OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 86, 1 (2024) 0305-9049
doi: 10.1111/obes.12578

A Brief History of General-to-specific Modelling*

DAVID F. HENDRY

Climate Econometrics, Nuffield College, University of Oxford, Oxford, UK
(e-mail: david.hendry@nuffield.ox.ac.uk)

Abstract

We review key stages in the development of general-to-specific modelling (*Gets*). Selecting a simplified model from a more general specification was initially implemented manually, then through computer programs to its present automated machine learning role to discover a viable empirical model. Throughout, *Gets* applications faced many criticisms, especially from accusations of 'data mining'—no longer pejorative—with other criticisms based on misunderstandings of the methodology, all now rebutted. A prior theoretical formulation can be retained unaltered while searching over more variables than the available sample size from non-stationary data to select congruent, encompassing relations with invariant parameters on valid conditioning variables.

I. Introduction

General-to-specific modelling, *Gets*, has a long history of important methodological developments, but has faced many criticisms, especially as being mindless 'data mining', now no longer pejorative, and other criticisms often based on misunderstandings of its properties.

The key stages in *Gets* development are reviewed, leading to its present automated machine learning role. *Gets* can search over more variables than the available sample size from wide-sense non-stationary data to select congruent, encompassing relations with invariant parameters on valid conditioning variables. This brief history draws on many sources where more detailed discussions of various aspects can be found (see Morgan, 1990, and Qin, 1993, 2013, for histories of econometrics and Klein, 1997, for a history of time series analysis). In particular, Ericsson (2004, 2021) discusses the writing of *Dynamic Econometrics* (Hendry, 1995) and the backgrounds to automated *Gets*, indicator saturation estimation, and the development of the implementing software (see Klein, 1987, Doornik and Hendry, 1999, and Renfro, 2009, for histories of econometric computing). Ericsson also reviews a number of my empirical modelling

JEL Classification numbers: C5, C18.

*Financial support from Nuffield College is gratefully acknowledged as are valuable comments on the paper by Anindya Banerjee, Jennie Castle, Jurgen Doornik, Neil Ericsson and Andrew Martinez. Many important contributions to research on *Gets* were made by (*inter alia*) Julia Campos, Jennifer Castle, Jurgen Doornik, Rob Engle, Neil Ericsson, Kevin Hoover, Søren Johansen, Katarina Juselius, Hans-Martin Krolzig, Grayham Mizon, Bent Nielsen, Peter Phillips, Felix Pretis, Jean François Richard and Aris Spanos. All calculations and graphs use *PcGive* and *OxMetrics*.

studies of housing and mortgage markets, consumers' expenditure, money demand, television advertising, and climate change, most jointly with co-authors, and provides an extensive bibliography. Even so, there is too large a number of important empirical applications by others to record here, hence the title 'brief'.

The structure of the paper is as follows. Section II discusses developments till the late 1970s of model selection by general-to-specific modelling (also known as the LSE approach). Section III links *Gets* to encompassing and the theory of reduction as its methodological basis. Section IV describes some of the important later developments of *Gets*, then section V considers the extensions of *Gets* to contracting and expanding multiple-path searches needed once the number of variables (n) to be searched over exceeds the available sample size (T). In particular, this extension is required for indicator saturation estimators. Section VI notes the essential contributions of computer software, and section VII concludes. Throughout, the main criticisms of 'data mining', as data-based model selection was often called, are addressed, as are counter criticisms of simply estimating a theory-specified model.

II. Early developments of *Gets*

This section focuses on three key aspects in the early development of Gets: its comparison with specific-to-general approaches, selection criteria and distinguishing endogenous dynamics from error autocorrelation (COMFAC).

General-to-specific versus specific-to-general

To date, model selection by a general-to-specific approach has been mainly time series focused, especially macroeconomics: see Hendry (2020). Much of the structure of econometrics was formalized by Haavelmo (1944), essentially assuming a viable economic theory context (see Moene and Rødseth, 1991, Spanos, 1989, and Hendry, Spanos and Ericsson, 1989).[1] Although Koopmans (1937) had noted the need for 'a reductionist approach of proceeding from general to simple' as discussed by Hendry and Morgan (1995), the *Gets* story really begins with Sargan (1957), who emphasized the dangers of over-simplification when model building.[2] Sargan was concerned that economic theory was too abstract given the complexities of data, so that large numbers of estimated regression equations excluded relevant variables, as well as misinterpreting tests that had been applied to many hypotheses. Nevertheless, Sargan was a lone voice against a deluge of existing criticisms of anything other than estimating a theory-based specification, facing the early 1940s John Maynard Keynes–Jan Tinbergen debate and 'measurement without theory' by Koopmans (1947) (on both, see Hendry and Morgan, 1995).

However, Anderson (1962) demonstrated that general-to-specific (*Gets*) dominated simple-to-general (*Sig*) when selecting the degree of a polynomial regression, and in Anderson (1971) showed the same conclusion applied to selecting the lag length

[1]However, in Haavelmo (1958), he complained about ''the shortcomings of basic economic theory'', p. 355.
[2]Denis Sargan was discussing a paper by Fisher (1956) in the *Oxford Bulletin*, also discussed by Albert Ando, Franco Modigliani, Milton Friedman, Trygve Haavelmo, Lawrence Klein and James Tobin, all six later Economics Nobel Laureates.

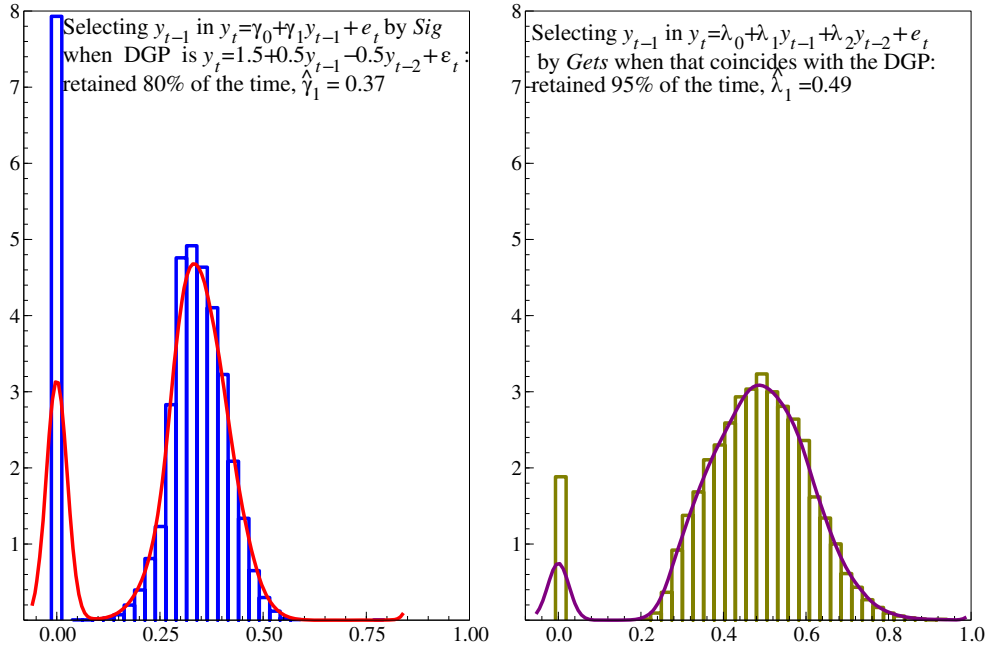

Figure 1. Simulation distributions at $\alpha = 5\%$ for $T = 50$ of (a) $\widetilde{\gamma}_1$ and (b) $\widehat{\lambda}_1$ when the DGP is $y_t = 1.5 + 0.5y_{t-1} - 0.5y_{t-2} + \epsilon_t$ where $\epsilon_t \sim \mathsf{IN}[0, 1]$, using $M = 10{,}000$ replications. [Colour figure can be viewed at wileyonlinelibrary.com]

in an autoregressive model. Figure 1a illustrates *Sig* by the distribution of $\widetilde{\gamma}_1$ in a simulation for $T = 50$ observations selecting y_{t-1} at $\alpha = 5\%$ commencing from the model $y_t = \gamma_0 + \gamma_1 y_{t-1} + u_t$ when the DGP is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2], \tag{1}$$

so includes y_{t-2} (with $\beta_1 = 0.5$, $\beta_2 = -0.5$), but the model does not. Then y_{t-1} was retained 80% of the time with $\widehat{\gamma}_1 = 0.37$. Panel (b) records *Gets* for the distribution of $\widehat{\lambda}_1$ when selecting y_{t-1} from:

$$y_t = \lambda_0 + \lambda_1 y_{t-1} + \lambda_2 y_{t-2} + e_t \tag{2}$$

which coincides with the DGP. Now y_{t-1} is retained 95% of the time with the average value of $\widehat{\lambda}_1 = 0.49$.

Hendry (2003) discussed Sargan (1957) as the stimulus to the central contributions in Sargan (1964), which had many key innovations, especially the use of equilibrium-correction mechanisms (EqCMs), but then called error correction, which assumed stationarity. Sargan followed up his ideas on 'data mining' in Sargan (1973), and Sargan (1981) considered how to choose between sets of regressors. Meantime, his doctoral student, Mizon (1977) extended *Gets* to model selection in general, so hypotheses need not be ordered as with selecting the lag length.

Selection criteria

Methods that aimed to select equations that minimized quadratic loss were proposed by James and Stein (1961) by shrinking parameter estimates in a general specification. Multi-collinearity had long been seen as a serious problem for estimation of general regressions, highlighted by Frisch (1934) (reprinted in Hendry and Morgan, 1995, chapter 23). There are potential benefits to shrinkage when estimating the β_i in (3):

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i z_{i,t} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right], \tag{3}$$

where the mean squared errors (MSEs) of the estimated parameters are:

$$\mathsf{MSE}_i = \mathsf{E}\left[\left(\widehat{\beta}_i - \beta_i\right)^2\right] = \left(\mathsf{E}\left[\widehat{\beta}_i\right] - \beta_i\right)^2 + \mathsf{V}\left[\widehat{\beta}_i\right]. \tag{4}$$

Assuming the $\widehat{\beta}_i$ are unbiased estimates, the MSE is the variance, so shrinkage becomes a trade-off between the resulting bias and reduced variance, as in later shrinkage methods like ridge regression (see Hoerl and Kennard, 1970). In extreme shrinkage, setting $\widehat{\beta}_i = 0$ has a MSE of β_i^2, so could be smaller than $\mathsf{V}[\widehat{\beta}_i]$ in (4) when β_i is small, and is bound to be smaller when $\beta_i = 0$. That same logic applies to *Gets* selection, which omits insignificant $z_{i,t}$. Since linear models like (3) are equivariant under linear (and therefore orthogonal) transforms, the choice of parameterization matters, and merits careful consideration, and misspecification tests should check the initial formulation sustains valid inference and has constant parameters.

Branch and bound algorithms (see e.g., Morrison *et al.,* 2016) evolved from the research by Land and Doig (1960) who had developed an automatic method for solving discrete programming problems. These could be adapted to finding a solution by 'tree' searching, branching, and pruning. In the case of selecting an empirical equation from a general specification, the criterion could be best fit subject to a penalty for the number of variables retained. The tree of all possible combinations of variables is usually searched by partitioning the search space into subsets (i.e., branches) to locate regions that need not be explored (pruning), so the problem becomes computationally feasible.

Stepwise methods, either adding the next most significant variable (*Sig*) or commencing from including all candidates and eliminating the next least significant, had been used for many years and in many disciplines (Leamer, 1978, calls them 'unwise'). They work well when all regressors have low intercorrelations or the number of variables, n, is small, but not otherwise as simulations like those by Murtaugh (1998) show. More generally, single-step, single-path searches are prone to fail both because sets of variables may need to be evaluated jointly, and path dependence may miss better routes. Although out of a historical time sequence, variants of stepwise regression with shrinkage have been developed since, including the least absolute shrinkage and selection operator (Lasso) proposed by Tibshirani (1996) and Efron *et al.* (2004), which have additional penalties for each extra parameter.

Coen, Gomme and Kendall (1969) developed an 'optimal' regression package, essentially finding the best fit within a set of possible regressors. The extensive discussion that

followed their paper raised many issues critical of their approach, including the need for misspecification tests, and also several that would later be salient, namely potential institutional change, and that a better model may not produce better forecasts. In an experimental setting, Cox and Snell (1974) distinguish between models for prediction and explanation, and propose estimating all possible models when that is feasible ($n \leq 15$), rejecting models that are invalid reductions of the full model (but without diagnostic testing).

Other model selection approaches were proposed by Akaike (1973, 1981) called the Akaike information criterion (AIC) that sought to find the 'optimal' forecasting relation from an infinite stationary autoregression; followed by the Schwarz Criterion by Schwarz (1978) (SC or BIC, for Bayesian information criterion), that ensured a consistent selection for finite constant parameters as the sample size diverged; Hannan and Quinn (HQ: 1979) who showed the minimum penalty to ensure a consistent selection;[3] and the posterior information criterion (PIC) by Phillips and Ploberger (1994). AIC, BIC and SC rarely included misspecification tests, so while useful for finding the best model when the DGP is a stationary autoregression, their application in other contexts does not ensure the selected model will be well specified.

COMFAC: endogenous dynamics versus error autocorrelation

Hendry (1970) had factorized system dynamics between variables and autoregressive errors using polynomial matrices, a precursor to COMFAC for partitioning between dynamics and error autocorrelation, which Hendry and Mizon (1978) explained because the theory was not published till Sargan (1980). The key issue was that in the econometrics literature, autoregressive errors were viewed as a generalization of random errors, whereas COMFAC showed they were a testable restriction on a more general dynamic equation. In the simplest setting:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + \epsilon_t, \tag{5}$$

where $\epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2]$ with $|\beta_2| < 1$, when $\beta_3 = -\beta_1\beta_2$, (5) can be rewritten as:

$$y_t = \beta_0 + \beta_1 z_t + u_t \text{ where } u_t = \rho u_{t-1} + \epsilon_t, \tag{6}$$

so $\rho = \beta_2$. The restriction $\beta_3 = -\beta_1\beta_2$ in (5) can be tested to determine whether it is a static equation with a first-order autoregressive error as in (6) or the more general dynamic equation. When there are longer lags and more variables, the testing and separating into endogenous dynamics and error autocorrelation is more complicated, hence the need for the special COMFAC algorithm Sargan created.

The distinction matters for policy since the mean lag response time for y_t when z_t is changed is zero in (6) resulting in a long-run response of β_1, whereas the mean lag is $m = (\kappa\beta_2 + \beta_3)/\kappa(1 - \beta_2)$ in (5), where the long-run change is $\kappa = (\beta_1 + \beta_3)/(1 - \beta_2)$. Alternative (5) can be expressed as the EqCM:

$$\Delta y_t = \beta_0 + \beta_1 \Delta z_t - (1 - \beta_2)(y_{t-1} - \kappa z_{t-1}) + \epsilon_t. \tag{7}$$

[3]The penalty at which a marginal variable is retained can be translated into the implicit significance level of the associated test: see, e.g., Campos, Hendry and Krolzig (2003).

When $\kappa = 1$ in (7), $m = (1 - \beta_1)/(1 - \beta_2)$ so any adjustment remaining after the initial β_1 is removed at $(1 - \beta_2)$ per period.

III. Encompassing and the theory of reduction

An intermediate stage in the development of *Gets* occurred with the formulation of encompassing and the theory of reduction. Theory and empirical practice proceeded in tandem.

Empirically, Davidson *et al.* (1978) (DHSY) sought to explain the plethora of different UK consumption function estimates by a more general formulation (see Ericsson, 2004, for more details). Almost simultaneously, Hendry (1979b) showed how predictive failure could arise from over-simplification when the distributions of the omitted variables shifted. Moreover, that source of predictive failure was avoidable by including those variables, whereas predictive failure from unanticipated structural breaks (such as oil crises, and pandemics) was unavoidable.

The theory of reduction provides the methodological basis of general-to-specific and originated with research at CORE by Richard (1980) and Florens and Mouchart (1980). A visit by Richard to LSE in the mid-1970s and my later visit to CORE led to our papers on the principles of model formulation in Hendry and Richard (1982, 1983). Available information was partitioned into the disjoint sets of past, present and future observations of the investigator's variables, other data of rival hypotheses and theory information. Each set had associated concepts, including innovation errors, exogeneity and congruence, parameter constancy, encompassing and data admissible (although congruence was only named as such in Hendry, 1985, and the clarification of 'Exogeneity' was not published till Engle, Hendry and Richard, 1983).

However, the intrusion of unit-root non-stationarity into econometrics was changing modelling practices. Disagreements at a 1975 conference (published by Sims, 1977) between myself and Granger and Newbold about differencing data before modelling led to Engle and Granger (1987), with EqCMs re-appearing as an equilibrium correction solution that embedded cointegration (see Hendry, 2004). The original cointegration estimator in Engle and Granger (1987) was based on static models, so like (6) enforced COMFAC. However, simulation experiments by Banerjee *et al.* (1986) revealed that such a *Sig* approach was less efficient than *Gets* from a more general EqCM model using the *PcGive* test for the existence of long-run relationships, now with different critical values under the null calculated by Ericsson and MacKinnon (2002).[4]

Nevertheless, opposition remained to *Gets* as 'mindless data mining' in the sense of finding adventitiously significant coefficients due to testing many effects and not accounting for that in significance levels (see Leamer, 1978). There were also claims about 'pretest biases' when selecting, as against estimating the correct model (Judge and Bock, 1978). However, 'pretest' effects are not much changed by *Gets* search because conditional coefficient distributions are similar when selecting from the DGP and a

[4]While writing a version of *PcGive* in the late 1980s, I experimented with an 'expert system' that made modelling decisions based on the empirical results available, but abandoned it as it was essentially a single path and did not perform well.

nesting general unrestricted model (GUM). Such criticisms were supported by simulation results like Lovell (1983) apparently showing 'data mining' was disastrous using inappropriate selection criteria like 'best fit'; and Denton (1985) generalizing that to criticising an 'industry'. Conversely, Campos and Ericsson (1999, Table 1) review four senses of 'pejorative data mining' (namely repeated testing, data interdependence, corroboration and over-fitting), and provide refutations of each in terms of 'constructive data mining' (e.g., through recursive estimation, testing super-exogeneity and encompassing, *Gets* modelling, and using highly informative data by not omitting crises and breaks).

Criticisms of not simply estimating a complete theoretical model of the phenomena of interest also included data inaccuracy, its intermittent revisions and small sample sizes, but not non-stationarity that endangered theoretical derivations most of which assumed stationarity. Many of those criticisms applied to all empirical modelling, but were less relevant to a structured approach like *Gets*, which avoided path dependence (see Pagan, 1987). By attacking model selection and general starting points (KISS, keep it simple stupid, later revised to keep it sophisticatedly simple, as many simple models are stupid: Zellner, 1992), essentially only theory-based empirical results were published. Thus, in a neat 'Catch 22', it became easy to criticize econometrics for adding nothing to knowledge, as with Summers (1991) 'scientific illusion in empirical macroeconomics', rebutted by Juselius (1993).

Another arena where empirical macromodels were being criticized was forecast failure, where 'naive' predictors outperformed large macro systems, exemplified by Nelson (1972). However, three crucial insights were lacking at the time. First, forecast failure will only occur rarely in a stationary DGP, even after a model is selected, as the in-sample fit will match the out-of-sample: see Miller (1978) and Hendry (1979a). Thus, for systematic forecast failure, parameter changes, location shifts, trend breaks or omitted non-linearities are needed, any of which could be due to included or incorrectly excluded relevant variables.

Second, many so-called 'naive' predictors are actually robust after shifts in that they quickly move back on track, unlike EqCMs that instead try to return to the previous equilibrium: see Hendry (2006) and Ericsson (2017). EqCMs continue to be used after forecast failure has occurred, despite that serious flaw. An extreme example is the Office of Budget Responsibility (OBR), which continued for a decade to make 5-year ahead forecasts in the first and third quarters of each year for the UK's productivity (measured as output per hour worked) that were dramatically above actual productivity, seemingly 'correcting' to the trend before 2011, as seen in the recreated Figure 2a based on Martinez *et al.* (2022). The smooth random walk forecasts in Panel (b) use local estimates over the previous 16 observations of the long-run equilibrium since the break. To mimic the OBR, both sets of forecasts are from an AR(1) model with a trend and a constant, re-estimated for each vintage. Smooth random walk forecasts were the best of the robust predictors considered by the authors, although most substantially outperformed the OBR.

Finally, every forecast failure entails a theory-model failure, especially for dynamic-stochastic general equilibrium systems (DSGEs). Hendry and Mizon (2014) prove that when distributions shift, conditional expectations are neither unbiased nor minimum mean-square error predictors, and the law of iterated expectations is invalid. Thus, forcing DSGE equations on data leads to a non-structural representation, exemplified by the 'breakdown' of the Bank of England quarterly econometric model (BEQEM: see

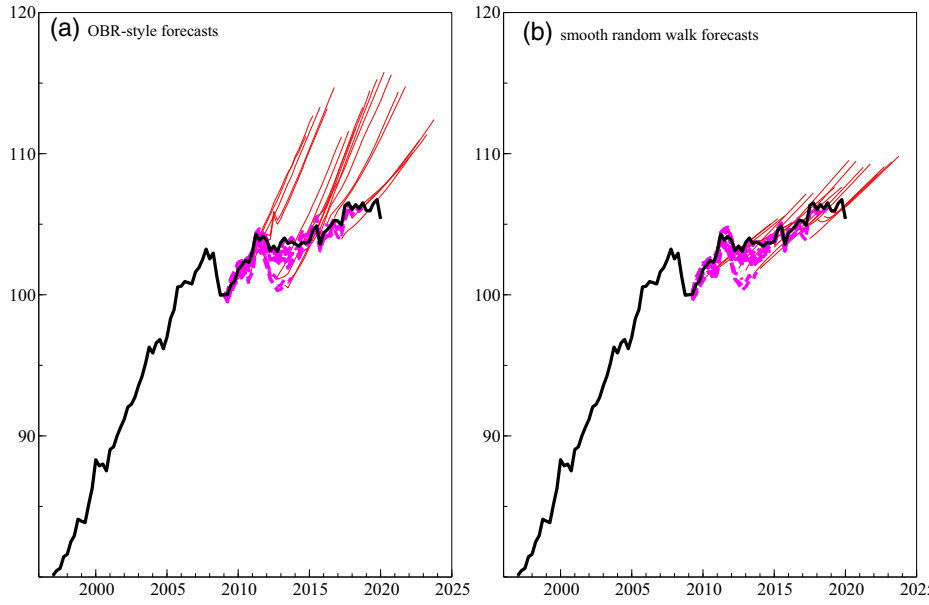

Figure 2. (a) OBR-style and (b) smooth random walk forecasts of UK productivity. The dashed lines are real-time data vintages starting in 2009(1), the thick black line shows the most recent actual values, and the thinner solid lines are forecasts starting in the first quarter of 2010 and updated every two quarters till March 2019. . [Colour figure can be viewed at wileyonlinelibrary.com]
Source: Martinez, Castle and Hendry (2022)

Harrison *et al.,* 2005) yet replaced by another DSGE (see Burgess *et al.,* 2013) which also fails on past data. Hendry (2018) proposes methods for deciding between alternative approaches, and Hendry and Muellbauer (2018) provide an extensive critique of Bank of England econometric modelling.

IV. Later developments of *Gets*

The state of the debate about *Gets* and model selection in general was reviewed by Gilbert (1986) in a counter critique to *Sig*-type approaches, comparing *Gets* favourably with the 'average economic regression' (AER: also see Hendry and Ericsson, 1991, and the discussion in Ericsson, 2004). The concept of encompassing was formalized by Mizon and Richard (1986), where a related development had been non-nested hypothesis tests (see Cox, 1962) that could be re-interpreted as error-variance encompassing (see Hendry, Marcellino and Mizon, 2008b, for recent overviews). Encompassing together with the theory of reduction in Hendry (1987) and proofs by White (1990) that with sufficiently rigorous testing, the selected model will converge to the DGP, provided a theoretical basis for model selection.[5] A local DGP (LDGP) arises from reductions of the set of variables

[5]To quote from White (1990), p. 381: ''Thus, we have an m-testing model selection procedure which rejects sufficiently mis-specified models and retains sufficiently correctly specified models with confidence approaching certainty as [the sample size becomes large]. . . . For this reason we wholeheartedly endorse progressive research strategies such as that of Hendry and Richard (1982) and Hendry (1987) for arriving at sufficiently well specified

that define their DGP, such as aggregation, marginalization, sequential factorization, etc., summarized in Hendry and Doornik (2014). Valid reductions entail no loss of information, and in fact correspond to the main concepts of econometrics, matching those of the data partition above, and also including sufficient statistics, Granger non-causality, cointegration, identified and invariant parameters (see Engle and Hendry, 1993).

The LDGP is the best representation one can obtain given the selection of variables, so must be the target for model selection, hence the general unrestricted model (GUM) should be the best approximation to the LDGP. All empirical models are implicit reductions of the DGP, though rarely viewed that way, and should parsimoniously encompass the LDGP by being congruent, as the DGP must be both congruent with itself and encompass all simplifications. A key difference between just selecting the 'best model' by some criterion and *Gets* is the role of misspecification tests in evaluating the latter to exclude non-congruent models. Rebuttals of criticisms of 'data mining' can pose the counter question of what to do if a model is rejected? It is not valid to infer the alternative to a test that rejects as there may be many other reasons for the rejection, and arbitrarily assuming one will not lead to productive modelling. Any empirical models that fail misspecification tests at stringent critical values must be non-congruent.

It is essential to carefully distinguish two roles of misspecification tests: an initial one-off application is to test the congruence of the general model, which if satisfied, means later use as diagnostics should avoid invalid reductions during simplification, 'not repeated testing', thereby improving the model selection algorithm (see Campos, Ericsson and Hendry, 2005 for a detailed discussion and an extensive bibliography).

After a Carnegie-Rochester conference debate between Hendry (1997) and Faust and Whiteman (1997), Hoover remained sceptical about *Gets* modelling, so with Perez sought to evaluate *Gets* by simulating its performance following a small number of different simplification paths. Published as Hoover and Perez (1999), their *Gets* algorithm actually showed excellent performance. This led to automated selection following many paths as in *PcGets* (Hendry and Krolzig, 1999), and more thoroughly with *Autometrics* by Doornik (2009). Moreover, Hoover and Perez (2000) argue that *Gets* successfully operationalizes model selection methodology: to quote 'intelligent data mining is an important element in empirical investigation in economics'. Indeed, Hendry and Krolzig (2004, 2005) show how powerful and versatile *PcGets* is, and that the discrimination between relevant and irrelevant variables is excellent even selecting at $\alpha = 5\%$, as seen from their Figure 2, reproduced as Figure 3 here. The DGP is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \sum_{r=1}^{10}\sum_{s=0}^{1} \gamma_{r(s+1)} z_{r,t-s} + \epsilon_t \;\text{ where }\; \epsilon_t \sim \mathsf{IN}[0, \sigma_\epsilon^2] \tag{8}$$

and $z_{r,t} \sim \mathsf{IN}[0, \sigma_{z_r}^2]$. $\gamma_r = r + 1$ for $r = 1, \ldots, 4$, & $\gamma_5 = 6$, with all other parameters equal to zero.

Doornik (2009) provides a comprehensive description of *Autometrics*, which has many clever features beyond the outline here. The automatic searches use a tree representation

characterizations of the DGP. We believe the m-testing framework set forth here can be a convenient vehicle for such strategies.'' Campos *et al.* (2003) show this also applies to *Gets*.

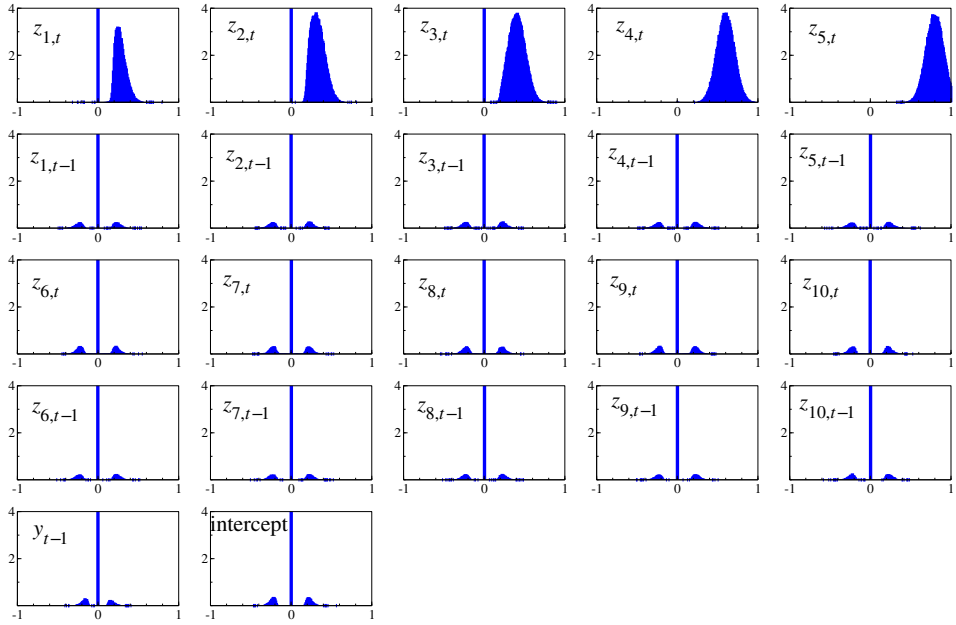

Figure 3. Unconditional distributions of $\{\widehat{\beta}_i\}$ in (8) from a simulation, selecting at $\alpha = 5\%$, using $M = 10{,}000$ replications. The vertical bar at 0 shows the cases where the estimated coefficient was insignificant: every graph y-axis maxima are set to 4, so distributions of chance significant cases can be seen, and all x-axes are $[-1, 1]$. . [Colour figure can be viewed at wileyonlinelibrary.com]
Source: Hendry and Krolzig (1999)

of the model space: this avoids estimating a model more than once, and only requires remembering one reduction path from the GUM (along which it is possible to backtrack if an encompassing test fails), and is able to implement shortcuts along a path. As the 2^n unique subsets of models are usually too large to search completely, pruning removes 'dead' branches when a reduction step fails diagnostic tests, with chopping permanently removing a path as no longer needing exploration when its starting point is highly insignificant. Bunching collects groups of branches, so their relevance can be tested by F-tests to enable block searches: all these decisions can be checked at several significance levels, including diagnostic checks on reductions. Information criteria are used to select between terminal undominated congruent models when more than one is found, and sometimes for lag-length selection. Thus, *Autometrics* is both a descendent of several streams of earlier research on model selection (branch and bound, information criteria, likelihood, mis-specification testing), and 'internal' innovations like encompassing (see Doornik, 2008), theory of reduction and indicator saturation (discussed below) while avoiding the mistakes of other methods like stepwise; and by handling distribution shifts and unit-root non-stationarities is a machine learning algorithm for time series models.

The empirical gauge (see Hendry and Santos, 2010) is the actual retention rate of irrelevant variables during selection (say 1%), and *Autometrics* has a gauge close to nominal significance for tight levels: Johansen and Nielsen (2016) derive the distribution of gauge estimates which corroborates that finding. The empirical potency is the actual retention rate of relevant variables, and is usually close to the power of a t-test on

that estimated parameter given its non-centrality. Hendry and Doornik (2014) compare *Autometrics* gauge and potency on the experiments designed by Hoover and Perez (1999) with selecting by stepwise regression, information criteria, Lasso and RETINA (see Perez-Amaral, Gallo and White, 2003) and show *Autometrics* generally outperforms those alternatives. Re-running the simulation experiments in Hendry and Krolzig (2005) over a range of negative and positive correlated regressors delivers similar results.

A different criticism by Leeb and Pötscher (2003) is that model selection is distortionary when eliminating low significance variables correlated with very significant: also see their less technical Leeb and Pötscher (2005). The GUM in their analysis coincides with the DGP, so is essentially about inference in the DGP. Eliminating low-significance variables is bound to occur as the cost of avoiding the more serious alternative of under-specification from omitting important influences not included in the GUM. *Gets* concerns searching for a model of the LDGP when many of its features are unknown, including lag lengths, functional forms, and especially the numbers, magnitudes and timings of shifts, as well as which variables matter, hence involves a process of discovery commencing from a GUM intended to be more general than the LDGP. Phillips (2005) has also implemented an automated approach to data-based model discovery, building on Phillips (1996).

V. Beyond *Gets* to model discovery from more variables than observations

Once the number of variables, n, in the GUM exceeds the available sample size T, the initial GUM cannot be estimated, and instead contracting and expanding multiple-path searches (CEMPS) are needed. Indicator saturation estimation (ISE) will always create $n > T$ when an indicator for every observation is included in the initial GUM. Impulse indicator saturation (IIS) was the first ISE after a serendipitous discovery of its feasibility when analysing US food data in Hendry (1999). Most investigators in the Magnus and Morgan 'experiment' had only used data after 1946, so to replicate their findings while retaining all the data, I created zero-one dummies to almost saturate the earlier period, selecting major outliers for 1931–33 and war-time rationing. Retaining just these led to a model that was constant over the whole period on a Chow (1960) test for the postwar data. Salkever (1976) had shown that the test was equivalent to testing the significance of impulse indicators over the relevant period. Thus, as many impulse indicators as observations had been entered in two large blocks: see Ericsson (2004) and Hendry (2009).

Formal proofs of the benign properties of 2-block IIS in simple DGPs were provided by Hendry, Johansen and Santos (2008a), generalized by Johansen and Nielsen (2009, 2013) to dynamic regressions, including for unit-root non-stationarity, who also show it is a robust method (see Huber, 1964). Castle, Doornik and Hendry (2012) illustrate how IIS can handle outliers, which may be drawn from 'fat-tailed distributions' including a t_3 with only two finite moments. Figure 4 shows the conditional distributions of parameter estimates in a model with t_3 errors for $T = 100$ selecting at $\alpha = 1\%$ with and without IIS for $n = 10, k = 5$ relevant ($\beta_i = 1,\ i = 1, \ldots, 5$) and 5 irrelevant ($\beta_i = 0,\ i = 6, \ldots, 10$) variables. The conditional distributions are tighter after IIS-based selection, and MSEs generally smaller, especially for irrelevant variables.

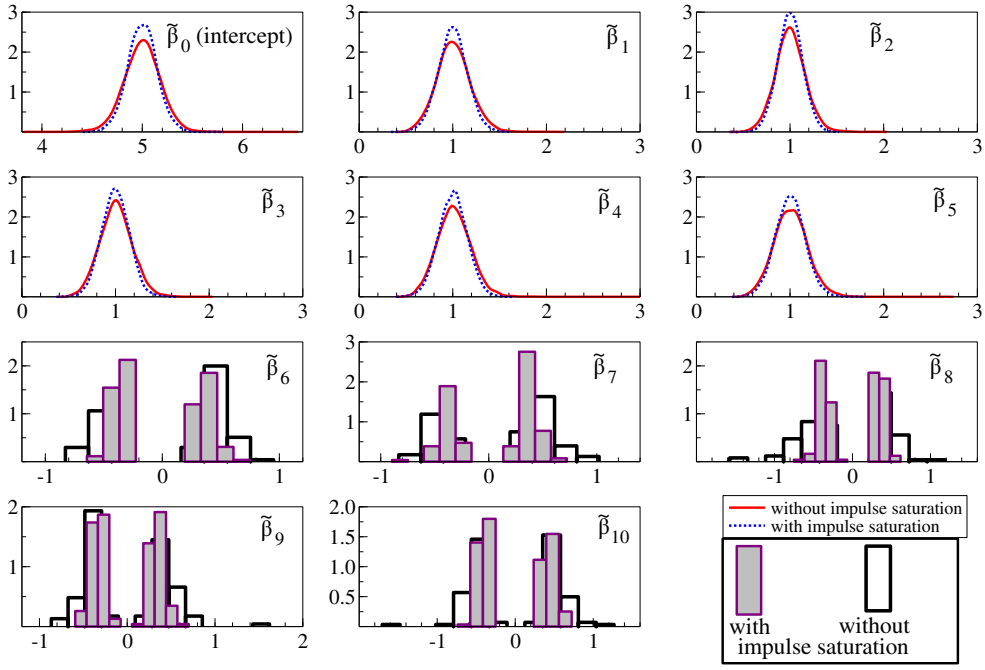

Figure 4. Distributions of estimates conditional on retention with (solid, shaded) and without (dotted, white) IIS when $n = 10$ with a t_3 error distribution selecting at $\alpha = 1\%$, $M = 10{,}000$. . [Colour figure can be viewed at wileyonlinelibrary.com]
Source: Castle *et al.* (2012)

The class of ISEs was extended to step-indicator saturation (SIS) then trend-indicator saturation (TIS) by Castle *et al.* (2015, 2019b), with a medical application of TIS in Walker *et al.* (2019), which elicited a favourable editorial by the BMJ. ISEs are versatile and can be generalized to designed indicators, as with ν-shaped to detect volcanic eruptions from the dendrochronological temperature record (see Pretis *et al.,* 2016). Parameter shifts can be detected using multiplicative-indicator saturation (see Castle, Doornik and Hendry, 2019a), where every regressor is interacted with step indicators; or after finding location shifts by SIS, interacting retained SIS indicators just with the lagged regressand can help discriminate between direct shifts and those induced by changes in dynamics (see Castle, Doornik and Hendry, 2023).

Hendry and Johansen (2015) show that a theory-based model can be retained as the null hypothesis while searching at tight significance levels over additional regressors, and by orthogonalization, this approach can deliver the same estimates as direct fitting, so *Gets* keeps the best of theory-based and data-based model selection, the former if no additional variables are retained, the latter by an improved model if some are incorporated.

VI. Computing

When commencing from very large GUMs, *Gets* is infeasible for humans alone and needs powerful software. Fortunately, this has been available for many years, first manually

in the computer program library *AUTOREG* (Hendry and Srba, 1980, where the name refers to autoregressive not automatic), followed by *PcGive*, then in automatic versions in *PcGets* (Hendry and Krolzig, 2001), and more recently in *Autometrics* (Doornik, 2009) and XLModeler, since extended to ISEs within *PcGive* (Doornik and Hendry, 2021, and *OxMetrics*, Doornik, 2021b).

Automatic selection with ISEs is now also implemented in an R package by Pretis, Reade and Sucarrat (2018) and in EViews. Simulations of how CEMPS performs in a variety of settings can be undertaken in the menu-driven program, *PcNaive*, Doornik and Hendry (2018), more generally in *Ox* (Doornik, 2021a), and in R, all of which allow generating artificial data facsimiles of that used for the empirical model.

VII. Conclusion of this brief history

Castle, Doornik and Hendry (2021) emphasize that successful empirical modelling requires discovering both the relevant determinants and the reliable observations from which to estimate models. They illustrate their robust model discovery approach on two popular test datasets. The GUM must include all substantively relevant variables (plus dynamics in time series), and any nonlinear reactions, have valid conditioning with invariant parameters, and handle distributional shifts, outliers and fat-tailed distributions.

Common modelling approaches—such as classical econometrics, information criteria, robust statistics, and non-parametric methods—each address a single 'problem', in contrast to automatic *Gets* modelling, re-framed as a discovery process and part of a progressive research strategy. Those four approaches can be characterized methodologically in a stylized manner as follows:

Classical econometrics: a covert and unstructured empirical discovery process, often theory based, seeking the **'best' estimates** from assuming: uncontaminated observations $\mathcal{T}$, the correct functional form for all n valid conditioning variables $\mathbf{z}_t$ with constant parameters $\boldsymbol{\beta}$, and independent of the errors $\{\epsilon_t\}$, where $\epsilon_t \sim \mathsf{IID}\left[0, \sigma_\epsilon^2\right]$ denoting independently and identically distributed. After estimation, compute tests to discover if there are departures from the assumptions, and if so, follow recipes for 'fixing' them.

Information criteria: discover the **'best' model** from assumed accurate data over $\mathcal{T}$, given the correct functional form for the complete set of n valid conditioning $\mathbf{z}_t$ with constant $\boldsymbol{\beta}$ and independent of the IID (white-noise) errors. But the selected 'best model' may be a poor approximation to the LDGP if not evaluated.

Robust statistics: discover the **'best' sample** by selecting over $\mathcal{T}$, given the correct functional form for the complete set of relevant valid conditioning variables $\mathbf{z}_t$ with constant $\boldsymbol{\beta}$ independent of the IID errors. All difficulties assumed away still need testing, and must be fixed if found.

Non-parametric methods: discover the **'best' functional form** or distribution, assuming accurate data, a correct set of valid conditioning variables $\mathbf{z}_t$ with constant $\boldsymbol{\beta}$ and independent of the IID errors.

Thus, each of these approaches essentially assumes away what the other approaches tackle, whereas successful empirical modelling requires tackling all the problems jointly, as with *Gets*.

We reviewed the key stages that have led to *Gets* automated machine learning role for the given non-stationary time series under analysis, able to search over more variables than the available sample size to select congruent, encompassing relations with invariant parameters on valid conditioning variables. The approach allows a theory model to be retained with the same estimates as directly fitting, yet stringently tested jointly against all the likely seminar questions of the type 'did you try x_t?' before the seminar and know the answer: win–win if either all the criteria are met or an improved model if not.

A surprising aspect of the *Gets* research program is why it is not more widely adopted. The criticisms of *Gets* have all been rebutted, yet the flaws of not doing *Gets* remain ignored. As cited decades ago in Hendry (1987), the Guinness advert 'I have never tried it because I don't like it' still seems to apply!

Final Manuscript Received: September 2023; Accepted 27September 2023

References

Akaike, H. (1973). 'Information theory and an extension of the maximum likelihood principle', in Petrov B. N. and Csaki F. (eds), *Second International Symposium on Information Theory*, Akademia Kiado, Budapest, pp. 267–281.

Akaike, H. (1981). 'Likelihood of a model and information criteria', *Journal of Econometrics*, Vol. 16, pp. 3–14. https://doi.org/10.1016/0304-4076(81)90071-3.

Anderson, T. W. (1962). 'The choice of the degree of a polynomial regression as a multiple-decision problem', *Annals of Mathematical Statistics*, Vol. 33, pp. 255–265 https://www.jstor.org/stable/2237656.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, John Wiley & Sons, New York.

Banerjee, A., Dolado, J. J., Hendry, D. F. and Smith, G. W. (1986). 'Exploring equilibrium relationships in econometrics through static models: some Monte Carlo evidence', *Oxford Bulletin of Economics and Statistics*, Vol. 48, pp. 253–277. https://doi.org/10.1111/j.1468-0084.1986.mp48003005.x.

Burgess, S., Fernandez-Corugedo, E., Groth, C., Harrison, R., Monti, F., Theodoridis, K. and Waldron, M. (2013). *The Bank of England's Forecasting Platform: COMPASS, MAPS, EASE and the Suite of Models*. Working paper no. 471 and Appendices, Bank of England, London.

Campos, J. and Ericsson, N. R. (1999). 'Constructive data mining: modeling consumers' expenditure in Venezuela', *Econometrics Journal*, Vol. 2, pp. 226–240. https://doi.org/10.1111/1368-423X.00029.

Campos, J., Ericsson, N. R. and Hendry, D. F. (2005). 'Editors' introduction', in Campos J., Ericsson N. R., and Hendry D. F. (eds), *Readings on General-to-Specific Modeling*, Edward Elgar, Cheltenham, pp. 1–81.

Campos, J., Hendry, D. F. and Krolzig, H.-M. (2003). 'Consistent model selection by an automatic *Gets* approach', *Oxford Bulletin of Economics and Statistics*, Vol. 65, pp. 803–819. https://doi.org/10.1046/j.0305-9049.2003.00092.x.

Castle, J. L., Doornik, J. A. and Hendry, D. F. (2012). 'Model selection when there are multiple breaks', *Journal of Econometrics*, Vol. 169, pp. 239–246. https://doi.org/10.1016/j.jeconom.2012.01.026.

Castle, J. L., Doornik, J. A. and Hendry, D. F. (2019a). *Multiplicative-Indicator Saturation*. Working paper, Nuffield College, Oxford University.

Castle, J. L., Doornik, J. A. and Hendry, D. F. (2021). 'Robust discovery of regression models', *Econometrics and Statistics*, Vol. 26, pp. 31–51. https://doi.org/10.1016/j.ecosta.2021.05.004.
Castle, J. L., Doornik, J. A., and Hendry, D. F. (2023). 'Improving models and forecasts after equilibrium-mean shifts', *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2023.09.006.
Castle, J. L., Doornik, J. A., Hendry, D. F. and Pretis, F. (2015). 'Detecting location shifts during model selection by step-indicator saturation', *Econometrics*, Vol. 3, pp. 240–264. https://doi.org/10.3390/econometrics3020240.
Castle, J. L., Doornik, J. A., Hendry, D. F. and Pretis, F. (2019b). *Trend-Indicator Saturation*. Working paper, Nuffield College, Oxford University.
Chow, G. C. (1960). 'Tests of equality between sets of coefficients in two linear regressions', *Econometrica*, Vol. 28, pp. 591–605. https://doi.org/10.2307/1910133.
Coen, P. G., Gomme, E. D. and Kendall, M. G. (1969). 'Lagged relationships in economic forecasting', *Journal of the Royal Statistical Society A*, Vol. 132, pp. 133–163 https://www.jstor.org/stable/2343782.
Cox, D. R. (1962). 'Further results on tests of separate families of hypotheses', *Journal of the Royal Statistical Society B*, Vol. 24, pp. 406–424. https://doi.org/10.1111/j.2517-6161.1962.tb00468.x.
Cox, D. R. and Snell, E. J. (1974). 'The choice of variables in observational studies', *Applied Statistics*, Vol. 23, pp. 51–59 https://www.jstor.org/stable/2347053.
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*, Vol. 88, pp. 661–692. https://doi.org/10.2307/2231972.
Denton, F. (1985). 'Data mining as an industry', *Review of Economics and Statistics*, Vol. 67, pp. 124–127 https://www.jstor.org/stable/1928442.
Doornik, J. A. (2008). 'Encompassing and automatic model selection', *Oxford Bulletin of Economics and Statistics*, Vol. 70, pp. 915–925. https://doi.org/10.1111/j.1468-0084.2008.00536.x.
Doornik, J. A. (2009). 'Autometrics', in Castle J. L. and Shephard N. (eds), *The Methodology and Practice of Econometrics*, Oxford University Press, Oxford, pp. 88–121. https://doi.org/10.1093/acprof:oso/9780199237197.003.0004.
Doornik, J. A. (2021a). *Object-Oriented Matrix Programming using Ox* 9th ed., Timberlake Consultants Press, London.
Doornik, J. A. (2021b). *OxMetrics: An Interface to Empirical Modelling* 9th ed., Timberlake Consultants Press, London.
Doornik, J. A. and Hendry, D. F. (1999). 'The impact of computational tools on time-series econometrics', in Coppock T. (ed), *Information Technology and Scholarship*, Oxford University Press, Oxford, pp. 257–269.
Doornik, J. A. and Hendry, D. F. (2018). *Interactive Monte Carlo Experimentation in Econometrics using PcNaive: Volume IV*, Timberlake Consultants Press, London.
Doornik, J. A. and Hendry, D. F. (2021). *Empirical Econometric Modelling using PcGive: Volume I* 9th ed., Timberlake Consultants Press, London.
Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). 'Least angle regression', *The Annals of Statistics*, Vol. 32, pp. 407–499. https://doi.org/10.1214/009053604000000067.
Engle, R. F. and Granger, C. W. J. (1987). 'Cointegration and error correction: representation, estimation and testing', *Econometrica*, Vol. 55, pp. 251–276 https://www.jstor.org/stable/1913236.
Engle, R. F. and Hendry, D. F. (1993). 'Testing super exogeneity and invariance in regression models', *Journal of Econometrics*, Vol. 56, pp. 119–139. https://doi.org/10.1016/0304-4076(93)90103-C.
Engle, R. F., Hendry, D. F. and Richard, J.-F. (1983). 'Exogeneity', *Econometrica*, Vol. 51, pp. 277–304. https://doi.org/10.2307/1911990.
Ericsson, N. R. (2004). 'The ET interview: Professor David F. Hendry', *Econometric Theory*, Vol. 20, pp. 745–806. https://doi.org/10.1017/S0266466604204078.
Ericsson, N. R. (2017). 'Economic forecasting in theory and practice: an interview with David F. Hendry', *International Journal of Forecasting*, Vol. 33, pp. 523–542. https://doi.org/10.1016/j.ijforecast.2016.10.001.

Ericsson, N. R. (2021). 'David F. Hendry (1944–)', in Cord R. (ed), *The Palgrave Companion to Oxford Economics*, Palgrave Macmillan, London, pp. 563–622. https://doi.org/10.1007/978-3-030-58471-9_24.
Ericsson, N. R. and MacKinnon, J. G. (2002). 'Distributions of error correction tests for cointegration', *Econometrics Journal*, Vol. 5, pp. 285–318. https://doi.org/10.1111/1368-423X.00085.
Faust, J. and Whiteman, C. H. (1997). 'General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model of the DGP: a translation and critique', *Carnegie–Rochester Conference Series on Public Policy*, Vol. 47, pp. 121–161. https://doi.org/10.1016/S0167-2231(98)00007-4.
Fisher, M. R. (1956). 'Exploration in saving behaviour', *Oxford Bulletin of Economics and Statistics*, Vol. 18, pp. 201–278. https://doi.org/10.1111/j.1468-0084.1956.mp18003001.x.
Florens, J.-P. and Mouchart, M. (1980). *Initial and Sequential Reduction of Bayesian Experiments*. Discussion paper 8015, CORE, Louvain-La-Neuve, Belgium.
Frisch, R. (1934). *Statistical Confluence Analysis by means of Complete Regression Systems*, University Institute of Economics, Oslo. https://doi.org/10.1017/CBO9781139170116.026.
Gilbert, C. L. (1986). 'Professor Hendry's econometric methodology', *Oxford Bulletin of Economics and Statistics*, Vol. 48, pp. 283–307. https://doi.org/10.1111/j.1468-0084.1986.mp48003007.x.
Haavelmo, T. (1944). 'The probability approach in econometrics', *Econometrica Supplement*, Vol. 12, pp. 1–118 https://www.jstor.org/stable/1906935.
Haavelmo, T. (1958). 'The role of the econometrician in the advancement of economic theory', *Econometrica*, Vol. 26, pp. 351–357 https://www.jstor.org/stable/1907616.
Hannan, E. J. and Quinn, B. G. (1979). 'The determination of the order of an autoregression', *Journal of the Royal Statistical Society B*, Vol. 41, pp. 190–195. https://doi.org/10.1111/j.2517-6161.1979.tb01072.x.
Harrison, R., Nikolov, K., Quinn, M., Ramsay, G., Scott, A. and Thomas, R. (2005). *The Bank of England Quarterly Model*. Research paper, 244pp, Bank of England, London.
Hendry, D. F. (1970). *The Estimation of Economic Models with Autoregressive Errors* (unpublished Ph.D. thesis), London: London University.
Hendry, D. F. (1979a). 'The behaviour of inconsistent instrumental variables estimators in dynamic systems with autocorrelated errors', *Journal of Econometrics*, Vol. 9, pp. 295–314. https://doi.org/10.1016/0304-4076(79)90076-9.
Hendry, D. F. (1979b). 'Predictive failure and econometric modelling in macro-economics: the transactions demand for money', in Ormerod P. (ed), *Economic Modelling*, Heinemann, London, pp. 217–242.
Hendry, D. F. (1985). 'Monetary economic myth and econometric reality', *Oxford Review of Economic Policy*, Vol. 1, pp. 72–84. https://doi.org/10.1093/oxrep/1.1.72.
Hendry, D. F. (1987). 'Econometric methodology: a personal perspective', in Bewley T. F. (ed), *Advances in Econometrics*, Cambridge University Press, Cambridge, pp. 29–48 https://www.cambridge.org/core/services/aop-cambridge-core/content/view/1253AF410FBF1C96AAFE54CD73BBEC62/9781139052078c10_p29-48_CBO.pdf/econometric-methodology-a-personal-perspective.pdf.
Hendry, D. F. (1995). *Dynamic Econometrics*, Oxford University Press, Oxford. https://doi.org/10.1093/0198283164.001.0001.
Hendry, D. F. (1997). 'On congruent econometric relations: a comment', *Carnegie–Rochester Conference Series on Public Policy*, Vol. 47, pp. 163–190. https://doi.org/10.1016/S0167-2231(98)00008-6.
Hendry, D. F. (1999). 'An econometric analysis of US food expenditure, 1931–1989', in Magnus J. R. and Morgan M. S. (eds), *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, John Wiley and Sons, Chichester, pp. 341–361.
Hendry, D. F. (2003). 'J. Denis Sargan and the origins of LSE econometric methodology', *Econometric Theory*, Vol. 19, pp. 457–480. https://doi.org/10.1017/S0266466603193061.
Hendry, D. F. (2004). 'The Nobel memorial prize for Clive W. J. Granger', *Scandinavian Journal of Economics*, Vol. 106, pp. 187–213. https://doi.org/10.1111/j.0347-0520.2004.00361.x.
Hendry, D. F. (2006). 'Robustifying forecasts from equilibrium-correction models', *Journal of Econometrics*, Vol. 135, pp. 399–426. https://doi.org/10.1016/j.jeconom.2005.07.029.
Hendry, D. F. (2009). 'The methodology of empirical econometric modeling: applied econometrics through the looking-glass', in Mills T. C. and Patterson K. D. (eds), *Palgrave Handbook of Econometrics*, Palgrave MacMillan, Basingstoke, pp. 3–67 https://link.springer.com/chapter/10.1057/9780230244405.

Hendry, D. F. (2018). 'Deciding between alternative approaches in macroeconomics', *International Journal of Forecasting*, Vol. 34, pp. 119–135, with 'Response to the Discussants', 142–146. https://doi.org/10.1016/j.ijforecast.2017.09.003.

Hendry, D. F. (2020). 'A short history of macro-econometric modelling', *Journal of Banking, Finance and Sustainable Development*, Vol. 1, pp. 1–32 http://www.collegepress.org.uk/jbfsd/.

Hendry, D. F. and Doornik, J. A. (2014). *Empirical Model Discovery and Theory Evaluation*, MIT Press, Cambridge, MA. https://doi.org/10.7551/mitpress/9780262028356.001.0001.

Hendry, D. F. and Ericsson, N. R. (1991). 'An econometric analysis of UK money demand in 'Monetary Trends in the United States and the United Kingdom' by Milton Friedman and Anna J. Schwartz', *American Economic Review*, Vol. 81, pp. 8–38 https://www.jstor.org/stable/2006786.

Hendry, D. F. and Johansen, S. (2015). 'Model discovery and Trygve Haavelmo's legacy', *Econometric Theory*, Vol. 31, pp. 93–114. https://doi.org/10.1017/S0266466614000218.

Hendry, D. F., Johansen, S. and Santos, C. (2008a). 'Automatic selection of indicators in a fully saturated regression', *Computational Statistics*, Vol. 23, 317–335, (Erratum, pp. 337–339). https://doi.org/10.1007/s00180-007-0054-z.

Hendry, D. F. and Krolzig, H.-M. (1999). 'Improving on 'Data mining reconsidered' by K. D. Hoover and S. J. Perez', *Econometrics Journal*, Vol. 2, pp. 202–219. https://doi.org/10.1111/1368-423X.00027.

Hendry, D. F. and Krolzig, H.-M. (2001). *Automatic Econometric Model Selection*, Timberlake Consultants Press, London.

Hendry, D. F. and Krolzig, H.-M. (2004). 'We ran one regression', *Oxford Bulletin of Economics and Statistics*, Vol. 66, pp. 799–810. https://doi.org/10.1111/j.1468-0084.2004.102_1.x.

Hendry, D. F. and Krolzig, H.-M. (2005). 'The properties of automatic Gets modelling', *Economic Journal*, Vol. 115, pp. C32–C61. https://doi.org/10.1111/j.0013-0133.2005.00979.x.

Hendry, D. F., Marcellino, M. and Mizon, G. E. (2008b). 'Guest editors' introduction to special issue on encompassing', *Oxford Bulletin of Economics and Statistics*, Vol. 70, pp. 715–719. https://doi.org/10.1111/j.1468-0084.2008.00527.x.

Hendry, D. F. and Mizon, G. E. (1978). 'Serial correlation as a convenient simplification, not a nuisance: a comment on a study of the demand for money by the Bank of England', *Economic Journal*, Vol. 88, pp. 549–563. https://doi.org/10.2307/2232053.

Hendry, D. F. and Mizon, G. E. (2014). 'Unpredictability in economic analysis, econometric modeling and forecasting', *Journal of Econometrics*, Vol. 182, pp. 186–195. https://doi.org/10.1016/j.jeconom.2014.04.017.

Hendry, D. F. and Morgan, M. S. (eds). (1995). *The Foundations of Econometric Analysis*, Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139170116.

Hendry, D. F. and Muellbauer, J. N. J. (2018). 'The future of macroeconomics: macro theory and models at the Bank of England', *Oxford Review of Economic Policy*, Vol. 34, pp. 287–328 https://academic.oup.com/oxrep/article/34/1-2/287/4781814.

Hendry, D. F. and Richard, J.-F. (1982). 'On the formulation of empirical models in dynamic econometrics', *Journal of Econometrics*, Vol. 20, pp. 3–33. https://doi.org/10.1016/0304-4076(82)90101-4.

Hendry, D. F. and Richard, J.-F. (1983). 'The econometric analysis of economic time series (with discussion)', *International Statistical Review*, Vol. 51, pp. 111–163 https://www.jstor.org/stable/1402738.

Hendry, D. F. and Santos, C. (2010). 'An automatic test of super exogeneity', in Watson M. W., Bollerslev T., and Russell J. (eds), *Volatility and Time Series Econometrics*, Oxford University Press, Oxford, pp. 164–193. https://doi.org/10.1093/acprof:oso/9780199549498.003.0009.

Hendry, D. F., Spanos, A. and Ericsson, N. R. (1989). 'The contributions to econometrics in Trygve Haavelmo's the probability approach in econometrics', *Sosialøkonomen*, Vol. 11, pp. 12–17.

Hendry, D. F. and Srba, F. (1980). 'AUTOREG: a computer program library for dynamic econometric models with autoregressive errors', *Journal of Econometrics*, Vol. 12, pp. 85–102. https://doi.org/10.1016/0304-4076(80)90054-8.

Hoerl, A. E. and Kennard, R. W. (1970). 'Ridge regression: biased estimation for nonorthogonal problems', *Technometrics*, Vol. 12, pp. 55–67 https://www.jstor.org/stable/1271436.

Hoover, K. D. and Perez, S. J. (1999). 'Data mining reconsidered: encompassing and the general-to-specific approach to specification search', *Econometrics Journal*, Vol. 2, pp. 167–191. https://doi.org/10.1111/1368-423X.00025.
Hoover, K. D. and Perez, S. J. (2000). 'Three attitudes towards data mining', *Journal of Economic Methodology*, Vol. 7, pp. 195–210. https://doi.org/10.1080/13501780050045083.
Huber, P. J. (1964). 'Robust estimation of a location parameter', *Annals of Mathematical Statistics*, Vol. 34, pp. 598–611. https://doi.org/10.1214/aoms/1177703732.
James, W. and Stein, C. (1961). 'Estimation with quadratic loss', in Neyman J. (ed), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, pp. 361–379.
Johansen, S. and Nielsen, B. (2009). 'An analysis of the indicator saturation estimator as a robust regression estimator', in Castle J. L. and Shephard N. (eds), *The Methodology and Practice of Econometrics*, Oxford University Press, Oxford, pp. 1–36. https://doi.org/10.1093/acprof:oso/9780199237197.003.0001.
Johansen, S. and Nielsen, B. (2013). 'Outlier detection in regression using an iterated one-step approximation to the Huber-skip estimator', *Econometrics*, Vol. 1, pp. 53–70. https://doi.org/10.3390/econometrics1010053.
Johansen, S. and Nielsen, B. (2016). 'Asymptotic theory of outlier detection algorithms for linear time series regression models', *Scandinavian Journal of Statistics*, Vol. 43, pp. 321–348. https://doi.org/10.1111/sjos.12174.
Judge, G. G. and Bock, M. E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, North Holland Publishing Company, Amsterdam.
Juselius, K. (1993). 'VAR modelling and Haavelmo's probability approach to econometrics', *Empirical Economics*, Vol. 18, pp. 595–622. https://doi.org/10.1007/BF01205413.
Klein, J. L. (1997). *Statistical Visions in Time*, Cambridge University Press, Cambridge.
Klein, L. R. (1987). *A history of computation in econometrics*, University of Pennsylvania, Mimeo.
Koopmans, T. C. (1937). *Linear Regression Analysis of Economic Time Series*, Netherlands Economic Institute, Haarlem.
Koopmans, T. C. (1947). 'Measurement without theory', *Review of Economics and Statistics*, Vol. 29, pp. 161–179 https://www.jstor.org/stable/1928627.
Land, A. H. and Doig, A. G. (1960). 'An automatic method for solving discrete programming problems', *Econometrica*, Vol. 28, pp. 497–520. https://doi.org/10.1016/j.disopt.2016.01.005.
Leamer, E. E. (1978). *Specification Searches. Ad-Hoc Inference with Non-Experimental Data*, John Wiley, New York.
Leeb, H. and Pötscher, B. M. (2003). 'The finite-sample distribution of post-model-selection estimators, and uniform versus non-uniform approximations', *Econometric Theory*, Vol. 19, pp. 100–142 http://www.jstor.org/stable/3533583.
Leeb, H. and Pötscher, B. M. (2005). 'Model selection and inference: facts and fiction', *Econometric Theory*, Vol. 21, pp. 21–59. https://doi.org/10.1017/S0266466605050036.
Lovell, M. C. (1983). 'Data mining', *Review of Economics and Statistics*, Vol. 65, pp. 1–12 https://www.jstor.org/stable/1924403.
Martinez, A. B., Castle, J. L. and Hendry, D. F. (2022). 'Smooth robust multi-horizon forecasts', *Advances in Econometrics*, Vol. 43A, pp. 143–165. https://doi.org/10.1108/S0731-905320210000043A008.
Miller, P. J. (1978). 'Forecasting with econometric methods: a comment', *Journal of Business*, Vol. 51, pp. 579–586 https://www.jstor.org/stable/2352648.
Mizon, G. E. (1977). 'Model selection procedures', in Artis M. J. and Nobay A. R. (eds), *Studies in Modern Economic Analysis*, Basil Blackwell, Oxford, pp. 97–120.
Mizon, G. E. and Richard, J.-F. (1986). 'The encompassing principle and its application to non-nested hypothesis tests', *Econometrica*, Vol. 54, pp. 657–678. https://doi.org/10.2307/1911313.
Moene, K. O. and Rødseth, A. (1991). 'Nobel laureate: Trygve Haavelmo', *Journal of Economic Perspectives*, Vol. 5, pp. 175–192 https://www.jstor.org/stable/1942803.
Morgan, M. S. (1990). *The History of Econometric Ideas*, Cambridge University Press, Cambridge.
Morrison, D. R., Jacobson, S. H., Sauppe, J. J. and Sewell, E. C. (2016). 'Branch-and-bound algorithms: a survey of recent advances in searching, branching, and pruning', *Discrete Optimization*, Vol. 19, pp. 79–102. https://doi.org/10.1016/j.disopt.2016.01.005.

Murtaugh, P. A. (1998). 'Methods of variable selection in regression modeling', *Communications in Statistics – Simulation and Computation*, Vol. 27, pp. 711–734. https://doi.org/10.1080/03610919808813505.

Nelson, C. R. (1972). 'The prediction performance of the FRB-MIT-PENN model of the US economy', *American Economic Review*, Vol. 62, pp. 902–917 https://www.jstor.org/stable/1815208.

Pagan, A. R. (1987). 'Three econometric methodologies: a critical appraisal', *Journal of Economic Surveys*, Vol. 1, pp. 3–24. https://doi.org/10.1111/j.1467-6419.1987.tb00022.x.

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*, Vol. 65, pp. 821–838. https://doi.org/10.1046/j.0305-9049.2003.00096.x.

Phillips, P. C. B. (1996). 'Econometric model determination', *Econometrica*, Vol. 64, pp. 763–812 https://www.jstor.org/stable/2171845.

Phillips, P. C. B. (2005). 'Automated discovery in econometrics', *Econometric Theory*, Vol. 21, pp. 3–20. https://doi.org/10.1017/S0266466605050024.

Phillips, P. C. B. and Ploberger, W. (1994). 'Posterior odds testing for a unit root with data-based model selection', *Econometric Theory*, Vol. 10, pp. 774–808. https://doi.org/10.1017/S026646660000877X.

Pretis, F., Reade, J. and Sucarrat, G. (2018). 'Automated general-to-specific (gets) regression modeling and indicator saturation methods for the detection of outliers and structural breaks', *Journal of Statistical Software*, Vol. 86 https://www.jstatsoft.org/article/view/v086i03.

Pretis, F., Schneider, L., Smerdon, J. E. and Hendry, D. F. (2016). 'Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation', *Journal of Economic Surveys*, Vol. 30, pp. 403–429. https://doi.org/10.1111/joes.12148.

Qin, D. (1993). *The Formation of Econometrics: A Historical Perspective*, Clarendon Press, Oxford.

Qin, D. (2013). *A History of Econometrics: The Reformation from the 1970s*, Clarendon Press, Oxford.

Renfro, C. G. (2009). *The Practice of Econometric Theory: An Examination of the Characteristics of Econometric Computing*, Springer, London.

Richard, J.-F. (1980). 'Models with several regimes and changes in exogeneity', *Review of Economic Studies*, Vol. 47, pp. 1–20 https://www.jstor.org/stable/2297101.

Salkever, D. S. (1976). 'The use of dummy variables to compute predictions, prediction errors and confidence intervals', *Journal of Econometrics*, Vol. 4, pp. 393–397. https://doi.org/10.1016/0304-4076(76)90027-0.

Sargan, J. D. (1957). 'The danger of over-simplification', *Oxford Bulletin of Economics and Statistics*, Vol. 19, pp. 171–178 https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0084.1957.mp19002009x.

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* Colston PapersVol. 16, Butterworth Co, London, pp. 25–63.

Sargan, J. D. (1973). 'Model building and data mining', in *Presented to the Association of University Teachers of Economics April 1973 Meeting*, Manchester. *Econometric Reviews*, (2001), **20**, 159–170. https://doi.org/10.1081/ETC-100103820.

Sargan, J. D. (1980). 'Some tests of dynamic specification for a single equation', *Econometrica*, Vol. 48, pp. 879–897 https://www.jstor.org/stable/1912938.

Sargan, J. D. (1981). *The Choice between Sets of Regressors*, Mimeo, Economics Department, London School of Economics. Published in *Econometric Reviews*, (2001), **20**, 171–186. https://doi.org/10.1081/ETC-100103821.

Schwarz, G. (1978). 'Estimating the dimension of a model', *Annals of Statistics*, Vol. 6, pp. 461–464 https://www.jstor.org/stable/2958889.

Sims, C. A. (ed). (1977). *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis.

Spanos, A. (1989). 'On re-reading Haavelmo: A retrospective view of econometric modeling', *Econometric Theory*, Vol. 5, pp. 405–429 https://www.jstor.org/stable/3532376.

Summers, L. H. (1991). 'The scientific illusion in empirical macroeconomics', *Scandinavian Journal of Economics*, Vol. 93, pp. 129–148 https://www.jstor.org/stable/3440321.

Tibshirani, R. (1996). 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society B*, Vol. 58, pp. 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Walker, A., Pretis, F., Powell-Smith, A. and Goldacre, B. (2019). 'Variation in responsiveness to warranted behaviour change among NHS clinicians: a novel implementation of change-detection methods in longitudinal prescribing data', *British Medical Journal*, Vol. 367, l5205 https://www.bmj.com/content/367/bmj.l5205.

White, H. (1990). 'A consistent model selection procedure based on m-testing', in Granger C. W. J. (ed), *Modelling Economic Series: Readings in Econometric Methodology*, Clarendon Press, Oxford, pp. 369–383.

Zellner, A. (1992). 'Statistics, science and public policy', *Journal of the American Statistical Association*, Vol. 87, pp. 1–6. https://doi.org/10.1080/01621459.1992.10475168.