Weak exogeneity in misspecified sequential models
Steel, M.F.J.

Publication date: 1989

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. Dec. 2018
No. 8942

WEAK EXOGENEITY IN MISSPECIFIED SEQUENTIAL MODELS

by Mark F.J. Steel

September, 1989
Weak Exogeneity in Misspecified Sequential Models

Mark F.J. Steel*
Tilburg University
August 1989

JEL code: 211

Keywords: time-series econometrics; weak exogeneity; sequential model; misspecification; Bayesian cut.

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 specification 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.

Acknowledgements:
Stimulating discussions with Jacek Osiewalski and Gary Koop proved very clarifying. Needless to say, all 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 research fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW).

* Department of Econometrics, Tilburg University. P.O. Box 90153, 5000 LE Tilburg, The Netherlands.
1. Introduction

The notion of exogeneity has, from the start, permeated the econometrics literature as it seems absolutely crucial to most of the empirical work in this area. Intuitively introduced in the early fifties by e.g. Koopmans (1950), Orcutt (1952), and Marschak (1953), a formalization of many concepts was given in Engle et al. (1983), based on the statistical literature concerning cuts [see Barndorff-Nielsen (1978) or Florens and Mouchart (1977)].

The importance of exogeneity for the practice of econometric modelling was, again, stressed in Hendry and Richard (1982, 1983) and has even found its way into textbooks [see e.g. Spanos (1986)]. 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, due to ignorance, confusion, lack of computing facilities, or bad luck, 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 much 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 (1974) or Mouchart and Rolin (1984)].

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 $z_t$

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

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

with the error terms independently and identically distributed according to a Normal law, denoted as

$$D(\epsilon_{1t}, \epsilon_{2t}) = f_{\text{N}}(\epsilon_{1t} | 0, \sigma_{11}, \sigma_{12} | 0, \sigma_{21}, \sigma_{22}) \tag{1.3}$$

Such a model as (1.1) - (1.3) 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 like (1.1) - (1.3) 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. Conditions ensuring weak exogeneity of $z_t$ for the purpose of inference on $\beta$ and $\sigma_{11}$ are $\sigma_{12} = 0$ and prior independence of the resulting parameterizations $\lambda_1$ and $\lambda_2$ in

$$D(y_t | z_t, I_{t-1}, \lambda_1) = f_{\text{N}}^1(y_t | \beta z_t, \sigma_{11}) \tag{1.4}$$

$$D(z_t | I_{t-1}, \lambda_2) = f_{\text{N}}^1(z_t | \delta_1 z_{t-1} + \delta_2 y_{t-1}, \sigma_{22}) \tag{1.5}$$

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 (1.5) or of "falsely" imposing $\delta_2 = 0$ [i.e. a noncausality restriction, as explained in
e.g. Florens and Mouchart (1982) and Engle et al. (1983)], the marginal process will become

\[ D(z_t | I_{t-1}, \mu_2) = f_N(z_t | (\delta_1 + \delta_2 \beta) z_{t-1}, \sigma_{22}^2 + \delta_2^2 \sigma_{11}^2) , \]  

(1.6)

with \( I_{t-1} = z_{t-1} \) and the parameterization is now changed to

\[ \mu_2 = (\delta_1 + \delta_2 \beta, \sigma_{22}^2 + \delta_2^2 \sigma_{11}^2) \]  

(1.7)

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 \( \lambda_1 \) and \( \mu_2 \) 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 incoherenty, 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 \( \delta_2 \) will be near zero, giving the data a chance to revise the prior idea that \( y_{t-1} \) does not matter in (1.5).

If enough data evidence is collected, we will eventually find out that (1.5) leads to a more useful window than (1.6) (exactly how this occurs is beyond this discussion), and in this sense (1.6) is misspecified. this is the light in which terms like "correct" or "misspecified" model should be seen here.
Also, from this point of view, it is obvious that $\mu_2$ will now no longer be prior independent from $\lambda_1$, in general, so that the misspecification has destroyed weak exogeneity.

Throughout the paper we assume that sufficient conditions hold to ensure weak exogeneity at the level of the "correct" model which implies e.g. that prior notions are formulated for all the parameters appearing at this level. 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 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 misspecification (one of our five types, to be introduced later) becomes irrelevant.

Section 2 describes the statistical framework and the various types of misspecification that we wish to consider. Section 3 briefly discusses Bayesian cuts and defines weak exogeneity, whereas the next section evaluates the consequences of different specification 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.
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 $w_t(t:1\rightarrow T)$:

$$w_t^1 = (w_1, \ldots, w_T)'$$

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

$$D[w_t^1 | w_0, \theta]$$

(2.2)

where $w_0$ denotes a (possibly infinitely dimensional) matrix of initial conditions and $\theta \in \Theta$ 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 $t : 1 \rightarrow T$ of sequential models

$$D(w_t | w_{t-1}, \theta)$$

(2.3)

where we have implicitly defined the full information set available at time $t$ by
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 \( \Theta \), denoted by \( D(\Theta|\mathbf{W}_0) \) as it will often depend on initial conditions, that may include e.g. a previous sample. We obtain

\[
D(w_t, \Theta|\mathbf{W}_{t-1}) = D(w_t|\mathbf{W}_{t-1}, \Theta)D(\Theta|\mathbf{W}_0),
\]

(2.5)

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, \( x_t \subseteq w_t \), that is of "manageable" size and includes all variables that we are interested in modelling. Typically, the sampling density for \( x_t \) will then only depend on a small subset of the entire past of the economy, say, \( I_{t-1} \subseteq \mathbf{W}_{t-1} \). Let us denote by \( \lambda = f(\Theta) \) a sufficient parameterization of this marginalized process, and we concentrate, therefore, on the marginal prior density for \( \lambda \), as derived from the overall prior assumptions on \( \Theta \). Our substantially reduced Bayesian model then becomes

\[
D(x_t, \lambda|I_{t-1}) = D(x_t|I_{t-1}, \lambda)D(\lambda|\mathbf{W}_0),
\]

(2.6)

where the status of \( I_{t-1} \) and \( \lambda \) is formally explained by the following assumptions, that implicitly define them in the order chosen in the text:

\[
x_t \perp \mathbf{W}_{t-1}|I_{t-1}, \Theta
\]

(2.7')

and

\[
x_t \perp \Theta|I_{t-1}, \lambda
\]

(2.7'')

where conditional independence of random variables, say, \( a \) and \( b \), given \( c \), is denoted by \( a \perp b|c \).
Remark that both assumptions are jointly equivalent to

\[ x_t \perp (W_{t-1}, \theta) | I_{t-1}, \lambda \]  \hspace{1cm} (2.7)

using the fundamental properties of conditional independence [see e.g. Mouchart and Rolin (1984)].

Within the constraints of the "window" [see Poirier (1988)] or likelihood function thus chosen (usually not in great conflict with one's prior ideas, nor with relevant modeling experience), 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 \( x_t \) (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 (1989) 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. Misspecification

The reduction steps involved in going from the DGP (2.3) to our reduced econometric sampling 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 misspecification. 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 misspecification, but we shall content ourselves with assuming that misspecification can occur in three basic guises.\(^1\)

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. If we partition

\[
\begin{bmatrix}
y_t \\
z_t
\end{bmatrix} = \begin{bmatrix}
y^*_t \\
z^*_t
\end{bmatrix},
\]

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

\[
\begin{bmatrix}
x_t \\
y_t \\
z_t
\end{bmatrix} = \begin{bmatrix}
x^*_t \\
y^*_t \\
z^*_t
\end{bmatrix},
\]

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 \(W_{t-1}\). 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 occur in either the conditional model for \(y_t\) given \(z_t\), or in the marginal model for \(z_t\), or in both. In fact, this implies a violation of assumption (2.7').

c. **Classification** misspecification, implying that we consider the exogeneity status of too small a subset of \(x_t\). If we partition

\[
\begin{bmatrix}
z_{1t} \\
z_{2t}
\end{bmatrix} = \begin{bmatrix}
z^*_1 \\
z^*_2
\end{bmatrix}.
\]
we then test for the exogeneity of $z_{2t}$, without, however, questioning the endogenous character of $z_{1t}$. Remark the inherently asymmetrical nature of this misspecification: if $z_t$ is chosen too large, we have a fair chance of detecting the endogeneity of some of its elements, provided the test conducted has sufficient power, but if we choose it too small, we have no way of assessing the adequacy of our decision.

These three basic forms of "invalid" specifications 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), but we feel clarity is served by considering their effects on (sufficient) conditions for exogeneity one by one.

The next section will provide a somewhat more formal framework for our discussion of weak exogeneity.

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') $\lambda_1$ and $\lambda_2$ are not "linked".

Under (i') and (ii') we can limit ourselves to only the conditional process for the purpose of inference on $\lambda_1$, which has inspired the definition of weak exogeneity in Engle et al. (1983).

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:

\[(i) \begin{align*}
\lambda_1 & \perp x_t | \lambda_1, z_t, I_{t-1} \\
\lambda_2 & \perp z_t | \lambda_2, I_{t-1}
\end{align*}\]

\[(ii) \lambda_1 \perp \lambda_2 | w_0\]

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 \( \psi = f(\lambda) \), 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 \( \psi \) if and only if there exists a parameterization \( \lambda = (\lambda_1, \lambda_2) \) such that (i) and (ii) hold, and

\[(iii) \ \psi \text{ is a function of } \lambda_1 \text{ alone.}\]

The sequential Bayesian cut implies that \( \lambda_1 \) and \( \lambda_2 \) will be independent a posteriori [see Florens and Mouchart (1985), Theorem 2.8], whereas (iii) ensures that we can conduct inference on \( \psi \) based on the posterior density of \( \lambda_1 \) alone, which will be given by

\[
D(\lambda_1 | W_T) \propto \prod_{t=1}^{T} D(y_t | z_t, I_{t-1}, \lambda_1) D(\lambda_1 | W_0). \tag{3.1}
\]

Clearly, the combination of both conditions is sufficient to validate inference on \( \psi \) based on only the conditional model and the prior distribution on \( \lambda_1 \).

In practice, we should like to verify whether weak exogeneity of certain variables holds. As (i) is just the defining characteristic of \( \lambda_1 \) and \( \lambda_2 \), 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 misspecification are on our exogeneity conclusions.

4. Consequences of Misspecification

4.1. Contemporaneous Misspecification

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

\[ D(x_t^*, \mu | I_{t-1}) = D(x_t^* | I_{t-1}, \mu) D(\mu | W_0) \]  

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

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

\[ x_t^* = \begin{bmatrix} y_t^* \\ z_t \end{bmatrix}, \]  

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

\[ D(y_t^* | z_t, I_{t-1}, \mu_1) = \int D(y_t^* | z_t, I_{t-1}, \lambda_1) dy_t^* \]  

from which we find that

\[ \mu_1 = f(\lambda_1), \]  

whereas a sufficient parameterization of the marginal process, say \( \mu_2 \), will not be affected by the misspecification, so that

\[ \mu_2 \]  

Given that independence between random variables also entails independence between any Borel measurable functions of these random variables [see e.g. Chung (1974), Th. 3.3.1], we can write
\[ \lambda_1 \perp \lambda_2 | \omega_0 = \mu_1 \perp \mu_2 | \omega_0 \],

so that, provided

(iv) \( \varphi = f(\mu_1) \),

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

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

\[ x^*_t = \begin{bmatrix} y_t \\ z^*_t \end{bmatrix}, \tag{4.6} \]

which means that the conditional process becomes

\[ D(y_t | z^*_t, I_{t-1}, \mu_1) = \int D(y_t | z_t, I_{t-1}, \lambda_1)D(z^*_t | z^*_t, I_{t-1}, \lambda_2^{**})dz^{**} \tag{4.7} \]

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

\[ \mu_1 = f(\lambda_1, \lambda_2^{**}) = f(\lambda_1, \lambda_2). \tag{4.8} \]

Conversely, the marginal sampling process for \( z^*_t \) is now

\[ D(z^*_t | I_{t-1}, \mu_2) = \int D(z_t | I_{t-1}, \lambda_2)dz^{**}, \tag{4.9} \]

implying that

\[ \mu_2 = f(\lambda_2) = , \text{ say, } \lambda_2^{**}. \tag{4.10} \]

From (4.8) and (4.10) we note that \( \lambda_2^{**} \) may very well link \( \mu_1 \) and \( \mu_2 \) even though \( \lambda_1 \) and \( \lambda_2 \) 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 \( \lambda_2 \) both \( \lambda_2^{*} \), sufficient for \( z^*_t \), and \( \lambda_2^{**} \), sufficient for \( z_t^{**} \) given \( z^*_t \), as used in (4.7) and (4.10), we can deduce from a cut at the DGP level that we also have
\[ \lambda_1 \perp \lambda_2 \mid w_0 \leftrightarrow \begin{cases} 
\text{(iia)} & \lambda_1 \perp \lambda_2^{**} \mid w_0 \\
\text{(iib)} & \lambda_1 \perp \lambda_2^{*} \mid \lambda_2^{**} \mid w_0 
\end{cases} \]

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

\[ (v) \quad \lambda_2^* \perp \lambda_2^{**} \mid w_0 \]

will be equivalent to

\[ \lambda_2^* \perp (\lambda_1, \lambda_2^{**}) \mid w_0 \], \hspace{1cm} (4.11) \]

which will, in its turn, imply for any bounded Borel function \( f \)

\[ \lambda_2^* \perp f(\lambda_1, \lambda_2^{**}) \mid w_0 \], \hspace{1cm} \text{or} \hspace{1cm} \mu_2 \perp \mu_1 \mid w_0 \]. \hspace{1cm} (4.12) \]

So, ultimately, \((ii) \wedge (v) \Rightarrow (4.12)\), which means that a cut will carry over from the DGP to the model with the \( z_t^{**} \) variables missing [see (4.6)] 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 \( \varphi \), our parameters of interest.
4.2. Lag Misspecification

Now, the Bayesian model under consideration is

$$D(x_t, \mu | I_{t-1}^*) = D(x_t | I_{t-1}^*, \mu) D(\mu | W_0) .$$  (4.13)

As briefly noted in Subsection 2.2.b this exclusion of important lagged (endogenous or exogenous) variables from the information set $I_{t-1}$ can occur in the conditional model, the marginal model, or both. For reasons of clarity, we shall only explicitly treat the first two possibilities here. Extension to the third one is straightforward.

In the case of an inadequate information set for the conditional model we focus on

$$D(y_t | z_t, I_{t-1}^*, \mu_1) = \int D(y_t | z_t, I_{t-1}, \lambda_1) D(I_{t-1}^* | I_{t-1}^*, z_t, \rho_1) dI_{t-1}^* .$$  (4.14)

In order to find out a bit more about the parameters $\rho_1$ of the conditional density of $I_{t-1}^*$, we start from the full DGP for the first $t$ observations [cf. (2.2)]

$$D(w_t | w_0, \theta) = D(w_t, w_{t-1} | w_0, \theta) ,$$  (4.15)

and we consider its reduction by marginalization to

$$D(z_t, I_{t-1}^1 | w_0, \rho) ,$$  (4.16)

where $I_{t-1}^1$ is the information set without the initial conditions and we note that if $I_{t-1}^1$ only contains lagged values of $z_t$ we obtain the result that $\rho = f(\lambda_2)$.

In the, more likely, case that $I_{t-1}^1$ only contains lags of $x_t$, $\rho$ will be a function of $\lambda$, whereas the general case implies that $\rho = f(\theta)$.

Now, assume that all of the falsely excluded variables are present in $I_{t-1}^1$, i.e. they do not pertain to the initial conditions, giving us from (4.16):

$$D(z_t, I_{t-1}^1, I_{t-1}^* | w_0, \rho) = D(I_{t-1}^* | I_{t-1}^*, z_t, \rho_1) D(z_t, I_{t-1}^1 | w_0, \rho_2) ,$$  (4.17)

implicitly defining $I_{t-1}^1 = (I_{t-1}^*, I_{t-1}^*)$ and $I_{t-1}^* = (I_{t-1}^*, W_0)$. 

We, thus, find that generally

\[ \varphi_1 = f(f(\phi)) = f(\phi), \]  

whereas if \( I_{t-1} = \lambda_1 \) only has lags of \( x_t \)

\[ \varphi_1 = f(f(\lambda)) = f(\lambda), \]  

and, finally, if \( I_{t-1} = \lambda_1 \) groups only lagged \( z_t \)

\[ \varphi_1 = f(f(\lambda_2)) = f(\lambda_2), \]  

leading to the following generic expressions for \( \mu_1 = f(\lambda_1, \varphi_1) \):

\[ \mu_1 = f(\theta) \]  

(4.21)

in the general case, and

\[ \mu_1 = f(\lambda) \]  

(4.22)

if either of the special cases applies.

The expressions (4.21) and (4.22) are instructive at this rather abstract level as they confound the information in \( \lambda_1 \) and \( \lambda_2 \) [and possibly even beyond that in (4.21)], 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

\[ (vi) \quad \varphi_1 \perp \lambda_2 | \lambda_1, W_0, \]  

this cut will carry over to the misspecified model, while for weak exogeneity we would naturally also require that \( \varphi \) can be retrieved from \( \mu_1 \), i.e. that (iv) holds.
The second case, with lag misspecification affecting only the marginal process, leads to considering

$$D(z_t | I_{t-1}^*, \mu_2) = \int D(z_t | I_{t-1}^*, \lambda_2)D(I_{t-1}^{**} | I_{t-1}^*, \lambda_1)\,dI_{t-1}^{**} \ . \quad (4.23)$$

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

$$D(W_{t-1}^1 | W_0, \theta)$$

and reduce this to

$$D(I_{t-1}^1 | W_0, \lambda_1^2) = D(I_{t-1}^{**} | I_{t-1}^*, \lambda_1)D(I_{t-1}^{**} | W_0, \lambda_2) \ . \quad (4.24)$$

again under the assumption that none of the relevant initial conditions are omitted. Clearly, now

$$\lambda_1 = f[f(\theta)] \ , \quad (4.25)$$

and, thus,

$$\mu_2 = f[\lambda_2, f[f(\theta)]] = f(\theta) \ , \quad (4.26)$$

so that independence of $\lambda_1$ and $\lambda_2$ will certainly not ensure a cut at the level of the misspecified model, except in certain special cases, e.g. when, in addition

$$(\text{vii}) \quad x_1 \perp \lambda_1 | \lambda_2, W_0 \ .$$

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 the initial conditions, only contains lags of variables in $z_t$, since $\lambda = f(\lambda_2)$ in that case. In contrast with the first case discussed in this subsection, conditions (i) - (iii) and (vii) directly lead to weak exogeneity for $\theta$ as $\mu_1 = \lambda_1$, i.e. condition (iv) is superfluous.
4.3. Misclassification

The third basic form of misspecification examined here is the too narrow a choice of the candidates for exogeneity. Subsection 2.2.c briefly discussed this issue.

Following (2.10), we partition the sequential sampling model as

\[ D(x_t | I_{t-1}, \lambda) = D(y_t | z_{1t}, z_{2t}, I_{t-1}, \lambda_1) \]
\[ D(z_{1t} | I_{t-1}, \lambda_1^2) \]
\[ D(z_{2t} | I_{t-1}, \lambda_2^2) \]

with \( \lambda_2 = (\lambda_1^2, \lambda_2^2) \), from which we easily deduce

\[ D(x_t | I_{t-1}, \lambda) = D((y_t, z_{1t}) | z_{2t}, I_{t-1}, \mu_1) \]
\[ D(z_{2t} | I_{t-1}, \lambda_2^2) \]

where

\[ \mu_1 = (\lambda_1^1, \lambda_2^1). \]

Now \( \mu_1 \) and \( \mu_2 = \lambda_2^2 \) will not always be independent under a cut at the DGP level, i.e. (ii): \( \lambda_1 \perp \lambda_2 | W_0 \). Even though (iv): \( \rho = f(\mu_1) \) will automatically be implied by (iii): \( \rho = f(\lambda_1^1) \) and the equivalence of \( \mu_1 \) and \( (\lambda_1^1, \lambda_2^2) \) in (4.29), we clearly see that weak exogeneity for the entire \( z \) vector does not entail weak exogeneity for subvectors of \( z \), since \( \mu_1 \) mixes \( \lambda_1 \) with part of the parameters in \( \lambda_2 \) and the latter could very well induce links between \( \mu_1 \) and \( \mu_2 \) that prevent a cut. Note that this issue was raised within a linear Normal framework in Engle et al. (1980), Steel (1987) and Steel and Richard (1989). A joint Bayesian cut is weaker than having cuts for both \( z_{1t} \) and \( z_{2t} \) in isolation as links between the processes in \( z_t \) are irrelevant for such a joint cut, whereas they can certainly prevent isolated or simple cuts. An obvious sufficient additional condition for a cut at the level of (4.28) is that
(viii) \( \lambda_2^1 \mid \lambda_2^2 \mid W_0 \),

using the same argument as in Subsection 4.1.

Note that (viii) implies that a Bayesian cut is operated by \( z_{1t} \mid z_{2t} \) and \( z_{2t} \), which exactly corresponds to our intuition regarding the effect of links between both \( z \) subvectors. We conclude that the combination of conditions (i), (ii), (iii) and (viii) is sufficient for weak exogeneity of \( z_{2t} \) with respect to \( \varphi \), though, of course, not necessary.

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

Table 1: Sufficient Conditions for Weak Exogeneity in the DGP [i.e. (i), (ii) and (iii)] to be Preserved in the Model:

<table>
<thead>
<tr>
<th>Type of Misspecification</th>
<th>Sufficient Additional Conditions</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Contemporaneous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a.1) in ( y_t )</td>
<td>(iv)</td>
<td>( \varphi = f(\mu_1) ) (cut preserved)</td>
</tr>
<tr>
<td>(a.2) in ( z_t )</td>
<td>(iv), (v)</td>
<td>cut in ( (z_t^*, z_t^{**}) ) through (v)</td>
</tr>
<tr>
<td>(b) Lag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b.1) in cond. process</td>
<td>(iv), (vi)</td>
<td>no clear interpretation for (vi)</td>
</tr>
<tr>
<td>(b.2) in marg. process</td>
<td>(vii)</td>
<td>( \mu_1 = \lambda_1 ), no clear interpretation but holds if ( I_{t-1}^1 ) only contains lagged ( z_t )</td>
</tr>
<tr>
<td>(c) Classification</td>
<td>(viii)</td>
<td>( \mu_1 = (\lambda_1, \lambda_2^1) ), cut in ( (z_{1t}, z_{2t}) )</td>
</tr>
</tbody>
</table>

5. Some Examples

In this section a number of simple examples will be presented to illustrate the main results of the previous discussion. For notation, definitions, and
properties of the density functions used, we refer to Appendix A of Drève and Richard (1983).

Example 1: Contemporaneous misspecification of $y_t$.

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

$$y_{1t} = \gamma y_{2t} + \beta z_t + \epsilon_{1t}$$

(5.1)

$$y_{2t} = \nu y_{1t} + \delta y_{1,t-1} + \epsilon_{2t}$$

(5.2)

$$z_t = \alpha z_{t-1} + \epsilon_{3t}$$

(5.3)

with Normal i.i.d. errors

$$D \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix} = \mathcal{N}(0, \Sigma), \forall t$$

(5.4)

where the PDS matrix $\Sigma = (\sigma_{ij}); i,j \in \{1,2,3\}$, and with the restriction that $\gamma \nu \neq 1$ (in order to have a solution). The reduced form then factorizes into

$$D \begin{bmatrix} y_{1t} \\ z_t \\ y_{1,t-1} \end{bmatrix} = \mathcal{N}(0, \Sigma_{yy})$$

(5.5)

with

$$\Sigma_{yy} = \begin{bmatrix} 1 & -\xi \\ -\nu & 1 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{11.3} & \sigma_{12.3} \\ \sigma_{21.3} & \sigma_{22.3} \end{bmatrix} \begin{bmatrix} 1 & -\nu \\ -\xi & 1 \end{bmatrix}^{-1}.$$
where \( \sigma_{ij,k} = \sigma_{ij} - \left( \sigma_{ik}\sigma_{jk}/\sigma_{kk} \right) \) (i, j \( \neq \) k), and the marginal process

\[
D(z_t | I_{t-1}, \lambda_2) = f_{N}(z_t | az_{t-1}, \sigma_{33}). \tag{5.6}
\]

If our parameters of interest are the structural coefficients in (5.1) and (5.2), i.e. \( \varphi = (\xi, \beta, \nu, \delta) \), it is obvious that only \( \delta \) can be retrieved from the conditional process (5.5) alone. In addition, our prior assumptions will naturally be formulated in terms of \( \varphi \), whereas e.g. prior independence of \( \varphi \) and \( \alpha \) will generally not induce prior independence of the coefficients in (5.5) and (5.6) [it suffices to look at the coefficients of \( z_{t-1} \) in (5.5), which involve \( \alpha \) and \( \sigma_{33} \)]. A set of sufficient conditions for weak exogeneity of \( z_t \) in (5.5) for \( \varphi \) is

\[
\sigma_{13} = \sigma_{23} = 0 \tag{5.7}
\]

in combination with prior independence of the form \( \lambda_1 \perp \lambda_2 | \lambda_0 \), i.e.

\[
(\xi, \beta, \nu, \delta, \sigma_{11}, \sigma_{12}, \sigma_{22}) \perp (\alpha, \sigma_{33}) \tag{5.8}
\]

given possible initial conditions.

If we now misspecify the model (5.1) - (5.4) under the exogeneity conditions (5.7) and (5.8) by leaving out one of the endogenous variables, e.g. \( y_{2t} \) (= \( y_{2t}^{**} \) in our generic notation), we are left with the following conditional process for \( y_{1t} \) (= \( y_{1t}^{**} \)):

\[
D(y_{1t} | z_t, I_{t-1}, \mu_1) = f_{N}(y_{1t} | \beta z_t + \delta y_{1,t-1}, \sigma_{33}),
\]

\[
\frac{1}{(1-\xi)^2} (\sigma_{11}^2 + 2\xi^2 \sigma_{12}^2 + \xi^4 \sigma_{22}^2), \tag{5.9}
\]

whereas (5.6) is unchanged. From (5.8) we deduce \( \mu_1 = f(\lambda_1) \perp (\alpha, \sigma_{33}) | \lambda_0 \) and it is clear that weak exogeneity of \( z_t \) still holds for any function of \( \mu_1 \) in spite of this form of misspecification, but not for \( \varphi \) that involves the structural parameters, unless e.g. we make the system triangular by \( \xi = 0 \) which would validate conditional inference on \( \beta \). Conversely, leaving out \( y_{1t} \) preserves weak exogeneity of \( z_t \) for \( \delta \) if we impose e.g. \( \nu = 0 \), i.e. induce the opposite type of triangularity. 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 $z_{2t}$:

\begin{align*}
y_t &= \beta z_{1t} + \gamma z_{2t} + \epsilon_{1t} \quad (5.10) \\
z_{1t} &= \alpha z_{1,t-1} + \epsilon_{2t} \quad (5.11) \\
z_{2t} &= \gamma z_{1,t-1} + \nu z_{2,t-1} + \epsilon_{3t} \quad (5.12)
\end{align*}

with the same stochastic assumptions as in (5.4), we can simply verify that weak exogeneity of $z_{1t}$ and $z_{2t}$ jointly in the conditional model for $\rho = f(\beta, \gamma, \sigma_{11})$ is assured by

\begin{align*}
\sigma_{12} &= \sigma_{13} = 0 \quad (5.13)
\end{align*}

and

\begin{align*}
\beta, \gamma, \sigma_{11} &\perp \alpha, \gamma, \sigma_{22}, \sigma_{23}, \sigma_{33} | \omega_0, \quad (5.14)
\end{align*}

which implies for the conditional model

\[ D(y_t | z_{1t}, z_{2t}, I_{t-1}, \lambda_1) = f_N(y_t | \beta z_{1t} + \gamma z_{2t}, \sigma_{11}) . \quad (5.15) \]

Now assume the misspecification of excluding $z_{2t}$ from the conditional model, which is equivalent to falsely restricting $\gamma$ to be zero. The conditional model, marginalized with respect to $z_{2t}$ now becomes [under (5.13)]

\[ D(y_t | z_{1t}, I_{t-1}, \mu_1) = f_N\left( y_t | \begin{bmatrix} \beta + \gamma \frac{\sigma_{23}}{\sigma_{22}} \end{bmatrix} z_{1t} + \gamma \left[ \begin{bmatrix} \sigma_{23} \\ \sigma_{33} \end{bmatrix} \right] z_{1,t-1} + \sigma_{11} \right) + \gamma \left[ \begin{bmatrix} \sigma_{23} \\ \sigma_{33} \end{bmatrix} \right] , \quad (5.16) \]

whereas the marginal process is

\[ D(z_{1t} | I_{t-1}, \mu_2) = f_N\left( z_{1t} | \alpha z_{1,t-1}, \sigma_{22} \right) , \quad (5.17) \]
implying that (5.14) does not generally lead to prior independence of \( \mu_1 \) and \( \mu_2 \). In a classical framework, we do obtain variation free parameterizations \( \mu_1 \) and \( \mu_2 \), but we cannot retrieve the structural parameters \( \beta \) and \( \gamma \) (which, presumably, are the parameters of interest) from (5.16) alone; indeed, we can't even retrieve them from (5.16) and (5.17) combined, due to the marginalized \( z_{2t} \).

So we do have a classical cut in (5.16) - (5.17), but not a Bayesian one. In either case, however, weak exogeneity for \( (\beta, \gamma) \) is precluded.

If we also impose lack of correlation between the errors of (5.11) and (5.12), i.e.

\[
\sigma_{23} = 0 ,
\]

then \( \mu_1 = (\beta, \gamma, \nu, \sigma_{11}^2, \sigma_{22}^2) \) and \( \mu_2 = (\alpha, \sigma_{22}) \) [compare our generic expressions in (4.8) and (4.10)] are still not necessarily prior independent under only (5.14). This is, however, achieved under the additional condition

\[
\gamma, \nu, \sigma_{23} \perp \sigma_{22} | W_0 ,
\]

which, in combination with (5.18), is sufficient for a Bayesian cut operated by \( z_{2t} | z_{1t} \) and \( z_{1t} \).

We note that sufficient conditions for a cut at "DGP" level [i.e. (5.10) - (5.12)] and one between the jointly exogenous variables as in (v), also ensure that weak exogeneity still exists if we, mistakenly, leave aside one of the contemporaneous exogenous variables. Remark that we can only find \( \beta \) from the resulting conditional model, since our misspecification in fact implies the value zero for \( \gamma \).

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 (5.15), but it would obviously affect the misspecified one in (5.16).
Example 3: Lag misspecification for the conditional process.

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

\[ y_t = \beta_1 z_{1,t} + \beta_2 z_{1,t-1} + \beta_3 z_{1,t-2} + \gamma_1 z_{2,t} + \gamma_2 z_{2,t-1} + \gamma_3 z_{2,t-2} + \varepsilon_1 t \]

\[ z_{1,t} = \alpha z_{1,t-1} + \varepsilon_{2,t} \]

\[ z_{2,t} = \xi z_{1,t-1} + \nu z_{2,t-1} + \varepsilon_{3,t} \]

in combination with i.i.d. Normal errors as in (5.4).

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 \( z_t = (z_{1,t}, z_{2,t})' \) in the conditional model for \( y_t \) involves zero covariances, i.e.

\[ \sigma_{12} = \sigma_{13} = 0, \]

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 \( \lambda_2 = (\alpha, \xi, \nu, \sigma_{22}, \sigma_{23}, \sigma_{33}) \) given the initial conditions \( W_0 \), i.e. condition (ii), and finally the requirement that the parameters of interest \( \rho \) can be retrieved from \( \lambda_1 \) alone [condition (iii)]. Under (5.23) the conditional model exactly coincides with the structural model in (5.20) 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 of the conditional process by excluding \( z_{2,t-1} \), i.e. by incorrectly restricting \( \gamma_2 \) to be zero, we are, in fact, considering the process \( D(y_t|z_t, I_{t-1}^*, \mu_1) \), where \( I_{t-1}^* = (z_{1,t-1}, z_{1,t-2}, z_{2,t-2}, z_{2,t-1}, W_0) \) and the excluded part of the information set \( I_{t-1}^* \) comprises \( z_{2,t-1} \). As in (4.14), we shall use the density \( D(I_{t-1}^{**}, | I_{t-1}^*, z_t, \nu_1) \) to marginalize out \( z_{2,t-1} \). From (5.21) and (5.22) we can, using the properties of Normal distributions, deduce the following process for \( I_{t-1}^{**} \), given \( I_{t-1}^* \) and \( z_t \) [compare (4.17)].
where $\sigma_{33.2} = \sigma_{33} - \sigma_{32}^2/\sigma_{22}$, and we note that $I_{t-1}^1$ only contains lagged $z_t$, so that $\rho_1 = f(\lambda_2)$ as in (4.20). The misspecified conditional model then becomes

$$
D(y_t | z_t, I_{t-1}^*, \mu_1) = f_N^1 \left[ \begin{array}{c} y_t \\
\beta_1 - \frac{\gamma_2 \sigma_{22}}{1+\nu^2} \\
\beta_2 + \frac{\gamma_2}{1+\nu^2} \left[ (1+\alpha\nu) \frac{\sigma_{32}}{\sigma_{22}} - \gamma_2 \right] \\
\beta_3 + \frac{\gamma_2}{1+\nu^2} \left[ \gamma - \sigma_{32} \right] \\
\gamma_3 + \frac{\gamma_2}{1+\nu^2} \end{array} \right] z_{1,t-1} + \frac{\gamma_2 \sigma_{33.2}}{1+\nu^2},
$$

from which it is clear that the dependence of $\rho_1$ on $\mu_2 = \lambda_2$ generally also induces links between $\mu_1$ (mixing $\lambda_1$ and $\rho_1$) and $\mu_2$, given that (ii) holds. This destroys the possibility of a Bayesian cut after misspecification.

Remark that none of the links between $\mu_1$ and $\mu_2$ are exact, and therefore a classical cut is operated by (5.25) and the marginal process for $z_t$. However, none of the structural coefficients in $\lambda_1$ can be retrieved from $\mu_1$, so that classical weak exogeneity cannot exist for $\rho = f(\lambda_1)$.

A Bayesian analysis takes both stochastic and exact links into account, and thus a Bayesian cut is formally precluded as the elements of $\mu_1$ will generally not be independent of $\mu_2$. The severe consequences of this type of misspecification are illustrated by the fact that $\mu_1$ and $\mu_2$ seem to be fundamentally intertwined. Trivially, if $\gamma_2 = 0$ is a valid restriction, weak exogeneity is preserved for $\rho = f(\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_3, \sigma_{11})$, but less trivial conditions for weak exogeneity in (5.25) are hard to find.
The sufficient condition (vi) takes the form

$$\begin{bmatrix}
\frac{1}{1+\nu^2} \left[ -\nu \frac{\sigma_{32}}{\sigma_{22}} (1+\alpha \nu) \frac{\sigma_{32}}{\sigma_{22}} - \gamma \nu, \gamma - \alpha \frac{\sigma_{32}}{\sigma_{22}} \nu, \sigma_{33} \right]
\end{bmatrix}$$

which seems very hard to satisfy under the additional condition of independence of $\lambda_1$ and $\lambda_2$ (which e.g. rules out cross-equation restrictions of the type that make $\rho_1$ a function of $\lambda_1 | W_0$). A rather trivial solution is to let the parameters in $\rho_1$ be either known constants or functions of the initial conditions $W_0$. If $z_{2t}$ is a deterministic ($\sigma_{32}=\sigma_{33}=0$) known ($\gamma$ and $\nu$ constant) function of $z_{1t-1}$ and $z_{2t-1}$, we have, e.g., weak exogeneity of $z_t$ for $\rho = f(\beta_1, \sigma_{11})$ 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 lag misspecification of the conditional model can have serious consequences for exogeneity conclusions.

Example 4: Lag misspecification for the marginal process.

Now the same model as in the previous example, i.e. (5.4), (5.20), (5.21) and (5.22) is used, again with the sufficient exogeneity conditions (5.23), (ii) and (iii), and the same $\lambda_1$ and $\lambda_2$.

We now consider what happens if we leave out $z_{2,t-1}$ in the equation for $z_{2,t}$, i.e. (5.22). This implies that we need $D(z_{2,t-1}|z_{1,t-1}, W_0, x_1)$ since $I_{t-1}^* = (z_{1,t-1}, W_0)$ and the excluded $I_{t-1}^{**}$ comprises only $z_{2,t-1}$. If we wish to marginalize the process for $z_t = (z_{1t}, z_{2t})'$ with respect to the past, we need to specify the relevant initial conditions, which are assumed (for simplicity) to be Normally distributed:

$$D\left[\begin{bmatrix} z_{1,0} \\ z_{2,0} \end{bmatrix} | W_0 \right] = \mathcal{N}\left( \begin{bmatrix} z_{1,0} \\ z_{2,0} \end{bmatrix}, \begin{bmatrix} \Gamma_{11} \\ \Gamma_{22} \end{bmatrix} \right).$$

Formally, the distribution for $z_t$ marginalized with respect to $I_{t-1}^1$ will then be given by ($\forall t \geq 1$)
$$D\left[ \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \mid W_0, \kappa \right] = f^2_N \left[ \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \mid A^t \begin{bmatrix} z_{1,0} \\ z_{2,0} \end{bmatrix} \right].$$

$$\Sigma_{22} + A\Sigma_{22}A' + \ldots + A^{t-1}\Sigma_{22}A^{t-1} + A^{t}I_{22}A'^{t}$$, \hspace{1cm} (5.27)

where

$$A = \begin{bmatrix} \alpha & 0 \\ \gamma & \nu \end{bmatrix}$$

and

$$\Sigma_{22} = \begin{bmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{bmatrix}.$$ 

from the model. Note that $\kappa = f(\lambda_2)$ alone, which already gives us assurance that weak exogeneity will be preserved \[\text{compare (4.24) - (4.26)}\]. In particular, if we 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 misspecified marginal process as in (4.23) will be approximated by

$$D\left[ \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \mid \Gamma_{t-1}, \mu_2 \right] = f^2_N \left[ \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \mid \begin{bmatrix} \alpha \\ \gamma + \frac{\sigma_{32}}{\sigma_{22}} (1-\alpha) \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} \right] \begin{bmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} + \frac{\nu^2}{(1-\nu)^2} \sigma_{33,2} \end{bmatrix}.$$ \hspace{1cm} (5.28)

Clearly, $\mu_2 = f(\lambda_2)$ as well\(^8\) and we obtain the result that weak exogeneity in the original model (5.4), (5.20) - (5.22) is not affected at all by this marginal lag misspecification; $z_{1t}$ and $z_{2t}$ are still weakly exogenous for any $\varphi = f(\lambda_1)$, since $\mu_1$ and $\lambda_1$ coincide.

Also, condition (vii) can easily be verified to hold here, as (in the limit)

$$\kappa_1 = \begin{bmatrix} \frac{\gamma}{1-\nu} + \frac{\sigma_{32}(1-\alpha)}{\sigma_{22}(1-\nu)} & \frac{\sigma_{33,2}}{(1-\nu)^2} \end{bmatrix}.$$
An important difference with the previous example is that here an own lag of a variable is neglected in its process. If a lag of $y_t$ would have mistakenly been deleted from the process for $z_{2t}$, then $x_1$ would have been a mixture of $\lambda_1$ and $\lambda_2$, just as $\rho_1$ involved both $\lambda_1$ and $\lambda_2$ in Example 3. A simple illustration of this is found in the introduction (Example 0). Of course, it is not necessarily the case that a mixture of information is avoided when a lagged value of a variable itself is neglected in its process. Take e.g. the case where the conditional model for $y_t | z_t$ should include $y_{t-1}$ and $y_{t-2}$, whereas $y_{t-1}$, say, is excluded by mistake. In order to obtain the model $D(y_{t-1} | z_{t-1}, y_{t-2}, y_{t-3}, w_0, \lambda_1)$, we can use the conditional model $D(y_{t-1} | z_{t-1}, y_{t-2}, y_{t-3}, w_0, \lambda_1)$ but then still need to marginalize out $y_{t-3}$ and, more importantly, we have to condition on $z_t$ instead of $z_{t-1}$, which will typically introduce a dependence of $\rho_1$ on $\lambda_2$ as well.

Of course, the model in (5.20) - (5.22) does favour exogeneity invariance with respect to the dynamic specification of the marginal process as lags of $y_t$ do not appear in the structural equations for $z_t$, but we do feel there is a fundamental asymmetry in both lag misspecifications.

If our information set $I_{t-1}$ (i.e. without the initial conditions) only involves lags of $z_t$, then in (4.24) $x$ will be a function of $\lambda_2$ alone and thus $\mu_2$ as well, making the exogeneity status of $z_t$ in the conditional model for $y_t$ robust with respect to marginal dynamic misspecification.

For lag misspecification of the conditional model itself, such a result cannot be derived, as $p$ in (4.17) parameterizes the joint process of $I_{t-1}$ and $z_t$, thus inducing a possible (and even likely) dependence on $\lambda_2$ even if $I_{t-1}^*$ involved only lagged $y_t$ and the rest of $I_{t-1}^1$ (i.e. the included bit) would exactly be the relevant information set to parameterize $I_{t-1}^*$ given $I_{t-1}$ in terms of $\lambda_1$.

Example 5: Misclassification.

Let us, again, focus on the model introduced in Example 3. However, instead of considering the joint weak exogeneity of $z_{1t}$ and $z_{2t}$, we now classify $z_{1t}$ as endogenous and examine the exogeneity of $z_{2t}$ only. Clearly, under (5.23), (ii) and (iii) we have joint exogeneity, so that this implies what we have called a classification misspecification in Subsection 2.2.c.
The relevant conditional model for $y_t$ and $z_{1t}$, now both assumed endogenous, will then be

$$D \left( \begin{bmatrix} y_t \\ z_{1t} \end{bmatrix} \mid z_{2t}, I_{t-1}, \mu_1 \right) = f_{N}^{2} \left( \begin{bmatrix} y_t \\ z_{1t} \end{bmatrix} \right) \left( \begin{bmatrix} \gamma_1 + \beta_1 \frac{\sigma_{32}}{\sigma_{33}} \\ \sigma_{32}/\sigma_{33} \end{bmatrix} z_{2t} \right) +$$

$$\begin{bmatrix} \beta_1 (\alpha - \gamma) \frac{\sigma_{32}}{\sigma_{33}} + \beta_2 \beta_3 \gamma_2 - \beta_1 \nu \frac{\sigma_{32}}{\sigma_{33}} \gamma_3 \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{1,t-2} \\ z_{2,t-1} \\ z_{2,t-2} \end{bmatrix},$$

$$\begin{bmatrix} \alpha - \gamma \frac{\sigma_{32}}{\sigma_{33}} \\ 0 & -\nu \frac{\sigma_{32}}{\sigma_{33}} \end{bmatrix} \begin{bmatrix} \sigma_{11} + \beta_1^2 \sigma_{22.3} \\ \beta_1 \sigma_{22.3} \\ 0 \end{bmatrix} \begin{bmatrix} \sigma_{22.3} \\ \sigma_{22.3} \end{bmatrix}.$$  \hspace{1cm} (5.29)

with $\sigma_{22.3} = \sigma_{22} / \sigma_{33}^2$, and using (5.23), i.e. $\sigma_{12} = \sigma_{13} = 0$, whereas the marginal model for $z_{2t}$ will be given by

$$D(z_{2t} \mid I_{t-1}, \mu_2) = f_{N}^{1}(z_{2t} \mid \gamma_{z1,t-1} + \nu z_{2,t-1}, \sigma_{33}).$$  \hspace{1cm} (5.30)

Under prior independence of $\lambda_1 = (\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \sigma_{11})$ and $\lambda_2 = (\alpha, \gamma, \nu, \sigma_{22}, \sigma_{32}, \sigma_{33})$, condition (ii), we clearly see that $\mu_1$ and $\mu_2$ are linked. There will typically be stochastic dependences between them and they are not even variation free, since the ratio of the coefficients of $z_{2t}$ and $z_{2,t-1}$ in the second equation of (5.29) is restricted to $-\nu$, where $\nu$ is an element of $\mu_2$ in (5.30). Thus, no cut is operated by (5.29) and (5.30), not even a classical one. Interestingly enough, we know that under (5.23), (ii) and (iii) $z_{1t}$ and $z_{2t}$ are jointly weakly exogenous (see Example 3), yet the same conditions do not suffice for weak exogeneity of $z_{2t}$ separately. This situation was alluded to in Subsection 4.3, where it was stated that removing the links existing between $z_{1t}$ and $z_{2t}$ should be sufficient to retrieve weak exogeneity in the misclassified model.
As in (4.27), we can parameterize the marginal process for $z_t$ in terms of $\lambda_2^1$, sufficient for $z_{1t}$ given $z_{2t}$ and $I_{t-1}$, and $\lambda_2^2$, sufficient for $z_{2t}$ given $I_{t-1}$. In this example, we obtain

$$\lambda_2^1 = \left[ \frac{\sigma_{32}}{\sigma_{33}}, \alpha - \xi \frac{\sigma_{32}}{\sigma_{33}}, -\nu \frac{\sigma_{32}}{\sigma_{33}}, \sigma_{22}, \sigma_{33} \right]$$

(5.31)

and

$$\lambda_2^2 = (\xi, \nu, \sigma_{33})$$

(5.32)

where $\lambda_2^2$, of course, coincides with $\mu_2$ (by definition). Condition (viii), i.e. a Bayesian cut operated by $z_{1t}|z_{2t}$ and $z_{2t}$, is now satisfied under the following sufficient conditions:

$$\sigma_{32} = 0$$

(5.33)

and

$$(\alpha, \sigma_{22}) \perp (\xi, \nu, \sigma_{33})|W_0.$$ 

(5.34)

Indeed, under these additional conditions we find that

$$\mu_1 = (\alpha \beta_1 + \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \alpha, \sigma_{11}, \nu, \sigma_{22}, \beta_1 \sigma_{22}, \beta_2 \sigma_{22}, \sigma_{22})$$

and

$$\mu_2 = (\xi, \nu, \sigma_{33})$$

are independent, i.e. that (5.29) and (5.30) are separated by a Bayesian cut. In addition, all structural parameters from $\lambda_1$ and $\lambda_2^1$ can be retrieved from $\mu_1$, so that $z_{2t}$ is weakly exogenous for $\varphi = f(\lambda_1, \lambda_2^1)$ in (5.29) if we impose (5.23) and (ii), as well as (5.33) and (5.34). The same conditions will also ensure weak exogeneity of $z_{1t}$ in the conditional process for $y_t$ and $z_{2t}$, but the crucial point is that if joint weak exogeneity of $z_{1t}$ and $z_{2t}$ was valid from the start, we should focus on the conditional model of $y_t$ given both $z_{1t}$ and $z_{2t}$ as we are then left with a simpler model (one equation in Example 3 versus two equations here) and, more importantly, it does not imply weak exogeneity of either $z_{1t}$ or $z_{2t}$ separately, unless we make additional assumptions, which are fully irrelevant if our interest is solely in $\lambda_1$ or its transformations [condition (iii)].
Finally, note that conditions (5.33) and (5.34) are sufficient for a Bayesian cut separating $z_{1t}$ and $z_{2t}$ (or the other factorization), but are certainly not necessary. Consider, for example, a prior structure for the marginal model of $z_t$ that falls within the (restrictive) natural-conjugate framework, which is only possible if the zero restriction implicit in (5.21) is not imposed [see e.g. Rothenberg (1963), Richard and Steel (1988), and Steel (1988)], i.e. if the model for $z_t$ becomes

$$D_z = f^2_{N}(z_{1t}, z_{2t}, \lambda_2),$$

where $\eta$ now replaces the zero restriction. This model then results in the following parameterization [compare (5.31)]:

$$\lambda_2 = \begin{bmatrix} \sigma_{32} & \alpha - \zeta \sigma_{32} \\ \sigma_{33} & \eta - \nu \sigma_{33} \end{bmatrix},$$

for $z_{1t}$ given $z_{2t}$ (and $I_{t-1}$), and $\lambda_2^2$ as in (5.32). A natural-conjugate prior structure for the parameters in (5.35) will have the Normal-inverted Wishart form:

$$D(\Pi | \Sigma_{22}) = f^{2 \times 2}_{MN}(\Pi | \Pi_0, \Sigma_{22}, \Theta N_0^{-1}),$$

$$D(\Sigma_{22}) = f^{2 \times 2}_{IW}(\Sigma_{22} | \Sigma_{22}^0, \nu_0),$$

where

$$\Pi = \begin{bmatrix} \alpha \\ \eta \end{bmatrix}.$$  

$\Pi_0$ is its prior mean, $N_0$ and $\Sigma_{22}^0$ are $2 \times 2$ PDS matrices, and $\nu_0 > 1$. If we postmultiply $\Pi$ by the matrix

$$\begin{bmatrix} 1 & 0 \\ -\sigma_{32} & 1 \\ \sigma_{33} & 0 \end{bmatrix},$$

we immediately obtain from (5.37) that

$$\begin{bmatrix} \alpha - \zeta \sigma_{32} \\ \eta - \nu \sigma_{33} \end{bmatrix}$$

given $\sigma_{32}/\sigma_{33}$ and $\sigma_{22.3}$ is independent from $(\zeta, \nu)$ given $\sigma_{33}$. As also, from the properties of inverted Wishart densities, independence between $(\sigma_{32}/\sigma_{33}, \sigma_{22.3})$ and $\sigma_{33}$.
holds, we can conclude that $\lambda_2^1 \perp \lambda_2^2 | \omega_0$ is automatically satisfied under natural-conjugate prior densities, without imposing (5.33) and (5.34). See Steel and Richard (1989) for a related discussion.

6. Implications for Econometric Modelling

In this section some of the conclusions of this paper will be grouped, particularly those with relevance for the practice of building econometric models.

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.

Indeed, reducing the size of the endogenous vector or the amount of lags included may, from a strictly statistical point of view, not be fully warranted, but if the parts excluded are "sufficiently" unimportant for the particular (economic) question that we want to use the model for, such reductions may greatly add to the clarity, the ease of communication or the computational simplicity of the model, without really affecting the salient features of the DGP.

Hendry and Richard (1982) formally examine this issue and arrive at a definition of a "tentatively adequate conditional data characterization" (TACD), which crucially involves the concept of weak exogeneity (among others). The question we attempt to address in this paper is how sensitive weak exogeneity properties that exist at the level of the DGP (or, more realistically, in a TACD) are with respect to blatantly invalid reductions, i.e. misspecifications of our econometric model.

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 $\varphi$ can be expressed as a function of $\mu_1$, 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 misspecification arises, but the model itself may not allow inference on $\varphi$, 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 $\mu_1$.

A related argument was found in the discussion of misclassification, where, again, a cut between the $z$ variables ensures that valid conditioning on a subset of them is preserved. There is, nevertheless, an important difference between both types of misspecification at the inference level. If we omit $z_t^{**}$ from the analysis, we introduce a relatively serious specification error that will often prevent valid inference on $\psi$. Consider Example 2, where dynamics were not introduced in the "correct" conditional model (5.15); if we do not incorporate $z_{2t}$, however, we have to include $z_{1,t-1}$ and $z_{2,t-1}'$ even under (5.18) and (5.19) which preserve a Bayesian cut. The structure of the model is thus affected by this misspecification, and valid inference on $\psi$ or $\sigma_{11}$ cannot be conducted, not even when we take the marginal process for $z_{1t}$ into consideration as well.

On the other hand, the misclassification in Example 5 does not distort the structure of the model, only the way it is factorized, and we note that under (5.33) and (5.34), sufficient conditions for ensuring a cut, we can retrieve all structural parameters, and inference on $\psi = f(\lambda_1, \lambda_2^1)$ is warranted. The crucial difference is that $\mu_1$ is a function of $(\lambda_1, \lambda_2^*)$ in (4.8), whereas it is equivalent to $(\lambda_1^1, \lambda_2^1)$ in (4.29), so that knowledge of $\mu_1$ directly translates into information on $\lambda_1$ and $\lambda_2^1$. The "correct" conditional model for $y_t$ given $z_t$ (and $I_{t-1}$) can still be derived from (5.29), so that misclassification is more an issue of computational complexity [two equations if e.g. (5.33) and (5.34) hold, and otherwise even three equations instead of just one] than of invalid inferences.

Misspecifying the lags included in the conditional model was seen to have potentially very serious consequences in the discussion of Examples 3 and 4. As we have to use $D(I_{t-1}^*, I_{t-1}^*, z_t, \rho_1)$ [in (4.14)], where $z_t$ is a conditioning variable, it is rather unlikely that $\rho_1$ should not be contaminated by $\lambda_2$. This is illustrated by the expressions in (4.18) - (4.20) and by Example 3, where,
barring rather trivial cases, a Bayesian cut is destroyed. Admittedly, if $I_{t-1}^1$ involves more than just lagged $z_t$ variables, we can possibly save a cut by introducing restrictions that make $\varphi_1$ a function of $\lambda_1$, but such cases seem very exceptional and highly unlikely to occur in practice. And even if a cut is preserved, we still face a misspecified conditional model. Consider Example 3, where a cut is operated if $\sigma_{32} = \sigma_{33} = 0$ and $\xi$ and $\nu$ are constants, but only $\beta_1$ and $\sigma_{11}$ can be retrieved from $\mu_t$. In the special case that $\nu = 1$, inference can also be conducted on $\sum_{i=2}^{3} \beta_i$ and $\sum_{i=1}^{3} \gamma_i$, but the misspecification prevents inference on the remaining structural coefficients themselves.

Such misspecification of the conditional model is not present if we only get the dynamics of the marginal process wrong. 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, except for the influence it may have on the power of formal tests for exogeneity, that we may wish to conduct. Such tests are found e.g. in Engle (1984) and Holly (1985) in a classical framework, whereas Bayesian extensions are discussed in Lubrano and Marimoutou (1988).

Summarizing, it seems that both misclassification and choosing a wrong lag structure for the marginal model are relatively benign types of misspecification, as at least the conditional model is not distorted, and we can think of fairly "common" conditions under which weak exogeneity of the DGP is preserved.

If we inadvertently leave out contemporaneous variables, we do misspecify the conditional model, but a Bayesian cut carries over if we omit part of $y_t$. 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, with the possible exception of misclassification where the best strategy seems to
be to test as many current conditioning variables as possible jointly for weak exogeneity.

Although the various types of misspecification discussed here 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, 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 cases of misclassification or 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 $\lambda_1$ alone, and $\lambda_1$ can be recovered from the parameterization of the joint process, say $\mu$. Recall that the conditional model is not subject to misspecification in these cases. The situation becomes much worse, however, with the other types of specification errors, as $\mu$ generally does not allow inference on $\varphi$ then, so that even from the joint model valid inference on $\varphi$ cannot be conducted.

Finally, the present analysis can be extended to cover combinations of the various types of misspecification, other forms of exogeneity (e.g. strong and super exogeneity), errors that affect the initial conditions $W_0$, more general models than the sequential (time-series) models used here, or situations that cannot be handled in terms of densities, but we feel that clarity was served by excluding these complications at this stage, whereas a number of important, and practically relevant, issues were addressed here in a relatively formal framework.
Footnotes

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 $\sigma$-fields, which is, in itself, already a (possibly false) assumption. The latter type of misspecification is not explicitly considered either.

2. Note that we shall use the same notation $\mu$ 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.

3. Whenever we introduce a parameterization in the text, like $\rho_1$ in (4.14), we shall always define this to be a sufficient parameter set for the process in which it appears.

4. Of course, the fact that the expression for this coefficient has $\alpha$ and $\sigma_{33}$ appearing in it, does not, in itself, imply that independence with respect to $\lambda_2$ will be lost. The natural-conjugate case discussed at the end of Example 5 illustrates this fact. However, such special cases are very limited in number and felt to be the exception rather than the rule.

5. Since the process for $z_t$ does not depend on lagged $y$'s, we have a noncausality condition and can even conclude that strong exogeneity [see Engle et al. (1983)] holds under (5.7) and (5.8).

6. Exact links between $\mu_1$ and $\mu_2$ can be introduced by imposing exclusion restrictions on the structural coefficients of the model; if e.g. $\beta_2$ and $\beta_3$ are zero, then the ratio of the coefficients of $z_{1,t-1}$ and $z_{2,t-1}$ in (5.25) is an exact function of the elements in $\mu_2$.

7. This follows from the fact that, in this example, $\rho_1$ is only a function of $\lambda_2$ as in (4.20), but is not generally the case.
8. This holds true whether we approximate the marginal process or not. In particular, it even holds under nonstationarity.


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


Osiewalski, J. and M.F.J. Steel, 1989, A Bayesian analysis of exogeneity in models pooling time-series and cross-section data, CentER discussion paper 8914 (Tilburg University, Tilburg).


Steel, M.F.J., 1988, Seemingly unrelated regression equation systems under diffuse stochastic prior information: A recursive analytical approach, CentER discussion paper 8805 (Tilburg University, Tilburg).

Steel, M.F.J. and J.F. Richard, 1989, Bayesian multivariate exogeneity analysis: An application to a UK money demand equation, CentER discussion paper 8929 (Tilburg University, Tilburg).
## Discussion Paper Series, CentER, Tilburg University, The Netherlands:

<table>
<thead>
<tr>
<th>No.</th>
<th>Author(s)</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>8801</td>
<td>Th. van de Klundert and F. van der Ploueg</td>
<td>Fiscal Policy and Finite Lives in Interdependent Economies with Real and Nominal Wage Rigidity</td>
</tr>
<tr>
<td>8802</td>
<td>J.R. Magnus and H. Pesaran</td>
<td>The Bias of Forecasts from a First-order Autoregression</td>
</tr>
<tr>
<td>8804</td>
<td>F. van der Ploeg and A.J. de Zeeuw</td>
<td>Perfect Equilibrium in a Model of Competitive Arms Accumulation</td>
</tr>
<tr>
<td>8805</td>
<td>M.F.J. Steel</td>
<td>Seemingly Unrelated Regression Equation Systems under Diffuse Stochastic Prior Information: A Recursive Analytical Approach</td>
</tr>
<tr>
<td>8806</td>
<td>Th. Ten Raa and E.N. Wolff</td>
<td>Secondary Products and the Measurement of Productivity Growth</td>
</tr>
<tr>
<td>8807</td>
<td>F. van der Ploeg</td>
<td>Monetary and Fiscal Policy in Interdependent Economies with Capital Accumulation, Death and Population Growth</td>
</tr>
<tr>
<td>8901</td>
<td>Th. Ten Raa and P. Kop Jansen</td>
<td>The Choice of Model in the Construction of Input-Output Coefficients Matrices</td>
</tr>
<tr>
<td>8902</td>
<td>Th. Nijman and F. Palm</td>
<td>Generalized Least Squares Estimation of Linear Models Containing Rational Future Expectations</td>
</tr>
<tr>
<td>8903</td>
<td>A. van Soest, I. Woittiez, A. Kapteyn</td>
<td>Labour Supply, Income Taxes and Hours Restrictions in The Netherlands</td>
</tr>
<tr>
<td>8904</td>
<td>F. van der Ploeg</td>
<td>Capital Accumulation, Inflation and Long-Run Conflict in International Objectives</td>
</tr>
<tr>
<td>8905</td>
<td>Th. van de Klundert and A. van Schaik</td>
<td>Unemployment Persistence and Loss of Productive Capacity: A Keynesian Approach</td>
</tr>
<tr>
<td>8907</td>
<td>J. Osiewalski</td>
<td>Posterior Densities for Nonlinear Regression with Equicorrelated Errors</td>
</tr>
<tr>
<td>8908</td>
<td>M.F.J. Steel</td>
<td>A Bayesian Analysis of Simultaneous Equation Models by Combining Recursive Analytical and Numerical Approaches</td>
</tr>
<tr>
<td>No.</td>
<td>Author(s)</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>8909</td>
<td>F. van der Ploeg</td>
<td>Two Essays on Political Economy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(i) The Political Economy of Overvaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii) Election Outcomes and the Stockmarket</td>
</tr>
<tr>
<td>8910</td>
<td>R. Gradus and A. de Zeeuw</td>
<td>Corporate Tax Rate Policy and Public and Private Employment</td>
</tr>
<tr>
<td>8911</td>
<td>A.P. Barten</td>
<td>Allais Characterisation of Preference Structures and the Structure of Demand</td>
</tr>
<tr>
<td>8912</td>
<td>K. Kamiya and A.J.J. Talman</td>
<td>Simplicial Algorithm to Find Zero Points of a Function with Special Structure on a Simplootope</td>
</tr>
<tr>
<td>8913</td>
<td>G. van der Laan and A.J.J. Talman</td>
<td>Price Rigidities and Rationing</td>
</tr>
<tr>
<td>8914</td>
<td>J. Osiewalski and M.F.J. Steel</td>
<td>A Bayesian Analysis of Exogeneity in Models Pooling Time-Series and Cross-Section Data</td>
</tr>
<tr>
<td>8915</td>
<td>R.P. Gilles, P.H. Ruys and J. Shou</td>
<td>On the Existence of Networks in Relational Models</td>
</tr>
<tr>
<td>8916</td>
<td>A. Kapteyn, P. Kooreman and A. van Soest</td>
<td>Quantity Rationing and Concavity in a Flexible Household Labor Supply Model</td>
</tr>
<tr>
<td>8917</td>
<td>F. Canova</td>
<td>Seasonalities in Foreign Exchange Markets</td>
</tr>
<tr>
<td>8918</td>
<td>F. van der Ploeg</td>
<td>Monetary Disinflation, Fiscal Expansion and the Current Account in an Interdependent World</td>
</tr>
<tr>
<td>8919</td>
<td>W. Bossert and F. Stehling</td>
<td>On the Uniqueness of Cardinally Interpreted Utility Functions</td>
</tr>
<tr>
<td>8920</td>
<td>F. van der Ploeg</td>
<td>Monetary Interdependence under Alternative Exchange-Rate Regimes</td>
</tr>
<tr>
<td>8921</td>
<td>D. Canning</td>
<td>Bottlenecks and Persistent Unemployment: Why Do Booms End?</td>
</tr>
<tr>
<td>8922</td>
<td>C. Fershtman and A. Fishman</td>
<td>Price Cycles and Booms: Dynamic Search Equilibrium</td>
</tr>
<tr>
<td>8923</td>
<td>M.B. Canzoneri and C.A. Rogers</td>
<td>Is the European Community an Optimal Currency Area? Optimal Tax Smoothing versus the Cost of Multiple Currencies</td>
</tr>
<tr>
<td>8924</td>
<td>F. Groot, C. Withagen and A. de Zeeuw</td>
<td>Theory of Natural Exhaustible Resources: The Cartel-Versus-Fringe Model Reconsidered</td>
</tr>
<tr>
<td>8925</td>
<td>O.P. Attanasio and G. Weber</td>
<td>Consumption, Productivity Growth and the Interest Rate</td>
</tr>
<tr>
<td>No.</td>
<td>Author(s)</td>
<td>Title</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8926</td>
<td>N. Rankin</td>
<td>Monetary and Fiscal Policy in a 'Hartian' Model of Imperfect Competition</td>
</tr>
<tr>
<td>8927</td>
<td>Th. van de Klundert</td>
<td>Reducing External Debt in a World with Imperfect Asset and Imperfect Commodity Substitution</td>
</tr>
<tr>
<td>8928</td>
<td>C. Dang</td>
<td>The $D_1$-Triangulation of $R^n$ for Simplicial Algorithms for Computing Solutions of Nonlinear Equations</td>
</tr>
<tr>
<td>8929</td>
<td>M.F.J. Steel and J.F. Richard</td>
<td>Bayesian Multivariate Exogeneity Analysis: An Application to a UK Money Demand Equation</td>
</tr>
<tr>
<td>8930</td>
<td>F. van der Ploeg</td>
<td>Fiscal Aspects of Monetary Integration in Europe</td>
</tr>
<tr>
<td>8931</td>
<td>H.A. Keuzenkamp</td>
<td>The Prehistory of Rational Expectations</td>
</tr>
<tr>
<td>8932</td>
<td>E. van Damme, R. Selten and E. Winter</td>
<td>Alternating Bid Bargaining with a Smallest Money Unit</td>
</tr>
<tr>
<td>8933</td>
<td>H. Carlsson and E. van Damme</td>
<td>Global Payoff Uncertainty and Risk Dominance</td>
</tr>
<tr>
<td>8934</td>
<td>H. Huizinga</td>
<td>National Tax Policies towards Product-Innovating Multinational Enterprises</td>
</tr>
<tr>
<td>8935</td>
<td>C. Dang and D. Talman</td>
<td>A New Triangulation of the Unit Simplex for Computing Economic Equilibria</td>
</tr>
<tr>
<td>8936</td>
<td>Th. Nijman and M. Verbeek</td>
<td>The Nonresponse Bias in the Analysis of the Determinants of Total Annual Expenditures of Households Based on Panel Data</td>
</tr>
<tr>
<td>8937</td>
<td>A.P. Barten</td>
<td>The Estimation of Mixed Demand Systems</td>
</tr>
<tr>
<td>8938</td>
<td>G. Marini</td>
<td>Monetary Shocks and the Nominal Interest Rate</td>
</tr>
<tr>
<td>8939</td>
<td>W. Güth and E. van Damme</td>
<td>Equilibrium Selection in the Spence Signaling Game</td>
</tr>
<tr>
<td>8940</td>
<td>G. Marini and P. Scaramozzino</td>
<td>Monopolistic Competition, Expected Inflation and Contract Length</td>
</tr>
<tr>
<td>8941</td>
<td>J.K. Dagsvik</td>
<td>The Generalized Extreme Value Random Utility Model for Continuous Choice</td>
</tr>
<tr>
<td>8942</td>
<td>M.F.J. Steel</td>
<td>Weak Exogeneity in Misspecified Sequential Models</td>
</tr>
</tbody>
</table>