Working Paper Series CENTER FOR FINANCIAL ECONOMETRICS

THE PROPERTIES OF AUTOMATIC *GETS* MODELLING

David F. Hendry
Hans-Martin Krolzig

The Properties of Automatic *Gets* Modelling

David F. Hendry and Hans-Martin Krolzig*

October 8, 2004

Abstract

After reviewing the simulation performance of general-to-specific automatic regression-model selection, as embodied in *PcGets*, we show how model selection can be non-distortionary: approximately unbiased 'selection estimates' are derived, with reported standard errors close to the sampling standard deviations of the estimated DGP parameters, and a near-unbiased goodness-of-fit measure. The handling of theory-based restrictions, non-stationarity, and problems posed by collinear data are considered. Finally, we consider how *PcGets* can handle three 'intractable' problems: more variables than observations in regression analysis; perfectly collinear regressors; and modelling simultaneous equations without *a priori* restrictions.

JEL Classification: C51, C22.

Keywords: Econometric methodology; model selection; general-to-specific; automatic modelling; encompassing; Monte Carlo experiments.

Model selection is an essential component of empirical research in all disciplines where *a priori* theory does not pre-define a complete and correct specification. Economics is surely such an empirical science, as macroeconomic processes are complicated, high-dimensional and non-stationary. Since any statistical test that leads to a decision involves selection, it is obvious that selection is ubiquitous in empirical economic research.

Unfortunately, the methodology of empirical modelling, and in particular of methods of model selection, are both subject to dispute. Even accusing a selection procedure of being 'data mining' can be sufficient to dismiss its conclusions as valueless. Yet a careful reading of the critical literature on that methodology reveals much assertion, but few analyses.

In part, that lacuna occurs because model selection theory poses considerable technical difficulties: all statistics for selecting models and evaluating their specifications have interdependent distributions, which are different under null and alternative, and altered by every modelling

*david.hendry@economics.ox.ac.uk

Sargan Lecture of the Royal Economic Society, 2004. We are indebted to Julia Campos, Dorian Owen and Jon Temple for helpful comments and suggestions. Financial support to the first author from the UK Economic and Social Research Council under a Professorial Research Fellowship, RES 051 270035, and grant RES 000 230539, is gratefully acknowledged.

decision. Fortunately, however, by enabling operational studies of selection strategies, recent advances in computer automation have allowed a fresh look at this old problem.

Two central notions are sampling variability and search. The former is manifest in the theory of statistics: different outcomes result from different samples. The costs of inference are well known to be falsely rejecting a true null hypothesis, and failing to reject an incorrect null—errors of type I and II respectively. Such costs are determined by the nature of the problem, in that they apply even when a complete and correct model is used initially, but it is not known to be the truth, so specification and mis-specification tests are applied. Thus, costs of inference are inevitable if tests have non-zero null-rejection frequencies and non-unit powers, even when commencing from the data generation process (DGP).

Costs of search are additional to these costs of inference, and arise when the initial model is more general than needed, perhaps in an attempt to characterize the data and avoid the converse costs of mis-specification. Search costs are under the control of an investigator, partly through the closeness of the initial model to the DGP, and partly through the efficiency of the search procedure. Despite initially disappointing results from computer simulations of some selection procedures (see e.g., Lovell, 1983), recent experiments applying automatic search algorithms to both the correct specification and highly over-parameterized models thereof have revealed that search costs are surprisingly small in relation to the costs of inference: see *inter alia* Hoover and Perez (1999, 2004), Hendry and Krolzig (1999, 2003) and Krolzig and Hendry (2001).

This transformation in success rates is due to many developments, of which the most important is searching all the feasible reduction paths when simplifying the general model. Such a thorough exploration avoids selecting a sub-optimal representation, which might be located if only a single search path is examined (such as successive deletion of the least significant variables). Moreover, 'path dependence' of model selection is thereby removed, and the selected model becomes objectively reproducible. Conversely, doing so much search may be thought to raise the probability of false selection by greatly increasing the number of tests conducted. A technical explanation for why that does not occur is offered in section 1, but a medical analogy would be that a sequence of procedures is often used to increase the probability of correctly diagnosing an illness. However, if there are N regressors, there are 2^N sub-models, but $N!$ possible paths, each of up to N steps, so even for moderate N, computer search is an essential tool: $N = 10$ induces more than 10^6 paths. In practice, fewer paths need to be explored, as some immediately lead to non-congruent models, and viable shortcuts are feasible as we will show. Many approaches to automatic selection are now being developed, and this exciting area has already delivered important general breakthroughs (see RETINA for selecting non-linear representations by Perez-Amaral, Gallo and White, 2003, 2004, evaluated by Castle, 2004; Phillips, 1995, 1996, 2003, for selecting forecasting models; Omtzig, 2002, and Kurcewicz and Mycielski, 2003, for selecting cointegrating relations; and the special issue on model selection edited by Haldrup, van Dijk and Hendry, 2003).

Here, we focus on general-to-specific modelling, denoted *Gets*, as embodied in the computer program *PcGets* (see Hendry and Krolzig, 2001), based on the theory of reduction (see e.g., Hendry, 1995, ch.9: Campos, Ericsson and Hendry, 2004, overview the literature).[1] The settings of *PcGets* were calibrated by Monte Carlo in Hendry and Krolzig (2003) to implement two pre-programmed strategies, called Liberal and Conservative, at approximately 5% and 1% per test respectively. To summarize the known properties of these two strategies in sifting relevant from irrelevant variables in econometric modelling:

(a) *PcGets* model selection is consistent (see Campos, Hendry and Krolzig, 2003);

(b) irrelevant variables are eliminated at roughly the significance level chosen by the user (see Hendry and Krolzig, 2003);

(c) relevant variables are retained with probabilities close to the theory maximum achievable when the DGP equation is known (see Hendry and Krolzig, 2003);

(d) automatic model selection is labour saving—and perhaps essential—when there are many candidate variables;

(e) applications to some earlier empirical studies either match, or even improve upon, their authors' findings (e.g., Davidson, Hendry, Srba and Yeo, 1978, and Hendry and Ericsson, 1991); and

(f) *PcGets* suggests ways of improving model selection by information criteria (see e.g., Schwarz, 1978, denoted *SIC*).

The structure of the paper is as follows. After outlining the *PcGets* selection algorithm in section 1, its finite-sample behaviour is re-examined in section 2 across a range of Monte Carlo experiments from Hendry and Krolzig (1999, 2003) and Krolzig and Hendry (2001). Next, section 3 investigates possible small-sample 'pre-test biases' and 'model-selection effects' for both estimators and tests in the Krolzig and Hendry (2001) experiment. Section 4 discusses how unbiased estimation can be obtained despite selection, with reported standard errors close to the sampling standard deviations of the corresponding coefficients in the estimated DGP equation. Section 5 briefly describes a 'non-expert' version, then looks at the impact of near-collinearity on selection probabilities. Section 6 comments on using economic theory based restrictions during modelling. Section 7 then considers three problems that initially seem intractable, but in fact can be tackled by a *Gets* approach. The first is model selection when confronting more regressors than observations; the second is perfectly collinear regressors; and the third is the selection of simultaneous equations models despite the absence of prior identification information (although none of these developments has been programmed yet). Section 8 concludes.

[1] *PcGets* is an Ox Package (see Doornik, 1999) implementing automatic *Gets* modelling for linear regression equations.

1 The selection algorithm

PcGets has six basic stages in its approach to selecting a parsimonious undominated representation of an initial general unrestricted model, denoted the GUM. The first stage concerns the formulation of the GUM; the second, the settings of the selection algorithm; the third determines the estimation and testing of the GUM; the fourth is a pre-search process; the fifth is the multi-path search procedure; and the sixth is post-search evaluation. The following description sketches the main steps involved: see Hendry and Krolzig (2001) for details.

(1) Formulate the GUM based on subject-matter theory, institutional knowledge, historical contingencies, data availability and measurement information, ensuring the resulting model encompasses previous evidence, with a relatively orthogonal parameterization of the N candidate regressors.

(2) Select the set of m mis-specification tests (e.g., residual autocorrelation etc.), their forms (e.g., r^{th}-order), and significance levels (generically denoted δ below); choose the desired information criterion (e.g., *SIC*) for final selection between mutually encompassing congruent models; and set the significance levels of all selection tests (generically denoted α below) to ensure the desired rejection frequencies under the null, perhaps by selecting one of the pre-set Liberal or Conservative strategies.

(3) 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.

(4) Undertake pre-search reductions at a loose significance level (these include lag-order selection, F-tests on successively shorter lag groups, and F-type tests for sequentially increasing blocks of omitted variables (a) adding from the smallest absolute t-values till the critical value is reached from below, and (b) removing variables with the largest t-values till the critical value of the remainder is reached from above); eliminate the resulting insignificant variables to reduce the search complexity, then estimate the new GUM as the baseline for the remaining stages.

(5) Multiple-path reduction searches now commence from each feasible initial deletion; the validity of each reduction is diagnostically checked to ensure the congruence of the final model; 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 the start of 5); when all paths have been explored and all distinct terminal models have been found, they are tested against their union to find an undominated encompassing contender; rejected models are removed, and the union of the 'surviving' terminal models becomes the smaller GUM of a repeated multi-path search iteration; then this entire search process (i.e., from

the start of 5) continues till a unique choice of *final* model emerges with n regressors, or the search converges to a set of mutually encompassing and undominated contenders, in which case all the selected models are reported, and a unique final choice made by the pre-selected information criterion.

(6) The significance of every variable in the final model is assessed in two over-lapping sub-samples to check the reliability of the selection.

Stage 1 is crucial: a poor general framework is unlikely to lead to a good final model choice. The consistency properties of *PcGets* have been established in Campos *et al.* (2003) for the setting where the DGP is nested in the GUM using the analysis in Hannan and Quinn (1979): as the sample size $T \to \infty$ and $\alpha \to 0$ at a suitable rate for a fixed N, the DGP equation is selected with probability unity. While such a situation is unlikely to occur in empirical practice, the GUM should at least be designed as a good approximation to the local DGP (i.e., the DGP in the space of the variables under analysis, denoted LDGP: see Hendry, 1995, and Bontemps and Mizon, 2003). More generally, the results in White (1990) establish that a consistent selection can result using *Gets* within a progressive research strategy based on rigorous mis-specification testing.

The mis-specification tests selected in stage 2 implicitly define the measure of congruence relevant to the empirical study. These are used once only to test the GUM; thereafter, their re-use as diagnostic tests is simply as a constraint on the reduction paths, ensuring that only congruent models are considered. Here, 'repeated testing' at most marginally affects the tests' behaviour, though there are some finite-sample effects from eliminating irrelevant variables (see sub-section 3.3). The selection criterion for breaking 'ties', and the chosen search strategy, should both be set as a function of the nature of the problem. That requires appraising the relative costs of retaining irrelevant, as against losing relevant, variables. Knowledge of the likely number, k, of relevant effects and their importance in the DGP can be beneficial here, although the absolute and relative numbers of candidate variables N and data points T, and the objectives of the analysis, also matter.

Stage 3 completes the definition of the approach, using OLS or IV estimators at present (in principle, any maximum likelihood method could be implemented), given the set of tests selected. If the GUM is congruent, reduction can proceed; if not, re-thinking seems advisable. *PcGets* can also be used to select the relevant instrumental variables, and check for the problem of weak instruments (compare Hall, Rudebusch and Wilcox, 1996; and see e.g., Staiger and Stock, 1997, and Mavroeidis, 2004).

Stage 4 is described in detail in Hendry and Krolzig (2001), as pre-search reductions play a useful role in simplifying many problems to a manageable size. Pre-search block reductions even at a significance level of 0.75 can eliminate many irrelevant variables, and are especially useful when the null is true since 68% of t-values are less than unity in absolute value under the

null: when $N = 40$, and $k = 7$ say, the number of feasible paths falls to around 10^5. Campos *et al.* (2003) show that pre-search tests can also improve model selection based on information criteria (related to the argument in Hansen, 1999). Outlier corrections can be selected if desired at any percentage of the equation standard error in the GUM.

The multi-path search procedure (stage 5), based on the pioneering study by Hoover and Perez (1999), is central to *PcGets* and is evaluated in Hendry and Krolzig (2003). Section 6 notes a setting where the multi-path search procedure can handle perfectly collinear specifications, so pre-search tests should not be used in that context.

Finally, Hendry and Krolzig (2004b) show that the sub-sample reliability assessment is dominated by choosing an appropriately smaller significance level for the full sample. This matches the findings in Lynch and Vital-Ahuja (1998), who show that 'selecting variables that are significant on all three splits (the two sub-samples and overall)' delivers no gain over simply using a smaller nominal size. The Lynch and Vital-Ahuja (1998) argument applies widely to 'hold back observations' approaches, and to (e.g.) Hoover and Perez (1999, 2004) who retain variables at the selection stage only if they are significant in two overlapping sub-samples. However, the efficiency loss seems to be small, so a sub-sample selection procedure is still offered.

Several changes to this basic algorithm have been implemented since Hendry and Krolzig (2001), so we briefly note these. Most only slightly altered the program's behaviour, reflecting how near the theoretical upper bound performance already is, and the degree of 'error correction' manifest in the experiments used to calibrate the program (when one procedure performed relatively poorly, another usually did well). Nevertheless, improvements potentially remain feasible in several directions.

First, some formulations were not previously envisaged, such as a model with long lags of a variable when only a few lags actually matter. When one, or a few, important effects are hidden in a morass of irrelevance, the pre-search block tests need not be appropriate. Consequently, as a check on the F test of all lags, *PcGets* also considers the significance of the largest t-test in the group, and only deletes the block if both are insignificant at loose significance levels. We now also use less stringent significance levels for the block tests than in Hendry and Krolzig (1999), where the overall procedure was notably under-sized under the null.

Secondly, the calibration of the mis-specification heteroskedasticity tests was poor in early experiments, but this transpired to be a problem with the degrees of freedom assumed for the reference distribution.[2] The corrected degrees of freedom lead to a substantial improvement in matching the reference distribution under the null as noted in Hendry and Krolzig (2003). Sub-section 3.3 below reports comparisons of the ARCH and White (1980) heteroskedasticity tests applied to the DGP, GUM, and finally-selected model.

[2]We are indebted to Dorian Owen for noting this mistake.

Thirdly, lag-order determination uses the combined 'top-down/bottom-up' approach explained in 4 (a), (b), complemented by an automatic Lagrange-multiplier test for potential omitted regressors.

Finally, we investigated the information in the ordered t^2-statistics in the GUM (denoted $\mathsf{t}^2_{(i)}$) to locate a cut-off between included and excluded variables. After ordering, $\mathsf{t}^2_{(1)} \geq \mathsf{t}^2_{(2)} \geq \cdots \geq \mathsf{t}^2_{(i)} \geq \cdots \geq \mathsf{t}^2_{(N)}$, so if the critical value per test is c_α, let n correspond to $\mathsf{t}^2_{(n)} \geq c_\alpha \geq \mathsf{t}^2_{(n+1)}$. Since such a procedure is only suitable for orthogonal problems, multi-path searches remain necessary in general (section 5.1 briefly addresses the near-collinearity issue). Importantly, however, the logic of this procedure helps explain why *PcGets* works well, since it reveals that only a single 'model-selection test' is used to select the n included variables, namely $\mathsf{t}^2_{(n)} \geq c_\alpha \geq \mathsf{t}^2_{(n+1)}$, so 'repeated testing' does not occur (especially not $N!$ tests). The overall retention of adventitiously significant variables depends only on $N - k$, for k variables in the DGP, and the significance level per test α, so on average, $(N-k)\,\alpha$ irrelevant variables are retained by chance: it should not depend on how the searches *per se* are conducted, which instead determine the efficiency of the algorithm in attaining the upper bound feasible in an orthogonal problem. Thus, despite the appearance that large numbers of tests are conducted on coefficients and residuals, the probability of false rejection of a congruent GUM by the mis-specification tests, under independence, is just $\mathsf{p}_\delta = 1 - (1-\delta)^m$. This is approximately 0.05 for $m = 5$ and $\delta = 0.01$. Further, $(N-k)\,\alpha \simeq 1$ for $N = 30,\ k = 10$ and $\alpha = 0.05$—so 19 out of 20 irrelevant regressors will be eliminated on average—and is just 0.2 for $\alpha = 0.01$, so all 20 are eliminated on 4 out of 5 occasions. The so-called 'size' of the search procedure is $\mathsf{p}_\alpha = 1 - (1-\alpha)^{N-k}$ which is 0.64 when $\alpha = 0.05$ and 0.18 when $\alpha = 0.01$, revealing the unhelpful nature of such a characterization (pre-search tests can improve performance relative to these expository baselines). Retention of the k correctly significant variables depends on their individual non-centralities, ψ_i, and c_α via $\mathsf{P}(|\mathsf{t}_i| > c_\alpha|\,\psi_i)$ (for a t-test), which also decreases with α, necessitating a careful choice of strategy, as noted below. However, retention rates can be improved if inclusion is known, or (e.g.) sign information is provided by theory and found acceptable for the available sample, since only one-sided critical values are then needed (see section 6).

2 Small-sample behaviour of *PcGets*

Table 1 summarizes the main features of the various Monte Carlo experiments conducted to date, and referred to below (HP, JEDC, S_0–S_4 and S^*_0–S^*_4 respectively denote Hoover and Perez, 1999, Krolzig and Hendry, 2001, and two variants of the *PcGets* calibration experiments in Hendry and Krolzig, 2003). We now summarize the operating characteristics of *PcGets* across the experiments in table 1.

Table 1 Monte Carlo designs.

Design	regressors	causal	nuisance	$\|t\|$-values	avg. $\|t\|$-value
HP0	41	0	41		
HP2*	41	1	40	5.77	5.77
HP2	41	1	40	11.34	11.34
HP7	41	3	38	(10.9, 16.7, 8.2)	11.93
JEDC	22	5	17	(2,3,4,6,8)	4.6
S_0	34	0	34		
S_2	34	8	26	(2,2,2,2,2,2,2,2)	2
S_3	34	8	26	(3,3,3,3,3,3,3,3)	3
S_4	34	8	26	(4,4,4,4,4,4,4,4)	4
S_0^*	42	0	42		
S_2^*	42	8	34	(2,2,2,2,2,2,2,2)	2
S_3^*	42	8	34	(3,3,3,3,3,3,3,3)	3
S_4^*	42	8	34	(4,4,4,4,4,4,4,4)	4

Figure 1 graphically illustrates four main aspects of calibration accuracy across all the Monte Carlo experiments to date, for both Conservative and Liberal strategies. Panel (a) concerns one sense of 'overfitting', namely potentially downward biased estimates of the equation standard error, $\widehat{\sigma}$, for the true value σ. This does not occur: the final average $\widehat{\sigma}$ is close to σ in all settings. The Liberal strategy has a slight downward bias (less than 5% of σ), whereas the Conservative is upward biased by a similar amount. Such behaviour is easily explained: the Conservative strategy is more likely to eliminate variables which matter somewhat, so fits worse than the GUM, which unbiasedly estimates σ; and the Liberal strategy is more likely to retain some variables which only matter by chance, but thereby slightly overfits. It must be stressed that *PcGets* model selection is not based on fit as a criterion at any stage, but a minimal congruent encompassing model will necessarily have the best fit at the chosen significance level. Equation (1) records the goodness-of-fit relationship between models of size s and $s+1$ for unbiased OLS estimators of σ (i.e., corrected for degrees of freedom):

$$\frac{\widehat{\sigma}_s^2}{\widehat{\sigma}_{s+1}^2} = 1 + \frac{t_{(s+1)}^2 - 1}{T - s}. \tag{1}$$

The probability under the null that $|t| > 2.5$ is 0.014 (when $T = 110$ and $s = 10$), so larger t^2-values will occur less than once in 70 draws under the null, yet even for such an unlikely event, the left-hand ratio in (1) would only be about 1.05, the upper bound shown in the graph.

Panel (b) shows the null rejection frequencies per test for both strategies across all experiments, with their intended significance levels of 5% and 1%. In no case are deviations substantial for the unweighted null rejection frequencies. When the reliability statistics (see stage 6) are taken into account, and translated into retention probabilities in a linear fashion, the con-

trol of the null rejection frequency is improved further. Thus, if a 50% reliability is found, the investigator is assumed to drop that variable on half the occasions. The resulting sub-sample reliability-weighted outcomes are close to their targets (denoted '(rel)' on the graphs). This second sense of overfitting only occurs to the controlled extent of adventitious significance at the rate $(N-k)\,\alpha$.

Panel (c) plots 'power', namely the average rejection frequency of the null for relevant variables using the nominal critical values.[3] The Conservative strategy naturally has no higher power than the Liberal, so reveals that the cost of avoiding spurious variables can be high in terms of missing variables that matter. The graphs also show the impact of the sub-sample reliability weightings on the resulting power, confirming that there is only a small effect, even at quite low powers where it should have most impact. Since $\mathsf{P}(|\mathsf{t}| > 2|\,\psi = 2) \simeq 0.5$, the 'powers' in the S_2 experiments are close to their theoretical upper bound, despite selection (similarly for the other S_j). Comparisons between neighbouring successive S_j and S_j^* experiments also show that the impact on 'power' of eight additional irrelevant variables is small, especially for the Liberal strategy.

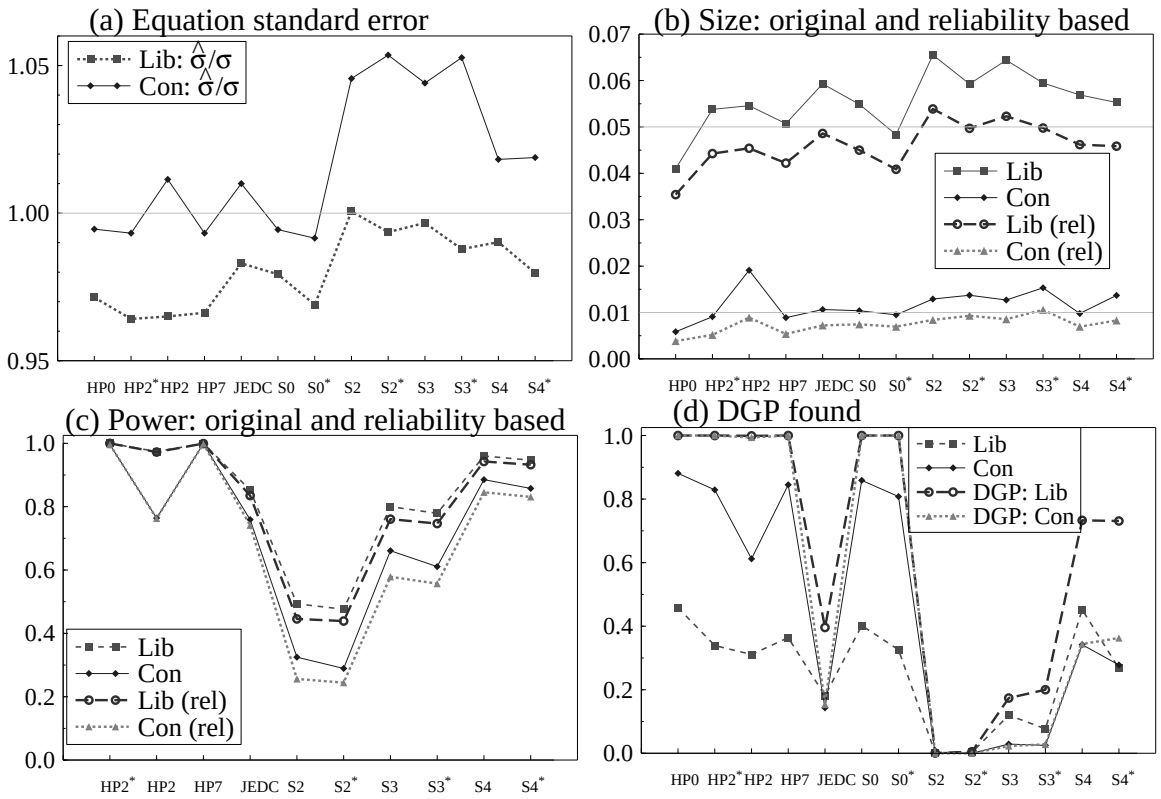

Figure 1 Overview of accuracy, null rejection frequency, power, and success.

Finally, figure 1(d) graphs the probabilities of locating the DGP, together with the corresponding outcomes when the search commences with the DGP itself treated as the GUM. The

[3]The simulation null rejection frequencies are sufficiently close to the nominal to use the latter, which matches what an empirical investigator would do in practice.

movements of the four lines are similar, and frequently the apparent problem for a search algorithm transpires to be a cost of inference not a cost of search, since the DGP is sometimes never retained even when it is the initial specification. The out-performance of commencing from the DGP in the Hoover–Perez experiments is owing to the high degree of over-parameterization and very large t-values on relevant variables, but even so, the Conservative strategy does a respectable job. When population t-values are 2 or 3, the Liberal strategy does best, and sometimes outperforms commencing from the DGP with a 1% significance level (S_3 and S_4). Notice also that the two strategies cannot be ranked on this fourth criterion: their relative performance depends on the unknown state of nature. Nevertheless, as Hendry and Krolzig (2001, Ch. 5) discuss, a user may be aware of the 'type' of problem being confronted, in which case, figure 1(d) shows the advantages of an appropriate choice of strategy combined with a good initial model specification.

These findings also confirm the closeness in practice of the strategies to their desired operating characteristics.

3 'Pre-test' and 'selection' effects in small samples

Statistical tests with non-degenerate null distributions have non-zero size, and (generally) non-unit power. Consequently, even if the local DGP were correctly specified *a priori* from economic theory, when an investigator did not know that the resulting model was 'true' – so sought to test hypotheses about its coefficients – then inferential mistakes can occur, the seriousness of which depend on the characteristics of the local DGP and the sample drawn. Should the selected model thereby differ from the DGP (with parameters β_i), it will deliver biased coefficient estimates $\widehat{\beta}_i$: $\mathsf{E}[\widehat{\beta}_i \neq \beta_i]$. This is called the 'pre-test' problem, since unbiased estimates could have been obtained from the unrestricted model by conducting no selection tests (see e.g., Judge and Bock, 1978). However, assuming that one both knows the truth, and knows that one does, so no testing is needed, is not a relevant benchmark in economics. Moreover, granted the arguments in Hendry and Krolzig (2003) against using alternatives such as Stein–James 'shrinkage', or even the general model, then some selection method is essential. In the following simulations, we also record the outcomes when commencing from the DGP to measure the additional costs of selection due to commencing from the GUM.

3.1 Selection effects on coefficient estimates

To investigate the impact of selection, we re-ran the Krolzig and Hendry (2001) experiments. As shown in table 2 (columns 3, 4, 8 and 9), unconditionally, coefficient estimates are downward biased (being a mix of $\widehat{\beta}_i$ and 0 when a regressor x_i is and is not retained). In this section, 'conditional' denotes conditional on a variable being retained in the selected model; whereas

Table 2 Coefficient estimates, estimated standard errors and standard deviations.

	DGP	Reduction of DGP				GUM	Reduction of GUM			
		unconditional (including zeros)		conditional (excluding zeros)			unconditional (including zeros)		conditional (excluding zeros)	
variable		LIB	CON	LIB	CON		LIB	CON	LIB	CON
Bias/SE_{Th}										
$Z_a\ (\psi = 2)$	0.041	-0.576	-1.076	0.859	1.242	0.042	-0.606	-1.072	0.852	1.221
$Z_b\ (\psi = 3)$	0.013	-0.300	-0.701	0.322	0.576	0.002	-0.291	-0.757	0.328	0.600
$Z_c\ (\psi = 4)$	-0.008	-0.035	-0.216	0.066	0.196	-0.006	-0.071	-0.266	0.093	0.215
$Z_d\ (\psi = 6)$	0.042	0.009	0.009	0.015	0.021	0.039	0.037	0.044	0.037	0.050
$Z_e\ (\psi = 8)$	0.033	-0.042	-0.041	-0.042	-0.041	0.005	0.034	0.026	0.034	0.026
SE/SE_{Th}										
$Z_a\ (\psi = 2)$	1.025	0.506	0.289	1.015	1.012	1.130	0.486	0.290	0.993	1.006
$Z_b\ (\psi = 3)$	1.019	0.832	0.659	1.023	1.024	1.124	0.811	0.625	0.996	1.004
$Z_c\ (\psi = 4)$	1.026	1.000	0.926	1.025	1.026	1.131	0.966	0.905	1.007	1.021
$Z_d\ (\psi = 6)$	1.023	1.029	1.036	1.030	1.038	1.129	1.007	1.026	1.007	1.027
$Z_e\ (\psi = 8)$	1.027	1.025	1.033	1.025	1.033	1.133	1.010	1.031	1.010	1.031
SD/SE_{Th}										
$Z_a\ (\psi = 2)$	1.028	1.503	1.499	0.659	0.611	1.145	1.508	1.497	0.702	0.624
$Z_b\ (\psi = 3)$	1.023	1.492	1.815	0.821	0.749	1.126	1.502	1.842	0.843	0.749
$Z_c\ (\psi = 4)$	1.034	1.133	1.509	0.950	0.894	1.146	1.254	1.584	0.984	0.899
$Z_d\ (\psi = 6)$	1.031	1.038	1.074	1.021	1.048	1.160	1.079	1.080	1.079	1.063
$Z_e\ (\psi = 8)$	1.057	1.004	1.020	1.004	1.020	1.187	1.108	1.099	1.108	1.099
RMSE/SE_{Th}										
$Z_a\ (\psi = 2)$	1.028	1.610	1.846	1.083	1.384	1.145	1.625	1.841	1.104	1.371
$Z_b\ (\psi = 3)$	1.023	1.522	1.946	0.882	0.945	1.126	1.530	1.992	0.905	0.960
$Z_c\ (\psi = 4)$	1.034	1.133	1.524	0.952	0.915	1.146	1.256	1.606	0.988	0.924
$Z_d\ (\psi = 6)$	1.032	1.038	1.074	1.021	1.041	1.160	1.079	1.080	1.079	1.064
$Z_e\ (\psi = 8)$	1.058	1.005	1.020	1.005	1.020	1.187	1.109	1.100	1.109	1.100
residuals										
$\widehat{\sigma}$	0.998	1.007	1.017			0.998	0.981	1.008		
% bias	-0.2%	0.7%	1.7%			-0.2%	-1.9%	0.8%		

Monte Carlo results for JEDC design (see table 1) with $T = 100$ observations and $M = 1000$ replications:

Bias	mean deviation of $\widehat{\beta}$ in MC from true β	SE_{Th}	true standard error of $\widehat{\beta}$ ($T^{-1/2}\sigma/\sigma_Z = 0.1$)
RMSE	root mean square error of $\widehat{\beta}$ in MC	SE	mean of reported standard errors
$\widehat{\sigma}$	estimated standard deviation of error term ($\sigma = 1$)	SD	standard deviation of $\widehat{\beta}$ in MC

'unconditional' denotes also including the coefficients imposed at zero. Figure 2 shows the unconditional distributions of the five relevant and 17 irrelevant regressors for the Liberal strategy.[4] These unconditional distributions illustrate the quality of the classification of variables into DGP variables (top row) and nuisance variables (all others). The non-zero-mass distribution of the DGP variables is truncated normal, but truncation does not affect variables with a population t-value greater than 4 in absolute value.

Conditional on being retained, the bias, reported standard errors (SE), standard deviations (SD) and root mean-square errors (RMSE) are shown in columns 5, 6, 10 and 11 of table 2, all relative to the theoretical standard errors (SE_{Th}). As expected, the coefficient estimates are now upward biased for smaller t-values ($|\mathsf{t}| \leq 3$), more so for the Conservative strategy, but are close to the population values for larger t-values. The Liberal strategy biases are under 10% for $|\mathsf{t}| > 3$.

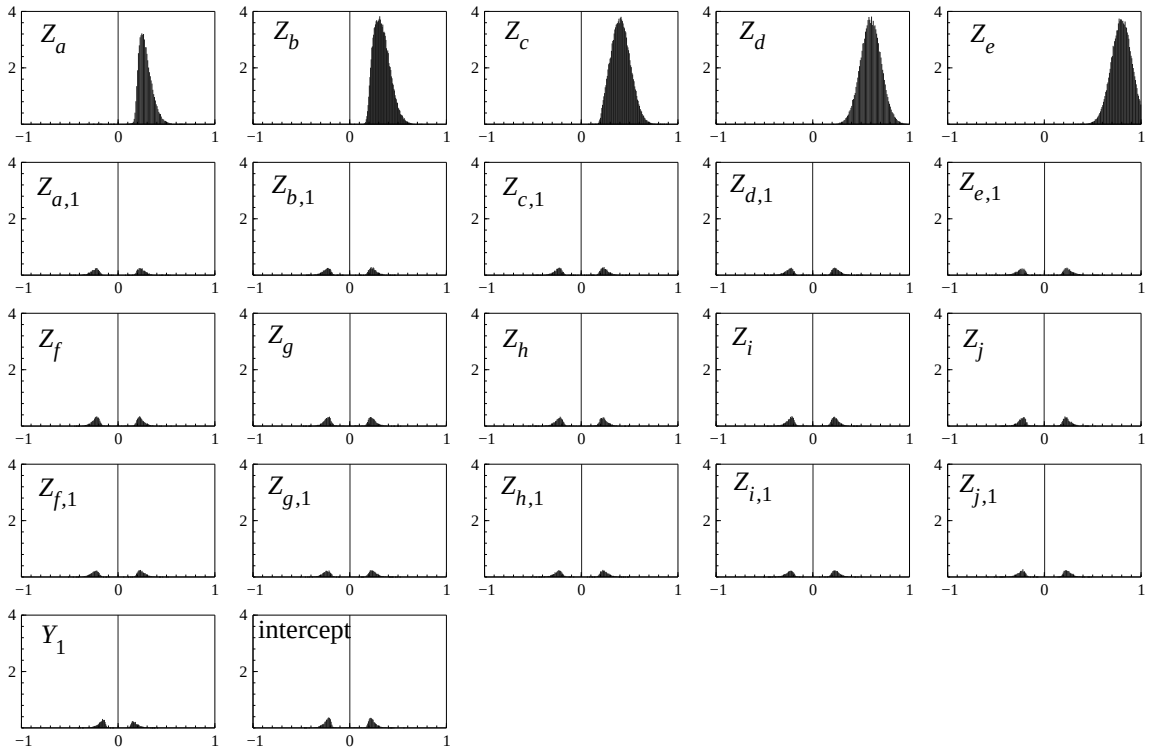

Figure 2 Unconditional distributions from the Liberal strategy.

Figure 3 records the corresponding conditional distributions (i.e., of retained regressors). Those for the non-DGP variables are bimodal and symmetric, except for the lagged endogenous variable, where the impact of the famous Hurwicz (1950) bias is clear.

The final important result is that these 'pre-test' effects are not, in any essential respects, changed by search. The coefficient biases are closely similar when commencing from the DGP

[4]The results for the Conservative strategy are similar, but the distributions of irrelevant variables are almost invisible, and so are not shown.

or the GUM for each strategy, both conditionally and unconditionally as table 2 confirms.

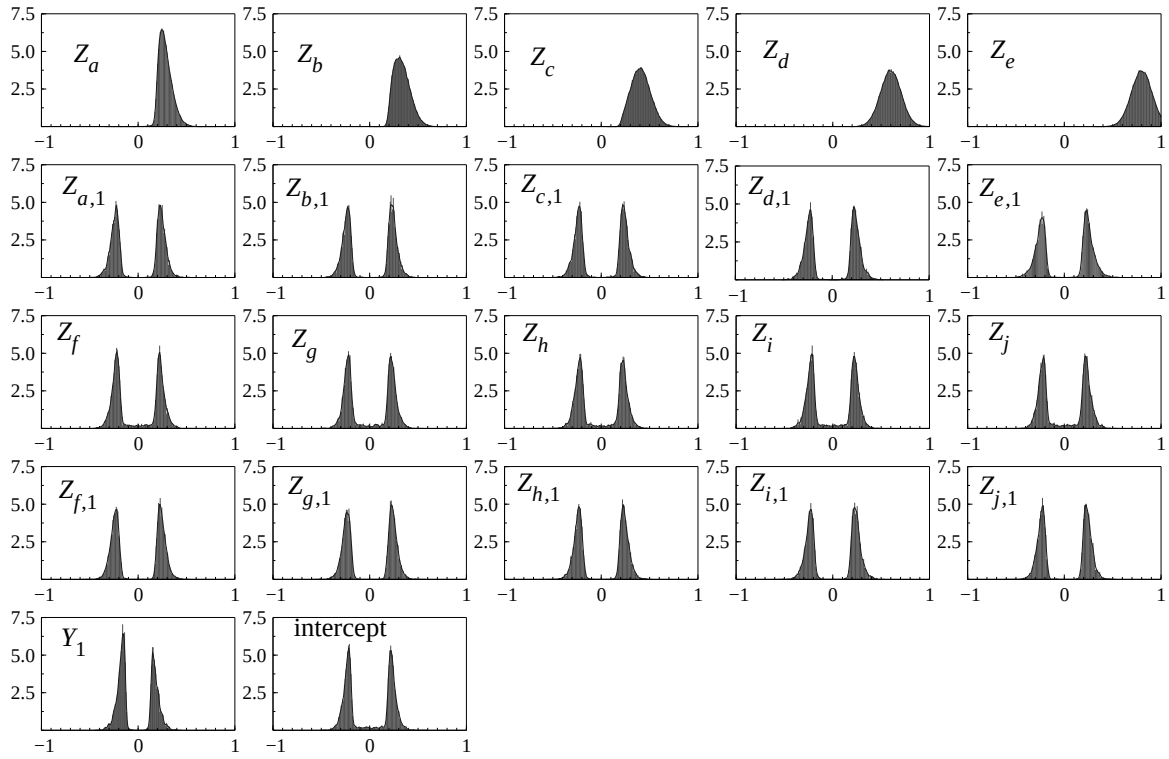

Figure 3 Conditional distributions (excluding zeros) from the Liberal strategy.

3.2 Selection effects on estimated standard errors and standard deviations

Table 2 shows that the estimated standard errors (SEs, namely those reported for the selected equation's coefficients) are close to providing unbiased estimates of the actual sampling standard deviations (SDs) for the estimated DGP. At first sight, that is an astonishing result, since the estimated uncertainty, despite having to select a DGP variable from a GUM, appears to reflect only the uncertainty due to estimating the DGP without selection. However, the intuition is simple: the SDs in the estimated DGP model are correctly estimated by the reported SEs (column 2); the latter are based on the estimated equation standard error ($\widehat{\sigma}$, which is close to σ on average as shown on the bottom row) times the associated square-root element from $(\mathbf{X}'\mathbf{X})^{-1}$; and that in turn is approximately the same in the selected model when the relevant variable is retained. Thus, similar SEs are reported.

Because parameter estimates restricted to zero have zero standard errors, unconditional SEs after selection are downwards biased, whereas the corresponding unconditional SDs are upward biased (being that of a mix of 0 and the variability in $\widehat{\beta}_i$). The probability p of retaining a variable with a population t^2-value of 4 is approximately 0.5, so the effects are largest at small population t-values. Indeed, the mean unconditional estimates and their SEs are approximately

p times the corresponding conditional estimates. However, the relevance of such unconditional 'sampling properties' is unclear in the context of model selection when the DGP is unknown. The elimination of insignificant variables is the objective of simplification in small samples, and the underlying state of nature is unknown (i.e., whether variables are relevant or irrelevant), so the cost of the bimodality of the unconditional selection distribution for relevant variables is a larger SD.

As noted earlier, in almost all cases, the estimated equation standard errors $\widehat{\sigma}$ are close to σ, so that *PcGets* does not 'overfit'. Rather, the Conservative strategy underfits by eliminating too many of the relevant regressors in its attempt to avoid adventitious significance, whereas the Liberal strategy performance depends on the number of irrelevant variables in the GUM, and can be either under or over σ. Indeed, so can the SEs and SDs, both conditional on retaining a variable, and unconditionally.

Overall, the results in this section seem to confirm using the Liberal strategy as the default option.

3.3 Selection effects on mis-specification tests

Another feature of interest is the impact of model selection on the outcomes of test statistics. In Krolzig and Hendry (2001), we have shown that, even in small samples ($T = 100$), the empirical distributions of the test statistics for no autocorrelation, normality and no structural break employed by *PcGets* are largely unaffected by the strongly-exogenous nuisance regressors. Here we consider selection effects on the two heteroskedasticity tests, recalibrated as noted. The graphs in figure 4 compare the ratios of actual sizes to nominal in the DGP, GUM and the selected model.[5]

The operational rules adopted were as follows. If the GUM showed no mis-specifications at 5%, then simplified models with diagnostic tests indicating an invalid reduction at 1% or less were rejected. If a mis-specification test of the GUM was significant at 1%, the test was dropped from the test battery. If the p-value of the mis-specification test was between 1% and 5%, the target significance level was reduced from 1% to 0.5%.

As can be seen from the graphs, there is little change in the rejection frequencies for quantiles above the nominal significance level, but an increasing impact as the quantile decreases. The latter effect is essentially bound to occur, since models with significant heteroskedasticity are selected against by construction. Nevertheless, the outcomes in these graphs do not represent a 'distortion' of the sampling properties: the key decision is taken at the level of the general model, and conditional on not rejecting there, no change should occur in that decision. At most nominal significance levels in the GUM, the tests have their anticipated operating char-

[5]The 1% level showed larger departures, but was imprecisely estimated given the rarity with which it occurred, and has been omitted from the graphs.

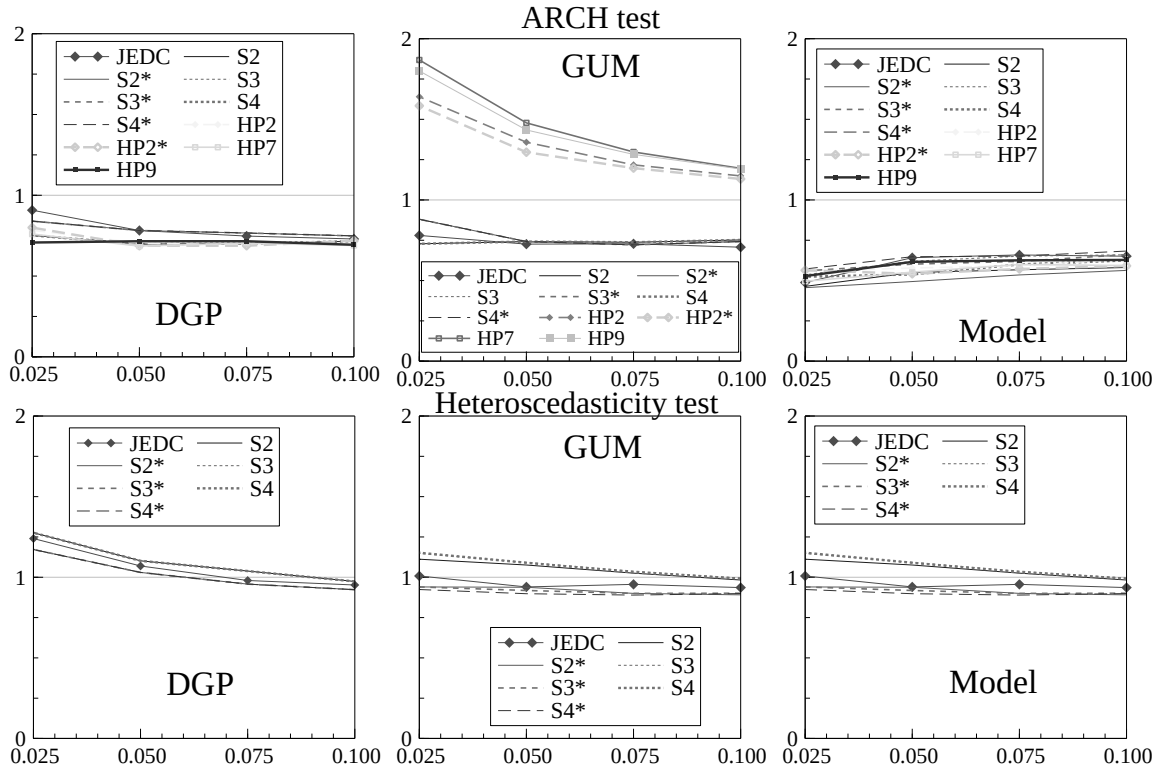

Figure 4 Ratios of test sizes to nominal in the DGP, GUM and selected model.

acteristics. However, the ARCH test was oversized at smaller significance levels in the HP experiments due to the heteroskedastic nuisance regressors affecting the residuals of the GUM.

4 Bias correction after model selection

The selection biases discussed in section 3.1 can be substantially corrected by an operational formula, which we now describe. Consequently, despite searching in a large model class, across different (unknown) states of nature in an orthogonal setting, the finally selected model can be modified to deliver nearly unbiased estimates and essentially unbiased standard errors for retained variables, with few adventitiously-significant effects—a performance close to that achievable when commencing from the local DGP.

4.1 Truncated-distribution approximations to the bias

We use the convenient approximations that for a given sample of size T:

$$\mathsf{t}_{\widehat{\beta}} = \frac{\widehat{\beta}}{\widehat{\sigma}_{\widehat{\beta}}} \simeq \frac{\widehat{\beta}}{\sigma_{\widehat{\beta}}},$$

where $\widehat{\sigma}_{\widehat{\beta}}$ is the coefficient SE, and $\sigma_{\widehat{\beta}} = \mathsf{E}[\widehat{\sigma}_{\widehat{\beta}}]$ is the population value, so:

$$\frac{\widehat{\beta}}{\sigma_{\widehat{\beta}}} \sim \mathsf{N}\left[\frac{\beta}{\sigma_{\widehat{\beta}}}, 1\right] = \mathsf{N}\left[\psi, 1\right],$$

when $\psi = \beta/\sigma_{\widehat{\beta}}$ is the non-centrality of the t-test. Let:

$$\phi\left(w\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}w^2\right) \text{ and } \Phi\left(w\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} \exp\left(-\frac{1}{2}x^2\right) \mathrm{d}x.$$

When $\psi > 0$, for one-sided truncation in a normal distribution (see e.g., Johnson and Kotz, 1970, ch. 13):

$$\mathsf{E}\left[\frac{\widehat{\beta}}{\sigma_{\widehat{\beta}}} \middle| \frac{\widehat{\beta}}{\sigma_{\widehat{\beta}}} \geq c_\alpha\right] = \psi + \frac{\phi\left(c_\alpha - \psi\right)}{1 - \Phi\left(c_\alpha - \psi\right)} = \psi + r\left(c_\alpha - \psi\right), \tag{2}$$

when $r\left(\cdot\right)$ is the inverse Mills ratio:

$$r\left(w\right) = \frac{\phi\left(w\right)}{1 - \Phi\left(w\right)},$$

so:

$$\mathsf{E}\left[\widehat{\beta} \middle| \widehat{\beta} \geq \sigma_{\widehat{\beta}} c_\alpha\right] = \beta + \sigma_{\widehat{\beta}} r\left(c_\alpha - \psi\right) = \beta\left(1 + \psi^{-1} r\left(c_\alpha - \psi\right)\right). \tag{3}$$

Table 3 shows the close correspondence between the mean observed outcomes for the relevant variables from the Monte Carlo experiments in Krolzig and Hendry (2001) (denoted $\widehat{\beta}_{MC}^{(c_\alpha)}$) and those implied by (3) using:

$$\widehat{\beta}_{Th}^{(c_\alpha)} = \beta\left(1 + \psi^{-1} r\left(c_\alpha - \psi\right)\right), \tag{4}$$

for one-sided truncation with $c_{0.05} = 2.0$ and $c_{0.01} = 2.625$. In all cases, the predicted bias $\widehat{\beta}_{Th}^{(c_\alpha)}$ closely matches that obtained in the Monte Carlo, $\widehat{\beta}_{MC}^{(c_\alpha)}$, at both 1% and 5%.

Table 3 Conditional coefficient estimates and theory predictions.

β	0.200	0.300	0.400	0.600	0.800
$\widehat{\beta}_{MC}^{(.05)}$	0.286	0.332	0.407	0.602	0.796
$\widehat{\beta}_{Th}^{(.05)}$	0.280	0.329	0.405	0.600	0.800
$\widehat{\beta}_{MC}^{(.01)}$	0.324	0.358	0.420	0.602	0.796
$\widehat{\beta}_{Th}^{(.01)}$	0.324	0.358	0.417	0.600	0.800

In practice, since $\psi = 0$ for the irrelevant variables, a doubly-truncated Gaussian density is required, where the central region is lost and only the tails retained. Again using $\mathsf{t}_{\widehat{\beta}} \simeq \widehat{\beta}/\sigma_{\widehat{\beta}} \sim \mathsf{N}\left[\psi, 1\right]$, the expectation of the truncated t-value is:

$$\psi^* = \mathsf{E}\left[\mathsf{t}_{\widehat{\beta}} \middle| |\mathsf{t}_{\widehat{\beta}}| > c_\alpha; \psi\right] = \psi + \frac{\phi\left(c_\alpha - \psi\right) - \phi\left(-c_\alpha - \psi\right)}{1 - \Phi\left(c_\alpha - \psi\right) + \Phi\left(-c_\alpha - \psi\right)} = \psi + r\left(\psi, c_\alpha\right). \tag{5}$$

For $\psi = 0$, the truncated distribution is symmetric around zero as illustrated in figure 5, so $\psi^* = 0$. However, even if (e.g.) $\psi = 1$ and $c_\alpha = 2$, the lower tail contributes almost nothing to the resulting mean, matching the closeness of (4) to the simulation outcomes for the non-central t-statistics.

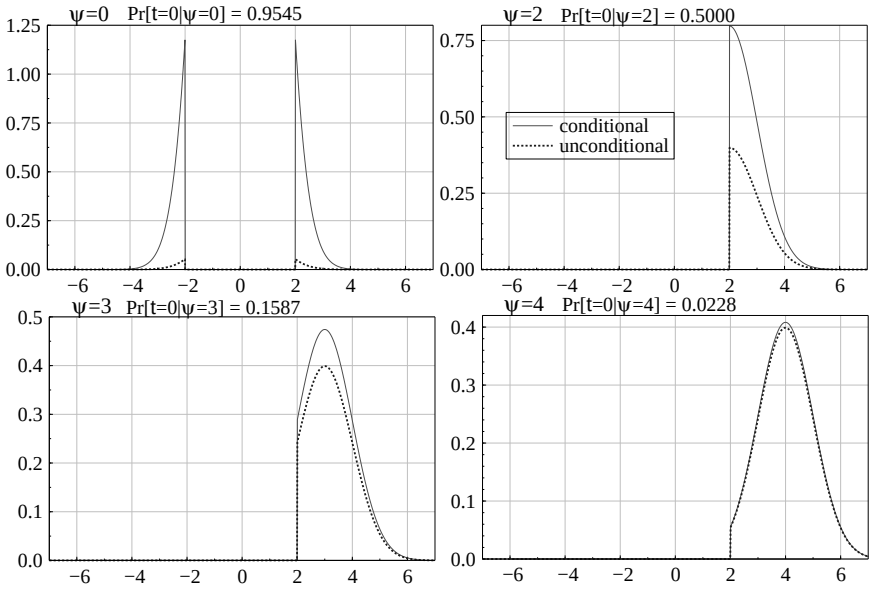

Figure 5 Probability densities of the conditional and unconditional coefficient estimates.

To implement the equivalent of (4) based on (5) requires an estimate $\widetilde{\psi}$ of the non-centrality ψ from the observed $\mathsf{t}_{\widehat{\beta}}$ statistic, which is then used in the correction formula for the estimated parameters. When a variable is retained because $|\mathsf{t}_{\widehat{\beta}}| > c_\alpha$, the selection bias is shown in (5) and hence:

$$\psi = \mathsf{E}\left[\mathsf{t}_{\widehat{\beta}} \middle| |\mathsf{t}_{\widehat{\beta}}| > c_\alpha; \psi\right] - r\left(\psi, c_\alpha\right). \tag{6}$$

This is a non-linear function of the unknown ψ, but could be solved by a step-wise iteration.[6] Figure 6 shows how non-linear the mapping is at $c_\alpha = 2$ as ψ varies, so such a bias correction cannot work perfectly because small variations in the estimate of ψ in some regions will induce large changes in the bias. Given an estimate $\widetilde{\psi}$ of ψ, then the bias-corrected parameter estimate is based on the inverse of (4) using (5), namely:

$$\overline{\beta} = \widehat{\beta}\left(\frac{\widetilde{\psi}}{\widetilde{\psi} + r\left(\widetilde{\psi}, c_\alpha\right)}\right). \tag{7}$$

We now consider these two steps in more detail.

[6]We tried several approaches, including a Newton and two Taylor approximations, shown in the simulation outcomes below.

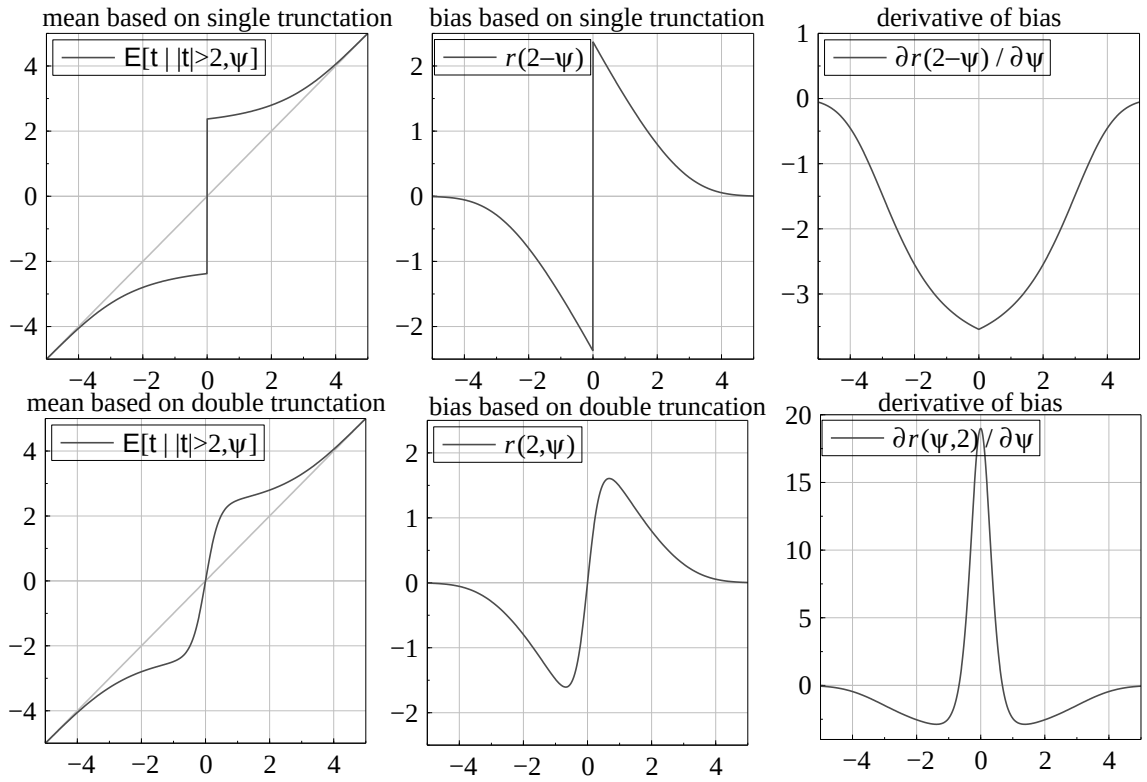

Figure 6 Non-linearity of the bias function.

4.2 Estimating the non-centrality

The first step in the iterative ψ-estimation procedure replaces the unknown $r(\psi, c_\alpha)$ in (6) by $r(\mathsf{t}_{\widehat{\beta}}, c_\alpha)$, and the expectation $\mathsf{E}[\mathsf{t}_{\widehat{\beta}} | \, |\mathsf{t}_{\widehat{\beta}}| > c_\alpha; \psi]$ by the observed value $\mathsf{t}_{\widehat{\beta}}$ (an unbiased estimate) to deliver:

$$\overline{\mathsf{t}}_{\widehat{\beta}} = \mathsf{t}_{\widehat{\beta}} - r\left(\mathsf{t}_{\widehat{\beta}}, c_\alpha\right), \tag{8}$$

then:

$$\widetilde{\psi} = \mathsf{t}_{\widehat{\beta}} - r\left(\overline{\mathsf{t}}_{\widehat{\beta}}, c_\alpha\right). \tag{9}$$

The Monte Carlo results in figure 7 below show that most, but not all, of the selection bias is corrected. However, any closer match for the smallest non-zero non-centrality considered here ($\psi = 2$) tends to induce over-correction at somewhat larger t-values. This would probably be exacerbated by matching at yet smaller non-centralities, given figure 6.

4.3 Correcting biases in $\widehat{\beta}$

The second step involves solving (7). Even if $\mathsf{E}[\widetilde{\psi}] = \psi$, $\overline{\beta}$ could be biased due to being a non-linear function of $\widetilde{\psi}$:

$$\overline{\beta} = \widehat{\beta}\left(\frac{\widetilde{\psi}}{\widetilde{\psi} + r\left(\widetilde{\psi}, c_\alpha\right)}\right). \tag{10}$$

An alternative formula for the denominator of (10) is:

$$\widetilde{\psi} + r\left(\overline{\mathsf{t}}_{\widehat{\beta}}, c_\alpha\right) = \mathsf{t}_{\widehat{\beta}}, \tag{11}$$

and this is used in the following derivation. Assuming $\widetilde{\psi}$ has been bias corrected such that $\mathsf{E}[\widetilde{\psi}] = \psi$, and as $\widehat{\beta} \simeq \sigma_{\widehat{\beta}} \mathsf{t}_{\widehat{\beta}}$, then from (10):

$$\overline{\beta} = \widehat{\beta}\frac{\widetilde{\psi}}{\mathsf{t}_{\widehat{\beta}}} \simeq \frac{\sigma_{\widehat{\beta}} \mathsf{t}_{\widehat{\beta}} \widetilde{\psi}}{\mathsf{t}_{\widehat{\beta}}} = \sigma_{\widehat{\beta}} \widetilde{\psi},$$

so:

$$\mathsf{E}\left[\ \overline{\beta}\ \right] \simeq \mathsf{E}\left[\sigma_{\widehat{\beta}} \widetilde{\psi}\right] = \sigma_{\widehat{\beta}} \psi = \beta. \tag{12}$$

However as:

$$\mathsf{E}\left[r\left(\overline{\mathsf{t}}_{\widehat{\beta}}, c_\alpha\right) \middle| \mathsf{t}_{\widehat{\beta}} \geq c_\alpha\right] \neq \mathsf{E}\left[r\left(\widetilde{\psi}, c_\alpha\right) \middle| \mathsf{t}_{\widehat{\beta}} \geq c_\alpha\right] \neq r\left(\psi, c_\alpha\right),$$

an additional bias may result from this step if other approximations to the denominator of (10) are used.

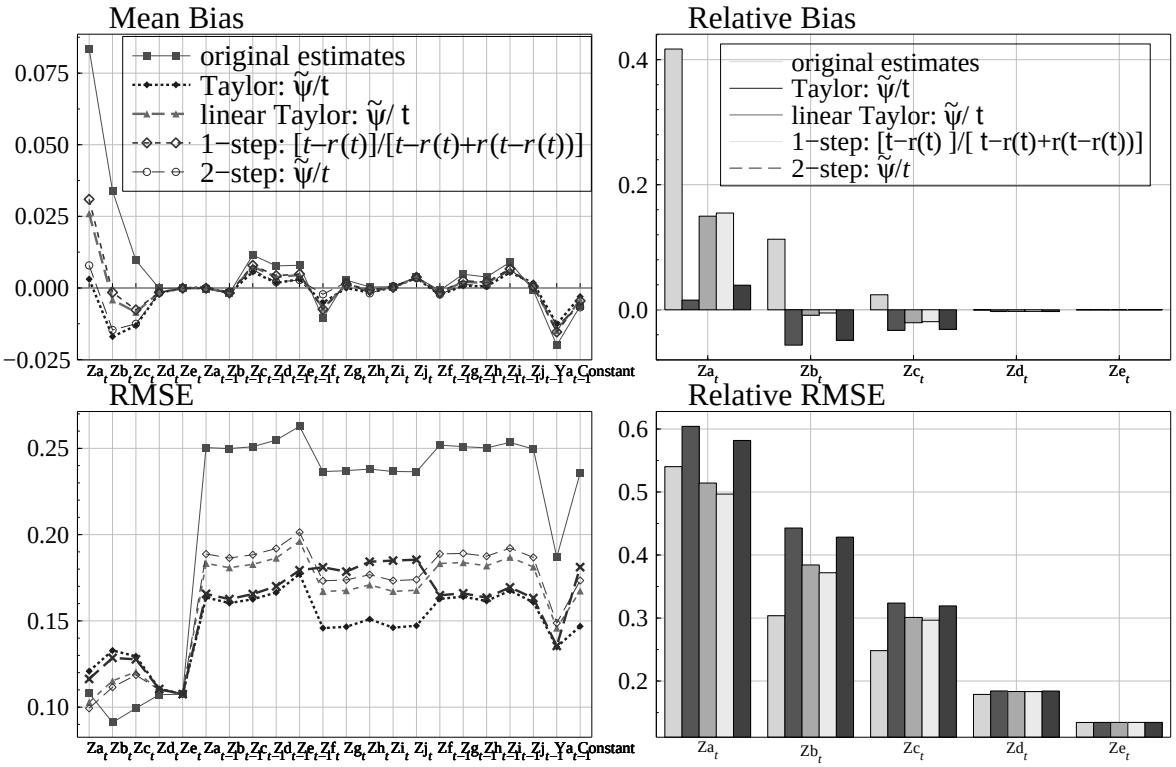

Figure 7 Biases and RMSEs of the original and adjusted estimates.

Figure 7 shows various comparisons of the biases and RMSEs for the original and adjusted estimates in the JEDC Monte Carlo experiments. We evaluated both 1-step and 2-step iterates and a Taylor approximation, in combination with the two main choices for the denominator

in (10). Figure 8 plots the comparative conditional distributions for the 2-step approximation using (11). Much of the bias for the retained relevant variables is corrected without too great an increase in their RMSEs. The next section discusses the impact of the various bias corrections on the retained irrelevant variables.

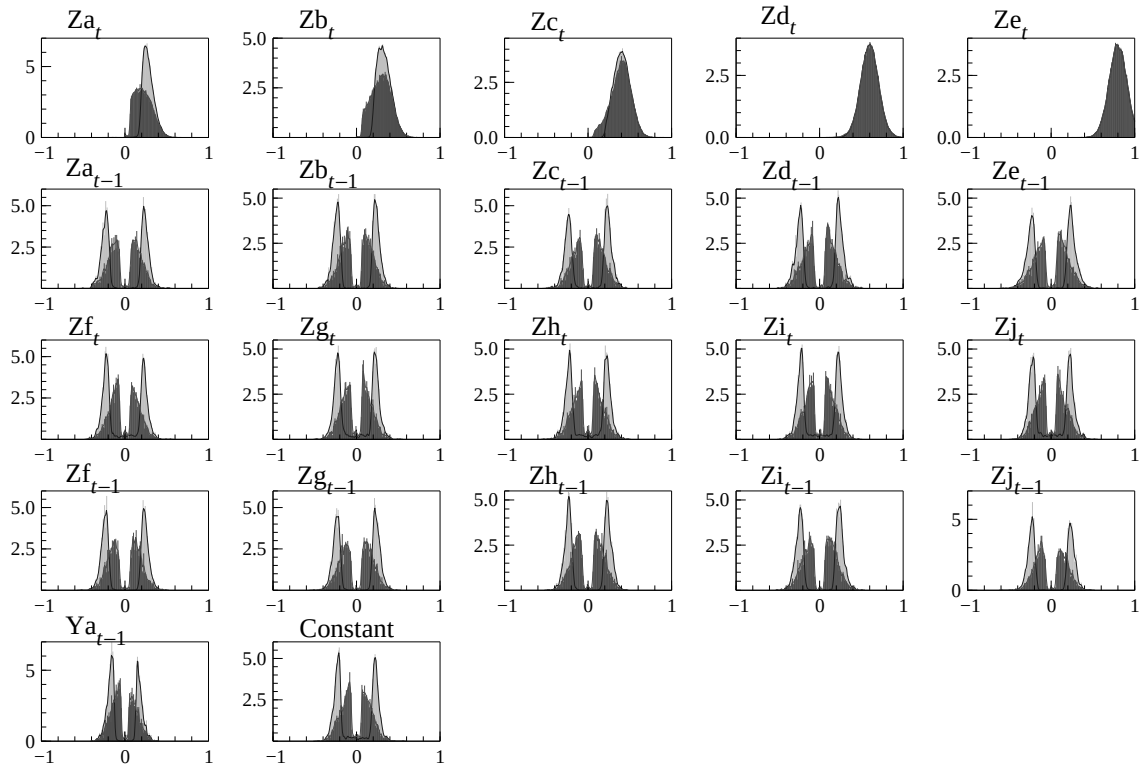

Figure 8 Conditional distributions of the original and 2-step adjusted estimates.

4.4 Bias correction impact on retained irrelevant variables

A surprising effect of the bias correction is its impact on the distributions of the coefficients of the retained irrelevant variables in the conditional model. While their unbiasedness is essentially unaffected by construction, figures 7 and 8 show the considerable reductions in their RMSEs, both from the iterated correction and a 'double correction' designed to substantially reduce RMSEs. As can be seen, reductions of around 25%–40% result, so the bias correction is beneficial from that aspect as well.

4.5 Bias correction impact on unconditional coefficients estimates

The analysis of the truncated t-values in (5) also gives useful insights into the biases of unconditional t-values (i.e., including zeros). Since:

$$\begin{aligned} \mathsf{E}\left[\mathsf{t}_{\widehat{\beta}}\,|\psi, c_\alpha\right] &= \mathsf{Pr}\left[|\mathsf{t}_{\widehat{\beta}}| > c_\alpha \middle| \psi\right] \mathsf{E}\left[\mathsf{t}_{\widehat{\beta}} \middle| |\mathsf{t}_{\widehat{\beta}}| > c_\alpha; \psi\right] \\ &= \left[1 - \Phi\left(c_\alpha - \psi\right) + \Phi\left(-c_\alpha - \psi\right)\right]\psi + \phi\left(c_\alpha - \psi\right) - \phi\left(-c_\alpha - \psi\right), \end{aligned}$$

the resulting bias has the opposite sign to the conditional bias $r(\psi, c_\alpha)$. The illustration in figure 9 for $c_\alpha = 2$ confirms the simulation results reported in table 2. Correcting for the bias in $\mathsf{E}[\mathsf{t}_{\widehat{\beta}}| \; |\mathsf{t}_{\widehat{\beta}}| > c_\alpha; \psi]$ therefore slightly increases the downward bias in the unconditional estimates of the coefficients. This, however, is not a major practical concern.

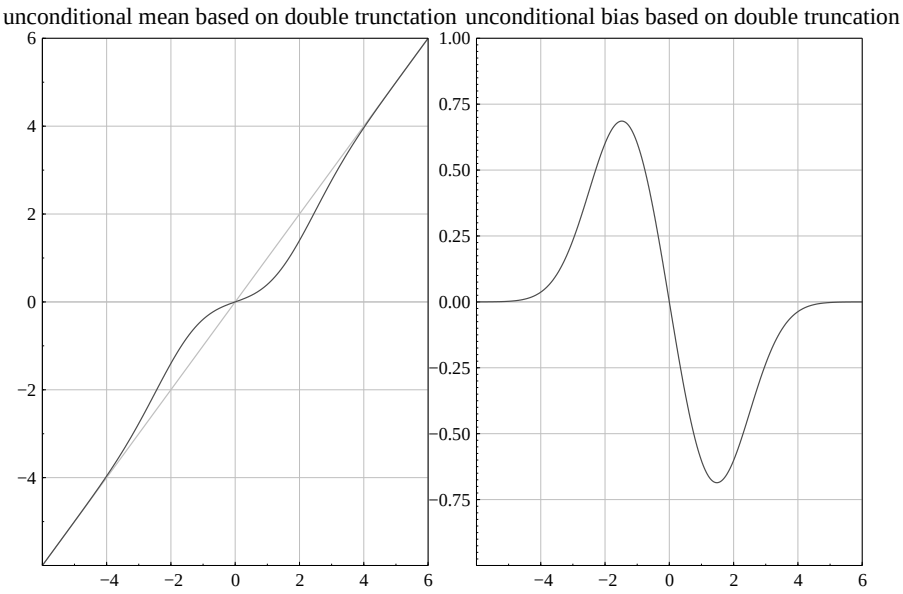

Figure 9 Bias of the unconditional coefficient estimates (including zeros).

4.6 Policy analysis implications

An important distinction must be made between the estimated orthogonalized congruent model, which delivers $\widehat{\beta}$ with near unbiased estimates of $\sigma_{\widehat{\beta}}$ and σ_ϵ, and the bias-corrected coefficients, $\overline{\beta}$ and their associated statistics. Re-calculating residuals, standard errors etc., using $\overline{\beta}$ has no theoretical foundation, and could introduce significant mis-specification diagnostics. However, for some purposes, near unbiased parameter estimates like $\overline{\beta}$ may be useful.

The main application for unbiased estimates is probably policy analysis, particularly estimating policy derivatives. If a variable is incorrectly excluded because of chance insignificance in a given sample, then a policy avenue may be missed, inducing opportunity costs: the Liberal strategy again seems preferred on this criterion. If an irrelevant variable is wrongly included, incorrect decisions could result—but the smaller coefficient induced by the bias-correction formula will decrease the chance of an 'over-reaction' by policy makers. In other words, when $\psi = 0$, although no bias arises, the bias correction reduces the anticipated impact of adventitiously-selected irrelevant variables, so the correction seems uniformly beneficial. If a variable is correctly included but with a biased coefficient, again incorrect policy could result, so unbiased coefficients seem valuable in this arena. Overall, therefore, given the bias correction procedure, the costs of missing relevant variables seem higher than those from adventitious significance.

5 Non-expert mode

To extend the practical realm of automatic selection, we have developed a mode where a non-expert user simply specifies the appropriate functions of the regressand and the basic regressors, then *PcGets* creates the GUM and selects a model. Thus, the input to 'quick modeller' is just the list of 'basic variables', $y_t, x_{1,t}, \ldots, x_{N,t}$. The maximum lag length is set from the data frequency given the sample size; the levels equation is estimated unrestrictedly; and the congruence of the resulting GUM is checked (Wooldridge, 1999, establishes the validity of the mis-specification tests for integrated data). Next, the *PcGive* unit-root test is computed (see e.g., Banerjee and Hendry, 1992, and Ericsson and MacKinnon, 2002), and the variables transformed to differences and any cointegration combination. Finally, that I(0) representation is re-estimated, and the usual procedures for selecting a parsimonious undominated model implemented.

On the data set from Hendry and Ericsson (1991), and just inputting the UK M1 variables $m - p$, y, Δp, R_{net}, with a maximum lag of 2 (as the data are seasonally adjusted) using the Liberal strategy with outlier correction, 'quick modeller' selects an improvement on their reported equation with $\widehat{\sigma} = 1.22\%$ from 4 variables and an impulse dummy, as against their 1.31% from 5 variables (and in seconds as against a large modelling time input!).

The main *caveats* of the present implementation are that (a) the user has to choose the functional form, which sits uneasily with the notion that they are not experts; and (b) the initial levels representation of the regressor set potentially departs strongly from orthogonality. Problem (a) can be addressed by an approach like Perez-Amaral *et al.* (2003, 2004) in their program RETINA, which automatically generates many non-linear transformations. We consider that a productive avenue to explore, especially given their finding that *PcGets* performs well on their empirical problem when the general GUM is used. The results in Castle (2004) illustrate the outcomes in some simulations, and perhaps suggest using a 'super-conservative' strategy for selecting the non-linear components, with a Liberal for the linear.

Problem (b) is considered in the next section, where we analyze the effects of near-collinearity on the selection properties of *PcGets*.

5.1 Collinearity

Perfect collinearity denotes an exact linear dependence between variables; perfect orthogonality denotes no linear dependencies; but any intermediate state depends on which 'version' of a model is inspected, as collinearity is not invariant under linear transforms. For example, inter-variable correlations above 0.99 can easily arise in systems with unit roots and drift, but there is little difficulty determining the relevance of variables as the DGP is isomorphic to an equilibrium-correction model. Conversely, when a conditional regression model is the DGP, for regressors from a bivariate normal distribution with a correlation of 0.99, there is almost

no hope of determining which variables matter. Moreover, orthogonal representations then just reveal that the variance of one transformed component is close to zero. The issue, therefore, is not just one of the capabilities of any selection algorithm, but of the properties of the DGP itself.

In empirical applications, the impact of collinearity will usually be manifest in the number of different terminal models located for encompassing comparisons. Since highly correlated variables may substitute for one another, the selection process can lead to set of final models where none clearly dominates all the others.[7] This information could still help guide selection when subject-matter knowledge is available.

The effects of collinearity on the selection properties of *PcGets* are now illustrated by a variation of the Monte Carlo experiments in Krolzig and Hendry (2001) (denoted as JEDC design in table 1). The DGP is a Gaussian regression model, where the strongly-exogenous variables are independent Gaussian AR(1) processes:

$$\begin{array}{rclrcl} y_t & = & \sum_{j=1}^{5} \beta_{j,0} z_{j,t} + u_t, & u_t & \sim & \mathsf{IN}\left[0, \sigma_u\right], \\ \mathbf{z}_t & = & \rho \mathbf{z}_{t-1} + \mathbf{v}_t, & \mathbf{v}_t & \sim & \mathsf{IN}_{10}\left[\mathbf{0}, \left(1-\rho^2\right)\sigma_v^2 \mathbf{I}_{10}\right] \text{ for } t = 1, \ldots, T, \end{array} \tag{13}$$

where $\mathbf{z}_t$ collects both the DGP and nuisance variables. The parameterization of the DGP is $\beta_{1,0} = 0.2$, $\beta_{2,0} = 0.3$, $\beta_{3,0} = 0.4$, $\beta_{4,0} = 0.6$, $\beta_{5,0} = 0.8$, with $|\rho| < 1$ and $\sigma_u^2 = \sigma_v^2 = 1$. The population t-value associated with regressor j is given by:

$$\mathsf{t}_j = \beta_j \sqrt{T} \frac{\sigma_z}{\sigma_u} = \beta_j \sqrt{T} \frac{\sqrt{1-\rho^2}\sigma_v}{\sqrt{1-\rho^2}\sigma_u} = \beta_j \sqrt{T}. \tag{14}$$

The DGP is designed to ensure invariant population t-values with increasing ρ, even though the entire second-moment matrix is becoming singular, and the data increasingly close to non-stationarity. For $T = 100$, the non-zero population t-values are therefore $2, 3, 4, 6, 8$, independently of ρ, although the approximation in (14) that $\sigma_z = \sigma_v$ becomes increasingly poor as ρ increases if $\mathbf{z}_0 = 0$.

The GUM is an *ADL*$(1, \ldots, 1)$ model, which includes as non-DGP variables the strongly-exogenous regressors $z_{6,t}, \ldots, z_{10,t}$ and the first lags of every variable, so $N = 22$, although k is only 5 from the first equation in (13):

$$y_t = \pi_{0,0} + \pi_{0,1} y_{t-1} + \sum_{k=1}^{10} \sum_{i=0}^{1} \pi_{k,i} z_{k,t-i} + w_t, \quad w_t \sim \mathsf{IN}\left[0, \sigma_w^2\right]. \tag{15}$$

In an alternative experiment, we also consider the orthogonal representation of (15) as a GUM:

$$y_t = \pi_{0,0} + \pi_{0,1} y_{t-1} + \sum_{k=1}^{10} \pi_k z_{k,t} + \sum_{k=1}^{10} \gamma_k \left(\rho z_{k,t} - z_{k,t-1}\right) + w_t, \quad w_t \sim \mathsf{IN}\left[0, \sigma_w^2\right]. \tag{16}$$

[7] An indirect cost of collinearity is that the t-values in the GUM are poor indicators of the importance of variables, so the initial ordered $\mathsf{t}^2_{(i)}$ cannot guide the selection of candidate variables for elimination.

In (15) as in (16), 17 of 22 regressors are 'nuisance'. The sample size T is just 100, and the number of replications M is 1000. In a third experiment, using (16), the sample size is adjusted for the time dependence of the regressors, so $T(\rho) = 100(1-\rho^2)^{-1}$.

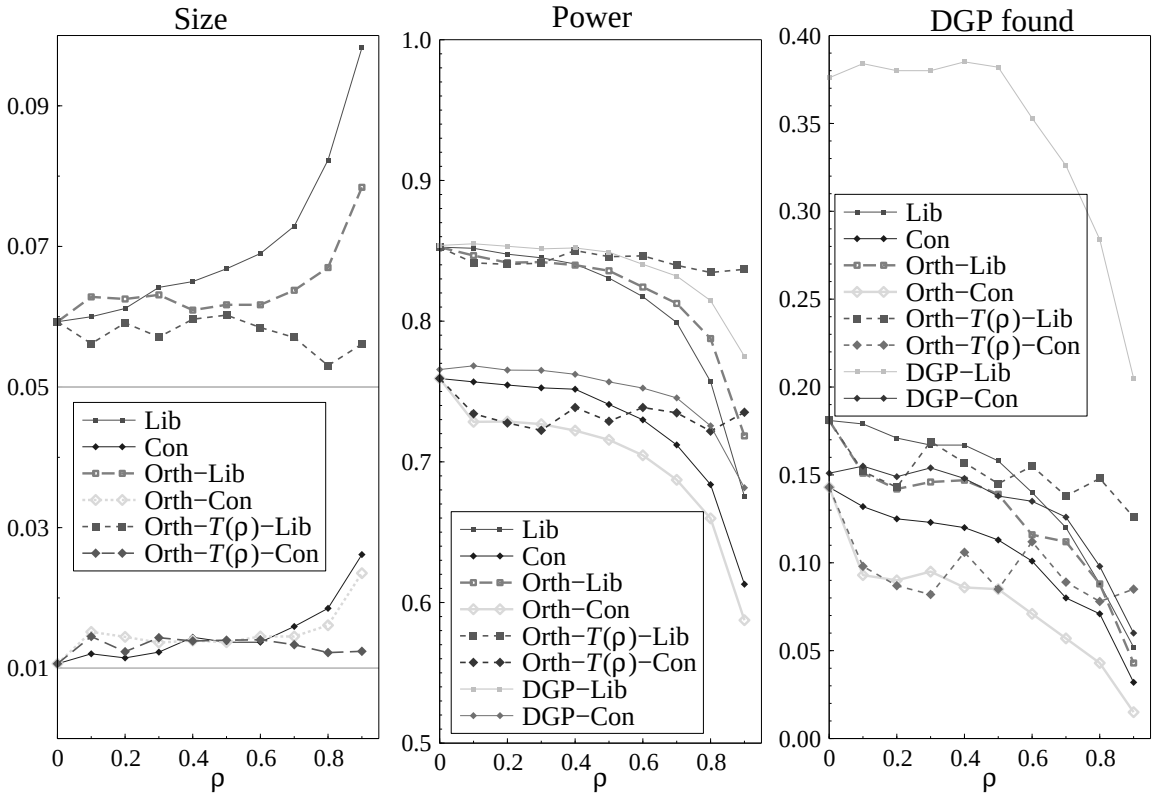

Figure 10 Selection properties of *PcGets* for varying ρ.

The Monte Carlo results are summarized in figure 10 which plots the null rejection frequency ('size'), the correct rejection frequency under the alternative ('power'), and the probability of finding the DGP with *PcGets* when commencing from (i) GUM (15) with $T = 100$; (ii) GUM (16) with $T = 100$; (iii) GUM (16) with $T(\rho)$; and (iv) from the DGP (13) with $T = 100$. The first and fourth experiments illustrate the effects of collinearity, namely a loss of power and (in (i), increasing size), as ρ moves towards unity. Starting from an orthogonalized GUM helps stabilize size and power, but not completely. However, size and power become ρ-invariant if the sample size is adjusted as in (iii), confirming that information loss is the problem, not just intercorrelations. The probability of locating the DGP falls sharply as ρ increases in all cases except (iii), where again it is stabilized. Except for the Liberal strategy commencing from the DGP for low collinearity, where there is a much higher probability of retaining the DGP, the costs of search in the other experiments are low compared to the costs of inference.

6 Theory restrictions

The general formulation of the model under analysis will almost always be suggested by economic theory (or, more generally, subject matter considerations), even if the specific implementation must also depend on institutions, historical contingencies, data availability, its measurement accuracy, and previous empirical evidence—hopefully encompassed by the GUM of the new specification. Parametric constraints that deliver a specific numerical combination of variables, such that the remainder ought then to be insignificant, are easily imposed.

Sign restriction information must first by tested in the GUM, since if it is rejected there, then no feasible congruent model satisfies that sign constraint. Thus, the researcher needs to re-think the theory and/or re-specify the GUM itself. However, if the sign restriction is accepted at the pre-assigned significance level, then it can be imposed during simplification as a constraint, precisely like the diagnostic tests, where a violation simply terminates a search path as inadmissible. Thus, the final model is guaranteed to both satisfy the constraints and be a valid, congruent reduction, that will parsimoniously encompass the GUM. Even so, one should always run the program unrestrictedly to check if the constraints hold anyway: if so, the best model has been found. If not, then it is worth recording the costs of the constraints even if an acceptable model satisfying them has been located. The advantages of such sign impositions are coherence with the theory and improved selection test power for a given size.

If there are competing theory models of a given variable, *PcGets* could be used to select the 'best representative' of each, conditional on the specifications of their information sets and their entailed GUMs. Then encompassing tests could be used to determine the relative performance of the selected candidates. This would automate the type of approach adopted by (e.g.) Bean (1981) and Ahumada (1985), and ensure an objective and reproducible outcome.

7 Tackling apparently intractable problems

We briefly discuss three problems that at first sight seem intractable, but in fact can be tackled by a *Gets* approach. The first is model selection when confronting more regressors than observations; the second is perfectly collinear regressors; and the third is the selection of simultaneous equations models despite the absence of any prior identification information.

7.1 Model selection confronting too many regressors

We have several times been asked about this 'singular case' by investigators who have had to confront an 'excess variables' problem when modelling, namely $N > T$. Some researchers seem to have tried many small blocks of variables in their search for significant regressors, but we doubt such a procedure will be effective, and have instead developed a variant of *Gets*:

see Hendry and Krolzig (2004a) for details, who apply that idea to selecting non-linear representations, and Hendry, Johansen and Santos (2004), who examine regressions with indicator dummies for every observation.

To illustrate, let y_t be an observed random variable where:

$$y_t \sim \mathsf{IN}\left[\mu + \sum_{i=1}^{k} \beta_i x_{i,t}, \sigma_v^2\right],$$

for $t = 1, \ldots, T$, where there are $k << T$ parameters of interest $\{\beta_i\}$. However, an investigator is uncertain about the specification, and wishes to regress y_t on $\{\mu, x_{i,t}, i = 1, \ldots, N > T\}$. A perfect fit will result if such a regression is tried, so nothing will be learned.

Consider adding fewer than half the variables, randomly selected (e.g., a third of the $x_{i,t}$ if $N/3 << T$) together with the intercept. A first-stage model is selected from this GUM using a relatively loose significance level (to compensate for the anticipated poor fit from omitted relevant variables), and that terminal model is stored (autocorrelation and heteroskedastic consistent estimated standard errors may be needed at these intermediate stages: see e.g., Andrews, 1991). Now enter the next third (1, $x_{i,t},\ i = N/3 + 1, \ldots, 2N/3$) and repeat, again storing the result. Finally, search the third set. Repeat for alternative selections, in every case storing which regressors are retained. Let $M < N$ denote the union of the variables in the terminal models. If $M > T$, repeat the process this far from partitions of the M variables, but at a more stringent significance level; and so on. Once $M << T$, formulate a new GUM where all these significant selected variables from the terminal models are combined, and re-select by a usual *PcGets* approach. Clearly, the 'perfect fit' problem does not arise. Moreover, αN irrelevant variables will be retained on average for a significance level α, so at the final stages α can be set to trade off adventitious significance against omitting relevant regressors.

7.2 Perfect collinearity

Despite the problems with near collinearity discussed above, perfect collinearity can be handled through the multi-path search process. To illustrate the general principles, economic theory is often unable to specify which lag transforms actually determine y_t: for example, when the maximum lag is known to be unity, only a subset of the level (z_t), its lag (z_{t-1}), difference (Δz_t), distributed lag (z_t, z_{t-1}), or moving average ($\overline{z}_t = z_t + z_{t-1}$) may be relevant. Despite the perfect collinearity, all four variables just noted can be entered, and if only multi-path searches are used, the correct combination can be selected by *PcGets*, subject to the usual sampling fluctuations. The following DGP is postulated:

$$y_t = \beta_0 + \beta_1 1_{\{z_t\}} z_t + \beta_2 1_{\{z_{t-1}\}} z_{t-1} + \beta_3 1_{\{\overline{z}_t\}} \overline{z}_t + \beta_4 1_{\{\Delta z_t\}} \Delta z_t + v_t \tag{17}$$

where $v_t \sim \mathsf{IN}\left[0, \sigma_v^2\right]$ and $1_{\{j\}}$ are indicator variables that take the value unity if the regressor in question enters the DGP, and are zero otherwise, with at most two being non-zero. The

investigator does not know which $1_{\{j\}}$ are unity, so formulates the GUM:

$$y_t = \gamma_0 + \gamma_1 z_t + \gamma_2 z_{t-1} + \gamma_3 \overline{z}_t + \gamma_4 \Delta z_t + u_t. \tag{18}$$

The regressors are perfectly collinear, and conventional inversion routines will arbitrarily assign singularity to some subset, usually that with the smallest second moments.

With multi-path searches, consider commencing a sequence which deletes each variable in turn, then every next variable, and so on: e.g., first drop Δz_t, which still leaves a collinear set; then $\overline{z}_t$ which is now a non-collinear set, and search; next, starting afresh from (18), drop z_t then z_{t-1}, and so on. Once a non-collinear set results, the usual algorithm can operate. If, for example, only $1_{\{\overline{z}_t\}} = 1$, then such a terminal model will be explored on some path, so should be selected if the power is adequate, and will parsimoniously dominate the other selections (e.g., on the Schwarz criterion). There are 11 possible models, including the null, of which only 6 are distinct. We conducted five one-off experiments on artificial data where, in turn, y_t depended on: (a) $\overline{z}_t$; (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}, $\overline{z}_t$ and Δz_t were entered as regressors in the GUM (18). *PcGets* is not yet programmed to follow all possible paths, but as a partial implementation by hand (namely searching once a non-collinear set was imposed, commencing from every path and using *SIC* to select between undominated choices) gave the correct answers in (a)–(d), but in (e), it selected the more orthogonal representation, $\overline{z}_t$ and Δz_t, which is equivalent. Pre-search tests designed to reduce the computational burden of path exploration would arbitrarily eliminate whatever variables the inversion routine treated as redundant, and possibly preclude finding a useful representation.

7.3 Simultaneous equations selection

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 could be lagged endogenous, as in a VAR): more details are provided in Hendry and Krolzig (2004a). What matters is the identification of the DGP representation: if that is identified, it can be found as a reduction, even if the identifying restrictions are not known; if it is not identified, then the program will revert to the 'reduced form'.

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 are constrained to be congruent as well. If congruence is accepted for the unrestricted representation, a parsimonious version of that system is selected by the usual *PcGets* approach, checking 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.

The key step is that each endogenous variable is now added *seriatim* to every other equation concomitant with dropping any regressor that is also present in that added variable's equation, checking 'just-identification' by the rank condition as implemented in *PcGive* by Hendry, Neale and Srba (1988). Then one checks what further reductions can be achieved in the regressor set of the first equation. There are two main possibilities in any equation for reductions in the regressor set:
(1) no additional reductions are found;
(2) one or more further reductions occur.
In the first case, the postulated equation is just-identified, so the choice just reverts to the system (reduced-form) equation. In the second case, the proposed equation must be over-identified, since there are several eliminated right-hand side regressors, and these must be a determinant of the added left-hand side variable by occurring significantly in its equation, thereby identifying that endogenous effect. The resulting restrictions are testable (e.g., by the test in Sargan, 1964).

Each equation is considered in turn in this instrumental variables approach. Since the rank condition is imposed as a constraint, the 'same equation' is not included twice, and the current 'partial structure' is always fully identified at every step. 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 (namely, 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. That states, but does not resolve, the problem which lies in available information, not the performance of any selection approach. As noted above, the choice of instruments can be made by *PcGets*, 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 even quantile regression), and to non-linear equations *inter alia* seem feasible in principle.

These three previously 'intractable' cases illustrate how a new tool can yield new insights: one might have suspected that regressions with $N > T$, or perfect collinearity, or simultaneous modelling in the absence of prior information, were all insoluble, but multi-path searches can resolve the choice of model in each case.

8 Conclusion

Model selection is an important part of a progressive research strategy, and itself is progressing rapidly. The automatic selection algorithm in *PcGets* provides a consistent selection like *SIC*,

but in finite samples, both ensures a congruent model and can out-perform in important special cases without *ad hoc* adjustments. Recent improvements have stabilized the null rejection frequency relative to the desired nominal significance level, and the power relative to that feasible when the DGP is the initial specification. The power performance on recent Monte Carlo experiments in orthogonal models is close to the upper bound of a scalar t-test at the given non-centrality from a known t-distribution.

Search *per se* does not seem to impose serious additional costs over those of inference (nor does mis-specification testing, as that is needed even when commencing from the DGP specification). The results confirm that 'pre-test' biases arise from simplifying the DGP, not from searching for it in an over-parameterized representation. The equation standard error is found within $\pm 5\%$ of the population value, depending on the strategy adopted, so *PcGets* has no substantive tendency to 'overfit'. Depending on the state of nature, *PcGets* can even have a higher probability of finding the DGP starting from the GUM using the Liberal strategy, than a researcher commencing from the DGP but selecting by the Conservative strategy. Such findings would have seemed astonishing in the aftermath of Lovell (1983), who reported that 'data mining' had a low probability of success, and both shows the progress achieved and serves to emphasize the importance of the choice of strategy for the underlying selection problem. Obtaining nearly unbiased estimates of the DGP parameters in selected models, with estimated standard errors that are close to those that would be reported for sampling standard deviations in the estimated DGP, might surprise even more. The key to such performance seems to lie in using a search algorithm that commences from a congruent orthogonal representation that nests the DGP, explores all feasible paths while retaining congruence for a given 'size' per candidate variable, and terminates with a dominant encompassing selection.

Non-orthogonal designs remain problematic, in that they can induce higher costs of search as well as of inference, and remain an area where expert knowledge will continue to prove valuable. Nevertheless, we have added a 'quick modeller' option for non-expert users, which initial experience suggests is able to outperform all but expert econometricians in selecting from an initial dynamic GUM that is possibly I(1). In models with many potential candidate variables, automatic selection is invaluable.

So what lies ahead? Certainly, the theoretical context assumed above of regression analysis with orthogonal strongly-exogenous regressors is far too simple to characterize real-world econometrics. Empirical researchers confront non-normal, mis-measured data, on evolving non-stationary dynamic and high-dimensional economies, with at best weakly exogenous, intercorrelated, conditioning variables. At a practical level, *Gets* is applicable to systems, such as vector autoregressions (see Krolzig, 2001, 2003a, 2003b), and to endogenous regressors when sufficient valid instruments exist. Moreover, Omtzig (2002) and Kurcewicz and Mycielski (2003) have proposed algorithms for automatic selection of cointegration vectors; and *Gets* approaches seem just as powerful a tool on cross-section problems, as demonstrated by Hoover

and Perez (2004) and Hendry and Krolzig (2004c). Even though bias corrections will be imperfect in many settings, simulation studies can reveal how well they do (or do not) perform. While it is usually infeasible to analytically derive either the conditional or unconditional distributions of the finally-selected model's parameter estimates, in some special cases under specific assumptions, important advances have been achieved: see *inter alia*, Pötscher (1991) and Leeb and Pötscher (2000).

As sketched above, selection with more candidate regressors than observations ($N > T$) is feasible when the DGP is estimable (with $k << T$ regressors). Simultaneous equations systems also pose less than insurmountable problems using a tool like *PcGets*. Developments like those in RETINA for creating and selecting functional form bode well, as do the ideas in Phillips (1995, 1996, 2003) for forecasting. Since applied researchers must often devote considerable effort to developing empirical representations, such labour-saving devices have much to offer. Automatic model selection could eventually replace 'hands-on' empirical research, but seems more likely to remain a complement to existing methods for the foreseeable future. Even in that role, by truncating the lower tail of the quality distribution, such procedures should improve the average quality of published models and direct researchers away from bad models that might otherwise have been selected. We remain confident that further developments will continue to improve the performance of, and widen the scope of application for, automatic modelling procedures.

Department of Economics, University of Oxford
Department of Economics, University of Kent

References

Ahumada, H. (1985). An encompassing test of two models of the balance of trade for Argentina. *Oxford Bulletin of Economics and Statistics*, **47**, 51–70.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**, 817–858.

Banerjee, A., and Hendry, D. F. (1992). Testing integration and cointegration: An overview. *Oxford Bulletin of Economics and Statistics*, **54**, 225–255.

Bean, C. R. (1981). An econometric model of investment in the United Kingdom. *Economic Journal*, **91**, 106–121.

Bontemps, C., and Mizon, G. E. (2003). Congruence and encompassing. In Stigum (2003), pp. 354–378.

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). Evaluating PcGets and RETINA as automatic model selection algorithms. Unpublished typescript, Economics Department, University of Oxford.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Doornik, J. A. (1999). *Object-Oriented Matrix Programming using Ox*. London: Timberlake Consultants Press. 3rd edition.

Ericsson, N. R., and MacKinnon, J. G. (2002). Distributions of error correction tests for cointegration. *Econometrics Journal*, **5**, 285–318.

Haldrup, N., van Dijk, H., and Hendry, D. F. (eds.)(2003). *Model Selection and Evaluation*. *Oxford Bulletin of Economics and Statistics*: 65.

Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1996). Judging instrument relevance in instrumental variables estimation. *International Economic Review*, **37**, 283–298.

Hannan, E. J., and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, **B, 41**, 190–195.

Hansen, B. E. (1999). Discussion of 'Data mining reconsidered'. *Econometrics Journal*, **2**, 26–40.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.

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. (1999). Improving on 'Data mining reconsidered' by K.D. Hoover and S.J. Perez. *Econometrics Journal*, **2**, 202–219.

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). Resolving three 'intractable' problems using a Gets

approach. Unpublished paper, Economics Department, University of Oxford.

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., 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.

Hoover, K. D., and Perez, S. J. (2004). Truth and robustness in cross-country growth regressions. *Oxford Bulletin of Economics and Statistics*, forthcoming.

Hurwicz, L. (1950). Least squares bias in time series. In Koopmans, T. C. (ed.), *Statistical Inference in Dynamic Economic Models*, No. 10 in Cowles Commission Monograph, Ch. 15. New York: John Wiley & Sons.

Johnson, N. L., and Kotz, S. (1970). *Continuous Univariate Distributions–1*. New York: John Wiley.

Judge, G. G., and Bock, M. E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. Amsterdam: North Holland Publishing Company.

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**, 769–801.

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.

Leeb, H., and Pötscher, B. M. (2000). The finite-sample distribution of post-model-selection estimators, and uniform versus non-uniform approximations. Economics working paper, 0004001, Department of Economics, Washington University.

Lovell, M. C. (1983). Data mining. *Review of Economics and Statistics*, **65**, 1–12.

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. (2004). Weak identification of forward-looking models in monetary economics. *Oxford Bulletin of Economics and Statistics*, forthcoming.

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. (1995). Automated forecasts of Asia-Pacific economic activity. *Asia-Pacific Economic Review*, **1**, 92–102.

Phillips, P. C. B. (1996). Econometric model determination. *Econometrica*, **64**, 763–812.

Phillips, P. C. B. (2003). Laws and limits of econometrics. *Economic Journal*, **113**, C26–C52.

Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory*, **7**, 163–185.

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.

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.

White, H. (1980). Non-linear regression on cross-section data. *Econometrica*, **48**, 721–746.

White, H. (1990). A consistent model selection. In Granger, C. W. J. (ed.), *Modelling Economic Series*, pp. 369–383. Oxford: Clarendon Press.

Wooldridge, J. M. (1999). Asymptotic properties of some specification tests in linear models

with integrated processes. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*, pp. 366–384. Oxford: Oxford University Press.