

Resolving Three ‘Intractable’ Problems using a *Gets* Approach

David F. Hendry and Hans-Martin Krolzig*

Economics Department, Oxford University, Manor Road, Oxford OX1 3UQ.

August 6, 2004

Abstract

The three ‘intractable’ problems are: selecting from a set of perfectly collinear variables the subset which actually enters the DGP; selecting a regression model when there are more variables than observations; and selecting a model when there are endogenous variables in simultaneous equations, but no *a priori* restrictions. Using a variant of *Gets*, each problem is embedded in the theory of reduction, and a viable approach shown, exploiting the properties of a new tool to solve previously intractable problems.

1 Introduction

We consider the application of general-to-specific (*Gets*) model selection procedures to three problems conventionally deemed intractable. First, when there are perfectly collinear variables; secondly, regression estimation with more variables n than observations T ; and thirdly selecting equations which are embedded in a simultaneous system, but no *a priori* restrictions are known. In each of these settings, the initial general unrestricted model (GUM) cannot be viably estimated at the outset. Instead, we consider ‘subset selection’ across combinations of candidate variables, each leading to a terminal model, such that the intermediate models are always identified and estimable. It is assumed throughout that the DGP, if known, is identifiable and estimable from the available data, and the only problem is selecting the relevant regressors from the candidate variables. For general analyses of *Gets*, see *inter alia* Hoover and Perez (1999), Krolzig and Hendry (2001), Hendry and Krolzig (2001, 2003, 2004a), Campos, Hendry and Krolzig (2003), Granger and Hendry (2004), and Campos, Ericsson and Hendry (2004). To understand how each of these problems can be tackled, *PcGets* can be viewed as a tool that extends the available technology and hence renders feasible previously intractable situations.

For one class of perfectly collinear variables, exploring all the feasible identified paths highlights which minimal subset actually enters the DGP. Commencing from each ‘redundant’ variable in turn, variables are eliminated till (a) a non-collinear group is found and then (b) only a significant subset remains as selected by the usual *PcGets* procedure. The artificial data illustration shown below is consistent with the feasibility of this approach.

When $n > T$, regression models based on subsets of $n_1 \leq T/2$ variables are explored *seriatim*, and a new joint model is formulated from the terminal models selected from these subset combinations, to which *PcGets* is then applied as usual. We show that the DGP can again be found by this approach, first for the specialized case of orthogonal regressors, then suggest how the analysis can be generalized to correlated variables. ‘Negative’ correlations are shown to be problematic, but recombining the initial blocks is likely to alleviate the main difficulties. Hendry, Johansen and Santos (2004) study in detail

*We are grateful to Oxford University Research Development Fund and the Economics Department for financial support, and to Grayham Mizon for helpful comments on an earlier draft.

the special case where indicator dummies for every observation are included in a locatio/scale model. We also comment on the application of our approach to selecting forecasting equations, and for selecting non-linear representations: Castle (2004) provides an application of these, which also compares RETINA (see Perez-Amaral, Gallo and White, 2003, 2004) with *PcGets*.

Thirdly, econometric model selection is apparently especially difficult when the parameters of interest are those of simultaneous equations: key issues concern endogeneity of potential regressors, the validity and ‘strength’ of putative instrumental variables, and the identification of the resulting equations. It transpires that conditional on knowing the set of all instruments (i.e., all weakly exogenous conditioning variables), jointly selecting identified equations for the endogenous variables and instruments can be viewed rather differently using this new tool, and we describe how that can be achieved. The key conceptual step is that simultaneity is a potential reduction on the parameter space of a system, not a generalization from a ‘reduced form’ to a simultaneous system. An earlier example of a related shift was the realization by Sargan (1980) that autoregressive residuals are a restriction on a dynamic equation and not a generalization. In fact, the idea that simultaneous equations models are reductions of conditional systems has been explored previously in the context of programming (see Hendry, Neale and Srba, 1988), modelling procedures (see Clements and Mizon, 1991), model evaluation (see Hendry and Mizon, 1993), and testing particular ‘structures’ from economic theory for being valid reductions (see Hall, Mizon and Welfe, 2000). However, the possibility of selecting simultaneous equations specifications in the absence of any prior information is more surprising.

The paper is organized as follows. Section 2 explores the problem of selecting from a set of perfectly collinear variables. Then section 3 looks at model selection when there are too many regressors for the available sample. Section 4 considers selecting simultaneous equations models, first formulating the setting, the role of identification in the DGP, and the choice of instruments for each endogenous variable, which then characterizes the system. Subsection 4.6 then explains how to construct an identified simultaneous representation, illustrated by a theoretical two-equation model, and an artificial data example. Section 5 concludes. An appendix records the current *PcGets* algorithm.

2 Perfectly collinear variables

Sometimes economic theory is unable to specify the precise relevant lag transformations, which might involve combinations of levels (z_t), differences (Δz_t), distributed lags (z_t, z_{t-1}), and moving averages ($\bar{z}_t = z_t + z_{t-1}$), etc. where a subset of these determines a variable y_t when the maximum lag is known to be one. Despite the perfect collinearity, all four variables just noted can be entered, and if only multi-path searches are used, the correct combination selected by *PcGets*, subject to the usual probabilities of retention of relevant and irrelevant variables.

Consider a setting in which the following data generation process (DGP) is postulated:

$$y_t = \beta_0 1_{\{z_t\}} z_t + \beta_1 1_{\{z_{t-1}\}} z_{t-1} + \beta_2 1_{\{\bar{z}_t\}} \bar{z}_t + \beta_3 1_{\{\Delta z_t\}} \Delta z_t + v_t \quad (1)$$

where $v_t \sim \text{IN}[0, \sigma_v^2]$ and $1_{\{j\}}$ are indicator variables that take the value unity if the regressor in question enters the DGP, and zero otherwise. At most two indicators $1_{\{j\}}$ are unity, but the investigator does not know which, so formulates the general unrestricted model (GUM):

$$y_t = \gamma_0 z_t + \gamma_1 z_{t-1} + \gamma_2 \bar{z}_t + \gamma_3 \Delta z_t + u_t. \quad (2)$$

The regressors in (2) are perfectly collinear, and conventional inversion routines will arbitrarily assign that singularity to some subset, not necessarily retaining the appropriate set.

Instead, consider a sequence of search paths commencing from every variable in turn, which is deleted, then every next variable is deleted, and so on. Thus, first drop Δz_t (say) which still leaves a collinear set, then \bar{z}_t which leaves a non-collinear set for which a conventional search can be conducted and a terminal model selected. Then starting afresh from (2), drop z_t then z_{t-1} and so on, such that 4 initial paths, 3 second paths for each of these, etc., are explored. In each case, once a non-collinear set results, the usual algorithm shown in the appendix can operate. Say, for example, only $1_{\{\Delta z_t\}} = 1$, then such a terminal model will be explored along several paths, should be selected on some when β_3 is sufficiently significant in the DGP, and will parsimoniously dominate the other selections (e.g., on the Schwarz, 1978, criterion, *SIC*). There are 11 possible models, including the null, of which only 6 are distinct.

We conducted five small one-off experiments on artificial data where, in turn y_t depended on: (a) \bar{z}_t ; (b) Δz_t ; (c) z_t ; (d) z_{t-1} ; and (e) z_t and Δz_t ; but in each case all of z_t , z_{t-1} , \bar{z}_t and Δz_t were entered as regressors in the GUM (2). *PcGets* is not yet programmed to follow all the feasible paths, but as a partial implementation (namely searching once a non-collinear set was imposed) gave the correct answers in (a), (b), (c) and (d), but in (e) selected the more orthogonal representation, \bar{z}_t and Δz_t , which is of course an equivalent representation. Selecting purely by *SIC* across all the possible models in each case would deliver essentially the same outcomes here. Indeed, for the current example, an investigator could simply estimate the conventional distributed lag model $y_t = \gamma_0 z_t + \gamma_1 z_{t-1} + u_t$ and decide ex post from the parameter estimates how to transform or simplify the model.

Nevertheless, this is our first example of how the new tool can yield new insights: one might have suspected that perfect collinearity was insoluble, but multi-path searches can resolve the choice. In retrospect, the solution is seen to be simply an automation of what many empirical modellers do in practice, a theme that will emerge again for the first case in the next section. More sophisticated aspects of the search process usually employed by *PcGets*, such as pre-search tests designed to reduce the computational burden of path exploration, would arbitrarily eliminate whatever variables the inversion routine happened to treat as redundant, and so preclude finding the most useful representation. Indeed, the outcome may depend on the order in which variables are entered in the regression. Thus, the pure multi-path search approach following from Hoover and Perez (1999) is seen as the key ingredient here.

3 Model selection confronting too many regressors

We have several times been asked about this rather different ‘singular case’ by investigators who have had to confront an ‘excess variables’ problem when modelling, namely $n > T$. Some researchers seem to have included small blocks of variables in their search for significant regressors, but we doubt such a procedure will be effective, and propose instead the variant of *Gets* now described.

3.1 Adding T indicator variables

We first note the behaviour for regressions which are ‘saturated’ by indicator variables, as studied in more detail by Hendry *et al.* (2004). Consider an observed random variable $y_t \sim \text{IN}[\mu, \sigma_v^2]$ for $t = 1, \dots, T$, where μ is the parameter of interest, but an investigator is completely uncertain where outliers (if any) may lurk. She therefore defines a saturating set of T indicator dummies $I_t = 1_{\{t=t_i\}}$, one for every t_i , and wishes to regress y_t on $\{\mu, I_t, t = 1, \dots, T\}$ (although any indicator could be dropped without loss). Clearly, a perfect fit will result from such a regression, with every indicator ‘infinitely significant’, so nothing is learned.

Consider instead entering half of the indicators (e.g., I_t for $t = 1, \dots, T/2$) together with the intercept, then that GUM is selected from, and the resulting terminal model is stored. Now enter the other half (I_t , $t = T/2 + 1, \dots, T$) and repeat, again storing the result. Finally, formulate a model where all significant selected indicators from the two terminal model are combined, and re-select. Clearly, the ‘perfect fit’ problem does not arise. Moreover, since αT outliers will occur on average for a significance level α , then αT indicators will be selected on average. It may be thought that the huge number of $T/2$ indicators entered in each stage might induce spurious significance, but all that happens is that those observations are ‘dummied out’ for estimating μ , which is then just the mean of the remaining sample. Thus, even in this extreme setting of saturating indicators, a feasible algorithm exists. Notice also that ‘repeated testing’ does not occur, in the sense that an indicator will be significant at level α if and only if there is an α -level outlier at that observation. Additional regressors entail an inability to add half the indicators at each stage, and may necessitate exploring combinations, but do not otherwise affect this analysis. Thus, the distribution of this procedure under the null can be established.

Interestingly, under the alternative, if μ takes two values μ and μ^* , pre and post T^* say, then providing each sub-sample includes indicators covering all of the break, then blocks of I_t will be significant with an average value equal to $\mu - \mu^*$, and thereby reveal a step shift. Moreover, if y_{t-1} had been included as a regressor, when there were no indicators, then its coefficient would reflect the step shift by being close to unity, removing the break except at its end points where impulses of roughly equal magnitude, opposite sign would be created: a conventional ‘outlier removal’ approach would then conclude with the incorrect model, albeit one which might be fine for forecasting. Amalgamating these blocks of indicators would clarify that there is a step shift, but no dynamics, in this simple case. Thus, there are clear uses for a ‘saturation’ approach. An example in a cross-section context is provided by Hendry and Krolzig (2004c).

3.2 $n > T$ regressors

Now consider the more general situation of two groups of variables relevant to determining a variable of interest y_t , denoted $\mathbf{x}_{i,t}$, for $i = 1, 2$, of dimensions $n_i \ll T$ respectively where $n = n_1 + n_2 > T$, but both (n_1 and n_2) are sufficiently smaller than T for regressions on them to be estimable. The analysis can be generalized for more groups, although the computational burden rises in a combinatorial fashion. Many of the elements in each set need not have an effect, but we assume components of each are relevant. For $t = 1, \dots, T$, let the DGP be:

$$y_t = \sum_{i=1}^2 \beta'_i \mathbf{x}_{i,t} + \epsilon_t \quad \text{where } \epsilon_t \sim \text{IN} [0, \sigma_\epsilon^2], \quad (3)$$

where the β_i contain many zeros, such that the remaining non-zero parameters number $k \ll T$. The central case is where $n_1 = n_2$, so two ‘general models’ are considered of the form ($\tilde{\cdot}$ denotes ‘claimed to be distributed as’):

$$y_t = \gamma'_j \mathbf{x}_{j,t} + u_{j,t} \quad \text{where } u_{j,t} \tilde{\cdot} \text{IN} [0, \sigma_{u_j}^2], \quad (4)$$

Then as neither $\beta_i = \mathbf{0}$, the selected model from each of (4) will not coincide with that selected from (3) when the latter is estimable. If (4) cannot be estimated, sub-divide further; however, two sets explains the logic. We assume that all the models are congruent against their own information sets, perhaps by design, otherwise consistent covariance estimators will need to be used initially (see White, 1980).

In more detail, first select the best model from:

$$y_t = \gamma'_1 \mathbf{x}_{1,t} + u_{1,t} \quad (5)$$

to get the first terminal model:

$$y_t = \boldsymbol{\lambda}'_1 \mathbf{x}_{1,t}^* + v_{1,t} \text{ where } v_{1,t} \underset{c}{\sim} \text{IN} [0, \sigma_{v_1}^2], \quad (6)$$

where $\mathbf{x}_{1,t}^*$ denotes the n_1^* retained components of $\mathbf{x}_{1,t}$ such that all elements of $\boldsymbol{\lambda}_1$ are ‘significantly’ non-zero at a relatively loose level δ .

Similarly for $\mathbf{x}_{2,t}$, commence from:

$$y_t = \boldsymbol{\gamma}'_2 \mathbf{x}_{2,t} + u_{2,t} \quad (7)$$

to get a second terminal model with n_2^* regressors:

$$y_t = \boldsymbol{\lambda}'_2 \mathbf{x}_{2,t}^* + v_{2,t} \text{ where } v_{2,t} \underset{c}{\sim} \text{IN} [0, \sigma_{v_2}^2]. \quad (8)$$

Now, assuming $n_1^* + n_2^* < T$, re-start the selection at the overall desired significance level $\alpha < \delta$ from:

$$y_t = \boldsymbol{\theta}'_1 \mathbf{x}_{1,t}^* + \boldsymbol{\theta}'_2 \mathbf{x}_{2,t}^* + \xi_t \text{ where } \xi_t \underset{c}{\sim} \text{IN} [0, \sigma_\xi^2], \quad (9)$$

to end with the final selection involving $n_1^{**} + n_2^{**}$ regressors:

$$y_t = \boldsymbol{\rho}'_1 \mathbf{x}_{1,t}^{**} + \boldsymbol{\rho}'_2 \mathbf{x}_{2,t}^{**} + \eta_t \text{ where } \eta_t \underset{c}{\sim} \text{IN} [0, \sigma_\eta^2]. \quad (10)$$

This provides the selection approximating (3).

An important variant is to ‘cross-over’ the variables. First partition each of $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ into two halves, producing four groups. Now select (say) the first halves of $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ as another GUM, then the second halves (assuming the ordering is arbitrary), then the cross-pairings. Add all the variables selected in the resulting terminal models from those searches as the next GUM. There are ${}_2C_4 = 6$ combinations [(1,2), (3,4); (1,3), (2,4); (1,4), (2,3)] to be investigated, but such a procedure is easily automated.

3.3 Properties of the selected model

If $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ are mutually orthogonal in the population, and (9) is feasible, then this procedure delivers the correct answer unless the significance of the relevant variables is close to marginal, so the improved fit of the combination is essential to retain them. Critical values will probably need to be loose in the early subset selections to avoid that problem. Conversely, stringent critical values will probably be needed at the final stage to avoid retaining too many variables that are significant adventitiously. If, say, $n_1 = n_2 = 100 < T = 150$, then a 1% level would only entail 2 irrelevant variables retained on average despite 200 candidate variables at the start. We first consider the IID case, so the sub-models are congruent but incomplete.

At stage 1, selecting from (5) when the DGP is as in (3), under the orthogonality condition that $\mathbf{M}_{1,2} = \text{E}[\mathbf{x}_{1,t} \mathbf{x}'_{2,t}] = \mathbf{0}$:

$$u_{1,t} = \boldsymbol{\beta}'_2 \mathbf{x}_{2,t} + \epsilon_t \quad (11)$$

so $\boldsymbol{\gamma}_1$ is unbiasedly estimated, but with the equation error variance of $\sigma_\epsilon^2 + \boldsymbol{\beta}'_2 \mathbf{M}_{2,2} \boldsymbol{\beta}_2$ under stationarity, where $\mathbf{M}_{2,2} = \text{E}[\mathbf{x}_{2,t} \mathbf{x}'_{2,t}]$. Thus, the primary problems are lack of significance of variables that matter due to ‘underfitting’, and retention of:

$$\alpha (n_1 - k_1)$$

irrelevant variables on average when a test of size α is used. Here, we imagine $\alpha = 0.1$ at stage 1, to minimize missing variables which matter. For example, if $n_1 = 100$ and $k_1 = 10$ (say), all

relevant variables with t-values in excess of about 1.65 in absolute value (as judged using $\sigma_{u_1}^2$) will be retained together with 9 irrelevant variables. When $\beta_2' M_{2,2} \beta_2$ is large compared to σ_e^2 even important variables might be missed, hence the need to include many different pairings of sets of regressors. One variant is to select the first subset by the simple correlations between each $x_{i,t}$ and y_t to ensure that all obviously important regressors are included from the outset, as is done in RETINA. A second variant worth exploring by simulation is then to retain from every terminal model for addition to the next step all variables that are highly significant, and therefore contribute importantly to the goodness of fit. Then a more stringent selection criterion could be used, reducing the chances of retaining spurious variables at intermediate stages. Such additions would cumulate as alternative blocks were investigated.

Similarly for selecting from the $\mathbf{x}_{2,t}$. Between the two intermediate terminal models, about 40 variables will result for the example in the combination of the terminal models:

$$y_t = \theta_1' \mathbf{x}_{1,t}^* + \theta_2' \mathbf{x}_{2,t}^* + \xi_t.$$

At stage 2, set $\alpha = 0.01$ (say), so only about 2 adventitiously-significant variables will on average be retained from the initial 200, whereas all relevant variables that have absolute t-values in excess of about 2.6 in the DGP will be retained. Alternatively, depending on the investigators loss function, $\alpha = 0.025$ would be closer to the value implicit in *SIC*, and should retain all variables with absolute t-values in excess of about 2.25 (thereby including potentially 5 irrelevant). Such trade-offs between size and power should be an explicit consideration in designing the selection strategy.

The third stage of cross-over is necessary for population-orthogonal variables, as these cannot be jointly orthogonal for $n > T$, and may deliver additional relevant variables in some terminal models.

If $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ are positively correlated, efficiency of selection is lower even if a regression can be conducted in a single stage, and so must be lower for the multi-stage process proposed here. Nevertheless, the procedure is likely to work quite well, since the intercorrelations should entail that proxy variables improve the fits at the intermediate stages, which should raise the probability of retaining the relevant variables within each subset. When the ‘correct’ regressors are included in the final model, proxies should be eliminated, subject to the usual caveats affecting collinear situations.

The most difficult case is when $\mathbf{x}_{1,t}$ and $\mathbf{x}_{2,t}$ are negatively correlated given that all the β_i parameters are positive, since then each sub-set needs to be included for the other to be included as well. In practice, some negative correlations are likely, with some near orthogonal and others positive, so cross-matching is needed to ensure all appropriate pairs at least are always jointly included (high-order interactions may be problematic).

On a 40-variable IID, orthogonal regressor illustrative artificial data set of 40 observations with 10 relevant regressors, spanning population t-values from 2–6, using the Liberal strategy, the least significant variable of the 10 relevant regressors was missed, and one irrelevant was retained when four combinations of two blocks of ten were used. However, only the most significant 7 relevant variables were retained if only two starting points of twenty variables were tried.

When the data are not IID, so sub-models may be non-congruent, HAC coefficient standard errors may be needed during intermediate stages to ensure that terminal models include all DGP-relevant variables, but should not be needed at the final selection from (9): see e.g., Andrews (1991).

3.4 Selecting forecasting models

An algorithm for $n \gg T$ allows an alternative approach to ‘factor forecasting’ methods, where an excess of variables is virtually bound to occur. The analysis in Hendry and Clements (2001) suggests that existing proposals such as Stock and Watson (1998) have potential drawbacks should any individual

variables matter for forecasting. The above selection procedure could be automated, although 6 sub-blocks or more might be computationally demanding. The risk of an excess of ‘spurious’ variables is less relevant in a forecasting context, since the factors anyway attribute some weight to most variables. An obvious additional variant in this setting is to include the first few factors as regressors in all initial models, thereby viewing the approach as one of detecting additional salient effects, as well as possibly replacing the general factors by a subset of ‘relevant’ regressors.

3.5 Selecting non-linear models

A relatively common approach in a non-linear setting (see Granger and Teräsvirta, 1993) is to fit non-linear models beginning from a previously selected linear representation. Such an approach is analogous to simple-to-general in two respects. First, moves between studies are almost bound to be simple to general, which has poor properties—and may be why empirical advances are so difficult. Secondly, however, any extension of a model should commence from a more general exemplar than the best selected earlier representative, otherwise in-built restrictions can preclude finding the appropriate generalization.

Instead, we propose commencing with a general polynomial approximation to the non-linearity (which needs to be identified, but could be cubic, or even rational ...). Add in the proposed logistic, squashing or desired functions one at a time, and test if they explain the non-linear components of the approximation. This avoids the lack of identification under the null, and also directly tests that the postulated functions are the correct ones. A mathematical outline takes the following form.

Let the DGP be:

$$y_t = f(z_{1,t} \dots z_{k,t}) + \epsilon_t \quad (12)$$

where $\epsilon_t \sim \text{IN}[0, \sigma_\epsilon^2]$. The GUM takes the form, under separability for some transformations $g_i(\cdot)$ (which includes classes like polynomials):

$$y_t = \sum_{i=1}^k \sum_{j=1}^{n_i} \beta_{i,j} h_{i,j}(z_{i,t}) + v_t \quad (13)$$

where all the basic relevant variables ($z_{1,t} \dots z_{k,t}$) are assumed to be known, but the most appropriate parsimonious functional forms are not. We only consider the case where the number of variables in the set $\{h_j(z_{i,t})\}$ is sufficiently less than the number of observations that the procedure of the earlier section is otiose; the generalization is otherwise clear, but probably tedious. In the case of polynomials, for example:

$$\sum_{j=1}^{n_i} \beta_{i,j} h_{i,j}(z_{i,t}) = \sum_{j=1}^{n_i} \beta_{i,j} z_{i,t}^j. \quad (14)$$

Naturally, the GUM must be tested for congruence before proceeding. This will ensure that inferences on possible reductions are validly undertaken: a mis-specified GUM would deliver inconsistent estimates and inappropriate standard errors, so the validity of reductions could not be judged. This remains a major advantage of commencing from a congruent GUM relative to other approaches.

However, while congruent, (13) may be highly over-parameterized, with a parsimonious approximation (e.g., by STAR functions) believed feasible of the form:

$$y_t = \sum_{i=1}^k \beta_{i,0} z_{i,t} + \sum_{i=1}^s \alpha_i g_i(z_{i,t}) + v_t. \quad (15)$$

The linear component has been explicitly delineated as the remaining terms are intended to capture the non-linearities evident in (13).

The next step is to test for a reduction of (13) to a linear representation:

$$y_t = \sum_{i=1}^k \beta_{i,0} z_{i,t} + v_t. \quad (16)$$

If that reduction is accepted, then (16) is the final model. If that reduction is rejected, however, non-linearity is established, so construct one of the functions $g_i(z_{i,t})$ corresponding to a non-linear component, and add it to the model in (13) as a ‘fixed regressor’ (i.e., one which *PcGets* must retain), to check if the optimized outcome allows the elimination of more than one $h_{i,j}(z_{i,t})$, so a reduction is indeed achieved. If the remaining terms are not linear, continue adding $g_i(z_{i,t})$ functions till (15) is obtained, or that plus additional non-linear terms.

Several potential issues arise. First, the $g_i(z_{i,t})$ may be perfectly collinear with the functions already in the model. Then, one of those could be eliminated (without loss), and a potential reduction investigated for the remaining $h_{i,j}(z_{i,t})$. Secondly, the parameters in $g_i(z_{i,t})$ may be imprecisely estimated until reductions are achieved. That spread of information will not matter, as *PcGets* will seek to eliminate only the redundant $h_{i,j}(z_{i,t})$, not the $g_i(z_{i,t})$ – if the parsimonious specification is correct, the final model (15) will be as well estimated as feasible. Thirdly, there may be no feasible reduction. This could occur for two entirely different, but easily distinguished, reasons. In the former, it happens because the $g_i(z_{i,t})$ are not the appropriate approximation, in which case the associated α_i estimate will be insignificant, so that avenue should be dropped as unproductive. In the latter, the associated α_i estimate is also significant in which case (13) was not sufficiently general as an initial GUM. In either case, crucial additional information is learned about the nature of the non-linearity that other methods seem unlikely to capture.

Experience to date with RETINA (see Perez-Amaral *et al.*, 2003, 2004) suggests that a very stringent initial significance level is desirable in deciding on the presence or absence of non-linearity, partly because so many non-linear terms may need to be tested. RETINA achieves that by using three sub-sample splits, but the evidence in Lynch and Vital-Ahuja (1998) and Hendry and Krolzig (2004b) suggests that is not an optimal use of the available information, albeit that the efficiency loss is not large. Equally, different significance levels may be useful for the non-linear and linear components, since avoiding overfitting for the former can produce low power for the latter: see Castle (2004), who also compares RETINA with *PcGets* for both linear and non-linear models using cross-section and time-series data sets, as well as by Monte Carlo.

4 Selecting simultaneous equations models

The properties of *Gets* derive from the theory of reduction (see Hendry, 1995), so our approach is to embed the selection of linear simultaneous systems in that theory, conditional on a prior division into endogenous and non-modelled variables (all of which might be lagged endogenous). First, the linear conditional statistical system (also called the ‘reduced form’) is formulated, noting that such a system is always identified. That system is then tested for congruence: once the initial system is congruent, all later selections can be chosen to be so as well. If congruence is accepted for the unrestricted representation, a parsimonious version of the system is then selected, ensuring that congruence is maintained, such that all right-hand side (regressor) variables are significant at the desired level in their associated equations: this step is to avoid later ‘spurious identification’ by excluding what are actually irrelevant regressors from some equations based on their (insignificant) presence in other equations. Then postulated endogenous variables are added *seriatim* to each equation in turn, matched by eliminating one

regressor, to see what further reductions can be achieved in the regressor set, contingent on maintaining identification. There are two main possibilities in any equation for further reductions in the regressor set:

- (1) no further reductions are found;
- (2) one or more further reduction are feasible.

The first case entails the existence of an equivalent equation to that already specified, so the choice reverts to the earlier system equation. The second case entails over-identification of the proposed equation, which restriction is then testable, since the eliminated right-hand side variable must be a determinant of the added left-hand side one, and is already significant, thereby identifying the endogenous effect.

Each equation is considered in turn in this instrumental-variables approach. Throughout, the rank condition (as implemented in *PcGive*: see Hendry *et al.*, 1988) is imposed as a constraint, so the ‘same equation’ is not included twice, and the current ‘partial structure’ is always fully identified at every step. Since the initial system is identified, all later selections must be as well. Here we use ‘structure’ (in quotation marks) to denote an equation with more than one endogenous variable, without any connotations that it really is structural (i.e., invariant to extensions of the information set for new variables, over time, and across regimes).

Weak instruments show up as a poorly determined initial system, or requiring a loose significance level for instruments to be retained: see section 4.5. That states, but does not resolve, the problem which lies in available information, not the performance of any selection algorithm or modelling approach. The choice of instruments can be made by *PcGets* (see Hendry and Krolzig, 2001, related to the approach in Hall, Rudebusch and Wilcox, 1996) both to determine their relevance for each endogenous variable, and to test for instrument mis-specification as part of the congruence check. Finally, while our approach is so far only worked out for zero restrictions on linear systems using instrumental variables, generalizations to other forms of restriction, other estimators such as maximum likelihood or quantile regression, and to non-linear equations *inter alia* seem feasible in principle.

Section 4.1 describes the notation for the system and its ‘structural’ representation; section 4.2 specifies the local data generation process, and section 4.3 considers its identification; then sections 4.4 and 4.5 respectively consider selecting the instrumental variables and the problem of weak instruments, allowing the formulation for jointly selecting the endogenous variables and instruments in each equation in section 4.6. A 2-equation example in sections 4.7 and 4.8 illustrates how the approach works both theoretically and on an artificial DGP.

4.1 Formulation

Consider a general unrestricted system of m endogenous variables \mathbf{y}_t where there are n candidate regressor variables $\mathbf{z}'_t = (z_{1,t} \dots z_{n,t})$ over a sample $t = 1, \dots, T > (n + m)$:

$$y_{i,t} = \sum_{j=1}^n \psi_{i,j} z_{j,t} + v_{i,t} \quad \text{where } v_{i,t} \sim \text{IN}[0, \omega_{ii}]. \quad (17)$$

Lagged variables may be included as $z_{i,t}$, but any contemporaneous regressor variables need to be weakly exogenous (see Koopmans, 1950b, and Engle, Hendry and Richard, 1983). An important special case of (17) is a vector autoregression (VAR) where the $\{z_{j,t}\}$ are all lagged $\{y_{i,t}\}$. Congruence requires the stated assumptions on the $\{v_{i,t}\}$, the constancy of the $\{\psi_{i,j}\}$, and the weak exogeneity of the $\{z_{j,t}\}$ for the parameters of interest, which for the $\{\psi_{i,j}\}$ (say) necessitates that $E[z_{j,t} v_{i,t}] = 0 \forall i, j$. No assumption is needed as to the relevance of the candidate regressor variables \mathbf{z}'_t in the system overall, or any equations in it. The system is presumed to have been reduced to a non-integrated (I(0))

representation so that conventional critical values can be used: see e.g., Omtzig (2002) and Kurcewicz and Mycielski (2003) for possible algorithms centered on the Johansen (1988) estimator. However, an adaptation of the approach described below is also applicable to identifying cointegration vectors.

Writing (17) as a unrestricted conditional system delivers:

$$\mathbf{y}_t = \mathbf{\Psi} \mathbf{z}_t + \mathbf{v}_t \text{ where } \mathbf{v}_t \sim \text{IN}_m [\mathbf{0}, \mathbf{\Omega}_v]. \quad (18)$$

The system in (18) is congruent, so matches the data evidence in all relevant respects. Letting $E[\cdot]$ denote an expectation, then $E[\mathbf{z}_t \mathbf{v}_t'] = \mathbf{0}$; and if $\mathbf{Y}' = (\mathbf{y}_1 \dots \mathbf{y}_T)$ and $\mathbf{Z}' = (\mathbf{z}_1 \dots \mathbf{z}_T)$:

$$\widehat{\mathbf{\Psi}} = \mathbf{Y}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \quad (19)$$

which satisfies $E[\widehat{\mathbf{\Psi}}] = \mathbf{\Psi}$ for fixed regressors, and is consistent otherwise. Also, when $V[\cdot]$ denotes a variance:

$$V[\widehat{\mathbf{\Psi}}] \simeq \mathbf{\Omega}_v \otimes (\mathbf{Z}' \mathbf{Z})^{-1} \quad (20)$$

where $\mathbf{A} \otimes \mathbf{B} = \{a_{i,j} \mathbf{B}\}$, and for $\widehat{\mathbf{v}}_t = \mathbf{y}_t - \widehat{\mathbf{\Psi}} \mathbf{z}_t$:

$$\widehat{\mathbf{\Omega}}_v = \frac{1}{T-n} \sum_{t=1}^T \widehat{\mathbf{v}}_t \widehat{\mathbf{v}}_t', \quad (21)$$

which is a consistent estimator of $\mathbf{\Omega}_v$.

4.2 The local data generation process

The i^{th} equation in the DGP may in fact contain endogenous regressors, and so only involve $r_i \leq n$ variables:

$$y_{i,t} + \sum_{j \in \{r_i\}} \beta_{i,j} y_{j,t} = \sum_{j \in \{r_i\}} \gamma_{i,j} z_{j,t} + \epsilon_{i,t} \text{ where } \epsilon_{i,t} \sim \text{IN} [0, \sigma_{ii}^2]. \quad (22)$$

The DGP equations are assumed to be identified when the appropriate restrictions are known (see section 4.3), but the investigator need not possess that information. Moreover, provided it is congruent, (22) need only be the local DGP in the space of $(\mathbf{y}_t : \mathbf{z}_t)$, not necessarily the ‘true’ generating process: see Bontemps and Mizon (2003).

Written as a ‘structure’, (22) is:

$$\mathbf{B} \mathbf{y}_t = \mathbf{C} \mathbf{z}_t + \epsilon_t \text{ where } \epsilon_t \sim \text{IN}_m [\mathbf{0}, \mathbf{\Sigma}_\epsilon]. \quad (23)$$

The DGP system is then formally the same as (18) above, written as:

$$\mathbf{y}_t = \mathbf{\Pi} \mathbf{z}_t + \mathbf{w}_t \text{ where } \mathbf{w}_t \sim \text{IN}_m [\mathbf{0}, \mathbf{\Omega}_w] \quad (24)$$

with:

$$\mathbf{B} \mathbf{\Pi} - \mathbf{C} = \mathbf{0}, \quad (25)$$

which has a unique solution from $\mathbf{\Pi}$ for \mathbf{B} , \mathbf{C} when fully identified, an issue addressed in the next section. Thus, (24) may entail restrictions on the $\mathbf{\Pi}$ matrix, whereas there are none imposed on $\mathbf{\Psi}$ in (18) as yet. However, given that the same set \mathbf{z}_t is used in both (18) and (24), $\widehat{\mathbf{\Psi}}$ is a consistent estimator of $\mathbf{\Pi}$.

4.3 Identification in the DGP

Identification, in the sense of uniqueness, of \mathbf{B} , \mathbf{C} from $\mathbf{\Pi}$ in DGPs like (23) has been extensively explored: see e.g. Koopmans (1949), Koopmans and Reiersøl (1950), Fisher (1966) and Rothenberg (1973) *inter alia*. Whether or not \mathbf{B} , \mathbf{C} can be recovered uniquely from $\mathbf{\Pi}$ is an intrinsic property of the DGP, and is not an aspect which can be altered by search. To ensure that no equation can be obtained as a linear combination of other equations requires the exclusion of some variables, and the inclusion of others in every equation. The order condition for identification simply concerns the number of unrestricted coefficients in \mathbf{B} , \mathbf{C} . Since $\mathbf{\Pi}$ is $m \times n$, no more than n unrestricted coefficients can enter any equation, easily checked by just counting the number of unrestricted elements in any postulated candidate for \mathbf{B} , \mathbf{C} . The order condition therefore just ensures a sufficient number of equations to solve (25). The well known rank condition checks that these equations are linearly independent, and so determines the extent to which each equation is or is not identified. In the literature, identification is a synonym for uniqueness, although extant usage also entails connotations of ‘interpretable in the light of economic theory’ and ‘corresponding to reality’ (as in ‘identify the demand curve’). We return to those additional requirements below.

We consider in turn the three main possibilities of lack of identification, just identification, and over identification: subsets could be (un)identified when other elements are (not), in which case the following comments apply to the appropriate set.

If no elements in \mathbf{B} , \mathbf{C} are identified by the DGP rank condition, then (24) is the least restricted, yet identified, representation. Of course, any just-identified representation with a form like (23) will also be minimal, so there is an equivalence class of such specifications with equal likelihood (see e.g., Rothenberg, 1971). In that case, there are no usable restrictions on the $\mathbf{\Pi}$ matrix, although reductions that eliminate regressor variables that are irrelevant to the entire system can still occur.

Since (24) is just identified even when there are no restrictions on $\mathbf{\Pi}$, commencing the statistical analysis with (24) ensures an identified initial representation. In section 4.6, we show that our approach imposes the rank condition in (25) throughout as a constraint, so search paths that violate (25) are discarded. A hidden possible cost is that the ‘most minimal’ representation may not be located, if the rank condition ‘barrier’ has to be temporarily relaxed to enter an appropriate part of parameter space, but we have not yet found examples where that might occur. As it explores all feasible paths, the algorithm should find and report all members of the equivalence class of just-identified representations, but obviously cannot select between these. The investigator can then choose that which is best ‘interpretable in the light of economic theory’.

If \mathbf{B} , \mathbf{C} are over identified by the DGP rank condition, then (24) is a unique representation for the given restrictions. However, as shown in Hendry, Lu and Mizon (2001), there may exist different sets of restrictions embodied in matrices \mathbf{B}^* , \mathbf{C}^* which are not linear transforms of \mathbf{B} , \mathbf{C} (precluded by their identifiability), but under which (23) is equally over identified. Thus, again an equivalence class of such specifications with equal likelihood can result: a given degree of over identification by itself does not ensure a unique model even when there is a unique DGP. One advantage of a search procedure is to highlight such cases, or indeed cases of near equivalence, whereas the existing approach via *a priori* specification ensures they cannot be located.

In practice, the rank condition for any equation can only be determined on a probability-one basis (as in *PcGive*, checked by setting each non-zero coefficient to unity plus a uniform random number: see Hendry *et al.*, 1988, and Doornik and Hendry, 2001). However, once any ‘structure’ has been selected to characterize the system, and all retained coefficients are significant, the rank condition is definitively satisfied. Moreover, the validity of the over-identified restrictions can be checked through parsimonious

encompassing of the system by the ‘structure’: when \mathcal{L} is the log-likelihood of the system, and \mathcal{L}_0 that of the ‘structural’ form, the test is $2(\mathcal{L} - \mathcal{L}_0) \sim \chi^2(s)$ for s restrictions (see Koopmans, 1950a, and Hendry and Mizon, 1993).

Other forms of restriction could identify ‘structural’ parameters that do not satisfy the rank condition, such as a diagonal error covariance matrix, or cross-equation links. In principle, these could be accommodated in an algorithm, but if present and ignored would probably lead to selecting a system rather than a model thereof. If reliable *a priori* restrictions are available, these could be easily incorporated into the selection.

4.4 Selecting instrumental variables

The choice of instruments for each endogenous variable can be made by *PcGets*, treating each equation as a valid conditional specification as in Hendry and Krolzig (2001) (also see the approach in Hall *et al.*, 1996). An efficient algorithm would take account of the system nature of the task, reflected in Ω_v , and include tests for cross-equation significance as well as within. Applied to several equations treated separately there is a loss of efficiency, but no issue of principle is raised, merely the lack at present of an operational program to implement such tests. To date, the properties of *PcGets* have primarily been established for single equations (see Hendry and Krolzig, 2003, for the latest overview), although there are also studies of its behaviour in VARs (see Brüggemann, Krolzig and Lütkepohl, 2002, and Krolzig, 2001, 2003b, 2003a), which show respectable performance both absolutely and relative to other selection devices.

A natural first step is to test for the relevance of the instruments as a group, namely a joint test of $H_0: \Psi = \mathbf{0}$ to ensure the significance of the system explanation. If $H_0: \Psi = \mathbf{0}$ cannot be rejected at a stringent significance level, then there is little to explain, irrespective of the form of model as a system or ‘structure’. This could be followed by tests of $H_{0,i}: \psi_i = \{\psi_{i,j}, j = 1, \dots, n\} = 0$ for $i = 1, \dots, m$ at much looser significance levels to ensure an adequate number of instruments exists in each equation.

A test for instrument mis-specification is also needed as part of the congruence check, to ensure that (e.g.) endogenous variables are not used as instruments. In the single-equation context, once equations are over identified, we currently use that based on Sargan (1964).

Once the instruments have been selected, a restricted system exists, which we write as:

$$y_{i,t} = \sum_{j=1}^{n_i} \phi_{i,j} z_{j,t} + u_{i,t} \quad \text{where } u_{i,t} \underset{\tilde{e}}{\sim} \text{IN}[0, \lambda_{ii}] \quad (26)$$

and $\underset{\tilde{e}}{\sim}$ denotes ‘is empirically distributed as’. All the estimated $\phi_{i,j}$ have been selected to be significant at the pre-assigned level for the problem at hand, or are needed to maintain congruence (which we ignore in the following). The number of restrictions imposed by simplification of insignificant regressors at this stage is:

$$N = \sum_{i=1}^m (n^* - n_i),$$

where $n^* \leq n$ regressors remain somewhere in the system, and $N > m$ is needed for over identification. Thus, $\Phi = \{\phi_{i,j}\}$ is a restricted matrix, and any simultaneity reductions must be consistent with its restrictions. Campos *et al.* (2003) establish the consistency of the selection algorithm in *PcGets* for a single conditional equation, and similar arguments seem to apply here: the critical values of the selection tests diverge at a suitable rate, as do the non-centralities of statistics for non-zero parameters. Thus, null effects are not retained, and non-null always found, so that $\hat{\Phi}$ will converge on Π as $T \rightarrow \infty$. The speed with which that will occur depends on m, n and both the choice of the significance level rule and

the non-centralities of the test statistics, and is open to Monte Carlo study independently of modelling simultaneity.

A potential drawback of first selecting the instruments as regressors for (26) is that the resulting estimates of the $\phi_{i,j}$ will be biased by selection effects, and some IVs relevant in the DGP will be omitted due to chance insignificance in the given data set. The former can be approximately corrected (see Hendry and Krolzig, 2004a), though the latter cannot, and would anyway affect simultaneous equations estimation where a priori information was available. In addition, some irrelevant regressors will be significant by chance, but that too would affect estimation without selection, and indeed the outcome of any test of over identification for the given data set. It may, therefore, be worth selecting a model for the initial system before instrument selection as well to check that such effects have not distorted the outcome.

4.5 Weak instruments

A ‘structure’ could be identified in principle, yet the available instruments may be so weak that for practical purposes, the uncertainty is close to unbounded. A large literature exists on this problem: see e.g., Staiger and Stock (1997), Zivot, Startz and Nelson (1998), Wang and Zivot (1998), Phillips (1989), Stock and Wright (2000), and Mavroeidis (2003). Again, the issue is intrinsic to the available information, and is not a problem created by model selection procedures.

Since the first step is to test for the relevance of the instruments as discussed in section 4.4, failure to reject $H_0: \Psi = \mathbf{0}$ would warn of seriously weak instruments, as would failure to reject any $\psi_i = 0$. Zivot *et al.* (1998) suggest that such testing of the significance of the ‘first-stage regression’ is better performed using Lagrange multiplier or Likelihood ratio than Wald statistics. Conversely, failing to eliminate irrelevant instruments will generate spurious identification, and could mislead the algorithm. Thus, there is a narrow path between these two difficulties, one that would in fact also be present when theory-based identification depended on the significance of an irrelevant regressor.

4.6 Jointly selecting endogenous variables and instruments

The rank condition in (25) is maintained throughout as a constraint on the selected model, in that only reductions consistent with (25) are explored. This implementation follows our earlier approach to enforcing sign restrictions in searches, where a test on the general model establishes their viability or otherwise, and thereafter, violations are treated as non-congruent so such paths searches are terminated. In fact, that approach is equivalent to how congruence is ensured with diagnostic tests. Initially, $\Pi = \Psi = \mathbf{C}$ so there are no constraints, and the system in (18) is easily tested for congruence either by system and/or single-equation diagnostics (see e.g., Doornik and Hendry, 2001). If congruence is accepted, for the set of tests deemed relevant by the investigator, reduction paths are explored to deliver a parsimonious system, using the well-established approach already operational in *PcGets*, and with the properties described in Hendry and Krolzig (2004a).

The resulting simplified selected equations are as in (26) with coefficients ϕ_i for $i = 1, \dots, m$. If precisely the same regressors occur in some equations, say the l^{th} and k^{th} , then a modified approach is required of testing for proportionality between ϕ_l and ϕ_k as the only feasible reduction. Otherwise, if the regressor sets differ, then those excluded from the l^{th} equation and included in the k^{th} can act as instruments for $y_{k,t}$ being added to the equation for $y_{l,t}$.

Several possibilities then present themselves. If the rank condition is violated, the proposed step is rejected. If the rank condition is accepted then we could have:

- a) $y_{k,t}$ is insignificant in the equation for $y_{l,t}$ both initially and after attempted simplification;

- b) $y_{k,t}$ is significant in the equation for $y_{l,t}$, but no reduction ensues after attempted simplification;
 c) $y_{k,t}$ is significant in the equation for $y_{l,t}$ and a reduction allows other regressors to be excluded.

In case a), the resulting equation is not of great interest, and the parsimonious system equation seems preferable.

Case b) is the most complicated as the ‘structural equation’ has more variables than its ‘reduced form’ suggesting some form of mis-specification.

Case c) is clear cut: the resulting equation is at least as parsimonious as the previous system representation, congruent, and identified, so would be selected. Of course, other identified representations might also be found; and in any case, whichever ones are found remain subject to needing ‘acceptable’ interpretations.

4.7 Two-equation example

Consider a DGP which is a conventional ‘structural’ supply-demand illustration with four instruments, where the first instrument is common to both equations:

$$\begin{aligned} y_{1,t} + \beta_{1,2}y_{2,t} &= \gamma_{1,1}z_{1,t} + \gamma_{1,2}z_{2,t} + \epsilon_{1,t} \\ y_{2,t} + \beta_{2,1}y_{1,t} &= \gamma_{2,1}z_{1,t} + \gamma_{2,3}z_{3,t} + \gamma_{2,4}z_{4,t} + \epsilon_{2,t} \end{aligned} \quad (27)$$

As in (23), $\epsilon_t \sim \text{IN}_2[\mathbf{0}, \Sigma_\epsilon]$, and all variables are $I(0)$ with \mathbf{z}_t strongly exogenous for $\{\beta_{i,j} : \gamma_{i,j}\}$. Then, given that $\beta_{1,2}\beta_{2,1} \neq 1$ and $\gamma_{1,3} = \gamma_{1,4} = \gamma_{2,2} = 0$, the two equations in (27) are uniquely identified by $\gamma_{1,2} \neq 0$, and $\gamma_{2,3} \neq 0$ or $\gamma_{2,4} \neq 0$ respectively. Their interpretation as supply-demand, or *vice versa*, depends on theory considerations; and their correspondence to reality is ensured here by their being the DGP. For interpretation, we take the first equation as demand, and the second as supply, with $y_{1,t}$ and $y_{2,t}$ denoting quantity and price respectively.

If both $\gamma_{2,3} = \gamma_{2,4} = 0$ then the first equation is not unique, as is the second if $\gamma_{1,2} = 0$ unless one of $\gamma_{1,1}$ or $\gamma_{1,2}$ is zero when the other is non-zero. Since the issue here is the feasibility of an automatic system to locate the ‘structural’ DGP representation without prior information, we exclude these cases from further consideration.

The restricted system representation is:

$$\begin{aligned} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} &= d \begin{pmatrix} \gamma_{1,1} - \beta_{1,2}\gamma_{2,1} & \gamma_{1,2} & -\beta_{1,2}\gamma_{2,3} & -\beta_{1,2}\gamma_{2,4} \\ \gamma_{2,1} - \beta_{2,1}\gamma_{1,1} & -\beta_{2,1}\gamma_{1,2} & \gamma_{2,3} & \gamma_{2,4} \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \\ z_{4,t} \end{pmatrix} \\ &+ d \begin{pmatrix} \epsilon_{1,t} - \beta_{1,2}\epsilon_{2,t} \\ \epsilon_{2,t} - \beta_{2,1}\epsilon_{1,t} \end{pmatrix} \end{aligned} \quad (28)$$

where:

$$d = \frac{1}{1 - \beta_{1,2}\beta_{2,1}}.$$

The unrestricted projection on the unmodelled variables is:

$$\begin{aligned} y_{1,t} &= \phi_{1,1}z_{1,t} + \phi_{1,2}z_{2,t} + \phi_{1,3}z_{3,t} + \phi_{1,4}z_{4,t} + w_{1,t} \\ y_{2,t} &= \phi_{2,1}z_{1,t} + \phi_{2,2}z_{2,t} + \phi_{2,3}z_{3,t} + \phi_{2,4}z_{4,t} + w_{2,t}. \end{aligned} \quad (29)$$

Here (29) is uniquely identified for all parameter values.

For simplicity, first consider the special case where $\beta_{2,1} = 0$, so the supply equation does not have contemporaneous ‘explanatory’ variables. Then from (28), $\phi_{2,2} = 0$, so $z_{2,t}$ will usually not enter the second equation, and the initial simplification will detect that outcome $\alpha\%$ of the time for a significance level α . Consider adding $y_{1,t}$ to the $y_{2,t}$ equation, instrumenting it by $z_{2,t}$, which must enter the first equation significantly or it would have been eliminated from the system. Since $\beta_{2,1} = 0$, $y_{1,t}$ is irrelevant and again the algorithm will detect that outcome $\alpha\%$ (subject to the adequacy of the test statistics at the relevant sample size). Thus, the second equation is confirmed as in (29) with $\phi_{2,2} = 0$.

If simplifications also occur in the first equation, then they provide immediate instruments for proceeding. The first reduction step on adding $y_{2,t}$ to the $y_{1,t}$ equation is to instrument it by the non-retained regressors in the first equation that are still retained in the $y_{2,t}$ equation. If a further reduction results on doing so, the postulated ‘structure’ is now over identified: *PcGets* is programmed in IV mode to automatically add eliminated regressors to the instrument set and compute the Sargan test of instrument validity (see Sargan, 1964), as well as check congruence. The search can continue from there till all remaining paths have been explored. If no further reductions ensue on search, store the model located at that stage, revert to the original system, and drop another regressor, again continuing till all paths have been explored.

When no regressor variables are excluded from the first equation in (29), it may be thought that the possibility of a simultaneous representation cannot be investigated as there are no available instruments to commence the search. However, proceed by dropping (say) the ‘least significant’ regressor $z_{j,t}$ in the first equation (other than $z_{2,t}$) on the possible hypothesis that its lower significance derives from entering indirectly (i.e., because of $\beta_{1,2} \neq 0$ rather than $\gamma_{1,4} \neq 0$), and instrument $y_{2,t}$ by that variable. Now the analysis is like that of the previous paragraph; additional simplifications lead to an over-identified representation, whereas no simplifications entail a just-identified equivalent equation. Section 4.8 provides a numerical illustration.

In practice, in every setting, all paths need to be explored since a regressor may be insignificant in an equation of the system precisely because it is a composite of the form $\gamma_{1,1} - \beta_{1,2}\gamma_{2,1}$ where cancellation occurs, yet be significant in the model of the system. The converse is also feasible.

Even when $\beta_{2,1} \neq 0$, a similar algorithm can be delineated although no $\phi_{i,j}$ need be zero. Indeed, all regressor would be potential instruments, so the only impact is that more paths need to be explored. As before, first drop the ‘least significant’ regressor $z_{j,t}$ in the first equation on the possible hypothesis that its low significance derives from entering indirectly (e.g., because of $\beta_{1,2} \neq 0$ rather than $\gamma_{1,4} \neq 0$), and instrument $y_{2,t}$ by that variable. Every path in which each $z_{j,t}$ is dropped in turn and $y_{2,t}$ entered in the $y_{1,t}$ equation can be followed, and the best outcome selected for the first equation in (29). If no further reductions ensue on search, revert to the original system and drop another regressor, continuing till all paths have been explored. Select the minimal dominant congruent representation from the resulting set of terminal models. Then continue with the second equation, conditional on not deleting regressors that are not retained in the first equation’s ‘structural’ representation. As a check on the models located, the roles of the two equations could be reversed, although a large number of possibilities rapidly results as m rises.

Many of the above steps involve single instrument estimation, and hence there are no finite-sample moments for the associated estimators, which could complicate inference, and especially Monte Carlo simulation studies (but see Sargan, 1982, and Hendry, 1990). Nevertheless, by treating simultaneity as a potential reduction on the parameter space, the proposed algorithm appears to be feasible and capable in principle of correctly determining the ‘structural’ representation in this 2-equation system even when the prior restrictions are unknown, subject to the usual *caveats* about the sizes and powers of the tests used, to which topic we now turn.

4.7.1 Size

If tests are conducted at a significance level of α , so use a critical value of c_α , then there is an $\alpha\%$ probability of retaining $z_{2,t}$ in the $y_{2,t}$ equation despite its actual irrelevance. Whether or not such an event occurs does not affect the outcome in the special case that $\beta_{2,1} = 0$, but more generally, Type I errors could limit the performance of all methods of modelling, unless the DGP is known.

Once $y_{2,t}$ is included in the $y_{1,t}$ equation, there is an $\alpha\%$ probability of retaining either of $z_{3,t}$ and $z_{4,t}$ in the $y_{1,t}$ equation. The extent of the ‘simultaneity’ reduction depends on such adventitious significance. Two irrelevant variables are retained with probability $(1 - \alpha)^2$ (under orthogonality) if other instruments exist. Controlling size would seem potentially more important in simultaneous systems than in conditional models, but the chance significance just noted would affect all approaches. For example, if the ‘structure’ was derived *a priori* and the over-identifying restrictions tested for their validity, then rejection would result when adventitious significance of irrelevant regressors occurs. Thus, the problem simply appears in a different guise.

4.7.2 Power

Conversely, let $\delta_\alpha^{(i)}$ denote the power for correctly rejecting that the i^{th} parameter is non-zero, for the ordering $(\phi_{1,1}, \phi_{1,2}, \phi_{1,3}, \phi_{1,4}, \phi_{2,1}, \phi_{2,2}, \phi_{2,3}, \phi_{2,4})$ so $(1 - \delta_\alpha^{(i)})$ is the type II error for a significance level of α . A failure to reject $\phi_{1,2} = 0$ would reduce the dimension of the system by the complete removal of $z_{2,t}$ in the first case discussed in section 4.7 when $\phi_{2,2} = 0$. That could matter greatly when $\beta_{2,1} \neq 0$, but not if $\beta_{2,1} = 0$.

Failing to reject both $\phi_{2,3} = 0$ and $\phi_{2,4} = 0$ would lead to an unidentifiable ‘structure’ for the first equation, and no reductions could be effected. Rejecting either one would allow a ‘just-identified’ equivalent representation with $y_{2,t}$ in the $y_{1,t}$ equation, and which ever of $z_{3,t}$ or $z_{4,t}$ mattered in the $y_{2,t}$ equation now eliminated from the first equation. Rejecting both should allow an over-identified equation to be located. Similar considerations apply to the second equation irrespective of the value of $\beta_{2,1}$.

4.7.3 Diagonal error covariance matrix

An example that could be problematic is when only one equation appears to be identified by the rank condition, but the error covariance matrix is in fact diagonal, as in:

$$\begin{aligned} y_{1,t} + \beta_{1,2}y_{2,t} &= \gamma_{1,1}z_{1,t} + \gamma_{1,2}z_{2,t} + \epsilon_{1,t} \\ y_{2,t} + \beta_{2,1}y_{1,t} &= \gamma_{2,1}z_{1,t} + \gamma_{2,2}z_{2,t} + \gamma_{2,3}z_{3,t} + \epsilon_{2,t} \end{aligned} \quad (30)$$

where Σ_ϵ is diagonal, so $E[\epsilon_{1,t}\epsilon_{2,t}] = 0$. The restricted system representation is:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = d \begin{pmatrix} \gamma_{1,1} - \beta_{1,2}\gamma_{2,1} & \gamma_{1,2} - \beta_{1,2}\gamma_{2,2} & -\beta_{1,2}\gamma_{2,3} \\ \gamma_{2,1} - \beta_{2,1}\gamma_{1,1} & \gamma_{2,2} - \beta_{2,1}\gamma_{1,2} & \gamma_{2,3} \end{pmatrix} \begin{pmatrix} z_{1,t} \\ z_{2,t} \\ z_{3,t} \end{pmatrix} + d \begin{pmatrix} \epsilon_{1,t} - \beta_{1,2}\epsilon_{2,t} \\ \epsilon_{2,t} - \beta_{2,1}\epsilon_{1,t} \end{pmatrix} \quad (31)$$

where $d = 1/(1 - \beta_{1,2}\beta_{2,1})$. Adding $y_{2,t}$ to the first equation will be acceptable when $z_{3,t}$ is the eliminated variable, albeit creating an equivalent representation. However, adding $y_{1,t}$ to the second equation will not effect any reduction that satisfies the rank condition, even though the second equation is identifiable using $E[\epsilon_{1,t}\epsilon_{2,t}] = 0$, conditional on the first being as in (30). A system-based maximum likelihood search procedure could handle such cases.

4.8 Artificial data example

A one-sample computer-generated DGP mimicking (27) and (29) was investigated as ‘proof of concept’. The DGP (for $T = 100$ observations) was:

$$\begin{pmatrix} 1 & -0.5 \\ 0 & 1 \end{pmatrix} \mathbf{y}_t = \begin{pmatrix} 0.2 & 0.2 & 0 & 0 \\ 0.6 & 0 & 0.3 & 0.3 \end{pmatrix} \mathbf{z}_t + \boldsymbol{\epsilon}_t \quad (32)$$

$$\boldsymbol{\epsilon}_t \sim \text{IN}_2 \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.1 & 0 \\ 0 & 0.2 \end{pmatrix} \right] \quad (33)$$

$$\mathbf{z}_t \sim \text{IN}_4 [\mathbf{0}, \mathbf{I}] \quad (34)$$

The first system equation involves all four $z_{i,t}$ and was selected as such by *PcGets*; the second system equation was also selected correctly omitting $z_{2,t}$.

Adding $y_{2,t}$ to the first equation, and dropping its least significant regressor ($z_{3,t}$) to use as an instrument, checking that it was significant in the $y_{2,t}$ system equation, produced the DGP equation (32). Adding the eliminated $z_{4,t}$ as an instrument confirmed that outcome, with an insignificant Sargan test. Conversely, trying to add $y_{1,t}$ to the $y_{2,t}$ equation, selecting $z_{2,t}$ as the instrument led back to the system equation, which here is the ‘structural’ equation. Finally, incorrectly treating $y_{2,t}$ as an instrument with $z_{2,t}$ treated as endogenous, led to rejection on the Sargan test.

Based on the Monte Carlo experiments for a VAR when the DGP is a recursive SVAR (a special case of the problem considered here), cross-correlation of the errors was found to hamper the performance of single-equation reduction procedures such as *PcGets* in its current form: see Krolzig (2003b). In this numerical example, altering the DGP to allow the $\epsilon_{i,t}$ to be correlated does not in fact change the findings. This is not simply because of the recursive form of \mathbf{B} in (32), as also allowing $\beta_{2,1} = 0.5$ again did not alter the results for the first equation, but produced an ‘equivalent’ (i.e., ‘structural’ just identified) solution for the second, as would be expected. Nevertheless, a full simulation analysis is obviously required to be definitive. In addition, that would allow an investigation of any potential path dependence on the order of exploration when each equation is considered in turn.

5 Conclusion

We have considered three problems that previously seemed intractable, namely selecting from a perfectly collinear set of regressors; selecting a regression when there are more regressors than observations; and selecting a simultaneous equations representation when there are no *a priori* restrictions available to the investigator. In each case, a variant of the general-to-simple (*Gets*) approach suggested a feasible solution.

For the perfectly collinear set of regressors, the key notion was that of multi-path searches which explored all the combinations of variables to locate the undominated model that parsimoniously encompasses the other possible models. This is a novel application of the development in Hoover and Perez (1999) which started the recent upsurge of interest in computer automated model selection. The artificial data example suggested the idea can work, although the size and power remain to be established when all paths are forcibly explored.

To select a regression when there are more regressors than observations requires a block implementation of multi-path searches, as well as search procedures within tentative models. Again, an artificial data example suggested the idea could work, but again the size and power remain to be established, as do generalizations to embed factor forecasting and non-linear models.

The key conceptual notion in the third case is that simultaneous equations selection are a reduction, so the usual problem of model identification—which arises when commencing from ‘structural relations’—is circumvented. When the DGP is identifiable, and every representation satisfies the identification condition as a constraint, the final selection is identified, and no less parsimonious than the initial system. The logic of the selection algorithm outlined above seems to offer a feasible solution to jointly selecting instruments and identified relations when such exist. The illustration is very simple, but confirms that the idea can work. Moreover, when there was no simultaneous ‘structural’ equation, *PcGets* did not select one in the artificial data example. For a larger number of endogenous variables, the set of valid structural reductions of the system could be large, making for a cumbersome procedure. However, that is intrinsic to the dimensionality, not to the approach proposed here, which rather offers the computational power to tackle such difficulties. Equally, a maximum likelihood approach would seem essential if interest was focused on more than one equation, and conceptually our approach could be extended from instrumental variables to full information maximum likelihood or other members of the estimator generating equation (see Hendry, 1976). The framework should also apply to other sources of endogeneity, such as measurement errors.

References

- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817–858.
- Bontemps, C., and Mizon, G. E. (2003). Congruence and encompassing. in Stigum (2003), pp. 354–378.
- Brüggemann, R., Krolzig, H.-M., and Lütkepohl, H. (2002). Comparison of model selection procedures for VAR processes. Mimeo, Humboldt–University, Berlin.
- Campos, J., Ericsson, N. R., and Hendry, D. F. (2004). Editors’ introduction. In Campos, J., Ericsson, N. R., and Hendry, D. F. (eds.), *Readings on General-to-Specific Modeling*. Cheltenham: Edward Elgar. Forthcoming.
- Campos, J., Hendry, D. F., and Krolzig, H.-M. (2003). Consistent model selection by an automatic Gets approach. *Oxford Bulletin of Economics and Statistics*, **65**, 803–819.
- Castle, J. (2004). A comparison of automatic model selection procedures: PcGets and RETINA. Unpublished typescript, Economics Department, University of Oxford.
- Clements, M. P., and Mizon, G. E. (1991). Empirical analysis of macroeconomic time series: VAR and structural models. *European Economic Review*, **35**, 887–932.
- Doornik, J. A., and Hendry, D. F. (2001). *Modelling Dynamic Systems using PcGive 10: Volume II*. London: Timberlake Consultants Press.
- Engle, R. F., Hendry, D. F., and Richard, J.-F. (1983). Exogeneity. *Econometrica*, **51**, 277–304. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000; and in Ericsson, N. R. and Irons, J. S. (eds.) *Testing Exogeneity*, Oxford: Oxford University Press, 1994.
- Fisher, F. M. (1966). *The Identification Problem in Econometrics*. New York: McGraw Hill.
- Granger, C. W. J., and Hendry, D. F. (2004). A dialogue concerning a new instrument for econometric modeling. *Econometric Theory*, **20**, forthcoming.
- Granger, C. W. J., and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

- Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1996). Judging instrument relevance in instrumental variables estimation. *International Economic Review*, **37**, 283–298.
- Hall, S. G., Mizon, G. E., and Welfe, A. (2000). Modelling economies in transition. An introduction. *Economic Modelling*, **17**, 339–357.
- Hendry, D. F. (1976). The structure of simultaneous equations estimators. *Journal of Econometrics*, **4**, 51–88. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.
- Hendry, D. F. (1990). Using PC-NAIVE in teaching econometrics. *Oxford Bulletin of Economics and Statistics*, **53**, 199–223.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.
- Hendry, D. F., and Clements, M. P. (2001). Forecasting using factor models. Mimeo, Economics Department, University of Oxford.
- Hendry, D. F., Johansen, S., and Santos, C. (2004). Selecting a regression saturated by indicators. Unpublished paper, Economics Department, University of Oxford.
- Hendry, D. F., and Krolzig, H.-M. (2001). *Automatic Econometric Model Selection*. London: Timberlake Consultants Press.
- Hendry, D. F., and Krolzig, H.-M. (2003). New developments in automatic general-to-specific modelling. in Stigum (2003), pp. 379–419.
- Hendry, D. F., and Krolzig, H.-M. (2004a). The properties of automatic Gets modelling. Sargan lecture, Royal Economic Society Conference, Swansea.
- Hendry, D. F., and Krolzig, H.-M. (2004b). Sub-sample model selection procedures in general-to-specific modelling. In Becker, R., and Hurn, S. (eds.), *Contemporary Issues in Economics and Econometrics: Theory and Application*, pp. 53–75. Cheltenham: Edward Elgar.
- Hendry, D. F., and Krolzig, H. M. (2004c). We ran one regression. *Oxford Bulletin of Economics and Statistics*, –. Forthcoming.
- Hendry, D. F., Lu, M., and Mizon, G. E. (2001). Model identification and non-unique structure. Unpublished paper, Economics Department, Oxford University.
- Hendry, D. F., and Mizon, G. E. (1993). Evaluating dynamic econometric models by encompassing the VAR. In Phillips, P. C. B. (ed.), *Models, Methods and Applications of Econometrics*, pp. 272–300. Oxford: Basil Blackwell.
- Hendry, D. F., Neale, A. J., and Srba, F. (1988). Econometric analysis of small linear systems using Pc-Fiml. *Journal of Econometrics*, **38**, 203–226.
- Hoover, K. D., and Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, **2**, 167–191.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254. Reprinted in R.F. Engle and C.W.J. Granger (eds), *Long-Run Economic Relationships*, Oxford: Oxford University Press, 1991, 131–52.
- Koopmans, T. C. (1949). Identification problems in economic model construction. *Econometrica*, **17**, 125–144. Reprinted with minor revisions in Hood, W. C. and Koopmans, T. C. (eds.) (1953), *Studies in Econometric Method*. Cowles Commission Monograph 14, New York: John Wiley & Sons.
- Koopmans, T. C. (ed.) (1950a). *Statistical Inference in Dynamic Economic Models*. No. 10 in Cowles Commission Monograph. New York: John Wiley & Sons.

- Koopmans, T. C. (1950b). When is an equation system complete for statistical purposes?. in *Statistical Inference in Dynamic Economic Models* (1950a), Ch. 17.
- Koopmans, T. C., and Reiersøl, O. (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics*, **21**, 165–181.
- Krolzig, H.-M. (2001). General-to-specific reductions in vector autoregressive processes. In Friedmann, R., Knüppel, L., and Lütkepohl, H. (eds.), *Econometric Studies - A Festschrift in Honour of Joachim Frohn*, pp. 129–157. Münster: LIT Verlag.
- Krolzig, H.-M. (2003a). Computer-automated single-equation general-to-specific reduction procedures for vector autoregressions. In Neck, R. (ed.), *Modelling and Control of Economic Systems*, pp. 137–142. Oxford: Pergamon Press.
- Krolzig, H.-M. (2003b). General-to-specific model selection procedures for structural vector autoregressions. *Oxford Bulletin of Economics and Statistics*, **65**, 769–802.
- Krolzig, H.-M., and Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, **25**, 831–866.
- Kurcewicz, M., and Mycielski, J. (2003). A specification search algorithm for cointegrated systems. Discussion paper, Statistics Department, Warsaw University.
- Lynch, A. W., and Vital-Ahuja, T. (1998). Can subsample evidence alleviate the data-snooping problem? A comparison to the maximal R^2 cutoff test. Discussion paper, Stern Business School, New York University.
- Mavroeidis, S. (2003). Weak identification of forward-looking models in monetary economics. Manuscript, University of Amsterdam.
- Omtzig, P. (2002). Automatic identification and restriction of the cointegration space. Thesis chapter, Economics Department, Copenhagen University.
- Perez-Amaral, T., Gallo, G. M., and White, H. (2003). A flexible tool for model building: the relevant transformation of the inputs network approach (RETINA). *Oxford Bulletin of Economics and Statistics*, **65**, 821–838.
- Perez-Amaral, T., Gallo, G. M., and White, H. (2004). A comparison of complementary automatic modelling methods: RETINA and PcGets. *Econometric Theory*, **20**, forthcoming.
- Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory*, *5*(2), 181–240.
- Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, **39**, 577–592.
- Rothenberg, T. J. (1973). *Efficient Estimation with A Priori Information*. No. 23 in Cowles Foundation Monograph. New Haven: Yale University Press.
- Sargan, J. D. (1964). Wages and prices in the United Kingdom: A study in econometric methodology (with discussion). In Hart, P. E., Mills, G., and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*, Vol. 16 of *Colston Papers*, pp. 25–63. London: Butterworth Co. Reprinted as pp. 275–314 in Hendry D. F. and Wallis K. F. (eds.) (1984). *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell, and as pp. 124–169 in Sargan J. D. (1988), *Contributions to Econometrics*, Vol. 1, Cambridge: Cambridge University Press.
- Sargan, J. D. (1980). Some tests of dynamic specification for a single equation. *Econometrica*, **48**, 879–897. Reprinted as pp. 191–212 in Sargan J. D. (1988), *Contributions to Econometrics*, Vol. 1, Cambridge: Cambridge University Press.
- Sargan, J. D. (1982). On Monte Carlo estimates of moments that are infinite. In Basman, R. L., and Rhodes, G. F. (eds.), *Advances in Econometrics: A Research Annual*, Vol. 1, pp. 267–299.

Greenwich, Connecticut: Jai Press Inc.

- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- Staiger, D., and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, **65**, 557–586.
- Stigum, B. P. (ed.)(2003). *Econometrics and the Philosophy of Economics*. Princeton: Princeton University Press.
- Stock, J. H., and Watson, M. W. (1998). Diffusion indices. Working paper, 6702, NBER, Washington.
- Stock, J. H., and Wright, J. H. (2000). GMM with weak identification. *Econometrica*, **68**, 1055–1096.
- Wang, J., and Zivot, E. (1998). Inference on a structural parameter in instrumental variables regression with weak instruments. *Econometrica*, **66**, 1389–1404.
- White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.
- Zivot, E., Startz, R., and Nelson, C. R. (1998). Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review*, **39**, 1119–1144.

6 Appendix: The conditional model selection algorithm

PcGets has four basic stages in its approach to selecting a parsimonious undominated representation of an overly general initial model, denoted the general unrestricted model (GUM). The first concerns the estimation and testing of the GUM and the settings of the algorithm (1–4 below), the second is the pre-search process (5–6); the third is the multi-path search procedure (7–13); and the fourth is the post-search evaluation (14). The following description sketches the main steps involved: see Hendry and Krolzig (2001) for details.

- (1) Formulate the GUM based on theory, institutional knowledge, historical contingencies, measurement information, ensuring the GUM encompasses previous evidence, while seeking a relatively orthogonal parameterization;
- (2) select the set of m mis-specification tests (e.g., residual autocorrelation) and their forms (e.g., of r^{th} -order), and the desired information criterion (e.g., *SIC*);
- (3) set the significance levels of all selection tests (generically denoted α below) and mis-specification tests (generically denoted δ below) to ensure the desired null rejection frequencies, perhaps by selecting one of the pre-set Liberal or Conservative strategies;
- (4) estimate the GUM appropriately (least squares—OLS—and instrumental variables—IV—are presently available), and check by the mis-specification tests that the GUM captures the essential characteristics of the data (denoted congruence), perhaps with outlier adjustments;
- (5) undertake pre-search reductions at a loose significance level (these include lag-order pre-selection, F-tests on successively shorter lag groups, and cumulative F-tests based on t-tests ordered from the smallest up, and the largest down);
- (6) eliminate the resulting insignificant variables to reduce the search complexity, then estimate the new GUM as the baseline for the remaining stages;
- (7) multiple path reduction searches now commence from each feasible initial deletion (to avoid path-dependent selections);
- (8) the validity of each reduction is diagnostically checked to ensure the congruence of the final model;
- (9) if all reductions and diagnostic tests are acceptable, and all remaining variables are significant (or further reductions induce mis-specifications), that model becomes a *terminal* selection, and the next path search commences (i.e., back to 7);
- (10) when all paths have been explored and all distinct terminal models have been found, they are tested against their union to find an encompassing contender;
- (11) rejected models are removed, and the union of the ‘surviving’ terminal models becomes the GUM of a repeated multi-path search iteration;
- (12) the entire search process (i.e., from 7) continues till a unique choice of *final* model emerges, or the search converges to a set of mutually encompassing and undominated contenders;
- (13) in that last case, all the selected models are reported, and a unique final choice made by the pre-selected information criterion;
- (14) the significance of every variable in the final model is assessed in two over-lapping sub-samples to check the reliability of the selection.