MONASH

WP 9/97

ISSN 1032-3813
ISBN 0 7326 1032 X

MONASH UNIVERSITY

ANCORA IMPARO

AUSTRALIA

TREND STABILITY AND STRUCTURAL CHANGE:

AN EXTENSION TO THE M1 FORECASTING COMPETITION

Brett Inder and Ralph Snyder

Working Paper 9/97
October 1997

**DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS**

Trend stability and structural change:

an extension to the M1 forecasting competition.

Brett Inder

Department of Econometrics, Monash University, Clayton, Victoria, Australia, 3185

Telephone: 613 9905 2303 Fax: 613 9905 5474

Email: inder@monash.edu.au

Ralph Snyder

Department of Econometrics, Monash University, Clayton, Victoria, Australia, 3185

Telephone: 613 9905 2366 Fax: 613 9905 5474

Email: ralph.snyder@buseco.monash.edu.au

Abstract

The global linear trend with autocorrelated disturbances is a surprising omission from the M1 competition. This approach to forecasting is therefore evaluated using the 51 non-seasonal series from the competition. It is contrasted with a fully optimized version of Holts trend corrected exponential smoothing. It is found that an adaptation of Holts method, in which the growth rate is restricted to be constant, performs almost as well as its traditional counterpart and usually out-performs the global linear trend with autoregressive disturbances. This therefore confirms the results from other studies which indicate that a long-term trend may be missing from many business and economic time series. An implication of this study is that business forecasters, when applying trend corrected exponential smoothing, should explore the possibility of eliminating the second smoothing constant.

Keywords: forecasting; trends; M1 competition; unit roots; exponential smoothing; structural change

1. INTRODUCTION

The linear trend with auto-correlated disturbances has a long history in the analysis of macro-economic time series (see Nelson and Plosser, 1982). It was seen as a way of superimposing a business cycle effect on a long term trend. It was recognized that a business cycle is not systematic enough to be modeled deterministically, and that it is overly ambitious to expect to be able to predict its effects in the medium to long term. The ability to allow for persistent deviations from trend in the short term, while returning to global trend in the longer term, seemed to be a particularly attractive feature of the model. It is surprising, therefore, that this approach was completely omitted from the M1 competition (Makridakis et al, 1982).

In contrast, Holts method (Holt, 1957) of forecasting, which is also based on the concept of linear trend line, was included in the M1 competition. In this method the trend line may be viewed as the stochastic counterpart of a first-order Taylors approximation. To illustrate, the line LT in Figure 1 approximates the path (dashed line) followed by the data at the point of time A. In this context the trend is often referred to as a local linear trend. Because it approximates the data path locally, it might be anticipated that a local trend yields reasonable short-term forecasts. Long-run forecasts appear to be more problematic, however, particularly when the data contains a strong business cycle.

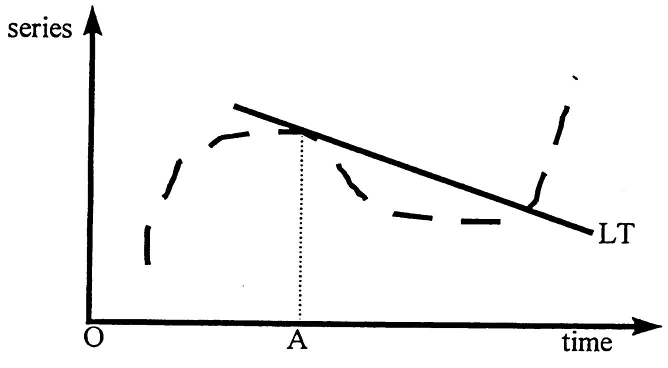

Figure 1 Local trend approximation to business cycle

Given the 'return to the mean' property of the global linear tend with autoregressive (AR) disturbances, it might be anticipated that its longer term forecasting capacity is better than that of Holts method. Lying beneath this supposition, however, is the implicit assumption that the data contains a long term trend, a point which can no longer be taken for granted following the work of Nelson and Plosser (1982) on macro-economic time series.

The primary purpose of this paper is to compare both approaches to forecasting with linear trends and to see whether our preconceptions hold up in practice. We use the M1 competition data set as our test bed. In the process we eliminate a gap in the M1 competition.

The following notational conventions are adopted in this paper:

a) We define a sequence of random variables $\{ y_t \}$ for t=1,...,T. The observed / sample value of y_t at time period t will be designated by $\bar{y}_t$.
b) A trial value for an unknown parameter θ is designated by $\bar{\theta}$. However the estimated value of the parameter, using a given estimation procedure, is represented by $\hat{\theta}$.
c) The unobservable random variables appearing in some of the statistical models will at times be assigned fixed values. For example, the assigned value of the random variable called a local level ℓ_t will be designated $\bar{\ell}_t$.

2. METHODOLOGY

The M1 data set contains 111 time series. Given the focus of this study on trends, all data deemed by Makridakis to contain a seasonal cycle was culled. 51 time series consisting of 20 annual, 14 quarterly and 17 monthly time series were left for analysis.

The model selection strategy considered in the M1 competition, based on the median absolute percentage prediction errors criterion, is adopted. Each series is divided into two parts: a 'fitting sample' $\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n$ and a 'prediction sample' $\bar{y}_{n+1}, \bar{y}_{n+2}, \ldots, \bar{y}_{n+h}$. Following the M1 convention we use $h = 6$ for annual series, $h = 8$ for quarterly series, and $h = 18$ for monthly series. Denoting the predictions made from the end of period n as $\hat{y}_{n+1}, \hat{y}_{n+2}, \ldots, \hat{y}_{n+h}$, the absolute percentage errors are calculated using the formulae $\mathrm{APE}_t = 100|\bar{y}_t - \hat{y}_t|/\bar{y}_t$ for $t = n+1, \ldots, n+h$. Let $\mathrm{APE}_{i,t}$ denote the absolute percentage error in period t for typical time series i. The median APE for period t is then defined as $\mathrm{MeAPE}_t = \mathrm{median}_{i \in I}\{\mathrm{APE}_{it}\}$ where I is the set of series indexes over which the median is found.

3. AUTOCORRELATED DISTURBANCES

We first thought it would be sensible to test whether there was any advantage in adding AR disturbances to a conventional linear trend line. A global linear trend (GT) had already been considered in the M1 competition, the results being reproduced in the second row of Table 1. We repeated the study for the global trend line and corroborated these results.

We then fitted the global linear trend with autocorrelated disturbances (GTAR). The model has the form $y_t = \beta_1 + \beta_2 t + u_t$ with disturbances $u_t = \phi u_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim NID(0, \sigma^2)$. The Beach and MacKinnon (1978) method was used to estimate the intercept β_1, the growth β_2, the autoregressive parameter ϕ and the standard deviation σ. Forecasts were generated with the formula $\hat{y}_{n+j} = \hat{\beta}_1 + \hat{\beta}_2(n+j) + \hat{\phi}^j \hat{u}_n$. The MeAPE's obtained for this method are summarized in row 3 of Table 1. Taken as a whole, the results indicate that the addition of an AR(1) component to the trend line can significantly improve forecasts in the shorter term.

As expected, the damping effect of the autocorrelation coefficient term in the forecasting equation eliminates this advantage at longer lead times.

	Lead Time									
Model	1	2	3	4	5	6	8	12	15	18
GT	8.5	9.1	10.7	13.5	17.9	17.5	19	14.2	13.8	19.8
GTAR	3.3	4.7	7.4	9.1	13	11	12.7	7.7	12.7	21.2

Table 1 MeAPE's: 51 series

4. LOCAL LINEAR TREND

The local linear trend (LT), a special case of the linear state space framework in Snyder (1985), is

$$y_t = \ell_{t-1} + \beta_{t-1} + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + \beta_{t-1} + \alpha_1 \varepsilon_t$$

$$\beta_t = \beta_{t-1} + \alpha_2 \varepsilon_t$$

where $\varepsilon_t \sim NID(0,\sigma^2)$. The local level ℓ_t and the local growth rate β_t are treated as stochastic quantities so that the model can embody the effects of structural change. The parameters α_1, α_2, commonly called the smoothing constants, determine the amount by which the level and growth rates are impacted by structural change.

The reduced form of this model can be obtained by eliminating the local level and local growth rate. It is $\nabla^2 y_t = \theta_2 \varepsilon_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$ where $\theta_1 = \alpha_1 + \alpha_2 - 2$ and $\theta_2 = 1 - \alpha_1$. This ARIMA(0,2,2) process underpins Holts trend corrected exponential smoothing (Box and Jenkins, 1976). Thus the local trend is a state space representation of the model for Holts method.

The unknown seed values of the state variables ℓ_0, β_0 and the smoothing parameters α_1, α_2 are chosen to minimize the sum of squared one-step ahead prediction errors. An adaptation of the algorithm in Ord, Koehler and Snyder (1997) is used. For each set of trial of values, denoted by $\overline{\ell}_0, \overline{\beta}_0, \overline{\alpha}_1, \overline{\alpha}_2$, the one-step ahead prediction errors are obtained recursively by applying Holts trend corrected exponential smoothing equations:

$$\overline{\varepsilon}_t = \overline{y}_t - \overline{\ell}_{t-1} - \overline{\beta}_{t-1}$$

$$\overline{\ell}_t = \overline{\ell}_{t-1} + \overline{\beta}_{t-1} + \overline{\alpha}_1 \overline{\varepsilon}_t$$

$$\overline{\beta}_t = \overline{\beta}_{t-1} + \overline{\alpha}_2 \varepsilon_t$$

together with the associated sum of squared errors $S = \sum_{t=1}^{n} \overline{\varepsilon}_t^2$. To emphasize the nature of the relationships involved here, the sum of squared errors function may be written as $S = S(\ell_0, \beta_0, \alpha_1, \alpha_2)$.

The errors are linear functions of the seed state vectors ℓ_0, β_0. Thus, for trial values $\overline{\alpha}_1$, $\overline{\alpha}_2$ of the smoothing parameters, conventional linear least squares methods can be applied to obtain the corresponding conditional estimates of the state vectors $\hat{\ell}_0(\overline{\alpha}_1, \overline{\alpha}_2)$, $\hat{\beta}_0(\overline{\alpha}_1, \overline{\alpha}_2)$. The state variable estimates may be substituted out of the sum of squared errors function to give the concentrated sum of squared errors function $S = S(\alpha_1, \alpha_2)$. The problem reduces to minimizing this concentrated sum of squared errors function with respect to the smoothing constants.

During the optimization, the values of the smoothing parameters are restricted to the unit square $0 \leq \alpha_1 \leq 1$ and $0 \leq \alpha_2 \leq 1$. By permitting both smoothing constants to equal 0, the possibility of 'no structural change' is taken into account. This special case, which corresponds to a global linear trend, is often ruled out in conventional implementations of trend corrected exponential smoothing. Common strategies used to estimate the seed level and growth, such as back-casting (Gardner, 1985), break down in this circumstance. In contrast, the sum of squared errors minimization strategy always works.

A larger region $\alpha_1 \geq 0$, $\alpha_2 \geq 0$ and $2\alpha_1 + \alpha_2 \leq 4$ could have been used instead of the unit square restrictions. These inequalities defining this region correspond to the invertibility conditions for the ARIMA(0,2,2) process (Box & Jenkins, 1976). We believe that the invertibility conditions are not widely implemented in conjunction with Holts method in practice. Furthermore it is difficult to make sense of values of the smoothing parameters outside the unit square. We therefore persisted with the unit square.

The first stage of the method of optimization involved a grid search: the SSE function $S = S(\alpha_1, \alpha_2)$ is not necessarily convex. The grid was defined over the closed unit interval [0,1] using increments of 0.1. The second stage involved the use of nonlinear optimization procedure (the Gauss constrained optimizer CO) seeded with the result from the grid search.

The resulting MeAPE's are shown in Table 2 together with those reported for Holts method in the M1 competition. We had originally expected only slight differences in both sets of results. The contrast, however, is quite marked at longer time horizons. The paper on the M1 competition is silent on implementation details such as

a) the method used for estimating the seed values

b) the constraints imposed on the smoothing parameters.

We therefore have been unable to explain this outcome. There seems to be no obvious pattern to the differences, and hence no clue as to where the discrepancy in implementation might be.

Method	Lead Time 1	2	3	4	5	6	8	12	15	18
M1	3.4	4.5	6.0	9.3	12.7	9.4	14.9	14.2	14.3	17.0
LT	3.6	4.1	6.1	7.9	11.7	8.0	15.7	12.8	20.4	24.4

Table 2 Holts method: all 51 series

5. LOCAL TREND AND GLOBAL TREND WITH AR DISTURBANCES

The global trend with AR disturbances and the local trend are two distinct adaptations of the concept of linear trend to take account of the phenomena of structural change. Our expectation was that both models would have a similar short term forecasting performance but that, in the longer term, the tendency of forecasts to return to the trend would work in favor of the global trend with AR disturbances. The results for all 51 series are shown in Table 3. The same results broken down into annual, quarterly and monthly series respectively, are shown in Table 4.

Method	Lead Time 1	2	3	4	5	6	8	12	15	18
GTAR	3.3	4.7	7.4	9.1	13	11	12.7	7.7	12.7	21.2
LT	3.6	4.1	6.1	7.9	11.7	8.0	15.7	12.8	20.4	24.4

Table 3 MeAPE's: all 51 series

Series	Model	Lead Time 1	2	3	4	5	6	8	12	15	18
Annual	GTAR	2.7	4	8.3	11.8	17.6	15.5				
	LT	2.4	1.8	3.4	6.9	11.6	7.5				
Quarterly	GTAR	3.6	4.5	5.8	8.3	12.8	8.3	10.5			
	LT	5.0	4.3	9.4	9.3	11.9	13.2	22.7			
Monthly	GTAR	1.7	7.3	4.1	5.7	8.8	9.6	12.8	7.7	12.7	21.2
	LT	2.1	6.7	6.6	8.2	9.9	6.9	8.4	12.8	20.4	24.4

Table 4 MeAPE's: analysis by data type

The results are mixed. There is some support for our hypothesis on quarterly and monthly data. Holts exponential trend corrected exponential smoothing, however, is superior on annual data.

6. LOCAL TREND WITH CONSTANT GROWTH

While fitting the local trend model it was noted that the estimate of α_2 was often equal to 0. It was therefore decided to consider, as a model in its own right, a local trend with a growth rate constrained to be constant (LTCG). The conventional local level model reduces to the form

$$y_t = \ell_{t-1} + \beta + \varepsilon_t$$

$$\ell_t = \ell_{t-1} + \beta + \alpha_1 \varepsilon_t$$

where β is the constant growth rate. The reduced form of this model is

$\nabla y_t = \beta + \theta \varepsilon_{t-1} + \varepsilon_t$ where $\theta = -(1-\alpha_1)$. The local trend with constant growth is a convenient state space representation of an ARIMA(0,1,1) model with drift.

Model	Lead Time 1	2	3	4	5	6	8	12	15	18
LTCG	27.5	23.5	25.5	21.6	23.5	23.5	25.8	23.5	23.5	23.5
LT	27.5	31.4	31.4	33.3	33.3	31.4	29.0	23.5	23.5	23.5
Ties	45.1	45.1	43.1	45.1	43.1	45.1	45.2	52.9	52.9	52.9

Table 5 Percentage 'wins': all 51 series

Model	Lead Time 1	2	3	4	5	6	8	12	15	18
LTCG	2.9	4.8	6.8	8.2	13.0	8.6	11.6	13.7	21.1	25.1
LT	3.6	4.1	6.1	7.9	11.7	8.0	15.7	12.8	20.4	24.4

Table 6 MeAPE's: all 51 series

Both forms of local trend are compared in Table 5. The second and third rows of this table show the proportion of 'wins' for the methods k=1,2 with a lead time j from the end of period n. This proportion is defined as:

$$p_{n+j}^{(k)} = \sum_{i \in I_{n+j}} \delta\left(\left| \hat{\varepsilon}_{i,n+j}^{(k)} \right| < \left| \hat{\varepsilon}_{i,n+j}^{(3-k)} \right| \right) \Big/ m_{n+j}$$

where $\delta(\bullet)$ denotes a Kronecker delta, $\hat{\varepsilon}_{i,n+j}^{(k)}$ denotes the j-step ahead error for the prediction made for series i at the end of period n with method k, I_{n+j} is the index set of series predicted in period $n+j$, and m_{n+j} is the number of series predicted in period $n+j$. The proportion of tied results, which largely reflect those results where the conventional local trend reduces to its constant growth counterpart, are shown in the fourth row. The results, shown in Table 6, indicate that there is little to be gained from a variable growth rate in a local trend model. Given that grid search optimization with respect to one parameter, rather than two parameters, involves much lower computational loads, this result indicates that business forecasters, particularly in mass forecasting situations such as inventory control, should give serious thought to testing and implementing this scheme where appropriate.

7. TREND STATIONARY AND DIFFERENCE STATIONARY PROCESSES

A comparison of the global trend with AR(1) disturbances and the local linear trend with constant growth rate is shown in Table 7. The local trend with constant growth rate has a tendency to do better in the shorter term. Its advantage, however, disappears at longer time horizons. The breakdown into annual, quarterly and monthly data in Table 8 reveals that the local trend with constant growth (LTCG) wins far more often at virtually all time horizons with annual and quarterly data. The comparison is much closer with monthly, but still the advantage is with the local trend with constant growth model for medium term forecast horizons, and the global trend with AR disturbances (GTAR) becomes more competitive with long term forecasts.

	Lead Time									
Model	1	2	3	4	5	6	8	12	15	18
GTAR	45.1	39.2	27.5	25.5	27.5	29.4	38.7	52.9	52.9	52.9
LTCG	49.0	51.0	66.7	70.6	64.7	64.7	61.3	47.1	47.1	47.1
Ties	5.9	9.8	5.9	3.9	7.8	5.9	0.0	0.0	0.0	0.0

Table 7 Percentage 'wins': all 51 series

Series	Model	Lead Time 1	2	3	4	5	6	8	12	15	18
Annual	GTAR	50.0	35.0	25.0	25.0	25.0	20.0				
	LTCG	45.0	45.0	70.0	70.0	55.0	65.0				
	Ties	5.0	20.0	5.0	5.0	20.0	15.0				
Quarterly	GTAR	42.9	21.4	7.1	14.3	14.3	28.6	28.6			
	LTCG	57.1	71.4	92.9	85.7	85.7	71.4	71.4			
	Ties	0.0	7.1	0.0	0.0	0.0	0.0	0.0			
Monthly	GTAR	41.2	58.8	47.1	35.3	41.2	41.2	47.1	52.9	52.9	52.9
	LTCG	47.1	41.2	41.2	58.8	58.8	58.8	52.9	47.1	47.1	47.1
	Ties	11.8	0.0	11.8	5.9	0.0	0.0	0.0	0.0	0.0	0.0

Table 8 Percentage 'wins': analysis by data type

Method	Lead Time 1	2	3	4	5	6	8	12	15	18
GTAR	3.3	4.7	7.4	9.1	13.0	11.0	12.7	7.7	12.7	21.2
LTCG	2.9	4.8	6.8	8.2	13.0	8.6	11.6	13.7	21.1	25.1

Table 9 MeAPE's: all 51 series

An interesting contrast is revealed, though, in Table 9. Despite the higher percentage wins of the LTCG model at most lead times, the MeAPE's are quite similar for forecast horizons up to 8 periods, and the GTAR model tends to have smaller MeAPE's thereafter. What seems to be happening here is that the global trend approach is rarely much inferior to the local trend model, but sometimes the LTCG model can give forecasts which are quite a long way from actual values. Perhaps this reflects the situation depicted in Figure 2 where the LTCG at time A fails to adapt properly in the longer term to business cycle effects in the data path represented by the dashed line.

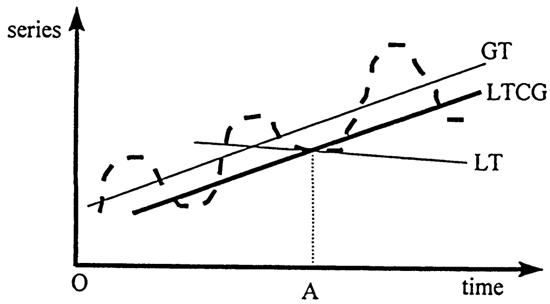

Figure 2 Local trend with constant growth

Having said that, the local trend constant growth model is still clearly superior to the global trend with autoregressive error model in most cases, especially for the quarterly and annual data.

8. CONCLUSIONS

It has been shown in this study that there can be significant gains from including an AR(1) disturbance with a global linear trend. Nevertheless, it was established that the local trend, another variation of the classical trend concept to allow for structural change, also performed remarkably well. In fact the special case of the local trend, with its growth rate restricted to be constant, can often out-perform a global trend with AR(1) disturbances.

The results imply that structural change is often of a form that makes questionable the concept of a long term secular trend in many business and economic time series. Structural change can have a significant impact on underlying levels, but not necessarily on growth rates. The case for the use of local trends with constant growth rates appears to be quite strong.

It remains true, however, that each series potentially has its own idiosyncrasies. As such it is still a good practice to trial many methods before choosing one with which to generate forecasts. The message from this study is that business forecasters should include Holts trend corrected exponential smoothing with the restriction $\alpha_2 = 0$ in their kit bag of techniques. Not only may this prove to give better forecasts than its traditional unrestricted counterpart, but it also simplifies and reduces computational loads for finding minimum sum of squared error estimates.

REFERENCES

Beach, C.M. and J.G. MacKinnon, 1978, A maximum likelihood prcoedure for regression with autocorrelated disturbances, *Econometrica* 46, 51-58.

Box, G.E.P. and G.M. Jenkins, 1976, *Time series analysis: forecasting and control* (Holden Day, San Francisco).

Gardner, E.S., 1985, Exponential smoothing: the state of the art, *Journal of Forecasting* 4, 1-28.

Holt, C.E., 1957, Forecasting trends and seasonals by exponentially weighted averages, *O.N.R. memorandum* 52, Carnegie Institute of Technology.

Makridakis, S. et. al., 1982, The accuracy of extrapolation (time series) methods: results of a forecasting competition, *Journal of Forecasting* 1, 111-153.

Nelson, C.R. and C.I. Plosser, 1982, Trends and random walks in macroeconomic time series, *Journal of Monetary Economics* 10, 139-162.

Ord, J.K., A.B. Koehler and R.D. Snyder, 1997, Estimation and prediction for a class of dynamic nonlinear statistical models, *Journal of the American Statistical Association*, forthcoming.

Snyder, R.D., 1985, Recursive estimation of dynamic linear statistical models, *Journal of the Royal Statistical Society* B47, 272-276.