International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Discussion

Response to the Discussants of 'Deciding between alternative approaches in macroeconomics'

David F. Hendry

Economics Department and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK

1. Introduction

I thank the discussants for their thoughtful analyses and numerous constructive suggestions. Their ideas and clarifications will help advance empirical modelling practice. Professor Proietti elegantly summarises my paper in his introduction, and Professor Perez-Quiros helpfully clarifies the alternative approaches in my paper by 'walking the reader' through the various stages of macroeconomic forecasting, noting the prevalence of both theory-driven small dynamic factor models, (see e.g., Stock and Watson, 2002), and data-driven specifications in large scale models as in Forni, Hallin, Lippi and Reichlin (2000).

Building on an existing body of research reviewed in Hendry and Doornik (2014), the aim of my paper was to draw together at a general level how to decide between approaches, then interested readers could consult more extensive explanations for whatever aspects mattered most to them. There are a number of publications concerning the details, with Monte Carlo simulation studies and theory analyses, albeit that many more remain to be undertaken, as well as applications to macroeconomics and a diverse range of fields including dendrochronology, volcanology and climatology.

To respond to the issues the discussants raised, Section 2 considers the different roles of strategy and tactics in model selection; Section 3 illustrates the combined theory-evidence approach when retaining a theory that transpires not to be a good guide to the finally selected model; and Section 4 briefly turns to the role of forecasting in model selection.

2. Strategy and tactics in model selection

Strategy and tactics both matter in model selection, howsoever that task is undertaken. The former was the focus of my paper, and even the choice of model selection algorithms was treated as part of tactics, so was not discussed at any length. The strategy is one of seeking to nest the local data generation process in a general starting model that retains available theory insights, and includes in an orthogonal form alternatives likely to be relevant in a wide-sense non-stationary world, then selects over the latter by multi-path block searches to discover what additional features matter, thereby evaluating the theory specification. Better tactics will undoubtedly evolve over time, and have done so already: compare model selection approaches before and after the introduction of indicator saturation for finding breaks.

Indicator saturation allows the design of indicators to match 'likely' break forms, providing a useful flexibility, as in impulse-indicator saturation (IIS) for outliers, step-indicator saturation (SIS) for location shifts, a $\boldsymbol{\nu}$ shape for the impacts on temperature of volcanic eruptions as in Pretis, Schneider, Smerdon and Hendry (2016), etc. Based on the results in Hendry and Krolzig (2005), collinearity is less problematic for SIS than it looked at first sight, encouraging us to investigate the properties of trend-indicator saturation. Moreover, the uncertainty around SIS-determined break dates can be estimated, as in Hendry and Pretis (2016), as can uncertainty bands for the trajectory of the mean: see Pretis (2015).

Although we have not yet tackled selecting shifts in second moments, that remains on the research agenda, and model selection of non-linearities goes some way towards that aim. Basis functions for approximating non-linearity are again an aspect of tactics, where future improvements are highly likely, and Professor Proietti's proposal of using B-splines for both shifts and non-linearity points a way ahead.

What Ericsson (2017) calls multiple-indicator saturation (MIS) – where every regressor is interacted with a saturating set of step indicators, so for k regressors and T observations, there are $k \times T$ candidate variables – now

E-mail address: david.hendry@nuffield.ox.ac.uk.

https://doi.org/10.1016/j.ijforecast.2017.09.004

enables the detection of changes in the parameters of regressors. Kitov and Tabor (unpublished) demonstrated the success of this approach, despite the high dimensionality of the set of candidate variables. To understand intuitively how SIS or MIS are able to detect location shifts or parameter changes, consider knowing where a single (moderately large magnitude) shift occurred, and splitting your data into sub-samples before and after that point. Then you would rightly be surprised if fitting your correctly specified model of the data generation process (DGP) separately to the different sub-samples did not deliver appropriately different estimates of their DGP parameters. Choosing that split by SIS or MIS will add variability from particular error draws around the break point, which may offset the apparent shift date temporarily by making it appear slightly before or after the actual occurrence, but the correct indicator, or one close to it, will accomplish almost the same task as knowing the timing, so will be the likely selection.

A surprising finding from our research is that model selection, appropriately conducted in a setting where the general unrestricted model (GUM) is sufficiently well specified to nest the DGP, is almost as good as selecting from that DGP at the same significance level. Of course, nesting the DGP, or the local DGP, is unlikely and the entailed LDGP may be a poor approximation to the actual DGP. Moreover, GUMs may be underspecified in many possible ways. Nevertheless, Castle, Doornik, and Hendry (2011) show in simulations that the costs from estimating the DGP can exceed those from selecting from an underspecified GUM, as measured by the RMSEs of coefficient estimates relative to the DGP parameters. Castle and Hendry (2014) also investigate the consequences for automatic model selection facing shifts when using indicator saturation, reinforcing the advantages of seeking to include all likely substantively-relevant variables.

3. Is it problematic to retain a poor theory?

An underspecified set of variables is most likely to arise when only a theory model is specified and estimated, eschewing the advice in the paper to retain that theory while searching over a much larger set of candidates. To illustrate the combined approach in that setting, I return to the Davidson, Hendry, Srba and Yeo (1978) (DHSY) study on quarterly UK data for constant price consumers' expenditure, C_t and real personal disposable income, I_t, over 1958(2)–1976(2). Starting from the simplest version of the permanent income hypothesis (PIH) current at the time of their study, namely $C_t = \beta_0 + \beta_1 I_t + \beta_2 C_{t-1} + e_t$, with added seasonal dummies, S_i, (equation (12) in DHSY), had *Autometrics* been available, could DHSY have found their model in an afternoon rather than several years, despite the initial theory being a poor guide to their finally selected model?

Re-estimating DHSY's PIH equation, but in a log-linear specification over the full sample yielded:

$$
\begin{aligned}
c_t &= \underset{(0.07)}{0.59}\, c_{t-1} + \underset{(0.054)}{0.31}\, i_t + \underset{(0.14)}{0.87} \\
&\quad - \underset{(0.007)}{0.12}\, S_{1,t} - \underset{(0.005)}{0.01}\, S_{2,t} - \underset{(0.003)}{0.03}\, S_{3,t} \qquad (1)\\
\widehat{\sigma} &= 1.0\% \;\; \mathsf{R}^2 = 0.995 \;\; \mathsf{F}_{ar}(5, 66) = 9.68^{**} \\
&\quad \mathsf{F}_{arch}(4, 69) = 2.79^{*} \\
\chi^2_{nd}(2) &= 5.14 \;\; \mathsf{F}_{het}(7, 69) = 3.97^{**} \;\; \mathsf{F}_{reset}(2, 69) = 0.57
\end{aligned}
$$

where estimated coefficient standard errors are shown in parentheses below estimated coefficients, $\widehat{\sigma}$ is the residual standard deviation, R^2 is the coefficient of multiple correlation, F_{ar} is a test for residual autocorrelation (see Godfrey, 1978), F_{arch} tests for autoregressive conditional heteroskedasticity (see Engle, 1982), F_{het} is a test for residual heteroskedasticity (see White, 1980), $\chi^2_{nd}(2)$ is a test for non-Normality (see Doornik and Hansen, 2008), and F_{reset} is the RESET test (see Ramsey, 1969). Thus, tests for residual autocorrelation, ARCH and heteroskedasticity all reject: see Fig. 2(a) for a graph of the resulting residuals.

The main result is the failure of that simple theory to characterize the evidence, so evaluation is accomplished, but there is no useful guidance on how to proceed towards a better formulation. Recipes for patching autocorrelation, heteroskedasticity, etc., do not take into account that rejection on such tests can arise from many failures of the assumptions needed for congruence, not necessarily the alternative hypothesis against which the test was designed to have power: see Mizon (1995) and Spanos (2017).

The DHSY data on c_t, i_t shown in Fig. 1(a) suggests that seasonal lags might matter (here 5 lags), and also including inflation, $\Delta_4 p_t$ and its lag, and their tax dummy, $\Delta_4 D_t$, creates the GUM. The continuous variables were orthogonalized against c_{t-1} and i_t, denoted $\widetilde{\ }$. Next, this GUM was estimated to check that the coefficient estimates of the retained variables in (1) were unaffected, as (2) confirms, where bold denotes coefficients of added variables that are individually significant at 1%.

$$
\begin{aligned}
c_t &= \underset{(0.04)}{0.59}\, c_{t-1} + \underset{(0.03)}{0.31}\, i_t + \underset{(0.09)}{0.88} \\
&\quad - \underset{(0.004)}{0.12}\, S_{1,t} - \underset{(0.003)}{0.01}\, S_{2,t} - \underset{(0.002)}{0.03}\, S_{3,t} \\
&\quad + \underset{(\mathbf{0.002})}{\mathbf{0.007}}\, \Delta_4 D_t - \underset{(0.09)}{0.04}\, \widetilde{c}_{t-2} + \underset{(0.09)}{0.05}\, \widetilde{c}_{t-3} \\
&\quad + \underset{(\mathbf{0.10})}{\mathbf{0.73}}\, \widetilde{c}_{t-4} - \underset{(0.12)}{0.02}\, \widetilde{c}_{t-5} + \underset{(\mathbf{0.06})}{\mathbf{0.16}}\, \widetilde{i}_{t-1} \\
&\quad - \underset{(0.05)}{0.031}\, \widetilde{i}_{t-2} + \underset{(0.05)}{0.04}\widetilde{i}_{t-3} - \underset{(0.05)}{0.10}\, \widetilde{i}_{t-4} \\
&\quad - \underset{(\mathbf{0.06})}{\mathbf{0.20}}\, \widetilde{i}_{t-5} - \underset{(\mathbf{0.08})}{\mathbf{0.33}}\widetilde{\Delta_4 p_t} + \underset{(0.09)}{0.19}\, \widetilde{\Delta_4 p_{t-1}} \\
&\widehat{\sigma} = 0.58\% \;\; \mathsf{R}^2 = 0.998 \\
&\quad \mathsf{F}_{ar}(5, 46) = 1.56 \;\; \mathsf{F}_{arch}(4, 61) = 2.98^{*} \qquad (2)\\
&\chi^2_{nd}(2) = 0.08 \;\; \mathsf{F}_{het}(31, 37) = 1.47 \\
&\quad \mathsf{F}_{reset}(2, 49) = 4.92^{*} \quad \mathsf{F}_{add}(12, 51) = 13.1^{**}.
\end{aligned}
$$

Please cite this article in press as: Hendry, D.F., Response to the Discussants of 'Deciding between alternative approaches in macroeconomics'. International Journal of Forecasting (2017), https://doi.org/10.1016/j.ijforecast.2017.09.004.