Forecasting economic and financial time-series
with non-linear models

Michael P. Clements
Department of Economics,
University of Warwick

Philip Hans Franses
Econometric Institute,
Erasmus University Rotterdam.

Norman R. Swanson*
Department of Economics,
Rutgers University.

October 1, 2003

Abstract

In this paper we discuss the current state-of-the-art in estimating, evaluating, and selecting among non-linear forecasting models for economic and financial time series. We review theoretical and empirical issues, including predictive density, interval and point evaluation and model selection, loss functions, data-mining, and aggregation. In addition, we argue that although the evidence in favor of constructing forecasts using non-linear models is rather sparse, there is reason to be optimistic. However, much remains to be done. Finally, we outline a variety of topics for future research, and discuss a number of areas which have received considerable attention in the recent literature, but where many questions remain.

1   Introduction

Whilst non-linear models are often used for a variety of purposes, one of their prime uses is for forecasting, and it is in terms of their forecasting performance that they are most often judged. However, a casual review of the literature suggests that often the forecasting performance of such models is not particularly good. Some studies find in favour, but equally there are studies in which their added complexity relative to rival linear models does not result in the expected gains in terms of forecast accuracy. Just over a decade ago, in their review of non-linear time series models, De Gooijer and Kumar (1992) concluded that there was no clear evidence in favour of non-linear over linear models in terms of forecast performance, and we suspect that the situation has not changed very much since then. It seems that we have not come very far in the area of non-linear forecast model construction.

We argue that the relatively poor forecasting performance of non-linear models calls for substantive further research in this area, given that one might feel uncomfortable asserting that non-linearities are unimportant in describing economic and financial phenomena. The problem may simply be that our non-linear models are not mimicing reality any better than simpler linear approximations, and in the next section we discuss this and related reasons why a good forecast performance 'across the board' may constitute something of a 'holy grail' for non-linear models.

We discuss the current state-of-the-art in non-linear modelling and forecasting, with particular emphasis placed on outlining a number of open issues. The topics we focus on include joint and conditional predictive density evaluation, loss functions, estimation and specification, and data-mining, amongst others. As such, this paper complements the rest of the papers in this special issue of the International Journal of Forecasting.

The rest of the paper is organized as follows. In Section 2 we discuss why one might want to consider non-linear models, and a number of reasons why their forecasting ability relative to linear models may not be as good as expected. In Section 3 we discuss recent theoretical and methodological issues to do with forecasting with non-linear models, many of which go beyond the traditional preoccupation with point forecasts to consider the whole predictive density. Section 4 highlights a number of empirical issues, and how these are dealt with in the papers collected in this issue. Concluding remarks are gathered in Section 5.

2   Why consider non-linear models?

Many of us believe that linear models ought to be a relatively poor way of capturing certain types of economic behaviour, or economic performance, at certain times. The obvious example would be a linear (e.g., Box-Jenkins ARMA) model of output growth in a Western economy subject to the business cycle, where the properties of output growth in recessions are in some ways quite different from expansions (e.g., Hamilton, 1989, Sichel, 1994, but references are innumerable). Output growth non-linearities can be characterised by the presence of two or more regimes (e.g. recessions and expansions), as can financial variables (periods of high and low volatility). Other types of non-linearity might include the possibility that the effects of shocks accumulate until a process "explodes" (self-exciting or catastrophic behaviour), as well as the notion that some variables are relevant for forecasting only once in a while (e.g., only when oil prices increase by a large amount do they have a significant effect on output growth, and therefore become useful for forecasting output growth).

In macroeconomics and finance theory a host of non-linear models are already in vogue. For example, almost all real business cycle models are highly non-linear. As well as which, bond pricing models, diffusion processes describing yield curves, and almost all other continuous time finance models are non-linear. The predominance of non-linear models in economics and finance is not inconsistent with the use of linear models by the applied practitioner, as such models can be viewed as reasonable approximations to the non-linear phenomenon of interest. Thus, from the perspective of forecasting, there is ample reason to continue to look at non-linear models. As our non-linear model estimation, selection and testing approaches become more sophisticated, one might expect to see their forecast performance improve commensurately. It is thus not surprising that non-linear models, ranging from regime-switching models, to neural networks and genetic algorithms, are receiving a great deal of attention in the literature.

On a more cautionary note, we review a number of factors which might count against the aforementioned improvement in the relative performance of non-linear models. Shortly after De Gooijer and Kumar, Granger and Teräsvirta (1993, ch. 9) (see also Teräsvirta and Anderson, 1992) in their review of smooth transition autoregressive (STAR) models of US industrial production, argue that the superior in-sample performance of such models will only be matched out-of-sample if that period contains 'non-linear features'. Similarly, Tong (1995,

p. 409 – 410) argues strongly that for non-linear models 'how well we can predict depends on where we are' and that there are 'windows of opportunity for substantial reduction in prediction errors'. This suggests that an important aspect of an evaluation of the forecasts from non-linear models relative to the linear AR models is to make the comparison in a way which highlights the favourable performance of the former for certain states, especially if it is forecasts in those states which are most valuable to the user of the forecasts. Clements and Smith (1999) compare the forecasting performance of empirical self-exciting threshold autoregressive (SETAR) models and AR models using simulation techniques which ensure that past non-linearities are present in the forecast period. See Tiao and Tsay (1994) for a four-regime TAR model applied to US GNP, and Boero and Marrocu (2004) for an application of regime-specific evaluation to exchange rate forecasts.

In addition, in the context of exchange rate prediction, Diebold and Nason (1990) give a number of reasons why non-linear models may fail to outperform linear models. One is that apparent non-linearities detected by tests for linearity are due to outliers or structural breaks, which cannot be readily exploited to improve out-of-sample performance, and may only be detected by careful analysis along the lines of Koop and Potter (2000), for example. They also suggest that conditional-mean non-linearities may be a feature of the data generating process (DGP), but may not be large enough to yield much of an improvement to forecasting, as well as the explanation that they are present and important, but that the wrong types of non-linear models have been used to try and capture them.

There is a view that, because some aspects of the economy or financial markets do indeed display non-linear behaviour, then neglecting these features in constructing forecasts would leave the end-user uneasy, feeling that the forecasts are in some sense "second best" . This follows from the belief that a good model for the in-sample data should also be a good out-of-sample forecasting model. As an example, an AR(2) model for US GNP growth may yield lower average squared errors than an artificial neural network with one hidden unit, but, knowing that the AR(2) model is incapable of capturing the distinct dynamics of expansions and contractions, is unlikely to engender confidence that one has a model that captures the true dynamics of the economy. But this is the crux of the matter – a good in-sample fit (here, capturing the distinct dynamics of the business cycle phases) does not necessarily translate in to a good out-of-sample performance relative to a model such as an AR. The sub-section below illustrates, and see also

Clements and Hendry (1999) for a more general discussion.

We take the view that, if one believes the underlying phenomenon is non-linear, it is worth considering a non-linear model, but warn against the expectation that such models will always do well – there are too many unknowns and the economic system is too complex to support the belief that simply generalising a linear model in one (simple!) direction, such as adding another regime, will necessarily improve matters. That said, it is natural to be unhappy with models which are obviously deficient in some respect, and to seek alternatives. The subsequent sections of this paper review some of the recent developments in model selection and empirical strategies, as well as the remaining papers in this issue, which take up the challenge of forecasting with non-linear models.

We end this section with two short illustrations of some of the difficulties that can arise. In the first, there is distinct regime-switching behaviour, but this does not contribute to an improved forecast performance. In the second, we discuss the relationship between output growth and the oil price.

2.1 Markov-switching models of US output growth

Clements and Krolzig (1998) present some theoretical explanations for why Markov-Switching (MS) models may not forecast much better than AR models, and apply their analysis to post War US output growth. The focus is on a two-regime Markov-Switching (MS) model:

$$\Delta y_t - \mu(s_t) = \alpha \left(\Delta y_{t-1} - \mu(s_{t-1}) \right) + u_t. \tag{1}$$

where $u_t \sim IN[0, \sigma_u^2]$. The conditional mean $\mu(s_t)$ switches between two states:

$$\mu(s_t) = \begin{cases} \mu_1 > 0 & \text{if } s_t = 1 \text{ ("expansion" or "boom")}, \\ \mu_2 < 0 & \text{if } s_t = 2 \text{ ("contraction" or "recession")}, \end{cases} \tag{2}$$

The description of a MS-AR model is completed by the specification of a model for the stochastic and unobservable regimes on which the parameters of the conditional process depend. Once a law has been specified for the states s_t, the evolution of regimes can be inferred from the data. The regime-generating process is assumed to be an ergodic Markov chain with a finite number of states $s_t = 1, 2$ (for a two-regime model), defined by the transition probabilities:

$$p_{ij} = \Pr(s_{t+1} = j \mid s_t = i), \qquad \sum_{j=1}^{2} p_{ij} = 1 \quad \forall i, j \in \{1, 2\}. \tag{3}$$

The model can be rewritten as the sum of two independent processes:

$$\Delta y_t - \mu_y = \mu_t + z_t,$$

where μ_y is the unconditional mean of Δy_t, such that $E[\mu_t] = E[z_t] = 0$. While the process z_t is Gaussian:

$$z_t = \alpha z_{t-1} + \epsilon_t, \quad \epsilon_t \sim IN\left(0, \sigma^2\right),$$

the other component, μ_t, represents the contribution of the Markov chain:

$$\mu_t = (\mu_2 - \mu_1)\zeta_t,$$

where $\zeta_t = 1 - \Pr(s_t = 2)$ if $s_t = 2$ and $-\Pr(s_t = 2)$ otherwise. $\Pr(s_t = 2) = p_{12}/(p_{12} + p_{21})$ is the unconditional probability of regime 2. Invoking the unrestricted VAR(1) representation of a Markov chain:

$$\zeta_t = (p_{11} + p_{22} - 1)\zeta_{t-1} + v_t,$$

then predictions of the hidden Markov chain are given by:

$$\hat{\zeta}_{T+h|T} = (p_{11} + p_{22} - 1)^h \hat{\zeta}_{T|T}$$

where $\hat{\zeta}_{T|T} = E[\zeta_T | Y_T] = \Pr(s_T = 2|Y_T) - \Pr(s_T = 2)$ is the filtered probability $\Pr(s_T = 2|Y_T)$ of being in regime 2 corrected for the unconditional probability. Thus, the conditional mean of Δy_{T+h} is given by $\Delta \hat{y}_{T+h|T} - \mu_y$ which equals:

$$
\begin{aligned}
& \hat{\mu}_{T+h|T} + \hat{z}_{T+hT} \\
&= (\mu_2 - \mu_1)(p_{11} + p_{22} - 1)^h \hat{\zeta}_{T|T} + \alpha^h \left[\Delta y_T - \mu_y - (\mu_2 - \mu_1)\hat{\zeta}_{T|T}\right] \\
&= \alpha^h \left[\Delta y_T - \mu_y\right] + (\mu_2 - \mu_1)\left[(p_{11} + p_{22} - 1)^h - \alpha^h\right]\hat{\zeta}_{T|T}.
\end{aligned}
\tag{4}
$$

The first term in (4) is the optimal prediction rule for a linear model, and the contribution of the Markov regime-switching structure is given by the term multiplied by $\hat{\zeta}_{T|T}$, where $\hat{\zeta}_{T|T}$ contains the information about the most recent regime at the time the forecast is made. Thus, the contribution of the non-linear part of (4) to the overall forecast depends on both the magnitude of the regime shifts, $|\mu_2 - \mu_1|$, and on the persistence of regime shifts $p_{11} + p_{22} - 1$ relative to the persistence of the Gaussian process, given by α.

Clements and Krolzig estimate $p_{11} + p_{22} - 1 = 0.65$, and the largest root of the AR polynomial to be 0.64, so that the second reason explains the success of the linear AR model in forecasting.

Since the predictive power of detected regime shifts is extremely small, $p_{11} + p_{22} - 1 \simeq \alpha$ in (4), the conditional expectation collapses to a linear prediction rule. Heuristically, the relative performance of non-linear regime-switching models would be expected to be better the more persistent the regimes. When the regimes are unpredictable, we can do no better than employing a simple linear model. See also Krolzig (2003), and Dacco and Satchell (1999) for an analysis of the effects of wrongly classifying the regime the process will be in.

A number of authors, such as Sensier, Artis, Osborn and Birchenhall (2004), attempt to predict business cycle regimes, and the transition probabilities in (3) can be made to depend on leading indicator variables, as a way of sharpening the forecasting ability of these models. Franses, Paap and Vroomen (2004) utilise information from extraneous variables to determine the regime in a novel approach to predicting the US unemployment rate.

2.2   Output growth and the oil price

Of obvious interest in the literature is the relationship between oil prices and the macroeconomy. To what extent did the OPEC oil price rises contribute to the recessions in the 70's, and might one expect reductions in prices to stimulate growth, although perhaps to a lesser extent? Hamilton (1983) originally proposed a linear relationship between oil prices and output growth for the US. This was challenged by Mork (1989), who suggested that the relationship was asymmetric, in that output growth responds negatively to oil price increases, but is unaffected by oil price declines. With the advantage of several more years of data, Hooker (1996) showed that the linear relationship proposed by Hamilton (1983) appears not to hold from 1973 onwards (the date of the first oil price hike!). However, he also cast doubt on the simple asymmetry hypothesis suggested by Mork (1989). More recently, Hamilton (1996) proposes relating output growth to the net increase in oil prices over the previous year, and constructs a variable that is the percentage change in the oil price in the current quarter over the previous year's high, when this is positive, and otherwise takes on the value zero. Thus, increases in the price of oil which simply reverse previous (within the preceding year) declines do not depress output growth. Recently, Hamilton (2000) has used a new flexible non-linear approach (Hamilton (2001) and Dahl and Hylleberg (2004)) to characterize the appropriate non-linear transform of the oil price.

Raymond and Rich (1997) have investigated the relationship between oil prices and the macroeconomy by including the net increase in the oil price in an MS model of US output for

the period 1951 – 1995. They are interested in whether the recurrent shifts between expansion and contraction identified by the MS model remain when oil prices have been included as an explanatory variable for the mean of output. Raymond and Rich (1997) conclude that 'while the behavior of oil prices has been a contributing factor to the mean of low growth phases of output, movements in oil prices generally have not been a principal determinant in the historical incidence of these phases ...' (p. 196). Further, Clements and Krolzig (2002) investigate whether oil prices can account for the asymmetry in the business cycle using the tests proposed in Clements and Krolzig (2003).

Clearly, the process of discovery of (an approximation to) the form of the non-linearity in the relationship between output growth and oil price changes has taken place over two decades, is far from simple, and has involved the application of state of the art econometric techniques. Perhaps this warns against expecting too much in the near future using 'canned' routines and models.

3   Theoretical and methodological issues

There are various theoretical issues involved in constructing non-linear models for forecasting. In this section we outline a number of these. An obvious starting point is which non-linear model to use, given the many possibilities that are available, even once we have determined the purpose to which it is to be put (here, forecasting). The different types of models often require different theoretical and empirical tools (see e.g. the recent surveys by Franses and van Dijk (2000) and van Dijk et al. (2002)). For example, closed form solutions exist for the conditional mean forecast for an MS process, but not for a threshold autoregressive process, requiring simulation or numerical methods in the latter case. Certain theoretical properties, such as stability and stationarity, and the persistence of shocks, are not always immediately evident.

The choice of model might be suggested by economic theory, and often by the requirement that the model is capable of generating the key characteristics of the data at hand. Of course, the issue of which characteristics often arises. For example, Pagan (1997a,b) and Harding and Pagan (2001) argue that non-linear models should be evaluated in terms of their ability to reproduce certain features of the classical cycle, rather than their ability to match the stationary

moments of the detrended growth cycle.[1]

An approach to tackling these issues is to use predictive density or distributional testing, as a means of establishing which of a number of candidate forecasting models has distributional features that most closely match the historical record. This could include, for example, finding out which of the models yields the best distributional or interval predictions. For example, in financial risk management interest often focuses on predicting a particular quantile (as the Value at Risk, VaR) but alternatively the entire conditional distribution of a variable may be of interest. Over the last few years, a new strand of literature addressing the issue of predictive density evaluation has arisen (see e.g., Bai (2001), Christoffersen (1998), Christoffersen, Hahn and Inoue (2001), Clements and Smith (2000, 2002), Diebold, Gunther and Tay (1998) (henceforth DGT), Diebold, Hahn and Tay (1999), Giacomini and White (2003) and Hong (2001)). The literature on the evaluation of predictive densities is largely concerned with testing the null of correct dynamic specification of an individual conditional distribution model. At the same time, the point forecast evaluation literature explicitly recognises that all the candidate models may be misspecified (see e.g. Corradi and Swanson (2002) and White (2000)). Corradi and Swanson (2003a) draw on elements from both types of papers in order to provide a test for choosing among competing predictive density models which may be misspecified. Giacomini and White (2003) tackle a similar problem, developing a framework that allows for the evaluation of both nested and nonnested models.

Many of the above papers seek to evaluate predictive densities by testing whether they have the property of correct (dynamic) specification. By making use of the probability integral transform, DGT suggest a simple and effective means by which predictive densities can be evaluated. Using the DGT terminology, if $p_t(y_t|\Omega_{t-1})$ is the "true" conditional distribution of $y_t|\Omega_{t-1}$, then the probability of observing a value of y_t no larger than that actually observed is a uniform random variable on $[0,1]$. Moreover, if we have a sequence of predictive densities, then the resulting sequence of probabilities should be identically independently distributed. Goodness of fit statistics can then be constructed that compare the empirical distribution

---

[1]As an aside, these papers have shown that the durations and amplitudes of expansions and contractions of the classical cycle can be reasonably well reproduced by simple random walk with drift models, where the ratio of the drift to the variance of the disturbance term is the crucial quantity. Non-linear models appear to add little over and above that which can be explained by the random walk with drift.

function of the probabilities to the 45 degree line, possibly taking into account that $p_t(y_t|\Omega_{t-1})$ contains parameters that have been estimated (see also Bai (2001), Diebold, Hahn and Tay (1999) and Hong (2001)).

The approach taken by Corradi and Swanson (2003a) differs from the DGT approach as they do not assume that any of the competing models are correctly specified. Thus, they posit that *all* models should be viewed as approximations to some unknown underlying data generating process. They proceed by "selecting" a conditional distribution model that provides the most accurate out-of-sample approximation of the true conditional distribution, allowing for misspecification under both the null and the alternative hypotheses. This is done using an extension of the conditional Kolmogorov test approach of Andrews (1997) due to Corradi and Swanson (2003b) that allows for the in-sample comparison of multiple misspecified models. More specifically, assume that the objective is to form parametric conditional distributions for a scalar random variable, y_{t+1}, given some vector of variables, $Z^t = (y_t, ..., y_{t-s_1+1}, X_t, ..., X_{t-s_2+1})$, $t = 1, ..., T$, where $s = \max\{s_1, s_2\}$, and X is a vector. Additionally, assume that $i = 1, ..., n$ models are estimated. Now, define the *recursive m*-estimator for the parameter vector associated with model i as:

$$\hat{\theta}_{i,t} = \arg \min_{i \in \Theta_i} \frac{1}{t} \sum_{j=s}^{} q_i(y_j, Z^{j-1}, \theta_i), \quad R \le t \le T - 1, \ i = 1, ..., n \tag{5}$$

and

$$\theta_i^\dagger = \arg \min_{i \in \Theta_i} E(q_i(y_j, Z^{j-1}, \theta_i)), \tag{6}$$

where q_i denotes the objective function for model i. Following standard practice (such as in the real-time forecasting literature), this estimator is first computed using R observations, then $R + 1$ observations, then $R + 2$, and so on until the last estimator is constructed using $T - 1$ observations, resulting in a sequence of $P = T - R$ estimators. In the current discussion, we focus on 1-step ahead prediction, so these estimators are then used to construct sequences of P 1-step ahead forecasts and associated forecast errors.

Now, define the group of conditional distribution models from which we want to make a selection as $F_1(u|Z^t, \theta_1^\dagger), ..., F_n(u|Z^t, \theta_n^\dagger)$, and define the true conditional distribution as $F_0(u|Z^t, \theta_0) = \Pr(y_{t+1} \le u|Z^t)$. Hereafter, assume that $q_i(y_t, Z^{t-1}, \theta_i) = -\ln f_i(y_t|Z^{t-1}, \theta_i)$, where $f_i(\cdot|\cdot, \theta_i)$ is the conditional density associated with F_i, $i = 1, ..., n$, so that $\theta_i^\dagger$ is the probability limit of a quasi maximum likelihood estimator (QMLE). If model i is correctly specified, then $\theta_i^\dagger = \theta_0$.

9

Now, $F_1(\cdot|\cdot, \theta_1^\dagger)$ is taken as the benchmark model, and the objective is to test whether some competitor model can provide a more accurate approximation of $F_0(\cdot|\cdot, \theta_0)$ than the benchmark. Assume that accuracy is measured using a distributional analog of mean square error. More precisely, the squared (approximation) error associated with model i, $i = 1, ..., n$, is measured in terms of the average over U of $E\left(F_i(u|Z^t, \theta_i^\dagger) - F_0(u|Z^t, \theta_0)\right)^2$, where $u \in U$, and U is a possibly unbounded set on the real line. The hypotheses of interest are:

$$H_0 : \max_{k=2,...,n} \int_U E\left(\left(F_1(u|Z^t, \theta_1^\dagger) - F_0(u|Z^t, \theta_0)\right)^2 - \left(F_k(u|Z^t, \theta_k^\dagger) - F_0(u|Z^t, \theta_0)\right)^2\right) \varphi(u)du \leq 0 \tag{7}$$

versus

$$H_A : \max_{k=2,...,n} \int_U E\left(\left(F_1(u|Z^t, \theta_1^\dagger) - F_0(u|Z^t, \theta_0)\right)^2 - \left(F_k(u|Z^t, \theta_k^\dagger) - F_0(u|Z^t, \theta_0)\right)^2\right) \varphi(u)du > 0, \tag{8}$$

where $\varphi(u) \geq 0$ and $\int_U \varphi(u) = 1$, $u \in U \subset \Re$, U possibly unbounded. Note that for a given u, we compare conditional distributions in terms of their (mean square) distance from the true distribution. The statistic is:

$$Z_P = \max_{k=2,...,n} \int_U Z_{P,u}(1,k) \, \varphi(u)du, \tag{9}$$

where

$$Z_{P,u}(1,k) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \left(\left(1\{y_{t+1} \leq u\} - F_1(u|Z^t, \hat{\theta}_{1,t})\right)^2 - \left(1\{y_{t+1} \leq u\} - F_k(u|Z^t, \hat{\theta}_{k,t})\right)^2\right). \tag{10}$$

Here, each model is estimated via QMLE, so that in terms of the above notation, $q_i = -\ln f_i$, where f_i is the conditional density associated with model i, and $\hat{\theta}_{i,t}$ is defined as $\hat{\theta}_{i,t} = \arg\max_{\theta_i \in \Theta_i} \frac{1}{t} \sum_{j=s}^{t} \ln f_i(y_j, Z^{j-1}, \theta_i)$, $R \leq t \leq T-1$, $i = 1, ..., n$. For further details, please refer to Corradi and Swanson (2003a).

Clearly, the above approach can be used to evaluate multiple non-linear forecasting models. Now, assume that focus centers on evaluating the joint dynamics of a non-linear prediction model, say in the form of a real business cycle (RBC) model. Corradi and Swanson (2003c) develop a statistic based on comparison of historical and simulated distributions. As the RBC data are simulated using estimated parameters (as well as previously calibrated parameters), the limiting distribution of their test statistic is a Gaussian process with a covariance kernel

10

that reflects the contribution of parameter estimation error. This limiting distribution is thus not nuisance parameter free, and critical values cannot be tabulated. In order to obtain valid asymptotic critical values, they suggest two block bootstrap procedures, each of which depends on the relative rate of growth of the actual and simulated sample size. In addition, the standard issue of singularity that arises when testing RBC models is circumvented by considering a subset of variables (and their lagged values) for which a non singular distribution exists. For example, in the case of RBC models driven by only one shock, say a technology shock, their approach can be used to evaluate the model's joint CDF of current and lagged output, including autocorrelation and second moment structures, etc. In the case of models driven by two shocks, say a technology and a preference shock, the approach can be used for the comparison of the joint CDF of current (and lagged) output and hours worked, say. In general, their testing framework can be used to address questions of the following sort: (i) For a given RBC model, what is the relative usefulness of different sets of calibrated parameters for mimicing different dynamic features of output growth? (ii) Given a fixed set of calibrated parameters, what is the relative performance of RBC models driven by shocks with a different marginal distribution?

More specifically, consider m RBC models, and assume that the variables of interest are output and lagged output. Now, set model 1 as the benchmark model. Let $\Delta \log X_t$, $t = 1, ..., T$ denote actual historical output (growth rates), and let $\Delta \log X_{j,n}$, $j = 1, ..., m$ and $n = 1, ..., S$, denote the output series simulated under model j, where S denotes the length of the simulated sample. Denote $\Delta \log X_{j,n}(\hat{\theta}_{j,T})$, $n = 1, ..., S$, $j = 1, ..., m$ to be a sample of length S drawn (simulated) from model j and evaluated at the parameters estimated under model j, where parameter estimation is done using the T available historical observations. Further, let $Y_t = (\Delta \log X_t, \Delta \log X_{t-1})$, $Y_{j,n}(\hat{\theta}_{j,T}) = (\Delta \log X_{j,n}(\hat{\theta}_{j,T}), \Delta \log X_{j,n-1}(\hat{\theta}_{j,T}))$, and let $F_0(u; \theta_0)$ denote the distribution of Y_t evaluated at u and $F_j(u; \theta_j^\dagger)$ denote the distribution of $Y_{j,n}(\theta_j^\dagger)$, where $\theta_j^\dagger$ is the probability limit of $\hat{\theta}_{j,T}$, taken as $T \to \infty$, and where $u \in U \subset \Re^2$, possibly unbounded. As above, accuracy is measured in terms of squared error. The squared (approximation) error associated with model j, $j = 1, ..., m$, is measured in terms of the (weighted) average over U of $\left(F_j(u; \theta_j^\dagger) - F_0(u; \theta_0) \right)^2$, where $u \in U$, and U is a possibly unbounded set on $\Re^2$. Thus, the rule is to choose Model 1 over Model 2 if

$$\int_U \left(F_1(u; \theta_1^\dagger) - F_0(u; \theta_0) \right)^2 (u) du < \int_U \left(F_2(u; \theta_2^\dagger) - F_0(u; \theta_0) \right)^2 (u) du,$$

11

where $\int_U w(u)du = 1$ and $w(u) \geq 0$ for all $u \in U \subset \Re^2$. For any evaluation point, this measure defines a norm and is a typical goodness of fit measure. Note that within our context, the hypotheses of interest are:

$$H_0 : \max_{j=2,...,m} \int_U \left[\left(F_0(u;\theta_0) - F_1(u;\theta_1^\dagger)\right)^2 - \left(F_0(u) - F_j(u;\theta_j^\dagger)\right)^2 \right] w(u)du \leq 0$$

and

$$H_A : \max_{j=2,...,m} \int_U \left[\left(F_0(u) - F_1(u;\theta_1^\dagger)\right)^2 - \left(F_0(u) - F_j(u;\theta_j^\dagger)\right)^2 \right] w(u)du > 0.$$

Thus, under H_0, no model can provide a better approximation (in a squared error sense) to the distribution of Y_t than the approximation provided by model 1. If interest focuses on confidence intervals, so that the objective is to "approximate" $\Pr(\underline{u} \leq Y_t \leq \overline{u})$, then the null and alternative hypotheses can be stated as:

$$H_0' : \max_{j=2,...,m} \left[\left(F_1(\overline{u};\theta_1^\dagger) - F_1(\underline{u};\theta_1^\dagger) - (F_0(\overline{u};\theta_0) - F_0(\underline{u};\theta_0))\right)^2 \right.$$
$$\left. - \left(F_j(\overline{u};\theta_j^\dagger) - F_j(\underline{u};\theta_j^\dagger) - (F_0(\overline{u};\theta_0) - F_0(\underline{u};\theta_0))\right)^2 \right] \leq 0.$$

versus

$$H_A' : \max_{j=2,...,m} \left[\left(F_1(\overline{u};\theta_1^\dagger) - F_1(\underline{u};\theta_1^\dagger) - (F_0(\overline{u};\theta_0) - F_0(\underline{u};\theta_0))\right)^2 \right.$$
$$\left. - \left(F_j(\overline{u};\theta_j^\dagger) - F_j(\underline{u};\theta_j^\dagger) - (F_0(\overline{u};\theta_0) - F_0(\underline{u};\theta_0))\right)^2 \right] > 0.$$

If interest focuses on testing the null of equal accuracy of two distribution models (analogous to the pairwise conditional mean comparison setup of Diebold and Mariano (1995)), we can simply state the hypotheses as:

$$H_0'' : \int_U \left[\left(F_0(u;\theta_0) - F_1(u;\theta_1^\dagger)\right)^2 - \left(F_0(u) - F_j(u;\theta_j^\dagger)\right)^2 \right] w(u)du = 0$$

versus

$$H_A'' : \int_U \left[\left(F_0(u) - F_1(u;\theta_1^\dagger)\right)^2 - \left(F_0(u) - F_j(u;\theta_j^\dagger)\right)^2 \right] w(u)du \neq 0,$$

In order to test H_0 versus H_A, the relevant test statistic is $\sqrt{T}Z_{T,S}$, where: [2]

$$Z_{T,S} = \max_{j=2,...,m} \int_U Z_{j,T,S}(u) \, w(u)du, \tag{11}$$

---

[2] H_0' versus H_A' and H_0'' versus H_A'' can be tested in a similar manner.

and

$$Z_{j,T,S}(u) = \left(\frac{1}{T} \sum_{t=1}^{T} 1\{Y_t \le u\} - \frac{1}{S} \sum_{n=1}^{S} 1\{Y_{1,n}(\hat{\theta}_{1,T}) \le u\} \right)^2$$
$$- \left(\frac{1}{T} \sum_{t=1}^{T} 1\{Y_t \le u\} - \frac{1}{S} \sum_{n=1}^{S} 1\{Y_{j,n}(\hat{\theta}_{j,T}) \le u\} \right)^2 ,$$

with $\hat{\theta}_{j,T}$ an estimator of $\theta_j^{\dagger}$ that satisfies Assumption 2 below. See Corradi and Swanson (2003c) for further details.

Another measure of distributional accuracy available in the literature (see e.g. White (1982) and Vuong (1989)), is the KLIC, according to which we should choose Model 1 over Model 2 if: [3]

$$E(\log f_1(Y_t; \theta_2^{\dagger}) - \log f_2(Y_t; \theta_2^{\dagger})) > 0.$$

The KLIC is a sensible measure of accuracy, as it chooses the model which on average gives higher probability to events which have actually occurred. Also, it leads to simple Likelihood Ratio tests. Interestingly, Fernandez-Villaverde and Rubio-Ramirez (2001) have shown that the best model under the KLIC is also the model with the highest posterior probability. The above approach is an alternative to the KLIC that should be viewed as complementary in some cases, and preferred in others. For example, if we are interested in measuring accuracy over a specific region, or in measuring accuracy for a given confidence interval, this cannot be done in an obvious manner using the KLIC, while it can easily done using our measure. As an illustration, assume that we wish to evaluate the accuracy of different models in approximating the probability that the rate of growth of output is say between 0.5% and 1.5%. We can do so quite easily using the squared error criterion, but not using the KLIC. Furthermore, we often do not have an explicit form for the density implied by the various models we are comparing. Of course, model comparison can be done using kernel density estimators, within the KLIC framework. However, this leads to tests with nonparametric rates (see e.g. Zheng (2000)). On the other hand, comparison via our squared error measure of accuracy is carried out using empirical distributions, so that resulting test statistics converge at parametric rates.

---

[3]Recently, Giacomini (2002) proposes an extension which uses a weighted (over Y_t) version of the KLIC, and Kitamura (2002) suggests a generalization for choosing among models that satisfy some conditional moment restrictions.

Point forecast production and evaluation continues to receive considerable attention, and can perhaps be viewed as a leading indicator for the predictive density literature. There are now a host of tests based on traditional squared error loss criteria, but in addition tests based on directional forecast accuracy and sign tests, tests of forecast encompassing, as well as measures and tests based on other loss functions. Swanson and White (1997a) survey a number of the important contributions, which include Chao, Corradi and Swanson (2001), Chatfield (1993), Clark and McCracken (2001), Clements and Hendry (1993), Diebold and Chen (1996), Diebold and Mariano (1995), Hansen (2001), Hansen, Heaton and Luttmer (1995), Hansen and Jeganathan (1997), Harvey, Leybourne and Newbold (1996,1998), Linton, Maasoumi and Whang (2003), McCracken (1999), Pesaran and Timmerman (1992,1994,2000), Stekler (1991,1994), West (1996,2001) and West and McCracken (2002), among others. A number of these papers emphasise the role of parameter estimation uncertainty in testing for equal forecast accuracy, or that one model forecast encompasses another, as well as the consequences of the models being nested. Nevertheless, much remains to be done with regard to making these tests applicable to non-linear models, and constructing tests using non-linear and/or nondifferentiable loss functions, as well as allowing for parameter estimation error and misspecification when the comparisons involve non-linear models.

The above paragraph notes the role of the loss function in determining how accuracy is to be assessed. However there is also the issue of whether the in-sample model estimation criterion and out-of-sample forecast accuracy criterion should be matched. For example, it is only recently that much attention has been given to the notion that the same loss function used in-sample for parameter estimation is often that which should be used out-of-sample for forecast evaluation. Much remains to be done in this area, although progress has been made, as discussed in Christoffersen and Diebold (1996,1997), Clements and Hendry (1996), Granger (1993,1999), and Weiss (1996). That said, it is often difficult to come up with asymptotically valid inferential strategies using standard estimation procedures (that essentially minimize one-step ahead errors) for many varieties of non-linear models, from smooth transition models to projection pursuit and wavelet models. In some contexts it is even difficult to establish the consistency of some econometric parameter estimates, such as cointegrating vectors in certain non-linear cointegration models, and threshold parameters in some types of regime-switching models. In sum, estimation and in-sample inference of non-linear forecasting models remains

a potentially difficult task, with much work remaining to be done. Nevertheless, there are many recent papers that propose novel approaches to estimation, such as the variety of new cross-validation related techniques in the area of neural nets.

A general problem with non-linear models is the 'curse of dimensionality' and the fact that such models tend to have a large number of parameters (at least relative to the available number of macroeconomic data points) – how to keep the number of parameters at a tractable level? This sort of issue is relevant to the specification of many varieties of non-linear models, including smooth transition models (see e.g. Granger and Teräsvirta (1993)) and neural network models (see e.g. Swanson and White (1995, 1997a,b)). A counter to the fear that such models may be overfitting in-sample – in the sense of picking up transient, accidental connections between variables – is of course to compare the models on out-of-sample performance, or on a 'hold-out' sample. The 'data-snooping procedures' developed by White (2000), and used in Sullivan, Timmermann and White (1999,2001,2003), can also be used, but this would appear to be an open area, with much room for advance. For example, the extension of the data-snooping methodology to multivariate models awaits attention, where the sheer magnitude of the problem can quickly grow out of hand.

Finally, although not germane to forecasting, some models have parameters that are readily interpretable, whilst others are less clear. The study by De Gooijer and Vidiella-i-Anguera (2004) extends 'non-linear cointegration' to allow the equilibrium, cointegrating relationship to depend upon the regime. Difficult issues arise when cointegration is no longer 'global', and the theory-justification for the long-run relationship is less clear. Clements and Galvão (2004) review specification and estimation procedures in systems when cointegration is 'global', and evaluate the forecasting ability of non-linear systems in the context of interest rate prediction, building on earlier contributions by Anderson (1997), van Dijk and Franses (2000) and Kunst (1992), inter alia. Using a variety of forecast evaluation methods, De Gooijer and Vidiella-i-Anguera establish the superiority of their model from a forecasting perspective.

Another new model is developed by Franses, Paap and Vroomen (2004), who propose a model in which a key autoregressive parameter depends on a leading indicator variable. This model captures the notion that some variables are relevant for forecasting only once in a while, and it mimics some of the ideas put forward in Franses and Paap (2002). The autoregressive parameter is constant unless a linear function of the leading indicator plus a disturbance term

exceeds a certain threshold level. The authors discuss issues relating to estimation, inference and forecasting, and the relationship of their model to existing non-linear models. The model is applied to forecasting unemployment, and is shown to be capable of capturing the sharp increases in unemployment in recessions, and to provide competitive forecasts compared to alternative models.

A number of possible approaches are available to account for the possibility that the parameters in forecasting models are changing over time. Taylor (2004) uses adaptive exponential smoothing methods that allow smoothing parameters to change over time, in order to adapt to changes in the characteristics of the time series. More specifically, he presents a new adaptive method for predicting the volatility in financial returns, where the smoothing parameter varies as a logistic function of user-specified variables. The approach is analogous to that used to model time-varying parameters in smooth transition GARCH models. These non-linear models allow the dynamics of the conditional variance model to be influenced by the sign and size of past shocks. These factors can also be used as transition variables in the new smooth transition exponential smoothing approach. Parameters are estimated for the method by minimising the sum of squared deviations between realised and forecast volatility. Using stock index data, the new method gives encouraging results when compared to fixed-parameter exponential smoothing and a variety of GARCH models.

Bradley and Jansen (2004) propose a model in which the dynamics that characterise stock returns are allowed to differ in periods following a large swing in stock returns – that is, a non-linear state-dependent model. Their approach allows them to test for the existence of non-linearities in returns, and to estimate the size of the shock that is required to cause the non-linear behaviour.

4   Empirical issues

In this section we discuss various practical issues. In contrast to linear models, the design of non-linear models for actual data and the estimation of parameters is less straightforward.

How should we select a model?[4] Should all the observations be used, or should models be

---

[4]There is a growing literature on nonparametric and semi-parametric forecasting methods and models. Traditionally this literature has focused on the conditional mean, but a number of recent papers have looked at

16

estimated and/or evaluated against specific (dynamic) features of the historical record? Should we split the sample into in- and out-of-sample periods, and if so, where should the split occur?

Boero and Marrocu (2004) provide evidence related to some of these questions. They analyse the out-of-sample performance of SETAR models relative to a linear AR and a GARCH model using daily data for the Euro effective exchange rate. Their evaluation is conducted on point, interval and density forecasts, unconditionally, over the whole forecast period, and conditional on specific regimes. Their results show that the GARCH model is better able to capture the distributional features of the series and to predict higher-order moments than the SETAR models. However, their results also indicate that the performance of the SETAR models improves significantly conditional on being in specific regimes. In a related study, Corradi and Swanson (2004) focus on various approaches to assessing predictive accuracy. One of the main conclusions of their study is that there are various easy to apply statistics that can be constructed using out-of-sample conditional-moment conditions, which are robust to the presence of dynamic misspecification. Because estimated models are approximations to the DGP and likely to be mis-specified in unknown ways, tests that are robust in this sense are obviously desirable. They provide an illustration of model selection via predictive ability testing involving the US money-income relation, and demonstrate the relevance of the various testing methods.

Next, which non-linear models should be entertained in any specific instance? This question can be answered by looking at the theoretical properties of the models, as well as the specific properties of the data under scrutiny. For example, Dahl and Hylleberg (2004) give a nice review of four non-linear models, namely, Hamilton's flexible non-linear regression model (Hamilton, 2001), artificial neural networks and two versions of the projection pursuit regression model. The forecasting performance of these four approaches is compared for U.S. industrial production and an unemployment rate series, in a 'real-time forecasting' exercise, whereby the specification of the model is chosen, and the parameters re-estimated at each step, as the forecast origin moves through the available sample. The paper provides some guidance for model selection,

---

nonparametric estimation of other aspects of conditional ditributions, such as quartiles, intervals and density regions: see e.g., De Gooijer, Gannoun and Zerom (2002), Matzner-Løber, Gannoun and De Gooijer (1998) and Samanta (1989), and the references therein. Developments in these areas look set to continue apace with the more parametric approaches considered in this issue.

and evaluates the resulting forecasts using standard MSE-related criteria as well as direction-of-change tests. Interestingly, they find evidence that some of the flexible non-linear regression models perform well relative to the non-linear benchmark.

Further, how can we reliably estimate the model parameters? Can we address the problem whereby we only find local optima? How can we select good starting values in non-linear optimization? Can we design methods to test the non-linear models, based on in-sample estimation and in-sample data? How should we aggregate and analyse our data? Some of these questions are examined in Van Dijk and Franses (2000), who consider daily, weekly and monthly data, and find distinct models for different temporal aggregation levels. But, what happens if we aggregate over the cross-section dimension? Marcellino (2004) argues that cross-section aggregation of countries, with constant weights over time, may produce 'smoother' series better suited for linear models, while aggregation with time-varying weights (and the presence of common shocks) is more likely to generate a role for non-linear modelling of the resultant series. Marcellino (2004) fits a variety of non-linear and time-varying models to aggregate EMU macroeconomic variables, and compares them with linear models. He assesses the quality of these models in a real-time forecasting framework. It is found that often non-linear models perform best.

Of course the bottom line is: Are there clear-cut examples where actual forecast improvements are delivered by non-linear models? Provision of such examples might serve to allay the fears held by many sceptics. Three nice recent examples of this sort are Clements and Galvão (2004), Sensier et al. (2004) and Gencay and Selcuk (2004), all of whom show the relevance of non-linear models for forecasting. The latter paper suggests that the use of non-linearities is crucial for making sensible statements about the tail behaviour of asset returns. In particular, Gencay and Selcuk (2004) investigate the relative performance of Value-at-Risk (VaR) models with the daily stock market returns of nine different emerging markets. In addition to well-known modeling approaches such as the variance-covariance method and historical simulation, they employ extreme value theory (EVT) to generate VaR estimates and provide the tail forecasts of daily returns at the 0.999 percentile along with 95 percent confidence intervals for stress testing purposes. The results indicate that EVT based VaR estimates are more accurate at higher quantiles. According to estimated Generalized Pareto Distribution parameters, certain moments of the return distributions do not exist in some countries. In addition, the

daily return distributions have different moment properties in their right and left tails. Therefore, risk and reward are not equally likely in these economies. The other two papers consider macroeconomic variables. Sensier et al. (2004) examine the role of domestic and international variables for predicting classical business cycles regimes in four European countries, where the regimes are classified as binary variables. One finding is that composite leading indicators and interest rates of Germany and the US have substantial predictive value. Clements and Galvao (2004) test whether there is non-linearity in the response of short and long-term interest rates to the spread. They assess the out-of-sample predictability of various models and find some evidence that non-linear models lead to more accurate short-horizon forecasts, especially of the spread. And, as mentioned, De Gooijer and Vidiella-i-Anguera (2004) report more marked gains to allowing for non-linearities.

5   Concluding remarks

In this paper we have summarized the state-of-the-art in forecast construction and evaluation for non-linear models, and in selection among alternative non-linear prediction models. We conclude that the day is still long off when simple, reliable and easy to use non-linear model specification, estimation and forecasting procedures will be readily available. Nevertheless, there are grounds for optimism. The papers in this issue suggest that careful application of existing techniques, and new models and tests, can result in significant advances in our understanding. Supposing that the world is inherently non-linear, then as computational capabilities increase, more complex models become amenable to analysis, allowing the possibility that future generations of models will significantly outperform linear models, especially if such models become truly multivariate.

Much remains to be done in the areas of specification, estimation, and testing, with important issues of non-differentiability, parameter-estimation error and data-mining remaining to be addressed. In addition, a further area for research (both empirical and theoretical) is the following. Suppose one has data with trend components, seasonality, non-linearity and outliers. How should one proceed? This is a complex issue, due to the possible inter-reactions between these elements, e.g., neglecting outliers may suggest non-linearity (see van Dijk et al. (1999)) and the trend can be intertwined with seasonality, as in periodic models of seasonality (see

Franses and Paap (2004) for an up-to-date survey). Much has been said about modelling each of these characteristics in isolation, but developing coherent strategies for such data remains an important task. It will be interesting to see to what extent developments in these areas give rise to tangible gains in terms of forecast performance – we remain hopeful that great strides will be made in the near future.

Biographies

Michael P. Clements is a Reader in the Department of Economics at the University of Warwick. His research interests include time-series modelling and forecasting. He has published in a variety of international journals, has co-authored two books on forecasting, and co-edited (with David F Hendry) of A Companion to Economic Forecasting 2002, Blackwells. He is also an editor of the International Journal of Forecasting.

Philip Hans Franses is Professor of Applied Econometrics and Professor of Marketing Research, both at the Erasmus University Rotterdam. He publishes on his research interests, which are applied econometrics, time series, forecasting, marketing research and empirical finance.

Norman Swanson, a 1994 University of California, San Diego Ph.D. is currently Associate Professor at Rutgers University. His research interests include forecasting, financial- and macro-econometrics. He is currently an associate editor of the Journal of Business and Economic Statistics, the International Journal of Forecasting, and Studies in Non-linear Dynamics and Econometrics. He has recently been awarded a National Science Foundation research grant entitled the Award for Young Researchers, and is a member of various professional organizations, including the Econometric Society, the American Statistical Association, and the American Economic Association. Swanson has recent publications in Journal of Development Economics, Journal of Econometrics, Review of Economics and Statistics, Journal of Business and Economic Statistics, Journal of the American Statistical Association, Journal of Time Series Analysis, Journal of Empirical Finance, and International Journal of Forecasting, among others. Further details about Swanson, including copies of all papers cited above, are available at his website: http://econweb.rutgers.edu/nswanson/

6   References

Anderson, H. M. (1997). Transaction costs and non-linear adjustment towards equilibrium in the US treasury bill market. Oxford Bulletin of Economics and Statistics, 59, 465—484.

Andrews, D.W.K., (1997), A conditional Kolmogorov test, Econometrica, 65, 1097-1128.

Bai, J., (2001), Testing parametric conditional distributions of dynamic models, Review of Economics and Statistics, forthcoming.

Boero, G. and E. Marrocu (2004), The performance of SETAR models: A regime conditional evaluation of point, interval and density forecasts, International Journal of Forecasting, 20, xxx-xxx.

Bradley, M.D. and D.W. Jansen (2004), Forecasting with a non-linear dynamic model of stock returns and industrial production, International Journal of Forecasting, 20, xxx-xxx.

Chatfield, C., (1993), Calculating interval forecasts, Journal of Business and Economic Statistics, 11, 121-135.

Christoffersen, P.F., (1998), Evaluating interval forecasts, International Economic Review, 39, 841-862.

Christoffersen, P.F. and F.X. Diebold, (1996), Further results on forecasting and model selection under asymmetric loss, Journal of Applied Econometrics, 11, 561-572.

Christoffersen, P.F. and F.X. Diebold, (1997), Optimal prediction under asymmetric loss, Econometric Theory, 13, 808-817.

Christoffersen, P.F., J. Hahn and A. Inoue, (2001), Testing and comparing Value-at-Risk measures, Journal of Empirical Finance, 8, 325-342.

Clark, T.E. and M.W. McCracken, (2001), Tests of equal forecast accuracy and encompassing for nested models, Journal of Econometrics, 105, 85-110.

Clements, M.P. and A.B. Galvão, (2004), A comparison of tests of non-linear cointegration with an application to the predictability of US interest rates using the term structure, International Journal of Forecasting, 2004, xxx-xxx.

Clements, M.P. and D.F. Hendry, (1993), On the limitations of comparing mean squared forecast errors: Comment, Journal of Forecasting, 12, 617-637.

Clements, M. P., and Hendry, D. F. (1996). Multi-step estimation for forecasting. Oxford Bulletin of Economics and Statistics, 58, 657—684.

Clements, M. P., and Hendry, D. F. (1999), Forecasting non-stationary economic time series. Cambridge, Mass.: MIT Press. The Zeuthen Lectures on Economic Forecasting.

Clements, M. P., and Krolzig, H.-M. (1998), A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP. Econometrics Journal, 1, C47—75.

Clements, M. P., and Krolzig, H.-M. (2002), Can oil shocks explain asymmetries in the US Business Cycle? . Empirical Economics, 27, 185—204.

Clements, M. P., and Krolzig, H.-M. (2003), Business cycle asymmetries: Characterisation and testing based on Markov-switching autoregressions. Journal of Business and Economic Statistics, 21, 196—211.

Clements, M. P., and Smith, J. (1999), A Monte Carlo study of the forecasting performance of empirical SETAR models. Journal of Applied Econometrics, 14, 124—141.

Clements, M.P. and J. Smith, (2000), Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment, Journal of Forecasting, 19, 255-276.

Clements, M.P. and J. Smith, (2002), Evaluating multivariate forecast densities: A comparison of two approaches, International Journal of Forecasting, 18, 397-407.

Corradi, V. and N.R. Swanson, (2002), A consistent test for out of sample non-linear predictive ability, Journal of Econometrics, 110, 353-381.

Corradi, V., N.R. Swanson, (2003a), The block bootstrap for recursive m-estimators with applications to predictive evaluation, Working Paper, Queen Mary, University of London, University of Exeter and Rutgers University.

Corradi, V., N.R. Swanson, (2003b), A test for comparing multiple misspecified conditional distribution models, Working Paper, Queen Mary, University of London and Rutgers University.

Corradi, V., N.R. Swanson, (2003c), Evaluation of dynamic stochastic general equilibrium models based on a distributional comparison of simulated and historical data, Journal of Econometrics, forthcoming.

Corradi, V. and N.R. Swanson (2004), Some recent developments in predictive accuracy testing with nested models and (generic) non-linear alternatives, International Journal of Forecasting, 20, xxx-xxx.

Dacco, R., and Satchell, S. (1999). Why do regime-switching models forecast so badly. Journal of Forecasting, 18, 1—16.

Dahl, C.M. and S. Hylleberg (2004), Flexible regression models and relative forecast performance, International Journal of Forecasting, 20, xxx-xxx.

De Gooijer, J.G., A. Gannoun and D. Zerom (2002), Mean squared error properties of the kernel-based multi-stage median predictor for time series. Statistics & Probability Letters, 56, 51-56.

De Gooijer, J.G. and K. Kumar (1992), Some recent developments in non-linear time series modelling, testing, and forecasting, International Journal of Forecasting, 8, 135-156.

De Gooijer, J.G. and A. Vidiella-i-Anguera (2004), Forecasting threshold cointegrated systems, International Journal of Forecasting, 20, xxx-xxx.

Diebold, F.X. and C. Chen, (1996), Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures, Journal of Econometrics, 70, 221-241.

Diebold, F.X., T. Gunther and A.S. Tay, (1998), Evaluating density forecasts with applications to finance and management, International Economic Review, 39, 863-883.

Diebold, F.X., J. Hahn and A.S. Tay, (1999), Multivariate density forecast evaluation and calibration in financial risk management: High frequency returns on foreign exchange, Review of Economics and Statistics, 81, 661-673.

Diebold, F.X., and R.S. Mariano, (1995), Comparing predictive accuracy, Journal of Business and Economic Statistics, 13, 253-263.

Diebold, F. X., and Nason, J. A. (1990). Nonparametric exchange rate prediction. Journal of International Economics, 28, 315—332.

Fernandez-Villaverde, J., and J.F. Rubio-Ramirez, (2001), Comparing dynamic equilibrium models to data, Manuscript, University of Pennsylvania.

Franses, P.H. and D.J.C. van Dijk (2000), Non-linear time series models in empirical finance, Cambridge: Cambridge University Press.

Franses, P.H. and R. Paap (2002), Censored latent effects autoregression, with an application to US unemployment, Journal of Applied Econometrics, 17, 347-366.

Franses, P.H. and R. Paap (2004), Periodic time series models, Oxford: Oxford University Press.

Franses, P.H., R. Paap and B. Vroomen (2004), Forecasting unemployment using an autoregression with censored latent effects parameters, International Journal of Forecasting, 20, xxx-xxx.

Gençay, R. and F. Selçuk (2004), Extreme value theory and Value-at-Risk: relative performance in emerging markets, International Journal of Forecasting, 20, xxx-xxx.

Giacomini, R. (2002), Comparing density forecasts via weighted Likelihood Ratio tests: Asymptotic and bootstrap methods, manuscript, University of California, San Diego.

Giacomini, R. and H. White (2003), Tests of conditional predictive ability, manuscript, University of California, San Diego.

Granger, C.W.J. and T. Teräsvirta (1993), Modelling non-linear economic relationships, New York, Oxford University Press.

Granger, C.W.J., (1999), Outline of forecast theory using generalized cost functions, Spanish Economic Review, 1, 161-173.

Granger, C.W.J. and T. Teräsvirta (1993), On the limitations of comparing Mean Squared Forecast Errors: Comment, Journal of Forecasting, 12, 651-652.

Hamilton, J. D. (1983). Oil and the Macroeconomy since World War II. Journal of Political Economy, 91, 228—248.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica, 57, 357—384.

Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. Journal of Monetary Economics, 38, 215—220.

Hamilton, J. D. (2000). What is an oil shock?. mimeo, Department of Economics, UCSD, La Jolla, California.

Hamilton, J. D. (2001). A parametric approach to flexible non-linear inference. Econometrica, 69, 537—573.

Hansen, L.P., J. Heaton and E.G.J, Luttmer (1995), Econometric evaluation of asset pricing models, Review of Financial Studies, 8, 237-279.

Hansen, L.P. and R. Jeganathan (1997), Assessing specification errors in stochastic discount factor models, Journal of Finance, 52, 557-590.

Hansen, P.R., (2001), An unbiased and powerful test for superior predictive ability, Manuscript, Brown University.

Harding, D., and Pagan, A. (2001). Dissecting the cycle: A methodological investigation. Journal of Monetary Economics, 49, 365-381.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13, 281—291.

Harvey, D. I., Leybourne, S., and Newbold, P. (1998). Tests for forecast encompassing. Journal of Business and Economic Statistics, 16, 254—259.

Hong, Y., (2001), Evaluation of out of sample probability density forecasts with applications to S&P 500 stock prices, Mimeo, Cornell University.

Hooker, M. A. (1996). Whatever happened to the oil price-macroeconomy relationship?. Journal of Monetary Economics, 38, 195—213.

Kitamura, Y., (2002), Econometric comparisons of conditional models, University of Pennsylvania, Mimeo.

Koop, G., and Potter, S. (2000). Non-linearity, structural breaks, or outliers in economic time series. In Barnett, W. A., Hendry, D. F., Hylleberg, S., Teräsvirta, T., Tjostheim, D., and Wurtz, A. (eds.), Non-linear Econometric Modelling in Time Series Analysis, Cambridge: Cambridge University Press, 61-78.

Krolzig, H.-M. (2003). Predicting markov-switching vector autoregressive processes. Journal of Forecasting. Forthcoming.

Kunst, R.M. (1992), Threshold cointegration in interest rates, Discussion Paper 92-26, Department of Economics, UC San Diego.

Linton, O., E. Maasoumi and Y.J. Whang, (2003), Consistent testing for stochastic dominance under general sampling schemes, Manuscript, LSE, Southern Methodist University and Ewha University.

Marcellino, M. (2004), Forecasting EMU macroeconomic variables, International Journal of Forecasting, 20, xxx-xxx.

Matzner-Løber, E., A. Gannoun and J.G. De Gooijer (1998), Nonparametric forecasting: a comparison of three kernel-based methods. Communications in Statistical Theory Methods, 27, 1593-1617.

McCracken, M.W. (1999), Asymptotics for out of sample tests of causality, Working Paper, Louisiana State University.

Mork, K. A. (1989). Oil and the Macroeconomy when prices go up and down: An extension of Hamilton's results. Journal of Political Economy, 97, 740—744.

Pagan, A. R. (1997a). Policy, theory and the Cycle. Oxford Review of Economic Policy, 13, 19—33.

Pagan, A. R. (1997b). Towards an understanding of some Business Cycle characteristics. Australian Economic Review, 30, 1—15.

Pesaran, M.H. and A.G. Timmerman, (1992), A simple nonparametric test of predictive performance, Journal of Business and Economic Statistics, 10, 461-465.

Pesaran, M.H. and A.G. Timmerman, (1994), A generalization of the non-parametric Henriksson-Merton test of market timing, Economics Letters, 44, 1-7.

Pesaran, M.H. and A.G. Timmerman, (2000), A recursive modelling approach to predicting UK stock returns. Economic Journal, 159-191.

Raymond, J.E., and Rich, R.W.(1997).Oil and the macroeconomy: A Markov state-switching approach. Journal of Money, Credit, and Banking, 29, 193—213.

Samanta, M. (1989). Nonparametric estimation of conditional quantiles. Statistics and & Probability Letters, 7, 407-412.

Sensier, M, M. Artis, D.R. Osborn, and C. Birchenhall (2004), Domestic and international influences on business cycle regimes in Europe, International Journal of Forecasting, 20, xxx-xxx.

Sichel, D. E. (1994). Inventories and the three phases of the business cycle. Journal of Business and Economic Statistics, 12, 269—277.

Stekler, H.O., (1991), Macroeconomic forecast evaluation techniques, International Journal of Forecasting, 7, 375-384.

Stekler, H.O., (1994), Are economic forecasts valuable?, Journal of Forecasting, 13, 495-505.

Sullivan, R., A.G. Timmerman and H. White, (1999), Data-snooping, technical trading rules and the bootstrap, Journal of Finance, 54, 1647-1692.

Sullivan, R., A.G. Timmerman and H. White, (2001), Dangers of data-driven inference: The case of calendar effects in stock returns, Journal of Econometrics, 249-286.

Sullivan, R., A.G. Timmerman and H. White, (2003), Forecast evaluation with shared data sets, International Journal of Forecasting, 19, 217-227.

Swanson, N.R. and H. White, (1995), A model selection approach to assessing the information in the term structure using linear models and artificial neural networks, Review of Economics and Statistics, 13, 265-279.

Swanson, N.R. and H. White, (1997a), Forecasting economic time series using adaptive versus nonadaptive and linear versus non-linear econometric models, International Journal of Forecasting, 13, 439-461.

Swanson, N.R. and H. White, (1997b), A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks, Review of Economics and Statistics, 79, 540-550.

Taylor, J.W. (2004), Volatility forecasting with smooth transition exponential smoothing, International Journal of Forecasting, 20, xxx-xxx.

Teräsvirta, T., and Anderson, H. M. (1992). Characterizing non-linearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics, 7,119— 139.

Terui, N and H.K. van Dijk (2002), Combined forecasts from linear and non-linear time series models, International Journal of Forecasting, 18, 421-438.

Tiao, G. C., and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. Journal of Forecasting, 13, 109—131.

Tong, H. (1995). A personal overview of non-linear time series analysis from a chaos perspective. Scandinavian Journal of Statistics, 22, 399—445.

van Dijk, D.J.C and P.H. Franses (2000), Non-linear error correction models for interest rates in The Netherlands, in Non-linear Econometric Modelling in Time Series Analysis (W. Barnett, D.F. Hendry, S. Hylleberg, T.Teräsvirta, D. Tjøstheim and A.W. Würtz, eds.), Cambridge: Cambridge University Press (2000), 203-227.

van Dijk, D.J.C., P.H. Franses and A Lucas (1999), "Testing for smooth transition non-linearity in the presence of outliers, Journal of Business and Economic Statistics, 17, 217-235.

van Dijk, D.J.C., P.H. Franses and R. Paap (2002), A non-linear long memory model, with an application to US unemployment, Journal of Econometrics, 110, 135-165.

van Dijk, D.J.C., P.H. Franses and T. Teräsvirta (2002), Smooth transition autoregressive models – a survey of recent developments, Econometric Reviews, 21, 1-47.

Vuong, Q. (1989), Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica, 57, 307-333.

Weiss, A., (1996) Estimating time series models using the relevant cost function, Journal of Applied Econometrics, 11, 539-560.

West, K., (1996), Asymptotic inference about predictive ability, Econometrica, 64, 1067-1084.

West, K. D., and McCracken, M. W. (2002). Inference about predictive ability. In Clements, M. P., and Hendry, D. F. (eds.), A Companion to Economic Forecasting, pp. 299—321: Oxford: Blackwells.

White, H., (2000), A reality check for data snooping, Econometrica, 68, 1097-1126.

Zheng, J.X. (2000), A consistent test of conditional parametric distributions, Econometric Theory, 16, 667-691.