ISSN 1440-771X

MONASH UNIVERSITY
AUSTRALIA

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Local Linear Forecasts Using Cubic Smoothing Splines

Rob J. Hyndman, Maxwell L King, Ivet Pitrun and Baki Billah

Working Paper 10/2002

Local linear forecasts using cubic smoothing splines

Rob J. Hyndman, Maxwell L. King, Ivet Pitrun, Baki Billah 27 August 2002

Abstract: We show how cubic smoothing splines fitted to univariate time series data can be used to obtain local linear forecasts. Our approach is based on a stochastic state space model which allows the use of a likelihood approach for estimating the smoothing parameter, and which enables easy construction of prediction intervals. We show that our model is a special case of an ARIMA(0,2,2) model and we provide a simple upper bound for the smoothing parameter to ensure an invertible model. We also show that the spline model is not a special case of Holt's local linear trend method. Finally we compare the spline forecasts with Holt's forecasts and those obtained from the full ARIMA(0,2,2) model, showing that the restricted parameter space does not impair forecast performance.

Key words: ARIMA models, exponential smoothing, Holt's local linear forecasts, maximum likelihood estimation, nonparametric regression, smoothing splines, state space model, stochastic trends.

JEL classification numbers: C13, C14, C22, C53.

*Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800. Correspondence to Rob.Hyndman@buseco.monash.edu.au.

1 Introduction

Suppose we observe a univariate time series $\{y_t\}$, $t = 1, \dots, n$, with non-linear trend. We are interested in forecasting the series by extrapolating the trend using a linear function estimated from the observed time series.

Linear trend extrapolation is very widely used and performs relatively well in practice. For example, Makridakis & Hibon (2000), Assimakopoulos & Nikolopoulos (2000), and Hyndman & Billah (2002) discuss the excellent performance of linear trend methods in the M3-competition. In this paper, we discuss a method for local linear extrapolation of a stochastic trend based on cubic smoothing splines.

For equally spaced time series, a cubic smoothing spline can be defined as the function $\hat{f}(t)$ which minimises

$$\sum_{t=1}^{n}(y_t - f(t))^2 + \lambda \int_{\mathcal{S}} [f''(u)]^2\, du \tag{1.1}$$

over all twice differentiable functions f on $\mathcal{S}$ where $[1, n] \subseteq \mathcal{S} \subseteq \mathbb{R}$. The smoothing parameter λ is controlling the "rate of exchange" between the residual error described by the sum of squared residuals and local variation represented by the square integral of the second derivative of f. For a given λ, fast algorithms for computing $\hat{f}(t)$ are described by Green and Silverman (1994). Large values of λ give $\hat{f}(t)$ close to linear while small values of λ give a very wiggly function $\hat{f}(t)$. In practice, λ is not generally known.

The solution to (1.1) consists of piecewise cubic polynomials joined at the times of observation, $t = 1, 2, \dots, n$. Furthermore, the solution has zero second derivative at $t = n$. Therefore, an extrapolation of $\hat{f}(t)$ for $t > n$ is linear. The linear extrapolation of $\hat{f}(t)$ provides our point forecasts.

We derive a new method for computing prediction intervals for these forecasts, utilizing a stochastic model formulation due to Wahba (1978) and Wecker and Ansley (1983). We also provide a new method for estimating the smoothing parameter λ.

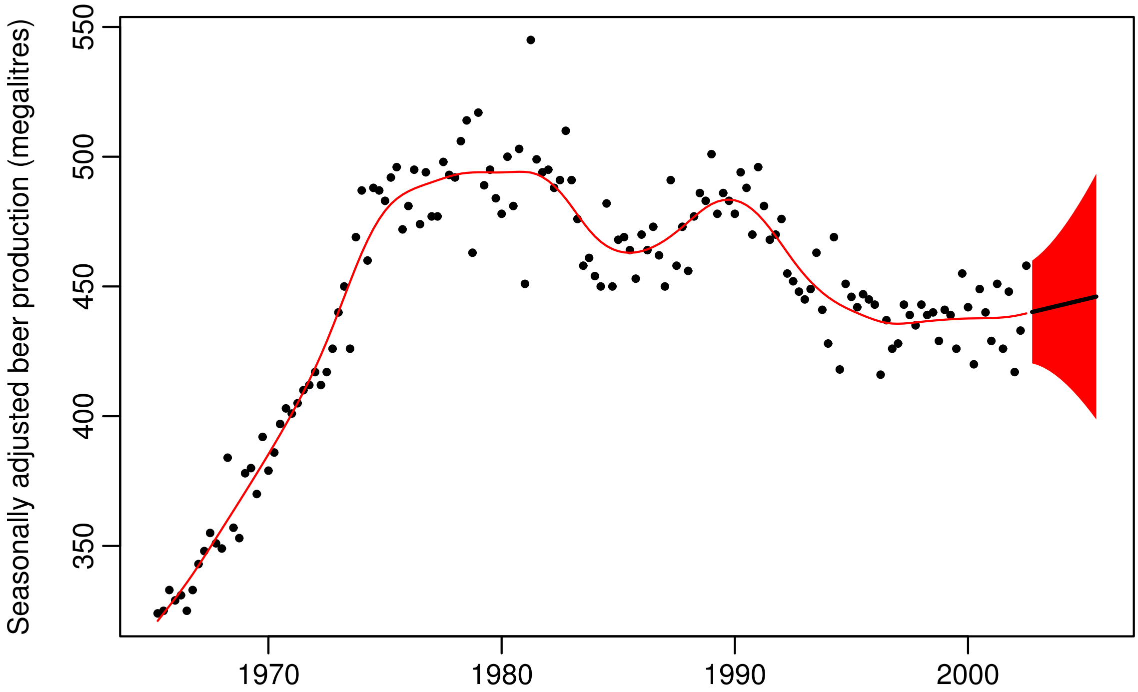

Figure 1: *Cubic spline forecasts of Australian quarterly beer production (seasonally adjusted) for September 2002 – June 2005, with 80% prediction intervals. The line through the historical data show the fitted cubic spline $\hat{f}(t)$; the forecasts are obtained by a linear extrapolation of $\hat{f}(t)$; the prediction intervals are obtained from the state space model described in Section 2. Here $\lambda = 232.2$.*

Figure 1 gives an example of our forecast procedure applied to seasonally adjusted Australian quarterly beer production (March 1965 – June 2002). The fitted spline curve is shown along with the associated linear forecast function and 80% prediction intervals. The methodology provides a smooth historical trend, a linear forecast function and prediction intervals.

Forecasts are usually made using models which give most weight to recent observations, and negligible weight to the distant past. This means that the smoothing parameter λ should not be too big for forecasting purposes. We make this explicit by finding the bounds on λ required for our model to be invertible. (Specifically, we find that $\lambda < 1.640519n^3$.)

Some linear forecast methods assume there is an underlying linear trend (e.g., a random walk with constant drift). We do not make this assumption. Our forecast function is linear, but the underlying trend $f(t)$ is allowed to be non-linear. Further, the possible future changes in trend direction are accommodated in our prediction intervals.

An alternative approach to local linear forecasting is to allow a deterministic non-linear trend. This is the approach followed by Nottingham and Cook (2001), for example. We prefer the stochastic trend approach as it allows the uncertainty in the trend to be explicitly allowed for in the measures of forecast uncertainty. A hybrid approach, combining both deterministic and stochastic trends, is provided through SEMIFAR models (see Beran and Ocker, 1999; and Beran and Feng, 2002).

Other local linear forecast models with stochastic trends include an ARIMA(0,2,2) model, Harvey's (1989) local linear growth model and the AN model of Hyndman, Koehler, Snyder & Grose (2002) which underlies Holt's (1957) linear trend method. In fact, these are all connected—Harvey's model is asymptotically equivalent to the AN model, and the AN model is a reparameterization of an ARIMA(0,2,2) model.

Our paper is structured as follows. Section 2 describes the stochastic model formulation for the cubic smoothing spline forecasts and Section 3 shows how to estimate the smoothing parameter. Simple expressions for obtaining point forecasts and prediction intervals are given in Section 4. In Section 5 we discuss the relationship between our model, an ARIMA(0,2,2) model and a state space model underlying Holt's linear trend forecasts. These relationships enable us to obtain the maximum bound for the smoothing parameter λ to ensure invertibility. Finally, in Section 6 we compare the forecasting performance of our model with other local linear forecasting models.

2 State space model

The definition of cubic smoothing splines given in Section 1 provides suitable point forecasts, but does not allow estimation of forecast uncertainty. To that end, we shall use the stochastic process formulation proposed by Wahba (1978) and developed in subsequent work of Wecker and Ansley (1983). We present Wecker and Ansley's state space model in the special case of cubic smoothing splines applied to equally spaced data.

First, we transform the observation time space to $[0,1]$ by defining the transformed observation times as $\{t_1,\ldots,t_n\}$ where $t_i = i/n$. Note that this transformation means that λ is rescaled also. Our transformed value of λ is $\lambda_* = n^{-3}\lambda$.

Then, for $i = 1, 2, \ldots$, define

$$g(t_i) = \tau \int_0^{t_i} (t_i - u)\, dW(u)$$

where $\tau > 0$ and $W(u)$ is a standard Wiener process. Also let

$$\boldsymbol{u}_i = \tau \begin{bmatrix} \int_{t_i-1}^{t_i} (t_i - u)\, dW(u) \\ W(t_i) - W(t_{i-1}) \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\alpha}_i = \begin{bmatrix} g(t_i) - g(t_1) \\ \tau(W(t_i) - W(t_1)) \end{bmatrix}.$$

Then we assume Y_i satisfies the state space model

$$\begin{aligned} Y_i &= \boldsymbol{s}_i'\boldsymbol{\beta} + (1,0)\boldsymbol{\alpha}_i + e_i, && (2.1)\\ \boldsymbol{\alpha}_i &= T_i\boldsymbol{\alpha}_{i-1} + \boldsymbol{u}_i, \qquad i = 1,\ldots,n && (2.2) \end{aligned}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is normally distributed with zero mean and covariance matrix cI,

$$T_i = \begin{bmatrix} 1 & i/n \\ 0 & 1 \end{bmatrix},$$

e_i are iid $N(0,\sigma^2)$ and $\boldsymbol{s}_i = (1, t_i)'$. The starting condition is $\boldsymbol{\alpha}_0 = (0,0)'$. The state $\boldsymbol{\alpha}_{i-1}$ is assumed independent of $\boldsymbol{u}_i$.

Wahba (1978) showed that

$$\lim_{c\to\infty} \mathrm{E}\left(\boldsymbol{s}_i'\boldsymbol{\beta} + (1,0)\boldsymbol{\alpha}_i \mid Y_1,\ldots,Y_n\right) \tag{2.3}$$

is the cubic smoothing spline $\hat{f}(t)$ with $\lambda_* = \sigma^2/\tau^2$. Thus $\hat{f}(t)$ is the mean of Y_t, and we can obtain point forecasts which extrapolate $\hat{f}(t)$ by applying the Kalman recursions to the state space model (2.1) and (2.2). Furthermore, we can also obtain forecast variances in this way.

However, a more direct approach is possible using a matrix formulation of the model. Let $\boldsymbol{Y} = (Y_1, \dots, Y_n)'$, $\boldsymbol{e} = (e_1, \dots, e_n)'$ and $\boldsymbol{g} = (g(t_1), \dots, g(t_n))'$. Then

$$\boldsymbol{Y} \quad = \quad S\boldsymbol{\beta} + \boldsymbol{g} + \boldsymbol{e} \tag{2.4}$$

where the ith row of S is $\boldsymbol{s}_i'$.

Proposition 1 *Let $\boldsymbol{Y}$ be given by (2.4). Then $\boldsymbol{Y}$ is normally distributed with mean $\mathbf{0}$ and covariance matrix*

$$\Omega = \sigma^2(cSS' + \lambda_*^{-1}\Sigma + I_n)$$

where c has been rescaled and where Σ is symmetric with the (j,k)th element on or above the diagonal given by

$$\Sigma_{jk} = \sigma^2 n^{-3} j^2(3k - j)/6, \qquad k \ge j.$$

That is

$$\Sigma = \frac{\sigma^2 n^{-3}}{6} \begin{bmatrix} 2 & 5 & 8 & \cdots & 3n-1 \\ 5 & 16 & 28 & & \vdots \\ 8 & 28 & 54 & & \\ \vdots & & & \ddots & \vdots \\ 3n-1 & \cdots & & \cdots & 2n^3 \end{bmatrix}.$$

We provide a proof for this result in the Appendix.

We shall use the stochastic formulation given by (2.4) and Proposition 1 to obtain point forecasts and prediction intervals.

3 Estimation

Estimates of the smoothing parameter λ_* can be obtained by maximizing the likelihood function of the model which is given by

$$\ell(\lambda_* \mid \boldsymbol{Y}) = |\Omega|^{-1/2}(\boldsymbol{Y}'\Omega^{-1}\boldsymbol{Y})^{-n/2}. \tag{3.1}$$

Let P be the upper-triangular matrix from the Choleski decomposition of $\sigma^2\Omega^{-1}$. (Note that P depends only on λ_*.) Then, we can write

$$|\Omega|^{-1/2} = \sigma^{-1}|P| \tag{3.2}$$

$$\text{and} \qquad (\boldsymbol{Y}'\Omega^{-1}\boldsymbol{Y})^{-n/2} = \sigma^n(\boldsymbol{Y}^{*\prime}\boldsymbol{Y}^*)^{-n/2} = \sigma^n\Big(\sum_{i=1}^n w_i^2\Big)^{-n/2} \tag{3.3}$$

where $\boldsymbol{Y}^* = P\boldsymbol{Y}$ and w_i is the ith element of $\boldsymbol{Y}^*$. Using (3.1)–(3.3), the log-likelihood is given by

$$\log \ell(\lambda_* \mid \boldsymbol{Y}) = (n-1)\log\sigma + \log|P| - \frac{n}{2}\log\Big(\sum_{i=1}^n w_i^2\Big). \tag{3.4}$$

Thus we can estimate λ_* by maximizing

$$\log|P| - \frac{n}{2}\log\Big(\sum_{i=1}^n w_i^2\Big).$$

This is a new method for selecting a bandwidth for smoothing splines, although it is similar in spirit to the likelihood-based method of Wecker and Ansley (1983). (Our method is much faster as we do not need to iteratively apply GLS estimation or the Kalman filter.)

4 Prediction

We now wish to use the fitted model to predict the next n_0 observations. We write them as the n_0-vector $\boldsymbol{Y}_0 = S_0\boldsymbol{\beta}+\boldsymbol{g}_0+\boldsymbol{e}_0$ where $\boldsymbol{Y}_0 = [Y_{n+1},\dots,Y_{n+n_0}]'$ and $\boldsymbol{g}_0$, $\boldsymbol{e}_0$ are defined analogously, and where S_0 has ith row $(1, t_{n+i})$, $i = 1,\dots,n_0$. We also define

Σ_0 as the symmetric $n_0 \times n_0$ matrix with (j,k)th element $\sigma^2 n^{-3}(n+j)^2(2n+3k-j)/6$ for $k \geq j$. It is assumed that $\boldsymbol{Y}_0$ has the same properties as the observed vector $\boldsymbol{Y}$. Then the variance-covariance of $\boldsymbol{Y}_0$ can be written as $\Omega_0 = \sigma^2(cS_0S_0' + \lambda_*^{-1}\Sigma_0 + I_{n_0})$.

To derive the best linear unbiased predictor for $\boldsymbol{Y}_0$ and the variance-covariance matrix of the associated prediction error, we first combine past and future values of $\{Y_t\}$ to obtain $\boldsymbol{Z} = [\boldsymbol{Y}', \boldsymbol{Y}_0']'$ with covariance matrix

$$E[\boldsymbol{Z}\boldsymbol{Z}'] = \begin{bmatrix} \Omega & U \\ U' & \Omega_0 \end{bmatrix} = \sigma^2(cS_1S_1' + \lambda_*^{-1}\Sigma_1 + I_{n+n_0})$$

where S_1 and Σ_1 are constructed analogously to S, S_0, Σ and Σ_0. Then, using standard results for conditional expectations of multivariate normal random variables (e.g., Rao, 1973, section 8a), we obtain

$$\mathrm{E}[\boldsymbol{Y}_0 \mid \boldsymbol{Y}] = U'\Omega^{-1}\boldsymbol{Y} \tag{4.1}$$

$$\text{and} \qquad \mathrm{Var}[\boldsymbol{Y}_0 \mid \boldsymbol{Y}] = \Omega_0 - U'\Omega^{-1}U. \tag{4.2}$$

Equations (4.1) and (4.2) allow point forecasts and associated prediction intervals to be easily computed. In particular, the h-step ahead point forecast $\hat{Y}_{n+h}$ is the hth element of $U'\Omega^{-1}\boldsymbol{Y}$, and its variance v_h is the hth diagonal element of the matrix $\Omega_0 - U'\Omega^{-1}U$. Since Y_t is assumed normal, prediction intervals can be constructed from these first two moments in the usual way. A 95% prediction interval is given by $\hat{Y}_{n+h} \pm 1.96\sqrt{v_h}$.

Note that these results assume that c, λ_* and σ^2 are known. In reality, c is any sufficiently large number (in the empirical calculations described in this paper we use $c = 100$), and the parameter λ_* can be estimated using the procedure described in Section 3.

To estimate σ^2, we first calculate one-step forecasts $\hat{Y}_t$ and associated "variances" v_t from (4.1) and (4.2) plugging in $\sigma^2 = 1$. This has no effect on the forecast means, but the forecast variances will be incorrect by a factor of σ^2. So σ^2 can be estimated as

$$\hat{\sigma}^2 = \sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2/v_t.$$

5 Comparisons with other approaches

The spline model described above gives local linear forecasts based on a stochastic trend. We now explore connections between this model and other models which also have stochastic trends and produce local linear forecast functions.

In particular, we look at the range of values for λ which will lead to an invertible model. Invertibility is a desirable property of a forecasting model because we want to avoid models where the distant past has a non-negligible effect on the present.

5.1 ARIMA(0,2,2) models

It is known (see Wecker and Ansley, 1983) that the cubic spline state space model described in Section 2 is equivalent to an ARIMA(0,2,2) model with some restrictions on parameters. However, no-one seems to have explicitly worked out the connection, or the implications it has for forecasting with the cubic spline model.

We define the ARIMA(0,2,2) model as

$$Y_t - 2Y_{t-1} + Y_{t-2} = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$$

where $\{\varepsilon_t\}$ is a Gaussian white noise process with variance σ_ε^2. For invertibility, we also require $|\theta_2| < 1$, $\theta_2 - \theta_1 < 1$ and $\theta_2 + \theta_1 < 1$ (Box, Jenkins & Reinsell, 1994). Then the ARIMA(0,2,2) forecast function is $\hat{Y}_{n+h} = \ell_n + b_n h$ where $\ell_n = Y_n - \theta_2\hat{e}_n$ and $b_n = Y_n - Y_{n-1} + \theta_1\hat{e}_n + \theta_2(\hat{e}_n + \hat{e}_{n-1})$. (Here, $\hat{e}_j$ denotes the jth residual.)

Now Brown and de Jong (2001) show that the cubic spline model can be written as an ARIMA(0,2,1) process observed with error:

$$Y_t = X_t + \eta_t, \qquad (1-B)^2X_t = \xi_t + \psi\xi_{t-1}$$

where $\psi = 2 - \sqrt{3}$, $(X_1, X_2 - X_1)$ is fully diffuse, and η_t and ξ_t are uncorrelated white noise series with means zero and variances σ^2 and τ^2 respectively. It is easy to show this is equivalent to an ARIMA(0,2,2) model with θ_2 obtained by solving the following quartic equation:

$$\theta_2^4 - c_1\theta_2^3 + c_2\theta_2^2 - c_1\theta_2 + 1 = 0,$$

$$\theta_1 = \frac{\theta_2}{1+\theta_2}(\psi/\lambda_* - 4)$$

and $\sigma_\varepsilon^2 = \sigma^2\lambda_*/\theta_2$, where

$$c_1 = 4 + (1+\psi^2)/\lambda_*, \qquad \text{and} \qquad c_2 = 6 - 2(1+4\psi+\psi^2)/\lambda_* + \psi^2/\lambda_*^2.$$

Numerical calculations show that the above quartic equation has at most one root which gives an invertible solution, and that an invertible solution is obtained if and only if $0 < \lambda_* < 1.640519$. Figure 2 shows the values of θ_1 and θ_2 as functions of λ_*. In the original time space (where observation times are $1, 2, \ldots, n$), the upper bound on λ is $1.640519n^3$. This upper bound on λ should be imposed whenever the spline model is used for *forecasting* purposes. If the model is simply used to describe the

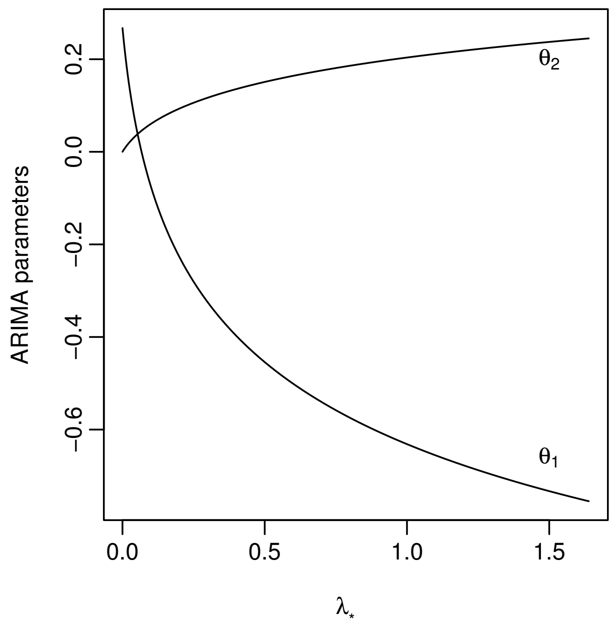

Figure 2: *The relationship between the ARIMA parameters θ_1 and θ_2 and the cubic spline parameter λ_*.*

historical trend, invertibility is not relevant and so the bound need not be imposed.

Note that the range of ARIMA(0,2,2) models that can be fitted in this way is greatly restricted, and that a wider range of models with linear forecast functions can be obtained by fitting a general ARIMA(0,2,2) model. In fact, Box, Jenkins & Reinsel (1994) show that all ARIMA(p,2,q) have forecast functions which are *asymptotically* linear (the "eventual forecast function"), and that the forecast function is *exactly* linear if and only if $p = 0$ and $q \leq 2$.

5.2 Holt's local linear forecasts

Holt's local trend method has been used in forecasting for many decades and it has proved remarkably versatile and useful. Point forecasts (see, e.g., Makridakis, Wheelwright and Hyndman, 1998, p.158) are given by $\hat{Y}_{n+h} = \ell_n + b_n h$ where ℓ_n and b_n are computed recursively as follows:

$$\ell_t = \alpha Y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{5.1}$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta) b_{t-1} \tag{5.2}$$

for $t = 2, \ldots, n$. Starting values for these recursions are often set to $\ell_1 = Y_1$ and $b_1 = Y_2 - Y_1$, although we choose the starting values optimally (see below).

The unobserved components ℓ_t and b_t represent the level and slope of the series at time t and α and β are constants. We normally restrict the parameters such that $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$.

Recently, Hyndman, Koehler, Snyder & Grose (2002) (hereafter HKSG) provided a general modelling framework for exponential smoothing methods, including Holt's method. This enables the forecasts to be obtained from a state space model, thus providing facilities for maximum likelihood estimation, calculation of prediction intervals, etc. HKSG actually provide two state space models for Holt's method, which give identical point forecasts but have different properties for high forecast moments. In this paper, we only consider the additive error version.

The model can be written as follows:

$$
\begin{aligned}
Y_t &= \ell_{t-1} + b_{t-1} + \varepsilon_t \\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t \\
b_t &= b_{t-1} + \alpha\beta\varepsilon_t
\end{aligned}
$$

where ℓ_t denotes the level at time t, b_t denotes the slope of the trend at time t, and ε_t is a Gaussian white noise process with zero mean and variance σ^2. We estimate the parameters α and β and the initial state vector $(\ell_0, b_0)'$ by maximizing the conditional likelihood as described in HKSG.

Hyndman, Koehler, Ord and Snyder (2001) show that the forecast mean of this model is identical to Holt's local trend forecast and the forecast variance of the model is

$$v_h = \sigma^2\Big[1 + \alpha^2(h-1)\Big\{1 + \beta h + \tfrac{1}{6}\beta^2 h(2h-1)\Big\}\Big].$$

Using this expressions, prediction intervals can be constructed in the usual way.

The above state space model underlying Holt's method is equivalent to an ARIMA(0,2,2) model where $\alpha = \theta_2 + 1$ and $\beta = (1 - \theta_1 - \theta_2)/(1 + \theta_2)$. In theory, the parameter space for (α, β) could be taken as the whole invertible region for the ARIMA model (in which case we would have $0 < \alpha < 2$ and $0 < \beta < 4/\alpha - 2$). However, it is usual to restrict the space further and require $0 < \alpha < 1$ and $0 < \beta < 1$ which leads to more interpretable models.

However, for the spline model, we found that $\theta_2 > 0$. Therefore, $\alpha > 1$ which means that the spline model falls outside the usual range of parameters considered for Holt's method. (We also found that $\beta > 1$ when $\lambda_* > 0.14514$.)

This means that Holt's method and the cubic spline model are both special but non-overlapping cases of the ARIMA(0,2,2) model.

6 Empirical comparison of models

Given that the cubic spline model is a special case of an ARIMA(0,2,2) model, it is interesting to see if the restricted parameter space results in poorer forecasting performance. We will also compare the forecasts from Holt's method based on a different and mutually exclusive subset of the parameter space of the ARIMA(0,2,2) model.

We compare the three models by applying them to the 645 annual series which were part of the M3 forecasting competition (Makridakis & Hibon, 2000). For each series, six observations were withheld at the end of the series for comparisons. The remaining observations were used for estimation of parameters.

For each series, we estimate the parameters using likelihood methods. We use the methods described in Sections 3 and 5.2 for the spline model and the state space model underlying Holt's method, and for the full ARIMA model we use the exact likelihood method of Gardner, Harvey and Phillips (1980) as implemented in the ts library distributed with R 1.5.1.

Then each model is used to forecast the remaining six observations in the series. The forecasts are compared by computing the Mean Absolute Percentage Error (MAPE) averaged across all series and the Coverage Percentage (CP) of the (nominally) 95% prediction intervals computed over all series.

As a further comparision, we also applied the local linear method of Nottingham and Cook (2001), although this assumed a deterministic trend rather than a stochastic trend.

The results are given in Tables 1 and 2 and highlight some interesting similarities and differences between methods.

	Forecast horizon					
Method	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Spline	9.8	23.0	26.8	32.0	37.6	41.9
ARIMA(0,2,2)	8.6	21.6	26.9	30.3	35.9	37.8
Holt	11.0	23.3	25.8	29.1	32.2	36.0
Local linear	11.8	23.1	26.3	28.4	32.6	36.9

Table 1: *Mean Absolute Percentage Error for each model, computed by averaging the absolute percentage error across all 645 annual series.*

	Forecast horizon					
Method	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
Spline	86.4	81.9	77.2	76.6	76.4	78.0
ARIMA(0,2,2)	84.3	80.8	79.1	78.3	77.7	78.9
Holt	80.2	71.3	65.3	60.2	58.9	58.0
Local linear	79.1	64.2	55.7	50.7	48.4	44.5

Table 2: *Coverage percent of the nominal 95% prediction intervals computed from each model. These are the percentage of actual observations within the prediction intervals across all 645 annual series.*

- All four methods have very similar performance for point forecasting. In particular, the restricted parameter spaces for the spline method and Holt's method do not result in much deterioration in forecast performance.
- The spline method and the ARIMA(0,2,2) method are very similar in coverage probabilities for prediction intervals. That these are much narrower than the nominal 95% probability is not surprising—similar results are standard in forecasting real data (see HKSG, for example).
- Holt's method does considerably worse than either spline or ARIMA models in terms of coverage probability. This is somewhat surprising. Comparable results in HKSG where a larger range of exponential smoothing models were used for these same data show average coverage probabilities around 82%. It seems that Holt's method is not so good as a general all-purpose forecast method for non-seasonal data.
- The local linear method has smaller coverage probabilities than any of the other methods. This is not surprising, as the method does not allow for a stochastically changing trend. Hence, the trend is assumed to be known into the future, and so the estimated future variation is smaller.

6.1 Conclusions

We have shown how cubic smoothing splines can be used to obtain local linear forecasts for a univariate time series. New results include a bound on the smoothing parameter to achieve invertibility, explicit and closed-form expressions for the point forecasts and prediction intervals, a new method for obtaining the smoothing parameter, and an empirical comparison with other local linear forecast methods.

Spline forecasts provide an alternative approach to ARIMA(0,2,2) models for local linear forecasting. The main advantage of the spline approach over the ARIMA approach is that it is directly associated with a smooth estimate of historical trend. This can aid interpretation of the historical data as well as provide information about the trend used in forecasting. For example, the smooth trend through the beer production data in Figure 1 clearly shows the trend away from beer in Australia since about 1975 (partly explained by an increase in wine consumption). It also shows a brief resurgence in beer production in the late 1980s (when Australian beer exports led to increased production), before the production settled down to the current level.

A common criticism of nonparametric methods in general, and cubic splines in particular, is that they can be considered as special cases of more general time series models (e.g., Brown and de Jong, 2001; and Harvey and Koopman, 2000). The (usually unstated) implication is that the more general model is better. We have shown that this restriction does not lead to much reduction in forecast performance, and so for forecasting purposes, the criticism is not valid.

Acknowledgements

Rob Hyndman would like to thank the Australian National University for their hospitality while much of this work was being done.

Appendix: Proof of Proposition

This result follows directly from the state space formulation except for the form of $\mathrm{Var}(\boldsymbol{g})$ which we write as $\lambda_*^{-1}\Sigma$.

Let $\Gamma_i(j) = \mathrm{E}(\boldsymbol{\alpha}_i\boldsymbol{\alpha}'_{i-j})$, $j = 0, 1, \ldots$. Note $\boldsymbol{\alpha}_i$ follows a vector autoregressive model of order one in (2.2). Thus we obtain the Yule-Walker equations (Reinsel, 1997)

$$\begin{aligned}\Gamma_0(0) &= 0 \\ \Gamma_i(0) &= V_i + T_i\Gamma_{i-1}(0)T_i', \qquad i = 1, \ldots, n \\ \Gamma_i(j) &= T_i\Gamma_{i-1}(j-1), \qquad j = 1, 2, \ldots.\end{aligned} \tag{A.1}$$

Note that $\Gamma_i(j) = 0$ if $j \geq i$. We can use these equations to iteratively calculate the values of $\Gamma_i(j)$ for $i = 1, \ldots, n$ and $j = 1, 2, \ldots$. Then the (i, j)th element of $\lambda_*^{-1}\Sigma$ is the top left element of $\Gamma_j(j-i)$ if $i \leq j$ and the top left element of $\Gamma_i(i-j)$ if $i \geq j$.

Now De Jong and Mazzi (2001) show that for any t_i where $0 < t_i < t_{i+1} < 1$ for $i = 1, 2, \ldots, n-1$, the covariance matrix of $\boldsymbol{u}_i$, which we denote by V_i, has (j, k)th entry

$$[V_i]_{jk} = \tau^2\int_{t_{i-1}}^{t_i}\frac{(t_i-u)^{2-j}(t_i-u)^{2-k}}{(2-j)!(2-k)!}\,du = \tau^2\frac{h_i^{5-j-k}}{(4-j-k+1)(2-j)!(2-k)!}. \tag{A.2}$$

where $h_i = t_{i+1} - t_i$. Thus, in this special case where $h_i = h = n^{-1}$, we have

$$V_i = \tau^2\begin{bmatrix} h^3/3 & h^2/2 \\ h^2/2 & h \end{bmatrix}. \tag{A.3}$$

By substituting (A.3) into (A.1), we can construct Σ.

First we show by induction that

$$\Gamma_i(0) = \tau^2\begin{bmatrix} i^3h^3/3 & i^2h^2/2 \\ i^2h^2/2 & ih \end{bmatrix}. \tag{A.4}$$

For $i = 0$, $\Gamma_0(0) = 0$, so (A.4) is true. Now assume (A.4) is true for $i = k$. Then from (A.1) we obtain

$$\begin{aligned}\Gamma_{k+1}(0) &= \tau^2 \begin{bmatrix} h^3/3 & h^2/2 \\ h^2/2 & h \end{bmatrix} + \tau^2 \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k^3h^3/3 & k^2h^2/2 \\ k^2h^2/2 & kh \end{bmatrix} \begin{bmatrix} 1 & 0 \\ h & 1 \end{bmatrix} \\ &= \tau^2 \begin{bmatrix} (k+1)^3h^3/3 & (k+1)^2h^2/2 \\ (k+1)^2h^2/2 & (k+1)h \end{bmatrix}.\end{aligned}$$

So (A.4) is true for $i = k+1$ and by induction is true for $i = 1, 2, 3, \ldots$. Now from (A.1) we have

$$\Gamma_i(j) = T\Gamma_{i-1}(j-1) = T^2\Gamma_{i-2}(j-2) = T^j\Gamma_{i-j}(0), \qquad \text{for } i \geq j,$$

and so $\Gamma_i(i-j) = T^{i-j}\Gamma_j(0)$. Thus

$$\Gamma_i(i-j) = \tau^2 \begin{bmatrix} 1 & (i-j)h \\ 0 & 1 \end{bmatrix} \begin{bmatrix} j^3h^3/3 & j^2h^2/2 \\ j^2h^2/2 & jh \end{bmatrix} = \begin{bmatrix} h^3j^2(3i-j)/6 & jh^2(2i-j)/2 \\ j^2h^2/2 & jh \end{bmatrix},$$

Thus Σ is symmetric with the (j,k)th element on or above the diagonal given by

$$\Sigma_{jk} = \Sigma_{kj} = \sigma^2h^3j^2(3k-j)/6, \qquad k \geq j$$

and so

$$\Sigma_{jk} = \sigma^2n^{-3}k^2(3j-k)/6$$

for $j \geq k$.

References

ASSIMAKOPOULOS, V. and K. NIKOLOPOULOS (2000) The theta model: a decomposition approach to forecasting. *International Journal of Forecasting* **16**, 521–530.

BERAN, J., and Y. FENG (2002) SEMIFAR models—a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics and Data Analysis*, **40**, 393–419.

BERAN, J., and D. OCKER (1999) SEMIFAR forecasts, with applications to foreign exchange rates. *J. Statistical Planning and Inference*, **80**, 137–153.

BOX, G.E.P., G.M. JENKINS, and G.C. REINSEL (1994) *Time Series Analysis: Forecasting and Control*, third edition, Prentice-Hall.

BROWN, P.E., and P. DE JONG (2001) Nonparametric smoothing using state space techniques. *Canadian J. Statistics*, **29**(1), 37–50.

DE JONG, P., and S. MAZZI (2001) Modeling and smoothing unequally spaced sequence data, *Stat. Inference Stoch. Process*, **4**(1), 53–71.

GARDNER, G, A.C. HARVEY, and G.D.A. PHILLIPS (1980) Algorithm AS154. An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Applied Statistics*, **29**, 311-322.

GREEN, P.J., and B.W. SILVERMAN (1994) *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman and Hall.

HARVEY, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

HARVEY, A.C. and S.J. KOOPMAN (2000) Signal extraction and the formulation of unobserved components models, *Econometrics Journal*, **3**, 84–107.

HOLT, C.C. (1957) “Forecasting seasonal and trends by exponentially weighted moving averages”, Office of Naval Research, Research Memorandum No. 52.

HYNDMAN, R.J., and B. BILLAH (2002) Unmasking the Theta method. *International J. Forecasting*. To appear.

HYNDMAN, R.J., A.B. KOEHLER, J.K. ORD, and R.D. SNYDER (2001) “Prediction intervals for exponential smoothing state space models." Working paper 11/2001, Department of Econometrics & Business Statistics, Monash University.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., Grose, S. (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International J. Forecasting*, **18**(3), 439–454.

Makridakis, S.G., and M. Hibon (2000) The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–476.

Nottingham, Q.J., and D.F. Cook (2001) Local linear regression for estimating time series data, *Computat. Stat. and Data Analysis*, **37**, 209–217.

Rao, C.R. (1973) *Linear statistical inference and its applications*, 2nd edition, John Wiley & Sons, New York.

Reinsel, G.C. (1997) *Elements of Multivariate Time Series Analysis*, second edition, Springer-Verlag: New York.

Wahba, G. (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression, *J. Royal Statist. Soc. B*, **40**, 364–372.

Wecker, W.E. and C.F. Ansley (1983) The signal extraction approach to nonlinear regression and spline smoothing, *J. Amer. Statist. Assoc.*, **78**, 81–89.