NATIONAL BANK OF POLAND
WORKING PAPER
No. 88

Algebraic Theory of Indentification in Parametric Models

Andrzej Kocięcki

Warsaw 2011

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

The author thanks Mateusz Pipeń for useful comments.

Design:

Oliwka s.c.

Layout and print:

NBP Printshop

Published by:

National Bank of Poland
Education and Publishing Department
00-919 Warszawa, 11/21 Świętokrzyska Street
phone: +48 22 653 23 35, fax +48 22 653 13 21

http://www.nbp.pl

Contents

List of appendices 4

Abstract 5

I. Introduction 6

II. Identification from an algebra standpoint 10

III. Survey of basic group theory 13

IV. Relationship between equivalence class and orbit 17

V. Examples 19

VI. Identification of the orbit space: a general view 22

VII. Conditions for identification of the orbit representatives space 25

VIII. Identification of the SEM 27

IX. Concluding remarks 34

Appendices 35

References 40

List of appendices

Appendix 1 (proof of proposition 1) 35
Appendix 2 (proof of lemma 3) 35
Appendix 3 (proof of proposition 2) 36
Appendix 4 (proof of proposition 3) 36
Appendix 5 (proof of proposition 4) 36
Appendix 6 (proof of lemma 4) 37
Appendix 7 (proof of proposition 5) 37
Appendix 8 (proof of proposition 6) 38
Appendix 9 38
Appendix 10 (Proof of lemma 5) 39

Abstract

The paper presents the problem of identification in parametric models from the algebraic point of view. We argue that it is not just another perspective but the proper one. That is using our approach we can see the very nature of the identification problem, which is slightly different than that suggested in the literature. In practice it means that in many models we can unambiguously estimate parameters that have been thought as unidentifiable. This is illustrated in the case of Simultaneous Equations Model (SEM), where our analysis leads to conclusion that existing identification conditions, although correct, are based on the inappropriate premise: only the structural parameters that are in one–to–one correspondence with the reduced form parameters are identified. We will show that this is not true. In fact there are other structural parameters, which are identified, but can not be uniquely recovered from the reduced form parameters. Although we apply our theory only to SEM, it can be used in many standard econometric models.

Keywords: Identification; Group theory; Orbits; Orbit representatives; Simultaneous Equations Model; Maximal Invariant

JEL Classification: C10; C18; C30

1

"*What we learn from our whole discussion and what has indeed become a guiding principle in modern mathematics is this lesson: Whenever you have to do with a structure–endowed entity* Σ *try to determine its group of automorphisms, the group of those element–wise transformations which leave all structural relations undisturbed. You can expect to gain a deep insight into the constitution of* Σ *in this way*", Hermann Weyl (1952), p. 144.

I. Introduction

Assume we have a parametric model. Being consistent with the classical literature on identification, we define a structure as given structural relationships (with all parameters assumed to be known) together with probability distribution for latent variables (with given parameters characterizing this distribution). Thus the formal description of a model is that it is a set of all possible structures. The structural relationships within model are determining relations between endogenous and exogenous variables. Since parameters of the probability distribution of latent variables are the integral part of a model and this probability distribution induces the probability distribution for the endogenous variables we have a first (informal) insight into the identification problem: "*anything is called identifiable that can be determined from a knowledge of the* [probability] *distribution of the endogenous variables*", Koopmans (1953), and "*anything not implied in this distribution is not a possible object of statistical inference*", Koopmans and Reiersøl (1950). However, Koopmans and Hood (1953), p. 126, go further and admit that since the reduced form parameters constitute a unique characterization of the distribution for observations "*they are a useful point of departure in establishing criteria of idenifiability*". The remark of Koopmans and Hood (1953) is so rooted in the econometric practice that today it sounds like an obvious triviality. In fact, the reduced form parameters became not only useful but essentially the only one point of departure to establish identification conditions for underlying structural models[1]. Our main practical contribution is to show that this strategy is not always sound. We argue that there are good reasons to analyze the identification problem in connection with basic structural model (instead of the reduced form). Among these reasons is the fact that reduced form models often lose important information about the structural model, which may be obtained when we scrutinize the structural model. Roughly

[1] This was advocated by Koopmans (1949): "*statistical inference, from observations to economic behavior parameters, can be made in two steps: inference from the observations to the parameters of the assumed joint distribution of the observations, and inference from that distribution to the parameters of the structural equations describing economic behavior. The latter problem of inference, described by the term "identification problem"*".

speaking, we may uniquely estimate more parameters of the underlying structural model than the reduced form model allows for. In other words, the reduced form view may blur the identification problem and taking the right perspective (i.e. structural model) may be rewarded in the sense that there may be more identifiable parameters than the reduced form model is able to produce.

Our view of the identification problem draws on its very nature and is consistent with informal descriptions mentioned in the beginning, provided that we properly understand what the probability distribution of the endogenous variables is. We must realize that the latter is connected with the structural model. Thus even though, the probability distributions (i.e. data sampling distributions) given the structural parameters and reduced form parameters are identical, we can not interchange them indifferently in the stage of identification analysis. Our understanding of the identification problem is this: we have a definite (structural) model which takes a form of the probability distribution for endogenous variables and must check whether the design of the model allows us to estimate all parameters uniquely. Thus if any structure (which is numerically parameterized structural model) within our model may be unambiguously recovered for every data then we are free of identification problems. If this is the case, then whatever criterion for the best structure we adopt, we are sure that all parameters in this structure may be uniquely retrieved.

The above heuristic description of the identification slightly differs from the common one. For example, according to Koopmans and Reiersøl (1950), identification is "*the problem of drawing inferences from the probability distribution of the observed variables to the underlying structure*". Almost identical statement begins the Rothenberg (1971) article. This suggests that there is a true structure which "generates" the probability distribution for observables[2]. In fact, this assumption is also explicitly adopted by Bowden (1973). Seeing in this light, identification conditions are a tool to guarantee that the true structure may be uncovered from the probability distribution for observations. We reject the above interpretation of the identification problem for two reasons. First of all, even if we consider an economic model as a genuine statement about some aspects of economic environment (realist's view), we do know that observations are not produced by some structure within our particular model. Secondly, we are leaning towards the view that economic science (understood as a condensed description of our sense impressions) has only (more or

[2] Koopmans et al. (1950), p. 63, explicitly state that there is a true structure. They use the term "structural equations" to describe "*representation according to economic* [implicitly, true] *structure*". Haavelmo (1944), p. 49, claims that "*we have to start out by an axiom, postulating that every set of observable variables has associated with it one particular "true", but unknown, probability law*" and "*our economic theory is indistinguishable from* (*and may even be equivalent to*) *the statement that the observable variables have the joint probability law*", ibid., p. 88.

less) instrumental character[3]. The model itself is an artificial invention and there is no true, hidden structure to be discovered. Of course we are mildly open to the realists' view since an economic model, being idealization, abstraction and theoretical isolation, can, in principle, capture "*small yet significant truths about the real world*", Mäki (2009). In fact, a model may be true (in some sense) thanks to its idealization and isolation. It is so because partial representations (about small slices of the economic world) may be true about those aspects of the world they are designated to represent, see Mäki (2010)[4]. But the truth–value of economic model is quite different from a view implicit in the citation from Koopmans and Reiersøl (1950).

The position maintained in this paper is that (to paraphrase the frequently cited statement of Kadane (1975)) the identification is an algebraic property of the underlying, structural model[5]. We replaced "likelihood" in the original statement of Kadane (1975) with a structural model. The latter is equivalent to the likelihood (in our framework), yet it emphasizes that we talk about particular presentation of the likelihood in terms of the structural equations (not the reduced form). It turns out that the language of algebra is very useful to describe properly and succinctly the core of identification. To this end, many notions from abstract algebra (particularly, the group theory) are introduced that build a self–consistent picture of the algebraic identification theory in parametric models.

The emergence of modern econometric identification theory is closely connected with the Simultaneous Equations Model (SEM). As a matter of fact, all econometrics textbooks (even those most recent) introduce young economists (and econometricians) to the identification problem on the basis of the SEM example. Thus, it should not be surprising that our theory is also explained with the help of SEM. Although we know why the reduced form SEM is identified, the literature does not answer the question: What does the identification of the reduced form SEM have to do with the identification of the prime object of inference i.e. the structural SEM? As painfully explained by e.g. Marschak (1953), Koopmans (1953), for many

[3] To describe this position most effectively we cite from two influential intrumentalists: "*In reality, the law always contains less than the fact itself, because it does not reproduce the fact as a whole but only in that aspect of it which is important for us, the rest being either intentionally or from necessity omitted*", Mach (1898) p. 193, and a model "*… is not, properly speaking, either true or false, it is, rather, something more or less well selected to stand for the reality it represents, and pictures that reality in a more or less precise, a more or less detailed manner*", Duhem (1962), p. 168.

[4] Theoretical assumptions of models (i.e. purposeful or deliberate falsehoods) by neutralizing various peripheral factors help us to isolate the fundamental relations (mechanism of interest) which are similar to the real relations in reality. Thus for a theory to be true (about isolated major forces, factors, relations etc.) it has to be comprised of the unrealistic assumptions. Similar reasoning is contained in Friedman (1953).

[5] This holds irrespective of whether we take Bayesian or non–Bayesian perspective provided that we define a model in appropriate way. However the present paper confines only to non–Bayesian model.

purposes, the reduced form SEM is useless and it is the structural SEM that preservers all theoretical information[6]. In fact, this is reflected in our position that identification conditions must be worked out for the structural not the reduced form model. We argue that we unnecessarily lose some information about the structural SEM when we rely on the identification of the reduced form SEM. Thus contrary to Koopmans and Hood (1953), we claim that the reduced form model is not so much useful starting point to resolve the identification problem, for there are many equally or more useful starting points. Indeed, we will show that there are many other forms of SEM (except the reduced form) that are also identified.

[6] The same insight inspired the Lucas' critique of the structural SEM. But the point is that, in general, we can not dispense with the structural model. It follows that if the identification problem looks different from the structural and the reduced form perspective, the structural one is appropriate.

2

II. Identification From an Algebra Standpoint

Let Y denote the sample space, which is a set of all $y \in \mathrm{Y}$ attainable by at least one structure within a model. A (parametric) structural model is a set $M = \{p(y,\theta) \mid \theta \in \Theta, y \in \mathrm{M}(\theta) \subseteq \mathrm{Y}\}$, where, without loss of generality, $p(y,\theta)$ is a probability density function with respect to Lebesgue measure on $\mathrm{M}(\theta)$ (i.e. for given θ, $p(y,\theta)$ is thought as a data sampling density), $\mathrm{M}(\theta)$ is a subset of the sample space that is permissible by a given structure $\theta \in \Theta$[7]. For simplicity we assume $\forall\, \theta \in \Theta$, $\mathrm{M}(\theta) = \mathrm{Y}$. For (any) fixed $y \in \mathrm{Y}$, define a function $p : \{y\} \times \Theta \to \mathrm{Im}(\Theta) \subseteq \mathbb{R}^+$, $\theta \mapsto p(y,\theta) \equiv p_y(\theta)$ (where $\mathrm{Im}(\Theta)$ denotes the image of $p_y(\Theta)$). For reference, $p_y(\theta)$ (or simply p_y) will be called the likelihood function. By construction, p_y is a surjective mapping (i.e. onto). We use the standard definition of identification (see Rothenberg (1971)): $\theta \in \Theta$ is identified if and only if (iff) for every $y \in \mathrm{Y}$, $p(y,\theta) = p(y,\overline{\theta}) \Rightarrow \theta = \overline{\theta}$. We find it useful to rewrite this as: $\theta \in \Theta$ is identified iff $p_y(\theta) = p_y(\overline{\theta}) \Rightarrow \theta = \overline{\theta}$. Strictly speaking, the latter is necessary for the original identification condition. However, since $y \in \mathrm{Y}$ is arbitrary, it is "empirically verifiable" that in standard situations there is a full equivalence between the above definitions[8]. Keeping in mind that $p_y(\theta)$ is surjective, it means that θ is identified iff $p_y(\theta)$ is the bijection (one–to–one correspondence) from θ onto $p_y(\theta)$. If θ is not identified then there is at least one other $\overline{\theta} \in \Theta$ ($\overline{\theta} \neq \theta$) such that $p_y(\theta) = p_y(\overline{\theta})$.

The important fact to notice is that any function (not only p_y) gives rise to an equivalence relation on its domain. In particular, the function p_y yields the equivalence relation on Θ by setting $\theta \sim_p \overline{\theta}$ iff $p_y(\theta) = p_y(\overline{\theta})$[9], which is easily recognized as formal description of the concept of observational equivalence used in standard identification theory. In algebra, the equivalence relation "$\sim_p$" is sometimes called the *equivalence kernel* of p_y. Note that we write "$\sim_p$" to emphasize that the equivalence relation is associated with p_y. In fact, "$\sim_p$" induces the equivalence relation on Θ and we say that there is an equivalence relation on Θ determined by p_y. Indeed, for given $\theta \in \Theta$ (and $y \in \mathrm{Y}$) leading to $r = p_y(\theta)$, the equivalence class of the element $\theta \in \Theta$ is the inverse image of $r \in \mathrm{Im}(\Theta)$ under $p_y(\cdot)$ (so called *fiber* of p_y over r) i.e. $p_y^{-1}(r) = \{\overline{\theta} \in \Theta \mid p_y(\overline{\theta}) = r\} = p_y^{-1}(p_y(\theta))$. Importantly, the set of all

---

[7] We characterize our model with the help of density function but not a probability measure for expository purposes. Of course doing this we assume that a density (with respect to Lebesgue measure) exists which is justified in case of many econometric models. The analysis based on probability measures would involve extra technical considerations concerning measurability and instead of the pure group theory we would need the topological group theory. This would make the paper less readable and obscure the main idea.

[8] A good illustrative example is the linear regression model: $\mathrm{y} = X\beta + \mathrm{e}$. Under the condition that X is of full column rank, if the model is identified for one particular y, X then it is identified for any other y, X. The non–identification arises only when X is not of full column rank. But this is excluded *a priori* from our considerations.

[9] One may easily check that this is indeed an equivalence relation which is symmetric, reflexive and transitive.

fibers is a partition of Θ i.e. $\Theta = \cup_{r \in \mathrm{Im}(\Theta)} p_y^{-1}(r)$, where $\{p_y^{-1}(r)\}$ is a collection of nonempty and mutually disjoint subsets of Θ. This means that every $\theta \in \Theta$ belongs to one and only one *fiber*. The equivalence class of the element $\theta \in \Theta$ is defined as $C_\theta = \{\overline{\theta} \in \Theta \mid p_y(\theta) = p_y(\overline{\theta})\} = p_y^{-1}(p_y(\theta))$ i.e. all elements $\overline{\theta} \in \Theta$ that belong to the *fiber* of $p_y(\theta)$ over r. In particular, $\overline{\theta} \in C_\theta$ iff $C_\theta = C_{\overline{\theta}}$. The set of all equivalence classes is known as the quotient set of Θ with respect to $\sim_p$ and will be denoted as $\Theta / \sim_p := \{C_\theta \mid \theta \in \Theta\}$. Let us define the canonical (natural) map $\pi : \Theta \to \Theta / \sim_p$, which sends each element $\theta \in \Theta$ to its equivalence class C_θ with respect to the relation $\sim_p$.

2

Lemma 1: *Let $\sim_p$ be an equivalence relation on Θ determined by p_y. If $\pi : \Theta \to \Theta / \sim_p$ is the canonical map then π is surjective and $\pi(\theta) = \pi(\overline{\theta})$ iff $\theta \sim_p \overline{\theta}$ for $\theta, \overline{\theta} \in \Theta$.*

Proof: see e.g. Bourbaki (1968), p. 115, MacLane and Birkhoff (1993), p. 33, Steinberger (1993), p. 8.

Remark 1: Lemma 1 means that every equivalence relation determined by p_y is the same as the equivalence relation determined by the canonical map with respect to $\sim_p$. Thus $\pi(\theta) = \pi(\overline{\theta}) \Leftrightarrow p_y(\theta) = p_y(\overline{\theta})$, and the problem of identification may be alternatively stated in terms of the canonical map (instead of p_y).

Lemma 2 (canonical decomposition): *Given a surjective map $p_y : \Theta \to \mathrm{Im}(\Theta)$ and the equivalence relation on Θ determined by p_y, i.e. $\sim_p$, we have a unique decomposition $p_y = h \circ \pi$, where $\pi : \Theta \to \Theta / \sim_p$ is the canonical map and $h : \Theta / \sim_p \to \mathrm{Im}(\Theta)$ (which is unique and induced by p_y). Moreover, h is the bijective map.*

Proof: see e.g. Jacobson (1985), pp. 13–14, MacLane and Birkhoff (1993), p. 35, Steinberger (1993), p. 9.

From lemma 1 (see also remark 1) we know that the likelihood function and canonical map with respect to this likelihood function determine the same equivalence relation on Θ. Thus we may consider the original identification problem confining ourselves only to the canonical map. The interesting question is whether there are other functions (except the canonical map) that determine the same equivalence relation as the likelihood function. Moreover, if there are such functions how we can construct them. To this end let us introduce a definition

Definition 1: *Two mappings $f : \Theta \to Y$ and $p : \Theta \to X$ determine the same equivalence relation on Θ, which we denote as $\sim_f \equiv \sim_p$, iff $f(\theta_1) = f(\theta_2) \Leftrightarrow p(\theta_1) = p(\theta_2)$ (or $\theta_1 \sim_f \theta_2 \Leftrightarrow \theta_1 \sim_p \theta_2$); $\theta_1, \theta_2 \in \Theta$.*

Proposition 1: *Two surjective maps* $f:\Theta \to Y$ *and* $p:\Theta \to X$ *determine the same equivalence relation on* Θ *iff there is a bijection* $h: X \to Y$ *such that* $f = h \circ p$. *Moreover,* $h = f \circ s$, *where* s *is a right inverse of* p.

Proof: see appendix 1.

Remark 2: In particular, putting $\pi:\Theta \to \Theta / \sim_p$ in place of $p:\Theta \to X$ and $p_y:\Theta \to \mathrm{Im}(\Theta)$ in place of $f:\Theta \to Y$ in the above proposition we arrive at the canonical decomposition (lemma 2). Then $\sim_p \equiv \sim_\pi$, i.e. lemma 1 follows. However, proposition 1 has more interesting applications and will be crucial in exploring identification problem.

Now we state a definition concerning the core of identification

Definition 2: *If the likelihood function* $p_y:\Theta \to \mathrm{Im}(\Theta)$ *may be uniquely decomposed as* $p_y = h \circ g$, *where* $h: X \to \mathrm{Im}(\Theta)$ *is a bijection and* $g:\Theta \to X$ *is a surjection, then* g *is called the identifying function and* X *is said to be identified. Furthermore, if* $g:\Theta \to X$ *is also a bijection (i.e.* p_y *is a bijection), then* Θ *is said to be identified.*

The definition of the identifying function is exactly the same as in Kadane (1975), for if g is the identifying function then $\forall\, \theta_1, \theta_2 \in \Theta$, $p_y(\theta_1) = p_y(\theta_2) \Leftrightarrow g(\theta_1) = g(\theta_2)$. In other words, $\sim_p \equiv \sim_g$. Note that in definition 2 what is identified is the whole space. In fact, in models where our theory applies there is no need to distinguish between local and global identification. When we say that some space is identified it means that elements of that space are globally identified. For example, a π function from the canonical decomposition is the identifying function and $\Theta / \sim_p$ is identified. Thus a set of all equivalence classes with respect to the relation $\sim_p$ is (globally) identified. However we will show that there are many other identified sets.

III. Survey of Basic Group Theory

This section contains some basic and more specialized facts from group theory (see books on group theory or algebra in our reference list for more details). A group G is a set with a binary operation $G \times G \to G$ that sends (g,h) (for $g,h \in G$) into $g \circ h$, with the following properties: 1) $\forall g \in G$, $e \circ g = g \circ e = g$ (e is an identity element of G) 2) $\forall g \in G$, there exists an inverse element $g^{-1} \in G$ satisfying $g \circ g^{-1} = g^{-1} \circ g = e$ and 3) $\forall g,h,u \in G$, $(g \circ h) \circ u = g \circ (h \circ u)$. "$\circ$" is a rule of composition of elements in G and will be termed as a binary operation (or, in short, an operation). A subset $K \subseteq G$ of a group G is called a subgroup if K with a binary operation from G is also a group. Each group G possesses a trivial subgroup, which is one–element set consisting only an identity element, and an improper subgroup which is G itself. If K is a subgroup of G we denote this fact as $K \le G$ ($K < G$ if K is a proper subset of G). Since elements of G form a set, all known operations on a set apply e.g. union and intersection of sets. In addition, due to group structure of G, we can define one more operation that is fundamental for many notions in group theory. Let H and K be two subsets of elements of a group G (H and K are called complexes), then we can define the operation $HK = \{h \circ k \mid h \in H; k \in K\} \subseteq G$, which is called the product of complexes (or Frobenius product). Implicitly, a product is a group operation in G. Thus HK is the collection of elements in G that are expressible (in at least one way) as a product of an element of H by an element of K. In general, if H, K and D are three subsets of elements (not necessarily groups) then $HK = D$, means that for every $h \in H, k \in K$ there is some element $d \in D$ such that $h \circ k = d$ and vice versa. Thus $HK = D$ means an equality of sets. Note that $K \le G$ iff $KK = K = K^{-1}$ ($K^{-1} = \{k^{-1} \mid k \in K\}$). Also if $K = \{k\}$ (or $H = \{h\}$), we will write $HK = Hk$ (or $HK = hK$). If G is a group then $hG = Gh = G$ iff $h \in G$. In general if $R \subseteq G$ (i.e. R is any subset of elements of a group G), then $RG = GR = G$. The sets like Hk or hK are of special importance. If $H \le K$ and $k \in K$, then $Hk = \{h \circ k \mid h \in H\}$ is called the right coset of H in K. Analogously, $kH = \{k \circ h \mid h \in H\}$ is called the left coset of H in K. The order of a group G is its cardinality and will be denoted as $|G|$, which is a common notation in algebra. We hope that such a notation will not introduce any confusions ($|G|$ has nothing to do with an absolute value or determinant of G). For any $K \le G$, $|G:K|$, i.e. the index of a subgroup K in a group G, is the number of distinct left or right cosets of K in G. Note that $|G:K| = 1$ iff $K = G$ and $|G:\{e\}| = |G|$.

Let G be a group and let Θ be a set. Consider the mapping $G \times \Theta \to \Theta$ which sends (g,θ) into $g \circ \theta$, where "$\circ$" is a binary operation. We say that G acts (or operates) on Θ (or that Θ is a G–set) if 1) $e \circ \theta = \theta$ for all $\theta \in \Theta$ (where e is an identity in G) and 2) $g_1 \circ (g_2 \circ \theta) = (g_1 * g_2) \circ \theta$ for all $g_1, g_2 \in G$ and $\theta \in \Theta$. The

3

binary operation "$*$" is an implicit operation in a group G. In general, the G–set itself may be the Cartesian product i.e. $\Theta = \Theta_1 \times \cdots \times \Theta_k$. In such a case, the action is defined as $g \circ (\theta_1,\ldots,\theta_k) = (g \circ_1 \theta_1,\ldots, g \circ_k \theta_k)$, for all $g \in G$ and $\theta_i \in \Theta_i$. Note that a binary operation may be distinct for every Θ_i. Actually, this is what is essential to develop the theory in our paper.

A group 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 $g \in G$ (which may be written using the set–theoretic equation as $\Theta = G\theta_0$). Of course, when $\Theta = \Theta_1 \times \cdots \times \Theta_k$, the transitivity may be defined in a natural way i.e. G acts transitively on $\Theta_1 \times \cdots \times \Theta_k$ if for each $(\theta_1^{(1)},\ldots,\theta_k^{(1)}),(\theta_1^{(2)},\ldots,\theta_k^{(2)}) \in \Theta_1 \times \cdots \times \Theta_k$, there is a $g \in G$ such that $(\theta_1^{(1)},\ldots,\theta_k^{(1)}) = g \circ (\theta_1^{(2)},\ldots,\theta_k^{(2)}) = (g \circ_1 \theta_1^{(2)},\ldots, g \circ_k \theta_k^{(2)})$. However, in this case the action need not be transitive even if G acts transitively (component–wise) on each Θ_i.

There are two basic notions connected with the theory of G–sets. The first one is the orbit. If G acts on Θ, then the subset $\mathrm{Orb}_\theta = \{g \circ \theta \mid g \in G\} \subseteq \Theta$ (for given $\theta \in \Theta$) is called the orbit of θ with respect to G. The basic facts about orbits are that $\theta \in \mathrm{Orb}_\theta$ (trivially) and $\overline{\theta} \in \mathrm{Orb}_\theta \Leftrightarrow \mathrm{Orb}_\theta = \mathrm{Orb}_{\overline{\theta}}$. Furthermore, $\theta \sim \overline{\theta} \Leftrightarrow \overline{\theta} = g \circ \theta$ (for some $g \in G$) is the equivalence relation. In general, if $\Theta = \Theta_1 \times \cdots \times \Theta_k$ we can generalize the concept of an orbit as $\mathrm{Orb}_{\theta_1,\ldots,\theta_k} = \{g \circ_1 \theta_1,\ldots, g \circ_k \theta_k \mid g \in G\}$. Note that within each orbit the action of the group is transitive (irrespective of whether Θ is the Cartesian product or not). That is why in older literature on groups, the orbits are simply called the transitive sets (or sets of transitivity).

The other notion that occupies central position in the theory of G–sets is the point stabilizer. For any given $\theta \in \Theta$, let us define $\mathrm{Stab}_\theta = \{g \in G \mid \theta = g \circ \theta\} \subseteq G$ and call it the point stabilizer of θ. The fundamental fact is that Stab_θ is a subgroup of G (i.e. $\mathrm{Stab}_\theta \le G$). Analogously as before, we shall extend the notion of point stabilizer to the case when G operates on $\Theta_1 \times \cdots \times \Theta_k$. To this end, let us define $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = \{g \in G \mid \theta_i = g \circ_i \theta_i; \forall i = 1,\ldots,k\}$ and call it the k–point stabilizer. In other words, $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = \mathrm{Stab}_{\theta_1} \cap \mathrm{Stab}_{\theta_2} \cap \ldots \cap \mathrm{Stab}_{\theta_k}$. It is clear that k–point stabilizer is invariant under the permutations of points e.g. $\mathrm{Stab}_{\theta_1,\theta_2} = \mathrm{Stab}_{\theta_2,\theta_1}$. Since $\mathrm{Stab}_{\theta_i} \le G$, for each $i = 1,\ldots,k$, and the intersection of subgroups is also a subgroup, we have $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} \le G$. Furthermore, if at least one $\mathrm{Stab}_{\theta_i} = \{e\}$ [10], then $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = \{e\}$ (since $\mathrm{Stab}_{\theta_1,\ldots,\theta_k}$ is non–empty as it is a group) and $\mathrm{Stab}_{\theta_1} \ge \mathrm{Stab}_{\theta_1,\theta_2} \ge \ldots \ge \mathrm{Stab}_{\theta_1,\ldots,\theta_k}$.

[10] If $\mathrm{Stab}_\theta = \{e\}$, for all $\theta \in \Theta$, we say that G acts freely on Θ.

Some caution should be reserved for the operation of G on k–point stabilizers space. Since $\mathrm{Stab}_{\theta_1,\ldots,\theta_k}$ is the intersection of groups, we must be sure that $g\mathrm{Stab}_{\theta_1,\ldots,\theta_k}$ is a well defined operation[11]. In general, the product of complexes is not well defined in similar situations since we only have an inclusion of the form $K(\mathrm{Stab}_{\theta_1} \cap \ldots \cap \mathrm{Stab}_{\theta_k}) \subseteq (K\,\mathrm{Stab}_{\theta_1}) \cap \ldots \cap (K\,\mathrm{Stab}_{\theta_k})$, where K is a subset of elements of G, see e.g. Scott (1987), p. 16. However, the equality holds when K is a single element from G. Thus $g\,\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = g(\mathrm{Stab}_{\theta_1} \cap \ldots \cap \mathrm{Stab}_{\theta_k}) = (g\mathrm{Stab}_{\theta_1}) \cap \ldots \cap (g\mathrm{Stab}_{\theta_k})$, for all $g \in G$ [12]. Note that in contrast to the operation of G on k–point orbits, since $g \in G$ and $\mathrm{Stab}_{\theta_i} < G$, the operation in $g\mathrm{Stab}_{\theta_i} := \{g \circ h \mid h \in \mathrm{Stab}_{\theta_i}\}$ is the same for each i, and it is an implicit operation in a group G. Furthermore, $g\mathrm{Stab}_{\theta_i}$ is recognized as the left coset of Stab_{θ_i} in G.

There is a well known connection between one–point orbits and stabilizers i.e. the fundamental orbit–stabilizer theorem: $|G : \mathrm{Stab}_\theta| = |\mathrm{Orb}_\theta|$. The following lemma generalizes this theorem in the case of the group action on the Cartesian product and gives a useful result on counting elements in the k–point orbit

Lemma 3:

a) $|G : \mathrm{Stab}_{\theta_1,\theta_2,\ldots,\theta_k}| = |\mathrm{Orb}_{\theta_1,\theta_2,\ldots,\theta_k}|$; (*orbit–k–point–stabilizer theorem*);

b) $|G : \mathrm{Stab}_{\theta_1,\theta_2,\ldots,\theta_k}| = |\mathrm{Orb}_{\theta_1}| \cdot |\mathrm{Stab}_{\theta_1}\theta_2| \cdot |\mathrm{Stab}_{\theta_1,\theta_2}\theta_3| \cdots |\mathrm{Stab}_{\theta_1,\ldots,\theta_{k-1}}\theta_k|$ *or*

$$|\mathrm{Orb}_{\theta_1,\theta_2,\ldots,\theta_k}| = |\mathrm{Orb}_{\theta_1}| \cdot |\mathrm{Stab}_{\theta_1}\theta_2| \cdot |\mathrm{Stab}_{\theta_1,\theta_2}\theta_3| \cdots |\mathrm{Stab}_{\theta_1,\ldots,\theta_{k-1}}\theta_k|;$$

where $\mathrm{Orb}_{\theta_1} \equiv G\theta_1 = \{g \circ \theta_1 \mid g \in G\}$ *and* $\mathrm{Stab}_{\theta_1,\ldots,\theta_{i-1}}\theta_i = \{g \circ \theta_i \mid g \in \mathrm{Stab}_{\theta_1,\ldots,\theta_{i-1}}\}$, *for* $i = 2,\ldots,k$, *is the orbit of* θ_i *with respect to* $\mathrm{Stab}_{\theta_1,\ldots,\theta_{i-1}}$.

Proof: see appendix 2.

In order to develop our theory we need the following definition

Definition 3: *The action of* G *on* $\Theta_1 \times \cdots \times \Theta_k$ *is orbit–regular if to any* $\theta^{(1)} = (\theta_1^{(1)},\ldots,\theta_k^{(1)})$ *and* $\theta^{(2)} = (\theta_1^{(2)},\ldots,\theta_k^{(2)})$ *belonging to the same orbit there corresponds exactly one* $g \in G$ *such that* $\theta^{(2)} = g \circ \theta^{(1)}$.

Note that in definition 3 the orbit is arbitrary, thus it holds for every orbit.

Proposition 2: *The action of* G *on* $\Theta_1 \times \cdots \times \Theta_k$ *is orbit–regular iff* $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = \{e\}$ *for every* $(\theta_1,\ldots,\theta_k) \in \Theta_1 \times \cdots \times \Theta_k$.

Proof: see appendix 3.

---

[11] Since $g \in G$ and the stabilizer is a subgroup of G, $g\mathrm{Stab}_{\theta_1,\ldots,\theta_k} := \{g \circ h \mid h \in \mathrm{Stab}_{\theta_1,\ldots,\theta_k}\}$ "inherits" the operation from a group G.

[12] The proof: $h \in g(\mathrm{Stab}_{\theta_1} \cap \ldots \cap \mathrm{Stab}_{\theta_k}) \Leftrightarrow g^{-1} \circ h \in (\mathrm{Stab}_{\theta_1} \cap \ldots \cap \mathrm{Stab}_{\theta_k}) \Leftrightarrow g^{-1} \circ h \in \mathrm{Stab}_{\theta_i} \Leftrightarrow h \in g\mathrm{Stab}_{\theta_i}$; $\forall i = 1,\ldots,k$; $\Leftrightarrow h \in (g\mathrm{Stab}_{\theta_1}) \cap \ldots \cap (g\mathrm{Stab}_{\theta_k})$.

Proposition 3: *If the action of G on $\Theta_1 \times \cdots \times \Theta_k$ is orbit–regular then* $|\,\mathrm{Orb}_{\theta_1,\ldots,\theta_k}| = |\,G\,|$; *i.e. each orbit has the same cardinality.*

Proof: see appendix 4.

Remark 3: If the action is not orbit–regular then the appropriate formula for counting elements in the orbit is given in lemma 3 b).

From now on, we use the simplified notation: $\theta := (\theta_1,\ldots,\theta_k)$, $\Theta := \Theta_1 \times \cdots \times \Theta_k$. As a consequence, $g \circ \theta := (g \circ_1 \theta_1,\ldots, g \circ_k \theta_k)$, $\mathrm{Orb}_\theta := \mathrm{Orb}_{\theta_1,\ldots,\theta_k}$, $\mathrm{Stab}_\theta := \mathrm{Stab}_{\theta_1,\ldots,\theta_k}$.

IV. Relationship Between Equivalence Class and Orbit

It turns out that there is a close connection between equivalence class and orbit. In fact, as the next section demonstrates, in a number of widely used econometric models, equivalence classes are simply orbits. This has far reaching consequences. We may ignore the characteristics of the likelihood function and concentrate our analytical efforts only on orbit properties. Thus when equivalence class is an orbit the approach to identification based on checking local properties of the likelihood (i.e. information matrix) is rather misplaced.

The following definition, which is fundamental in statistical invariance theory, is also quite important for arguments in the present paper

Definition 4: *A function* $f:\Theta \to Y$ *is said to be invariant under some action of a group* G *on* Θ *(in short,* $G-$*invariant) if* $f(\theta)=f(g\circ\theta)$ *for any* $g\in G$, $\theta\in\Theta$. *Moreover, a function* $f:\Theta\to Y$ *is called maximal* $G-$*invariant if* f *is* $G-$*invariant and for any* $\theta_1,\theta_2\in\Theta$, $f(\theta_1)=f(\theta_2)$ *implies* $\theta_1=g\circ\theta_2$ *for some* $g\in G$ *i.e.* θ_1 *and* θ_2 *lie on the same orbit.*

The next proposition is a key result in this section

Proposition 4: *Suppose the likelihood function* $p_y:\Theta\to \mathrm{Im}(\Theta)$ *is* $G-$*invariant. Then the equivalence class* $C_\theta=\{\overline{\theta}\in\Theta \mid p_y(\theta)=p_y(\overline{\theta})\}$ *is a disjoint union of orbits (one of which is* $\mathrm{Orb}_\theta=\{g\circ\theta \mid g\in G\}$*). If* $p_y:\Theta\to\mathrm{Im}(\Theta)$ *is maximal* $G-$*invariant then* $C_\theta=\mathrm{Orb}_\theta$.

Proof: see appendix 5.

Remark 4: Usually the proof that $C_\theta=\mathrm{Orb}_\theta$ will proceed in two steps. First we shall show that the likelihood is $G-$invariant. Then we use the proof by *reductio ad absurdum*: we assume $C_\theta\neq\mathrm{Orb}_\theta$, which means that C_θ contains at least two orbits, say Orb_{θ_1} and Orb_{θ_2} ($\mathrm{Orb}_{\theta_1}\neq\mathrm{Orb}_{\theta_2}\Rightarrow \mathrm{Orb}_{\theta_1}\cap\mathrm{Orb}_{\theta_2}=\varnothing$), and choose some $\theta_1\in\mathrm{Orb}_{\theta_1}$ and $\theta_2\in\mathrm{Orb}_{\theta_2}$. If $p_y(\theta_1)=p_y(\theta_2)$ implies $\theta_1=g\circ\theta_2$ for some $g\in G$, then $\mathrm{Orb}_{\theta_1}=\mathrm{Orb}_{\theta_2}$. The last statement contradicts $\mathrm{Orb}_{\theta_1}\neq\mathrm{Orb}_{\theta_2}$, therefore $C_\theta=\mathrm{Orb}_\theta$. The issue whether $p_y(\theta_1)=p_y(\theta_2)$ implies $\theta_1=g\circ\theta_2$ may be addressed with several methods. One option is to use theorem 4 in Rothenberg (1971). To this end we should no longer treat the data as given and explicitly introduce the sample space Y. Thus we work with the data sampling density $p(y,\theta)$ indexed by the parameter. Now, if it happens that $h(\theta)=E(f(y))$ for some functions h and f (where E denotes expectation), then $p_y(\theta_1)=p_y(\theta_2)\equiv p(y,\theta_1)=p(y,\theta_2)\Rightarrow \int_Y f(y)p(y,\theta_1)(dy)=\int_Y f(y)p(y,\theta_2)(dy)\Rightarrow h(\theta_1)=h(\theta_2)$. If we manage to prove $h(\theta_1)=h(\theta_2)\Rightarrow\theta_1=g\circ\theta_2$, then $C_\theta = \mathrm{Orb}_\theta$. Usually, $h(\theta_1) = h(\theta_2) \Rightarrow \theta_1 = g \circ \theta_2$ is easier to demonstrate than the original problem (i.e. $p_y(\theta_1) = p_y(\theta_2) \Rightarrow \theta_1 = g \circ \theta_2$). A second alternative is to use some integral transform of the probability density function e.g. characteristic function. That is, we can try to check whether $\phi(\theta_1) = \phi(\theta_2) \Rightarrow \theta_1 = g \circ \theta_2$, where $\phi(\theta)$ is some integral transform of $p_y(\theta)$ e.g. the characteristic function. Again, the latter implication may be less difficult to prove than the original problem.

Remark 5: The well known result is that any G–invariant function must be a function of some maximal G–invariant, see e.g. Lehmann (1986), p. 285. Since the maximal G–invariant takes distinct values on distinct orbits, it provides an orbit index. Thus given the G–invariant likelihood $p_y : \Theta \to \mathrm{Im}(\Theta)$, there exists a k function such that $p_y(\theta) = k(f(\theta))$, where f is maximal G–invariant. Now, if k turns out to be a bijection then $\forall\, \theta_1, \theta_2 \in \Theta$, $p_y(\theta_1) = p_y(\theta_2) \Leftrightarrow f(\theta_1) = f(\theta_2)$ $\Leftrightarrow \theta_1 = g \circ \theta_2$. Thus the question of whether $C_\theta = \mathrm{Orb}_\theta$ leads naturally to the question about the existence of the bijective k mapping between some maximal G–invariant and the likelihood function. It follows that proposition 4 may be weakened to the extent that if $p_y = k \circ f$ is a function of some maximal G–invariant f and k is a bijection, then $C_\theta = \mathrm{Orb}_\theta$.

V. EXAMPLES

This section provides some models in which the equivalence classes are generated by the operation of a group on parameter spaces. We selected models on the basis of two premises. First, to illustrate the fact that nice algebraic structures characterize very popular models and the group theory applies quite naturally and commonly. Secondly, to demonstrate that the concept of group action accommodates quite large specific operations i.e. from an algebraic perspective many apparently distinct models are, in fact, very similar. In all examples the fact that $C_\theta = \mathrm{Orb}_\theta$ may be established by the methods explained in remark 4. Let us begin with a basic, pedagogical example

Example 1 (Artificial but commonly stated to explain the identification problem)

$$y_t = \beta_1 + \beta_2 + \varepsilon_t \tag{1}$$

where y_t is a one–dimensional endogenous variable, $\beta_1, \beta_2 \in \mathbb{R}$ and $\varepsilon_t : (1\times 1) \sim i.i.d.\ N(0,\sigma^2)$. Let $\theta = (\beta_1, \beta_2, \sigma^2) \in \Theta$, then $C_\theta = \mathrm{Orb}_\theta = \{g \circ_1 \beta_1, g \circ_2 \beta_2, g \circ_3 \sigma^2 \mid g \in \mathbb{R}\}$, where $g \circ_1 \beta_1 := \beta_1 + g$, $g \circ_2 \beta_2 := \beta_2 - g$ and $g \circ_3 \sigma^2 := \sigma^2$ [13]. Note that the operating group is real numbers with an addition as the group operation. Such a group will be denoted as $(\mathbb{R},+)$. It is easily verified that $(\mathbb{R},+)$ acts on Θ (by checking two conditions that characterize a group action)[14].

Example 2 (Multiple indicators and multiple causes of a single latent variable)

$$y_t = \beta y_t^* + u_t \tag{2}$$

$$y_t^* = \alpha_1 x_{1t} + \ldots + \alpha_k x_{kt} + \varepsilon_t$$

This model was explicitly introduced by Jöreskog and Goldberger (1975). Let y_t be a one–dimensional endogenous variable, y_t^* is a scalar latent variable, $u_t : (1\times 1) \sim i.i.d.\ N(0,\sigma_u^2)$, $\varepsilon_t : (1\times 1) \sim i.i.d.\ N(0,\sigma_\varepsilon^2)$, $\mathrm{cov}(\varepsilon_t, u_t) = 0$ and $x_{1t},\ldots,x_{kt}$ are exogenous causes. Let $\theta = (\beta, \alpha_1, \ldots, \alpha_k, \sigma_u^2, \sigma_\varepsilon^2) \in \Theta$. Assuming $\beta \neq 0$, then $C_\theta = \mathrm{Orb}_\theta = \{g \circ_1 \beta, g \circ_2 (\alpha_1,\ldots,\alpha_k), g \circ_3 \sigma_\varepsilon^2, g \circ_4 \sigma_u^2 \mid g \in \mathbb{R} \setminus \{0\}\}$, where $g \circ_1 \beta := \frac{1}{g}\beta$, $g \circ_2 (\alpha_1,\ldots,\alpha_k) := (g\alpha_1,\ldots,g\alpha_k)$, $g \circ_3 \sigma_\varepsilon^2 := g^2\sigma_\varepsilon^2$ and $g \circ_4 \sigma_u^2 := \sigma_u^2$. Note that this time, the operating group is real numbers excluding 0 with a group operation of the usual multiplication. Such a group will be denoted as $(\mathbb{R},\times)$. It is easily verified that $(\mathbb{R},\times)$ acts on Θ. Thus C_θ is an orbit of $(\beta, \alpha_1, \ldots, \alpha_k, \sigma_\varepsilon^2, \sigma_u^2)$.

---

[13] The latter action is called the trivial action in which an orbit is one–element subset i.e. $\mathrm{Orb}_\theta = \{\theta\}$ ($\forall \theta \in \Theta$), and we say that θ is a fixed point with respect to the action of a group.

[14] In fact this example is not so far from reality. Similar form of non–identification appears in the following model (see e.g. Prakasa Rao (1992), p. 159). Suppose X_1 and X_2 are independently distributed with the exponential density $p(x) = \lambda_i \exp\{-\lambda_i x\}$ (for $i = 1,2$ and $x > 0$). Then $Y = \max\{X_1, X_2\}$ has density $p(y) = \lambda \exp\{-\lambda y\}$, where $\lambda = \lambda_1 + \lambda_2$. Clearly, $\lambda_1 + g$ and $\lambda_2 - g$ ($g \in \mathbb{R}$) result in the same distribution.

5

Example 3 (Finite Mixture Models (FMM))

$$pdf(y_t) = p_1 \cdot (2\pi)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(y_t - \mu_1)^2\} + p_2 \cdot (2\pi)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(y_t - \mu_2)^2\} \qquad (3)$$

where y_t is a one–dimensional endogenous variable, and $0 \leq p_i \leq 1$, $p_1 + p_2 = 1$, $\mu_1, \mu_2 \in \mathbb{R}$. For obvious reasons we assume $\mu_1 \neq \mu_2$. Let $\theta = (p_1, \mu_1, p_2, \mu_2) \in \Theta$, then $C_\theta = \mathrm{Orb}_\theta = \{g \circ (p_1, \mu_1, p_2, \mu_2) \mid g \in S_2\}$, where S_2 denotes the symmetric group of degree 2 (in general, S_n is the group of permutations which has $n!$ elements i.e. $|S_n| = n!$) and $g \circ (p_1, \mu_1, p_2, \mu_2) := (p_{g(1)}, \mu_{g(1)}, p_{g(2)}, \mu_{g(2)})$. Clearly, S_2 operates on indices by permuting them.

In the remaining examples the operating group will be either the general linear group or its subgroup i.e. an orthogonal group. It is assumed that the group operation is always the usual matrix multiplication. We begin with a model which to a large extent stimulated the formal identification theory

Example 4 (Simultaneous Equations Model (SEM))

$$Ay_t + Bx_t = u_t \qquad (4)$$

where y_t is an $(m \times 1)$ vector of endogenous variables, x_t is a $(k \times 1)$ vector of exogenous variables, $u_t : (m \times 1) \sim \textit{i.i.d.}\ N(0, \Sigma)$, and the coefficients matrices $A : (m \times m)$ (nonsingular) and $B : (m \times k)$. Let $\theta = (A, B, \Sigma) \in \Theta$, then $C_\theta = \mathrm{Orb}_\theta = \{g \circ_1 A, g \circ_2 B, g \circ_3 \Sigma \mid g \in GL_m\}$, where $g \circ_1 A := gA$, $g \circ_2 B := gB$ and $g \circ_3 \Sigma := g\Sigma g'$ and GL_m is the general linear group of $m \times m$ real matrices i.e. $GL_m = \{g \in \mathbb{R}^{m \times m} \mid \det(g) \neq 0\}$. It is easily verified that GL_m operates on Θ (by checking two conditions that characterize a group action). We note that $C_\theta = \mathrm{Orb}_\theta$ was demonstrated by Koopmans et al. (1950), pp. 74–76.

Example 5 (Structural VAR (SVAR))

$$Ay_t + Fy_{(-t)} = \varepsilon_t \qquad (5)$$

where y_t is an $(m \times 1)$ vector of endogenous variables, $y_{(-t)}$ is a $(k \times 1)$ vector of lagged endogenous variables, $\varepsilon_t : (m \times 1) \sim \textit{i.i.d.}\ N(0, \mathrm{I}_m)$, and the coefficients matrices $A : (m \times m)$ (nonsingular) and $F : (m \times k)$. Let $\theta = (A, F) \in \Theta$, then $C_\theta = \mathrm{Orb}_\theta = \{g \circ_1 A, g \circ_2 F \mid g \in O_m\}$, where $g \circ_1 A := gA$ and $g \circ_2 F := gF$ and O_m is the orthogonal group of $m \times m$ matrices i.e. $O_m = \{g \in \mathbb{R}^{m \times m} \mid g'g = gg' = \mathrm{I}_m\} < GL_m$. It is easily verified that O_m operates on Θ.

Example 6 (Error Correction Model (ECM))

$$\triangle y_t + \alpha\beta y_{t-1} + \Gamma \triangle y_{(-t)} = u_t \qquad (6)$$

where $\triangle$ is a difference operator, y_t is an $(m \times 1)$ vector of endogenous variables, $y_{(-t)}$ is a $(k \times 1)$ vector of lagged endogenous variables, $u_t : (m \times 1) \sim \textit{i.i.d.}\ N(0, \Sigma)$, and matrices of coefficients $\Gamma : (m \times k)$, $\alpha : (m \times r)$, $\mathrm{rank}(\alpha) = r$ and $\beta : (r \times m)$,

$\mathrm{rank}(\beta) = r$ (where $r \le m$). Let us decompose $\beta = [\lambda \vdots \eta]$ and assume $\lambda : (r \times r)$ is nonsingular. Let $\theta = (\alpha, \beta, \Gamma, \Sigma) \in \Theta$, then $C_\theta = \mathrm{Orb}_\theta = \{g \circ_1 \alpha, g \circ_2 \beta, g \circ_3 \Gamma, g \circ_4 \Sigma \mid g \in GL_r\}$, where $g \circ_1 \alpha := \alpha g^{-1}$, $g \circ_2 \beta := g\beta$, $g \circ_3 \Gamma := \Gamma$ and $g \circ_4 \Sigma := \Sigma$. It is easily verified that GL_r operates on Θ. Thus, in fact, C_θ is an orbit of $(\alpha, \beta, \Gamma, \Sigma)$. We note that an analogous group operation generates the equivalence class in the observable index models (see Sargent and Sims (1977), Sims (1981)), multivariate autoregressive index models (see Reinsel (1983)) and nested reduced–rank autoregressive models (see Ahn and Reinsel (1988)).

Example 7 (Factor model)

$$y_t = \Lambda f_t + \varepsilon_t \tag{7}$$

where y_t is an $(n \times 1)$ vector of endogenous variables, $\Lambda : (n \times k)$ is a matrix of factor loadings with $\mathrm{rank}(\Lambda) = k \le n$, $f_t : (k \times 1) \sim i.i.d.\ N(0, \Omega)$ (common factors), $\varepsilon_t : (n \times 1) \sim i.i.d.\ N(0, D_\varepsilon)$, where D_ε is diagonal with strictly positive elements. Moreover we assume f_t and ε_t are independent. Let us decompose $\Lambda = [\Psi' \vdots \Upsilon']'$ and assume $\Psi : (k \times k)$ is nonsingular. Let $\theta = (\Lambda, \Omega, \Sigma) \in \Theta$. Provided that D_ε is identifiable[15] then $C_\theta = \mathrm{Orb}_\theta = \{g \circ_1 \Lambda, g \circ_2 \Omega, g \circ_3 \Sigma \mid g \in GL_k\}$, where $g \circ_1 \Lambda := \Lambda g^{-1}$, $g \circ_2 \Omega := g \Omega g'$ and $g \circ_3 \Sigma := \Sigma$. It is easily verified that GL_k operates on Θ. Obviously, if $\Omega = \mathrm{I}_k$, then it is O_k (i.e. orthogonal group) that acts on Θ (in an analogous manner). Hence C_θ is an orbit of $(\Lambda, \Omega, \Sigma)$.

Since the above examples constitute well known models, a G–invariance of the likelihood function in any case is almost self–evident. In general, this may not be the case. However, the necessary and sufficient conditions for the likelihood to be G–invariant may be obtained using results of Brillinger (1963) and Fraser (1967). In addition, Brillinger (1963) gave two methods for constructing the group action (if the likelihood is G–invariant).

5

[15] Sufficient condition for identification of D_ε is given in Anderson and Rubin (1956), theorem 5.1. See also Koopmans and Reiersøl (1950) for other sufficient condition. In general, D_ε can not be identified when k is large in comparison with n. For example using sufficient condition of Anderson and Rubin (1956), D_ε might be unidentified when $k > \frac{1}{2}(n-1)$.

VI. Identification of the Orbit Space: a General View

If we confine ourselves to examples from section V, we may say that orbit space is identified. However, although orbits are point elements in the orbit space they are not those points that we are looking for (actually, orbits are subsets of parameter space). The "points" that we are interested in are the parameter points in the Euclidean spaces. If we manage to isolate one point in every orbit then we obtain an index set for the orbits. Using the group theory terminology, those parameter points may be called the orbit representatives

Definition 5: *Let* Θ *be a* G *–set. A set of orbit representatives is a subset* $\Lambda \subseteq \Theta$ *such that a) if two distinct* $\lambda_1, \lambda_2 \in \Lambda$ *then* $\mathrm{Orb}_{\lambda_1} \cap \mathrm{Orb}_{\lambda_2} = \varnothing$ *and b)* $\Theta = \cup_{\lambda \in \Lambda} \mathrm{Orb}_{\lambda}$.

The idea is that if we take one parameter point (i.e. representative) from each orbit, we obtain a "catalog of unique names" for all orbits. Since the space of orbits forms a partition of the whole parameter space, a "catalog of names" exhausts the whole parameter space. Every parameter in the parameter space is cataloged under a unique "name" and those "names" are written in terms of parameter points. Moreover, there is a one–to–one correspondence between orbits and their representatives (i.e. "names"). We no longer have to work with orbits. Their "names" are sufficient for us. Thus we arrive at the following definition

Definition 6: *An identifying rule is any rule that allows us to choose a unique representative from every orbit.*

Such a rule must guarantee that there is one and only one element in every orbit that obeys the identifying rule. Of course every element from the given orbit may be a representative of that orbit. The point is that we have to provide the rule that allows us to pick some element from an orbit in an unambiguous way. Note that we talk about the situation when there is a rule that allows for a unique choice of the representative but this has nothing to do with imposing any restrictions on the parameter space. An identifying rule is not arbitrary if the model is constructed in such a way that every orbit is in fact a single–element set (e.g. standard linear regression model). Otherwise, an identifying rule is arbitrary and there is necessarily more than one rule. We emphasize that any identifying rule that leads to the choice of a unique representative in every orbit serves the purpose i.e. we can not say that any identifying rule is better than any other (valid) one. However, some identifying rules may be more useful than their alternatives for the particular inferential problem.

Let us formalize the concept of the identifying rule. To this end assume that $C_\theta = \mathrm{Orb}_\theta$. Every identifying rule will materialize through some function $f:\Theta \to \Lambda$ ($\Lambda \subseteq \Theta$ denotes the set of orbit representatives) with two properties: 1) $f(\theta) \in \mathrm{Orb}_\theta$, for each $\theta \in \Theta$ and 2) $f(\theta) = \lambda$, for each $\theta \in \mathrm{Orb}_\theta$, i.e. f is constant on orbits hence G–invariant. Note that since $f(\theta) = \lambda \in \mathrm{Orb}_\theta$ we must have $\lambda = g \circ \theta$ (for some $g \in G$), hence without loss of generality we may take $f(\theta) = g \circ \theta$, for some $g \in G$. Of course f is surjective by construction. For future reference we will simply call $f:\Theta \to \Lambda$, an identifying rule. The following lemma gives various properties of f and the spaces on which it operates

Lemma 4: *Provided that $C_\theta = \mathrm{Orb}_\theta$ and $f:\Theta \to \Lambda$ is an identifying rule, we have*:

a) $p_y:\Theta \to \mathrm{Im}(\Theta)$ *and $f:\Theta \to \Lambda$ determine the same equivalence relation on Θ i.e.* $\sim_p \equiv \sim_f$.

b) *the space of orbit representatives i.e. Λ, is identified and $f:\Theta \to \Lambda$ is the identifying function.*

c) f *is maximal G–invariant.*

Proof: see appendix 6.

The above results suggest that given $C_\theta = \mathrm{Orb}_\theta$, the application of any identifying rule results in the identified space of orbit representatives. Since $\sim_p \equiv \sim_f$ (by lemma 4 a)), if $f:\Theta \to \Lambda$ is a bijection, then $\theta_1 \sim_f \theta_2 \Leftrightarrow \theta_1 \sim_p \theta_2 \Leftrightarrow \theta_1 = \theta_2$ i.e. we arrive at the identification on the primary space of parameters i.e. Θ. The problem is that the mapping f is only surjective. Evidently to identify Θ we should impose some restrictions on the parameter space i.e. to work with the restricted model $\Theta_r \subset \Theta$. Whether we require $f:\Theta \to \Lambda$ to be a bijection depends on the inferential problem. In fact, in some cases identification of Λ will suffice.

Choosing identifying rule amounts to choosing the basic parameterization of the original model. Since the term model is reserved for the family of data sampling densities we introduce the notion of the functional model. The latter is a model as we usually think of when we say a model. Expressions (1) to (7) in section V are functional models. Each functional model may be symbolically denoted as $\varrho(y,\theta,\varepsilon) = 0$, where ε signifies the random error component. The functional model induces uniquely the likelihood function. This suggests the following

Definition 7: *Let $f:\Theta \to \Lambda$ be any identifying rule. Let $\varrho(y,\theta,\varepsilon) = 0$ be the original functional model with its likelihood function $p_y(\theta)$. Then $\varrho(y,f(\theta),\varepsilon) = 0$ is called the canonical form of the original functional model and its likelihood is $p_y(\lambda) \equiv p_y(f(\theta))$.*

6

In line with classical notions in the identification theory, when a model is parameterized with $f:\Theta \to \Lambda$ (i.e. identifying rule) we say that the model is identified (in Λ–parameterization). Hence any canonical form of the functional model is identified.

VII. Conditions for Identification of the Orbit Representatives Space

In the previous section we introduced the notion of the identifying rule. The question of practical interest is when a given rule is identifying. That is we need a condition to check that an application of the given rule will guarantee that in every orbit there is one and only one element that is consistent with this rule. To save the space, we continue to denote $\theta := (\theta_1, \ldots, \theta_k)$, $\Theta := \Theta_1 \times \ldots \times \Theta_k$ with all consequences for actions, stabilizers, orbits etc.

Any identifying rule leads to a statement: if you confine yourself to checking the particular subset of the original parameter space Θ, which was denoted by Λ, it turns out that each orbit contains exactly one element that belongs to Λ. Thus, essentially, any identifying rule is a kind of restriction of the parameter space. However, we emphasize that identifying rule is not a restriction in the strict sense, for to find the orbit representative we do not have to impose any restrictions on Θ at all[16]. Let us denote the subset of the parameter space by $\Theta_r \subset \Theta$ (where the subscript r stands for a quasi–restriction nature of the orbit representatives space). That is we simply put $\Lambda = \Theta_r$. We must ensure that in every orbit there is one and only one element that belongs to Θ_r. If this is the case, the given rule is identifying. Otherwise a rule is not identifying.

Without loss of generality let us focus on any orbit and denote it simply as Orb_θ. Assume that there is some $\theta_r \in \Theta_r$ that belongs to Orb_θ. In such a case we obtain $\mathrm{Orb}_\theta = \mathrm{Orb}_{\theta_r}$ (so as for every $\theta \in \mathrm{Orb}_\theta$ there is a $g \in G$ such that $\theta = g \circ \theta_r$). In fact all elements in Orb_θ are represented as $g \circ \theta_r$ for some $g \in G$. That is as g runs over G, $g \circ \theta_r$ runs over all elements in Orb_{θ_r} ($= \mathrm{Orb}_\theta$). Now it may happen that in Orb_{θ_r} there is at least one other $\overline{\theta}_r \in \Theta_r$. If this is the case then the subset Θ_r is not restrictive enough to guarantee that in every orbit there is only one element that belongs to Θ_r. Let us define $\Theta_r^* = \Theta_r \cap \mathrm{Orb}_{\theta_r}$ (i.e. a set of those elements in the orbit that also belong to Θ_r). By the transitivity of G in Orb_{θ_r}, all elements in Θ_r^* must be represented as $g \circ \theta_r$ for some $g \in G$. Let us define $S = \{g \in G \mid g \circ \theta_r \in \Theta_r^*\}$. We have the basic result on identification

Proposition 5: *Assume the action of G on Θ is orbit–regular, then $|S| = |\Theta_r^*|$. In particular, $|S| = 1 \Leftrightarrow S = \{e\} \Leftrightarrow \Theta_r^* = \Theta_r \cap \mathrm{Orb}_{\theta_r} = \{\theta_r\}$ (where e is the identity element in a group G). If $S = \{e\}$, each orbit may be trivially partitioned into the singletons as* $\mathrm{Orb}_{\theta_r} = \theta_r \cup g_1 \circ \theta_r \cup g_2 \circ \theta_r \cup \cdots = \theta_r \cup (\cup_{g \in G \setminus \{e\}} g \circ \theta_r)$.

Proof: see appendix 7.

---

[16] For example, as will be clear later, the reduced form parameters of SEM are orbit representatives but they do not entail any restrictions on the structural form parameters.

Intuition behind the above proposition is as follows. If Θ_r is chosen so as Θ_r^* is a singleton and because $\theta_r \in \mathrm{Orb}_{\theta_r}$ and $\theta_r \in \Theta_r$, we must have $\Theta_r^* = \{\theta_r\}$. In other words, in every orbit there is one and only one element (i.e. θ_r) that belongs to the subset Θ_r. If the action is orbit–regular then $g \circ \theta_r = \theta_r \Rightarrow g = \{e\}$ and every $g \in G \setminus \{e\}$ (all g's except the identity element) moves θ_r to some $\overline{\theta} \in \mathrm{Orb}_{\theta_r}$ ($\overline{\theta} \neq \theta_r$), i.e. $\overline{\theta} = g \circ \theta_r$. Of course $\overline{\theta} \not\subset \Theta_r^*$ since $\Theta_r^* = \{\theta_r\}$. In such a case, θ_r may serve as the representative for $\mathrm{Orb}_\theta = \mathrm{Orb}_{\theta_r}$ and the given quasi–restriction $\Theta_r \subset \Theta$ may be thought as an identifying rule. Lastly, by the symmetry argument, if the subset $\Theta_r \subset \Theta$ is chosen so as the given Orb_θ contains exactly one element $\theta_r \in \Theta_r$, then every other orbit also contains only one element that belongs to Θ_r (since θ_r was arbitrary).

Proposition 5 gives a criterion to check if the given rule is identifying when orbit–regularity holds. There is also one other useful criterion to check the validity of the identifying rule. As explained in the preceding section, every identifying rule is essentially some function $f : \Theta \to \Lambda$. Lemma 4 c) shows that every identifying rule must be such that f is maximal G–invariant. The question of great importance is whether the converse holds. To this end we have a proposition

Proposition 6: *Assume that $C_\theta = \mathrm{Orb}_\theta$ and $f : \Theta \to X$ is a maximal G–invariant surjective function, where $X \subseteq \Theta$. Then f is an identifying function i.e. X is identified.*

Proof: see appendix 8.

7

Although f from the above proposition is an identifying function it need not be an identifying rule. For instance in our example 4 (SEM), $f_1(A,B,\Sigma) = (A^{-1}B, A^{-1}\Sigma A'^{-1})$ is the identifying function but not the identifying rule because $(A^{-1}B, A^{-1}\Sigma A'^{-1})$ does not lie on the orbit $\mathrm{Orb}_{A,B,\Sigma}$. On the other hand $f_2(A,B,\Sigma) = (\mathrm{I}_m, A^{-1}B, A^{-1}\Sigma A'^{-1})$ is the identifying rule since $f_2(A,B,\Sigma) = A^{-1} \circ (A,B,\Sigma)$, where $A^{-1} \in GL_m$. Taking this remark into account we have a useful defining property of the identifying rule

Corollary 1: *Given $C_\theta = \mathrm{Orb}_\theta$, let $f : \Theta \to X$ be maximal G–invariant, surjective function such that $f(\theta) = x \in \mathrm{Orb}_\theta$, for each $\theta \in \Theta$. Then f is an identifying rule.*

Corollary 1 constitutes an easy working criterion to decide whether the given rule is identifying or not. In fact it is more general and powerful then the criterion in proposition 5.

VIII. Identification of the SEM

In general, an introduction of restrictions into a model may be direct or indirect. The direct method (to introduce restrictions) does not refer to the orbit representatives space, whereas in the indirect method the orbit representative space plays a crucial role. In the direct method we simply choose the restriction $\Theta_r \subset \Theta$ (in the strict sense), which implies that every orbit Orb_{θ_r} ($\theta_r \in \Theta_r$) is a singleton. We do not consider the direct method in our paper. In the indirect method we first provide the identifying rule that leads to choosing some space of orbit representatives i.e. Λ. Given Λ, it is only in the next step when we impose restrictions on Θ. That is we impose restrictions $\Theta_r \subset \Theta$ so as the map $f : \Theta_r \to \Lambda$ is a bijection. An example of the indirect method is an introduction of sufficient number of restrictions in order that the mapping between the reduced form and the structural form parameters in SEM is one–to–one correspondence.

In fact our general strategy to identify the parameter space is a creative elaboration of the existing methodology (which, for reference, will be called the traditional approach). In the traditional approach we apply only one identifying rule: choosing the reduced form parameters which are unique orbit representatives. Our algebraic insight into the identification problem suggests that we can use any identifying rule, because any such a rule allows us to pick a unique element in every orbit. The merits of our approach follow from the fact that, in general, it is the parameters space (not the orbit representatives space) that we are interested in. But the conditions for a bijection between the parameter space and orbit representatives space (i.e. $f : \Theta \to \Lambda$) depend on the algebraic structure of the latter (i.e. Λ). In fact, as will be clear later, there may be less restrictive identifying rules than the traditional identifying rule (i.e. choosing reduced form parameters) in the sense that they require smaller number of restrictions imposed on Θ to have a bijection $f : \Theta \to \Lambda$. To explain this issue carefully there is no better option than to resort to the familiar SEM example. Although our discussion will be confined to SEM, the method proper may be applied to all examples in section V (in general, in all cases when equivalence classes are equal to orbits).

It is instructive to begin with a description of the SEM (our example 4) in terms of the algebraic language that was introduced earlier. To this end let us define the following spaces: $\mathbb{R}_*^{m\times m}$: the space of $m \times m$ nonsingular matrices, $\mathbb{R}^{m\times k}$: the space of $m \times k$ matrices and $\Im_m$: the space of $m \times m$ positive definite symmetric matrices, LT_m^+ (UT_m^+): the space of $m \times m$ lower (upper) triangular matrices with positive diagonal elements, LT_m^1 (UT_m^1): the space of $m \times m$ lower (upper) triangular matrices with ones on the diagonal. Furthermore, O_m and GL_m is the orthogonal and the general linear group, respectively (see section V). Note that LT_m^+, UT_m^+, LT_m^1, UT_m^1 and O_m are proper subgroups of GL_m.

8

As shown in our example 4, the equivalence class of each $(A,B,\Sigma)\in\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m$ is just the orbit of A,B,Σ with respect to GL_m i.e. $C_{A,B,\Sigma}=\mathrm{Orb}_{A,B,\Sigma}$. Thus the quotient set of $\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m$ with respect to $\sim_p$ i.e. $(\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m)/\sim_p$, is just the orbit space. The latter will be denoted as $GL_m\setminus(\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m)$. Hence the canonical map in our case is the function $\pi:(\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m)\to GL_m\setminus(\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m)$ defined as $\pi(A,B,\Sigma)=\mathrm{Orb}_{A,B,\Sigma}:=\{gA,gB,g\Sigma g'\mid g\in GL_m\}$. Moreover, the likelihood function $p_y(\cdot)$ obeys the following canonical decomposition: $p_y(A,B,\Sigma)=h(\pi(A,B,\Sigma))=h(\mathrm{Orb}_{A,B,\Sigma})$, where h is the bijective map. It follows that orbit space is identified. Although GL_m operates transitively both on $\mathbb{R}_*^{m\times m}$ and $\Im_m$, as taken individually, GL_m operates intransitively on $\mathbb{R}^{m\times k}$. In the latter case, the orbit is a subspace of $\mathbb{R}^{m\times k}$ which may be thought as the set of matrices whose every row belongs to the row space of the given $B\in\mathbb{R}^{m\times k}$ (i.e. all linear combinations of the rows of B)[17]. Needless to say, the action of GL_m on $\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m$ is intransitive. On the other hand, the action of GL_m on $\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m$ is orbit–regular, thus each orbit $\mathrm{Orb}_{A,B,\Sigma}$ for $(A,B,\Sigma)\in\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m$ has the same (infinite) order $|GL_m|=\infty$. To demonstrate orbit–regularity note that $\mathrm{Stab}_A=\{g\in G\mid gA=A\}=\{\mathrm{I}_m\}$[18] ($\mathrm{I}_m$ is the identity element in GL_m under the operation of matrix multiplication), but $\mathrm{Stab}_A=\{\mathrm{I}_m\}\Rightarrow\mathrm{Stab}_{A,B,\Sigma}=\{\mathrm{I}_m\}$ (which follows from the properties of stabilizer mentioned in section III).

Before we account for our general approach to identify parameters space, we shall outline the traditional approach with a group–theoretic flavor. The orbit containing any $(A,B,\Sigma)\in\mathbb{R}_*^{m\times m}\times\mathbb{R}^{m\times k}\times\Im_m$ may be written as

$$\begin{aligned}\mathrm{Orb}_{A,B,\Sigma}&:=\{gA,gB,g\Sigma g'\mid g\in GL_m\}=\\&=\{(gA)A^{-1}A,(gA)A^{-1}B,(gA)A^{-1}\Sigma A'^{-1}(gA)'\mid g\in GL_m\}\end{aligned}\tag{8}$$

Since $GL_mA=GL_m$ (because $A\in GL_m$) we have

$$\begin{aligned}\mathrm{Orb}_{A,B,\Sigma}&:=\{gA^{-1}A,gA^{-1}B,gA^{-1}\Sigma A'^{-1}g'\mid g\in GL_m\}=\\&=\{g\,\mathrm{I}_m,gA^{-1}B,gA^{-1}\Sigma A'^{-1}g'\mid g\in GL_m\}\end{aligned}\tag{9}$$

The above equality means that the orbit containing the given structural coefficients (A,B,Σ) also contains the reduced form coefficients $(\mathrm{I}_m,A^{-1}B,A^{-1}\Sigma A'^{-1})$. Thus from section III we know that $\mathrm{Orb}_{A,B,\Sigma}=\mathrm{Orb}_{\mathrm{I}_m,A^{-1}B,A^{-1}\Sigma A'^{-1}}$. Using the notation from section

[17] Such a space will be denoted as $\Re(B)$.

[18] $gA=A\Rightarrow gAA^{-1}=AA^{-1}\Rightarrow g=\mathrm{I}_m\Rightarrow\mathrm{Stab}_A=\{\mathrm{I}_m\}$

VII, let us denote the reduced form representative as $\theta_r = (\mathrm{I}_m, A^{-1}B, A^{-1}\Sigma A'^{-1})$. Evidently, $\Theta_r = \{\mathrm{I}_m\} \times \mathbb{R}^{m\times k} \times \Im_m$. Then $\Theta_r^* = \Theta_r \cap \mathrm{Orb}_{\theta_r} = \{\mathrm{I}_m\} \times \Re(B) \times \Im_m$, where $\Re(B)$ denotes the space of all $(m \times k)$ matrices in which every row belongs to the row space of B. It is easy to check $S = \{g \in GL_m \mid (g\,\mathrm{I}_m, gA^{-1}B, gA^{-1}\Sigma A'^{-1}g') \in \Theta_r^*\} = \{\mathrm{I}_m\}$. It is so because $g\,\mathrm{I}_m = \mathrm{I}_m \Rightarrow g = \mathrm{I}_m$. Thus by proposition 5, an action of any $g \neq \mathrm{I}_m$ moves $(\mathrm{I}_m, A^{-1}B, A^{-1}\Sigma A'^{-1})$ to an element of $\mathrm{Orb}_{A,B,\Sigma}$ that certainly does not have I_m in the first component position. Hence, the reduced form coefficients may serve well as the representative for every orbit i.e. the rule that we choose the reduced form parameters in every orbit is identifying. By lemma 4, the likelihood $p_y(A,B,\Sigma)$ and the identifying (surjective) function $f(A,B,\Sigma) = A^{-1} \circ (A,B,\Sigma) := (\mathrm{I}_m, A^{-1}B, A^{-1}\Sigma A'^{-1})$ determine the same equivalence relation on $\mathbb{R}_*^{m\times m} \times \mathbb{R}^{m\times k} \times \Im_m$ and the space of reduced form parameters is identified. However, it suggests that from the group–theoretic point of view the reduced form parameters are identified because they represent every orbit uniquely. In contrast, the traditional perspective on this issue is that the reduced form coefficients are identified since they are population moments. That is the identification is equalized to the complete characterization of the sampling probability distribution. Our attitude is that this traditional perspective is very narrow and imposes artificial restraints on how we can deal with econometric models to avoid the identification problems. Of course the conditions for identification of the space $\mathbb{R}_*^{m\times m} \times \mathbb{R}^{m\times k} \times \Im_m$ (to have a bijection $(A,B,\Sigma) \mapsto (\mathrm{I}_m, A^{-1}B, A^{-1}\Sigma A'^{-1})$) are well known and constitute the solution of the identification problem within the traditional approach.

Now we are in a position to explain some generalization of the traditional approach. The reduced form representative i.e. $(\mathrm{I}_m, A^{-1}B, A^{-1}\Sigma A'^{-1})$, leads to the following canonical form of the SEM

$$y_t + A^{-1}Bx_t = u_t \tag{10}$$

where $u_t : (m\times 1) \sim i.i.d.\ N(0, A^{-1}\Sigma A'^{-1})$. It was derived using the fact that each A possesses a unique inverse (since $A \in \mathbb{R}_*^{m\times m}$). Interestingly this strategy can be used in a number of variants. For example, by analogy, let us exploit the fact that every $\Sigma \in \Im_m$ also possesses an inverse, which is unique if we decide a priori about its particular structure. For example, using the Choleski decomposition of Σ we have $R^{-1}\Sigma R'^{-1} = \mathrm{I}_m$ (where $\Sigma = RR'$ and $R \in LT_m^+ < GL_m$). Then

$$\begin{aligned}\mathrm{Orb}_{A,B,\Sigma} &:= \{gA, gB, g\Sigma g' \mid g \in GL_m\} = \{gA, gB, gR\,\mathrm{I}_m R'g' \mid g \in GL_m\} = \\ &= \{(gR)R^{-1}A, (gR)R^{-1}B, (gR)\,\mathrm{I}_m (gR)' \mid g \in GL_m\}\end{aligned} \tag{11}$$

From the fact that $GL_m R = GL_m$ (since $R \in LT_m^+ < GL_m$) it follows

8

$$\text{Orb}_{A,B,\Sigma} := \{gR^{-1}A, gR^{-1}B, gg' \mid g \in GL_m\} = \text{Orb}_{R^{-1}A,R^{-1}B,\text{I}_m} \tag{12}$$

We argue that $(R^{-1}A, R^{-1}B, \text{I}_m)$ is a valid orbit representative which results in the following canonical form of the SEM

$$R^{-1}Ay_t + R^{-1}Bx_t = u_t \tag{13}$$

where $u_t : (m \times 1) \sim i.i.d.\ N(0, \text{I}_m)$ and $\Sigma = RR'$. Using notation from section VII, let us denote $\theta_r = (R^{-1}A, R^{-1}B, \text{I}_m)$. It follows that the structure of our representative is $\Theta_r = \{R^{-1}A, R^{-1}B, \text{I}_m \mid R^{-1} \in LT_m^+, A \in \mathbb{R}_*^{m\times m}, B \in \mathbb{R}^{m\times k}\}$. In other words we could write $\Theta_r = \mathbb{R}_*^{m\times m} \times \mathbb{R}^{m\times k} \times \{\text{I}_m\}$ plus an extra condition that $R^{-1} \in LT_m^+$ (because $R \in LT_m^+$). Let us signify this by expanding the parameter space so as $\theta_r = (R^{-1}, R^{-1}A, R^{-1}B, \text{I}_m)$ and $\Theta_r = LT_m^+ \times \mathbb{R}_*^{m\times m} \times \mathbb{R}^{m\times k} \times \{\text{I}_m\}$. Analogously, we can write $\text{Orb}_{\theta_r} = \text{Orb}_{R^{-1},R^{-1}A,R^{-1}B,\text{I}_m} := \{gR^{-1}, gR^{-1}A, gR^{-1}B, gg' \mid g \in GL_m\}$. Note that the action of GL_m on LT_m^+ is implicit in the action of GL_m on $\mathbb{R}_*^{m\times m}$ i.e. $gR^{-1}A$. Ultimately, with this parameter space augmentation we have $\Theta_r^* = \Theta_r \cap \text{Orb}_{\theta_r} = LT_m^+ \times \mathbb{R}_*^{m\times m} \times \Re(B) \times \{\text{I}_m\}$. In appendix 9 we show that $S = \{g \in GL_m \mid (gR^{-1}, gR^{-1}A, gR^{-1}B, gg') \in \Theta_r^*\} = \{\text{I}_m\}$. Consequently by proposition 5, though the orbit $\text{Orb}_{A,B,\Sigma}$ contains $|GL_m|$ elements, there is exactly one element that admits the structure of $(R^{-1}A, R^{-1}B, \text{I}_m)$. This element is just $(R^{-1}A, R^{-1}B, \text{I}_m)$. The latter is equally good representative for the orbit as the reduced form parameters[19].

Alternatively, the proof that $(R^{-1}A, R^{-1}B, \text{I}_m)$ is valid orbit representative may rely on corollary 1. Let us define the function $f(A,B,\Sigma) = \tau(\Sigma) \circ (A,B,\Sigma) := (R^{-1}A, R^{-1}B, \text{I}_m)$. Where $\tau(\Sigma) = R^{-1}$, $\Sigma = RR'$ and $R \in LT_m^+$. Note that $f(A,B,\Sigma) \in \text{Orb}_{A,B,\Sigma}$ and surjection of $f(A,B,\Sigma)$ trivially holds. We must only show that $f(A,B,\Sigma)$ is maximal G–invariant. Assume we have two elements in the orbit $\text{Orb}_{A,B,\Sigma}$: (A,B,Σ) and $(\bar{A},\bar{B},\bar{\Sigma}) = (g_1A, g_1B, g_1\Sigma g_1')$ for some $g_1 \in GL_m$. To prove that $f(A,B,\Sigma)$ is G–invariant we have to show that $f(\bar{A},\bar{B},\bar{\Sigma}) = \tau(\bar{\Sigma}) \circ (\bar{A},\bar{B},\bar{\Sigma}) := (\bar{R}^{-1}\bar{A}, \bar{R}^{-1}\bar{B}, \text{I}_m) = f(A,B,\Sigma)$. Obviously $\tau(\bar{\Sigma}) = \bar{R}^{-1}$ where $\bar{\Sigma} = \bar{R}\bar{R}'$ and $\bar{R} \in LT_m^+$. Since $\Sigma = RR'$ we have $\bar{\Sigma} = \bar{R}\bar{R}' = g_1RR'g_1'$. By Vinograd's theorem it follows $\bar{R} = g_1RQ$ for some $Q \in O_m$. Now it should be noted that $\bar{R} = g_1RQ$ can not hold for arbitrary $g_1 \in GL_m$. To see this write $\bar{R} = g_1RQ$ equivalently as $LT_m^+ \supseteq GL_mLT_m^+W$, where W is some subset of O_m. But

[19] Note that although the representative $(R^{-1}A, R^{-1}B, \text{I}_m)$ has identity matrix in a position attributed to the covariance matrix (and it looks like the SVAR model), it does not mean that the restriction is really imposed. It happens so only by applying the algebraic manipulations on the orbit, but, in fact, the covariance is not restricted at all (i.e. is still "there"). That is provided that R^{-1} in $(R^{-1}A, R^{-1}B, \text{I}_m)$ is unique we can get the covariance back using the Choleski decomposition i.e. $\Sigma = RR'$. Conditions for uniqueness of R^{-1} are given right below.

$GL_m LT_m^+ W = GL_m W = GL_m$, thus we arrive at the contradiction $LT_m^+ \supseteq GL_m$. In fact we can prove that $\bar{R} = g_1 RQ$ for every $g_1 \in GL_m$ implies $g_1 \in LT_m^+$ and $Q = \mathrm{I}_m$. By contradiction assume $g_1 \in LT_m^+$ but $Q \neq \mathrm{I}_m$, then $R^{-1} g_1^{-1} \bar{R} = Q$ and $R^{-1} g_1^{-1} \bar{R} \in LT_m^+$, thus $Q \in (O_m \cap LT_m^+) = \{\mathrm{I}_m\}$ (a contradiction). Similar reasoning applies assuming $Q = \mathrm{I}_m$ but $g_1 \notin LT_m^+$. Lastly when $g_1 \notin LT_m^+$ and $Q \neq \mathrm{I}_m$, then it is easily to show that $RQ \notin LT_m^+$ and for every $g_1 \notin LT_m^+$ the product $g_1 RQ$ can not belong to LT_m^+ (a contradiction). Thus $\bar{R} = g_1 RQ = g_1 R$, where $g_1 \in LT_m^+$. Inserting $\bar{R} = g_1 R$, $\bar{A} = g_1 A$ and $\bar{B} = g_1 B$ into the function f we get $f(\bar{A}, \bar{B}, \bar{\Sigma}) = (\bar{R}^{-1}\bar{A}, \bar{R}^{-1}\bar{B}, \mathrm{I}_m) = (R^{-1} g_1^{-1} g_1 A, R^{-1} g_1^{-1} g_1 B, \mathrm{I}_m) = (R^{-1}A, R^{-1}B, \mathrm{I}_m)$. Hence f is G–invariant. On the other hand, assume $f(\bar{A}, \bar{B}, \bar{\Sigma}) = \tau(\bar{\Sigma}) \circ (\bar{A}, \bar{B}, \bar{\Sigma}) = f(A, B, \Sigma)$ $= \tau(\Sigma) \circ (A, B, \Sigma)$. Then $(\bar{A}, \bar{B}, \bar{\Sigma}) = (\tau(\bar{\Sigma}))^{-1} \circ \tau(\Sigma) \circ (A, B, \Sigma)$. Since $(\tau(\bar{\Sigma}))^{-1} \circ \tau(\Sigma) = \bar{R} R^{-1} \in GL_m$, we showed that $(\bar{A}, \bar{B}, \bar{\Sigma})$ and (A, B, Σ) lie on the same orbit. Therefore, f is maximal G–invariant, which proves that $(R^{-1}A, R^{-1}B, \mathrm{I}_m)$ is the orbit representative.

Of course having the representative $(R^{-1}A, R^{-1}B, \mathrm{I}_m)$ of $\mathrm{Orb}_{A,B,\Sigma}$ we can not obtain uniquely R, A, B. This is analogous to the problem of deriving A, B, Σ from the traditional representative of the orbit i.e. reduced form coefficients. In order to do so we should impose some restrictions on A, B, Σ (which was earlier termed as the indirect method to identify the parameter space). Note however that in order to obtain unique R, A, B from the representative $(R^{-1}A, R^{-1}B, \mathrm{I}_m)$ it suffices to impose only $\frac{1}{2} m(m+1)$ restrictions, for we have the following lemma

Lemma 5: *Assume* $A_1, A_2 \in UT_m^1$; $R_1^{-1}, R_2^{-1} \in LT_m^+$ *and* $B_1, B_2 \in \mathbb{R}^{m \times k}$, *then we have*: $(R_1^{-1}A_1, R_1^{-1}B_1, \mathrm{I}_m) = (R_2^{-1}A_2, R_2^{-1}B_2, \mathrm{I}_m) \Rightarrow R_1 = R_2$, $A_1 = A_2$ *and* $B_1 = B_2$.

Proof: see appendix 10.

Therefore if we restrict $A \in UT_m^1$ then we can uniquely get R, A, B from the representative $(R^{-1}A, R^{-1}B, \mathrm{I}_m)$. Moreover since R matrix is connected with the unique Choleski decomposition of Σ, we obtain the latter as $\Sigma = RR'$.

Remark 6: To obtain unique A, B, Σ from the representative $(R^{-1}A, R^{-1}B, \mathrm{I}_m)$ it suffices to impose only $\frac{1}{2} m(m+1)$ restrictions. In contrast, to make a unique transformation from the reduced form coefficients representative to A, B, Σ we must provide m^2 restrictions (including normalization), which is the necessary condition for identification. Since $\frac{1}{2} m(m+1) < m^2$ (for $m \geq 2$), the gain in using the representative $(R^{-1}A, R^{-1}B, \mathrm{I}_m)$ is evident. It is clear that what the necessary identification condition is depends on the particular orbit representatives structure. As a matter of fact, different structures of orbit representatives may entail different "necessary conditions" for identification (i.e. to make a unique transformation from the representative to the coefficients in a basic space Θ), which may be less demanding than those connected with the traditional approach. Thus the crucial point is that the representative should be chosen purposely: different representatives may be useful in different inferential problems[20].

Indeed there are many other valid orbit representatives for SEM. For example, instead of finding inverses of some parameters matrices, we may simply apply some matrix decompositions to certain parameters matrices. To this end let us use the so–called LU factorization in the context of A matrix i.e. $A = LU$, where $L \in LT_m^+$, $U \in UT_m^1$. Since $A \in \mathbb{R}_*^{m\times m}$ is subject to the unique LU factorization[21], we obtain

$$\begin{aligned}\mathrm{Orb}_{A,B,\Sigma} &= \{gA, gB, g\Sigma g' \mid g \in GL_m\} = \{gLU, gB, g\Sigma g' \mid g \in GL_m\} = \\ &= \{(gL)U, (gL)L^{-1}B, (gL)L^{-1}\Sigma L'^{-1}(gL)' \mid g \in GL_m\}\end{aligned} \tag{14}$$

As before we get $GL_m L = GL_m$ ($L \in LT_m^+ < GL_m$), hence

$$\mathrm{Orb}_{A,B,\Sigma} = \{gU, gL^{-1}B, gL^{-1}\Sigma L'^{-1}g' \mid g \in GL_m\} = \mathrm{Orb}_{U,L^{-1}B,L^{-1}\Sigma L'^{-1}} \tag{15}$$

It is easily to demonstrate that the orbit $\mathrm{Orb}_{A,B,\Sigma}$ contains only one element that preserves the structure of $(U, L^{-1}B, L^{-1}\Sigma L'^{-1})$. By application of the notation from section VII, we have $\theta_r = (U, L^{-1}B, L^{-1}\Sigma L'^{-1})$ and $\Theta_r = \{U, L^{-1}B, L^{-1}\Sigma L'^{-1} \mid U \in UT_m^1, L^{-1} \in LT_m^+, B \in \mathbb{R}^{m\times k}, \Sigma \in \Im_m\}$. The latter may be written as $\Theta_r = UT_m^1 \times \mathbb{R}^{m\times k} \times \Im_m$ together with an extra condition $L^{-1} \in LT_m^+$ (since $L \in LT_m^+$). As before we rewrite our problem with the help of the parameter space augmentation: $\theta_r = (L^{-1}, U, L^{-1}B, L^{-1}\Sigma L'^{-1})$, $\Theta_r = LT_m^+ \times UT_m^1 \times \mathbb{R}^{m\times k} \times \Im_m$ and $\mathrm{Orb}_{\theta_r} = \mathrm{Orb}_{L^{-1},U,L^{-1}B,L^{-1}\Sigma L'^{-1}} = \{gL^{-1}, gU, gL^{-1}B, gL^{-1}\Sigma L'^{-1}g' \mid g \in GL_m\}$. Note that the operation of GL_m on LT_m^+ is implicit in the operation of GL_m on $\mathbb{R}^{m\times k}$ i.e. $gL^{-1}B$. We easily find $\Theta_r^* = \Theta_r \cap \mathrm{Orb}_{\theta_r} = LT_m^+ \times UT_m^1 \times \Re(B) \times \Im_m$. It can be shown that $S = \{g \in GL_m \mid (gL^{-1}, gU, gL^{-1}B, gL^{-1}\Sigma L'^{-1}g') \in \Theta_r^*\} = \{\mathrm{I}_m\}$[22]. Thus $(U, L^{-1}B, L^{-1}\Sigma L'^{-1})$ is an unambiguous representative of the orbit containing (A, B, Σ) (as is the reduced form parameters). Of course to obtain (A, B, Σ) from the orbit representative $(U, L^{-1}B, L^{-1}\Sigma L'^{-1})$ we shall impose some restrictions on the latter. But contrary to

---

[20] In fact, it is possible that in some cases there may exist a structure of the orbit representatives such that Θ is identified without any restrictions imposed on the latter (i.e. $f : \Theta \to \Lambda$ is a bijection).

[21] In fact, for uniqueness of LU decomposition (besides $A \in \mathbb{R}_*^{m\times m}$) we shall also assume that all the leading principal submatrices of A are nonsingular, see e.g. Harville (1997), pp. 227–228. However, this restriction is immaterial for us since our point is only to demonstrate our approach. Actually any other (and less demanding) matrix decomposition applied to A would serve the purpose. For example, the discussion to follow may be based on QR decomposition i.e. $A = QR$, where $Q \in O_m$ and $R \in UT_m^+$.

[22] To this end note that by lemma A1, $gU \in UT_m^1 \Rightarrow g \in UT_m^1$, $gL^{-1} \in LT_m^+ \Rightarrow g \in LT_m^+$ and $UT_m^1 \cap LT_m^+ = \{\mathrm{I}_m\}$.

8

the reduced form parameters representative we shall introduce only $\frac{1}{2}m(m+1)$ restrictions. For example, if $B=[B_1 \vdots B_2]$ and $B_1 \in UT_m^1$, then we can uniquely retrieve U,L,B,Σ (thus A,B,Σ) from the orbit representative $(U,L^{-1}B,L^{-1}\Sigma L'^{-1})$ (the proof proceeds analogously as in lemma 5).

We showed that application of LU decomposition of A and Choleski decomposition of Σ result in the unique orbit representatives. We further demonstrated that these two types of orbit representatives require only $\frac{1}{2}m(m+1)$ restrictions to identify the original parameter space. However, it is evident that those restrictions were "very special". In fact, they conform to some group structure of matrices (e.g. triangular matrices). These kinds of restrictions allow for an easy and direct proof of identifiability. In general, there is a need to develop necessary and sufficient conditions in the situation when restrictions are introduced more freely. That is the analogous results to those that provide the conditions to obtain unique structural parameters from the reduced form parameters subject to some restrictions on the structural parameters. Note however that such conditions are to be specialized for the given structure of orbit representatives. Since our article has been focused on fundamentals of our idea, we postpone a derivation of those results to another study.

IX. Concluding Remarks

We showed that in many econometric models the underlying (observational) equivalence class (i.e. a set of those parameters that imply the same probability distribution for observables) has certain algebraic structure. That is the equivalence class is generated by some group operation on parameter space. We exploited this fact to propose an algebraic insight into the identification problem. Careful analysis provided many fresh results and remarks on the nature of the identification in parametric models. For example we showed that except the reduced form SEM there are many other canonical forms of SEM that are also identified. This fact is potentially of great importance because for many other standard models the same is true. To our knowledge this remark was missed in the econometric literature. We think that an algebraic perspective sheds new light on the true nature of the identification problem.

In the course of our analysis we came up with two criterions to check if the given canonical form of the model is identified. In particular one of them states that in many standard models (like those listed in section V), the given form of the model is canonical (i.e. identified) if the parameterization of this model enjoys the maximal $G-$invariance property. This result may be used in a number of models to design the canonical forms other than the reduced form that are identified. The importance of this follows from the fact that different canonical forms require different necessary conditions to get uniquely the parameters of the original model's form. In the case of SEM we showed that when we use the canonical form other than the reduced form we must provide strictly less restrictions than it is the case when using the reduced form as the canonical form.

Although the leading example was SEM, it is obvious that our approach applies to many other econometric models. Some of them were explicitly mentioned in section V, but the list could be easily broadened.

APPENDICES

Appendix 1 (proof of proposition 1):

The "if" part: $\theta_1 \sim_f \theta_2 \Leftrightarrow f(\theta_1) = f(\theta_2) \Leftrightarrow h(p(\theta_1)) = h(p(\theta_2)) \Leftrightarrow p(\theta_1) = p(\theta_2)$ (h is a bijection) $\Leftrightarrow \theta_1 \sim_p \theta_2$. The "only if" part: We need to show that for any $\theta \in \Theta$ we may construct $f(\theta) = h(p(\theta))$. Choose $\theta \in p^{-1}(x)$, then $f(\theta) = h(x)$. Note that h depends on θ only through x. In order that the mapping is well defined we have to show that for any $\theta_1, \theta_2 \in p^{-1}(x)$ we do have $f(\theta_1) = f(\theta_2)$. But $\theta_1, \theta_2 \in p^{-1}(x)$ means $\theta_1 \sim_p \theta_2$ which is equivalent to $\theta_1 \sim_f \theta_2$ by hypothesis, thus $f(\theta_1) = f(\theta_2)$. We should only demonstrate that h is a bijection. Since f is surjecitve then for each $y \in Y$ there is $\theta \in \Theta$ such that $y = f(\theta) = h(p(\theta))$, thus to each $y \in Y$ there corresponds $p(\theta) \in X$ and h is surjective. To prove that h is also injective, let us choose $\theta_1, \theta_2 \in p^{-1}(x)$ (for some $x \in X$), then $f(\theta_1) = h(p(\theta_1)) = h(p(\theta_2)) = f(\theta_2)$. But since p and f determine the same equivalence relation then $f(\theta_1) = f(\theta_2) \Leftrightarrow p(\theta_1) = p(\theta_2)$, thus $h(p(\theta_1)) = h(p(\theta_2)) \Leftrightarrow p(\theta_1) = p(\theta_2)$. The last assertion proves that h is injective. Lastly, to prove the expression for h, first note that since p is surjective it possesses a right inverse, which we denote as s. Then $h = h \circ (p \circ s) = (h \circ p) \circ s = f \circ s$.

Appendix 2 (proof of lemma 3):

Proof of a) The proof is almost standard and amounts to demonstrating that there is a bijection between the set of left cosets of $\mathrm{Stab}_{\theta_1,\theta_2,\ldots,\theta_k}$ in G and elements in $\mathrm{Orb}_{\theta_1,\theta_2,\ldots,\theta_k}$ i.e. the map $\mu : g\mathrm{Stab}_{\theta_1,\ldots,\theta_k} \mapsto g \circ (\theta_1,\ldots,\theta_k)$ (for all $g \in G$) is a well defined bijection. It is understood that the operation is component–wise i.e. $g \circ (\theta_1,\ldots,\theta_k) = (g \circ_1 \theta_1,\ldots, g \circ_k \theta_k)$ and as argued in section III, $g\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = (g\mathrm{Stab}_{\theta_1}) \cap \ldots \cap (g\mathrm{Stab}_{\theta_k})$. We sketch the proof and focus only on its nonstandard elements. If $g_1, g_2 \in G$ belong to the same left coset of $\mathrm{Stab}_{\theta_1,\ldots,\theta_k}$ in G, then there is a $h \in \mathrm{Stab}_{\theta_1,\ldots,\theta_k}$ such that $g_1 = g_2 \circ h$, thus $g_1 \circ (\theta_1,\ldots,\theta_k) = g_2 \circ h \circ (\theta_1,\ldots,\theta_k) = g_2 \circ (\theta_1,\ldots,\theta_k)$. Hence the map μ is well defined. It is also surjective, which follows from the definition of the map. The map μ is injective since for any $g_1, g_2 \in G$, $g_1 \circ (\theta_1,\ldots,\theta_k) = g_2 \circ (\theta_1,\ldots,\theta_k)$ implies $g_2^{-1} \circ g_1 \circ (\theta_1,\ldots,\theta_k) = (\theta_1,\ldots,\theta_k)$, therefore $g_2^{-1} \circ g_1 \in \mathrm{Stab}_{\theta_1,\ldots,\theta_k}$ and $g_1 \in g_2\mathrm{Stab}_{\theta_1,\ldots,\theta_k}$. The last result implies $g_1 \in g_2\mathrm{Stab}_{\theta_i}$, for all i, and from the properties of cosets we have $g_1\mathrm{Stab}_{\theta_i} = g_2\mathrm{Stab}_{\theta_i}$. Thus ultimately we obtain $g_1\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = g_2\mathrm{Stab}_{\theta_1,\ldots,\theta_k}$.

Proof of b) For the case of two–point stabilizer (i.e. $\mathrm{Stab}_{\theta_1,\theta_2}$), see e.g. Wielandt (1964), proposition 3.3. The proof for the three–point stabilizer is as follows. Using the similar reasoning as in the proof a) we can demonstrate that there is a bijection between the left cosets of $\mathrm{Stab}_{\theta_1,\theta_2}$ in Stab_{θ_1} and the elements in the orbit of θ_2 with respect to Stab_{θ_1} i.e. the map $\mu : g\mathrm{Stab}_{\theta_1,\theta_2} \mapsto g \circ \theta_2$ (for all $g \in \mathrm{Stab}_{\theta_1}$) is a well defined bijection. Therefore $|\mathrm{Stab}_{\theta_1} : \mathrm{Stab}_{\theta_1,\theta_2}| = |\mathrm{Stab}_{\theta_1}\theta_2|$. By the same sort of argument we also obtain $|\mathrm{Stab}_{\theta_1,\theta_2} : \mathrm{Stab}_{\theta_1,\theta_2,\theta_3}| = |\mathrm{Stab}_{\theta_1,\theta_2}\theta_3|$. Since $|\mathrm{Stab}_{\theta_1} : \mathrm{Stab}_{\theta_1,\theta_2,\theta_3}| = |\mathrm{Stab}_{\theta_1} : \mathrm{Stab}_{\theta_1,\theta_2}| \cdot |\mathrm{Stab}_{\theta_1,\theta_2} : \mathrm{Stab}_{\theta_1,\theta_2,\theta_3}|$, see e.g. Hall (1959), p. 12, we get $|\mathrm{Stab}_{\theta_1} : \mathrm{Stab}_{\theta_1,\theta_2,\theta_3}| = |\mathrm{Stab}_{\theta_1}\theta_2| \cdot |\mathrm{Stab}_{\theta_1,\theta_2}\theta_3|$. By the standard orbit–one–point–stabilizer theorem we get $|G : \mathrm{Stab}_{\theta_1}| = |\mathrm{Orb}_{\theta_1}|$. Lastly from $|G : \mathrm{Stab}_{\theta_1}| \cdot |\mathrm{Stab}_{\theta_1} : \mathrm{Stab}_{\theta_1,\theta_2,\theta_3}| = |G : \mathrm{Stab}_{\theta_1,\theta_2,\theta_3}|$ (again see e.g. Hall (1959), p. 12) we arrive at $|G : \mathrm{Stab}_{\theta_1,\theta_2,\theta_3}| = |\mathrm{Orb}_{\theta_1}| \cdot |\mathrm{Stab}_{\theta_1}\theta_2| \cdot |\mathrm{Stab}_{\theta_1,\theta_2}\theta_3|$, which is the formula in the case of three–point stabilizer. The result for general (finite) k–point stabilizer follows by mathematical induction.

Appendix 3 (proof of proposition 2):

Setting $\theta = (\theta_1, \ldots, \theta_k) = \theta^{(1)} = \theta^{(2)}$ in definition 3, we have $\theta = g \circ \theta$ i.e. $g \in \mathrm{Stab}_{\theta_1,\ldots,\theta_k}$. One particular g that solves the equation is e, and from definition of orbit–regularity there is only one such a g, thus $g = e$ i.e. $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = \{e\}$. On the other hand, let us choose any $\theta = (\theta_1, \ldots, \theta_k) \in \mathrm{Orb}_{\theta_1,\ldots,\theta_k}$ then $g_2 \circ \theta = g_1 \circ \theta \Leftrightarrow \theta = g_2^{-1} \circ g_1 \circ \theta \Leftrightarrow g_2^{-1} \circ g_1 \in \mathrm{Stab}_{\theta_1,\ldots,\theta_k}$. But $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = \{e\}$, thus $g_2 = g_1$. It remains to show that if $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = \{e\}$ for some $\theta = (\theta_1, \ldots, \theta_k) \in \mathrm{Orb}_{\theta_1,\ldots,\theta_k}$, then $\mathrm{Stab}_{\bar{\theta}_1,\ldots,\bar{\theta}_k} = \{e\}$ for all $\bar{\theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_k) \in \mathrm{Orb}_{\theta_1,\ldots,\theta_k}$. To this end note that any $\bar{\theta} \in \mathrm{Orb}_{\theta_1,\ldots,\theta_k}$ may be represented as $\bar{\theta} = g \circ \theta$. Using the fact that k–point stabilizers of θ and $\bar{\theta}$ are conjugate, which means $\mathrm{Stab}_{\bar{\theta}_1,\ldots,\bar{\theta}_k} = g\mathrm{Stab}_{\theta_1,\ldots,\theta_k} g^{-1}$[23], we obtain $\mathrm{Stab}_{\bar{\theta}_1,\ldots,\bar{\theta}_k} = g \circ e \circ g^{-1} = g \circ g^{-1} = e$. Since the choice of the particular orbit was arbitrary, the result holds for all $(\theta_1, \ldots, \theta_k) \in \Theta_1 \times \cdots \times \Theta_k$.

Appendix 4 (proof of proposition 3):

By the orbit–k–point–stabilizer theorem (see lemma 3 a)), we get $|G : \mathrm{Stab}_{\theta_1,\ldots,\theta_k}| = |\mathrm{Orb}_{\theta_1,\ldots,\theta_k}|$. But the action is orbit–regular, hence $\mathrm{Stab}_{\theta_1,\ldots,\theta_k} = \{e\}$. Since $(\theta_1, \ldots, \theta_k)$ is arbitrary, we have $|G| = |\mathrm{Orb}_{\theta_1,\ldots,\theta_k}|$, for every $(\theta_1, \ldots, \theta_k) \in \Theta_1 \times \cdots \times \Theta_k$.

Appendix 5 (proof of proposition 4):

We have $C_\theta = p_y^{-1}(p_y(\theta)) = p_y^{-1}(p_y(g \circ \theta)) = C_{g \circ \theta}$. If follows $g \circ \theta \in C_\theta$ for each $g \in G$, hence $\mathrm{Orb}_\theta \subseteq C_\theta$. As a next step we will show that C_θ is G–stable subset of

[23] This is a standard result when Θ is not a Cartesian product (see e.g. Alperin and Bell (1995), p. 29). It may be shown that it holds also for the Cartesian product case.

Θ (i.e. $\overline{\theta} \in C_\theta \Rightarrow g \circ \overline{\theta} \in C_\theta$, for every $g \in G$). If $\overline{\theta} \in C_\theta$ then $C_{\overline{\theta}} = C_\theta$ and $C_{\overline{\theta}} = C_{g \circ \overline{\theta}}$. Thus we obtain $\overline{\theta} \in C_\theta \Rightarrow g \circ \overline{\theta} \in C_\theta$ for every $g \in G$. Hence for any $\overline{\theta} \in C_\theta$, $\mathrm{Orb}_{\overline{\theta}} \subseteq C_\theta$, hence $\cup_{\overline{\theta} \in C_\theta} \mathrm{Orb}_{\overline{\theta}} \subseteq C_\theta$. On the other hand, if $\overline{\theta} \in C_\theta$ then $\overline{\theta} \in \mathrm{Orb}_{\overline{\theta}}$ by definition, hence $C_\theta \subseteq \cup_{\overline{\theta} \in C_\theta} \mathrm{Orb}_{\overline{\theta}}$. As a consequence $C_\theta = \cup_{\overline{\theta} \in C_\theta} \mathrm{Orb}_{\overline{\theta}} = \cup_{\overline{\theta} \in \Delta} \mathrm{Orb}_{\overline{\theta}}$ (where Δ denotes the index set of distinct orbits contained in C_θ i.e. for all $\overline{\theta}_1, \overline{\theta}_2 \in \Delta$ and $\overline{\theta}_1 \neq \overline{\theta}_2$ we have $\mathrm{Orb}_{\overline{\theta}_1} \cap \mathrm{Orb}_{\overline{\theta}_2} = \varnothing$). Thus, any G–stable subset of Θ is a disjoint union of orbits. Suppose that p_y is maximal G–invariant, then by definition $\theta^* \in C_\theta = \{\overline{\theta} \in \Theta \mid p_y(\theta) = p_y(\overline{\theta})\}$ $\Rightarrow \theta^* \in \{\overline{\theta} \in \Theta \mid \overline{\theta} = g \circ \theta, g \in G\}$. But the latter set is recognized as Orb_θ. Therefore $C_\theta \subseteq \mathrm{Orb}_\theta$. Since we have already established that $\mathrm{Orb}_\theta \subseteq C_\theta$, it follows $C_\theta = \mathrm{Orb}_\theta$.

Appendix 6 (proof of lemma 4):

Proof of *a*) By the canonical decomposition there exists a factorization $p_y = h^* \circ \pi^*$, where $h^* : \Theta / \sim_p \to \mathrm{Im}(\Theta)$ (a bijection) and $\pi^* : \Theta \to \Theta / \sim_p$ is the canonical map (a surjection). Since by assumption $C_\theta = \mathrm{Orb}_\theta$, in the canonical map, we can replace the quotient set $\Theta / \sim_p$ with the orbit space $G \setminus \Theta$. Thus $\pi^* : \Theta \to G \setminus \Theta$ (i.e. $\pi^*(\theta) = \mathrm{Orb}_\theta$) and $h^* : G \setminus \Theta \to \mathrm{Im}(\Theta)$. Let us define the map $k : G \setminus \Theta \to \Lambda$ (i.e. $k(\mathrm{Orb}_\theta) = \lambda$). Clearly, k is a bijection because in every orbit there is exactly one orbit representative, thus we may write the canonical decomposition as follows $p_y = h^* \circ k^{-1} \circ k \circ \pi^*$. Let us denote $f = k \circ \pi^*$ and $h = h^* \circ k^{-1}$. Note that $f = k \circ \pi^*$ is just the canonical decomposition of $f : \Theta \to \Lambda$. Thus we arrive at the decomposition $p_y = h \circ f$, where $f : \Theta \to \Lambda$ (which is surjective). Furthermore, since h^* and k^{-1} are bijections, $h : \Lambda \to \mathrm{Im}(\Theta)$ is a bijection, too. Hence by proposition 1, $\sim_p \equiv \sim_f$.

Proof of *b*) Since h in $p_y = h \circ f$ is bijective it follows by definition 2 that Λ is identified and $f : \Theta \to \Lambda$ is the identifying function.

Proof of *c*) Having a unique decomposition $f = k \circ \pi^*$, where k is a bijection, we obtain $\forall \theta_1, \theta_2 \in \Theta$, $f(\theta_1) = k(\pi^*(\theta_1)) = k(\pi^*(\theta_2)) = f(\theta_2) \Leftrightarrow \pi^*(\theta_1) = \pi^*(\theta_2)$ (k is a bijection) $\Leftrightarrow \theta_1 = g \circ \theta_2$ for some $g \in G$ ($\pi^* : \Theta \to G \setminus \Theta$ is maximal G–invariant).

Appendix 7 (proof of proposition 5):

Define the mapping $\eta : S \to S\theta_r$ i.e. $\eta : g \mapsto g \circ \theta_r$ for every $g \in S$ (and fixed θ_r). Then η is surjective by construction. Assume $g_1 \circ \theta_r = g_2 \circ \theta_r$; $g_1, g_2 \in S$. It follows $g_2^{-1} \circ g_1 \circ \theta_r = \theta_r \Rightarrow g_2^{-1} \circ g_1 \in \mathrm{Stab}_{\theta_r}$. But by the orbit–regularity (see proposition 2), we have $\mathrm{Stab}_{\theta_r} = \{e\}$, hence $g_2^{-1} \circ g_1 = e \Rightarrow g_1 = g_2$ (i.e. η is injective). Thus η is the bijection. Noting that $S\theta_r = \Theta_r^*$, we get $|S| = |\Theta_r^*|$. By definition, $\theta_r \in \mathrm{Orb}_{\theta_r}$ and

$\theta_r \in \Theta_r$ hence $\theta_r \in \Theta_r \cap \mathrm{Orb}_{\theta_r}$. Obviously, $|S| = 1 \Leftrightarrow \Theta_r^* = \{\theta_r\}$ and $S = \{e\} \Rightarrow |S| = 1$. Moreover, if $|S| = 1$ then $S = \{g^*\}$ and we must have $g^* \circ \theta_r = \theta_r$. By proposition 2, it follows, $g^* \circ \theta_r = \theta_r \Rightarrow g^* = e$ (i.e. $S = \{e\}$). As each orbit is the set of transitivity and by orbit–regularity, every element in the orbit, say $\overline{\theta}$, must be represented as $\overline{\theta} = g \circ \theta_r$ for unique $g \in G$ (i.e. given $g_1, g_2 \in G$ such that $g_1 \neq g_2$ we have $g_1 \circ \theta_r \neq g_2 \circ \theta_r$). It follows that all elements of G, except the identity element, move θ_r to distinct elements in the orbit. Thus Orb_θ may be trivially partitioned into the singletons, one of which is θ_r. Of course, by proposition 3, $|\mathrm{Orb}_\theta| = |G|$.

Appendix 8 (proof of proposition 6):

The proof is similar to that of lemma 4. Since we assume $C_\theta = \mathrm{Orb}_\theta$, by the canonical decomposition, there exists a factorization $p_y = h^* \circ \pi^*$, where $h^* : G \setminus \Theta \to \mathrm{Im}(\Theta)$ (a bijection) and $\pi^* : \Theta \to G \setminus \Theta$ (i.e. $\pi^*(\theta) = \mathrm{Orb}_\theta$) (a surjection). As a next step, we show that equivalence class of $f : \Theta \to X$ is equal to orbit. By the maximal G–invariance property of f we have $C_\theta = \{\overline{\theta} \in \Theta \mid f(\theta) = f(\overline{\theta})\} = \{\overline{\theta} \in \Theta \mid \overline{\theta} = g \circ \theta, g \in G\} = \mathrm{Orb}_\theta$. Thus in the context of f we can also apply the canonical decomposition replacing equivalence class with the orbit. This results in the factorization $f = k \circ \pi^*$, where $\pi^* : \Theta \to G \setminus \Theta$ is surjective (recall that $G \setminus \Theta$ denotes the orbit space) and $k : G \setminus \Theta \to X$ (a bijection). Note that $\pi^* : \Theta \to G \setminus \Theta$ is the same in the canonical decomposition of p_y and f. Combining these two canonical decompositions we obtain $p_y = h^* \circ k^{-1} \circ k \circ \pi^* := h^* \circ k^{-1} \circ f$. Denoting $h = h^* \circ k^{-1}$, we arrive at the decomposition $p_y = h \circ f$. Clearly, h is bijective (f is surjective by hypothesis). Then using definition 2 it follows that X is identified and $f : \Theta \to X$ is the identifying function.

Appendix 9:

In order to prove $S = \{g \in GL_m \mid (gR^{-1}, gR^{-1}A, gR^{-1}B, gg') \in \Theta_r^*\} = \{\mathrm{I}_m\}$, where $\Theta_r^* = LT_m^+ \times \mathbb{R}_*^{m\times m} \times \Re(B) \times \{\mathrm{I}_m\}$, we need the following instrumental result

Lemma A1: *Let G be a group, then $g \in G \Rightarrow h \circ g \in G$ iff $h \in G$.*

Proof: Given $g \in G$, if $h \circ g \in G$, then there is $g_1 \in G$ such that $h \circ g = g_1 \Rightarrow h = g_1 \circ g^{-1} \in G$ (the last assertion follows since G is a group). On the other hand if $h \in G$ and $g \in G$ then $h \circ g \in G$ trivially.

If $(gR^{-1}, gR^{-1}A, gR^{-1}B, gg') \in \Theta_r^*$, we evidently must have $gg' = \mathrm{I}_m \Rightarrow g \in O_m$ and $gR^{-1} \in LT_m^+ \Rightarrow g \in LT_m^+$ (by lemma A1). Since $O_m \cap LT_m^+ = \{\mathrm{I}_m\}$, the needed result follows.

Appendix 10 (Proof of lemma 5):

Assume $A \in UT_m^1$ and $(R_1^{-1}A_1, R_1^{-1}B_1, \mathrm{I}_m) = (R_2^{-1}A_2, R_2^{-1}B_2, \mathrm{I}_m)$. Then $R_1^{-1}A_1 = R_2^{-1}A_2$ $\Rightarrow R_2R_1^{-1} = A_2A_1^{-1}$. Since $R_2R_1^{-1} \in LT_m^+$ and $A_2A_1^{-1} \in UT_m^1$ (because $UT_m^1, LT_m^+ < GL_m$) we have $R_2R_1^{-1} \in UT_m^1 \cap LT_m^+$, $A_2A_1^{-1} \in UT_m^1 \cap LT_m^+$. But $UT_m^1 \cap LT_m^+ = \{\mathrm{I}_m\}$, thus we must have $R_2R_1^{-1} = \mathrm{I}_m$ and $A_2A_1^{-1} = \mathrm{I}_m$. Ultimately, $R_1 = R_2$, $A_1 = A_2$ and $R_1^{-1}B_1 = R_2^{-1}B_2 = R_1^{-1}B_2 \Rightarrow B_1 = B_2$.

REFERENCES:

Ahn, S.K., and G.C. Reinsel (1988), "Nested Reduced–Rank Autoregressive Models for Multiple Time Series", *Journal of the American Statistical Association*, 83, pp. 849–856.

Alperin, J.L., and R.B. Bell (1995), *Groups and Representations*, Springer–Verlag.

Anderson, T.W. and H. Rubin (1956), "Statistical Inference in Factor Analysis", in: J. Neyman, ed., *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 5, Contributions to Econometrics, Industrial Research, and Psychometry*, University of California Press, California, Berkeley.

Bourbaki, N. (1968), *Elements of Mathematics*: *Theory of Sets*, Addison–Wesley Pub. Co. (Hermann, Paris).

Bowden, R. (1973), "The Theory of Parametric Identification", *Econometrica*, 41, pp. 1069–1074.

Brillinger, D.R. (1963), "Necessary and Sufficient Conditions for a Statistical Problem to be Invariant Under a Lie Group", *The Annals of Mathematical Statistics*, 34, pp. 492–500.

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

Fraser, D.A.S. (1967), "Statistical Models and Invariance", *The Annals of Mathematical Statistics*, 38, pp. 1061–1067.

Friedman, M. (1953), "The Methodology of Positive Economics", in: *Essays in Positive Economics*, The University of Chicago Press.

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

Hall, M. (1959), *The Theory of Groups*, The Macmillan Company.

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

Jacobson, N. (1985), *Basic Algebra I*, W.H. Freeman & Co.

Jöreskog K.G., and A.S. Goldberger (1975), "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable", *Journal of the American Statistical Association*, 70, pp. 631–639.

Kadane, J.B. (1975), "The Role of Identification in Bayesian Theory", in: S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics*, North–Holland Pub. Co.

Koopmans, T.C. (1949), "Identification Problems in Economic Model Construction", *Econometrica*, 17, pp. 125–144.

Koopmans, T.C. (1953), "Identification Problems in Economic Model Construction", in: Wm. C. Hood and T.C. Koopmans, eds., *Studies in Econometric Method*, Cowles Commission Monograph No. 14, John Wiley & Sons.

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

Koopmans, T.C., and O. Reiersøl (1950), "The Identification of Structural Characteristics", *The Annals of Mathematical Statistics*, 21, pp. 165–181.

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

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

Mach, E. (1898), *Popular Scientific Lectures*, The Open Court Pub. Co.

MacLane, S, and G. Birkhoff (1993), *Algebra*, *third edition*, AMS Chelsea Pub.

Mäki, U. (2009), "Realistic Realism about Unrealistic Models", in: H. Kincaid and D. Ross, eds., *Oxford Handbook of the Philosophy of Economics*, Oxford University Press.

Mäki, U. (2010), "Models and the Locus of Their Truth", forthcoming, *Synthese*.

Marschak, J. (1953), "Economic Measurements for Policy and Prediction", in: Wm. C. Hood and T.C. Koopmans, eds., *Studies in Econometric Method*, Cowles Commission Monograph No. 14, John Wiley & Sons.

Prakasa Rao, B.L.S. (1992), *Identifiability in Stochastic Models*: *Characterization of Probability Distributions*, Academic Press, Inc.

Reinsel, G.C. (1983), "Some Results on Multivariate Autoregressive Index Models", *Biometrika*, 70, pp. 145–156.

Rothenberg, T.J. (1971), "Identification in Parametric Models", *Econometrica*, 39, pp. 577–591.

Sargent, T.J., and C.A. Sims (1977), "Business Cycle Modeling Without Pretending to Have Too Much a Priori Economic Theory", in: C.A. Sims, ed., *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis.

Scott, W.R. (1987), *Group Theory*, Dover Pub., Inc.

Sims, C.A. (1981), "An Autoregressive Index Model for the U.S., 1948–1975", in: J. Kmenta and J.B. Ramsey, eds., *Large–Scale Macro–Econometric Models*: *Theory and Practice*, North–Holland Pub. Co.

Steinberger, M. (1993), *Algebra*, Prindle, Weber & Schmidt Pub. Co.

Weyl, H. (1952), *Symmetry*, Princeton University Press.

Wielandt, H. (1964), *Finite Permutation Groups*, Academic Press, Inc.