

IMA(1,1) as a new benchmark for forecast evaluation

Philip Hans Franses

Econometric Institute

Erasmus School of Economics

EI2019-28

Abstract

Many forecasting studies compare the forecast accuracy of new methods or models against a benchmark model. Often, this benchmark is the random walk model. In this note I argue that for various reasons an IMA(1,1) model is a better benchmark in many cases.

Key words: One-step-ahead forecasts; Benchmark model

JEL codes: C53

Correspondence: PH Franses, Econometric Institute, Erasmus School of Economics, POB 1738, NL-3000 DR Rotterdam, the Netherlands, franses@ese.eur.nl. Thanks to Rob Hyndman for helpful suggestions.

Introduction

It is common practice to compare the forecast performance of a new model or method with that of a benchmark model. This holds in particular in these days where many new and advanced econometric models are put forward, like various versions of dynamic factor models and where many studies emerge using novel machine learning methods, see Kim and Swanson (2018) for a recent extensive survey and application.

Typically one chooses as the benchmark for one-step-ahead forecasts a simple autoregressive time series model, and most often one seems to choose for a random walk model. When y_t denotes a time series to be predicted, then the random walk forecast for $t + 1$ is

$$\hat{y}_{t+1|t} = y_t$$

which is based on the random walk model

$$y_t = y_{t-1} + \varepsilon_t$$

where ε_t is a mean-zero white noise process with variance σ_ε^2 . One motivation to consider this model is of course that there is no parameter to estimate, and hence there is no effort involved to create this forecast.

In many situations, however, the random walk model rarely fits the actual data. For financial time series one may perhaps encounter this model as it associates with a weak-form efficient market, but for many other time series like in macroeconomics or business, the random walk model does not provide a good fit. It is therefore that in this note I propose to replace the random walk benchmark model by another model, which has more face value for a wider range of economic variables. This new benchmark model is the Integrated Moving Average model of order (1,1) [with acronym: IMA(1,1)], which looks like

$$y_t = y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (1)$$

This IMA(1,1) basically is a random walk model with an additional lagged error term $\theta\varepsilon_{t-1}$. The θ parameter, which can be positive or negative and which is usually bounded by -1 and 1, in this IMA(1,1) model can be estimated using Maximum Likelihood or Iterative Least Squares. As an example Nelson and Plosser (1982) and Rossana and Seater (1995) find much empirical evidence of this model for a range of macroeconomic variables.

Writing

$$u_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

then the variance of u_t , γ_0^u , is

$$\gamma_0^u = (1 + \theta^2)\sigma_\varepsilon^2$$

using the methods outlined in Chapter 3 of Franses, van Dijk and Opschoor (2014), and the first-order autocovariance, γ_1^u , is

$$\gamma_1^u = \theta\sigma_\varepsilon^2$$

This makes that the first-order autocorrelation of u_t , ρ_1^u , is

$$\rho_1^u = \frac{\gamma_1^u}{\gamma_0^u} = \frac{\theta}{1 + \theta^2}$$

When $\theta > 0$, then $\rho_1^u > 0$, and when $\theta < 0$, then $\rho_1^u < 0$.

In this note I will show that the IMA(1,1) model follows naturally in a variety of settings. First, there will be some theoretical arguments. Next, I provide two additional, empirics-based, arguments. The last section concludes.

How can an IMA(1,1) model arise?

This section shows that an IMA(1,1) can follow from temporal aggregation of a random walk process, that it can follow from a simple basic structural model, that it associates with a time series process which experiences permanent and immediate shocks, and that it can be viewed as a simple and sensible forecasting updating process associated with exponential smoothing.

Aggregation of a random walk

Suppose that there is a variable y_τ where τ is of a higher frequency than t . For example, τ amounts to months, where t can concern years. Suppose further that the variable at the higher frequency τ obeys a random walk model, that is,

$$y_\tau = y_{\tau-1} + \varepsilon_\tau$$

where ε_τ is a mean-zero white noise process with some variance. Suppose that this high frequency random walk is temporally aggregated to a variable with frequency t , and suppose that this aggregation involves m steps. So, aggregation from months to years implies that $m = 12$. Working (1960) shows that such temporal aggregation results in the following model:

$$y_t = y_{t-1} + u_t$$

where the first order autocorrelation of u_t , say, ρ_1^u is the only non-zero valued autocorrelation, and this autocorrelation is

$$\rho_1^u = \frac{m^2 - 1}{2(2m^2 + 1)}$$

When $m \rightarrow \infty$, $\rho_1^u \rightarrow \frac{1}{4}$. When $m = 2$, $\rho_1^u = \frac{1}{6}$. In other words, aggregation of a high frequency random walk leads to an IMA(1,1) model with a positive valued θ .

Basic structural model

Consider the basic structural time series model (Harvey, 1989)

$$y_t = \mu_{t-1} + \varepsilon_t$$

with

$$\mu_t = \mu_{t-1} + \beta \varepsilon_t$$

Writing the latter expression as

$$\mu_t = \frac{\beta \varepsilon_t}{1 - L}$$

where L is the familiar lag operator, then we have

$$y_t = \frac{\beta \varepsilon_{t-1}}{1 - L} + \varepsilon_t$$

Multiplying both sides with $1 - L$ and ordering the variables gives the joint expression for y_t :

$$y_t = y_{t-1} + \varepsilon_t + (\beta - 1)\varepsilon_{t-1}$$

Here the IMA(1,1) model in (1) appears with $\theta = \beta - 1$. The MA(1) parameter θ is negative when $\beta < 1$, and it is positive when $\beta > 1$. Note that when the error source in the two equations of the basic structural model is not the same ε_t , that then still the IMA(1,1) model appears, see Harvey and Koopman (2000).

Permanent and temporary shocks

Another but related way to arrive at an IMA(1,1) model is given by the following. Suppose that a time series can be decomposed into a part with permanent shocks and a part with only transitory shocks, like

$$y_t = \frac{v_t}{1-L} + w_t$$

As such, the white-noise shocks v_t with variance σ_v^2 have a permanent effect, because of the $1-L$ operator, and the white noise shocks w_t with variance σ_w^2 have a temporary (immediate) effect. Multiplying both sides with $1-L$ results in

$$(1-L)y_t = v_t + (1-L)w_t$$

This is

$$y_t = y_{t-1} + u_t$$

with the variance of u_t equal to

$$\gamma_0^u = \sigma_v^2 + 2\sigma_w^2$$

The first-order autocovariance is equal to

$$\gamma_1^u = -\sigma_w^2$$

and hence

$$\rho_1^u = \frac{-\sigma_w^2}{\sigma_v^2 + 2\sigma_w^2}$$

which is non-zero and negative because of the positive-valued variance σ_w^2 .

Forecast updates

A final simple motivation to favor an IMA(1,1) model as a benchmark is because it can be written as a simple random walk forecast update but now where past forecast errors are accommodated, where still the prediction interval can simply be computed (Chatfield, 1993). Consider again

$$y_t = y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

The one-step-ahead forecast is based on

$$\hat{y}_{t+1|t} = y_t + \theta \varepsilon_t$$

The error term can be viewed as the forecast error from the previous forecast, that is

$$\varepsilon_t = y_t - \hat{y}_{t|t-1}$$

Hence,

$$\hat{y}_{t+1|t} = y_t + \theta(y_t - \hat{y}_{t|t-1})$$

There are now four possible cases in terms of forecast updates, and these depend on the sign of θ and on the sign of $y_t - \hat{y}_{t|t-1}$. Note that the latter expression associates with a so-called simple exponential smoothing model (Chatfield et al., 2001).

Further arguments

Two further arguments which would make the IMA(1,1) model a better benchmark are the following. First, as Hyndman and Billah (2003) show, the IMA(1,1) model has the same forecasting function as the so-called “Theta” method, proposed in Assimakopoulos and Nikolopoulos (2000). This Theta method is empirically relevant as it seems to come out as the winner in various forecast contests, most notably the so-called M3 and M4 forecast competitions, see Makridakis and Hibon (2000), and Makridakis et al (2018), respectively.

Finally, an IMA(1,1) process can have autocorrelations that associate with long memory. At the same time, long memory associates with aggregation across time series variables (Granger, 1980) and structural breaks (Granger and Hyung, 2004). Consider again,

$$y_t = y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

Using the lag operator, this can be written as

$$(1 - L)y_t = (1 + \theta L)\varepsilon_t$$

And hence

$$\frac{1 - L}{1 + \theta L}y_t = \varepsilon_t$$

This can be written as

$$(1 - L)(y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 y_{t-3} + \dots) = \varepsilon_t$$

or

$$y_t - (\theta + 1)y_{t-1} + (\theta^2 + \theta)y_{t-2} - (\theta^3 + \theta^2)y_{t-3} + \dots = \varepsilon_t$$

Put simpler, the approximate infinite autoregression reads as

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \cdots + \varepsilon_t$$

with

$$\begin{aligned}\alpha_1 &= \theta + 1 \\ \alpha_2 &= -(\theta^2 + \theta) \\ \alpha_3 &= \theta^3 + \theta^2 \\ \alpha_4 &= -(\theta^4 + \theta^3) \\ &\dots\end{aligned}$$

Now consider the fractionally integrated model

$$(1 - L)^d y_t = \varepsilon_t$$

with $0 < d < 1$, see Granger and Joyeux (1980). Franses, van Dijk and Opschoor (2014, page 91) show that this can be written again as an infinite autoregression

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \cdots + \varepsilon_t$$

where now

$$\begin{aligned}\alpha_1 &= d \\ \alpha_2 &= \frac{d(1-d)}{2!} \\ \alpha_3 &= \frac{d(1-d)(2-d)}{3!} \\ \alpha_4 &= \frac{d(1-d)(2-d)(3-d)}{4!} \\ &\dots\end{aligned}$$

For particular values of θ and d , the patterns of the autoregressive parameters of the IMA(1,1) and the fractionally integrated process can look very similar. Consider for example Figure 1 which gives the first 10 autoregressive parameters, that is α_1 to α_{10} for $\theta = -0.9$ and $d = 0.3$.

Conclusion

In this note I proposed to replace the random walk benchmark model in forecast evaluations by another model, which has more face value for many economic variables. This new benchmark model is the Integrated Moving Average model of order (1,1). I have put forward six arguments why this IMA(1,1) model is a suitable benchmark model in practice.

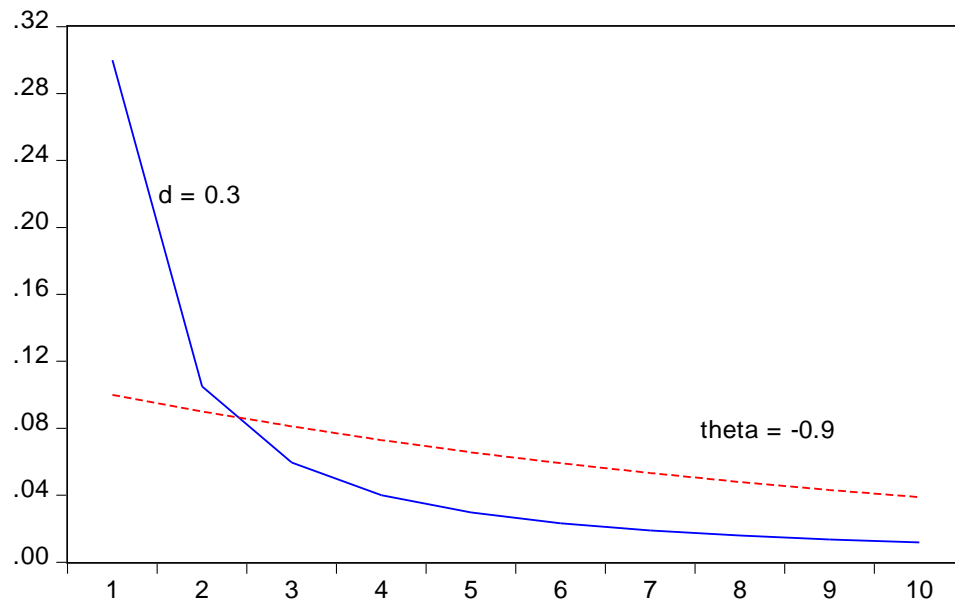


Figure 1: The first 10 autoregressive parameters in an approximate autoregressive model, that is α_1 to α_{10} for $\theta = -0.9$ and $d = 0.3$.

References

- Assimakopoulos, V. and K. Nikolopoulos (2000), The theta model: A decomposition approach to forecasting, *International Journal of Forecasting*, 16, 521-530.
- Chatfield, C. (1993), Calculating interval forecasts (with discussion), *Journal of Business and Economic Statistics*, 11, 121-144.
- Chatfield, C., A.B. Koehler, J.K. Ord, and R.D. Snyder (2001), A new look at models for exponential smoothing, *The Statistician*, 50, 146-159.
- Franses, P.H., D. van Dijk, and A. Opschoor (2014), *Time Series Models for Business and Economic Forecasting*, Cambridge UK: Cambridge University Press.
- Granger, C.W.J. (1980), Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics*, 14, 227-238.
- Granger, C.W.J. and N. Hyung (2004), Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance*, 11, 399-421.
- Granger, C.W.J. and R. Joyeux (1980), An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1, 15-39.
- Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge UK: 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.
- Hyndman, R. J. and B. Billah (2003), Unmasking the Theta method, *International Journal of Forecasting*, 19, 187-290.

Kim, H.H. and N.R. Swanson (2018), Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods, *International Journal of Forecasting*, 34, 339-354.

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

Makridakis, S., E. Spiliotis and V. Assimakopoulos (2018), The M4 competition: Results, findings, conclusion and way forward, *International Journal of Forecasting*, 34, 802-808.

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

Rossana, R. and J. Seater (1995), Temporal aggregation and economic time series, *Journal of Business and Economic Statistics*, 13, 441-451.

Working, H. (1960), Note on the correlation of first differences of averages in a random chain, *Econometrica*, 28, 916-918.