

# Evaluating PcGets and RETINA as Automatic Model Selection Algorithms

Jennifer Castle\*  
Nuffield College, Oxford

August 11, 2004

## Abstract

The paper describes RETINA and briefly discusses how it works and what the performance claims are. Their Matlab implementation of the code is explained, then the two programs are compared on the data in Perez-Amaral, Gallo and White (2004), ‘A Comparison of Complementary Automatic Modelling Methods: RETINA and PcGets’, and that used by Hoover and Perez (1999). Monte Carlo simulation results assess the size and power properties of the PcGets model selection process in the presence of non-linear functions.

## 1 Introduction

Advances in automatic model selection procedures have been swift in recent years with impressive results. Despite the controversy surrounding many model selection strategies and the multitude of diverse approaches in the literature, automatic procedures have been developed which show remarkably ‘good’ properties. Two such procedures are PcGets and RETINA. PcGets is based on a general-to-specific search strategy, starting with a general model capturing the underlying characteristics of the data and testing downwards, ensuring validity of the reductions at each stage to result in a congruent parsimonious undominated model. RETINA differs from PcGets in that the general-to-specific methodology is not its main tenet. The model uses a specific-to-general approach whereby variables are added into the model depending on a given criteria. RETINA aims to identify a parsimonious

---

\*I am indebted to David Hendry for helpful comments and suggestions and to Teodosio Perez-Amaral, Giampiero Gallo, Halbert White and Christian Brownlees for the RETINA code. Financial support from the Economic and Social Research Council under grant RES 015 270035 is gratefully acknowledged.

set of variables that are likely to be relevant for predicting out-of-sample. Both approaches have elements in common although they differ in many ways including their objectives, general strategy, selection criteria and the use of sub-samples. Both have a high level of success; automatic model selection procedures bring about gains in terms of power and efficiency that had previously been inconceivable.

The structure of the paper is as follows. Section 2 outlines the RETINA algorithm and discusses its objectives and performance claims. Section 3 briefly outlines the PcGets algorithm and notes the development that PcGets can handle more variables than observations. Section 4 reviews the differences between the two approaches. Section 5 examines various applications including the results from Perez-Amaral, Gallo and White (2004) which compares PcGets and RETINA on cross sectional data and a comparison of the two procedures on time series data including data from Hoover and Perez (1999) and an artificial GUM and DGP as outlined in Krolzig and Hendry (2001). Section 6 presents a range of Monte Carlo experiments assessing the size and power of automatic model selection procedures when there are non-linear functions in the GUM, both when the non-linear functions enter the DGP and when the DGP is linear. Finally section 7 concludes, reviewing the importance of automatic model selection procedures and assessing their use in a non-linear framework.

## 2 RETINA

RETINA (Relevant Transformation of the Inputs Network Approach) is a method of model selection along the lines of neural network models and is designed to identify a parsimonious set of regressors to predict out-of-sample. The model is outlined in Perez-Amaral, Gallo and White (2003). RETINA is useful when the functional form of the conditional mean of the dependent variable is unknown, testing for non-linearities and interaction effects within the model procedure. Concavity of the likelihood is achieved by imposing linearity in the parameters. The method relies on a sub-sample cross validation scheme to ensure parsimony. Rather than doing an exhaustive model search, which would require the evaluation of  $2^m$  models, the number of models is narrowed down by including variables sequentially in rank order. Collinearity is controlled by ensuring the  $R^2$  between the included variables and the additional variable lies below a specified threshold parameter,  $\lambda$ . This section describes RETINA and briefly discusses how it works.

## 2.1 The Selection Algorithm

There are 4 main stages in the RETINA algorithm, including data building and sorting, isolating a candidate model, the search strategy and model selection. Initially, a set of potentially relevant variables,  $\mathbf{X}$ , that are thought to contain information about the conditional mean of the dependent variable,  $y$ , are identified and labelled “level 0 transforms”. A set of transformed variables,  $\zeta(\mathbf{X}) = \{W_1, \dots, W_M\}$ , are generated and denoted “level 1 transforms”. There are numerous possible transformations which capture both non-linearities and interactions and RETINA uses the transformations given by  $\zeta(\mathbf{X}) = X_{i,h}^\alpha X_{i,j}^\beta$  for  $\alpha, \beta = -1, 0, 1$ , which results in a set of  $M$  potential predictors, i.e. squares, inverses, squared inverses, cross products, cross inverses and ratios.<sup>1</sup> Note that further iterations of the transformations can be implemented and appended to the level 0 and level 1 transforms.

The sample is divided into 3 disjoint sub-samples. The observations in the first sub-sample,  $W_j$  for  $j = 1, \dots, M$  are ranked according to a relevance measure. RETINA uses the sample correlation with  $y$ , denoted  $|\hat{\rho}_j|$ . Starting with a constant and the first variable,  $W_1$ , which has the highest absolute sample correlation, successive variables are added in their ranked order until the  $R^2$  of the regression of the last added variable on the subset already included lies below a specified threshold parameter,  $\lambda$  ( $0 \leq \lambda \leq 1$ ), i.e. include  $W_j$  if  $R_j^2 \leq \lambda_p$  where  $R_j^2$  is derived from the regression of  $W_{j,i} = \beta_0 + \sum_{k=1}^{j-1} \beta_k W_{k,i} + \epsilon_i$  for  $i = 1, \dots, N$ . This will result in a set of variables selected on the basis of  $\lambda_p$ . Repeating the process for a grid of values for  $\lambda$ , a set of models will be retained, denoted by  $\zeta_p(\mathbf{X})$  for  $p = 1, \dots, \nu$ , where  $p$  determines the value of  $\lambda$ . The  $\nu$  models are estimated in sub-sample 1 and cross-validated in sub-sample 2 using Mean Square Prediction Error to determine the ‘candidate’ model. The  $\lambda$  varies in increments of approximately 0.1, resulting in  $\nu \approx 9$ . The optimal  $\lambda^*$  is chosen on the basis of MSPE and the model given by  $\lambda^*$  is denoted the local best model (i.e. has the lowest MSPE in sub-sample 2).

The search strategy aims to select a more parsimonious model by searching over all other possible models. Using the regressors in the local best model, the regressors are added in sequentially, starting with the highest ranked variable and estimated over the second sub-sample, resulting in  $S$  models where  $S$  = the number of regressors in the local best model. These regressors are also ranked based on correlations in the second sub-sample and included sequentially, giving  $S - 1$  models if the rankings differ. The AIC is computed for these models on sub-sample 3 and the resulting preferred model is chosen on the basis of lowest AIC. The process is repeated over different combinations of the sub-samples resulting in 6 repeti-

<sup>1</sup>Some transformations may be excluded in practice because of numerical problems. For example, the inverse of  $X_j$  will be excluded if at least one observation is equal to 0.

tions. A candidate model is chosen for each ordering and the preferred model is chosen on the basis of out-of-sample AIC. Parsimony is achieved by creating new models with a different number of regressors using the regressors of a candidate model derived from MSPE criteria and by judging the models using the AIC on the third sub-sample. The use of disjoint subsamples is a powerful mechanism for controlling size.

RETINA cannot handle more level 1 transformations than observations and so the regressors in the candidate model set depend on the order of inclusion of the transformations. This is determined by absolute correlations with  $y$ . The program can be run using all level 1 transformations including those that are linearly dependent (note that the candidate model will only contain regressors that are linearly independent because of the collinearity index) ensuring that any transformation can be potentially included. In the situation where there are more level 1 transformations than observations a regressor will only be included if it is ranked in the first  $N - 1$  variables. RETINA can also be run just using the level 0 regressors, providing a base model to check the properties under the null for experiments in which the DGP contains no non-linear terms. A third setting for RETINA enables the program to be run excluding all the cross terms, i.e. the level 1 transformations are given by  $\zeta(\mathbf{X}) = X_{i,h}^\alpha X_{i,h}^\beta$  for  $\alpha, \beta = -1, 0, 1$ , thereby removing interactions but allowing for non-linearities in individual regressors.

RETINA is available as Matlab code and full instructions are given as to how to implement the code. The code can be downloaded from [www.ds.unifi.it/ctb/projects.html](http://www.ds.unifi.it/ctb/projects.html).

### 3 PcGets

PcGets is a procedure for automatic model selection which is designed to select a parsimonious undominated representation of a general initial model (denoted the GUM). Krolzig and Hendry (2001) examine the properties of this model selection procedure in a linear framework. The data generating process is found almost as often commencing from a general model as from the DGP itself and false rejection frequencies of null hypotheses can be controlled, correct rejections of alternatives are close to the theoretical upper bound and model selection is consistent. The selected estimates have the appropriate reported standard errors and can be bias corrected if desired. PcGets is described briefly below. See Hendry and Krolzig (2001) for details.

### 3.1 The Selection Algorithm

There are 4 stages in the PcGets algorithm, including estimation and testing of the GUM, the pre-search process, the multipath search procedure and finally post-search evaluation. PcGets uses a general-to-specific strategy in which the GUM is formulated based on theory, previous evidence and institutional knowledge, which is sufficiently general to nest the DGP. The GUM is tested for congruence and pre-search reduction tests are undertaken to remove insignificant variables. Note that this is done at loose significance levels. Then an exhaustive search across all paths is undertaken from each feasible deletion, checking the diagnostics at each reduction to ensure congruence. The resulting models are denoted terminal models, which are tested against their union until a unique undominated congruent model is selected.<sup>2</sup> Finally, two overlapping sub-samples are used to validate the significance of variables retained in the final model.

Various search strategies are available in PcGets including the liberal strategy, the conservative strategy and the expert users strategy. The liberal strategy reduces the non-selection probability whereas the conservative strategy reduces the non-deletion probability. A conservative strategy is recommended if there are highly significant variables amongst many insignificant variables. If there are fewer regressors and the significant variables have smaller  $t$ -values, a liberal strategy is suggested. Both are examined in the empirical examples below.

An important development is that PcGets can handle more variables,  $n$ , than observations,  $T$ . This enables a greater flexibility when examining non-linear models with a larger number of potential regressors. The GUM cannot be estimated initially. Instead, the variables are divided into  $J$  groups (we take  $J = 2$  in this exposition but the analysis is easily generalized) in which the dimensions of the two groups are strictly smaller than  $T$ . PcGets selects the first terminal model from the first subset of variables and likewise the second terminal model from the second subset. The 2 terminal models are combined and used as the GUM in the second estimation stage by PcGets, resulting in a final terminal model. The critical values are adjusted at each stage, starting with fairly loose values to ensure marginally relevant variables are retained, and undertaking the final stage with stringent critical values due to the large number of initial variables that are irrelevant. If the variables are not orthogonal, it is recommended that the variables are ‘crossed-over’. This requires the 2 groups of variables,  $\mathbf{x}_{1,t}$  and  $\mathbf{x}_{2,t}$ , to be partitioned into 2 halves,  $\mathbf{x}_{j,t}^a$  and  $\mathbf{x}_{j,t}^b$  for  $j = 1, 2$ , and cross-paired, giving  ${}_2C_4$  combinations which are denoted the GUMs for the search procedure outlined above.<sup>3</sup>

<sup>2</sup>In the case in which a unique model does not emerge and the models are mutually encompassing and undominated, selection of the preferred model is made on the basis of selection criterion.

<sup>3</sup>The ‘crossed pairs’ would include  $\mathbf{x}_1^a \mathbf{x}_1^b$ ,  $\mathbf{x}_1^a \mathbf{x}_2^a$ ,  $\mathbf{x}_1^a \mathbf{x}_2^b$ ,  $\mathbf{x}_1^b \mathbf{x}_2^a$ ,  $\mathbf{x}_1^b \mathbf{x}_2^b$  and  $\mathbf{x}_2^a \mathbf{x}_2^b$  for  $J = 2$ .

It is also worth noting the distinction between the costs of search and the costs of inference. The costs of search arise because the GUM is overly general and is tested down to find a congruent undominated model. The costs of inference, on the other hand, are unavoidable and are due to non-zero significance levels. They arise because the true specification is unknown and must therefore be tested for, even if the econometrician starts with the DGP. The costs of search are small in comparison to the costs of inference, refuting many criticisms aimed at model selection strategies.

## 4 A comparison of the two approaches

Whilst RETINA and PcGets have many similar characteristics, a brief look at their selection algorithms reveals fundamentally different search strategies. A comparison of the two programs is provided in Perez-Amaral, Gallo and White (2004). This section briefly assesses the differences between the two models, noting potential criticisms and counter arguments.

Firstly, RETINA was developed with the aim of finding a model that has good out-of-sample predictive ability whereas PcGets selects a congruent dominant in-sample model, aiming to locate the DGP which is nested within the GUM. The problem with selecting models for forecasting purposes is that forecasting models require entirely different characteristics to in-sample models. For time series models in particular, a good forecasting model is one that is robust to breaks, see Clements and Hendry (1999). In fact, models that are non-linear in the parameters seem likely to be less robust than linear models because the interactions between variables are also likely to be subject to breaks. Whilst RETINA does not insure against non-stationarities and structural breaks the properties of RETINA as an out-of-sample forecasting tool are shown to be successful in cross sectional applications. The problems of breaks in time series tend to be mitigated in cross sections and so the two approaches are complementary, depending on the problem at hand.

The question regarding which non-linear terms to include depends on the degree of interpretability that the modeller wishes to retain. PcGets specifies the GUM based on the econometrician's knowledge and experience, institutional knowledge, past evidence and economic theory. If the econometrician deems any non-linear terms to be relevant then they should be included in the GUM. On the other hand, RETINA automates this decision, including all transformations that have been specified by the program. There are advantages to both procedures although a degree of economic interpretation is lost in RETINA. As the goal is out-of-sample prediction this property is not as fundamental as it is for PcGets. The development of the 'Quick Modeller' which automatically selects the lag length for the vari-

ables in the GUM has advanced the automatic procedures embodied in PcGets, and automation in specifying non-linear terms in the GUM may well prove to be a fruitful avenue of research. As there are an infinite number of non-linear transforms an arbitrary cut-off point will always be needed. More research into the types of functions that best approximate economic data, such as squashing functions, polynomials, Fourier transforms and many others, is also needed.

RETINA uses a specific-to-general approach in which variables are added into the model sequentially whereas PcGets uses a general-to-specific approach that tests downwards from the GUM. There are numerous problems with the simple-to-general approach, see Hendry (1995) for a discussion. The specific-to-general methodology refutes the ideas behind encompassing and models chosen by RETINA could easily miss the DGP. The theory of model selection for forecasting does not currently shed any light on whether a specific-to-general approach could be plausible for a model aiming to predict out-of-sample, although intuition suggests a general-to-specific approach would be most advantageous.

PcGets performs an exhaustive search ensuring that all paths are checked whereas RETINA uses a selective search determined by correlations. This is particularly relevant when there are more potential regressors than observations as the candidate models will depend on the order of inclusion of variables. Thus, RETINA could miss potentially relevant variables. RETINA does emphasise the problems of collinearity by controlling for the collinearity of variables. Obviously, an orthogonal specification of the GUM is preferable in PcGets but the program can handle collinearity, although there is a loss in power and the size grows. See Hendry and Krolzig (2004) for Monte Carlo evidence on collinearity. For forecasting time series, controlling for collinearity as RETINA does will certainly help due to the interactions of collinearity with breaks.

RETINA uses 3 disjoint sub-samples which are used to cross validate the variables selected. The PcGets approach performs a post selection reliability check using sub-samples to evaluate whether significance is substantive or due to chance. There is a clear size gain in the RETINA procedure due to the use of sub-samples. However, Lynch and Vital-Ahuja (2003) show that the use of sub-samples for model selection delivers no gain over using a smaller nominal size. This is because there is a trade-off between size and power. The power function is highly non-linear and depends on the DGP. Thus, the key question is whether a size reduction for the whole sample which is equivalent to the size gain due to the use of sub-samples will result in an increase or decrease in power. Hendry and Krolzig (2003b) examine the case of non-overlapping sub-samples. With 3 equal partitions as in RETINA, defining  $t_0$  as the full sample  $t$ -value and  $t_j$  as the  $j^{\text{th}}$  sub-sample

$t$ -value:

$$t_0 \simeq 0.58 \sum_{j=1}^3 t_j.$$

Hence, with marginally significant sub-sample  $t$ -values of 2, the full sample  $t$ -value will be 3.5. The use of sub-sample rules in model selection procedures does imply that it will be more difficult to find the DGP. A direct relationship between the average squared  $t$ -value on the full sample and sub-samples is given by:

$$E[t_j^2] \simeq \frac{1}{J} \psi^2$$

where  $\psi$  is the non-centrality parameter. There is a reduction in the information content of the sub-sample  $t$ -tests which Hendry and Krolzig (2003b) refer to as the ‘curse of sub-samples’.

When the data is non-stationary the results will differ. If the breakpoints are known substantial gains could be achieved by analysing sub-sample information. On the other hand, the best in-sample fit in one sub-sample may well differ to that in a second sub-sample due to structural breaks. In this case the use of cross-validation would be more appropriate for post-sample evaluation as opposed to informing model selection.

Whilst PcGets controls the size and power of the test battery, RETINA does not address the question of costs of search and costs of inference. Krolzig and Hendry (2001) examine the properties of PcGets and show that, even for small samples, the size and power of the  $t$ -tests are nearly at their theoretical levels based on a  $t$ -test from the true model. Distinguishing between individual significance levels,  $\eta$ , and the overall significance level of a test battery of  $n$  mis-specification tests, the overall rejection probability under the null is given by  $1 - (1 - \eta)^n$ , which can be controlled. RETINA does not employ any hypothesis testing. As no diagnostics are performed the final model could be mis-specified. Variables are chosen on the basis of out-of-sample MSPE rather than statistical tests, reflecting the differing objectives of the two procedures.

## 5 Applications

This section assesses some evidence on RETINA and PcGets. Firstly, the results of PcGets and RETINA on the cross section data reported in Perez-Amaral, Gallo and White (2004) (hereafter referred to as PGW) are summarised. Two times series applications are also assessed, including data from Hoover and Perez (1999) and an artificial DGP taken from Krolzig and Hendry (2001).



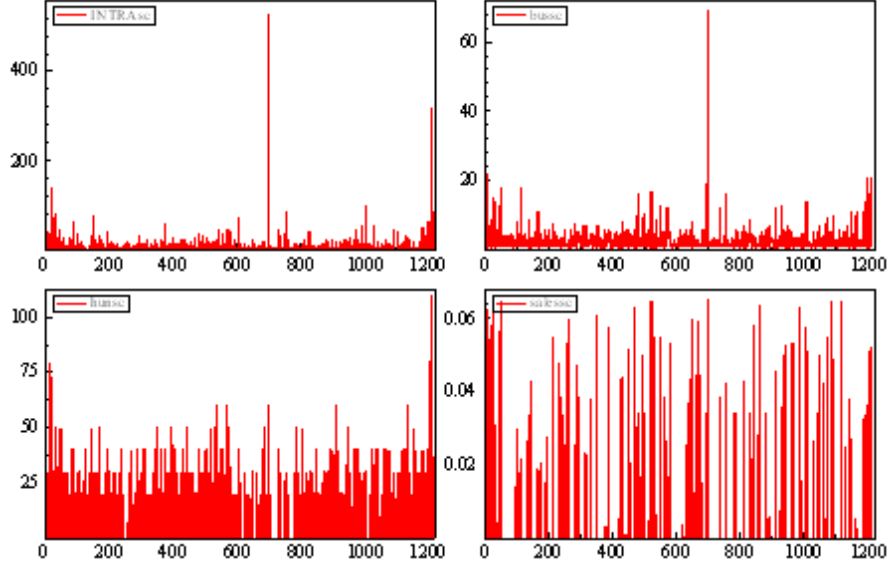


Figure 1: Cross sectional data including *INTRA*, *bus*, *hun* and *sales*.

### 5.1 Cross Sectional Data: The demand for business toll telephone services.

PGW assess the two programs using data from a cross section of US firms regarding the demand for business toll telephone services in 1997. The authors model the duration of intra-lata (Local Access and Transport Area) calls with the explanatory variables including the number of business lines (*bus*), hunting lines (*hun*), sales of the company (*sales*), number of employees working locally (*emt*), total number of employees for the business (*emh*), physical size of the business proxied by ft<sup>2</sup> of the premises (*sqft*) and population of the business area location (*pop*). The sample size is 1217 and the data are initially rescaled to avoid large differences in the magnitudes of the variables. The data are recorded in figures (1) and (2). Observations are randomly assigned to the three sub-samples and the models are assessed on the basis of two criteria; AIC and Cross-Validated Mean Square Prediction Error (CMSPE).<sup>4</sup>

Prior to implementing the non-linear transforms, the linear model was com-

<sup>4</sup>CMSPE is computed by using two of the three sub-samples for estimation and the third for cross-validation. The overall CMSPE is obtained by summing the MSPEs from each of the three rotations.

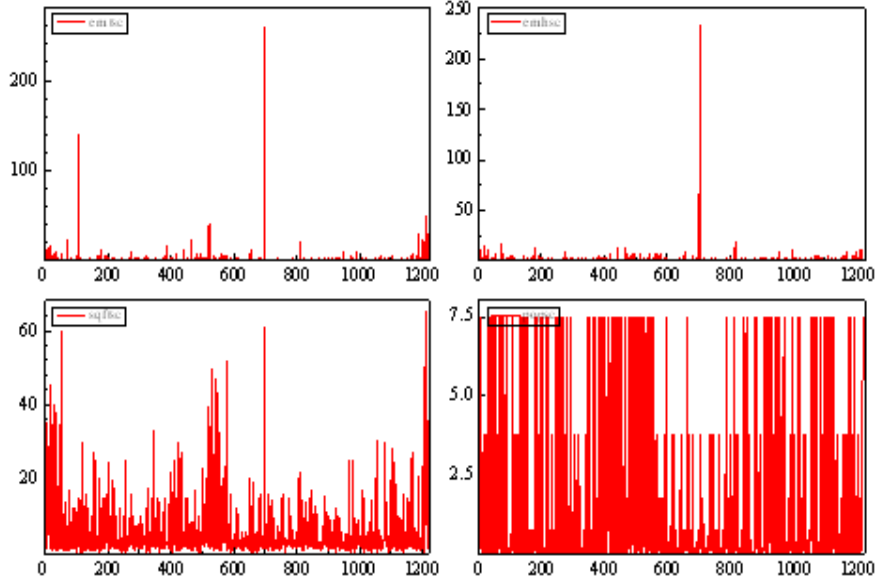


Figure 2: Cross sectional data including *emt*, *emh*, *sqft* and *pop*.

pared. Four models were computed, including a simple linear regression, two PcGets models using the liberal and conservative strategies respectively and RETINA in levels. The GUM consisted of a constant and the 7 explanatory variables. PGW's results are reported in table 1.

RETINA performs best on out-of-sample prediction criterion whereas the liberal PcGets model has the lowest AIC. The results highlight many problems with the data. In comparison to the conservative strategy, RETINA is retaining *hun* rather than *emt*, which dramatically improves CMSPE but worsens AIC. There are two substantial outliers in *emt* (see figure 2, panel a) at observations 107 and 701 which are driving the results. Rather than include intercept dummies for these observations they are removed from the data set, reducing the sample size to 1215.<sup>5</sup> Secondly, the non-constancy of the data is likely to lead to the problem of nonsense regressions. A plot of *INTRA* when it is ordered by *emt* clearly shows the data is non-constant. Whilst the linear models are unlikely to pass all diagnostics because of the non-linearities and interaction terms that are found to be significant, PcGets aims to find a congruent model given the GUM. The GUM should initially

<sup>5</sup>The two outliers were removed rather than dummied out to ensure the CMSPE that was computed was an accurate estimate of the forecast error.

be tested and when found to fail the diagnostics as in this case the procedure should be to re-formulate the GUM rather than proceed with the search strategy. Hence, the results for the linear models should be considered as hypothetical comparisons only.

Results for the linear models excluding the 2 outliers are reported in table 2. The models differ substantially from those reported in table 1. Firstly, the adjusted  $R^2$  is reduced by approximately 0.22 for all models, indicating that the 2 outliers were driving the results. In the benchmark linear model, the coefficients on *emt* and *emh* have changed substantially. This has a strong impact on the models chosen by PcGets and RETINA. The inclusion of *emh* (with a  $t$ -statistic of 7 when *emt* is included) is solely driven by the outliers and becomes insignificant when these are removed. It was effectively acting as an intercept dummy. The significance of *emt* substantially increases when the outliers are removed.

The CMSPE and AIC results now provide a much more coherent explanation of the data. Both PcGets strategies retain the same variables. The PcGets model marginally outperforms RETINA on both CMSPE and AIC criterion due to the inclusion of *sqft* which is highly significant. Note that the RETINA model selection procedure excludes a variable that has a  $t$ -statistic of 6.3, suggesting that the size properties of RETINA are very stringent. The use of sub-samples controls the size and hence achieves parsimony but it does have a detrimental effect on retaining highly significant variables.

To examine the performance of the programs when the non-linear functions were generated, table 3 reports the models selected with the GUM including all level 1 transformations. Including the non-linear and interaction terms improves the performance of all models in comparison to table 2. Parsimony is lost moving to the non-linear model for PcGets but not for RETINA. With 7 explanatory variables, the GUM will contain 113 variables including a constant, 7 levels, 7 squares, 7 inverses, 7 squared inverses,  ${}_7C_2 = 21$  cross products,  ${}_7C_2 = 21$  cross inverses and  ${}_7P_2 = 42$  ratios. However, there are numerical problems with some transformations, namely *bus* and *hwn* contain some zeros and therefore ratios and inverses cannot be computed. Hence, 27 level 1 transformations are excluded, resulting in a GUM containing 86 variables.

RETINA retains just 4 variables including a constant, *bus*, *emh* and  $(bus \times sqft)$ . PcGets retains more variables, including a number of squares, inverses products and ratios.<sup>6</sup> PcGets does outperform RETINA both in terms of CMSPE and AIC, even though RETINA is designed to forecast whereas PcGets is not primarily designed as a forecasting tool. The non-linear and interaction terms yield improvements for both models, suggesting that a move towards models which are linear in

---

<sup>6</sup>Full details of all the models are available on request.

the parameters but with non-linear functions should be considered in applications of this type. The minimum AIC falls from 5.4 to 4.8 when moving from a linear to non-linear specification and forecast accuracy increases. The non-linear models reported in PGW are reproduced in table 4 for comparison. The substantial forecasting gains exhibited in this table are reduced by removing the 2 outliers. This is driven by the non-linear terms acting as intercept dummies for the two outliers. By removing these, fewer non-linear terms are retained by RETINA but more variables are retained by PcGets.

The costs of searching for a non-linear model are small for RETINA and suggest that if it is unknown whether the DGP will contain non-linearities and interaction terms the use of RETINA will be informative. The costs of PcGets appear slightly higher in terms of size, but this can be controlled. Monte Carlo evidence on the size properties of PcGets when there are non-linearities is examined in section 6.

## 5.2 Time Series Data: Hoover and Perez (1999)

To assess both programs on time series data we apply both RETINA and PcGets to a drawing from the Hoover and Perez (1999) data set. Hendry and Krolzig (1999) undertake Monte Carlo experiments to assess how well PcGets recovers the DGP. These results are also briefly presented. The data is outlined in table 5. The DGP for various specifications are given in equations (1) to (15). Note that models  $y2'$ ,  $y7'$ ,  $y8'$  and  $y9'$  are equivalent to models  $y2$ ,  $y7$ ,  $y8$ , and  $y9$  with lags eliminating the autoregressive parameter in the error process.

$$y1_t = 130u_t \quad (1)$$

$$y2_t = 130u_t^* \quad (2)$$

$$y2'_t = 0.75y'_{2,t-1} + 85.99u_t \quad (3)$$

$$\ln y3_t = 0.395 \ln y3_{t-1} + 0.3995 \ln y3_{t-2} + 0.00172u_t \quad (4)$$

$$y4_t = 1.33\Delta FM1DQ_t + 9.73u_t \quad (5)$$

$$y5_t = -0.046\Delta^2 GGEQ_t + 0.11u_t \quad (6)$$

$$y6_t = 0.67\Delta FM1DQ_t - 0.023\Delta^2 GGEQ_t + 4.92u_t \quad (7)$$

$$y6A_t = 0.67\Delta FM1DQ_t - 0.32\Delta^2 GGEQ_t + 4.92u_t \quad (8)$$

$$y6B_t = 0.67\Delta FM1DQ_t - 0.65\Delta^2 GGEQ_t + 4.92u_t \quad (9)$$

$$y7_t = 1.33\Delta FM1DQ_t + 9.73u_t^* \quad (10)$$

$$y7'_t = 0.75y'_{7,t-1} + 1.33\Delta FM1DQ_t - 0.9975\Delta FM1DQ_{t-1} + 9.73u_t \quad (11)$$

$$y8_t = -0.046\Delta^2GGEQ_t + 0.11u_t^* \quad (12)$$

$$y8'_t = 0.75y8'_{t-1} - 0.046\Delta^2GGEQ_t + 0.00345\Delta^2GGEQ_{t-1} + 0.073u_t \quad (13)$$

$$y9_t = 0.67\Delta FM1DQ_t - 0.023\Delta^2GGEQ_t + 4.92u_t^* \quad (14)$$

$$y9'_t = 0.75y9'_{t-1} + 0.67\Delta FM1DQ_t - 0.5025\Delta FM1DQ_{t-1} - 0.023\Delta^2GGEQ_t + 0.01725\Delta^2GGEQ_{t-1} + 3.25u_t \quad (15)$$

where:

$$u_t \sim N[0, 1] \quad (16)$$

$$u_t^* = 0.75u_{t-1}^* + u_t\sqrt{7/4}. \quad (17)$$

The GUM for RETINA and PcGets in levels includes a constant, variables dated  $t$  and  $t - 1$  of  $\Delta DCOINC$ ,  $\Delta^2GD$ ,  $\Delta^2GGEQ$ ,  $\Delta GGFEQ$ ,  $\Delta^2GGFR$ ,  $\Delta GNPQ$ ,  $\Delta GYDQ$ ,  $\Delta GPIQ$ ,  $\Delta^2FMRR$ ,  $\Delta^2FMBA$ ,  $\Delta FM1DQ$ ,  $\Delta FM2DQ$ ,  $\Delta FSDJ$ ,  $\Delta FYAAAC$ ,  $\Delta LHC$ ,  $\Delta LHUR$ ,  $\Delta MU$ ,  $\Delta^2MO$ ,  $\Delta GCQ$ , and the lagged dependent variable giving a total of 40 regressors. The sample extends from 1960q2-1995q1, resulting in 140 observations. The GUM for RETINA and PcGets with the level 1 transformations includes the GUM as outlined for levels and the level 1 transformations given by the squares, inverses, and inverses of the squares. The interaction terms including cross products, cross inverses and ratios are excluded due to the size of the data set. In total there are 154 regressors in the GUM for the non-linear model. Including the interaction terms would increase the number of regressors to 334. The 154 regressors are partitioned into 4 groups and all pairwise combinations of the groups are implemented as GUMs, resulting in 6 combinations. The specific models are then combined to produce the final GUM. As there is some evidence of heteroscedasticity in the errors, the selection was checked by switching off the diagnostics to enable a clearer comparison with RETINA. If the results differ substantially, both models are reported.

The resulting model specifications are reported in tables 6 to 17 for the various model specifications listed above. As well as assessing how well each selection procedure finds the DGP, the models are compared on information criteria ( $AIC$ ) and out-of-sample forecasting ability using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The forecasts were derived from the full-sample model specification that was fitted over the first 100 observations and 1-step static forecasts were computed over the last 40 observations. As the model specification uses ex-post information the forecast errors will be biased downwards.

Table 18 summarises the results. ‘Relevant’ refers to the number of relevant variables retained and ‘irrelevant’ refers to the number of irrelevant variables that

are retained. Note that RETINA always retains the constant and therefore it will always retain 1 irrelevant variable given the DGP specifications.

As a general overview of the results, RETINA shows no evidence of overfitting and the models rarely retain any non-linear terms that are not included in the DGP. The sub-sample cross-validation method used in RETINA is successful in avoiding the problems of retaining too many non-linear terms. In particular, one concern was that RETINA uses 3 sub-samples for cross-validation which would indicate that a large sample is needed. However, even with only 140 observations, the results are promising. One must restrict the number of non-linear transformations used in smaller sample sizes though. PcGets tends to retain more irrelevant variables, although less are retained when the significance levels are adjusted and the diagnostics are switched off.

PcGets tends to overfit slightly, although the number of variables retained is close to what would be expected given the size and power of the program under the liberal and conservative strategies. Even with the substantially larger GUM for the non-linear models, PcGets does not retain many more regressors than in the linear case. This appears to be driven by the use of pairwise groupings for the GUM. It is essential that the data passes diagnostics as non-linear terms may be kept to ensure congruence, even if they are insignificant.

Overall, very few non-linear terms were retained. The data may need to be standardized, for example by taking logs, as the non-linear terms are very erratic and may be acting like dummies. Also, we have not looked at the interaction terms, cross products, cross inverses and ratios, which may exhibit different behaviour.

To understand the results in context, the results of Hendry and Krolzig (1999) are briefly reported in table 19, which look at 1000 replications of the DGP with a sample size of 100 for equations (1), (2), (10) and (14). Table 2 in Hendry and Krolzig (1999) reports the full results. In model  $y7$ , the DGP is found 59% and 11% of the time at 0.01 and 0.05 significance levels respectively. The drawing examined above found the DGP using the conservative strategy. The probability of including non-DGP variables is 41% at the 0.01 significance level, and accordingly, no non-DGP variables were retained in the model. However, in the case where non-linear variables were included in the GUM, non-DGP variables were retained. Section 6.2 examines the results of similar Monte Carlo experiments in this case.

To conclude, the results of both PcGets and RETINA are promising. RETINA tends to underfit, which may be useful for forecasting purposes as parsimonious models may be more robust. This is not necessarily the case, a less parsimonious model could be made robust by differencing for example. Simplicity in itself is not a necessary element of forecasting models but robustness is. Most importantly, RETINA does not retain numerous irrelevant non-linear terms which is useful if testing whether a model may contain non-linear functions or otherwise. The prop-

erties of PcGets are well documented and results reported here accord with the Monte Carlo evidence on PcGets.

### 5.3 Artificial Time Series Data: Krolzig and Hendry (2001)

As well as examining the DGP outlined in Hoover and Perez (1999), a white noise DGP is appraised. We follow the DGP outlined in Krolzig and Hendry (2001) in which the orthogonality of regressors enables an easier analysis and the problems of heteroscedasticity that arise in the Hoover and Perez DGP are avoided. The DGP is given by:

$$y_t = \sum_{k=1}^5 \beta_{k,0} x_{k,t} + \varepsilon_t, \quad \varepsilon_t \sim \text{IN}[0, 1] \quad (18)$$

$$x_t = v_t, \quad v_t \sim \text{IN}_{10}[0, \mathbf{I}_{10}] \quad \text{for } t = 1, \dots, T. \quad (19)$$

where  $\beta_{1,0} = 2/\sqrt{T}$ ,  $\beta_{2,0} = 3/\sqrt{T}$ ,  $\beta_{3,0} = 4/\sqrt{T}$ ,  $\beta_{4,0} = 6/\sqrt{T}$ ,  $\beta_{5,0} = 8/\sqrt{T}$ . Two sample sizes are examined, including  $T = 100$  and  $T = 1000$ . Because of orthogonality, the non-zero population  $t$ -values are 2,3,4,6 and 8.

Prior to including the non-linear transformations, the levels GUM is an ADL(1,1) model which includes as non-DGP variables the lagged endogenous variable  $y_{t-1}$ , the strongly exogenous variables  $x_{6,t}, \dots, x_{10,t}$  and the 1<sup>st</sup> lags of all regressors:

$$y_t = \pi_{0,1} y_{t-1} + \sum_{k=1}^{10} \sum_{i=0}^1 \pi_{k,i} x_{k,t-i} + \pi_{0,0} + u_t, \quad u_t \sim \text{IN}[0, \sigma^2] \quad (20)$$

In equation (20) 17 out of the 22 regressors are nuisance.

We then examine the case in which the non-linear terms and interaction terms are included in the GUM. Due to the number of level 1 regressors that will be generated, we discard the non-linear transformations for the strongly exogenous variables  $x_{6,t}, \dots, x_{10,t}$ . The GUM is given by:

$$\begin{aligned} y_t = & \pi_{0,1} y_{t-1} + \pi_{0,0} + \sum_{k=1}^{10} \sum_{i=0}^1 \pi_{k,i} x_{k,t-i} + \sum_{k=1}^5 \sum_{i=0}^1 \gamma_{k,i} x_{k,t-i}^2 \\ & + \sum_{k=1}^5 \sum_{i=0}^1 \delta_{k,i} \left( \frac{1}{x_k} \right)_{t-i} + \sum_{k=1}^5 \sum_{i=0}^1 \theta_{k,i} \left( \frac{1}{x_k^2} \right)_{t-i} \\ & + \sum_{j=1}^4 \sum_{l=j+1}^5 \sum_{i=0}^1 \kappa_{k,i} (x_j x_l)_{t-i} + \sum_{j=1}^4 \sum_{l=j+1}^5 \sum_{i=0}^1 \mu_{k,i} \left( \frac{1}{x_j x_l} \right)_{t-i} \\ & + \sum_{j=1}^4 \sum_{l=j+1}^5 \sum_{i=0}^1 \phi_{k,i} \left( \frac{x_j}{x_l} \right)_{t-i} + u_t, \quad u_t \sim \text{IN}[0, \sigma^2] \end{aligned} \quad (21)$$

In equation (21) there are a total of 112 regressors, 107 of which are nuisance. These include a constant, the lagged dependent variable, 10 ‘levels’ variables and their lags and the non-linear terms including 5 squared regressors, 5 inverses, 5 squared inverses, 10 cross products, 10 cross inverses and 10 ratios, plus the lags of all non-linear terms. Note that RETINA will also include the non-linear transformations of the lagged dependent variable, resulting in 18 additional regressors. Additional terms could be included by implementing all  ${}_5P_2$  ratios. For the sample size of 100, PcGets will be run using 6 GUMs. The 112 regressors will be partitioned into 4 groups of 28 and all possible pairwise combinations of these groups shall be implemented as the GUM. The resulting specific models are then merged and used as the GUM for the final search procedure.

Table 20 compares the results of PcGets and RETINA using the GUM outlined in equation (20). A linear OLS model is included as a benchmark and the results of both the liberal and conservative strategies of PcGets are reported. Evaluating the models in-sample, the PcGets liberal strategy delivers the model with the lowest AIC, both for  $T = 100$  and  $T = 1000$ . This model also retains the most relevant variables (4 out of 5 for both sample sizes). A summary of the number of relevant and irrelevant variables that are retained is given in table 22. RETINA delivers a more parsimonious model, retaining just  $x_{4,t}$  and  $x_{5,t}$  for both sample sizes.

None of the models exactly identify the DGP but all models retain the two most significant variables in the DGP. All of the coefficient estimates for  $x_{5,t}$  differ to their populations values by more than 1 standard error (other than the OLS model for  $T = 100$ ), with smaller estimates for sample size 100 and larger estimates than the population values for sample size 1000. Only the liberal strategy of PcGets for the large sample size picks up  $x_{3,t}$  with a population  $t$ -value of 4, which is somewhat surprising. Monte Carlo evidence from Krolzig and Hendry (2001) shows that, for  $\alpha = 0.05$ ,  $\eta = 0.01$  and  $T = 100$ , the probability of including  $x_{3,t}$  with a population  $t$ -value of 4 lies above 95%.<sup>7</sup>

As RETINA is primarily a forecasting tool, the RMSE and the MAPE is also reported. The model specification derived from the full sample was used, but the coefficients were re-estimated over the in-sample period (i.e.  $T = 67$  and 667) and 1-step static forecasts were computed over the last third of the sample. Note that this is a purely theoretical exercise. A true forecasting model would be developed using only ex-ante information. Our forecast errors will be biased downwards because the model specifications were chosen ex-post. RETINA does use out-of-sample forecasts when looking at the sub-sample local best models. However, to ensure a fair comparison the forecasts from the final best model delivered by

---

<sup>7</sup> $\alpha$  is the significance level of the simplification tests whereas  $\eta$  is the significance level of the diagnostic tests.



RETINA are reported.

For a sample size of 100, the liberal and conservative strategies of PcGets perform best on RMSE and MAPE respectively. For the larger sample size of 1000, RETINA delivers the best forecasts. Note that there is no lagged information in the DGP and so forecasting the DGP given in equation (18) will be difficult unless the contemporaneous exogenous variables are assumed to be known.

Examining the case in which non-linear terms are included, the results of testing from the GUM outlined in equation (21) are reported in table 21. On the basis of minimum AIC, the liberal strategy of PcGets produces the best model at both sample sizes. Interestingly, more variables are retained in the smaller sample size. 3 non-linear terms are retained but  $x_{3,t}$  is not retained. For the larger sample, only 4 variables are retained including 1 non-linear term. RETINA only retains the most significant variable (population  $t$ -value of 8) at both sample sizes. Comparing this with table 20, the inclusion of level 1 transformations results in  $x_{4,t}$  not being retained in the selected model.

With regard to forecasting properties, for a sample size of 1000, the conservative strategy of PcGets delivers the lowest RMSE and the smallest MAPE. For the smaller sample size, the liberal strategy of PcGets delivers the lowest RMSE but RETINA produces the best forecasts in terms of MAPE.

To summarize, RETINA tends to select more parsimonious models than PcGets. This is driven by the use of disjoint sub-samples and the search strategy stage in which variables from the local best model are sequentially added in and chosen on the basis of AIC. Note that AIC does tend to overselect asymptotically. RETINA did not retain any non-linear transformations, suggesting that the costs of using such a model when the DGP is linear are very small. As such, the program is useful in indicating whether a non-linear model should be considered when modelling. PcGets tended to retain more relevant variables than RETINA. The costs of search are easily computed for PcGets. The number of variables retained in selection  $t$ -testing is given by:

$$n = \sum_{i=0}^k i \frac{k!}{i! (k-i)!} \alpha^i (1-\alpha)^{k-i} = ka.$$

Hence, when  $\alpha = 0.05$ ,  $n = 1.1$  for  $k = 22$  but  $n = 5.6$  for  $k = 112$ . A larger GUM (due to the inclusion of non-linear terms) will lead to more irrelevant variables being retained. Campos, Hendry and Krolzig (2003) observe the need for significance levels to vary with both sample size and the number of candidate variables. One interesting feature of the results is that more variables were retained in the smaller sample size, suggesting that small sample properties outweigh the problems of retaining irrelevant variables on average.

## 6 Monte Carlo Results for Non-Linear models using PcGets

Whilst the above experiments are informative for the individual problems assessed, a formal evaluation of the properties of non-linear model selection is needed. This section aims to measure the ‘size’ and ‘power’ properties of the PcGets model selection process along the lines of Krolzig and Hendry (2001) when non-linear terms enter the GUM. Evidence in section 5 suggests that the inclusion of a substantial number of non-linear terms in the GUM may result in over-parameterization. This could be driven by ‘inference fragility’, whereby the model is sensitive to keeping one variable which may have low significance. Once the marginal variable is removed the significance of other correlated variables tends to fall. Note that some non-linear terms may be acting like intercept dummies. There are a multitude of non-linear and interaction terms that could be examined but these experiments concentrate on those terms that produce the level 1 transformations for RETINA.

### 6.1 Design of the Monte Carlo

The design of the Monte Carlo experiment is kept simple for tractability and is outlined in section 5.3. Three cases are analysed below. These include the linear DGP and GUM given in equations (18), (19) and (20). This experiment is identical to Krolzig and Hendry (2001) and the results are reported for comparison with the non-linear models. Secondly, we retain the linear DGP in equations (18) and (19) but assess PcGets on a non-linear GUM similar to equation (21). The inverses of the squares are excluded from the GUM as the inverse of the square of a WN process with mean 0 will contain very large values, resulting in numerical problems. We also exclude the 1<sup>st</sup> lags of the level 1 transformations to reduce the parameters in the GUM. The GUM is given in equation (22) and there are a total of 62 regressors, 57 of which are nuisance.

$$\begin{aligned}
 y_t = & \pi_{0,1}y_{t-1} + \pi_{0,0} + \sum_{k=1}^{10} \sum_{i=0}^1 \pi_{k,i}x_{k,t-i} + \sum_{k=1}^5 \gamma_{k,t}x_{k,t}^2 \\
 & + \sum_{k=1}^5 \delta_{k,t} \left( \frac{1}{x_k} \right)_t + \sum_{j=1}^4 \sum_{l=j+1}^5 \kappa_{k,t} (x_j x_l)_t \\
 & + \sum_{j=1}^4 \sum_{l=j+1}^5 \mu_{k,t} \left( \frac{1}{x_j x_l} \right)_t + \sum_{j=1}^4 \sum_{l=j+1}^5 \phi_{k,t} \left( \frac{x_j}{x_l} \right)_t + u_t, \quad (22) \\
 u_t \sim & \text{IN} [0, \sigma^2]
 \end{aligned}$$

Finally, we assess the properties of the PcGets model selection procedure when there are non-linearities and interaction terms in the DGP. Maintaining orthogonality of the regressors enables easier analysis and so, with 5 white noise processes, the non-linear terms include a square and 2 cross products, resulting in the DGP given in equations (23) and (24).

$$y_t = \sum_{k=1}^5 \beta_{k,0} x_{k,t} + \beta_{6,0} (x_1)_t^2 + \beta_{7,0} (x_2 x_3)_t + \beta_{8,0} (x_4 x_5)_t + \varepsilon_t, \quad (23)$$

$$\varepsilon_t \sim \text{IN}[0, 1]$$

$$v_t = v_t, \quad v_t \sim \text{IN}_5[0, \mathbf{I}_5] \quad \text{for } t = 1, \dots, T. \quad (24)$$

We can determine the  $\beta$ 's by specifying the  $t$ -values and backing out the  $\beta$ 's due to orthogonality. A  $t$ -test of  $H_0 : \beta_k = 0$  versus  $H_1 : \beta_k \neq 0$  is given by:

$$t_k = \frac{\hat{\beta}_k}{\hat{\sigma}_{\beta_k}}.$$

Because of orthogonality,  $\hat{\beta}_k = (X'X)^{-1} (X'y) = \hat{\sigma}_{x_k y} / \hat{\sigma}_{x_k}^2$  and  $\hat{\sigma}_{\beta_k}^2 = \hat{\sigma}_{\varepsilon}^2 (X'X)^{-1} = \hat{\sigma}_{\varepsilon}^2 / (T \hat{\sigma}_{x_k}^2)$ . This implies:

$$t_k = \frac{\hat{\beta}_k}{\hat{\sigma}_{\beta_k}} = \frac{\hat{\beta}_k}{\sqrt{\hat{\sigma}_{\varepsilon}^2 / (T \hat{\sigma}_{x_k}^2)}} = \sqrt{T} \hat{\beta}_k \frac{\hat{\sigma}_{x_k}}{\hat{\sigma}_{\varepsilon}}.$$

As  $\hat{\sigma}_{\varepsilon} = 1$  and  $\hat{\sigma}_{x_k}$  for  $x_k$ ,  $t$ -values of 2,3,4,6 and 8 result in  $\beta_{1,0} = 2/\sqrt{T}$ ,  $\beta_{2,0} = 3/\sqrt{T}$ ,  $\beta_{3,0} = 4/\sqrt{T}$ ,  $\beta_{4,0} = 6/\sqrt{T}$ , and  $\beta_{5,0} = 8/\sqrt{T}$  respectively.

For  $\beta_{6,0}$  we need to determine  $\hat{\sigma}_{x_k}^2$ . This is given by:

$$\begin{aligned} \hat{\sigma}_{x_k}^2 &= E \left[ \sum_{t=1}^T (x_{1t})^2 \right] = 3T^2 \hat{\sigma}_{x_k}^4 \\ \therefore t_k &= \sqrt{3T} \hat{\beta}_k \end{aligned}$$

and so we shall set  $\beta_{6,0} = 4/\sqrt{3T}$  for a  $t$ -value of 4.

For  $\beta_{7,0}$  and  $\beta_{8,0}$ :

$$\begin{aligned} \hat{\sigma}_{x_k x_j} &= E \left[ \sum_{t=1}^T (x_{1t} x_{2t}) \right] = T \\ \therefore t_k &= \sqrt{T} \hat{\beta}_k \end{aligned}$$

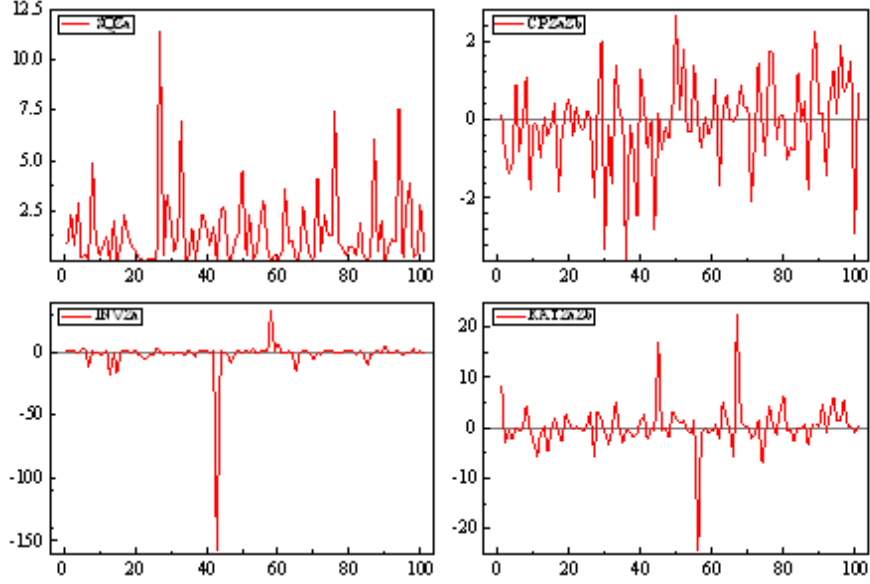


Figure 3: Non-linear transformations of a white noise process including the square, the cross product of two WN processes, the inverse and the ratio of two WN processes.

and so we shall set  $\beta_{7,0} = 3/\sqrt{T}$  and  $\beta_{8,0} = 4/\sqrt{T}$  for  $t$ -values of 3 and 4 respectively. Note that various  $t$ -values were assessed for the non-linear terms but were not reported for brevity.

To characterize the non-linear terms, figure 3 records the square of a WN process, the cross product of two WN processes, the inverse and the ratio of two WN processes in panels a to d respectively.

## 6.2 Results for the Monte Carlo Experiments

Prior to reporting the results for the non-linear models, results from Krolzig and Hendry (2001) are briefly reported. As noted in their paper, the portmanteau and hetero- $x$  diagnostic tests do not exhibit satisfactory behaviour and are therefore excluded from the test battery that PcGets uses. The empirical distributions of the test statistics are unaffected by strongly exogenous nuisance regressors. Earlier experiments suggested that the calibration of the heteroscedasticity tests was poor but a correction for degrees of freedom leads to a substantial improvement, see Hendry

and Krolzig (2003a).<sup>8</sup> The DGP contains 5 variables and the GUM contains 22. The number of replications,  $M$ , is 1000. An overview of the selection probabilities is given in table 23 along with the results for the non-linear experiments. DGP found refers to the probability that the exact DGP is found. The liberal strategy finds the DGP marginally more often than the conservative strategy, which has more difficulty retaining significant variables with  $t$ -statistics of 2 or 3. The smaller sample size has a higher probability of finding the DGP. The liberal strategy has a much higher probability of retaining non-DGP variables compared to the conservative strategy. The trade-off is that the probability that a DGP variable is not included is higher for the conservative strategy. The final 2 rows refer to whether the DGP and the specific model is dominated. The probability that the specific model is dominated by the true model is almost 0 for all models, suggesting that the models derived by PcGets are good representations of the true model.

Table 24 records the selection probabilities for the non-central  $t$ -statistics given a linear GUM and DGP, along with the results from a simulated model when the DGP is known. There is very little difference in selection probabilities between searching for the DGP using PcGets and starting with a known DGP. There is a small difference for  $t$ -values of 2 with small sample sizes when PcGets actually has a higher retention rate than starting with the DGP. The last 4 columns record the average coefficients with standard errors in parentheses. All of the specific model coefficients lie within 1 standard error of the true coefficients, other than the conservative model for a  $t$ -value of 2 at  $T = 1000$ . The results are promising and show that the costs of search are small compared to the costs of inference. Figure 4 plots the probability of retaining variables in the GUM. The size for all non-DGP variables is fairly constant all probabilities lie well below 0.1, suggesting that starting with an overly general GUM will not be costly in the linear model.

We next examine the case using the same DGP but the GUM is extended to include non-linear terms. Results are recorded in table 25 and figure 5 plots the retention probabilities of the variables in the GUM.<sup>9</sup> The results are encouraging. The inclusion of a substantial number of non-linear and interaction terms does not dramatically alter the results and does not lead to the problem of ‘inference fragility’ which was a possibility. The exact DGP is found less often than the linear case in accordance with a larger GUM (62 variables as opposed to 22 variables). Also, the probability of not retaining a DGP variable increases, as does the probability of including non-DGP variables. A sensible way to proceed when analysing

---

<sup>8</sup>Diagnostic tests are not reported but can be found in Krolzig and Hendry (2001) and Hendry and Krolzig (2004).

<sup>9</sup>The number ordering on the  $x$ -axis proceeds in the order; lagged dependent variable, levels and their lags, squares, inverses, cross products, cross inverses, ratios and finally the constant. The interaction terms are ordered  $x_1x_2, x_1x_3, \dots, x_2x_3, \dots$  etc. in logical order.

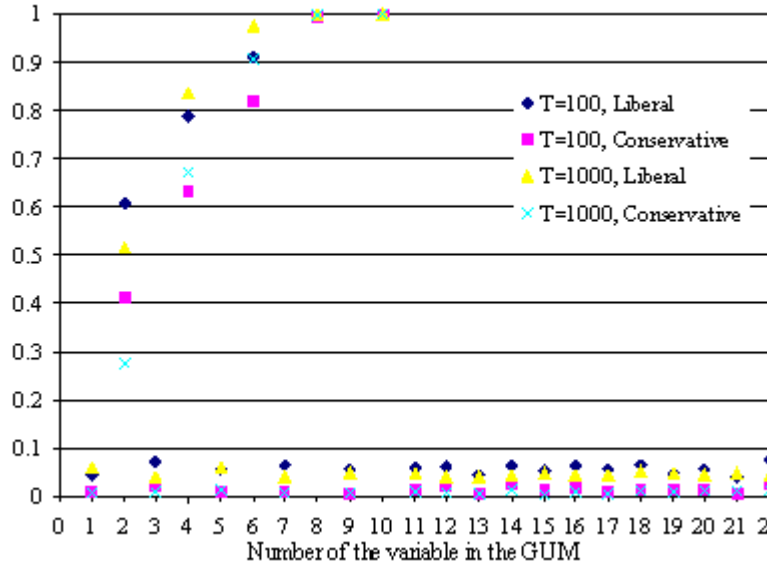


Figure 4: Probability of retaining the variable in the specific model, linear GUM and linear DGP.

this type of problem would be to tighten significance levels in direct correlation to the number of variables in the GUM. See Campos, Hendry and Krolzig (2003) for a discussion on optimal selection strategies. There is no evidence of different behaviour for non-linear terms in which a large number of variables are retained due to their interdependence.

Finally we assess the case in which there are non-linear terms contained in the DGP. Results are recorded in table 26 and figure 6 plots the retention probabilities as before. The non-linear terms in the DGP are numbered 22 for  $x_1^2$ , 36 for  $x_2x_3$  and 41 for  $x_4x_5$ . The results are very similar to those for the linear DGP, although the probability that a DGP variable is not retained is higher. Comparing the non-linear terms with  $t$ -statistics of 4 to the linear variable with a  $t$ -statistic of 4, there does appear to be some differences between the regressors, with the cross product having the highest retention probability. Comparing the results to simulations given the known DGP, for smaller  $t$ -values there are some costs of search. However, the behaviour of the non-linear variables is not dramatically different to that of the linear variables.

The Monte Carlo evidence suggests that we can treat problems that are non-linear in the variables (but linear in the parameters) in exactly the same way as

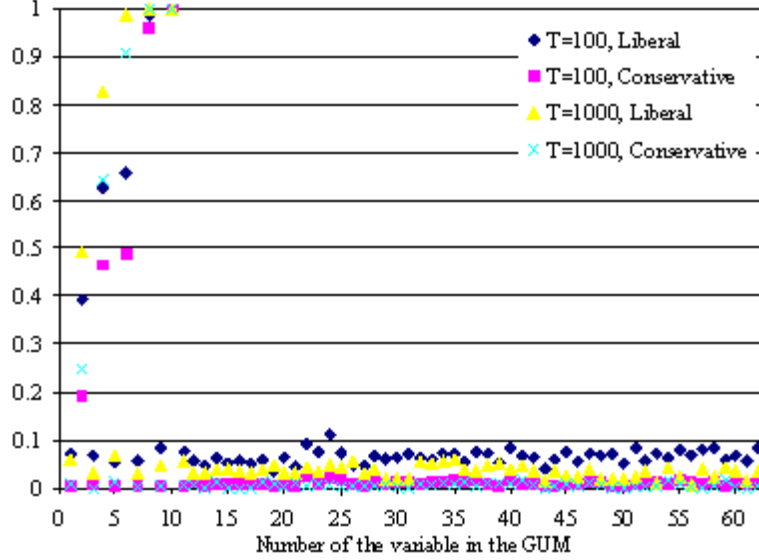


Figure 5: Probability of retaining the variable in the specific model, non-linear GUM and linear DGP.

we treat linear problems although the size of the tests will need to be controlled if starting with an overly general GUM. There is no strong evidence of inference fragility but both the cross section and time series applications suggest that certain types of non-linear functions may exhibit different behaviour. The results presented in this section are entirely dependent on the DGPs examined and a further avenue of research would be to assess the Monte Carlo evidence for non-linear models that are not based on white noise processes.

## 7 Conclusions

The aim of the paper was to assess two automatic model selection procedures, PcGets and RETINA, and undertake a horse race between the two programs to see how they perform on both cross section and time series data. The results are promising; automated methods of model selection have a high level of success and should be dominant in econometric modelling in the future. RETINA is a method of model selection designed primarily to forecast and its predominant feature is the parsimony it can achieve from a highly over-parameterized GUM. PcGets aims to

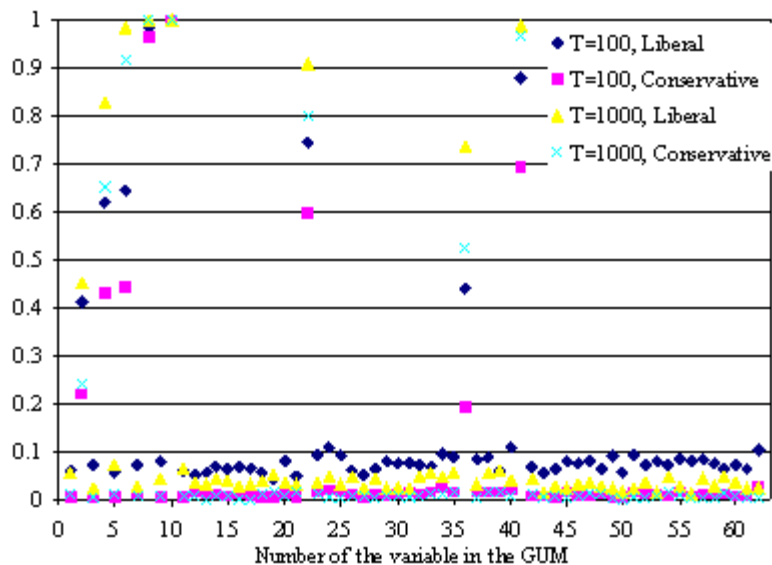


Figure 6: Probability of retaining the variable in the specific model, non-linear GUM and non-linear DGP.

find a congruent undominated representation of an overly general model. RETINA achieves parsimony via the use of 3 disjoint sub-samples and in all applications considered delivered a more parsimonious model than PcGets. The trade-off is that it often fails to retain highly significant variables. Both programs are useful for modelling and forecasting. The ease with which the significance of non-linearities can be tested with relatively low cost is most useful. The cross section results highlight the need to check for outliers. Non-linear functions may simply be reflecting a few outliers and vastly different results may be obtained by removing these observations. The Monte Carlo evidence shows that the size and power properties of PcGets does not differ substantially due to the inclusion of non-linear functions of white noise processes, although one must control the size. Further investigation into the types of non-linear functions that should be tested for is needed.

## 8 References

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**,



pp.803-819.

Clements, M.P., and Hendry, D.F. (1999). *Forecasting Non-stationary Economic Time Series*. Cambridge, Massachusetts: The MIT Press.

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

Hendry, D.F., and Krolzig, H.-M. (1999). 'Improving on 'Data mining reconsidered' by K. D. Hoover and S. J. Perez'. *Econometrics Journal*, **2**, pp.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. (2003a). 'New developments in automatic general-to-specific modelling'. in Stigum (2003), pp. 379–419.

Hendry, D.F., and Krolzig, H.-M. (2003b). 'Sub-sample Model Selection Procedures in *Gets* Modelling'. Nuffield College Working Paper, 2003/17.

Hendry, D.F., and Krolzig, H.-M. (2004). 'The Properties of Automatic *Gets* Modelling'. Sargan Lecture of the Royal Economic Society, 2004.

Hoover, K. D., and Perez, S. J. (1999). 'Data mining reconsidered: Encompassing and the general-to-specific approach to specification search'. *Econometrics Journal*, **2**, pp.167-191.

Krolzig, H.-M. and Hendry, D.F. (2001). 'Computer automation of general-to-specific model selection procedures'. *Journal of Economic Dynamics and Control*, **25**, pp.831-866.

Lynch, A.W., and Vital-Ahuja, T. (2003). '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.

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**, pp. 821–838.

Perez-Amaral, T., Gallo, G.M. and White, H. (2004). 'A Comparison of Complementary Automatic Modelling Methods: RETINA and PcGets'. *Econometric Theory*, -, forthcoming.

Stigum, B. P. (ed.)(2003). *Econometrics and the Philosophy of Economics*. Princeton: Princeton University Press.

	Linear Model (Benchmark)	PcGets (Liberal)	PcGets (Conservative)	RETINA (Level 0 regressors)
CMSPE	909.88	896.11	903.08	770.86
AIC	5.443	5.440	5.446	5.459
Adj. $R^2$	0.603	0.603	0.600	0.595
Parameters	8	6	5	5
Const	−3.945 (0.717)	−3.907 (0.678)	−3.278 (0.649)	−4.086 (0.684)
Bus	2.508 (0.207)	2.492 (0.183)	2.579 (0.182)	2.571 (0.185)
Hun	0.089 (0.031)	0.091 (0.030)		0.083 (0.030)
Sales	−7.972 (41.852)			
Emt	0.472 (0.093)	0.472 (0.093)	0.458 (0.093)	
Emh	0.930 (0.132)	0.933 (0.130)	0.921 (0.131)	1.438 (0.085)
Sqft	0.450 (0.061)	0.450 (0.061)	0.462 (0.061)	0.481 (0.061)
Pop	0.033 (0.184)			

Table 1: RETINA and PcGets: Linear models from PGW.

	Linear Model (Benchmark)	PcGets (Lib and Cons)	RETINA (Level 0 regressors)
CMSPE	772.595	732.852	754.414
AIC	5.390	5.387	5.417
Adj. $R^2$	0.385	0.386	0.366
Parameters	8	5	4
Const	−3.365 (0.709)	−3.403 (0.667)	−2.044 (0.641)
Bus	2.374 (0.202)	2.301 (0.178)	2.461 (0.179)
Hun	0.096 (0.031)	0.096 (0.029)	0.106 (0.029)
Sales	−16.017 (41.23)		
Emt	1.682 (0.179)	1.522 (0.122)	1.730 (0.119)
Emh	−0.429 (0.360)		
Sqft	0.389 (0.061)	0.380 (0.060)	
Pop	0.039 (0.179)		

Table 2: RETINA and PcGets: Linear models for PGW excluding 2 outliers.

	PcGets (Liberal)	PcGets (Conservative)	RETINA (AIC)
CMSPE	694.315	719.138	722.839
AIC	4.799	4.812	5.258
Adj. $R^2$	0.666	0.661	0.460
Parameters	33	29	4

Table 3: RETINA and PcGets: Non-linear models from PGW excluding two outliers.

	PcGets (Liberal)	PcGets (Conservative)	RETINA (AIC)	RETINA (CMSPE)
CMSPE	507.42	498.63	572.01	518.00
AIC	4.839	4.839	4.932	4.947
Adj. $R^2$	0.785	0.784	0.757	0.756
Parameters	19	18	9	9

Table 4: RETINA and PcGets: Non-linear models from PGW.

Variable	Times differenced for Stationarity	Name
Index of four coincident indicators	1	DCOINC
GNP price deflator	2	GD
Government purchases of goods and services	2	GGEQ
Federal purchases of goods and services	1	GGFEQ
Federal government receipts	2	GGFR
GNP	1	GNPQ
Disposable personal income	1	GYDQ
Gross private domestic investment	1	GPIQ
Total member bank reserves	2	FMRR
Monetary base (federal reserve bank of St. Louis)	2	FMBASE
M1	1	FM1DQ
M2	1	FM2DQ
Dow Jones stock price	1	FSDJ
Moody's AAA corporate bond yield	1	FYAAAC
Labour force (16 years+, civilian)	1	LHC
Unemployment rate	1	LHUR
Unfilled orders (manufacturing, all industries)	1	MU
New orders (manufacturing, all industries)	2	MO
Personal consumption expenditure	1	GCQ

Table 5: Hoover and Perez (1999) data set. Sample: 1959q1-1995q1.

	<i>y1</i>				
	Linear Models		Non-linear Models		
	PcGets Conservative	RETINA	PcGets Conservative	PcGets Cons. exc. diags	RETINA
Parameters retained	3	2	8	1	2
Adj. $R^2$	0.2836	0.2174	0.3915	0.2210	0.2174
$AIC$	9.9771	10.0585	9.8527	10.0511	10.0585
$\hat{\sigma}$	145.17	151.74	134.08	151.71	151.74
RMSE	150.3	153.4	130.74	153.4	153.4
MAPE	165.36	166.26	326.8	164.59	166.26
Const		−1.9981 (12.82)			−1.9981 (12.82)
$\Delta y_{t-1}$	−0.4395 (0.073)	−0.4728 (0.075)	−0.3756 (0.068)	−0.4713 (0.075)	−0.4728 (0.075)
$\Delta DCOINC_t$			−6566.05 (2173.65)		
$\Delta DCOINC_{t-1}$	7319.924 (2430.58)		7857.69 (2465.68)		
$\Delta^2 GGEQ_{t-1}$			−3002.94 (1005.92)		
$\Delta GNPQ_{t-1}$	−8491.573 (2311.95)		−8024.86 (2241.43)		
$\Delta GCQ_t$			8644.30 (1969.64)		
$\left[ \frac{1}{(\Delta^2 GGEQ)^2} \right]_{t-1}$			−0.00001 (0.0000)		
$(\Delta FSDJ)_t^2$			−3837.43 (1525.09)		

Table 6: Results for Hoover and Perez (1999) Model 1.

	$y2$			$y2'$		
	Linear Models PeGets Conservative	RETINA 2	Non-linear Models PeGets Conservative	RETINA 2	Linear Models PeGets Conservative	Non-linear Models PeGets Conservative
Parameters retained	0	2	0	2	0	0
Adj. $R^2$	0.00679	0.0439	0.00679	0.0439	0.00679	0.00679
$AIC$	10.34432	10.3204	10.34432	10.3204	9.78470	10.34432
$\hat{\sigma}$	176.29	172.968	176.29	172.968	133.27	176.29
RMSE	-	172.61	-	172.61	-	-
MAPE	-	123.35	-	123.35	-	-
Const	-3.5194 (14.62)		-3.5194 (14.62)		-3.5194 (14.62)	-3.5194 (14.62)
$\Delta^2 GD_t$	-9528.73 (3506)		-9528.73 (3506)		-9528.73 (3506)	-9528.73 (3506)

Table 7: Results for Hoover and Perez (1999) Model 2.

	$y3$			
	Linear Models		Non-linear Models	
	PcGets Conservative	RETINA	PcGets Conservative	RETINA
Parameters retained	2	2	2	2
Adj. $R^2$	0.2296	0.2156	0.2527	0.2156
$AIC$	-12.7193	-12.7054	-12.7498	-12.7054
$\hat{\sigma}$	0.0017	0.0017	0.0017	0.0017
RMSE	0.0017	0.0017	0.0017	0.0017
MAPE	142.47	134.20	134.86	134.20
Const		-0.00002 (0.0001)		-0.00002 (0.0001)
$y3_{t-1}$	-0.46216 (0.075)	-0.4728 (0.076)	-0.4609 (0.074)	-0.4728 (0.076)
$\Delta GNPQ_{t-1}$	-0.0213 (0.013)			
$\Delta^2 GD$			-0.0930 (0.035)	

Table 8: Results for Hoover and Perez (1999) Model 3.

	$y4$			
	Linear Models		Non-linear Models	
	PcGets Conservative	RETINA	PcGets Conservative	RETINA
Parameters retained	5	3	8	2
Adj. $R^2$	0.5716	0.4384	0.6195	0.4158
$AIC$	4.8445	5.1014	4.7462	5.1338
$\hat{\sigma}$	11.075	12.6807	10.437	12.933
RMSE	11.966	13.354	10.930	13.600
MAPE	425.66	340.38	287.11	295.08
Const		2.3182 (1.406)		0.1016 (1.129)
$y4_{t-1}$	-0.4192 (0.074)		-0.4183 (0.070)	
$\Delta FM1DQ_t$	870.692 (93.893)	843.633 (82.46)	1008.74 (94.73)	840.699 (84.10)
$\Delta FM1DQ_{t-1}$	419.812 (113.632)		326.448 (109.45)	
$\Delta GNPQ_{t-1}$	-678.563 (178.411)	-305.678 (119.5)	-767.272 (171.73)	
$\Delta DCOINC_{t-1}$	572.174 (187.244)		671.472 (180.31)	
$\Delta^2 FMBASE_t$			-482.087 (146.27)	
$\left[\frac{1}{\Delta FYAAAC}\right]_t$			0.00896 (0.00332)	
$\left[\frac{1}{(\Delta^2 GGEQ)^2}\right]_{t-1}$			-0.000 [t=-2.44] 0.000	

Table 9: Results for Hoover and Perez (1999) Model 4.

	$y5$				
	Linear Models			Non-linear Models	
	PcGets Cons (with diags)	PcGets Cons (without diags)	RETINA	PcGets Conservative	RETINA
Parameters retained	8	6	2	10	2
Adj. $R^2$	0.7951	0.7831	0.6984	0.8188	0.6984
$AIC$	-3.5549	-3.5114	-3.2096	-3.6643	-3.20964
$\hat{\sigma}$	0.1644	0.1692	0.1995	0.1547	0.199506
RMSE	0.1885	0.1984	0.1949	0.1814	0.1949
MAPE	67.448	69.447	59.323	79.089	59.323
Const			-0.1228 (0.017)	-0.1304 (0.028)	-0.1228 (0.017)
$\Delta GGFEQ_t$	-6.861 (1.440)	-7.4972 (1.463)	-15.7400 (0.876)	-6.2898 (1.375)	-15.7400 (0.876)
$\Delta GGFEQ_{t-1}$	-6.203 (1.513)	-6.2784 (1.556)		-7.1747 (1.454)	
$\Delta^2 GGEQ_t$	-22.207 (2.953)	-20.754 (2.993)		-23.265 (2.827)	
$\Delta^2 GGEQ_{t-1}$	-5.017 (1.430)	-5.3250 (1.465)		-6.0158 (1.357)	
$\Delta GNPQ_{t-1}$	-29.453 (4.886)	-18.025 (1.954)		-18.333 (3.347)	
$\Delta GPIQ_{t-1}$	4.025 (0.717)	2.6698 (0.471)		1.6698 (0.542)	
$\Delta FM1DQ_t$	-2.661 (1.126)				
$\Delta DCOINC_{t-1}$				16.086 (3.590)	
$\Delta LHUR_{t-1}$				1.3465 (0.477)	
$\Delta GCQ_{t-1}$	10.189 (3.757)				
$[\Delta^2 MO]_{t-1}^2$				12.659 (5.470)	

Table 10: Results for Hoover and Perez (1999) Model 5.



	$y6$			
	Linear Models		Non-linear Models	
	PcGets Conservative	RETINA	PcGets Conservative	RETINA
Parameters retained	5	3	8	2
Adj. $R^2$	0.5695	0.4356	0.6173	0.4126
$AIC$	3.4829	3.7398	3.3856	3.7728
$\hat{\sigma}$	5.6065	6.4193	5.2859	6.5491
RMSE	6.0419	6.748	5.5183	6.8675
MAPE	429.71	347.76	348.24	299.29
Const		1.1082 (0.712)		-0.0217 (0.572)
$y6_{t-1}$	-0.4180 (0.074)		-0.4172 (0.070)	
$\Delta FM1DQ_t$	434.616 (47.515)	424.420 (41.75)	504.273 (47.965)	422.924 (42.59)
$\Delta FM1DQ_{t-1}$	214.150 (57.482)		167.099 (55.437)	
$\Delta GNPQ_{t-1}$	-348.077 (90.238)	-155.817 (60.48)	-392.799 (86.908)	
$\Delta DCOINC_{t-1}$	286.058 (94.755)		336.151 (91.290)	
$\Delta^2 FMBASE_t$			-243.164 (74.075)	
$\left[\frac{1}{\Delta FYAAAC}\right]_t$			0.0045 (0.002)	
$\left[\frac{1}{(\Delta^2 GGEQ)^2}\right]_{t-1}$			-0.000 [t=-2.44] (0.000)	

Table 11: Results for Hoover and Perez (1999) Model 6.

	<i>y6A</i>				<i>y6B</i>			
	Linear Models		Non-linear Models		Linear Models		Non-linear Models	
	PcGets Conservative	RETINA	PcGets Conservative	RETINA	PcGets Conservative	RETINA	PcGets Conservative	RETINA
Parameters retained	7	3	9	2	8	3	9	2
Adj. $R^2$	0.5748	0.4399	0.6174	0.3705	0.6259	0.5071	0.64553	0.2929
$AIC$	3.5565	3.8045	3.4642	3.9143	3.6272	3.8686	3.57988	4.2224
$\hat{\sigma}$	5.7771	6.6303	5.4796	7.0290	5.9647	6.8463	5.80591	8.1999
RMSE	6.3523	6.7929	6.0368	7.2789	6.8371	6.8626		8.6690
MAPE	171.84	235.93	135.39	212.38	85.321	88.519		106.37
Const		-0.7363 (0.581)		-0.9607 (0.614)		-1.5790 (0.600)		-2.0040 (0.716)
$y6A_{t-1} y6B_{t-1}$	-0.3749 (0.076)		-0.3633 (0.072)		-0.1875 (0.066)		-0.1835 (0.064)	
$\Delta FM1DQ_t$	400.098 (48.978)	414.862 (43.11)	463.476 (49.363)	415.912 (45.71)	488.325 (43.179)	406.128 (44.52)	499.356 (42.203)	408.118 (53.32)
$\Delta FM1DQ_{t-1}$	212.413 (60.044)		166.752 (58.169)					
$\Delta GNPQ_{t-1}$	-411.142 (95.359)		-407.984 90.453		-471.901 (98.352)		-470.607 (95.735)	
$\Delta GGFEQ_t$		-123.828 (29.11)				-234.696 (30.06)		
$\Delta GGFEQ_{t-1}$	-160.081 (33.112)		-174.070 (31.604)		-246.902 (38.331)		-252.963 (37.369)	
$\Delta^2 GGFEQ_t$	-218.657 (53.770)		-242.687 (51.421)		-523.428 (155.888)		-525.420 (54.405)	
$\Delta^2 GGFEQ_{t-1}$					-151.222 (50.140)		-145.957 (48.840)	
$\Delta DCOINC_{t-1}$	277.306 (100.945)		282.847 (95.853)		318.916 (102.561)		327.721 (99.877)	
$\Delta^2 FMBASE_t$			-265.824 (76.724)		-293.256 (81.383)		-325.692 (80.011)	
$\left[\frac{1}{(\Delta^2 GGFEQ)^2}\right]_{t-1}$			-0.000 [t=-2.69] (0.000)				-0.000 [t=-2.88] (0.000)	

Table 12: Results for Hoover and Perez (1999) Model 6A and 6B.

	$y7$			$y7'$		
	Linear Models PcGets Conservative	RETINA	Non-linear Models PcGets Conservative	RETINA	Linear Models PcGets Conservative	Non-linear Models PcGets Conservative RETINA
Parameters retained	1	2	3	2	8	12
Adj. $R^2$	0.4335	0.4295	0.4608	0.4295	0.7298	0.7676
$AIC$	5.1848	5.1989	5.1494	5.1989	4.9532	4.8290
$\hat{\sigma}$	13.314	13.3613	12.989	13.3613	11.576	10.736
RMSE	13.492	13.490	13.291	13.490	11.972	10.456
MAPE	110.63	110.92	122.79	110.92	192.19	131.20
Const	-0.1648 (1.166)			-0.1648 (1.166)	2.6033 (1.240)	1.5865 (1.307)
$y7_{t-1}   y7'_{t-1}$					0.6150 (0.040)	0.5655 (0.040)
$\Delta FM1DQ_t$	889.949 (83.834)	893.014 (86.88)	916.760 (83.253)	893.014 (86.88)	379.568 (99.157)	332.59 (95.883)
$\Delta DCOINC_{t-1}$					665.023 (199.097)	987.46 (212.006)
$\Delta GNPQ_{t-1}$					-789.581 (192.370)	-1082.89 (184.81)
$\Delta^2 FMRRR_t$					225.464 (80.922)	
$\Delta^2 FMBASE_t$					-561.009 (188.456)	
$\Delta FM2DQ_t$					577.627 (131.617)	534.914 (132.932)
$\Delta FM2DQ_{t-1}$					-426.361 (128.451)	-455.358 (132.760)
$\Delta GPIQ_t$						-76.287 (23.702)

Table 13: Results for Hoover and Perez (1999) Model 7.

cont...	$y7$		$y7'$	
	Linear Models PcGets Conservative	Non-linear Models PcGets Conservative	Linear Models PcGets Conservative	Non-linear Models PcGets Conservative
$\Delta MU_{t-1}$	RETINA	RETINA	RETINA	RETINA
$\Delta^2 MO_{t-1}$				
$\Delta GCQ$				
$[\Delta GNPQ]_{t-1}^2$		-13065.8 (5383.9)		-65.217 (33.289)
$\left[\frac{1}{\Delta FMIDQ}\right]_{t-1}$		0.0022 (0.001)		81.706 (31.289)
$\left[\frac{1}{(\Delta^2 GGEQ)^2}\right]_{t-1}$				499.780 (147.396)
$\left[\frac{1}{(\Delta FYAAC)^2}\right]_t$				
				-0.000 [t=-3.69] (0.000)
				0.000 [t=2.88] (0.000)

Table 14: Model 7 continued.

	$y_8$			$y_8'$		
	Linear Models PcGets * Conservative	RETINA	Non-linear Models PcGets Conservative	Linear Models PcGets ** Conservative	RETINA	Non-linear Models PcGets Conservative RETINA
Parameters retained	8	2	9	10	4	3
Adj. $R^2$	0.7837	0.6940	0.8021	0.9960	0.9931	0.9929
$AIC$	-3.5455	-3.2397	-3.6279	-4.0686	-3.5672	-3.5402
$\hat{\sigma}$	0.1652	0.1965	0.1580	0.1264	0.1658	0.1685
RMSE	0.1806	0.1881	0.1704	0.1536	0.1636	0.1702
MAPE	107.83	121.75	107.71	34.128	35.872	39.135
Const	-0.1245 (0.017)	-0.1245 (0.017)	-0.1245 (0.017)	-0.0535 (0.018)	-0.1122 (0.015)	-0.0554 (0.018)
$y_8^{t-1}   y_8'^{t-1}$	0.3305 (0.075)		0.3294 (0.074)	0.7479 (0.004)	0.7517 (0.006)	0.7500 (0.008)
$\Delta GGFEEQ_t$	-6.0541 (1.500)	-15.3424 (0.863)	-6.970 (1.448)	-7.3129 (1.095)	-15.6629 (0.754)	-7.4439 (1.055)
$\Delta GGFEEQ_{t-1}$	-3.7580 (1.613)		-3.1978 (1.546)	-5.8385 (1.170)	1.7857 (0.745)	-5.3758 (1.127)
$\Delta^2 GD_t$	-9.7245 (3.441)			-5.8977 (2.635)		
$\Delta^2 GGEQ_t$	-22.795 (3.077)		-21.104 (2.973)	-20.9329 (2.266)		-20.7494 (2.176)
$\Delta^2 GGEQ_{t-1}$				-4.3091 (1.096)		-3.7859 (1.071)
$\Delta GNPQ_{t-1}$	-23.458 (4.979)		-19.496 (3.003)	-10.0155 (2.336)		-15.1067 (2.436)
$\Delta GPIQ_{t-1}$	2.8991 (0.735)		1.4407 (0.468)	1.5530 (0.446)		2.2013 (0.441)
$\Delta GCQ_{t-1}$	9.4878 (3.572)					

Table 15: Results for Hoover and Perez (1999) Model 8. \* indicates that switching the diagnostics off results in a model that excludes  $GCQ_{t-1}$  but is otherwise identical. \*\* indicates that switching the diagnostics off results in a model that excludes  $FMIDQ_{t-1}$  but is otherwise identical.

cont...	$y_8$		$y_{8'}$	
	Linear Models PcGets * Conservative	Non-linear Models PcGets Conservative	Linear Models PcGets ** Conservative	Non-linear Models PcGets Conservative
$\Delta DCOINC_{t-1}$	RETINA	RETINA	RETINA	RETINA
$\Delta^2 MO_{t-1}$		8.0640 (2.792)		1.0757 (0.337)
$\Delta FM1DQ_t$		1.5179 (0.417)	-1.4061 (0.826)	227.070 (75.440)
$[\Delta DCOINC]_{t-1}^2$				
$\left[ \frac{1}{(\Delta^2 GGEQ)^2} \right]_{t-1}$		-0.000 [t=-2.963] (0.000)		

Table 16: Model 8 continued.

	$y_9$				$y_9'$			
	Linear Models		Non-linear Models		Linear Models		Non-linear Models	
	PcGets Conservative	RETINA	PcGets Conservative	RETINA	PcGets Conservative	RETINA	PcGets Conservative	RETINA
Parameters retained	1	2	1	2	3	2	6	2
Adj. $R^2$	0.4311	0.4273	0.4311	0.4273	0.9671	0.9638	0.9711	0.9638
$AIC$	3.8206	3.8344	3.8206	3.8344	3.6628	3.7517	3.5523	3.7517
$\hat{\sigma}$	6.7312	6.7538	6.7312	6.7538	6.1768	6.4802	5.7846	6.4802
RMSE	6.7943	6.7910	6.7943	6.7910	6.5240	6.6519	5.8428	6.6519
MAPE	110.19	110.66	110.19	110.66	235.35	415.00	950.94	415.00
Const	-0.1564 (0.590)		-0.1564 (0.590)		0.0346 (0.561)		0.0346 (0.561)	
$y_9'_{t-1} \mid y_9'_{t-1}$					0.7418 (0.011)		0.7398 (0.011)	
$\Delta FM1DQ_t$	446.468 (42.384)	449.377 (43.92)	446.468 (42.384)	449.377 (43.92)	339.176 (102.617)		279.946 (97.621)	
$\Delta DCOINC_{t-1}$					-377.690 (98.032)		-423.297 (93.638)	
$\Delta GNPQ_{t-1}$							-213.336 (76.257)	
$\Delta^2 FMBASE_t$							193.918 (59.694)	
$\Delta GCQ_t$							-0.000 [ $t=-3.25$ ] (0.000)	
$\left[ \frac{1}{(\Delta^2 GGEQ)^2} \right]_{t-1}$								

Table 17: Results for Hoover and Perez (1999) Model 9.

Model		Linear Models		Non-linear Models	
		PcGets (Conservative)	RETINA	PcGets (Conservative)	RETINA
$y1$ [0]	Relevant	-	-	-	-
	Irrelevant	3	2	8 (1)	2
$y2$ [0]	Relevant	-	-	-	-
	Irrelevant	0	2	0	2
$y2'$ [1]	Relevant	0	0	0	0
	Irrelevant	0	2	0	2
$\ln y3$ [1]*	Relevant	1	1	1	1
	Irrelevant	1	1	1	1
$y4$ [1]	Relevant	1	1	1	1
	Irrelevant	4	2	7	1
$y5$ [1]	Relevant	1 (1)	0	1	0
	Irrelevant	7 (5)	2	9	2
$y6$ [2]	Relevant	1	1	1	1
	Irrelevant	4	2	7	1
$y6A$ [2]	Relevant	2	1	0	1
	Irrelevant	5	2	7	1
$y6B$ [2]	Relevant	2	1	2	1
	Irrelevant	6	2	7	1
$y7$ [1]	Relevant	1	1	1	1
	Irrelevant	0	1	2	1
$y7'$ [3]	Relevant	2	2	2	1
	Irrelevant	6	1	10	1
$y8$ [1]	Relevant	1	0	1	0
	Irrelevant	7	2	8	2
$y8'$ [3]	Relevant	3	1	3	1
	Irrelevant	7	3	7	2
$y9$ [2]	Relevant	1	1	1	1
	Irrelevant	0	1	0	1
$y9'$ [5]	Relevant	1	1	1	1
	Irrelevant	2	1	5	1

Table 18: RETINA and PcGets: Summary of results for DGP based on Hoover and Perez (1999). [.] denotes the number of variables in the DGP. (.) denotes results for models in which the diagnostics were switched off. \* indicates that the second lag of the dependent variable in the DGP was not included in the GUM.



	y1	y2	y2	y2	y2	y7	y7	y9	y9	y9
Significance level	0.01	0.01	0.01	0.05	0.05	0.01	0.05	0.01	0.01	0.05
Pre-search reduction tests	-	-	yes	-	yes	-	-	-	yes	-
DGP found	0.9720	0.6020	0.8520	0.0840	0.7690	0.5900	0.1050	0.0000	0.0000	0.0000
Non-DGP var. included	0.0280	0.3980	0.1480	0.9160	0.2310	0.4100	0.8950	0.4200	0.1480	0.8930
DGP var. not included	0.0000	0.0000	0.0000	0.0000	0.0000	0.0030	0.0020	0.9970	0.9990	0.9890
DGP is dominated	0.0260	0.3830	0.1030	0.9080	0.2100	0.3900	0.8900	0.9250	0.9180	0.9680
Specific is dominated	0.0020	0.0150	0.0450	0.0080	0.0210	0.0200	0.0050	0.0010	0.0300	0.0000

Table 19: Hendry and Krolzig (1999); Simulation results for Hoover and Perez with PcGets.



	Linear Model (Benchmark)		PcGets (Liberal)		PcGets (Conservative)		RETINA (Level 0 regressors)	
Sample	100	1000	100	1000	100	1000	100	1000
Parameters	22	22	7	5	5	3	3	3
Adj. $R^2$	0.5775	0.1211	0.5656	0.1208	0.5164	0.1076	0.4415	0.1018
$AIC$	0.2752	0.0068	<i>0.1787</i>	<i>-0.0097</i>	0.2673	0.0033	0.3921	0.0098
$\hat{\sigma}$	1.0427	0.9926	1.0572	0.9927	1.1155	1.0002	1.1988	1.0034
RMSE	1.3008	1.0245	<i>1.0873</i>	1.0173	1.1423	0.9945	1.2539	<i>0.9919</i>
MAPE	149.81	172.92	111.34	181.36	<i>105.48</i>	114.42	126.9	<i>111.8</i>
Const.	−0.2075 (0.128)	0.0140 (0.032)	−0.3169 (0.110)		−0.2730 (0.115)		−0.2383 (0.122)	0.0086 (0.032)
$Y_{t-1}$	−0.1492 (0.108)	−0.0165 (0.032)						
$x_{1,t}$	0.2139 (0.101)	0.0954 (0.033)	0.2189 (0.093)	0.0890 (0.032)	0.0828 (0.032)			
$x_{1,t-1}$	−0.1244 (0.103)	−0.0489 (0.033)						
$x_{2,t}$	0.1881 (0.125)	0.0672 (0.032)	0.3687 (0.103)		0.3316 (0.108)			
$x_{2,t-1}$	0.2666 (0.123)	−0.0018 (0.032)	0.2563 (0.105)					
$x_{3,t}$	0.2532 (0.136)	0.0935 (0.032)	0.0974 (0.032)					
$x_{3,t-1}$	0.1653 (0.141)	−0.0681 (0.032)						
$x_{4,t}$	0.6704 (0.124)	0.1738 (0.030)	0.7328 (0.114)	0.1703 (0.030)	0.6959 (0.117)	0.1675 (0.030)	0.7172 (0.126)	0.1687 (0.030)
$x_{4,t-1}$	0.4667 (0.150)	−0.0137 (0.030)	0.3181 (0.115)		0.2679 (0.117)			
$x_{5,t}$	0.7223 (0.104)	0.2947 (0.030)	0.6393 (0.093)	0.2970 (0.030)	0.6881 (0.098)	0.2864 (0.030)	0.6889 (0.105)	0.2834 (0.030)
$x_{5,t-1}$	−0.0046 (0.124)	−0.0244 (0.032)						
$x_{6,t}$	−0.1798 (0.115)	0.0131 (0.033)						
$x_{6,t-1}$	−0.0681 (0.111)	0.0369 (0.033)						
$x_{7,t}$	0.0532 (0.142)	−0.0228 (0.032)						
$x_{7,t-1}$	−0.0799 (0.139)	−0.0899 (0.033)	−0.0905 (0.032)					
$x_{8,t}$	−0.0080 (0.115)	−0.0265 (0.031)						
$x_{8,t-1}$	0.0557 (0.116)	−0.0313 (0.031)						
$x_{9,t}$	0.1389 (0.123)	0.0169 (0.032)						
$x_{9,t-1}$	0.0896 (0.126)	0.0293 (0.032)						
$x_{10,t}$	0.1830 (0.123)	0.0275 (0.031)						
$x_{10,t-1}$	−0.1021 (0.128)	0.0312 (0.032)	42					

Table 20: RETINA and PcGets: linear models for Krolzig and Hendry (2001).

	PcGets (Liberal)		PcGets (Conservative)		RETINA (Level 1 regressors)	
Sample	100	1000	100	1000	100	1000
Parameters	8	4	3	3	2	2
Adj. $R^2$	0.6009	0.1154	0.4763	0.1076	0.2626	0.0744
$AIC$	<i>0.1031</i>	<i>-0.0045</i>	0.3277	0.0033	0.6602	0.0388
$\hat{\sigma}$	1.0133	0.9958	1.1608	1.0002	1.3774	1.0186
RMSE	<i>1.1126</i>	1.0173	1.1973	<i>1.0104</i>	1.3601	1.0400
MAPE	117.89	179.25	111.47	<i>165.15</i>	<i>99.46</i>	168.33
Const.					-0.1154 (0.139)	0.0140 (0.032)
$x_{1,t}$				0.0828 (0.032)		
$x_{2,t}$	0.4008 (0.100)					
$x_{2,t-1}$	0.3575 (0.104)					
$x_{3,t}$			0.4102 (0.127)			
$x_{4,t}$	0.7832 (0.108)	0.1695 (0.030)	0.7152 (0.121)	0.1675 (0.030)		
$x_{4,t-1}$	0.2915 (0.108)					
$x_{5,t}$	0.6410 (0.090)	0.2799 (0.030)	0.7560 (0.102)	0.2864 (0.030)	0.7228 (0.120)	0.2768 (0.031)
$x_{7,t-1}$		-0.0904 (0.032)				
$(x_1)_{t-1}^2$	-0.1450 (0.044)					
$\left(\frac{1}{x_3 x_5}\right)_t$	0.0088 (0.004)					
$\left(\frac{x_2}{x_3}\right)_{t-1}$	0.0227 (0.007)					
$(x_2 x_5)_{t-1}$		-0.0889 (0.030)				

Table 21: RETINA and PcGets: Non-linear models for Krolzig and Hendry (2001).

Sample	PcGets (Lib)		PcGets (Cons)		RETINA	
	100	1000	100	1000	100	1000
Linear Models						
Total variables retained	7	5	5	3	3	3
Relevant variables retained	4	4	3	3	2	2
Irrelevant variables retained	3	1	2	0	1	1
Non-Linear Models						
Total variables retained	8	4	3	3	2	2
Relevant variables retained	3	2	3	3	1	1
Irrelevant variables retained	5	2	0	0	1	1

Table 22: RETINA and PcGets: Summary of results for DGP based on Krolzig and Hendry (2001).

Strategy	Linear GUM, Linear DGP				Non-lin GUM, Linear DGP				Non-lin GUM, Non-lin DGP			
	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons
Sample Size	100	100	1000	1000	100	100	1000	1000	100	100	1000	1000
DGP found	0.201	0.182	0.180	0.141	0.031	0.031	0.062	0.112	0.010	0.004	0.045	0.055
Non-DGP var. included	0.590	0.167	0.556	0.146	0.817	0.275	0.834	0.288	0.844	0.336	0.830	0.295
DGP var. not included	0.557	0.777	0.583	0.841	0.831	0.943	0.600	0.852	0.937	0.990	0.765	0.938
DGP is dominated	0.713	0.676	0.798	0.795	0.523	0.363	0.890	0.776	0.475	0.241	0.818	0.658
Specific is dominated	0.005	0.008	0.001	0.000	0.000	0.002	0.002	0.004	0.000	0.000	0.003	0.000

Table 23: Summary statistics regarding the probability that PcGets finds the DGP.

Strategy	<i>t</i> -Test: simulated				PcGets, Linear GUM				PcGets average coefficient			
	0.05	0.01	0.05	0.01	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons
Sample	100	100	1000	1000	100	100	1000	1000	100	100	1000	1000
<i>t</i> = 2	0.473	0.214	0.505	0.272	0.608	0.413	0.517	0.276	0.254 (0.082)	0.277 (0.083)	0.089 (0.032)	0.103 (0.032)
<i>t</i> = 3	0.823	0.633	0.861	0.671	0.790	0.632	0.837	0.671	0.334 (0.097)	0.362 (0.097)	0.106 (0.031)	0.115 (0.032)
<i>t</i> = 4	0.966	0.897	0.987	0.927	0.911	0.821	0.978	0.907	0.430 (0.110)	0.460 (0.111)	0.127 (0.032)	0.129 (0.032)
<i>t</i> = 6	1.000	0.995	1.000	1.000	1.000	0.998	1.000	1.000	0.605 (0.103)	0.610 (0.105)	0.188 (0.030)	0.189 (0.030)
<i>t</i> = 8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.802 (0.087)	0.805 (0.089)	0.253 (0.030)	0.252 (0.030)

Table 24: Power and size: PcGets selection probabilities with a linear GUM.

Strategy	PcGets, Non-linear GUM				PcGets average coefficient			
	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons
Sample	100	100	1000	1000	100	100	1000	1000
$t = 2$	0.393	0.193	0.494	0.250	0.280 (0.082)	0.292 (0.083)	0.089 (0.032)	0.103 (0.032)
$t = 3$	0.628	0.465	0.828	0.644	0.366 (0.095)	0.393 (0.099)	0.106 (0.031)	0.116 (0.032)
$t = 4$	0.657	0.490	0.986	0.908	0.472 (0.108)	0.503 (0.111)	0.128 (0.032)	0.130 (0.032)
$t = 6$	0.987	0.960	1.000	1.000	0.600 (0.102)	0.607 (0.108)	0.188 (0.030)	0.189 (0.030)
$t = 8$	1.000	1.000	1.000	1.000	0.798 (0.085)	0.798 (0.091)	0.256 (0.030)	0.255 (0.030)

Table 25: Power and size: PcGets selection probabilities with a non-linear GUM and linear DGP.

Strategy	$t$ -Test: simulated						PcGets, Non-linear GUM						PcGets average coefficient					
	0.05	0.01	0.05	0.01	1000	0.01	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons	Lib	Cons
	100	100	1000	1000	1000	1000	100	100	1000	1000	100	1000	100	100	1000	1000	1000	1000
$t = 2$	0.465	0.239	0.507	0.285	0.285	0.285	0.414	0.219	0.453	0.240	0.309	0.341	0.309	0.341	0.090	0.101	0.090	0.101
$t = 3$	0.805	0.589	0.848	0.659	0.659	0.659	0.621	0.432	0.828	0.653	0.354	0.376	0.354	0.376	0.108	0.118	0.108	0.118
$t = 4$	0.960	0.864	0.979	0.912	0.912	0.912	0.645	0.443	0.985	0.915	0.443	0.457	0.443	0.457	0.129	0.132	0.129	0.132
$t = 6$	1.000	0.997	1.000	1.000	1.000	1.000	0.986	0.966	1.000	1.000	0.597	0.611	0.597	0.611	0.187	0.186	0.187	0.186
$t = 8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.796	0.793	0.796	0.793	0.255	0.254	0.255	0.254
$t = 4 \left( x_1^2 \right)$	0.934	0.828	0.982	0.921	0.921	0.921	0.746	0.595	0.910	0.800	0.208	0.222	0.208	0.222	0.074	0.076	0.074	0.076
$t = 3 \left( x_2 x_3 \right)$	0.782	0.566	0.855	0.670	0.670	0.670	0.441	0.192	0.735	0.523	0.376	0.391	0.376	0.391	0.107	0.118	0.107	0.118
$t = 4 \left( x_4 x_5 \right)$	0.940	0.838	0.971	0.906	0.906	0.906	0.881	0.692	0.990	0.968	0.385	0.396	0.385	0.396	0.126	0.127	0.126	0.127

Table 26: Power and size: PcGets selection probabilities with a non-linear GUM and non-linear DGP.