

FCBM

CBM
R

CBM
8414
1989

14/14



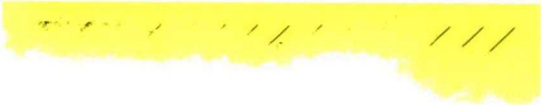
* C I N 0 0 9 9 3 *



No. 8914

**A BAYESIAN ANALYSIS OF
EXOGENEITY IN MODELS POOLING
TIME-SERIES AND CROSS-SECTION DATA**

by Jacek Osiewalski and
Mark F.J. Steel



March, 1989

A Bayesian Analysis of Exogeneity in Models Pooling
Time-Series and Cross-Section Data

Jacek Osiewalski
Academy of Economics, Kraków

Mark F.J. Steel
Tilburg University

JEL code: 211

Keywords: exogeneity; Bayesian cuts; time-series of cross-sections; prediction; noncausality

Abstract: The analysis of exogeneity in econometric time-series models as formalized in the seminal paper by Engle *et al.* (1983) is extended to cover a more general class of models, including error-components and spatial autoregression models. The Bayesian framework adopted here allows us to take full advantage of a number of statistical tools, related to the reduction of Bayesian experiments, and motivates a careful consideration of prediction issues, leading to a concept of predictive exogeneity. We also adapt and slightly weaken the formal definitions of weak and strong exogeneity introduced in Engle *et al.* (1983), and provide a naturally nested set of definitions for exogeneity. Four examples highlight the main implications of our analysis for econometric modelling.

Acknowledgements: The present paper was written under visiting fellowships from the Netherlands Organization for Scientific Research (N.W.O.) and Center, Tilburg University, for the first author, and under a fellowship of the Royal Netherlands Academy of Arts and Sciences (K.N.A.W.) for the second author. Useful discussions with Luc Bauwens, Jean-Pierre Florens, Michel Mouchart, Theo Nijman, and Jean-Marie Rolin are gratefully acknowledged. Of course, the usual disclaimer applies.

1. Introduction

The role of exogeneity as a fundamental concept in econometrics seems well established by now. Ever since the early fifties [see e.g. Koopmans (1950), Orcutt (1952), and Marschak (1953)] it has received considerable attention from the econometrics profession, which became aware of the necessarily ubiquitous nature of exogeneity assumptions in empirical work. These largely intuitive notions were formalized and clearly defined in the seminal paper by Engle et al. (1983), henceforth denoted by EHR, based on the statistical concept of classical cut as described in Barndorff-Nielsen (1978). A number of formal statistical tools prove extremely useful for defining exogeneity in EHR, who provide an extensive application to dynamic simultaneous equation models.

Econometric practice, however, is more and more inspired by microeconomic theory, leading to models with usually more than one observational dimension and somewhat different structures than the pure time-series examples found in EHR. Apart from the rapidly growing use of such micro models, they can, in some cases, lead to direct economic interpretations of statistical properties like exogeneity, as is briefly mentioned in Section 6. Therefore, we shall explicitly extend the framework used in EHR to cover such models, focusing in particular on error-components and spatial autoregression models. A second major difference with EHR is the essentially Bayesian character of the exposition, drawing on a strong statistical tradition in Bayesian cuts and reduction of Bayesian experiments, developed in a series of papers by Florens and Mouchart, henceforth denoted by FM (1977, 1980a, 1980b, 1982, 1985). We distinguish four different types of Bayesian cuts, depending on the model and the time period for which they are defined, which leads to somewhat weaker exogeneity conditions than found in EHR. It also provides a natural framework for the treatment of models outside the time-series domain.

Note that the use of the conditional model as opposed to the joint one has important advantages, both from the point of view of simplicity of presentation and computation, and also for the sake of robustness with respect

to misspecifications of the marginal process. In addition, the actual modelling of this marginal process may be problematic, in practice, due to the absence of relevant economic theories (just think of prices, interest rates, income variables, etc.). The intuitive content of weak exogeneity is the same as in EHR, namely a set of conditions sufficient for conducting inference on the parameters of interest using only the conditional model.

The formal definition, however, is different from the one adopted in EHR in two respects (apart from the Bayesian versus classical background). Our definition does not necessarily hold for each period in time, only for the sample period observed, and is weaker in that sense. In addition, the definition used here applies to both sequential models (as in EHR) and to joint or initial models (to be defined in Subsection 2.2), and is broader than its EHR counterpart in that sense.

In view of the "predictivist" tendencies in Bayesian econometrics [see e.g. Poirier (1988)], strengthened by de Finetti's (1974, 1975) representation theorem, we define a concept of predictive exogeneity as a set of sufficient conditions that validate forecasting solely on the basis of the conditional model and the posterior density of its parameters. Parameters of interest do not appear in this definition, which clearly illustrates that we can still validly predict with the conditional model, even though our (structural) parameters of interest can not be recovered from it.

We then define a third type of exogeneity that validates both inference and forecasting conditionally on a subset of the variables, as the union of both weak and predictive exogeneity. The result of combining these sets of conditions is called strong exogeneity, as it has the same intuitive meaning as its namesake in EHR, but its formal definition is, of course, different: again, it is both weaker (not for all time periods) and broader (for sequential and initial models).

Thus, our three definitions require rather weak conditions and are directly suited for the treatment of models pooling time-series and cross-section data. Also, particular attention is paid to the issue of prediction, to which we devote a special type of exogeneity, quite distinct from the property of noncausality, used in EHR to extend weak to strong exogeneity, and which has no exogeneity message in itself.

In the type of error-components and spatial models considered here it is not uncommon (as opposed to current practice in time-series models) to assume independence over time in the sampling process. In this special case it is found that prior independence is enough for predictive exogeneity, whereas strong exogeneity holds if, in addition, the parameters of interest can be deduced from those of the conditional model. Remark that independence is also the only case where the ordering of the observations becomes irrelevant and similar tools can be used over individuals as over time. If, however, individuals are linked, these dependences are usually not unidirectional, so that sequential concepts do not naturally apply. In the examples of Section 5 we shall illustrate that such individual links can often prevent a cut and destroy exogeneity.

The paper is organized in the following way: in Section 2 we discuss Bayesian cuts and noncausality, whereas Sections 3 and 4 apply these concepts to, respectively, the posterior and the predictive analysis, defining our three types of exogeneity in the process. Section 5 contains four examples, two of which are directly taken from the literature: the stochastic production function analysed in Zellner et al. (1966), and a model describing schooling, occupation, and income with unobserved individual effects found in Chamberlain and Griliches (1975). The final section groups some conclusions.

The notation used is kept in close correspondence with the conventions adopted in EHR and in the FM papers, whereas for density functions we refer the reader to Appendix A in Drèze and Richard (1983). The basic statistical tools for the reduction of Bayesian experiments are conditional independence, measurable separability, and strong identification, which are discussed in detail in Mouchart and Rolin (1984). Here only the first of these concepts is used extensively (the second one is briefly mentioned in Subsections 2.3 and 4.3) and the emphasis is not on statistical rigour (for this we refer to the paper by FM), but rather on the application to econometric models, so that proofs are often only sketched.

2. Bayesian Cuts

2.1. Preliminaries

In order to define and discuss the various concepts of cut and noncausality that we shall use in this paper, we first establish some notational conventions. As we noted in the introduction, we shall focus upon models with two observational dimensions, one of which we shall usually call "time", denoted by an index $t = 1, \dots, T$, whereas the other observational dimension will be assumed to be over "individuals" $i = 1, \dots, N$. Although this nomenclature is not crucial at all, since we can easily think of other names for the observational dimensions,¹⁾ there is an important difference between both dimensions as a well-defined natural ordering exists in time, but usually not over individuals. This, of course, implies that both indices are not generally interchangeable, except in special cases of independence between observations. Therefore, we shall require different types of cuts for different observational dimensions. Such definitions already exist in the statistical literature, e.g. in FM (1980a,b; 1985), which we shall simply adapt for our purposes and state with brief comments in this section.

Let us assume that we possess observations on m variables for N individuals at each time period t , i.e. we are implicitly restricting both m and N to be constant over time, without, however, preventing things like rotating samples in panel data. We group the observations at time t in an mN -dimensional vector x_t , ordered as

$$x_t = \begin{bmatrix} x_{1t}^{(1)} \\ \vdots \\ x_{Nt}^{(1)} \\ \vdots \\ x_{1t}^{(m)} \\ \vdots \\ x_{Nt}^{(m)} \end{bmatrix} = \begin{bmatrix} x_t^{(1)} \\ \vdots \\ x_t^{(m)} \end{bmatrix}, \quad (2.1)$$

where superscripts in parentheses denote the different variables observed. The time index t runs from 1 to T for the sample and from $T+1$ to $T+s$ if we wish to forecast s periods ahead. In line with the notation in EHR, let us define the matrix

$$X_t^r = (x_r \dots x_t)' \text{ for } r \leq t, \quad (2.2)$$

which implies $X_t^r \in \mathbb{R}^{(t-r+1) \times mN}$, and denote the full information set available at time t by

$$X_{t-1} = \begin{bmatrix} X_0 \\ X_{t-1}^1 \end{bmatrix}, \quad (2.3)$$

where X_0 is a matrix of possibly infinite dimension describing all initial conditions relevant to the process. As discussed in Richard (1979) and Engle *et al.* (1980), their treatment can be rather difficult, but, for our present purposes, we shall consider them given, like in EHR.

Next, we partition x_t into

$$x_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix} \quad (2.4)$$

with $y_t = (x_t^{(1)'} \dots x_t^{(p)'})'$, containing the first p variables and $z_t = (x_t^{(p+1)'} \dots x_t^{(m)'})'$, grouping the last q variables ($p+q = m$), and conformably, $\forall t$,

$$X_t = (Y_t \ Z_t) = \begin{bmatrix} Y_0 \\ Y_t^1 \\ Z_0 \\ Z_t^1 \end{bmatrix}. \quad (2.5)$$

It is obvious from the above that a pure time-series framework as in EHR corresponds to $N=1$, whereas a strictly cross-section analysis implies that $T=1$.

Although much of the statistical literature was written in terms of conditional independence among σ -fields, we shall, for expository purposes, remain in the less general framework of probability densities [indicated by $D(\cdot|\cdot)$], as this is much closer to the existing econometrics tradition. Conditional independence of random variables, say, d and e , given g , is denoted by $d \perp e|g$.

2.2. Cuts and Noncausality: Definitions

As noted in the previous subsection, the entire analysis will be conducted conditionally on the initial conditions X_0 , so that our joint data density becomes

$$D(X_T^1|X_0, \theta), \quad (2.6)$$

where $\theta \in \Theta$ denotes a vector of parameters. The model in (2.6) is called an "initial model" in FM (1985) and it serves to define the concept of initial cut. We shall also define a global cut on (2.6), a terminology found in FM (1980a).

If we concentrate on the factorization of (2.6) into a product of "sequential models":

$$D(X_T^1|X_0, \theta) = \prod_{t=1}^T D(x_t|x_{t-1}, \theta), \quad (2.7)$$

it is natural to define sequential and one-shot cuts, again based on the FM terminology.

In a Bayesian framework we wish to treat both observations and parameters in a symmetric way and, therefore, extend the sampling theory model in

(2.6) or (2.7) with a so-called prior probability on the parameter space Θ , denoted by $D(\theta|X_0)$ as it may very well depend on the initial conditions (which can e.g. include a previous sample). The Bayesian model then becomes

$$D(X_T^1, \theta | X_0) = D(X_T^1 | X_0, \theta) D(\theta | X_0), \quad (2.8)$$

from which both posterior densities $D(\theta|X_T)$ and predictive densities $D(X_T^1|X_0)$ can be derived, as well as their reductions obtained by marginalizing and conditioning.

To complete our description of the model in (2.8), we can consider any one-to-one transformation or reparameterization from θ to some parameter vector $a \in A$, if the latter is more appropriate for our purposes at hand.

Let us now focus on the concept of cut. In general, a cut is defined in terms of a factorization of the likelihood function of x into a conditional process of y given z , for which $b = b(a) \in B$ is a sufficient parameter vector, and a marginal process describing z , which has $c = c(a) \in C$ as a sufficient parameterization. In the absence of links between b and c , we can then limit ourselves to only the conditional process for the purpose of conducting inference on b . This is exactly the formalization of the intuitive exogeneity concept which was given in EHR and which led to their definition of weak exogeneity.

In a Bayesian analysis, the absence of links between b and c has a very natural and direct interpretation in terms of prior independence, which, of course, implies "equivalence" of $D(b, c | X_0)$ and $D(b | X_0) \cdot D(c | X_0)$ [i.e. the joint prior distribution has the same null sets as the product of the marginal distributions; see Florens et al. (1987)], which, in its turn, implies the classical concept of "variation free"-ness. The latter concept was used in EHR, but seems, at least to the present authors, a much less natural requirement than prior independence,²⁾ which also explicitly allows for stochastic links between both sets of parameters and is often very easy to check (however, see footnote 6).

A Bayesian cut is thus characterized by both a factorization of the likelihood function and prior independence between the parameters in both processes, which implies a total separation of both types of information

and naturally leads to posterior independence, so that for inference on b we do not have to specify the marginal process with parameters c , nor the prior distribution of c .

Depending upon the situation at hand, we can either focus on initial [see (2.6)] or sequential [see (2.7)] models, and we also distinguish between cuts that hold for all time periods or just for one. This classification gives us four possible cuts, which are formally defined in Table 1.

Table 1. Types of Bayesian Cuts

applies to: holds for:	sequential model (2.7)	initial model (2.6)
all $t \geq 1$	sequential cut $b \perp c X_0$ $a \perp z_t c, X_{t-1}$ $a \perp x_t b, z_t, X_{t-1}$	initial cut $b \perp c X_0$ $a \perp Z_T^1 c, X_0$ $a \perp X_T^1 b, Z_T^1, X_0$
one t	one-shot cut (over individuals) $b \perp c X_{t-1}$ $a \perp z_t c, X_{t-1}$ $a \perp x_t b, z_t, X_{t-1}$	global cut ($t=T$) $b \perp c X_0$ $a \perp Z_T^1 c, X_0$ $a \perp X_T^1 b, Z_T^1, X_0$

The situations and models in which each of these four cuts is most appropriate will be discussed in Subsection 3.3.

Finally, we define the concept of noncausality,³⁾ a version of which was used in EHR to characterize "strong exogeneity". We say that y does not cause z given v if and only if $\forall t \geq 1$

$$z_{t+1} \perp y_t^1 | z_t^1, X_0, v. \quad (2.9)$$

which, basically, states that the past of y does not influence the marginal process of z , once we know the past of z , the initial conditions, and the parameters in v . Note that EHR use the special case of (2.9) where $v=a$, which is called Granger noncausality.

2.3. Connections Between Cuts

If we know how the four Bayesian cuts of the previous subsection are interrelated, we might use this to our advantage both in understanding and in applying the theory.

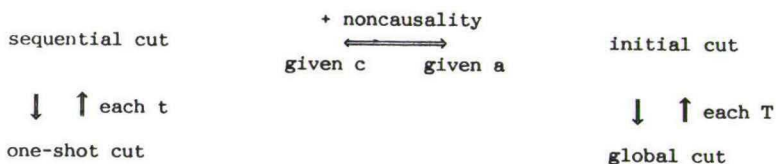
FM (1980a) show that the relations between sequential and initial cuts can be described in terms of noncausality. In particular, they prove that

- (i) a sequential cut and noncausality given c is exactly equivalent to
- (ii) an initial cut and noncausality given a (i.e. Granger noncausality).

Using the posterior independence of b and c for any Bayesian cut [FM (1985), Theorem 2.8] we can prove that a sequential cut implies a one-shot cut over individuals (vt), whereas the converse holds by definition if we have the one-shot cut for all t .

Similarly, an initial cut implies a global cut, and is only equivalent to it if the global cut holds for all sample periods T .

To summarize, we can identify the following relationships:



Clearly, if we are in a pure cross-section framework ($t=T=1$), the concept of noncausality, which is specifically time-oriented, becomes vacuous, and all cuts coincide, which can be seen directly from the definitions in Table 1.

Note that a sequential cut and Granger noncausality do not generally lead to an initial cut, i.e. to separation into data densities. The latter separation is of interest as it greatly facilitates prediction (discussed here in Subsection 4), which formed the object of the definition of strong exogeneity in EHR. FM (1980a) show that such a separation is then achieved under the additional condition of measurable separability, a rather technical requirement which, roughly, says there should be no information "in common" between the parameter set α and X_t , given (c, Z_t^1, X_0) . Florens et al. (1987) stress that this is a very weak condition indeed, as it is always verified under equivalence of probabilities (in other words, when the joint distribution has the same null sets as the product of the marginal distributions).

3. Posterior Analysis: Weak Exogeneity

3.1. Posterior Inference

In Subsection 2.2 we introduced the full Bayesian model (2.8), from which the posterior distribution is simply obtained by conditioning on the observed sample, giving

$$D(\theta|X_T) \propto D(X_T^1|X_0, \theta)D(\theta|X_0). \quad (3.1)$$

If there exists some reparameterization from θ to $a = (b, c)$ which operates a Bayesian cut, then b and c are independent a posteriori and any posterior inference on b or some function thereof can be conducted solely on the basis of the conditional process and the prior distribution of b .

In particular, under a global cut for a given sample period T , we obtain

$$D(b|X_T) \propto D(Y_T^1|Z_T^1, X_0, b)D(b|X_0), \quad (3.2)$$

which clearly shows that there is no need to use the full data density $D(X_T^1|X_0, a)$, nor the joint prior density $D(a|X_0)$ for the purpose of inference on b (or any function of b).

An initial cut will give us (3.2) for any $T \geq 1$, by definition.

In the sequential model, the weaker condition that does the job is to have a one-shot cut over individuals at each $t \in [1, T]$, which implies

$$D(b|X_T) \propto \prod_{t=1}^T D(y_t|z_t, X_{t-1}, b)D(b|X_0), \quad (3.3)$$

whereas the stronger condition is a sequential cut, leading to a complete separation of information for all sample periods $T \geq 1$.

This explains the crucial role of Bayesian cuts in any discussion of conducting inference based on a reduction of the joint model, although cuts are not strictly necessary as is stressed in the discussion of "mutual

exogeneity" in FM (1985). Such ideas were already present in FM (1977) and cuts were also used in EHR, albeit in a classical framework. The next subsection will address the issue in more detail.

3.2. A Description of Weak Exogeneity

We shall adhere to the terminology of EHR in using the name "weak exogeneity" for the concept that validates inference based solely on the conditional model, without taking the marginal model in consideration. Obviously, it then becomes a crucial issue what exactly we want to learn from this conditional model, which leads EHR to introduce the concept of parameters of interest as those parameters that have a specific meaning. The latter may e.g. be derived from economic theory in which case inference on them may constitute evidence in favour of or against certain theories concerning the outside world. An alternative source of interest might be their relative stability in changing environments, a vital requirement for conducting "out-of-sample" policy simulations and predictions.

Combining the issue of parameters of interest, say $\Psi = f(a)$, with a complete separation of prior and sample information, i.e. a Bayesian cut, we are led to the following definition of weak exogeneity in our Bayesian framework, where we define z to be the relevant set of observations, i.e. z_t in the sequential model (2.7) and Z_t^1 in the initial model (2.6).

Definition 1: z is weakly exogenous over the given sample period $[1, T]$ for Ψ if and only if there exists a reparameterization from θ to $a = (b, c)$ such that

- (i) a one-shot cut over individuals holds for each $t \in [1, T]$ or a global cut holds (for $t=T$).
- (ii) $\Psi = f(b)$.

Condition (ii) makes sure that we can infer about the parameters of interest from the parameterization of the conditional model only, whereas condition (i) gives us the required separation of information between the conditional and marginal processes.

If we strengthen (i) to a sequential or initial cut, combined with (ii), then z is weakly exogenous for Ψ for any sample period, like in EHR.

However, if all we want is valid inference on Ψ from the conditional model over the fixed sample period $[1, T]$, conditions (i) and (ii) are sufficient.

Finally, note that if we wish to consider subperiods of $[1, T]$, e.g. for parameter stability tests, we can still validly use the conditional model under a one-shot cut for each $t \in [1, T]$, but not under a global cut for T . In the latter case, an initial cut seems more appropriate.

Note, finally, that, as mentioned in EHR, a cut is not necessary for obtaining full efficiency from the conditional model, as e.g. we can have prior independence between Ψ and c , but not between b and c , yet in some cases no loss of information may occur.

Even in the case where we are interested in the entire parameter set b , a Bayesian cut is not strictly necessary, as is clearly stated e.g. in FM (1985), where the concept of "mutual exogeneity", introduced in FM (1977), is used, which is weaker than a Bayesian cut, but validates inference for functions of b on the basis of the conditional model. Unfortunately, it depends on the specific form of the prior density $D(c|b, X_0)$, whereas a cut is much more robust with respect to the prior structure, as it only requires prior independence of b and c .

Therefore, we prefer to use the stronger concept of Bayesian cut, as in EHR.

Let us stress, however, that the notions of cut and weak exogeneity used here are not necessary, but only sufficient for our purposes of valid inference based on the conditional model alone.

3.3. Which Cut for Which Model?

The definition of weak exogeneity given in the previous subsection was based on four different types of Bayesian cuts. This subsection will investigate which type of cut is called for in a number of different situations.

Pure time-series data ($i=N=1$) have a natural ordering in which information is accumulated, and, thus, a sequential cut seems most appropriate here.

Cross-section data ($t=T=1$) lack such a natural ordering of observations and if observations are dependent, this is usually not in a unidirectional way. We are then led to consider a one-shot cut over the sample size N .

Time-series of cross-sections or, more generally, models with two observational dimensions are the focus of this paper. Here we can use a sequential cut over time, a one-shot cut over individuals or a global cut only, depending on the structure of our model.

If we have independence over time, i.e. the following holds

$$D(x_t | X_{t-1}, a) = D(x_t | a), \quad \forall t, \quad (3.4)$$

then under the additional conditions:⁴⁾

$$D(x_t | a) = D(y_t | z_t, b) D(z_t | c), \quad \forall t, \quad (3.5)$$

$$D(a | X_0) = D(b | X_0) D(c | X_0), \quad (3.6)$$

we obtain both

- (i) an initial cut and noncausality given a , and (since it is equivalent, see Subsection 2.3)
- (ii) a sequential cut and noncausality given c .

This gives us separation into sample densities and validates (provided $\Psi = f(b)$) both inference from the conditional model and its use for forecasting (see Section 4). Interestingly, under independence the (time) ordering becomes irrelevant and we can simply use the same approach for models with independence over individuals (just interchange the meaning of the indices).

Such an approach can be applied to e.g. the following models with two error components:

- models with time effects and independence over time (see our example in Subsection 5.2);
- models with individual effects and independence over individuals (see our example in Subsection 5.4).

In models with a time effect without independence over time, we should look for a sequential cut, while for prediction purposes we then still have to check noncausality given c , as will be discussed in Section 4. In other cases, like models with three error components, only a global cut can be examined.

4. Predictive Analysis and Strong Exogeneity

4.1. Forecasting and Predictive Exogeneity

Let us first consider forecasting over time, whereas we shall briefly mention forecasting over individuals at the end of this section.

We are basically interested in "conditional forecasting", i.e. forecasting certain values for the variables we condition on and then use the conditional model for predicting future values of the other variables.

Specifically, if we wish to forecast s periods ahead (over all N individuals for each period), we start from the following Bayesian model in terms of the original parameterization θ

$$D(X_{T+s}^1 | X_0, \theta) = D(X_{T+s}^{T+1} | X_T, \theta) D(X_T^1 | X_0, \theta) D(\theta | X_0), \quad (4.1)$$

where $D(X_T^1 | X_0, \theta)$ and $D(\theta | X_0)$ have already been introduced in Subsection 2.2 as the sample data density and the prior density, respectively. The post-sample predictive density is then obtained by weighting the other factor in (4.1) using the posterior density

$$D(X_{T+s}^{T+1} | X_T) = \int_{\theta} D(X_{T+s}^{T+1} | X_T, \theta) D(\theta | X_T) d\theta. \quad (4.2)$$

Using a partitioning similar to (2.5) we are now really interested in the conditional predictive density

$$D(Y_{T+s}^{T+1} | Z_{T+s}^{T+1}, X_T) = \frac{D(X_{T+s}^{T+1} | X_T)}{\int D(X_{T+s}^{T+1} | X_T) dY_{T+s}^{T+1}}, \quad (4.3)$$

where the last integration is over the "forecasting" sample space of Y_{T+s}^{T+1} . This density (4.3) can be obtained without considering the marginal process for Z_{T+s}^{T+1} (i.e. for purposes of conditional forecasting) under the following (sufficient) conditions, which will define the concept of predictive exogeneity:

Definition 2: z is predictively exogenous over the forecasting period $[T+1, T+s]$ if and only if there exists a reparameterization from θ to a_* = (b_*, c_*) such that

$$(i) \quad b_* \perp c_* | X_T$$

$$(ii) \quad a_* \perp Z_{T+s}^{T+1} | c_*, X_T$$

$$(iii) \quad a_* \perp X_{T+s}^{T+1} | b_*, Z_{T+s}^{T+1}, X_T$$

which, in terms of densities, implies that (4.3) becomes

$$D(Y_{T+s}^{T+1} | Z_{T+s}^{T+1}, X_T) = \int_{B_*} D(Y_{T+s}^{T+1} | Z_{T+s}^{T+1}, X_T, b_*) D(b_* | X_T) db_* \quad (4.4)$$

if $b_* \in B_*$. This means that we can focus on the conditional part only, and treat Z_{T+s}^{T+1} as if it were fixed. Treating X_T as initial conditions for the forecasting period, we can, of course, interpret (i)-(iii) as a global cut for the period we want to predict.

The posterior independence between b_* and c_* [condition (i)] is implied by any Bayesian cut over the observed sample period in this parameterization (b_*, c_*) [see FM (1985)], but such a Bayesian cut is not strictly necessary, as (i) is also satisfied under the weaker conditions of prior independence and "parameter unrelatedness" of b_* and c_* as in Basu (1977). Remark that the posterior density of b_* is now not used for inference purposes on Ψ (it just serves as a "weighting function"), and therefore the issue of parameters of interest does not appear in Definition 2.

4.2. Strong Exogeneity

We have seen that the concept of predictive exogeneity is sufficient for purely predictive issues, which occur rather often in Bayesian analyses [see e.g. Zellner (1971), Zellner and Hong (1988), Chib et al. (1988), and Poirier (1988)]. If, however, we wish to combine both inferences on Ψ on the basis of the observed sample and prediction over some future period, the conditions mentioned in Definition 2 are no longer sufficient. In particular, we need to join the concepts of weak exogeneity for Ψ over the

observations and that of predictive exogeneity over the forecasting period. The union of both sets of conditions under the same parameterization (i.e. $b = b_*$ and $c = c_*$) will define strong exogeneity. The equality of parameterizations basically implies that we use the same parameterization of the conditional model for both inference and prediction and seems a practical requirement (although not a logical one!).

Definition 3: z is strongly exogenous for inference on Ψ over the sample period $[1, T]$ and for prediction over $[T+1, T+s]$ if and only if there exists a reparameterization from θ to $a = (b, c)$ such that

- (i) a one-shot cut over individuals holds for each $t \in [1, T]$ or a global cut holds for the sample period ($t=T$)
- (ii) $\Psi = f(b)$
- (iii) $a \perp Z_{T+s}^{T+1} | c, X_T$
- (iv) $a \perp X_{T+s}^{T+1} | b, Z_{T+s}^{T+1}, X_T$.

Of course, condition (i) of the present definition ensures also that condition (i) of Definition 2, i.e. posterior independence, is satisfied and its extension to a sequential or an initial cut again allows for valid inference on Ψ based on any sample period, i.e. for any $T \geq 1$.

4.3. Cuts and Prediction

Some implications of sequential and initial cuts will be examined in the prediction context.

One-shot and global cuts are, in principle, formulated for just one time period, so that these are not very natural concepts to consider for prediction, unless we explicitly define the latter to hold over the forecasting period (remember that predictive exogeneity can be interpreted as a global cut for the forecasting period). Several considerations here are well-known and documented in e.g. FM (1980b); however, we adapt them to our particular framework.

Consider first a sequential cut and note that this is sufficient for one-period ahead forecasting [compare FM (1980b)].

In other words, z is predictively exogenous over $[T+1, T+1]$ ($s=1$), since (in terms of densities)

$$D(z_{T+1} | X_T, a) = D(z_{T+1} | X_T, c) \quad (4.5)$$

$$D(y_{T+1} | z_{T+1}, X_T, a) = D(y_{T+1} | z_{T+1}, X_T, b) \quad (4.6)$$

and we have posterior independence

$$D(a | X_T) = D(b | X_T) D(c | X_T), \quad (4.7)$$

which corresponds to the conditions in Definition 2 for $s=1$. If, in addition, $\Psi = f(b)$ alone, then a sequential cut even gives us strong exogeneity for inference on Ψ based on $[1, T]$ and prediction to period $T+1$ for any $T \geq 1$.

If we wish to predict further ahead, a sequential cut by itself is not enough; we shall also use the noncausality conditions that lead to the equivalence of sequential and initial cuts (and thus to a separation into data densities). We need, in particular, that the following holds

$$D(Z_{T+s}^{T+1} | X_T, a) = D(Z_{T+s}^{T+1} | X_T, c) = \prod_{j=1}^s D(z_{T+j} | X_{T+j-1}, c) \quad (4.8)$$

and

$$D(Y_{T+s}^{T+1} | Z_{T+s}^{T+1}, X_T, a) = D(Y_{T+s}^{T+1} | Z_{T+s}^{T+1}, X_T, b) = \prod_{j=1}^s D(y_{T+j} | z_{T+j}, X_{T+j-1}, b) \quad (4.9)$$

which are both satisfied under noncausality given c , i.e. from (2.9) for all $t \geq 1$:

$$D(z_{t+1} | X_t, c) = D(z_{t+1} | Z_t^1, X_0, c). \quad (4.10)$$

The proof of this assertion is simply obtained after writing out the density for X_{T+s}^{T+1} given X_T and a .

Under an initial cut conditions (i) and (iii) of Definition 2 can easily be shown to hold for all pairs (T,s) . We also obtain that

$$D(Z_{T+s}^{T+1} | Z_T^1, X_0, a) = D(Z_{T+s}^{T+1} | Z_T^1, X_0, c) \quad (4.11)$$

which implies that c is sufficient for Z_{T+s}^{T+1} given its own past (and X_0) and this is not equivalent to (ii) in Definition 2. Let us now add the condition of noncausality given a (Granger), which can also be written as [see FM (1982)]

$$Z_{t+1}^t \perp Y_t^1 | Z_t^1, X_0, a, \quad (4.12)$$

implying for any pair (T,s)

$$D(Z_{T+s}^{T+1} | X_T, a) = D(Z_{T+s}^{T+1} | Z_T^1, X_0, a), \quad (4.13)$$

which, combined with (4.11), shows that c is a sufficient parameterization for the process of Z_{T+s}^{T+1} given the whole sample of observations X_T , i.e. condition (ii).

Thus, under an initial cut and Granger noncausality (i.e. given a) or, equivalently, under a sequential cut combined with noncausality given c , z is predictively exogenous for all pairs (T,s) . This means we can focus on the conditional model for prediction purposes based on any sample of observations and for any future period.

If, in addition, our parameters of interest Ψ are a function of b only, z is strongly exogenous for any $T \geq 1$ and for any $s \geq 1$, validating also inference on Ψ . Remark that the extension to all future periods instead of just a fixed s , as in the original Definition 3, is a direct consequence of the noncausality conditions, combined with initial or sequential cuts, of course, which, themselves, extend the strong exogeneity concept to any sample period T .

This, in our framework, strongest version of strong exogeneity is in fact the definition adopted in EHR, although they use the combination of sequential cut and Granger noncausality.

As briefly mentioned in Subsection 2.3, this is only equivalent to the strong version of our definition if we make an additional assumption. FM (1980a) introduce the very weak concept of measurable separability to bridge this theoretical gap, and note that such a condition often holds in practice. In fact, the issue is so subtle that even Basu (1955) overlooked it, until it was corrected by himself in Basu (1958). It is, however, not surprising that EHR use the Granger concept as they operate in a sampling-theory framework, i.e. without defining a measure over the parameter space, so that everything is naturally stated given the whole set of parameters α and it becomes notoriously difficult to get rid of "nuisance parameters" (b), as was described in Basu (1977).

As stated in Subsection 3.3, an independent sampling process together with (3.5) and (3.6) will give us e.g. a sequential cut and noncausality given c , and, therefore, predictive exogeneity for all T and s . Of course, it then depends on our parameters of interest whether this also guarantees strong exogeneity for all pairs (T,s) .

Under independence over time, its specific feature of possessing a fixed ordering no longer matters. Thus, in the case of independence over individuals (even with any form of dependence over time for each individual), the approach discussed here for forecasting over time also applies to predicting over individuals by simply interpreting t as the index for individuals.

5. Examples

5.1. A Cobb-Douglas Production Function

In Zellner et al. (1966), henceforth denoted by ZKD, we find an interesting production function model of the Cobb-Douglas type, where profits are formulated in a stochastic way and firms maximize the expected value of profits. This model will be used here to exemplify some of the issues that can arise in the practical application of concepts like Bayesian cut and exogeneity, and to stress the role of the prior distribution in distinguishing these concepts from their classical counterparts.

In their notation, the production function of firm i is given by

$$X_i = AL_i^{\alpha_1} K_i^{\alpha_2} e^{u_{0i}}, \quad (5.1)$$

where X_i , L_i and K_i denote output, labour and capital inputs, respectively, for firm i , whereas the stochastics are introduced through u_{0i} , reflecting unpredictable disturbances, assumed Normally distributed with variance σ_{00} .

Denoting by p , w and r the respective prices of output, labour and capital, we can write the resulting model for $i : 1 \rightarrow n$ as

$$x_{0i} - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \alpha_0 = u_{0i} \quad (5.2)$$

$$(\alpha_1 - 1)x_{1i} + \alpha_2 x_{2i} - k_1 = u_{1i} \quad (5.3)$$

$$\alpha_1 x_{1i} + (\alpha_2 - 1)x_{2i} - k_2 = u_{2i}, \quad (5.4)$$

with⁵⁾

$$x_{0i} = \log X_i$$

$$x_{1i} = \log L_i$$

$$x_{2i} = \log K_i$$

$$\alpha_0 = \log A$$

$$k_1 = \log \frac{wR_1}{A\alpha_1^p} - \frac{\sigma_{00}}{2}$$

$$k_2 = \log \frac{rR_2}{A\alpha_2^p} - \frac{\sigma_{00}}{2},$$

where R_j and u_{ji} reflect, respectively, systematic (if $R_j \neq 1$) and random errors made in satisfying the j th first order condition ($j = 1, 2$). The vectors $u_i = (u_{0i} \ u_{1i} \ u_{2i})'$ are independently identically distributed (i.i.d.) and joint Normality is assumed with covariance matrix $\Sigma = (\sigma_{kl})$ ($k, l = 0, 1, 2$), where rather convincing economic reasons exist for constraining σ_{01} and σ_{02} to zero, as they represent covariances between "acts of nature" in u_{0i} and "human errors" in u_{1i} and u_{2i} .

If we impose these zero restrictions, we have nine parameters left in the model, namely $\theta = (\alpha_0, \alpha_1, \alpha_2, \sigma_{00}, \sigma_{11}, \sigma_{12}, \sigma_{22}, k_1, k_2)$ which is exactly the number of free parameters in an unrestricted trivariate Gaussian process. The difference with the latter as it was used e.g. in Example 3.1 of EHR (but then bivariate) is that here we start out from a structural form of the model, instead of its reduced form. The presence of the two structural parameters α_1 and α_2 is, however, compensated by the two zero restrictions on the error covariances so that the model is not overparameterized, nor restricted, and a (classical) cut seems easy to obtain.

The model in (5.2)-(5.4), together with the assumptions on u_i , implies the following factorization in terms of the original parameterization θ

$$D(x_{0i} | x_{1i}, x_{2i}, \theta_1) = f_N^1(x_{0i} | \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i}, \sigma_{00}) \quad (5.5)$$

$$D\left(\begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix} | \theta_2\right) = f_N^2\left(\begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix} | B^{-1} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}, B^{-1} \Sigma_* B^{-1}\right), \quad (5.6)$$

where

$$\Sigma = \begin{bmatrix} \sigma_{00} & 0 & 0 \\ 0 & \sigma_{11} & \sigma_{12} \\ 0 & \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma_{00} & 0 \\ 0 & \Sigma_*$$

$$B = \begin{bmatrix} \alpha_1^{-1} & \alpha_2 \\ \alpha_1 & \alpha_2^{-1} \end{bmatrix}.$$

Since $\vartheta_1 = (\alpha_0, \alpha_1, \alpha_2, \sigma_{00})$ and $\vartheta_2 = (\alpha_1, \alpha_2, k_1, k_2, \Sigma_*)$ are not variation-free, this factorization does not operate a classical cut, nor a Bayesian one. Consider, however, the following reparameterization from ϑ to a

$$\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \rightarrow \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = B^{-1} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \quad (5.7)$$

$$\Sigma_* \rightarrow \tilde{\Sigma} = B^{-1} \Sigma_* B^{-1}, \quad (5.8)$$

giving us $b = \vartheta_1 = (\alpha_0, \alpha_1, \alpha_2, \sigma_{00})$ in the conditional model (5.5), as before, but $c = (d_1, d_2, \tilde{\Sigma})$ as a sufficient parameter set for the marginal model (5.6). Remark that b and c are variation free, leading to a classical cut and possibly a Bayesian one, provided b and c are prior independent. As discussed in Subsection 4.3, independence over individuals allows us to use the same approach as for independence over time, just by renaming the indices. Moreover, provided we impose prior independence as in (3.6)

$$D(a|X_0) = D(b|X_0)D(c|X_0),$$

we obtain an initial cut and Granger noncausality (or, equivalently, a sequential cut and noncausality given c) over individual firms. This validates both the use of (5.5) for inference about any function of b (e.g. the returns to scale parameter $\omega = \alpha_1 + \alpha_2$) and its use for conditional predictions. Under (3.6) we thus always have predictive exogeneity and both weak and strong exogeneity for $\Psi = f(b)$ over any sample of individual firms.

ZKD, however, remain within the original ϑ parameterization, which precludes a cut, yet they find that the conditional model suffices for efficient

inference on θ_1 , which concurs with our conclusions above based on the concept of Bayesian cut (in terms of a).

The reason for this lies in the fact that they choose a Jeffreys' type prior structure (with independence between location and scale parameters), which takes the form

$$D(\theta|X_0) \propto |B|^{-1} \sigma_{00}^{-1} |\Sigma_*|^{-3/2}, \quad (5.9)$$

and which corresponds to the following prior measure in terms of a

$$D(a|X_0) \propto \sigma_{00}^{-1} |\Sigma|^{-3/2}. \quad (5.10)$$

Remark that (5.10) does not link the parameters b and c of both processes, which accounts for the full separation of information.⁶⁾ This is a direct consequence of the use of Jeffreys' rule, which is invariant with respect to differentiable one-to-one transformations of θ [see Zellner (1971, p. 49)],⁷⁾ and would not be discovered with a general prior structure on θ . In particular, any prior density on θ implying stochastic links between b and c could invalidate ZKD's conclusions in their Section 5 as a Bayesian cut would no longer be operated. A classical cut, however, completely ignores such stochastic links and would still exist (conform to the results in Section 4 of ZKD).

The advantage of working in a parameterization for which a cut obtains is that from the outset we know which prior structure [namely the one in (3.6)] will validate focusing on the conditional model only, without first having to conduct our posterior analysis on the whole joint model.

5.2. A Simple Model with Time Effects

5.2.1. The Reduced Form

Let us consider a very simple error-components model with only time-specific effects, where observations on $x_{it} = (y_{it} \ z_{it})'$ are generated independently over time by

$$x_{it} = \mu + w_t + v_{it}, \quad (5.11)$$

where $w_t \perp v_{it}$ with

$$w_t \sim f_N^2(w_t | 0, B),$$

and

$$v_{it} \sim f_N^2(v_{it} | 0, C)$$

independently over individuals. Covariance matrices B and C are both of dimension 2×2 and, respectively, PSDS and PDS. Using the notational convention (2.1)-(2.4), we can write

$$x_t \sim f_N^{2N}(x_t | \mu \otimes e_N, B \otimes e_N e_N' + C \otimes I_N), \quad (5.12)$$

where e_N is an N-dimensional vector of ones. Partitioning

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \text{ and } C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix},$$

we factorize as follows:

$$D(y_t | z_t, b) = f_N^N(y_t | \gamma_1 z_t + \gamma_2 \bar{z}_t e_N + \gamma_3 e_N, \sigma^2 I_N + d e_N e_N') \quad (5.13)$$

$$D(z_t | c) = f_N^N(z_t | \mu_2 e_N, b_{22} e_N e_N' + c_{22} I_N), \quad (5.14)$$

with

$$\gamma_1 = \frac{c_{12}}{c_{22}}, \gamma_2 = N \frac{b_{12} - \gamma_1 b_{22}}{c_{22} + N b_{22}}, \gamma_3 = \mu_1 - (\gamma_1 + \gamma_2) \mu_2, \sigma^2 = c_{11} - \frac{c_{12}^2}{c_{22}},$$

$$d = b_{11} - \frac{N b_{12}^2 + 2 b_{12} c_{12} - b_{22} \frac{c_{12}^2}{c_{22}}}{c_{22} + N b_{22}},$$

and defining the mean of z_t over individuals at time t as

$$\bar{z}_t = \frac{1}{N} e_N' z_t.$$

Let us now examine the effect of various restrictions imposed on B.

Restriction 1.1: $b_{12} = 0$ and $b_{11}, b_{22} \geq 0$, which implies independence of the (possible) time effects in y_{it} and z_{it} , and leads to the following exact restriction

$$\gamma_2 = -\gamma_1 \frac{Nb_{22}}{c_{22} + Nb_{22}}, \quad (5.15)$$

linking the parameters in (5.13) and (5.14), thus preventing any cut (even a classical one) and any of the three types of exogeneity.

Restriction 1.2: $b_{11} = b_{12} = 0$ and $b_{22} \geq 0$, allowing only for a time effect on z_{it} in the joint process. We now have two exact restrictions, namely (5.15) and also

$$d = \gamma_1^2 \frac{b_{22}c_{22}}{c_{22} + Nb_{22}}, \quad (5.16)$$

so that, again, a cut is impossible.

Restriction 1.3: $b_{22} = b_{12} = 0$ and $b_{11} \geq 0$, which means the time effect can only influence y_{it} (it is now absent from the marginal process). This gives us $\gamma_2 = 0$, whereas we also lose the first part of the covariance matrix in (5.14). These two restrictions on the original parameters translate into two zero restrictions on the parameters in (5.13) and (5.14) and leave the latter variation-free. If our prior distribution also induces stochastic independence as in (3.6), then, in view of the independence over time as in (3.4), we have all types of cuts and z_t is predictively exogenous over the entire future. Strong exogeneity will depend on the parameters of interest; if we focus on e.g. the regression coefficient γ_1 , then z_t is also strongly exogenous.

Now compare these results with a very simple time-series model obtained by either "collapsing" or "aggregating" the previous one, namely the bivariate Normal model as in Example 3.1 of EHR:

$$\begin{pmatrix} y_{.t} \\ z_{.t} \end{pmatrix} \sim f_N^2 \left[\begin{pmatrix} y_{.t} \\ z_{.t} \end{pmatrix} \middle| \begin{pmatrix} \mu_{*1} \\ \mu_{*2} \end{pmatrix}, \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} \right], \quad (5.17)$$

where, in the case of "collapsing" (i.e. $i = N = 1$) we have implicitly defined $y_{.t} = y_{1t}$, $z_{.t} = z_{1t}$, $\mu_{*j} = \mu_j$ ($j = 1, 2$), and $\Omega = (\omega_{k\ell}) = B + C$ ($k, \ell = 1, 2$), whereas in the case of aggregating over individuals the interpretation is $y_{.t} = e_N' y_t$, $z_{.t} = e_N' z_t$, $\mu_{*j} = N\mu_j$ ($j = 1, 2$), and $\Omega = N^2B + NC$. As now both error components become indistinguishable, we have the usual result of variation-free parameters in the $y_{.t}|z_{.t}$ and $z_{.t}$ processes, as in EHR, even under restrictions on B , such as Restrictions 1.1 and 1.2 that prevent both classical and Bayesian cuts in the error-components model. This absence of exact links is, of course, not a sufficient, but merely a necessary condition for a Bayesian cut.

This example illustrated some of the differences that may occur between pure time-series models and error-components models, where links between individuals may have an additional (and decisive) influence on the existence of a cut.

5.2.2. The Structural Form

Consider now the simple structural model that has the model in 5.2.1 as its reduced form:

$$y_{it} = \beta z_{it} + \alpha_t + u_{1it} \quad (5.18)$$

$$z_{it} = \eta_t + u_{2it} \quad (5.19)$$

with $\alpha_t, \eta_t \perp u_{it}$ if $u_{it} = (u_{1it} \ u_{2it})'$ and where

$$\begin{pmatrix} \alpha_t \\ \eta_t \end{pmatrix} \sim f_N^2 \left[\begin{pmatrix} \alpha_t \\ \eta_t \end{pmatrix} \middle| \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha\alpha} & \sigma_{\alpha\eta} \\ \sigma_{\eta\alpha} & \sigma_{\eta\eta} \end{pmatrix} \right] \quad (5.20)$$

$$u_{it} \sim f_N^2(u_{it} | 0, \Sigma) \quad (5.21)$$

with $\Sigma = (\sigma_{k\ell})$ for $k, \ell = 1, 2$.

In terms of the reduced form, this implies

$$\mu = 0 \quad (5.22)$$

$$B = L' \begin{bmatrix} \sigma_{\alpha\alpha} & \sigma_{\alpha\eta} \\ \sigma_{\eta\alpha} & \sigma_{\eta\eta} \end{bmatrix} L \quad (5.23)$$

$$C = L' \Sigma L \quad (5.24)$$

with the triangular matrix

$$L = \begin{bmatrix} 1 & 0 \\ \beta & 1 \end{bmatrix}. \quad (5.25)$$

Considering the factorization in (5.13) and (5.14) we now obtain

$$\gamma_1 = \beta + \frac{\sigma_{12}}{\sigma_{22}}, \quad \gamma_2 = N \frac{\sigma_{\alpha\eta} - \frac{\sigma_{12}}{\sigma_{22}} \sigma_{\eta\eta}}{\sigma_{22} + N\sigma_{\eta\eta}}$$

$$\gamma_3 = 0, \quad \sigma^2 = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}, \quad d = \sigma_{\alpha\alpha} - \frac{N\sigma_{\alpha\eta}^2 + 2\sigma_{\alpha\eta}\sigma_{12} - \frac{\sigma_{12}^2}{\sigma_{22}} \sigma_{\eta\eta}}{\sigma_{22} + N\sigma_{\eta\eta}}$$

for the parameters in the conditional model, whereas the marginal one will have zero mean and covariance matrix $(\sigma_{\eta\eta} e_N e_N' + \sigma_{22} I_N)$. Remark that our structural model has seven parameters, whereas six parameters suffice to describe the reduced form (with zero mean), so we expect that at least some restrictions can be accommodated without destroying the possibility of having a cut. Let us examine various situations.

Restriction 2.1: $\sigma_{\alpha\eta} = 0$.

This somewhat changes the expressions for γ_2 and d , without, however, inducing any exact links between the parameterizations of both models.

Provided we have prior independence between $(\gamma_1, \gamma_2, \sigma^2, d)$ and $(\sigma_{\eta\eta}, \sigma_{22})$, we obtain any Bayesian cut in view of the static character of the model as in (3.4), and we have predictive exogeneity of z_t for all sample and future periods. If our interest focuses on β , however, we have no strong (nor weak) exogeneity, as β cannot be recovered from $(\gamma_1, \gamma_2, d, \sigma^2)$.

Restriction 2.2: $\sigma_{12} = 0$.

This restriction affects all parameters in the conditional model, again without introducing any exact links. Under prior independence as in (3.6) we obtain predictive exogeneity, and, as now $\gamma_1 = \beta$, we even have strong exogeneity for β and all pairs (T,s).

Remark that this is the situation referred to in Mundlak (1978), who argues that under correlation between α_t and z_{it} one should include a term in \bar{z}_t in the structural equation for y_t [obtained from (5.18)] before estimation [see also Hsiao (1986)]. Indeed, we are then, implicitly, using our conditional model in (5.13), from which we can safely infer on β , as we have shown in our framework.

Restriction 2.3: $\sigma_{12} = \sigma_{\alpha\eta} = 0$.

The combination of both restrictions gives us the situation where $\gamma_2 = 0$, without destroying "variation-freeness", whereas also $\gamma_1 = \beta$. Under prior independence we now have strong exogeneity for β and the conditional model we should use here simplifies exactly to the structural equation for y_t .

In this example we can, of course, find the same conclusions as in the reduced form model since both are statistically equivalent. The, somewhat artificial, restriction $\beta = -\sigma_{\alpha\eta}/\sigma_{\eta\eta}$ will, for example, prevent a cut as it is equivalent to Restriction 1.1. It is, however, important to realize that restrictions naturally imposed on the structural form can have rather different consequences than those restrictions one might consider in the reduced form. The restriction $b_{12} = 0$ in the reduced form serves to prevent a cut (unless e.g. $b_{22} = 0$ as well), whereas imposing $\sigma_{12} = 0$ in the structural model has the, rather opposite, effect of making weak and strong exogeneity possible for the structural coefficient β .

5.3. A Spatial Autoregressive Structure

Here we start from the simple dynamic model for time-series data used in Example 3.2 of EHR:

$$y_t = \beta z_t + \epsilon_{1t} \quad (5.26)$$

$$z_t = \delta_1 z_{t-1} + \delta_2 y_{t-1} + \epsilon_{2t} \quad (5.27)$$

where $(\epsilon_{1t} \ \epsilon_{2t})'$ is i.i.d. Normally distributed with mean zero and covariance matrix $\Sigma = (\sigma_{ij})$ ($i, j = 1, 2$).

Let us now generalize the scalars y_t, z_t, ϵ_{1t} and ϵ_{2t} to N -dimensional vectors, where each element refers to a spatial (i.e. geographical) location, e.g. N regions within a country, at time t . The parameters β, δ_1 and δ_2 remain scalars and are constant over time and space. Now we assume that disturbances remain independent over time, but not over locations, as we might feel that some external factors (e.g. weather or soil conditions in an agricultural context or, more generally, things like railway connections or road accessibility) could induce a connection between disturbances in adjoining regions.

Let us, therefore, suppose the following autoregressive spatial process for $E_t = (\epsilon_{1t} \ \epsilon_{2t})'$:

$$E_t = WE_t R + V_t, \quad (5.28)$$

where

$$V_t \sim f_{MN}^{N \times 2}(V_t | 0, \Omega \otimes I_N) \quad (5.29)$$

and we partition

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{22} \\ \omega_{21} & \omega_{12} \end{bmatrix},$$

conformably with Σ .

Here R is a (2×2) matrix generalizing the usual scalar autoregression parameter,⁸⁾ and W is a weighting matrix of dimension $N \times N$ with elements w_{ij} ($i, j = 1, \dots, N$) representing the "connectedness" of the spatial locations with $w_{ii} = 0$ and w_{ij} for $i \neq j$ indicating the degree of dependence among locations. The elements of the matrix $R = (r_{ij})$ ($i, j = 1, 2$) allow us to interconnect these dependences across equations (if r_{12} or $r_{21} \neq 0$). Under the assumption that $|I_{2N} - (R' \Omega W)| \neq 0$, we can deduce from (5.28) and (5.29) that

$$E_t \sim f_{MN}^{N \times 2}(E_t | 0, \Sigma_*) \quad (5.30)$$

with

$$\Sigma_* = [I_{2N} - (R' \Omega W)]^{-1} (\Omega \Omega I_N) [I_{2N} - (R \Omega W')]^{-1} \quad (5.31)$$

partitioned into $N \times N$ blocks as

$$\Sigma_* = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Things simplify if we take R and Ω to be diagonal, which means that Σ_* will have a block-diagonal structure with $\Sigma_{12} = \Sigma_{21} = 0$, and for $i = 1, 2$

$$\Sigma_{ii} = \omega_{ii} [(I_N - r_{ii} W)' (I_N - r_{ii} W)]^{-1} \quad (5.32)$$

The reduced form of the model then factorizes as

$$D(y_t | z_t, X_{t-1}, b) = f_N^N(y_t | \beta z_t, \Sigma_{11}) \quad (5.33)$$

$$D(z_t | X_{t-1}, c) = f_N^N(z_t | \delta_1 z_{t-1} + \delta_2 y_{t-1}, \Sigma_{22}), \quad (5.34)$$

which, superficially, looks like just a multivariate version of Example 3.2 in EHR under $\sigma_{12} = 0$, where a sequential (classical) cut is operated and z_t is found to be weakly exogenous for β . There is, however, a difference emanating from the structure of Σ_{ii} ($i = 1, 2$) in (5.32) involving W , which is sometimes assumed known, but often depends on some additional

parameter ζ , for example $w_{ij} = \exp(-\zeta d_{ij})$ where d_{ij} denotes geographical distance (for $i \neq j$), as e.g. discussed in Anselin (1980).

Consider both of these situations.

Case 1: W is known; now, provided $b = (\beta, \omega_{11}, r_{11})$ and $c = (\delta_1, \delta_2, \omega_{22}, r_{22})$ are prior independent, we have a sequential Bayesian cut and z_t is weakly exogenous for β over all sample periods T and predictively exogenous for $T+1$ only. If, additionally, $\delta_2 = 0$ (noncausality given c), then we have predictive (and strong) exogeneity for all future periods.

Case 2: $W = f(\zeta)$, which implies that now both b and c contain the same parameter ζ . This prevents any type of cut (unless one of the r_{ii} is zero, which would imply that either y_t or z_t in the structural model has no spatial links) even though the error terms ϵ_{1t} and ϵ_{2t} are uncorrelated. z_t is still a valid instrument, but it is not weakly exogenous (nor predictively exogenous) since the information on ζ does not separate over the marginal and conditional processes.

Both cases formally reduce to the EHR example in the pure time-series framework (i.e. $i = N = 1$) where $\Sigma_* = \Sigma = \Omega$, and after aggregation over regions whenever the columns of W sum to one, which leads to

$$\Sigma = N(I_2 - R')^{-1} \Omega (I_2 - R)^{-1}.$$

In such a "collapsed" or aggregated version, we know from EHR that the condition $\sigma_{12} = 0$ is sufficient for (classical) weak exogeneity of z_t for β . Adding prior independence of (β, σ_{11}) and $(\delta_1, \delta_2, \sigma_{22})$ in our framework, we have a sequential cut and the same exogeneity conclusions as in Case 1 above.

This example highlights the fact that pure time-series and "time-series of cross-sections" models can require rather different conditions to operate a cut, as in the latter type of models a cut over time must simultaneously be a cut over (possibly connected) sample units. In other words, not only the time dimension can prevent a cut, but also the features of the model

in the other observational dimension (e.g. space), where the marginal process may contain some information regarding the conditional one.

5.4. An Example of Individual Effects from the Literature

This subsection is devoted to an error-components model with individual effects as introduced in Chamberlain and Griliches (1975) who use a pure cross-section (sibling) data set in which, however, they distinguish two observational dimensions: families, denoted by i , and brothers within each family, denoted by t . The model can be described in the following triangular form, using the notation found in Hsiao (1986)

$$y_{1it} = \beta_1' x_{it} + d_1 h_{it} + u_{1it}, \quad (5.35)$$

$$y_{2it} = \gamma_{21} y_{1it} + \beta_2' x_{it} + d_2 h_{it} + u_{2it}, \quad (5.36)$$

$$y_{3it} = \gamma_{31} y_{1it} + \beta_3' x_{it} + d_3 h_{it} + u_{3it}, \quad (5.37)$$

where y_{jit} denote, respectively, observations on grade of schooling ($j=1$), occupational standing ($j=2$), and income ($j=3$). Apart from the observed variables in x_{it} it is believed that the unobserved "ability" h_{it} influences all three y variables in the model. Its structure incorporates an effect proper to the family (i), and is given as

$$h_{it} = \alpha_i + \omega_{it}, \quad (5.38)$$

with both components of zero mean and α_i independently identically distributed (i.i.d.) across i with variance normalized to unity since the scale of h_{it} is indeterminate, whereas ω_{it} is i.i.d. over both i and t with variance σ_ω^2 . Also, we assume that $E(\alpha_i \omega_{it}) = 0, \forall i, t$.

In addition, we simplify matters by taking $E(u_{jit} u_{kit}) = 0, j \neq k$, and $E(u_{jit}^2) = \sigma_{u_j}^2$ ($j, k = 1, 2, 3$).

Assuming, finally, an i.d.d. Normal structure, we obtain the following factorization of the reduced form:

$$D(y_{1i} | X_i, c) = f_N^T(y_{1i} | X_i, \beta_1, \tilde{\sigma}_{11} I_T + \hat{\sigma}_{11} e_T e_T') \quad (5.39)$$

$$\begin{aligned}
D \begin{bmatrix} y_{2i} \\ y_{3i} \end{bmatrix} \left| x_i, y_{1i}, b \right. &= f_N^{2T} \begin{bmatrix} y_{2i} \\ y_{3i} \end{bmatrix} \left[(I_2 \otimes X_i) \delta + (I_2 \otimes T e_T \bar{x}_i) \zeta + (I_2 \otimes y_{1i}) \eta \right. \\
&\left. + (I_2 \otimes T e_T \bar{y}_{1i}) \lambda, \tilde{\Sigma}_{22} \otimes I_T + \hat{\Sigma}_{22} \otimes e_T e_T' \right], \quad (5.40)
\end{aligned}$$

where we have defined

$$X_i' = (x_{i1} \dots x_{iT}),$$

$$\bar{x}_i = \frac{1}{T} e_T' X_i,$$

and

$$\bar{y}_{1i} = \frac{1}{T} e_T' y_{1i}.$$

Some calculations show that the parameters in b and c are related to the ones in the structural form in the following way:

$$\delta = \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix} - \frac{d_1 \sigma_\omega^2}{\sigma_{u_1}^2 + d_1^2 \sigma_\omega^2} \begin{bmatrix} d_2 \\ d_3 \end{bmatrix} \otimes \beta_1, \quad (5.41)$$

$$\zeta = \frac{-d_1 \sigma_{u_1}^2}{(\sigma_{u_1}^2 + d_1^2 \sigma_\omega^2)(\sigma_{u_1}^2 + d_1^2(\sigma_\omega^2 + T))} \begin{bmatrix} d_2 \\ d_3 \end{bmatrix} \otimes \beta_1, \quad (5.42)$$

$$\eta = \begin{bmatrix} \gamma_{21} \\ \gamma_{31} \end{bmatrix} + \frac{d_1 \sigma_\omega^2}{\sigma_{u_1}^2 + d_1^2 \sigma_\omega^2} \begin{bmatrix} d_2 \\ d_3 \end{bmatrix}, \quad (5.43)$$

$$\lambda = \frac{d_1 \sigma_{u_1}^2}{(\sigma_{u_1}^2 + d_1^2 \sigma_\omega^2)(\sigma_{u_1}^2 + d_1^2(\sigma_\omega^2 + T))} \begin{bmatrix} d_2 \\ d_3 \end{bmatrix}, \quad (5.44)$$

$$\tilde{\Sigma}_{22} = \begin{bmatrix} \sigma_{u_2}^2 & 0 \\ 0 & \sigma_{u_3}^2 \end{bmatrix} + \frac{\sigma_\omega^2 \sigma_{u_1}^2}{\sigma_{u_1}^2 + d_1^2 \sigma_\omega^2} \begin{bmatrix} d_2^2 & d_2 d_3 \\ d_3 d_2 & d_3^2 \end{bmatrix}, \quad (5.45)$$

$$\hat{\Sigma}_{22} = \frac{\sigma_{u_1}^4}{(\sigma_{u_1}^2 + d_1^2 \sigma_\omega^2)(\sigma_{u_1}^2 + d_1^2(\sigma_\omega^2 + T))} \begin{bmatrix} d_2^2 & d_2 d_3 \\ d_3 d_2 & d_3^2 \end{bmatrix} \quad (5.46)$$

for b, and

$$\tilde{\sigma}_{11} = \sigma_{u_1}^2 + d_1^2 \sigma_\omega^2. \quad (5.47)$$

$$\hat{\sigma}_{11} = d_1^2 \quad (5.48)$$

for the parameters in c.

Since the reduced form can be fully described by $(3k+9)$ parameters, and (5.39) and (5.40) count $(5k+12)$ different parameters, some restrictions must be implicitly present. In particular, we find

$$\Sigma = -\lambda \otimes \beta_1 \quad (5.49)$$

and

$$\hat{\sigma}_{11}^{-\frac{1}{2}} \hat{\Sigma}_{22} = \sigma_{u_1}^2 \lambda (d_2 \ d_3). \quad (5.50)$$

which prevents any type of cut as parameters in b and c are not variation-free.

Let us now consider the effect of various restrictions on the exogeneity status of y_{1i} in (5.40).

Restriction 3.1: $\sigma_\omega^2 = 0$, which means that both brothers of the same family have exactly the same ability, or $h_{it} = \alpha_i$. The expressions in (5.41)-(5.48) will simplify; in particular, we now obtain that $\delta = (\beta_2' \ \beta_3')$ and $\eta = (\gamma_{21} \ \gamma_{31})'$, but the restrictions in (5.49) and (5.50) remain, so that a cut is impossible.

Restriction 3.2: $T = 1$, in which case the model "collapses" to only the family dimension. The result of this restriction is that (5.39) and (5.40) now become much simpler:

$$D(y_{1i} | X_i, c_*) = f_N^1(y_{1i} | X_i \beta_1, \sigma_{11}^*) \quad (5.51)$$

$$D \begin{pmatrix} y_{2i} \\ y_{3i} \end{pmatrix} \left| X_i, y_{1i}, b_* \right. = f_N^2 \begin{pmatrix} y_{2i} \\ y_{3i} \end{pmatrix} \left| (I_2 \otimes X_i) \delta_* + y_{1i} \eta_*, \Sigma_{22}^* \right. \quad (5.52)$$

with $\sigma_{11}^* = \tilde{\sigma}_{11} + \hat{\sigma}_{11}$, $\delta_* = \delta + \zeta$, $\eta_* = \eta + \lambda$, and $\Sigma_{22}^* = \tilde{\Sigma}_{22} + \hat{\Sigma}_{22}$.

The parameter set (b_*, c_*) has the same number of free parameters as the corresponding reduced form (namely $3k+6$) and there are no exact restrictions linking b_* and c_* . Given the independence across families i , and if b_* and c_* are prior independent, we conclude that y_{1i} is predictively exogenous for any family i outside the sample based on any set of observations. If, however, our parameters of interest are in the set $(\gamma_{21}, \gamma_{31}, \beta_2, \beta_3)$ we do not have weak or strong exogeneity, as these structural parameters cannot be retrieved from b_* only.

Restriction 3.3: $d_1 = 0$, or no effect of ability on schooling in (5.35),⁹⁾ leading to both ζ and λ being zero, whereas $\delta = (\beta_2' \beta_3')$ and $\eta = (\gamma_{21} \gamma_{31})'$. The parameters are now variation-free, and under prior independence we have predictive exogeneity as well as both weak and strong exogeneity for $(\gamma_{21}, \gamma_{31}, \beta_2, \beta_3)$ for any sample and any out-of-sample predictions of individuals (families).

Restriction 3.4: $(d_2 \ d_3) = 0$, implying that ability does not affect the structural equations for occupation and income. Now both error components are only present in the marginal model, as opposed to the previous case, but the exogeneity conclusions are the same as in the version with $d_1 = 0$.

Note that this model differs in two important respects from the time effects model presented in Subsection 5.2. Firstly, the presence of explanatory variables x_{it} in all three equations (5.35)-(5.37) leads to restrictions of the form (5.49), and, secondly, we use the same unobserved variable h_{it} in all three equations. The latter fact imposes a special structure on the error components; in particular, the matrix corresponding to the individual effects becomes singular (of the form kk' , where k is a vector). This results in links as in (5.50) and contrasts with the general covariance matrix B in (5.11).

Indeed, if we impose B to be of a similar structure (i.e. $b_{12} = \sqrt{b_{11}b_{22}}$) in Subsection 5.2, we implicitly introduce the restriction

$$\gamma_2^2 = N^2 \frac{b_{22}}{c_{22}(c_{22} + Nb_{22})} d, \quad (5.53)$$

which links the parameters in both processes, and is an exact counterpart of (5.50).

6. Concluding Remarks

In the present paper, the concept of exogeneity so frequently used by applied econometricians, either consciously or not, is examined within a Bayesian framework and for models with two observational dimensions.

The latter implies that the existing predominant emphasis on macroeconomic models¹⁰⁾ in this type of literature [see e.g. Koopmans (1950), Marschak (1953), Sims (1972) and EHR] is here broadened somewhat towards including the analysis of models emanating from the microeconomic tradition, such as error-components or spatial models. With the steadily growing availability of e.g. panel data sets, such models seem to gain rapidly in popularity within the economics profession.

An additional reason for considering exogeneity issues in micromodels is that the latter are often relatively close to the theory, as they usually don't suffer from aggregation or feedback issues that might blur the theory implications at the macro level. This means that exogeneity conditions may have a very direct theoretical interpretation, as e.g. in the ZKD production function described in Subsection 5.1, where theoretical grounds exist for imposing zero covariances that imply exogeneity.

Another example is the use of the life-cycle theory in intertemporal labour supply and consumption models in MaCurdy (1983). The outcomes of exogeneity tests may then prove to be more informative about certain theory assumptions than in many macro models.

The whole issue of testing for exogeneity was not taken up in this paper, but classical analyses can be found e.g. in Wu (1973, 1983), Holly (1985) and Smith (1988).

With this somewhat extended class of models in mind, we feel it is natural to distinguish between four different types of cuts, introduced in Subsection 2.2 and borrowed from the more statistical literature in FM (1980a, 1985). Having these different definitions of cut directly allows for more general models than just time-series models and also leads to somewhat weaker definitions than found in EHR, conveying the same intuitive ideas. A cut is certainly not necessary for "not losing any relevant information", but it proves a very useful concept, as it is rather robust with

respect to the prior distribution [compare the remarks concerning mutual exogeneity in FM (1985) and our Subsection 3.2], from which it only requires independence, and it clearly shows under which prior assumptions we can concentrate on the conditional model only (compare the ZKD example in Subsection 5.1), although we have to be careful when using improper prior measures (as explained in footnote 6).

An essentially Bayesian approach is adopted in this paper as it seems natural to formally treat all information in a symmetric way and thus require a full separation of both prior and sample information over marginal and conditional processes. It also becomes very straightforward to deal with nuisance parameters, as discussed in Basu (1977), which means we can define concepts like noncausality given a subset of all parameters, and we can also consider e.g. predictive densities.

The latter leads to a definition of predictive exogeneity, which seems natural in a Bayesian context and is often the only possible type of exogeneity if we start from structural models (see Subsections 5.2.2 and 5.4). The entire issue of parameters of interest is simply irrelevant for purely predictive purposes.

The three concepts of exogeneity used here are nested in the sense that strong exogeneity is just the union of both predictive and weak exogeneity on the same parameterizations.

The latter requirement was added for the sake of presentation, but is by no means a necessity. It may very well occur that we have a specific interest in some (structural) parameterization a for inference purposes, but that a different parameterization a_* , e.g. of a reduced form or just in terms of regression coefficients, could prove most useful for prediction. A generalization of Definition 3 in this sense is, of course, straightforward if we wish to cover such cases.

We noted that prediction issues in a sequential context crucially change if we go from one-period ahead forecasting to predictions further in the future, and we also examined some consequences of treating static rather than dynamic models, since models with independence (over time or over individuals) seem to appear rather frequently in the literature (compare also Subsections 5.1, 5.2 and 5.4).

Remark that exogeneity concepts as defined here (or, for that matter, in EHR) validate the use of the conditional model, either for inference, prediction, or both, but not the use of the structural model, nor of some particular estimation technique [see e.g. EHR and Steel (1987)]. Only if the stochastics of z are either absent or completely separated from those governing the process of the variables we wish to explain (e.g. $\sigma_{01} = \sigma_{02} = 0$ in Subsection 5.1, Restriction 2.3 in Subsection 5.2.2, or Restriction 3.3 in Subsection 5.4) will the conditional model and the relevant equation(s) of the structural model coincide, but even then a cut may be prevented (see e.g. Case 2 in Subsection 5.3). One should, therefore, bear in mind that e.g. weak exogeneity and the valid use of OLS in the structural equation are concepts of a very different nature.

The examples in Section 5 also serve to illustrate several other issues. In particular, imposing restrictions on the reduced form can have rather different exogeneity consequences than restricting structural parameters, as was seen in Subsections 5.2.1 versus 5.2.2.

Also, the presence of two observational dimensions can often prevent a cut, which does exist if we collapse or aggregate to just one of the dimensions. This occurred in Subsections 5.2.1, 5.3 and 5.4. In error-components models the inclusion of additional regressors into the system as in Subsection 5.4 compared to 5.2 can destroy exogeneity properties, whereas a similar effect can be brought on by using the same unobserved variable throughout the system. An example of the latter issue can also be found by comparing Subsections 5.2 and 5.4.

Needless to say, numerous extensions of the basic framework of this paper could be considered. Let us just name a few, without claiming to be exhaustive in any way.

- (i) The treatment of initial conditions (X_0) could be formalized, following the suggestions in Richard (1979) and Engle et al. (1980).
- (ii) The problems (hinted at in footnote 6) with verifying the absence of prior links under improper prior measures should most certainly receive due attention. The use of group analysis, as in Dawid et al. (1973), may be instrumental here.

- (iii) Parameter constancy under changes in the distribution of the conditioning variables z is crucial for policy simulations and leads to a definition of "super exogeneity" in EHR. An explicit discussion of such a concept in our framework seems both feasible and of some interest.
- (iv) Extending the framework in Subsection 2.1 to include also models where the number of individuals N changes over t might be useful, as it would allow us to treat e.g. the "unbalanced" version of the model in Subsection 5.4, i.e. where a subset of the families observed have more than two brothers (e.g. $t = 3, 4$ and 5).

Nevertheless, it is felt that the present analysis, however incomplete, already gives some useful ideas about the specific exogeneity issues arising in models that do not fit the pure time-series framework.

Footnotes

- 1) This is illustrated by some of the examples in the sequel; for the model in Subsection 5.3 i denotes geographical location, and the one in Subsection 5.4 attributes to index t the role of indicating the t^{th} member of some family i , as here one could possibly consider the birth-order of members as inducing a natural ordering.
- 2) See also the discussion in Basu (1977) who uses the term "variation independent" instead of "variation free".
- 3) The term "noncausality" was finally adopted by FM (1985), as it seems ubiquitous in the econometrics literature [see e.g. Zellner (1979), Granger (1980), and Geweke (1984)]. Previously, other terms were used, such as "transitivity" in FM (1980a), which is more general and closer to the statistics tradition, and "self-predictivity" in FM (1980b, Section 4). The latter denominations have the advantage of stressing the difference with the notion of causality in philosophy.
- 4) Under independence as in (3.4), the initial conditions do not affect the sampling process but can only contribute to our prior information.
- 5) Note that the definitions of k_1 and k_2 have an extra term α_0 in ZKD. This seems, however, a misprint (as then $k_1 = k'_1$ and $k_2 = k'_2$ in their notation), as well as their use of R_1 in the equation for k_2 .
- 6) Since θ_1 and θ_2 have some parameters in common on which the prior measure in (5.9) is not uniform, it is rather easy to see that (5.9) prevents a Bayesian cut in θ . However, under improper prior densities of the type used in (5.9) or (5.10), often expressing "diffuse" prior ideas, the factorization into marginal and conditional priors is not well-defined, and the whole issue of prior independence as in (3.6) becomes very tricky since the conditioning event is often not σ -finite (i.e. has infinite probability mass). The use of improper prior dis-

tributions has been shown in Stone and Dawid (1972) to lead to so-called "marginalization paradoxes", analyzed by group theory in Dawid et al. (1973).

- 7) This is, of course, not invariance of the prior density itself, but rather of the rule used to construct the prior distribution. For an extensive discussion of invariance concepts see Hartigan (1964), who calls this particular property "Q-labelling invariance".
- 8) The formulation in (5.28) is a generalization of the single-equation process

$$\epsilon_{1t} = \rho W \epsilon_{1t} + v_{1t}$$

with ρ scalar, as usually adopted in spatial econometrics in univariate cases. Some references in this field are Cliff and Ord (1973), Ord (1975), Hordijk (1979) and Anselin (1980).

- 9) Incidentally, this restriction seems to be corroborated by a classical analysis of the Gorseline (1932) data in Chamberlain and Griliches (1975). Remark that they only treat the "balanced" case in which each family (i) has the same number of brothers (T=2), which is also the case that fits directly into our framework, as outlined in Subsection 2.1.
- 10) With a few notable exceptions, e.g. Smith and Blundell (1986).

References

- Anselin, L., 1980, Estimation methods for spatial autoregressive structures. A study in spatial econometrics, Regional science dissertation & monograph series no. 8 (Cornell University, Ithaca).
- Barndorff-Nielsen, O., 1978, Information and exponential families in statistical theory (Wiley, New York).
- Basu, D., 1955, On statistics independent of a complete sufficient statistic, Sankhya 15, 377-380.
- Basu, D., 1958, On statistics independent of a sufficient statistic, Sankhya 20, 223-226.
- Basu, D., 1977, On the elimination of nuisance parameters, Journal of the American Statistical Association 72, 355-366.
- Chamberlain, G. and Z. Griliches, 1975, Unobservables with a variance-components structure: Ability, schooling and the economic success of brothers, International Economic Review 16, 422-449.
- Chib, S., R.C. Tiwari and S.R. Jammalamadaka, 1988, Bayes prediction in regressions with elliptical errors, Journal of Econometrics 38, 349-360.
- Cliff, A. and K. Ord, 1973, Spatial autocorrelation (Pion, London).
- Dawid, A.P., M. Stone and J.V. Zidek, 1973, Marginalization paradoxes in Bayesian and structural inference, Journal of the Royal Statistical Society B 35, 189-223.
- de Finetti, B., 1974, Theory of Probability, Vol. 1 (Wiley, New York).
- de Finetti, B., 1975, Theory of Probability, Vol. 2 (Wiley, New York).

- Drèze, J.H. and J.F. Richard, 1983, Bayesian analysis of simultaneous equation systems, Ch. 9 in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics, Vol. 1 (North-Holland, Amsterdam).
- Engle, R.F., D.F. Hendry and J.F. Richard, 1980, Exogeneity, causality and structural invariance in econometric modelling, CORE discussion paper no. 8038 (Université Catholique de Louvain, Louvain-la-Neuve).
- Engle, R.F., D.F. Hendry and J.F. Richard, 1983, Exogeneity, Econometrica 51, 277-304.
- Florens, J.P. and M. Mouchart, 1977, Reduction of Bayesian experiments, CORE discussion paper no. 7737 (Université Catholique de Louvain, Louvain-la-Neuve).
- Florens, J.P. and M. Mouchart, 1980a, Initial and sequential reduction of Bayesian experiments, CORE discussion paper no. 8015 (Université Catholique de Louvain, Louvain-la-Neuve).
- Florens, J.P. and M. Mouchart, 1980b, Conditioning in econometric models, CORE discussion paper no. 8042 (Université Catholique de Louvain, Louvain-la-Neuve).
- Florens, J.P. and M. Mouchart, 1982, A note on noncausality, Econometrica 50, 583-591.
- Florens, J.P. and M. Mouchart, 1985, Conditioning in dynamic models, Journal of Time Series Analysis 6, 15-34.
- Florens, J.P., M. Mouchart and J.M. Rolin, 1987, Noncausality and marginal Markov processes, CORE discussion paper no. 8706 (Université Catholique de Louvain, Louvain-la-Neuve).
- Geweke, J., 1984, Inference and causality in economic time series models, Ch. 18 in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam).

Gorseline, D.E., 1932, The effect of schooling upon income (Indiana University Press, Bloomington).

Granger, C.W.J., 1980, Testing for causality: A personal viewpoint, Journal of Economic Dynamics and Control 2, 329-352.

Hartigan, J.A., 1964, Invariant prior distributions, Annals of Mathematical Statistics 35, 836-845.

Holly, A., 1985, Testing for exogeneity: A survey, cahier no. 8506 (Ecole des Hautes Etudes Commerciales, Lausanne).

Hordijk, L., 1979, Problems in estimating econometric relations in space, Papers of the Regional Science Association 42, 99-115.

Hsiao, C., 1986, Analysis of panel data (Cambridge University Press, Cambridge).

Koopmans, T.C., 1950, When is an equation system complete for statistical purposes?, Ch. 17 in: T.C. Koopmans, ed., Statistical inference in dynamic economic models (Wiley, New York).

MaCurdy, T.E., 1983, A simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty, International Economic Review 24, 265-289.

Marschak, J., 1953, Economic measurements for policy and prediction, Ch. 1 in: W.C. Hood and T.C. Koopmans, eds., Studies in econometric method (Wiley, New York).

Mouchart, M. and J.M. Rolin, 1984, A note on conditional independence with statistical applications, Statistica 44, 557-584.

Mundlak, Y., 1978, On the pooling of time series and cross section data, Econometrica 46, 69-85.

Orcutt, G.H., 1952, Towards a partial redirection of econometrics, Review of Economics and Statistics 34, 195-213.

Ord, K., 1975, Estimation methods for models of spatial interaction, Journal of the American Statistical Association 70, 120-126.

Poirier, D.J., 1988, Frequentist and subjectivist perspectives on the problems of model building in economics, Economic Perspectives 2, 121-144.

Richard, J.F., 1979, Exogeneity, inference and prediction in so-called incomplete dynamic simultaneous equation models, CORE discussion paper no. 7922 (Université Catholique de Louvain, Louvain-la-Neuve).

Sims, C.A., 1972, Money, income and causality, American Economic Review 62, 540-552.

Smith, R.J., 1988, Asymptotically optimal tests using limited information and testing for weak exogeneity, mimeo (University of Manchester, Manchester).

Smith, R.J. and R.W. Blundell, 1986, An exogeneity test for a simultaneous equation Tobit model with an application to labour supply, Econometrica 54, 679-685.

Steel, M.F.J., 1987, A Bayesian analysis of multivariate exogeneity: A Monte Carlo approach, Nouvelle série no. 168 (Université Catholique de Louvain, Louvain-la-Neuve).

Stone, M. and A.P. Dawid, 1972, Un-Bayesian implications of improper Bayes inference in routine statistical problems, Biometrika 59, 369-375.

Wu, D.M., 1973, Alternative tests of independence between stochastic regressors and disturbances, Econometrica 41, 733-750.

Wu, D.M., 1983, Tests of causality, predeterminedness and exogeneity, International Economic Review 24, 547-558.

Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York).

Zellner, A., 1979, Causality and econometrics, Ch. 1 in: K. Brunner and A.H. Meltzer, eds., Three aspects of policy and policymaking (North-Holland, Amsterdam).

Zellner, A. and C. Hong, 1988, Forecasting international growth rates using Bayesian shrinkage and other procedures, Journal of Econometrics 40, 183-202.

Zellner, A., J. Kmenta and J.H. Drèze, 1966, Specification and estimation of Cobb-Douglas production function models, Econometrica 34, 784-795.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

No.	Author(s)	Title
8801	Th. van de Klundert and F. van der Ploeg	Fiscal Policy and Finite Lives in Interdependent Economies with Real and Nominal Wage Rigidity
8802	J.R. Magnus and B. Pesaran	The Bias of Forecasts from a First-order Autoregression
8803	A.A. Weber	The Credibility of Monetary Policies, Policy-makers' Reputation and the EMS-Hypothesis: Empirical Evidence from 13 Countries
8804	F. van der Ploeg and A.J. de Zeeuw	Perfect Equilibrium in a Model of Competitive Arms Accumulation
8805	M.F.J. Steel	Seemingly Unrelated Regression Equation Systems under Diffuse Stochastic Prior Information: A Recursive Analytical Approach
8806	Th. Ten Raa and E.N. Wolff	Secondary Products and the Measurement of Productivity Growth
8807	F. van der Ploeg	Monetary and Fiscal Policy in Interdependent Economies with Capital Accumulation, Death and Population Growth
8901	Th. Ten Raa and P. Kop Jansen	The Choice of Model in the Construction of Input-Output Coefficient Matrices
8902	Th. Nijman and F. Palm	Generalized Least Squares Estimation of Linear Models Containing Rational Future Expectations
8903	A. van Soest, I. Woittiez, A. Kapteyn	Labour Supply, Income taxes and Hours Restrictions in The Netherlands
8904	F. van der Ploeg	Capital Accumulation, Inflation and Long-Run Conflict in International Objectives
8905	Th. van de Klundert and A. van Schaik	Unemployment Persistence and Loss of Productive Capacity: a Keynesian Approach
8906	A.J. Markink and F. van der Ploeg	Dynamic Policy Simulation of Linear Models with Rational Expectations of Future Events: A Computer Package
8907	J. Osiewalski	Posterior Densities for Nonlinear Regression with Equicorrelated Errors
8908	M.F.J. Steel	A Bayesian Analysis of Simultaneous Equation Models by Combining Recursive Analytical and Numerical Approaches

- 8909 F. van der Ploeg Two Essays on Political Economy
(i) The Political Economy of Overvaluation
(ii) Election Outcomes and the Stockmarket
- 8910 R. Gradus and Corporate Tax Rate Policy and Public
A. de Zeeuw and Private Employment
- 8911 A.P. Barten Allais Characterisation of Preference
Structures and the Structure of Demand
- 8912 K. Kamiya and Simplicial Algorithm to Find Zero Points
A.J.J. Talman of a Function with Special Structure on a
Simplotope
- 8913 G. van der Laan and Price Rigidities and Rationing
A.J.J. Talman
- 8914 J. Osiewalski and A Bayesian Analysis of Exogeneity in Models
M.F.J. Steel Pooling Time-Series and Cross-Section Data

P.O. BOX 90153. 5000 LE TILBURG. THE NETHERLANDS

Bibliotheek K. U. Brabant



17 000 01113878 2