Sigma_{22}^* - C_{21}C_{21}' = \Sigma_{22}^* - C_{21}C_{11}'(C_{11}C_{11}')^{-1}C_{11}C_{21}' \equiv \Sigma_{22}^* - \Sigma_{21}^*\Sigma_{11}^{*-1}\Sigma_{12}^*$$

and from standard results on partitioned inverse we have:

$$C_{22}C^{21} = -C_{22}C_{22}^{-1}C_{21}C_{11}^{-1} = -C_{21}C_{11}^{-1} = -C_{21}C_{11}'(C_{11}C_{11}')^{-1} \equiv -\Sigma_{21}^*\Sigma_{11}^{*-1}$$

In other words:

$$x_2^* \sim N(\mu_2 + \Sigma_{21}^*\Sigma_{11}^{*-1}(x_1^* - \mu_1), \Sigma_{22}^* - \Sigma_{21}^*\Sigma_{11}^{*-1}\Sigma_{12}^*)$$

Bearing in mind the decomposition of singular covariance matrix (2), and letting $\lambda \to 0$ we get $\lambda C_{22}C_{22}' \to \mathbf{0}$, $\Sigma_{22}^* \to C_{21}C_{21}' = \Sigma_{22}$, $\Sigma_{21}^*\Sigma_{11}^{*-1}\Sigma_{12}^* \equiv \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ hence $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \to \mathbf{0}$. Therefore we may state the limiting singularity ($\lambda \to 0$) in which case, pdf of x_2 given x_1 is increasingly concentrated around $\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$. On the other hand, if $\lambda = 0$ we obtain $x_2^* \equiv x_2$, $\Sigma^* \equiv \Sigma$, $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \mathbf{0}$, and conclude that x_2 given x_1 has degenerated distribution i.e. $x_2 = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$ with probability 1. It should not be a surprise since from the fact that $\text{rank}(\Sigma) = k$ and Σ_{11} is nonsingular it follows $\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \mathbf{0}$ – see e.g. Harville (1997) p. 110.

APPENDIX 2:

In this appendix we shall prove the proposition 1. To this end we write the general formula for $m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$ using the Lebesgue–Stieltjes integral:

$$m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) =$$
$$= \int L(\mathbf{Y} \mid \Omega_{k11}, \Omega_{k21.1}, B_k, B_{m-k}, \Omega_{k22.1} = \mathbf{0})(d\Pi_0(\Omega_{k21.1}, \Omega_{k11}, B_{m-k}, B_k \mid \Omega_{k22.1} = \mathbf{0})) =$$
$$= \int L(\mathbf{Y}_k \mid B_k, \Omega_{k11}) \times$$
$$\times \left\{ \int L(\mathbf{Y}_{m-k} \mid \mathbf{Y}_k, B_k, B_{m-k}, \Omega_{k22.1} = \mathbf{0}, \Omega_{k21.1})(d\Pi_0(B_{m-k}, \Omega_{k21.1} \mid \Omega_{k11}, B_k, \Omega_{k22.1} = \mathbf{0})) \right\} (d\Pi_0(\Omega_{k11}, B_k))$$

where $\Pi_0(\cdot)$ denotes prior probability measure (distribution function) and we assumed $\Pi_0(\Omega_{k11}, B_k \mid \Omega_{k22.1} = \mathbf{0}) = \Pi_0(\Omega_{k11}, B_k)$ (but this is of no special importance). Since the hypothesis $\Omega_{k22.1} = \mathbf{0}$ implies $Y_{m-k}' - B_{m-k}X' - \Omega_{k21.1}(Y_k' - B_kX') = \mathbf{0}$, the prior $\Pi_0(B_{m-k}, \Omega_{k21.1} \mid \Omega_{k11}, B_k, \Omega_{k22.1} = \mathbf{0})$ must be singular with respect to Lebesgue measure for $B_{m-k}, \Omega_{k21.1}$. It means that in order to preserve the coherence this prior must be concentrated along the hyperplane $Y_{m-k}' - B_{m-k}X' - \Omega_{k21.1}(Y_k' - B_kX') = \mathbf{0}$. According to lemma 1 and (13) we have:

$$L(\mathbf{Y}_{m-k} \mid \mathbf{Y}_k, B_k, B_{m-k}, \Omega_{k22.1} = \mathbf{0}, \Omega_{k21.1}) \equiv \mathbf{1}_{Y'_{m-k} = B_{m-k}X' + \Omega_{k21.1}(Y'_k - B_k X')}(\mathbf{Y}_{m-k})$$

hence:

$$m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) =$$
$$= \int L(\mathbf{Y}_k \mid B_k, \Omega_{k11}) \times$$
$$\times \left\{ \int \mathbf{1}_{Y'_{m-k} = B_{m-k}X' + \Omega_{k21.1}(Y'_k - B_k X')}(\mathbf{Y}_{m-k})(d\Pi_0(B_{m-k}, \Omega_{k21.1} \mid \Omega_{k11}, B_k, \Omega_{k22.1} = \mathbf{0})) \right\} (d\Pi_0(\Omega_{k11}, B_k)) =$$
$$= \int L(\mathbf{Y}_k \mid B_k, \Omega_{k11}) \times$$
$$\times \left\{ \int_{Y'_{m-k} - B_{m-k}X' - \Omega_{k21.1}(Y'_k - B_k X') = \mathbf{0}} (d\Pi_0(B_{m-k}, \Omega_{k21.1} \mid \Omega_{k11}, B_k, \Omega_{k22.1} = \mathbf{0})) \right\} (d\Pi_0(\Omega_{k11}, B_k))$$

Let us write the constraint $Y'_{m-k} - B_{m-k}X' - \Omega_{k21.1}(Y'_k - B_k X') = \mathbf{0}$ as:

$$\mathrm{F}\, Z = Y_{m-k} \qquad \text{(A2.1)}$$

where $Z = \begin{bmatrix} B'_{m-k} \\ \Omega'_{k21.1} \end{bmatrix}$ is $(k+l)\times(m-k)$, $\mathrm{F} = [X \vdots (Y_k - XB'_k)]$ is $T \times (k+l)$. We may perceive $\mathrm{F}\, Z = Y_{m-k}$ as a linear system of equations in Z. Consider the case[9] in which $T > k + l$. As long as Y_k is not identically zero we have $\mathrm{rank}(\mathrm{F}) = k + l$. Since by the null hypothesis (i.e. covariance singularity) the above linear system (in Z) must be consistent (i.e. to have at least one solution) our goal is to conveniently parameterize this total set of solutions. As is known any solution to (A2.1) is a sum of any particular solution to (A2.1) and a solution to homogenous system $\mathrm{F}\, Z = \mathbf{0}$. However since F is full column rank (since $T > k + l$) the only solution to $\mathrm{F}\, Z = \mathbf{0}$ is the trivial one i.e. $Z = \mathbf{0}$. It follows that all solutions to (A2.1) are of the form:

$$Z = \mathrm{F}^{-} Y_{m-k} \qquad \text{(A2.2)}$$

where F^{-} is any generalized inverse of F. On the other hand we know that having any particular inverse of F, say F^{*}, all generalized inverses of F may be generated as (see e.g. Rao (1973) p. 25):

$$\mathrm{F}^{-} = \mathrm{F}^{*} + U - \mathrm{F}^{*}\, \mathrm{F}\, U\, \mathrm{F}\, \mathrm{F}^{*} \qquad \text{(A2.3)}$$

where $U : (k+l) \times T$ is arbitrary. In our case, since $\mathrm{rank}(\mathrm{F}) = k + l$ we may take $\mathrm{F}^{*} = (\mathrm{F}'\,\mathrm{F})^{-1}\,\mathrm{F}'$ (i.e. the left inverse of F) and (A2.3) reads:

$$\mathrm{F}^{-} = (\mathrm{F}'\,\mathrm{F})^{-1}\,\mathrm{F}' + U(\mathbf{I}_T - \mathrm{F}(\mathrm{F}'\,\mathrm{F})^{-1}\,\mathrm{F}') \qquad \text{(A2.4)}$$

and inserting (A2.4) into (A2.2) we obtain:

$$Z = (\mathrm{F}'\,\mathrm{F})^{-1}\,\mathrm{F}'\,Y_{m-k} + U(\mathbf{I}_T - \mathrm{F}(\mathrm{F}'\,\mathrm{F})^{-1}\,\mathrm{F}')Y_{m-k} \qquad \text{(A2.5)}$$

Accordingly, since Z are our coefficients we parameterized Z in terms of the arbitrary matrix U. As U runs over $\mathbb{R}^{(k+l)\times T}$, Z runs over all solutions to (A2.1).

[9] The case $T \le k + l$ will be omitted here because of its small importance from practical point of view. However it does not change the ultimate result but only the way we reach (the same) conclusion.

However, since under the null hypothesis $\Omega_{k22.1}=\mathbf{0}$ the system (A2.1) must be consistent we know that Y_{m-k} lies in the column space of F. This means that $\mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}'Y_{m-k}=Y_{m-k}$. Therefore utilizing this fact in (A2.5) we get:

$$Z=(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}'Y_{m-k}+U(\mathbf{I}_T-\mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}')Y_{m-k}=(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}'Y_{m-k}$$

thus the only solution to the system (A2.1) is $Z=(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}'Y_{m-k}$. Consequently the region of integration in $m(\mathbf{Y}\mid\Omega_{k22.1}=\mathbf{0})$ becomes degenerated as:

$$\begin{bmatrix}B'_{m-k}\\ \Omega'_{k21.1}\end{bmatrix}=\begin{bmatrix}\tilde{B}'_{m-k}\\ \tilde{\Omega}'_{k21.1}\end{bmatrix}\equiv(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}'Y_{m-k}$$

thus $\Pi_0(B_{m-k}=\tilde{B}_{m-k},\Omega_{k21.1}=\tilde{\Omega}_{k21.1}\mid\Omega_{k11},B_k,\Omega_{k22.1}=\mathbf{0})=1$, and:

$$\int_{Y'_{m-k}-B_{m-k}X'-\Omega_{k21.1}(Y'_k-B_kX')=\mathbf{0}}(d\Pi_0(B_{m-k},\Omega_{k21.1}\mid\Omega_{k11},B_k,\Omega_{k22.1}=\mathbf{0}))=$$

$$=\int\mathbf{1}_{Y'_{m-k}=B_{m-k}X'+\Omega_{k21.1}(Y'_k-B_kX')}(\mathbf{Y}_{m-k})(d\Pi_0(B_{m-k}=\tilde{B}_{m-k},\Omega_{k21.1}=\tilde{\Omega}_{k21.1}\mid\Omega_{k11},B_k,\Omega_{k22.1}=\mathbf{0}))=$$

$$=\mathbf{1}_{Y'_{m-k}=\tilde{B}_{m-k}X'+\tilde{\Omega}_{k21.1}(Y'_k-B_kX')}(\mathbf{Y}_{m-k})=\mathbf{1}_{Y'_{m-k}[\mathbf{I}_T-\mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}']=\mathbf{0}}(\mathbf{Y}_{m-k})$$

As a result:

$$m(\mathbf{Y}\mid\Omega_{k22.1}=\mathbf{0})=\int\mathbf{1}_{Y'_{m-k}[\mathbf{I}_T-\mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}']=\mathbf{0}}(\mathbf{Y}_{m-k})\times L(\mathbf{Y}_k\mid B_k,\Omega_{k11})(d\Pi_0(\Omega_{k11},B_k))=$$

$$=\int_{Y'_{m-k}[\mathbf{I}_T-\mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}']=\mathbf{0}}L(\mathbf{Y}_k\mid B_k,\Omega_{k11})p(\Omega_{k11},B_k)(d\Omega_{k11})(dB_k)$$

(in the last equality we assumed the absolute continuity of $\Pi_0(\Omega_{k11},B_k)$).

APPENDIX 3:

Our goal is to show that under the standard conjugate Normal–Wishart prior given in section VI, *BF* testing the null hypothesis $\Omega_{k22.1}=\mathbf{0}$ (if it is true) converges in probability to 0 (thus finds more evidence for the false alternative hypothesis). By proposition 1:

$$m(\mathbf{Y}\mid\Omega_{k22.1}=\mathbf{0})=\int_{Y'_{m-k}[\mathbf{I}_T-\mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}']=\mathbf{0}}L(\mathbf{Y}_k\mid B_k,\Omega_{k11})p(\Omega_{k11},B_k)(d\Omega_{k11})(dB_k)$$

As we emphasized in the text, the integration support constraint of the above integral introduces an enormous complication. In particular even adopting the standard Normal–Wishart prior it seems extremely difficult to analytically evaluate $m(\mathbf{Y}\mid\Omega_{k22.1}=\mathbf{0})$. However as our goal is to show an inconsistency of the underlying *BF*, we will be satisfied with the following remark. Since $\{B_k:Y'_{m-k}[\mathbf{I}_T-\mathrm{F}(\mathrm{F}'\mathrm{F})^{-1}\mathrm{F}']=\mathbf{0}\}\subset\{B_k:B_k=\mathbb{R}^{k\times l}\}$ we conclude:

$$m(\mathbf{Y}\mid\Omega_{k22.1}=\mathbf{0})\le\int L(\mathbf{Y}_k\mid B_k,\Omega_{k11})p(\Omega_{k11},B_k)(d\Omega_{k11})(dB_k)$$

where in the upper bound, the region of integration is not constrained at all. Let us denote the latter as $m_*(\mathbf{Y}\mid\Omega_{k22.1}=\mathbf{0})$. Inserting $L(\mathbf{Y}_k\mid B_k,\Omega_{k11})$ as given in (11) and

$p(\Omega_{k11}, B_k) = p(B_k \mid \Omega_{k11}) p(\Omega_{k11})$, where $p(B_k \mid \Omega_{k11})$, $p(\Omega_{k11})$ is given by (20) and (24) respectively, into $m_*(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0})$ we can easily integrate out Ω_{k11} and B_k:

$$m_*(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) = \pi^{-\frac{1}{2}kT} \Gamma_k(\tfrac{T+\nu-m+k}{2})[\Gamma_k(\tfrac{\nu-m+k}{2})]^{-1} \left|\overline{Q}_{k11}\right|^{\frac{1}{2}(\nu-m+k)} \left|\overline{\mathrm{Z}}\right|^{-\frac{1}{2}k} \left|X'X + \overline{\mathrm{Z}}^{-1}\right|^{-\frac{1}{2}k} \times$$
$$\times \left|\overline{Q}_{k11} + (\hat{B}_k - \overline{B}_k)((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}(\hat{B}_k - \overline{B}_k)' + (Y_k' - \hat{B}_k X')(Y_k' - \hat{B}_k X')'\right|^{-\frac{1}{2}(T+\nu)+\frac{1}{2}(m-k)}$$

Next, the *PPD* under the alternative $\text{rank}(\Omega) = m$ i.e. $m(\mathbf{Y} \mid \text{rank}(\Omega) = m)$ has already been derived in lemma 2; 2), hence it is a matter of simple algebraic manipulation that:

$$BF = m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) / m(\mathbf{Y} \mid \text{rank}(\Omega) = m) \le m_*(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) / m(\mathbf{Y} \mid \text{rank}(\Omega) = m) =$$
$$= \frac{\pi^{\frac{1}{2}(m-k)T} \left|\overline{Q}_{k11}\right|^{\frac{1}{2}(k-m)} \left|\overline{\mathrm{Z}}\right|^{\frac{1}{2}(m-k)} \left|X'X + \overline{\mathrm{Z}}^{-1}\right|^{\frac{1}{2}(m-k)}}{[\Gamma_{m-k}(\frac{\nu}{2})]^{-1} \Gamma_{m-k}(\frac{T+\nu}{2}) \left|\overline{Q}_{k22.1}\right|^{\frac{1}{2}\nu} \left|\mathrm{U}_k + \overline{Q}_{k22.1}\right|^{-\frac{1}{2}(T+\nu)}} \times$$
$$\times \left|\overline{Q}_{k11} + (\hat{B}_k - \overline{B}_k)((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}(\hat{B}_k - \overline{B}_k)' + (Y_k' - \hat{B}_k X')(Y_k' - \hat{B}_k X')'\right|^{\frac{1}{2}(m-k)} =$$
$$= \frac{\pi^{\frac{1}{2}(m-k)T} \left|\overline{Q}_{k11}\right|^{\frac{1}{2}(k-m)} \left|\overline{\mathrm{Z}}\right|^{\frac{1}{2}(m-k)} \left|X'X + \overline{\mathrm{Z}}^{-1}\right|^{\frac{1}{2}(m-k)} T^{\frac{1}{2}k(m-k)}}{[\Gamma_{m-k}(\frac{\nu}{2})]^{-1} \Gamma_{m-k}(\frac{T+\nu}{2}) \left|\overline{Q}_{k22.1}\right|^{\frac{1}{2}\nu} \left|\mathrm{U}_k + \overline{Q}_{k22.1}\right|^{-\frac{1}{2}(T+\nu)}} \times$$
$$\times \left|\tfrac{1}{T}\overline{Q}_{k11} + (\hat{B}_k - \overline{B}_k)\tfrac{1}{T}((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}(\hat{B}_k - \overline{B}_k)' + \tfrac{1}{T}(Y_k' - \hat{B}_k X')(Y_k' - \hat{B}_k X')'\right|^{\frac{1}{2}(m-k)} =$$
$$= \frac{\pi^{\frac{1}{2}(m-k)T} \left|\overline{Q}_{k11}\right|^{\frac{1}{2}(k-m)} \left|\overline{\mathrm{Z}}\right|^{\frac{1}{2}(m-k)} \left|X'X + \overline{\mathrm{Z}}^{-1}\right|^{\frac{1}{2}(m-k)} T^{\frac{1}{2}k(m-k)}}{[\Gamma_{m-k}(\frac{\nu}{2})]^{-1} \Gamma_{m-k}(\frac{T+\nu}{2}) \left|\overline{Q}_{k22.1}\right|^{\frac{1}{2}\nu} \left|\mathrm{U}_k + \overline{Q}_{k22.1}\right|^{-\frac{1}{2}(T+\nu)}} \times \left|\tfrac{1}{T}(Y_k' - \hat{B}_k X')(Y_k' - \hat{B}_k X')'\right|^{\frac{1}{2}(m-k)}$$

where the last equality follows since $(\hat{B}_k - \overline{B}_k)\frac{1}{T}((X'X)^{-1} + \overline{\mathrm{Z}})^{-1}(\hat{B}_k - \overline{B}_k)' = O_{\mathbf{p}}(\frac{1}{T})$. To better see this, by assumption A1, for large T, we have:

$$\tfrac{1}{T}((X'X)^{-1} + \overline{\mathrm{Z}})^{-1} = \tfrac{1}{T}\overline{\mathrm{Z}}^{-1} - \tfrac{1}{T}\overline{\mathrm{Z}}^{-1}(X'X + \overline{\mathrm{Z}}^{-1})^{-1}\overline{\mathrm{Z}}^{-1} \sim \tfrac{1}{T}\overline{\mathrm{Z}}^{-1} - \tfrac{1}{T}\overline{\mathrm{Z}}^{-1}(B_T \mathbf{Q} B_T + \overline{\mathrm{Z}}^{-1})^{-1}\overline{\mathrm{Z}}^{-1}$$
$$= \tfrac{1}{T}\overline{\mathrm{Z}}^{-1} - \tfrac{1}{T}\overline{\mathrm{Z}}^{-1} B_T^{-1}(\mathbf{Q} + B_T^{-1}\overline{\mathrm{Z}}^{-1} B_T^{-1})^{-1} B_T^{-1}\overline{\mathrm{Z}}^{-1} \to \tfrac{1}{T}\overline{\mathrm{Z}}^{-1} - \tfrac{1}{T}\overline{\mathrm{Z}}^{-1} B_T^{-1}\mathbf{Q}^{-1} B_T^{-1}\overline{\mathrm{Z}}^{-1} \to \mathbf{0}$$

as $B_T = diag(T^{\alpha_1}, T^{\alpha_2}, \ldots, T^{\alpha_l})$ and $\alpha_i \ge \frac{1}{2}$ (see assumption A1).

Note also that for large T, by assumption A1 we have:

$$\left|X'X + \overline{\mathrm{Z}}^{-1}\right| \sim \left|B_T \mathbf{Q} B_T + \overline{\mathrm{Z}}^{-1}\right| = \left|B_T\right|^2 \left|\mathbf{Q} + B_T^{-1}\overline{\mathrm{Z}}^{-1} B_T^{-1}\right| = \left|B_T\right|^2 \left|\mathbf{Q}\right| \le T^{2l\alpha_{\max}} \left|\mathbf{Q}\right|$$

where $\alpha_{\max} = \max_i\{\alpha_i\}$ for $i = 1, 2, \ldots, l$. Recall that by assumption A1, α_i are positive (If the data are stationary then $\alpha_{\max} = \alpha_1 = \ldots = \alpha_l = \frac{1}{2}$). Anyway, for large T, $\left|X'X + \overline{\mathrm{Z}}^{-1}\right|^{\frac{1}{2}(m-k)} \le T^{l\alpha_{\max}(m-k)} \left|\mathbf{Q}\right|^{\frac{1}{2}(m-k)}$. Furthermore, let us denote (for obvious reasons) $\hat{\Omega}_{k11} = \frac{1}{T}(Y_k' - \hat{B}_k X')(Y_k' - \hat{B}_k X')'$. Then:

$$m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) / m(\mathbf{Y} \mid \text{rank}(\Omega) = m) \le$$
$$\le \frac{\pi^{\frac{1}{2}(m-k)T} \left|\overline{Q}_{k11}\right|^{\frac{1}{2}(k-m)} \left|\overline{\mathrm{Z}}\right|^{\frac{1}{2}(m-k)} \left|\mathbf{Q}\right|^{\frac{1}{2}(m-k)} T^{\frac{1}{2}k(m-k)+l\alpha_{\max}(m-k)} \left|\hat{\Omega}_{k11}\right|^{\frac{1}{2}(m-k)}}{[\Gamma_{m-k}(\frac{\nu}{2})]^{-1} \Gamma_{m-k}(\frac{T+\nu}{2}) \left|\overline{Q}_{k22.1}\right|^{\frac{1}{2}\nu} \left|\mathrm{U}_k + \overline{Q}_{k22.1}\right|^{-\frac{1}{2}(T+\nu)}}$$

$$= \frac{\pi^{\frac{1}{2}(m-k)T}\left|\bar{Q}_{k11}\right|^{\frac{1}{2}(k-m)}\left|\bar{\mathrm{Z}}\right|^{\frac{1}{2}(m-k)}\left|\mathbf{Q}\right|^{\frac{1}{2}(m-k)}T^{\frac{1}{2}k(m-k)+l\alpha_{\max}(m-k)}\left|\hat{\Omega}_{k11}\right|^{\frac{1}{2}(m-k)}}{[\Gamma_{m-k}(\frac{\nu}{2})]^{-1}\Gamma_{m-k}(\frac{T+\nu}{2})\left|\bar{Q}_{k22.1}\right|^{\frac{1}{2}\nu}T^{-\frac{1}{2}(T+\nu)(m-k)}\left|\frac{1}{T}\mathrm{U}_k+\frac{1}{T}\bar{Q}_{k22.1}\right|^{-\frac{1}{2}(T+\nu)}}$$

Now, we derive an asymptotic bound for all terms that depend on T (except $\left|\frac{1}{T}\mathrm{U}_k+\frac{1}{T}\bar{Q}_{k22.1}\right|^{-\frac{1}{2}(T+\nu)}$):

$$\pi^{\frac{1}{2}(m-k)T}T^{(m-k)(\frac{1}{2}k+l\alpha_{\max}+\frac{1}{2}T+\frac{1}{2}\nu)}\,/\,\Gamma_{m-k}(\tfrac{T+\nu}{2}) = \pi^{-\frac{1}{4}(m-k)(m-k-1)+\frac{1}{2}(m-k)T}\prod\nolimits_{i=1}^{m-k}T^{(\frac{1}{2}k+l\alpha_{\max}+\frac{1}{2}T+\frac{1}{2}\nu)}/\Gamma(\tfrac{T+\nu-i+1}{2})$$
$$\sim \pi^{-\frac{1}{4}(m-k)(m-k-1)+\frac{1}{2}(m-k)T}(2\pi)^{-\frac{1}{2}(m-k)}\prod\nolimits_{i=1}^{m-k}T^{(\frac{1}{2}k+l\alpha_{\max}+\frac{1}{2}T+\frac{1}{2}\nu)}\exp\{\tfrac{T+\nu-i+1}{2}\}(\tfrac{T+\nu-i+1}{2})^{-\frac{1}{2}(T+\nu-i)}$$
$$< \pi^{-\frac{1}{4}(m-k)(m-k-1)}(2\pi)^{-\frac{1}{2}(m-k)}e^{(m-k)T+\frac{1}{2}(\nu-1)(m-k)+T}$$

hence for large T:

$$m(\mathbf{Y}\mid\Omega_{k22.1}=\mathbf{0})\,/\,m(\mathbf{Y}\mid\mathrm{rank}(\Omega)=m)\le C\times\left|\mathbf{Q}\right|^{\frac{1}{2}(m-k)}e^{T(m-k+1)}\left|\tfrac{1}{T}\mathrm{U}_k+\tfrac{1}{T}\bar{Q}_{k22.1}\right|^{\frac{1}{2}(T+\nu)}$$

where C denotes all terms that are constant (and bounded). Since by assumption A1 $\mathbf{Q}=O_{\mathbf{p}}(1)$, the term $\left|\mathbf{Q}\right|^{\frac{1}{2}(m-k)}$ may be also absorbed into "C" as it will be dominated by $e^{T(m-k+1)}\left|\frac{1}{T}\mathrm{U}_k+\frac{1}{T}\bar{Q}_{k22.1}\right|^{\frac{1}{2}(T+\nu)}$. The key question is how $\frac{1}{T}\mathrm{U}_k$ behaves for large T. To this end let us write:

$$\tfrac{1}{T}\mathrm{U}_k+\tfrac{1}{T}\bar{Q}_{k22.1}=\hat{\Omega}_{k22.1}+\tfrac{1}{T}E'_{m-k}E_k(\tfrac{1}{T}E'_kE_k)^{-1}\tfrac{1}{T}E'_kE_{m-k}+\underbrace{\Omega^*_{k21.1}D_k\tfrac{1}{T}((X'X)^{-1}+\bar{\mathrm{Z}})^{-1}D'_k\Omega'^*_{k21.1}}_{O_{\mathbf{p}}(\frac{1}{T})}+$$
$$-(\tfrac{1}{T}E'_kE_{m-k}+\underbrace{D_k\tfrac{1}{T}((X'X)^{-1}+\bar{\mathrm{Z}})^{-1}D'_k\Omega'^*_{k21.1}}_{O_{\mathbf{p}}(\frac{1}{T})}+\tfrac{1}{T}\bar{Q}'_{k21})'\times$$
$$\times(\tfrac{1}{T}E'_kE_k+\underbrace{D_k\tfrac{1}{T}((X'X)^{-1}+\bar{\mathrm{Z}})^{-1}D'_k}_{O_{\mathbf{p}}(\frac{1}{T})}+\tfrac{1}{T}\bar{Q}_{k11})^{-1}\times$$
$$\times(\tfrac{1}{T}E'_kE_{m-k}+\underbrace{D_k\tfrac{1}{T}((X'X)^{-1}+\bar{\mathrm{Z}})^{-1}D'_k\Omega'^*_{k21.1}}_{O_{\mathbf{p}}(\frac{1}{T})}+\tfrac{1}{T}\bar{Q}'_{k21})+$$
$$+\underbrace{(D_{m-k}-\Omega^*_{k21.1}D_k)\tfrac{1}{T}((X'X)^{-1}+\bar{\mathrm{Z}})^{-1}(D_{m-k}-\Omega^*_{k21.1}D_k)'}_{O_{\mathbf{p}}(\frac{1}{T})}+\tfrac{1}{T}\bar{Q}_{k21}\bar{Q}^{-1}_{k11}\bar{Q}'_{k21}+\tfrac{1}{T}\bar{Q}_{k22.1}\to$$
$$\to\tfrac{1}{T}E'_{m-k}E_k(\tfrac{1}{T}E'_kE_k)^{-1}\tfrac{1}{T}E'_kE_{m-k}-(\tfrac{1}{T}E'_kE_{m-k})'(\tfrac{1}{T}E'_kE_k)^{-1}(\tfrac{1}{T}E'_kE_{m-k})+O_{\mathbf{p}}(\tfrac{1}{T})+\hat{\Omega}_{k22.1}=$$
$$=O_{\mathbf{p}}(\tfrac{1}{T})+\hat{\Omega}_{k22.1}$$

(for a definition of all terms comprising U_k see lemma 2). Note that we exploited the fact that under the null hypothesis $\Omega_{k22.1}=\mathbf{0}$, Ω_{k11} is nonsingular and since $\frac{1}{T}E'_kE_k$ converges in probability to Ω_{k11}, the inverse exists: $(\frac{1}{T}E'_kE_k)^{+}\equiv(\frac{1}{T}E'_kE_k)^{-1}$. But as we pointed out in section VII, if the null hypothesis is true i.e. $\Omega_{k22.1}=\mathbf{0}$, $\hat{\Omega}_{k22.1}=\frac{1}{T}\times o_{\mathbf{p}}(1)\times o_{\mathbf{p}}(1)=o_{\mathbf{p}}(\frac{1}{T})$ which converges to 0 very fast, hence $\frac{1}{T}\mathrm{U}_k+\frac{1}{T}\bar{Q}_{k22.1}=O_{\mathbf{p}}(\frac{1}{T})+o_{\mathbf{p}}(\frac{1}{T})=\frac{1}{T}O_{\mathbf{p}}(1)=o_{\mathbf{p}}(1)$. Next, if $\frac{1}{T}\mathrm{U}_k+\frac{1}{T}\bar{Q}_{k22.1}=o_{\mathbf{p}}(1)$ then all eigenvalues of $\frac{1}{T}\mathrm{U}_k+\frac{1}{T}\bar{Q}_{k22.1}$ are also $o_{\mathbf{p}}(1)$. Let us denote them as λ_i ($i=1,\ldots,m-k$) and we have:

$$\left|\tfrac{1}{T}\mathrm{U}_k+\tfrac{1}{T}\bar{Q}_{k22.1}\right|^{\frac{1}{2}(T+\nu)}=\prod\nolimits_{i=1}^{m-k}\lambda_i^{\frac{1}{2}(T+\nu)}\le\lambda_{\max}^{\frac{1}{2}(T+\nu)(m-k)}$$

where $\lambda_{\max} = \max_i \lambda_i$ and of course $\lambda_{\max} = o_{\mathbf{p}}(1)$. Lastly, we have:

$$0 < m(\mathbf{Y} \mid \Omega_{k22.1} = \mathbf{0}) / m(\mathbf{Y} \mid \text{rank}(\Omega) = m) \leq C \times |\mathbf{Q}|^{\frac{1}{2}(m-k)} e^{T(m-k+1)} \lambda_{\max}^{\frac{1}{2}(T+\nu)(m-k)} =$$

$$= C \times |\mathbf{Q}|^{\frac{1}{2}(m-k)} e^{T(m-k+1)+\frac{1}{2}(T+\nu)(m-k)\ln(\lambda_{\max})} = C \times O_{\mathbf{p}}(1) \times e^{T[m-k+1+\frac{1}{2}(m-k)\ln(\lambda_{\max})]+\frac{1}{2}\nu(m-k)\ln(\lambda_{\max})}$$

as $\lambda_{\max}$ converges in probability to 0, $\ln(\lambda_{\max})$ tends to minus infinity, thus $e^{T[m-k+1+\frac{1}{2}(m-k)\ln(\lambda_{\max})]+\frac{1}{2}\nu(m-k)\ln(\lambda_{\max})}$ becomes extremely small. This proves the proposition.

APPENDIX 4:

We have to analyze the asymptotic behavior ($T \to \infty$) of the following expression:

$$BF_k^* \equiv m(\mathbf{Y} \mid \Omega_{k22.1} \approx \mathbf{0}) / m(\mathbf{Y} \mid \text{rank}(\Omega) = m) \,|_{\bar{Q}^*_{k22.1} = \bar{Q}_{k22.1}} =$$

$$= \lim_{\nu_* \to \infty} \Gamma_{m-k}(\tfrac{T+\nu_*}{2})[\Gamma_{m-k}(\tfrac{\nu_*}{2})]^{-1} \Gamma_{m-k}(\tfrac{\nu}{2})[\Gamma_{m-k}(\tfrac{T+\nu}{2})]^{-1} \left|\bar{Q}_{k22.1}\right|^{\frac{1}{2}(\nu_*-\nu)} \left|\mathrm{U}_k + \bar{Q}_{k22.1}\right|^{-\frac{1}{2}(\nu_*-\nu)}$$

For simplicity we omit the limit operator from the above, but it is present implicitly as we let $\nu_* = \alpha T$ which diverges to infinity with the number of observations. First we prove a) i.e. we consider the case $\Omega_{k22.1} = \mathbf{0}$. If $\Omega_{k22.1} = \mathbf{0}$, then from appendix 3 we know that $\frac{1}{T}\mathrm{U}_k + \frac{1}{T}\bar{Q}_{k22.1} = O_{\mathbf{p}}(\frac{1}{T})$. Let us write the latter as $\frac{1}{T}\mathrm{U}_k + \frac{1}{T}\bar{Q}_{k22.1} = \frac{1}{T}\mathbf{R}_k$ where $\mathbf{R}_k = O_{\mathbf{p}}(1)$ and $\mathbf{R}_k$ is positive definite. First we obtain the following expansion:

$$\ln[\Gamma_{m-k}(\tfrac{T+\nu_*}{2}) / \Gamma_{m-k}(\tfrac{T+\nu}{2})] \equiv \ln[\Gamma_{m-k}(\tfrac{T+\alpha T}{2}) / \Gamma_{m-k}(\tfrac{T+\nu}{2})] \sim$$

$$\sim -\tfrac{\alpha T(m-k)}{2} + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln(\tfrac{T}{2}) + (\tfrac{T(1+\alpha)(m-k)}{2} - \tfrac{(m-k)(m-k+1)}{4})\ln(1+\alpha) \qquad \text{(A4.1)}$$

Hence:

$$\ln BF_k^* = \ln\Gamma_{m-k}(\tfrac{\nu}{2}) - \ln\Gamma_{m-k}(\tfrac{\alpha T}{2}) + (\tfrac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| +$$
$$-\tfrac{\alpha T(m-k)}{2} + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln(\tfrac{T}{2}) + (\tfrac{T(1+\alpha)(m-k)}{2} - \tfrac{(m-k)(m-k+1)}{4})\ln(1+\alpha)$$
$$-\tfrac{(m-k)(\alpha T-\nu)}{2}\ln(T) - \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_k + \tfrac{1}{T}\bar{Q}_{k22.1}\right| =$$
$$= const - \ln\Gamma_{m-k}(\tfrac{\alpha T}{2}) + (\tfrac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| +$$
$$-\tfrac{\alpha T(m-k)}{2} + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln(\tfrac{1}{2}) + \tfrac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) - \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_k + \tfrac{1}{T}\bar{Q}_{k22.1}\right| =$$
$$= const - \ln\Gamma_{m-k}(\tfrac{\alpha T}{2}) + (\tfrac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| +$$
$$-\tfrac{\alpha T(m-k)}{2} + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln(\tfrac{1}{2}) + \tfrac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) - \tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_k\right| + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln T \geq$$
$$\geq const - (m-k)\ln\Gamma(\tfrac{\alpha T}{2}) + (\tfrac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| +$$
$$-\tfrac{\alpha T(m-k)}{2} + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln(\tfrac{1}{2}) + \tfrac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) - \tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_k\right| + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln T$$

where, from now on, the term *const* does not depend on T and is bounded. Expanding $\Gamma(\frac{\alpha T}{2})$ for large T:

$$\ln BF_k^* \geq const - \tfrac{(m-k)(\alpha T-1)}{2}\ln(\tfrac{\alpha T}{2}) + \tfrac{\alpha T(m-k)}{2} + (\tfrac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| +$$
$$-\tfrac{\alpha T(m-k)}{2} + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln(\tfrac{1}{2}) + \tfrac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) - \tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_k\right| + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln T =$$
$$= const - \tfrac{(m-k)(\alpha T-1)}{2}\ln(\tfrac{\alpha T}{2}) + (\tfrac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| +$$
$$+\tfrac{(\alpha T-\nu)(m-k)}{2}\ln(\tfrac{1}{2}) + \tfrac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) - \tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_k\right| + \tfrac{(\alpha T-\nu)(m-k)}{2}\ln T \geq$$

$\geq const - \frac{(m-k)(\alpha T-1)}{2}\ln(\frac{\alpha T}{2}) + (\frac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| +$
$+\frac{(\alpha T-1)(m-k)}{2}\ln(\frac{1}{2}) + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) - \frac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_k\right| + \frac{(\alpha T-\nu)(m-k)}{2}\ln T =$
$= const - \frac{(m-k)(\alpha T-1)}{2}\ln(\alpha T) + (\frac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) - \frac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_k\right| + \frac{(\alpha T-\nu)(m-k)}{2}\ln T$
$= const - \frac{(m-k)(\alpha T-1)}{2}\ln(\alpha) + (\frac{\alpha T-\nu}{2})\ln[\frac{|\bar{Q}_{k22.1}|}{|\mathbf{R}_k|}] + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) - \frac{(\nu-1)(m-k)}{2}\ln T \geq$
$\geq const - \frac{\alpha T(m-k)}{2}\ln(\alpha) + \frac{\alpha T}{2}\ln[\frac{|\bar{Q}_{k22.1}|}{|\mathbf{R}_k|}] + \frac{\alpha T(m-k)}{2}\ln(1+\alpha) - \frac{(\nu-1)(m-k)}{2}\ln T =$
$= const + \frac{\alpha T}{2}\ln[(1+\frac{1}{\alpha})^{m-k}\frac{|\bar{Q}_{k22.1}|}{|\mathbf{R}_k|}] - \frac{(\nu-1)(m-k)}{2}\ln T \geq$
$\geq const + \frac{\alpha T}{2}\ln[(1+\frac{1}{\alpha})\frac{|\bar{Q}_{k22.1}|}{|\mathbf{R}_k|}] - \frac{(\nu-1)(m-k)}{2}\ln T$

But from elementary calculus we know that with $T \to \infty$, $\frac{\alpha T}{2}\ln[(1+\frac{1}{\alpha})\frac{|\bar{Q}_{k22.1}|}{|\mathbf{R}_k|}] - \frac{(\nu-1)(m-k)}{2}\ln T \to \infty$ provided that $(1+\frac{1}{\alpha})\frac{|\bar{Q}_{k22.1}|}{|\mathbf{R}_k|} > 1$, or $\alpha < (\frac{|\mathbf{R}_k|}{|\bar{Q}_{k22.1}|} - 1)^{-1}$. Thus if we set $\nu_* = \alpha T$ where α fulfills the last condition and the covariance matrix is singular of rank k (i.e. $\Omega_{k22.1} = \mathbf{0}$), then as $T \to \infty$, we shall certainly accept the true null hypothesis since $\ln BF_k^* \to \infty$.

Now we prove b). If Ω is nonsingular and bearing in mind the asymptotic result from appendix 3, we have $\frac{1}{T}\mathrm{U}_k + \frac{1}{T}\bar{Q}_{k22.1} = \hat{\Omega}_{k22.1} + O_{\mathbf{p}}(\frac{1}{T}) = \hat{\Omega}_{k22.1} + o_{\mathbf{p}}(1)$. But given assumption A2, $\hat{\Omega}_{k22.1} \xrightarrow{\mathbf{p}} \Omega_{k22.1}$, where $\Omega_{k22.1}$ denotes the (pseudo) true value. We also have:

$\left|\frac{1}{T}\mathrm{U}_k + \frac{1}{T}\bar{Q}_{k22.1}\right|^{-\frac{1}{2}(\nu_*-\nu)} = \left|\frac{1}{T}\mathrm{U}_k + \frac{1}{T}\bar{Q}_{k22.1}\right|^{-\frac{1}{2}(\alpha T-\nu)} \sim \left|\frac{1}{T}\mathrm{U}_k\right|^{-\frac{1}{2}(\alpha T-\nu)}\mathrm{etr}\{-\frac{\alpha}{2}\bar{Q}_{k22.1}(\frac{1}{T}\mathrm{U}_k)^{-1}\}$

but in the case b) $\frac{1}{T}\mathrm{U}_k \xrightarrow{\mathbf{p}} \Omega_{k22.1}$, thus:

$\left|\frac{1}{T}\mathrm{U} + \frac{1}{T}\bar{Q}_{k22.1}\right|^{-\frac{1}{2}(\nu_*-\nu)} \xrightarrow{\mathbf{p}} \left|\Omega_{k22.1}\right|^{-\frac{1}{2}(\alpha T-\nu)}\mathrm{etr}\{-\frac{\alpha}{2}\bar{Q}_{k22.1}\Omega_{k22.1}^{-1}\}$

and utilizing (A4.1) we get:

$\ln BF_k^* = \ln\Gamma_{m-k}(\frac{\nu}{2}) - \ln\Gamma_{m-k}(\frac{\alpha T}{2}) + (\frac{\alpha T-\nu}{2})\ln\left|\bar{Q}_{k22.1}\right| +$
$-\frac{\alpha T(m-k)}{2} + \frac{(\alpha T-\nu)(m-k)}{2}\ln(\frac{T}{2}) + (\frac{T(1+\alpha)(m-k)}{2} - \frac{(m-k)(m-k+1)}{4})\ln(1+\alpha)$
$-\frac{(m-k)(\alpha T-\nu)}{2}\ln(T) - \frac{\alpha T-\nu}{2}\ln\left|\Omega_{k22.1}\right| - \frac{\alpha}{2}\mathrm{tr}\{\bar{Q}_{k22.1}\Omega_{k22.1}^{-1}\} \leq$
$\leq const - \ln\Gamma_{m-k}(\frac{\alpha T}{2}) + (\frac{\alpha T-\nu}{2})\ln[\frac{|\bar{Q}_{k22.1}|}{|\Omega_{k22.1}|}] - \frac{\alpha T(m-k)}{2} + \frac{(\alpha T-\nu)(m-k)}{2}\ln(\frac{1}{2}) + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha)$
$\leq const - (m-k)\ln\Gamma(\frac{\alpha T-m+k+1}{2}) + (\frac{\alpha T-\nu}{2})\ln[\frac{|\bar{Q}_{k22.1}|}{|\Omega_{k22.1}|}] - \frac{\alpha T(m-k)}{2} + \frac{(\alpha T-\nu)(m-k)}{2}\ln(\frac{1}{2}) + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha)$

Using Stirling's approximation for gamma function we obtain:

$\ln BF_k^* \leq const - \frac{(\alpha T-m+k)(m-k)}{2}\ln(\frac{\alpha T-m+k+1}{2}) + \frac{(m-k)(\alpha T-m+k+1)}{2} + (\frac{\alpha T-\nu}{2})\ln[\frac{|\bar{Q}_{k22.1}|}{|\Omega_{k22.1}|}] - \frac{\alpha T(m-k)}{2}$
$+\frac{(\alpha T-\nu)(m-k)}{2}\ln(\frac{1}{2}) + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) =$
$= const - \frac{(\alpha T-m+k)(m-k)}{2}\ln(\frac{\alpha T-m+k+1}{2}) + (\frac{\alpha T-\nu}{2})\ln[\frac{|\bar{Q}_{k22.1}|}{|\Omega_{k22.1}|}] + \frac{(\alpha T-\nu)(m-k)}{2}\ln(\frac{1}{2}) + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha)$
$\leq const - \frac{(\alpha T-\nu)(m-k)}{2}\ln(\frac{\alpha T-\nu}{2}) + (\frac{\alpha T-\nu}{2})\ln[\frac{|\bar{Q}_{k22.1}|}{|\Omega_{k22.1}|}] + \frac{(\alpha T-\nu)(m-k)}{2}\ln(\frac{1}{2}) + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) =$
$= const - \frac{(\alpha T-\nu)(m-k)}{2}\ln[(\alpha T-\nu)(\frac{|\Omega_{k22.1}|}{|\bar{Q}_{k22.1}|})^{\frac{1}{m-k}}] + \frac{T(1+\alpha)(m-k)}{2}\ln(1+\alpha) =$

$= const - \frac{T}{2}\ln[\frac{(\alpha T-\nu)^{\alpha(m-k)}}{(1+\alpha)^{(1+\alpha)(m-k)}}(\frac{|\Omega_{k22.1}|}{|\bar{Q}_{k22.1}|})^{\alpha}] + \frac{\nu(m-k)}{2}\ln[(\alpha T-\nu)(\frac{|\Omega_{k22.1}|}{|\bar{Q}_{k22.1}|})^{\frac{1}{m-k}}]$

Now, from calculus we conclude that the upper bound diverges to minus infinity provided that $\frac{(\alpha T-\nu)^{\alpha(m-k)}}{(1+\alpha)^{(1+\alpha)(m-k)}}(\frac{|\Omega_{k22.1}|}{|\bar{Q}_{k22.1}|})^{\alpha} > 1$. But this is not a stringent condition since the left hand of this inequality converges to infinity as $T \to \infty$ for any fixed $\alpha > 0$. Thus for large T, any $\alpha > 0$ will suffice to make $\ln BF_k^* \to -\infty$ (if Ω is nonsingular).

To prove c) we shall show consistency of BF testing $\text{rank}(\Omega) = i$ vs $\text{rank}(\Omega) = j$ where $i > j$ and $i = 2,...,m-1$ (the null is $\text{rank}(\Omega) = i$ and is true). We obtain:

$$\frac{m(\mathbf{Y} \mid \Omega_{i22.1} \approx \mathbf{0})}{m(\mathbf{Y} \mid \Omega_{j22.1} \approx \mathbf{0})} = \frac{m(\mathbf{Y} \mid \Omega_{i22.1} \approx \mathbf{0})}{m(\mathbf{Y} \mid \text{rank}(\Omega) = m)} \Big/ \frac{m(\mathbf{Y} \mid \Omega_{j22.1} \approx \mathbf{0})}{m(\mathbf{Y} \mid \text{rank}(\Omega) = m)} = \frac{BF_i^*}{BF_j^*} =$$

$$= \frac{\prod_{l=m-i+1}^{m-j}\Gamma(\frac{\alpha T}{2} - \frac{(l-1)}{2}) \times \prod_{l=m-i+1}^{m-j}\Gamma(\frac{T+\nu}{2} - \frac{(l-1)}{2})}{\prod_{l=m-i+1}^{m-j}\Gamma(\frac{\nu}{2} - \frac{(l-1)}{2}) \times \prod_{l=m-i+1}^{m-j}\Gamma(\frac{T+\alpha T}{2} - \frac{(l-1)}{2})} \frac{\left|\bar{Q}_{i22.1}\right|^{\frac{1}{2}(\alpha T-\nu)}\left|\mathrm{U}_i + \bar{Q}_{i22.1}\right|^{-\frac{1}{2}(\alpha T-\nu)}}{\left|\bar{Q}_{j22.1}\right|^{\frac{1}{2}(\alpha T-\nu)}\left|\mathrm{U}_j + \bar{Q}_{j22.1}\right|^{-\frac{1}{2}(\alpha T-\nu)}} \geq$$

$$\geq \frac{\Gamma(\frac{\alpha T}{2} - \frac{(m-j-1)}{2})\Gamma(\frac{T+\nu}{2} - \frac{(m-j-1)}{2})}{\Gamma(\frac{\nu}{2} - \frac{(m-i)}{2})\Gamma(\frac{T+\alpha T}{2} - \frac{(m-i)}{2})} \frac{\left|\bar{Q}_{i22.1}\right|^{\frac{1}{2}(\alpha T-\nu)}\left|\mathrm{U}_i + \bar{Q}_{i22.1}\right|^{-\frac{1}{2}(\alpha T-\nu)}}{\left|\bar{Q}_{j22.1}\right|^{\frac{1}{2}(\alpha T-\nu)}\left|\mathrm{U}_j + \bar{Q}_{j22.1}\right|^{-\frac{1}{2}(\alpha T-\nu)}}$$

Approximating gamma function for large T we have:

$$\ln[\Gamma(\tfrac{T+\nu}{2} - \tfrac{(m-j-1)}{2}) / \Gamma(\tfrac{T+\alpha T}{2} - \tfrac{(m-i)}{2})] \sim \tfrac{\alpha T}{2} - (\tfrac{\alpha T}{2} - \tfrac{(\nu+j-i+1)}{2})\ln(\tfrac{T}{2}) - (\tfrac{(1+\alpha)T}{2} + \tfrac{(-m+i-1)}{2})\ln(1+\alpha)$$

$$\ln\Gamma(\tfrac{\alpha T}{2} - \tfrac{(m-j-1)}{2}) \sim -\tfrac{\alpha T}{2} + (\tfrac{\alpha T}{2} - \tfrac{(m-j-1)}{2} - \tfrac{1}{2})\ln(\tfrac{\alpha T}{2})$$

thus:

$$\ln m(\mathbf{Y} \mid \Omega_{i22.1} \approx \mathbf{0}) - \ln m(\mathbf{Y} \mid \Omega_{j22.1} \approx \mathbf{0}) \geq$$

$$\geq const + (\tfrac{\nu-m+2j-i+1}{2})\ln(\tfrac{T}{2}) - \tfrac{(1+\alpha)T}{2}\ln(1+\alpha) + \tfrac{\alpha T}{2}\ln(\alpha)$$

$$+ \tfrac{\alpha T-\nu}{2}\ln[\left|\bar{Q}_{i22.1}\right| / \left|\bar{Q}_{j22.1}\right|]$$

$$- \tfrac{(m-i)(\alpha T-\nu)}{2}\ln(T) - \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_i + \tfrac{1}{T}\bar{Q}_{i22.1}\right|$$

$$+ \tfrac{(m-j)(\alpha T-\nu)}{2}\ln(T) + \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_j + \tfrac{1}{T}\bar{Q}_{j22.1}\right| =$$

$$= const + (\tfrac{\nu-m+2j-i+1}{2} + \tfrac{(i-j)(\alpha T-\nu)}{2})\ln(T) - \tfrac{(1+\alpha)T}{2}\ln(1+\alpha) + \tfrac{\alpha T}{2}\ln(\alpha) + \tfrac{\alpha T-\nu}{2}\ln[\left|\bar{Q}_{i22.1}\right| / \left|\bar{Q}_{j22.1}\right|]$$

$$- \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_i + \tfrac{1}{T}\bar{Q}_{i22.1}\right| + \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_j + \tfrac{1}{T}\bar{Q}_{j22.1}\right|$$

$$\geq const + (\tfrac{\nu-m+2j-i+1}{2} + \tfrac{\alpha T-\nu}{2})\ln(T) - \tfrac{(1+\alpha)T}{2}\ln(1+\alpha) + \tfrac{\alpha T-\nu}{2}\ln(\alpha) + \tfrac{\alpha T-\nu}{2}\ln[\left|\bar{Q}_{i22.1}\right| / \left|\bar{Q}_{j22.1}\right|]$$

$$- \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_i + \tfrac{1}{T}\bar{Q}_{i22.1}\right| + \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_j + \tfrac{1}{T}\bar{Q}_{j22.1}\right|$$

If $\text{rank}(\Omega) = i$ then from appendix 3, for large T:

$$- \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_i + \tfrac{1}{T}\bar{Q}_{i22.1}\right| = -\tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(\tfrac{1}{T}) + \hat{\Omega}_{i22.1}\right| \equiv -\tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_i\right| + \tfrac{(\alpha T-\nu)(m-i)}{2}\ln T$$

Of course for large T we also have:

$$\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_j + \tfrac{1}{T}\bar{Q}_{j22.1}\right| = \tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(\tfrac{1}{T}) + \hat{\Omega}_{j22.1}\right|$$

To find out what the asymptotic behavior of the latter is, let us write $\hat{\Omega}_{j22.1}$ in terms of its spectral decomposition: $\hat{\Omega}_{j22.1} = H'\hat{D}_j H$, where $\hat{D}_j = diag\{\frac{1}{T}\lambda_1, \frac{1}{T}\lambda_2, \ldots, \frac{1}{T}\lambda_{m-j}\}$, λ_i are eigenvalues of $T\hat{\Omega}_{j22.1}$, and $H'H = HH' = \mathbf{I}_{m-j}$. Then:

$$\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\,\mathrm{U}_j + \tfrac{1}{T}\bar{Q}_{j22.1}\right| = \tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(\tfrac{1}{T}) + \hat{\Omega}_{j22.1}\right| + \tfrac{\alpha T-\nu}{2}\ln\left|H'H\right| =$$

$$= \tfrac{\alpha T-\nu}{2}\ln\left|HO_{\mathbf{p}}(\tfrac{1}{T})H' + diag\{\tfrac{1}{T}\lambda_1, \tfrac{1}{T}\lambda_2, \ldots, \tfrac{1}{T}\lambda_{m-j}\}\right| = \tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(\tfrac{1}{T}) + diag\{\tfrac{1}{T}\lambda_1, \tfrac{1}{T}\lambda_2, \ldots, \tfrac{1}{T}\lambda_{m-j}\}\right|$$

since $\hat{\Omega}_{j22.1} \xrightarrow{\mathbf{p}} \Omega_{j22.1}$ and $\mathrm{rank}(\Omega_{j22.1}) = i - j$ (to prove the last fact note that as long as Ω is positive semidefinite it follows $\mathrm{rank}(\Omega) = \mathrm{rank}(\Omega_{j11}) + \mathrm{rank}(\Omega_{j22.1})$ see e.g. Carlson et al. (1974)) we conclude that $i - j$ eigenvalues $\frac{1}{T}\lambda_i$ will converge to positive values whereas the remaining $(m - i)$ will converge to 0, hence:

$$\tfrac{\alpha T-\nu}{2}\ln\det[O_{\mathbf{p}}(\tfrac{1}{T}) + diag\{\underbrace{O_{\mathbf{p}}(1), O_{\mathbf{p}}(1), \ldots, O_{\mathbf{p}}(1)}_{i-j}, \underbrace{O_{\mathbf{p}}(\tfrac{1}{T}), \ldots, O_{\mathbf{p}}(\tfrac{1}{T})}_{m-i}\}] =$$

$$= \tfrac{\alpha T-\nu}{2}\ln\det[U'diag\{\underbrace{O_{\mathbf{p}}(1), O_{\mathbf{p}}(1), \ldots, O_{\mathbf{p}}(1)}_{i-j}, \underbrace{O_{\mathbf{p}}(\tfrac{1}{T}), \ldots, O_{\mathbf{p}}(\tfrac{1}{T})}_{m-i}\}U] =$$

(U is an orthogonal matrix that diagonalizes the first term in determinant i.e. $O_{\mathbf{p}}(\frac{1}{T})$)

$$= \tfrac{\alpha T-\nu}{2}\ln\det[diag\{\underbrace{O_{\mathbf{p}}(1), \ldots, O_{\mathbf{p}}(1)}_{i-j}\}] + \tfrac{\alpha T-\nu}{2}\ln\det[diag\{\underbrace{O_{\mathbf{p}}(\tfrac{1}{T}), \ldots, O_{\mathbf{p}}(\tfrac{1}{T})}_{m-i}\}] =$$

$$= \tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(1)\right| + \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}O_{\mathbf{p}}(1)\right| = \tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(1)\right| - \tfrac{(\alpha T-\nu)(m-i)}{2}\ln T$$

taking into account to the above result we obtain:

$$-\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\,\mathrm{U}_i + \tfrac{1}{T}\bar{Q}_{i22.1}\right| + \tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\,\mathrm{U}_j + \tfrac{1}{T}\bar{Q}_{j22.1}\right| =$$

$$-\tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_i\right| + \tfrac{(\alpha T-\nu)(m-i)}{2}\ln T + \tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(1)\right| - \tfrac{(\alpha T-\nu)(m-i)}{2}\ln T = \tfrac{\alpha T-\nu}{2}\ln O_{\mathbf{p}}(1)$$

(we used the fact that $\mathbf{R}_i = O_{\mathbf{p}}(1)$, and note that the terms bounded in probability are also strictly positive), finally:

$$\ln m(\mathbf{Y} \mid \Omega_{i22.1} \approx \mathbf{0}) - \ln m(\mathbf{Y} \mid \Omega_{j22.1} \approx \mathbf{0}) \geq$$

$$\geq const + (\tfrac{\nu-m+2j-i+1}{2} + \tfrac{\alpha T-\nu}{2})\ln(T) - \tfrac{(1+\alpha)T}{2}\ln(1+\alpha) + \tfrac{\alpha T-\nu}{2}\ln(\alpha) + \tfrac{\alpha T-\nu}{2}\ln O_{\mathbf{p}}(1)$$

$$\geq const + (\tfrac{\nu-2m}{2} + \tfrac{\alpha T-\nu}{2})\ln(T) - \tfrac{T}{2}\ln(1+\alpha) + \tfrac{\alpha T-\nu}{2}\ln O_{\mathbf{p}}(1)$$

$$= const + \tfrac{\alpha T-2m}{2}\ln(T) - \tfrac{T}{2}\ln(1+\alpha) + \tfrac{\alpha T-2m}{2}\ln O_{\mathbf{p}}(1) =$$

$$= const + \tfrac{\alpha T-2m}{2}\ln(T \times O_{\mathbf{p}}(1)) - \tfrac{T}{2}\ln(1+\alpha)$$

since the terms under "$O_{\mathbf{p}}(1)$" are strictly positive, the first term will dominate the lower bound provided that $\alpha > \frac{2m}{T}$. If this condition holds then $\ln BF_i^* - \ln BF_j^*$ diverges to infinity for $T \to \infty$.

Lastly we prove d) i.e. consistency of BF testing $\mathrm{rank}(\Omega) = i$ vs $\mathrm{rank}(\Omega) = j$ where $i > j$ and $i = 2, \ldots, m-1$. This time the true hypothesis is $\mathrm{rank}(\Omega) = j$. As before we put $\nu_* = \alpha T$ and we can bound BF as:

$$\frac{m(\mathbf{Y} \mid \Omega_{i22.1} \approx \mathbf{0})}{m(\mathbf{Y} \mid \Omega_{j22.1} \approx \mathbf{0})} \leq \frac{\Gamma(\frac{\alpha T}{2} - \frac{(m-i)}{2})\Gamma(\frac{T+\nu}{2} - \frac{(m-i)}{2})}{\Gamma(\frac{\nu}{2} - \frac{(m-j-1)}{2})\Gamma(\frac{T+\alpha T}{2} - \frac{(m-j-1)}{2})} \frac{\left|\bar{Q}_{i22.1}\right|^{\frac{1}{2}(\alpha T-\nu)} \left|\mathrm{U}_i + \bar{Q}_{i22.1}\right|^{-\frac{1}{2}(\alpha T-\nu)}}{\left|\bar{Q}_{j22.1}\right|^{\frac{1}{2}(\alpha T-\nu)} \left|\mathrm{U}_j + \bar{Q}_{j22.1}\right|^{-\frac{1}{2}(\alpha T-\nu)}}$$

Expanding gamma functions we obtain:

$$\ln[\Gamma(\tfrac{T}{2}-\tfrac{(m-\nu-i)}{2})\,/\,\Gamma(\tfrac{(1+\alpha)T}{2}-\tfrac{(m-j-1)}{2})] \sim \tfrac{\alpha T}{2}-(\tfrac{\alpha T}{2}-\tfrac{(\nu+i-j-1)}{2})\ln(\tfrac{T}{2})-(\tfrac{(1+\alpha)T}{2}+\tfrac{j-m}{2})\ln(1+\alpha)$$

$$\ln\Gamma(\tfrac{\alpha T}{2}-\tfrac{(m-i)}{2}) \sim -\tfrac{\alpha T}{2}+(\tfrac{\alpha T}{2}-\tfrac{(m-i)}{2}-\tfrac{1}{2})\ln(\tfrac{\alpha T}{2})$$

thus:

$$\ln[\Gamma(\tfrac{T}{2}-\tfrac{(m-\nu-i)}{2})\,/\,\Gamma(\tfrac{(1+\alpha)T}{2}-\tfrac{(m-j-1)}{2})]+\ln\Gamma(\tfrac{\alpha T}{2}-\tfrac{(m-i)}{2}) \sim$$

$$\sim const+\tfrac{\nu+2i-j-m-2}{2}\ln(\tfrac{T}{2})-\tfrac{(1+\alpha)T}{2}\ln(1+\alpha)+\tfrac{\alpha T}{2}\ln(\alpha)$$

and:

$$\ln m(\mathbf{Y}\mid\Omega_{i22.1}\approx\mathbf{0})-\ln m(\mathbf{Y}\mid\Omega_{j22.1}\approx\mathbf{0})\leq$$

$$\leq const+\tfrac{\nu+2i-j-m-2}{2}\ln(\tfrac{T}{2})-\tfrac{(1+\alpha)T}{2}\ln(1+\alpha)+\tfrac{\alpha T}{2}\ln(\alpha)+\tfrac{(i-j)(\alpha T-\nu)}{2}\ln(T)$$

$$+\tfrac{\alpha T-\nu}{2}\ln[\left|\bar{Q}_{i22.1}\right|/\left|\bar{Q}_{j22.1}\right|]-\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_i+\tfrac{1}{T}\bar{Q}_{i22.1}\right|+\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_j+\tfrac{1}{T}\bar{Q}_{j22.1}\right|$$

As $\mathrm{rank}(\Omega)=j$, analogously as in appendix 3 we conclude:

$$\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_j+\tfrac{1}{T}\bar{Q}_{j22.1}\right|=\tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_j\right|-\tfrac{(\alpha T-\nu)(m-j)}{2}\ln T$$

where $\mathbf{R}_j=O_{\mathbf{p}}(1)$ and is positive definite, and:

$$-\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_i+\tfrac{1}{T}\bar{Q}_{i22.1}\right|=-\tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(\tfrac{1}{T})+\hat{\Omega}_{i22.1}\right|$$

where $\hat{\Omega}_{i22.1}=\frac{1}{T}E'_{m-i}E_{m-i}-\frac{1}{T}E'_{m-i}E_i(\frac{1}{T}E'_iE_i)^{+}\frac{1}{T}E'_iE_{m-i}$. Note that since $i>j$ and we assume $\mathrm{rank}(\Omega)=j$, Ω_{i11} must be singular and of course $\mathrm{rank}(\Omega_{i11})=j$. Since $\hat{\Omega}_{i11}\equiv\frac{1}{T}E'_iE_i\xrightarrow{\mathbf{p}}\Omega_{i11}$, for large T, we would have problems with inverting $\frac{1}{T}E'_iE_i$. This explains why in lemma 2 we work with M–P inverse. Furthermore we know $\hat{\Omega}_{i22.1}\xrightarrow{\mathbf{p}}\Omega_{i22.1}\geq 0$ (nonnegative definiteness is proved e.g. in Albert (1969)), but as $\mathrm{rank}(\Omega)=\mathrm{rank}(\Omega_{i11})+\mathrm{rank}(\Omega_{i22.1})$ we have $\mathrm{rank}(\Omega_{i22.1})=j-j=0$ i.e. $\Omega_{i22.1}=\mathbf{0}$, so that $\hat{\Omega}_{i22.1}\xrightarrow{\mathbf{p}}\mathbf{0}$. It implies that $\hat{\Omega}_{i22.1}$ is of smaller order than $O_{\mathbf{p}}(\frac{1}{T})$ (see appendix 3). Thus we can write:

$$-\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_i+\tfrac{1}{T}\bar{Q}_{i22.1}\right|=-\tfrac{\alpha T-\nu}{2}\ln\left|O_{\mathbf{p}}(\tfrac{1}{T})+\hat{\Omega}_{i22.1}\right|=-\tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_i\right|+\tfrac{(\alpha T-\nu)(m-i)}{2}\ln T$$

where $\mathbf{R}_i=O_{\mathbf{p}}(1)$ and is positive definite. Hence:

$$\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_j+\tfrac{1}{T}\bar{Q}_{j22.1}\right|-\tfrac{\alpha T-\nu}{2}\ln\left|\tfrac{1}{T}\mathrm{U}_i+\tfrac{1}{T}\bar{Q}_{i22.1}\right|=$$

$$=\tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_j\right|-\tfrac{(\alpha T-\nu)(m-j)}{2}\ln T-\tfrac{\alpha T-\nu}{2}\ln\left|\mathbf{R}_i\right|+\tfrac{(\alpha T-\nu)(m-i)}{2}\ln T=\tfrac{\alpha T-\nu}{2}\ln[\tfrac{\left|\mathbf{R}_j\right|}{\left|\mathbf{R}_i\right|}]-\tfrac{(i-j)(\alpha T-\nu)}{2}\ln T$$

inserting the above result into the upper bound we get:

$$\ln m(\mathbf{Y}\mid\Omega_{i22.1}\approx\mathbf{0})-\ln m(\mathbf{Y}\mid\Omega_{j22.1}\approx\mathbf{0})\leq$$

$$\leq const+\tfrac{\nu+2i-j-m-2}{2}\ln(\tfrac{T}{2})-\tfrac{(1+\alpha)T}{2}\ln(1+\alpha)+\tfrac{\alpha T}{2}\ln(\alpha)+\tfrac{\alpha T-\nu}{2}\ln[\tfrac{\left|\mathbf{R}_j\right|}{\left|\mathbf{R}_i\right|}\tfrac{\left|\bar{Q}_{i22.1}\right|}{\left|\bar{Q}_{j22.1}\right|}]$$

$$\leq const+\tfrac{\nu+2i-j-m-2}{2}\ln(\tfrac{T}{2})+\tfrac{\alpha T-\nu}{2}\ln[\tfrac{\alpha}{1+\alpha}\tfrac{\left|\mathbf{R}_j\right|}{\left|\mathbf{R}_i\right|}\tfrac{\left|\bar{Q}_{i22.1}\right|}{\left|\bar{Q}_{j22.1}\right|}]$$

the upper bound will diverge to minus infinity with $T \to \infty$ provided that $\alpha > \frac{\nu}{T}$ and $\alpha < (\frac{|\mathbf{R}_j|}{|\mathbf{R}_i|}\frac{|\bar{Q}_{i22.1}|}{|\bar{Q}_{j22.1}|} - 1)^{-1}$. Since $\mathbf{R}_j \equiv \mathrm{U}_j + \bar{Q}_{j22.1}$ and $\mathbf{R}_i \equiv \mathrm{U}_i + \bar{Q}_{i22.1}$ we obtain the condition $\frac{\nu}{T} < \alpha < (\frac{|\mathrm{U}_j + \bar{Q}_{j22.1}|}{|\mathrm{U}_i + \bar{Q}_{i22.1}|}\frac{|\bar{Q}_{i22.1}|}{|\bar{Q}_{j22.1}|} - 1)^{-1}$.

REFERENCES:

Albert, A. (1969), "Conditions for Positive and Nonnegative Definiteness in Terms of Pseudoinverses", *SIAM Journal on Applied Mathematics*, 17, pp. 434–440.

Altug, S. (1989), "Time–to–Build and Aggregate Fluctuations: Some New Evidence", *International Economic Review*, 30, pp. 889–920.

Alvarez–Lois, P., R. Harrison, L. Piscitelli and A. Scott (2005), "Taking DSGE Models to the Policy Environment", *Bank of England*, *Working Paper*.

An, S. and F. Schorfheide (2007), "Bayesian Analysis of DSGE Models", *Econometric Reviews*, 26, pp. 113–172.

Andrews, D.W.K. (1999), "Estimation When a Parameter is on a Boundary", *Econometrica*, 67, pp. 1341–1383.

Berger, J.O. and M. Delampady (1987), "Testing Precise Hypotheses", with discussion, *Statistical Science*, 2, pp. 317–352.

Bierens, H.J. (2007), "Econometric Analysis of Linearized Singular Dynamic Stochastic General Equilibrium Models", *Journal of Econometrics*, 136, pp. 595–627.

Bierens, H.J. and N.R. Swanson (2000), "The Econometric Consequence of the Ceteris Paribus Condition in Economic Theory", *Journal of Econometrics*, 95, pp. 223–253.

Boivin, J. and M.P. Giannoni (2006), "DSGE Models in a Data–Rich Environment", *Working Paper*.

Braithwaite, R.B. (1953), *Scientific Explanation*, Cambridge University Press.

Canova, F. (2007), *Methods for Applied Macroeconomic Research*, Princeton University Press.

Carlson, D., E. Haynsworth and T. Markham (1974), "A Generalization of the Schur Complement by Means of the Moore–Penrose Inverse", *SIAM Journal on Applied Mathematics*, 26, pp. 169–175.

Chari, V.V., P.J. Kehoe and E.R. McGrattan (2005), "A Critique of Structural VARs Using Business Cycle Theory", *Federal Reserve Bank of Minneapolis*, *Working Paper 631*.

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press.

Duhem, P. (1962), *The Aim and Structure of Physical Theory*, Atheneum, New York.

Eichenbaum M. (1996), "Some Comments on the Role of Econometrics in Economic Theory", *Federal Reserve Bank of Chicago Economic Perspectives*, 20, pp. 22–31.

Fernández–Villaverde, J., J.F. Rubio–Ramírez and T.J. Sargent (2005), "A,B,C's (and D)'s for Understanding VARS", *New York University*, *Working Paper*.

Geweke, J. (1999), "Computational Experiments and Reality", *University of Minnesota*, *Working Paper*.

Haavelmo, T. (1944), "The Probability Approach in Econometrics", *Econometrica*, 12, *Supplement*, pp. 1–115.

Hansen, G.D. (1985), "Indivisible Labor and the Business Cycle", *Journal of Monetary Economics*, 16, pp. 309–327.

Harville, D.A. (1997), *Matrix Algebra From a Statistician's Perspective*, Springer–Verlag.

Ireland, P.N. (2004), “A Method for Taking Models to the Data”, *Journal of Economic Dynamics and Control*, 28, pp. 1205–1226.

Jeffreys, H. (1961), *Theory of Probability*, Oxford University Press.

Khatri, C.G. (1968), “Some Results for Singular Normal Multivariate Regression Models”, *Sankhyā* , *series A*, 30, pp. 267–280.

King, R.G. (1995), “Quantitative Theory and Econometrics”, *Federal Reserve Bank of Richmond Economic Quarterly*, *vol. 81/3*.

King, R.G. and M.W. Watson (1996), “Money, Prices, Interest Rates and the Business Cycle”, *Review of Economics and Statistics*, 78, pp. 35–53.

Koopmans, T.C., H. Rubin and R.B. Leipnik (1950), “Measuring the Equation Systems of Dynamic Economics”, in: *Statistical Inference in Dynamic Economic Models*, *Cowles Commission Monograph No. 10*, John Wiley & Sons.

Koopmans, T.C. and Wm.C. Hood (1953), “The Estimation of Simultaneous Linear Economic Relationships”, in: Wm.C. Hood and T.C. Koopmans, eds., *Studies in Econometric Method*, *Cowles Commission Monograph No. 14*, John Wiley & Sons.

Kotz, S., N. Balakrishnan and N.L. Johnson (2000), *Continuous Multivariate Distributions*, *vol. I*: *Models and Applications*, John Wiley & Sons.

Kydland, F.E. and E.C. Prescott (1982), “Time to Build and Aggregate Fluctuations”, *Econometrica*, 50, pp. 1345–1370.

Landon–Lane, J. (2002), “Evaluating Dynamic Stochastic General Equilibrium Models using Likelihood Methods”, *Rutgers University*, *Working Paper*.

Lukomski, J. (1939), “On Some Properties of Multidimensional Distributions”, *Annals of Mathematical Statistics*, 10, pp. 236–246.

McGrattan, E.R., R. Rogerson and R. Wright (1997), “An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy”, *International Economic Review*, 38, pp. 267–290.

Pollard D. (2002), *A User’s Guide to Measure Theoretic Probability*, Cambridge University Press.

Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*, John Wiley & Sons.

Ravenna F. (2007), “Vector Autoregressions and Reduced Form Representations of DSGE Models”, *Journal of Monetary Economics*, 54, pp. 2048–2064.

Ruge–Murcia, F.J. (2003), “Methods to Estimate Dynamic Stochastic General Equilibrium Models”, *CIREQ*, *Cahier 17–2003*.

Rust, J. (1988), “Comment on Poirier: The Subjective Perspective of a “Spiritual Bayesian””, *Journal of Economic Perspectives*, 2, pp. 145–151.

Schorfheide, F. (2000), “Loss Function–Based Evaluation of DSGE Models”, *Journal of Applied Econometrics*, 15, pp. 645–670.

Sims, C.A. (2006), “Improving Monetary Policy Models”, *CEPS working paper No. 128*.

Smets, F. and R. Wouters (2003), “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area”, *Journal of the European Economic Association*, 1, pp. 1123–1175.

Teicher, H. (1960), “On the Mixture of Distributions”, *Annals of Mathematical Statistics*, 31, pp. 55–73.

Tinbergen, J. (1940), “Econometric Business Cycle Research”, *Review of Economic Studies*, 7, pp. 73–90.

Tintner, G. (1946), “Multiple Regression for Systems of Equations”, *Econometrica*, 14, pp. 5–36.

Watson, M.W. (1993), “Measures of Fit for Calibrated Models”, *Journal of Political Economy*, 101, pp. 1011–1041.