ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 27 — 1992

Weak Exogeneity in Overreduced Sequential Models

Mark F. J. STEEL *

ABSTRACT. – In a formal statistical framework we examine the consequences of misspecifying sequential econometric models for their weak exogeneity properties. Modelling is viewed as a reduction sequence and the Bayesian paradigm is adopted. The central aim is inference on certain parameters of interest, which, in practice, almost invariably relies on some weak exogeneity assumption. Sufficient conditions for the preservation of weak exogeneity despite reduction errors are given, and several examples illustrate these findings. We conclude that some types of misspecification are much more dangerous than others, and we advocate attentive dynamic modelling.

Exogénéité faible dans les modèles séquentiels sous erreurs de spécification

RÉSUMÉ. – Dans un cadre statistique formalisé, nous examinons les conséquences d'une mauvaise spécification des modèles séquentiels pour leurs propriétés d'exogénéité faible. On donne des conditions suffisantes pour la conservation de l'exogénéité faible malgré les erreurs de réduction. On en conclut que certains types d'erreurs de spécification sont beaucoup plus dangereux que d'autres et en particulier qu'il faut apporter beaucoup de soins à la modélisation dynamique.

---

* Mark F. J. STEEL: Department of Economics, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands.

Stimulating discussions with Jayasri Dutta, Gary Koop, Jacek Osiewalski and Dale Poirier proved very clarifying. In addition, two anonymous referees provided very helpful comments. Needless to say, any errors are my own. Part of this work was completed at the Institute of Statistics and Decision Sciences, Duke University, and their hospitality as well as a travel grant from the Netherlands Organization for Scientific Research (NWO) are gratefully acknowledged. The author is supported by a senior research fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW).

1 Introduction

The importance of exogeneity for the practice of econometric modelling was stressed in ENGLE *et al.* [1983] and HENDRY and RICHARD [1982, 1983]. Many facets of the concepts have been illuminated in the existing literature: ENGLE *et al.* [1983] apply it to dynamic simultaneous equation models, whereas SMITH and BLUNDELL [1986] consider a simultaneous tobit model, and OSIEWALSKI and STEEL [1989] explore the implications for models pooling time-series and cross section data.

However, all these analyses start from the assumption that the models used are "correctly" specified. As we know that the latter is bound to be a very heroic assumption indeed in econometrics, we feel it serves a purpose to explicitly deviate from the "axiom of correct specification" (see LEAMER [1978]). In particular, we set out to examine here exactly what happens to weak exogeneity properties of a statistical model when we happen to misspecify that model. Of course, our intuition may tell us that something will go wrong, but, in our opinion, it is of some use to evaluate the consequences of various types of misspecification in a somewhat formalized framework.

This framework is constructed using the Bayesian paradigm, as we feel this provides a more natural way of looking at exogeneity, taking both exact and stochastic links between the parameters into account (with possibly some abuse of the term, we consider exact links to include inequality restrictions). It also allows us to use the concept of conditional independence, which gives access to a powerful toolbox existing in probability theory (see e.g. CHUNG [1968], MOUCHART and ROLIN [1984] or FLORENS *et al.* [1990]).

Now let us more formally define what misspecification means to a Bayesian. The relevance of this question may be illustrated by the following very simple example.

EXAMPLE 0: Bayesian misspecification.

Let us consider Example 3.2 from ENGLE *et al.* [1983], where the following model is formulated for the scalar variables y_t and z_t

$$y_t = \beta z_t + \varepsilon_{1t} \tag{1.1}$$

$$z_t = \delta_1 z_{t-1} + \delta_2 y_{t-1} + \varepsilon_{2t} \tag{1.2}$$

with the error terms independently and identically distributed according to a Normal law with mean zero and covariance matrix $\Sigma = (\sigma_{ij})$; $i, j \in \{1, 2\}$, denoted as $f_N^2(\varepsilon_t \mid 0, \Sigma)$. Such a model is **not** perceived here as reflecting an objective "truth", corresponding to some mechanism in the outside world, but merely as a useful way of looking at observables, a useful "window" in POIRIER's [1988] terminology. There are many other windows that we can think of, and whenever a model is postulated here as the "correct" specification, this just means that it is the **most useful window** for our purposes at hand.

The concept of weak exogeneity will be explained in Section 3, but, roughly, we can say it boils down to the absence of links (*i. e.* independence) between the parameterizations of the conditional process for y_t given z_t and the marginal process for z_t, combined with the possibility to retrieve our parameters of interest from those of the conditional model. Such parameters of interest only have a meaning relative to a particular model. Here they will generally be defined in the context of the most useful window. In this example β and σ_{11} are obvious candidates. Conditions ensuring weak exogeneity of z_t for the purpose of inference on β and σ_{11} are $\sigma_{12}=0$ and prior independence of the resulting parameterizations λ_1 and λ_2 in

$$\text{(1.3)} \qquad D(y_t \mid z_t, I_{t-1}, \lambda_1) = f_N^1(y_t \mid \beta z_t, \sigma_{11})$$

$$\text{(1.4)} \qquad D(z_t \mid I_{t-1}, \lambda_2) = f_N^1(z_t \mid \delta_1 z_{t-1} + \delta_2 y_{t-1}, \sigma_{22}),$$

where $I_{t-1} = (z_{t-1}, y_{t-1})$ is the information set relevant at time t.

If we now introduce the "misspecification" of leaving out y_{t-1} in (1.4) or of "falsely" imposing $\delta_2 = 0$ (*i. e.* a noncausality restriction, as explained in e. g. FLORENS and MOUCHARD [1982] and ENGLE *et al.* [1983]), the marginal process will become

$$\text{(1.5)} \qquad D(z_t \mid I_{t-1}^*, \mu_2) = f_N^1(z_t \mid (\delta_1 + \delta_2 \beta) z_{t-1}, \sigma_{22} + \delta_2^2 \sigma_{11}),$$

with $I_{t-1}^* = z_{t-1}$ and the parameterization is now changed to

$$\text{(1.6)} \qquad \mu_2 = (\delta_1 + \delta_2 \beta, \sigma_{22} + \delta_2^2 \sigma_{11})$$

instead of $\lambda_2 = (\delta_1, \delta_2, \sigma_{22})$. The window used is now reduced to a subset of the most useful one and we distinguish two ways of interpreting this reduction.

One way is to view $\delta_2 = 0$ as an **exact prior restriction**, which implies that prior independence of λ_1 and μ_2 still holds, and that weak exogeneity carries over from the original model to the reduced one. In this case, we can't really talk of misspecification as the reduction of the window was induced by our prior ideas, which can hardly be labelled as "wrong"; they are what they are, and, barring incoherency, we should not tamper with them. Taken strictly, this, of course, implies that expanding the window, even in the face of very strong disagreement between sample and prior information, is **always** ruled out. Apart from being in flagrant contradiction with the actual practice of econometric modelling, this interpretation does not seem warranted from any pragmatic view on methodology as it essentially prevents critical experimentation.

We, therefore, propose to follow the suggestion in LEAMER [1978], LINDLEY [1982], SMITH [1984] and POIRIER [1988] to always retain the possibility of redefining the window, if necessary, by not being too dogmatic in the interpretation of the prior. LINDLEY [1982] introduced the recipe to avoid literally assigning zero prior probability to any open set as **Cromwell's rule** and POIRIER [1988] ranks it as his fifth pragmatic principle of model building. For our purposes here, this rule implies that $\delta_2 = 0$ is not formally treated as a prior restriction, forever excluding the possibility to enlarge the window, but just reflecting our prior belief that δ_2 will be **near** zero, giving the data a chance to revise the prior idea that y_{t-1} does not matter in (1.4). If enough data evidence is collected, we will eventually find out that (1.4) leads to a **more useful** window than (1.5) (exactly how this occurs is beyond this discussion), and in this sense (1.5) is misspecified. This is the light in which terms like "correct" or "misspecified" model should be seen here. More in particular, we shall focus on misspecification through invalid reductions. [1] From this point of view, it is obvious that μ_2 will now no longer be prior independent from λ_1, in general, so that the misspecification has destroyed weak exogeneity for the parameters of interest, defined in the context of the "correctly" reduced model.

Throughout the paper we assume that sufficient conditions hold to ensure weak exogeneity at the level of the "correct" model. For simplicity, we have limited our examples to the Gaussian domain, which means that zero covariances typically appear in these sufficient conditions. However, covariance restrictions should not automatically be assimilated to exogeneity conditions, since they are often not sufficient if we leave either the Gaussian or the time-series framework (for the latter, see OSIEWALSKI and STEEL [1989]) and they are certainly not necessary conditions (see e.g. STEEL [1987]). In fact, even a Bayesian cut (see Section 3) is not strictly necessary to avoid loss of information by using only the conditional model, but the definition of weak exogeneity is, nevertheless, based on the concept of cut, as this greatly facilitates the analysis (see the discussion of "mutual exogeneity" in FLORENS and MOUCHART [1985]).

Finally, it is supposed here that a formal examination of weak exogeneity is conducted, *i.e.* based on the full model for both y_t and z_t. If only an informal test of exogeneity is performed, based on the stability of the inference on the conditional model in a changing environment, as explained in ENGLE *et al.* [1983], we are not required to specify the marginal process, and, thus, its possible overreduction (one of our four types, to be introduced later) becomes irrelevant.

Section 2 describes the statistical framework and the various types of overreduction that we wish to consider. Section 3 briefly discusses Bayesian cuts and defines weak exogeneity, whereas Section 4 evaluates the consequences of different reduction errors for these exogeneity properties. A fifth section seeks to illuminate matters by providing simple examples, and a final section groups some conclusions for the practice of econometric modelling.

---

1. Needless to say, there are other forms of misspecification. In particular, one could falsely assume certain functional forms (e.g. linearity) or certain stochastic characteristics (e.g. Normality). However, as we do not make such assumptions at the theoretical level, we abstract from these sources of misspecification. Incidentally, we reason in terms of densities instead of general σ-fields, which is, in itself, already a (possibly false) assumption. The latter type of misspecification is not explicitly considered either.

2 The Statistical Model

2.1 The Reduction Sequence

In line with the methodology set forth in HENDRY and RICHARD [1982, 1983], we view the process of econometric modelling as a series of successive reduction steps through marginalization and conditionalization of an immensely complicated process that is supposed to jointly dictate the behaviour of all variables we can possibly think of. We hasten to add that such a system, often referred to as the **"data generating process"** (DGP), is, of course, but a convenient fiction that is not claimed to have any tangible existence. It is, however, of great value in clearly formulating our ideas.

If we assume that such a DGP can conveniently be characterized in terms of **densities** (denoted by D) defined on a matrix of T observations for all the variables in the (large) vector $w_t\,(t: 1 \rightarrow \mathrm{T})$:

(2.1) $$\mathrm{W}_\mathrm{T}^1 = (w_1, \ldots, w_\mathrm{T})',$$

we can assimilate the DGP to the following joint data density:

(2.2) $$\mathrm{D}(\mathrm{W}_\mathrm{T}^1 \mid \mathrm{W}_0, \vartheta),$$

where W_0 denotes a (possibly infinitely dimensional) matrix of initial conditions and $\vartheta \in \Theta$ is a sufficient parameterization of the process.

In this paper we shall focus upon the sequential representation of our **sampling theory model** in (2.2), knowing that the latter can be written as a product over $t: 1 \rightarrow \mathrm{T}$ of **sequential models**

(2.3) $$\mathrm{D}(w_t \mid \mathrm{W}_{t-1}, \vartheta),$$

where we have implicitly defined the full information set available at time t by

(2.4) $$\mathrm{W}_{t-1} = (\mathrm{W}_0, w_1, \ldots, w_{t-1})'.$$

In a **Bayesian** framework we opt for full symmetry between observations and parameters and, therefore, extend (2.3) with a so-called **prior** probability on the parameter space Θ, denoted by $\mathrm{D}(\vartheta \mid \mathrm{W}_0)$ as it will often depend on initial conditions, that may include e.g. a previous sample. We obtain

(2.5) $$\mathrm{D}(w_t, \vartheta \mid \mathrm{W}_{t-1}) = \mathrm{D}(w_t \mid \mathrm{W}_{t-1}, \vartheta)\,\mathrm{D}(\vartheta \mid \mathrm{W}_0),$$

where our prior density should not depend on any observations in the sample under consideration.

In practice, the Bayesian model in (2.5) will be far too large as w_t denotes all the variables in our DGP, so that we consider **marginalizing** w_t to end up with a subset of variables, say, $x_t \subset w_t$, that is of "manageable" size and includes all variables that we are interested in modelling.

Let us take x_t to describe e.g. a sector of the economy and we assume that, if we define $w_t' = (x_t' \ \tilde{w}_t')$, then

(*a*) $$x_t \perp \tilde{w}_t \mid W_{t-1}, \vartheta$$

will hold, so that we can concentrate on x_t instead of the full vector w_t.[2] Note that conditional independence of random variables, say *a* and *b*, given *c*, is denoted by $a \perp b \mid c$. Typically, the sampling density for x_t will then only depend on a small subset of the entire past of the economy, say I_{t-1}, where $W_0 \subset I_{t-1} \subset W_{t-1}$, so that we consider only $D(x_t, \vartheta \mid I_{t-1})$, which is warranted if:

(*b*) $$x_t \perp W_{t-1} \mid I_{t-1}, \vartheta.$$

We then reparameterize from ϑ to $(\lambda, \tilde{\vartheta})$ such that λ is sufficient for the process of x_t and $\tilde{\vartheta}$ for $\tilde{w}_t$, *i.e.*

(*c*.1) $$x_t \perp \vartheta \mid I_{t-1}, \lambda$$

(*c*.2) $$\tilde{w}_t \perp \vartheta \mid W_{t-1}, \tilde{\vartheta},$$

and we finally assume that λ and $\tilde{\vartheta}$ are prior independent

(*d*) $$\lambda \perp \tilde{\vartheta} \mid W_0,$$

so that we can focus our attention on the **reduced** (marginalized) Bayesian model

(2.6) $$D(x_t, \lambda \mid I_{t-1}) = D(x_t \mid I_{t-1}, \lambda)\, D(\lambda \mid W_0),$$

for the purpose of inference on λ. We can, of course, short-circuit this derivation by directly specifying $D(x_t \mid I_{t-1}, \lambda)$ as our "distribution of interest" but we feel parameters of interest are usually defined at the earlier level of $D(x_t \mid \tilde{w}_t, W_{t-1}, \vartheta)$ and are thus only contained in the reduced model if

(2.7) $$x_t \perp (\tilde{w}_t, W_{t-1}, \vartheta) \mid I_{t-1}, \lambda,$$

which is equivalent to (*a*), (*b*) and (*c*.1) jointly. Finally, assumptions (*c*.2) and (*d*) ensure that no information about λ can be obtained from the process for $\tilde{w}_t$.

Let us now consider further reduction by **conditioning**. Within the constraints of the "window" thus chosen, one would often (if x_t contains more variables than we explicitly wish to model) look for so-called **exogeneity conditions**, *i.e.* conditions that validate treating part of the variables in x_t

2. In fact, this reflects assigning a "structural" status to the parameters of the conditional process $D(x_t \mid \tilde{w}_t, W_{t-1}, \vartheta)$ rather than those of the marginal process $D(x_t \mid W_{t-1}, \vartheta)$.

(say, $z_t \subset x_t$) as "given" for the purpose of either estimation *per se*, estimation plus conditional forecasting or estimation plus policy predictions. The concepts referred to are weak, strong and super exogeneity, respectively, as explained in ENGLE *et al.* [1983] within a classical context. See STEEL and RICHARD [1991] or OSIEWALSKI and STEEL [1989] for a Bayesian discussion of exogeneity. Section 3 will address the issue briefly, tailored to the particular problem at hand and focusing on weak exogeneity.

2.2 Overreduction

The reduction steps involved in going from the DGP (2.3) to our reduced econometric model used in (2.6) might, of course, entail some loss of information. In practice, however, a "small" loss of information is often compensated by a positive evaluation of parsimony, which enhances communication, interpretation and computational facility of our models. In addition, empirical experience suggests that parsimonious models often display particularly stable characteristics and usually forecast much better than their large unrestricted counterparts.

However, the situation can easily arise that one imposes "invalid" reductions, in the sense that an **essential** part of the information present in the DGP is not communicated to the econometric model. In that case one really chooses a "less useful window" to view the world and we then talk of **overreduction**. Especially if one uses a less methodical approach to modelling or if one opts for the specific-to-general route, this is certainly not an unlikely event.

We shall, at this stage, not address the question of how to measure the size of the (inevitable) information loss, nor when to consider it large enough to talk of overreduction, but we shall content ourselves with stating that it can lead to two basic consequences:

a. **Contemporaneous** misspecification; *i.e.* x_t lacks one or more variables that are present in w_t and crucially influence the processes we set out to examine (marginal overreduction). If we partition

$$x_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix}, \tag{2.8}$$

where y_t groups the variables we are actually modelling and z_t are variables we would prefer to treat as exogenous, this misspecification can affect either y_t, z_t, or both.

Of course, the variables we are ultimately interested in will appear in y_t, but the latter should also include any variable that is determined jointly with the variables of interest. The exogeneity status of z_t will be the focus of our analysis here. If we denote by * the variables that are actually included in our misspecified model and by ** those that are falsely excluded, we partition

$$(2.9) \qquad x_t=\begin{pmatrix} x_t^* \\ x_t^{**} \end{pmatrix}, \quad y_t=\begin{pmatrix} y_t^* \\ y_t^{**} \end{pmatrix} \quad \text{and} \quad z_t=\begin{pmatrix} z_t^* \\ z_t^{**} \end{pmatrix}.$$

In terms of x_t^* assumption (*a*) now no longer holds, but it does for the full x_t vector.

b. **Lag** misspecification, where the information set I_{t-1}^* does not contain all the (important) information pertaining to the process for x_t that is present in W_{t-1} (overreduction of the conditioning set). Here we use a similar notation as in *a.*: I_{t-1} contains I_{t-1}^* and I_{t-1}^{**}, where I_{t-1}^* lacks important variables or important lags of variables, which are grouped in I_{t-1}^{**}. Note that this type of misspecification may affect either the conditional model for y_t given z_t, or the marginal model for z_t, or both. In fact, this implies a violation of assumption (*b*) if we replace I_{t-1} by I_{t-1}^*.

These two basic forms of invalid reductions of our likelihood function or window will be the focus of our attention in the sequel. Of course, any combination of misspecification forms is possible (and maybe even likely in practice),[3] but we feel clarity is served by considering their effects on (sufficient) conditions for exogeneity one by one.

3 Bayesian Sequential Cuts and Weak Exogeneity

As was discussed in detail in FLORENS and MOUCHART [1985] and in ENGLE *et al.* [1983], the statistical concept of **cut** is of crucial value when considering exogeneity. Therefore, we shall briefly describe cuts in the framework of (reduced) sequential models as in (2.6). Following the partitioning in (2.8), there are two major characteristics of cuts in general:

(i′) the likelihood function factorizes into a conditional part $y_t | z_t$ for which $\lambda_1 = f(\lambda) \in \Lambda_1$ is a sufficient parameterization and a marginal process for z_t with $\lambda_2 = f(\lambda) \in \Lambda_2$ as a sufficient parameter vector, and

(ii′) λ_1 and λ_2 are not "linked".

Under (i′) and (ii′) we can limit ourselves to only the conditional process for the purpose of inference on λ_1, which has inspired the definition of **weak exogeneity** in ENGLE *et al.* [1983].

3. In particular, omitted lagged variables that are crucial for the conditional model will often be crucial for the marginal model as well.

In a Bayesian analysis, (ii′) has a very natural and direct interpretation in terms of prior independence, which implies the classical concept of **variation free** parameters, as used in ENGLE *et al.* [1983] (*i.e.* $(\lambda_1, \lambda_2) \in \Lambda_1 \times \Lambda_2$, called "variation independence" by BASU [1977]).

Clearly, a Bayesian cut thus requires a complete separation of **both** sample and prior information between the conditional and the marginal process.

A **sequential Bayesian cut** is then formally defined by:

(i) $$\begin{cases} \lambda \perp x_t \mid \lambda_1, z_t, I_{t-1} \\ \lambda \perp z_t \mid \lambda_2, I_{t-1} \end{cases}$$

(ii) $$\lambda_1 \perp \lambda_2 \mid W_0$$

as e.g. in FLORENS and MOUCHART [1985].

Following ENGLE *et al.* [1983], we then define weak exogeneity with respect to the **parameters of interest** $\varphi = f(\lambda)$, *i.e.* those parameters that possess a specific meaning to the model user, as:

z_t is **weakly exogenous** in the process for y_t over the sample period for the purpose of inference on φ if and only if there exists a parameterization $\lambda = (\lambda_1, \lambda_2)$ such that (i) and (ii) hold for $t: 1 \to T$, and

(iii) φ is a function of λ_1 alone.

The sequential Bayesian cut implies that λ_1 and λ_2 will be independent *a posteriori* (see FLORENS and MOUCHART [1985], Theorem 2.8), whereas (iii) ensures that we can conduct inference on φ based on the posterior density of λ_1 alone, which will be given by

(3.1) $$D(\lambda_1 \mid W_T) \propto \prod_{t=1}^{T} D(y_t \mid z_t, I_{t-1}, \lambda_1) D(\lambda_1 \mid W_0).$$

Clearly, the combination of conditions (i)-(iii) is sufficient to validate inference on φ based on only the conditional model and the prior distribution on λ_1.

In practice, we should like to **verify** whether weak exogeneity of certain variables holds. As (i) is just the defining characteristic of λ_1 and λ_2, and (iii) is easy to check, the real test for weak exogeneity will be (ii). If we are working within a **misspecified** model, however, we shall test (ii) for a **different** set of parameters. The object of the next section is to find out what the implications of the various types of overreduction are for our exogeneity conclusions.

4 Consequences of Overreduction

4.1. Contemporaneous Misspecification

In this case, as defined in Subsection 2.2.*a*, we consider the model

$$D(x_t^*, \mu \mid I_{t-1}) = D(x_t^* \mid I_{t-1}, \mu)\, D(\mu \mid W_0), \quad (4.1)$$

instead of (2.6), where μ is a sufficient parameterization for the sequential sampling model of x_t^*, given I_{t-1}, and $D(\mu \mid W_0)$ is the prior density on μ implied by the overall prior density $D(\vartheta \mid W_0)$.[4]

If we first assume that we have left out some variables the endogeneity of which is not under scrutiny, *i.e.*

$$x_t^* = \begin{pmatrix} y_t^* \\ z_t \end{pmatrix}, \quad (4.2)$$

then the conditional sampling model is given by [using (2.9)]

$$D(y_t^* \mid z_t, I_{t-1}, \mu_1) = \int D(y_t \mid z_t, I_{t-1}, \lambda_1)\, dy_t^{**}, \quad (4.3)$$

from which we find that

$$\mu_1 = f(\lambda_1), \quad (4.4)$$

whereas a sufficient parameterization of the marginal process, say μ_2, will not be affected by the misspecification, so that $\mu_2 = \lambda_2$. Given that independence between random variables also entails independence between any Borel measurable functions of these random variables (see e.g. CHUNG [1968], Th. 3.3.1), we can write

$$\lambda_1 \perp \lambda_2 \mid W_0 \Rightarrow \mu_1 \perp \mu_2 \mid W_0,$$

so that, **provided**

$$\varphi = f(\mu_1), \quad \text{(iv)}$$

weak exogeneity will be present in the misspecified model, if it exists in the (reduced) DGP. However, finding it in the overreduced model does **not** automatically imply that it holds at the DGP level.

4. Note that we shall use the same notation μ for the parameters of any misspecified model, no matter what the source of misspecification is. The latter will become obvious from the context and this practice allows us to economize on an already heavy notation.

Now assume that the left out variables are possible candidates for exogeneity status, *i.e.*

(4.5)
$$x_t^* = \begin{pmatrix} y_t \\ z_t^* \end{pmatrix},$$

which means that the conditional process becomes

(4.6)
$$\mathrm{D}(y_t \mid z_t^*, \mathrm{I}_{t-1}, \mu_1) = \int \mathrm{D}(y_t \mid z_t, \mathrm{I}_{t-1}, \lambda_1)\, \mathrm{D}(z_t^{**} \mid z_t^*, \mathrm{I}_{t-1}, \lambda_2^{**})\, dz_t^{**}$$

with $\lambda_2^{**} = f(\lambda_2)$, so that

(4.7)
$$\mu_1 = f(\lambda_1, \lambda_2^{**}) = f(\lambda_1, \lambda_2).$$

Conversely, the marginal sampling process for z_t^* is now

(4.8)
$$\mathrm{D}(z_t^* \mid \mathrm{I}_{t-1}, \mu_2) = \int \mathrm{D}(z_t \mid \mathrm{I}_{t-1}, \lambda_2)\, dz_t^{**},$$

implying that

(4.9)
$$\mu_2 = f(\lambda_2) = \text{ , say, } \lambda_2^*.$$

From (4.7) and (4.9) we note that λ_2^{**} may very well link μ_1 and μ_2 even though λ_1 and λ_2 are independent. It is, however, possible to find at least some additional sufficient conditions under which weak exogeneity of the DGP carries over to the misspecified model. If we consider as functions of λ_2 both λ_2^*, sufficient for z_t^*, and λ_2^{**}, sufficient for z_t^{**} given z_t^*, we can deduce from a cut at the DGP level that we also have

$$\lambda_1 \perp \lambda_2 \mid \mathrm{W}_0 \Leftrightarrow \begin{cases} (\text{ii}.a) & \lambda_1 \perp \lambda_2^{**} \mid \mathrm{W}_0 \\ (\text{ii}.b) & \lambda_1 \perp \lambda_2^* \mid \lambda_2^{**}, \mathrm{W}_0 \end{cases}$$

using e.g. Lemma 0.4 in FLORENS and MOUCHART [1977], and where the implication from right to left stems from the fact that $(\lambda_2^*, \lambda_2^{**})$ is a one-to-one transformation (a partition) and not just any function of λ_2. With a slight abuse of notation, we denote this by $\lambda_2 = (\lambda_2^*, \lambda_2^{**})$. Now, combining (ii.*b*) with a Bayesian cut at the level of z_t into $z_t^{**} \mid z_t^*$ and z_t^*, *i.e.*

(v)
$$\lambda_2^* \perp \lambda_2^{**} \mid \mathrm{W}_0,$$

will be equivalent to

(4.10)
$$\lambda_2^* \perp (\lambda_1, \lambda_2^{**}) \mid \mathrm{W}_0,$$

which will, in its turn, imply for any Borel function f

(4.11)
$$\lambda_2^* \perp f(\lambda_1, \lambda_2^{**}) \mid \mathrm{W}_0, \qquad \text{or} \qquad \mu_2 \perp \mu_1 \mid \mathrm{W}_0.$$

So, ultimately, (ii)+(v) $\Rightarrow$ (4.11), which means that a cut will carry over from the DGP to the model with the z_t^{**} variables missing [see (4.5)] under the additional condition (v) that a cut exists within the z process. Of course, (v) is only a **sufficient** condition, and by no means necessary, so that finding a cut in the model does **not** mean that (ii) and (v) are always implied.

However, rejecting a cut will lead us to conclude that either (ii) or (v) (or both) are invalid assumptions to make at the DGP level.

For **weak exogeneity** we also need that condition (iv) holds, *i.e.* $\varphi = f(\mu_1)$ only, so that the combination of (i), (ii), (iv) and (v) is sufficient for weak exogeneity of z_t^* in the misspecified model with respect to φ, our parameters of interest.

4.2. Lag Misspecification

Now, the Bayesian model under consideration is

$$D(x_t, \mu \mid I_{t-1}^*) = D(x_t \mid I_{t-1}^*, \mu) D(\mu \mid W_0). \tag{4.12}$$

The invalid reduction of the information set will generally affect both conditional and marginal models, but in order to separate the consequences of lag misspecification in both, we will first study the case where $D(z_t \mid I_{t-1}, \lambda_2) = D(z_t \mid I_{t-1}^*, \mu_2)$ and $\mu_2 = \lambda_2$ so that only the conditional model is really misspecified as a result of the overreduction. Afterwards, we examine the opposite case where $D(y_t \mid z_t, I_{t-1}, \lambda_1) = D(y_t \mid z_t, I_{t-1}^*, \mu_1)$ and $\mu_1 = \lambda_1$, thus isolating the effect of lag misspecification in the marginal model (see Example 0). As the consequences of both types of overreduction are rather different, we feel we should avoid compounding both. Simply combining the sufficient conditions in both cases will give us a sufficient condition for preserving weak exogeneity in the general case where both conditional and marginal processes are affected by the overreduction of the information set.

In the case where I_{t-1}^* is only an inadequate information set for the conditional model we focus on

$$D(y_t \mid z_t, I_{t-1}^*, \mu_1) = \int D(y_t \mid z_t, I_{t-1}, \lambda_1) D(I_{t-1}^{**} \mid I_{t-1}^*, z_t, \rho_1) dI_{t-1}^{**}. \tag{4.13}$$

In order to find out a bit more about the parameters[5] ρ_1 of the conditional density of I_{t-1}^{**}, we start from the full DGP for the first t observations [cf. (2.2)]

$$D(W_t^1 \mid W_0, \vartheta) = D(w_t, W_{t-1}^1 \mid W_0, \vartheta), \tag{4.14}$$

and we consider its reduction by marginalization to

$$D(z_t, I_{t-1}^1 \mid W_0, \rho), \tag{4.15}$$

5. Whenever we introduce a parameterization in the text, like ρ_1 in (4.13), we shall always define this to be a sufficient parameter set for the process in which it appears.

where I^1_{t-1} is the information set without the initial conditions and we note that if I^1_{t-1} only groups lagged values of z_t we obtain the result that $\rho = f(\lambda_2)$. In the, more likely, case that I^1_{t-1} only contains lags of x_t, ρ will be a function of λ, whereas the general case implies that $\rho = f(\vartheta)$.

Now, assume that all of the falsely excluded variables are present in I^1_{t-1}, *i.e.* they do not pertain to the initial conditions, giving us from (4.15):

(4.16) $$D(z_t, I^{1*}_{t-1}, I^{**}_{t-1} \mid W_0, \rho) = D(I^{**}_{t-1} \mid I^*_{t-1}, z_t, \rho_1)\, D(z_t, I^{1*}_{t-1} \mid W_0, \rho_2),$$

implicity defining $I^1_{t-1} = (I^{1*}_{t-1}, I^{**}_{t-1})$ and $I^*_{t-1} = (I^{1*}_{t-1}, W_0)$. We, thus, find that generally $\rho_1 = f(f(\vartheta)) = f(\vartheta)$, whereas if I^1_{t-1} only has lags of $x_t: \rho_1 = f(f(\lambda)) = f(\lambda)$, and finally, if I^1_{t-1} groups only lagged $z_t: \rho_1 = f(f(\lambda_2)) = f(\lambda_2)$, leading to the following generic expressions for $\mu_1 = f(\lambda_1, \rho_1)$:

(4.17) $$\mu_1 = f(\vartheta)$$

in the general case, and

(4.18) $$\mu_1 = f(\lambda)$$

if either of the special cases applies. The expressions (4.17) and (4.18) are instructive at this rather abstract level as they confound the information in λ_1 and λ_2 [and possibly even beyond that in (4.17)], so that we conclude that cuts at the DGP level can easily be destroyed if we falsely omit important lagged variables from the information set in our sequential model. This is certainly a strong case for **very careful dynamic modelling** and, in particular, a methodical general-to-specific approach. Of course, one could think of sufficient conditions under which a Bayesian cut will be preserved in the face of this type of misspecification, although such conditions seem to lack a clear interpretation. In particular, if we have a cut at the DGP level combined with

(vi) $$\rho_1 \perp \lambda_2 \mid \lambda_1, W_0,$$

this cut will carry over to the overreduced model. One interesting, though rather impractical, case in which (vi) can be satisfied is where I_{t-1} constitutes the **entire** past of x_t, *i.e.* X_{t-1} in a notation analogous to (2.4). If we then falsely omit y_{t-1} and we have a condition of **noncausality** given ρ_1 (see FLORENS and MOUCHART [1985]), defined as

(4.19) $$y_{t-1} \perp (z_t, \ldots, z_T) \mid z_{t-1}, X_{t-2}, \rho_1,$$

then $D(I^{**}_{t-1} \mid I^*_{t-1}, z_t, \rho_1) = D(y_{t-1} \mid z_{t-1}, X_{t-2}, z_t, \rho_1)$ will coincide with the conditional model $D(y_{t-1} \mid z_{t-1}, X_{t-2}, \lambda_1)$ and thus $\rho_1 = \lambda_1$. A cut is then preserved, but weak exogeneity of z_t for φ in (4.13) will still require that φ can be retrieved from μ_1, *i.e.* that (iv) holds.

The second case, with the lag misspecification affecting only the marginal process, leads to considering

(4.20) $$D(z_t \mid I^*_{t-1}, \mu_2) = \int D(z_t \mid I_{t-1}, \lambda_2)\, D(I^{**}_{t-1} \mid I^*_{t-1}, \kappa_1)\, dI^{**}_{t-1}.$$

Now, consider the data density for the observations until $t-1$:

$$D(W_{t-1}^1 \mid W_0, \vartheta)$$

and reduce this to

$$D(I_{t-1}^1 \mid W_0, \kappa) = D(I_{t-1}^{**} \mid I_{t-1}^{*}, \kappa_1) D(I_{t-1}^{1*} \mid W_0, \kappa_2), \tag{4.21}$$

again under the assumption that none of the relevant initial conditions are omitted. Clearly, now $\kappa_1 = f[f(\vartheta)]$, and, thus,

$$\mu_2 = f\{\lambda_2, f[f(\vartheta)]\} = f(\vartheta), \tag{4.22}$$

so that independence of λ_1 and λ_2 will certainly **not** ensure a cut at the level of the misspecified model, except in certain special cases, e. g. when, in addition

$$\kappa_1 \perp \lambda_1 \mid \lambda_2, W_0. \tag{vii}$$

Again, the latter condition is not clearly interpretable in terms of properties of the DGP. It is, however, **always** satisfied if I_{t-1}^1, *i. e.* the information set without initial conditions, only contains lags of variables in z_t, since $\kappa = f(\lambda_2)$ in that case. In contrast with the first case discussed in this subsection, conditions (i)-(iii) and (vii) directly lead to weak exogeneity for φ as $\mu_1 = \lambda_1$, *i. e.* condition (iv) is superfluous.

To summarize, Table 1 gives an overview of some conditions that ensure that weak exogeneity properties in the DGP will carry over to the overreduced model.

TABLE 1

Sufficient Conditions for Weak Exogeneity in the DGP [*i. e.* **(i), (ii)** ***and*** **(iii)**] ***to be Preserved in the Model***

Type of Misspecification	Sufficient Additional Conditions	Remarks
(*a*) Contemporaneous		
(*a*.1) in y_t	(iv)	$\varphi = f(\mu_1)$ (cut preserved)
(*a*.2) in z_t	(iv), (v)	cut in (z_t^*, z_t^{**}) through (v)
(*b*) Lag		
(*b*.1) in cond. process	(iv), (vi)	holds in special case with non-causality
(*b*.2) in marg. process	(vii)	$\mu_1 = \lambda_1$, holds if I_{t-1}^1 only contains lagged z_t

5 Some Examples

EXAMPLE 1: Contemporaneous Misspecification of y_t.

Consider the following model for the three scalar variables y_{1t}, y_{2t} and z_t:

$$y_{1t} = \zeta y_{2t} + \beta z_t + \varepsilon_{1t} \tag{5.1}$$

$$y_{2t} = \nu y_{1t} + \delta y_{1,t-1} + \varepsilon_{2t} \tag{5.2}$$

$$z_t = \alpha z_{t-1} + \varepsilon_{3t} \tag{5.3}$$

with Normal i.i.d. errors on $\varepsilon_t = (\varepsilon_{1t}\varepsilon_{2t}\varepsilon_{3t})'$ with zero mean and a PDS covariance matrix $\Sigma = (\sigma_{ij})$; $i, j \in \{1, 2, 3\}$, and under the restriction that $\zeta\nu \neq 1$ (in order to have a solution). The reduced form then factorizes into

$$D\left(\begin{matrix} y_{1t} \\ y_{2t} \end{matrix} \middle| z_t, I_{t-1}, \lambda_1\right) \tag{5.4}$$

$$= f_N^2 \left[\begin{matrix} y_{1t} \\ y_{2t} \end{matrix} \middle| \frac{1}{1-\zeta\nu} \left\{ \begin{pmatrix} \beta + \frac{\sigma_{13}}{\sigma_{33}} + \zeta \frac{\sigma_{23}}{\sigma_{33}} \\ \beta\nu + \nu \frac{\sigma_{13}}{\sigma_{33}} + \frac{\sigma_{23}}{\sigma_{33}} \end{pmatrix} z_t + \begin{pmatrix} \delta\zeta & -\alpha\left(\frac{\sigma_{13}}{\sigma_{33}} + \zeta\frac{\sigma_{23}}{\sigma_{33}}\right) \\ \delta & -\alpha\left(\nu\frac{\sigma_{13}}{\sigma_{33}} + \frac{\sigma_{23}}{\sigma_{33}}\right) \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ z_{t-1} \end{pmatrix} \right\}, \Sigma_{yy.z} \right],$$

with

$$\Sigma_{yy.z} = \begin{pmatrix} 1 & -\zeta \\ -\nu & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{11.3} & \sigma_{12.3} \\ \sigma_{21.3} & \sigma_{22.3} \end{pmatrix} \begin{pmatrix} 1 & -\nu \\ -\zeta & 1 \end{pmatrix}^{-1},$$

where $\sigma_{ij.k} = \sigma_{ij} - (\sigma_{ik}\sigma_{jk}/\sigma_{kk})$ $(i, j \neq k)$, and the marginal process

$$D(z_t | I_{t-1}, \lambda_2) = f_N^1(z_t | \alpha z_{t-1}, \sigma_{33}). \tag{5.5}$$

If our parameter of interest is one of the structural coefficients in (5.1), e.g. $\varphi = \beta$, it is obvious that φ can not be retrieved from the conditional process (5.4) alone. In addition, our prior assumptions will naturally be formulated in terms of λ, whereas prior independence of λ_1 and λ_2 will generally not induce prior independence of the coefficients in (5.4) and (5.5) [it suffices to look at the coefficients of z_{t-1} in (5.4), which involve[6]

6. Of course, the fact that the expression for this coefficient has α and σ_{33} appearing in it, does not, in itself, imply that independence with respect to λ_2 will be lost (the natural-conjugate case for Normal linear models is an example in point).

both α and σ_{33}]. A set of sufficient conditions for weak[7] exogeneity of z_t in (5.4) for φ is

$$\sigma_{13}=\sigma_{23}=0 \tag{5.6}$$

in combination with prior independence of the form $\lambda_1 \perp \lambda_2 \mid W_0$, *i.e.*

$$(\zeta, \beta, \nu, \delta, \sigma_{11}, \sigma_{12}, \sigma_{22}) \perp (\alpha, \sigma_{33}) \mid W_0. \tag{5.7}$$

If we now misspecify the model (5.1)-(5.3) under the exogeneity conditions (5.6) and (5.7) by leaving out one of the endogenous variables, e. g. y_{2t} ($=y_t^{**}$ in our generic notation), we are left with the following conditional process for y_{1t} ($=y_t^*$):

$$\begin{aligned} D(y_{1t} \mid z_t, I_{t-1}, \mu_1) = f_N^1\Big(y_{1t} \Big| & \frac{1}{1-\zeta\nu}(\beta z_t + \delta\zeta y_{1,t-1}), \\ & \frac{1}{(1-\zeta\nu)^2}(\sigma_{11}+2\zeta\sigma_{12}+\zeta^2\sigma_{22})\Big), \end{aligned} \tag{5.8}$$

whereas (5.5) is unchanged. From (5.7) we deduce

$$\mu_1 = f(\lambda_1) \perp (\alpha, \sigma_{33}) \mid W_0$$

and it is clear that weak exogeneity of z_t still holds for any function of μ_1 in spite of this form of overreduction, but not for φ which involves the stuctural parameter β, unless e. g. we make the system triangular by $\zeta=0$. Clearly, the condition that $\varphi=f(\mu_1)$, *i.e.* condition (iv), is a crucial one in this class of misspecification.

EXAMPLE 2: Contemporaneous Misspecification of z_t

In the model for scalar y_t, z_{1t}, and z_{2t}:

$$y_t = \beta z_{1t} + \gamma z_{2t} + \varepsilon_{1t} \tag{5.9}$$

$$z_{1t} = \alpha z_{1,t-1} + \varepsilon_{2t} \tag{5.10}$$

$$z_{2t} = \zeta z_{1,t-1} + \nu z_{2,t-1} + \varepsilon_{3t}, \tag{5.11}$$

with the same stochastic assumptions as in Example 1, we can simply verify that weak exogeneity of z_{1t} and z_{2t} jointly in the conditional model for $\varphi = f(\beta, \gamma, \sigma_{11}) =$, say, β is assured by

$$\sigma_{12}=\sigma_{13}=0 \tag{5.12}$$

and

$$(\beta, \gamma, \sigma_{11}) \perp (\alpha, \zeta, \nu, \sigma_{22}, \sigma_{23}, \sigma_{33}) \mid W_0, \tag{5.13}$$

7. Since the process for z_t does not depend on lagged y's, we have a non-causality condition and can even conclude that strong exogeneity (*see* ENGLE *et al.* [1983]) holds under (5.6) and (5.7).

which implies for the conditional model

$$(5.14)\qquad D(y_t | z_{1t}, z_{2t}, I_{t-1}, \lambda_1) = f^1_N(y_t | \beta z_{1t} + \gamma z_{2t}, \sigma_{11}).$$

Now assume the misspecification of excluding z_{2t} from the conditional model, which is equivalent to falsely restricting γ to be zero. The conditional model, marginalized with respect to z_{2t} now becomes [under (5.12)]

$$(5.15)\quad D(y_t | z_{1t}, I_{t-1}, \mu_1) = f^1_N\left(y_t \middle| \left(\beta + \gamma \frac{\sigma_{23}}{\sigma_{22}}\right) z_{1t} + \gamma\left(\zeta - \alpha \frac{\sigma_{23}}{\sigma_{22}}\right) z_{1,t-1} + \gamma\nu z_{2,t-1}, \sigma_{11} + \gamma^2 \sigma_{33.2}\right),$$

whereas the marginal process is

$$(5.16)\qquad D(z_{1t} | I_{t-1}, \mu_2) = f^1_N(z_{1t} | \alpha z_{1,t-1}, \sigma_{22}).$$

implying that (5.13) does not generally lead to prior independence of μ_1 and μ_2. In a classical framework, we do obtain variation free parameterizations μ_1 and μ_2, but we cannot retrieve the structural parameter β from (5.15) alone; indeed, we can't even retrieve β from (5.15) and (5.16) **combined**, due to the marginalized z_{2t}. So we do have a classical cut in (5.15)-(5.16), but not a Bayesian one. In either case, however, weak exogeneity for $\varphi = \beta$ is precluded.

If we also impose lack of correlation between the errors of (5.10) and (5.11), *i.e.* $\sigma_{23} = 0$, then $\mu_1 = (\beta, \gamma\zeta, \gamma\nu, \sigma_{11} + \gamma^2 \sigma_{33})$ and $\mu_2 = (\alpha, \sigma_{22})$ [compare our generic expressions in (4.7) and (4.9)] are still not necessarily prior independent under only (5.13). This is, however, achieved under the additional condition

$$(5.17)\qquad (\zeta, \nu, \sigma_{33}) \perp (\alpha, \sigma_{22}) | W_0,$$

which, in combination with $\sigma_{23} = 0$, is sufficient for a Bayesian cut operated by $z_{2t} | z_{1t}$ and z_{1t}. Now, weak exogeneity for β is preserved.

In order not to confound contemporaneous and lag misspecification, we have assumed here that a large enough information set I_{t-1} was used throughout, *i.e.* containing at least $z_{1,t-1}$ and $z_{2,t-1}$. Failure to do so would not change the conditional model in (5.14), but it would obviously affect the misspecified one in (5.15).

EXAMPLE 3: General Lag Misspecification.

Let us now consider as the "correctly" reduced model:

$$(5.18)\qquad y_t = \beta_1 z_{1t} + \beta_2 z_{1,t-1} + \beta_3 z_{1,t-2} + \gamma_1 z_{2t} + \gamma_2 z_{2,t-1} + \gamma_3 z_{2,t-2} + \varepsilon_{1t}$$

$$(5.19)\qquad z_{1t} = \alpha z_{1,t-1} + \varepsilon_{2t}$$

$$(5.20)\qquad z_{2t} = \zeta z_{1,t-1} + \nu z_{2,t-1} + \varepsilon_{3t},$$

in combination with i.i.d. Normal errors on ε_t.

As usual in such a simple Normal framework, where independence and lack of correlation are equivalent, a set of sufficient conditions for the joint weak exogeneity of the element of $z_t=(z_{1t} z_{2t})'$ in the conditional model for y_t involves zero covariances, *i.e.*

$$\sigma_{12}=\sigma_{13}=0, \tag{5.21}$$

as well as prior independence of the resulting parameterizations $\lambda_1=(\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \sigma_{11})$ and $\lambda_2=(\alpha, \zeta, \nu, \sigma_{22}, \sigma_{23}, \sigma_{33})$ given the initial conditions W_0, *i.e.* condition (ii), and finally the requirement that the parameters of interest φ can be retrieved from λ_1 alone [condition (iii)]. Under (5.21) the conditional model exactly coincides with the structural model in (5.18) which is, incidentally, by no means a necessary nor a sufficient condition for weak exogeneity. It merely reflects the independent error structure.

If we now misspecify the information set by excluding $z_{2,t-1}$, *i.e.* by incorrectly restricting γ_2 and ν to be zero, we are, in fact, considering the process $D(x_t | I^*_{t-1}, \mu)$, where $I^*_{t-1}=(z_{1,t-1}, z_{1,t-2}, z_{2,t-2}, W_0)$ and the excluded part of the information set I^{**}_{t-1} comprises $z_{2,t-1}$.

As in (4.13), we shall use the density $D(I^{**}_{t-1} | I^*_{t-1}, z_t, \rho_1)$ to marginalize out $z_{2,t-1}$ from the conditional process, and we note that I^1_{t-1} only contains lagged z_t, so that $\rho_1=f(\lambda_2)$. The misspecified conditional model then becomes

$$\begin{aligned} D(y_t | z_t, I^*_{t-1}, \mu_1)=f^1_N\Bigg(y_t \Bigg| \left(\beta_1-\frac{\gamma_2 \nu}{1+\nu^2}\frac{\sigma_{32}}{\sigma_{22}}\right)z_{1t} \\ +\left(\beta_2+\frac{\gamma_2}{1+\nu^2}\left((1+\alpha\nu)\frac{\sigma_{32}}{\sigma_{22}}-\zeta\nu\right)\right)z_{1,t-1} \\ +\left(\beta_3+\frac{\gamma_2}{1+\nu^2}\left(\zeta-\alpha\frac{\sigma_{32}}{\sigma_{22}}\right)\right)z_{1,t-2}+\left(\gamma_1+\frac{\gamma_2\nu}{1+\nu^2}\right)z_{2t} \\ +\left(\gamma_3+\frac{\gamma_2\nu}{1+\nu^2}\right)z_{2,t-2}, \sigma_{11}+\gamma_2^2\frac{\sigma_{33.2}}{1+\nu^2}\Bigg). \end{aligned} \tag{5.22}$$

We now consider what happens to the marginal process if we leave out $z_{2,t-1}$ in the equation for z_{2t}, *i.e.* (5.20). This implies that we need $D(z_{2,t-1} | z_{1,t-1}, W_0, \kappa_1)$ since $I^*_{t-1}=(z_{1,t-1}, W_0)$ and the excluded I^{**}_{t-1} comprises only $z_{2,t-1}$. Thus I^1_{t-1} only contains lags of z_t, which already assures that μ_2 only involves λ_2. If we wish to do the formal analysis, we need to marginalize the process for $z_t=(z_{1t} z_{2t})'$ with respect to the past, specifying the distribution for z_0, which is assumed (for simplicity) to be

Normal. If we then impose stationarity of the z_t process, *i.e.* we incorporate the prior restrictions that $|\nu|<1$ and $|\alpha|<1$, then, for large t, the over-reduced marginal process as in (4.20) will be approximated by

$$(5.23)\quad D\left(\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} \middle| I^*_{t-1}, \mu_2\right)$$

$$= f_N^2\left(\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} \middle| \begin{bmatrix} \alpha \\ \frac{1}{1-\nu}\left(\zeta + \frac{\sigma_{32}\nu}{\sigma_{22}}(1-\alpha)\right) \end{bmatrix} z_{1,t-1}, \begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} + \frac{\nu^2}{(1-\nu)^2}\sigma_{33.2} \end{pmatrix}\right)$$

Clearly, $\mu_2 = f(\lambda_2)$ as expected.[8]

However, it is clear that the dependence of ρ_1 on λ_2 generally induces links between μ_1 (mixing λ_1 and ρ_1) and $\mu_2 = f(\lambda_2)$, given that (ii) holds. This destroys the possibility of a Bayesian cut after overreduction. Remark that none of the links between μ_1 and μ_2 are exact here, and therefore a classical cut is operated by (5.22) and (5.23). Nevertheless, no structural coefficient in λ_1 can be retrieved from μ_1, so that classical weak exogeneity cannot exist for $\varphi = f(\lambda_1)$. A Bayesian analysis takes both stochastic and exact links into account, and thus a Bayesian cut is formally precluded as the elements of μ_1 will generally not be independent of μ_2. The severe consequences of this type of misspecification are illustrated by the fact that μ_1 and μ_2 seem to be fundamentally intertwined. Nontrivial conditions for weak exogeneity in (5.22) are hard to find. If z_{2t} is a deterministic ($\sigma_{32} = \sigma_{33} = 0$) known ($\zeta$ and ν constant) function of $z_{1,t-1}$ and $z_{2,t-1}$, we have, *e.g.*, weak exogeneity of z_t for $\varphi = f(\beta_1, \sigma_{11})$ in the misspecified model. Although the latter could occur in the case of e.g. accounting identities or defining equations, it should be clear from this example that general lag misspecification can have serious consequences for exogeneity conclusions.

Example 4: Lag Misspecification for the Individual Processes.

a. The Conditional Process:

The same model as in the previous example is used, again with the sufficient exogeneity conditions (5.21), (ii) and (iii), and the same λ_1, but now with $\nu = 0$ [*i.e.* $\mu_2 = \lambda_2 = (\alpha, \zeta, \sigma_{22}, \sigma_{23}, \sigma_{33})$], so that leaving out $z_{2,t-1}$ will only affect the conditional model. From (5.22) we immediately obtain that

$$\mu_1 = \left(\beta_1, \beta_2 + \gamma_2\frac{\sigma_{32}}{\sigma_{22}}, \beta_3 + \gamma_2\left(\zeta - \alpha\frac{\sigma_{32}}{\sigma_{22}}\right), \gamma_1, \gamma_3, \sigma_{11} + \gamma_2^2\sigma_{33.2}\right)$$

8. This holds true whether we approximate the marginal process or not. In particular, it even holds under nonstationarity.

and λ_2 are not prior independent, although they are variation free. This implies there is no Bayesian cut and weak exogeneity of z_t is destroyed, without much hope of restoring it through additional conditions (again, making z_{2t} a deterministic known function is sufficient, but seems a very strong requirement indeed). Interestingly, a classical analysis will be led astray in accepting weak exogeneity of z_t for $\varphi = f(\beta_1, \gamma_1, \gamma_3)$, even though some (stochastic) information concerning μ_1 will generally be present in λ_2.

b. The Marginal Process:

We now add the restriction $\gamma_2 = 0$ to the model of Example 3, still assuming (5.21), (ii) and (iii). In this case $\mu_1 = \lambda_1 = (\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_3, \sigma_{11})$ and the overreduction of the information set to I^*_{t-1} will not affect the conditional model. From (5.23) we obtain the result that weak exogeneity in the original model (5.18)-(5.20) with $\gamma_2 = 0$ is not affected at all by this marginal lag misspecification; z_t is still weakly exogenous for any $\varphi = f(\lambda_1)$, since μ_1 and λ_1 coincide. A simple illustration of a case where I^1_{t-1} does contain lagged y_t was provided in the Introduction (Example 0).

Comparison with Example 3 reveals that the main problem causing a loss of weak exogeneity there is situated in lag misspecification of the conditional process.

6 Implications for Econometric Modelling

Although we are in the, perhaps not very enviable, position in econometrics that every model we specify is bound to exhibit some degree of misspecification, it is important to realize that some forms of misspecification are much more dangerous than others, given the specific questions that we wish to address.

Unwarranted omission of contemporaneous endogenous variables can occur if we are only interested in a subset of the y_t vector (contained in y^*_t). Provided our parameters of interest φ can be expressed as a function of μ_1, the parameters of the conditional process for y^*_t, weak exogeneity of z_t is not affected. A cut at DGP level is always carried over when such overreduction arises, but the model itself may not allow inference on φ, as was illustrated in Example 1.

If we leave out some of the contemporaneous conditioning variables it often implies that we can no longer validly condition on the remaining variables in z_t, since links between the included z^*_t and the excluded z^{**}_t will typically prevent a Bayesian cut after z^{**}_t is integrated out of the conditional process. Rejecting weak exogeneity of z^*_t does not imply that we can reject weak exogeneity of the full z_t vector, which is, in a sense, a weaker condition, as the interior links **within** the z_t process are then completely irrelevant. Obviously, removing those interior links is sufficient to save weak exogeneity of the included variables in z_t^* for the parameters in μ_1 (not necessarily for φ).

Misspecifying the lags included in the conditional model was seen to have potentially very serious consequences. As we have to use $D(I_{t-1}^{**} \mid I_{t-1}^{*}, z_t, \rho_1)$ [in (4.13)], where z_t is a conditioning variable, it is rather unlikely that ρ_1 should not be contaminated by λ_2. This is illustrated by the expressions following (4.16) and by Examples 3 and 4*a*, where, barring rather trivial cases, a Bayesian cut is destroyed. And even if a cut is preserved, we still face a misspecified conditional model.

Such misspecification of the conditional model is not present if we only get the dynamics of the marginal process wrong (see Example 4*b*). The only issue then is to preserve a Bayesian cut, which always goes through if I_{t-1}^{1} contains only lags of z_t. In this case, misspecifying these lags in the marginal model does not affect the weak exogeneity of z_t, nor is the conditional model misspecified. Therefore, this special case seems quite harmless.

Thus, the most vicious type of misspecification seems to reside in wrong dynamics for the conditional model, as we then generally face both problems of inference on the structural parameters and a loss of cut. We suspect these consequences to be very pervasive in practice, and we feel this situation presents a powerful argument in favour of extremely careful dynamic modelling, particularly of the conditional model. In addition, the general-to-specific methodology in econometrics (see e.g. HENDRY and RICHARD [1983]) combined with meticulous testing of each reduction that is implemented seems, of course, the best way to guard against the types of misspecification discussed here.

Although the various types of overreduction have rather different consequences, they all generally lead to a loss of weak exogeneity, provided we introduce additional conditions (see Table 1). Without sufficient additional restrictions we are led to falsely rejecting weak exogeneity (assuming our tests are powerful enough to pick up the deviation from weak exogeneity) and, thus, to a joint treatment of y_t and z_t (y_t^* or z_t^* in case of contemporaneous misspecification). Of course, this unnecessarily complicates the analysis, but in the case of dynamic misspecification of the marginal process no further problems occur: we can infer on the parameters of interest as they are [by assumption (iii)] a function of λ_1 alone, and λ_1 can be recovered from the parameterization of the joint process, μ. Recall that the conditional model is **not** subject to misspecification in this case.

The situation becomes much worse, however, with the other types of specification errors, as μ generally does not allow inference on φ then, so that **even from the joint model** valid inference on φ **cannot** be conducted.

• References

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

Chung, K. L. (1968). – *A course in probability theory*, Harcourt, New York.

Engle, R. F., Hendry, D. F. and Richard, J. F. (1983). – "Exogeneity", *Econometrica*, 51, pp. 277-304.

Florens, J. P. and Mouchart, M. (1977). – "Reduction of Bayesian Experiments", CORE discussion paper 7737, Université Catholique de Louvain, Louvain-la-Neuve.

Florens, J. P. and Mouchart, M. (1982). – "A note on Noncausality", *Econometrica*, 50, pp. 583-591.

Florens, J. P. and Mouchart, M. (1985). – "Conditioning in Dynamic Models", *Journal of Time Series Analysis*, 6, pp. 15-34.

Florens, J. P., Mouchart, M. and Rolin, J. M. (1990). – *Elements of Bayesian Statistics*, Marcel Dekker, New York.

Hendry, D. F. and Richard, J. F. (1982). – "On the Formulation of Empirical Models in Dynamic Econometrics", *Journal of Econometrics*, 20, pp. 3-33.

Hendry, D. F. and Richard, J. F. (1983). – "The Econometric Analysis of Economic Time Series", *International Statistical Review*, 51, pp. 111-163.

Leamer, E. E. (1978). – *Specification searches: Ad hoc inference with nonexperimental data*, Wiley, New York.

Lindley, D. V. (1982). – "The Bayesian Approach to Statistics", in: J. Tiago de Oliveira and B. Epstein, eds., *Some recent advances in statistics*, Academic Press, London.

Mouchart, M. and Rolin, J. M. (1984). – "A Note on Conditional Independence with Statistical Applications", *Statistica*, 44, pp. 557-584.

Osiewalski, J. and Steel, M. F. J. (1989). – "A Bayesian Analysis of Exogeneity in Models Pooling Time-series and Cross-section Data", CentER discussion paper 8914, Tilburg University, Tilburg.

Poirier, D. J. (1988). – "Frequentist and Subjectivist Perspectives on the Problems of Model Building in Economics", *Economic Perspectives*, 2, pp. 121-144.

Smith, A. F. M. (1984). – "Bayesian Statistics", *Journal of the Royal Statistical Society*, series A, 147, pp. 245-259.

Smith, R. J. and Blundell, R. W. (1986). – "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labour Supply", *Econometrica*, 54, pp. 679-685.

Steel, M. F. J. (1987). – *A Bayesian analysis of multivariate exogeneity: A Monte Carlo approach*, Nouvelle Série 168, Université Catholique de Louvain, Louvain-la-Neuve.

Steel, M. F. J. and Richard, J. F. (1991). – "Bayesian Multivariate Exogeneity Analysis: An Application to a UK Money Demand Equation", *Journal of Econometrics*, 49, pp. 239-274.