Orbital Priors for Time-Series Models

Kociecki, Andrzej

National Bank of Poland

23 November 2012

ORBITAL PRIORS FOR TIME–SERIES MODELS

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

First version: September 2011
This version: November 2012

Abstract: We propose the unified approach to construct the non–informative prior for time–series econometric models that are invariant under some group of transformations. We show that this invariance property characterizes some of the most popular models hence the applicability of the proposed framework is quite general. The suggested prior enjoys many desirable properties both from the Bayesian and non–Bayesian perspective. We provide detailed derivations of our prior in many standard time–series models including, AutoRegressions (AR), Vector AutoRegressions (VAR), Structural VAR and Error Correction Models (ECM).

KEYWORDS: Bayesian; Model invariance; Groups; Free group action; Orbit; Right Haar measure; Orbital decomposition; Maximal invariant; Cross section; Intersubjective prior; Vector AutoRegression (VAR); Structural VAR; Error Correction Model (ECM).

JEL Classification: C10, C11, C18, C32.

I. Introduction

Assume you have a model. If it turns out that some 1–1 transformation of the data preserves the original structure of the model we say that the model is invariant under this transformation. For example, let the model be $y \sim N(0, \sigma^2)$, where $y \in \mathbb{R}$ is the data and $N(a, b)$ denotes Normal distribution with mean a and variance b. Assume that we consider different scale of measurement of the data. It amounts to multiplying y by some positive constant g. Then $gy \sim N(0, (g\sigma)^2)$. Note that the transformed data possesses the same model structure i.e. it is still normally distributed with mean 0 and some (strictly positive) variance. In such a situation it is reasonable to assume that the estimation technique should be characterized by some sort of invariance with respect to this data transformation. Indeed, every estimation technique which is dependent on the scale of data measurement is highly suspect because the scale of measurement should not affect the inference. This does not prove that invariant statistical procedures are optimal in any sense but in fact there is a large statistics literature that demonstrates superiority of invariant statistical procedures applied to invariant models under a number of optimality criterions in various theoretical frameworks, see e.g. Eaton (1989), Lehmann (1986), Lehmann and Casella (1998) and numerous references therein.

The above discussion concerns the non–Bayesian estimation techniques. The Bayesian invariance considerations may be also relevant. For example, in the above example when we transform the data $y \mapsto gy$ there is an induced transformation of the parameter $\sigma^2 \mapsto g^2\sigma^2$. Thus it is sensible to ask e.g. whether your non–informative prior for σ^2 has also some sort of invariance properties. The idea of invariant priors is deeply rooted in Bayesian analysis and the best known example is the famous Jeffreys' prior. The latter is invariant under the 1–1 transformations of the parameter. However the focus of this paper is quite distinct from the invariance satisfied by the Jeffreys' prior. We deal with Bayesian inference that is invariant under 1–1 transformations of the data.

As in case of non–Bayesian estimation techniques it turns out that invariance arguments in the context of the sample space lead to some interesting optimality results of Bayesian inference. Thus although the sample space invariance arguments proper are not fully compelling, see e.g. Berger (1985), pp. 86–87, they do lead to priors that have remarkable and desirable properties. In particular under various theoretical setups it is possible to construct a non–informative prior for an invariant model, which we decided to call the intersubjective prior[1]. By the latter we mean the prior that possesses attributes of non–informative prior but in addition has many other desirable characteristics both from the Bayesian perspective (because e.g. it avoids some paradoxes in the spirit of Stone (1976), Dawid et al. (1973), Eaton and Freedman (2004)), and non–Bayesian perspective (because such a prior leads to reconciling frequentists and Bayesians since it works "as if" there is no prior at all[2]). To this end we introduce the so–called orbital prior that satisfies desiderata for such an intersubjective prior.

When a group acts transitively[3] on the parameter space then there is a number of compelling arguments to use the prior which is induced by the right invariant Haar measure on this underlying group and treating it as the intersubjective prior. Unfortunately when a group acts intransitively on the parameter space, which is true in the context of almost all time–series models applied in economics, then perhaps surprisingly the invariance arguments are not very decisive to suggest a formally justified intersubjective prior. However when some extra condition is satisfied (which we termed a free action of a group on the sample and parameter space) then there is a candidate for such a prior i.e. orbital prior.

Priors designed on the basis of invariance principles are especially useful in the context of economic time–series models. The reason is that implementing such priors does not require the stationarity assumption provided that appropriate parameterization of a model is chosen. We think this argument is not sufficiently emphasized in the literature. For example, the Jeffreys' prior in the strict form needs the computation of the expectation of the data unconditional second moments. This raises the issue whether conditional (on initial observations) or exact likelihood should be used. Uhlig (1994) showed that this choice really matters. In general when the data are nonstationary there is a fundamental problem with the existence of the exact likelihood or equivalently unconditional distribution for the initial observations. Though it should be admitted that many methods to mitigate this problem were proposed (Uhlig (1994) and Kleibergen and van Dijk (1994) cover many of them),

[1] We borrowed the term "intersubjective" from Dawid (1982), but with different connotations.

[2] For example, if the prior leads to the exact probability matching i.e. frequentist coverage sets are equal to Bayesian credible regions.

[3] Roughly speaking, it is the case if the parameter space is in 1–1 correspondence with the underlying group acting in a model. This excludes the case when a model is "too big" in relation to a group (seen as a space). See section II for the mathematical definition.

they should be considered only as informal ways to compute the Jeffreys' prior. Needless to say, the Jeffreys' prior is much delicate concept and its adoption even in regression models entails many ad hoc features. In fact, this was noticed by Jeffreys (1961) himself, pp. 182–183, 192, 360. Therefore the abstract mathematical treatment of non–informative prior presented in this paper is justified by the fact that the notion of non–informativeness is a very subtle one. Seemingly intuitive and compelling ignorance priors often turn out to be unacceptable when sufficiently scrutinized. For example, in time–series models, the curious thing about the Jeffreys' prior is that it depends on the sample size, see e.g. Phillips (1991). Moreover it does so in such a way that unreasonable weight is put on explosive cases. For this reason the use of Jeffreys' prior was criticized e.g. by Sims (1991), Leamer (1991), Poirier (1991), Koop and Steel (1991). The latter authors even indicated "*the inadequacy of Jeffreys prior for time–series models*".

Although we agree that sometimes the Jeffreys' prior may be useful in that it can penalize the non–identified parameter subspace in the parameter space, see e.g. Kleibergen and van Dijk (1994), Chao and Phillips (1998). However it does not change our impression that the motivation for Jeffreys' prior in time–series models is weak. Invariance under reparameterizations of a model i.e. one–to–one transformations of the parameters, sounds reasonable. However the ultimate properties or "real" effects of the Jeffreys' prior on the data analysis are not equally reasonable and may be easily questioned from several perspectives. See e.g. Ni and Sun (2003), Berger and Sun (2006,2007) and Eaton and Freedman (2004) for some recent critique.

Our position is that if a model shows some group invariant structure it is reasonable to exploit this in construction of the prior. Hence the prior we propose is a kind of the logical prior: it follows from a model but also by contemplating the group of transformations that might be relevant for the problem at hand. Evidently such a prior is both model and group specific. Staying within a given model and assuming different group of transformations usually entails different orbital priors and consequently different posteriors. This is a consequence of our implicit assumption that the group of transformation is an integral part of a model. Although such an inferential framework (slightly) violate the Likelihood Principle (LP), even worse LP violation concerns the Jeffreys' or Bernardo's priors, see e.g. Koop and Steel

(1991,1992), Lindley (1979), Poirier (1992)[4]. For in contrast to these priors, in time–series models, the orbital prior at least does not depend functionally on either the data or the sample size. The latter undesirable property may be called the serious LP violation.

The approach presented in this paper is very close to that in Chang and Eaves (1990) in that we use the same decomposition of the parameter space. However suggested elicitation of the prior on this decomposition is different, so is our motivation. As far as the latter is concerned our goal is to derive some non–informative priors that will be useful in invariant econometric time–series models. Also, the framework of Chamberlain and Moreira (2009) has a close contact with the approach suggested in our paper. It may be shown that their recommendation when dealing with Panel Data Models is just the application of the more general framework presented in this paper i.e. decomposing parameter space in accordance with orbital decomposition and assigning the orbital prior.

II. General Setup and Notation

All results in our paper are restricted to invariant models with respect to some group of transformations. See e.g. Lehmann (1986), chapter 6, for the theory and our assumption 1 for the mathematical definition. By G we will denote this underlying group acting in a model. We assume that G is a locally compact topological group. Basic material on groups, group actions and other related notions may be found in Eaton (1989). By e we denote the identity element in a group G. We will extensively use the concepts of Haar measures and integrals. Traditional reference is Nachbin (1965), but Eaton (1989) and Wijsman (1990) are also useful.

We will not differentiate between groups and its domain spaces. Thus $GL_m = \{g \in \mathbb{R}^{m\times m} \mid \det(g) \neq 0\}$ signifies both the general linear group with matrix multiplication as a group composition, and (seen as a space) the space of $m \times m$ nonsingular matrices. Analogous remark relates to LT_m^+ (UT_m^+): the group (= space) of $m \times m$ lower (upper) triangular matrices with positive elements on the diagonal; and $O_m = \{g \in \mathbb{R}^{m\times m} \mid g'g = gg' = \mathrm{I}_m\}$: the group (space) of orthogonal matrices ($\mathrm{I}_m : (m \times m)$ is the identity matrix). Obviously, a group composition in LT_m^+, UT_m^+

[4] It is not altogether obvious whether LP must hold in time–series models or, in general, within a non–experimental science like economics, see e.g. Akaike (1980), Geisser (1984), Berger and Wolpert (1988), p. 194, Hill (1988), Lane (1988), Poirier (1991).

and O_m is the usual matrix multiplication. Lastly, a space of $m\times m$ positive definite symmetric matrices will be denoted as PD_m, and $\mathbf{1}_n$ will signify a $1\times n$ vector of 1's.

Let Y be a random variable (with realization y) taking on values in $\mathcal{Y}$ (a sample space). Let $\mathcal{P}=\{P_\theta \mid \theta\in\Theta\}$ be a family of probability measures on $\mathcal{Y}$ indexed by the parameter $\theta\in\Theta$ i.e. a model. We assume that P_θ has a density $p(y\mid\theta)$ with respect to some dominating measure $\lambda_{\mathcal{Y}}$. Since we are working in the Bayesian framework in addition to a model we need a prior. The latter will be systematically denoted as π and it is understood that π is $\sigma-$finite measure.

We use the symbol "$\circ$" to denote the abstract group operation on some sets. On the other hand the group composition will not be symbolically distinguished from the usual matrix operation e.g. $g\circ\theta$ but gh, for $g,h\in G$ (group) and $\theta\in\Theta$ (set).

Special role in our applications plays the Affine group i.e. $AL_m=\{g=(w,k)\mid w\in GL_m, k\in\mathbb{R}^{m\times 1}\}$ with the following properties. The group composition is $g_2g_1=(w_2,k_2)(w_1,k_1)=(w_2w_1,w_2k_1+k_2)$, the identity element is $e=(\mathrm{I}_m,0)$ and inverse element of $g=(w,k)$ is $g^{-1}=(w^{-1},-w^{-1}k)$ so as $g^{-1}g=gg^{-1}=(\mathrm{I}_m,0)$.

With abuse of notation, but following Eaton (1989) and a large body of the literature, both an action of a group on $\mathcal{Y}$ and its induced action on Θ will be denoted as $g\circ y$ and $g\circ\theta$, respectively.

We will use the following unified notation, ν_G: the right invariant Haar measure on a group G; $\lambda_{\mathcal{S}}$: a $\sigma-$finite measure on a space $\mathcal{S}$. In particular the Lebesgue measure on a space $\mathcal{S}$ will be denoted as (ds), where $s\in\mathcal{S}$. We shall denote σ–algebra of Borel subsets of a space $\mathcal{S}$ as $\mathcal{B}_{\mathcal{S}}$.

The right invariant Haar measure on G satisfies $\nu_G(B\overline{g})=\nu_G(B)$, for all $B\in\mathcal{B}_G$, $\overline{g}\in G$, where $B\overline{g}=\{g\overline{g}\mid g\in B\}$, and leads to the right invariant integral on G i.e. $\int_G f(g\overline{g}^{-1})\nu_G(dg)=\int_G f(g)\nu_G(dg)$, for each (fixed) $\overline{g}\in G$ and all integrable f. Less formal way to write it is $\nu_G(d(g\overline{g}))=\nu_G(dg)$.

We say that G acts transitively on Θ if for each $\theta_1,\theta_2\in\Theta$ there is a $g\in G$ such that $\theta_2=g\circ\theta_1$. In other words, transitivity means that given $\theta_0\in\Theta$, every $\theta\in\Theta$ can be represented as $\theta=g\circ\theta_0$, for some $g\in G$.

We need two basic notions connected with group theory. The first one is the orbit. If G acts on some space $\mathcal{S}$, then the subset $\mathrm{Orb}_s=\{g\circ s\mid g\in G\}\subseteq\mathcal{S}$ (for given $s\in\mathcal{S}$) is called the orbit of s with respect to G. The other notion that occupies central position in our paper is the stabilizer. For any given $s\in\mathcal{S}$, let us define $\text{Stab}_s = \{g \in G \mid g \circ s = s\} \subseteq G$ and call it the stabilizer of s. The fundamental fact is that Stab_s is a subgroup of G. When $\text{Stab}_s = \{e\}$, $\forall \; s \in \mathcal{S}$, we say that a group G acts freely on $\mathcal{S}$, or there is a free action of G on $\mathcal{S}$ (recall that e denotes the identity element in G).

The function $f : \mathcal{S} \to \mathcal{X}$ is invariant under the action of some group G, or in short $G-$invariant, if $f(s) = f(g \circ s)$, for each $s \in \mathcal{S}$, $g \in G$. The function $f : \mathcal{S} \to \mathcal{X}$ is called maximal $G-$invariant if it is $G-$invariant and $f(s_1) = f(s_2)$ implies $s_1 = g \circ s_2$, for some $g \in G$.

The assumption that a model $\mathcal{P}$ is $G-$invariant reads

Assumption 1 (model G–invariance): $P_{g \circ \theta}(Y \in gB) = P_\theta(Y \in B)$ for all $g \in G$, where $B \in \mathcal{B}_\mathcal{Y}$ and $gB = \{g \circ y \mid y \in B\}$.

Assumption 1 holds in standard statistical models. However it is less known that this assumption is also valid in standard econometric models like univariate AR, VAR, Structural VAR (SVAR), Error Correction Models (ECM), Linear State–Space Models, Linear Panel Data Models (see e.g. Chamberlain and Moreira (2009)) and Instrumental Variables Model (see e.g. Chamberlain (2007)). This forms the natural basis for our approach.

We say that the prior measure is relatively invariant if $\pi(gB) = \chi(g) \cdot \pi(B)$, for all $B \in \mathcal{B}_\Theta$, $g \in G$ (notation gB is explained in definition 1) and $\chi : G \to \mathbb{R}^+$ is the multiplier which is a continuous function such that $\chi(g_1 g_2) = \chi(g_1)\chi(g_2)$, for all $g_1, g_2 \in G$, see e.g. Wijsman (1990), pp. 127–130. Equivalent definition of the relative invariance is that for all integrable f one has $\int_\Theta f(g^{-1} \circ \theta)\pi(d\theta) = \chi(g) \cdot \int_\Theta f(\theta)\pi(d\theta)$.

III. The Prior Under the Model Invariance

When a model is $G-$invariant it seems reasonable to restrict our considerations to the $G-$invariant posterior inference

Definition 1: *A Posterior* $\Pi(\cdot \mid y)$ *is said to be* G*–invariant if* $\Pi(\theta \in gB \mid g \circ y) = \Pi(\theta \in B \mid y)$, *for all* $g \in G$, $B \in \mathcal{B}_\Theta$, where $gB = \{g \circ \theta \mid \theta \in B\}$.

The $G-$invariance of the posterior in the context of $G-$invariant model was motivated and applied e.g. by Stone (1970), Dawid et al. (1973), Dawid (2006), Berger (1985), sections 3.3, 6.1 and p. 413, Lo and Cabrera (1987), Eaton (1989), pp. 48–50, Eaton and Sudderth (2002, 2004), Helland (2010), p. 80. The point is that having $G-$invariant posterior our inference for θ will be invariant under the simultaneous action of G on the sample and parameter space. This is a very intuitive requirement when a model is $G-$invariant. However the assumption that posterior is $G-$invariant is weaker than it might appear for we have the following

Lemma 1: *Let a model be G –invariant. Then a posterior is G –invariant iff the prior measure is relatively invariant.*

Proof: The sufficiency is proved by Stone (1970), under certain weak conditions. The necessary part is easy to prove, see e.g. Eaton (1989), p. 49.

This gives us the first criterion for selection of appropriate prior i.e. necessary condition. Unfortunately it turns out that the criterion is not particularly useful in itself because many priors agree with it. However, the posterior $G-$invariance constitutes a useful starting point to narrow down the possible choice of the prior in a more general prior setup. The apparently obvious "narrowing" strategy is to choose the (left) invariant prior measure i.e. $\pi(gB)=\pi(B)$; for all $B\in\mathcal{B}_\Theta$, $g\in G$, i.e. the relatively invariant prior measure with a multiplier $\chi(g)\equiv 1$. Such a prior guarantees that the posterior will be $G-$invariant yet in some cases there is no ambiguity what the invariant prior should be. Technically speaking if a group G acts transitively on Θ, under some further regularity conditions the only prior measure (up to a constant) which satisfies $\pi(gB)=\pi(B)$ is the Jeffreys' prior, see e.g. Dawid (2006), Eaton and Sudderth (2010), Helland (2010), p. 81. Unfortunately, practically in all econometric models the transitivity assumption is violated and there are many (left) invariant prior measures. Our aim is to propose general method to construct the relatively invariant prior that will be useful in practical situations. Doing this we shall incorporate other reasonable criteria and arguments that will make our prior setup more convincing.

IV. ORBITAL DECOMPOSITION

The crucial assumption in our framework is that the underlying group G acts freely on Θ

Assumption 2 ($G-\Theta$ freeness): $\mathrm{Stab}_{\theta}=\{e\}$; $\forall\ \ \theta\in\Theta$.

It turns out that in many time–series econometric models assumption 2 will be satisfied given the appropriate reparameterization of a model. This will be demonstrated for every particular model that we encounter later.

From assumption 2 it follows that the parameter space may be factorized

$$\Theta = G\times\mathcal{Z} \tag{1}$$

where $\mathcal{Z}$ is a global cross section (or in short a cross section) which is a subset of Θ that intersects each orbit Orb_{θ} in exactly one point, see e.g. Wijsman (1986). Thus a (global) cross section is in one–to–one correspondence with the orbit space. There are two ways to read factorization (1). Either $\theta\leftrightarrow(g,z)$ or $\theta = g\circ z$; $\theta\in\Theta, g\in G, z\in\mathcal{Z}$. The first variant only emphasizes that there is a bijection between every θ and (g,z). The second one signifies that every $\theta\in\Theta$ may be obtained by the action of (unique) element of a group $g\in G$ on some (unique) element of a cross section $z\in\mathcal{Z}$. These two "interpretations" will be useful in various contexts. After Barndorff–Nielsen et al. (1989), we shall call (1) the orbital decomposition of Θ (in short, the orbital decomposition). In the sequel we will use the notation $z(\theta)\in\mathcal{Z}$ for $\{z\}=\mathrm{Orb}_{\theta}\cap\mathcal{Z}$, to emphasize the dependence of a cross section z on θ. A fact of key importance is that having the orbital decomposition, a group G acts on $G\times\mathcal{Z}$ according to the rule $\overline{g}\circ(g,z):=(\overline{g}g,z)$, for every $\overline{g}\in G$, see e.g. Wijsman (1986,1990) i.e. the action of G on $\mathcal{Z}$ is trivial.

Intuition behind (1) is as follows. Any cross section indexes orbits i.e. $\mathrm{Orb}_{\theta}\cap\mathcal{Z}=\{z(\theta)\}$ (a singleton). It amounts to saying that each Orb_{θ}, for every $\theta\in\Theta$, contains one and only one element from $\mathcal{Z}$, which is $z(\theta)$. Since, by definition, G is transitive within each orbit, every $\theta^{*}\in\mathrm{Orb}_{\theta}$ such that $\theta^{*}\neq\theta$ may be represented as $\theta^{*}=g\circ z(\theta)$, where $z(\theta)\in\mathrm{Orb}_{\theta}$. Moreover, by $G-\Theta$ freeness assumption, g in $\theta^{*}=g\circ z(\theta)$ is unique. In general, every $\theta\in\Theta$ may be obtained by identifying the index of the orbit where θ lies i.e. z, and then finding the position of θ in that orbit i.e. g. Hence the latter gives coordinates within an orbit.

In the context of the sample space we assume the analogous

Assumption 3 ($G-\mathcal{Y}$ freeness): $\mathrm{Stab}_y = \{e\}$; $\forall \ y \in \mathcal{Y}$.

Assumption 3 will automatically hold in our case for almost all values of the data (see lemma 5). Assumption 3 implies existence of the orbital decomposition on the sample space

$$\mathcal{Y} = G \times \mathcal{U} \tag{2}$$

Where $\mathcal{U}$ is a (global) cross section on $\mathcal{Y}$. Hence there is a bijection $y \leftrightarrow (g,u)$, $g \in G$, $u \in \mathcal{U}$, and a group G acts on $G \times \mathcal{U}$ according to the rule $\overline{g} \circ (g,u) := (\overline{g}g, u)$.

Since we work in the Bayesian framework the more important for our development is the orbital decomposition on the parameter space (1). Its counterpart on the sample space (2) will play an instrumental role. However it is important to check that assumption 3 is satisfied because without it some nice consequences of the orbital prior as listed in section VI are simply untrue. Suffice it to say, proofs in e.g. Stein (1965), Stone (1970), Dawid et al. (1973), Severini et al. (2002) rely heavily on this assumption.

To make orbital decomposition (1) operational in practice we should consider its modification. Let us define a maximal invariant $t : \Theta \to \mathcal{T}$ i.e. $t(\theta) = t$ (note that with abuse of notation t also denotes the image of θ under t). Usually one chooses a maximal invariant so as it is easy to work with $\mathcal{T}$ ("nice" subset of $\mathbb{R}^n$). Following Wijsman (1986) we introduce a bijection $s : \mathcal{T} \to \mathcal{Z}$, $s(t(\theta)) = z(\theta)$. Note that such a bijection exists since in fact a cross section is also a maximal invariant and any one–to–one correspondence with maximal invariant is also maximal invariant, see e.g. Wijsman (1986). The difference between cross section and a maximal invariant is that $z(\theta)$ (besides it is a maximal invariant) must be a point in Orb_θ so that $\theta = g \circ z(\theta)$ for some $g \in G$, whereas a maximal invariant $t(\theta)$ does not have to be the point in Orb_θ. However $t(\theta)$ can be "lifted" to the orbit Orb_θ by a bijective map $s : \mathcal{T} \to \mathcal{Z}$ so as $\theta = g \circ s(t(\theta))$, for some $g \in G$. The following example (from Wijsman (1986)) illustrates the difference between z and t. Consider the simultaneous action of $G = GL_m$ on two spaces of symmetric positive definite matrices defined as $g \circ (S_1, S_2) := (gS_1g', gS_2g')$, for every $S_1, S_2 \in PD_m$. As is known there is some $\overline{g} \in G$ such that $\overline{g}S_1\overline{g}' = \mathrm{I}_m$ and $\overline{g}S_2\overline{g}' = \Lambda$ where $\Lambda = diag(\lambda_1, \ldots, \lambda_m)$ with $\lambda_1 > \ldots > \lambda_m > 0$. Denoting $\theta = (S_1, S_2)$ we have $\overline{g} \circ \theta = (\mathrm{I}_m, \Lambda)$. Since (I_m, Λ) lies in Orb_θ and takes different values in different orbits it is a cross section i.e. $z(\theta) = (\mathrm{I}_m, \Lambda)$. Thus $\mathcal{Z} = \{(\mathrm{I}_m, \Lambda) \mid \Lambda = diag(\lambda_1, \ldots, \lambda_m), \lambda_1 > \ldots > \lambda_m > 0\}$. On the other hand the maximal invariant will be $t(\theta) = (\lambda_1, \lambda_2, \ldots, \lambda_m)$, which are in fact characteristic values of $S_1^{-1}S_2$. Therefore $\mathcal{T} = \{\lambda_1, \lambda_2, \ldots, \lambda_m \mid \lambda_1 > \ldots > \lambda_m > 0\}$. The bijection $s : \mathcal{T} \to \mathcal{Z}$ is obvious.

Taking into account the above discussion we have two versions of the orbital decomposition

$$\Theta = G \times \mathcal{Z} = G \times \mathcal{T} \tag{3}$$

Thus there is a bijection $\theta \leftrightarrow (g, z) \leftrightarrow (g, t)$ and each $\theta \in \Theta$ may be uniquely written as $\theta = g \circ z(\theta) = g \circ s(t(\theta))$ for some $g \in G$ and $z \in \mathcal{Z}$. Since both $t(\theta)$ and $z(\theta)$ are maximal invariant, when $\theta = (g, z) = (g, t)$ then $\overline{g} \circ \theta := (\overline{g}g, z) = (\overline{g}g, t)$, for each $\overline{g} \in G$.

To deal with the orbital decomposition $\Theta = G \times \mathcal{T}$ we must ensure the bimeasurability condition i.e. that there is a one–to–one correspondence between Borel subsets in Θ and those in $G \times \mathcal{T}$. There are many ways to accomplish it. One possibility is to adopt conditions listed in theorem 10.1.2 in Farrell (1985). Other option is to assume that the action of G on Θ is proper (see e.g. Andersson (1982)). However we assume the following regularity condition which ensures we can work in practical situations where Jacobian derivations are needed.

Regularity condition (RC) (Wijsman (1986, 1990)): *Let the spaces Θ, $\mathcal{T}$ and a group G be differentiable (of order 1) manifolds with group action $(g, \theta) \mapsto g \circ \theta$ differentiable. Moreover, assume a bijective map $s : \mathcal{T} \to \mathcal{Z}$ is differentiable. Define $\varphi : G \times \mathcal{T} \to \Theta$ as $\varphi(g, t) = g \circ s(t)$, which is also differentiable and bijective. Assume φ has a positive Jacobian at the point (e, t), where e is the identity element in a group G.*

To build further intuition we have a simple example (that will be referred to later)

Example 1: Assume we have some O_m-invariant model (the specific action on the sample space does not concern us). The parameter space is $\Theta = \mathbb{R}^{m\times m}$ such that $\det(\theta) \neq 0$, for each $\theta \in \Theta$ (in other words $\Theta = GL_m$). The induced action on the parameter space is defined as $g \circ \theta := g\theta$, for each $g \in O_m$ (i.e. usual matrix multiplication). Clearly the $G-\Theta$ freeness holds i.e. $\mathrm{Stab}_\theta = \{g \in O_m \mid g\theta = \theta\} = \{\mathrm{I}_m\}$ ($=$the identity in O_m) since $g\theta = \theta$ $\Rightarrow g\theta\theta^{-1} = \theta\theta^{-1} \Rightarrow g = \mathrm{I}_m$. Thus there is an orbital decomposition of $\Theta = O_m \times \mathcal{T}$ $= O_m \times \mathcal{Z}$. In fact for every $\theta \in \Theta$ we can take $\theta = g \circ z(\theta) := gz(\theta)$, $g \in O_m$ $z(\theta) \in UT_m^+$ so as $\mathcal{Z} = UT_m^+$. This is a known decomposition in matrix theory called the QR decomposition. In our simple case $\mathcal{T} = \mathcal{Z}$, thus $s : \mathcal{T} \to \mathcal{Z}$ is the identity function. To show that $z \equiv z(\theta)$ is maximal O_m-invariant note that $\theta = gz(\theta) \Rightarrow \overline{g}\theta = \overline{g}gz(\theta)$ for every fixed $\overline{g} \in O_m$. On the other hand the orbital decomposition of $\overline{g}\theta$ is $\overline{g}\theta = \tilde{g}z(\overline{g}\theta)$ for some $\tilde{g} \in O_m$. This gives $\overline{g}gz(\theta) = \tilde{g}z(\overline{g}\theta) \Rightarrow \tilde{g}^{-1}\overline{g}g = z(\overline{g}\theta)(z(\theta))^{-1}$. But $\tilde{g}^{-1}\overline{g}g \in O_m$ and $z(\overline{g}\theta)(z(\theta))^{-1} \in UT_m^+$ and since $O_m \cap UT_m^+ = \{\mathrm{I}_m\}$ it follows $z(\overline{g}\theta)(z(\theta))^{-1} = \mathrm{I}_m$ $\Rightarrow z(\theta) = z(\overline{g}\theta)$ (this proves that z is O_m-invariant). On the other hand assume $z(\theta) = z(\overline{\theta})$. Denote by $g \in O_m$ the element such that $\theta = gz(\theta)$ and by $\overline{g} \in O_m$ the element such that $\overline{\theta} = \overline{g}z(\overline{\theta})$. Then $z(\theta) = z(\overline{\theta})$ $\Rightarrow g^{-1}\theta = z(\theta) = z(\overline{\theta}) = \overline{g}^{-1}\overline{\theta}$ $\Rightarrow \overline{g}g^{-1}\theta = \overline{\theta}$. Putting $g_* = \overline{g}g^{-1} \in O_m$ we have $g_*\theta = \overline{\theta}$, which proves maximal O_m-invariance of z. Note that the number of functionally independent elements in $z = t$ is $\frac{1}{2}m(m+1)$. The support for $\frac{1}{2}m(m-1)$ elements is $\mathbb{R}$ and for the remaining m elements is $\mathbb{R}^+$. Every two distinct $\frac{1}{2}m(m+1)-$tuples in $\mathbb{R}^{\frac{1}{2}m(m-1)} \times (\mathbb{R}^+)^m$ represent different orbits.

V. Properties of the Orbital Prior

It is easy to guess that the orbital decomposition is a key factorization in our approach. In fact our general recommendation is to elicit the prior for components in the orbital decomposition. This justifies calling our prior the orbital prior. In this section we explore various properties of this (orbital) prior.

The orbital prior will be useful in all models that are invariant under some group of transformations when, for some reasons, an improper or partially improper prior π is needed[5]. In such a case it is recommended 1) first to apply the orbital factorization on Θ i.e. $\theta \leftrightarrow (g,t)$ and 2) to elicit the improper prior as a product of

[5] By partially improper prior we mean for example the prior pdf $p(\theta_1,\theta_2) = p(\theta_1 \mid \theta_2)p(\theta_2)$, where $p(\theta_1 \mid \theta_2)$ is proper, whereas $p(\theta_2)$ is not.

independent marginal priors for g and t i.e. $\pi(dg,dt)=\pi(dg)\pi(dt)$, where $\pi(dg)$ is the right invariant Haar measure on G and $\pi(dt)$ may be any marginal prior (proper or improper). We notice that such a recommendation is "almost" explicitly given by Chamberlain and Moreira (2009). When $\pi(dt)$ is proper then the resultant joint posterior will be also proper irrespective of whether the marginal $\pi(dg)$ is proper or not (see lemma 3). Since the latter is the right invariant Haar measure on G it may be proper if and only if G is a compact group e.g. an orthogonal group, see Eaton (1989), p. 8. However in many applications G is not compact i.e. general linear or Affine groups.

The RC implies that there is a diffeomorphism between Θ and $G\times\mathcal{T}$, but also a homeomorphism between Θ and $G\times\mathcal{T}$ (i.e. both φ and φ^{-1} are continuous) and that the action of G on Θ is proper, see Wijsman (1986,1990). By assumption that G is a locally compact topological group there is a unique (up to a constant) right invariant Haar measure on G i.e. ν_G. The homeomorphism between Θ and $G\times\mathcal{T}$ implies that we can define a measure $\pi(B):=(\nu_G\otimes\lambda_{\mathcal{T}})\varphi^{-1}(B)$, $B\in\mathcal{B}_\Theta$, which is a prior on Θ induced by the product measure $\nu_G\otimes\lambda_{\mathcal{T}}$. Since, by the orbital decomposition, each $\theta\in\Theta$ may be represented as $\theta=g\circ s(t(\theta))=\varphi(g,t)$ for some $g\in G$, for all integrable f we have

$$\int_{\theta\in B} f(\theta)\pi(d\theta)=\int_{(g,t)\in\varphi^{-1}B} f(\varphi(g,t))\nu_G(dg)\lambda_{\mathcal{T}}(dt) \tag{4}$$

Previously we emphasized that when a model is $G-$invariant as a minimal requirement we should postulate the posterior $G-$invariance. By lemma 1 this is equivalent to adopting the relatively invariant prior. The next lemma demonstrates that when we work with a prior π which is induced by the orbital decomposition then this is the case

Lemma 2: *The product measure $\nu_G\otimes\lambda_{\mathcal{T}}$ and the prior π induced by $\nu_G\otimes\lambda_{\mathcal{T}}$ are relatively invariant.*

Proof: see appendix 1.

There is an apparent non–uniqueness in the orbital decomposition which calls for some clarification. Assume we have chosen some maximal invariant t and the corresponding z, which leads to the orbital decomposition $\theta=g\circ z$. The latter is unique but for the given z. Of course a cross section is non–unique in general. In fact any other cross section will be obtained as $\overline{g} \circ z$ for some $\overline{g} \in G$. Then we get the alternative orbital decomposition $\theta = g\overline{g}^{-1}\overline{g} \circ z = g_1 \circ z_1$, where $g_1 = g\overline{g}^{-1} \in G$, $z_1 = \overline{g} \circ z$. The question is whether the induced measure π for θ will be different in the case of two orbital decompositions. The next proposition is important since it shows that in a well defined sense the induced prior π will be independent of the particular cross section on Θ. Thus we can choose the one that is most convenient to work with. To fully grasp the meaning of the next proposition note that $z_1 = \overline{g} \circ z = \overline{g} \circ s(t) \equiv \overline{s}(t)$. Hence although z_1 and z are different cross sections they are both the functions of the same underlying maximal invariant t (which is common for them). Intuitively you may think of t as the functionally independent elements of z_1 and z.

Proposition 1: *Assume we have two orbital decompositions with cross sections* $z : \Theta \to \mathcal{Z}$ *and* $z_1 : \Theta \to \mathcal{Z}_1$ *i.e.* $\theta = g \circ z \equiv \varphi(g,t)$ *and* $\theta = g \circ z_1 \equiv \varphi_1(g,t)$, *then* $(\nu_G \otimes \lambda_T)\varphi^{-1}(B) = (\nu_G \otimes \lambda_T)\varphi_1^{-1}(B) := \pi(B)$, *for* $B \in \mathcal{B}_\Theta$.

Proof: see appendix 2.

Example 1 (cont.): Assume that instead of the cross section $z \in UT_m^+$ we take $z_1 : \Theta \to \mathcal{Z}_1 = PD_m$ such that $z_1(\theta) = (\theta'\theta)^{\frac{1}{2}}$, where $(\theta'\theta)^{\frac{1}{2}}$ denotes the square root of the positive definite $\theta'\theta$. The maximal $O_m -$invariance of z_1 is easy to prove ($O_m -$invariance is trivial and maximal $O_m -$invariance follows by Vinograd's theorem). In fact $\theta = gz_1$ for $g \in O_m$, $z_1 \in PD_m$ is the so–called polar decomposition of θ. Since $(\theta'\theta)^{\frac{1}{2}} = (z'g'gz)^{\frac{1}{2}} = (z'z)^{\frac{1}{2}}$ we have $z = z(z'z)^{-\frac{1}{2}}(z'z)^{\frac{1}{2}} = g(z'z)^{\frac{1}{2}} = g(\theta'\theta)^{\frac{1}{2}} =$ $= gz_1$, where $g = z(z'z)^{-\frac{1}{2}} \in O_m$. Since $z = gz_1 \Leftrightarrow z_1 = g^{-1}z$, denoting $\overline{g} = g^{-1}$ we have $z_1 = \overline{g}z$. Recall that in our simple case $t \equiv z \in UT_m^+$ and indeed z_1 is a function of t.

$$\int f(\theta)\pi(d\theta) = \int f(gt)\nu_{O_m}(dg)(dt) = \int f(gz)\nu_{O_m}(dg)(dt) =$$
$$= \int f(g\overline{g}^{-1}z_1)\nu_{O_m}(dg)(dt) = \int f(gz_1)\nu_{O_m}(dg)(dt) = \int f(g\overline{s}(t))\nu_{O_m}(dg)(dt)$$

where we used (4) and the fact that ν_{O_m} is the right invariant Haar measure. Therefore we end up with the same induced prior measure π on Θ.

We have remarkable consequence of the orbital decomposition

Lemma 3 (Zidek (1969)): $m(y) = \int_\Theta p(y \mid \theta)\pi(d\theta) = \int_{G\times T} p(y \mid g,t)\nu_G(dg)\lambda_T(dt) < \infty$; a.e. $[\lambda_y]$, *provided that* $\int_T \lambda_T(dt) < \infty$.

This is a very interesting property of the orbital prior: It does not matter for the marginal data density finiteness or posterior existence whether the right invariant Haar measure is proper or not. As long as the marginal prior for t is proper so is the posterior. We mention that if $\mathcal{Z}$ is empty i.e. $\Theta = G$ (this is the case when G acts transitively and freely on Θ), then $\int_{\Theta} p(y \mid \theta)\pi(d\theta) = \int_{G} p(y \mid g \circ \theta_0)\nu_G(dg) < \infty$; a.e. $[\lambda_y]$, where $\theta_0 \in \Theta$ is arbitrary and fixed, see Bondar (1977) for the proof and Kocięcki (2011) for some clarification. Hence working with right invariant Haar measure as a prior in the case $\Theta = G$ (which will not be a probability measure if G is non–compact) implies the existence of the posterior.

VI. Motivations for Obrital Prior

Working with the orbital decomposition $\Theta = G \times \mathcal{T}$ i.e. $\theta = (g, t)$, and assigning the measure $\nu_G \otimes \lambda_{\mathcal{T}}$ to $G \times \mathcal{T}$ i.e. $\pi(dg, dt) = \pi(dg)\pi(dt)$, which implicitly induces the prior measure $\pi(d\theta)$, turns out to be sensible for several reasons. In fact the list of arguments in favor is quite long.

Assume first that $\mathcal{T}$ in $\Theta = G \times \mathcal{T}$ is empty, that is $\Theta = G$. It is the case when a group G acts transitively and freely on Θ. Then

a) Stein (1965) demonstrated that we get the exact probability matching for θ i.e. the coverage frequentist probability is equal to the credible Bayesian probability, and for many functions of θ, see e.g. Berger and Sun (2007,2008) and Dawid (2007). In the context of prediction, Severini et al. (2002) and Eaton and Sudderth (2004) showed that exact probability matching holds also for certain predictive regions. Needless to say, for many non–Bayesians the exact probability matching is equivalent to applying the non–informative prior within Bayesian model (from their perspective the prior works “as if” there is no prior at all).

b) Many Bayesians think of the improper non–informative priors as an approximation to some proper priors. We argued earlier that in the case of $G-$invariant model the posterior should be also $G-$invariant. Stone (1970) showed that there is a sequence of proper priors applied to $G-$invariant model that corresponds in the limit to $G-$invariant posterior if and only if this $G-$invariant posterior was derived under the right invariant Haar measure ν_G.

c) Any posterior that is not derived under the right invariant Haar prior must be strongly inconsistent in the Stone’s (1976) sense, see Eaton and Sudderth (2004). Kocięcki (2011) showed that with the $G-\Theta$ freeness assumption the posterior inference under ν_G is not strongly inconsistent in the Stone's (1976) sense and is coherent in the sense of Heath and Sudderth (1978) and De Finetti (i.e. Dutch book).

d) Standard results indicate that when a decision problem is invariant, the Bayes estimators derived under the right invariant Haar prior are the best (minimum risk) invariant decision rules, see e.g. Berger (1985), section 6.6.2. Eaton and Sudderth (2001) showed that posterior predictive distributions computed with a right invariant Haar prior entail best invariant decision rules using a number of loss functions. Moreover, any predictive distribution that is not based on the right invariant Haar prior is incoherent in the sense of Heath and Sudderth (1978) and strongly inconsistent in the Stone's (1976) sense, see e.g. Eaton and Sudderth (1998,1999), Eaton (2008).

e) The posterior under ν_G is identical to Fraser's (1968) structural (fiducial) distribution, see Hora and Buehler (1966).

When $\mathcal{T}$ is not empty so as $\Theta = G \times \mathcal{T}$, then

a) We avoid the marginalization paradox for a component parameter t in $\theta = (g,t)$. Dawid et al. (1973) showed that when we adopt the prior which is a product of the right invariant Haar measure on G i.e. ν_G, and an arbitrary prior on $\mathcal{T}$, then the marginalization paradoxes for t will not appear[6].

b) In models where t is a scalar maximal invariant, Datta and Ghosh (1995) demonstrated superiority of the prior $\pi(dg,dt) = \pi(dg)\pi(dt)$[7]. In particular, they showed that such a prior is probability matching for t up to order $O(\frac{1}{N})$, where N denotes the sample size. Interestingly, they showed that Bernardo's reference prior may not be probability matching for t and suffer from the marginalization paradoxes. The same drawbacks concern the Jeffreys' prior, see e.g. Datta and Ghosh (1995), Berger and Sun (2006,2007,2008).

---

[6] There may be still paradoxes, which however will occur only if we contemplate different groups under which the model is still invariant, see Dawid et al. (1973). However, the invariance argument proper is most convincing only if we know a priori which form of invariance our model should preserve i.e. if we regard the underlying group of transformations as an integral part of a model. Then we must a priori decide what specific group acts on the sample space and there is little sense to consider other groups. In that case, there will be no marginalization paradox. See Bunke (1975) for similar remarks. Thus in practical cases one should use the largest group under which the model is invariant and which accommodates all sensible invariance considerations.

[7] To be precise, they use the marginal prior on $\mathcal{T}$ as suggested by Chang and Eaves (1990), but it may be shown that all conclusions stated in Datta and Ghosh (1995) are valid for arbitrary measure on $\mathcal{T}$.

c) Stein's (1965) results concerning equality of Bayesian credibility region with the frequentist confidence region were somewhat generalized to the case where $\mathcal{T}$ is not empty by Chang and Villegas (1986).

d) In a problem that is invariant under a transformation of some group, the risk function of any equivariant estimator is constant on each orbit in Θ, see e.g. Lehmann and Casella (1998), corollary 2.13. Even if the action of a group G is not transitive on Θ, the class of all equivariant estimators is small enough to propose some optimal estimator using other (perhaps ad–hoc) rules, see e.g. Berger (1985), p. 397, Helland (2010), p. 92. Zidek (1969) points out that the consequence of the risk constancy on every orbit is that we should explicitly specify the prior on orbits, that is $\pi(dt)$. He showed that under fairly general conditions, if we accompany any prior on orbits i.e. $\pi(dt)$, with the right invariant Haar measure i.e. ν_G, what really matters is $\pi(dt)$. That is the risk depends only on $\pi(dt)$. Related results in Chamberlain (2007), theorem 6.1, Chamberlain and Moreira (2009), proposition 4, establish the fact that, under some conditions, Bayes decision rules under orbital decomposition on the parameter space, have minimax property (provided that a group G is compact). In particular, Bayes decision rule depends only on the prior $\pi(dt)$ and we can replace averaging with respect to the measure ν_G with the fixed value $g \in G$ that leads to the maximum risk. This emphasizes the fundamental meaning of the orbital decomposition $\theta = (g, t)$.

e) In fact the crucial importance of t for inference is noticed by many researchers. The component t in the orbital decomposition of θ is called permissible (using terminology of Helland (2010), pp. 83–85) or natural (using terminology of McCullagh (2002)). Moreover t is also recommended as "appropriate" functions of parameters by Fraser (1968). According to model reduction policy of Helland (2010), p. 94, a model must be reduced to t and McCullagh (2002) agrees.

f) The distribution of a maximal invariant (and hence any invariant) on the sample space i.e. u in the orbital decomposition (2), depends only on maximal invariant on the parameter space i.e. t, see e.g. Berger (1985), p. 403. Moreover, if the group acts transitively and freely on the parameter space (i.e. $\Theta = G$) then the maximal invariant of the sample space is ancillary i.e. does not depend on any parameter, see e.g. Lehmann (1986), Ch. 10, theorem 1. Since all invariant tests are some function of maximal invariant, whose distribution depends only on maximal invariant on the parameter space, it appears that decomposition of parameter space into the maximal invariant and a group element has some extra merits. We may use this decomposition strategy in order to compare Bayesian and non–Bayesian tests e.g. how prior distribution of maximal invariant on the parameter space affects the invariant tests, whether there is a prior of maximal invariant on the parameter space that makes Bayesian and non–Bayesian tests agree with each other. In general we can try to find specific priors for maximal invariant t that has interesting non–Bayesian consequences. For the latter application see e.g. Chamberlain and Moreira (2009). Moreover, under mild regularity conditions, the likelihood ratio tests in invariant testing problem depend only on maximal invariant on the sample space, which, in turn, is entirely influenced by maximal invariant on the parameter space i.e. t, see e.g. Lehmann (1986), Eaton (1989). Lastly, denoting $\psi(t)$ any real function of t, under some conditions, there exists essentially unique unbiased estimator of $\psi(t)$ with minimum risk for all $t \in \mathcal{T}$ (Fraser (1956), Basu (1977)). This strengthens our recommendation for application of the orbital decomposition (when possible) by emphasizing an intersubjective character of the orbital prior.

g) The orbital decomposition on the parameter space is analogous to that used in the structural and/or structured (functional) models of Fraser (1968, 1979). The difference is that we use this decomposition on the parameter space whereas Fraser applies it to the (objective) error space, which he finds fundamental in his approach.

VII. Intersubjective Orbital Prior

Although $\pi(dt)$ of the measure $\pi(dg,dt) = \pi(dg)\pi(dt)$ may be problem–specific (proper or improper) there is a need for having some default prior in our setup. This was termed by us as the intersubjective prior which in fact is a different name for a non–informative prior.

Definition 2: *The intersubjective prior on $G \times \mathcal{T}$ is a product of the right invariant Haar measure on G i.e. $\nu_G(dg)$, and the Lebesgue measure on $\mathcal{T}$ i.e. (dt). The implied measure on Θ will be referred to as the intersubjective orbital prior.*

Therefore the intersubjective prior on $G \times \mathcal{T}$ will be $\pi(dg,dt) = \nu_G(dg)(dt)$, and the intersubjective orbital prior $\pi(d\theta)$ will be found by computing the Jacobian of the bijective transformation $\varphi : G \times \mathcal{T} \to \Theta$ under $\pi(dg,dt) = \nu_G(dg)(dt)$.

Example 1 (cont.): Assume we have chosen the cross section $z = t \in UT_m^+$. This leads to orbital decomposition $\theta = (g,t)$, where $g \in O_m$. The intersubjective prior on $G \times \mathcal{T}$ will be $\pi(dg,dt) = \nu_{O_m}(dg)(dt)$, where $\nu_{O_m}(dg)$ is the right invariant Haar measure on O_m and (dt) the Lebesgue measure on UT_m^+. Since O_m is compact, $\nu_{O_m}(dg)$ is also the left invariant Haar measure, hence it is common to say that $\nu_{O_m}(dg)$ is just invariant Haar measure on O_m. The induced intersubjective prior on Θ is computed by derivation of the Jacobian $J(g,t \to \theta)$ under the measure $\pi(dg,dt) = \nu_{O_m}(dg)(dt)$. Since $J(\theta \to g,t) = \prod_{i=1}^m t_{ii}^{m-i}\,\nu_{O_m}(dg)(dt)$ (where t_{ii} are diagonal elements in $t \in UT_m^+$), we have $(d\theta) = \prod_{i=1}^m t_{ii}^{m-i}\,\nu_{O_m}(dg)(dt)$ or $\prod_{i=1}^m t_{ii}^{-m+i}\,(d\theta) = \nu_{O_m}(dg)(dt)$. We need only to express t_{ii} in terms of θ. In fact t_{ii} are implicit functions of θ. Thus the intersubjective orbital prior on Θ is $\pi(d\theta) = \prod_{i=1}^m (t(\theta))_{ii}^{-m+i}\,(d\theta)$.

In a common nomenclature the Lebesgue measure is termed as the flat prior. Seeing in this light, the orbital decomposition gives justification for using flat priors but only in the context of maximal invariants t (or a cross section z). Intuitively the parameter space may be decomposed as a disjoint union of orbits. It means that each $\theta \in \Theta$ belongs to one and only one orbit. Usually each orbit (seen as a subspace of the parameter space Θ) will contain infinite number of parameters. However a cross section z (or equivalently maximal invariant t) is in 1–1 correspondence with orbits. Thus adopting a flat prior for t amounts to saying that all orbits are equally probable (i.e. different t's represent different orbits). That is we find no reason why the unknown ("true") value of the parameter may belong to the particular orbit and not the other one. This is the principle of insufficient reason in the purest mathematical form because orbits may also be perceived as abstract elements of the orbit space[8]. To put it other way, one must be reminded of the inherent property of the orbital decomposition. If you apply some group of transformation to the data i.e. $\overline{g} \circ y$; for some $\overline{g} \in G$, which implies the induced action of the group on the parameter space i.e. $\overline{g} \circ \theta$, then it has no effect on the cross section z (or maximal invariant t) since $\overline{g} \circ \theta := (\overline{g}g, z) = (\overline{g}g, t)$ for each $\overline{g} \in G$. That is z (or t) in the orbital decomposition of θ is the same as z (or t) in the orbital decomposition of $\overline{g} \circ \theta$. The latter property realizes us that flatness assumption for t will be invariant

[8] In fact our whole analysis equivalently may be based on abstract (Bourbaki) approach in which instead of cross section there is a topological quotient space of orbits $\Theta / G := \{\mathrm{Orb}_\theta \mid \theta \in \Theta\}$, see e.g. Zidek (1969), Andersson (1982), Eaton (1989), Wijsman (1990).

under any transformation of the data. This is the ideal situation for the application of the insufficient reason principle. On the other hand note the following subtle point. The invariance arguments applied only to the parameter space (in short, parameter invariance) lead to Jeffreys' prior which is the solution to the problem of finding the mathematical form of the prior measure that is invariant to any 1–1 reparameterizations. As a matter of fact, in the case of parameter invariance the principle of insufficient reason is not applicable and the Jeffreys' prior was the consequence of it. In contrast, invariance arguments applied to the sample space (in short, sample invariance) and representation of the parameter space with the orbital decomposition allows us to assign the flat prior for t which will be invariant under this form of invariance (i.e. sample invariance).

VIII. UNIVARIATE AR(p) MODEL

This section begins derivations of the orbital prior for most popular time–series models that enjoy invariance property. Consider the simple AR(p) model

$$y_t = c + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \sigma \cdot \varepsilon_t \tag{5}$$

where $\varepsilon_t \sim N(0,1)$. The most natural group of transformations is the affine group i.e. $G = AL_1$ [9]. That is we contemplate the following AL_1 action on the sample space

$$g \circ y \equiv (w,k) \circ y := (wy_1 + k, \ldots, wy_T + k) \tag{6}$$

where $w > 0$ and $k \in \mathbb{R}$. Using definition of G–invariance we can show that the AR(p) model is AL_1–invariant with the induced action on the parameter space

[9] We think that the Affine group is the most appropriate in the context of univariate and multivariate economic time–series models. It is the case when some of the variables are in logs and the remaining ones are not and we change the measurement units of all variables. For example, assume we have the original data $\begin{bmatrix} y_{1,t} \\ \log(y_{2,t}) \end{bmatrix}$ and we change the scale of measurement as follows $\begin{bmatrix} by_{1,t} \\ \log(ly_{2,t}) \end{bmatrix} = \begin{bmatrix} b & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ \log(y_{2,t}) \end{bmatrix} + \begin{bmatrix} 0 \\ \ln(l) \end{bmatrix}$, where $b \in \mathbb{R} \setminus \{0\}$ and $l > 0$. Denote $w = \begin{bmatrix} b & 0 \\ 0 & 1 \end{bmatrix}$ and $k = \begin{bmatrix} 0 \\ \ln(l) \end{bmatrix}$, then we have the following group action $\begin{bmatrix} y_{1,t} \\ \log(y_{2,t}) \end{bmatrix} \mapsto (w,k) \circ \begin{bmatrix} y_{1,t} \\ \log(y_{2,t}) \end{bmatrix} := w \begin{bmatrix} y_{1,t} \\ \log(y_{2,t}) \end{bmatrix} + k$, where (w,k) is an element of the Affine group AL_2.

$$(w,k)\circ(c,\beta_1,\ldots,\beta_p,\sigma):=(wc+(1-\sum_{i=1}^{p}\beta_i)k,\beta_1,\ldots,\beta_p,w\sigma) \tag{7}$$

The proof is essentially particular case of the proof of lemma 4, whence it is omitted. Note that the action of AL_1 on autoregressive parameters is trivial. The $G-\mathcal{Y}$ freeness is satisfied a.e. $[\lambda_{\mathcal{Y}}]$, when $T\geq 2$ (see lemma 5). On the other hand, the action AL_1 on the parameter space is free provided that $\sum_{i=1}^{p}\beta_i\neq 1$. Hence we must only exclude the zero measure hyperplane $\sum_{i=1}^{p}\beta_i=1$ from the parameter space but the part of parameter space that implies existence of explosive growth or oscillations makes no problem. Therefore $G-\Theta$ freeness holds almost everywhere in the parameter space and we can apply the orbital decomposition to $\theta=(c,\beta_1,\ldots,\beta_p,\sigma)$. One particular cross section is $z(\theta)=(0,\beta_1,\ldots,\beta_p,1)$ so that $t(\theta)=(\beta_1,\ldots,\beta_p)$. Remember that any cross section z serves the purpose because the induced prior $\pi(d\theta)$ is independent of this choice (proposition 1).

The intersubjective orbital prior will be a product of the right invariant Haar measure on $G=AL_1$ and the Lebesgue measure for $t(\theta)$. Since the former is $\nu_{AL_1}(dg)=w^{-1}(dw)(dk)$ we have

$$\pi(dg,dt)=w^{-1}(dw)(dk)(d\beta_1)\ldots(d\beta_p) \tag{8}$$

Since the orbital decomposition is a 1–1 correspondence, by writing $(c,\beta_1,,\ldots,\beta_p,\sigma)=(\sigma,c\cdot(1-\sum_{i=1}^{p}\beta_i)^{-1})\circ(0,\beta_1,\ldots,\beta_p,1)$, we identify $w\equiv\sigma$ and $k\equiv c\cdot(1-\sum_{i=1}^{p}\beta_i)^{-1}$ in $\theta=g\circ z(\theta)\equiv(w,k)\circ z(\theta)$. So we can write (8) as

$$\pi(dg,dt)\equiv\pi(d\sigma,dk,d\beta_1,\ldots,d\beta_p)=\sigma^{-1}(d\sigma)(dk)(d\beta_1)\ldots(d\beta_p) \tag{9}$$

Note that for stationary AR(p) model, k has a clear interpretation, namely the unconditional mean of a process. The rationale for eliciting prior for k instead of a constant c was given e.g. by Schotman and van Dijk (1991b), Zivot (1994). Incidentally, (9) is exactly the flat prior suggested by Zivot (1994) (his expression (5)).

Changing variables from $(\sigma,k,\beta_1,\ldots,\beta_p)$ to $(\sigma,c,\beta_1,\ldots,\beta_p)$ with the Jacobian $|1-\sum_{i=1}^{p}\beta_i|^{-1}$ we get the intersubjective prior on the original space induced by $\pi(dg,dt)$

$$\pi(d\sigma, dc, d\beta_1, \ldots, d\beta_p) \propto \sigma^{-1} \mid 1 - \sum_{i=1}^{p} \beta_i \mid^{-1} (d\sigma)(dc)(d\beta_1)\ldots(d\beta_p) \tag{10}$$

Interestingly, exactly the same prior (but differently motivated) was proposed by Sims (1991). The very important property of (10) is that in AR(1) model it implies probability matching for β_1 up to order $O(\frac{1}{T})$, where T denotes the sample size (see section VI).

Needless to say, the prior avoids marginalization paradoxes for autoregressive coefficients i.e. $t(\theta) = (\beta_1, \ldots, \beta_p)$. Moreover, note that in standard location–scale case (all $\beta_i = 0$ in (5)), the intersubjective orbital prior is $\pi(d\sigma, dc) \propto \sigma^{-1}(d\sigma)(dc)$, which is the Jeffreys' prior (when mean and variance are treated as independent of each other) and seems to be acceptable by many Bayesians and non–Bayesians (because it leads to convergence of Bayesian and frequentist solutions to the same inferential problem).

The appearance of $\mid 1 - \sum_{i=1}^{p} \beta_i \mid^{-1}$ in the prior (10) for the model (5) makes the simulation from the implied posterior more demanding. However our orbital prior is very suggestive and treats $(\gamma, \beta_1, \ldots, \beta_p, \sigma)$, where $\gamma = c \cdot (1 - \sum_{i=1}^{p} \beta_i)^{-1}$ as an "appropriate" parameter space. Therefore the orbital prior finds it natural to use the following parameterization of AR(p) model instead of (5)

$$(y_t - \gamma) = \beta_1(y_{t-1} - \gamma) + \cdots + \beta_p(y_{t-p} - \gamma) + \sigma \cdot \varepsilon_t \tag{11}$$

The model (11) is called the non–linear reduced form of (5) by Zivot (1994). Of course the interpretation of γ as the unconditional mean of the data is valid only for stationary AR(p) model. Otherwise it is just the "Greek letter". We notice that the parameterization (11) was also considered more convenient than (5) by Schotman and van Dijk (1991a,1991b), Zivot (1994).

It may be shown that the model (11) is AL_1-invariant under the action (6) (on the sample space) with the induced action on the parameter space

$$g \circ (\gamma, \beta_1, \ldots, \beta_p, \sigma) \equiv (w, k) \circ (\gamma, \beta_1, \ldots, \beta_p, \sigma) := (w\gamma + k, \beta_1, \ldots, \beta_p, w\sigma) \tag{12}$$

In contrast to specification (5), in the case of model (11) the $G - \Theta$ freeness always holds. The $G - \mathcal{Y}$ freeness is satisfied a.e. $[\lambda_{\mathcal{Y}}]$, when $T \geq 2$ (see lemma 5). Denoting $\theta = (\gamma, \beta_1, \ldots, \beta_p, \sigma)$ we have that $t(\theta) = (\beta_1, \ldots, \beta_p)$ and the corresponding cross section is $z(\theta) = s(t(\theta)) = (0, \beta_1, \ldots, \beta_p, 1)$.

The orbital decomposition is

$$(\gamma,\beta_1,,\dots,\beta_p,\sigma) = (\sigma,\gamma)\circ(0,\beta_1,\dots,\beta_p,1) \tag{13}$$

Identifying $w \equiv \sigma$ and $k \equiv \gamma$ in $\theta = g \circ z(\theta) \equiv (w,k)\circ z(\theta)$, we have the following intersubjective orbital prior for the model (11)

$$\pi(dg,dt) \equiv \pi(d\theta) = \pi(d\sigma,d\gamma,d\beta_1,\dots,d\beta_p) = \sigma^{-1}(d\sigma)(d\gamma)(d\beta_1)\dots(d\beta_p) \tag{14}$$

In general in all cases to be analyzed it seems to be a good practice to work with the parameterization of the model which comprises explicitly a group element of the orbital decomposition (in the case (11) this is $g = (\sigma, c\cdot(1-\sum_{i=1}^{p}\beta_i)^{-1})$). The main reason for this is the facilitation of the posterior sampling.

IX. Structural VAR Model

Consider the following version of the Structural VAR (SVAR) model

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \Lambda\varepsilon_t\,; \qquad t = 1,\dots,T\,. \tag{15}$$

Where $y_t \in \mathbb{R}^{m\times 1}$, $A_i : (m\times m)$, $c \in \mathbb{R}^{m\times 1}$, $\Lambda \in GL_m$ and $\varepsilon_t \mid y_{t-1},\dots \sim N(0,\mathrm{I}_m)$. Denote $y = [y_1\ y_2\ \dots\ y_T]$. The most natural action is that of AL_m e.g. think of the situation when variables in a vector y_t may or may not be in logs and we change the measurement units of all variables, see footnote 9. That is let $g = (w,k) \in AL_m$, where $w \in GL_m$, $k \in \mathbb{R}^{m\times 1}$, act on the sample space as

$$g\circ y \equiv (w,k)\circ y := (wy_1 + k,\dots,wy_T + k) = wy + k\cdot \mathbf{1}_T \tag{16}$$

Lemma 4: *The SVAR model* (15) *is* $AL_m -$ *invariant and the induced action on the parameter space is defined as* $g\circ(c,A_1,\dots,A_p,\Lambda) \equiv (w,k)\circ(c,A_1,\dots,A_p,\Lambda)$ $:= (wc + w(\mathrm{I}_m - \sum_{i=1}^{p} A_i)w^{-1}k, wA_1w^{-1},\dots,wA_pw^{-1},w\Lambda)$

Proof: see appendix 3.

To apply some of our results (e.g. those concerning the marginalization paradoxes) we must be sure that assumption 3 is satisfied

Lemma 5: *Let* $G = AL_m$ *act on the sample space as* $g \circ y := (wy_1 + k, \ldots, wy_T + k)$ $= wy + k \cdot \mathbf{1}_T$. *The* $G - \mathcal{Y}$ *freeness holds if all rows of* $[y', \mathbf{1}_T']'$ *are linearly independent.*

Proof: see appendix 4.

Thus when $T \geq m + 1$ then $G - \mathcal{Y}$ freeness is satisfied a.e. $[\lambda_{\mathcal{Y}}]$.

Far more important for our approach is the $G - \Theta$ freeness. Hence we have

Lemma 6: *In the case of SVAR model* (15)*:*

a) *When* $\text{rank}(\mathrm{I}_m - \sum_{i=1}^{p} A_i) = m$, *the* $G - \Theta$ *freeness assumption is satisfied.*

b) *When* $0 \leq \text{rank}(\mathrm{I}_m - \sum_{i=1}^{p} A_i) < m$, *the* $G - \Theta$ *freeness assumption is violated.*

Proof: see appendix 5.

There are several comments on lemma 6. First, when we exclude a priori parameter values that satisfy $0 \leq \text{rank}(\mathrm{I}_m - \sum_{i=1}^{p} A_i) < m$, then we are left with stationarity assumption and the $G - \Theta$ freeness holds. Second, some may question to attribute special importance to subsets of parameter space of measure zero (this is the position of Sims (1988)). Then the $G - \Theta$ freeness holds for almost all $[\pi]$ values of parameters. Third, if we need to impose the exact restriction of cointegration we should introduce this explicitly and work with e.g. Error Correction Model, to be discussed later (for which the $G - \Theta$ freeness assumption holds). Fourth, if $\mathrm{I}_m - \sum_{i=1}^{p} A_i = 0$, the appropriate treatment of the model amounts to using SVAR model in data first differences and the problem disappears i.e. SVAR in differences will fulfill the $G - \Theta$ freeness assumption. In sum, we find lemma 6 very interesting for the following reason. Although we did not explicitly consider the stationarity aspects of the SVAR model, those considerations were brought out in the course of our analysis in their full intensity. One must also be reminded that many of the alternative non–informative priors suggested in the literature do have similar restrictions on its use. For example, both Bernardo's reference prior and Jeffreys' prior possess inherent dichotomy with respect to stationarity assumption i.e. to obtain those priors you should a priori decide whether the data are stationary or not. Moreover the treatment of initial observations constitutes a great challenge. Also, the existence of cointegration requires extra considerations when applying Bernardo's reference prior or Jeffreys' prior.

Keeping in mind the above discussion we should further assume that $\mathrm{rank}(\mathrm{I}_m - \sum_{i=1}^{p} A_i) = m$. Denoting $\theta = (c, A_1, \ldots, A_p, \Lambda)$ we have

Lemma 7: $t(\theta) = (\Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda, \ldots, \Lambda^{-1}A_p\Lambda)$ *is maximal* AL_m *– invariant.*

Proof: see appendix 6.

It is a good place to remind the reader of the consequences of proposition 1 in the context of lemma 7. Any cross section will be a function of t. The latter consists of $m \times (mp)$ functionally independent elements on $\mathbb{R}^{m \times (mp)}$ and the induced prior measure π will be the same under any choice of cross section.

Let us denote $t_{A_i} = \Lambda^{-1}A_i\Lambda \in \mathbb{R}^{m \times m}$, for $i = 1, \ldots, p$. The intersubjective orbital prior will be a product of the right invariant Haar measure on $G = AL_m$ and Lebesgue measure on $t(\theta) = (t_{A_1}, \ldots, t_{A_p})$. Since the former is $\nu_{AL_m}(dg) = |\, w \,|^{-m}\, (dw)(dk)$ (see e.g. Eaton (1989), p. 11), we have

$$\pi(dg, dt) = |\, w \,|^{-m}\, (dw)(dk)(dt_{A_1}) \ldots (dt_{A_p}) \tag{17}$$

Having t we can easily derive a cross section

$$z(\theta) = s(t(\theta)) = (0, \Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda, \ldots, \Lambda^{-1}A_p\Lambda, \mathrm{I}_m) \tag{18}$$

Then one may check that

$$(c, A_1, \ldots, A_p, \Lambda) = (\Lambda, k_*) \circ (0, \Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda, \ldots, \Lambda^{-1}A_p\Lambda, \mathrm{I}_m) \tag{19}$$

where $k_* = \Lambda(\mathrm{I}_m - \sum_{i=1}^{p} A_i)^{-1}\Lambda^{-1}c$. Since the orbital decomposition is a 1–1 correspondence, we can identify $w \equiv \Lambda$ and $k \equiv k_*$. Hence we can write (17) as

$$\pi(dg, dt) = |\, \Lambda \,|^{-m}\, (d\Lambda)(dk_*)(dt_{A_1}) \ldots (dt_{A_p}) \tag{20}$$

Changing variables from $(\Lambda, k_*, t_{A_1}, \ldots, t_{A_p})$ to $(\Lambda, c, A_1, \ldots, A_1)$ with the Jacobian $|\det(\mathrm{I}_m - \sum_{i=1}^{p} A_i)|^{-1}$ we obtain the prior on the original space

$$\pi(d\Lambda, dc, dA_1, \ldots, dA_p) \propto |\, \Lambda \,|^{-m} |\det(\mathrm{I}_m - \sum_{i=1}^{p} A_i)|^{-1}\, (d\Lambda)(dc)(dA_1) \ldots (dA_p) \tag{21}$$

The intersubjective orbital prior for (15) i.e. (21), has evident drawback due to appearance of $|\det(\mathrm{I}_m - \sum_{i=1}^{p} A_i)|^{-1}$. Simulation from the posterior of SVAR under the prior (21) will be necessarily more difficult than that in the framework presented in Sims and Zha (1998). However one should be reminded that the prior (21) possesses intersubjective characteristics whereas the "non–informative" priors for SVAR proposed in the literature have no fundamental justifications. In fact it appears that the literature on the Bayesian SVAR models has not developed any standards for the non–informative priors yet, see e.g. ad hoc solutions in Sims and Zha (1999). For example a flat prior for $A_0, c, A_1, \dots, A_p$ (where $A_0 = \Lambda^{-1}$) was used by Sims and Zha (1999), Zha (1999) with the motivation to eliminate the discrepancy between posterior mode and ML estimates. The same flat prior appears also in Waggoner and Zha (2003) and Hamilton et al. (2007). We think that the rationale for the orbital prior is much deeper than that for the flat prior.

One possibility to avoid all problems connected with adoption of (21) is to use the following SVAR parameterization

$$(y_t - \gamma) = A_1(y_{t-1} - \gamma) + A_2(y_{t-2} - \gamma) + \cdots + A_p(y_{t-p} - \gamma) + \Lambda\varepsilon_t \tag{22}$$

where γ has interpretation of unconditional mean for stationary data but remains just a "Greek letter" otherwise. Assuming (16) one may show that the model (22) is $AL_m -$invariant with the induced action of AL_m on the parameter space[10]

$$g \circ (\gamma, A_1, \dots, A_p, \Lambda) \equiv (w,k) \circ (\gamma, A_1, \dots, A_p, \Lambda) := (w\gamma + k, wA_1w^{-1}, \dots, wA_pw^{-1}, w\Lambda) \tag{23}$$

Interesting fact about parameterization (22) is that $G - \Theta$ freeness holds without any qualification, hence the orbital decomposition $\Theta = G \times \mathcal{T}$ is valid for stationary and non–stationary data and in the presence of the cointegration ($G - \mathcal{Y}$ freeness is satisfied under condition in lemma 5). Denoting $\theta = (\gamma, A_1, \dots, A_p, \Lambda)$ we still have that $t(\theta) = (\Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda, \dots, \Lambda^{-1}A_p\Lambda)$ is maximal $AL_m -$invariant, with the corresponding cross section $z(\theta) = s(t(\theta)) = (0, \Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda, \dots, \Lambda^{-1}A_p\Lambda, \mathrm{I}_m)$. The orbital decomposition is

[10] The proof is a simple modification of that of lemma 4. In general, to save the space we omit proofs of model invariance in each particular case, since those proofs are quite similar to that of lemma 4.

$$(\gamma, A_1, \ldots, A_p, \Lambda) = (\Lambda, \gamma) \circ (0, \Lambda^{-1} A_1 \Lambda, \Lambda^{-1} A_2 \Lambda, \ldots, \Lambda^{-1} A_p \Lambda, \mathrm{I}_m) \tag{24}$$

We identify $w \equiv \Lambda$ and $k \equiv \gamma$, hence the intersubjective orbital prior on $G \times \mathcal{T}$ reads

$$\pi(d\Lambda, d\gamma, dt_{A_1}, \ldots, dt_{A_p}) \propto |\Lambda|^{-m} (d\Lambda)(d\gamma)(dt_{A_1})\ldots(dt_{A_p}) \tag{25}$$

Or equivalently (by changing variables)

$$\pi(d\Lambda, d\gamma, dA_1, \ldots, dA_p) \propto |\Lambda|^{-m} (d\Lambda)(d\gamma)(A_1)\ldots(A_p) \tag{26}$$

which is the intersubjective orbital prior on Θ i.e. for the model (22).

For completeness of our results we also provide the treatment of the following (most popular) SVAR specification

$$A_0 y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t \tag{27}$$

where $A_0 \in GL_m$ and $\varepsilon_t \mid y_{t-1}, \ldots \sim N(0, \mathrm{I}_m)$. To save the space we give the following results without proofs. The model (27) is AL_m – invariant under (16) with the group action on the parameter space defined as

$$\begin{aligned} g \circ (c, A_0, A_1, \ldots, A_p) &\equiv (w, k) \circ (c, A_0, A_1, \ldots, A_p) \\ &:= (c + [A_0 - \textstyle\sum_{i=1}^{p} A_i] w^{-1} k, A_0 w^{-1}, A_1 w^{-1}, \ldots, A_p w^{-1}) \end{aligned} \tag{28}$$

for each $g \in AL_m$. The AL_m acts freely on the parameter space provided that $\mathrm{rank}(A_0 - \sum_{i=1}^{p} A_i) = m$. Denoting $\theta = (c, A_0, A_1, \ldots, A_p)$, we have $t(\theta) = (A_1 A_0^{-1}, \ldots, A_p A_0^{-1})$ and the corresponding cross section $z(\theta) = (0, \mathrm{I}_m, A_1 A_0^{-1}, \ldots, A_p A_0^{-1})$ which induces the orbital decomposition

$$(c, A_0, A_1, \ldots, A_p) = (A_0^{-1}, (A_0 - \textstyle\sum_{i=1}^{p} A_i)^{-1} c) \circ (0, \mathrm{I}_m, A_1 A_0^{-1}, \ldots, A_p A_0^{-1}) \tag{29}$$

We easily identify $w \equiv A_0^{-1}$ and $k \equiv (A_0 - \sum_{i=1}^{p} A_i)^{-1} c$. Since the intersubjective orbital prior on $G \times \mathcal{T}$ is

$$\pi(dw,dk,dt_{A_1},\ldots,dt_{A_p}) \propto |\, w \,|^{-m}\ (dw)(dk)(dt_{A_1})\ldots(dt_{A_p}) \tag{30}$$

where $t_{A_i} = A_i A_0^{-1}$, by changing variables from $w,k,t_{A_1},\ldots,t_{A_p}$ to $c,A_0,A_1,\ldots,A_p$ we have the induced intersubjective orbital prior for the model (27)

$$\pi(dc,dA_0,dA_1,\ldots,dA_p) \propto |\, A_0 \,|^{-m(p+1)} |\, \det(A_0 - \sum_{i=1}^{p} A_i) \,|^{-1}\ (dc)(dA_0)(dA_1)\ldots(dA_p) \tag{31}$$

Lastly, using the parameterization

$$A_0(y_t - \gamma) = A_1(y_{t-1} - \gamma) + A_2(y_{t-2} - \gamma) + \cdots + A_p(y_{t-p} - \gamma) + \varepsilon_t \tag{32}$$

And assuming (16) we have the induced action on the parameter space

$$g \circ (\gamma, A_0, A_1,\ldots,A_p) \equiv (w,k) \circ (\gamma, A_0, A_1,\ldots,A_p) := (w\gamma + k, A_0 w^{-1}, A_1 w^{-1},\ldots,A_p w^{-1}) \tag{33}$$

for each $g \in AL_m$. Denoting $\theta = (\gamma, A_0, A_1,\ldots,A_p)$ we have $t(\theta) = (A_1 A_0^{-1},\ldots,A_p A_0^{-1})$ and the corresponding cross section $z(\theta) = (0, \mathrm{I}_m, A_1 A_0^{-1},\ldots,A_p A_0^{-1})$. Since $G - \Theta$ freeness in the case of (32) is always satisfied we can write the unique orbital decomposition as

$$\theta = g \circ z(\theta) \equiv (w,k) \circ z(\theta) = (A_0^{-1}, \gamma) \circ (0, \mathrm{I}_m, A_1 A_0^{-1},\ldots,A_p A_0^{-1}) \tag{34}$$

Therefore we identify $w \equiv A_0^{-1}$, $k \equiv \gamma$ and the intersubjective orbital prior for the model (32) is

$$\pi(d\gamma,dA_0,dA_1,\ldots,dA_p) \propto |\, A_0 \,|^{-m(p+1)}\ (d\gamma)(dA_0)(dA_1)\ldots(dA_p) \tag{35}$$

X. VAR MODEL

Consider the SVAR model (15) with the restriction $\Lambda \in LT_m^+$

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \Lambda \varepsilon_t \tag{36}$$

Equivalent way to write (36) is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \upsilon_t \tag{37}$$

where $\upsilon_t \mid y_{t-1},\ldots \sim N(0,\Sigma)$ and $\Sigma \in PD_m$ is in 1–1 correspondence with $\Lambda \in LT_m^+$ through the Choleski decomposition $\Sigma = \Lambda\Lambda'$. That is we arrive at the VAR model. For mathematical reasons we prefer to work with (36) than with (37)[11]. Since presently $\Lambda \in LT_m^+$ (in the case of SVAR, $\Lambda \in GL_m$) the group AL_m is too large for our problem. Indeed, if $g = (w,k) \in AL_m$, where $w \in GL_m$ and $k \in \mathbb{R}^{m\times 1}$, acts in the model (36) then this destroys the model structure. That is after the action of AL_m on the parameter space, Λ no longer belongs to LT_m^+ but to GL_m. We should consider the subset of AL_m consisting of elements $g = (w,k)$, where $w \in LT_m^+$ and $k \in \mathbb{R}^{m\times 1}$. In fact such a subset is a subgroup of AL_m to be denoted as AL_m^Δ. Interestingly most results from the previous section concerning the model (15) remain valid without any modification. In particular, the model (36) is $AL_m^\Delta-$invariant, with the same action induced on the parameter space

$$
\begin{aligned}
g \circ (c, A_1,\ldots,A_p,\Lambda) &\equiv (w,k)\circ(c,A_1,\ldots,A_p,\Lambda) \\
&:= (wc + w(\mathrm{I}_m - \textstyle\sum_{i=1}^p A_i)w^{-1}k, wA_1w^{-1},\ldots,wA_pw^{-1},w\Lambda) \qquad (38)
\end{aligned}
$$

Note that this time $g = (w,k) \in AL_m^\Delta$. Further, when $T \geq m+1$ then $G-\mathcal{Y}$ freeness is satisfied a.e. $[\lambda_{\mathcal{Y}}]$. When $\mathrm{rank}(\mathrm{I}_m - \sum_{i=1}^p A_i) = m$, the $G-\Theta$ freeness assumption holds. Lastly $t(c,A_1,\ldots,A_p,\Lambda) = (\Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda,\ldots,\Lambda^{-1}A_p\Lambda)$ is maximal $AL_m^\Delta-$invariant

We continue to assume $\mathrm{rank}(\mathrm{I}_m - \sum_{i=1}^p A_i) = m$. Using the notation from the previous section i.e. $t_{A_i} = \Lambda^{-1}A_i\Lambda$, for $i = 1,\ldots,p$, the intersubjective orbital prior will be a product measure of the right invariant Haar measure on $G = AL_m^\Delta$ and the Lebesgue measure for $t = (t_{A_1},\ldots,t_{A_p})$. It may be shown (e.g. using slight modification of derivations in Eaton (1989), pp. 10–11 and 16–17) that $\nu_{AL_m^\Delta}(dg) = \prod_{i=1}^m w_{ii}^{-m+i-1}(dw)(dk)$, where w_{ii} are diagonal elements of $w \in LT_m^+$. As a result

$$
\pi(dg,dt) = \prod_{i=1}^m w_{ii}^{-m+i-1}(dw)(dk)(dt_{A_1})\ldots(dt_{A_p}) \qquad (39)
$$

[11] Traditional VAR specification (37) (although AL_m-invariant) is cumbersome for developing invariance arguments since the underlying stabilizer Stab_θ is not only an identity element in AL_m but is not even a compact space. To our knowledge there is no mathematical theory to accommodate this case. Suffice it to say, in such a case nothing of the kind of the orbital decomposition exists.

As before, having t we can easily derive a cross section

$$z(c, A_1, \ldots, A_p, \Lambda) = s(t(c, A_1, \ldots, A_p, \Lambda)) = (0, \Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda, \ldots, \Lambda^{-1}A_p\Lambda, \mathrm{I}_m) \quad (40)$$

Then one may check that

$$(c, A_1, \ldots, A_p, \Lambda) = (\Lambda, k_*) \circ (0, \Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda, \ldots, \Lambda^{-1}A_p\Lambda, \mathrm{I}_m) \quad (41)$$

where $k_* = \Lambda(\mathrm{I}_m - \sum_{i=1}^{p} A_i)^{-1}\Lambda^{-1}c$. Since the orbital decomposition is a 1–1 correspondence, we can identify $w \equiv \Lambda$ and $k \equiv k_*$. Hence we can write

$$\pi(dg, dt) = \prod_{i=1}^{m} \Lambda_{ii}^{-m+i-1}(d\Lambda)(dk_*)(dt_{A_1})\ldots(dt_{A_p}) \quad (42)$$

We can change variables from $(\Lambda, k_*, t_{A_1}, \ldots, t_{A_p})$ to that of the traditional VAR parameterization $(\Sigma, c, A_1, \ldots, A_p)$. Noting that the Jacobian $J(\Lambda \to \Sigma) = = 2^{-m} \cdot \prod_{i=1}^{m} \Lambda_{ii}^{-m+i-1} = 2^{-m} \cdot \prod_{i=1}^{m} | \Sigma^{[1 \mapsto i]} |^{-\frac{1}{2}}$, where $\Sigma^{[1 \mapsto i]} := (\sigma_{jk})$; $j, k = 1, \ldots, i$ ($\Sigma^{[1 \mapsto i]}$ is a leading principal submatrix of Σ consisting the first i rows and columns of Σ), we can easily find the intersubjective orbital prior for (37)

$$\pi(d\Sigma, dc, dA_1, \ldots, dA_p) \propto | \det(\mathrm{I}_m - \sum_{i=1}^{p} A_i) |^{-1} \cdot \prod_{i=1}^{m} | \Sigma^{[1 \mapsto i]} |^{-1}(d\Sigma)(dc)(dA_1)\ldots(dA_p) \quad (43)$$

As in the case of SVAR model we may write VAR model so as to reduce to a minimum the sampling problems. Let us write the VAR model (36) as

$$(y_t - \gamma) = A_1(y_{t-1} - \gamma) + A_2(y_{t-2} - \gamma) + \cdots + A_p(y_{t-p} - \gamma) + \Lambda\varepsilon_t \quad (44)$$

where $\Lambda \in LT_m^+$. All remarks which were expressed concerning (22) apply also to (44) i.e. $G - \Theta$ freeness always holds, the action on the parameter space is the same as (23) with the modification that $g = (w, k) \in AL_m^{\Delta}$, the cross section is the same etc. We only confine to providing the intersubjective orbital prior for (44), which is

$$\pi(d\Lambda, d\gamma, dA_1, \ldots, dA_p) \propto \prod_{i=1}^{m} \Lambda_{ii}^{-m+i-1}(d\Lambda)(d\gamma)(A_1)\ldots(A_p) \quad (45)$$

Changing variables from Λ to $\Sigma = \Lambda\Lambda'$ we get the intersubjective orbital prior for the standard VAR specification

$$(y_t - \gamma) = A_1(y_{t-1} - \gamma) + A_2(y_{t-2} - \gamma) + \cdots + A_p(y_{t-p} - \gamma) + \upsilon_t \tag{46}$$

where $\upsilon_t \mid y_{t-1}, \ldots \sim N(0, \Sigma)$, which reads

$$\pi(d\Sigma, d\gamma, dA_1, \ldots, dA_p) \propto \prod_{i=1}^{m} |\, \Sigma^{[1 \mapsto i]} \,|^{-1} \, (d\Sigma)(d\gamma)(A_1)\ldots(A_p) \tag{47}$$

The model (46) was called the steady–state VAR by Villani (2009), thus the "marginal" prior for Σ i.e. $\prod_{i=1}^{m} |\, \Sigma^{[1 \mapsto i]} \,|^{-1}$, could be used in the framework of Villani (2009) (instead of the Jeffreys' prior adopted by him). Although (47) looks unhandy from the computational point of view, a method to sample from the joint posterior of the normal model under the prior (47) may be constructed using some results in Berger and Sun (2007). For example, the full conditional posterior of Σ under the prior (47) allows for the exact sampling from (which facilitates the Gibbs sampling). Hence the sampling algorithm for the model (46) given by Villani (2009) requires only slight modification.

We note in passing that the "marginal" prior for Σ i.e. $\prod_{i=1}^{m} |\, \Sigma^{[1 \mapsto i]} \,|^{-1}$, is exactly the prior recommended by Eaton and Sudderth (2010), proposition 4.1, for an $m-$variate normal model with mean 0 and covariance Σ (written in a more elegant form). However their motivation for this prior was based on coherence requirements in the sense of Heath and Sudderth (1978) and Stone's (1976) strong inconsistency arguments. See also Kocięcki (2011) for some further discussion.

As a final digression concerning VAR model, we note that if we accept the orbital decomposition of the parameter space as the "appropriate" parameterization then we should design a version of Minnesota prior for maximal invariants $t_{A_i} = \Lambda^{-1} A_i \Lambda$, for $i = 1, \ldots, p$ (instead of A_i in the original Minnesota prior, see e.g. Litterman (1986)). This prior for t_{A_i} could be accompanied with the "marginal" prior for Σ i.e. $\prod_{i=1}^{m} |\, \Sigma^{[1 \mapsto i]} \,|^{-1}$. At the conceptual level it looks promising, since the original Minnesota prior needs for scaling the prior of each A_i by variances of the error components, which appear as the ratios of diagonal elements from Σ, see e.g. Litterman (1986). But Λ is the Choleski square root of Σ i.e. $\Sigma = \Lambda\Lambda'$, hence eliciting the prior for $\Lambda^{-1} A_i \Lambda$ we implicitly scale the prior by ratios of variance components in a way that is dictated by our orbital decomposition. Whether it could lead to better forecasting is an interesting empirical question.

XI. Error Correction Model

Consider the following Error Correction Model (ECM)

$$(\Delta y_t - \gamma) = \alpha\beta y_{t-1} + A_1(\Delta y_{t-1} - \gamma) + \cdots + A_p(\Delta y_{t-p} - \gamma) + \upsilon_t \tag{48}$$

where $\alpha : (m \times r)$ with $\text{rank}(\alpha) = r$, $\beta : (r \times m)$ with $\text{rank}(\beta) = r$, and $\upsilon_t \mid y_{t-1}, \ldots \sim N(0, \Sigma)$. In particular, Bayesian analysis of this specification of ECM was recently given by Villani (2009) who called it the steady–state Vector ECM and found some compelling arguments for it (e.g. the model (48) allows for explicit modeling of the growth rates since $E(\Delta y_t) = \gamma$).

In accordance with recent Bayesian cointegration literature represented by Strachan and Inder (2004), Villani (2005) and Koop et al. (2006), we should impose semiorthogonal restrictions on cointegrating vectors i.e. $\beta\beta' = \mathrm{I}_r$. This is a consequence of considering the row space of β as the basic object of inference. Denote the latter as $\rho = sp\{\beta\}$ which is an element of the so–called Grassmann manifold. The restriction $\beta\beta' = \mathrm{I}_r$ and the model structure implies that the only sensible group transformation on the sample space is O_m. It is so because 1) there is no traditional constant in the specification (48) hence the largest sensible group is GL_m and 2) as will be clear in a moment, the induced action of a group GL_m on the cointegrating vectors is $\beta g'$, for any $g \in GL_m$. But the latter will remain to be semiorthogonal only if $g \in O_m$.

Consider the following reparameterization of (48)

$$(\Delta y_t - \gamma) = \alpha\beta y_{t-1} + A_1(\Delta y_{t-1} - \gamma) + \cdots + A_p(\Delta y_{t-p} - \gamma) + HD_\lambda \cdot \varepsilon_t \tag{49}$$

where $\varepsilon_t \mid y_{t-1}, \ldots \sim N(0, \mathrm{I}_m)$, H is an orthogonal matrix with positive elements on the diagonal, $D_\lambda = diag(\lambda_1, \ldots, \lambda_m)$ with $\lambda_m > \cdots > \lambda_1 > 0$. Noting that $\text{cov}(\upsilon_t) \equiv \Sigma = HD_\lambda^2 H'$, there is a 1–1 correspondence between Σ and H, D_λ.

It may be shown that the model (49) is O_m–invariant with the induced action on the parameter space defined as

$$g \circ (\gamma, \alpha, \beta, A_1, \ldots, A_p, H, D_\lambda) := (g\gamma, g\alpha, \beta g', gA_1 g', \ldots, gA_p g', gH, D_\lambda) \tag{50}$$

for all $g \in O_m$. The $G - \Theta$ freeness assumption is always satisfied (proof: $gH = H \Rightarrow gHH' = HH' \Rightarrow g = \mathrm{I}_m$, an identity element in O_m) and $G - \mathcal{Y}$ freeness holds provided that all rows of y are linearly independent (which is the case for almost all y when $T \geq m$). Furthermore

Lemma 8: $t(\gamma, \alpha, \beta, A_1, \ldots, A_p, H, D_\lambda) = (H'\gamma, H'\alpha, \beta H, H'A_1 H, \ldots, H'A_p H, D_\lambda)$ *is maximal* O_m *– invariant.*

Proof: see appendix 7.

The corresponding cross section is

$$\begin{aligned} z(\gamma, \alpha, \beta, A_1, \ldots, A_p, H, D_\lambda) &= s(t(\gamma, \alpha, \beta, A_1, \ldots, A_p, H, D_\lambda)) = \\ &= (H'\gamma, H'\alpha, \beta H, H'A_1 H, \ldots, H'A_p H, \mathrm{I}_m, D_\lambda) \end{aligned} \tag{51}$$

Noting that

$$(\gamma, \alpha, \beta, A_1, \ldots, A_p, H, D_\lambda) = H \circ (H'\gamma, H'\alpha, \beta H, H'A_1 H, \ldots, H'A_p H, \mathrm{I}_m, D_\lambda) \tag{52}$$

and by uniqueness of the orbital decomposition i.e. $\theta = g \circ z(\theta)$, we can easily identify the group element $g \equiv H$.

Let us denote $t_\gamma = H'\gamma$, $t_\alpha = H'\alpha$, $t_\beta = \beta H$, $t_{A_i} = H'A_i H$. Note that $t_\beta t_\beta' = \mathrm{I}_r$ and due to these restrictions it makes no sense to define the Lebesgue measure (dt_β) on $\mathbb{R}^{r \times m}$ (which is equal to zero in such a case). We can do that only for functionally independent elements in t_β. But following James (1954) it is easier to work with differential forms than to choose independent elements in t_β. We can define the (left and right) invariant probability measure on the space $\{t_\beta \in \mathbb{R}^{r \times m} \mid t_\beta t_\beta' = \mathrm{I}_r\}$ which will be denoted as $[dt_\beta]$, see e.g. Muirhead (1982), pp. 69–72. The latter possesses natural interpretation of the "flat" or uniform invariant measure that will satisfy $\int_{t_\beta t_\beta' = \mathrm{I}_r} [dt_\beta] = 1$. Note that $[dt_\beta] = [d\beta]$, where $[d\beta]$ is the invariant probability measure on the space $\{\beta \in \mathbb{R}^{r \times m} \mid \beta\beta' = \mathrm{I}_r\}$, that is $\int_{\beta\beta' = \mathrm{I}_r} [d\beta] = 1$. Importantly using $[dt_\beta] = [d\beta]$ as the "marginal" prior probability measure is essentially equivalent to imposing a uniform prior on the Grassmann manifold, see e.g. Strachan and Inder (2004). Hence we are fully consistent with the recent cointegration literature. Bearing in mind the above discussion the intersubjective orbital measure on $G \times \mathcal{T}$ is given by the following product measure

$$\pi(dg, dt) = \nu_{O_m}(dH)(dt_\gamma)(dt_\alpha)[dt_\beta](dt_{A_1})\ldots(dt_{A_p})(dD_\lambda) \tag{53}$$

Since $\Sigma = HD_\lambda^2 H'$ we can change variables from $H, t_\gamma, t_\alpha, t_\beta, t_{A_1}, \ldots, t_{A_p}, D_\lambda$ to $\gamma, \alpha, \beta, A_1, \ldots, A_p, \Sigma$ with the Jacobian $2^{-m} |\Sigma|^{-\frac{1}{2}} \prod_{i>j} (\chi_i - \chi_j)^{-1}$, where χ_i are eigenvalues of Σ ordered as $\chi_m > \ldots > \chi_1 > 0$. Hence the intersubjective orbital prior on Θ is

$$\pi(d\gamma, d\alpha, d\beta, dA_1, \ldots, dA_p, d\Sigma) \propto |\Sigma|^{-\frac{1}{2}} \prod_{i>j} (\chi_i - \chi_j)^{-1} (d\gamma)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma) \tag{54}$$

The "marginal" prior for Σ i.e. $|\Sigma|^{-\frac{1}{2}} \prod_{i>j} (\chi_i - \chi_j)^{-1}$, is very similar to Bernardo's reference prior for multivariate normal model obtained by Yang and Berger (1994), yet in such a form it probably did not appear in the literature. The Bernardo's reference prior results if instead of the Lebesgue measure (dD_λ) we use the "marginal" prior for D_λ in the form $|D_\lambda|^{-1}(dD_\lambda)$. In such a case

$$\pi(d\gamma, d\alpha, d\beta, dA_1, \ldots, dA_p, d\Sigma) \propto |\Sigma|^{-1} \prod_{i>j} (\chi_i - \chi_j)^{-1} (d\gamma)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma) \tag{55}$$

Then the "marginal" prior for Σ i.e. $|\Sigma|^{-1} \prod_{i>j} (\chi_i - \chi_j)^{-1}$, is exactly the Bernardo's reference prior for multivariate normal model as suggested by Yang and Berger (1994). Thus although the Bernardo's reference prior is not the intersubjective orbital prior for ECM it is the orbital prior in general. In fact since the measure for D_λ comprises the element in the product measure for a cross section, its particular form is irrelevant to avoid the marginalization paradoxes i.e. whatever measure for D_λ we use we are free of the marginalization paradox. Overall, our theory finds some further rationale for the Bernardo's reference prior is the context of ECM. On the other hand, the Jeffreys' prior for Σ i.e. $|\Sigma|^{-\frac{1}{2}(m+1)}$, has no fundamental justification. Interestingly, Ni and Sun (2003,2005) using various evaluation criteria found that the Bernardo's reference prior dominates the Jeffreys' prior in VAR models. Whether such a conclusion is correct in the case of ECM requires serious investigation. Note however that all arguments presented in section VI "speak" in favor of the Bernardo's reference prior.

The algorithms to sample from the full conditional posterior under the prior (54) or (55), were given e.g. by Ni and Sun (2003,2005), Berger and Sun (2007).

Now consider the following traditional specification of ECM

$$\Delta y_t = c + \alpha\beta y_{t-1} + A_1\Delta y_{t-1} + \cdots + A_p\Delta y_{t-p} + \upsilon_t \tag{56}$$

where, as before, $\beta\beta' = \mathrm{I}_m$ and $\upsilon_t \mid y_{t-1},\ldots \sim N(0,\Sigma)$. Replacing υ_t with $HD_\lambda \cdot \varepsilon_t$ it may be shown that the largest group under which the model (56) is invariant and the stabilizer is a compact space is O_m. Hence the model (56) is O_m–invariant with the induced action on the parameter space

$$g \circ (c,\alpha,\beta,A_1,\ldots,A_p,H,D_\lambda) := (gc, g\alpha, \beta g', gA_1g',\ldots,gA_pg', gH, D_\lambda) \tag{57}$$

for all $g \in O_m$. Note that the action on the parameter space is the same as in the case of (49) (which is (50) if you put $c = \gamma$). As a consequence all results and remarks concerning the model (49) apply here (reading $c = \gamma$) and need not be repeated.

On the other hand consider the ECM specification

$$(\Delta y_t - \gamma) = \alpha(\beta y_{t-1} - \mu) + A_1(\Delta y_{t-1} - \gamma) + \cdots + A_p(\Delta y_{t-p} - \gamma) + \upsilon_t \tag{58}$$

where $\beta\beta' = \mathrm{I}_m$ and $\upsilon_t \mid y_{t-1},\ldots \sim N(0,\Sigma)$. This is the general specification of the ECM preferred by Clements and Hendry (1999) since it allows for explicit modeling of the growth rates (i.e. $E(\Delta y_t) = \gamma$) and means of the cointegrating relations (i.e. $E(\beta y_{t-1}) = \mu$). Again one may show that the model is O_m–invariant with the induced action on the parameter space defined as

$$g \circ (\gamma,\mu,\alpha,\beta,A_1,\ldots,A_p,H,D_\lambda) := (g\gamma,\mu,g\alpha,\beta g', gA_1g',\ldots,gA_pg', gH, D_\lambda) \tag{59}$$

for all $g \in O_m$ [12]. Since the model fulfills the $G - \Theta$ freeness assumption (always), we can write the orbital decomposition

[12] O_m is the largest group under which the model (58) is invariant yet the stabilizer is a compact space.

$$(\gamma,\mu,\alpha,\beta,A_1,\ldots,A_p,H,D_\lambda)=H\circ(H'\gamma,\mu,H'\alpha,\beta H,H'A_1H,\ldots,H'A_pH,\mathrm{I}_m,D_\lambda) \quad (60)$$

To save the space we give the ultimate result i.e. intersubjective orbital prior for (58)[13]

$$\pi(d\gamma,d\mu,d\alpha,d\beta,dA_1,\ldots,dA_p,d\Sigma)\propto$$
$$\propto|\Sigma|^{-\frac{1}{2}}\prod\nolimits_{i>j}(\chi_i-\chi_j)^{-1}(d\gamma)(d\mu)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma) \quad (61)$$

where the notation is explained earlier.

We note in passing that the Bernardo's reference prior

$$\pi(d\gamma,d\mu,d\alpha,d\beta,dA_1,\ldots,dA_p,d\Sigma)\propto$$
$$\propto|\Sigma|^{-1}\prod\nolimits_{i>j}(\chi_i-\chi_j)^{-1}(d\gamma)(d\mu)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma) \quad (62)$$

although not the intersubjective orbital prior, is the orbital prior for the model (58).

XII. Conclusion and Summary

The main motivation for this work was to propose the alternative non–informative prior in the context of general time–series models, since both the Jeffreys' and Bernardo's reference prior are not well suited for this purpose. To this end, we exploited the fact that many standard econometric time–series models are invariant under some group of transformations (in short, they are invariant models).

We presented a unified approach to eliciting non–informative or partially non–informative prior for invariant models. We recommended to apply the orbital decomposition on the parameter space. The latter comprises two components: group element and maximal invariant. When dealing with invariant models and there is a need for non–informative or partially non–informative prior we suggested independent joint prior for a group element and a maximal invariant, which we called the orbital prior. Such a theoretical construct was seriously motivated and resemblance with the framework of Chamberlain and Moreira (2009) was indicated.

[13] Although β and γ in (58) are connected by the equation $\beta\gamma=0$ (see e.g. Clements and Hendry (1999), p. 153), the elicitation of the independent prior for β,γ is in line with that suggested by Villani (2009), provided that we project the marginal prior e.g. for γ, down to the subspace $\beta\gamma=0$.

To apply the orbital prior in practice, the original parameterization of a model should be appropriately chosen. It follows that the natural parameterization is motivated by invariance arguments. The reason is that in some parameterizations the crucial assumption (i.e. $G-\Theta$ freeness) may be violated, so as the orbital decomposition does not exist, or the “wrong” parameterization entails difficult sampling from the resultant posterior. This fact may be considered as a drawback of our approach.

Special case of orbital prior is the intersubjective orbital prior, which in fact is the other name for the non–informative prior in invariant models. The intersubjective orbital prior is the prior induced by the product measure of the right invariant Haar measure for a group element and the Lebesgue measure for maximal invariant. Such a choice was justified intuitively and theoretically.

We derived the orbital and intersubjective orbital prior in many specific models including AR, VAR, SVAR and ECM. However the latter list could be broadened by e.g. Linear State–Space , Linear Panel Data or Instrumental Variables models. The invariance of the last two models was noticed by Chamberlain and Moreira (2009) and Chamberlain (2007), respectively. In some sense they exploited this invariance to propose “appropriate” prior setup, which shares some common ground with the framework of the present paper.

In order to balance the orbital prior with other alternatives we should mention that whereas the orbital prior may be used only in models that are invariant under some group of transformations, the Jeffreys’ or Bernardo’s reference prior may be in principle used (also) in other situations.

As a useful form of summary, for reader’s convenience, appendix 8 contains all orbital and intersubjective orbital priors derived for specific models considered in the paper.

REFERENCES:

Akaike, H. (1980), "Likelihood and the Bayes Procedure", in: J.M. Bernardo et al., eds., *Bayesian Statistics*, Valencia University Press.

Andersson, S. (1982), "Distributions of Maximal Invariants Using Quotient Measures", *The Annals of Statistics*, 10, pp. 955–961.

Barndorff–Nielsen, O.E., P. Blaesild and P.S. Eriksen (1989), *Decomposition and Invariance of Measures and Statistical Transformation Models*, Lecture Notes in Statistics, No 58, Springer–Verlag, Berlin.

Basu, D. (1977), "On the Elimination of Nuisance Parameters", *Journal of the American Statistical Association*, 72, pp. 355–366.

Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, *Second edition*, Springer–Verlag, New York.

Berger, J.O. and D. Sun (2006), "Objective Priors for the Bivariate Normal Model With Multivariate Generalizations", *Duke University ISDS Technical Report 2007–06*.

Berger, J.O. and D. Sun (2007), "Objective Bayesian Analysis for the Multivariate Normal Model", in: J.M. Bernardo et al., eds., *Bayesian Statistics 8*, Oxford University Press, Oxford.

Berger, J.O. and D. Sun (2008), "Objective Priors for the Bivariate Normal Model", *The Annals of Statistics*, 36, pp. 963–982.

Berger J.O. and R.L. Wolpert (1988), *The Likelihood Principle* (*second edition*), Institute of Mathematical Statistics, Hayward, California.

Bondar, J.V. (1977), "A Conditional Confidence Principle", *The Annals of Statistics*, 5, pp. 881–891.

Bunke, H. (1975), "Statistical Inference: Fiducial and Structural vs. Likelihood", *Statistics*, 6, pp. 667–676.

Chamberlain, G. (2007), "Decision Theory Applied to an Instrumental Variables Model", *Econometrica*, 75, pp. 609–652.

Chamberlain, G. and M.J. Moreira (2009), "Decision Theory Applied to a Linear Panel Data Model", *Econometrica*, 77, pp. 107–133.

Chang, T. and C. Villegas (1986), "On a Theorem of Stein Relating Bayesian and Classical Inferences in Group Models", *The Canadian Journal of Statistics*, 14, pp. 289–296.

Chang, T. and D. Eaves (1990), "Reference Priors for the Orbit in a Group Model", *The Annals of Statistics*, 18, pp. 1595–1614.

Chao, J.C. and P.C.B. Phillips (1998), "Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model Using the Jeffreys Prior", *Journal of Econometrics*, 87, pp. 49–86.

Clements, M.P. and D.F. Hendry (1999), *Forecasting Non–Stationary Economic Time Series*, MIT Press, Cambridge, Massachusetts.

Datta, G.S. and J.K. Ghosh (1995), "Noninformative Priors for Maximal Invariant Parameter in Group Models", *Test*, 4, pp. 95–114.

Dawid, A.P. (1982), "Intersubjective Statistical Models", in: G. Koch and F. Spizzichino, eds., *Exchangeability in Probability and Statistics*, North–Holland Pub. Co.

Dawid, A.P. (2006), "Invariant Prior Distributions", in: S. Kotz et al., eds. *Encyclopedia of Statistical Sciences*, *Second edition*, *vol. 6*, John Wiley & Sons, New Jersey.

Dawid, A.P. (2007), "Comment" on "Objective Bayesian Analysis for the Multivariate Normal Model", by J.O. Berger and D. Sun, in: J.M. Bernardo et al., eds., *Bayesian Statistics 8*, Oxford University Press, Oxford.

Dawid, A.P., M. Stone and J.V. Zidek (1973), "Marginalization Paradoxes in Bayesian and Structural Inference", *Journal of the Royal Statistical Society, series B*, 35, pp. 189–213.

Eaton, M.L. (1989), *Group Invariance Applications in Statistics*, Regional Conference Series in Probability and Statistics, vol. 1, Institute of Mathematical Statistics, Hayward, California.

Eaton, M.L. (2008), "Dutch Book in Simple Multivariate Normal Prediction: Another Look", in: D. Nolan and T. Speed, eds., *Probability and Statistics*: *Essays in Honor of David A. Freedman,* Institute of Mathematical Statistics, Beachwood, Ohio.

Eaton, M.L. and D.A. Freedman (2004), "Dutch Book Against Some 'Objective' Priors", *Bernoulli*, 10, pp. 861–872.

Eaton, M.L. and W.D. Sudderth (1998), "A New Predictive Distribution for Normal Multivariate Linear Models", *Sankhyā* , *series A*, 60, pp. 363–382.

Eaton, M.L. and W.D. Sudderth (1999), "Consistency and Strong Inconsistency of Group–Invariant Predictive Inferences", *Bernoulli*, 5, pp. 833–854.

Eaton, M.L. and W.D. Sudderth (2001), "Best Invariant Predictive Distributions", in: M. Viana and D. Richards, eds., *Algebraic Methods in Statistics and Probability*, American Mathematical Society, Providence.

Eaton, M.L. and W.D. Sudderth (2002), "Group Invariant Inference and Right Haar Measure", *Journal of Statistical Planning and Inference*, 103, pp. 87–99.

Eaton, M.L. and W.D. Sudderth (2004), "Properties of Right Haar Predictive Inference", *Sankhyā* , 66, pp. 487–512.

Eaton, M.L. and W.D. Sudderth (2010), "Invariance of Posterior Distributions Under Reparametrization", *Sankhyā* , *series A*, 72, pp. 101–118.

Farrell, R.H. (1985), *Multivariate Calculation*, *Use of the Continuous Groups*, Springer–Verlag, New York.

Fraser, D.A.S. (1956), "Sufficient Statistics with Nuisance Parameters", *The Annals of Mathematical Statistics*, 27, pp. 838–842.

Fraser, D.A.S. (1968), *The Structure of Inference*, John Wiley & Sons, New York.

Fraser, D.A.S. (1979), *Inference and Linear Models*, McGraw–Hill, New York.

Geisser, S. (1984), "On Prior Distributions for Binary Trials", *American Statistician*, 38, pp. 244–251.

Hamilton, J.D., D.F. Waggoner and T. Zha (2007), "Normalization in Econometrics", *Econometric Reviews*, 26, pp. 221–253.

Heath, D. and W.D. Sudderth (1978), "On Finitely Additive Priors, Coherence and Extended Admissibility", *The Annals of Statistics*, 6, pp. 333–345.

Helland, I.S. (2010), *Steps Towards a Unified Basis for Scientific Models and Methods*, World Scientific Pub., Singapore.

Hill, B.M. (1988), "*Discussion by Professor Bruce M. Hill*" in, Berger J.O. and R.L. Wolpert, *The Likelihood Principle* (*second edition*), Institute of Mathematical Statistics, Hayward, California.

Hora, R.B. and R.J. Buehler (1966), "Fiducial Theory and Invariant Estimation", *The Annals of Mathematical Statistics*, 37, pp. 643–656.

James, A.T. (1954), "Normal Multivariate Analysis and the Orthogonal Group", *The Annals of Mathematical Statistics*, 25, pp. 40–75.

Jeffreys, H. (1961), *Theory of Probability*, *3–rd edition*, Oxford University Press.

Kleibergen, F. and H.K. van Dijk (1994), "On the Shape of the Likelihood/Posterior in Cointegration Models", *Econometric Theory*, 10, pp. 514–551.

Kocięcki, A. (2011), "Some Remarks on Consistency and Strong Inconsistency of Bayesian Inference", *MPRA Paper 28731*, *University Library of Munich*.

Koop, G. and M.F.J. Steel (1991), "A Comment" on "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends, by P.C.B. Phillips", *Journal of Applied Econometrics*, 6, pp. 365–370.

Koop, G. and M.F.J. Steel (1992), "A Comment" on "On the Development of Reference Priors" by J.O. Berger and J.M. Bernardo, in: J.M. Bernardo et al., eds. *Bayesian Statistics 4*, Clarendon Press, Oxford.

Koop, G., R. Strachan, H.K. van Dijk and M. Villani (2006), "Bayesian Approaches to Cointegration", in T.C. Mills and K. Patterson, eds., *The Palgrave Handbook of Econometrics*, *Vol.* 1: *Econometric Theory*, Palgrave–Macmillan, Basingstoke.

Lane, D.A. (1988), "*Discussion by Professor David A. Lane*" in, Berger J.O. and R.L. Wolpert, *The Likelihood Principle* (*second edition*), Institute of Mathematical Statistics, Hayward, California

Leamer, E.E. (1991), "Comment" on "To Criticize the Critics", *Journal of Applied Econometrics*, 6, pp. 371–373.

Lehmann, E.L. (1986), *Testing Statistical Hypotheses*, *Second edition*, Springer–Verlag, New York.

Lehmann, E.L. and G. Casella (1998), *Theory of Point Estimation*, *2–nd edition*, Springer–Verlag, New York.

Lindley, D. (1979), "A Comment" on "Reference Posterior Distributions for Bayesian Inference" by J.M. Bernardo, *Journal of the Royal Statistical Society*, *series B*, 41, pp. 133–134.

Litterman, R.B. (1986), "Forecasting with Bayesian Vector Autoregressions – Five Years of Experience", *Journal of Business and Economic Statistics,* 4, pp. 25–38.

Lo, A.Y and J. Cabrera (1987), "Bayes Procedures for Rotationally Symmetric Models on the Sphere", *The Annals of Statistics*, 15, pp. 1257–1268.

McCullagh, P. (2002), "What Is a Statistical Model?" with discussion, *The Annals of Statistics*, 30, pp. 1225–1310.

Muirhead, R.J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, New York.

Nachbin, L. (1965), *The Haar Integral*, D. Van Nostrand Company, Princeton, New Jersey.

Ni, S. and D. Sun (2003), "Noninformative Priors and Frequentist Risks of Bayesian Estimators of Vector–Autoregressive Models", *Journal of Econometrics*, 115, pp. 159–197.

Ni, S. and D. Sun (2005), "Bayesian Estimates for Vector Autoregressive Models", *Journal of Business and Economic Statistics*, 23, pp. 105–117.

Phillips, P.C.B. (1991), "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends", *Journal of Applied Econometrics*, 6, pp. 333–364.

Poirier, D.J. (1991), "A Comment" on "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends", *Journal of Applied Econometrics*, 6, pp. 381–386.

Poirier, D.J. (1992), “A Comment” on “On the Development of Reference Priors” by J.O. Berger and J.M. Bernardo, in: J.M. Bernardo et al., eds. *Bayesian Statistics 4*, Clarendon Press, Oxford.

Schotman, P.C. and H.K. van Dijk (1991a), “On Bayesian Routes to Unit Roots”, *Journal of Applied Econometrics*, 6, pp. 387–401.

Schotman, P.C. and H.K. van Dijk (1991b), “A Bayesian Analysis of the Unit Root in Real Exchange Rates”, *Journal of Econometrics*, 49, pp. 195–238.

Severini, T.A., R. Mukerjee and M. Ghosh (2002), “On an Exact Probability Matching Property of Right–Invariant Priors”, *Biometrika*, 89, pp. 952–957.

Sims, C.A. (1988), “Bayesian Skepticism on Unit Root Econometrics”, *Journal of Economic Dynamics and Control*, 12, pp. 463–474.

Sims, C.A. (1991), “Comment by Christopher A. Sims on ‘To Criticize the Critics’, by Peter C.B. Phillips”, *Journal of Applied Econometrics*, 6, pp. 423–434.

Sims, C.A. and T. Zha (1998), “Bayesian Methods for Dynamic Multivariate Models”, *International Economic Review*, 39, pp. 949–968.

Sims, C.A. and T. Zha (1999), “Error Bands for Impulse Responses”, *Econometrica*, 67, pp. 1113–1155.

Stein, C. (1965), “Approximation of Improper Prior Measures by Prior Probability Measures”, in: J. Neyman and L.M. Le Cam, eds., *Bernoulli*, *Bayes*, *Laplace*: *Anniversary Volume*, Springer–Verlag, New York.

Stone, M. (1970), “Necessary and Sufficient Condition for Convergence in Probability to Invariant Posterior Distributions”, *The Annals of Mathematical Statistics*, 41, pp. 1349–1353.

Stone, M. (1976), “Strong Inconsistency from Uniform Priors” with discussion, *Journal of the American Statistical Association*, 71, pp. 114–125.

Strachan, R.W. and B. Inder (2004), “Bayesian Analysis of the Error Correction Model”, *Journal of Econometrics*, 123, pp. 307–325.

Uhlig, H. (1994), “On Jeffreys Prior When Using the Exact Likelihood Function”, *Econometric Theory*, 10, pp. 633–644.

Villani, M. (2005), “Bayesian Reference Analysis of Cointegration”, *Econometric Theory*, 21, pp. 326–357.

Villani, M. (2009), “Steady–State Priors for Vector Autoregressions”, *Journal of Applied Econometrics*, 24, pp. 630–650.

Waggoner, D.F. and T. Zha (2003), “Likelihood Preserving Normalization in Multiple Equation Models”, *Journal of Econometrics*, 114, pp. 329–347.

Wijsman, R. A. (1986), “Global Cross Sections as a Tool for Factorization of Measures and Distribution of Maximal Invariants”, *Sankhyā* , *series A*, 48, pp. 1–42.

Wijsman, R. A. (1990), *Invariant Measures on Groups and Their Use in Statistics*, Institute of Mathematical Statistics Lecture Notes–Monograph Series, vol. 14, Hayward, California.

Yang, R. and J.O. Berger (1994), “Estimation of a Covariance Matrix Using the Reference Prior”, *The Annals of Statistics*, 22, pp. 1195–1211.

Zha, T. (1999), “Block Recursion and Structural Vector Autoregressions”, *Journal of Econometrics*, 90, pp. 291–316.

Zidek, J.V. (1969), "A Representation of Bayes Invariant Procedures in Terms of Haar Measure", *Annals of the Institute of Statistical Mathematics*, 21, pp. 291–308.

Zivot, E. (1994), "A Bayesian Analysis of the Unit Root Hypothesis Within An Unobserved Components Model", *Econometric Theory*, 10, pp. 552–578.

APPENDICES

Appendix 1 (proof of lemma 2):

Let us denote $f^*(g,t) = f(\varphi(g,t))$. Since G acts trivially on the maximal invariant i.e. $\mathcal{T}$, we have $\theta = (g,t) \Rightarrow \overline{g} \circ \theta := (\overline{g}g, t)$. Then

$$\int_\Theta f(\overline{g} \circ \theta)\pi(d\theta) =$$
$$= \int_{G\times\mathcal{T}} f(\overline{g} \circ \varphi(g,t))\nu_G(dg)\lambda_{\mathcal{T}}(dt)$$
$$= \int_{G\times\mathcal{T}} f^*(\overline{g}g,t)\nu_G(dg)\lambda_{\mathcal{T}}(dt)$$
$$= \Delta_l(\overline{g}^{-1}) \cdot \int_{G\times\mathcal{T}} f^*(g,t)\nu_G(dg)\lambda_{\mathcal{T}}(dt) \qquad \text{[see e.g. Nachbin (1965), p. 78]}$$
$$= \Delta_l(\overline{g}^{-1}) \cdot \int_\Theta f(\theta)\pi(d\theta)$$

where $\Delta_l(\cdot)$ is the (left–hand) modulus of G (see e.g. Wijsman (1990), p. 122, Nachbin (1965), p. 78). Since $\Delta_l(\cdot)$ is a continuous function that satisfies $\Delta_l(g_1 g_2) = \Delta_l(g_1)\Delta_l(g_2)$ (e.g. Wijsman (1990), p. 122), it is a multiplier (see section II for the definition of multiplier). Hence $\nu_G \otimes \lambda_{\mathcal{T}}$ is relatively invariant (equality of lines 3 and 4) and so is π (equality of lines 1 and 5).

Appendix 2 (proof of proposition 1):

Let $z(\theta)$ be a given element of a cross section $\mathcal{Z}$ in the orbit Orb_θ. Assume we have chosen another cross section $\mathcal{Z}_1$ which has an element $z_1(\theta)$ in Orb_θ. Since each orbit is a transitive set and the action of G on Θ is free we must have $z_1(\theta) = \overline{g} \circ z(\theta) = \overline{g} \circ s(t(\theta))$, where $\overline{g} \in G$ is unique and fixed. Let us omit in the notation the dependence of z_1, z, t on θ e.g. $z_1 = \overline{g} \circ z$. Then we have two orbital decompositions $\theta = g \circ z$ and $\theta = g \circ z_1$ (it is understood that for given θ, $g \in G$ is not the same in the two decompositions). Note that $z = s(t)$ and $z_1 = \overline{g} \circ s(t) \equiv \overline{s}(t)$ (they both are bijective functions of the same maximal invariant t). Using RC we can write $\theta = g \circ z$ as $\theta = \varphi(g,t)$ and $\theta = g \circ z_1$ as $\theta = \varphi_1(g,t)$. Then for all integrable f and $B \in \mathcal{B}_\Theta$ and using (4) we have

$\int_{\theta \in B} f(\theta)\pi(d\theta) =$

$= \int_{\varphi(g,t)\in B} f(\varphi(g,t))\nu_G(dg)\lambda_T(dt)$ [since φ is bijective $(g,t) \in \varphi^{-1}B \Leftrightarrow \varphi(g,t) \in B$]

$= \int_{g\circ z \in B} f(g \circ z)\nu_G(dg)\lambda_T(dt)$

$= \int_{g\overline{g}^{-1}\circ z_1 \in B} f(g\overline{g}^{-1} \circ z_1)\nu_G(dg)\lambda_T(dt)$ [since $z_1 = \overline{g} \circ z \Leftrightarrow z = \overline{g}^{-1} \circ z_1$]

$= \int_{g\circ z_1 \in B} f(g \circ z_1)\nu_G(dg)\lambda_T(dt)$ [by definition of ν_G, see section II]

$= \int_{\varphi_1(g,t)\in B} f(\varphi_1(g,t))\nu_G(dg)\lambda_T(dt)$

In particular when f is an indicator function of a set $B \in \mathcal{B}_\Theta$ i.e. $f(\varphi_1(g,t)) = \mathbf{1}_B(\varphi_1(g,t))$, $f(\varphi(g,t)) = \mathbf{1}_B(\varphi(g,t))$ and $f(\theta) = \mathbf{1}_B(\theta)$ then $(\nu_G \otimes \lambda_T)\varphi^{-1}(B) = (\nu_G \otimes \lambda_T)\varphi_1^{-1}(B) := \pi(B)$, for any $B \in \mathcal{B}_\Theta$.

Appendix 3 (proof of lemma 4):

Let us rewrite the model G–invariance condition $P_{g\circ\theta}(Y \in gB) = P_\theta(Y \in B)$ as $P_{g\circ\theta}(Y \in B) = P_\theta(Y \in g^{-1}B)$, where $Y = [Y_1,\ldots,Y_T]$. By proving the latter we explicitly derive the induced action of $G = AL_m$ on the parameter space. Denoting y a realization of Y we have

$P_\theta(Y \in g^{-1}B) = \int_{Y\in g^{-1}B} (2\pi)^{-\frac{1}{2}mT} \mid \Lambda\Lambda' \mid^{-\frac{1}{2}T} \times$

$\times \prod_{t=1}^{T} \exp\{-\frac{1}{2}(y_t - c - A_1 y_{t-1} - \cdots - A_p y_{t-p})'(\Lambda\Lambda')^{-1}(y_t - c - A_1 y_{t-1} - \cdots - A_p y_{t-p})\}(dy)$

$= P_\theta(g \circ Y \in B) = \int_{g\circ Y\in B} (2\pi)^{-\frac{1}{2}mT} \mid \Lambda\Lambda' \mid^{-\frac{1}{2}T} \times$

$\times \prod_{t=1}^{T} \exp\{-\frac{1}{2}(g^{-1}g \circ y_t - c - A_1[g^{-1}g \circ y_{t-1}] - \cdots - A_p[g^{-1}g \circ y_{t-p}])'(\Lambda\Lambda')^{-1} \times$

$\times (g^{-1}g \circ y_t - c - A_1[g^{-1}g \circ y_{t-1}] - \cdots - A_p[g^{-1}g \circ y_{t-p}])\}(dy)$

Let us introduce the random variables $Z = [Z_1, \ldots, Z_T]$ defined as $Z = g \circ Y = [wY_1 + k, \ldots, wY_T + k]$. Taking into account the Jacobian $J(y \to z) = |w|^{-T}$ (z is a realization of Z) we get

$P_\theta(Y \in g^{-1}B) = P_\theta(g \circ Y \in B) = P_{\theta^*}(Z \in B) = \int_{Z \in B} (2\pi)^{-\frac{1}{2}mT} \mid (w\Lambda)(w\Lambda)' \mid^{-\frac{1}{2}T} \times$

$\times \prod_{t=1}^{T} \exp\{-\frac{1}{2}(g^{-1} \circ z_t - c - A_1[g^{-1} \circ z_{t-1}] - \cdots - A_p[g^{-1} \circ z_{t-p}])'(\Lambda\Lambda')^{-1} \times$

$(g^{-1} \circ z_t - c - A_1[g^{-1} \circ z_{t-1}] - \cdots - A_p[g^{-1} \circ z_{t-p}])'\}(dz)$

Since $g^{-1} = (w^{-1}, -w^{-1}k)$ it follows $g^{-1} \circ z_{t-i} := w^{-1}z_{t-i} - w^{-1}k$ for $i = 0,1,\ldots,p$. Inserting the latter into the above pdf and rearranging we obtain

$P_{\theta^*}(Z \in B) = \int_{Z \in B} (2\pi)^{-\frac{1}{2}mT} \mid \Lambda^*\Lambda^{*\prime} \mid^{-\frac{1}{2}T} \times$

$\times \prod_{t=1}^{T} \exp\{-\frac{1}{2}(z_t - c^* - A_1^* z_{t-1} - \cdots - A_p^* z_{t-p})'(\Lambda^*\Lambda^{*\prime})^{-1}(z_t - c^* - A_1^* z_{t-1} - \cdots - A_p^* z_{t-p})'\}(dz)$

where $c^* = wc + w(\mathrm{I}_m - \sum_{i=1}^{p} A_i)w^{-1}k$, $A_i^* = wA_iw^{-1}$; for $i = 1,\ldots,p$, and $\Lambda^* = w\Lambda$

Hence as AL_m acts on the sample space the structure of SVAR model is preserved. We must only show that the transformed parameters θ^* conform to some action of AL_m on the parameter space i.e. $\theta^* = g \circ \theta \equiv g \circ (c, A_1, \ldots, A_p, \Lambda)$. To this end we must check that the operation $g \circ (c, A_1, \ldots, A_p, \Lambda)$ is a left group action of AL_m on Θ. First since the identity element in AL_m is $e = (\mathrm{I}_m, 0)$ we get

$(\mathrm{I}_m, 0) \circ (c, A_1, \ldots, A_p, \Lambda) := (\mathrm{I}_m c + \mathrm{I}_m(\mathrm{I}_m - \sum_{i=1}^{p} A_i)\mathrm{I}_m 0, \mathrm{I}_m A_1 \mathrm{I}_m, \ldots, \mathrm{I}_m A_p \mathrm{I}_m, \mathrm{I}_m \Lambda) =$

$= (c, A_1, \ldots, A_p, \Lambda)$

The second defining property of the left group action is that

$g_1(g_2 \circ (c, A_1, \ldots, A_p, \Lambda)) = (g_1 g_2) \circ (c, A_1, \ldots, A_p, \Lambda)$ for every $g_1, g_2 \in AL_m$. Indeed this holds but we omit the proof since it is a routine exercise.

Putting $Z \equiv Y$ in $P_{\theta^* = g \circ \theta}(Z \in B)$ (Z and Y are equivalent symbols for a random variable in $\mathbb{R}^m$ that obeys the SVAR process) we conclude that SVAR model is AL_m–invariant with the action on the sample space $y \mapsto g \circ y := (wy_1 + k, \ldots, wy_T + k)$ and the induced action on the parameter space

$(c, A_1, \ldots, A_p, \Lambda) \mapsto g \circ (c, A_1, \ldots, A_p, \Lambda) := (wc + w(\mathrm{I}_m - \sum_{i=1}^{p} A_i)w^{-1}k, wA_1w^{-1}, \ldots, wA_pw^{-1}, w\Lambda)$

Appendix 4 (proof of lemma 5):

We have to show that $\mathrm{Stab}_y = \{g \in AL_m \mid g \circ y = y\} = \{e\}$ if the rows $[y', \mathbf{1}_T']'$ are linearly independent. Equivalently, since the identity element in AL_m is $(\mathrm{I}_m, 0)$, $\mathrm{Stab}_y = \{g = (w,k), w \in GL_m, k \in \mathbb{R}^m \mid wy + k \cdot \mathbf{1}_T = y\} = \{(\mathrm{I}_m, 0)\}$. We note

$wy + k \cdot \mathbf{1}_T = y \Leftrightarrow \begin{bmatrix} w & k \\ 0_{1\times m} & 1 \end{bmatrix}\begin{bmatrix} y \\ \mathbf{1}_T \end{bmatrix} = \begin{bmatrix} y \\ \mathbf{1}_T \end{bmatrix}$, where $w \in GL_m, k \in \mathbb{R}^m$. Immediate conclusion is that when all rows of $[y', \mathbf{1}_T']'$ are linearly independent then $[y', \mathbf{1}_T']'$ possesses its right inverse which gives

$$\begin{bmatrix} w & k \\ 0_{1\times m} & 1 \end{bmatrix}\begin{bmatrix} y \\ \mathbf{1}_T \end{bmatrix} = \begin{bmatrix} y \\ \mathbf{1}_T \end{bmatrix} \Rightarrow \begin{bmatrix} w & k \\ 0_{1\times m} & 1 \end{bmatrix} = \mathrm{I}_{m+1} \Rightarrow w = \mathrm{I}_m \text{ and } k = 0$$

Appendix 5 (proof of lemma 6):

We shall analyze the stabilizer in each case. Since

$$\mathrm{Stab}_{c,A_1,\ldots,A_p,\Lambda} = \{g \in AL_m \mid g \circ (c, A_1, \ldots, A_p, \Lambda) = (c, A_1, \ldots, A_p, \Lambda)\} =$$
$$= \{w \in GL_m, k \in \mathbb{R}^m \mid (wc + w(\mathrm{I}_m - \textstyle\sum_{i=1}^p A_i)w^{-1}k, wA_1w^{-1}, \ldots, wA_pw^{-1}, w\Lambda) = (c, A_1, \ldots, A_p, \Lambda)\}$$

we get $w\Lambda = \Lambda \Rightarrow w\Lambda\Lambda^{-1} = \Lambda\Lambda^{-1} \Rightarrow w = \mathrm{I}_m$. Inserting $w = \mathrm{I}_m$ into the first parameter component we obtain $c + (\mathrm{I}_m - \sum_{i=1}^p A_i)k = c \Rightarrow (\mathrm{I}_m - \sum_{i=1}^p A_i)k = 0$. If $(\mathrm{I}_m - \sum_{i=1}^p A_i)$ is nonsingular i.e. $\mathrm{rank}(\mathrm{I}_m - \sum_{i=1}^p A_i) = m$, then $k = 0$. Hence $\mathrm{Stab}_{c,A_1,\ldots,A_p,\Lambda}$ is the singleton $g = (\mathrm{I}_m, 0)$, which is the identity element in AL_m and lemma a) follows.

When $\mathrm{I}_m - \sum_{i=1}^p A_i = 0$, the equation $(\mathrm{I}_m - \sum_{i=1}^p A_i)k = 0$ is satisfied for any $k \in \mathbb{R}^m$. In this case the stabilizer comprises all $g \in (\mathrm{I}_m, \mathbb{R}^m)$ which is not equal to identity element in AL_m. Lastly if $1 \le \mathrm{rank}(\mathrm{I}_m - \sum_{i=1}^p A_i) = r < m$, we can write $\mathrm{I}_m - \sum_{i=1}^p A_i = \alpha\beta$, where $\alpha : (m \times r)$ is of full column rank and $\beta : (r \times m)$ is of full row rank. In such a case $(\mathrm{I}_m - \sum_{i=1}^p A_i)k = 0 \Rightarrow \beta k = 0$. Note that β constitute cointegrating vectors. Then $\beta k = 0 \Rightarrow k = 0$ provided that only the vector $k = 0$ lies in the null space of β (recall that the null space of a matrix A is $\mathrm{null}\{A\} = \{x \mid Ax = 0\}$). This is the case if $\dim\{\mathrm{null}\{\beta\}\} = 0$. But since $1 \le \mathrm{rank}(\beta) = r < m$, $\dim\{\mathrm{null}\{\beta\}\} = m - r > 0$. Thus there must be other vectors except $k = 0$ that lie in $\mathrm{null}\{\beta\}$. Overall, $\mathrm{Stab}_{c,A_1,\ldots,A_p,\Lambda}$ in the case $0 \le \mathrm{rank}(\mathrm{I}_m - \sum_{i=1}^p A_i) < m$ can not be equal to $(\mathrm{I}_m, 0)$, the identity element in AL_m. This proves b).

Appendix 6 (proof of lemma 7):

Take two elements in the same orbit e.g. $c, A_1, \ldots, A_p, \Lambda$ and $g \circ (c, A_1, \ldots, A_p, \Lambda) := (wc + w(\mathrm{I}_m - \sum_{i=1}^p A_i)w^{-1}k, wA_1w^{-1}, \ldots, wA_pw^{-1}, w\Lambda)$. By construction $t(c, A_1, \ldots, A_p, \Lambda) = (\Lambda^{-1}A_1\Lambda, \Lambda^{-1}A_2\Lambda, \ldots, \Lambda^{-1}A_p\Lambda)$ and

$$t(wc + w(\mathrm{I}_m - \textstyle\sum_{i=1}^p A_i)w^{-1}k, wA_1w^{-1}, \ldots, wA_pw^{-1}, w\Lambda) = ((w\Lambda)^{-1}wA_1w^{-1}w\Lambda, \ldots, (w\Lambda)^{-1}wA_pw^{-1}w\Lambda) =$$

$= (\Lambda^{-1}A_1\Lambda, \dots, \Lambda^{-1}A_p\Lambda)$. Hence t is $AL_m -$invariant. On the other hand assume $t(c, A_1, \dots, A_p, \Lambda) = t(\overline{c}, \overline{A}_1, \dots, \overline{A}_p, \overline{\Lambda})$, it follows $\Lambda^{-1}A_i\Lambda = \overline{\Lambda}^{-1}\overline{A}_i\overline{\Lambda} \Rightarrow \overline{A}_i = \overline{\Lambda}\Lambda^{-1}A_i(\overline{\Lambda}\Lambda^{-1})^{-1}$. We must decide whether there is a $g \in AL_m$ such that $g \circ (c, A_1, \dots, A_p, \Lambda) = (\overline{c}, \overline{A}_1, \dots, \overline{A}_p, \overline{\Lambda})$. Putting $g = (\overline{\Lambda}\Lambda^{-1}, k_*)$, where $k_* = \overline{\Lambda}\Lambda^{-1}(\mathrm{I}_m - \sum_{i=1}^{p} A_i)^{-1}(\Lambda\overline{\Lambda}^{-1}\overline{c} - c)$, one can check that $g \circ (c, A_1, \dots, A_p, \Lambda) \equiv (\overline{\Lambda}\Lambda^{-1}, k_*) \circ (c, A_1, \dots, A_p, \Lambda) = (\overline{c}, \overline{A}_1, \dots, \overline{A}_p, \overline{\Lambda})$. We conclude that $c, A_1, \dots, A_p, \Lambda$ and $\overline{c}, \overline{A}_1, \dots, \overline{A}_p, \overline{\Lambda}$ lie on the same orbit, which proves maximal $AL_m -$invariance.

Appendix 7 (proof of lemma 8):

Take two elements in the same orbit e.g. $\gamma, \alpha, \beta, A_1, \dots, A_p, H, D_\lambda$ and $g \circ (\gamma, \alpha, \beta, A_1, \dots, A_p, H, D_\lambda) := (g\gamma, g\alpha, \beta g', gA_1g', \dots, gA_pg', gH, D_\lambda)$. By construction $t(\gamma, \alpha, \beta, A_1, \dots, A_p, H, D_\lambda) = (H'\gamma, H'\alpha, \beta H, H'A_1H, \dots, H'A_pH, D_\lambda)$ and

$t(g\gamma, g\alpha, \beta g', gA_1g', \dots, gA_pg', gH, D_\lambda) =$
$= ((gH)'g\gamma, (gH)'g\alpha, \beta g'(gH), (gH)'gA_1g'(gH), \dots, (gH)'gA_pg'(gH), D_\lambda) =$
$= (H'\gamma, H'\alpha, \beta H, H'A_1H, \dots, H'A_pH, D_\lambda)$

Hence t is $O_m -$invariant. On the other hand assume $t(\gamma, \alpha, \beta, A_1, \dots, A_p, H, D_\lambda) = t(\overline{\gamma}, \overline{\alpha}, \overline{\beta}, \overline{A}_1, \dots, \overline{A}_p, \overline{H}, \overline{D}_\lambda)$, it follows $H'\gamma = \overline{H}'\overline{\gamma}$ $\Rightarrow \overline{\gamma} = \overline{H}H'\gamma$. Let us denote $h = \overline{H}H'$. Hence $\overline{\gamma} = h\gamma$. By applying the same procedure to all components in $t(\cdot)$ we have $\overline{\alpha} = h\alpha$, $\overline{\beta} = \beta h'$ and $\overline{A}_i = hA_ih'$. Thus using the definition of group action on the parameter space in our case we obtain $(\overline{\gamma}, \overline{\alpha}, \overline{\beta}, \overline{A}_1, \dots, \overline{A}_p, \overline{H}, \overline{D}_\lambda) = h \circ (\gamma, \alpha, \beta, A_1, \dots, A_p, H, D_\lambda) := (h\gamma, h\alpha, \beta h', hA_1h', \dots, hA_ph', hH, D_\lambda)$ since $h \in O_m$ we conclude that $(\overline{\gamma}, \overline{\alpha}, \overline{\beta}, \overline{A}_1, \dots, \overline{A}_p, \overline{H}, \overline{D}_\lambda)$ and $(\gamma, \alpha, \beta, A_1, \dots, A_p, H, D_\lambda)$ lie on the same orbit, which proves maximal $O_m -$invariance.

Appendix 8 (Orbital priors for all models considered in the paper):

Table 1: AR(p) model

Specification	Intersubjective orbital prior	Orbital prior	Validity
$y_t = c + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \sigma \cdot \varepsilon_t$	$\pi(d\sigma, dc, d\beta_1, \ldots, d\beta_p) \propto \sigma^{-1} \mid 1 - \sum_{i=1}^{p} \beta_i \mid^{-1} (d\sigma)(dc)(d\beta_1)\ldots(d\beta_p)$	$\pi(d\sigma, dc, d\beta_1, \ldots, d\beta_p) \propto \sigma^{-1} \mid 1 - \sum_{i=1}^{p} \beta_i \mid^{-1} (d\sigma)(dc) \cdot \pi(d\beta_1, \ldots, d\beta_p)$	$\sum_{i=1}^{p} \beta_i \neq 1$
$(y_t - \gamma) = \beta_1(y_{t-1} - \gamma) + \cdots + \beta_p(y_{t-p} - \gamma) + \sigma \cdot \varepsilon_t$	$\pi(d\sigma, d\gamma, d\beta_1, \ldots, d\beta_p) \propto \sigma^{-1}(d\sigma)(d\gamma)(d\beta_1)\ldots(d\beta_p)$	$\pi(d\sigma, d\gamma, d\beta_1, \ldots, d\beta_p) \propto \sigma^{-1}(d\sigma)(d\gamma) \cdot \pi(d\beta_1, \ldots, d\beta_p)$	always

$\varepsilon_t \mid y_{t-1}, \ldots \sim N(0,1)$

Table 2: VAR(p) model

Specification	Intersubjective orbital prior	Orbital prior	Validity
$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \upsilon_t$	$\pi(d\Sigma, dc, dA_1, \ldots, dA_p) \propto \mid \det(\mathrm{I}_m - \sum_{i=1}^{p} A_i) \mid^{-1}$ $\prod_{i=1}^{m} \mid \Sigma^{[1\mapsto i]} \mid^{-1} (d\Sigma)(dc)(dA_1)\ldots(dA_p)$	$\pi(d\Sigma, dc, dt_{A_1}, \ldots, dt_{A_p}) \propto \mid \det(\mathrm{I}_m - \sum_{i=1}^{p} t_{A_i}) \mid^{-1}$ $\prod_{i=1}^{m} \mid \Sigma^{[1\mapsto i]} \mid^{-1} (d\Sigma)(dc) \cdot \pi(dt_{A_1}, \ldots, dt_{A_p})$	$\mathrm{rank}(\mathrm{I}_m - \sum_{i=1}^{p} A_i) = m$
$y_t - \gamma = A_1(y_{t-1} - \gamma) + \cdots + A_p(y_{t-p} - \gamma) + \upsilon_t$	$\pi(d\Sigma, d\gamma, dA_1, \ldots, dA_p) \propto \prod_{i=1}^{m} \mid \Sigma^{[1\mapsto i]} \mid^{-1} (d\Sigma)(d\gamma)(A_1)\ldots(A_p)$	$\pi(d\Sigma, d\gamma, dt_{A_1}, \ldots, dt_{A_p}) \propto$ $\prod_{i=1}^{m} \mid \Sigma^{[1\mapsto i]} \mid^{-1} (d\Sigma)(d\gamma) \cdot \pi(dt_{A_1}, \ldots, dt_{A_p})$	always

$\upsilon_t \mid y_{t-1}, \ldots \sim N(0, \Sigma)$, $t_{A_i} = \Lambda^{-1} A_i \Lambda$, for $i = 1, \ldots, p$, where $\Lambda \in LT_m^{+}$ comes from the Choleski decomposition $\Sigma = \Lambda\Lambda'$, $\Sigma^{[1\mapsto i]}$ is a leading principal submatrix of Σ consisting the first i rows and columns of Σ.

Table 3: SVAR(p) model

Specification	Intersubjective orbital prior	Orbital prior	Validity
$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \Lambda\varepsilon_t$	$\pi(d\Lambda, dc, dA_1, \ldots, dA_p) \propto$ $\mid \Lambda \mid^{-m} \mid \det(\mathrm{I}_m - \sum_{i=1}^{p} A_i) \mid^{-1} (d\Lambda)(dc)(dA_1)\ldots(dA_p)$	$\pi(d\Lambda, dc, dt_{A_1}, \ldots, dt_{A_p}) \propto$ $\mid \Lambda \mid^{-m} \mid \det(\mathrm{I}_m - \sum_{i=1}^{p} t_{A_i}) \mid^{-1} (d\Lambda)(dc) \cdot \pi(dt_{A_1}, \ldots, dt_{A_p})$	$\mathrm{rank}(\mathrm{I}_m - \sum_{i=1}^{p} A_i) = m$
$y_t - \gamma = A_1(y_{t-1} - \gamma) + \cdots + A_p(y_{t-p} - \gamma) + \Lambda\varepsilon_t$	$\pi(d\Lambda, d\gamma, dA_1, \ldots, dA_p) \propto \mid \Lambda \mid^{-m} (d\Lambda)(d\gamma)(A_1)\ldots(A_p)$	$\pi(d\Lambda, d\gamma, dt_{A_1}, \ldots, dt_{A_p}) \propto \mid \Lambda \mid^{-m} (d\Lambda)(d\gamma) \cdot \pi(dt_{A_1}, \ldots, dt_{A_p})$	always
$A_0 y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t$	$\pi(dc, dA_0, dA_1, \ldots, dA_p) \propto$ $\mid A_0 \mid^{-m(p+1)} \mid \det(A_0 - \sum_{i=1}^{p} A_i) \mid^{-1} (dc)(dA_0)(dA_1)\ldots(dA_p)$	$\pi(dc, dA_0, d\tilde{t}_{A_1}, \ldots, d\tilde{t}_{A_p}) \propto$ $\mid A_0 \mid^{-(m+1)} \mid \det(\mathrm{I}_m - \sum_{i=1}^{p} \tilde{t}_{A_i}) \mid^{-1} (dc)(dA_0) \cdot \pi(d\tilde{t}_{A_1}, \ldots, d\tilde{t}_{A_p})$	$\mathrm{rank}(A_0 - \sum_{i=1}^{p} A_i) = m$
$A_0(y_t - \gamma) = A_1(y_{t-1} - \gamma) + \cdots + A_p(y_{t-p} - \gamma) + \varepsilon_t$	$\pi(d\gamma, dA_0, dA_1, \ldots, dA_p) \propto \mid A_0 \mid^{-m(p+1)} (d\gamma)(dA_0)(dA_1)\ldots(dA_p)$	$\pi(d\gamma, dA_0, d\tilde{t}_{A_1}, \ldots, d\tilde{t}_{A_p}) \propto$ $\mid A_0 \mid^{-m} (d\gamma)(dA_0) \cdot \pi(d\tilde{t}_{A_1}, \ldots, d\tilde{t}_{A_p})$	always

$\varepsilon_t \mid y_{t-1}, \ldots \sim N(0, \mathrm{I}_m)$, $t_{A_i} = \Lambda^{-1} A_i \Lambda$, $\tilde{t}_{A_i} = A_i A_0^{-1}$; for $i = 1, \ldots, p$, $\Lambda \in GL_m$, $A_0 \in GL_m$.

Table 4: ECM(p)

Specification	Intersubjective orbital prior	Orbital prior	Validity
$\Delta y_t = c + \alpha\beta y_{t-1} + A_1\Delta y_{t-1} + \cdots + A_p\Delta y_{t-p} + \upsilon_t$	$\pi(dc, d\alpha, d\beta, dA_1,\ldots, dA_p, d\Sigma) \propto$ $\mid\Sigma\mid^{-\frac{1}{2}} \prod_{i>j}(\chi_i - \chi_j)^{-1}(dc)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma)$	$\pi(dH, dD_\lambda, dt_c, dt_\alpha, dt_\beta, dt_{A_1},\ldots, dt_{A_p}) \propto$ $\nu_{O_m}(dH)\cdot\pi(dD_\lambda, dt_c, dt_\alpha, dt_\beta, dt_{A_1},\ldots, dt_{A_p})$ In particular (Bernardo's reference prior): $\pi(dc, d\alpha, d\beta, dA_1,\ldots, dA_p, d\Sigma) \propto$ $\mid\Sigma\mid^{-1} \prod_{i>j}(\chi_i - \chi_j)^{-1}(dc)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma)$	always
$(\Delta y_t - \gamma) = \alpha\beta y_{t-1} + A_1(\Delta y_{t-1} - \gamma) + \cdots + A_p(\Delta y_{t-p} - \gamma) + \upsilon_t$	$\pi(d\gamma, d\alpha, d\beta, dA_1,\ldots, dA_p, d\Sigma) \propto$ $\mid\Sigma\mid^{-\frac{1}{2}} \prod_{i>j}(\chi_i - \chi_j)^{-1}(d\gamma)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma)$	$\pi(dH, dD_\lambda, dt_\gamma, dt_\alpha, dt_\beta, dt_{A_1},\ldots, dt_{A_p}) \propto$ $\nu_{O_m}(dH)\cdot\pi(dD_\lambda, dt_\gamma, dt_\alpha, dt_\beta, dt_{A_1},\ldots, dt_{A_p})$ In particular (Bernardo's reference prior): $\pi(d\gamma, d\alpha, d\beta, dA_1,\ldots, dA_p, d\Sigma) \propto$ $\mid\Sigma\mid^{-1} \prod_{i>j}(\chi_i - \chi_j)^{-1}(d\gamma)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma)$	always
$(\Delta y_t - \gamma) = \alpha(\beta y_{t-1} - \mu) + A_1(\Delta y_{t-1} - \gamma) + \cdots + A_p(\Delta y_{t-p} - \gamma) + \upsilon_t$	$\pi(d\gamma, d\mu, d\alpha, d\beta, dA_1,\ldots, dA_p, d\Sigma) \propto$ $\mid\Sigma\mid^{-\frac{1}{2}} \prod_{i>j}(\chi_i - \chi_j)^{-1}(d\gamma)(d\mu)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma)$	$\pi(dH, dD_\lambda, d\mu, dt_\gamma, dt_\alpha, dt_\beta, dt_{A_1},\ldots, dt_{A_p}) \propto$ $\nu_{O_m}(dH)\cdot\pi(dD_\lambda, d\mu, dt_\gamma, dt_\alpha, dt_\beta, dt_{A_1},\ldots, dt_{A_p})$ In particular (Bernardo's reference prior): $\pi(d\gamma, d\mu, d\alpha, d\beta, dA_1,\ldots, dA_p, d\Sigma) \propto$ $\mid\Sigma\mid^{-1} \prod_{i>j}(\chi_i - \chi_j)^{-1}(d\gamma)(d\mu)(d\alpha)[d\beta](dA_1)\ldots(dA_p)(d\Sigma)$	always

$\upsilon_t \mid y_{t-1},\ldots \sim N(0,\Sigma)$, $\beta\beta' = \mathrm{I}_r$, $\chi_m > \ldots > \chi_1 > 0$ are eigenvalues of Σ, $H \in O_m$ and diagonal D_λ comes from the spectral decomposition $\Sigma = HD_\lambda^2 H'$, $t_c = H'c$, $t_\gamma = H'\gamma$, $t_\alpha = H'\alpha$, $t_\beta = \beta H$, $t_{A_i} = H'A_i H$; for $i = 1,\ldots,p$, $[dx]$ is invariant probability measure on the space $\{x \in \mathbb{R}^{r\times m} \mid xx' = \mathrm{I}_r\}$.